การเปรียบเทียบประสิทธิภาพในการทำนายผลการปรับความไม่สมดุลของข้อมูลในการจำแนกด้วยเทคนิคการทำเหมืองข้อมูล

PERFORMANCE COMPARISON IN PREDICTION OF IMBALANCED DATA IN DATA MINING CLASSIFICATION

พัชรียา ทองพูล

พิมพ์ชนก จำเรือง

รมย์นลิน บุญฤทธิ์

ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรบัณฑิต (สถิติประยุกต์)

ภาควิชาสถิติ คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2561

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

PERFORMANCE COMPARISON IN PREDICTION OF IMBALANCED DATA IN DATA MINING CLASSIFICATION

PACHAREEYA THONGPOOL

PIMCHANOK JAMRUENG

ROMNALIN BOONRIT

A SPECIAL PROBLEM SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR

THE DEGREE OF BACHELOR OF SCIENCE (APPLIED STATISTICS)

DEPARTMENT OF STATISTICS, FACULTY OF SCIENCE

KING MONGKUT'S INTITUTE OF TECHNOLOGY LADKRABANG

ACADEMIC YEAR 2018

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ	การเปรียบเทียบประสิทธิภาพในการทำนายผลการปรับความไม่สมดุลของข้อมูลในการจำแนกด้วยเทคนิคการทำเหมืองข้อมูล Performance Comparison in Prediction of Imbalanced Data in Data Mining Classification
ชื่อนักศึกษา	นางสาวพัชรียา ทองพูล รหัสนักศึกษา 58051277 นางสาวพิมพ์ชนก จำเรือง รหัสนักศึกษา 58051281 นางสาวรมย์นลิน บุญฤทธิ์ รหัสนักศึกษา 58051303
ปริญญา	วิทยาศาสตรบัณฑิต (สถิติประยุกต์)
ภาควิชา	สถิติ
ปีการศึกษา	2561
อาจารย์ที่ปรึกษา	รองศาสตราจารย์สายชล สินสมบูรณ์ทอง

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง อนุมัติให้ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต (สถิติประยุกต์) ประจำปีการศึกษา 2561

คณะกรรมการสอบ	ลายมือชื่อ
ผศ.วราพร เหลือสินทรัพย์ ประธานกรรมการ	
ดร.พรรณทิพา วาณิชย์จิรัฐติกาล กรรมการ	
รองศาสตราจารย์ สายชล สินสมบูรณ์ทอง กรรมการและอาจารย์ที่ปรึกษา	

ลิขสิทธิ์ของคณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ	การเปรียบเทียบประสิทธิภาพในการทำนายผลการปรับความไม่สมดุลของข้อมูลในการจำแนกด้วยเทคนิคการทำเหมืองข้อมูล		
ชื่อนักศึกษา	นางสาวพัชรียา	ทองพูล	รหัสนักศึกษา 58051277
	นางสาวพิมพ์ชนก	จำเรือง	รหัสนักศึกษา 58051281
	นางสาวรมย์นลิน	บุญฤทธิ์	รหัสนักศึกษา 58051303
ปริญญา	วิทยาศาสตรบัณฑิต (สถิติประยุกต์)		
ภาควิชา	สถิติ		
ปีการศึกษา	2561		
อาจารย์ที่ปรึกษา	รองศาสตราจารย์สายชล สินสมบูรณ์ทอง		

บทคัดย่อ

งานวิจัยนี้มีนี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการปรับข้อมูลที่ไม่สมดุล 4 วิธี คือ วิธีการสุ่มเกิน วิธีการสุ่มเกินโดยเทคนิค SMOTE วิธีการสุ่มลด และวิธีการสุ่มผสมผสาน โดยวิธีการจำแนก 4 วิธีคือ วิธีเพื่อนบ้านใกล้สุด k ตัว วิธีต้นไม้ตัดสินใจ วิธีโครงข่ายประสาทเทียม และวิธีซัพพอร์ตเวกเตอร์แมชชีน ว่าวิธีใดมีประสิทธิภาพในการจำแนกและการปรับความไม่สมดุลดีที่สุด โดยพิจารณาจากค่าความถูกต้อง ค่าความไว ค่าความจำเพาะ และค่าคลาดเคลื่อนกำสองเฉลี่ย โดยแบ่งข้อมูลเป็นชุดข้อมูลเรียนรู้ ชุดข้อมูลตรวจสอบความถูกต้อง และชุดข้อมูลทดสอบ ในอัตราส่วน 70, 20 และ 10 ตามลำดับ โดยการค้นคว้าและศึกษาในการหาข้อมูลที่ไม่สมดุล 3 ชุด คือชุดข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก มีจำนวนกลุ่มส่วนน้อยร้อยละ 4.49 ชุดข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้า มีจำนวนกลุ่มส่วนน้อยร้อยละ 10 และชุดข้อมูลคุณภาพไวน์แดง มีจำนวนกลุ่มส่วนน้อยร้อยละ 19.52 โดยใช้โปรแกรม WEKA และใช้ตารางเลขสุ่มในการปรับข้อมูลให้สมดุล

จากการเปรียบเทียบข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก วิธีที่มีประสิทธิภาพสูงสุดคือวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ชุดข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้า วิธีที่มีประสิทธิภาพสูงสุดคือวิธีเพื่อนบ้านใกล้สุด k ตัวโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ส่วนชุดข้อมูลคุณภาพไวน์แดง วิธีที่มีประสิทธิภาพสูงสุดคือวิธีซัพพอร์ตเวกเตอร์แมชชีนโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด

คำสำคัญ : วิธีการสุ่มเกิน วิธีการสุ่มเกินโดยเทคนิค SMOTE วิธีการสุ่มลด วิธีการสุ่มผสมผสาน วิธีเพื่อนบ้านใกล้สุด k ตัว วิธีต้นไม้ตัดสินใจ วิธีโครงข่ายประสาทเทียม วิธีซัพพอร์ตเวกเตอร์แมชชีน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Performance Comparison in Prediction of Imbalanced Data in Data Mining Classification	
Students	PACHAREEYA THONGPOOL	Student ID 58051277
	PIMCHANOK JAMRUENG	Student ID 58051281
	ROMNALIN BOONRIT	Student ID 58051303
Degree	Bachelor of Science (Applied Statistics)	
Department	Statistics	
Faculty	Science	
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)	
Academic Year	2018	
Advisor	Assoc.Prof.Saichon Sinsomboonthong	

ABSTRACT

We compared the performance of four classification methods: k- nearest neighbors, decision trees, artificial neural networks and support vector machines, on imbalanced data that consisted of over sampling, synthetic minority over sampling technique, under sampling and hybrid methods. Metrics were accuracy, sensitivity, specificity and mean squared error. The data sets were auditory perception in children with otitis media with effusion (OME), credit card balance and red wine quality. Each of these data sets was divided into three smaller sets: training, validating, and testing sets at a proportions in the ratio of 70:20:10 using The University of Waikato's WEKA program and Table of Random Numbers

When we compared the OME data set, the best classification method was the k-nearest neighbor in imbalanced data, adapting the synthetic minority over sampling technique. For the credit card data sets, the best classification methods was the same. For the wine data sets, the best method was the support vector machine in imbalanced data adapting under sampling.

Keywords : over sampling; synthetic minority over sampling technique; under sampling; hybrid methods k-nearest neighbor; decision tree; artificial neural network; support vector machine

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ปัญหาพิเศษฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี มีความละเอียดและความถูกต้องในเนื้อหา เนื่องด้วยการให้ความช่วยเหลือแนะนำของ รศ.สายชล สินสมบูรณ์ทอง ซึ่งเป็นอาจารย์ที่ปรึกษา ได้ให้ความรู้ ให้คำปรึกษา ให้ความช่วยเหลือ และตรวจแก้ไขในการทำปัญหาพิเศษจนสำเร็จลุล่วงไปได้ด้วยดี จึงขอกราบขอบพระคุณไว้ ณ โอกาสนี้

ขอขอบคุณ ผศ.วราพร เหลือสินทรัพย์ และ ดร.พรรณทิพา วาณิชย์จิรัฐติกาล ที่เป็นอาจารย์คณะกรรมการ ซึ่งได้กรุณาตรวจแก้ไขปัญหาพิเศษฉบับนี้ให้ถูกต้องสมบูรณ์ยิ่งขึ้น ให้คำแนะนำเกี่ยวกับปัญหาพิเศษฉบับนี้ทั้งหมด ทั้งแนวคิด วิธีการ และข้อมูลทุกอย่างที่เป็นประโยชน์ และขอขอบคุณอาจารย์ภาคสถิติทุกท่าน ที่ได้ประสิทธิ์ประสาทวิชาความรู้ พร้อมทั้งคำแนะนำและช่วยเหลือในเรื่องต่างๆตลอดมา

สุดท้ายนี้ผู้จัดทำขอขอบคุณ บิดา มารดา เพื่อน พี่ น้อง ทุกๆท่าน ที่เป็นกำลังใจ สนับสนุนในด้านกำลังทรัพย์ และให้ความช่วยเหลือในด้านต่าง ๆ ตลอดมา จนปัญหาพิเศษฉบับนี้สำเร็จลุล่วงด้วยดี

พัชรียา ทองพูล
พิมพ์ชนก จำเรือง
รมย์นลิน บุญฤทธิ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	ฉ
สารบัญรูป	ฌ
บทที่ 1 บทนำ	1
1.1 ที่มาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ที่ศึกษา	2
1.3 ประโยชน์ที่คาดว่าจะได้รับ	2
1.4 ขอบเขตการศึกษา	3
1.5 นิยามศัพท์เฉพาะ	3
1.6 ประชากรและตัวอย่าง	4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	5
2.1 ข้อมูลที่ไม่สมดุล	5
2.2 การปรับความไม่ดุลของข้อมูล	7
2.3 วิธีการจำแนก	9
2.7 การเปรียบเทียบประสิทธิภาพของวิธีการจำแนก	16
2.8 งานวิจัยที่เกี่ยวข้อง	17
บทที่ 3 วิธีการดำเนินงานวิจัย	20
3.1 การเก็บรวบรวมข้อมูล	20
3.2 ตัวแปรและนิยาม	20
3.3 เครื่องมือที่ใช้ในการวิจัย	22
3.4 ขั้นตอนในการดำเนินงานวิจัย	22
บทที่ 4 ผลการวิเคราะห์ข้อมูล	25
4.1 ผลการปรับความไม่สมดุล 4 วิธี	26
4.2 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัวโดยใช้ขั้นตอนวิธี IBk	27
4.3 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจโดยใช้ขั้นตอนวิธี J48	51

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

หน้า

4.4 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียมโดยใช้ขั้นตอนวิธี Multilayer Perceptron 103

4.5 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีนโดยใช้ขั้นตอนวิธี SMO 131

4.6 ผลการเปรียบเทียบประสิทธิภาพของวิธีการจำแนก 4 วิธี และการปรับความไม่สมดุล 4 วิธี 155

บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ 188

5.1 สรุปผลการวิจัย 188

5.2 อภิปรายผล 190

5.3 ข้อเสนอแนะ 190

บรรณานุกรม 192

ภาคผนวก ก 195

ภาคผนวก ข 208

ภาคผนวก ค 225

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่		หน้า
2.1	ชุดข้อมูล A ที่ยังไม่ได้ทำการปรับความสมดุล	6
2.2	ชุดข้อมูล A หลังจากทำการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน	6
2.3	ชุดข้อมูล A หลังจากทำการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินโดยเทคนิค SMOTE	7
2.4	ชุดข้อมูล A หลังจากทำการปรับความไม่สมดุลด้วยวิธีการสุ่มลด	8
2.5	ชุดข้อมูล A หลังจากทำการปรับความไม่สมดุลด้วยวิธีการผสมผสาน	9
2.6	เมทริกซ์ความสับสน	16
4.1	ผลการปรับความไม่สมดุลของข้อมูล 4 วิธี	26
4.2	ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกโดยวิธีการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน เมื่อ Random Seed 10, 20 และ 30	155
4.3	ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกโดยวิธีการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE เมื่อ Random Seed 10, 20 และ 30	156
4.4	ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกโดยวิธีการปรับความไม่สมดุลด้วยวิธีการสุ่มลด เมื่อ Random Seed 10, 20 และ 30	157
4.5	ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกโดยวิธีการปรับความไม่สมดุลด้วยวิธีการผสมผสานเมื่อ Random Seed 10, 20 และ 30	158
4.6	ผลการวิเคราะห์ข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้าโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน เมื่อ Random Seed 10, 20 และ 30	159
4.7	ผลการวิเคราะห์ข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้า โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE เมื่อ Random Seed 10, 20 และ 30	160
4.8	ผลการวิเคราะห์ข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้าโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด เมื่อ Random Seed 10, 20 และ 30	161
4.9	ผลการวิเคราะห์ข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้าโดยการปรับความไม่สมดุลด้วยวิธีการผสมผสาน เมื่อ Random Seed 10, 20 และ 30	162
4.10	ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยวิธีการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน เมื่อ Random Seed 10, 20 และ 30	163

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง(ต่อ)

ตารางที่		หน้า
4.11	ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยวิธีการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE เมื่อ Random Seed 10, 20 และ 30	164
4.12	ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยวิธีการปรับความไม่สมดุลด้วยวิธีการสุ่มลด เมื่อ Random Seed 10, 20 และ 30	165
4.13	ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยวิธีการปรับความไม่สมดุลด้วยวิธีการผสมผสาน เมื่อ Random Seed 10, 20 และ 30	166
4.14	ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกโดยการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว เมื่อ Random Seed 10, 20 และ 30	167
4.15	ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยการจำแนกด้วยวิธีต้นไม้ตัดสินใจ เมื่อ Random Seed 10, 20 และ 30	168
4.16	ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยการจำแนกด้วยวิธีโครงข่ายประสาทเทียม เมื่อ Random Seed 10, 20 และ 30	169
4.17	ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกโดยการจำแนกด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน เมื่อ Random Seed 10, 20 และ 30	170
4.18	ผลการวิเคราะห์ข้อมูลยอดเงินคงเหลือในบัตรเครดิตของลูกค้าโดยการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว เมื่อ Random Seed 10, 20 และ 30	171
4.19	ผลการวิเคราะห์ข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้าโดยการจำแนกด้วยวิธีต้นไม้ตัดสินใจ เมื่อ Random Seed 10, 20 และ 30	172
4.20	ผลการวิเคราะห์ข้อมูลยอดเงินคงเหลือในบัตรเครดิตของลูกค้าโดยการจำแนกด้วยวิธีโครงข่ายประสาทเทียม เมื่อ Random Seed 10, 20 และ 30	173
4.21	ผลการวิเคราะห์ข้อมูลยอดเงินคงเหลือในบัตรเครดิตของลูกค้าโดยการจำแนกด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน เมื่อ Random Seed 10, 20 และ 30	174
4.22	ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว เมื่อ Random Seed 10, 20 และ 30	175
4.23	ผลการวิเคราะหข้อมูลคุณภาพไวน์แดงโดยการจำแนกด้วยวิธีต้นไม้ตัดสินใจ เมื่อ Random Seed 10, 20 และ 30	176
4.24	ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยการจำแนกด้วยวิธีโครงข่ายประสาทเทียม เมื่อ Random Seed 10, 20 และ 30	177
4.25	ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยการจำแนกด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน เมื่อ Random Seed 10, 20 และ 30	178

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง(ต่อ)

ตารางที่ **หน้า**

4.26 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือน้ำหนวก เมื่อ Random Seed 10, 20 และ 30..... 179

4.27 ผลการวิเคราะห์ชุดข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10, 20 และ 30..... 182

4.28 ผลการวิเคราะห์ชุดข้อมูลคุณภาพไวน์แดง เมื่อ Random Seed 10, 20 และ 30.. 185

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่		หน้า
2.1	ตัวอย่างวิธีเพื่อนบ้านใกล้สุด k ตัว	10
2.2	โครงสร้างต้นไม้ตัดสินใจ	10
2.3	โครงข่ายของเซลล์ประสาท	12
2.4	ระบบโครงข่ายประสาทเทียม	13
2.5	การปรับความชันของเส้นแบ่งแล้วทำให้ได้ระยะขอบที่มากที่สุด	15
2.6	ตัวอย่างการจำแนกข้อมูลของซัพพอร์ตเวกเตอร์แมชชีนโพลิโนเมียลเคอร์เนล	15
4.1	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10	27
4.2	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20	27
4.3	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30	28
4.4	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตลูกค้า เมื่อ Random Seed 10	29
4.5	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20	29
4.6	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30	30
4.7	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 10	31
4.8	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 20	31
4.9	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 30	32
4.10	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10	33
4.11	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20	33

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่		หน้า
4.12	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30	34
4.13	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10	35
4.14	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20	35
4.15	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IB ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30	36
4.16	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 10	37
4.17	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 20	37
4.18	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 30	38
4.19	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10	39
4.20	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20	39
4.21	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30	40
4.22	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10	41
4.23	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20	41
4.24	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30	42
4.25	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 10	43
4.26	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 20	43

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่		หน้า
4.27	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 30	44
4.28	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10	45
4.29	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20	45
4.30	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30	46
4.31	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10	47
4.32	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20	47
4.33	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30	48
4.34	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 10	49
4.35	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 20	49
4.36	ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 30	50
4.37	ต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกเมื่อ RandomSeed10	51
4.38	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10	52
4.39	ต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกเมื่อ Random Seed 20	52
4.40	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20	54
4.41	ต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกเมื่อ Random Seed 30	54

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่		หน้า
4.42	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30	56
4.43	ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10	57
4.44	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10	58
4.45	ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20	58
4.46	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20	59
4.47	ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30	60
4.48	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30	61
4.49	ต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 10	61
4.50	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 10	62
4.51	ต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 20	62
4.52	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 20	63
4.53	ต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 30	64
4.54	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 30	65
4.55	ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10	66
4.56	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10	67
4.57	ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20	67
4.58	ผลการวิเคราะห์ข้อมูลด้วยต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20	68

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่		หน้า
4.59	ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกเมื่อ Random Seed 30	68
4.60	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30	69
4.61	ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10	70
4.62	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10	71
4.63	ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20	71
4.64	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J4 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20	72
4.65	ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30	73
4.66	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30	74
4.67	ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 10	75
4.68	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 10	76
4.69	ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 20	76
4.70	ผลการวิเคราะห์ข้อมูลด้วยต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 20	77
4.71	ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 30	77
4.72	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 30	78
4.73	ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10	79
4.74	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10	79
4.75	ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกเมื่อ Random Seed 20	80

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่		หน้า
4.76	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20	80
4.77	ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30	81
4.78	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30	81
4.79	ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10	82
4.80	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10	82
4.81	ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20	83
4.82	ผลการวิเคราะห์ข้อมูลด้วยต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20	83
4.83	ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30	84
4.84	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30	84
4.85	ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 10	85
4.86	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 10	86
4.87	ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 20	86
4.88	ผลการวิเคราะห์ข้อมูลด้วยต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 20	87
4.89	ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 30	87
4.90	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 30	88
4.91	ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10	89
4.92	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10	90

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่		หน้า
4.93	ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20	91
4.94	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20	92
4.95	ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30	92
4.96	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30	93
4.97	ต้นไม้ตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10	94
4.98	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10	95
4.99	ต้นไม้ตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20	95
4.100	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เครดิต เมื่อ Random Seed 20	96
4.101	ต้นไม้ตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30	97
4.102	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30	98
4.103	ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 10	99
4.104	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 10	99
4.105	ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 20	100
4.106	ผลการวิเคราะห์ข้อมูลด้วยต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 20	100
4.107	ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 30	101
4.108	ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 30	102
4.109	โครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือน้ำหนวก เมื่อ Random Seed 10, 20 และ 30	103
4.110	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหู ชั้นกลางหรือหูน้ำหนวก เมื่อRandom Seed 10	104

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่		หน้า
4.111	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20.....	104
4.112	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่ง ในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30.....	105
4.113	โครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10, 20 และ 30.....	106
4.114	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10.....	106
4.115	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตขอลูกค้า เมื่อ Random Seed 20.....	107
4.116	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30.....	107
4.117	โครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 10, 20 และ 30.....	108
4.118	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 10.....	108
4.119	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 20.....	109
4.120	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 30.....	109
4.121	โครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10, 20 และ 30.....	110
4.122	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10.....	111
4.123	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20.....	111
4.124	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อRandom Seed 30.....	112
4.125	โครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10, 20 และ30.....	113
4.126	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10.....	113
4.127	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20.....	114

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่		หน้า
4.128	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30	114
4.129	โครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 10, 20 และ 30	115
4.130	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 10	115
4.131	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 20	116
4.132	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 30	116
4.133	โครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10, 20 และ 30	117
4.134	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10	118
4.135	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20	118
4.136	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาท การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30	119
4.137	โครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10, 20 และ30	120
4.138	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10	120
4.139	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20	121
4.140	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30	121
4.141	โครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 10, 20 และ 30	122
4.142	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 10	122
4.143	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 20	123
4.144	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 30	123

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่		หน้า
4.145	โครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10, 20 และ 30	124
4.146	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่ง ในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10	125
4.147	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20	125
4.148	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30	126
4.149	โครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10, 20 และ 30	127
4.150	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10	127
4.151	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20	128
4.152	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้าเมื่อ Random Seed 30	128
4.153	โครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 10, 20 และ 30	129
4.154	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 10	129
4.155	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 20	130
4.156	ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 30	130
4.157	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10	131
4.158	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20	131
4.159	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30	132
4.160	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10	133

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่		หน้า
4.161	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20.....	133
4.162	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30.....	134
4.163	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 10.....	135
4.164	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 20.....	135
4.165	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 30.....	136
4.166	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10.....	137
4.167	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20.....	137
4.168	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30.....	138
4.169	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10.....	139
4.170	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20.....	139
4.171	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30.....	140
4.172	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 10.....	141
4.173	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 20.....	141
4.174	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 30.....	142
4.175	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMOการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10.....	143

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่		หน้า
4.176	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20	143
4.177	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30	144
4.178	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10	145
4.179	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20	145
4.180	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30	146
4.181	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 10	147
4.182	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 20	147
4.183	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 30	148
4.184	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10	149
4.185	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMOการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20	149
4.186	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30	150
4.187	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10	151
4.188	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20	151
4.189	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30	152
4.190	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 10	153

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป(ต่อ)

รูปที่		หน้า
4.191	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 20.....	153
4.192	ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 30.....	154

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ในปัจจุบันมีปัญหาการแบ่งข้อมูลที่กำลังได้รับความสนใจคือ ปัญหาการแบ่งกลุ่มข้อมูลที่ไม่สมดุล (Imbalanced Datasets) ซึ่งเกิดจากการที่มีข้อมูล 2 กลุ่ม หรือมากกว่า 2 กลุ่ม โดยข้อมูลที่เป็นกลุ่มส่วนมาก (Majority) จะมีข้อมูลจำนวนมากกว่า ในขณะเดียวกันข้อมูลกลุ่มส่วนน้อย (Minority) จะมีข้อมูลจำนวนน้อยกว่า ทั้งนี้เนื่องจากโดยธรรมชาติของความเป็นจริง การที่จะกำหนดให้ขนาดของข้อมูลในกลุ่มส่วนมากและกลุ่มส่วนน้อยมีขนาดที่เท่าเทียมกันเพื่อการสอนหรือการจัดกลุ่มข้อมูลนั้นเป็นเรื่องยากหรืออาจจะเป็นไปไม่ได้ ดังนั้นจึงเป็นปัญหาที่ท้าทายและมีความยากมากสำหรับการหาขั้นตอนวิธีที่เหมาะสมสำหรับการแบ่งกลุ่มข้อมูลที่ไม่สมดุล ทั้งนี้เนื่องจากถ้านำข้อมูลทั้งสองชุดเข้าสู่ขั้นตอนวิธีพร้อมกันทั้งหมด จะทำให้ผลการแบ่งกลุ่มข้อมูลเกิดความผิดพลาด กล่าวคือ ข้อมูลที่อยู่ในกลุ่มส่วนน้อยถูกครอบงำหรือจะถูกจัดให้ไปอยู่ในกลุ่มส่วนมากทั้งหมด ซึ่งจะนำไปสู่ปัญหาที่เรียกว่า ปัญหาการจำแนกข้อมูลผิดกลุ่ม (Misclassification) (เบญจภรณ์ และคณะ, 2557)

จากการศึกษางานวิจัยที่เกี่ยวข้อง วีระยุทธ และคณะ (2557) ได้ทำการศึกษางานวิจัยเกี่ยวกับการรักษาซ้ำของผู้ป่วยจิตเภท เรื่องการพัฒนาแบบจำลองเพื่อการพยากรณ์การรักษาซ้ำของผู้ป่วยโรคจิตเภทโดยเทคนิคเหมืองข้อมูล งานวิจัยฉบับนี้มีวัตถุประสงค์เพื่อสร้างแบบจำลองในการพยากรณ์ระยะเวลาการกลับมารักษาซ้ำ จากผลการวิจัยสามารถสรุปได้ว่าข้อมูลการรักษาของผู้ป่วยโรคจิตเภททางการแพทย์มีข้อมูลที่มีความผิดปกติและมีความไม่สมดุลของข้อมูล (Imbalanced Data Set) ซึ่งผู้วิจัยได้ใช้วิธีการกรองด้วยซัพพอร์ตเวกเตอร์แมชชีนแล้วใช้วิธีการสุ่มเกินโดยเทคนิค SMOTE และวิธีการสุ่มลด ในการแก้ปัญหา แล้วนำข้อมูลที่สมดุลแล้วมาจำแนกพบว่าวิธีการกรองแล้วใช้การสุ่มเกินโดยเทคนิค SMOTE สามารถเพิ่มประสิทธิภาพให้แบบจำลองเพิ่มขึ้น โดยค่าความถูกต้องเพิ่มขึ้นเฉลี่ยร้อยละ 46.36 ค่าความไวเพิ่มขึ้นเฉลี่ยร้อยละ 20.05 และค่าความจำเพาะเพิ่มขึ้นร้อยละ 32.69 ซึ่งมากกว่าวิธีการกรองแล้วใช้วิธีการสุ่มลด (Under Sampling) เชาวนันท์ และคณะ (2556) ได้ทำการศึกษางานวิจัยเกี่ยวกับมะเร็งปากมดลูก เรื่องแบบจำลองการทำนายผลการรักษาผู้ป่วยมะเร็งปากมดลูกด้วยวิธีโครงข่ายประสาทเทียม งานวิจัยนี้นำเสนอแบบจำลองเพื่อทำนายผลการรักษาผู้ป่วยมะเร็งปากมดลูกที่เข้ารับการรักษาด้วยวิธีการฉายรังสีโดยการประยุกต์ใช้วิธีโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ ด้วยการนำขั้นตอนวิธีการเรียนรู้แบบมีค่าใช้จ่าย (Cost-Sensitive Learning : CSL) และวิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-sampling Technique: SMOTE) เมื่อเปรียบเทียบประสิทธิภาพการทำนายพบว่าวิธีโครงข่ายประสาทเทียมที่มีการแก้ปัญหาความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มโดยเทคนิค SMOTE ของชุดข้อมูลตัวอย่างมีประสิทธิภาพการทำนายด้วยค่าความถูกต้อง 81.71% ค่าความไว 94.47% และค่าความจำเพาะ 55.47% สูงกว่าวิธีโครงข่ายประสาทเทียมที่มีการแก้ปัญหาความไม่สมดุลของข้อมูลด้วยวิธีการเรียนรู้แบบมีค่าใช้จ่าย เบญจภรณ์ และคณะ (2557) ได้ทำการศึกษางานวิจัยเกี่ยวกับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิธีการแบ่งกลุ่มข้อมูลที่ไม่สมดุล เรื่องวิธีการที่เหมาะสมสำหรับการแบ่งกลุ่มข้อมูลที่ไม่สมดุลสูง งานวิจัยนี้ได้นำเสนอวิธีการเลือกคุณลักษณะสำหรับข้อมูลไม่สมดุลที่มีจำนวนมิติข้อมูลที่สูง เพื่อลดจำนวนมิติข้อมูลที่ซ้ำซ้อนและเพิ่มประสิทธิภาพของการจำแนกด้วยการใช้ข้อมูลเดิมที่ไม่ได้ทำการปรับปรุง และวิธีการสุ่มแบบผสมผสาน (Hybrid Methods) จากนั้นค้นหาคุณลักษณะย่อยด้วยขั้นตอนวิธีทางพันธุกรรมร่วมกับโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ ทำการเปรียบเทียบประสิทธิภาพของวิธีการที่นำเสนอโดยใช้ค่าความถูกต้อง ค่าเฉลี่ยเรขาคณิต และความถูกต้องของการจำแนกข้อมูลกลุ่มส่วนน้อย รวมถึงเวลาที่ใช้ในการประมวลผล ทำการทดสอบกับ 3 ตัวจำแนกข้อมูล ได้แก่ โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ โครงข่ายฟังก์ชันรัศมีฐาน และซัพพอร์ตเวกเตอร์แมชชีน ผลปรากฏว่าตัวจำแนกข้อมูลทั้ง 3 วิธี ให้ค่าความถูกต้อง ค่าเฉลี่ยเรขาคณิต โดยส่วนมากของการจำแนกชุดข้อมูลของวิธีการสุ่มแบบผสมผสานดีกว่าการใช้ข้อมูลเดิมที่ไม่ได้ทำการปรับปรุง McCarthy et al. (2005) ได้ทำ การศึกษาเกี่ยวกับความสามารถในการแก้จำแนกกลุ่ม เรื่องการทดสอบความสามารถในการจำแนกข้อมูลที่ค้นพบได้ยากด้วยขั้นตอนวิธี (Algorithm) วิธีการเรียนรู้แบบมีค่าใช้จ่าย (Cost-Sensitive) และเทคนิคการสุ่ม 2 เทคนิค ได้แก่ วิธีการสุ่มเกิน (Over-Sampling) และวิธีการสุ่มลด (Under-Sampling) ผลจากการทดลองสามารถสรุปได้ว่า โดยรวมแล้ววิธีการเรียนรู้แบบมีค่าใช้จ่ายมีประสิทธิภาพดีกว่าวิธีการสุ่มเกินและวิธีการสุ่มลด

ดังนั้นปัญหาพิเศษฉบับนี้ทำการศึกษาการปรับความไม่สมดุลของข้อมูลด้วยวิธีต่างๆ 4 วิธี คือ วิธีการสุ่มเกิน (Over Sampling) วิธีการสุ่มเกินโดยใช้เทคนิค SMOTE (Synthetic Minority Over-Sampling Technique) วิธีการสุ่มลด (Under Sampling) และวิธีการผสมผสาน (Hybrid Methods) ทำการจำแนก (Classification) ด้วยวิธีต่างๆ 4 วิธี คือ วิธีเพื่อนบ้านใกล้สุด k ตัว (k-Nearest Neighbor) วิธีต้นไม้ตัดสินใจ (Decision Tree) วิธีโครงข่ายประสาทเทียม (Artificial Neural Network) และวิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) เพื่อต้องการเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลทั้ง 4 วิธี ว่าวิธีใดมีประสิทธิภาพและเหมาะสมกับรูปแบบของชุดข้อมูล โดยใช้การทดสอบเปรียบเทียบประสิทธิภาพด้วยค่าความถูกต้อง (Accuracy) ค่าความไว (Sensitivity) ค่าความจำเพาะ (Specificity) และค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE)

1.2 วัตถุประสงค์ที่ศึกษา

1. เพื่อเปรียบเทียบประสิทธิภาพในการปรับข้อมูลที่ไม่สมดุล 4 วิธี เมื่อมีการจำแนกข้อมูลแต่ละวิธี

2. เพื่อเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูล 4 วิธี เมื่อมีการปรับข้อมูลที่ไม่สมดุลแต่ละวิธี

3. เพื่อเปรียบเทียบประสิทธิภาพในภาพรวมของการจำแนกข้อมูล 4 วิธี และการปรับข้อมูลที่ไม่สมดุล 4 วิธี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3 ประโยชน์ที่คาดว่าจะได้รับ

สามารถนำวิธีที่ได้ไปใช้เป็นแนวทางในการเลือกวิธีการปรับข้อมูลที่ไม่สมดุลเพื่อนำไปจำแนกข้อมูลให้มีประสิทธิภาพมากยิ่งขึ้น และช่วยลดข้อผิดพลาดในการเลือกใช้วิธีการปรับข้อมูลที่ไม่สมดุลและการจำแนกข้อมูล

1.4 ขอบเขตการศึกษา

ในการทำปัญหาพิเศษได้ทำการปรับข้อมูลเพื่อให้เกิดความสมดุล 4 วิธี คือ วิธีการสุ่มเกิน (Over Sampling) วิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-Sampling Technique) วิธีการสุ่มลด (Under Sampling) และวิธีการผสมผสาน (Hybrid Methods) ทำการแบ่งชุดข้อมูลอย่างสุ่มด้วยโปรแกรม WEKA ออกเป็น 3 ส่วน โดยมีอัตราส่วน 70:20:10 ในข้อมูลส่วนที่ 1 ข้อมูลเรียนรู้ (Training data) นำไปสร้างตัวแบบ (Model) ร้อยละ 70 ข้อมูลส่วนที่ 2 ข้อมูลตรวจสอบความถูกต้อง (Validation data) นำข้อมูลไปประเมินความผิดพลาดของตัวแบบ ร้อยละ 20 และข้อมูลส่วนที่ 3 ข้อมูลทดสอบ (Testing Data) นำไปทดสอบตัวแบบ ร้อยละ 10 โดยการกำหนดตัวสร้างเลขสุ่มเทียม (Random Seed) เป็น 10, 20 และ 30 (พนิดา และคณะ, 2560) ซึ่งจะใช้วิธีการทดสอบประกอบด้วย 4 วิธี คือ วิธีเพื่อนบ้านใกล้สุด k ตัว ใช้ขั้นตอนวิธีชนิด IBk วิธีต้นไม้ตัดสินใจ ใช้ขั้นตอนวิธีชนิด J48 (C4.5) วิธีโครงข่ายประสาทเทียม ใช้ขั้นตอนวิธีชนิดเพอร์เซปตรอนแบบหลายชั้น โดยชั้นซ่อน 1 ชั้น และวิธีซัพพอร์ตเวกเตอร์แมชชีน ใช้ขั้นตอนวิธีชนิด SMO เคอร์เนลแบบโพลิโนเมียลเคอร์เนล ในการเปรียบเทียบประสิทธิภาพของการจำแนกทั้ง 4 วิธี ใช้ค่าความถูกต้อง (Accuracy) ค่าความไว (Sensitivity) ค่าความจำเพาะ (Specificity) และค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) ทำการเก็บรวบรวมข้อมูลเป็นจำนวน 3 ชุด คือ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) และคุณภาพไวน์แดง (Red Wine Quality)

1.5 นิยามศัพท์เฉพาะ

บัตรเครดิต (Credit Card) คือ บัตรที่เราต่างคุ้นเคยกันเป็นอย่างดี โดยเฉพาะในปัจจุบันที่หลายๆ คนเลือกที่จะใช้บัตรเครดิตแทนเงินสด เพราะสามารถพกพาได้สะดวก มีเพียงบัตรเดียวก็สามารถกดเงินได้ทันทีโดยไม่ต้องจ่ายเงินสด โดยบัตรเครดิตที่เรารู้จักกันนี้คือบัตรซึ่งเราจะได้รับวงเงินอนุมัติจากทางธนาคาร หรือพูดง่ายๆ ว่าเป็นการยืมเงินมาจากธนาคารเพื่อชำระสินค้าก่อนในวงเงินที่ทางธนาคารกำหนดไว้สำหรับแต่ละบุคคล แล้วจึงผ่อนชำระคืนในภายหลัง โดยหักออกจากบัญชีของลูกค้าเป็นรายเดือน (เกร็ดความรู้, 2556)

ภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Otitis Media with Effusion) หรือเรียกย่อว่า โอเอ็มอี เป็นภาวะที่มีน้ำขังอยู่ในหูชั้นกลางโดยที่ไม่มีอาการแสดงของการอักเสบหรือติดเชื้อ ผู้ป่วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มักจะมีอาการหูอื้อ การได้ยินลดลง แต่ไม่มีอาการปวดหูและไม่มีไข้ เมื่อตรวจดูในหูจะไม่พบการบวมแดงของแก้วหู แต่จะมีการขยับของเยื่อแก้วหูลดลง เพราะมีน้ำขังอยู่ด้านหลัง ภาวะนี้มักพบในเด็กที่มีปัจจัยเสี่ยง เช่น ปากแหว่งเพดานโหว่ กลุ่มอาการของดาวน์ซินโดรม มีโครงสร้างใบหน้าที่ผิดปกติ (Medthai, 2552)

ไวน์แดง (Red Wine) คือ ไวน์ (Wine) ซึ่งเป็นเครื่องดื่มแอลกอฮอล์ (Alcoholic Beverage) เป็นไวน์ที่ทำจากองุ่นดำ โดยมีสีของเปลือกองุ่น คือแอนโทไซยานิน (Anthocyanin) เป็นตัวทำให้ไวน์มีสีแดง ซึ่งต้องนำองุ่นที่บดแล้วมาหมัก (Fermentation) ทั้งเปลือกที่อุณหภูมิที่ 24–29 องศาเซลเซียส เป็นเวลา 4–6 วัน จนกระทั่งค่อยๆ ลดอุณหภูมิเป็น 2–4 องศาเซลเซียส การนำองุ่นมาหมักทั้งเปลือก จะเป็นการสกัดสี กลิ่นรส และแทนนิน ออกจากเปลือกองุ่น หลังจากหมักเสร็จแล้วก็แยกเปลือกออกจากไวน์ (Wordpress, 2558)

ตัวสร้างเลขสุ่มเทียม (Random Seed) หรือตัวสร้างบิตสุ่มแบบกำหนดได้ (Deterministic Random Bit Generator : DRBG) เป็นขั้นตอนวิธีสำหรับใช้ในการสร้างลำดับของตัวเลขที่มีความใกล้เคียงกับสมบัติของการสุ่ม ถึงแม้ว่าลำดับตัวเลขที่ได้จากขั้นตอนวิธีตัวสร้างเลขสุ่มเทียมนี้จะใกล้เคียงกับลำดับเลขสุ่มแท้จริงมากแค่ไหน แต่ก็ไม่ได้เป็นลำดับตัวเลขแบบสุ่มที่แท้จริงเนื่องจากลำดับตัวเลขที่ได้จากตัวสร้างเลขสุ่มเทียมทั้งหมดได้มาจากกลุ่มเล็กๆ ของค่าเริ่มต้นที่เรากำหนดให้เป็นตัวตั้งต้น (Seed) ของตัวสร้างเลขสุ่มเทียม (พนิดา และคณะ, 2560)

1.6 ประชากรและตัวอย่าง

1.6.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME)

ประชากร คือ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก ประเทศอังกฤษ

ตัวอย่าง คือ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก จากภาควิชาสรีรวิทยา มหาวิทยาลัยออกซ์ฟอร์ด ประเทศอังกฤษ ในวันที่ 9 มกราคม 2016

1.6.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance)

ประชากร คือ ยอดเงินคงเหลือในบัตรเครดิตของลูกค้า ประเทศสหรัฐอเมริกา

ตัวอย่าง คือ ยอดเงินคงเหลือในบัตรเครดิตของลูกค้า สถาบันการเงินในนิวยอร์ก ประเทศสหรัฐอเมริกา ในวันที่ 19 ตุลาคม 2017

1.6.3 คุณภาพไวน์แดง (Red Wine Quality)

ประชากร คือ คุณภาพไวน์แดง ทางตอนเหนือ ประเทศโปรตุเกส

ตัวอย่าง คือ คุณภาพไวน์แดง จากบริษัท Vinho Verde ทางตอนเหนือ ประเทศโปรตุเกส ในวันที่ 11 ตุลาคม 2009

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้ ผู้วิจัยได้นำเสนอเนื้อหาที่เน้นถึงทฤษฎีและงานวิจัยที่เกี่ยวข้องโดยมีรายละเอียดของเนื้อหาประกอบด้วยหัวข้อย่อย 5 หัวข้อ ดังนี้

2.1 ข้อมูลที่ไม่สมดุล (Imbalanced Datasets)

2.2 การปรับความไม่สมดุลของข้อมูล (Imbalanced Adjustment of Datasets)

- 2.2.1 วิธีการสุ่มเกิน
- 2.2.2 วิธีการสุ่มเกินโดยเทคนิค SMOTE
- 2.2.3 วิธีการสุ่มลด
- 2.2.4 วิธีการผสมผสาน

2.3 วิธีการจำแนก (Classification)

- 2.3.1 วิธีเพื่อนบ้านใกล้สุด k ตัว
- 2.3.2 วิธีต้นไม้ตัดสินใจ
- 2.3.3 วิธีโครงข่ายประสาทเทียม
- 2.3.4 วิธีซัพพอร์ตเวกเตอร์แมชชีน

2.4 การเปรียบเทียบประสิทธิภาพของวิธีการจำแนก

2.5 งานวิจัยที่เกี่ยวข้อง

2.1 ข้อมูลที่ไม่สมดุล (Imbalanced Datasets)

ข้อมูลที่ไม่สมดุล (Imbalanced Datasets) หมายถึง ข้อมูลที่มีการกระจายตัวที่ไม่เท่าเทียมกัน หรืออาจจะหมายถึง ข้อมูลซึ่งอัตราจำนวนสมาชิกในกลุ่มส่วนมากและกลุ่มส่วนน้อยมีจำนวนไม่เท่ากัน เช่น 100:1, 1000:1 หรือ 10000:1 เป็นต้น ยกตัวอย่างเช่น ข้อมูลผู้ป่วยโรคมะเร็งชนิดหนึ่ง ปัญหานี้คำตอบที่เราต้องการคือ ผู้ป่วยรายนั้นเป็นมะเร็ง (Positive) หรือไม่เป็นมะเร็ง (Negative) ซึ่งข้อมูลผู้ป่วยที่ไม่เป็นโรคมะเร็ง (กลุ่มส่วนมาก) อาจจะมีข้อมูลหลายหมื่นคน แต่ข้อมูลผู้เป็นโรคมะเร็งอาจมีเพียงประมาณหลักร้อยคน (กลุ่มส่วนน้อย) ดังนั้น ถ้าเรานำข้อมูลทั้งสองกลุ่มมาสอนเพื่อแบ่งกลุ่มข้อมูลพร้อมกันทั้งหมด จะพบว่าคำตอบของการพยากรณ์เพื่อการแบ่งกลุ่มข้อมูลทุกๆ ข้อมูลจะถูกจัดเป็นกลุ่มผู้ป่วยที่ไม่เป็นโรคมะเร็งทั้งหมด (Positive Class) (เบญจภรณ์ และคณะ, 2557)

2.2 การปรับความไม่สมดุลของข้อมูล (Imbalanced Adjustment of Datasets)

สำหรับขั้นตอนวิธีที่ใช้สำหรับปรับข้อมูลที่ไม่สมดุล จะประกอบด้วย 4 วิธีการหลักๆ คือ วิธีการสุ่มเกิน วิธีการสุ่มเกินโดยเทคนิค SMOTE วิธีการสุ่มลด และวิธีการผสมผสาน

2.2.1 วิธีการสุ่มเกิน (Over Sampling)

วิธีการสุ่มเกินเป็นเทคนิคหรือวิธีที่ใช้ในการเพิ่มข้อมูลที่อยู่ในกลุ่มส่วนน้อยให้มีจำนวนใกล้เคียงหรือเท่ากับจำนวนข้อมูลในกลุ่มส่วนมาก โดยการสุ่มเกินเพื่อเพิ่มข้อมูลให้กลุ่มส่วนน้อยจะเพิ่มข้อมูลโดยการสุ่มเลือกข้อมูลจากข้อมูลเดิม (กีระชาติ, 2559)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างที่ 2.1 สมมติว่ามีชุดข้อมูล A ซึ่งมีจำนวนกลุ่ม (Class) ทั้งหมด 2 กลุ่ม ได้แก่ กลุ่ม No และกลุ่ม Yes โดยข้อมูลในกลุ่ม Yes มีจำนวนทั้งหมด 3 ข้อมูล และข้อมูลในกลุ่ม No มีจำนวนทั้งหมด 7 ข้อมูล แสดงดังตารางที่ 2.1 (Hogan, 2013)

ตารางที่ 2.1 ชุดข้อมูล A ที่ยังไม่ได้ทำการปรับความสมดุล

ลำดับที่	X1	X2	X3	X4	X5	Y
1	30	low	35	coherent	1	No
2	30	low	35	incoherent	4	No
3	30	low	40	coherent	0	No
4	30	low	40	incoherent	1	No
5	30	low	45	coherent	2	No
6	30	low	45	incoherent	2	No
7	60	low	40	coherent	9	Yes
8	60	low	40	incoherent	8	Yes
9	60	low	45	coherent	5	No
10	60	low	45	incoherent	9	Yes

ตารางที่ 2.2 ชุดข้อมูล A หลังจากทำการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน (Over Sampling)

ลำดับที่	X1	X2	X3	X4	X5	Y
1	30	low	35	coherent	1	no
2	30	low	35	incoherent	4	no
3	30	low	40	coherent	0	no
4	30	low	40	incoherent	1	no
5	30	low	45	coherent	2	no
6	30	low	45	incoherent	2	no
7	60	low	40	coherent	9	Yes
8	60	low	40	incoherent	8	Yes
9	60	low	45	coherent	5	no
10	60	low	45	incoherent	9	Yes
11	**60**	**low**	**40**	**coherent**	**9**	**Yes**
12	**60**	**low**	**40**	**incoherent**	**8**	**Yes**
13	**60**	**low**	**45**	**incoherent**	**9**	**Yes**

จากตารางที่ 2.2 เพิ่มจำนวนกลุ่มส่วนน้อยมา 3 ลำดับ ได้แก่ ลำดับที่ 11, 12 และ 13 โดยลำดับที่ 11 ได้มาจากลำดับที่ 7 ลำดับที่ 12 ได้มาจากลำดับที่ 8 และลำดับที่ 13 ได้มาจากลำดับที่ 10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.2 วิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-Sampling Technique)

วิธีการสุ่มเกินโดยเทคนิค SMOTE เป็นการสุ่มสร้างข้อมูลจากกลุ่มส่วนน้อยตามจำนวนที่กำหนด โดยจะสร้างข้อมูลสังเคราะห์จากข้อมูลตัวอย่างด้วยการวัดระยะห่างจากจุดข้อมูลตัวอย่างไปยังจุดข้อมูลใกล้เคียง แล้วสุ่มสร้างข้อมูลสังเคราะห์ขึ้นโดยข้อมูลสังเคราะห์ที่สร้างขึ้นจะอยู่ภายในระยะห่างจากจุดข้อมูลตัวอย่างไปยังจุดข้อมูลเพื่อนบ้านซึ่งจุดใหม่ที่สร้างขึ้นมาแสดงดังสมการที่ 2.1 (ภรันยา, 2559)

$$X_{new} = X_i + (\hat{X}_i - X_i) \times \delta \qquad (2.1)$$

โดยที่ X_{new} คือ ข้อมูลใหม่

X_i คือ ข้อมูลที่สุ่มในตอนแรก

$\hat{X}_i$ คือ ข้อมูลที่สุ่มมาอีก เช่น สุ่มมาอีก 5 จุด

δ คือ ค่าสุ่มตั้งแต่ 0-1

ตัวอย่างที่ 2.2 จากวิธีการสุ่มเกินถ้าเลือกใช้เทคนิค SMOTE เทคนิคนี้จะทำการเลือกข้อมูลที่อยู่ในกลุ่ม Yes ขึ้นมา แล้วทำการเลือกข้อมูลเพื่อนบ้านที่อยู่ในกลุ่มเดียวกัน แล้วคำนวณระยะห่างระหว่างจุดสองจุด ทำการสุ่มเพิ่มค่าขึ้นเล็กน้อยหรือลดค่าลงเล็กน้อย โดยเพิ่มข้อมูลเป็นจำนวน 3 ค่า โดยใช้ข้อมูลจากลำดับที่ 7, 8 และ 10 เป็นข้อมูลตั้งต้น เพื่อทำให้ชุดข้อมูล A จากตารางที่ 2.1 มีความสมดุลเกิดขึ้น จะได้ชุดข้อมูลใหม่โดยมีข้อมูลทั้งหมด 13 ค่า แบ่งเป็นข้อมูลในกลุ่ม No จำนวน 7 ค่า และข้อมูลในกลุ่ม Yes จำนวน 6 ค่า แสดงดังตารางที่ 2.3 (Hogan, 2013)

ตารางที่ 2.3 ชุดข้อมูล A หลังจากทำการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินโดยเทคนิค SMOTE

ลำดับที่	X1	X2	X3	X4	X5	Y
1	30.0	low	35.0	coherent	1.0	No
2	30.0	low	35.0	incoherent	4.0	No
3	30.0	low	40.0	coherent	0.0	No
4	30.0	low	40.0	incoherent	1.0	No
5	30.0	low	45.0	coherent	2.0	No
6	30.0	low	45.0	incoherent	2.0	No
7	60.0	low	40.0	coherent	9.0	Yes
8	60.0	low	40.0	incoherent	8.0	Yes
9	60.0	low	45.0	coherent	5.0	No
10	60.0	low	45.0	incoherent	9.0	Yes

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.3 ชุดข้อมูล A หลังจากทำการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินโดยเทคนิค SMOTE (ต่อ)

ลำดับที่	X1	X2	X3	X4	X5	Y
11	60.0	low	43.0317260	incoherent	9.0	Yes
12	60.0	low	40.0	incoherent	8.38518918	Yes
13	60.0	low	44.8845907	incoherent	8.82402319	Yes

2.2.3 วิธีการสุ่มลด (Under Sampling)

วิธีการสุ่มลดเป็นการลดจำนวนของข้อมูลกลุ่มส่วนมากจนกระทั่งจำนวนข้อมูลของกลุ่มส่วนมากใกล้เคียงหรือเท่ากับจำนวนกลุ่มส่วนน้อย โดยทำการสุ่มเลือกข้อมูลของกลุ่มส่วนน้อยทำเช่นนี้ไปเรื่อยๆ จนกระทั่งจำนวนข้อมูลในกลุ่มส่วนมากใกล้เคียงหรือเท่ากับจำนวนกลุ่มส่วนน้อย (กีระชาติ, 2559)

ตัวอย่างที่ 2.3 จากวิธีการสุ่มลดนี้จะทำการลดจำนวนข้อมูลที่อยู่ในกลุ่ม No โดยลดข้อมูลเป็นจำนวน 4 ค่า เพื่อทำให้ชุดข้อมูล A จากตารางที่ 2.1 มีความสมดุลเกิดขึ้น สมมติว่าทำการสุ่มลดจำนวนข้อมูลจากกลุ่มส่วนมากลงเป็นจำนวน 4 ค่า (ลดข้อมูลลำดับที่ 2, 3 และ 9) จะได้ชุดข้อมูลใหม่โดยมีข้อมูลทั้งหมด 7 ค่า แบ่งเป็นข้อมูลในกลุ่ม Yes จำนวน 3 ค่า และข้อมูลในกลุ่ม No จำนวน 4 ค่า แสดงดังตารางที่ 2.4 (Hogan, 2013)

ตารางที่ 2.4 ชุดข้อมูล A หลังจากทำการปรับความไม่สมดุลด้วยวิธีการสุ่มลด (Under Sampling)

ลำดับที่	X1	X2	X3	X4	X5	Y
1	30	low	45	incoherent	2	No
2	30	low	35	coherent	1	No
3	30	low	40	incoherent	1	No
4	30	low	45	coherent	2	No
5	60	low	40	coherent	9	Yes
6	60	low	40	incoherent	8	Yes
7	60	low	45	incoherent	9	Yes

2.2.4 วิธีการผสมผสาน (Hybrid Methods)

วิธีการผสมผสานเป็นวิธีการที่นำวิธีสุ่มเกินและวิธีสุ่มลดมาทำงานร่วมกันโดยการใช้วิธีนี้จะเป็นการสุ่มลดจำนวนข้อมูลจากกลุ่มส่วนมากและทำการสุ่มเพิ่มข้อมูลในกลุ่มส่วนน้อย ให้จำนวนข้อมูลจากทั้งสองกลุ่มมีจำนวนใกล้เคียงกันหรือเท่ากัน (กีระชาติ, 2559)

ตัวอย่างที่ 2.4 จากวิธีการผสมผสาน วิธีนี้จะทำการสุ่มลดข้อมูลจากกลุ่มส่วนมากลงจำนวน 2 ค่า และสุ่มเพิ่มข้อมูลจากกลุ่มส่วนน้อยเพิ่มขึ้นจำนวน 2 ค่า เพื่อให้ชุดข้อมูล A จากตารางที่ 2.1 มีความสมดุลเกิดขึ้น สมมติว่าทำการสุ่มลดจำนวนข้อมูลจากกลุ่มส่วนมากลงจำนวน 2 ค่า (สุ่มลดข้อมูลในลำดับที่ 3 และ 4) และทำการสุ่มเพิ่มข้อมูลจากกลุ่มส่วนน้อยด้วยการสุ่มเกินจากข้อมูลเดิมจำนวน

เอกสารนี้มีลิขสิทธิ์ ห้ามมิให้ทำการคัดลอก ทำซ้ำ ดัดแปลง เผยแพร่ หรือนำไปใช้ประโยชน์ทางการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3 ค่า (สุ่มเพิ่มข้อมูลจากลำดับที่ 8) จะได้จำนวนข้อมูลทั้งหมด 10 ค่า โดยเป็นข้อมูลจากกลุ่ม Yes มีทั้งหมด 5 ค่า และจำนวนข้อมูลทั้งหมดจากกลุ่ม No มีทั้งหมด 5 ค่า แสดงดังตารางที่ 2.5 (Hogan, 2013)

ตารางที่ 2.5 ชุดข้อมูล A หลังจากทำการปรับความไม่สมดุลด้วยวิธีการผสมผสาน

ลำดับที่	X1	X2	X3	X4	X5	Y
1	30	low	45	coherent	2	No
2	30	low	45	coherent	2	No
3	30	low	35	incoherent	4	No
4	30	low	35	coherent	1	No
5	60	low	45	coherent	5	No
6	60	low	40	coherent	9	Yes
7	60	low	40	incoherent	8	Yes
8	60	low	45	incoherent	9	Yes
9	60	low	40	incoherent	8	Yes
10	60	low	40	incoherent	8	Yes

2.3 วิธีการจำแนก (Classification)

เป็นวิธีในการจำแนกข้อมูลด้วยคุณลักษณะต่างๆ ที่ได้มีการกำหนดไว้แล้ว วิธีนี้เหมาะกับการสร้างตัวแบบเพื่อการพยากรณ์ค่าข้อมูล (Predictive Modeling) ในอนาคตจากการที่ได้จำแนกข้อมูลตัวอย่างไว้แล้ว ซึ่งในลักษณะดังกล่าวเรียกว่า การเรียนรู้แบบมีผู้สอน (Supervised Learning) วิธีการจำแนกเป็นกระบวนการสร้างตัวแบบเพื่อจัดข้อมูลให้อยู่ในกลุ่มที่กำหนด ตัวอย่างเช่น การแบ่งประเภทลูกค้าว่าเชื่อถือได้หรือไม่ ซึ่งเป็นการสร้างตัวแบบโดยการเรียนรู้จากข้อมูลที่ได้กำหนดไว้เรียบร้อยแล้ว (พนิดา และคณะ, 2560)

2.3.1 วิธีเพื่อนบ้านใกล้สุด k ตัว (k-Nearest Neighbor)

วิธีเพื่อนบ้านใกล้สุด k ตัว เป็นวิธีการที่ได้รับความนิยมในการใช้งานเป็นอย่างมาก สาเหตุเนื่องจากเป็นวิธีการที่ง่ายและมีประสิทธิภาพ ซึ่งสามารถนำไปประยุกต์ใช้กับงานได้อย่างหลากหลาย เช่น งานทางด้านการจำแนก รวมถึงงานทางด้านการแทนที่ข้อมูลที่สูญหาย (Missing Values Imputation) ซึ่งมีวิธีการดำเนินการดังนี้ (พนิดา และคณะ, 2560)

1. กำหนดค่า k เพื่อใช้พิจารณาสมาชิกที่อยู่ใกล้กันมากที่สุด เช่น k = 3 คือจะพิจารณาเฉพาะข้อมูล 3 ตัวแรกที่อยู่ใกล้กับจุดที่ต้องการทำนาย

2. คำนวณหาระยะห่างระหว่างข้อมูลตัวอย่างที่สนใจกับข้อมูลอื่นๆ ทุกตัวด้วยระยะห่างยูคลิเดียน (Euclidian Distance) จากสมการที่ 2.2

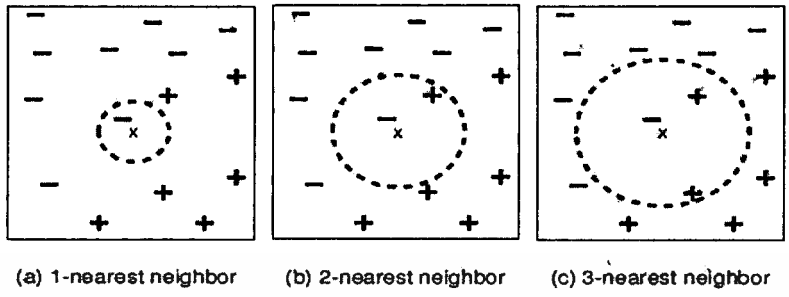
$$D_{Euclidian}(x_i, y_i) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \qquad (2.2)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่ $D_{Euclidian}(x_i, y_i)$ คือ ระยะห่างระหว่างตัวอย่าง x_i กับตัวอย่าง y_i

k คือ คุณลักษณะทั้งหมดของตัวอย่าง

3. เลือกค่าข้อมูลที่มีค่าระยะห่างน้อยที่สุด k ตัว เพื่อนำมาพิจารณาหาคำตอบ ดังรูปที่ 2.1



รูปที่ 2.1 ตัวอย่างวิธีเพื่อนบ้านใกล้สุด k ตัว

(a) วิธีเพื่อนบ้านใกล้สุด k ตัว โดยพิจารณาจากข้อมูล 1 ตัว

(b) วิธีเพื่อนบ้านใกล้สุด k ตัว โดยพิจารณาจากข้อมูล 2 ตัว

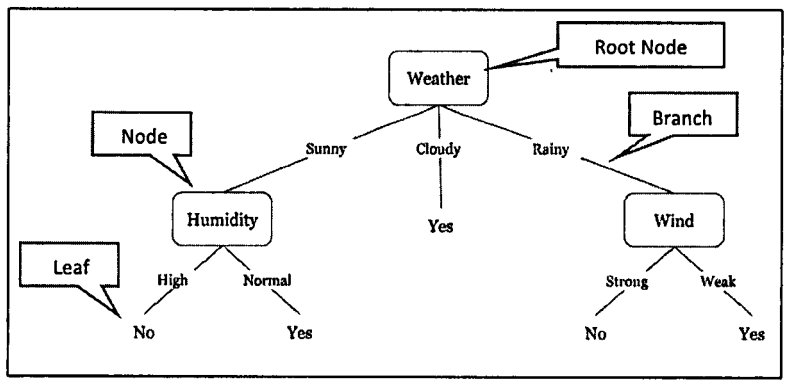
(c) วิธีเพื่อนบ้านใกล้สุด k ตัว โดยพิจารณาจากข้อมูล 3 ตัว

2.3.1.1 ขั้นตอนวิธี IBk

IBk เป็นฟังก์ชันหลัก ซึ่งเป็นพื้นฐานของวิธีเพื่อนบ้านใกล้สุด k ตัว อย่างไรก็ตาม ขั้นตอนวิธี IBk ยังสามารถกำหนดน้ำหนัก ระยะห่างและทางเลือก (Option) เพื่อกำหนดค่า k โดยใช้การตรวจสอบไขว้ (Cross-Validation) (สุรวัชร และสายชล, 2560)

2.3.2 วิธีต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจเป็นเทคนิคที่ให้ผลลัพธ์ในลักษณะของโครงสร้างต้นไม้ ซึ่งเมื่อมีข้อมูลที่ต้องการจำแนกก็จะนำคุณลักษณะต่าง ๆ ของข้อมูลนั้นไปเทียบกับเส้นทางในต้นไม้ จนกระทั่งกลุ่ม (Class) ปลายทางซึ่งก็คือกลุ่มของข้อมูลที่เหมือนกัน ภายในต้นไม้จะประกอบไปด้วยโหนด (Node) ซึ่งแต่ละโหนดจะมีคุณลักษณะเป็นตัวทดสอบกิ่งของต้นไม้ (Branch) ซึ่งแสดงถึงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกเลือกทดสอบ และใบ (Leaf) ซึ่งเป็นสิ่งที่อยู่ล่างสุดของต้นไม้ตัดสินใจแสดงถึงกลุ่ม (Class) ก็คือผลลัพธ์ที่ได้จากการทำนาย โหนดที่อยู่บนสุดของต้นไม้เรียกว่าโหนดราก (Root Node) โครงสร้างของต้นไม้ตัดสินใจแสดง ดังรูปที่ 2.2 (ภรันยา, 2559)



รูปที่ 2.2 โครงสร้างต้นไม้ตัดสินใจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ต้นไม้ตัดสินใจมีค่าวัดความสามารถของการจำแนกของแต่ละคุณลักษณะหรือปัจจัย ดังต่อไปนี้

Gini Index คือ ค่าที่บ่งบอกว่าคุณลักษณะหรือปัจจัยใดควรนำมาใช้เป็นคุณลักษณะในการจำแนกของขั้นตอนวิธี J48 และ CART ดังสมการที่ 2.3

$$\text{Gini}(t_i) = 1 - \sum_{i=1}^{N}[p(t_i)]^2 \qquad (2.3)$$

ค่าคาดคะเนของข้อมูล (Entropy) เป็นค่าที่แยกโดยใช้ลักษณะประจำของขั้นตอนวิธี ID3 ดังสมการที่ 2.4

$$\text{Entropy}(t_i) = 1 - \sum_{i=1}^{N}[p(t_i)]\log_2 p(t_i) \qquad (2.4)$$

โดยที่ t_i คือ คุณลักษณะที่นำมาวัดค่าคาดคะเนของข้อมูล

$p(t_i)$ คือ สัดส่วนของจำนวนสมาชิกของกลุ่ม i กับจำนวนสมาชิกทั้งหมดของกลุ่มตัวอย่าง

2.3.2.1 ขั้นตอนวิธี J48 (C4.5)

J48 หรือ C4.5 เป็นขั้นตอนวิธีในการสร้างต้นไม้ตัดสินใจจากกลุ่มของข้อมูลฝึกสอนโดยใช้ความถูกต้องของแต่ละคุณลักษณะของข้อมูลเพื่อใช้เป็นการตัดสินใจจำแนกข้อมูลกลุ่มย่อยๆ โดยพิจารณาจากค่าความแตกต่างในค่าคาดคะเนของข้อมูล (Entropy) ผลลัพธ์จากการเลือกคุณลักษณะสำหรับจำแนกข้อมูลด้วยค่าข้อมูลผลกำไรปรกติมาตรฐาน (Normalized Information Gain) ที่สูงสุดนั้นคือการสร้างการตัดสินใจ (ภรันยา, 2559)

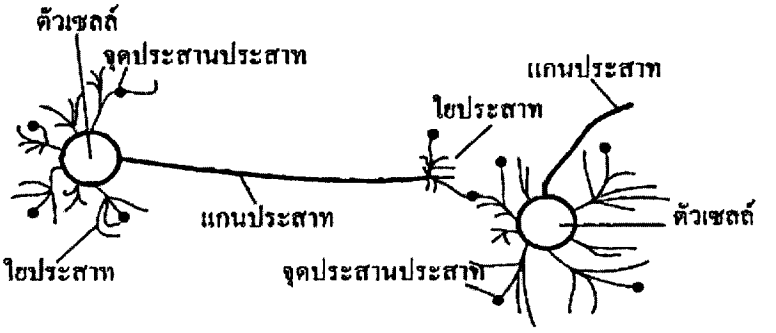
2.3.3 วิธีโครงข่ายประสาทเทียม (Artificial Neural Network)

เป็นเทคโนโลยีที่มาจากงานวิจัยด้านปัญญาประดิษฐ์ (Artificial Intelligence : AI) เพื่อใช้ในการคำนวณค่าฟังก์ชันจากกลุ่มข้อมูล วิธีโครงข่ายประสาทเทียมเป็นวิธีการที่ให้เครื่องเรียนรู้จากตัวอย่างต้นแบบ แล้วฝึกให้ระบบได้รู้จักคิดแก้ปัญหาที่กว้างขึ้นได้ ในโครงสร้างของโครงข่ายประสาทเทียมประกอบด้วยโหนด (Node) สำหรับข้อมูลเข้า (Input Data) และข้อมูลออก (Output Data) และการประมวลผลกระจายอยู่ในโครงสร้างเป็นชั้น ๆ ได้แก่ ชั้นข้อมูลเข้า (Input Layer) ชั้นข้อมูลออก (Output Layer) และชั้นซ่อน (Hidden Layer) การประมวลผลของโครงข่ายประสาทเทียมจะอาศัยการส่งการทำงานผ่านโหนดต่าง ๆ ในชั้น (Layers) (ทิพย์ธิดา, 2555)

2.3.3.1 ความรู้พื้นฐานของระบบประสาท (Neural System Knowledge)

โครงข่ายประสาทเทียมได้ถูกพัฒนาขึ้นโดยอาศัยหลักการทำงานของสมองมนุษย์ ซึ่งสมองประกอบด้วยหน่วยประมวลผลพื้นฐานที่เรียกว่า นิวรอล (เซลล์ประสาท หรือ neuron) ภายในสมองประกอบด้วยนิวรอลจำนวนมหาศาล (ประมาณ 10^{11}) ประกอบขึ้นด้วยส่วนสำคัญ 3 ส่วน คือ ใยประสาท (Dendrite) ตัวเซลล์ (Soma) และแกนประสาท (Axon) ซึ่งแบ่งออกเป็น 4 บริเวณด้วยกัน ดังรูปที่ 2.3 (วราภรณ์ และคณะ, 2558) คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.3 โครงข่ายของเซลล์ประสาท

1. บริเวณนำกระแสประสาทเข้า (Input Region) เป็นบริเวณที่จะมีการนำกระแสประสาท (Nerve Impulse) จากเซลล์ประสาทอื่นเข้ามาภายในตัวเซลล์โดยผ่านทางใยประสาทซึ่งมีลักษณะแตกเป็นกิ่งก้านคล้ายต้นไม้และมีจำนวนตั้งแต่ 1 ใยขึ้นไป

2. บริเวณรวมกระแสประสาท (Integration) เป็นบริเวณที่มีการรวมกระแสประสาทก่อนที่จะเข้าสู่บริเวณการนำกระแสประสาทรวมออกจากเซลล์

3. บริเวณการนำกระแสประสาทรวมออกจากเซลล์ (Conduction Region) เป็นบริเวณที่จะนำกระแสประสาทรวมออกจากเซลล์ โดยใช้เส้นใยประสาทเป็นทางผ่าน ซึ่งมีเพียง 1 เส้นใยเท่านั้น

4. บริเวณนำกระแสประสาทออก (Output Region) เป็นบริเวณส่วนปลายของเส้นใยประสาทที่มีการแตกแขนง ใช้ในการถ่ายทอดกระแสประสาทข้ามเซลล์ไปยังเซลล์ประสาทอื่นโดยผ่านทางใยประสาทของเซลล์ประสาทนั้น

2.3.3.2 การเรียนรู้ของโครงข่ายประสาทเทียม (Artificial Neural Network Learning)

การเรียนรู้ของโครงข่ายประสาทเทียมจะมีประสิทธิภาพเพียงใดนั้นขึ้นอยู่กับค่าถ่วงน้ำหนัก (Weight) ของโครงข่ายที่ทำการออกแบบซึ่งการฝึกหัด (Training) โครงข่ายคือการหาค่าถ่วงน้ำหนักที่เหมาะสมให้กับโครงข่ายนั้น ๆ โดยทั่วไปสามารถจำแนกวิธีการเรียนรู้ของโครงข่ายประสาทประสาทเทียมได้เป็น 2 ประเภทคือ การเรียนรู้แบบมีผู้สอนและการเรียนรู้แบบไม่มีผู้สอน

1. การเรียนรู้แบบมีการสอน (Supervised Learning) เป็นการใช้รูปแบบ (Pattern) ของชุดข้อมูล ตัวแปรข้อมูลเข้าและตัวแปรข้อมูลออกสอนโครงข่ายให้เรียนรู้ความสัมพันธ์ระหว่างตัวแปรข้อมูลเข้าและตัวแปรข้อมูลออก โดยโครงข่ายแบบจำลอง ANN จะปรับค่าน้ำหนักที่เชื่อมโยงเพื่อลดความผิดพลาดในการทำนายค่าข้อมูลออก

2. การเรียนรู้แบบไม่มีการสอน (Unsupervised Learning) จะไม่มีการตรวจสอบคำตอบว่าถูกหรือผิด ซึ่งโครงข่าย ANN จะจัดเรียงโครงสร้างด้วยตัวเองตามลักษณะของข้อมูล โดยจะสร้างค่าน้ำหนัก (Weight) ที่เชื่อมโยงตัวแปรข้อมูลเข้าว่าข้อมูลมีลักษณะที่ควรจะจัดรวมเข้ากลุ่มใด หลังจากนั้นจึงมีการปรับคุณลักษณะของกลุ่ม โดยการนำลักษณะของข้อมูลใหม่นี้มาช่วยกำหนดแนวทางการจัดด้วยในการตัดสินว่าข้อมูลผลลัพธ์ที่ได้นี้ควรจัดรวมเข้ากลุ่มใด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

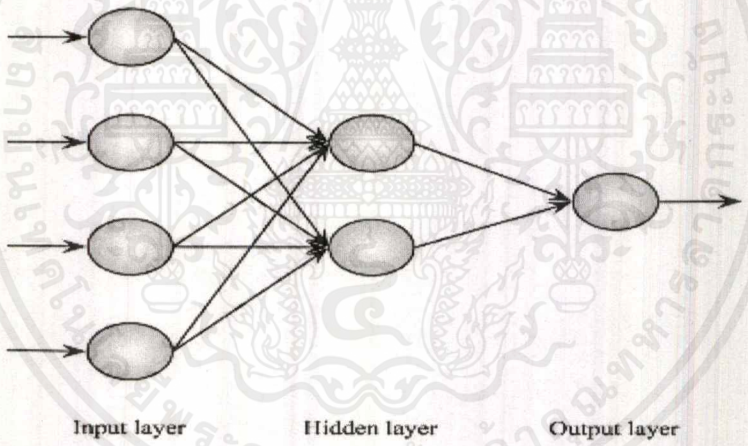
2.3.3.3 แบบจำลองโครงข่ายประสาทเทียม ((Artificial Neural Network Model)

โครงสร้างของโครงข่ายประสาทเทียมประกอบด้วย 3 องค์ประกอบ ดังนี้ (วราภรณ์ และคณะ, 2558)

1. ชั้นข้อมูลเข้า (Input Layer) เป็นชั้นแรกในโครงสร้างของแบบจำลองโครงข่ายประสาทเทียม จะประกอบด้วยนิวรอนรับข้อมูลเข้า (Input Neuron) จำนวนนิวรอนรับข้อมูลเข้าจะเท่ากับจำนวนของข้อมูลออก โดยทำหน้าที่รับข้อมูลเข้าสู่แบบจำลองโครงข่ายประสาทเทียม

2. ชั้นข้อมูลออก (Output Layer) เป็นชั้นสุดท้ายในโครงสร้างของแบบจำลองโครงข่ายประสาทเทียม จะประกอบด้วยนิวรอนแสดงผลข้อมูล (Output Neuron) จำนวนนิวรอนของข้อมูลเข้าจะเท่ากับจำนวนของข้อมูลออกของตัวแปรทำหน้าที่แสดงผลลัพธ์ (Output) ของแบบจำลอง

3. ชั้นซ่อน (Hidden Layer) จะอยู่ระหว่างชั้นข้อมูลเข้าและชั้นข้อมูลออก ประกอบด้วยนิวรอนในชั้นซ่อน (Hidden Neuron) ทำหน้าที่ในการรวมข้อมูลเข้าที่คูณด้วยน้ำหนัก (Weight) ที่แปรค่าเพื่อแปลงให้เป็นข้อมูลออก (Output) และส่งไปยังชั้นข้อมูลออกโดยไม่มีกฎตายตัวในการกำหนดจำนวนชั้นซ่อนและนิวรอนที่เหมาะสม



รูปที่ 2.4 ระบบโครงข่ายประสาทเทียม

2.3.3.4 ขั้นตอนวิธีเพอร์เซปตรอนหลายชั้น (Multilayer Perceptron)

ประกอบด้วยชั้นข้อมูลเข้า (Input Layer) ชั้นซ่อน (Hidden Layer) และชั้นข้อมูลออก (Output Layer) ในแต่ละชั้นของโครงข่ายประสาทเทียมจะประกอบด้วยหน่วยประมวลผลหรือโหนด ขั้นตอนวิธีแบบแพร่ย้อนกลับ (back propagation algorithm) เป็นขั้นตอนวิธีที่ใช้ในการฝึกสอนโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น ใช้สำหรับปรับค่าน้ำหนักเส้นเชื่อมโยงของโหนดต่างๆ ให้มีค่าที่เหมาะสม (เชาวนันท์ และคณะ, 2556)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.4 วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)

ซัพพอร์ตเวกเตอร์แมชชีนหรือ SVM เป็นขั้นตอนวิธีในการคัดแยกกลุ่มเพื่อจัดประเภทหรือจำแนกประเภทข้อมูลที่มีการนำมาใช้ในด้านการประมวลผลภาพ เป็นวิธีการจำแนกข้อมูลที่อาศัยระนาบการตัดสินใจที่เรียกว่า ระนาบเกิน (Hyperplane) มาใช้ในการจำแนกข้อมูลโดยใช้สมการเส้นตรงในการแบ่งข้อมูลออกเป็น 2 กลุ่ม แยกออกจากกัน ซัพพอร์ตเวกเตอร์แมชชีนมีรูปแบบในการเรียนรู้เป็นกระบวนการเลือกแบบจำลองที่เหมาะสมที่สุด จะทำให้ได้ค่าที่เหมาะสมที่สุดเป็นคำตอบ ดังนั้นซัพพอร์ตเวกเตอร์แมชชีนจึงเป็นที่นิยมและเริ่มนำไปใช้ในงานด้านการรู้จำรูปแบบซึ่งจะเลือกใช้ซัพพอร์ตเวกเตอร์แมชชีนแบบแบ่งกลุ่ม (รติพร, 2557)

ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการแบ่งกลุ่มข้อมูลนั้นจะใช้ระนาบเกินที่เหมาะสมที่สุด (Optimal Hyperplane) ในการแบ่งกลุ่ม ในการสร้างระนาบเกินที่ใช้ในการแบ่งกลุ่มข้อมูลสามารถสร้างได้หลายแบบ แต่จะมีระนาบเกินที่เหมาะสมที่สุดเพียงระนาบเดียวเท่านั้นที่สามารถรักษาระยะห่างมากที่สุดระหว่างข้อมูล 2 กลุ่ม ที่ใกล้กันมากที่สุดได้

กำหนดให้ (x_i,y_i), ... ,(x_n,y_n) เมื่อ $x \in R^m$, $y \in \{-1,1\}$ เป็นตัวอย่างที่ใช้สำหรับการสอน

โดยที่ n คือ จำนวนข้อมูลตัวอย่าง

m คือ จำนวนมิติของข้อมูลเข้า

x คือ ข้อมูลนำเข้า

y คือ ประเภทหรือกลุ่มของข้อมูล ซึ่งประกอบด้วย 2 กลุ่ม มีค่า + 1 หรือ – 1 (+1 = “ข้อมูลบวก”, -1 = “ข้อมูลลบ”)

สำหรับปัญหาเชิงเส้น ข้อมูลมิติขนาดสูงได้ถูกแบ่งเป็น 2 กลุ่ม โดยใช้ระนาบตัดสินใจพิจารณาชุดของกลุ่มข้อมูล X โดยที่กำหนดให้กลุ่มข้อมูล X_1 เป็นข้อมูล X_i ที่มีค่าเป็นบวก และ X_2 เป็นข้อมูล X_i ที่มีค่าเป็นลบ การสร้างระนาบตัดสินใจเพื่อแบ่งแยกกลุ่มข้อมูลสามารถคำนวณได้ดังสมการที่ 2.5

$$(w * x_1) + b > 0 \text{ ถ้า } y_i = +1 \text{ และ } (w * x_2) + b < 0 \text{ ถ้า } y_i = -1 \qquad (2.5)$$

โดยที่ w คือ เวกเตอร์น้ำหนัก

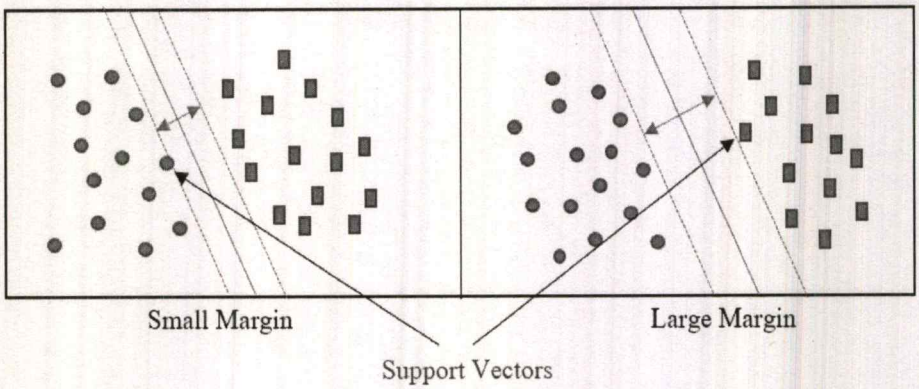
x_1 คือ เวกเตอร์ข้อมูลที่มีค่าเป็นบวก

x_2 คือ เวกเตอร์ข้อมูลที่มีค่าเป็นลบ

b คือ ค่าเอนเอียง

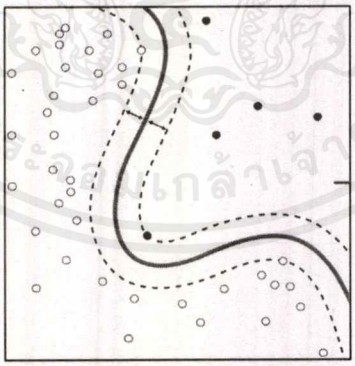
ในการหาระนาบเกินที่เหมาะสมที่สุด จะทำการหาตำแหน่งของซัพพอร์ตเวกเตอร์ (Support Vector) เพื่อใช้เป็นตัวแทนของกลุ่มข้อมูลทั้งชุด ในการพิจารณาเกณฑ์แบ่งกลุ่มโดยอาศัยหลักการคือ จะใช้ระนาบเกินที่เป็นระยะห่างมากที่สุดระหว่างข้อมูล 2 กลุ่ม ที่อยู่ใกล้กันมากที่สุดเพียงระนาบเดียวเท่านั้น ในทางทฤษฎีจะต้องไม่มีข้อมูลเกินเข้ามาในระหว่างขอบระนาบทั้งสอง จากนั้นจึงหาระนาบที่รักษาระยะห่างจากขอบมากที่สุด (Maximum Margin) และถือว่าระนาบดังกล่าวคือระนาบสำหรับการแบ่งกลุ่มที่เหมาะที่สุดแสดงดังรูปที่ 2.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.5 การปรับความชันของเส้นแบ่งแล้วทำให้ได้ระยะขอบที่มากที่สุด

จากที่กล่าวข้างต้นเป็นการแบ่งกลุ่มข้อมูลด้วยระนาบการตัดสินใจแบบเชิงเส้นเท่านั้น โดย SVM มีฟังก์ชันเคอร์เนล (Kernel Function) แบบอื่นให้ผู้ใช้สามารถประยุกต์ใช้ในการแก้ปัญหาได้หลายวิธี ดังนั้นเพื่อให้ขั้นตอนวิธีดังกล่าวสามารถแบ่งแยกกลุ่มข้อมูลที่มีลักษณะไม่เป็นเชิงเส้น (Nonlinear Dataset) จะต้องแปลงกลุ่มข้อมูลตัวอย่างไปสู่มิติที่สูงขึ้น (Higher Dimensional Space) ซึ่งเรียกว่า พื้นที่คุณลักษณะ (Feature Space) โดยการแปลงดังกล่าวจะกระทำผ่านฟังก์ชันที่ไม่เป็นเชิงเส้นและสร้างฟังก์ชันวัดระยะห่างที่เรียกว่า ฟังก์ชันเคอร์เนล (Kernel Function) บนพื้นที่คุณลักษณะ ซึ่งเหมาะสมสำหรับข้อมูลที่มีมิติข้อมูลสูง โดยมีวัตถุประสงค์ที่จะพยายามจะทำการลดความผิดพลาดในการทำนายกลุ่มข้อมูล (Minimize Error) พร้อมกับเพิ่มระยะแยกแยะโดยพยายามสร้างเส้นแบ่งกลางระหว่างกลุ่มใหม่ให้มีระยะห่างระหว่างขอบเขตทั้งสองกลุ่มมากที่สุด (Maximized Margin) ใช้สำหรับข้อมูลที่มีลักษณะมิติของข้อมูลที่สูงมาก



รูปที่ 2.6 ตัวอย่างการจำแนกข้อมูลของซัพพอร์ตเวกเตอร์แมชชีนโพลิโนเมียลเคอร์เนล

2.3.4.1 ขั้นตอนวิธี SMO (Sequential Minimal Optimization)

เป็นขั้นตอนวิธีสำหรับซัพพอร์ตเวกเตอร์แมชชีนในกรณีที่ปัญหาพหุนามกำลังสองขนาดใหญ่ถูกแตกออกเป็นชุดปัญหาเล็กที่สุดเท่าที่จะเป็นไปได้ซึ่งสามารถถูกแก้ไขได้ SMO สามารถจัดการกับชุดข้อมูลที่ใช้สอนที่มีขนาดใหญ่ภายใต้เวลาที่เหมาะสม (คณาวุฒิ, 2551)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4 การเปรียบเทียบประสิทธิภาพของวิธีการจำแนก

เมทริกซ์ความสับสน (Confusion Matrix) คือ ตารางสรุปจำนวนข้อมูลที่ตัวแบบมีการจำแนกได้ถูกต้องและไม่ถูกต้อง ดังแสดงในตารางที่ 2.6 จากตารางที่ 2.6 สามารถนำข้อมูลในตารางมาใช้ในการคำนวณการวัดประสิทธิภาพของตัวแบบการทำนาย

ตารางที่ 2.6 เมทริกซ์ความสับสน

ค่าจริง	ค่าทำนาย	
	คำตอบเป็นบวก	คำตอบที่เป็นลบ
คำตอบเป็นบวก	TP (True Positive)	FN (False Negative)
คำตอบที่เป็นลบ	FP (False Positive)	TN (True Negative)

โดยที่ บวกจริง (True Positive : TP) คือ ค่าความถูกต้องในการจำแนกข้อมูล ซึ่งมีค่าที่แท้จริงอยู่ในกลุ่มบวก และผลการทำนายว่าอยู่ในกลุ่มบวก

ลบจริง (True Negative : TN) คือ ค่าความถูกต้องในการจำแนกข้อมูล ซึ่งมีค่าที่แท้จริงอยู่ในกลุ่มลบ และผลการทำนายว่าอยู่ในกลุ่มลบ

บวกเท็จ (False Positive : FP) คือ ค่าความผิดพลาดในการจำแนกข้อมูล ซึ่งมีค่าที่แท้จริงอยู่ในกลุ่มลบ แต่ผลการทำนายว่าอยู่ในกลุ่มบวก

ลบเท็จ (False Negative : FN) คือ ค่าความผิดพลาดในการจำแนกข้อมูล ซึ่งมีค่าที่แท้จริงอยู่ในกลุ่มบวก แต่ผลการทำนายว่าอยู่ในกลุ่มลบ

2.4.1 ค่าความถูกต้อง (Accuracy) คือ การแสดงการวัดที่ได้มีความถูกต้องในรูปอัตราส่วน (สุรวัชร และสายชล, 2560)

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100$$

2.4.2 ค่าความไว (Sensitivity หรือ True Positive Rate : TPR) คือสัดส่วนของผลบวกที่เป็นจริงสำหรับภาวะนั้น ๆ (กีระชาติ, 2559)

$$TPR = \frac{TP}{TP+FN}$$

2.4.3 ค่าความจำเพาะ (Specificity หรือ True Negative Rate : TNR) คือสัดส่วนของผลลบที่เป็นจริงสำหรับภาวะนั้น ๆ (กีระชาติ, 2559)

$$TNR = \frac{TN}{TN+FP}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4.4 ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) คือ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยใช้หลักการเดียวกันกับการหาค่าความแปรปรวนในทางสถิติ การวัดค่าคลาดเคลื่อนด้วยวิธีนี้จะได้ค่าคลาดเคลื่อนที่สูง เนื่องจากเป็นการนำค่าคลาดเคลื่อน ณ เวลาใด ๆ มายกกำลังสองก่อนที่จะหาผลรวม แล้วจึงนำมาหาค่าเฉลี่ยอีกครั้งหนึ่ง นั่นคือ ค่า MSE ยิ่งน้อย หมายถึง การทำนายยิ่งแม่นยำ มีสูตรในการคำนวณดังนี้ (สุรวัชร และสายชล, 2560)

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}$$

โดยที่ y_i คือ ค่าจริง

$\hat{y}_i$ คือ ค่าทำนาย

2.5 งานวิจัยที่เกี่ยวข้อง

วีระยุทธ และคณะ (2557) ได้ทำการศึกษางานวิจัยเกี่ยวกับการรักษาซ้ำของผู้ป่วยจิตเภท เรื่องการพัฒนาแบบจำลองเพื่อการพยากรณ์การรักษาซ้ำของผู้ป่วยโรคจิตเภทโดยเทคนิคเหมืองข้อมูล งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองในการพยากรณ์ระยะเวลาการรักษาซ้ำของผู้ป่วยโรคจิตเภทโดยเทคนิคเหมืองข้อมูลจากฐานข้อมูลโรงพยาบาลพระศรีมหาโพธิ์ จังหวัดอุบลราชธานี จำนวน 2831 ระเบียน โดยการแบ่งข้อมูลออกเป็น 2 กลุ่ม คือ กลุ่ม 0 เป็นกลุ่มของผู้ป่วยมารับการรักษาซ้ำ 1 ถึง 28 วัน ส่วนกลุ่ม 1 เป็นกลุ่มของผู้ป่วยมารับการรักษาซ้ำตั้งแต่ 29 ถึง 90 วัน แบบจำลองนี้สามารถช่วยในการวางแผนการรักษาของแพทย์ ผลการศึกษาพบว่าข้อมูลมีค่าผิดปกติและมีความไม่สมดุลของกลุ่มในข้อมูล โดยมีจำนวนกลุ่มหนึ่งมากกว่าอีกกลุ่มหนึ่งเป็นจำนวนมาก ผู้วิจัยจึงได้ทำการแก้ปัญหาโดยคัดกรองเอาค่าผิดปกติออกด้วยเทคนิค Support Vector Machine และปรับความสมดุลของข้อมูลด้วยวิธี Synthetic Minority Over-sampling Technique (SMOTE) แล้วพัฒนาแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจ (C4.5) นาอีฟเบส์ (Naïve Bayes) และ PART Decision List นอกจากนั้นผู้วิจัยยังได้ใช้ 10-fold cross validation ในการแบ่งข้อมูลออกเป็นชุดข้อมูลสอนและชุดข้อมูลทดสอบและได้ใช้ค่าความถูกต้อง ค่าความไว และค่าความจำเพาะในการแสดงประสิทธิภาพในการพยากรณ์ของแบบจำลองอีกด้วย ผลการทดลองประสิทธิภาพในการพยากรณ์ของแบบจำลองพบว่า เทคนิค PART Decision List สามารถพยากรณ์ได้ดีกว่าต้นไม้ตัดสินใจ (C4.5) และนาอีฟเบส์ (Naïve Bayes) ซึ่งมีค่าความถูกต้อง (Accuracy) ร้อยละ 92.98 ค่าความไว (Sensitivity) ร้อยละ 92.95 และค่าความจำเพาะ (Specificity) ร้อยละ 93.02 ในการแก้ปัญหาจากการทดลอง ผู้วิจัยพบว่าการกรองแล้วใช้วิธีการทำ SMOTE สามารถเพิ่มประสิทธิภาพให้แบบจำลองเพิ่มขึ้น โดยค่าความถูกต้องเพิ่มขึ้นเฉลี่ยร้อยละ 46.36 ค่าความไวเพิ่มขึ้นเฉลี่ยร้อยละ 20.05 และค่าความจำเพาะเพิ่มขึ้นร้อยละ 32.69 ซึ่งมากกว่าการกรองแล้วใช้วิธีการสุ่มลด (Under Sampling) นอกจากนี้ PART Decision List สามารถเพิ่มประสิทธิภาพให้แบบจำลองเพิ่มขึ้น โดยค่าความถูกต้องเพิ่มขึ้นเฉลี่ยร้อยละ 27.39 ค่าความไวเพิ่มขึ้นเฉลี่ยร้อยละ 10.45 และค่าความจำเพาะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพิ่มขึ้นร้อยละ 63.81 ทำให้เทคนิค PART Decision List มีประสิทธิภาพในการทำนายได้สูงที่สุด โดยมีค่าความถูกต้องร้อยละ 92.98 ค่าความไวร้อยละ 93.17 และค่าความจำเพาะร้อยละ 92.79

เชาวนันท์ และคณะ (2556) ศึกษาเรื่องแบบจำลองการทำนายผลการรักษาผู้ป่วยมะเร็งปากมดลูกด้วยโครงข่ายประสาทเทียม โดยมีวัตถุประสงค์เพื่อสร้างแบบจำลองในการทำนายผลการรักษาผู้ป่วยมะเร็งปากมดลูกที่เข้ารับการรักษาด้วยวิธีการฉายรังสี โดยการประยุกต์ใช้โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ และได้ทำการวิเคราะห์ความสัมพันธ์ของปัจจัยด้วยวิธีแบบลำดับขั้น ผลการศึกษาพบว่าปัจจัยที่มีความสัมพันธ์ต่อผลการรักษาประกอบด้วย ระยะของโรค อายุ ขนาดของก้อนเนื้อ ประเภทของเซลล์มะเร็ง และน้ำหนักตัว จากการศึกษางานวิจัยที่เกี่ยวข้องพบว่าระดับฮีโมโกลบินมีผลตอบสนองต่อการรักษา ดังนั้นจึงได้เพิ่มระดับฮีโมโกลบินรวมกับปัจจัยที่ได้จากการวิเคราะห์รวมเป็น 6 ปัจจัย ในการสร้างแบบจำลองด้วยแบบจำลองโครงข่ายประสาทเทียม ผลการศึกษาพบว่าแบบจำลองที่นำเสนอถึงแม้จะมีประสิทธิภาพสูง แต่ค่าความจำเพาะมีค่าเป็นร้อยละ 0 ซึ่งจากการวิเคราะห์พบว่าปัญหาเกิดจากความไม่สมดุลของข้อมูลซึ่งมีข้อมูลกลุ่มหนึ่งมากกว่าอีกกลุ่มเป็นจำนวนมาก ดังนั้นผู้วิจัยจึงได้ทำการแก้ปัญหาปรับความไม่สมดุลของข้อมูลด้วยวิธี Cost-Sensitive Learning : CSL และวิธี Synthetic Minority Over-sampling Technique : SMOTE และทำการสร้างแบบจำลองข้อมูลโครงข่ายประสาทเทียมและแบบจำลองด้วยการถดถอยลอจิสติก เมื่อเปรียบเทียบประสิทธิภาพการทำนายพบว่าโครงข่ายประสาทเทียมที่มีการแก้ปัญหาความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเพิ่มชุดข้อมูลตัวอย่าง (SMOTE) มีประสิทธิภาพการทำนายด้วยค่าความถูกต้องร้อยละ 81.71 ค่าความไวร้อยละ 94.47 และค่าความจำเพาะร้อยละ 55.47 สูงกว่าวิธีการถดถอยลอจิสติกที่มีการแก้ปัญหาความไม่สมดุลของข้อมูลด้วยวิธี CSL ซึ่งมีค่าความถูกต้องร้อยละ 81.00 ค่าความไวร้อยละ 84.52 และค่าความจำเพาะร้อยละ 30.66

Rahman and Davis (2013) ได้ทำการศึกษางานวิจัยที่เกี่ยวข้องกลุ่มไม่สมดุล เรื่องการจัดการปัญหาเกี่ยวกับกลุ่มไม่สมดุลในข้อมูลทางการแพทย์ โดยใช้เทคนิค FURIA14 และ Decision Tree ปรับความสมดุลของข้อมูลด้วยวิธี SMOTE พบว่าวิธีต้นไม้ตัดสินใจ (Decision Tree) มีประสิทธิภาพในการทำนายที่ดีกว่าวิธีอื่น โดยมีค่าความถูกต้องร้อยละ 85.78 ค่าความไวร้อยละ 84.21 และค่าความจำเพาะร้อยละ 87.34 หลังจากปรับความไม่สมดุลด้วยวิธี SMOTE พบว่าค่าความถูกต้องเพิ่มขึ้นร้อยละ 6.19 ค่าความไวเพิ่มขึ้นร้อยละ 64.21 และค่าความจำเพาะลดลงร้อยละ 2.42

เบญจภรณ์ และคณะ (2557) ศึกษาเรื่องวิธีการที่เหมาะสมสำหรับการแบ่งกลุ่มข้อมูลที่ไม่สมดุลสูง โดยมีวัตถุประสงค์เพื่อให้เกิดอัตราความผิดพลาดในการเรียนรู้ข้อมูลน้อยที่สุดโดยในงานวิจัยนี้ได้ทำการปรับปรุงฟังก์ชันความผิดพลาดที่ใช้วัดอัตราความผิดพลาดสำหรับชุดข้อมูลที่ไม่สมดุลได้อย่างเหมาะสม ซึ่งฟังก์ชันความผิดพลาดส่วนใหญ่จะใช้ค่าน้ำหนักที่เท่ากันทุกกลุ่ม จากที่ทราบกันโดยทั่วไป ข้อมูลที่ไม่สมดุลหมายถึงชุดข้อมูลที่มีจำนวนสมาชิกของกลุ่มส่วนมากและกลุ่มส่วนน้อยจำนวนไม่เท่ากัน ดังนั้นหากใช้ค่าน้ำหนักเท่ากันทุกกลุ่มจะทำให้การจัดกลุ่มไม่เหมาะสม และปัญหาของการเรียนรู้รูปแบบข้อมูลของชุดข้อมูลไม่สมดุลส่วนใหญ่ พบว่าข้อมูลในกลุ่มส่วนน้อยถูกครอบงำด้วยข้อมูลของกลุ่มส่วนมาก จึงเป็นผลให้เกิดความเอนเอียงในการจำแนกข้อมูล ทำให้ข้อมูลในกลุ่มส่วนน้อยเกิดความผิดพลาดในการจำแนกมากกว่ากลุ่มส่วนมาก จากที่กล่าวมางานวิจัยนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จึงได้ทำการหาพารามิเตอร์ที่เหมาะสมสำหรับปรับปรุงฟังก์ชันความผิดพลาดเฉลี่ยกำลังสอง โดยวิธีการที่นำเสนอได้นำอัตราการซ้อนทับกันของข้อมูลและอัตราความไม่สมดุลของข้อมูลมาใช้ร่วมในการปรับปรุงด้วย สำหรับขั้นตอนวิธีในการเรียนรู้ข้อมูลได้ใช้โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับและฟังก์ชันความผิดพลาดที่ทำการปรับปรุง

McCarthy et al. (2005) ได้ทำการศึกษาเกี่ยวกับความสามารถในการแก้จำแนก เรื่องการทดสอบความสามารถในการจำแนกข้อมูลที่ค้นพบได้ยากด้วยวิธีการเรียนรู้แบบมีค่าใช้จ่าย (Cost-Sensitive) และเทคนิคการสุ่ม 2 เทคนิค ได้แก่ วิธีการสุ่มเกิน (Over-Sampling) และวิธีการสุ่มลด (Under-Sampling) ใช้ข้อมูลที่มีอัตราส่วนของจำนวนข้อมูลที่ค้นพบได้ยากต่อจำนวนข้อมูลปกติน้อยที่สุดอยู่ที่ร้อยละ 4 และมากที่สุดอยู่ที่ร้อยละ 50 เพื่อทำการเปรียบเทียบประสิทธิภาพ ผลจากการทดลองสามารถสรุปได้ว่าผลที่ได้ไม่ชัดเจนพอที่จะสรุปว่าวิธีใดมีประสิทธิภาพดีที่สุด แต่พบว่าวิธีการเรียนรู้แบบมีค่าใช้จ่ายมีประสิทธิภาพดีกว่าวิธีการสุ่มเกินหรือวิธีการสุ่มลด และยังให้คำแนะนำในการวิจัยว่าถ้าหากทำการทดสอบประสิทธิภาพของวิธีการสุ่มเกินและการสุ่มลดไม่ควรนำไปเปรียบเทียบวิธีการเรียนรู้แบบมีค่าใช้จ่าย เนื่องจากประสิทธิภาพในการจำแนกข้อมูลใกล้เคียงกัน

He and Ghodsi (2010) ได้ทำการศึกษาเกี่ยวกับการจำแนกข้อมูลที่ค้นพบได้ยากจากข้อมูลที่มีขนาดต่างกันมาก (Imbalanced Datasets) โดยใช้วิธีที่ได้มีการพัฒนามาจากซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) ซึ่งได้ทำการทดสอบด้วยชุดข้อมูลทั้งหมด 7 ชุด และใช้มาตราวัด 3 ชนิด ได้แก่ F-measure, G-mean และ AUR-ROC โดยทำการเปรียบเทียบวิธีที่ทำการพัฒนาเปรียบเทียบกับวิธีการสุ่มลด (Under-Sampling) ผลการทดสอบปรากฏว่าวิธีที่พัฒนามาจากซัพพอร์ตเวกเตอร์แมชชีนให้ผลลัพธ์ที่ดีกว่าวิธีซัพพอร์ตเวกเตอร์แมชชีนปกติและโดยรวมแล้วมีประสิทธิภาพดีกว่าวิธีการสุ่มลด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

วิธีการดำเนินงานวิจัย

3.1 การเก็บรวบรวมข้อมูล

ทำการค้นหาและการศึกษาข้อมูลที่ไม่สมดุล จากเว็บไซต์ Vincentarelbundock โดยได้ข้อมูลมา 2 ชุด และจากเว็บไซต์ UCI โดยได้ข้อมูลมา 1 ชุด คือ

1. การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทั้งหมด 890 ค่า พบค่าในกลุ่มส่วนมาก (Majority) 850 ค่า คิดเป็นร้อยละ 95.51 และค่าในกลุ่มส่วนน้อย (Minority) 40 ค่า คิดเป็นร้อยละ 4.49 (Hogan, 2013)

2. ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทั้งหมด 400 ค่า พบค่าในกลุ่มส่วนมาก (Majority) จำนวน 360 ค่า คิดเป็นร้อยละ 90 และค่าในกลุ่มส่วนน้อย (Minority) จำนวน 40 ค่า คิดเป็นร้อยละ 10 (James et al., 2013)

3. คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทั้งหมด 999 ค่า พบค่าในกลุ่มส่วนมาก (Majority) 804 ค่า คิดเป็นร้อยละ 80.48 และค่าในกลุ่มส่วนน้อย (Minority) 195 ค่า คิดเป็นร้อยละ 19.52 (Cortez, 2009)

3.2 ตัวแปรและนิยาม

3.2.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME)

จำนวนข้อมูลทั้งหมด 890 ค่า พบค่าในกลุ่มส่วนมากได้ยินเสียงไม่ปกติ (Majority) จำนวน 850 ค่า คิดเป็นร้อยละ 95.51 และค่าในกลุ่มส่วนน้อยได้ยินเสียงปกติ (Minority) จำนวน 40 ค่า คิดเป็นร้อยละ 4.49 (Hogan, 2013)

ตัวแปรอิสระ

X_1 แทน อายุ (หน่วย : เดือน)

X_2 แทน อัตราความเสี่ยงการเป็นโรคหูชั้นกลางอักเสบ (low คือความเสี่ยงน้อย, high คือความเสี่ยงมาก และ N/A คือไม่มีความเสี่ยง)

X_3 แทน ความดังที่ได้รับ (หน่วย : เดซิเบล)

X_4 แทน สัญญาณในการกระตุ้น (coherent คือสัญญาณต่อเนื่อง และ incoherent คือสัญญาณไม่ต่อเนื่อง)

X_5 แทน จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง (หน่วย : ครั้ง)

ตัวแปรตาม

Y แทน การรับรู้ทางหู (No คือได้ยินเสียงไม่ปกติ และ Yes คือได้ยินเสียงปกติ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance)

จำนวนข้อมูลทั้งหมด 400 ค่า พบค่าในกลุ่มส่วนมากไม่ชำระเงิน (Majority) จำนวน 360 ค่า คิดเป็นร้อยละ 90 และค่าในกลุ่มส่วนน้อยชำระเงิน (Minority) จำนวน 40 ค่า คิดเป็นร้อยละ 10 (James et al., 2013)

ตัวแปรอิสระ

X_1 แทน รายได้ต่อเดือน (หน่วย : สกุลเงินดอลลาร์สหรัฐอเมริกา (USD))

X_2 แทน วงเงินในบัตรเครดิต (หน่วย : สกุลเงินดอลลาร์สหรัฐอเมริกา (USD))

X_3 แทน จำนวนครั้งในการใช้บัตรเครดิตต่อปี (หน่วย : ครั้ง)

X_4 แทน จำนวนบัตรเครดิตที่มี (หน่วย : บัตร)

X_5 แทน อายุ (หน่วย : ปี)

X_6 แทน จำนวนปีในการศึกษา (หน่วย : ปี)

X_7 แทน เพศ (Male คือเพศชาย และ Female คือเพศหญิง)

X_8 แทน การศึกษา (Yes คือกำลังศึกษา และ No คือสำเร็จการศึกษา)

X_9 แทน สถานภาพสมรส (Yes คือแต่งงานแล้ว และ No คือโสด)

X_{10} แทน ยอดเงินคงเหลือในบัตรเครดิต (หน่วย : สกุลเงินดอลลาร์สหรัฐอเมริกา (USD))

ตัวแปรตาม

Y แทน การชำระเงินในเดือนถัดไป (No คือไม่ชำระเงิน และ Yes คือชำระเงิน)

3.2.3 คุณภาพไวน์แดง (Red Wine Quality)

จำนวนชุดข้อมูลทั้งหมด 999 ค่า พบค่าในกลุ่มส่วนมากคุณภาพดี (Majority) 804 ค่า คิดเป็นร้อยละ 80.48 และค่าในกลุ่มส่วนน้อยคุณภาพไม่ดี (Minority) 195 ค่า คิดเป็นร้อยละ 19.52 (Cortez, 2009)

ตัวแปรอิสระ

X_1 แทน ค่าความเป็นกรด (หน่วย : pH)

X_2 แทน ค่าความเป็นกรดระเหย (หน่วย : pH)

X_3 แทน ค่ากรดซิตริก (หน่วย : pH)

X_4 แทน ค่าน้ำตาลรีดิวซ์ (หน่วย : กรัม)

X_5 แทน ค่าคลอไรด์ (หน่วย : กรัม)

X_6 แทน ค่าซัลเฟอร์ไดออกไซด์ (หน่วย : กรัม)

X_7 แทน ค่าซัลเฟอร์ไดออกไซด์ทั้งหมด (หน่วย : กรัม)

X_8 แทน ค่าความหนาแน่น (หน่วย : กิโลกรัมต่อลูกบาศก์เมตร)

X_9 แทน ค่าความเป็นกรดหรือด่าง (หน่วย : pH)

X_{10} แทน ค่าเกลือของกรดซัลฟิวริก (หน่วย : pH)

X_{11} แทน ค่าแอลกอฮอล์ (หน่วย : กรัม)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X_{12} แทน คะแนนคุณภาพ (หน่วย : คะแนน)

ตัวแปรตาม

Y แทน คุณภาพไวน์แดง (No คือคุณภาพไม่ดี และ Yes คือคุณภาพดี)

3.3 เครื่องมือที่ใช้ในงานวิจัย

โปรแกรม WEKA เวอร์ชั่น 3.9.2 และ Microsoft Excel เวอร์ชั่น 2016

3.4 ขั้นตอนในการดำเนินงานวิจัย

3.4.1 การปรับข้อมูลให้มีความสมดุล

สำหรับขั้นตอนนี้จะทำการปรับข้อมูลให้มีความสมดุลด้วย 4 วิธี

3.4.1.1 วิธีการสุ่มเกิน (Over Sampling)

ทำการสุ่มข้อมูลในกลุ่มส่วนน้อยเพื่อสร้างข้อมูลใหม่ของกลุ่มส่วนน้อยให้มีจำนวนเพิ่มมากขึ้นให้ใกล้เคียงหรือเท่ากับจำนวนในกลุ่มส่วนมาก โดยจะใช้วิธีการสุ่มกลุ่มตัวอย่างอย่างง่าย (Simple random sampling) (กีระชาติ, 2559)

3.4.1.2 วิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-Sampling Technique)

จะทำการสุ่มสร้างข้อมูลจากกลุ่มส่วนน้อยตามจำนวนที่กำหนด โดยการวัดระยะห่างจากจุดข้อมูลตัวอย่างไปยังจุดข้อมูลใกล้เคียง แล้วสุ่มสร้างข้อมูลสังเคราะห์ขึ้นให้ใกล้เคียงกับกลุ่มส่วนมาก (ภรันยา, 2559)

3.4.1.3 วิธีการสุ่มลด (Under Sampling)

การปรับข้อมูลให้มีความสมดุลด้วยวิธีการสุ่มลดจำนวนข้อมูลจากกลุ่มส่วนมากลงเพื่อทำให้จำนวนข้อมูลระหว่างกลุ่มส่วนมากและกลุ่มส่วนน้อยมีจำนวนใกล้เคียงกันมากขึ้น (กีระชาติ, 2559)

3.4.1.4 วิธีการผสมผสาน (Hybrid Methods)

เป็นวิธีการที่นำวิธีสุ่มเกินและวิธีสุ่มลดมาทำงานร่วมกันโดยการใช้วิธีนี้จะเป็นการสุ่มลดจำนวนข้อมูลจากกลุ่มส่วนมาก และทำการสุ่มเพิ่มข้อมูลในกลุ่มส่วนน้อย ให้จำนวนข้อมูลจากทั้งสองกลุ่มมีจำนวนใกล้เคียงกันหรือเท่ากัน (กีระชาติ, 2559)

3.4.2 วิธีการแบ่งข้อมูล

ทำการแบ่งชุดข้อมูลโดยโปรแกรม WEKA เวอร์ชั่น 3.9.2 ในอัตราส่วน 70:20:10 ส่วนที่ 1 ข้อมูลเรียนรู้ (Training data) นำไปสร้างตัวแบบ (Model) ร้อยละ 70 ข้อมูลส่วนที่ 2 ข้อมูลตรวจสอบความถูกต้อง (Validation data) นำไปประเมินความผิดพลาดของตัวแบบ ร้อยละ 30 และข้อมูลส่วนที่ 3 ข้อมูลทดสอบ (Testing data) นำไปทดสอบตัวแบบร้อยละ 10 (พนิดา และคณะ, 2560)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. เปิด WEKA Explorer Open File (นามสกุล .csv) เลือกไฟล์ชุดข้อมูล

2. กดปุ่ม Choose → filters → unsupervised → instance → Resample

3. กดช่อง Resample - S 1 -Z 100.0 → ในช่อง invertSelection เป็น False → ในช่อง noReplacement เป็น True → กำหนดค่า randomSeed เท่ากับ 10 → ในช่อง sampleSizePercent กำหนดให้เป็น 70 คือ ร้อยละ 70

4. กด Apply จะแสดงค่าข้อมูล ร้อยละ 70 จำนวน 1155 ค่า → กด Save → กด Undo

5. กดช่อง Resample - S 10 → ในช่อง invertSelection เป็น True → ในช่อง noReplacement เป็น True → กำหนดค่า randomSeed เท่ากับ 10 → ในช่อง sampleSizePercent กำหนดให้เป็น 70 เพื่อต้องการแบ่งข้อมูล ร้อยละ 30 จากร้อยละ70

6. กด Apply จะได้ข้อมูล ร้อยละ 30 จำนวน 495 ค่า

7. กดช่อง Resample - S 10 → ในช่อง invertSelection เป็น False → ในช่อง noReplacement เป็น True → กำหนดค่า randomSeed เท่ากับ 10 → ในช่อง sampleSizePercent กำหนดให้เป็น 66.67 เพื่อต้องการแบ่งข้อมูล ร้อยละ 20 จากร้อยละ 30

8. กด Apply จะแสดงค่าข้อมูล ร้อยละ 20 จำนวน 330 ค่า → กด Save → กด Undo

9. กดช่อง Resample - S 10 → ในช่อง invertSelection เป็น True → ในช่อง noReplacement เป็น True → กำหนดค่า randomSeed เท่ากับ 10 → ในช่อง sampleSizePercent กำหนดให้เป็น 66.67 เพื่อต้องการแบ่งข้อมูล ร้อยละ 10 จากร้อยละ 30

10. กด Apply จะแสดงค่าข้อมูล ร้อยละ 10 จำนวน 165 ค่า → กด Save → กด Undo

3.4.3 การศึกษาขั้นตอนวิธี (Algorithm)

3.4.3.1 วิธีเพื่อนบ้านใกล้สุด k ตัว (k-Nearest Neighbor)

เป็นวิธีการที่ได้รับความนิยมในการใช้งานอย่างมาก สาเหตุเนื่องจากเป็นวิธีการที่ง่ายและมีประสิทธิภาพซึ่งสามารถนำไปประยุกต์ใช้กับงานได้อย่างหลากหลาย เช่น งานทางด้านการจำแนก (Classification) โดยในงานวิจัยใช้ขั้นตอนวิธี IBk (สุรวัชร และสายชล, 2560)

3.4.3.2 วิธีต้นไม้ตัดสินใจ (Decision Tree)

เป็นการนำข้อมูลมาสร้างแบบจำลอง มีลักษณะเป็นโครงสร้างเหมือนต้นไม้เป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning) การสร้างแบบจำลองต้นไม้ตัดสินใจจะทำการคัดเลือกคุณลักษณะที่มีความสัมพันธ์กับกลุ่มมากที่สุดขึ้นมาเป็นโหนดบนสุดของต้นไม้ (Root Node) หลังจากนั้นก็จะหาคุณลักษณะถัดไปเรื่อยๆ ในการหาความสัมพันธ์ของคุณลักษณะนี้จะใช้ขั้นตอนวิธี J48 (C4.5) (วีระยุทธ และคณะ, 2557)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.3.3 วิธีโครงข่ายประสาทเทียม (Artificial Neural Network)

ใช้ขั้นตอนวิธีชนิดเพอร์เซปตรอนหลายชั้น (Multilayer Perceptron) โดยกำหนดค่าอัตราการเรียนรู้ (Learning Rate) เป็น 0.1 ค่าโมเมนตัม (Momentum) เป็น 0.9 จำนวนรอบการสอน (Training Time) 20,000 รอบ การวิจัยครั้งนี้ใช้ขั้นตอนวิธีของวิธีโครงข่ายประสาทเทียมชนิดเพอร์เซปตรอนหลายชั้นที่มีชั้นซ่อน (Hidden Layer) 1 ชั้น แม้ว่าโครงสร้างโครงข่ายประสาทเทียมที่ซับซ้อนสามารถมีชั้นซ่อนมากกว่า 1 ชั้น แต่ในทางปฏิบัติการกำหนดชั้นซ่อน 1 ชั้น ก็เพียงพอต่อการวิเคราะห์ข้อมูล (สายชล, 2561)

3.4.3.4 วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)

เป็นเทคนิคที่ใช้ในการแก้ปัญหาทางด้านการรู้จำรูปแบบข้อมูลโดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้โดยเน้นไปยังเส้นแบ่งแยกแยะกลุ่มข้อมูลได้ดีที่สุด (Optimal Separating Hyperplane) สำหรับรากฐานเดิมของซัพพอร์ตเวกเตอร์แมชชีนถูกนำมาใช้กับข้อมูลที่เป็นเชิงเส้น แต่ในความเป็นจริงแล้วข้อมูลที่นำมาใช้ในระบบการสอนให้ระบบเรียนรู้ส่วนใหญ่มักเป็นข้อมูลแบบไม่เป็นเชิงเส้น ซึ่งสามารถแก้ปัญหาดังกล่าวด้วยเคอร์เนลที่เป็นการเปลี่ยนแปลงมิติของข้อมูลให้สูงขึ้น เพื่อช่วยในการเรียงตัวของข้อมูลใหม่ โดยงานวิจัยนี้ได้ใช้ขั้นตอนวิธี SMO ชนิดโพลิโนเมียลเคอร์เนล (Polynomial Kernel) (สุรเดช และคณะ, 2554)

3.4.4 การเปรียบเทียบประสิทธิภาพของวิธีการจำแนก

นำผลการวิเคราะห์ของแต่ละวิธีทั้ง 4 วิธี มาเปรียบเทียบประสิทธิภาพ

3.4.4.1 เมทริกซ์ความสับสน (Confusion Matrix) เป็นรูปแบบตารางที่เฉพาะเจาะจงที่นำผลลัพธ์จากการทำนายมาใส่ในรูปตารางเมทริกซ์ ซึ่งจะช่วยให้ง่ายต่อการมองเห็นค่าทำนายของขั้นตอนวิธี

3.4.4.2 ค่าความถูกต้อง (Accuracy) คือการแสดงการวัดที่ได้มีความถูกต้องในรูปอัตราส่วน (สุรวัชร และสายชล, 2560)

3.4.4.3 ค่าความไว (Sensitivity หรือ True Positive Rate : TPR) คือสัดส่วนของผลบวกที่เป็นจริงสำหรับภาวะนั้น ๆ (กีระชาติ, 2559)

3.4.4.4 ค่าความจำเพาะ (Specificity หรือ True Negative Rate : TNR) คือสัดส่วนของผลลบที่เป็นจริงสำหรับภาวะนั้น ๆ (กีระชาติ, 2559)

3.4.4.3 ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) เป็นมาตรวัดการประเมินค่าได้ดี เนื่องจากค่าคลาดเคลื่อนกำลังสองเฉลี่ยประกอบด้วยทั้งความเอนเอียงและความแปรปรวน (พนิดา และคณะ, 2560)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการวิเคราะห์ข้อมูล

งานวิจัยครั้งนี้ผู้วิจัยใช้การจำแนกโดยใช้วิธีการทำเหมืองข้อมูล โดยนำชุดข้อมูลที่ค้นคว้ามาจำนวน 3 ชุด มาทำการวิเคราะห์ข้อมูล ซึ่งจะสุ่มแบ่งข้อมูลออกเป็น 3 ส่วน คือ ส่วนที่ 1 ข้อมูลเรียนรู้ (Training data) นำไปสร้างตัวแบบ (Model) ร้อยละ 70 ข้อมูลส่วนที่ 2 ข้อมูลตรวจสอบความถูกต้อง (Validation data) นำไปประเมินความผิดพลาดของตัวแบบ ร้อยละ 20 และข้อมูลส่วนที่ 3 ข้อมูลทดสอบ (Testing data) นำไปทดสอบตัวแบบ ร้อยละ 10 และผู้วิจัยได้นำวิธีเหล่านี้มาประยุกต์กับงานวิจัย โดยนำมาเปรียบเทียบประสิทธิภาพผลการจำแนกในการทำวิจัยครั้งนี้ ผู้วิจัยสนใจศึกษาการจำแนก (Classification) เพื่อเปรียบเทียบประสิทธิภาพวิธีการจำแนกมาจากค่าความถูกต้อง (Accuracy) ค่าความไว (Sensitivity) ค่าความจำเพาะ (Specificity) และค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) ซึ่งวิธีที่ใช้ในการทดสอบครั้งนี้มี 4 วิธี ผลการวิเคราะห์ข้อมูลแบ่งเป็นหัวข้อดังนี้

4.1 ผลการปรับความไม่สมดุลของข้อมูล 4 วิธี

4.2 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk

4.3 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48

4.4 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม โดยใช้ขั้นตอนวิธี Multilayer Perceptron

4.5 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO

4.6 ผลการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกทั้ง 4 วิธี และการปรับความไม่สมดุลทั้ง 4 วิธี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1 ผลการปรับความไม่สมดุลของข้อมูล 4 วิธี

จำนวนข้อมูลก่อนปรับ หลังปรับ และร้อยละหลังปรับความสมดุลของวิธีการปรับความไม่สมดุล 4 วิธี คือ วิธีการสุ่มเกิน (Over Sampling) วิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-Sampling) วิธีการสุ่มลด (Under Sampling) และวิธีกาผสมผสาน (Hybrid Methods) ดังแสดงในตารางที่ 4.1

ตารางที่ 4.1 ผลการปรับความไม่สมดุลของข้อมูล 4 วิธี

วิธีการปรับ		ชุดข้อมูล		
		OME	Credit	Red Wine
วิธีการสุ่มเกิน (Over Sampling)	ก่อนปรับ	890	400	999
	หลังปรับ	1650	710	1594
	ร้อยละ 70	1155	497	1115
	ร้อยละ 20	330	142	319
	ร้อยละ 10	165	71	160
วิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-Sampling)	ก่อนปรับ	890	400	999
	หลังปรับ	1690	720	1584
	ร้อยละ 70	1183	504	1108
	ร้อยละ 20	338	144	317
	ร้อยละ 10	169	72	159
วิธีการสุ่มลด (Under Sampling)	ก่อนปรับ	890	400	999
	หลังปรับ	100	100	390
	ร้อยละ 70	70	70	273
	ร้อยละ 20	20	20	78
	ร้อยละ 10	10	10	39
วิธีการผสมผสาน (Hybrid Methods)	ก่อนปรับ	890	400	999
	หลังปรับ	890	400	998
	ร้อยละ 70	623	280	698
	ร้อยละ 20	178	80	200
	ร้อยละ 10	89	40	100

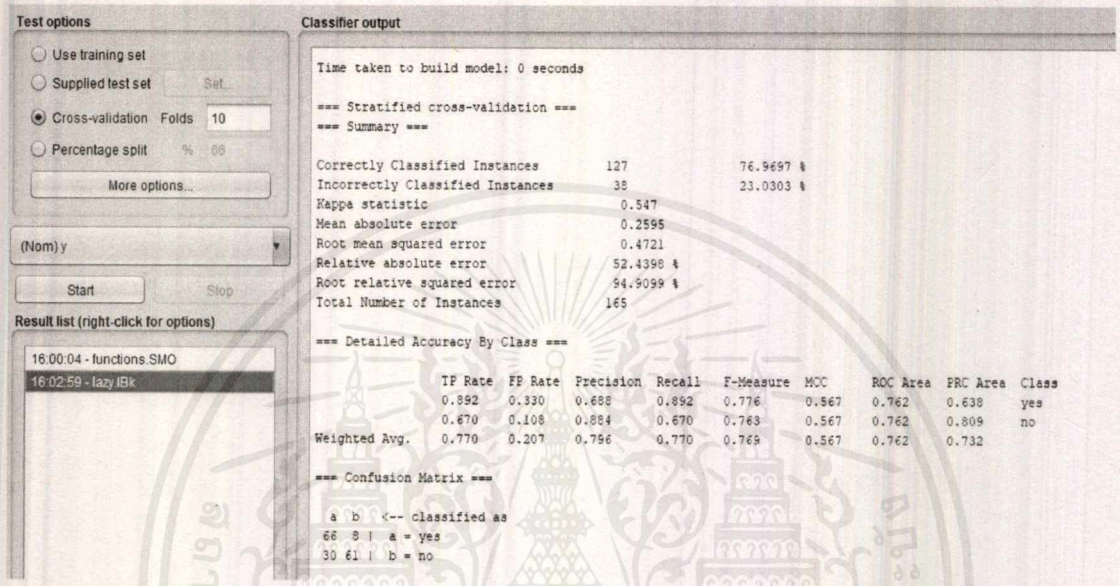
จากตารางที่ 4.1 จะพบว่าข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก มีจำนวนข้อมูลทั้งหมด 890 ค่า หลังการปรับด้วยวิธีการสุ่มเกิน มีจำนวนข้อมูล 1650 ค่า แบ่งข้อมูลทั้งหมดออกเป็นข้อมูลเรียนรู้ร้อยละ 70 จำนวน 1150 ค่า ข้อมูลตรวจสอบความถูกต้องร้อยละ 20 จำนวน 330 ค่า และข้อมูลทดสอบร้อยละ 10 จำนวน 165 ค่า วิธีการอื่นๆ ก็เช่นเดียวกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

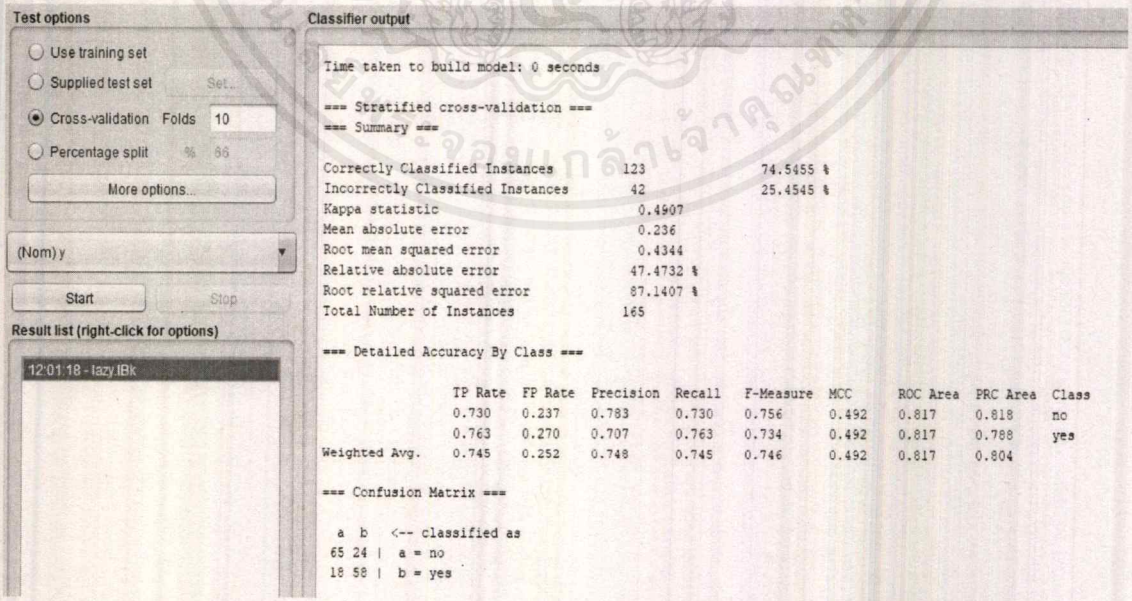
4.2 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk

4.2.1 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน (Over Sampling)

4.2.1.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทดสอบ 165 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

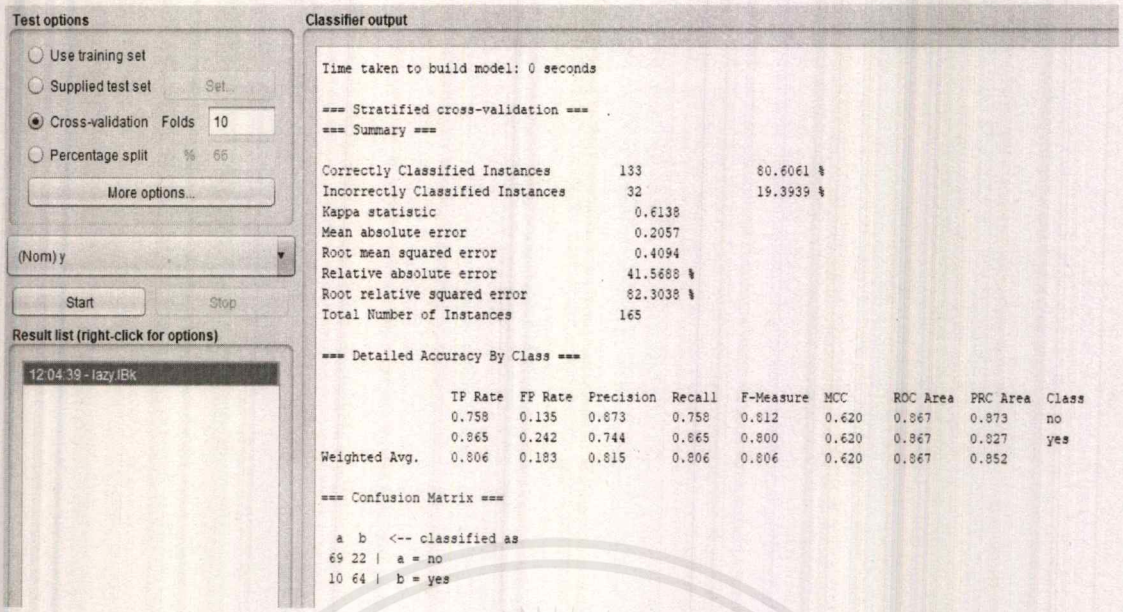


รูปที่ 4.1 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10



รูปที่ 4.2 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.3 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูปที่ 4.1 ถึง 4.3 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยวิธีสุ่มเกิน (Over Sampling) การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

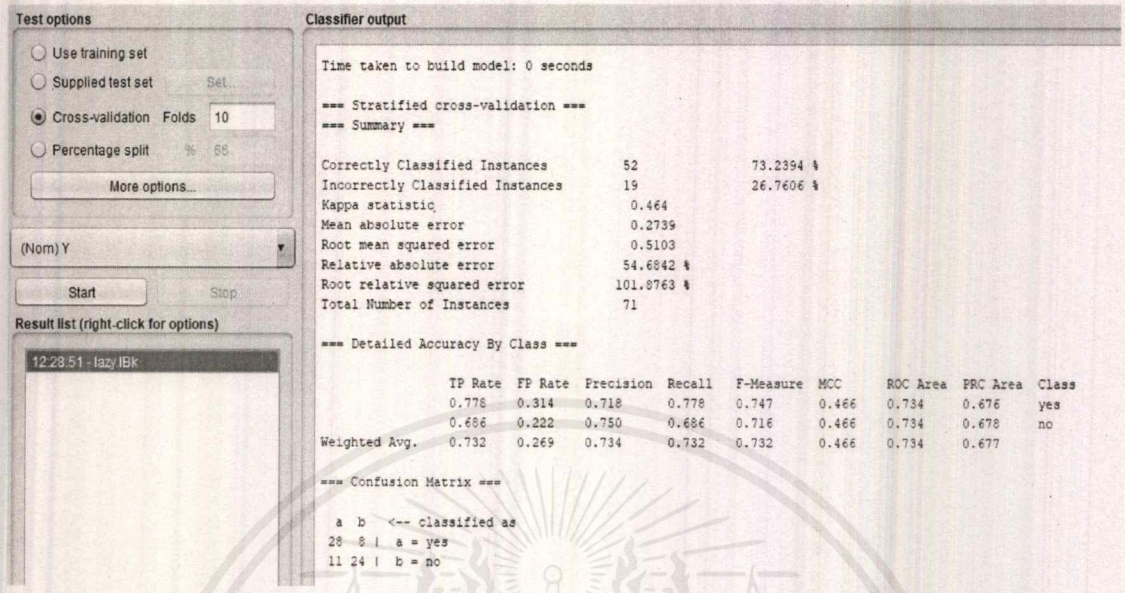
การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 165 คน มีค่าทำนายความถูกต้องจำนวน 127 คน คิดเป็นร้อยละ 76.9697 โดยมีค่าความไว = 0.8919 ค่าความจำเพาะ = 0.6703 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4721)^2$ = 0.2229

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 165 คน มีค่าทำนายความถูกต้องจำนวน 123 คน คิดเป็นร้อยละ 74.5455 โดยมีค่าความไว = 0.7303 ค่าความจำเพาะ = 0.7632 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4344)^2$ = 0.1887

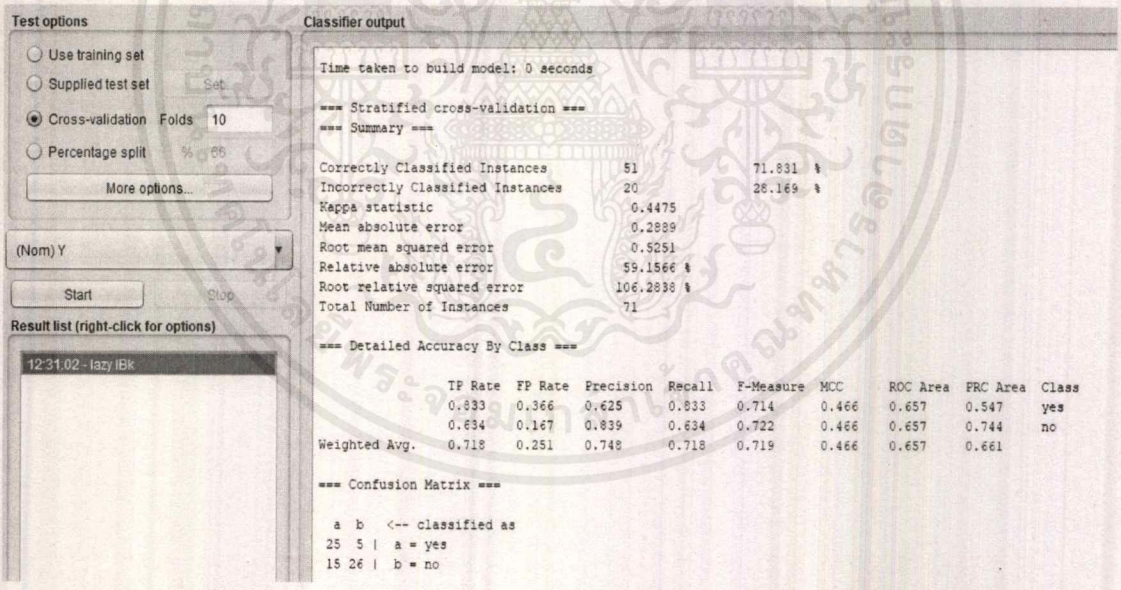
การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 165 คน มีค่าทำนายความถูกต้องจำนวน 133 คน คิดเป็นร้อยละ 80.6061 โดยมีค่าความไว = 0.7582 ค่าความจำเพาะ = 0.8649 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4094)^2$ = 0.1676

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.1.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทดสอบ 71 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

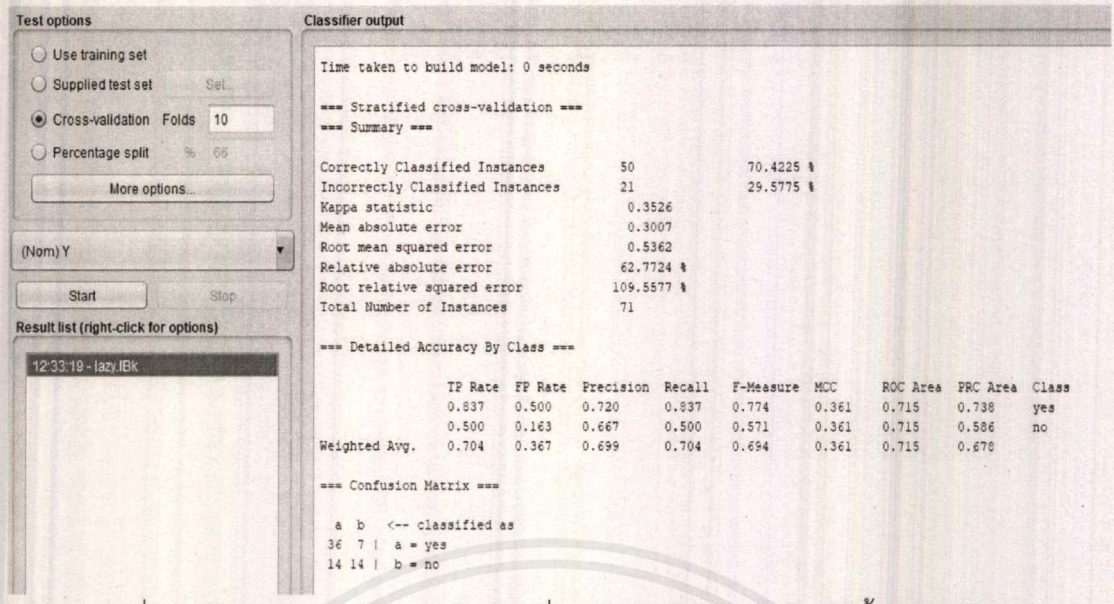


รูปที่ 4.4 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10



รูปที่ 4.5 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.6 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูปที่ 4.4 ถึง 4.6 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยวิธีสุ่มเกิน (Over Sampling) ยอดคงเหลือในบัตรเครดิตของลูกค้าโดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

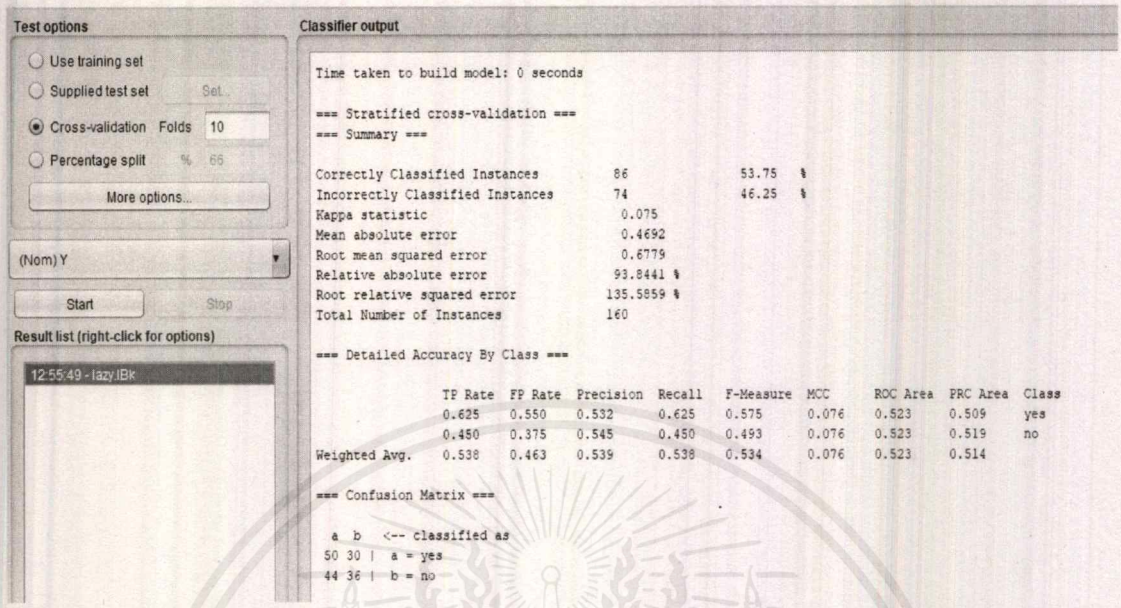
การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 71 คน มีค่าทำนายความถูกต้องจำนวน 52 คน คิดเป็นร้อยละ 73.2394 โดยมีค่าความไว = 0.7778 ค่าความจำเพาะ = 0.6856 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5103)^2$ = 0.2604

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 71 คน มีค่าทำนายความถูกต้องจำนวน 51 คน คิดเป็นร้อยละ 71.8310 โดยมีค่าความไว = 0.8333 ค่าความจำเพาะ = 0.6342 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5251)^2$ = 0.2757

การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 71 คน มีค่าทำนายความถูกต้องจำนวน 50 คน คิดเป็นร้อยละ 70.4225 โดยมีค่าความไว = 0.8372 ค่าความจำเพาะ = 0.5000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5362)^2$ = 0.2875

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.1.3 คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทดสอบ 160 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) Y

Start Stop

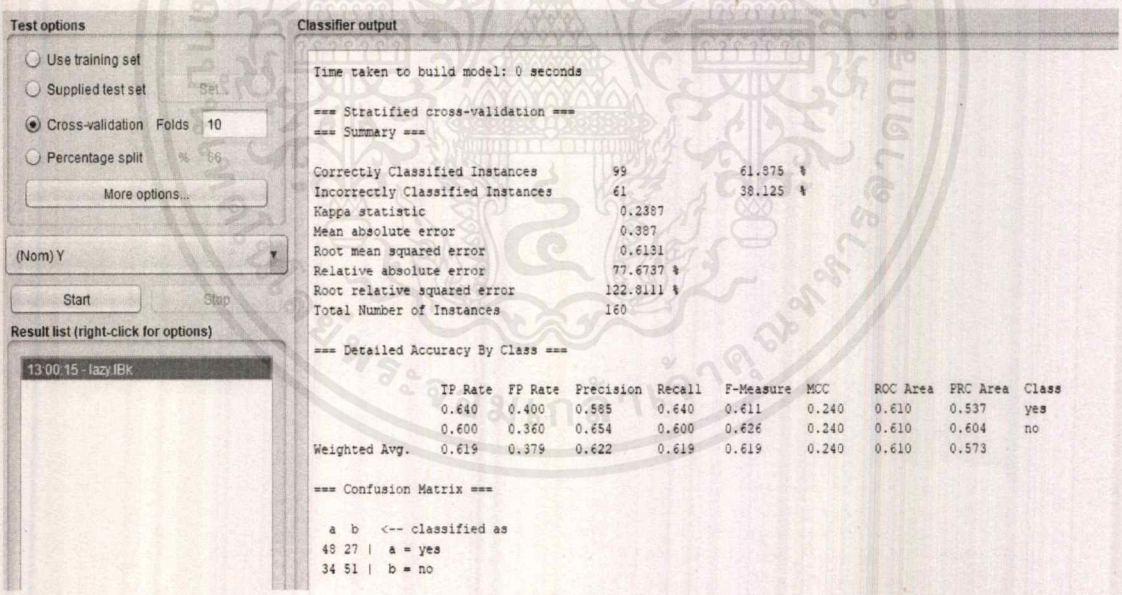
Result list (right-click for options)

12:55:49 - lazy.IBk

Classifier output

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          86               53.75  %
Incorrectly Classified Instances        74               46.25  %
Kappa statistic                          0.075
Mean absolute error                      0.4692
Root mean squared error                  0.6779
Relative absolute error                 93.8441 %
Root relative squared error            135.5859 %
Total Number of Instances              160

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
               0.625     0.550     0.532     0.625     0.575     0.076     0.523     0.509     yes
               0.450     0.375     0.545     0.450     0.493     0.076     0.523     0.519     no
Weighted Avg.  0.538     0.463     0.539     0.538     0.534     0.076     0.523     0.514

=== Confusion Matrix ===

  a  b   <-- classified as
 50 30 |  a = yes
 44 36 |  b = no
```

รูปที่ 4.7 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 10



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) Y

Start Stop

Result list (right-click for options)

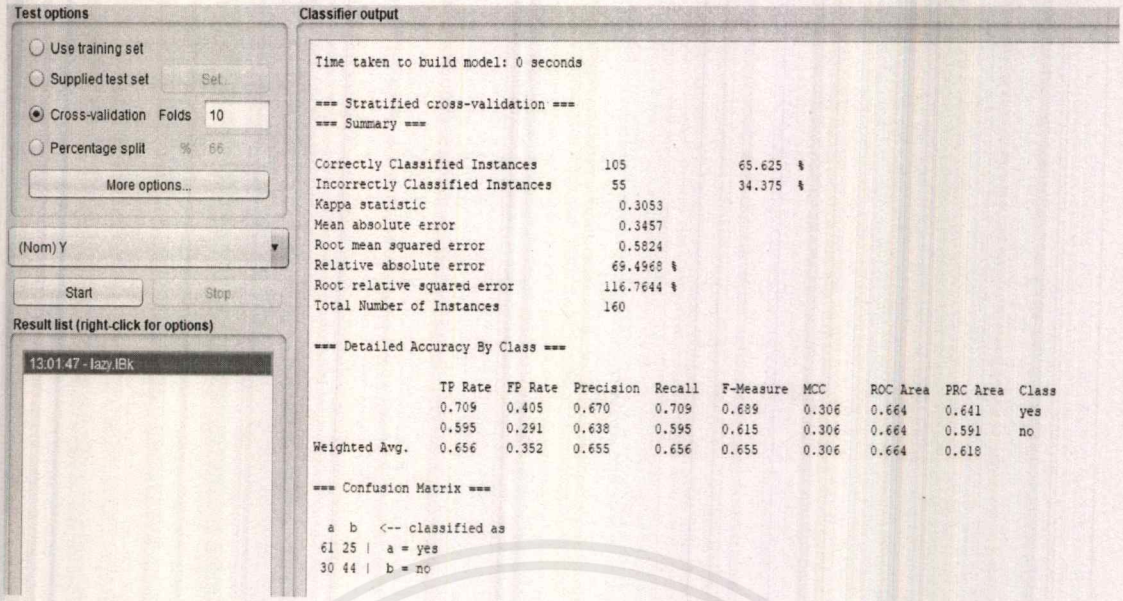
13:00:15 - lazy.IBk

Classifier output

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          99               61.875 %
Incorrectly Classified Instances        61               38.125 %
Kappa statistic                          0.2387
Mean absolute error                      0.387
Root mean squared error                  0.6131
Relative absolute error                 77.6737 %
Root relative squared error            122.8111 %
Total Number of Instances              160

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
               0.640     0.400     0.585     0.640     0.611     0.240     0.610     0.537     yes
               0.600     0.360     0.654     0.600     0.626     0.240     0.610     0.604     no
Weighted Avg.  0.619     0.379     0.622     0.619     0.619     0.240     0.610     0.573

=== Confusion Matrix ===

  a  b   <-- classified as
 48 27 |  a = yes
 34 51 |  b = no
```

รูปที่ 4.8 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.9 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูปที่ 4.7 ถึง 4.9 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k โดยการปรับความไม่สมดุลด้วยวิธีสุ่มเกิน (Over Sampling) คุณภาพไวน์แดง โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 160 ขวด มีค่าทำนายความถูกต้องจำนวน 86 ขวด คิดเป็นร้อยละ 53.75 โดยมีค่าความไว = 0.6250 ค่าความจำเพาะ = 0.4500 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6779)^2$ = 0.4596

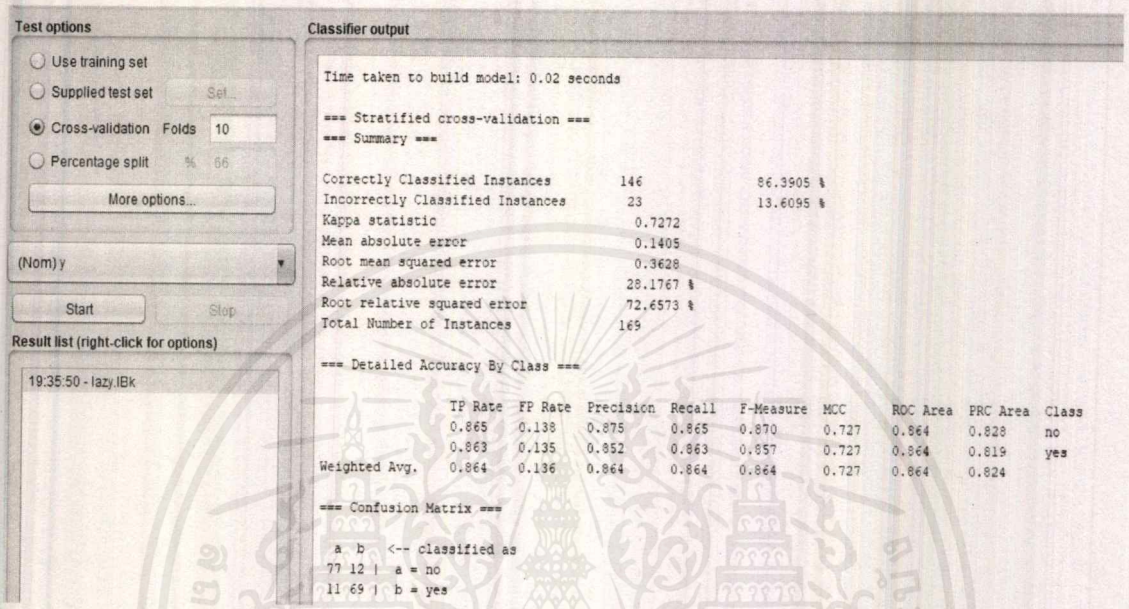
การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 160 ขวด มีค่าทำนายความถูกต้องจำนวน 99 ขวด คิดเป็นร้อยละ 61.875 โดยมีค่าความไว = 0.6400 ค่าความจำเพาะ = 0.6000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6131)^2$ = 0.3759

การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 160 ขวด มีค่าทำนายความถูกต้องจำนวน 105 ขวด คิดเป็นร้อยละ 65.625 โดยมีค่าความไว = 0.7093 ค่าความจำเพาะ = 0.5946 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5824)^2$ = 0.3392

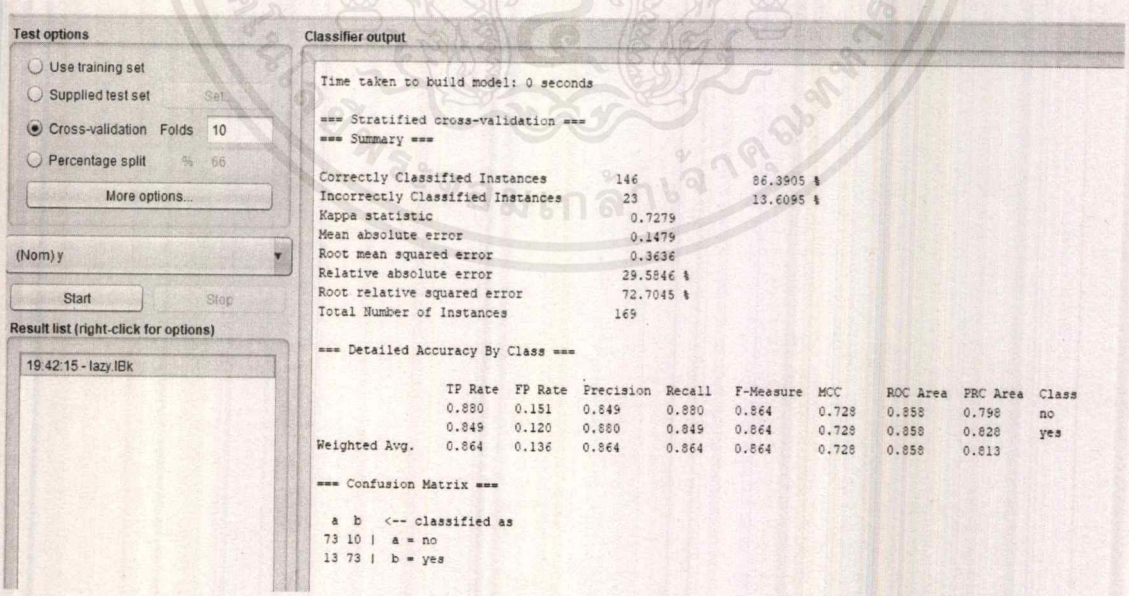
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-Sampling)

4.2.2.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทดสอบ 169 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

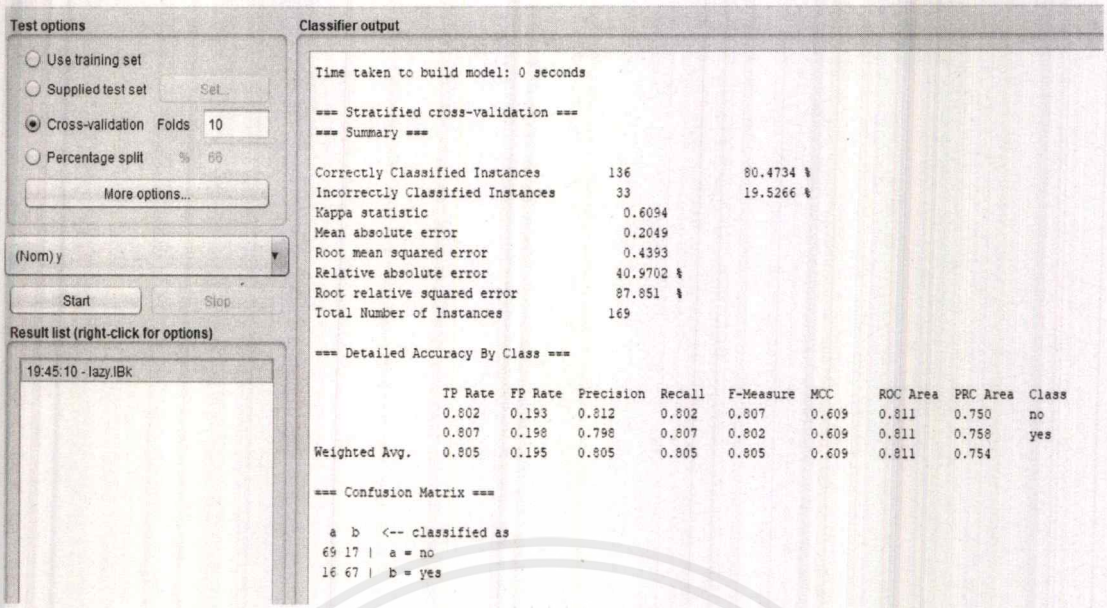


รูปที่ 4.10 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10



รูปที่ 4.11 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.12 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูป 4.10 ถึง 4.12 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-Sampling) การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

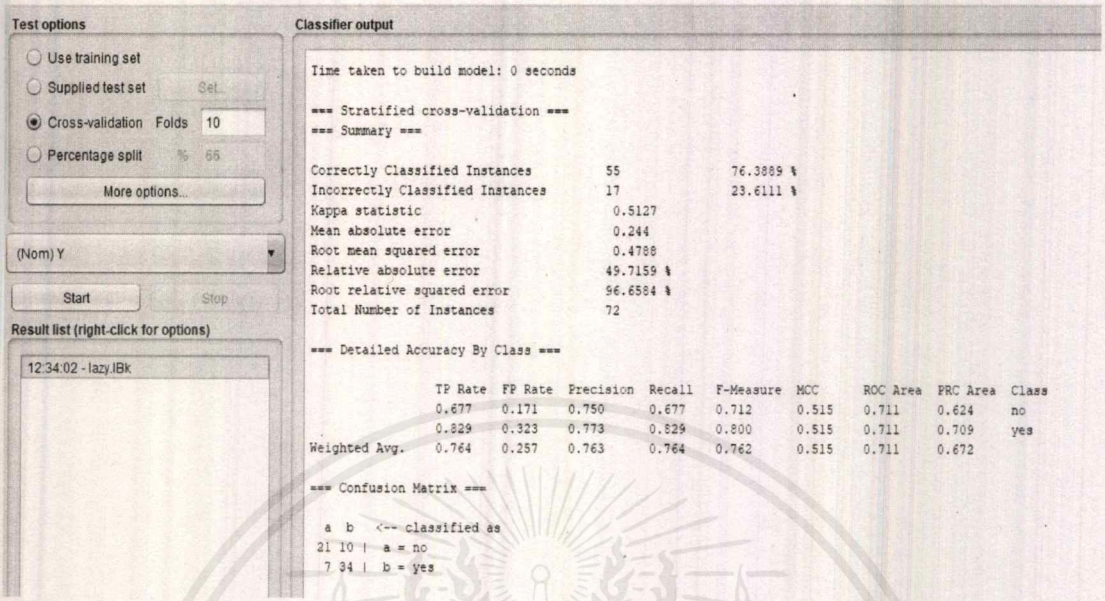
การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 169 คน มีค่าทำนายความถูกต้องจำนวน 146 คน คิดเป็นร้อยละ 86.3905 โดยมีค่าความไว = 0.8651 ค่าความจำเพาะ = 0.8625 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.3628)^2$ = 0.1316

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 169 คน มีค่าทำนายความถูกต้องจำนวน 146 คน คิดเป็นร้อยละ 86.3905 โดยมีค่าความไว = 0.8795 ค่าความจำเพาะ = 0.8488 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.3636)^2$ = 0.1322

การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 169 คน มีค่าทำนายความถูกต้องจำนวน 136 คน คิดเป็นร้อยละ 80.4734 โดยมีค่าความไว = 0.8023 ค่าความจำเพาะ = 0.8072 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4393)^2$ = 0.1930

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทดสอบ 72 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) Y

Start Stop

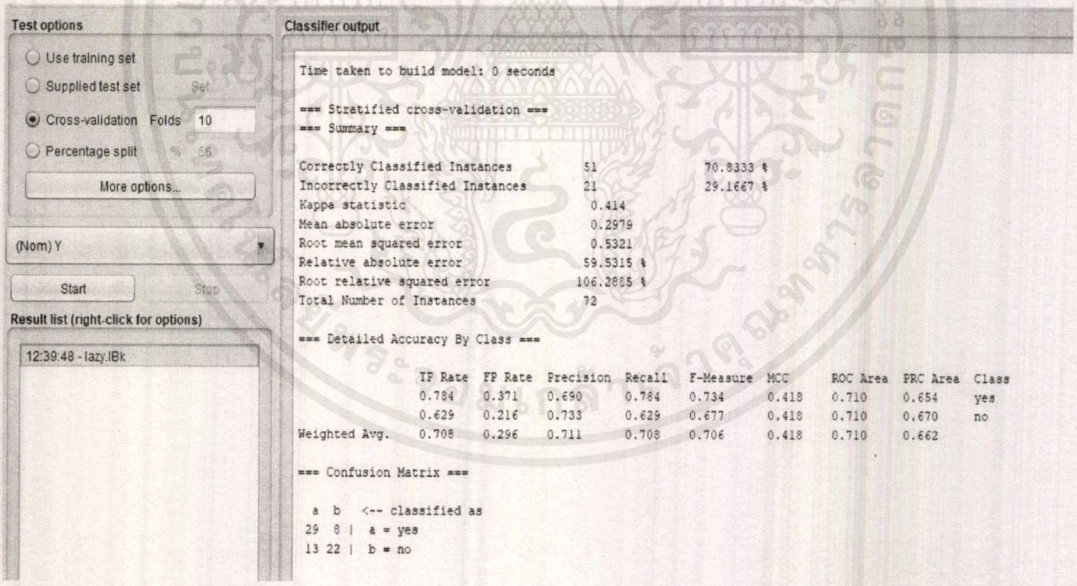
Result list (right-click for options)

12:34:02 - lazy.IBk

Classifier output

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          55               76.3889 %
Incorrectly Classified Instances        17               23.6111 %
Kappa statistic                          0.5127
Mean absolute error                      0.244
Root mean squared error                  0.4788
Relative absolute error                 49.7159 %
Root relative squared error             96.6584 %
Total Number of Instances               72

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
               0.677     0.171     0.750      0.677     0.712      0.515     0.711     0.624     no
               0.829     0.323     0.773      0.829     0.800      0.515     0.711     0.709     yes
Weighted Avg.  0.764     0.257     0.763      0.764     0.762      0.515     0.711     0.672

=== Confusion Matrix ===

  a  b   <-- classified as
 21 10 |  a = no
  7 34 |  b = yes
```

รูปที่ 4.13 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) Y

Start Stop

Result list (right-click for options)

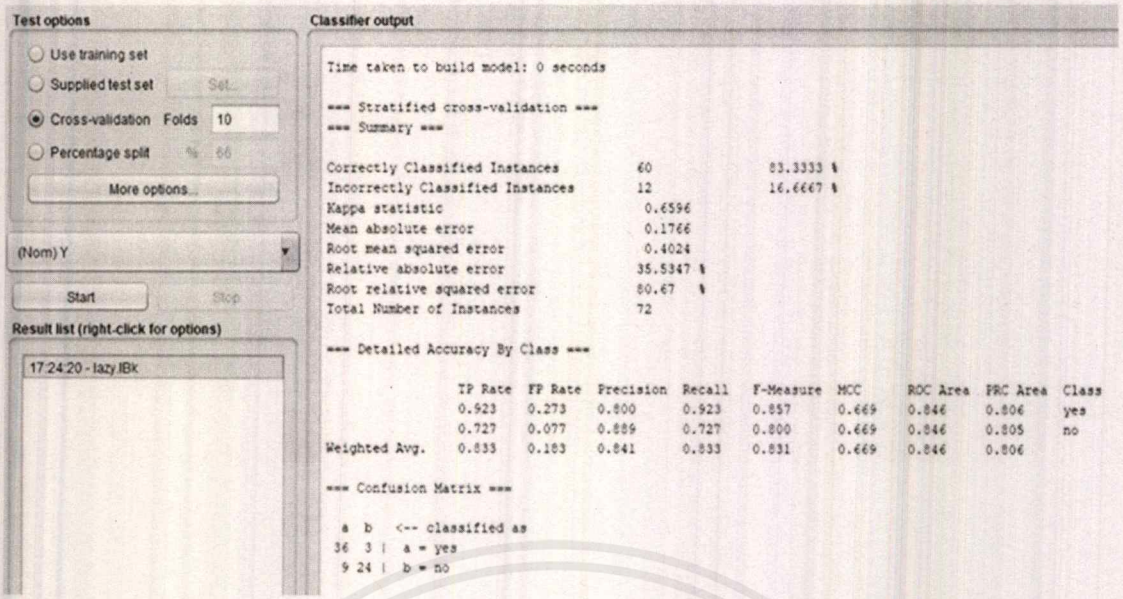
12:39:48 - lazy.IBk

Classifier output

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          51               70.8333 %
Incorrectly Classified Instances        21               29.1667 %
Kappa statistic                          0.414
Mean absolute error                      0.2979
Root mean squared error                  0.5321
Relative absolute error                 59.5315 %
Root relative squared error            106.2885 %
Total Number of Instances               72

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
               0.784     0.371     0.690      0.784     0.734      0.418     0.710     0.654     yes
               0.629     0.216     0.733      0.629     0.677      0.418     0.710     0.670     no
Weighted Avg.  0.708     0.296     0.711      0.708     0.706      0.418     0.710     0.662

=== Confusion Matrix ===

  a  b   <-- classified as
 29  8 |  a = yes
 13 22 |  b = no
```

รูปที่ 4.14 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.15 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูปที่ 4.13 ถึง 4.15 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลวิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-Sampling) ยอดคงเหลือในบัตรเครดิตของลูกค้า โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

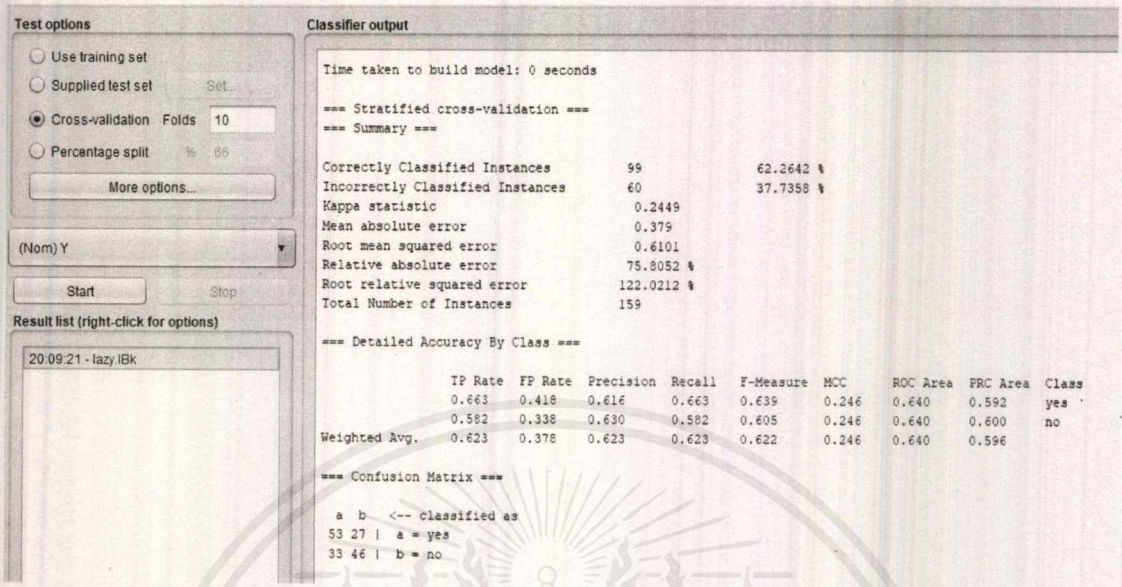
การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 72 คน มีค่าทำนายความถูกต้องจำนวน 55 คน คิดเป็นร้อยละ 76.3889 โดยมีค่าความไว = 0.6774 ค่าความจำเพาะ = 0.8293 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4788)^2$ = 0.2292

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 72 คน มีค่าทำนายความถูกต้องจำนวน 51 คน คิดเป็นร้อยละ 70.8333 โดยมีค่าความไว = 0.7838 ค่าความจำเพาะ = 0.6286 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5321)^2$ = 0.2831

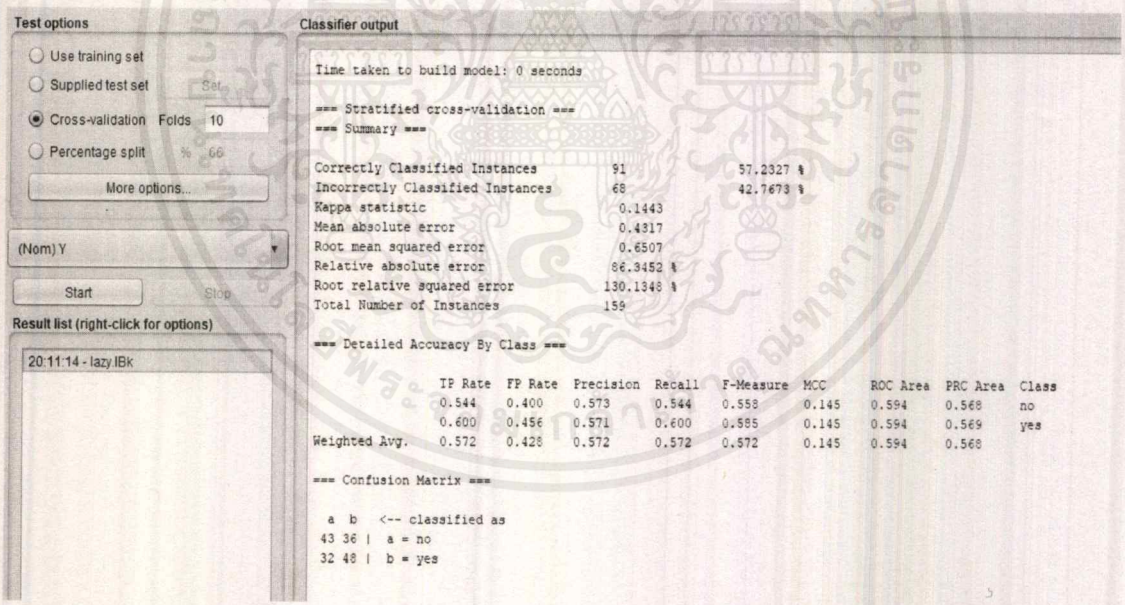
การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 72 คน มีค่าทำนายความถูกต้องจำนวน 60 คน คิดเป็นร้อยละ 83.3333 โดยมีค่าความไว = 0.9231 ค่าความจำเพาะ = 0.7273 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4024)^2$ = 0.1619

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2.3 คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทดสอบ 159 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

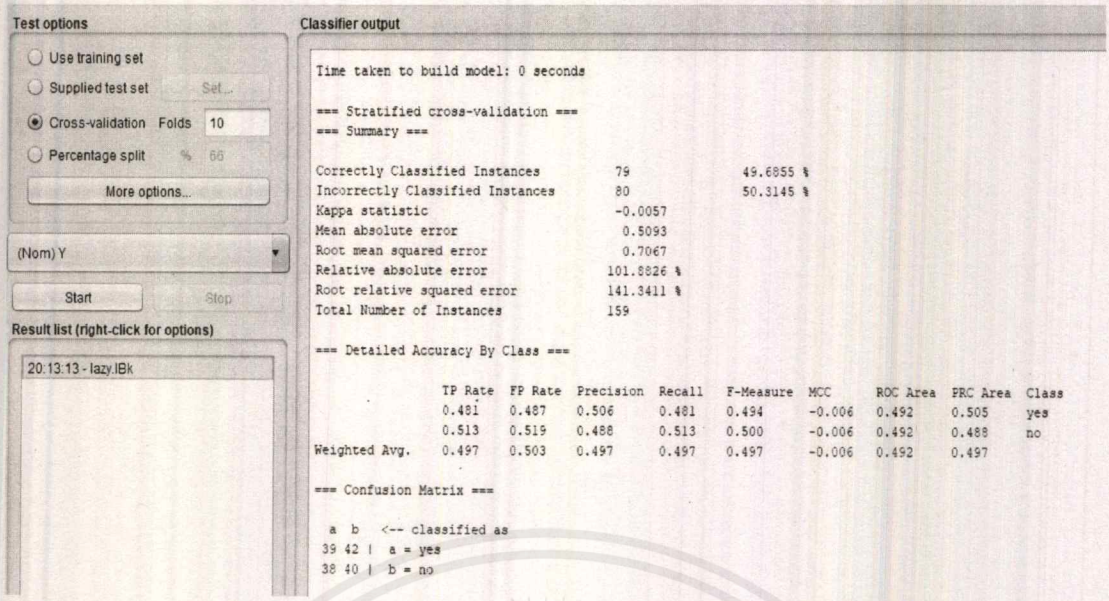


รูปที่ 4.16 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 10



รูปที่ 4.17 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.18 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูปที่ 4.16 ถึง 4.18 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-Sampling) คุณภาพไวน์แดง โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 159 ขวด มีค่าทำนายความถูกต้องจำนวน 99 ขวด คิดเป็นร้อยละ 62.2642 โดยมีค่าความไว = 0.6625 ค่าความจำเพาะ = 0.5822 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6101)^2$ = 0.3722

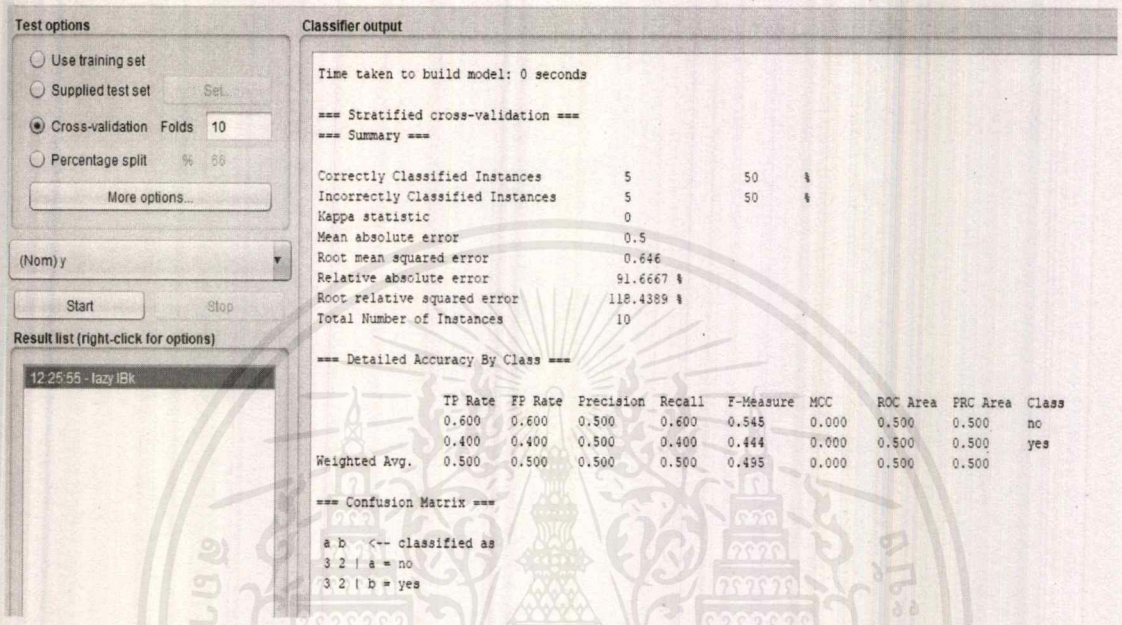
การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 159 ขวด มีค่าทำนายความถูกต้องจำนวน 91 ขวด คิดเป็นร้อยละ 57.2327 โดยมีค่าความไว = 0.5443 ค่าความจำเพาะ = 0.6000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6507)^2$ = 0.4234

การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 159 ขวด มีค่าทำนายความถูกต้องจำนวน 79 ขวด คิดเป็นร้อยละ 49.6855 โดยมีค่าความไว = 0.4815 ค่าความจำเพาะ = 0.5128 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.7067)^2$ = 0.4994

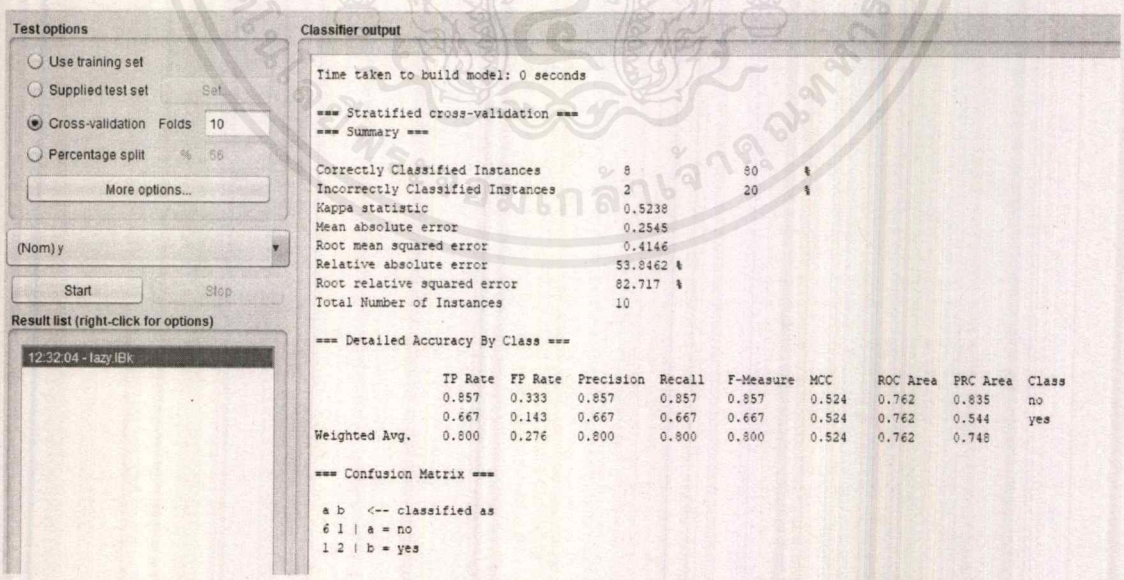
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.3 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด (Under Sampling)

4.2.3.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทดสอบ 10 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

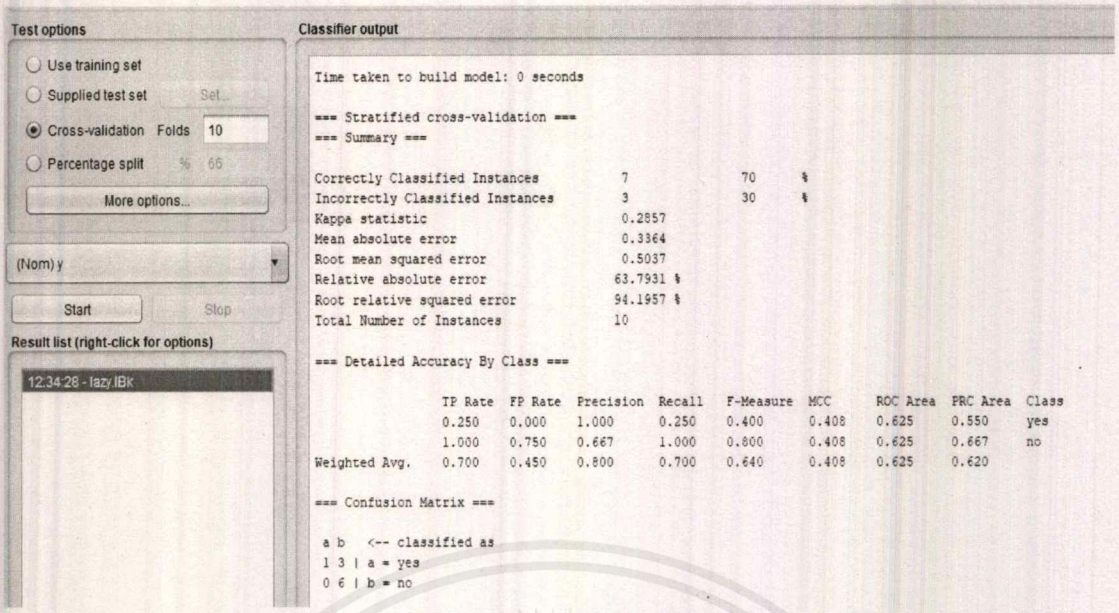


รูปที่ 4.19 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10



รูปที่ 4.20 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.21 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูปที่ 4.19 ถึง 4.21 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด (Under Sampling) การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

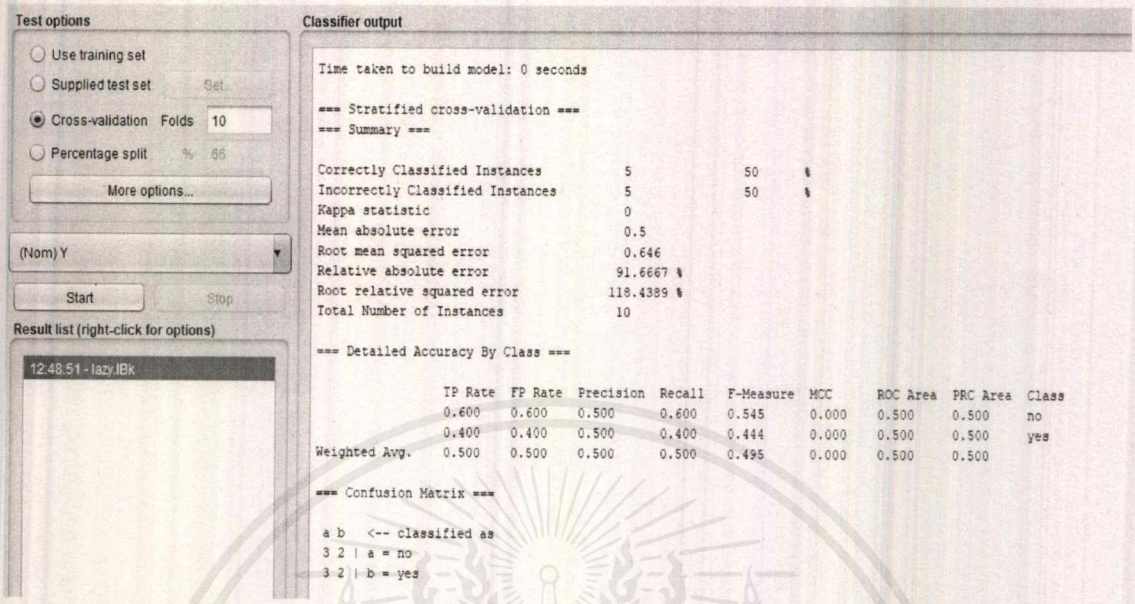
การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 5 คน คิดเป็นร้อยละ 50 โดยมีค่าความไว = 0.6000 ค่าความจำเพาะ = 0.4000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.646)^2$ = 0.4173

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 8 คน คิดเป็นร้อยละ 80 โดยมีค่าความไว = 0.8571 ค่าความจำเพาะ = 0.6667 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4146)^2$ = 0.1719

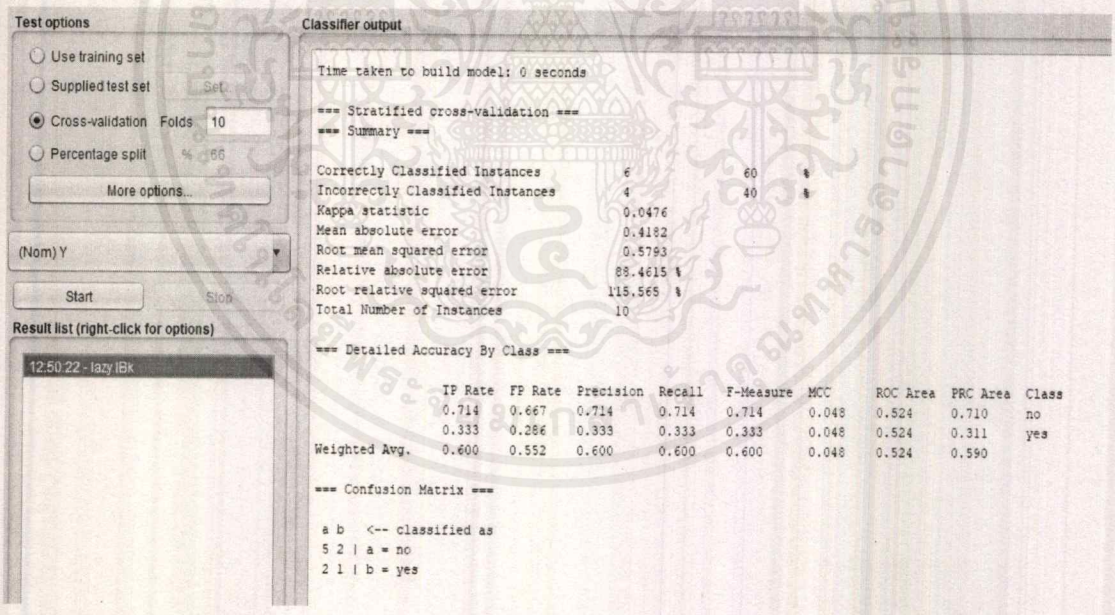
การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 7 คน คิดเป็นร้อยละ 70 โดยมีค่าความไว = 0.2500 ค่าความจำเพาะ = 1 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5037)^2$ = 0.2537

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.3.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทดสอบ 10 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

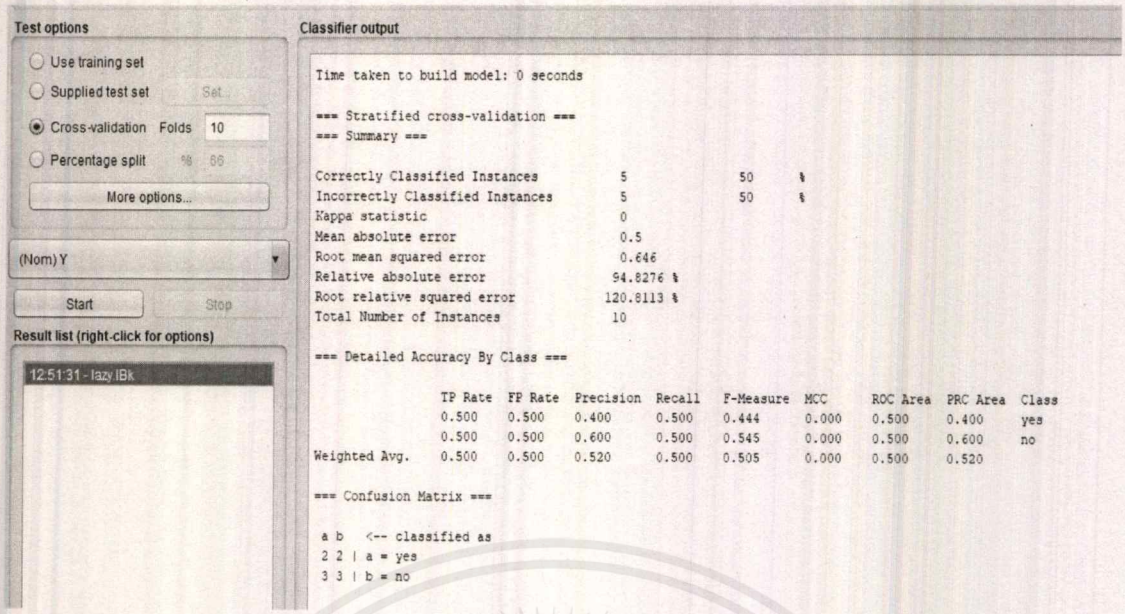


รูปที่ 4.22 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10



รูปที่ 4.23 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.24 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูปที่ 4.22 ถึง 4.24 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด (Under Sampling) ยอดคงเหลือในบัตรเครดิตของลูกค้า โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

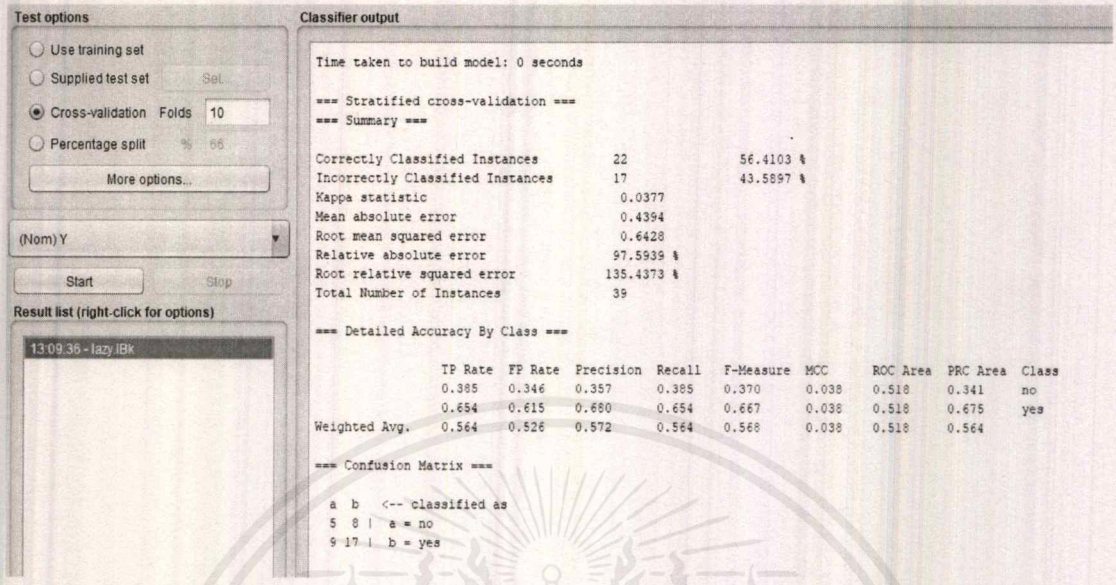
การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 5 คน คิดเป็นร้อยละ 50 โดยมีค่าความไว = 0.6000 ค่าความจำเพาะ = 0.4000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.646)^2$ = 0.4173

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 6 คน คิดเป็นร้อยละ 60 โดยมีค่าความไว = 0.7142 ค่าความจำเพาะ = 0.3333 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5793)^2$ = 0.3356

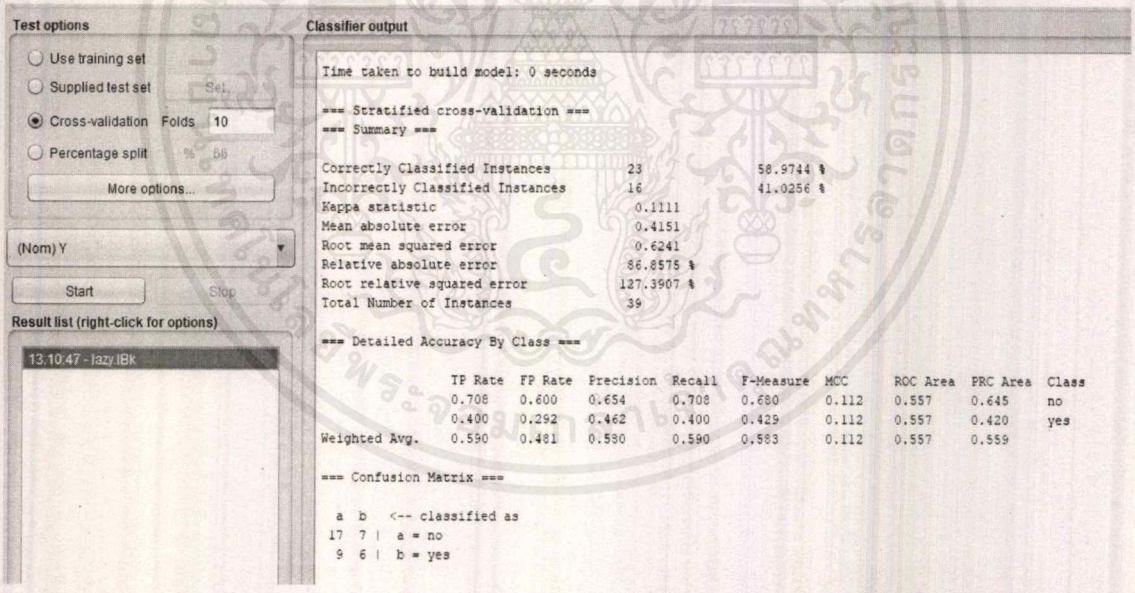
การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 5 คน คิดเป็นร้อยละ 50 โดยมีค่าความไว = 0.5000 ค่าความจำเพาะ = 0.5000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.646)^2$ = 0.4173

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.3.3 คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทดสอบ 39 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

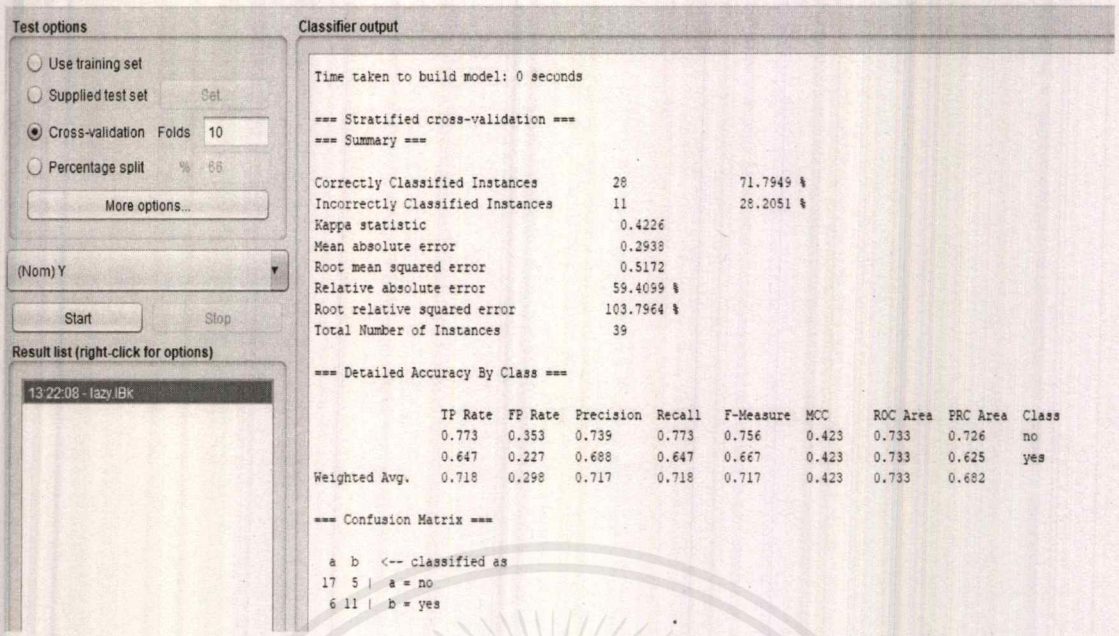


รูปที่ 4.25 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 10



รูปที่ 4.26 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.27 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูปที่ 4.25 ถึง 4.27 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด (Under Sampling) คุณภาพไวน์แดง โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 22 ขวด คิดเป็นร้อยละ 56.4103 โดยมีค่าความไว = 0.3846 ค่าความจำเพาะ = 0.6538 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6428)^2$ = 0.4132

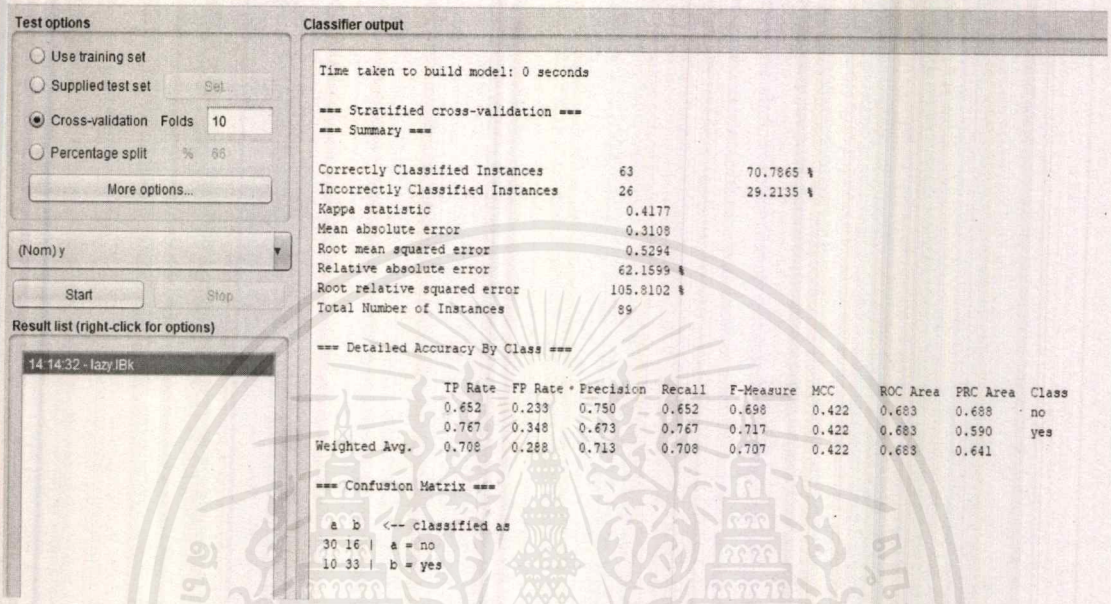
การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 23 ขวด คิดเป็นร้อยละ 58.9744 โดยมีค่าความไว = 0.7083 ค่าความจำเพาะ = 0.4000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6241)^2$ = 0.3895

การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 28 ขวด คิดเป็นร้อยละ 71.7949 โดยมีค่าความไว = 0.7727 ค่าความจำเพาะ = 0.6471 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5172)^2$ = 0.2675

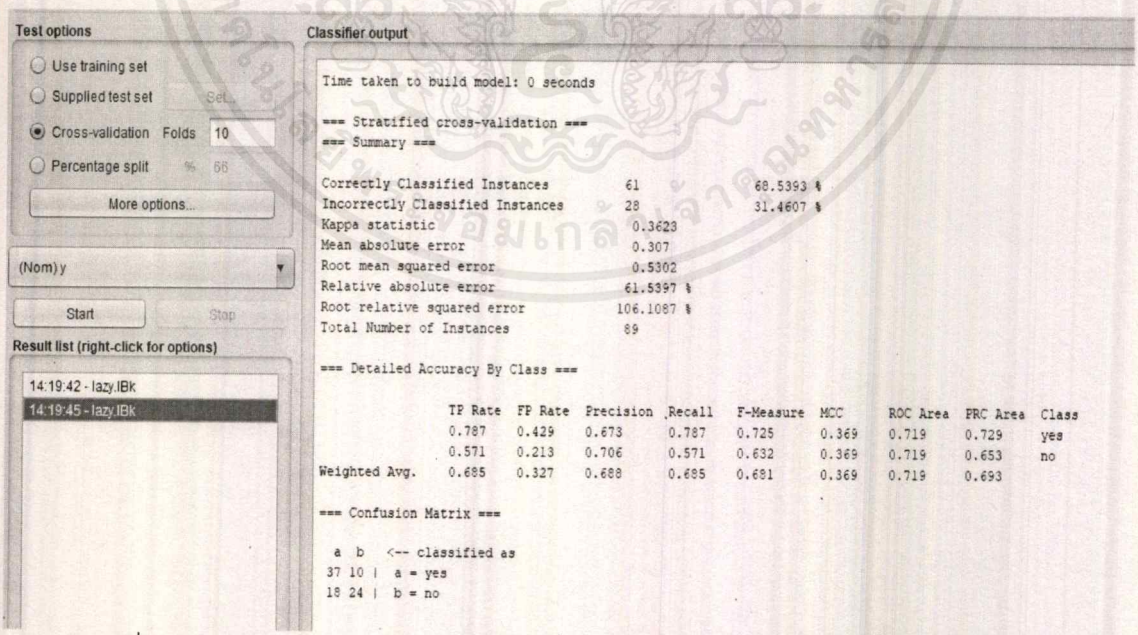
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.4 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยวิธีการผสมผสาน (Hybrid Methods)

4.2.4.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทดสอบ 89 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

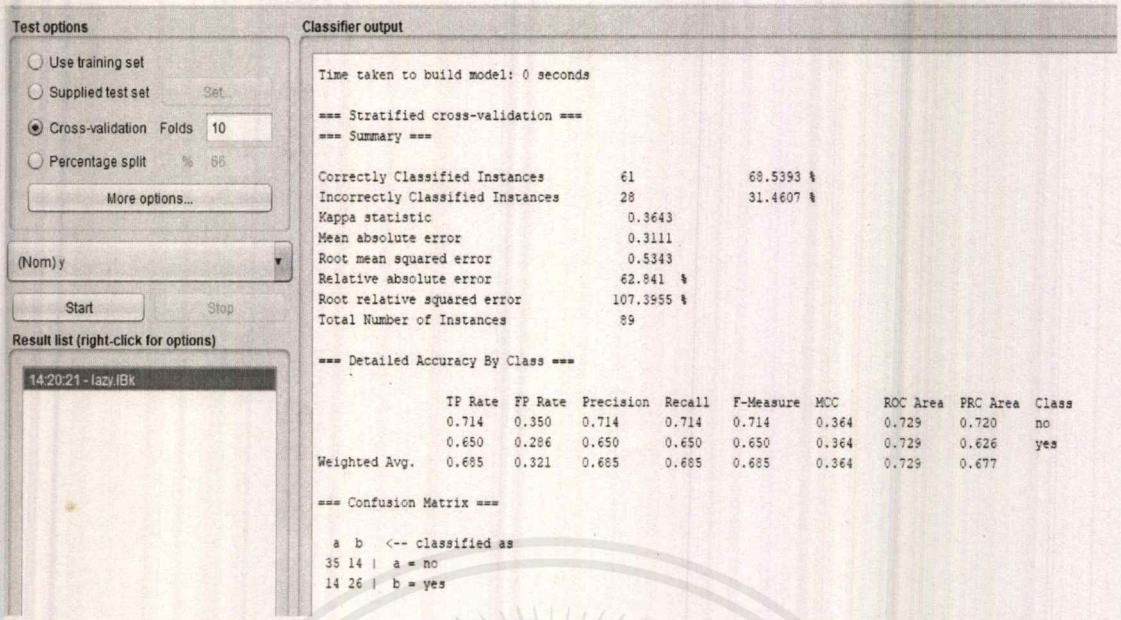


รูปที่ 4.28 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10



รูปที่ 4.29 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.30 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูปที่ 4.27 ถึง 4.30 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยด้วยวิธีการผสมผสาน (Hybrid Methods) การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

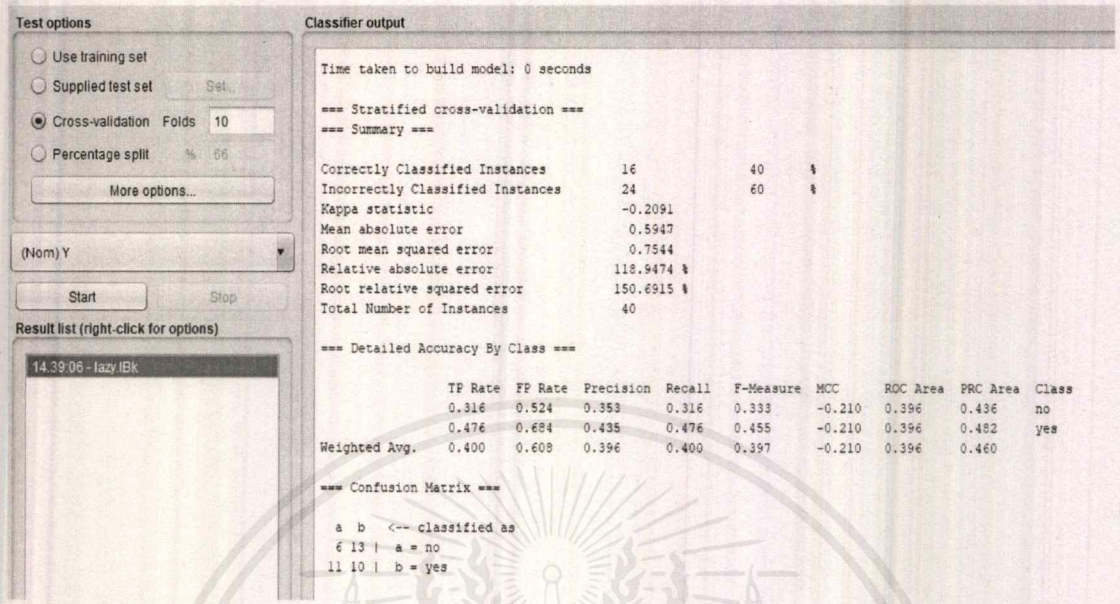
การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 89 คน มีค่าทำนายความถูกต้องจำนวน 63 คน คิดเป็นร้อยละ 70.7865 โดยมีค่าความไว = 0.6522 ค่าความจำเพาะ = 0.7674 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5294)^2$ = 0.2803

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 89 คน มีค่าทำนายความถูกต้องจำนวน 61 คน คิดเป็นร้อยละ 68.5393 โดยมีค่าความไว = 0.7872 ค่าความจำเพาะ = 0.5714 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5302)^2$ = 0.2811

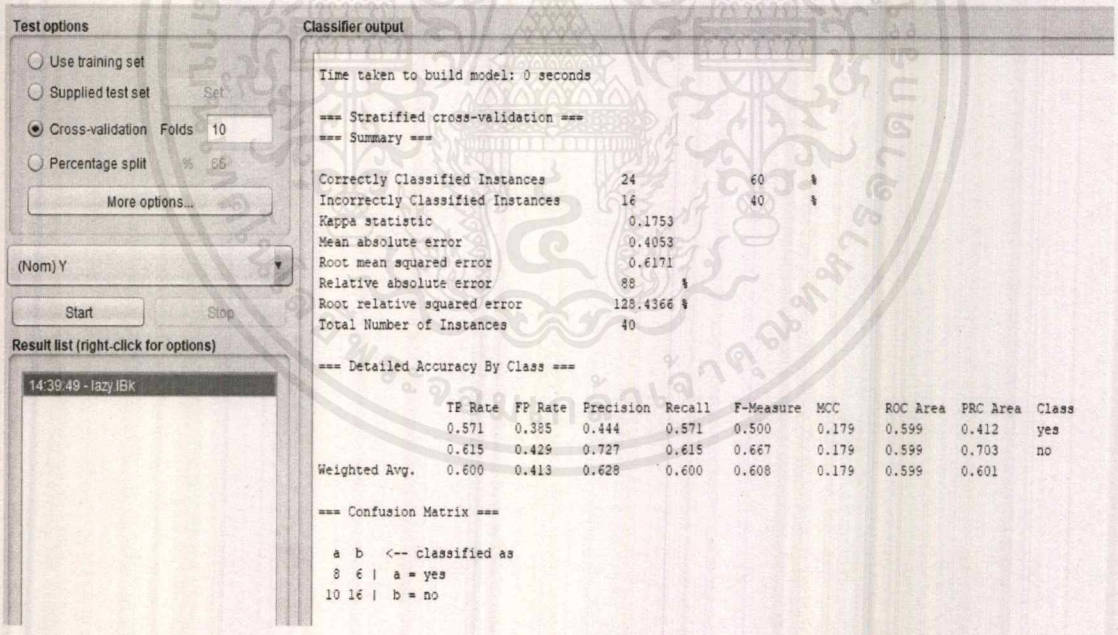
การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 89 คน มีค่าทำนายความถูกต้องจำนวน 61 คน คิดเป็นร้อยละ 68.5393 โดยมีค่าความไว = 0.7143 ค่าความจำเพาะ = 0.6500 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5343)^2$ = 0.2855

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.4.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทดสอบ 40 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

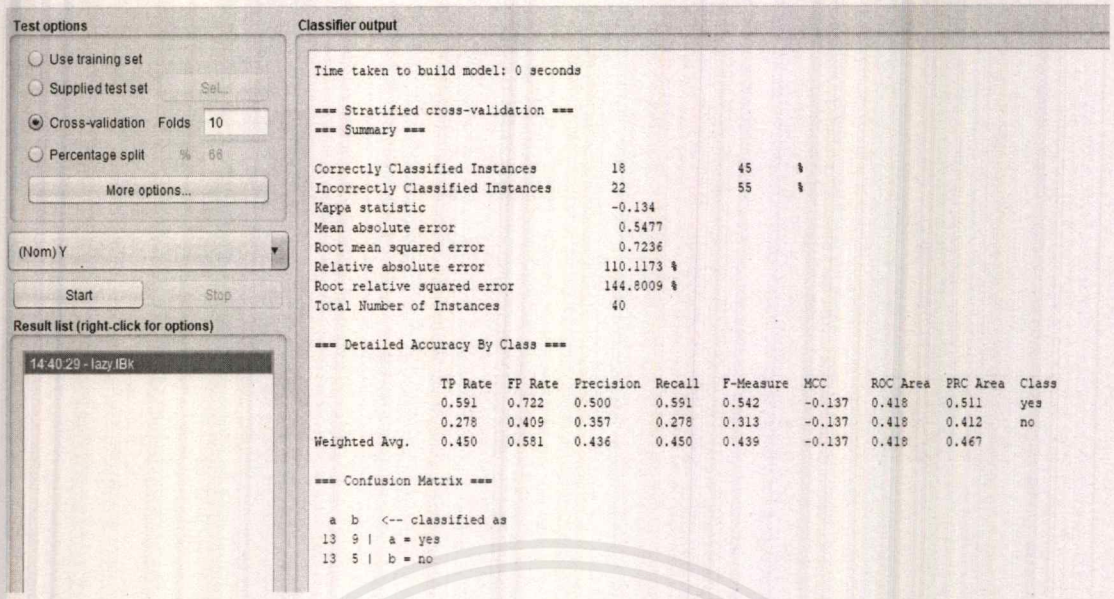


รูปที่ 4.31 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10



รูปที่ 4.32 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.33 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูปที่ 4.31 ถึง 4.33 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยวิธีการผสมผสาน (Hybrid Metods) โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

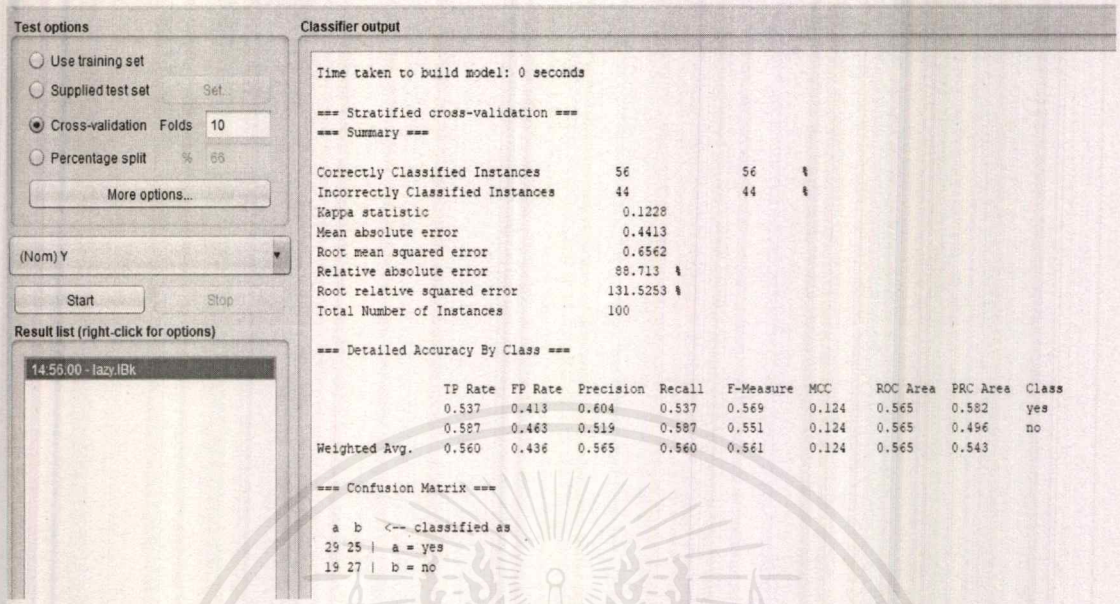
การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 40 คน มีค่าทำนายความถูกต้องจำนวน 16 คน คิดเป็นร้อยละ 40 โดยมีค่าความไว = 0.3158 ค่าความจำเพาะ = 0.4762 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.7544)^2$ = 0.5691

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 40 คน มีค่าทำนายความถูกต้องจำนวน 24 คน คิดเป็นร้อยละ 60 โดยมีค่าความไว = 0.5714 ค่าความจำเพาะ = 0.6154 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6171)^2$ = 0.3808

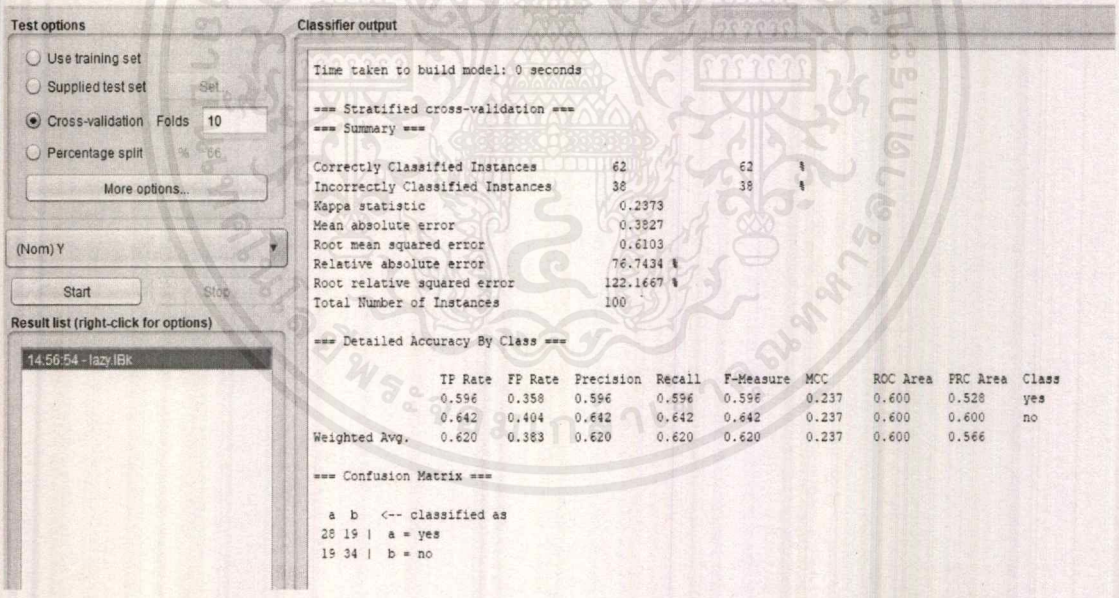
การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 40 คน มีค่าทำนายความถูกต้องจำนวน 18 คน คิดเป็นร้อยละ 45 โดยมีค่าความไว = 0.5909 ค่าความจำเพาะ = 0.2778 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.7236)^2$ = 0.5236

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.4.3 คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทดสอบ 100 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

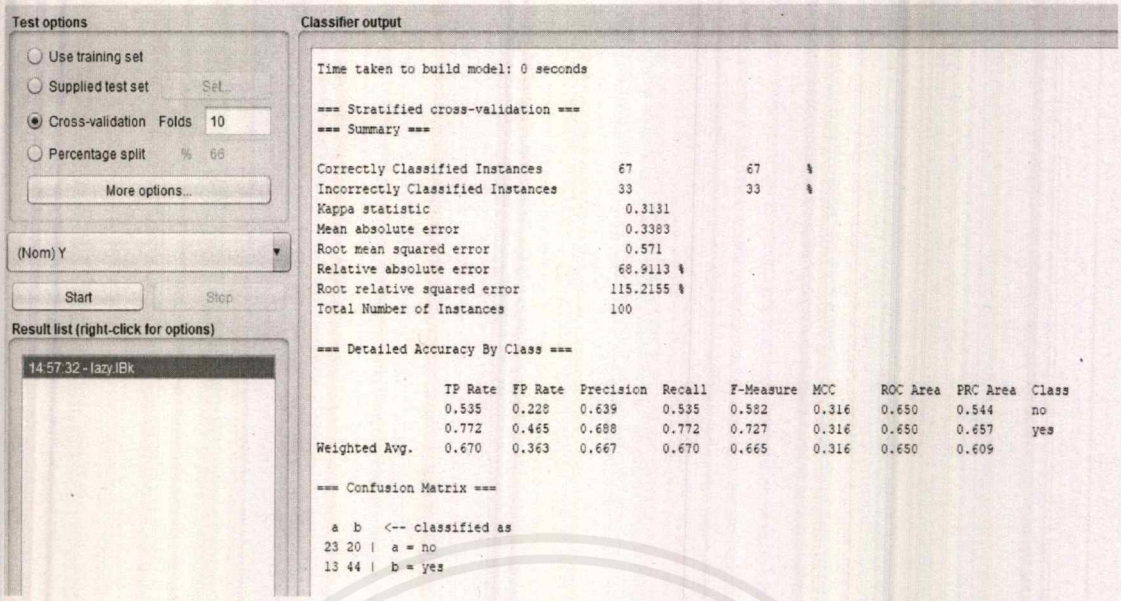


รูปที่ 4.34 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 10



รูปที่ 4.35 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.36 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยใช้ขั้นตอนวิธี IBk คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูปที่ 4.34 ถึง 4.36 ผลการวิเคราะห์ข้อมูลด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยวิธีการผสมผสาน (Hybrid Metods) คุณภาพไวน์แดง โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 100 ขวด มีค่าทำนายความถูกต้องจำนวน 56 ขวด คิดเป็นร้อยละ 56 โดยมีค่าความไว = 0.5370 ค่าความจำเพาะ = 0.5870 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6562)^2$ = 0.4306

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 100 ขวด มีค่าทำนายความถูกต้องจำนวน 62 ขวด คิดเป็นร้อยละ 62 โดยมีค่าความไว = 0.5957 ค่าความจำเพาะ = 0.6415 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6103)^2$ = 0.3725

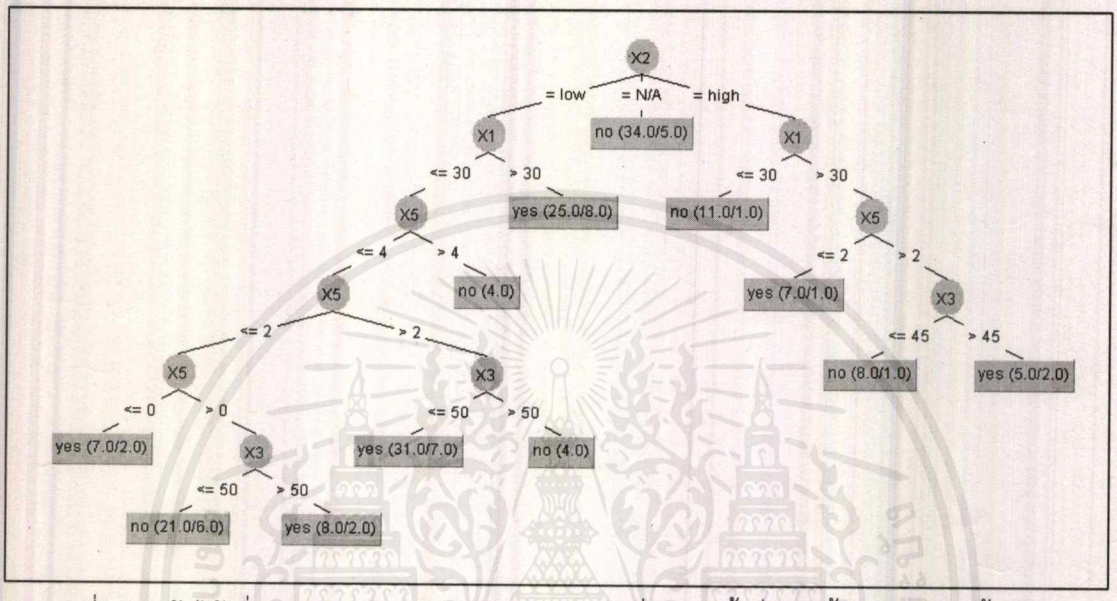
การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 100 ขวด มีค่าทำนายความถูกต้องจำนวน 67 ขวด คิดเป็นร้อยละ 67 โดยมีค่าความไว = 0.5349 ค่าความจำเพาะ = 0.7719 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5710)^2$ = 0.3260

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48

4.3.1 ผลการวิเคราะห์ข้อมูลด้วยวิธีวิธีต้นไม้ตัดสินใจ โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน (Over Sampling)

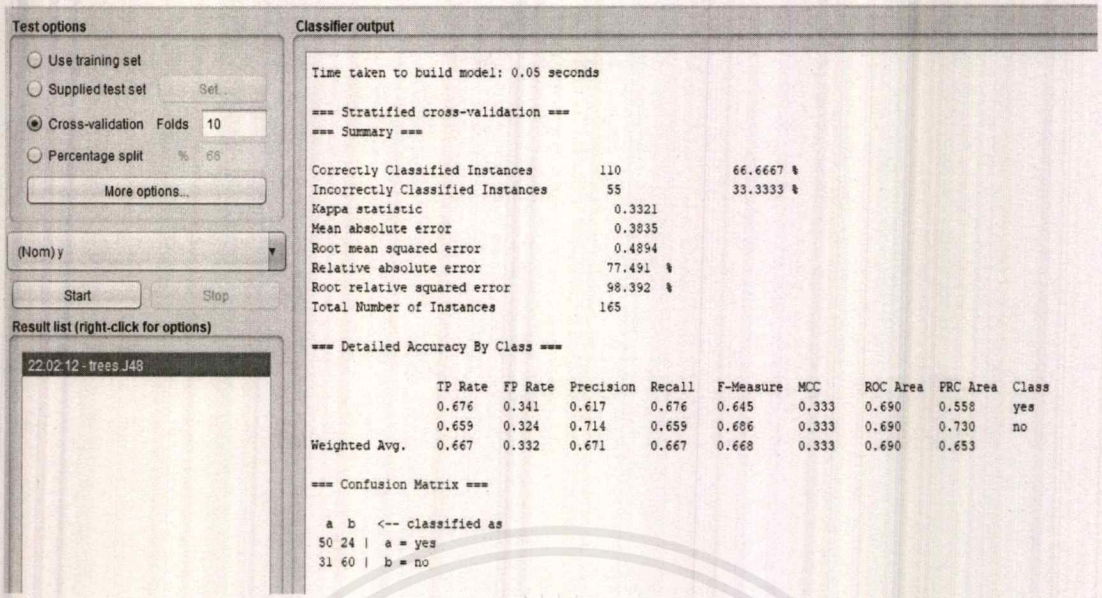
4.3.1.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทดสอบ 165 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.37 ต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10

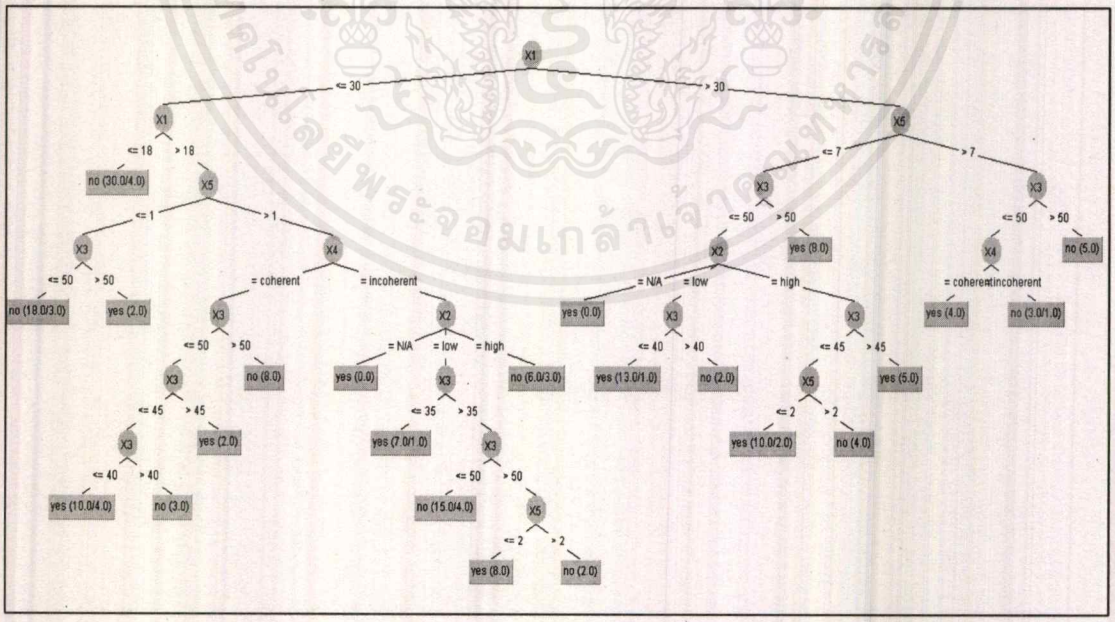
จากรูป 4.37 แสดงต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โหนดรากแบ่งโดยตัวแปรทำนายคือ X_2 = อัตราความเสี่ยงการเป็นโรคหูชั้นกลางอักเสบ ถ้าผลเป็น N/A (คือไม่มีความเสี่ยง) คือได้ยินเสียงไม่ปกติ จำนวน 34 คน ถ้าผลเป็น high (คือความเสี่ยงมาก) ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = อายุ ถ้าผลน้อยกว่าหรือเท่ากับ 30 เดือน คือได้ยินเสียงไม่ปกติ จำนวน 11 คน ถ้าผลมากกว่า 30 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าเท่ากับ 2 ครั้ง คือได้ยินเสียงปกติ จำนวน 7 คน ถ้าผลมากกว่า 2 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลน้อยกว่าเท่ากับ 45 เดซิเบล คือได้ยินเสียงไม่ปกติ ถ้าผลมากกกว่า 45 เดซิเบล คือได้ยินเสียงปกติ ถ้าผลเป็ณ low (คือความเสี่ยงน้อย) ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = อายุ ถ้าผลมากกว่า 30 เดือน คือได้ยินเสียงปกติ จำนวน 25 คน ถ้าผลน้อยกว่าหรือเท่ากับ 30 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลมากกว่า 4 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 4 คน ถ้าผลน้อยกว่าเท่ากับ 4 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลมากกว่า 2 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลน้อยกว่าเท่ากับ 50 เดซิเบล คือได้ยินเสียงปกติ จำนวน 31 คน ถ้าผลมากกกว่า 50 เดซิเบล คือได้ยินเสียงไม่ปกติ จำนวน 4 คน ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าเท่ากับ 2 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าเท่ากับ 0 ครั้ง คือได้ยินเสียงปกติ จำนวน 7 คน ถ้าผลมากกว่า 0 ตัดสินใจด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลน้อยกว่าเท่ากับ 50 เดซิเบล คือได้ยินเสียงไม่ปกติ จำนวน 21 คน ถ้าผลมากกว่า 50 เดซิเบล คือได้ยินเสียง จำนวน 8 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.38 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10

จากรูป 4.38 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 165 คน มีค่าทำนายความถูกต้องจำนวน 110 คน คิดเป็นร้อยละ 66.6667 โดยมีค่าความไว = 0.6757 ค่าความจำเพาะ = 0.6593 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4894)^2$ = 0.2395

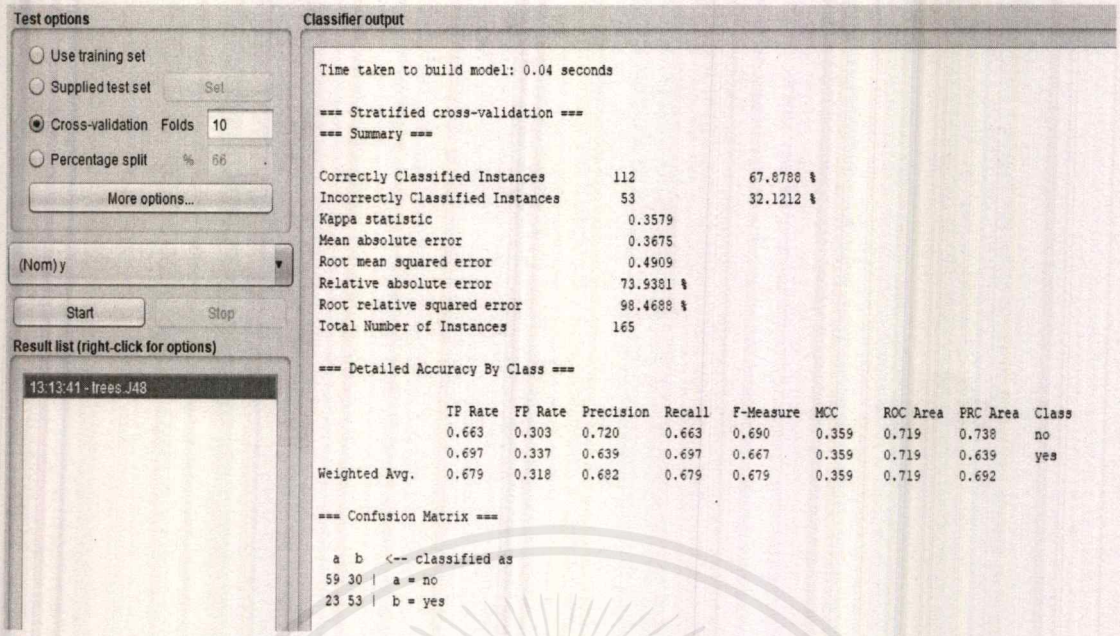


รูปที่ 4.39 ต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

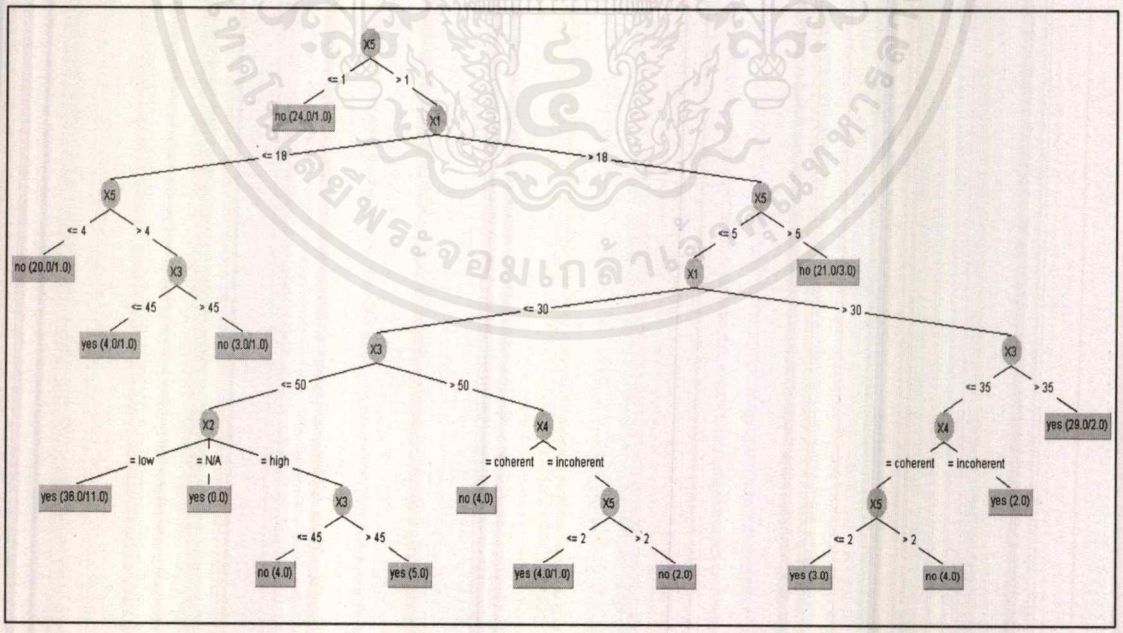
จากรูป 4.39 แสดงต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โหนดรากแบ่งโดยตัวแปรทำนายคือ X_1 = อายุ ถ้าผลมากกว่า 30 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลมากกว่า 7 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลมากกกว่า 50 เดซิเบล คือได้ยินเสียงไม่ปกติ จำนวน 5 คน ถ้าผลน้อยกว่าเท่ากับ 50 เดซิเบล ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = สัญญาณในการกระตุ้น ถ้าผลเท่ากับ coherent (คือสัญญาณต่อเนื่อง) คือได้ยินเสียงปกติ จำนวน 4 คน ถ้าผลเท่ากับ incoherent (คือสัญญาณไม่ต่อเนื่อง) คือได้ยินเสียงปกติ จำนวน 3 คน ถ้าผลน้อยกว่าเท่ากับ 7 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลมากกกว่า 50 เดซิเบล คือได้ยินเสียงปกติ จำนวน 8 คน ถ้าผลน้อยกว่าเท่ากับ 50 เดซิเบล ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = อัตราความเสี่ยงการเป็นโรคหูชั้นกลางอักเสบ ถ้าผลเท่ากับ N/A (คือไม่มีความเสี่ยง) คือได้ยินเสียงปกติ จำนวน 0 คน ถ้าผลเท่ากับ low (คือความเสี่ยงน้อย) ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลมากกกว่า 40 เดซิเบล คือได้ยินเสียงไม่ปกติ จำนวน 2 คน ถ้าผลน้อยกว่าเท่ากับ 40 เดซิเบล คือได้ยินเสียงปกติ จำนวน 13 คน ถ้าผลเท่ากับ high (คือความเสี่ยงมาก) ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลมากกกว่า 45 เดซิเบล คือได้ยินเสียงปกติ จำนวน 5 คน ถ้าผลน้อยกว่าเท่ากับ 45 เดซิเบล ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลมากกว่า 2 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 4 คน ถ้าผลน้อยกว่าเท่ากับ 2 ครั้ง คือได้ยินเสียงปกติ จำนวน 10 คน ถ้าผลน้อยกว่าเท่ากับ 30 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = อายุ ถ้าผลน้อยกว่าเท่ากับ 18 เดือน คือได้ยินเสียงไม่ปกติ จำนวน 30 คน ถ้าผลมากกว่า 18 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าเท่ากับ 1 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลมากกกว่า 50 เดซิเบล คือได้ยินเสียงปกติ จำนวน 2 คน ถ้าผลน้อยกว่าเท่ากับ 50 เดซิเบล คือได้ยินเสียงไม่ปกติ จำนวน 18 คน ถ้ามากกว่า 1 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = สัญญาณในการกระตุ้น ถ้าผลเท่ากับ coherent (คือสัญญาณต่อเนื่อง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลมากกว่า 50 เดซิเบล คือได้ยินเสียงไม่ปกติ จำนวน 8 คน ถ้าผลน้อยกว่าเท่ากับ 50 เดซิเบล ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลมากกว่า 45 เดซิเบล คือได้ยินเสียงปกติ จำนวน 2 คน ถ้าผลน้อยกว่าเท่ากับ 45 เดซิเบล ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลมากกว่า 40 เดซิเบล คือได้ยินเสียงไม่ปกติ จำนวน 3 คน ถ้าผลน้อยกว่าเท่ากับ 40 เดซิเบล คือได้ยินเสียงปกติ จำนวน 10 คน ถ้าผลเท่ากับ incoherent (คือสัญญาณไม่ต่อเนื่อง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = อัตราความเสี่ยงการเป็นโรคหูชั้นกลางอักเสบ ถ้าผลเท่ากับ N/A (ไม่มีความเสี่ยง) คือได้ยินเสียงปกติ จำนวน 0 คน ถ้าผลเท่ากับ high (คือความเสี่ยงมาก) คือได้ยินเสียงไม่ปกติ จำนวน 6 คน ถ้าผลเท่ากับ low (คือความเสี่ยงน้อย) ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลน้อยกว่าเท่ากับ 35 เดซิเบล คือได้ยินเสียงปกติ จำนวน 7 คน ถ้าผลมากกว่า 35 เดซิเบล ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลน้อยกว่าเท่ากับ 50 เดซิเบล คือได้ยินเสียงไม่ปกติ จำนวน 15 คน ถ้าผลมากกว่า 50 เดซิเบล ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าเท่ากับ 2 ครั้ง คือได้ยินเสียงปกติ จำนวน 8 คน ถ้าผลมากกว่า 2 ครั้ง คือได้ยินเสียงม่ปกติ จำนวน 2 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.40 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

จากรูป 4.40 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 165 คน มีค่าทำนายความถูกต้องจำนวน 112 คน คิดเป็นร้อยละ 67.8788 โดยมีค่าความไว = 0.6629 ค่าความจำเพาะ = 0.6974 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4909)^2$ = 0.2410

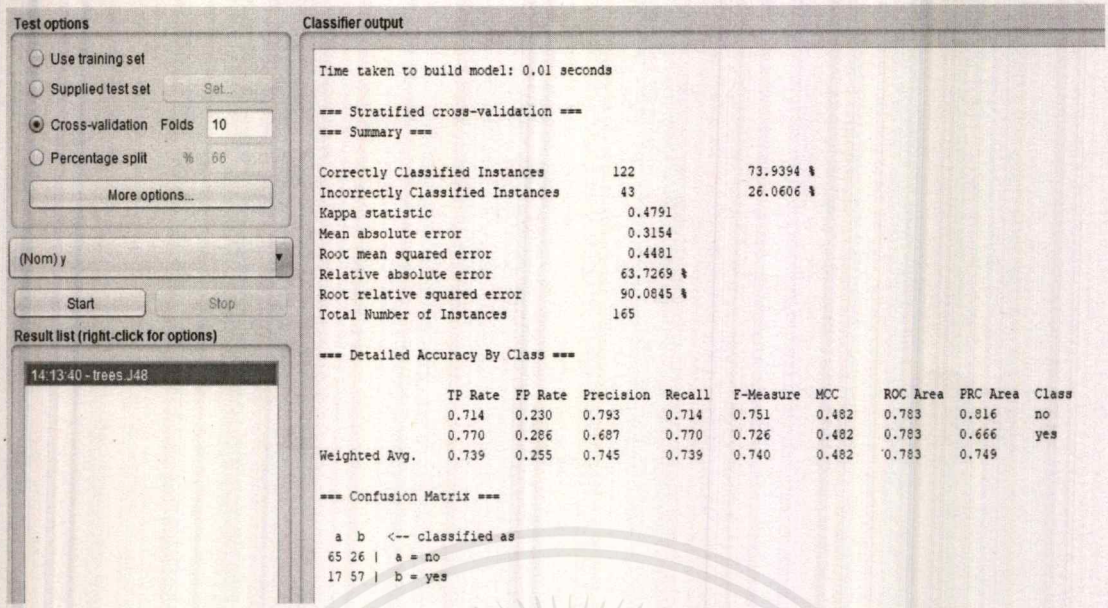


รูปที่ 4.41 ต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูป 4.41 แสดงต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โหนดรากแบ่งโดยตัวแปรทำนายคือ X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าเท่ากับ 1 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 24 คน ถ้าผลมากกว่า 1 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = อายุ ถ้าผลน้อยกว่าเท่ากับ 18 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าเท่ากับ 4 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 20 คน ถ้าผลมากกว่า 4 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลน้อยกว่าเท่ากับ 45 เดซิเบล คือได้ยินเสียงปกติ จำนวน 4 คน ถ้าผลมากกว่า 45 เดซิเบล คือได้ยินเสียงไม่ปกติ จำนวน 3 คน ถ้าผลมากกว่า 18 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 =จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลมากกว่า 5 ครั้ง คือได้ยินเสียงไม่ปกติ จำานวน 21 คน ถ้าผลน้อยกว่าเท่ากับ 5 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = อายุ ถ้าผลมากกว่า 30 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลมากกว่า 35 เดซิเบล คือได้ยินเสียงปกติ จำนวน 29 คน ถ้าผลน้อยกว่าเท่ากับ 35 เดซิเบล ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = สัญญาณในการกระตุ้น ถ้าผลเท่ากับ incoherent คือได้ยินเสียงปกติ จำนวน 2 คน ถ้าผลเท่ากับ coherent ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 =จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลมากกว่า 2 ครั้ง คือได้ยินเสียงไม่ปกติ จำานวน 4 คน ถ้าผลน้อยกว่าเท่ากับ 2 ครั้ง คือได้ยินเสียงปกติ จำนวน 3 คน ถ้าผลน้อยกว่าเท่ากับ 30 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลมากกว่า 50 เดซิเบล ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = สัญญาณในการกระตุ้น ถ้าผลเท่ากับ coherent คือได้ยินเสียงไม่ปกติ จำนวน 4 คนถ้าผลเท่ากับ incoherent ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 =จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลมากกว่า 2 ครั้ง คือได้ยินเสียงไม่ปกติ จำานวน 2 คน ถ้าผลน้อยกว่าเท่ากับ 2 ครั้ง คือได้ยินเสียงปกติ จำนวน 4 คน ถ้าผลน้อยกว่าเท่ากับ 50 เดซิเบล ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = อัตราความเสี่ยงการเป็นโรคหูชั้นกลางอักเสบ ถ้าผลเท่ากับ N/A คือได้ยินเสียงปกติ จำนวน 0 คน ถ้าผลเท่ากับ low คือได้ยินเสียงปกติ จำนวน 36 คน ถ้าผลเท่ากับ high ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลมากกว่า 45 เดซิเบล คือได้ยินเสียงปกติ จำนวน 5 คน ถ้าผลน้อยกว่าเท่ากับ 45 เดซิเบล คือได้ยินเสียงไม่ปกติ จำนวน 4 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

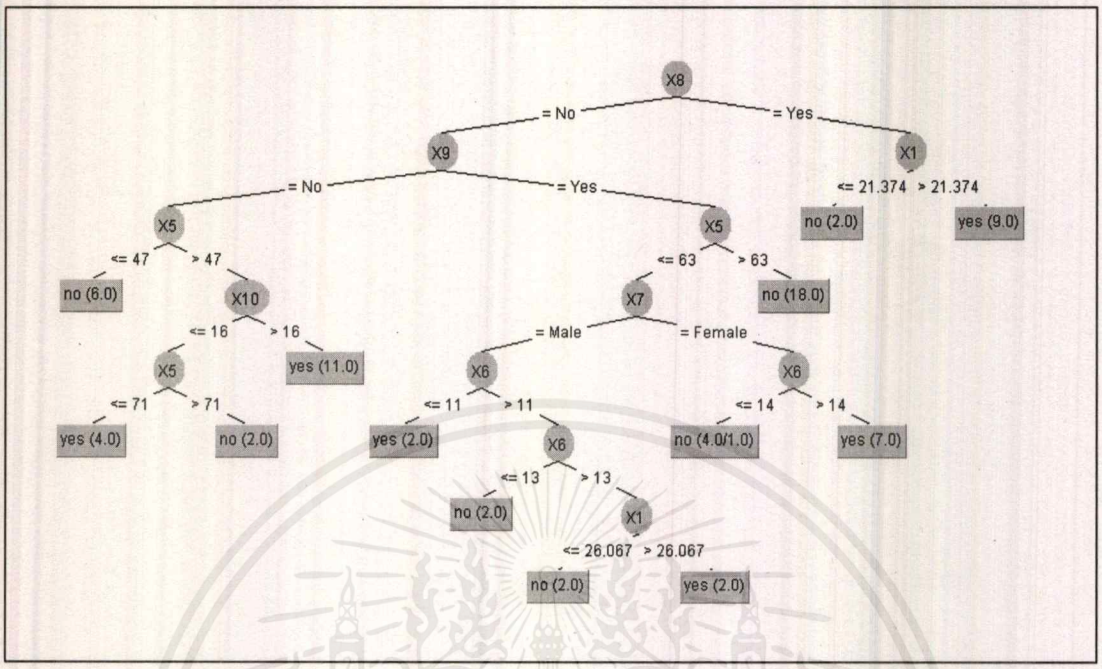


รูปที่ 4.42 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูป 4.42 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 165 คน มีค่าทำนายความถูกต้องจำนวน 122 คน คิดเป็นร้อยละ 73.9394 โดยมีค่าความไว = 0.7143 ค่าความจำเพาะ = 0.7703 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4481)^2$ = 0.2008

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

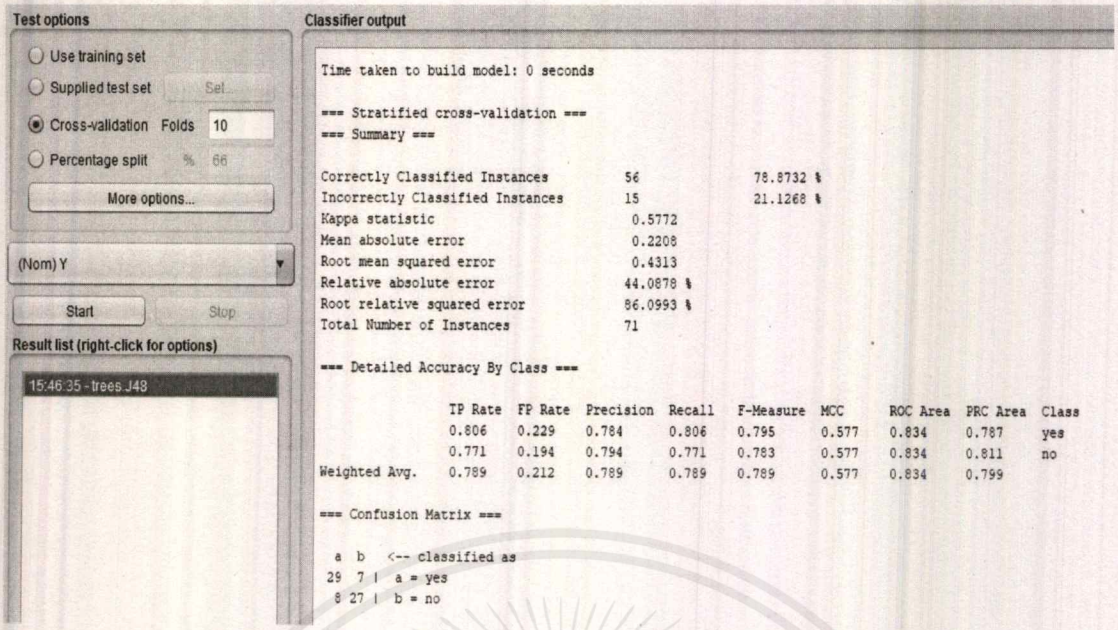
4.3.1.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทดสอบ 71 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.43 ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10

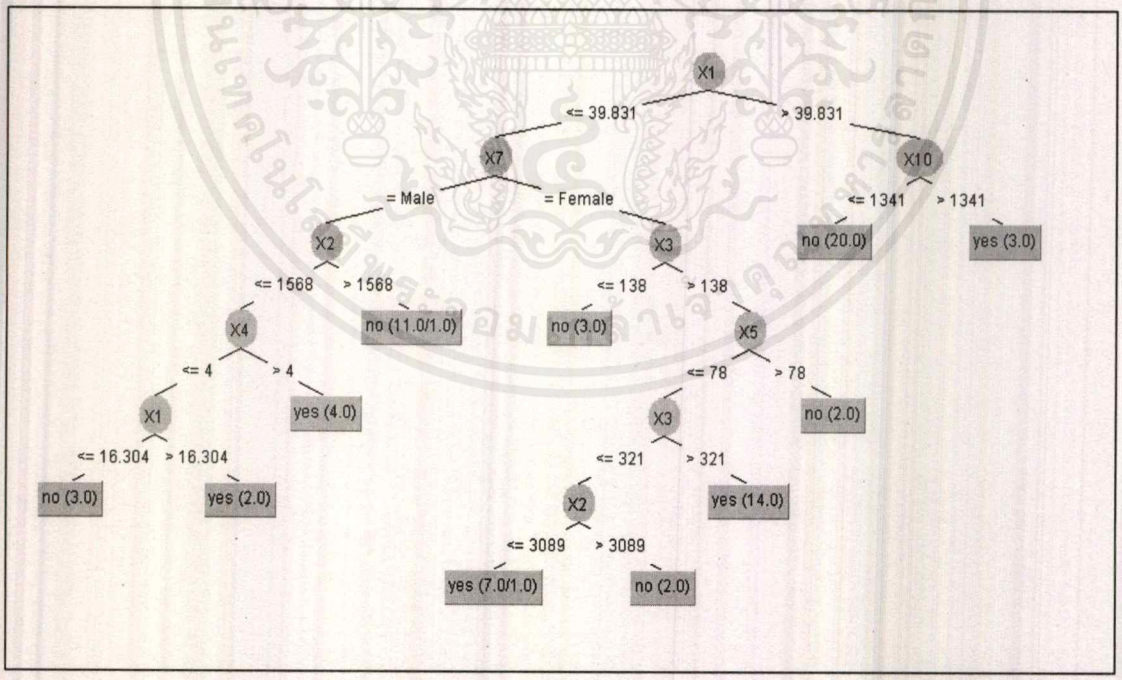
จากรูป 4.43 แสดงต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า โหนดรากแบ่งโดยตัวแปรทำนายคือ X_8 = การศึกษา ถ้าผลเท่ากับ yes (คือกำลังศึกษา) ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = รายได้ต่อเดือน ถ้าผลน้อยกว่าเท่ากับ 21.374 ดอลลาร์ คือไม่ชำระ จำนวน 2 คน ถ้าผลมากกว่า 21.374 ดอลลาร์ คือชำระ จำนวน 9 คน ถ้าผลเท่ากับ no (คือสำเร็จการศึกษา) ตัดสินใจต่อด้วยตัวแปรทำนาย X_9 = สถานภาพสมรส ถ้าผลเท่ากับ no (คือโสด) ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = อายุ ถ้าผลน้อยกว่าเท่ากับ 47 ปี คือไม่ชำระ จำนวน 6 คน ถ้าผลมากกว่า 10 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_{10} = ยอดเงินคงเหลือในบัตรเครดิต ถ้าผลมากกว่า 16 ดอลลาร์ คือชำระ จำนวน 11 คน ถ้าผลน้อยกว่าเท่ากับ 16 ดอลลาร์ ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = อายุ ถ้าผลน้อยกว่าเท่ากับ 71 ปี คือชำระ จำนวน 4 คน ถ้าผลมากกว่า 71 ปี คือไม่ชำระ จำนวน 2 คน ถ้าผลเท่ากับ yes (แต่งงานแล้ว) ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = อายุ ถ้าผลมากกว่า 63 ปี คือไม่ชำระ จำนวน 18 คน ถ้าผลน้อยกว่าเท่ากับ 63 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_7 = เพศ ถ้าผลเท่ากับ Female (เพศหญิง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = จำนวนปีในการศึกษา ถ้าผลน้อยกว่าเท่ากับ 14 ปี คือไม่ชำระ จำนวน 4 คน ถ้าผลมากกว่า 14 ปี คือชำระ จำนวน 7 คน ถ้าผลเท่ากับ Male (เพศชาย) ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = จำนวนปีในการศึกษา ถ้าผลน้อยกว่าเท่ากับ 11 ปี คือชำระ จำนวน 2 คน ถ้าผลมากกว่า 11 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = จำนวนปีในการศึกษา ถ้าผลน้อยกว่าเท่ากับ 13 ปี คือไม่ชำระ จำนวน 2 คน ถ้าผลมากกว่า 13 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = รายได้ต่อเดือน ถ้าผลน้อยกว่าเท่ากับ 26.067 ดอลลาร์ คือไม่ชำระ จำนวน 2 คน ถ้าผลมากกว่า 26.067 ดอลลาร์ คือชำระ จำนวน 2 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.44 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10

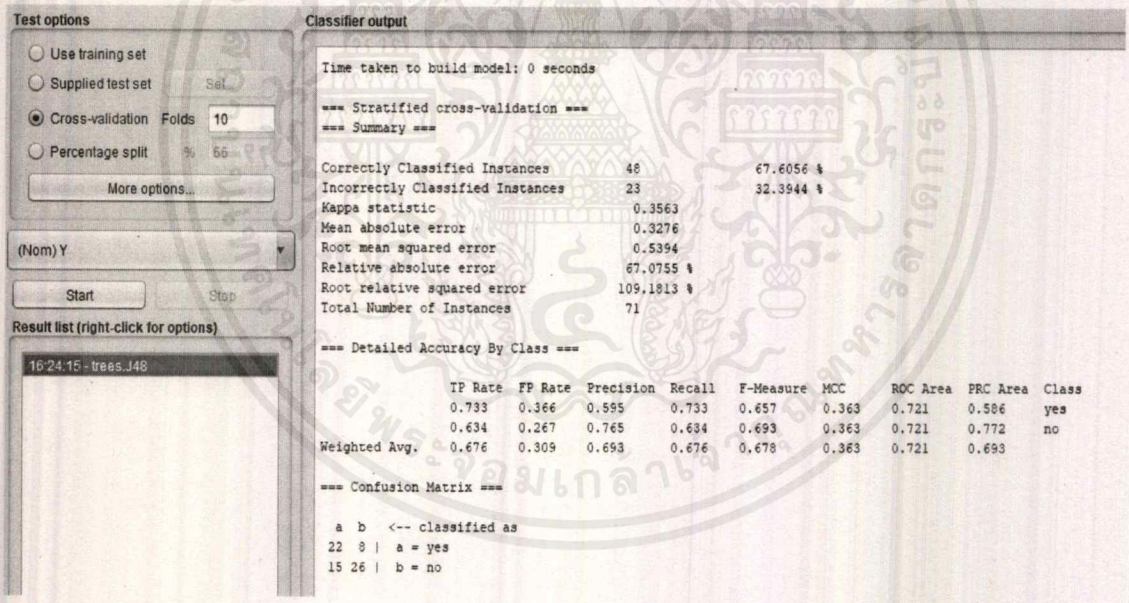
จากรูป 4.44 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 71 คน มีค่าทำนายความถูกต้องจำนวน 56 คน คิดเป็นร้อยละ 78.8732 โดยมีค่าความไว = 0.8056 ค่าความจำเพาะ = 0.7714 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4313)^2$ = 0.1860



รูปที่ 4.45 ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

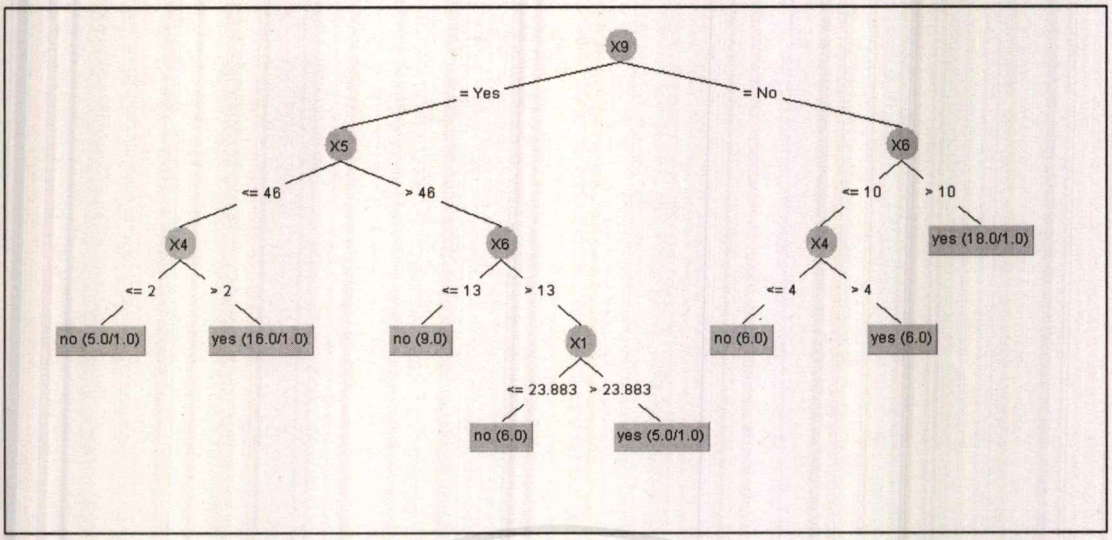
จากรูป 4.45 แสดงต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า โหนดราก แบ่งโดยตัวแปรทำนายคือ X_1 = รายได้ต่อเดือน ถ้าผลมากกว่า 39.831 ดอลลาร์ ตัดสินใจต่อด้วยตัวแปรทำนาย X_{10} = ยอดเงินคงเหลือในบัตรเครดิต ถ้าผลมากกว่า 1341 ดอลลาร์ คือชำระ จำนวน 3 คน ถ้าผลน้อยกว่าเท่ากับ 1341 ดอลลาร์ คือไม่ชำระ จำนวน 20 คน ถ้าผลน้อยกว่าเท่ากับ 39.831 ดอลลาร์ ตัดสินใจต่อด้วยตัวแปรทำนาย X_7 = เพศ ถ้าผลเท่ากับ Male (เพศชาย) ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = วงเงินในบัตรเครดิต ถ้าผลมากกว่า 1568 ดอลลาร์ คือไม่ชำระ จำนวน 11 คน ถ้าผลน้อยกว่าเท่ากับ 1568 ดอลลาร์ ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = จำนวนบัตรเครดิตที่มี ถ้าผลมากกว่า 4 บัตร คือชำระ จำนวน 4 คน ถ้าผลน้อยกว่าเท่ากับ 4 บัตร ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = รายได้ต่อเดือน ถ้าผลมากกว่า 16.304 ดอลลาร์ คือชำระ จำนวน 2 คน ถ้าผลน้อยกว่าเท่ากับ 16.304 ดอลลาร์ คือไม่ชำระ จำนวน 3 คน ถ้าผลเท่ากับ Female (เพศหญิง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = จำนวนครั้งในการใช้บัตรเครดิตต่อปี ถ้าผลน้อยกว่าเท่ากับ 138 ครั้ง คือไม่ชำระ จำนวน 3 คน ถ้าผลมากกว่า 138 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = อายุ ถ้าผลมากกว่า 78 ปี คือไม่ชำระ จำนวน 2 คน ถ้าผลน้อยกว่าเท่ากับ 78 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = จำนวนครั้งในการใช้บัตรเครดิตต่อปี ถ้าผลมากกว่า 321 ครั้ง คือชำระ จำนวน 14 คน ถ้าผลน้อยกว่าเท่ากับ 321 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = วงเงินในบัตรเครดิต ถ้าผลมากกว่า 3089 ดอลลาร์ คือไม่ชำระ จำนวน 2 คน ถ้าผลน้อยกว่าเท่ากับ 3089 ดอลลาร์ คือชำระ จำนวน 7 คน



รูปที่ 4.46 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

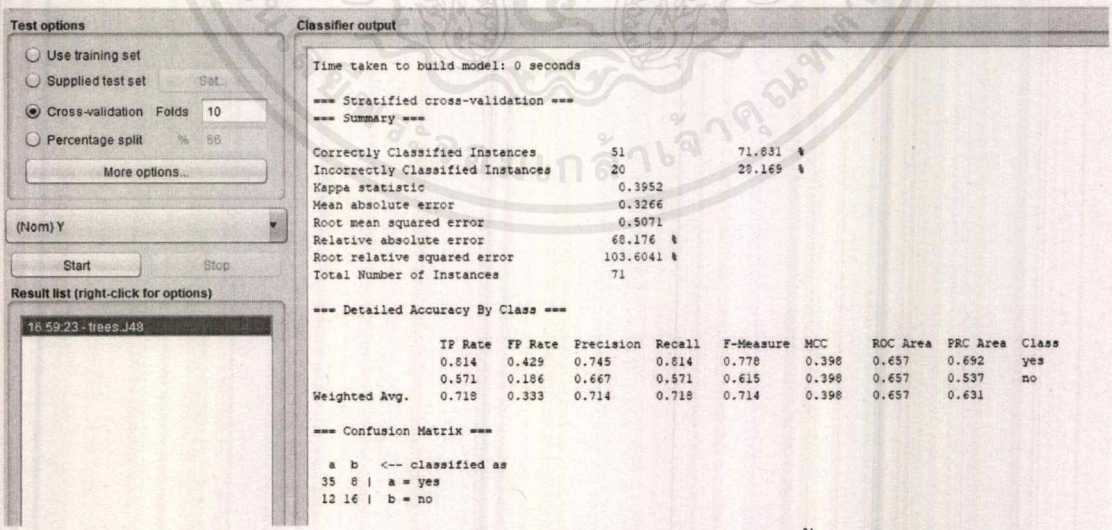
จากรูป 4.46 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 71 คน มีค่าทำนายความถูกต้องจำนวน 48 คน คิดเป็นร้อยละ 67.6056 โดยมีค่าความไว = 0.7333 ค่าความจำเพาะ = 0.6341 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5394)^2$ = 0.2910

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.47 ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูป 4.47 แสดงต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า โหนดรากแบ่งโดยตัวแปรทำนายคือ X_9 = สถานภาพสมรส ถ้าผลเท่ากับ no (คือโสด) ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = จำนวนปีในการศึกษา ถ้าผลมากกว่า 10 ปี คือชำระ จำนวน 18 คน ถ้าผลน้อยกว่าเท่ากับ 10 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = จำนวนบัตรเครดิตที่มี ถ้าผลมากกว่า 4 บัตร คือชำระ จำนวน 6 คน ถ้าผลน้อยกว่าเท่ากับ 4 บัตร คือไม่ชำระ จำนวน 6 คน ถ้าผลเท่ากับ yes (แต่งงานแล้ว) ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = อายุ ถ้าผลน้อยกว่าเท่ากับ 46 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = จำนวนบัตรเครดิตที่มี ถ้าผลมากกว่า 2 บัตร คือชำระ จำนวน 16 คน ถ้าผลน้อยกว่าเท่ากับ 2 บัตร คือไม่ชำระ จำนวน 5 คน ถ้าผลมากกว่า 46 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = จำนวนปีในการศึกษา ถ้าผลน้อยกว่าเท่ากับ 13 ปี คือไม่ชำระ จำนวน 9 คน ถ้าผลมากกว่า 13 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = รายได้ต่อเดือน ถ้าผลมากกว่า 23.883 ดอลลาร์ คือชำระ จำนวน 5 คน ถ้าผลน้อยกว่าเท่ากับ 23.883 ดอลลาร์ คือไม่ชำระ จำนวน 6 คน

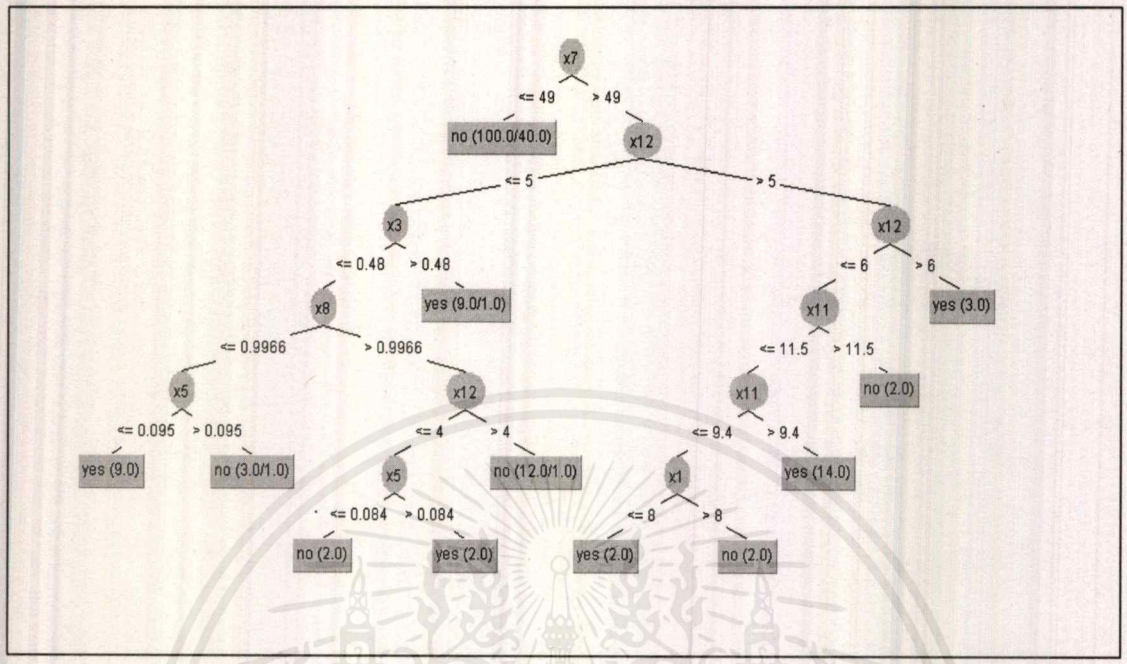


รูปที่ 4.48 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48
ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูป 4.48 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 71 คน มีค่าทำนายความถูกต้องจำนวน 51 คน คิดเป็นร้อยละ 71.831 โดยมีค่าความไว = 0.8140 ค่าความจำเพาะ = 0.5714 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5071)^2$ = 0.2572

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

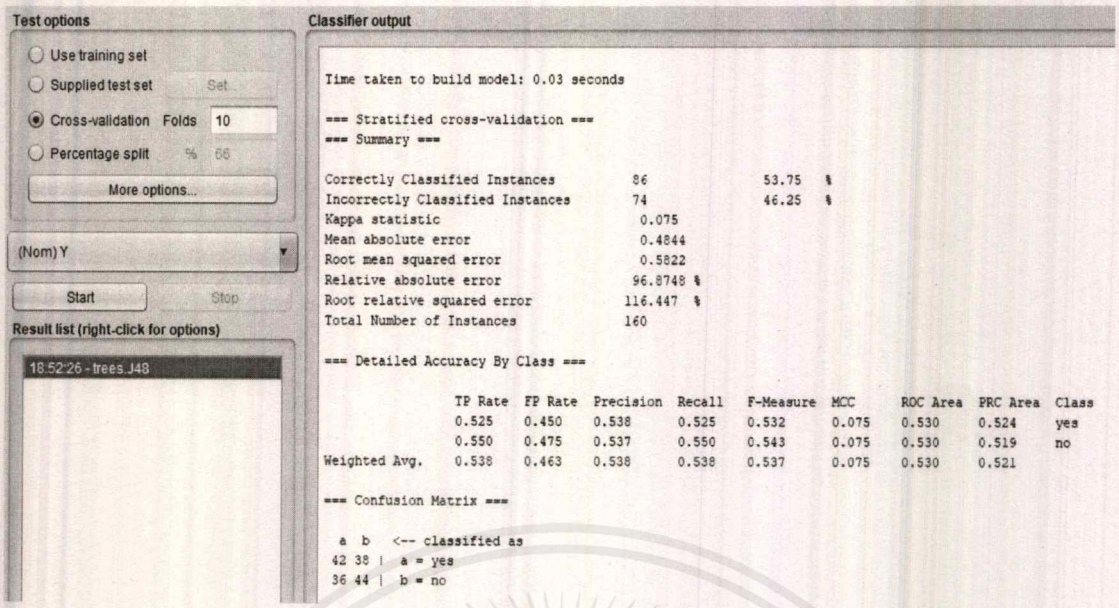
4.3.1.3 คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทดสอบ 160 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.49 ต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 10

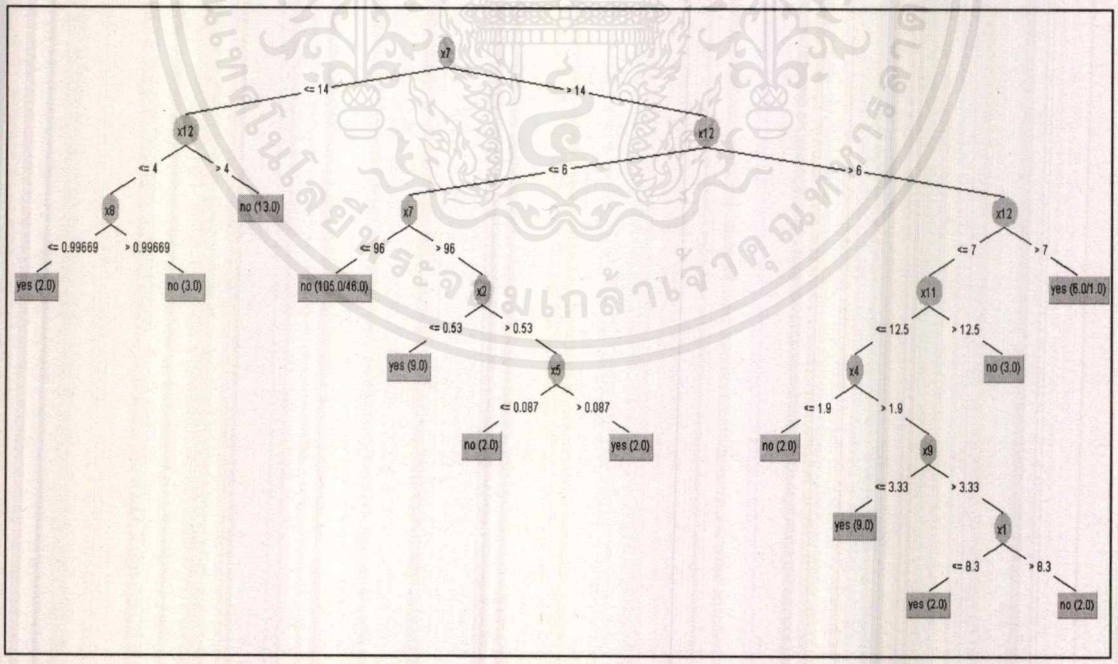
จากรูป 4.49 แสดงต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง โหนดรากแบ่งโดยตัวแปรทำนายคือ X_7 = ค่าซัลเฟอร์ไดออกไซด์ทั้งหมด ถ้าผลน้อยกว่าเท่ากับ 49 กรัม คือคุณภาพไม่ดี จำนวน 100 ขวด ถ้าผลมากกว่า 49 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_{12} = คะแนนคุณภาพ ถ้าผลมากกว่า 5 คะแนน ตัดสินใจต่อด้วยตัวแปรทำนาย X_{12} = คะแนนคุณภาพ ถ้าผลมากกว่า 6 คะแนน คือคุณภาพดี จำนวน 3 ขวด ถ้าผลน้อยกว่าเท่ากับ 6 คะแนน ตัดสินใจต่อด้วยตัวแปรทำนาย X_{11} = ค่าแอลกอฮอล์ ถ้าผลมากกว่า 11.5 กรัม คือคุณภาพไม่ดี จำนวน 2 ขวด ถ้าผลน้อยกว่าเท่ากับ 11.5 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_{11} = ค่าแอลกอฮอล์ ถ้าผลมากกว่า 9.4 กรัม คือคุณภาพดี จำนวน 14 ขวด ถ้าผลน้อยกว่าเท่ากับ 9.4 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = ค่าความเป็นกรด ถ้าผลมากกว่า 8 pH คือคุณภาพไม่ดี จำนวน 2 ขวด ถ้าผลน้อยกว่าเท่ากับ 8 pH คือคุณภาพดี จำนวน 2 ขวด ถ้าผลน้อยกว่าเท่ากับ 5 คะแนน ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ค่ากรดซิตริก ถ้าผลมากกว่า 0.48 pH คือคุณภาพดี จำนวน 9 ขวด ถ้าผลน้อยกว่าเท่ากับ 0.48 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_8 = ค่าความหนาแน่น ถ้าผลน้อยกว่าเท่ากับ 0.9966 กิโลกรัมต่อลูกบาศก์เมตร ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = ค่าคลอไรด์ ถ้าผลน้อยกว่าเท่ากับ 0.095 กรัม คือคุณภาพดี จำนวน 9 ขวด ถ้าผลมากกว่า 0.095 กรัม คือคุณภาพไม่ดี จำนวน 3 ขวด ถ้าผลมากกว่า 0.9966 กิโลกรัมต่อลูกบาศก์เมตร ตัดสินใจต่อด้วยตัวแปรทำนาย X_{12} = คะแนนคุณภาพ ถ้าผลมากกว่า 4 คะแนน คือคุณภาพไม่ดี จำนวน 12 ขวด ถ้าผลน้อยกว่าเท่ากับ 4 คะแนน ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = ค่าคลอไรด์ ถ้าผลน้อยกว่าเท่ากับ 0.084 กรัม คือคุณภาพไม่ดี จำนวน 2 ขวด ถ้าผลมากกว่า 0.084 กรัม คือคุณภาพดี จำนวน 2 ขวด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.50 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 10

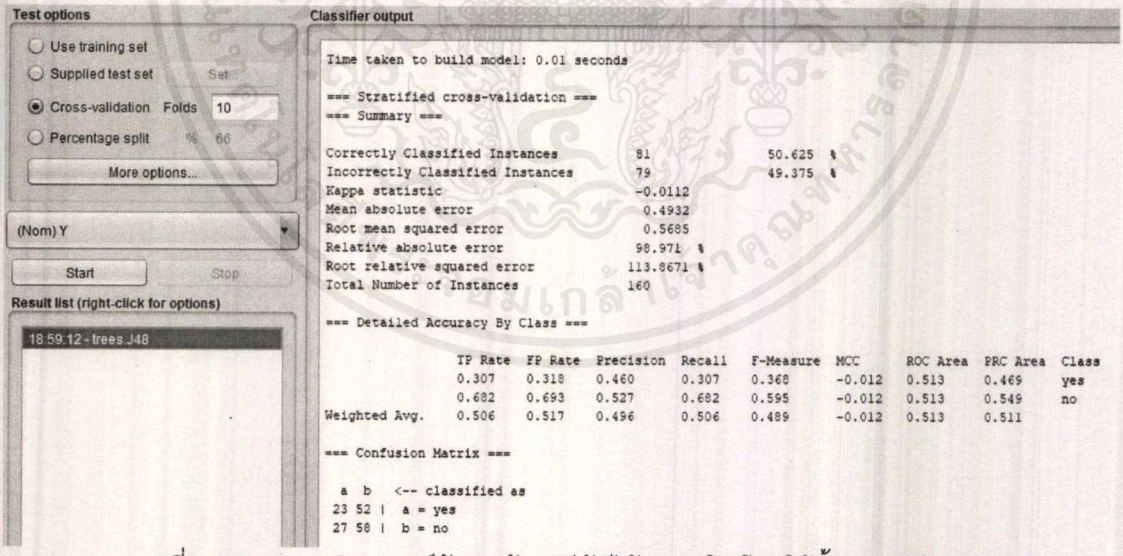
จากรูป 4.50 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 160 ขวด มีค่าทำนายความถูกต้องจำนวน 86 ขวด คิดเป็นร้อยละ 53.75 โดยมีค่าความไว = 0.5250 ค่าความจำเพาะ = 0.5500 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5822)^2$ = 0.3390



รูปที่ 4.51 ต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

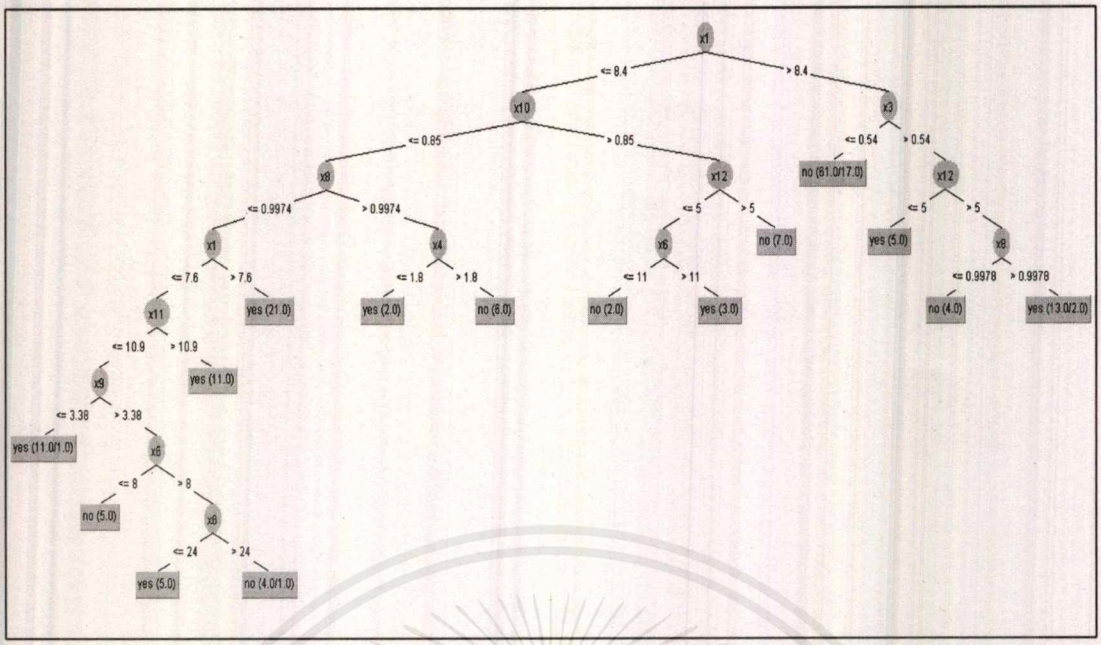
จากรูป 4.51 แสดงต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง โหนดรากแบ่งโดยตัวแปรทำนาย คือ X_7 = ค่าซัลเฟอร์ไดออกไซด์ทั้งหมด ถ้าผลน้อยกว่าเท่ากับ 14 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_{12} = คะแนนคุณภาพ ถ้าผลมากกว่า 4 คะแนน คือคุณภาพไม่ดี จำนวน 3 ขวด ถ้าผลน้อยกว่าเท่ากับ 4 คะแนน ตัดสินใจต่อด้วยตัวแปรทำนาย X_8 = ค่าความหนาแน่น ถ้าผลน้อยกว่าเท่ากับ 0.99669 กิโลกรัมต่อลูกบาศก์เมตร คือคุณภาพดี จำนวน 2 ขวด ถ้าผลมากกว่า 0.99669 กิโลกรัมต่อลูกบาศก์เมตร คือคุณภาพไม่ดี จำนวน 3 ขวด ถ้าผลมากกว่าเท่ากับ 14 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_{12} = คะแนนคุณภาพ ถ้าผลน้อยกว่าเท่ากับ 6 คะแนน ตัดสินใจต่อด้วยตัวแปรทำนาย X_7 = ค่าซัลเฟอร์ไดออกไซด์ทั้งหมด ถ้าผลน้อยกว่าเท่ากับ 96 กรัม คือคุณภาพไม่ดี จำนวน 105 ขวด ถ้าผลมากกว่าเท่ากับ 96 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 =ค่าความเป็นกรดระเหย ถ้าผลน้อยกว่าเท่ากับ 0.53 pH คือคุณภาพดี จำนวน 9 ขวด ถ้าผลมากกว่า 0.53 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = ค่าคลอไรด์ ถ้าผลน้อยกว่าเท่ากับ 0.087 กรัม คือคุณภาพไม่ดี จำนวน 2 ขวด ถ้าผลมากกว่า 0.087 กรัม คือคุณภาพดี จำนวน 2 ขวด ถ้าผลมากกว่า 6 คะแนน ตัดสินใจต่อด้วยตัวแปรทำนาย X_{12} = คะแนนคุณภาพ ถ้าผลมากกว่า 7 คะแนน คือคุณภาพดี จำนวน 6 ขวด ถ้าผลน้อยกว่าเท่ากับ 7 คะแนน ตัดสินใจต่อด้วยตัวแปรทำนาย X_{11} = ค่าแอลกอฮอล์ ถ้าผลมากกว่า 12.5 กรัม คือคุณภาพไม่ดี จำนวน 3 ขวด ถ้าผลน้อยกว่าเท่ากับ 12.5 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = ค่าน้ำตาลรีดิวซ์ ถ้าผลน้อยกว่าเท่ากับ 1.9 กรัม คือคุณภาพไม่ดี จำนวน 2 ขวด ถ้าผลมากกว่า 1.9 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_9 = ค่าความเป็นกรดหรือด่าง ถ้าผลน้อยกว่าเท่ากับ 3.33 pH คือคุณภาพไม่ดี จำนวน 9 ขวด ถ้าผลมากกว่า 3.33 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = ค่าความเป็นกรด ถ้าผลน้อยกว่าเท่ากับ 8.3 pH คือคุณภาพดี จำนวน 2 ขวด ถ้าผลมากกว่า 8.3 pH คือคุณภาพไม่ดี จำนวน 2 ขวด



รูปที่ 4.52 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 20

จากรูป 4.52 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 160 ขวด มีค่าทำนายความถูกต้องจำนวน 81 ขวด คิดเป็นร้อยละ 50.625 โดยมีค่าความไว = 0.3067 ค่าความจำเพาะ = 0.6824 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5685)^2$ = 0.3232

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

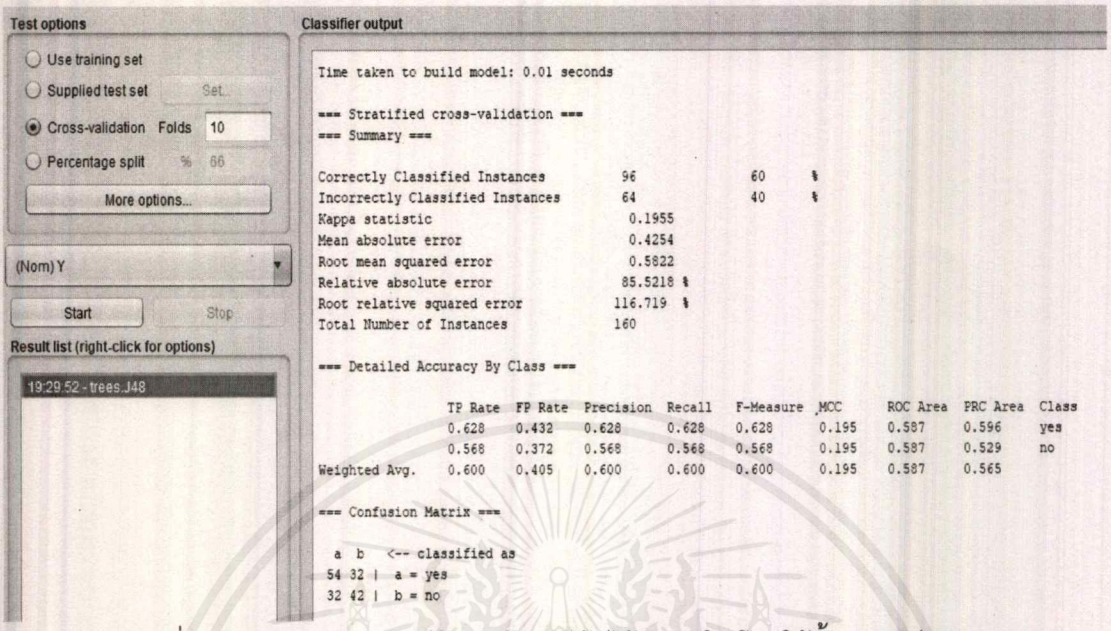


รูปที่ 4.53 ต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูป 4.53 แสดงต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง โหนดรากแบ่งโดยตัวแปรทำนาย คือ X_1 = ค่าความเป็นกรด ถ้าผลมากกว่า 8.4 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ค่ากรดซิตริก ถ้าผลน้อยกว่าเท่ากับ 0.54 pH คือคุณภาพไม่ดี จำนวน 61 ขวด ถ้าผลมากกว่า 0.54 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_{12} = คะแนนคุณภาพ ถ้าผลน้อยกว่าเท่ากับ 5 คะแนน คือคุณภาพดี จำนวน 5 ขวด ถ้าผลมากกว่า 5 คะแนน ตัดสินใจต่อด้วยตัวแปรทำนาย X_8 = ค่าความหนาแน่น ถ้าผลน้อยกว่าเท่ากับ 0.9978 กิโลกรัมต่อลูกบาศก์เมตร คือคุณภาพไม่ดี จำนวน 4 ขวด ถ้าผลมากกว่า 0.9978 กิโลกรัมต่อลูกบาศก์เมตร คือคุณภาพดี จำนวน 13 ขวด ถ้าผลน้อยกว่าเท่ากับ 8.4 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_{10} = ค่าเกลือของกรดซัลฟิวริก ถ้าผลมากกว่า 0.85 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_{12} = คะแนนคุณภาพ ถ้าผลมากกว่า 5 คะแนน คือคุณภาพไม่ดี จำนวน 7 ขวด ถ้าผลน้อยกว่าเท่ากับ 5 คะแนน ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = ค่าซัลเฟอร์ไดออกไซด์ ถ้าผลมากกว่า 11 กรัม คือคุณภาพดี จำนวน 3 ขวด ถ้าผลน้อยกว่าเท่ากับ 11 กรัม คือคุณภาพไม่ดี จำนวน 2 ขวด ถ้าผลน้อยกว่าเท่ากับ 0.85 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_8 = ค่าความหนาแน่น ถ้าผลมากกว่า 0.9974 กิโลกรัมต่อลูกบาศก์เมตร ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = ค่าน้ำตาลรีดิวซ์ ถ้าผลน้อยกว่าเท่ากับ 1.8 กรัม คือคุณภาพดี จำนวน 2 ขวด ถ้าผลมากกว่า 1.8 กรัม คือคุณภาพไม่ดี จำนวน 6 ขวด ถ้าผลน้อยกว่าเท่ากับ 0.9974 กิโลกรัมต่อลูกบาศก์เมตร ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = ค่าความเป็นกรด ถ้าผลมากกว่า 7.6 pH คือคุณภาพดี จำนวน 21 ขวด ถ้าผลน้อยกว่าเท่ากับ 7.6 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_{11} = ค่าแอลกอฮอล์ ถ้าผลมากกว่า 10.9 กรัม คือคุณภาพดี จำนวน 11 ขวด ถ้าผลน้อยกว่าเท่ากับ 10.9 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_9 = ค่าความเป็นกรดหรือด่าง ถ้าผลน้อยกว่าเท่ากับ 3.38 pH คือคุณภาพดี จำนวน 11 ขวด ถ้าผลมากกว่า 3.38 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = ค่าซัลเฟอร์ไดออกไซด์ ถ้าผลน้อยกว่าเท่ากับ 8 กรัม คือคุณภาพไม่ดี จำนวน 5 ขวด ถ้าผลมากกว่า 8 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = ค่าซัลเฟอร์ ได

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ออกไซด์ ถ้าผลมากกว่า 24 กรัม คือคุณภาพไม่ดี จำนวน 4 ขวด ถ้าผลน้อยกว่าเท่ากับ 24 กรัม คือคุณภาพดี จำนวน 5 ขวด



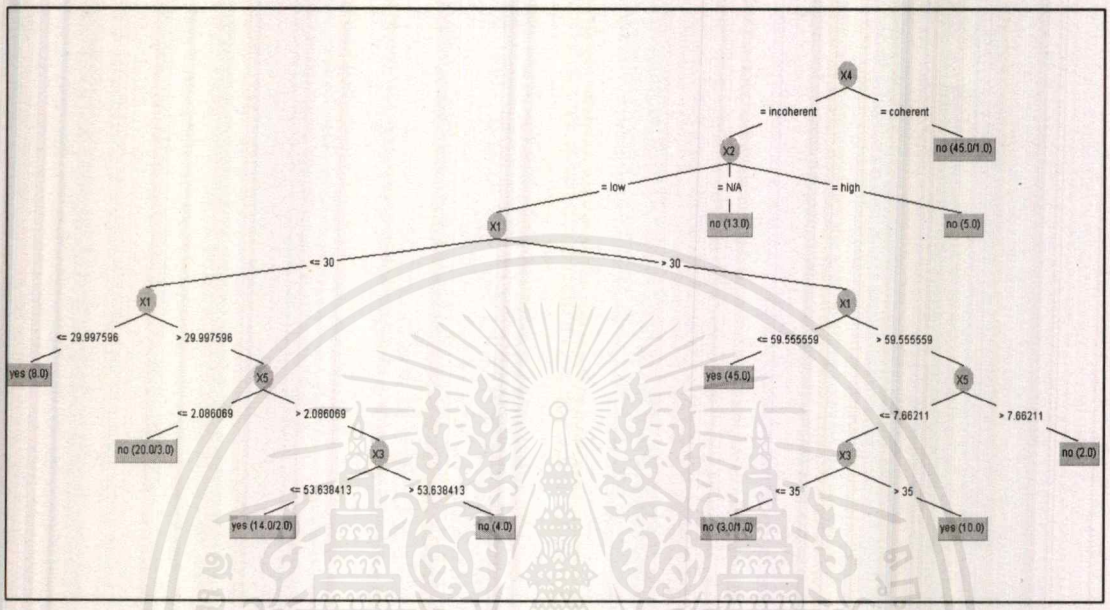
รูปที่ 4.54 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูป 4.54 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 160 ขวด มีค่าทำนายความถูกต้องจำนวน 96 ขวด คิดเป็นร้อยละ 60 โดยมีค่าความไว = 0.6279 ค่าความจำเพาะ = 0.5676 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5822)^2$ = 0.3390

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.2 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-Sampling)

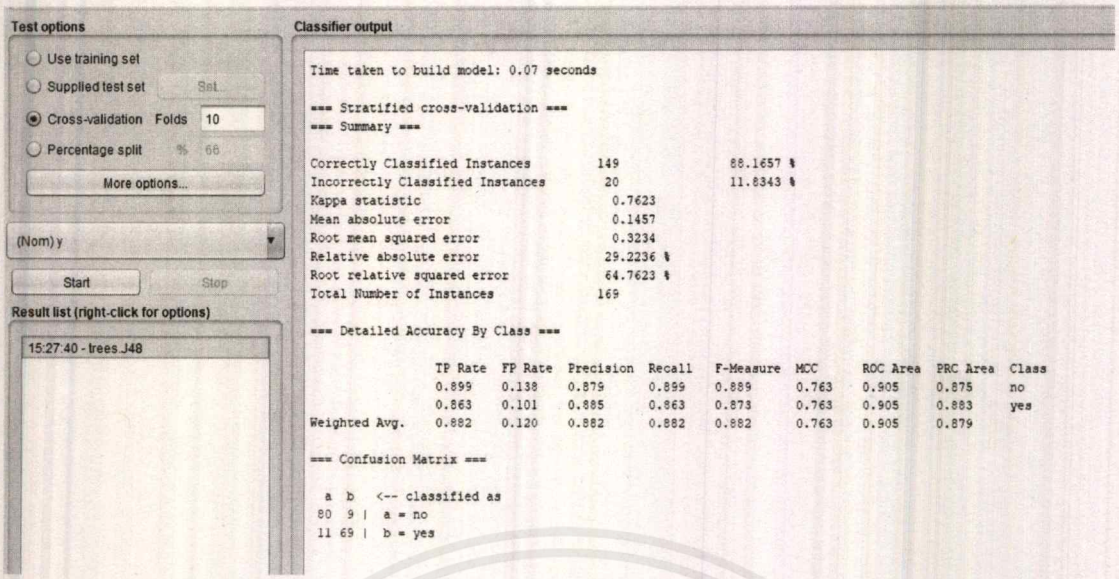
4.3.2.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทดสอบ 169 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.55 ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10

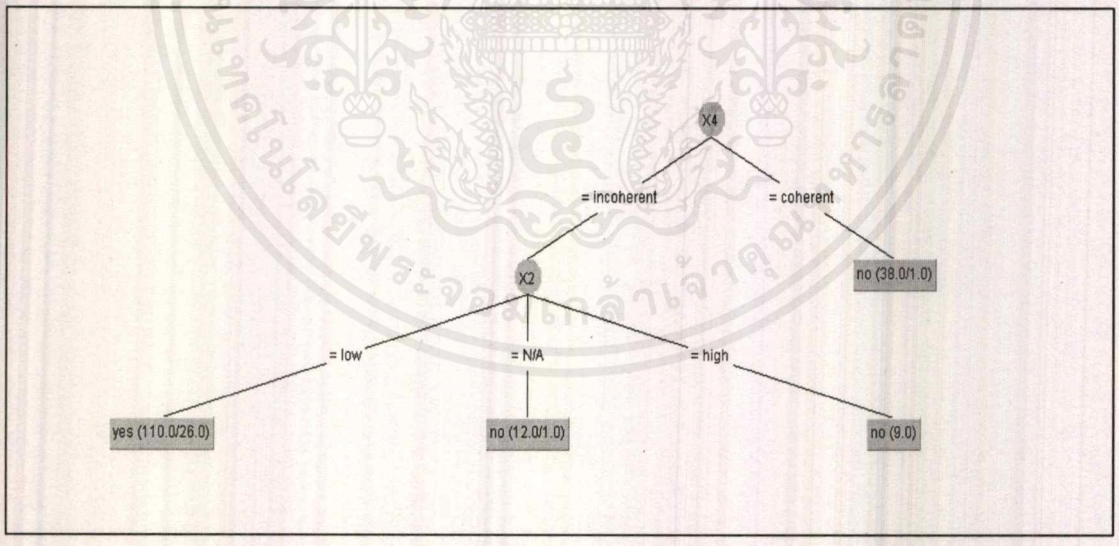
จากรูป 4.55 แสดงต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โหนดรากแบ่งโดยตัวแปรทำนาย X_4 = สัญญาณในการกระตุ้น ถ้าผลเท่ากับ coherent (สัญญาณต่อเนื่อง) คือได้ยินเสียงไม่ปกติ จำนวน 45 คน ถ้าผลเท่ากับ incoherent (สัญญาณไม่ต่อเนื่อง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = อัตราความเสี่ยงการเป็นโรคหูชั้นกลาง ถ้าผลเท่ากับ high (ความเสี่ยงมาก) คือได้ยินเสียงไม่ปกติ จำนวน 5 คน ถ้าผลเท่ากับ N/A (ไม่มีความเสี่ยง) คือได้ยินเสียงไม่ปกติ จำนวน 13 คน และถ้าผลเท่ากับ low (ความเสี่ยงน้อย) ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = อายุ ถ้าผลน้อยกว่าเท่ากับ 30 เดือน และน้อยกว่าเท่ากับ 29.997596 เดือน คือได้ยินเสียงปกติ จำนวน 8 คน แต่ถ้าผลมากกว่า 29.997596 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าเท่ากับ 2.086069 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 20 คน ถ้าผลมากกว่า 2.086069 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลน้อยกว่าเท่ากับ 53.6384123 เดซิเบล คือได้ยินเสียงปกติ จำนวน 14 คน ถ้าผลมากกว่าเท่ากับ 53.6384123 คือได้ยินเสียงไม่ปกติ จำนวน 4 คน จากตัวแปรทำนาย X_1 = อายุ ถ้าผลมากกว่า 30 เดือน และน้อยกว่าเท่ากับ59.555559 เดือน คือ ได้ยินเสียงปกติ จำนวน 45 คน แต่ถ้าผลมากกว่า 59.555559 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลมากกว่า 7.66211 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 2 คน ถ้าผลน้อยกว่า 7.66211 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลน้อยกว่าเท่ากับ 35 เดซิเบล คือได้ยินเสียงไม่ปกติ จำนวน 3 คน ถ้าผลมากกว่าเท่ากับ 35 เดซิเบล คือ ได้ยินเสียงปกติ จำนวน 10 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.56 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10

จากรูป 4.56 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 169 คน มีค่าทำนายความถูกต้องจำนวน 149 คน คิดเป็นร้อยละ 88.1657 โดยมีค่าความไว = 0.8989 ค่าความจำเพาะ = 0.8625 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.3234)^2$ = 0.1046

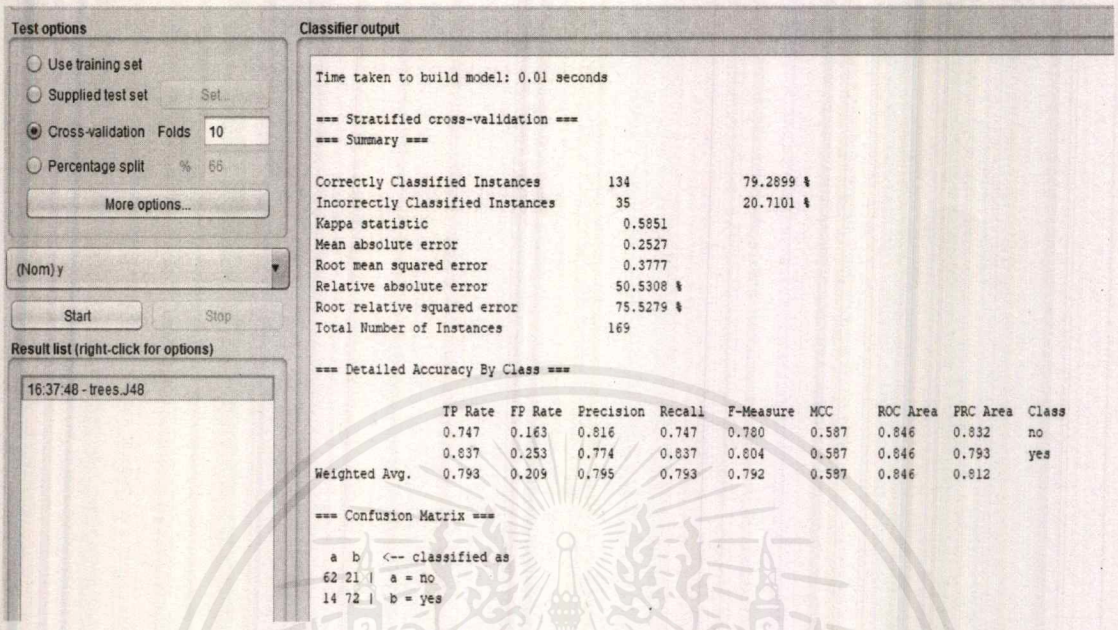


รูปที่ 4.57 ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

จากรูป 4.57 แสดงต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โหนดรากแบ่งโดยตัวแปรทำนายคือ X_4 = สัญญาณในการกระตุ้น ถ้าผลเท่ากับ coherent (สัญญาณต่อเนื่อง) คือได้ยินเสียงไม่ปกติ จำนวน 38 คน ถ้าผลเท่ากับ incoherent (สัญญาณไม่ต่อเนื่อง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = อัตราความเสี่ยงการเป็นโรคหูชั้นกลาง

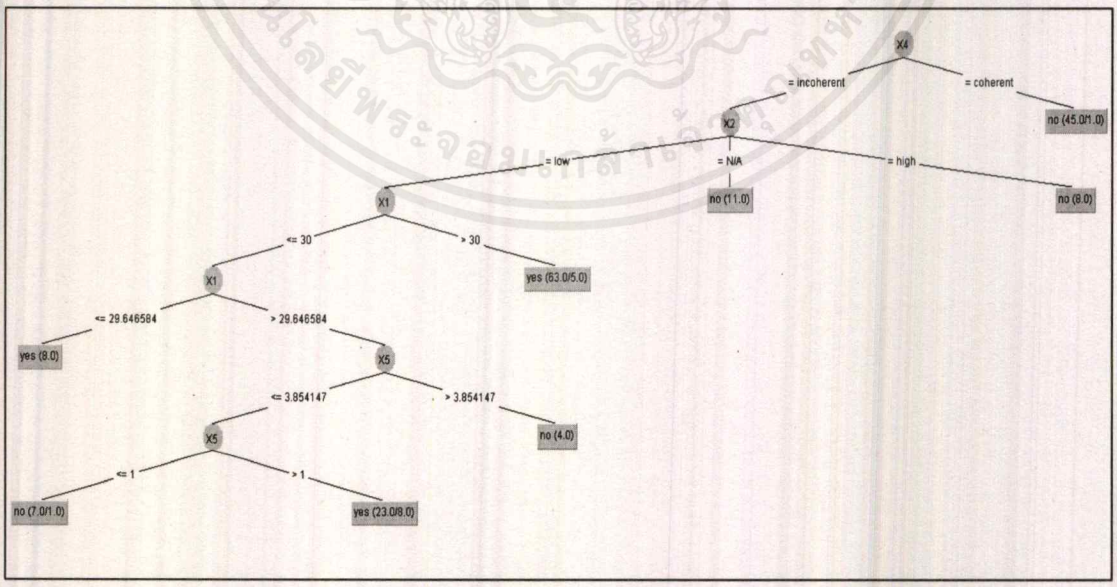
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อักเสบ ถ้าผลเท่ากับ low (ความเสี่ยงน้อย) คือได้ยินเสียงปกติ (จำนวน 110 คน) ถ้าผลเท่ากับ N/A (ไม่มีความเสี่ยง) คือ ได้ยินเสียงไม่ปกติ จำนวน 12 คน ถ้าผลเท่ากับ high (ความเสี่ยงมาก) คือได้ยินเสียงไม่ปกติ จำนวน 9 คน



รูปที่ 4.58 ผลการวิเคราะห์ข้อมูลด้วยต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

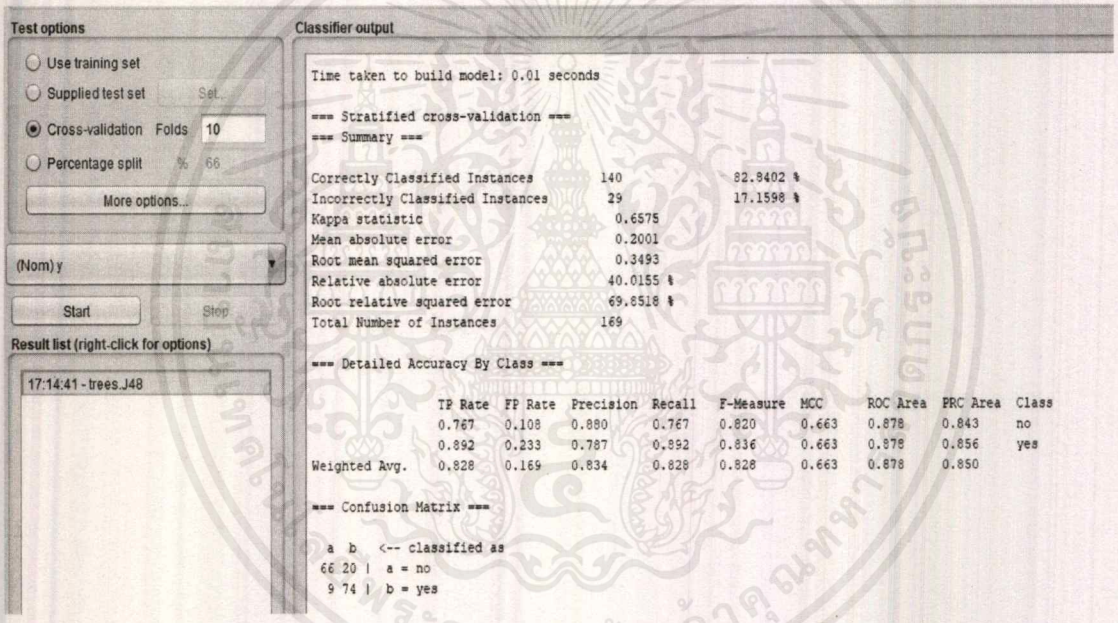
จากรูป 4.58 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 169 คน มีค่าทำนายความถูกต้องจำนวน 134 คน คิดเป็นร้อยละ 79.2899 โดยมีค่าความไว = 0.7470 ค่าความจำเพาะ = 0.8372 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.3777)^2$ = 0.1427



รูปที่ 4.59 ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูป 4.59 แสดงต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โหนดรากแบ่งโดยตัวแปรทำนายคือ แบ่งโดยตัวแปรทำนายคือ X_4 = สัญญาณในการกระตุ้น ถ้าผลเท่ากับ coherent (สัญญาณต่อเนื่อง) คือได้ยินเสียงไม่ปกติ จำนวน 45 คน ถ้าผลเท่ากับ incoherent (สัญญาณไม่ต่อเนื่อง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = อัตราความเสี่ยงการเป็นโรคหูชั้นกลางอักเสบ ถ้าผลเท่ากับ high (ความเสี่ยงมาก) คือได้ยินเสียงไม่ปกติ จำนวน 8 คน ถ้าผลเท่ากับ N/A (ไม่มีความเสี่ยง) คือได้ยินเสียงไม่ปกติ จำนวน 11 คน ถ้าผลเป็นความเสี่ยงน้อย (low) ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = อายุ ถ้าผลน้อยกว่าเท่ากับ 30 เดือน และน้อยกว่าเท่ากับ 29.646584 เดือน คือได้ยินเสียงปกติ จำนวน 8 คน แต่ถ้าผลมากกว่า 29.646584 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลมากกว่า 3.854147 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 4 คน ถ้าน้อยกว่าเท่ากับ 3.854147 ครั้ง และน้อยกว่าเท่ากับ 1 คือได้ยินเสียงไม่ปกติ จำนวน 7 คน แต่ถ้าผลมากกว่า 1 คือ ได้ยินเสียงปกติ จำนวน 23 คน

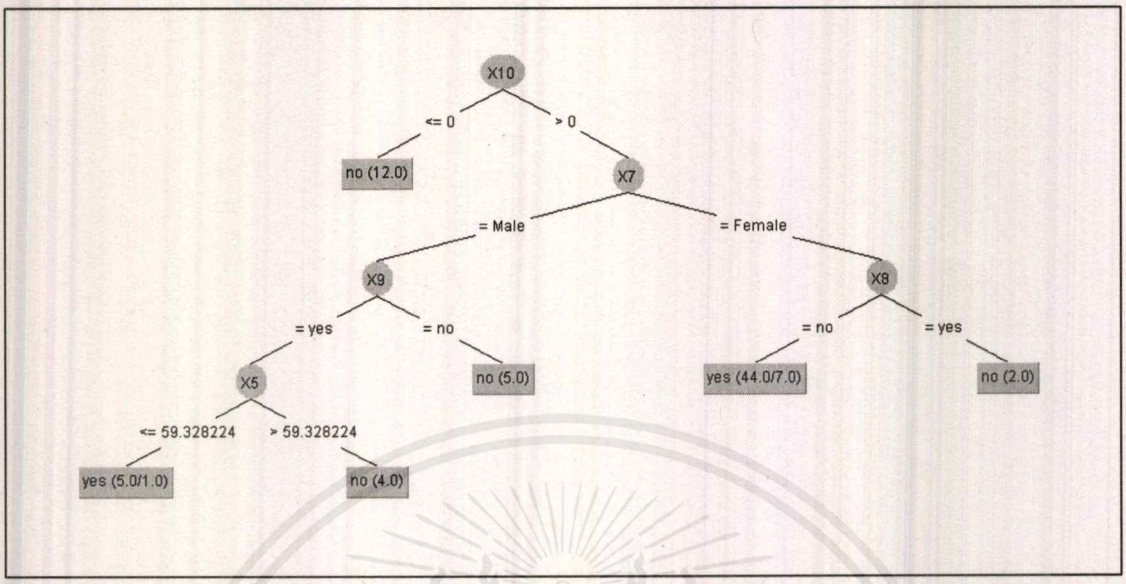


รูปที่ 4.60 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูป 4.60 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 169 คน มีค่าทำนายความถูกต้องจำนวน 140 คน คิดเป็นร้อยละ 82.8402 โดยมีค่าความไว = 0.7674 ค่าความจำเพาะ = 0.8916 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.3493)^2$ = 0.1220

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

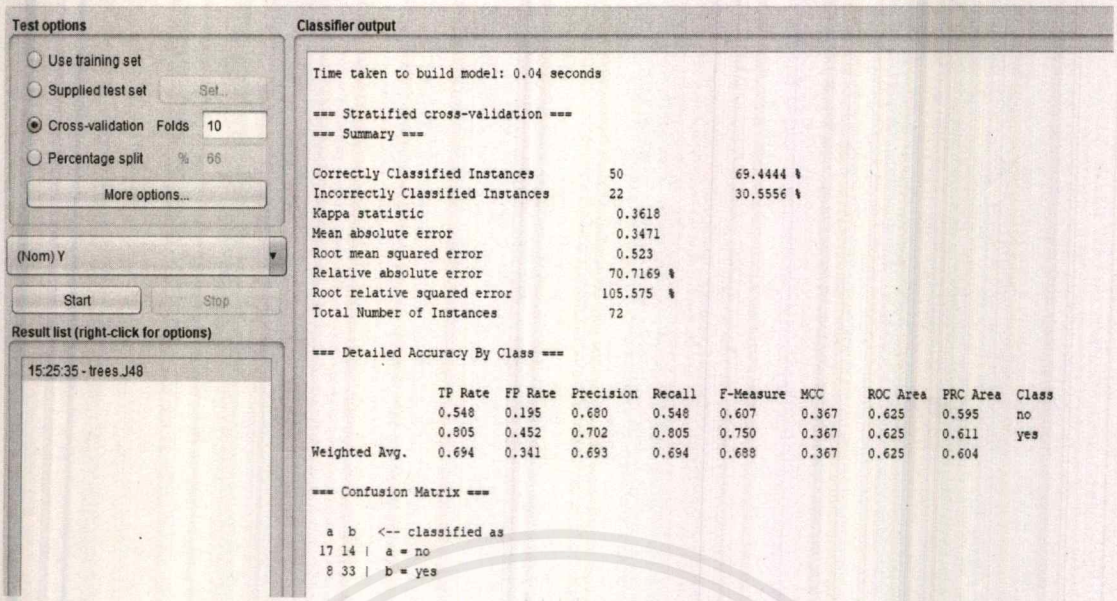
4.3.2.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทดสอบ 72 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.61 ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10

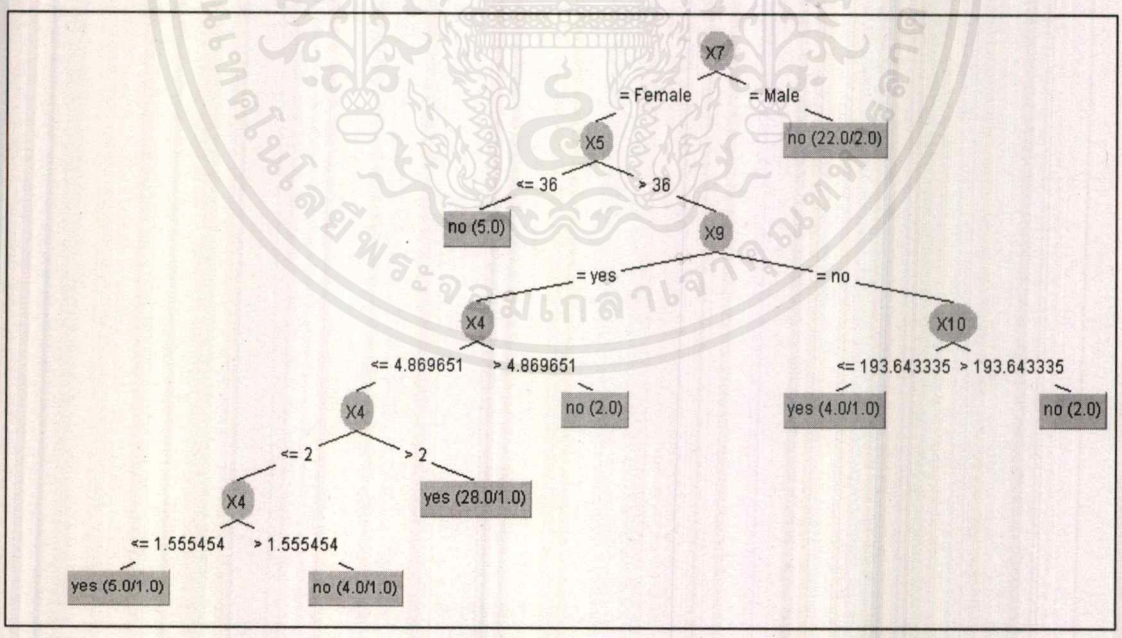
จากรูป 4.61 แสดงต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า โหนดรากแบ่งโดยตัวแปรทำนายคือ X_{10} = ยอดเงินคงเหลือในบัตรเครดิต ถ้าผลน้อยกว่าเท่ากับ 0 ดอลลาร์ คือไม่ชำระ จำนวน 12 คน ถ้าผลมากกว่า 0 ดอลลาร์ ตัดสินใจต่อด้วยตัวแปรทำนาย X_7 = เพศ ถ้าผลเป็น Female (เพศหญิง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_8 = การศึกษา ถ้าผลเป็น no (สำเร็จการศึกษา) คือชำระ จำนวน 44 คน ถ้าผลเป็น yes (กำลังศึกษา) คือไม่ชำระ จำนวน 2 คน ตัดสินใจต่อด้วยตัวแปรทำนาย X_7 = เพศ ถ้าผลเป็น Male (เพศชาย) ตัดสินใจต่อด้วยตัวแปรทำนาย X_9 = สถานภาพสมรส ถ้าผลเป็น no (โสด) คือไม่ชำระ (จำนวน 5 คน) ถ้าผลเป็น yes (แต่งงานแล้ว) ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = อายุ ถ้าผลน้อยกว่าเท่ากับ 59.328224 ปี คือชำระ จำนวน 5 คน ถ้าผลมากกว่า 59.328224 ปี คือไม่ชำระ จำนวน 4 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.62 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10

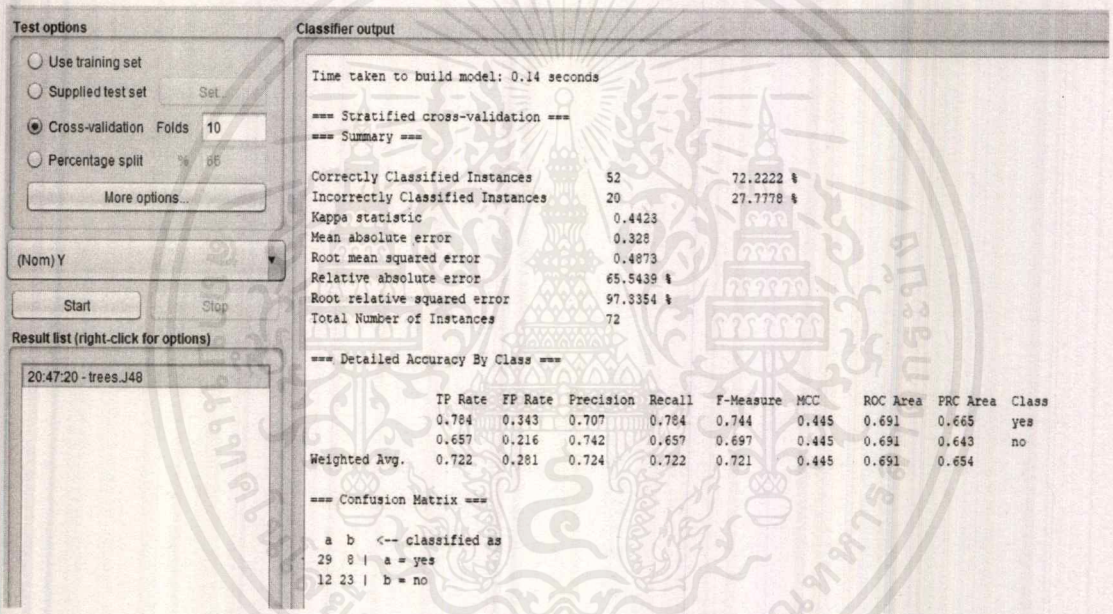
จากรูป 4.62 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 72 คน มีค่าทำนายความถูกต้องจำนวน 50 คน คิดเป็นร้อยละ 69.4444 โดยมีค่าความไว = 0.5484 ค่าความจำเพาะ = 0.8049 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.523)^2$ = 0.2735



รูปที่ 4.63 ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

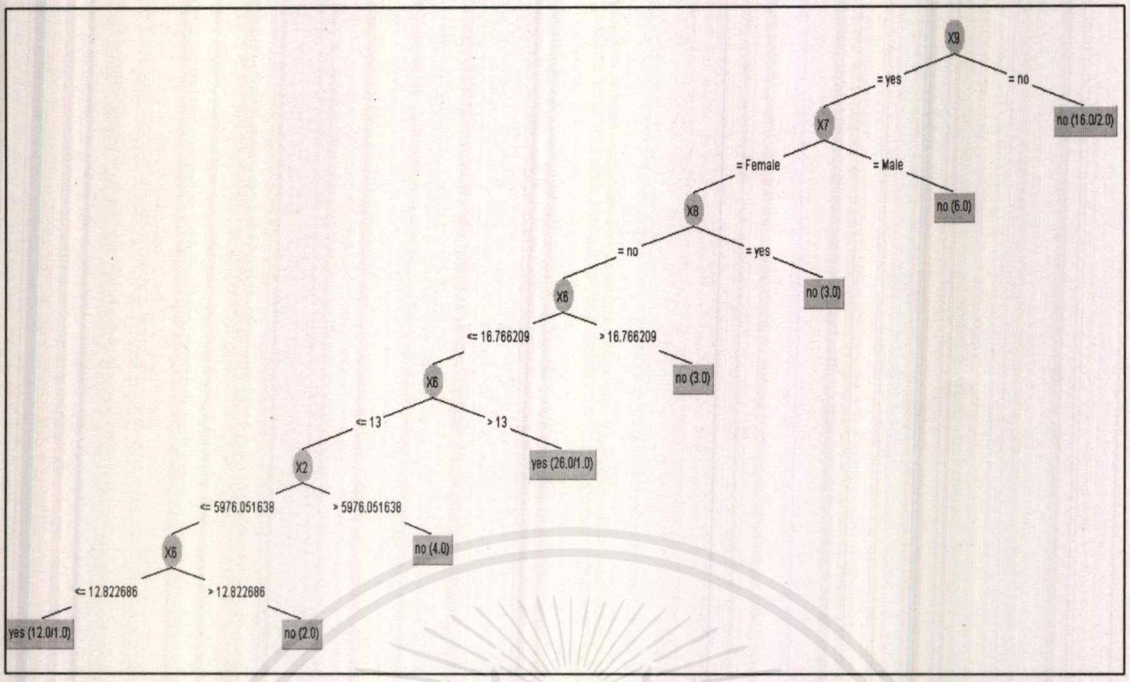
จากรูป 4.63 แสดงต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า โหนดราก แบ่งโดยตัวแปรทำนายคือ X_7 = เพศ ถ้าผลเป็น Male (เพศชาย) คือไม่ชำระ (จำนวน 22 คน) ถ้าผลเป็น Female (เพศหญิง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = อายุ ถ้าผลน้อยกว่าเท่ากับ 35 ปี คือไม่ชำระ จำนวน 5 คน ถ้าผลมากกว่า 35 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_9 = สถานภาพสมรส ถ้าผลเป็น no (โสด) ตัดสินใจด้วยตัวแปรทำนาย X_{10} = ยอดเงินคงเหลือในบัตรเครดิต ถ้าผลน้อยกว่าเท่ากับ 193.643335 ดอลลาร์ คือชำระ จำนวน 4 คน ถ้าผลมากกว่า 193.643335 ดอลลาร์ คือไม่ชำระ จำนวน 2 คน ตัดสินใจต่อด้วยตัวแปรทำนาย X_9 = สถานภาพสมรส ถ้าผลเป็น yes (แต่งงานแล้ว) ตัดสินใจต่อด้วยตัวแปร X_4 = จำนวนบัตรเครดิตที่มี ถ้าผลมากกว่า 4.869651 บัตร คือไม่ชำระ (จำนวน 2 คน) ถ้าผลน้อยกว่าเท่ากับ 4.869651 บัตร ตัดสินใจต่อด้วยตัวแปร X_4 = จำนวนบัตรเครดิตที่มี ถ้าผลมากกว่า 2 บัตร คือชำระ จำนวน 28 คน ถ้าผลน้อยกว่าเท่ากับ 2 บัตร ตัดสินใจต่อด้วยตัวแปร X_4 = จำนวนบัตรเครดิตที่มี ถ้าผลน้อยกว่าเท่ากับ 1.555454 บัตร คือชำระ จำนวน 5 คน ถ้าผลมากกว่า 1.555454 คือไม่ชำระ จำนวน 4 คน



รูปที่ 4.64 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

จากรูป 4.64 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 72 คน มีค่าทำนายความถูกต้องจำนวน 52 คน คิดเป็นร้อยละ 72.2222 โดยมีค่าความไว = 0.7838 ค่าความจำเพาะ = 0.6571 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4873)^2$ = 0.2375

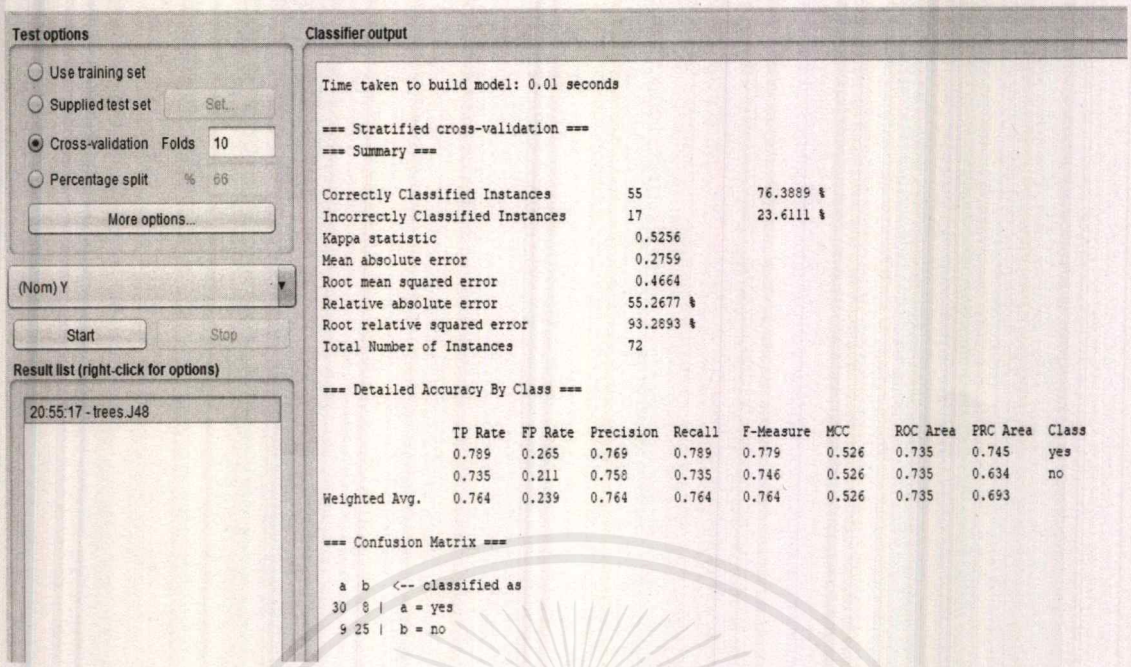
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.65 ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูป 4.65 แสดงต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า โหนดรากแบ่งโดยตัวแปรทำนายคือ X_9 = สถานภาพสมรส ถ้าผลเป็น no (โสด) คือไม่ชำระ (จำนวน 16 คน) ถ้าผลเป็น yes (แต่งงานแล้ว) ตัดสินใจต่อด้วยตัวแปรทำนาย X_7 = เพศ ถ้าผลเป็น Male (เพศชาย) คือไม่ชำระ (จำนวน 6 คน) ถ้าผลเป็น Female (เพศหญิง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_8 = การศึกษา ถ้าผลเป็น yes (กำลังศึกษา) คือไม่ชำระ (จำนวน 3 คน) ถ้าผลเป็น no (สำเร็จการศึกษา) ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = จำนวนปีในการศึกษา ถ้าผลมากกว่า 16.766209 ปี คือไม่ชำระ (จำนวน 3 คน) ถ้าผลน้อยกว่าเท่ากับ 16.766209 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = จำนวนปีในการศึกษา ถ้าผลมากกว่า 13 ปี คือชำระ (จำนวน 26 คน) ถ้าผลน้อยกว่าเท่ากับ 13 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = วงเงินในบัตรเครดิต ถ้าผลมากกว่า 5976.051638 ดอลลาร์ คือไม่ชำระ (จำนวน 4 คน) ถ้าผลน้อยกว่าเท่ากับ 5976.051638 ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = จำนวนปีในการศึกษา ถ้าผลน้อยกว่าเท่ากับ 12.822686 ปี คือชำระ (จำนวน 12 คน) ถ้าผลมากกว่า 12.822686 ปี คือไม่ชำระ (จำนวน 2 คน)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

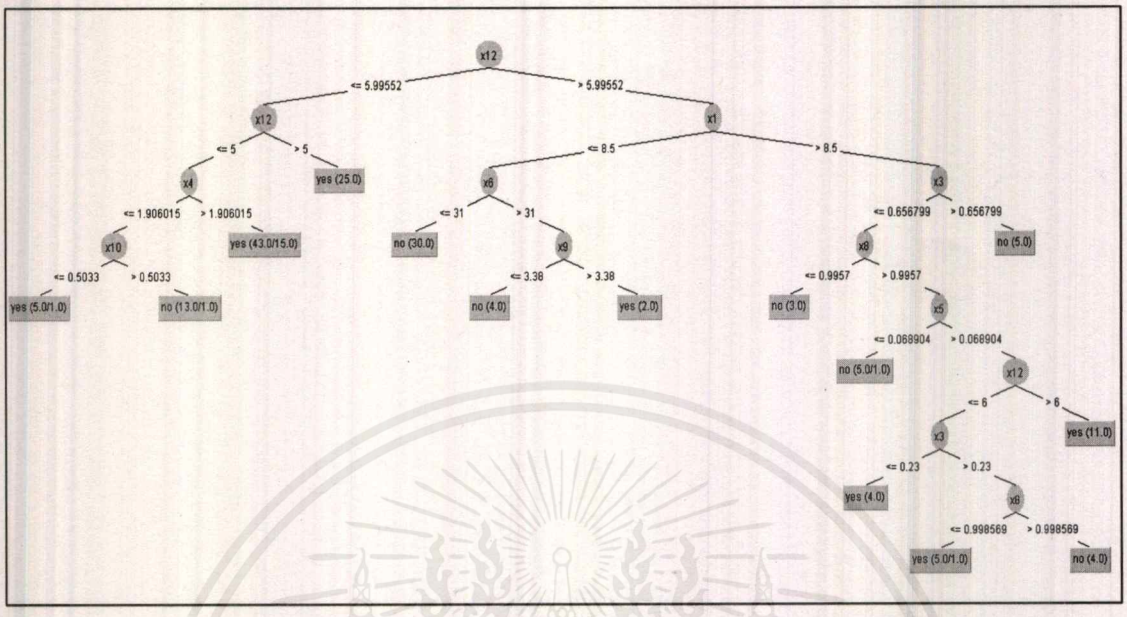


รูปที่ 4.66 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูป 4.66 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 72 คน มีค่าทำนายความถูกต้องจำนวน 41 คน คิดเป็นร้อยละ 56.9444 โดยมีค่าความไว = 0.5263 ค่าความจำเพาะ = 0.6176 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6024)^2$ = 0.3629

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

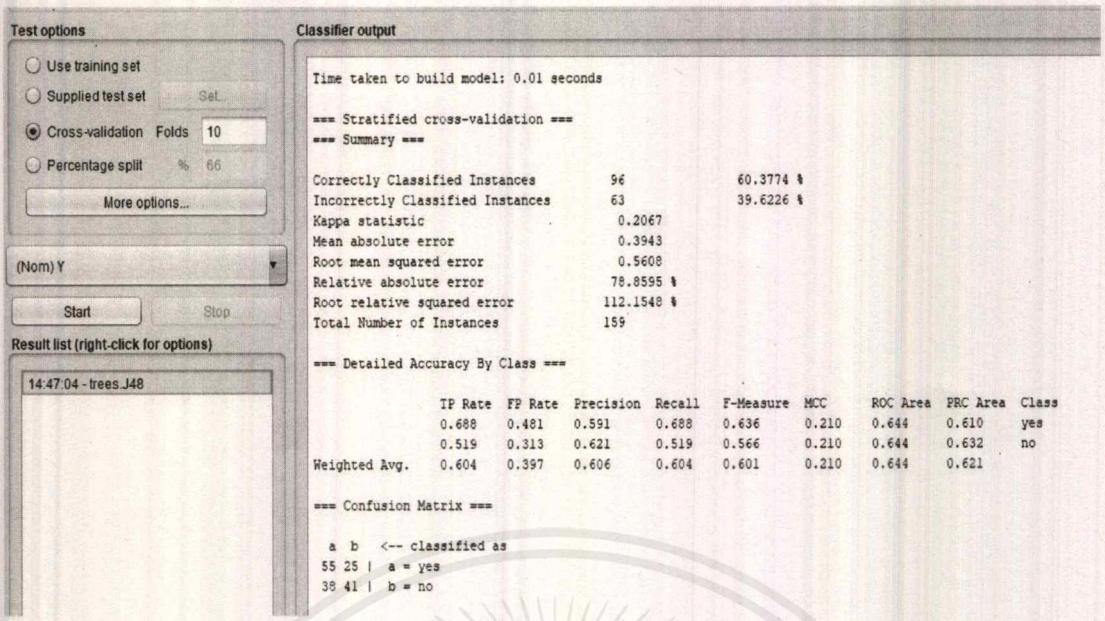
4.3.2.3 คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทดสอบ 159 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.67 ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 10

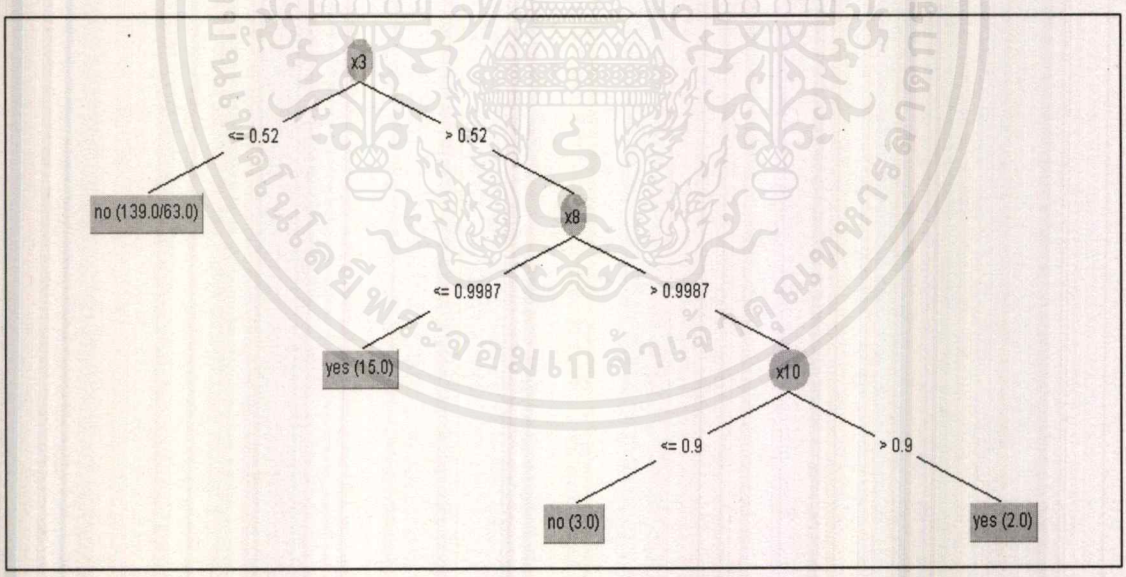
จากรูป 4.67 แสดงต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง โหนดรากแบ่งโดยตัวแปรทำนายคือ X_{12} = คะแนนคุณภาพ ถ้าผลน้อยกว่าเท่ากับ 5.99552 คะแนน ตัดสินใจต่อด้วยตัวแปรทำนาย X_{12} = คะแนนคุณภาพ คือคุณภาพดี จำนวน 25 ขวด ถ้าผลน้อยกว่าเท่ากับ 5 คะแนน ตัดสินต่อด้วยตัวแปรทำนาย X_4 = ค่าน้ำตาลรีดิวซ์ ถ้าผลมากกว่า 1.906015 กรัม คือคุณภาพดี จำนวน 43 ขวด ถ้าผลน้อยกว่าเท่ากับ 1.906015 กรัม ตัดสินต่อด้วยตัวแปรทำนาย X_{10} = ค่าเกลือของกรดซัลฟิวริก ถ้าผลมากกว่า 0.5033 pH คือคุณภาพไม่ดี จำนวน 13 ขวด ถ้าผลน้อยกว่า 0.5033 pH คือคุณภาพดี จำนวน 5 ขวด ตัดสินใจต่อด้วยตัวแปร X_{12} = คะแนนคุณภาพ ถ้าผลมากกว่า 5.99552 คะแนน ตัดสินต่อด้วยตัวแปรทำนาย X_1 = ค่าความเป็นกรด ถ้าผลน้อยกว่าเท่ากับ 8.5 pH ตัดสินใจต่อด้วยตัวแปร X_6 = ค่าซัลเฟอร์ไดออกไซด์ ถ้าผลน้อยกว่าเท่ากับ 31 กรัม คือคุณภาพไม่ดี จำนวน 30 ขวด ถ้าผลมากกว่า 31 กรัม ตัดสินใจต่อด้วยตัวแปร X_9 = ค่าความเป็นกรดหรือด่าง ถ้าผลน้อยกว่าเท่ากับ 3.38 pH คือคุณภาพไม่ดี จำนวน 4 ขวด ถ้าผลมากกว่า 3.38 pH คือคุณภาพดี จำนวน 2 ขวด ตัดสินต่อด้วยตัวแปรทำนาย X_1 = ค่าความเป็นกรด ถ้าผลมากกว่า 8.5 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ค่ากรดซิตริก ถ้าผลมากกว่า 0.656799 pH คือคุณภาพไม่ดี จำนวน 5 ขวด ถ้าผลน้อยกว่า 0.656799 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_8 = ค่าความหนาแน่น ถ้าผลน้อยกว่าเท่ากับ 0.9957 กิโลกรัมต่อลูกบาศก์เมตร คือคุณภาพไม่ดี จำนวน 3 ขวด ถ้าผลมากกว่า 0.9957 กิโลกรัมต่อลูกบาศก์เมตร ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = ค่าคลอไรด์ ถ้าผลน้อยกว่าเท่ากับ 0.068904 กรัม คือคุณภาพไม่ดี จำนวน 5 ขวด ถ้าผลมากกว่า 0.068904 กรัม ตัดสินต่อด้วยตัวแปรทำนาย X_{12} = คะแนนคุณภาพ ถ้าผลมากกว่า 6 คะแนน คือคุณภาพดี จำนวน 11 ขวด ถ้าผลน้อยกว่าเท่ากับ 6 คะแนน ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ค่ากรดซิตริก ถ้าผลน้อยกว่าเท่ากับ 0.23 pH คือคุณภาพดี จำนวน 4 ขวด ถ้าผลมากกว่า0.23 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_8 = ค่าความหนาแน่น ถ้าผลน้อยกว่าเท่ากับ 0.998569 กิโลกรัมต่อลูกบาศก์เมตร คือคุณภาพดี จำนวน 5 ขวด ถ้าผลมากกว่า 0.998569 กิโลกรัมต่อลูกบาศก์เมตร คือคุณภาพไม่ดี จำนวน 4 ขวด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.68 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 10

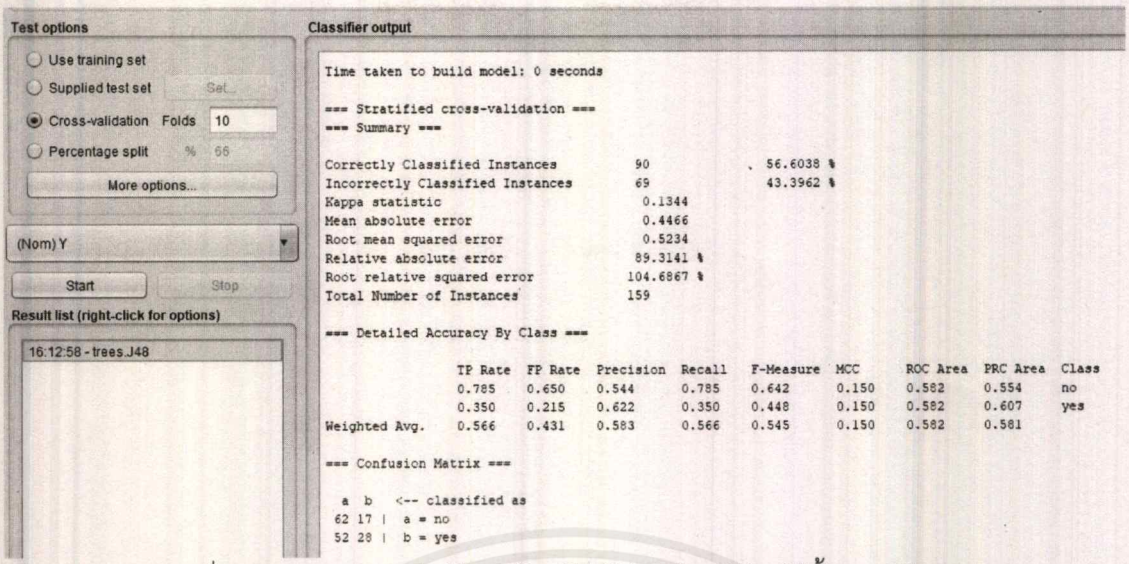
จากรูป 4.68 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 159 คน มีค่าทำนายความถูกต้องจำนวน 96 คน คิดเป็นร้อยละ 60.3774 โดยมีค่าความไว = 0.6875 ค่าความจำเพาะ = 0.5190 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5608)^2$ = 0.3145



รูปที่ 4.69 ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 20

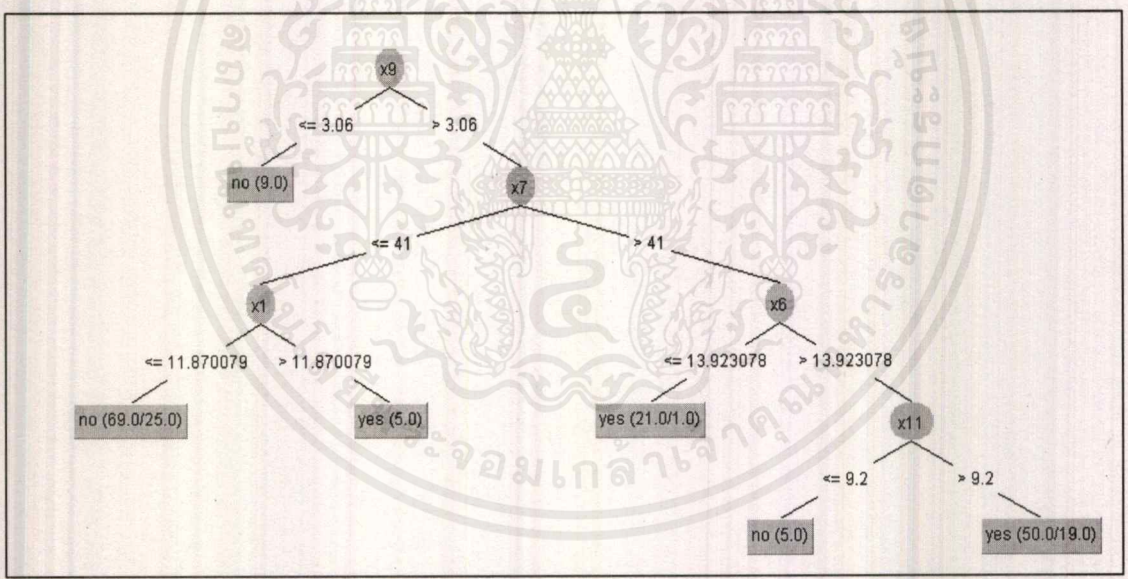
จากรูป 4.69 แสดงต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง โหนดรากแบ่งโดยตัวแปรทำนายคือ X_3 = ค่ากรดซิตริก ถ้าผลน้อยกว่าเท่ากับ 0.52 pH คือคุณภาพไม่ดี จำนวน 139 ขวด ถ้าผลมากกว่า 0.52 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_8 = ค่าความหนาแน่น ถ้าผลน้อยกว่าเท่ากับ 0.9987 กิโลกรัมต่อลูกบาศก์เมตร คือคุณภาพดี จำนวน 15 ขวด ถ้าผลมากกว่า 0.9987 กิโลกรัมต่อลูกบาศก์เมตร ตัดสินใจต่อด้วยตัวแปรทำนาย X_{10} = ค่าเกลือของกรดซัลฟิวริก ถ้าผลน้อยกว่าเท่ากับ 0.9 pH คือคุณภาพไม่ดี จำนวน 3 ขวด ถ้าผลมากกว่า 0.9 pH คือคุณภาพดี จำนวน 2 ขวด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.70 ผลการวิเคราะห์ข้อมูลด้วยต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 20

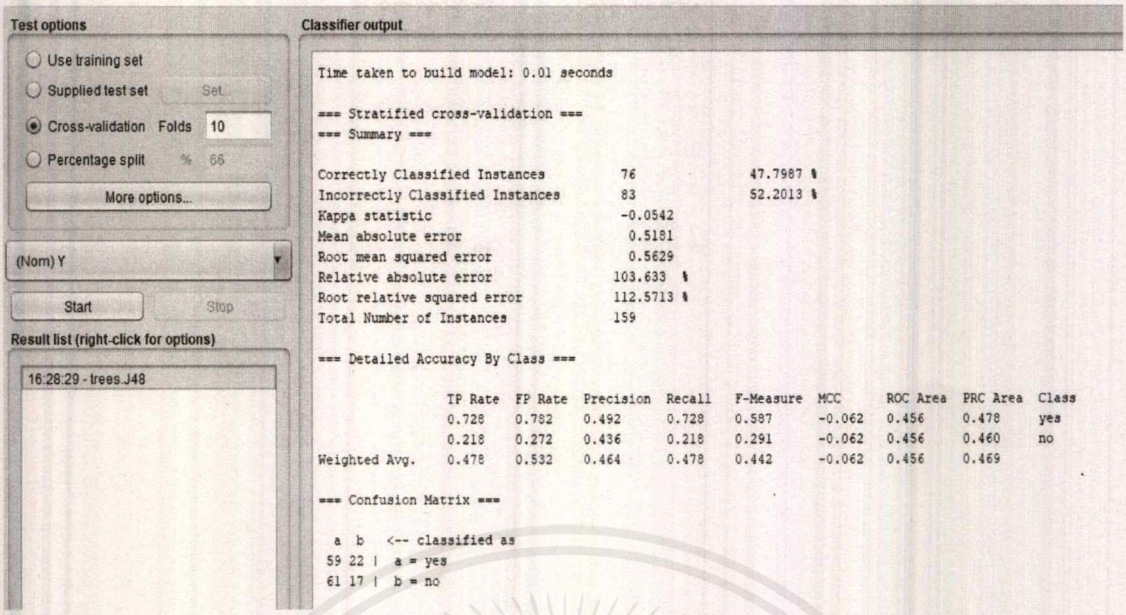
จากรูป 4.68 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 159 คน มีค่าทำนายความถูกต้องจำนวน 90 คน คิดเป็นร้อยละ 56.6038 โดยมีค่าความไว = 0.7848 ค่าความจำเพาะ = 0.3500 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5234)^2$ = 0.2739



รูปที่ 4.71 ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูป 4.71 แสดงต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง โหนดรากแบ่งโดยตัวแปรทำนายคือ X_9 = ค่าความเป็นกรดหรือด่าง ถ้าผลน้อยกว่าเท่ากับ 3.06 pH คือคุณภาพไม่ดี จำนวน 9 ขวด ถ้าผลมากกว่า 3.06 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_7 = ค่าซัลเฟอร์ไดออกไซด์ทั้งหมด ถ้าผลน้อยกว่าเท่ากับ 41 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = ค่าความเป็นกรด ถ้าผลน้อยกว่าเท่ากับ 11.870079 pH คือคุณภาพไม่ดี จำนวน 69 ขวด ถ้าผลมากกว่า 11.870079 pH คือคุณภาพดี จำนวน 5 ขวด ตัดสินใจต่อด้วยตัวแปรทำนาย X_7 = ค่าซัลเฟอร์ไดออกไซด์ทั้งหมด ถ้าผลมากกว่า 41 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = ค่าซัลเฟอร์ไดออกไซด์ ถ้าผลน้อยกว่าเท่ากับ 13.923078 กรัม คือคุณภาพดี จำนวน 21 ขวด ถ้าผลมากกว่า 13.923078 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_{11} = ค่าแอกอฮอล์ ถ้าผลน้อยกว่าเท่ากับ 9.2 กรัม คือคุณภาพไม่ดี จำนวน 5 ขวด ถ้าผลมากกว่า 9.2 กรัม คือคุณภาพดี จำนวน 50 ขวด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



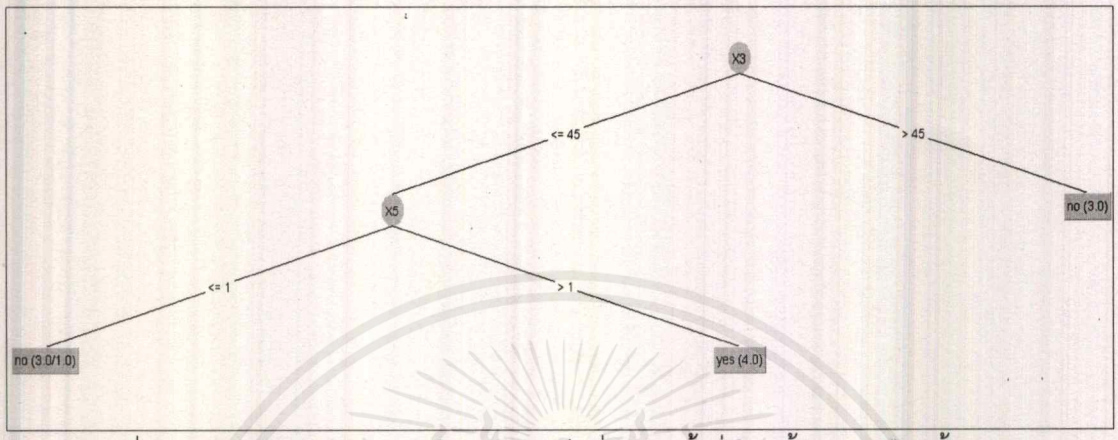
รูปที่ 4.72 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูป 4.72 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 159 คน มีค่าทำนายความถูกต้องจำนวน 76 คน คิดเป็นร้อยละ 47.7987 โดยมีค่าความไว = 0.7284 ค่าความจำเพาะ = 0.2180 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5629)^2$ = 0.3169

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

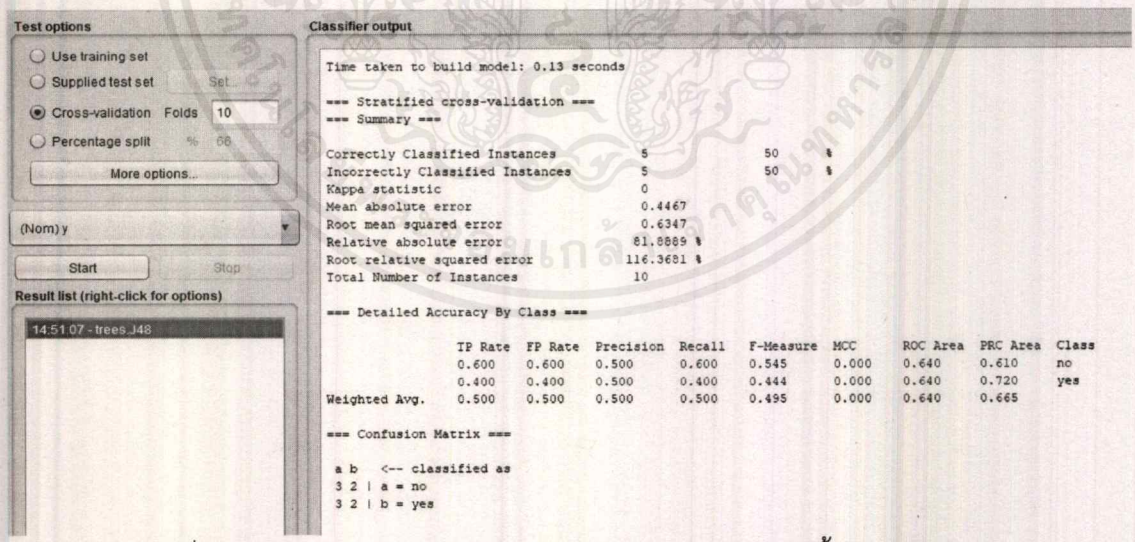
4.3.3 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้มไม้ตัดสินใจ โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด (Under Sampling)

4.3.3.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทดสอบ 10 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.73 ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10

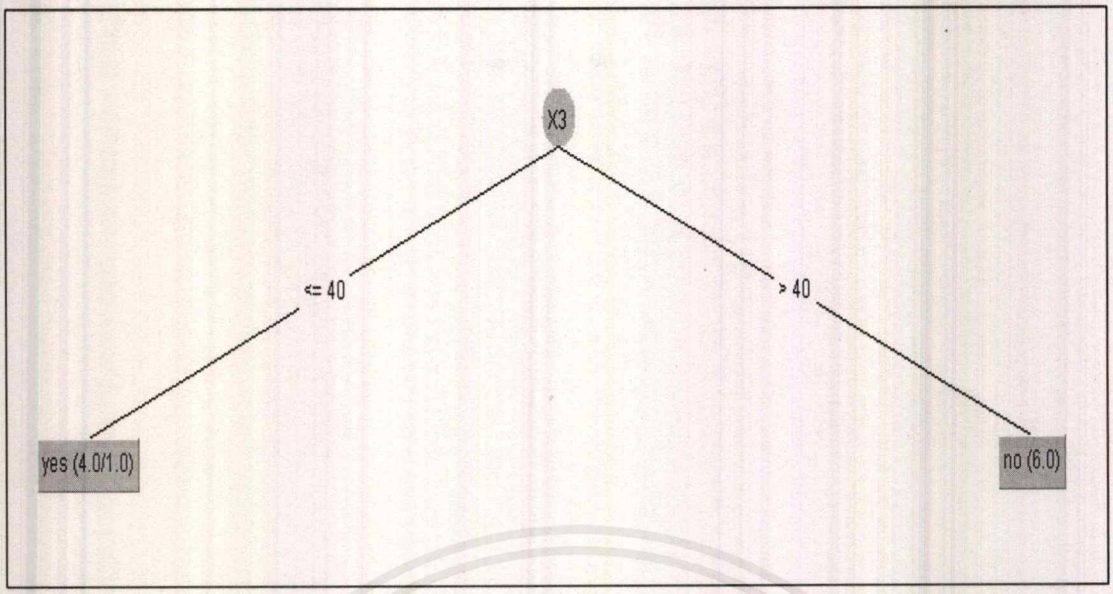
จากรูป 4.73 แสดงต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โหนดรากแบ่งโดยตัวแปรทำนายคือ X_3 = ความดังที่ได้รับ ถ้าผลมากกว่า 45 เดซิเบล คือได้ยินเสียงไม่ปกติ จำนวน 3 คน ถ้าผลน้อยกว่าหรือเท่ากับ 45 เดซิเบล ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าหรือเท่ากับ 1 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 3 คน ถ้าผลมาก 1 ครั้ง คือได้ยินเสียงปกติ จำนวน 4 คน



รูปที่ 4.74 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10

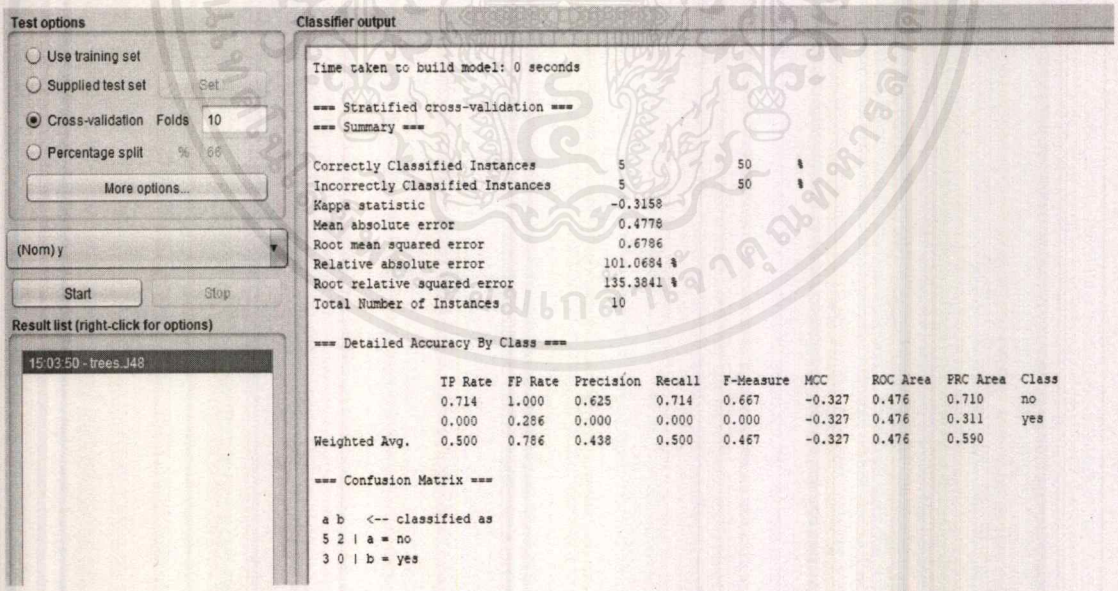
จากรูป 4.74 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 5 คน คิดเป็นร้อยละ 50 โดยมีค่าความไว = 0.6000 ค่าความจำเพาะ = 0.4000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6347)^2$ = 0.4028

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.75 ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

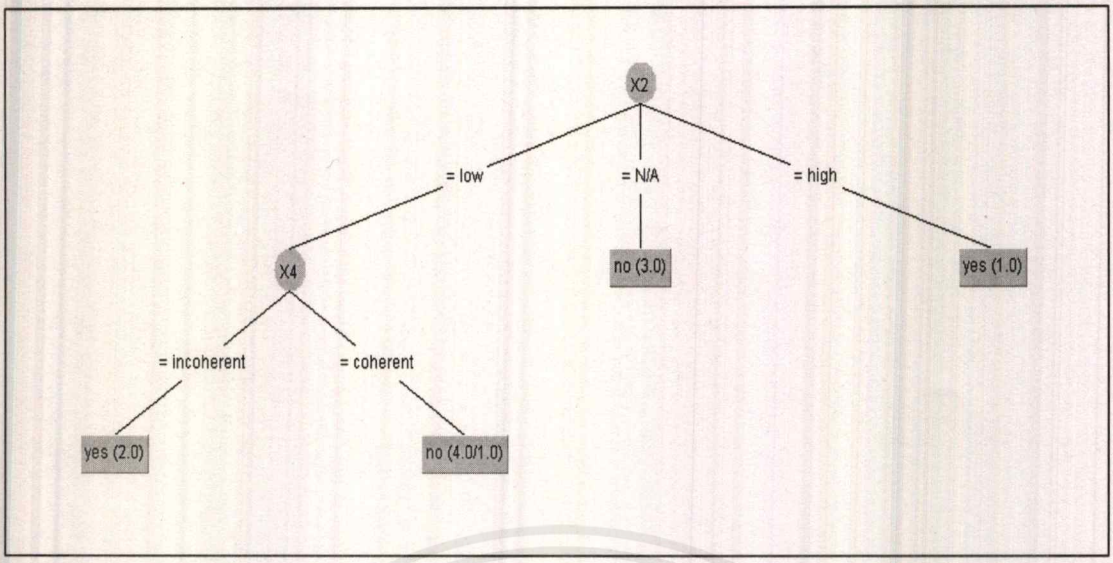
จากรูป 4.75 แสดงต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โหนดรากแบ่งโดยตัวแปรทำนายคือ X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าเท่ากับ 5 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 7 คน ถ้าผลมากกว่า 5 ครั้ง คือ ได้ยินเสียงปกติ จำนวน 3 คน



รูปที่ 4.76 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

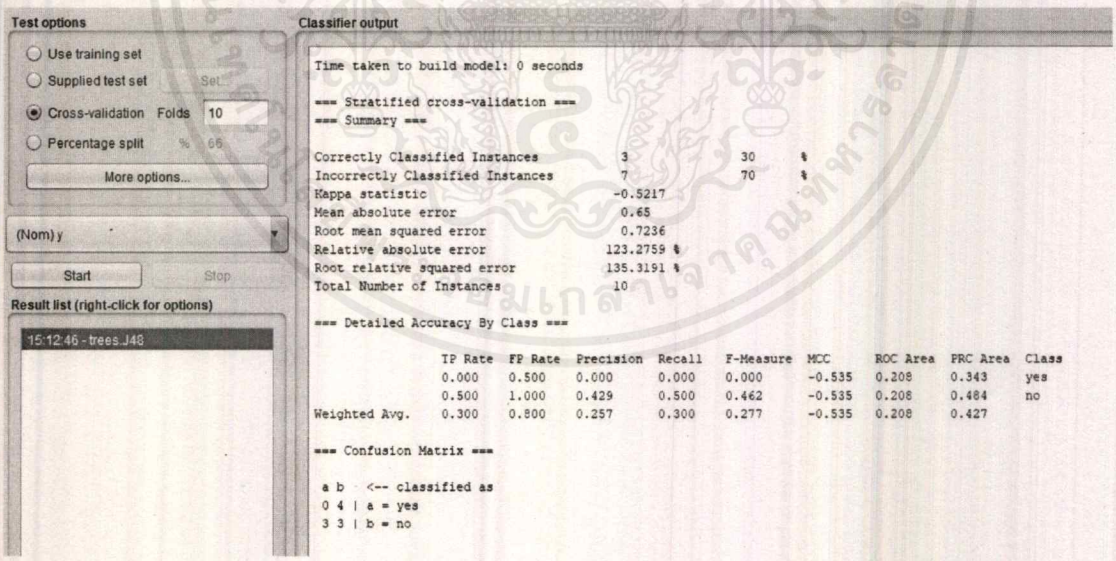
จากรูป 4.76 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 5 คน คิดเป็นร้อยละ 50 โดยมีค่าความไว = 0.7143 ค่าความจำเพาะ = 0.0000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6786)^2$ = 0.4605

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.77 ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูป 4.77 แสดงต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โหนดรากแบ่งโดยตัวแปรทำนายคือ X_2 = อัตราความเสี่ยงการเป็นโรคหูชั้นกลางอักเสบ ถ้าผลเท่ากับ N/A คือไม่มีความเสี่ยง จำนวน 2 คน ถ้าผลเท่ากับ high คือความเสี่ยงมาก จำนวน 2 คน ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = อายุ ถ้าผลน้อยกว่าเท่ากับ 30 ปี คือคุณภาพไม่ดี จำนวน 3 คน ถ้าผลมากกว่า 30 ปี คือคุณภาพดี จำนวน 3 คน

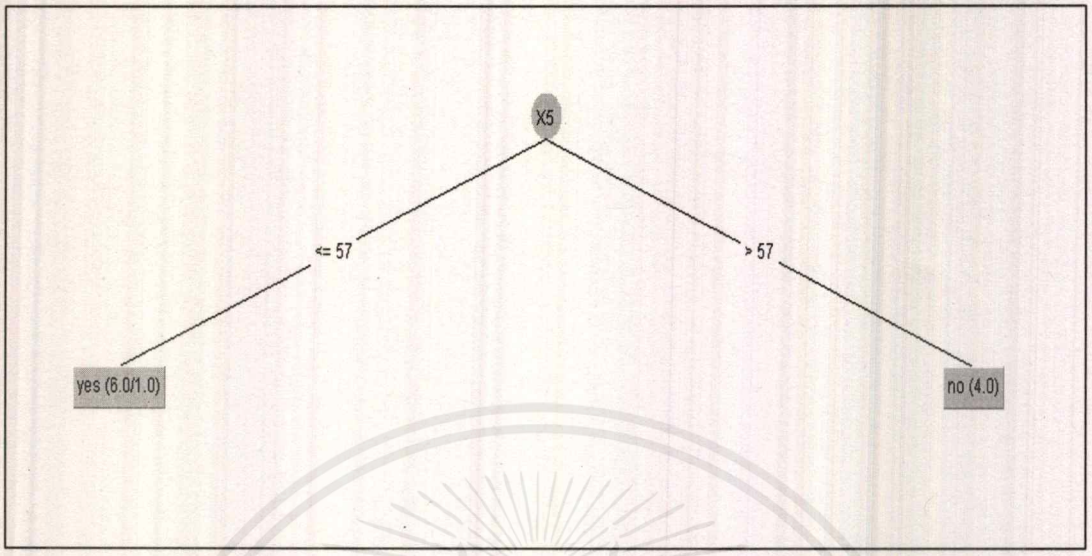


รูปที่ 4.78 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูป 4.78 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 3 คน คิดเป็นร้อยละ 30 โดยมีค่าความไว = 0.0000 ค่าความจำเพาะ = 0.5000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.7236)^2$ = 0.5236

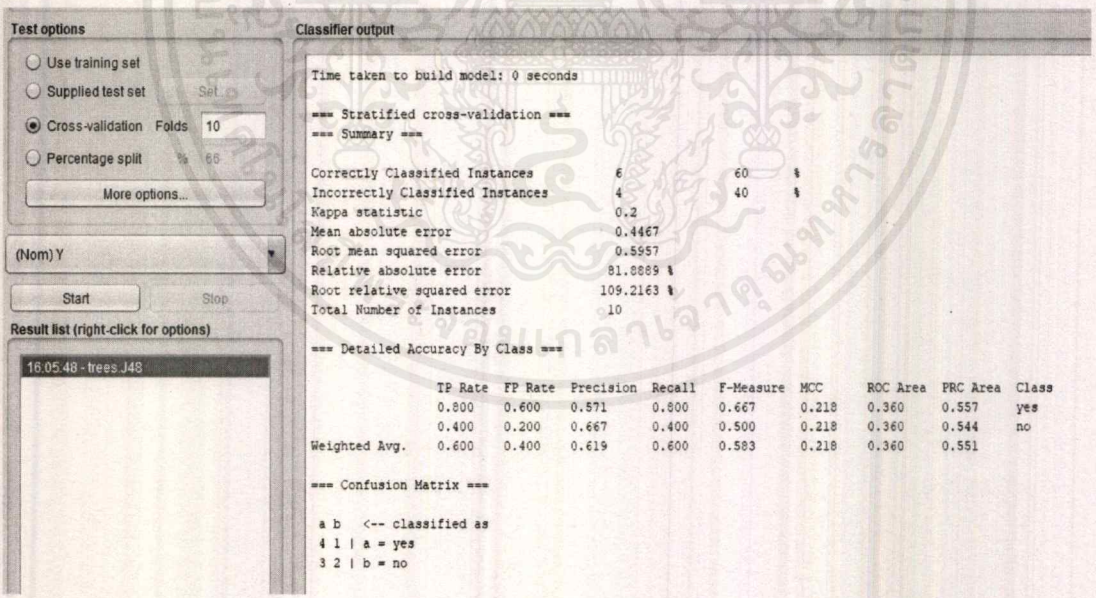
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.3.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทดสอบ 10 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.79 ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10

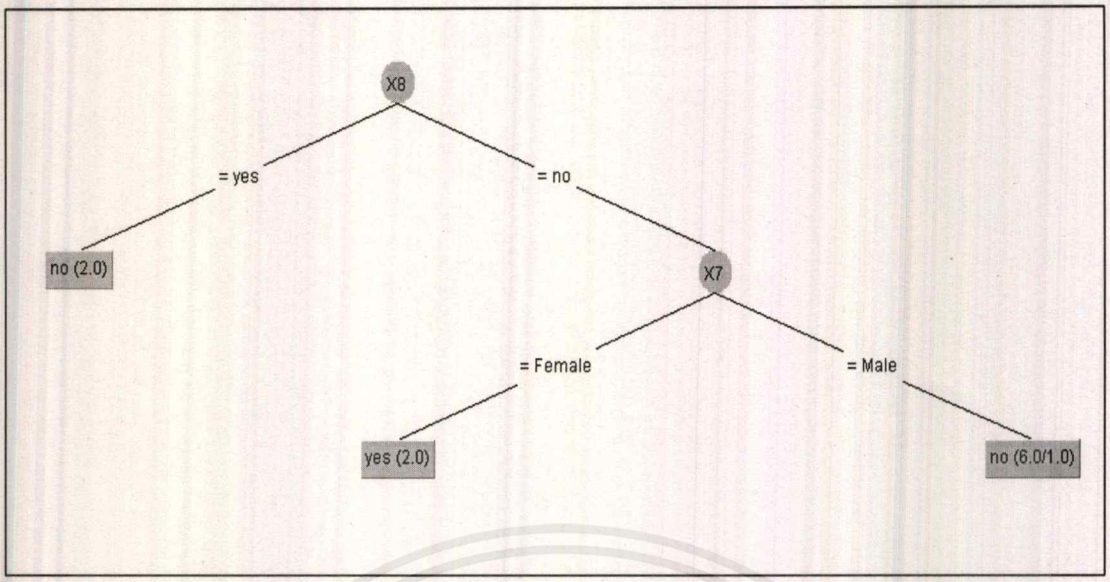
จากรูป 4.79 แสดงต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า โหนดรากแบ่งโดยตัวแปรทำนายคือ X_5 = อายุ ถ้าผลน้อยกว่าหรือเท่ากับ 57 ปี คือ ชำระ จำนวน 6 คน ถ้าผลมากกว่า 57 ปี คือ ไม่ชำระ จำนวน 4 คน



รูปที่ 4.80 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10

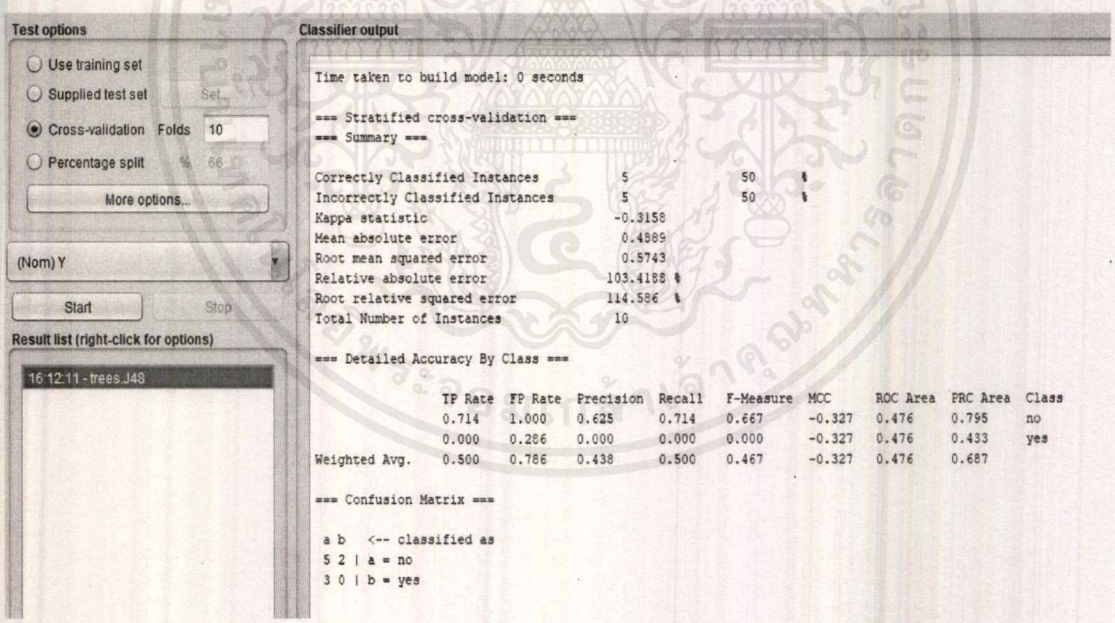
จากรูป 4.80 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 6 คน คิดเป็นร้อยละ 60 โดยมีค่าความไว = 0.8000 ค่าความจำเพาะ = 0.4000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5957)^2$ = 0.3549

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.81 ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10

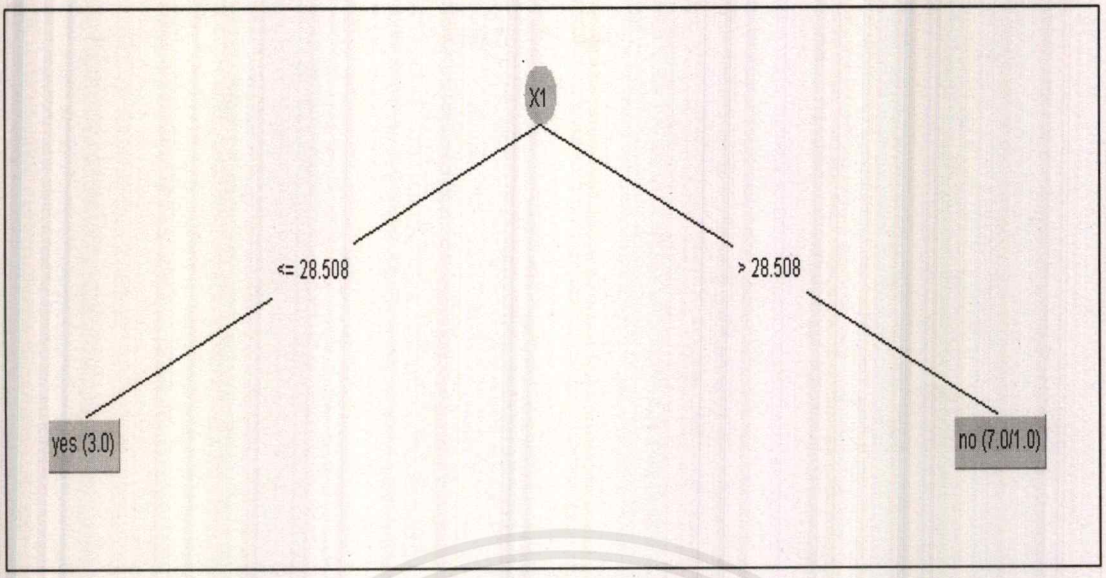
จากรูป 4.81 แสดงต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า โหนดราก แบ่งโดยตัวแปรทำนายคือ X_8 = การศึกษา ถ้าผลเป็น Yes (กำลังศึกษา) คือไม่ชำระ จำนวน 2 คน ถ้าผลเป็น No (สำเร็จการศึกษา) ตัดสินใจต่อด้วยตัวแปรทำนาย X_7 = เพศ ถ้าผลเป็น Female (เพศหญิง) คือชำระ จำนวน 2 คน ถ้าผลเป็น Male (เพศชาย) คือไม่ชำระ จำนวน 6 คน



รูปที่ 4.82 ผลการวิเคราะห์ข้อมูลด้วยต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

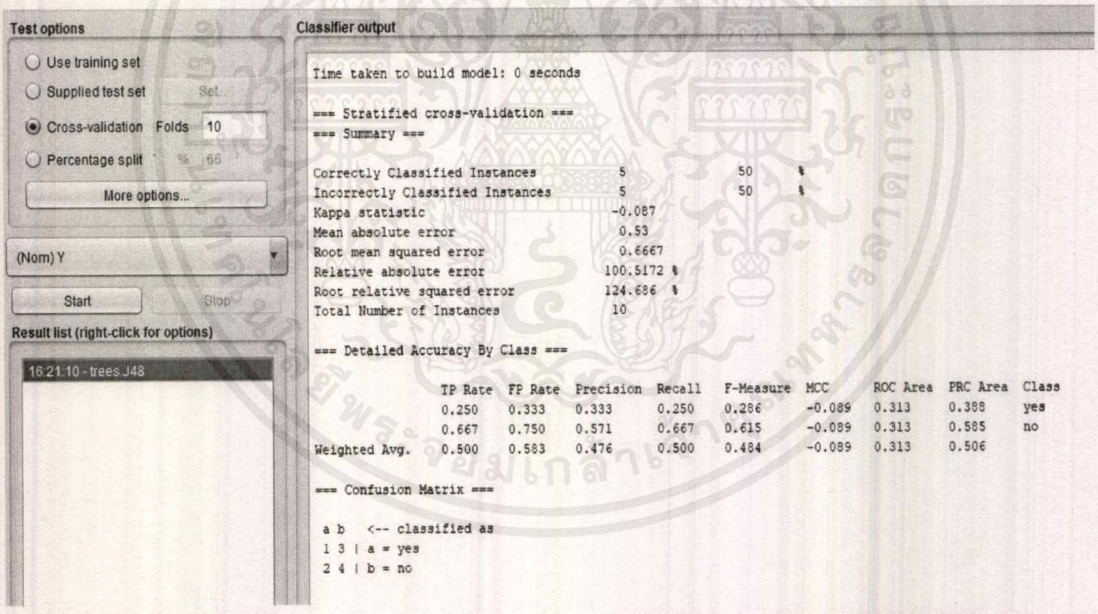
จากรูป 4.82 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 5 คน คิดเป็นร้อยละ 50 โดยมีค่าความไว = 0.7142 ค่าความจำเพาะ = 0.0000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5743)^2$ = 0.3298

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.83 ต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูป 4.83 แสดงต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า โหนดราก แบ่งโดยตัวแปรทำนายคือ X_1 = รายได้ต่อเดือน ถ้าผลน้อยกว่าหรือเท่ากับ 28.508 ดอลลาร์ คือชำระ จำนวน 3 คน ถ้าผลมากกว่า 28.508 ดอลลาร์ คือไม่ชำระ จำนวน 7 คน

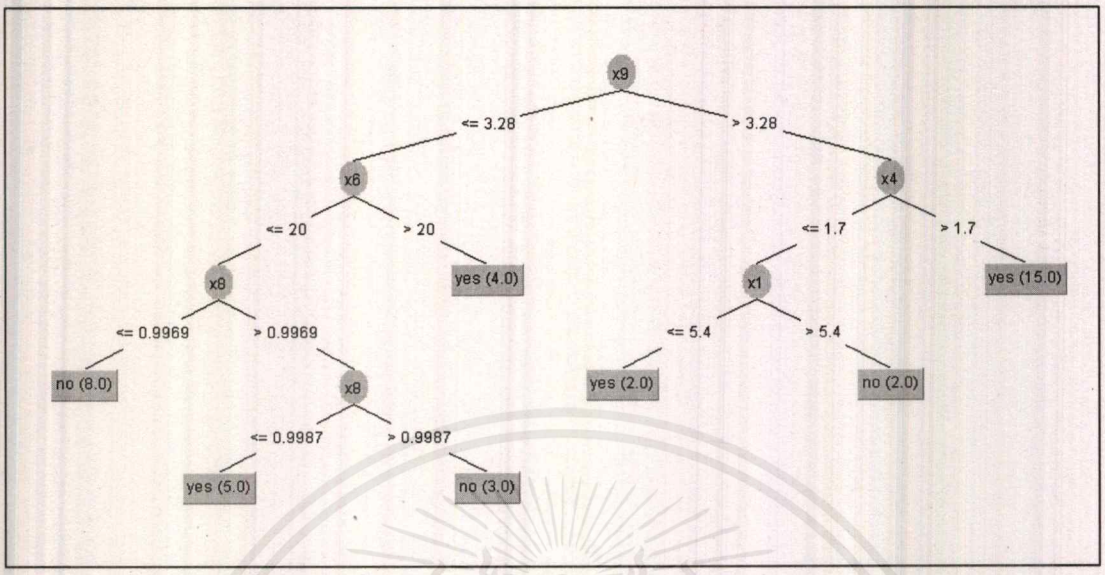


รูปที่ 4.84 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูป 4.84 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 5 คน คิดเป็นร้อยละ 50 โดยมีค่าความไว = 0.2500 ค่าความจำเพาะ = 0.6667 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6667)^2$ = 0.4445

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

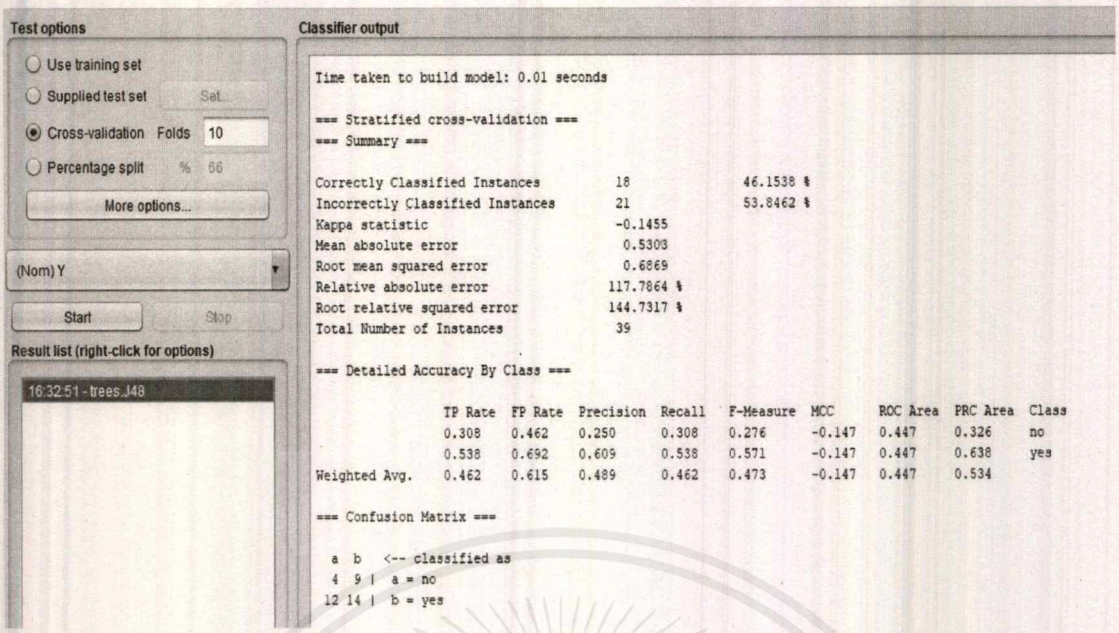
4.3.3.3 คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทดสอบ 39 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.85 ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 10

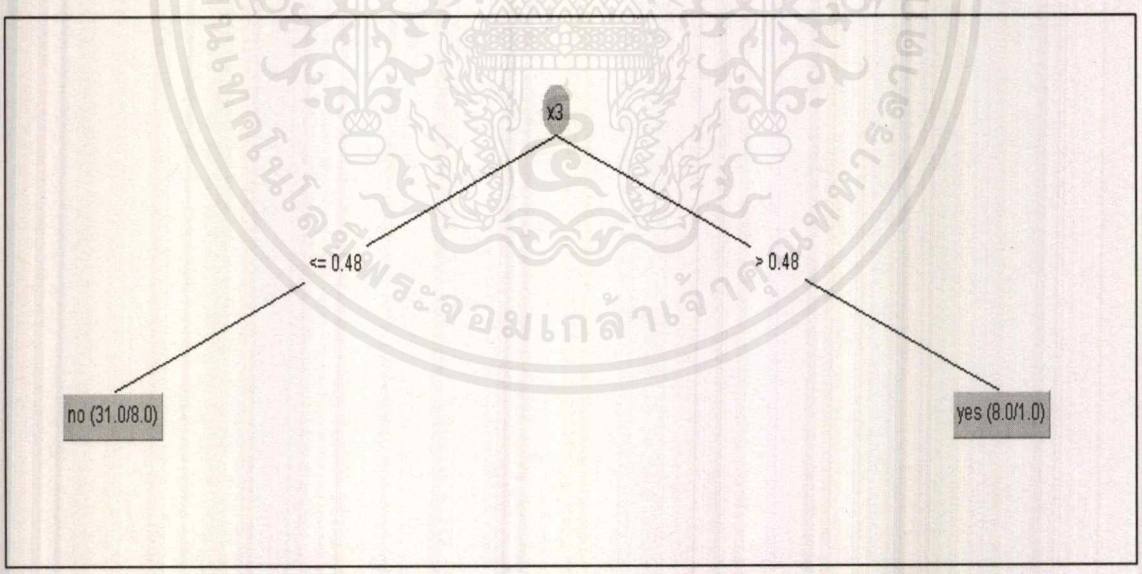
จากรูป 4.85 แสดงต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง โหนดรากแบ่งโดยตัวแปรทำนาย คือ X_9 = ค่าความเป็นกรดหรือด่าง ถ้าผลน้อยกว่าหรือเท่ากับ 3.28 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = ค่าซัลเฟอร์ไดออกไซด์ ถ้าผลมากกว่า 20 กรัม คือคุณภาพดี จำนวน 4 ขวด ถ้าผลน้อยกว่าหรือเท่ากับ 20 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_8 = ค่าความหนาแน่น ถ้าผลน้อยกว่าหรือเท่ากับ 0.9969 คือคุณภาพไม่ดี จำนวน 8 ขวด ถ้าผลมากกว่า 0.9969 ตัดสินใจต่อด้วยตัวแปรทำนาย X_8 = ค่าความหนาแน่น ถ้าผลน้อยกว่าหรือเท่ากับ 0.9987 คือคุณภาพดี จำนวน 5 ขวด ถ้าผลมากกว่า 0.9987 คือคุณภาพไม่ดี จำนวน 3 ขวด ตัดสินใจต่อด้วยตัวแปรทำนาย X_9 = ค่าความเป็นกรดหรือด่าง ถ้าผลมากกว่า 3.28 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = ค่าน้ำตาลรีดิวซ์ ถ้าผลมากกว่า 1.7 กรัม คือคุณภาพดี จำนวน 15 ขวด ถ้าผลน้อยกว่าหรือเท่ากับ 1.7 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = ค่าความเป็นกรด ถ้าผลน้อยกว่าหรือเท่ากับ 5.4 pH คือคุณภาพดี จำนวน 2 ขวด ถ้าผลมากกว่า 5.4 pH คือคุณภาพไม่ดี จำนวน 2 ขวด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.86 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 10

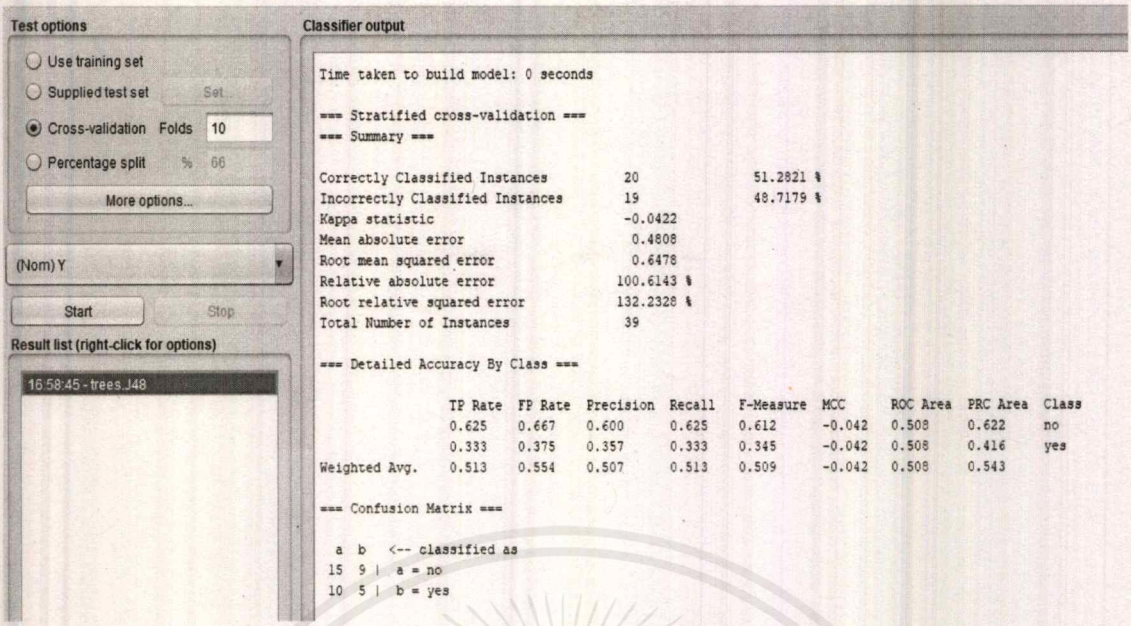
จากรูป 4.86 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 39 คน มีค่าทำนายความถูกต้องจำนวน 18 คน คิดเป็นร้อยละ 46.1538 โดยมีค่าความไว = 0.3077 ค่าความจำเพาะ = 0.5384 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6869)^2$ = 0.4718



รูปที่ 4.87 ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 20

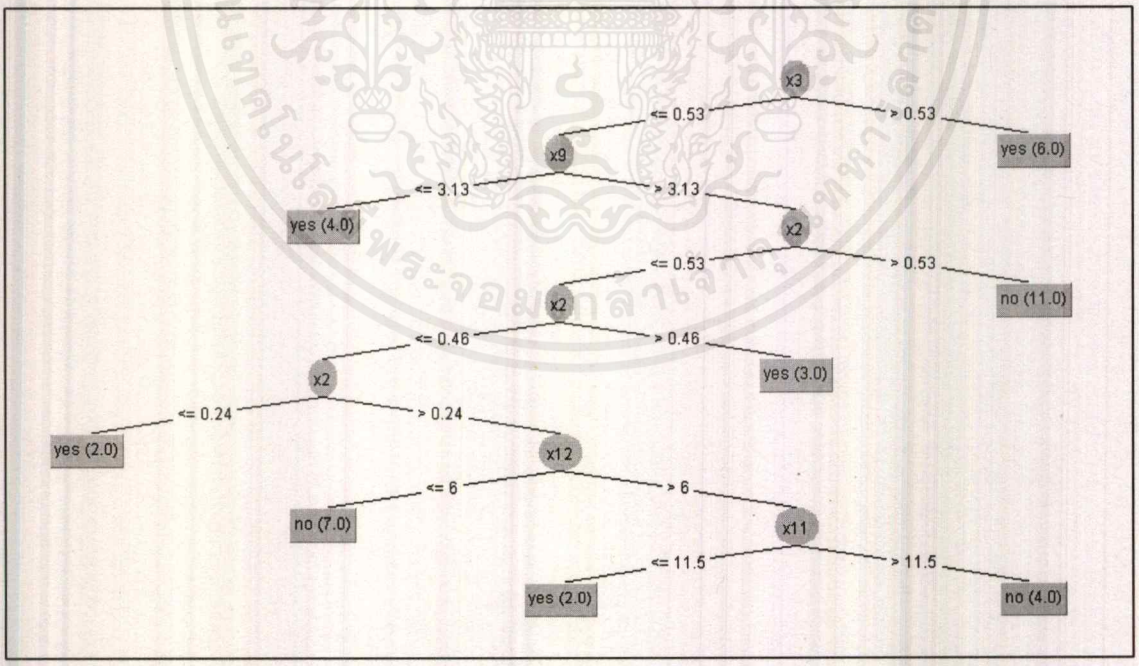
จากรูป 4.87 แสดงต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง โหนดรากแบ่งโดยตัวแปรทำนาย คือ X_3 = ค่ากรดซิตริก ถ้าผลน้อยกว่าหรือเท่ากับ 0.48 คือคุณภาพไม่ดี จำนวน 31 ขวด ถ้าผลมากกว่า 0.48 คือคุณภาพดี จำนวน 8 ขวด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.88 ผลการวิเคราะห์ข้อมูลด้วยต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 20

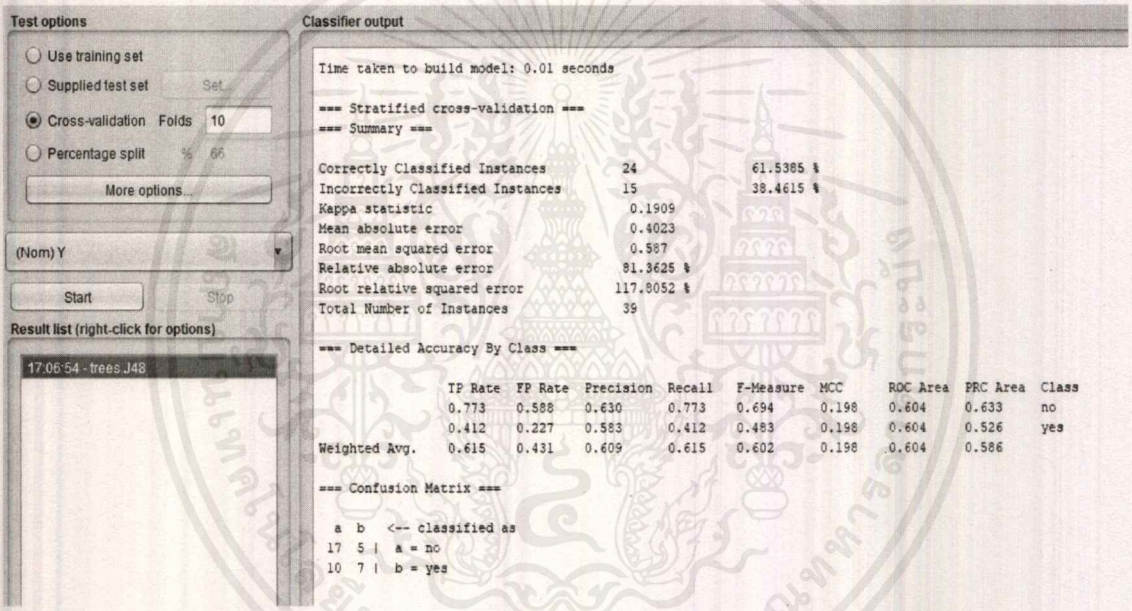
จากรูป 4.88 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 39 คน มีค่าทำนายความถูกต้องจำนวน 20 คน คิดเป็นร้อยละ 51.2821 โดยมีค่าความไว = 0.6250 ค่าความจำเพาะ = 0.3333 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6478)^2$ = 0.4196



รูปที่ 4.89 ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 30

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูป 4.89 แสดงต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง โหนดรากแบ่งโดยตัวแปรทำนาย คือ X_3 = ค่ากรดซิตริก ถ้าผลมากกว่า 0.53 pH คือคุณภาพดี จำนวน 6 ขวด ถ้าผลน้อยกว่าหรือเท่ากับ 0.53 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_9 = ค่าความเป็นกรดหรือด่าง ถ้าผลน้อยกว่าหรือเท่ากับ 3.13 pH คือคุณภาพดี จำนวน 4 ขวด ถ้าผลมากกว่า 3.13 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = ค่าความเป็นกรดระเหย ถ้าผลมากกว่า 0.53 pH คือคุณภาพไม่ดี จำนวน 11 ขวด ถ้าผลน้อยกว่าหรือเท่ากับ 0.53 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = ค่าความเป็นกรดระเหย ถ้าผลมากกว่า 0.46 pH คือคุณภาพดี จำนวน 3 ขวด ถ้าผลน้อยกว่าหรือเท่ากับ 0.46 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = ค่าความเป็นกรดระเหย ถ้าผลน้อยกว่าหรือเท่ากับ 0.24 pH คือคุณภาพดี จำนวน 2 ขวด ถ้าผลมากกว่า 0.24 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_{12} = คะแนนคุณภาพ ถ้าผลน้อยกว่าหรือเท่ากับ 6 คะแนน คือคุณภาพไม่ดี จำนวน 7 ขวด ถ้าผลมากกว่า 6 คะแนน ตัดสินใจต่อด้วยตัวแปรทำนาย X_{11} = ค่าแอลกอฮอล์ ถ้าผลน้อยกว่าหรือเท่ากับ 11.5 กรัม คือคุณภาพดี จำนวน 2 ขวด ถ้าผลมากกว่า 11.5 กรัม คือคุณภาพไม่ดี จำนวน 4 ขวด



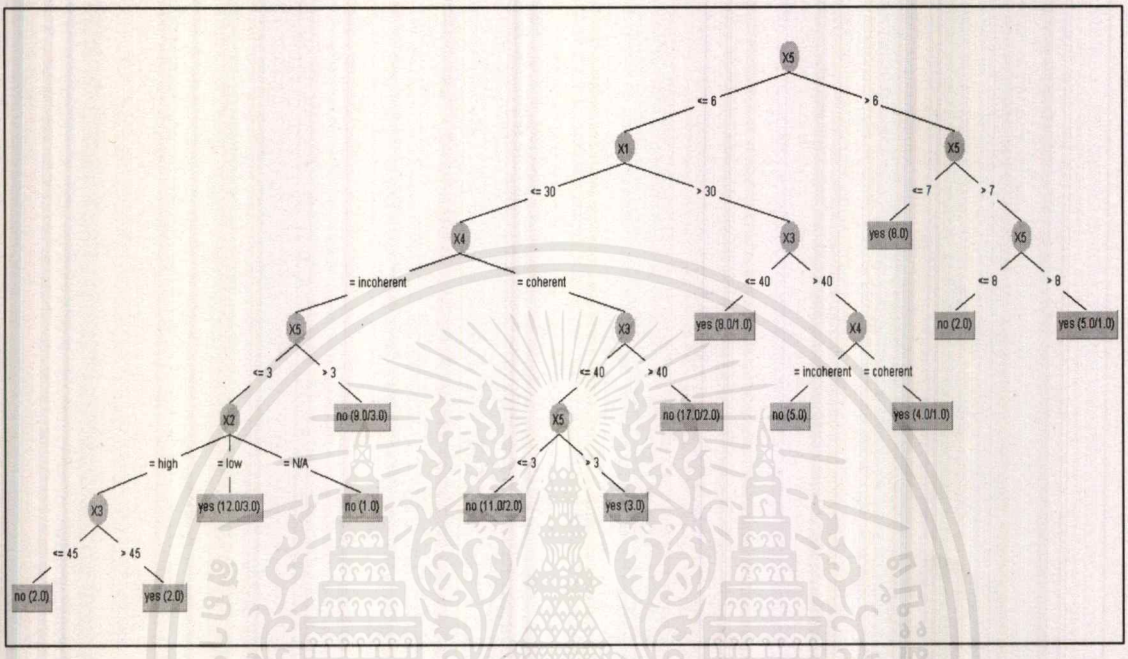
รูปที่ 4.90 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูป 4.90 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 39 คน มีค่าทำนายความถูกต้องจำนวน 24 คน คิดเป็นร้อยละ 61.5385 โดยมีค่าความไว = 0.7727 ค่าความจำเพาะ = 0.4118 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5870)^2$ = 0.3446

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.4 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มผสมผสาน (Hybrid Metods)

4.3.4.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทดสอบ 89 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

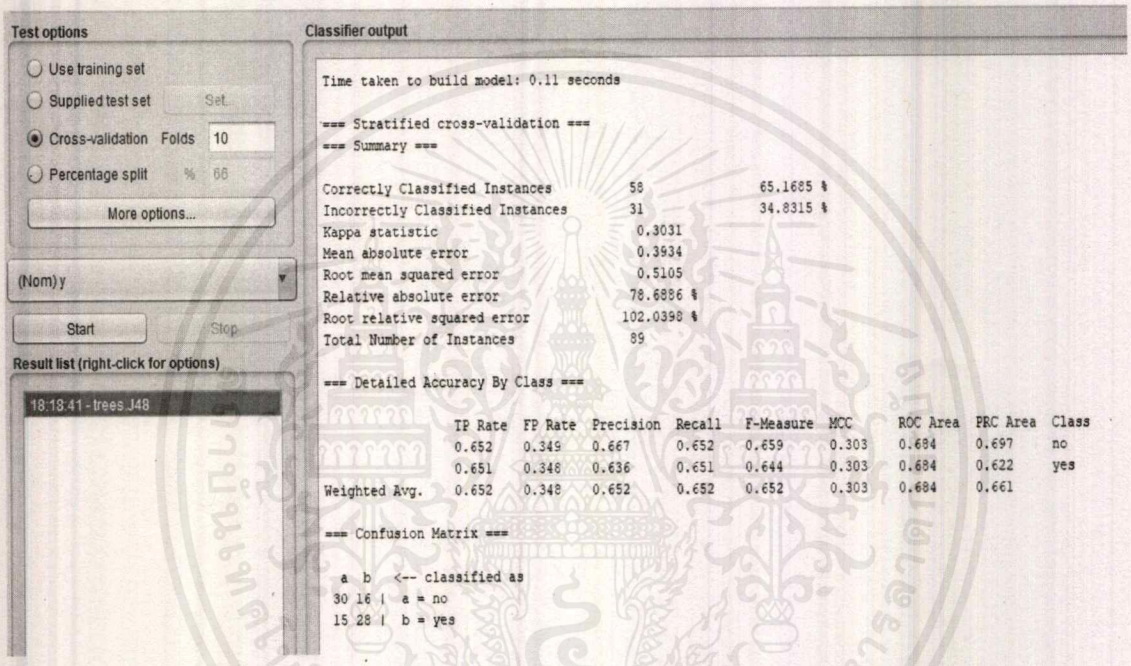


รูปที่ 4.91 ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10

จากรูป 4.91 แสดงต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โหนดรากแบ่งโดยตัวแปรทำนายคือ X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลมากกว่า 6 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าหรือเท่ากับ 7 ครั้ง คือได้ยินเสียงปกติ จำนวน 8 คน ถ้าผลมากกว่า 7 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าหรือเท่ากับ 8 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 2 คน ถ้าผลมากกว่า 8 ครั้ง คือได้ยินเสียงปกติ(จำนวน 5 คน ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าหรือเท่ากับ 6 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = อายุ ถ้าผลมากกว่า 30 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลน้อยกว่าหรือเท่ากับ 40 เดซิเบล คือได้ยินเสียงปกติ ถ้าผลมากกว่า 40 เดซิเบล ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = สัญญาณในการกระตุ้น ถ้าผลเป็น incoherent (สัญญาณไม่ต่อเนื่อง) คือได้ยินเสียงไม่ปกติ จำนวน 5 คน ถ้าผลเป็น coherent (สัญญาณต่อเนื่อง) คือได้ยินเสียงปกติ จำนวน 4 คน ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = อายุ ถ้าผลน้อยกว่าหรือเท่ากับ 30 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = สัญญาณในการกระตุ้น ถ้าผลเป็น coherent (สัญญาณต่อเนื่อง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลมากกว่า 40 เดซิเบล คือได้ยินเสียงไม่ปกติจำนวน 17 คน ถ้าผลน้อยกว่าหรือเท่ากับ 40 เดซิเบล ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าหรือเท่ากับ 3 ครั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

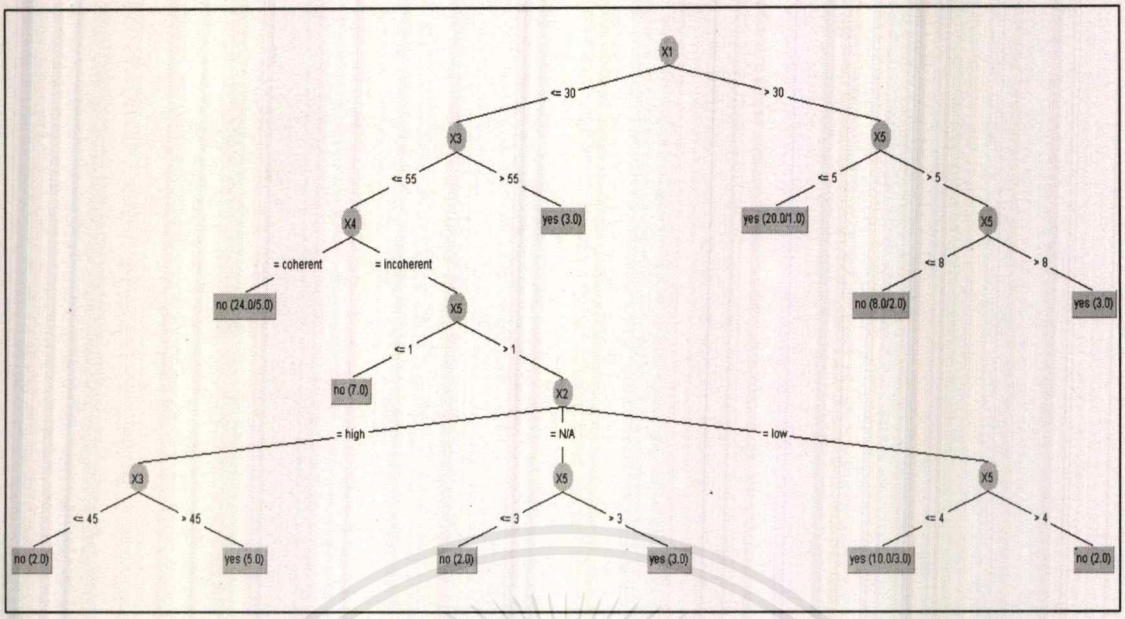
คือได้ยินเสียงไม่ปกติ (จำนวน 11 คน) ถ้าผลมากกว่า 3 ครั้ง คือได้ยินเสียงปกติ จำนวน 3 คน ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = สัญญาณในการกระตุ้น ถ้าผลเป็น incoherent (สัญญาณไม่ต่อเนื่อง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลมากกว่า 3 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 9 คน ถ้าผลน้อยกว่าหรือเท่ากับ 3 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = อัตราความเสี่ยงการเป็นโรคหูชั้นกลางอักเสบ ถ้าผลเป็น low (ความเสี่ยงน้อย) คือได้ยินเสียงปกติ จำนวน 12 คน ถ้าผลเป็น N/A (ไม่มีความเสี่ยง) คือได้ยินเสียงไม่ปกติ จำนวน 1 คน ถ้าผลเป็น high (ความเสี่ยงมาก) ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลน้อยกว่าหรือเท่ากับ 45 เดซิเบล คือได้ยินเสียงไม่ปกติ (จำนวน 2 คน) ถ้าผลมากกล่า 45 เดซิเบล คือได้ยินเสียงปกติ จำนวน 2 คน



รูปที่ 4.92 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10

จากรูป 4.92 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 89 คน มีค่าทำนายความถูกต้องจำนวน 58 คน คิดเป็นร้อยละ 65.1685 โดยมีค่าความไว = 0.6522 ค่าความจำเพาะ = 0.6511 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5105)^2$ = 0.2606

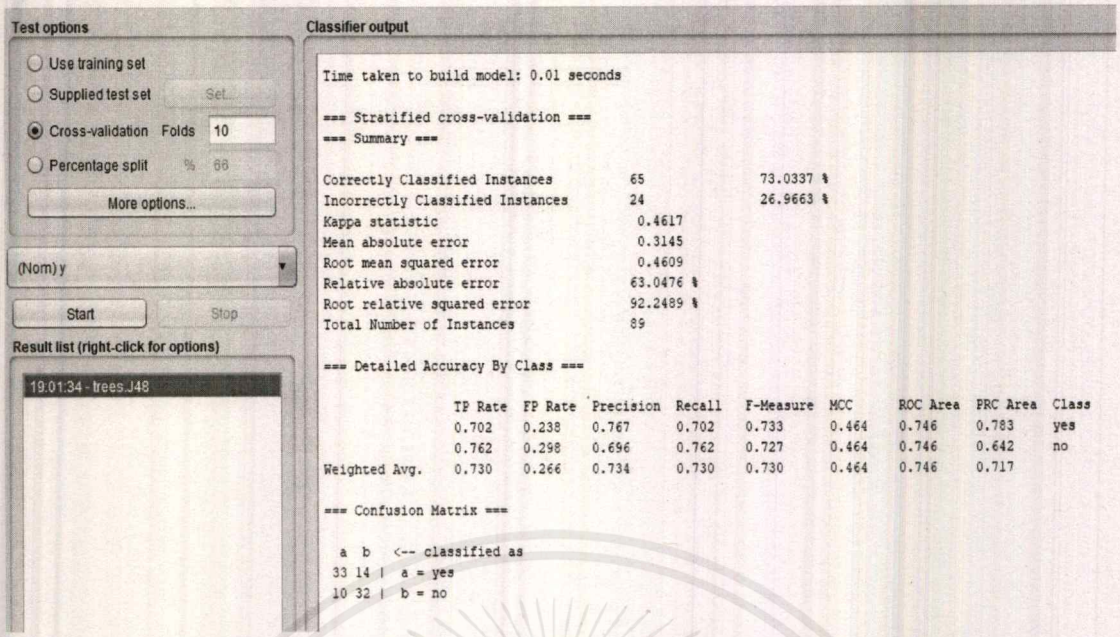
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.93 ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

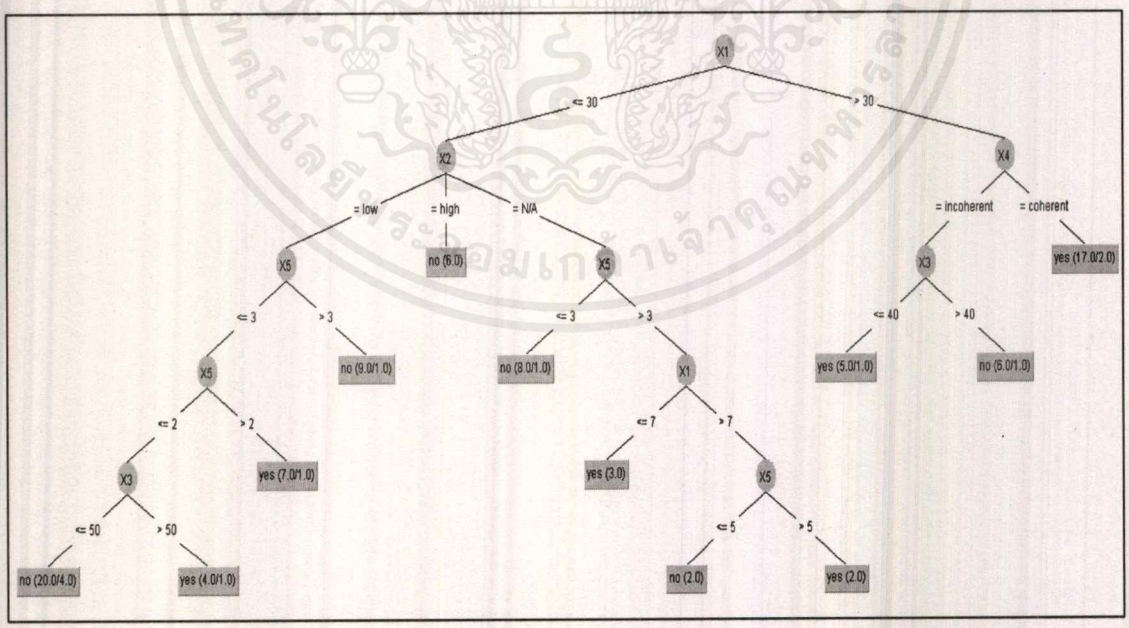
จากรูป 4.93 แสดงต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โหนดรากแบ่งโดยตัวแปรทำนาย X_1 = อายุ ถ้าผลมากกว่า 30 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าหรือเท่ากับ 5 ครั้ง คือได้ยินเสียงปกติ จำนวน 20 คน ถ้าผลมากกว่า 5 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าหรือเท่ากับ 8 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 8 คน ถ้าผลมากกว่า 8 ครั้ง คือได้ยินเสียงปกติ จำนวน 3 คน ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = อายุ ถ้าผลน้อยกว่าหรือเท่ากับ 30 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลมากกว่า 55 เดซิเบล คือได้ยินเสียงปกติ จำนวน 3 คน ถ้าผลน้อยกว่าหรือเท่ากับ 55 เดซิเบล ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = สัญญาณในการกระตุ้น ถ้าผลเป็น coherent (สัญญาณต่อเนื่อง) คือได้ยินเสียงไม่ปกติ จำนวน 24 คน ถ้าผลเป็น incoherent (สัญญาณไม่ต่อเนื่อง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าหรือเท่ากับ 1 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 7 คน ถ้าผลมากกว่า 1 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = อัตราความเสี่ยงการเป็นโรคหูชั้นกลางอักเสบ ถ้าผลเป็น high (ความเสี่ยงมาก) ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลน้อยกว่าหรือเท่ากับ 45 เดซิเบล คือได้ยินเสียงไม่ปกติ จำนวน 2 คน ถ้าผลมากกว่า 45 เดซิเบล คือได้ยินเสียงปกติ จำนวน 5 คน ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = อัตราความเสี่ยงการเป็นโรคหูชั้นกลางอักเสบ ถ้าผลเป็น N/A (ไม่มีความเสี่ยง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าหรือเท่ากับ 3 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 2 คน ถ้ามากกว่า 3 ครั้ง คือได้ยินเสียงปกติ จำนวน 3 คน ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = อัตราความเสี่ยงการเป็นโรคหูชั้นกลางอักเสบ ถ้าผลเป็น low (ความเสี่ยงน้อย) ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าหรือเท่ากับ 4 ครั้ง คือได้ยินเสียงปกติ จำนวน 10 คน ถ้ามากกว่า 4 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 2 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.94 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

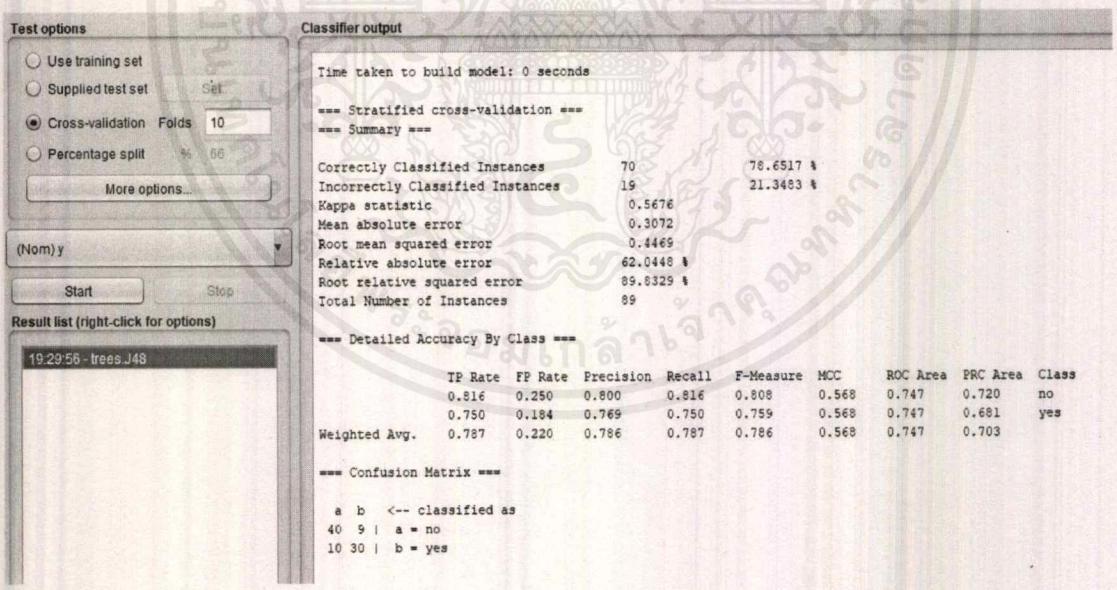
จากรูป 4.94 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 89 คน มีค่าทำนายความถูกต้องจำนวน 65 คน คิดเป็นร้อยละ 73.0337 โดยมีค่าความไว = 0.7021 ค่าความจำเพาะ = 0.7619 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4609)^2$ = 0.2124



รูปที่ 4.95 ต้นไม้ตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูป 4.95 แสดงต้นไม้เพื่อการตัดสินใจ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โหนรากแบ่งโดยตัวแปรทำนายคือ X_1 = อายุ ถ้าผลมากกว่า 30 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = สัญญาณในการกระตุ้น ถ้าผลเป็น coherent (สัญญาณต่อเนื่อง) คือได้ยินเสียงปกติ จำนวน 17 คน ถ้าผลเป็น incoherent (สัญญาณไม่ต่อเนื่อง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลน้อยกว่าหรือเท่ากับ 40 เดซิเบล คือได้ยินเสียงปกติ จำนวน 5 คน ถ้าผลมากกว่า 40 เดซิเบล คือได้ยินเสียงไม่ปกติ จำนวน 6 คน ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = อายุ ถ้าผลมากกว่า 30 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = อัตราความเสี่ยงการเป็นโรคหูชั้นกลางอักเสบ ถ้าผลเป็น high (ความเสี่ยงมาก) คือได้ยินเสียงไม่ปกติ จำนวน 6 คน เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = อัตราความเสี่ยงการเป็นโรคหูชั้นกลางอักเสบ ถ้าผลเป็น low (ความเสี่ยงน้อย) ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลมากกว่า 3 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 9 คน ถ้าผลน้อยกว่าหรือเท่ากับ 3 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลมากกว่า 2 ครั้ง คือได้ยินเสียงปกติ จำนวน 7 คน ถ้าผลน้อยกว่าหรือเท่ากับ 2 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ความดังที่ได้รับ ถ้าผลน้อยกว่าหรือเท่ากับ 50 เดซิเบล คือได้ยินเสียงไม่ปกติ จำนวน 2 คน ถ้าผลมากกว่า 50 เดซิเบล คือได้ยินเสียงปกติ จำนวน 4 คน ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = อัตราความเสี่ยงการเป็นโรคหูชั้นกลางอักเสบ ถ้าผลเป็น N/A (ไม่มีความเสี่ยง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าหรือเท่ากับ 3 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 8 คน ถ้าผลมากกว่า 3 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = อายุ ถ้าผลน้อยกว่าหรือเท่ากับ 7 เดือน คือได้ยินเสียงปกติ จำนวน 3 คน ถ้าผลมากกว่า 7 เดือน ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง ถ้าผลน้อยกว่าหรือเท่ากับ 5 ครั้ง คือได้ยินเสียงไม่ปกติ จำนวน 2 คน ถ้ามากกว่า 5 ครั้ง คือได้ยินเสียงปกติ จำนวน 2 คน

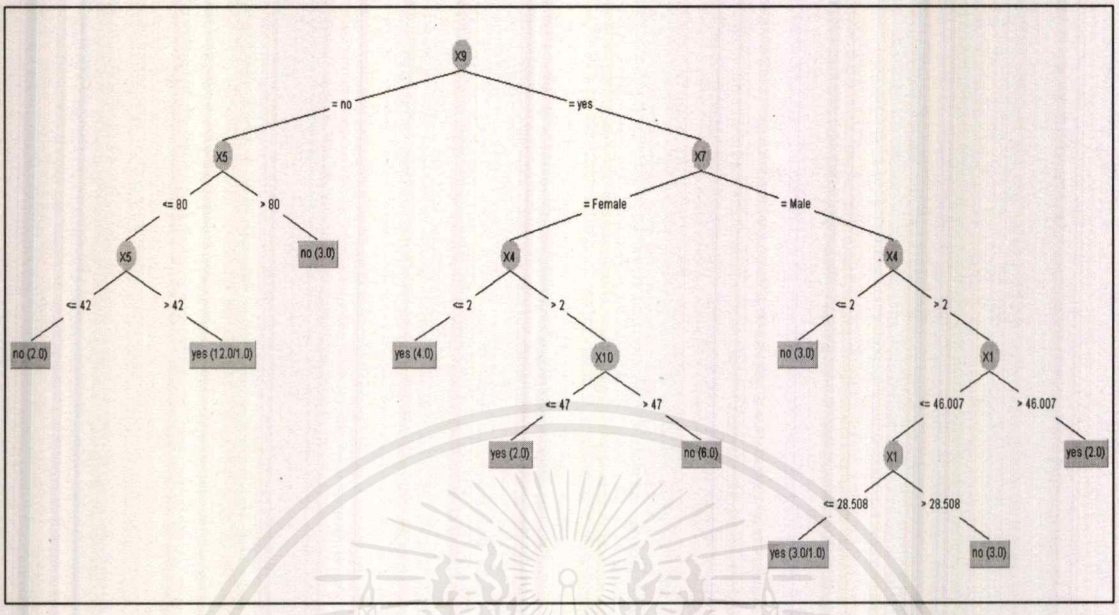


รูปที่ 4.96 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูป 4.96 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 89 คน มีค่าทำนายความถูกต้องจำนวน 70 คน คิดเป็นร้อยละ 78.6517 โดยมีค่าความไว = 0.8163 ค่าความจำเพาะ = 0.7500 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4469)^2$ = 0.1997

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

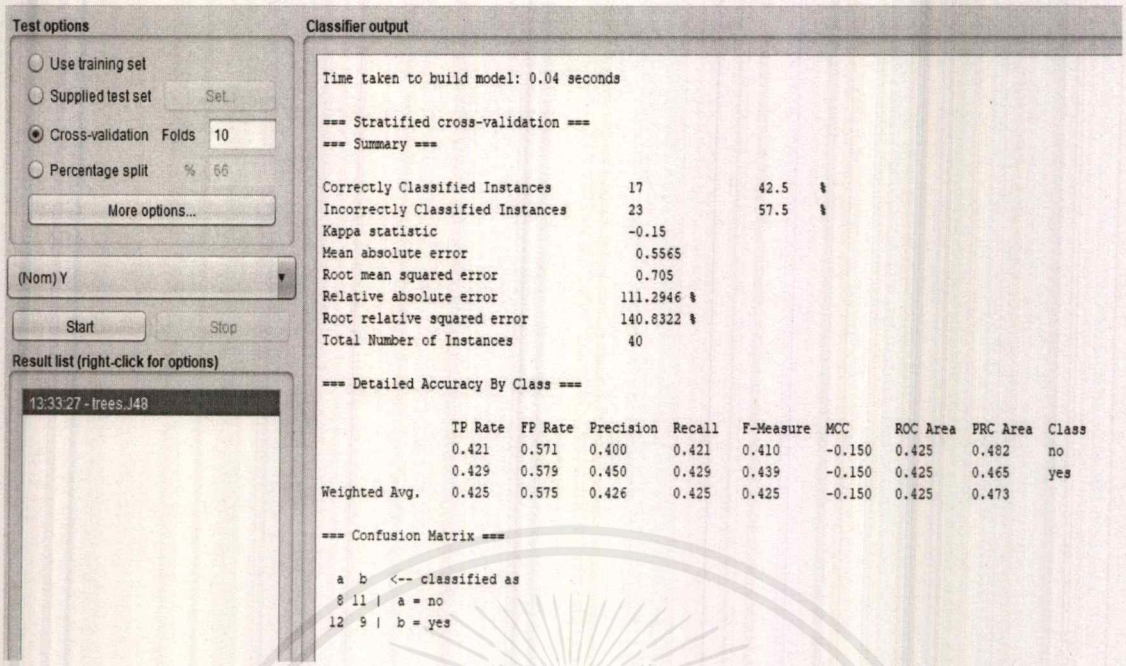
4.3.4.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทดสอบ 40 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.97 ต้นไม้ตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10

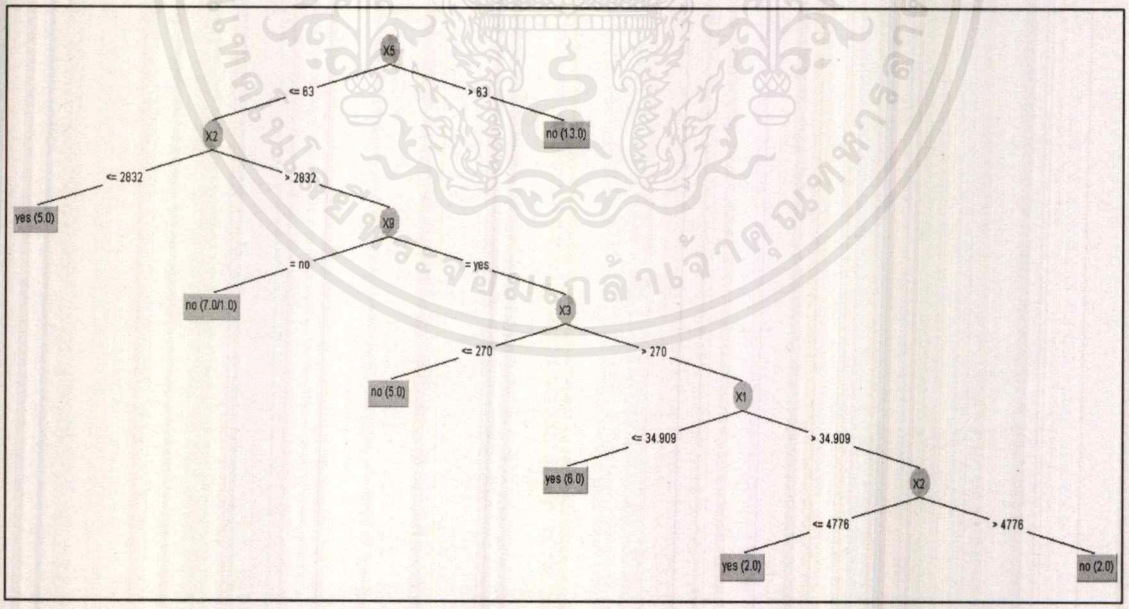
จากรูป 4.97 แสดงต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า โหนดรากแบ่งโดยตัวแปรทำนายคือ X_9 = สถานภาพสมรส ถ้าผลเป็น no (โสด) ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = อายุ ถ้าผลมากกว่า 80 ปี คือไม่ชำระ จำนวน 3 คน ถ้าผลน้อยกว่าหรือเท่ากับ 80 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = อายุ ถ้าผลน้อยกว่าหรือเท่ากับ 42 ปี คือไม่ชำระ จำนวน 2 คน ถ้าผลมากกว่า 42 ปี คือชำระ จำนวน 12 คน ตัดสินใจต่อด้วยตัวแปรทำนาย X_9 = สถานภาพสมรส ถ้าผลเป็น yes (แต่งงานแล้ว) ตัดสินใจต่อด้วยตัวแปรทำนาย X_7 = เพศ ถ้าผลเป็น Female (เพศหญิง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = จำนวนบัตรเครดิตที่มี ถ้าผลน้อยกว่าหรือเท่ากับ 2 บัตร คือชำระ จำนวน 4 คน ถ้าผลมากกว่า 2 บัตร ตัดสินใจต่อด้วยตัวแปรทำนาย X_{10} = ยอดเงินเหลือในบัตรเครดิต ถ้าผลน้อยกว่าหรือเท่ากับ 47 ดอลลาร์ คือชำระ จำนวน 2 คน ถ้าผลมากกว่า 47 ดอลลาร์ คือไม่ชำระ จำนวน 6 คน ตัดสินใจต่อด้วยตัวแปรทำนาย X_7 = เพศ ถ้าผลเป็น male (เพศชาย) ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = จำนวนบัตรเครดิตที่มี ถ้าผลน้อยกว่าหรือเท่ากับ 2 บัตร คือไม่ชำระ จำนวน 3 คน ถ้าผลมากกว่า 2 บัตร ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = รายได้ต่อเดือน ถ้าผลมากกว่า 46.007 ดอลลาร์ คือชำระ จำนวน 2 คน ถ้าผลน้อยกว่าหรือเท่ากับ 46.007 ดอลลาร์ ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = รายได้ต่อเดือน ถ้าผลน้อยกว่าหรือเท่ากับ 28.508 ดอลลาร์ คือชำระ จำนวน 3 ค ถ้าผลมากกว่า 28.508 ดอลลาร์ คือไม่ชำระ จำนวน 3 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.98 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10

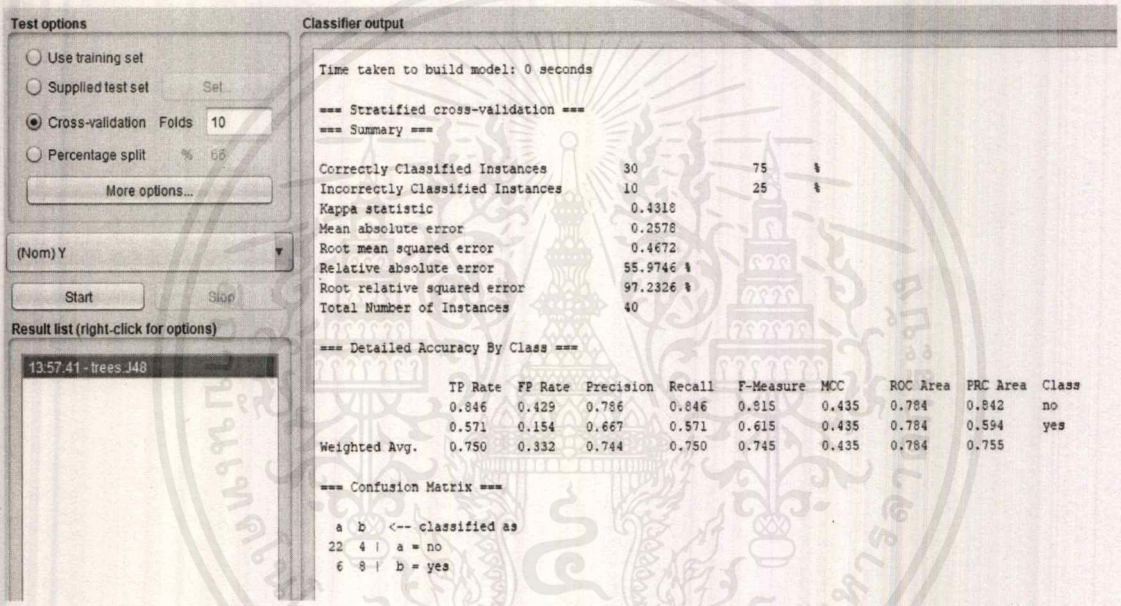
จากรูป 4.98 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 40 คน มีค่าทำนายความถูกต้องจำนวน 17 คน คิดเป็นร้อยละ 42.5 โดยมีค่าความไว = 0.4210 ค่าความจำเพาะ = 0.4286 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.7050)^2$ = 0.4970



รูปที่ 4.99 ต้นไม้ตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

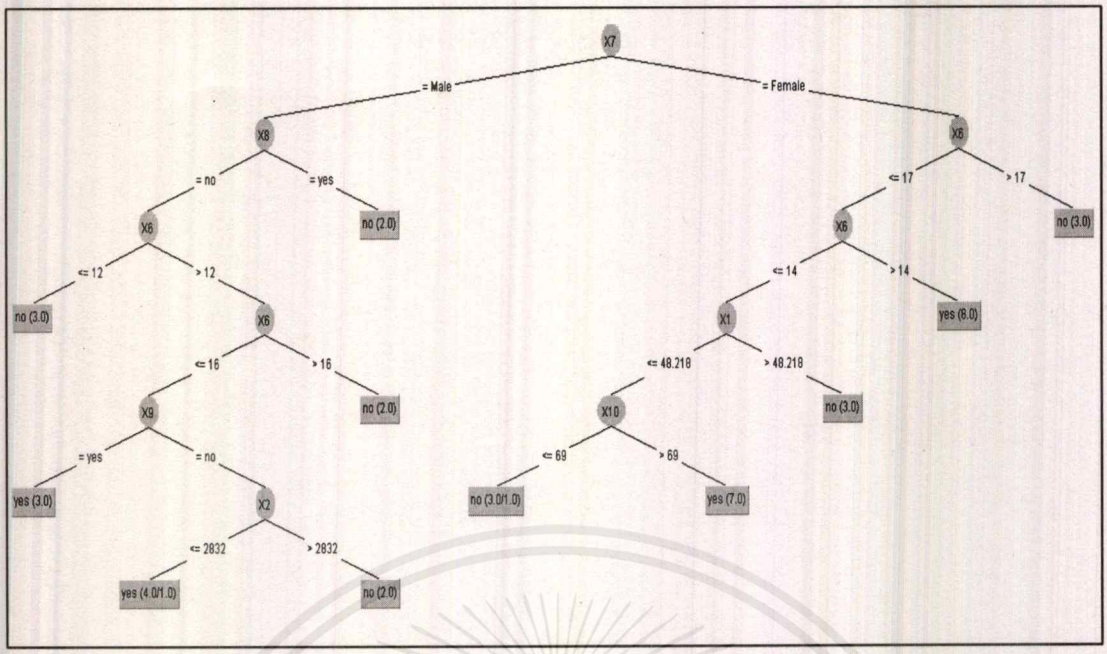
จากรูป 4.99 แสดงต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า โหนดรากแบ่งโดยตัวแปรทำนายคือ X_5 = อายุ ถ้าผลมากกว่า 63 ปี คือไม่ชำระ จำนวน 13 คน ถ้าผลน้อยกว่าหรือเท่ากับ 63 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = วงเงินในบัตรเครดิต ถ้าผลน้อยกว่าหรือเท่ากับ 2832 ดอลลาร์ คือชำระ จำนวน 5 คน ถ้าผลมากกว่า 2832 ดอลลาร์ ตัดสินใจต่อด้วยตัวแปรทำนาย X_9 = สถานภาพสมรส ถ้าผลเป็น no (โสด) คือไม่ชำระ จำนวน 7 คน ถ้าผลเป็น yes (แต่งงานแล้ว) ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = จำนวนครั้งในการใช้บัตรเครดิตต่อปี ถ้าผลน้อยกว่าหรือเท่ากับ 270 ครั้ง คือไม่ชำระ จำนวน 5 คน ถ้าผลมากกว่า 270 ครั้ง ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = รายได้ต่อเดือน ถ้าผลน้อยกว่าหรือเท่ากับ 34.909 ดอลลาร์ คือชำระ จำนวน 6 คน ถ้าผลมากกว่า 34.909 ดอลลาร์ คือไม่ชำระ จำนวน 3 คน ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = วงเงินในบัตรเครดิต ถ้าผลน้อยกว่าหรือเท่ากับ 4776 ดอลลาร์ คือชำระ จำนวน 2 คน ถ้าผลมากกว่า 4776 ดอลลาร์ คือไม่เชำระ จำนวน 2 คน



รูปที่ 4.100 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เครดิต เมื่อ Random Seed 20

จากรูป 4.100 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 40 คน มีค่าทำนายความถูกต้องจำนวน 30 คน คิดเป็นร้อยละ 75 โดยมีค่าความไว = 0.8462 ค่าความจำเพาะ = 0.5714 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4672)^2$ = 0.2183

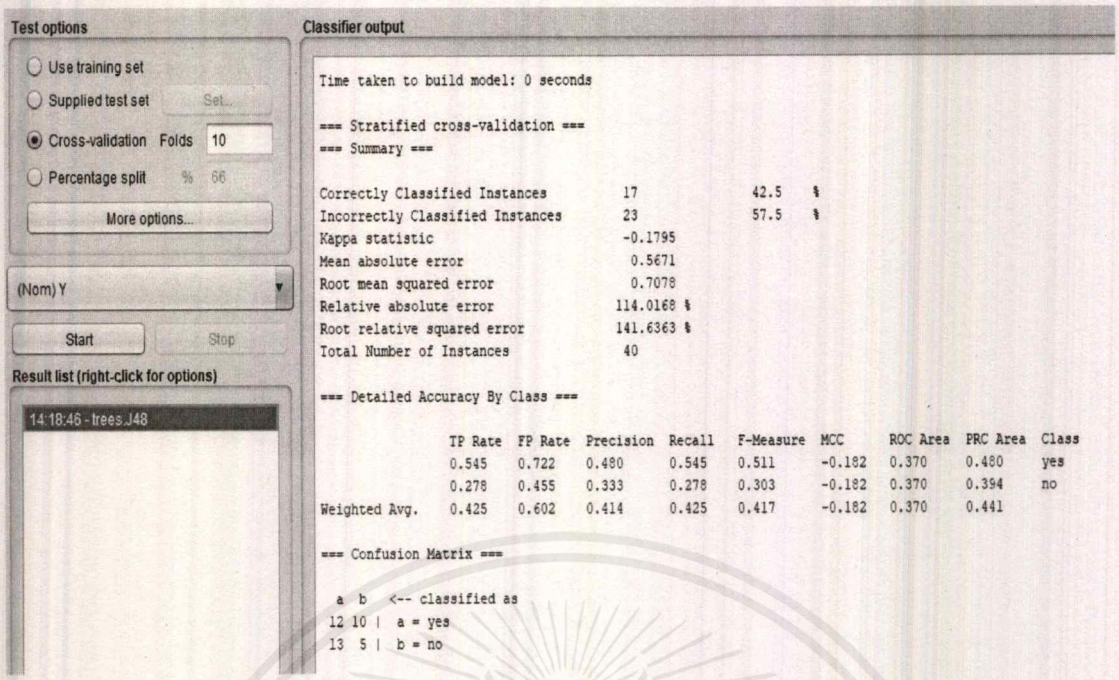
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.101 ต้นไม้ตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูป 4.101 แสดงต้นไม้เพื่อการตัดสินใจ ยอดคงเหลือในบัตรเครดิตของลูกค้า โหนดรากแบ่งโดยตัวแปรทำนายคือ X_8 = การศึกษา ถ้าผลเป็น yes (กำลังศึกษา) คือไม่ชำระ จำนวน 2 คน ถ้าผลเป็น no (สำเร็จการศึกษาแล้ว) ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = จำนวนปีในการศึกษา ถ้าผลน้อยกว่าหรือเท่ากับ 12 ปี คือไม่ชำระ จำนวน 3 คน ถ้าผลมากกว่า 12 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = จำนวนปีในการศึกษา ถ้าผลมากกว่า 16 ปี คือไม่ชำระ จำนวน 2 คน ถ้าผลน้อยกว่าหรือเท่ากับ 16 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_9 = สถานภาพสมรส ถ้าผลเป็น yes (แต่งงานแล้ว) คือชำระ จำนวน 3 คน ถ้าผลเป็น no (โสด) ตัดสินใจต่อด้วยตัวแปรทำนาย X_2 = วงเงินในบัตรเครดิต ถ้าผลน้อยกว่าหรือเท่ากับ 2832 ดอลลาร์ คือชำระ จำนวน 4 คน ถ้าผลมากกว่า 2832 ดอลลาร์ คือไม่ชำระ จำนวน 2 คน ตัดสินใจต่อด้วยตัวแปรทำนาย X_7 = เพศ ถ้าผลเป็น Female (เพศหญิง) ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = จำนวนปีในการศึกษา ถ้าผลมากกว่า 17 ปี คือไม่ชำระ จำนวน 3 คน ถ้าผลน้อยกว่าหรือเท่ากับ 17 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = จำนวนปีในการศึกษา ถ้าผลมากกว่า 14 ปี คือชำระ จำนวน 8 คน ถ้าผลน้อยกว่าหรือเท่ากับ 14 ปี ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = รายได้ต่อเดือน ถ้าผลมากกว่า 48.218 ดอลลาร์ คือไม่ชำระ จำนวน 3 คน ถ้าผลน้อยกว่าหรือเท่ากับ 48.218 ดอลลาร์ ตัดสินใจต่อด้วยตัวแปรทำนาย X_{10} = ยอดเงินเหลือในบัตรเครดิต ถ้าผลน้อยกว่าหรือเท่ากับ 69 ดอลลาร์ คือไม่ชำระ จำนวน 3 คน ถ้าผลมากกว่า 69 ดอลลาร์ คือชำระ จำนวน 7 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

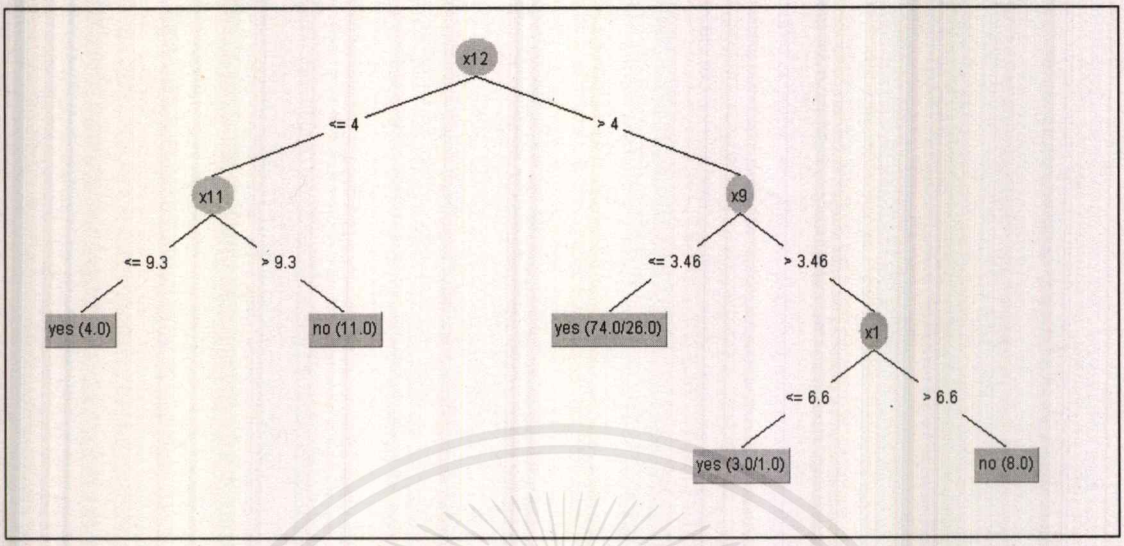


รูปที่ 4.102 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูป 4.102 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 40 คน มีค่าทำนายความถูกต้องจำนวน 17 คน คิดเป็นร้อยละ 42.5 โดยมีค่าความไว = 0.5454 ค่าความจำเพาะ = 0.2778 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.7078)^2$ = 0.5010

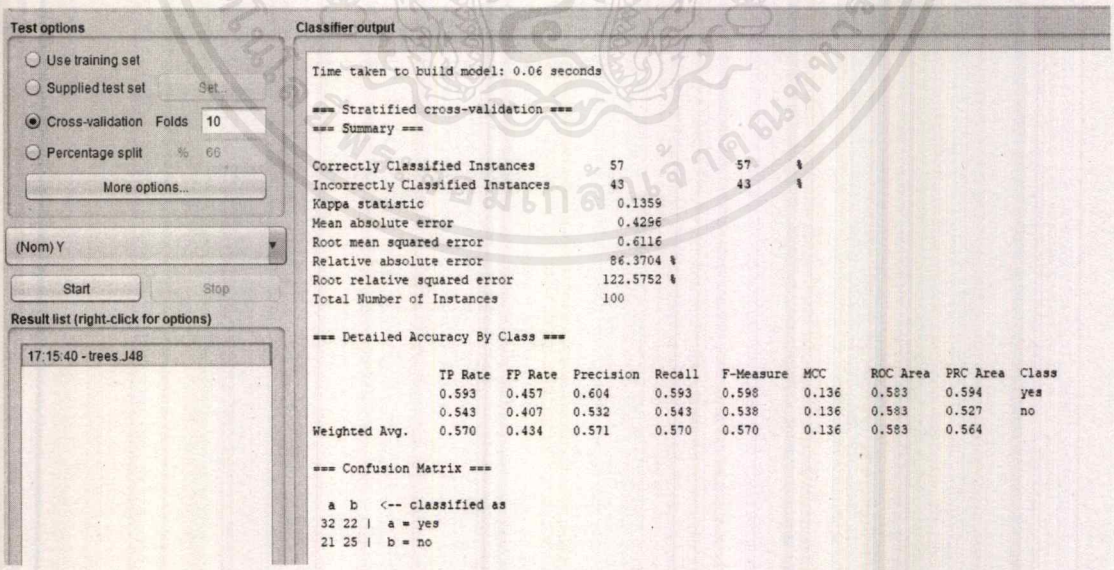
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.4.3 คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทดสอบ 100 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.103 ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 10

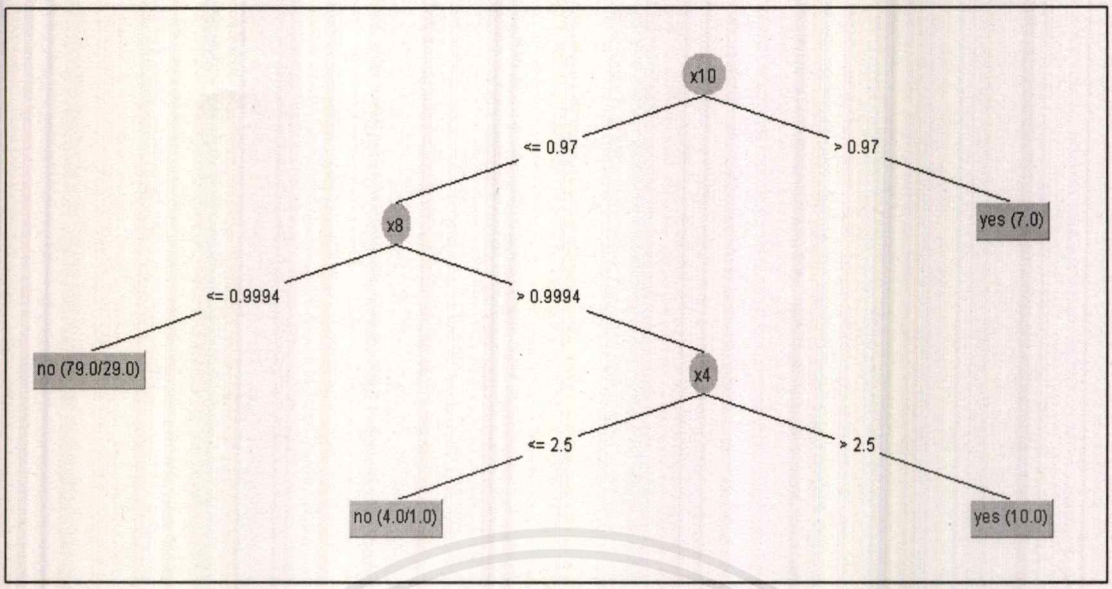
จากรูป 4.103 แสดงต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง โหนดรากแบ่งโดยตัวแปรทำนายคือ X_{12} = คะแนนคุณภาพ ถ้าผลน้อยกว่าเท่ากับ 4 คะแนน ตัดสินใจต่อด้วยตัวแปรทำนาย X_{11} = ค่าแอลกอฮอล์ ถ้าผลน้อยกวาเท่ากับ 9.3 กรัม คือคุณภาพดี จำนวน 4 ขวด ถ้าผลมากกว่า 9.3 กรัม คือคุณภาพไม่ดี จำนวน 11 ขวด ตัดสินใจต่อด้วยตัวแปรทำนาย X_{12} = คะแนนคุณภาพ ถ้าผลมากกว่า 4 คะแนน ตัดสินใจต่อด้วยตัวแปรทำนาย X_9 = ค่าความเป็นกรดหรือด่าง ถ้าน้อยกว่าเท่ากับ 3.46 pH คือคุณภาพดี จำนวน 74 ขวด ถ้าผลมากกว่า 3.46 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_1 = ค่าความเป็นกรด ถ้าผลน้อยกว่าเท่ากับ 6.6 pH คือคุณภาพดี จำนวน 3 ขวด ถ้าผลมากกว่า 6.6 pH คือคุณภาพไม่ดี จำนวน 8 ขวด



รูปที่ 4.104 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 10

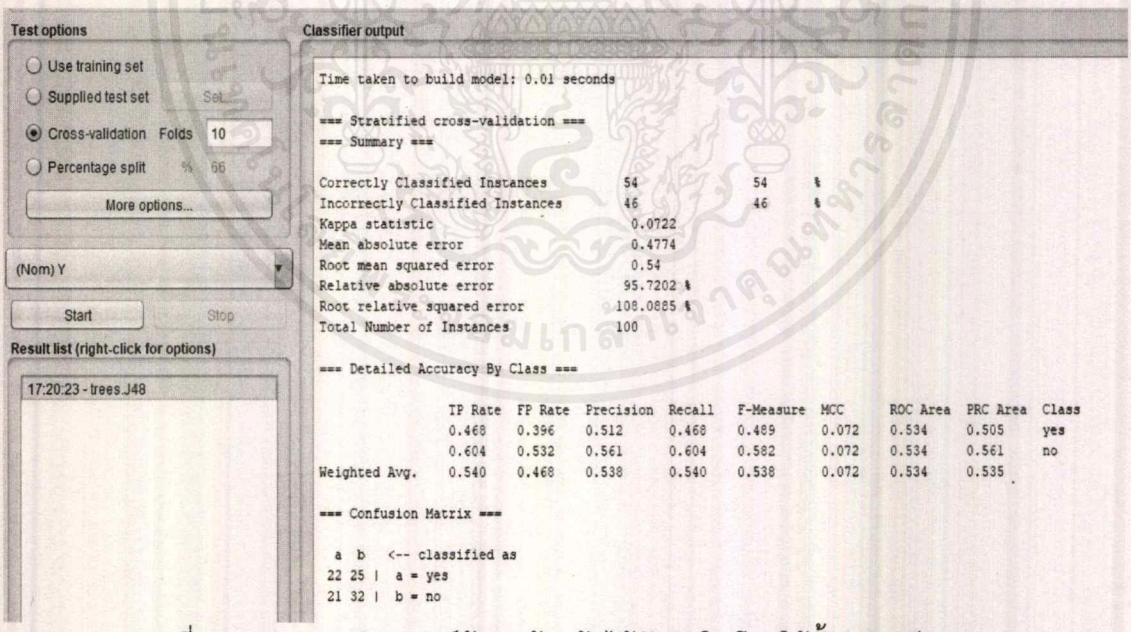
จากรูป 4.104 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 100 คน มีค่าทำนายความถูกต้องจำนวน 57 คน คิดเป็นร้อยละ 57 โดยมีค่าความไว = 0.5926 ค่าความจำเพาะ = 0.5435 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6116)^2$ = 0.3741

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.105 ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 20

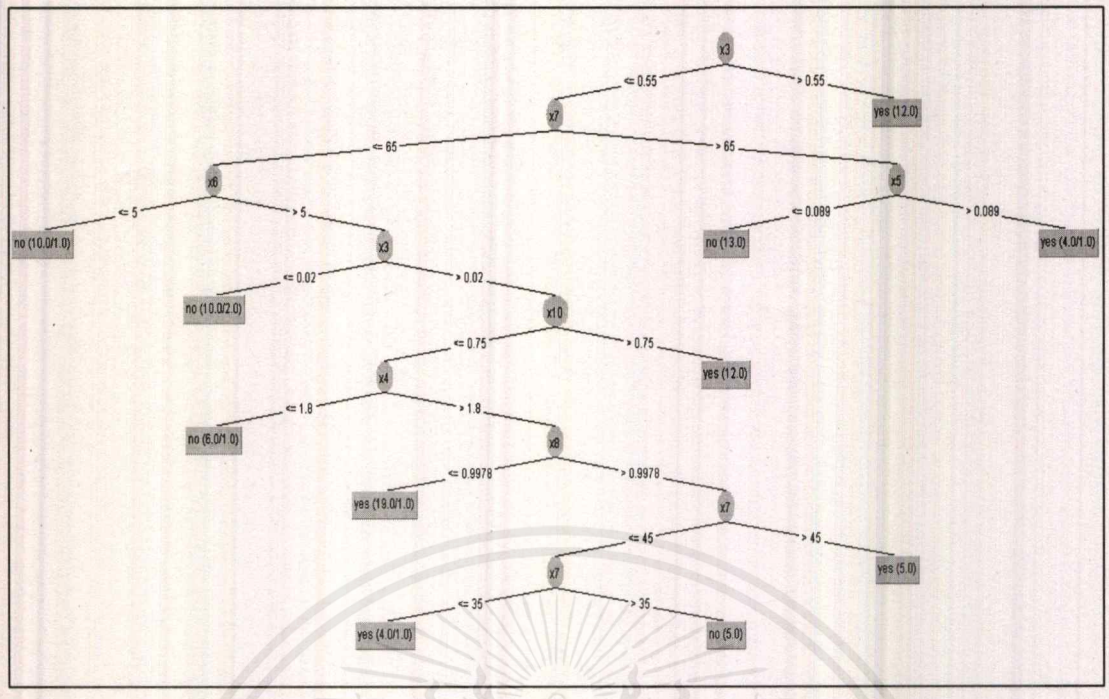
จากรูป 4.105 แสดงต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง โหนดรากแบ่งโดยตัวแปรทำนายคือ X_{10} = ค่าเกลือของกรดซัลฟิวริก ถ้าผลมากกว่า 0.97 pH คือคุณภาพดี จำนวน 7 ขวด ถ้าผลน้อยกว่าเท่ากับ 0.97 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_8 = ค่าความหนาแน่น ถ้าผลน้อยกว่าเท่ากับ 0.9994 กิโลกรัมต่อลูกบาศก์เมตร คือคุณภาพไม่ดี จำนวน 79 ขวด ถ้าผลมากกว่า 0.9994 กิโลกรัมต่อลูกบาศก์เมตร ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = ค่าน้ำตาลรีดิวซ์ ถ้าผลน้อยกว่าเท่ากับ 2.5 กรัม คือคุณภาพไม่ดี จำนวน 4 ขวด ถ้าผลมากกว่า 2.5 กรัม คือคุณภาพดี จำนวน 10 ขวด



รูปที่ 4.106 ผลการวิเคราะห์ข้อมูลด้วยต้นไม้ตัดสินใจ โดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 20

จากรูป 4.106 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 100 คน มีค่าทำนายความถูกต้องจำนวน 54 คน คิดเป็นร้อยละ 54 โดยมีค่าความไว = 0.4681 ค่าความจำเพาะ = 0.6038 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.54)^2$ = 0.2916

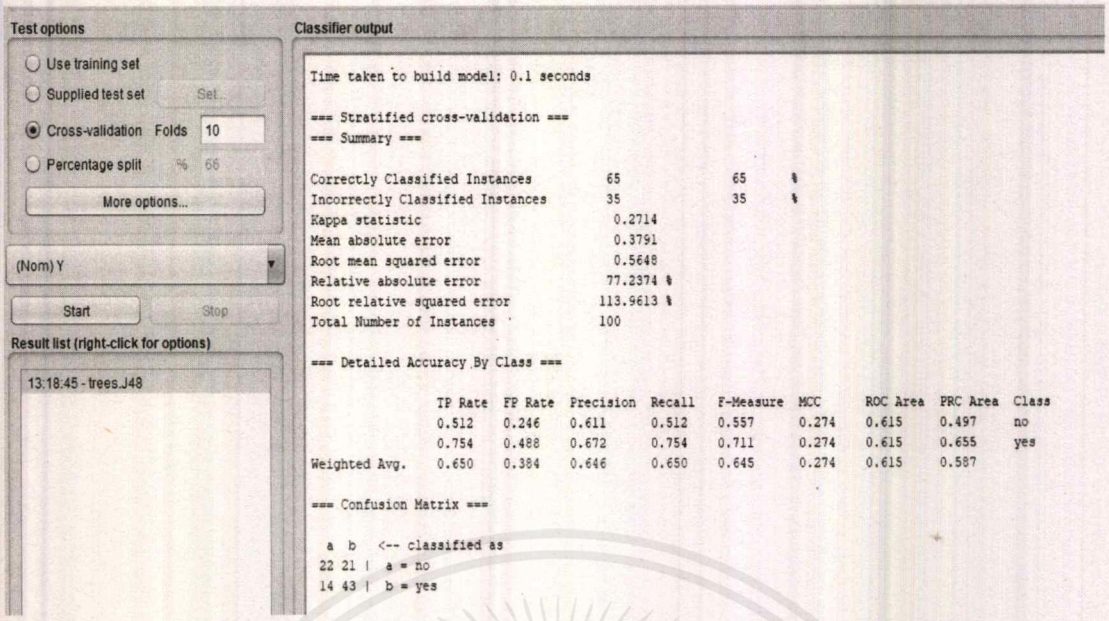
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.107 ต้นไม้ตัดสินใจ คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูป 4.107 แสดงต้นไม้เพื่อการตัดสินใจ คุณภาพไวน์แดง โหนดรากแบ่งโดยตัวแปรทำนายคือ X_3 = ค่ากรดซิตริก ถ้าผลมากกว่า 0.55 pH คือคุณภาพดี จำนวน 12 ขวด ถ้าผลน้อยกว่าเท่ากับ 0.55 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_7 = ค่าซัลเฟอร์ไดออกไซด์ทั้งหมด ถ้าผลมากกว่า 65 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_5 = ค่าคลอไรด์ ถ้าน้อยกว่าเท่ากับ 0.089 กรัม คือคุณภาพไม่ดี จำนวน 13 ขวด ถ้าผลมากกว่า 0.089 กรัม คือคุณภาพดี จำนวน 4 ขวด ตัดสินใจต่อด้วยตัวแปรทำนาย X_7 = ค่าซัลเฟอร์ไดออกไซด์ทั้งหมด ถ้าผลน้อยกว่าเท่ากับ 65 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_6 = ค่าซัลเฟอรไดออกไซด์ ถ้าผลน้อยกว่าเท่ากับ 5 กรัม คือคุณภาพไม่ดี จำนวน 10 ขวด ถ้าผลมากกว่า 5 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_3 = ค่ากรดซิตริก ถ้าผลน้อยกว่าเท่ากับ 0.02 pH คือคุณภาพไม่ดี จำนวน 10 ขวด ถ้าผลมากกว่า 0.02 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_{10} = ค่าเกลือของกรดซัลฟิวริก ถ้าผลมากกว่า 0.75 pH คือคุณภาพดี จำนวน 12 ขวด ถ้าผลน้อยกว่าเท่ากับ 0.75 pH ตัดสินใจต่อด้วยตัวแปรทำนาย X_4 = ค่าน้ำตาลรีดิวซ์ ถ้าผลน้อยกว่าเท่ากับ 1.8 กรัม คือคุณภาพไม่ดี จำนวน 6 ขวด ถ้าผลมากกว่า 1.8 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_8 = ค่าความหนาแน่น ถ้าผลน้อยกว่าเท่ากับ 0.9978 กิโลกรัมต่อลูกบาศก์เมตร คือคุณภาพดี จำนวน 19 ขวด ถ้าผลมากกว่า 0.9978 กิโลกรัมต่อลูกบาศก์เมตร ตัดสินใจต่อด้วยตัวแปรทำนาย X_7 = ค่าซัลเฟอร์ไดออกไซด์ทั้งหมด ถ้าผลมากกว่า 45 กรัม คือคุณภาพดี จำนวน 5 ขวด ถ้าผลน้อยกว่าเท่ากับ 45 กรัม ตัดสินใจต่อด้วยตัวแปรทำนาย X_7 = ค่าซัลเฟอร์ไดออกไซด์ทั้งหมด ถ้าผลน้อยกว่าเท่ากับ 35 กรัม คือคุณภาพดี จำนวน 4 ขวด ถ้าผลมากกว่า 35 กรัม คือคุณภาพไม่ดี จำนวน 5 ขวด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.108 ผลการวิเคราะห์ข้อมูลด้วยวิธีต้นไม้ตัดสินใจ ตัวโดยใช้ขั้นตอนวิธี J48 คุณภาพไวน์แดง เมื่อ Random Seed 30

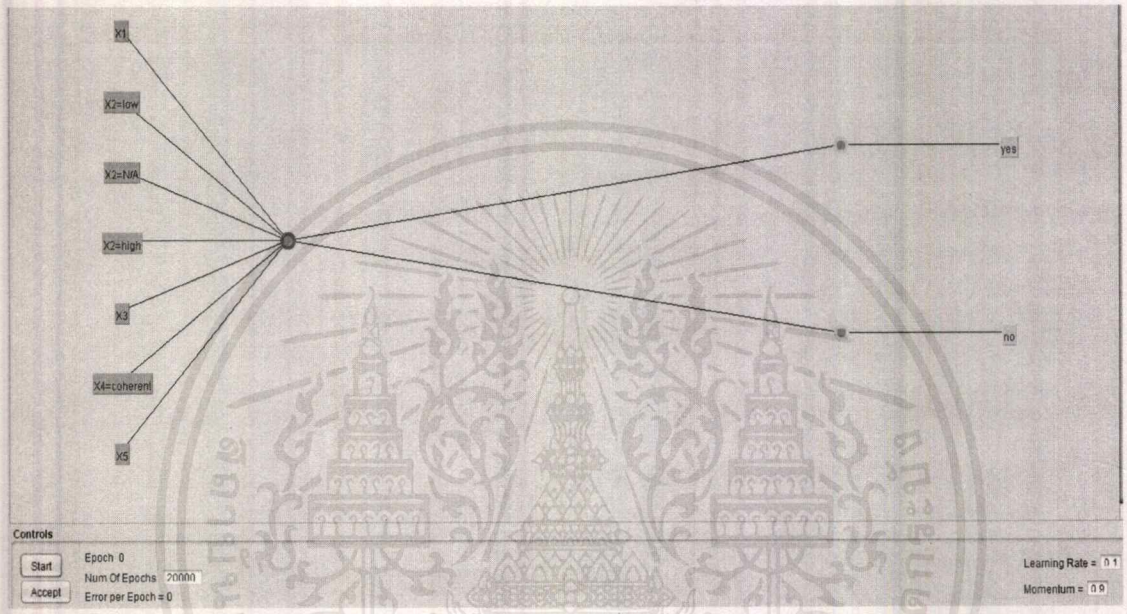
จากรูป 4.108 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 100 คน มีค่าทำนายความถูกต้องจำนวน 65 คน คิดเป็นร้อยละ 65 โดยมีค่าความไว = 0.5116 ค่าความจำเพาะ = 0.7544 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย $(0.5648)^2 = 0.3190$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม โดยใช้ขั้นตอนวิธี Multilayer Perceptron

4.4.1 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน (Over Sampling)

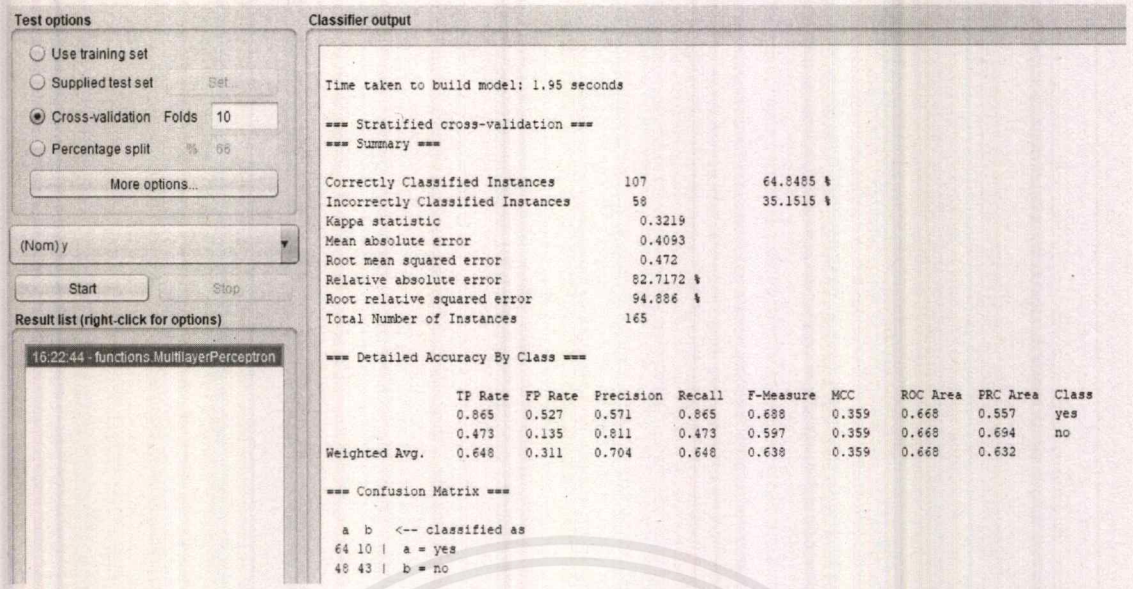
4.4.1.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทดสอบ 165 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.109 โครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือน้ำหนวก เมื่อ Random Seed 10, 20 และ 30

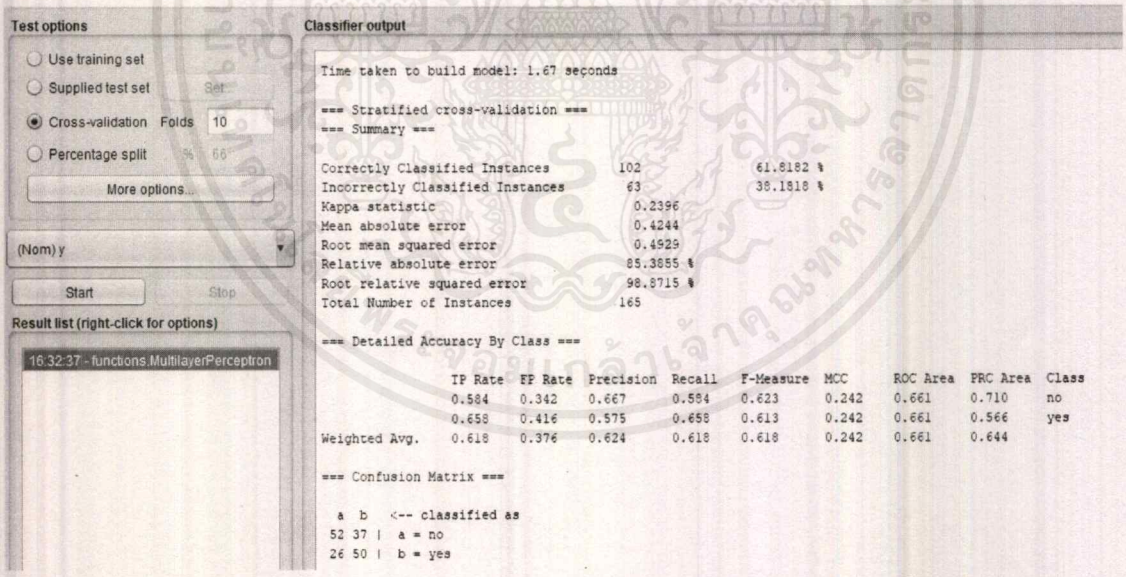
จากรูป 4.109 โครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก ชั้นข้อมูลเข้า อยู่ด้านซ้ายสุดซึ่งมีคุณลักษณะทั้งหมด 5 ตัวแปร (X_1-X_5) ชั้นซ่อน มี 1 ชั้นซ่อน ชั้นข้อมูลออกหรือชั้นข้อมูลผลลัพธ์ อยู่ด้านขวามือ มีคำตอบคือ Yes กับ No Epoch จะรันการทำงานตั้งแต่รอบที่ 1 ถึง 20,000 Error per Epoch แสดงความคลาดเคลื่อนต่อรอบ คือ Learning Rate เท่ากับ 0.1 และ Momentum เท่ากับ 0.9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.110 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10

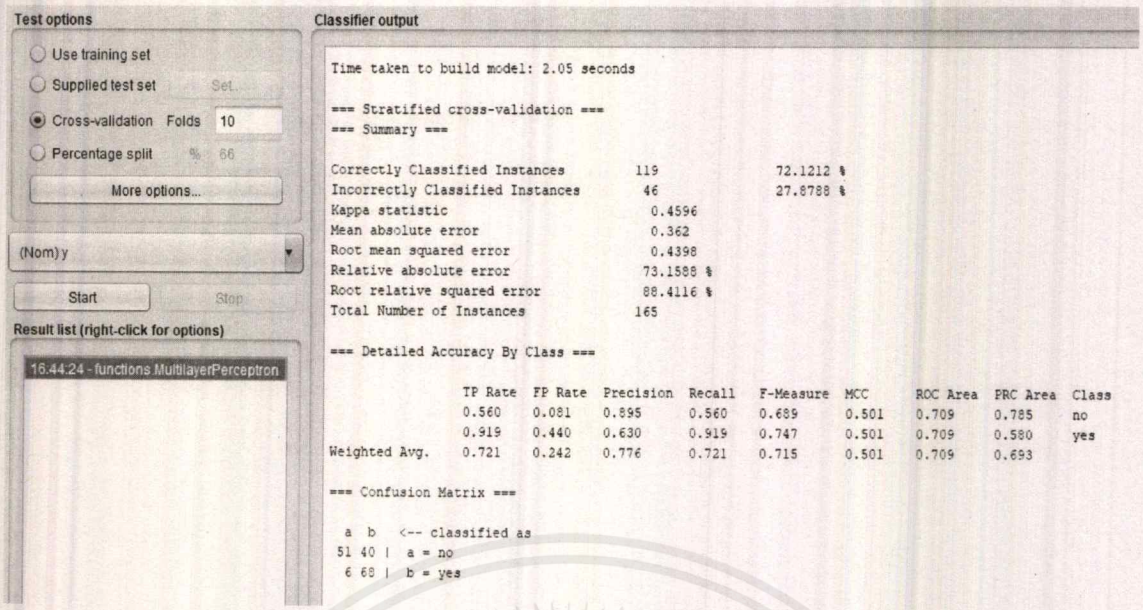
จากรูป 4.110 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 165 คน มีค่าทำนายความถูกต้องจำนวน 107 คน คิดเป็นร้อยละ 64.8485 โดยมีค่าความไว = 0.8649 ค่าความจำเพาะ = 0.4725 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.472)^2$ = 0.2228



รูปที่ 4.111 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

จากรูป 4.111 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 165 คน มีค่าทำนายความถูกต้องจำนวน 102 คน คิดเป็นร้อยละ 61.8182 โดยมีค่าความไว = 0.5843 ค่าความจำเพาะ = 0.6579 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4929)^2$ = 0.2430

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

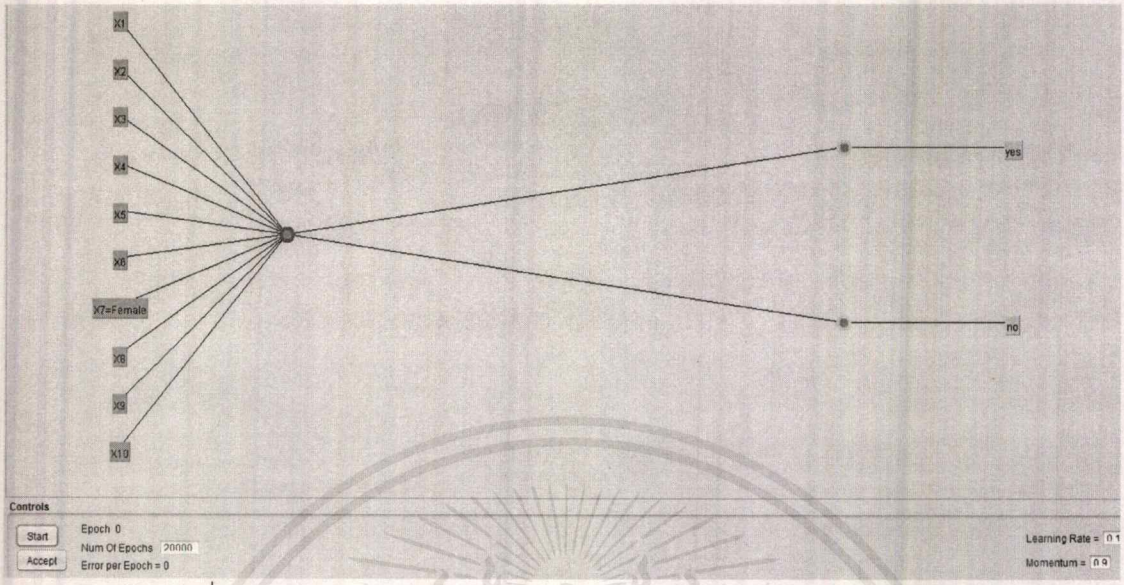


รูปที่ 4.112 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูป 4.112 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 165 คน มีค่าทำนายความถูกต้องจำนวน 119 คน คิดเป็นร้อยละ 72.1212 โดยมีค่าความไว = 0.5604 ค่าความจำเพาะ = 0.9189 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4398)^2$ = 0.1934

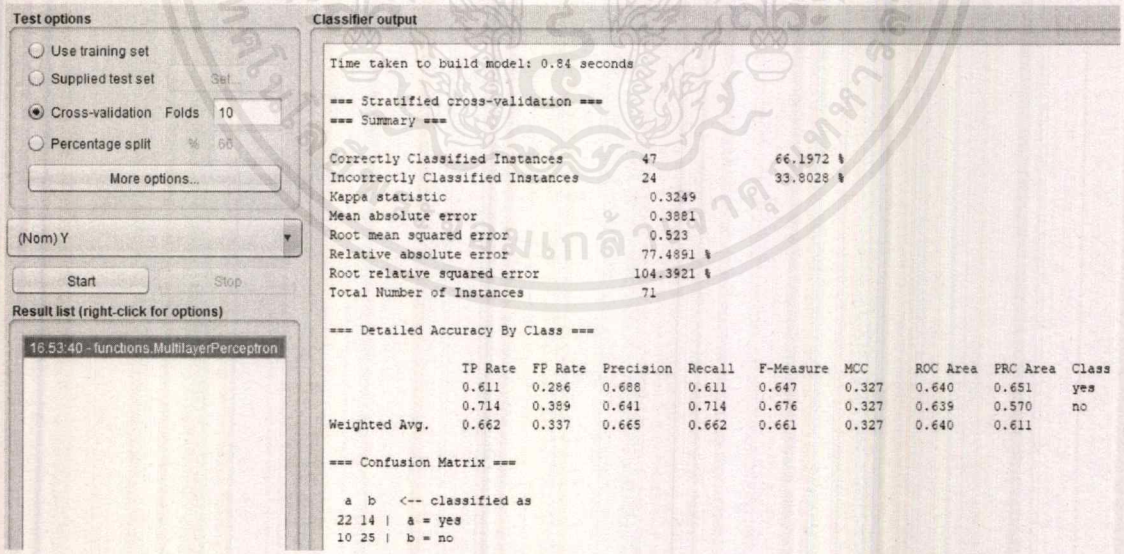
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.1.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทดสอบ 71 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.113 โครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10, 20 และ 30

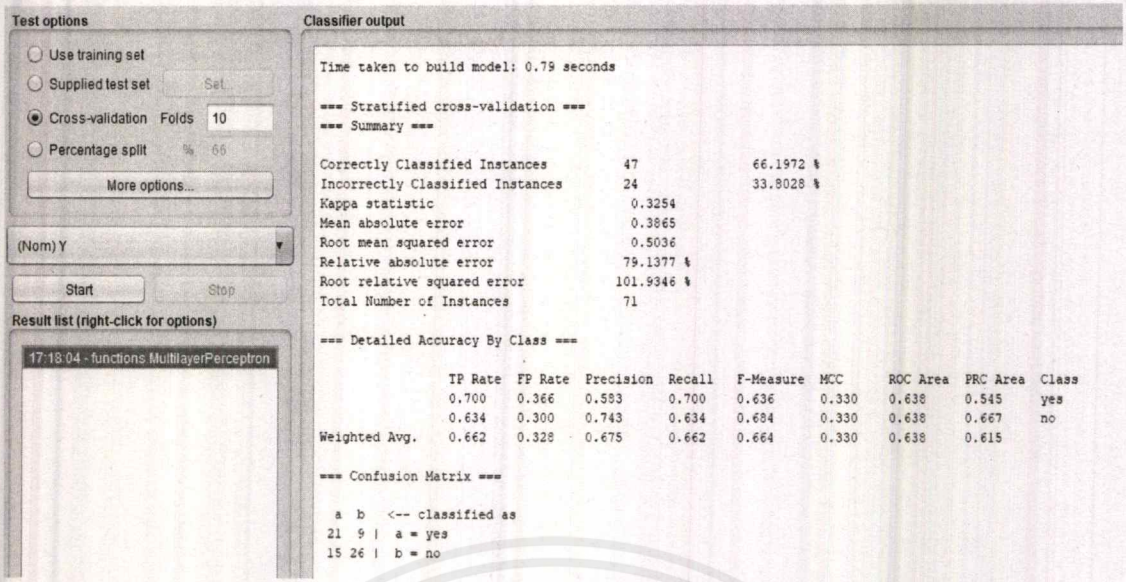
จากรูป 4.113 โครงข่ายประสาท ยอดคงเหลือในบัตรเครดิตของลูกค้า ชั้นข้อมูลเข้า อยู่ด้านซ้ายสุดซึ่งมีคุณลักษณะทั้งหมด 5 ตัวแปร (X_1-X_5) ชั้นซ่อน มี 1 ชั้นซ่อน ชั้นข้อมูลออกหรือชั้นข้อมูลผลลัพธ์ อยู่ด้านขวามือมีคำตอบคือ Yes กับ No Epoch จะรันการทำงานตั้งแต่รอบที่ 1 ถึง 20,000 Error per Epoch แสดงความคลาดเคลื่อนต่อรอบ คือ Learning Rate เท่ากับ 0.1 และ Momentum เท่ากับ 0.9



รูปที่ 4.114 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10

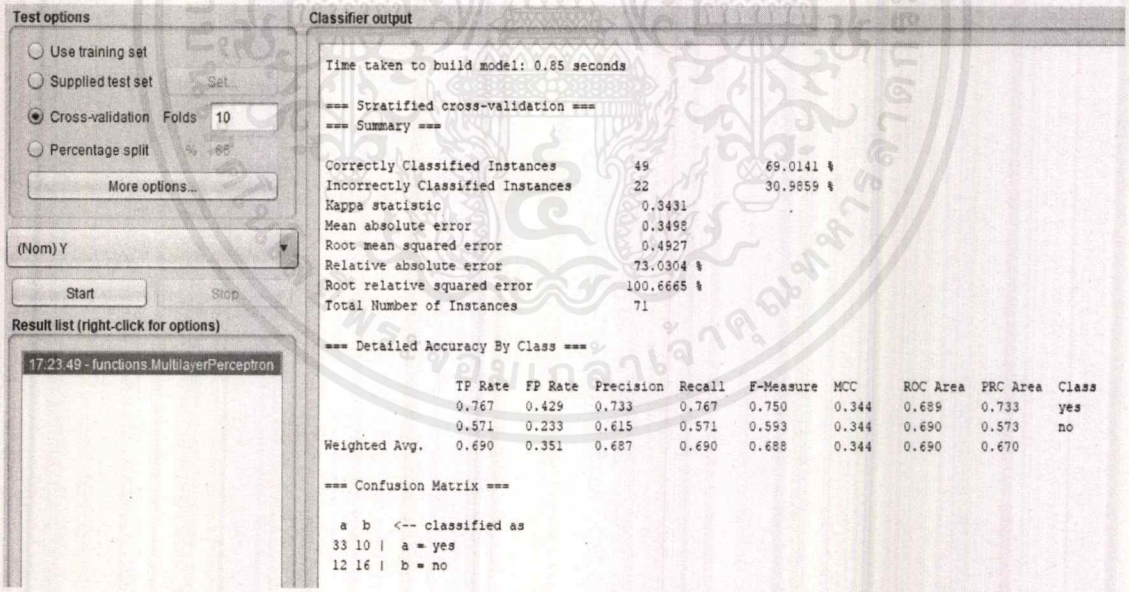
จากรูป 4.114 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 71 คน มีค่าทำนายความถูกต้องจำนวน 47 คน คิดเป็นร้อยละ 66.1972 โดยมีค่าความไว = 0.6111 ค่าความจำเพาะ = 0.7143 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.523)^2$ = 0.2735

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.115 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

จากรูป 4.115 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 71 คน มีค่าทำนายความถูกต้องจำนวน 47 คน คิดเป็นร้อยละ 66.1972 โดยมีค่าความไว = 0.7000 ค่าความจำเพาะ = 0.6341 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5036)^2$ = 0.2536

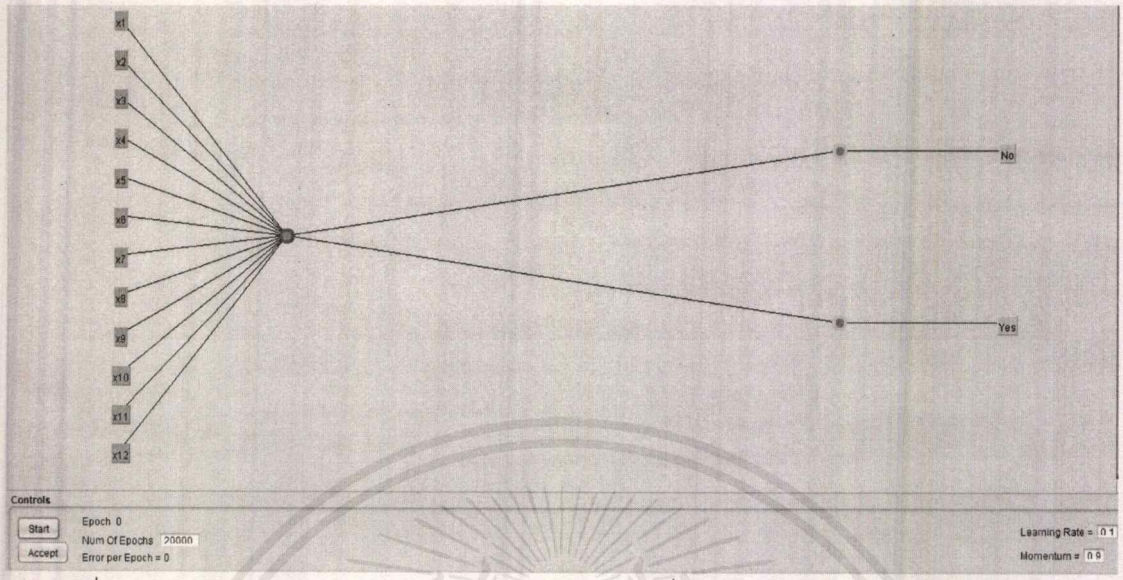


รูปที่ 4.116 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูป 4.116 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 71 คน มีค่าทำนายความถูกต้องจำนวน 49 คน คิดเป็นร้อยละ 69.0141 โดยมีค่าความไว = 0.7674 ค่าความจำเพาะ = 0.5714 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4927)^2$ = 0.2428

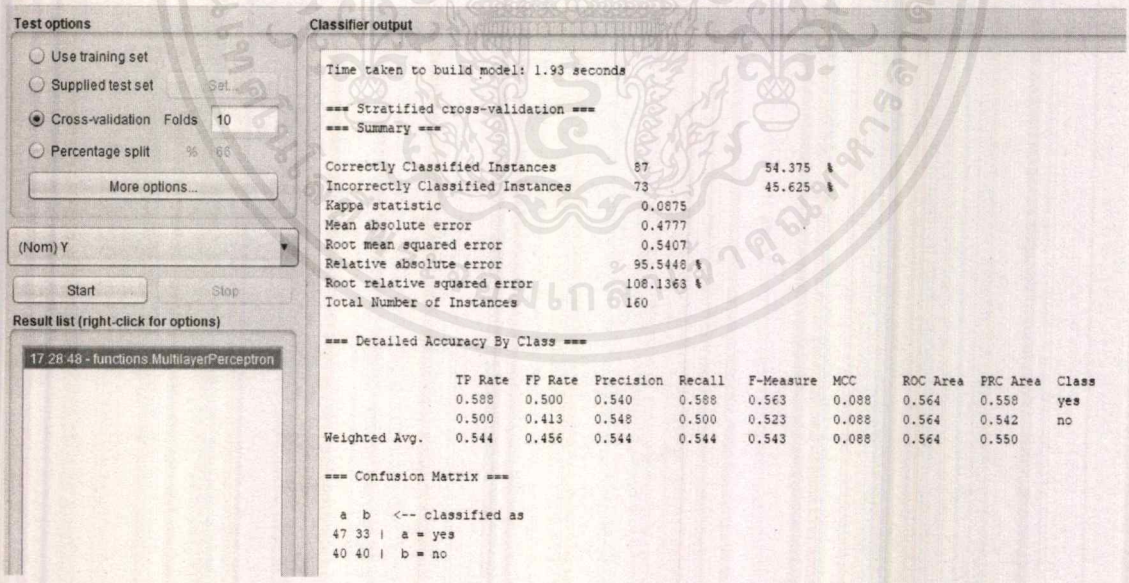
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.1.3 คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทดสอบ 160 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.117 โครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 10, 20 และ 30

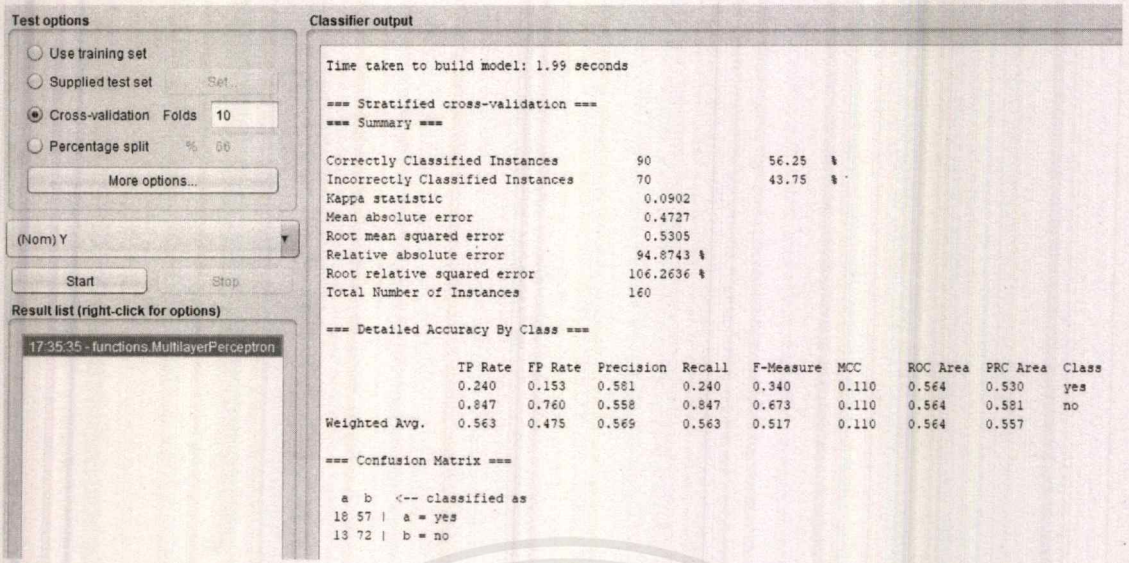
จากรูป 4.117 โครงข่ายประสาทเทียม คุณภาพไวน์แดง ชั้นข้อมูลเข้า อยู่ด้านซ้ายสุดซึ่งมีคุณลักษณะทั้งหมด 5 ตัวแปร (X_1-X_5) ชั้นซ่อน มี 1 ชั้นซ่อน ชั้นข้อมูลออกหรือชั้นข้อมูลผลลัพธ์ อยู่ด้านขวามือ มีคำตอบคือ Yes กับ No Epoch จะรันการทำงานตั้งแต่รอบที่ 1 ถึง 20,000 Error per Epoch แสดงความคลาดเคลื่อนต่อรอบ คือ Learning Rate เท่ากับ 0.1 และ Momentum เท่ากับ 0.9



รูปที่ 4.118 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 10

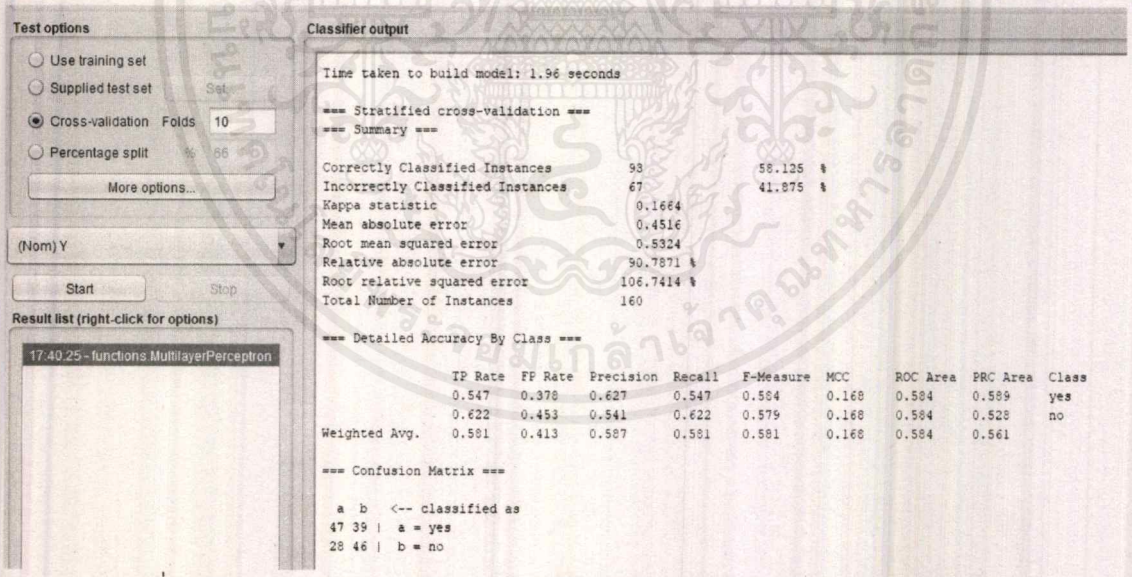
จากรูป 4.118 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 160 ขวด มีค่าทำนายความถูกต้องจำนวน 87 ขวด คิดเป็นร้อยละ 54.375 โดยมีค่าความไว = 0.5875 ค่าความจำเพาะ = 0.5000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = (0.5407)2 = 0.2924

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.119 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 20

จากรูป 4.119 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 160 ขวด มีค่าทำนายความถูกต้องจำนวน 90 ขวด คิดเป็นร้อยละ 56.25 โดยมีค่าความไว = 0.2400 ค่าความจำเพาะ = 0.8471 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5305)^2$ = 0.2814



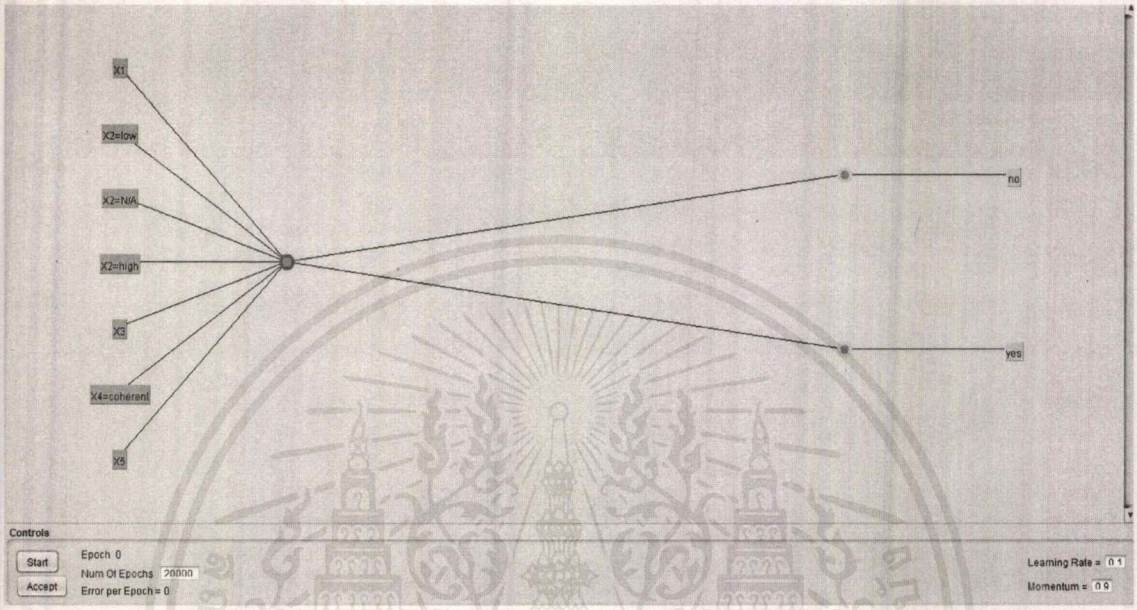
รูปที่ 4.120 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูป 4.120 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 160 ขวด มีค่าทำนายความถูกต้องจำนวน 93 ขวด คิดเป็นร้อยละ 58.125 โดยมีค่าความไว = 0.5465 ค่าความจำเพาะ = 0.6216 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5324)^2$ = 0.2834

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.2 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-Sampling)

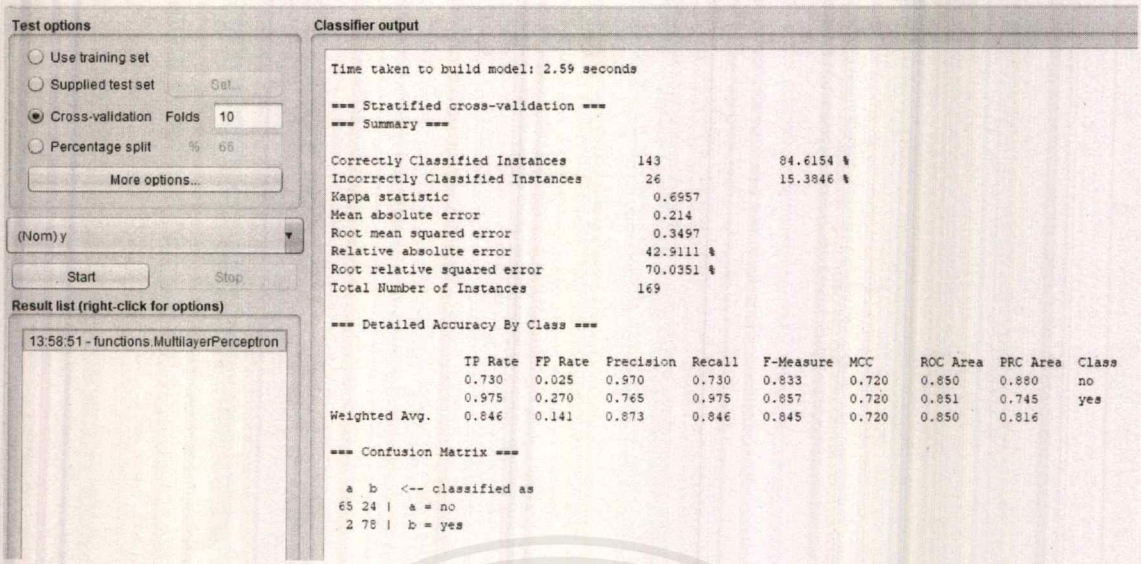
4.4.2.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทดสอบ 169 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.121 โครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10, 20 และ 30

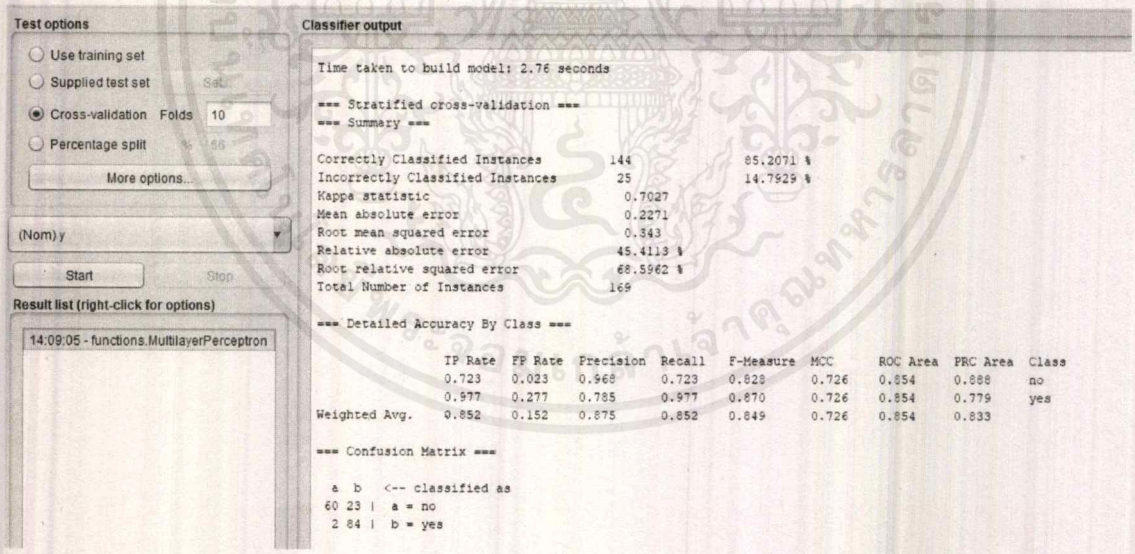
จากรูป 4.121 โครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก ชั้นข้อมูลเข้า อยู่ด้านซ้ายสุดซึ่งมีคุณลักษณะทั้งหมด 5 ตัวแปร (X1-X5) ชั้นซ่อน มี 1 ชั้นซ่อน ชั้นข้อมูลออกหรือชั้นข้อมูลผลลัพธ์ อยู่ด้านขวามือ มีคำตอบคือ Yes กับ No Epoch จะรันการทำงานตั้งแต่รอบที่ 1 ถึง 20,000 Error per Epoch แสดงความคลาดเคลื่อนต่อรอบ คือ Learning Rate เท่ากับ 0.1 และ Momentum เท่ากับ 0.9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.122 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10

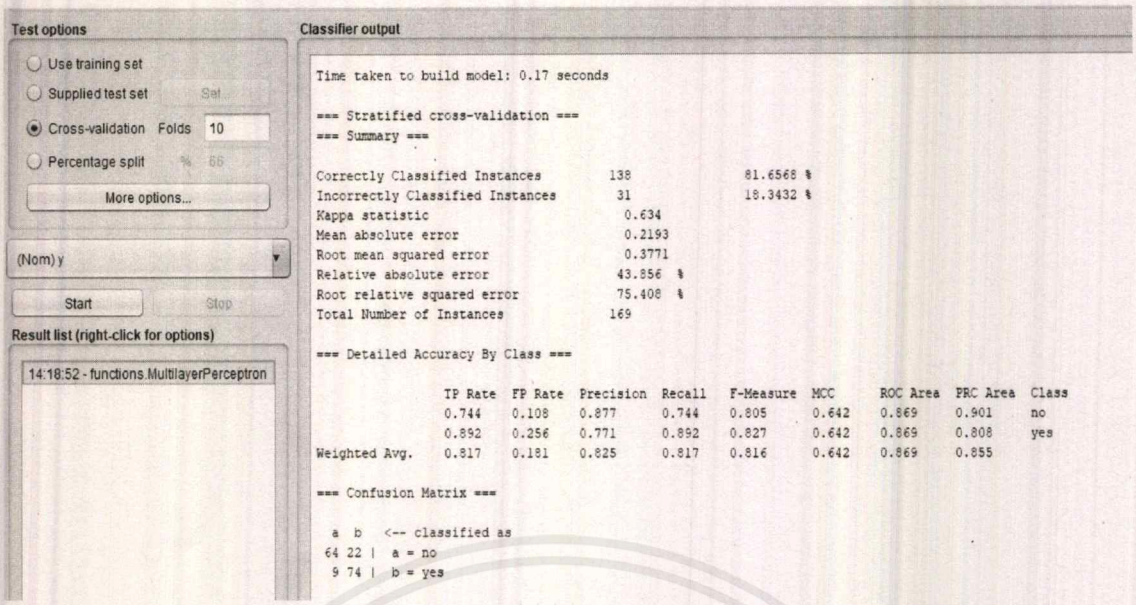
จากรูป 4.122 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 169 คน มีค่าทำนายความถูกต้องจำนวน 143 คน คิดเป็นร้อยละ 84.6154 โดยมีค่าความไว = 0.7303 ค่าความจำเพาะ = 0.9750 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.3497)^2$ = 0.1223



รูปที่ 4.123 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

จากรูป 4.123 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 169 คน มีค่าทำนายความถูกต้องจำนวน 144 คน คิดเป็นร้อยละ 85.2071 โดยมีค่าความไว = 0.7229 ค่าความจำเพาะ = 0.9767 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.343)^2$ = 0.1176

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

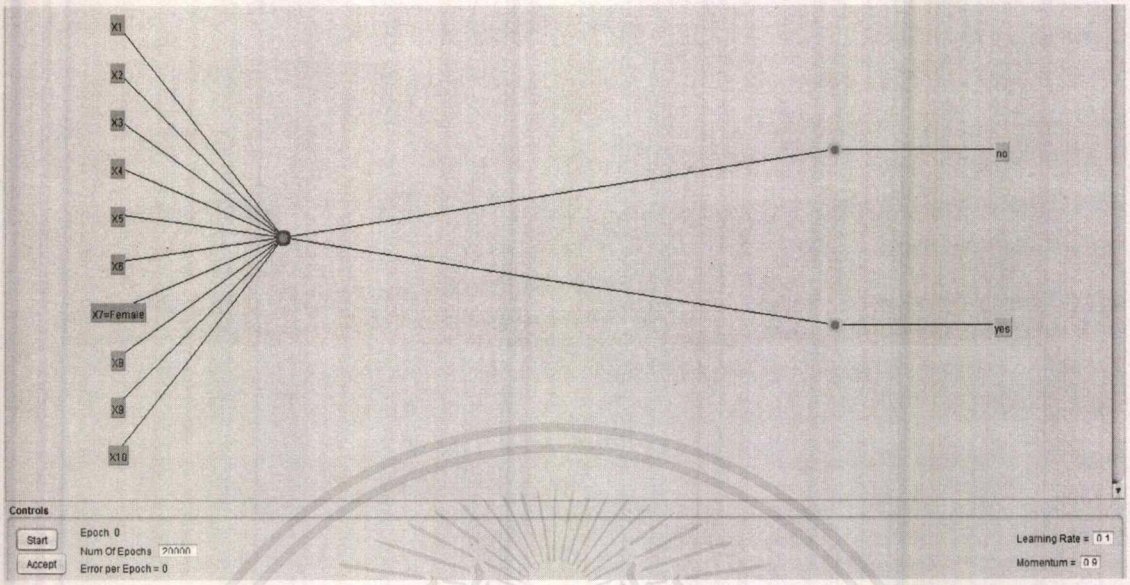


รูปที่ 4.124 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูป 4.124 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 169 คน มีค่าทำนายความถูกต้องจำนวน 138 คน คิดเป็นร้อยละ 81.6568 โดยมีค่าความไว = 0.7442 ค่าความจำเพาะ = 0.8916 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.3771)^2$ = 0.1422

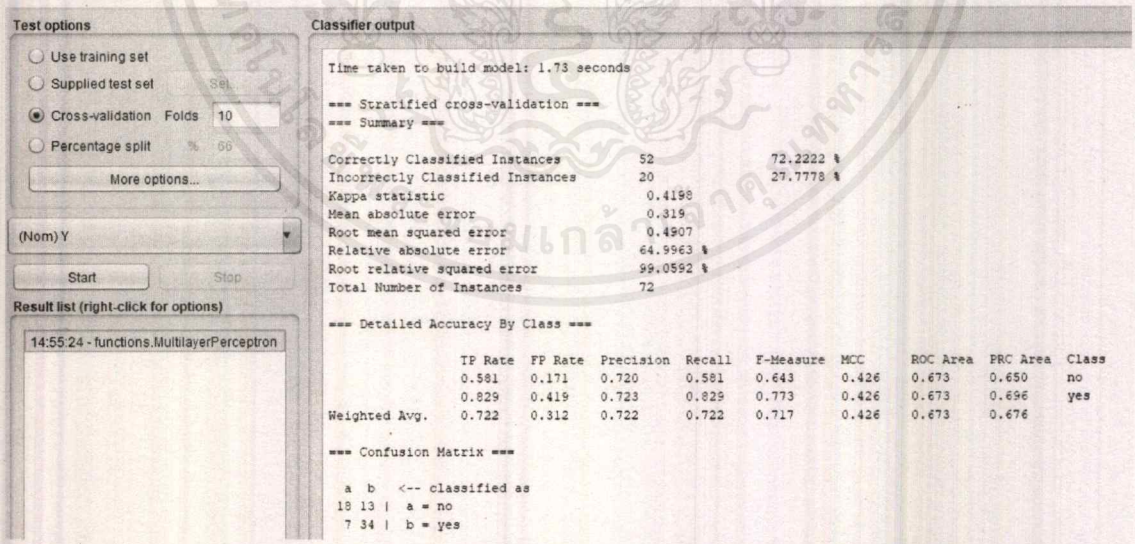
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.2.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทดสอบ 72 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.125 โครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10, 20 และ30

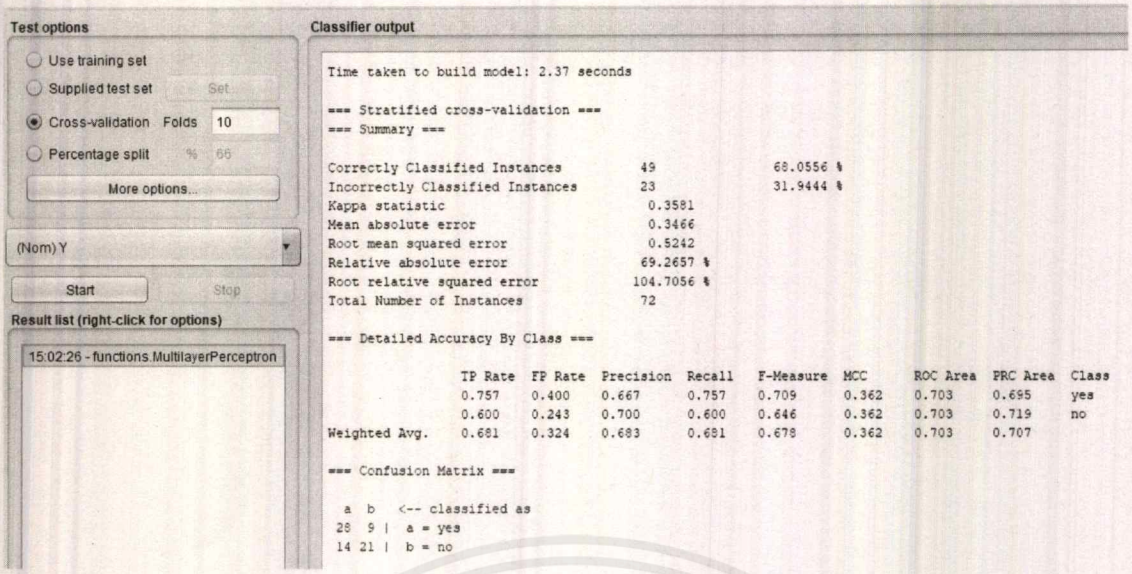
จากรูป 4.125 โครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า ชั้นข้อมูลเข้าอยู่ด้านซ้ายสุดซึ่งมีคุณลักษณะทั้งหมด 5 ตัวแปร (X_1-X_5) ชั้นซ่อน มี 1 ชั้นซ่อน ชั้นข้อมูลออกหรือชั้นข้อมูลผลลัพธ์ อยู่ด้านขวามือ มีคำตอบคือ Yes กับ No Epoch จะรันการทำงานตั้งแต่รอบที่ 1 ถึง 20,000 Error per Epoch แสดงความคลาดเคลื่อนต่อรอบ คือ Learning Rate เท่ากับ 0.1 และ Momentum เท่ากับ 0.9



รูปที่ 4.126 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10

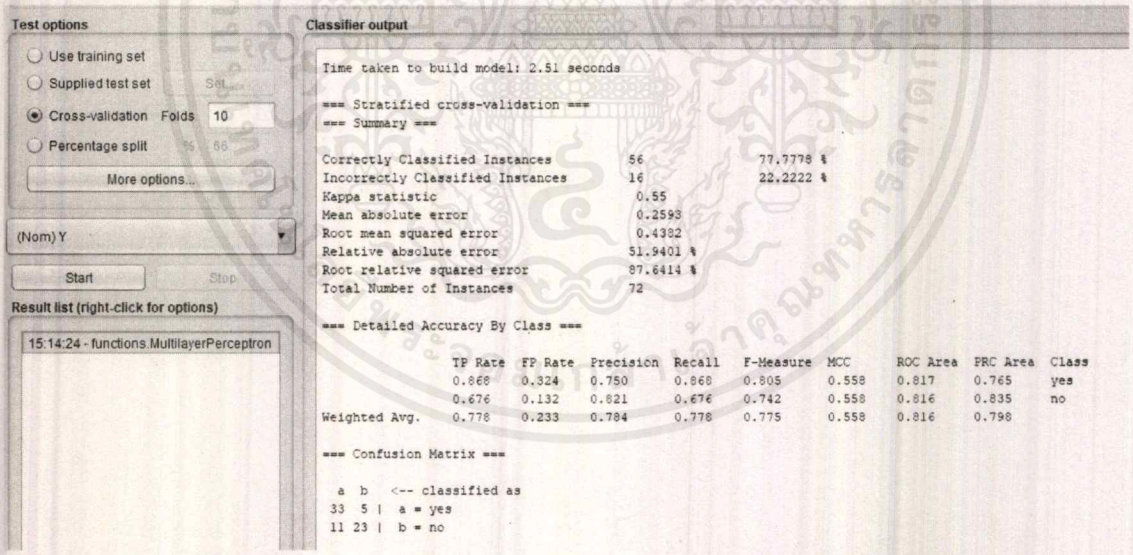
จากรูป 4.126 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 72 คน มีค่าทำนายความถูกต้องจำนวน 52 คน คิดเป็นร้อยละ 72.2222 โดยมีค่าความไว = 0.5806 ค่าความจำเพาะ = 0.8293 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4907)^2$ = 0.2408

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.127 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

จากรูป 4.127 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 72 คน มีค่าทำนายความถูกต้องจำนวน 49 คน คิดเป็นร้อยละ 68.0556 โดยมีค่าความไว = 0.7568 ค่าความจำเพาะ = 0.6000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5242)^2$ = 0.2748

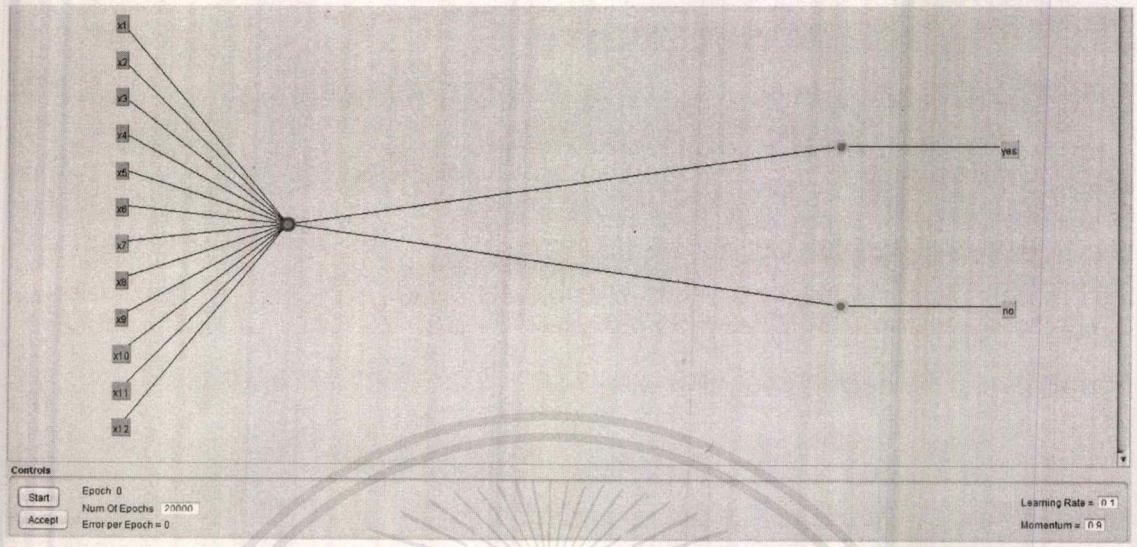


รูปที่ 4.128 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูป 4.128 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 72 คน มีค่าทำนายความถูกต้องจำนวน 56 คน คิดเป็นร้อยละ 77.7778 โดยมีค่าความไว = 0.8684 ค่าความจำเพาะ = 0.6765 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4382)^2$ = 0.1920

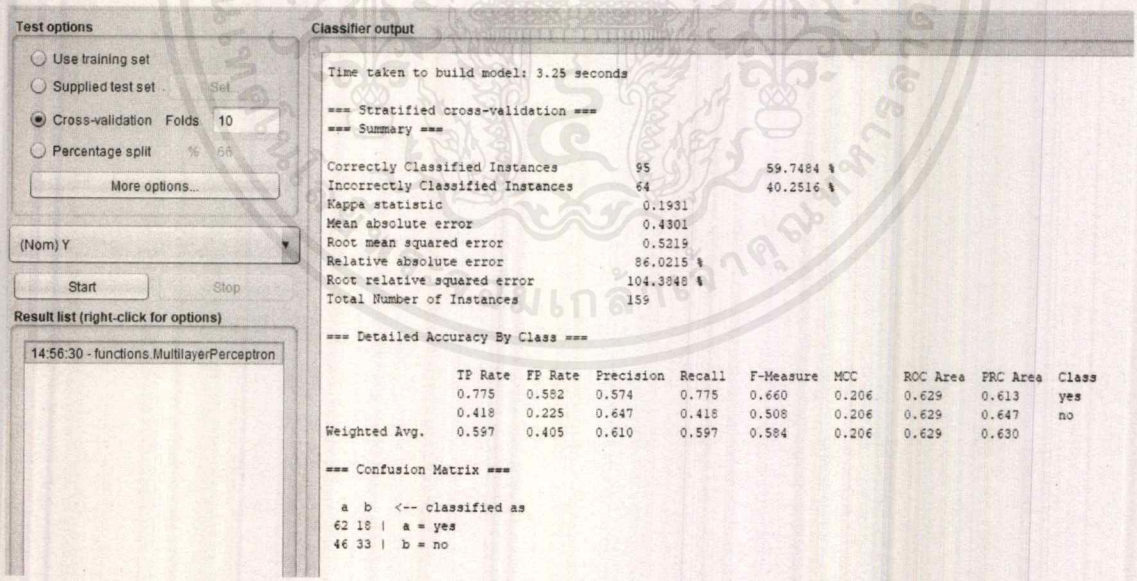
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.2.3 คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทดสอบ 159 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.129 โครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 10, 20 และ 30

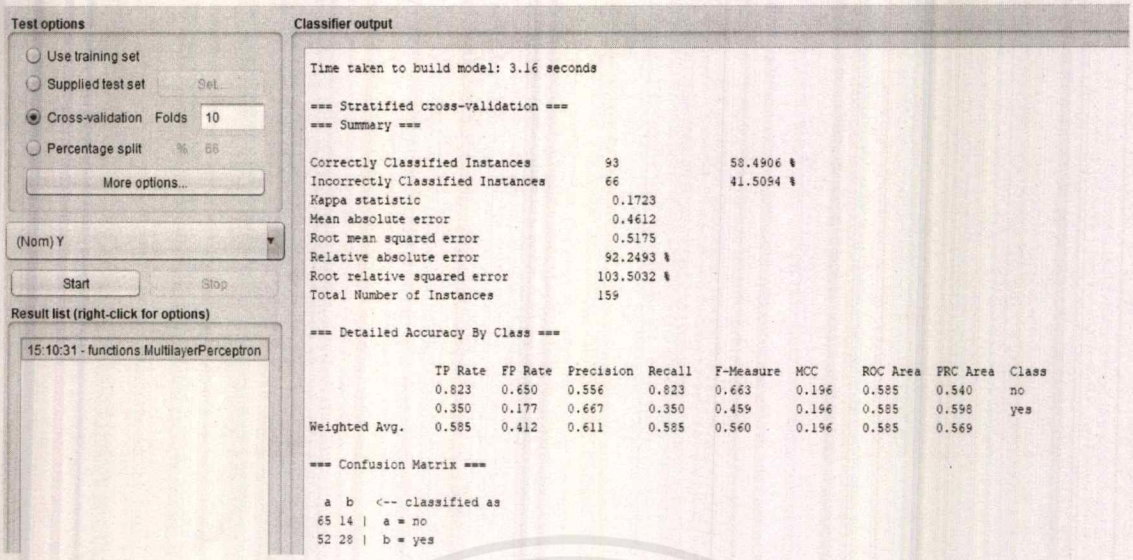
จากรูป 4.129 โครงข่ายประสาท คุณภาพไวน์แดง ชั้นข้อมูลเข้า อยู่ด้านซ้ายสุดซึ่งมีคุณลักษณะทั้งหมด 5 ตัวแปร (X_1-X_5) ชั้นซ่อน มี 1 ชั้นซ่อน ชั้นข้อมูลออกหรือชั้นข้อมูลผลลัพธ์ อยู่ด้านขวามือมีคำตอบคือ Yes กับ No Epoch จะรันการทำงานตั้งแต่รอบที่ 1 ถึง 20,000 Error per Epoch แสดงความคลาดเคลื่อนต่อรอบ คือ Learning Rate เท่ากับ 0.1 และ Momentum เท่ากับ 0.9



รูปที่ 4.130 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 10

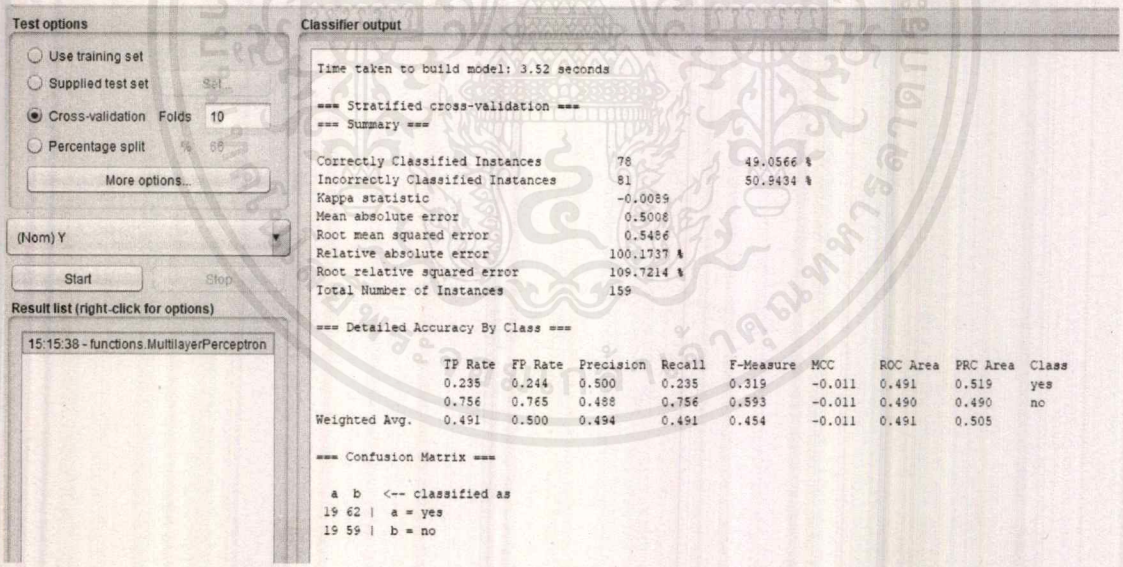
จากรูป 4.130 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 159 ขวด มีค่าทำนายความถูกต้องจำนวน 95 ขวด คิดเป็นร้อยละ 59.7484 โดยมีค่าความไว = 0.7750 ค่าความจำเพาะ = 0.4177 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5219)^2$ = 0.2724

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.131 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 20

จากรูป 4.131 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 159 ขวด มีค่าทำนายความถูกต้องจำนวน 93 ขวด คิดเป็นร้อยละ 58.4906 โดยมีค่าความไว = 0.8228 ค่าความจำเพาะ = 0.3500 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5175)^2$ = 0.2678



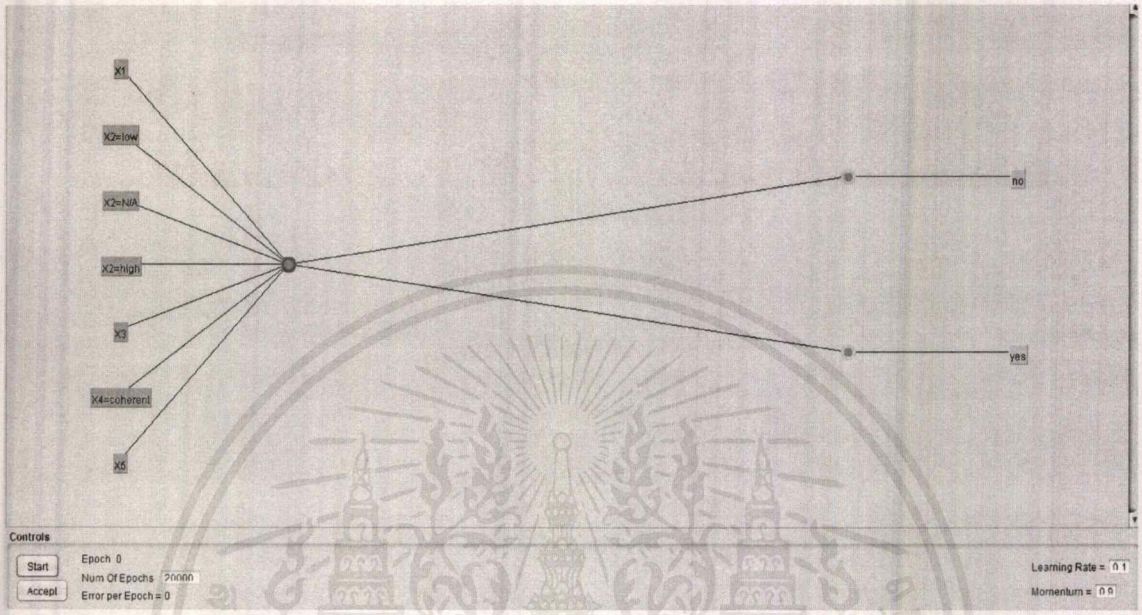
รูปที่ 4.132 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูป 4.132 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 159 ขวด มีค่าทำนายความถูกต้องจำนวน 78 ขวด คิดเป็นร้อยละ 49.0566 โดยมีค่าความไว = 0.2346 ค่าความจำเพาะ = 0.7564 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5486)^2$ = 0.3010

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.3 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาท โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด (Under Sampling)

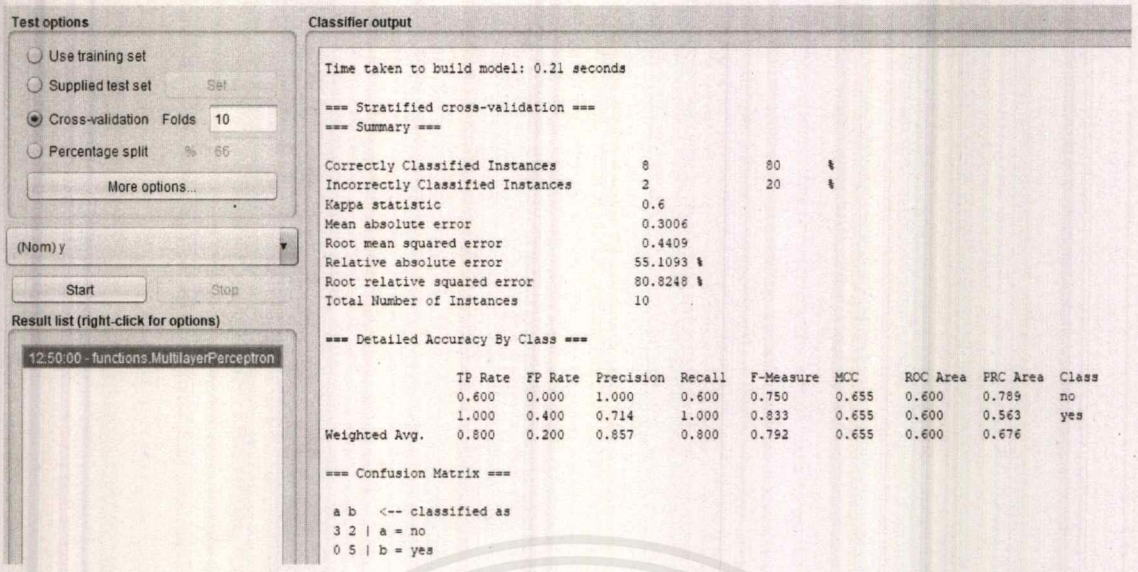
4.4.3.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทดสอบ 10 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.133 โครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10, 20 และ 30

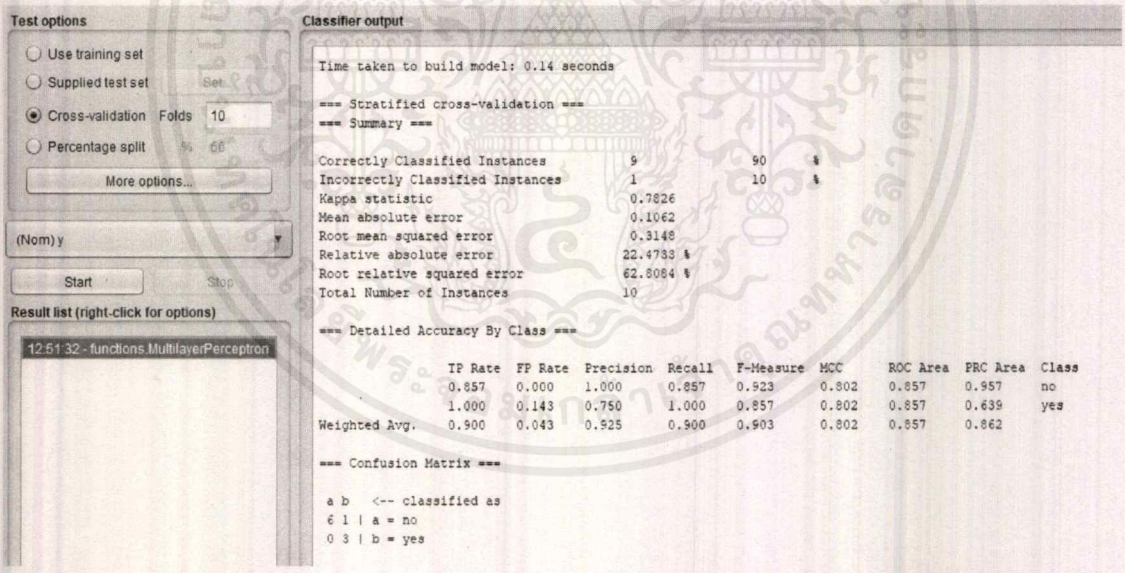
จากรูป 4.133 โครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก ชั้นข้อมูลเข้า อยู่ด้านซ้ายสุดซึ่งมีคุณลักษณะทั้งหมด 5 ตัวแปร (X1-X5) ชั้นซ่อน มี 1 ชั้นซ่อน ชั้นข้อมูลออกหรือชั้นข้อมูลผลลัพธ์ อยู่ด้านขวามือ มีคำตอบคือ Yes กับ No Epoch จะรันการทำงานตั้งแต่รอบที่ 1 ถึง 20,000 Error per Epoch แสดงความคลาดเคลื่อนต่อรอบ คือ Learning Rate เท่ากับ 0.1 และ Momentum เท่ากับ 0.9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.134 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10

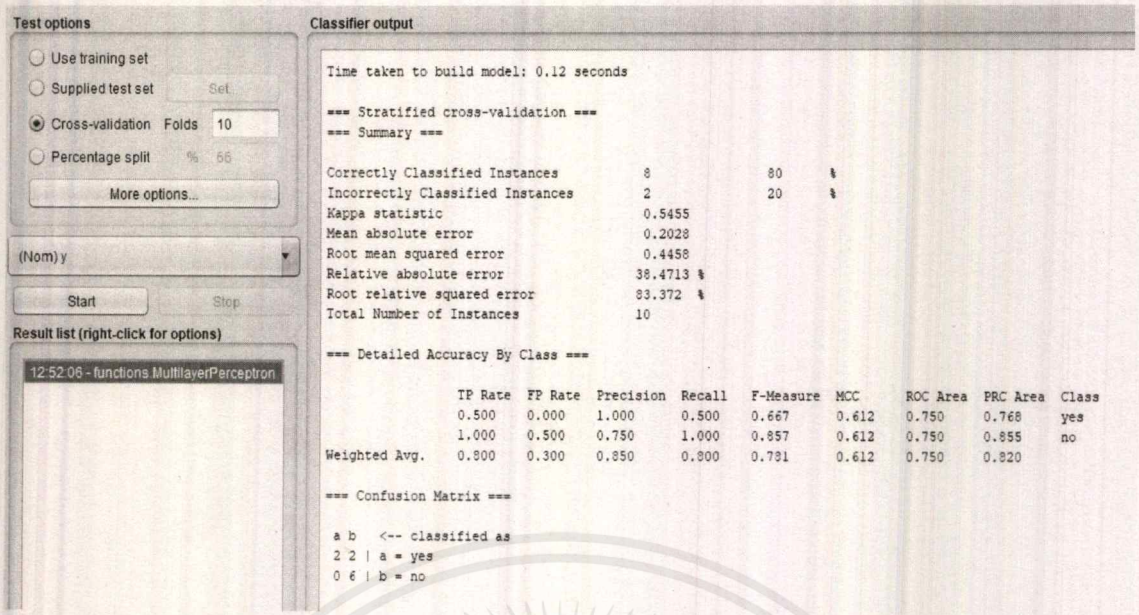
จากรูป 4.134 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 8 คน คิดเป็นร้อยละ 80 โดยมีค่าความไว = 0.6000 ค่าความจำเพาะ = 1 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4409)^2$ = 0.1944



รูปที่ 4.135 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

จากรูป 4.135 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 9 คน คิดเป็นร้อยละ 90 โดยมีค่าความไว = 0.8571 ค่าความจำเพาะ = 1 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.3148)^2$ = 0.0991

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

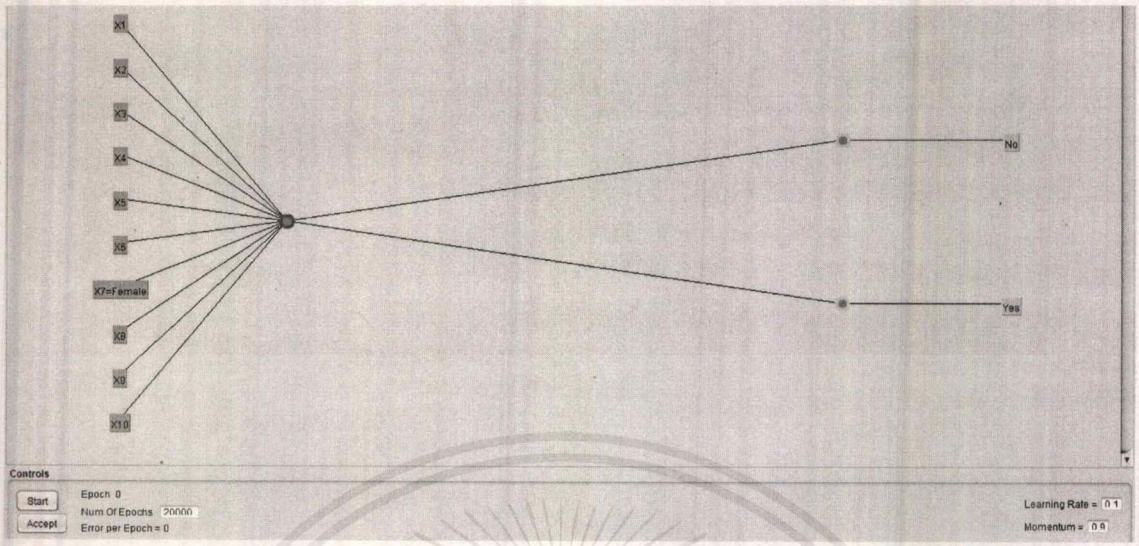


รูปที่ 4.136 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูป 4.136 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 8 คน คิดเป็นร้อยละ 80 โดยมีค่าความไว = 0.5000 ค่าความจำเพาะ = 1 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4458)^2$ = 0.1987

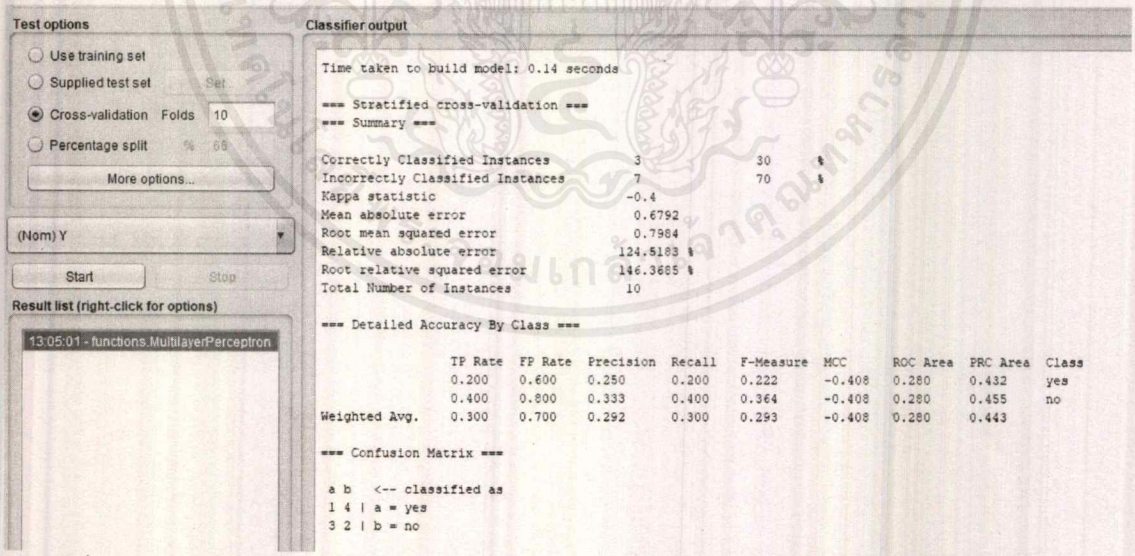
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.3.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทดสอบ 10 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.137 โครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10, 20 และ30

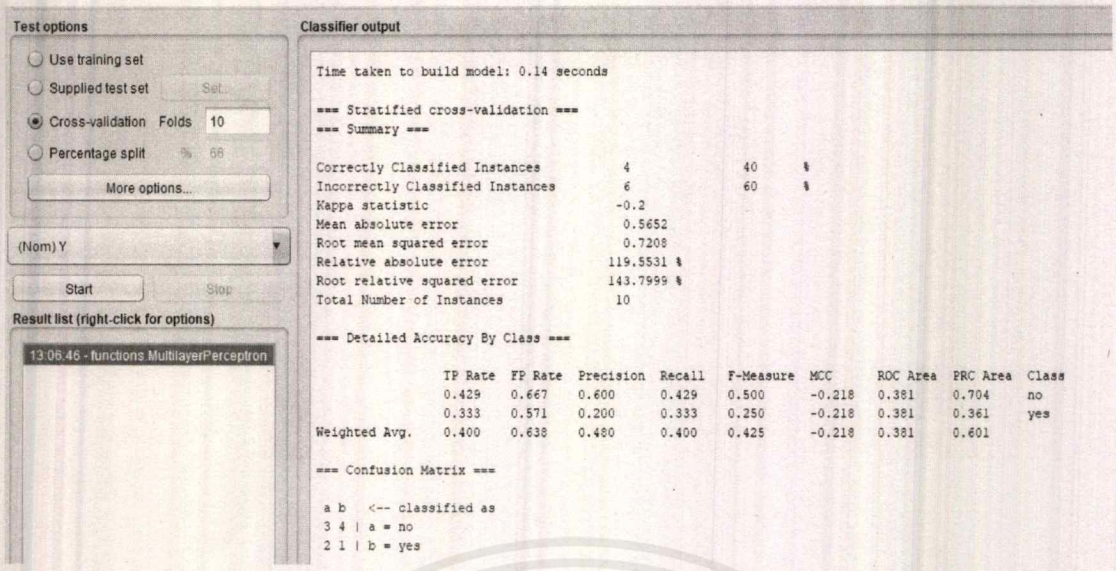
จากรูป 4.137 โครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า ชั้นข้อมูลเข้าอยู่ด้านซ้ายสุดซึ่งมีคุณลักษณะทั้งหมด 5 ตัวแปร (X_1-X_5) ชั้นซ่อน มี 1 ชั้นซ่อน ชั้นข้อมูลออกหรือชั้นข้อมูลผลลัพธ์ อยู่ด้านขวามือ มีคำตอบคือ Yes กับ No Epoch จะรันการทำงานตั้งแต่รอบที่ 1 ถึง 20,000 Error per Epoch แสดงความคลาดเคลื่อนต่อรอบ คือ Learning Rate เท่ากับ 0.1 และ Momentum เท่ากับ 0.9



รูปที่ 4.138 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10

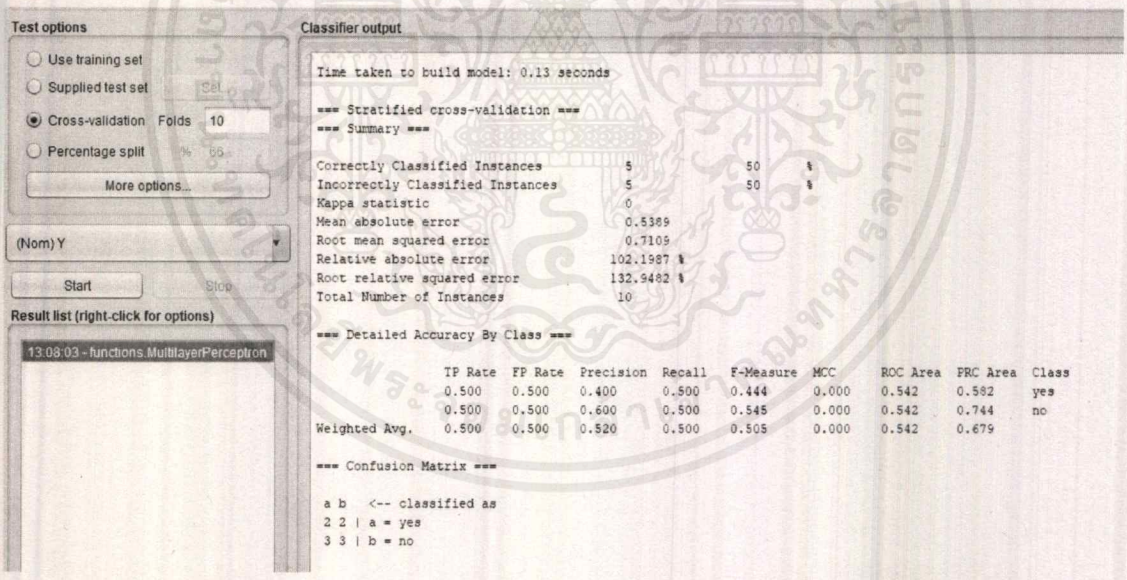
จากรูป 4.138 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 3 คน คิดเป็นร้อยละ 30 โดยมีค่าความไว = 0.2000 ค่าความจำเพาะ = 0.4000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.7984)^2$ = 0.6374

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.139 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

จากรูป 4.139 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 4 คน คิดเป็นร้อยละ 40 โดยมีค่าความไว = 0.4286 ค่าความจำเพาะ = 0.3333 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.7208)^2$ = 0.5196

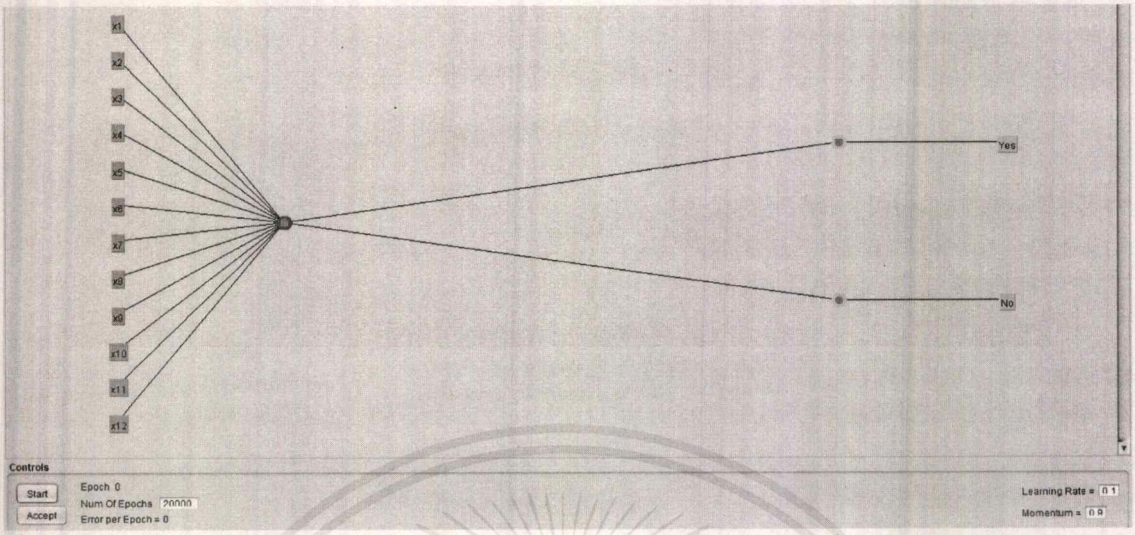


รูปที่ 4.140 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูป 4.140 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 5 คน คิดเป็นร้อยละ 50 โดยมีค่าความไว = 0.5000 ค่าความจำเพาะ = 0.5000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.7109)^2$ = 0.5054

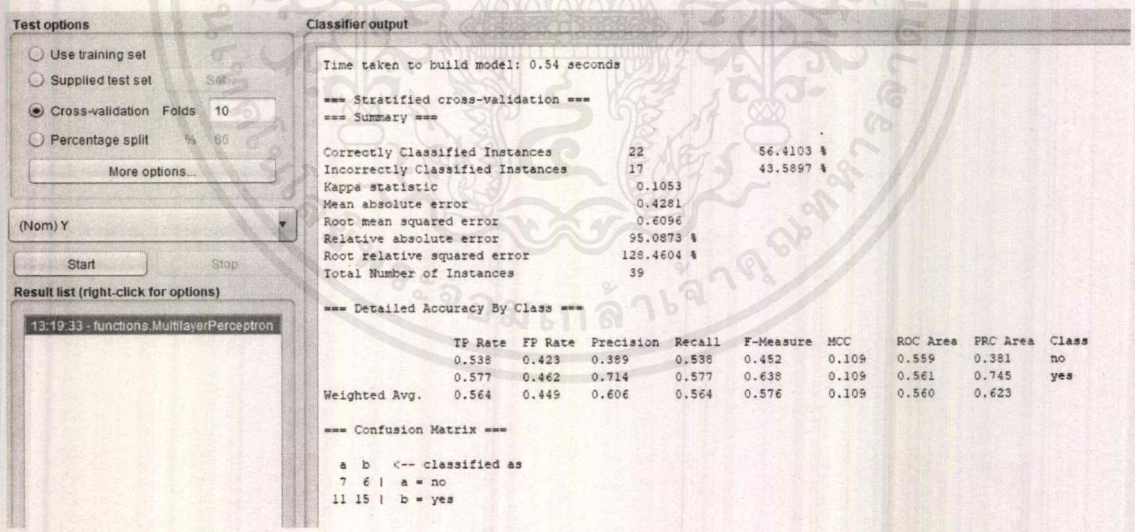
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.3.3 คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทดสอบ 39 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.141 โครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 10, 20 และ 30

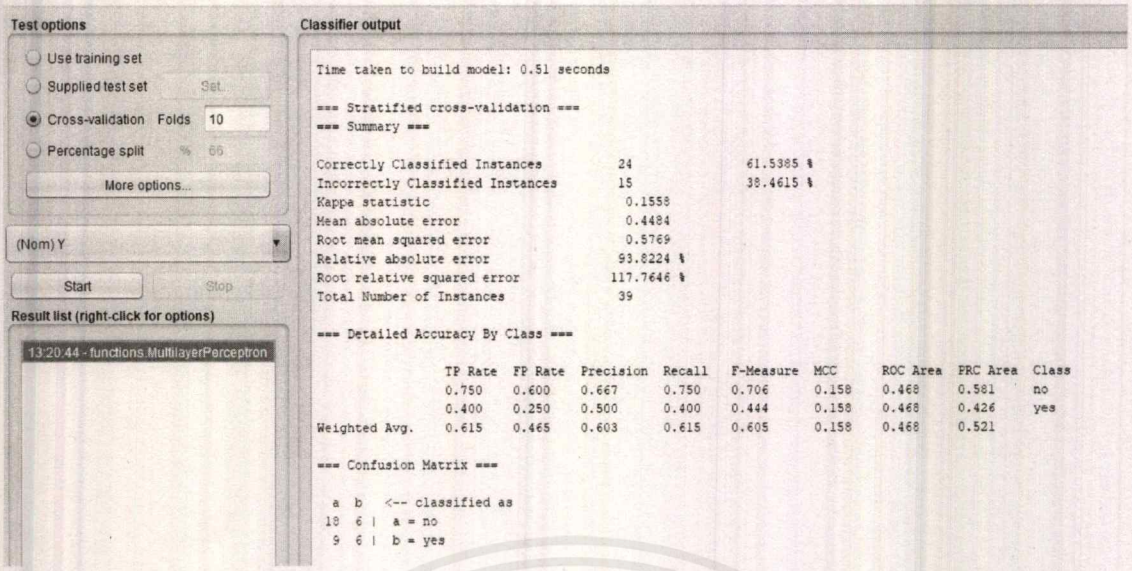
จากรูป 4.141 โครงข่ายประสาทเทียม คุณภาพไวน์แดง ชั้นข้อมูลเข้า อยู่ด้านซ้ายสุดซึ่งมีคุณลักษณะทั้งหมด 5 ตัวแปร (X_1-X_5) ชั้นซ่อน มี 1 ชั้นซ่อน ชั้นข้อมูลออกหรือชั้นข้อมูลผลลัพธ์ อยู่ด้านขวามือ มีคำตอบคือ Yes กับ No Epoch จะรันการทำงานตั้งแต่รอบที่ 1 ถึง 20,000 Error per Epoch แสดงความคลาดเคลื่อนต่อรอบ คือ Learning Rate เท่ากับ 0.1 และ Momentum เท่ากับ 0.9



รูปที่ 4.142 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 10

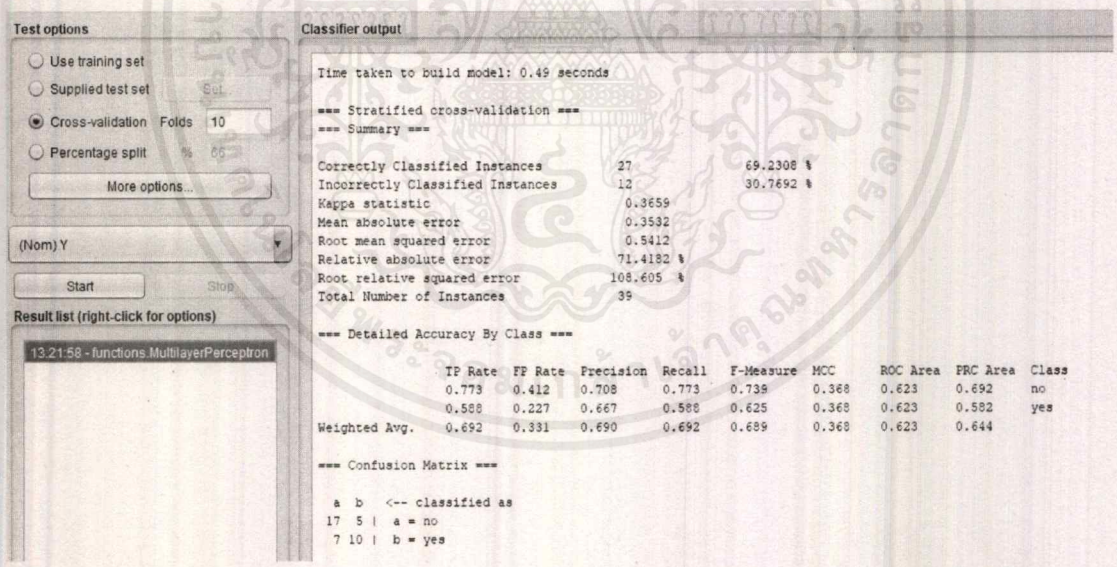
จากรูป 4.142 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 39 ขวด มีค่าทำนายความถูกต้องจำนวน 22 ขวด คิดเป็นร้อยละ 56.4103 โดยมีค่าความไว = 0.5385 ค่าความจำเพาะ = 0.5769 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6096)^2$ = 0.3716

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.143 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 20

จากรูป 4.143 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 39 ขวด มีค่าทำนายความถูกต้องจำนวน 24 ขวด คิดเป็นร้อยละ 61.5385 โดยมีค่าความไว = 0.7500 ค่าความจำเพาะ = 0.4000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5769)^2$ = 0.3328



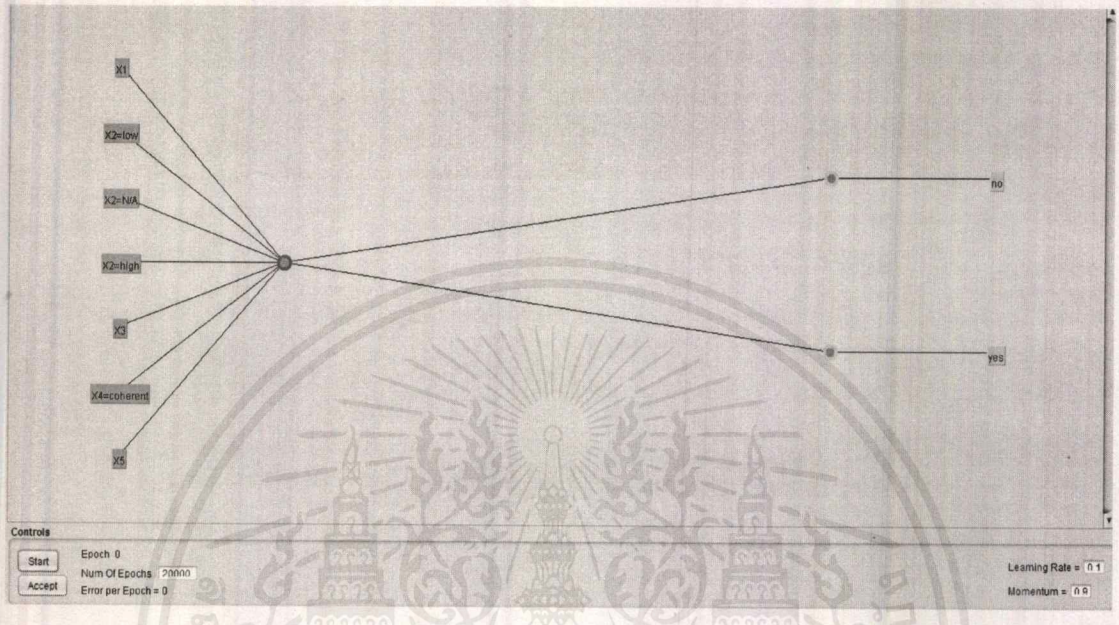
รูปที่ 4.144 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูป 4.144 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 39 ขวด มีค่าทำนายความถูกต้องจำนวน 27 ขวด คิดเป็นร้อยละ 69.2308 โดยมีค่าความไว = 0.7727 ค่าความจำเพาะ = 0.5882 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5412)^2$ = 0.2929

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.4 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม โดยการปรับความไม่สมดุลด้วยวิธีการผสมผสาน (Hybrid Methods)

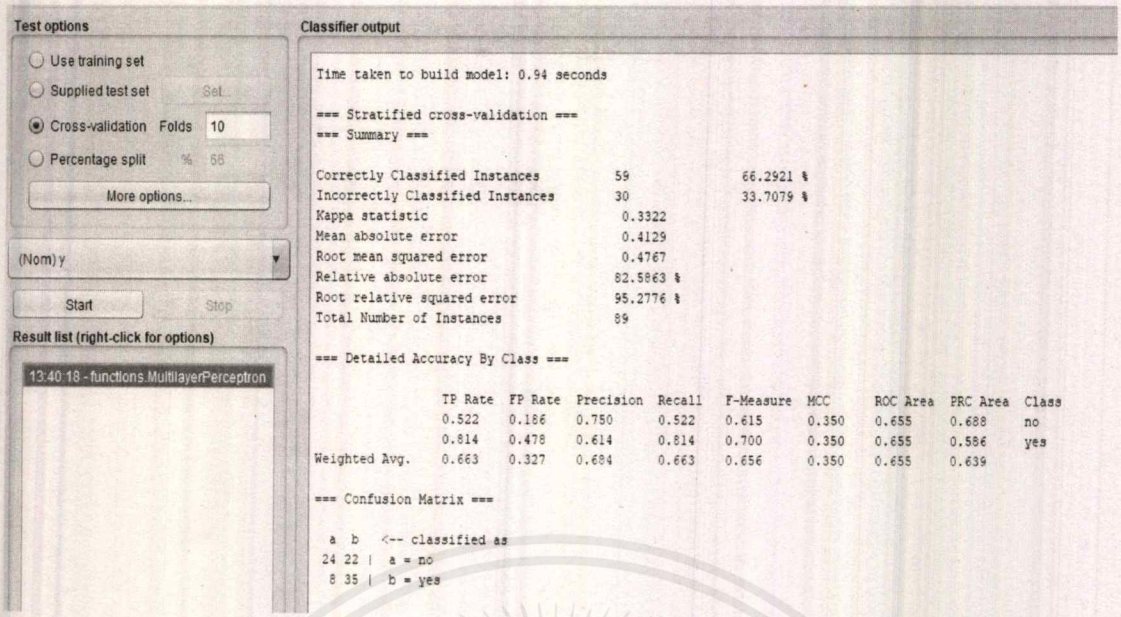
4.4.4.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทดสอบ 89 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.145 โครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10, 20 และ 30

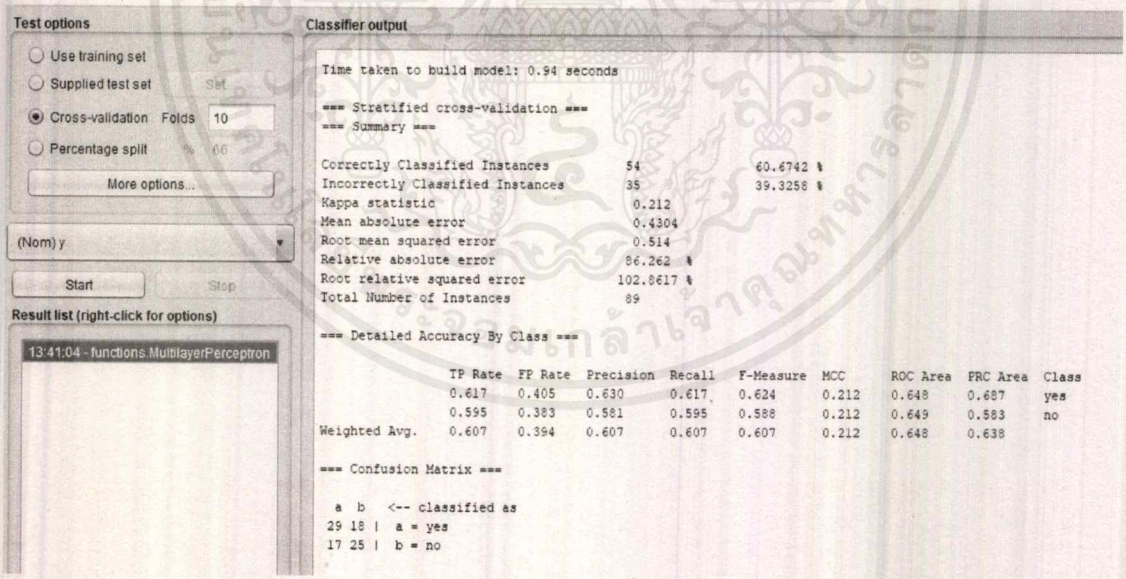
จากรูป 4.145 โครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก ชั้นข้อมูลเข้า อยู่ด้านซ้ายสุดซึ่งมีคุณลักษณะทั้งหมด 5 ตัวแปร (X1-X5) ชั้นซ่อน มี 1 ชั้นซ่อน ชั้นข้อมูลออกหรือชั้นข้อมูลผลลัพธ์ อยู่ด้านขวามือ มีคำตอบคือ Yes กับ No Epoch จะรันการทำงานตั้งแต่รอบที่ 1 ถึง 20,000 Error per Epoch แสดงความคลาดเคลื่อนต่อรอบ คือ Learning Rate เท่ากับ 0.1 และ Momentum เท่ากับ 0.9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.146 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10

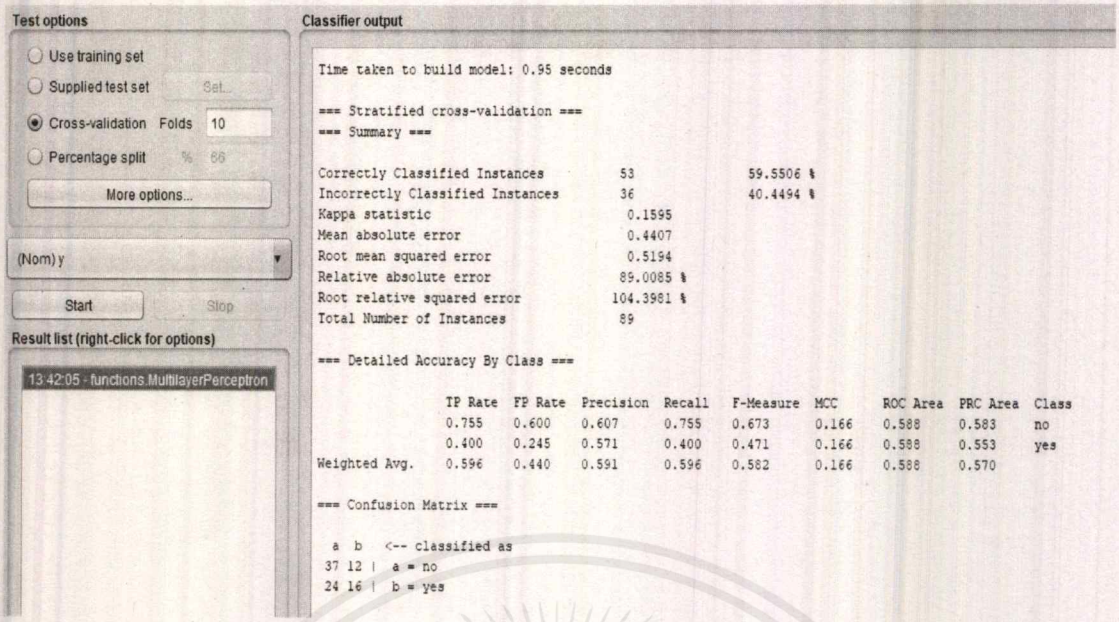
จากรูป 4.146 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 89 คน มีค่าทำนายความถูกต้องจำนวน 59 คน คิดเป็นร้อยละ 66.2921 โดยมีค่าความไว = 0.5217 ค่าความจำเพาะ = 0.8140 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4767)^2$ = 0.2272



รูปที่ 4.147 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

จากรูป 4.147 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 89 คน มีค่าทำนายความถูกต้องจำนวน 54 คน คิดเป็นร้อยละ 60.6742 โดยมีค่าความไว = 0.6170 ค่าความจำเพาะ = 0.5952 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5140)^2$ = 0.2642

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

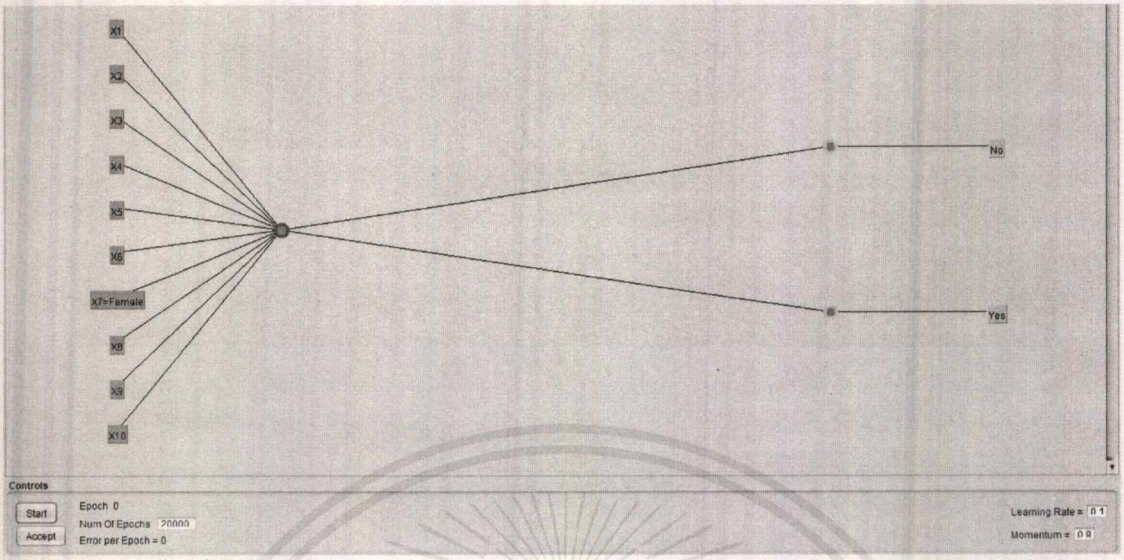


รูปที่ 4.148 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูป 4.148 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 89 คน มีค่าทำนายความถูกต้องจำนวน 53 คน คิดเป็นร้อยละ 59.5506 โดยมีค่าความไว = 0.7551 ค่าความจำเพาะ = 0.4000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5194)^2$ = 0.2698

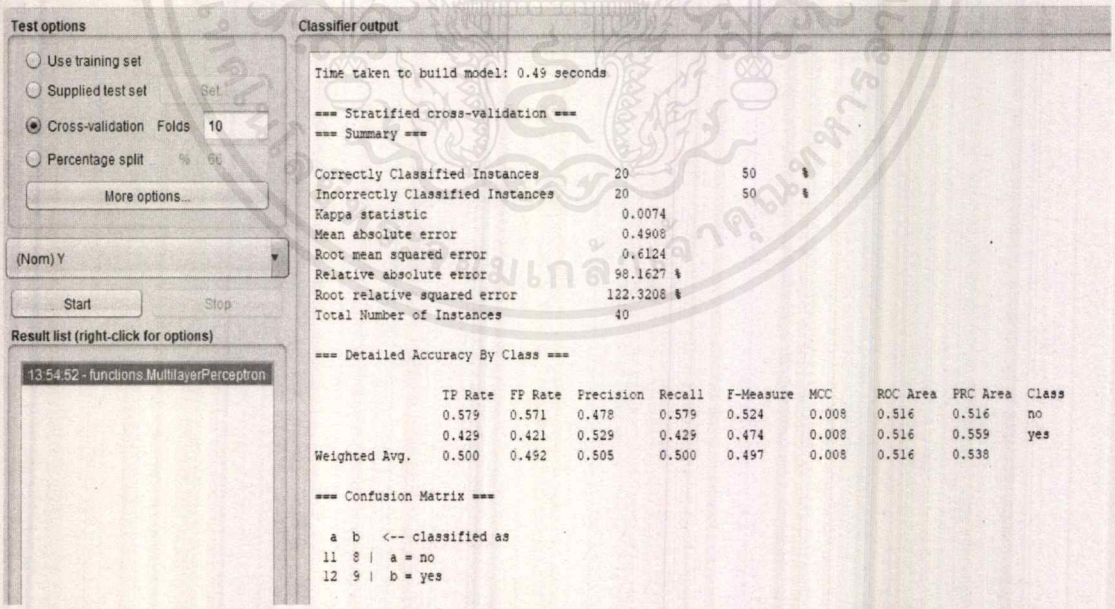
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.4.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทดสอบ 40 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.149 โครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10, 20 และ30

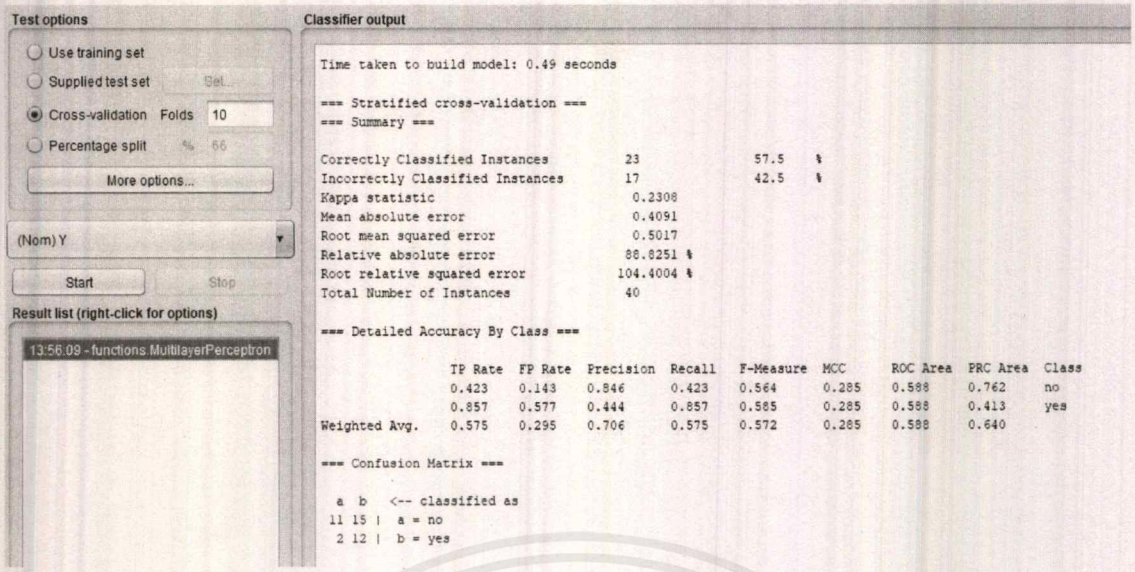
จากรูป 4.149 โครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า ชั้นข้อมูลเข้าอยู่ด้านซ้ายสุดซึ่งมีคุณลักษณะทั้งหมด 5 ตัวแปร (X_1-X_5) ชั้นซ่อน มี 1 ชั้นซ่อน ชั้นข้อมูลออกหรือชั้นข้อมูลผลลัพธ์ อยู่ด้านขวามือ มีคำตอบคือ Yes กับ No Epoch จะรันการทำงานตั้งแต่รอบที่ 1 ถึง 20,000 Error per Epoch แสดงความคลาดเคลื่อนต่อรอบ คือ Learning Rate เท่ากับ 0.1 และ Momentum เท่ากับ 0.9



รูปที่ 4.150 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้าเมื่อ Random Seed 10

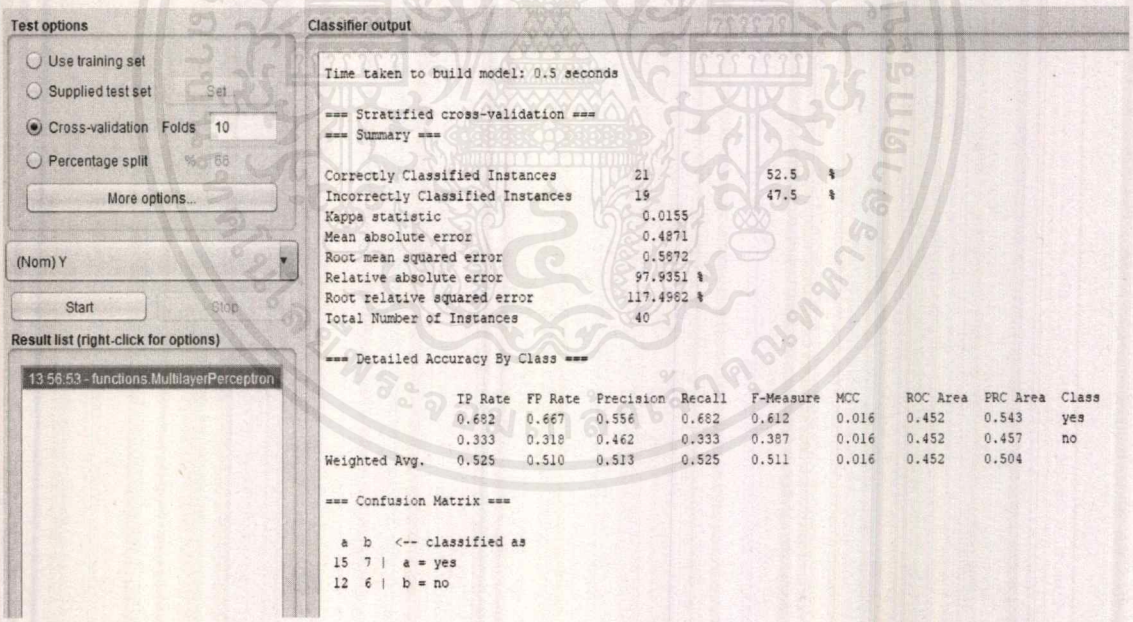
จากรูป 4.150 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 40 คน มีค่าทำนายความถูกต้องจำนวน 20 คน คิดเป็นร้อยละ 50 โดยมีค่าความไว = 0.5789 ค่าความจำเพาะ = 0.4286 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6124)^2$ = 0.3750

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.151 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

จากรูป 4.151 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 40 คน มีค่าทำนายความถูกต้องจำนวน 23 คน คิดเป็นร้อยละ 57.5 โดยมีค่าความไว = 0.4231 ค่าความจำเพาะ = 0.8571 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5017)^2$ = 0.2517

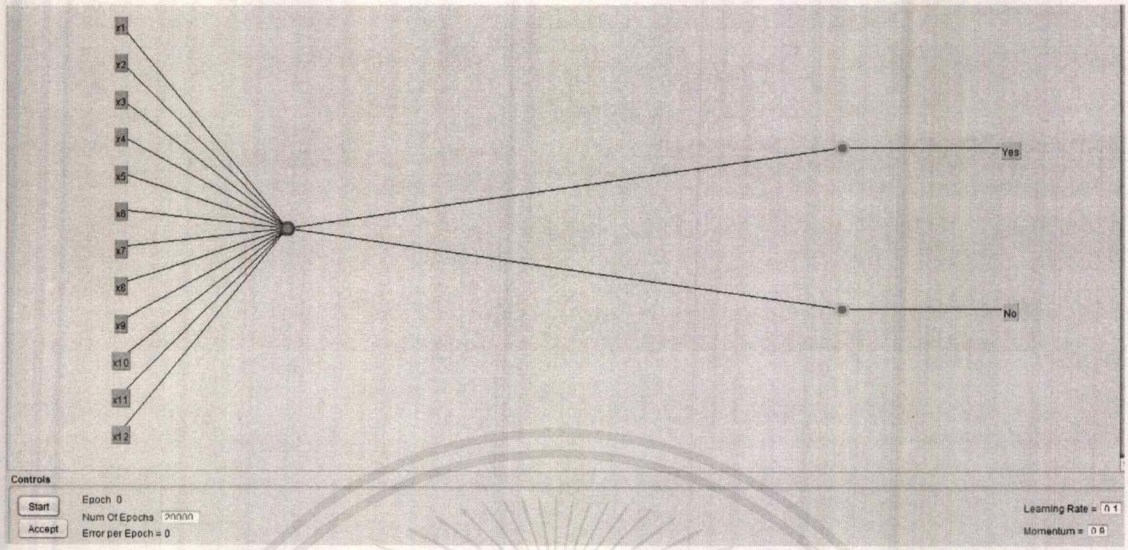


รูปที่ 4.152 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม ยอดคงเหลือในบัตรเครดิตของลูกค้าเมื่อ Random Seed 30

จากรูป 4.152 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 40 คน มีค่าทำนายความถูกต้องจำนวน 21 คน คิดเป็นร้อยละ 52.5 โดยมีค่าความไว = 0.6818 ค่าความจำเพาะ = 0.3333 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5872)^2$ = 0.3448

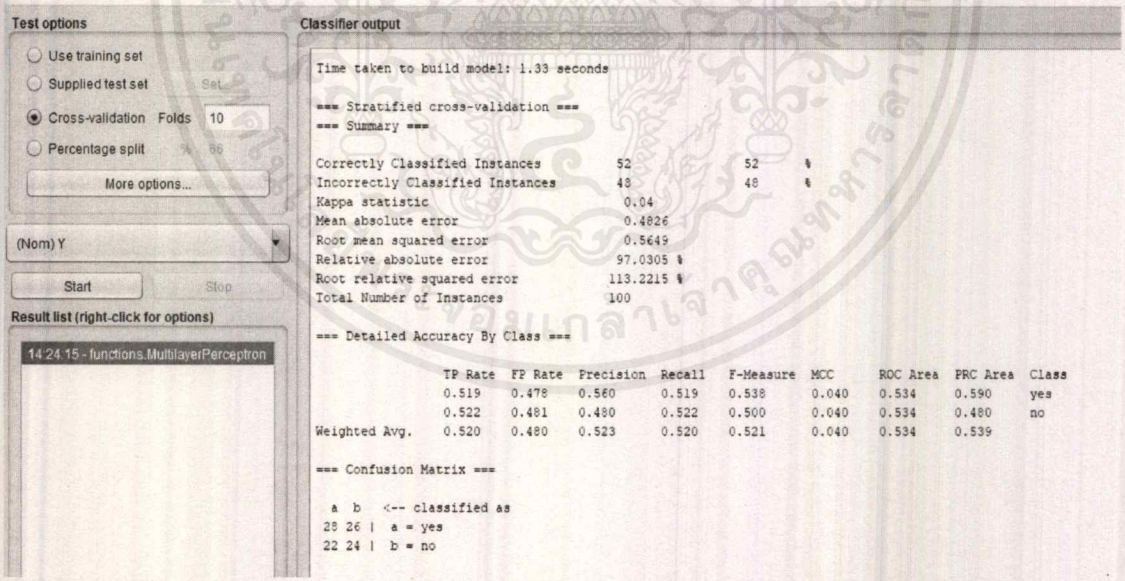
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4.4.3 คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทดสอบ 100 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



รูปที่ 4.153 โครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 10, 20 และ 30

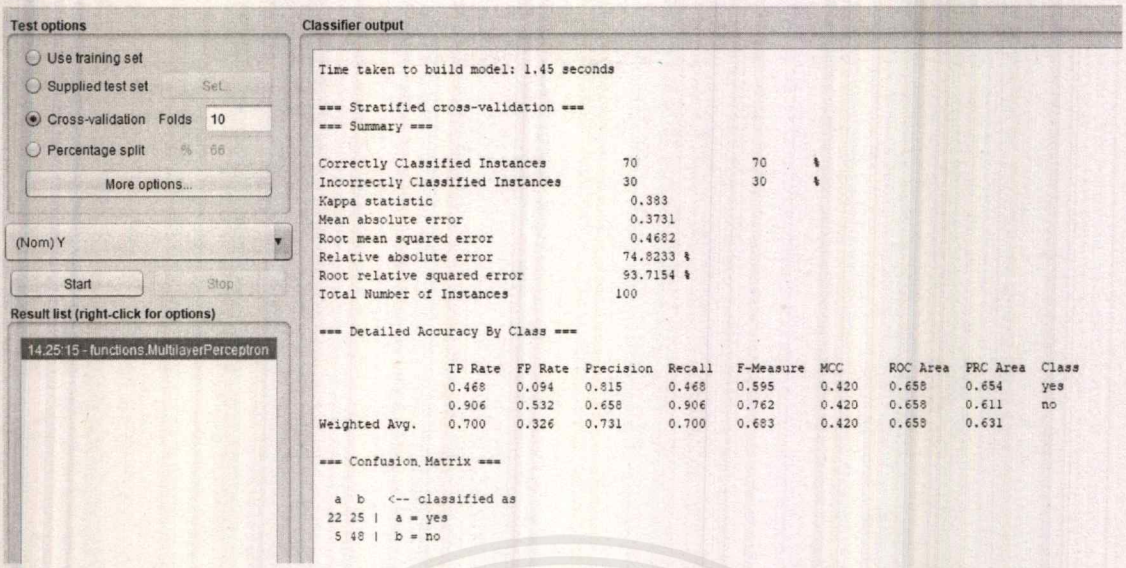
จากรูป 4.153 โครงข่ายประสาทเทียม คุณภาพไวน์แดง ชั้นข้อมูลเข้า อยู่ด้านซ้ายสุดซึ่งมีคุณลักษณะทั้งหมด 5 ตัวแปร (X_1-X_5) ชั้นซ่อน มี 1 ชั้นซ่อน ชั้นข้อมูลออกหรือชั้นข้อมูลผลลัพธ์ อยู่ด้านขวามือ มีคำตอบคือ Yes กับ No Epoch จะรันการทำงานตั้งแต่รอบที่ 1 ถึง 20,000 Error per Epoch แสดงความคลาดเคลื่อนต่อรอบ คือ Learning Rate เท่ากับ 0.1 และ Momentum เท่ากับ 0.9



รูปที่ 4.154 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 10

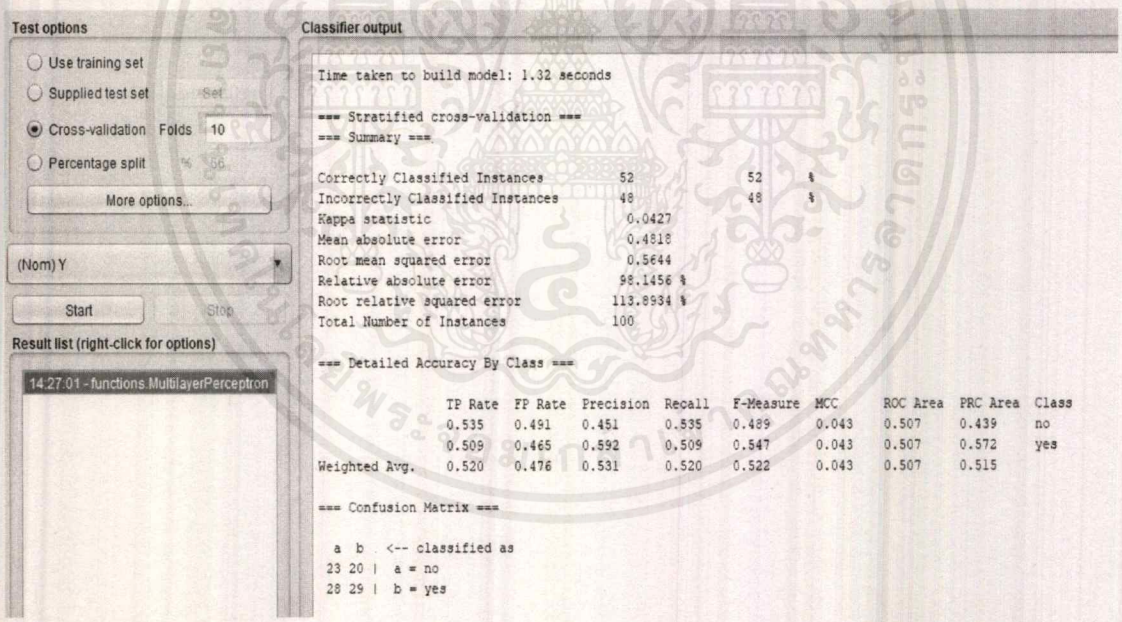
จากรูป 4.154 สุ่มครั้งที่ 1 Random Seed = 10 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 100 ขวด มีค่าทำนายความถูกต้องจำนวน 52 ขวด คิดเป็นร้อยละ 52 โดยมีค่าความไว = 0.5185 ค่าความจำเพาะ = 0.5217 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5649)^2$ = 0.3191

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.155 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 20

จากรูป 4.155 สุ่มครั้งที่ 2 Random Seed = 20 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 100 ขวด มีค่าทำนายความถูกต้องจำนวน 70 ขวด คิดเป็นร้อยละ 70 โดยมีค่าความไว = 0.4680 ค่าความจำเพาะ = 0.9057 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4682)^2$ = 0.2192



รูปที่ 4.156 ผลการวิเคราะห์ข้อมูลด้วยวิธีโครงข่ายประสาทเทียม คุณภาพไวน์แดง เมื่อ Random Seed 30

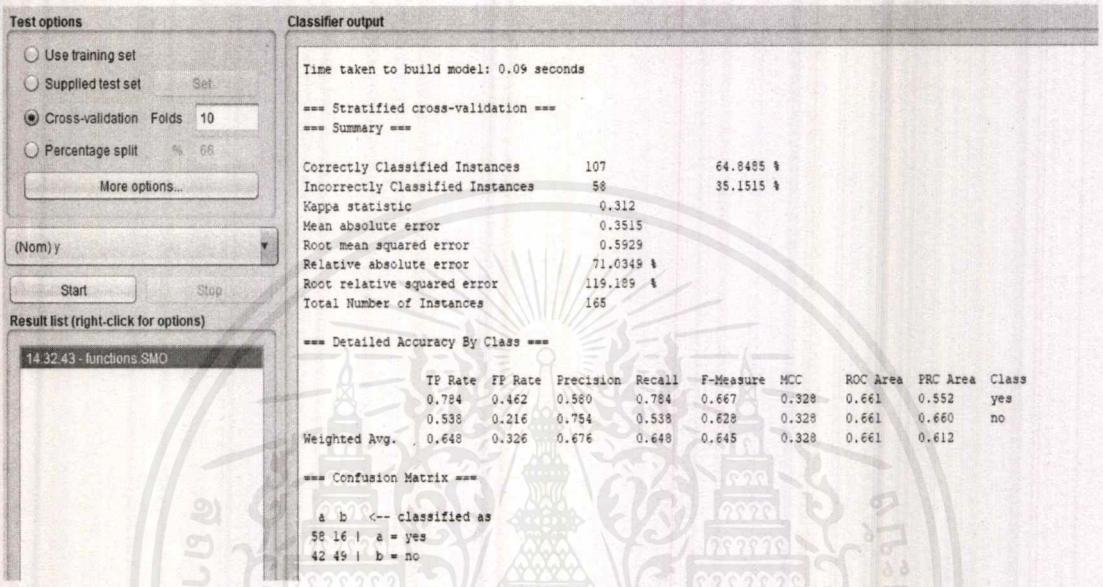
จากรูป 4.156 สุ่มครั้งที่ 3 Random Seed = 30 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 100 ขวด มีค่าทำนายความถูกต้องจำนวน 52 ขวด คิดเป็นร้อยละ 52 โดยมีค่าความไว = 0.5349 ค่าความจำเพาะ = 0.5087 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5644)^2$ = 0.3185

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO

4.5.1 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน (Over Sampling)

4.5.1.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทดสอบ 165 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) y

Start Stop

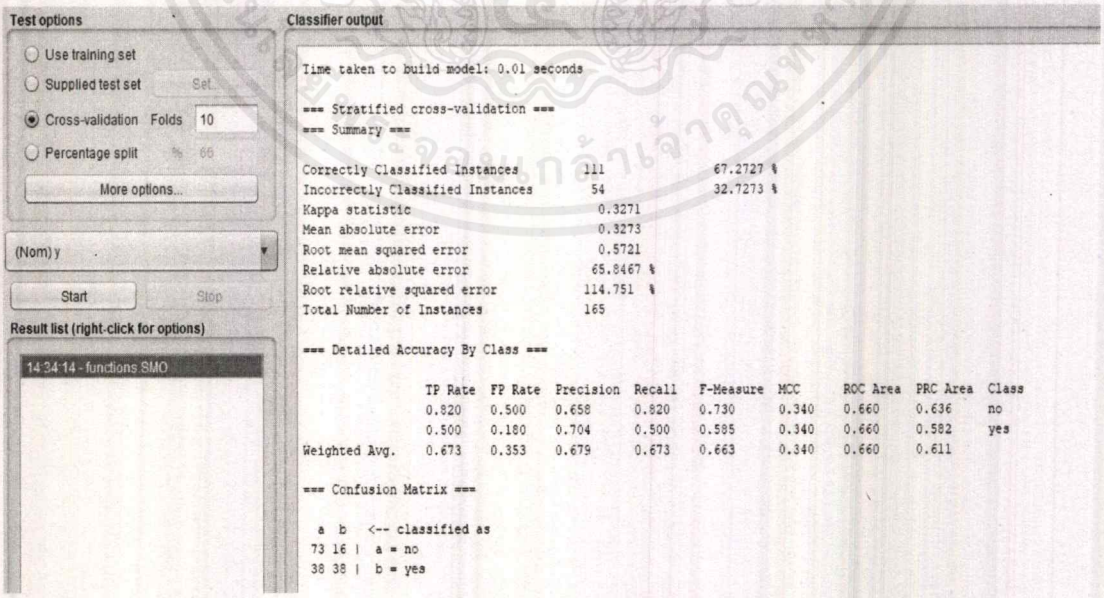
Result list (right-click for options)

14:32:43 - functions.SMO

Classifier output

```
Time taken to build model: 0.09 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         107               64.8485 %
Incorrectly Classified Instances        58               35.1515 %
Kappa statistic                          0.312
Mean absolute error                      0.3515
Root mean squared error                  0.5929
Relative absolute error                 71.0349 %
Root relative squared error            119.189  %
Total Number of Instances              165

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.784     0.462     0.580     0.784     0.667      0.328     0.661     0.552     yes
                 0.538     0.216     0.754     0.538     0.628      0.328     0.661     0.660     no
Weighted Avg.    0.648     0.326     0.676     0.648     0.645      0.328     0.661     0.612

=== Confusion Matrix ===

  a  b   <-- classified as
 58 16 |  a = yes
 42 49 |  b = no
```

รูปที่ 4.157 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) y

Start Stop

Result list (right-click for options)

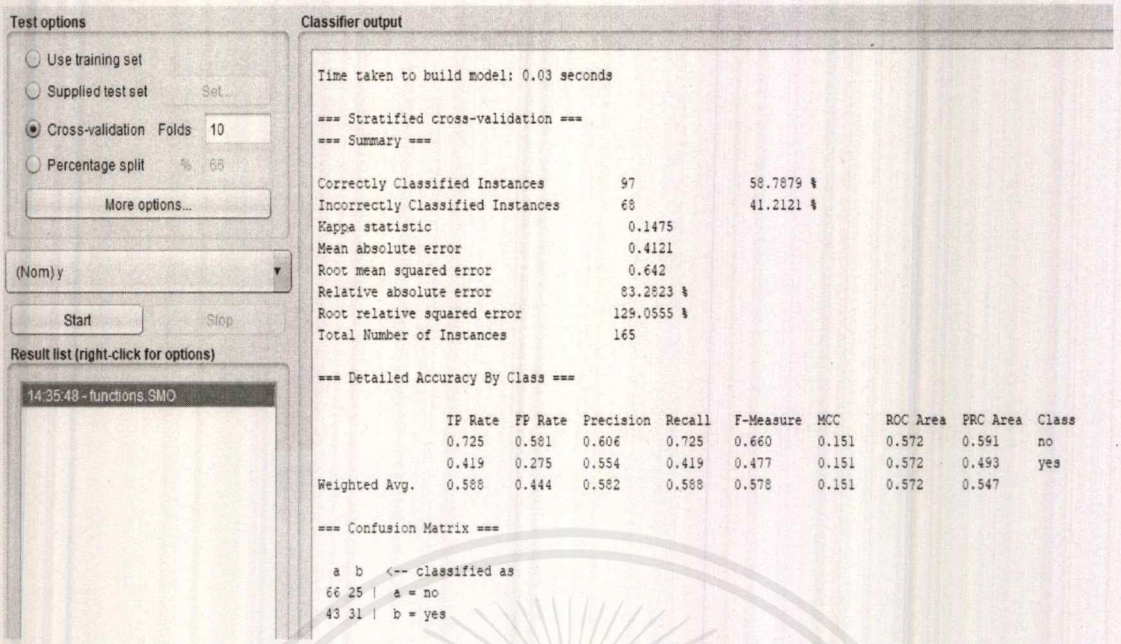
14:34:14 - functions.SMO

Classifier output

```
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         111               67.2727 %
Incorrectly Classified Instances        54               32.7273 %
Kappa statistic                          0.3271
Mean absolute error                      0.3273
Root mean squared error                  0.5721
Relative absolute error                 65.8467 %
Root relative squared error            114.751  %
Total Number of Instances              165

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.820     0.500     0.658     0.820     0.730      0.340     0.660     0.636     no
                 0.500     0.180     0.704     0.500     0.585      0.340     0.660     0.582     yes
Weighted Avg.    0.673     0.353     0.679     0.673     0.663      0.340     0.660     0.611

=== Confusion Matrix ===

  a  b   <-- classified as
 73 16 |  a = no
 38 38 |  b = yes
```

รูปที่ 4.158 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.159 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูปที่ 4.157 ถึง 4.159 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน (Over Sampling) การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

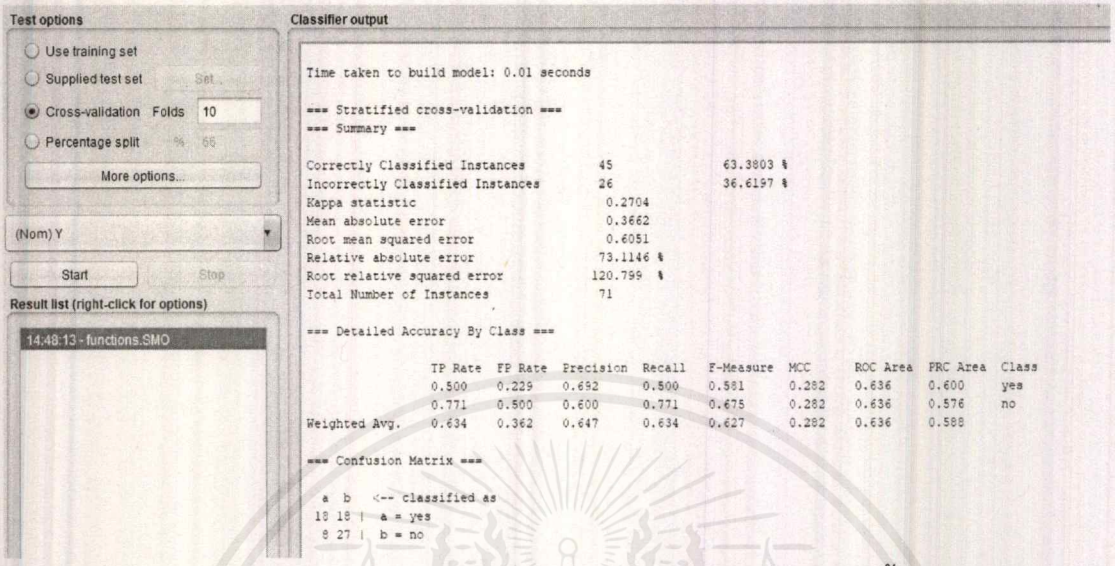
การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 165 คน มีค่าทำนายความถูกต้องจำนวน 107 คน คิดเป็นร้อยละ 64.8485 โดยมีค่าความไว = 0.7838 ค่าความจำเพาะ = 0.5385 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5929)^2$ = 0.3515

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 165 คน มีค่าทำนายความถูกต้องจำนวน 111 คน คิดเป็นร้อยละ 67.2727 โดยมีค่าความไว = 0.8202 ค่าความจำเพาะ = 0.5000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5721)^2$ = 0.3273

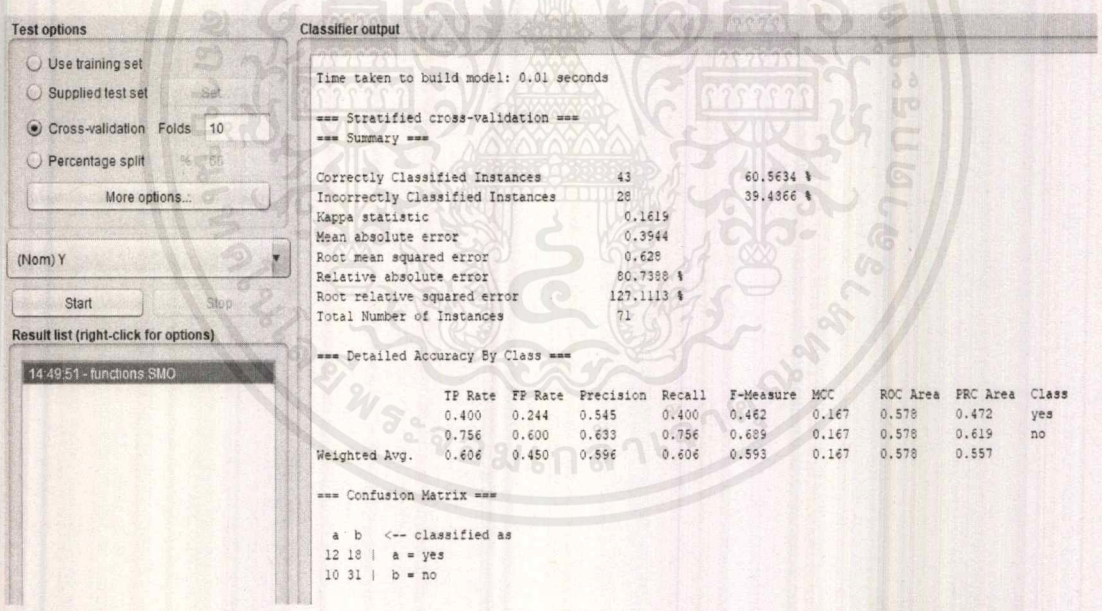
การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 165 คน มีค่าทำนายความถูกต้องจำนวน 97 คน คิดเป็นร้อยละ 58.7879 โดยมีค่าความไว = 0.7253 ค่าความจำเพาะ = 0.4189 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.642)^2$ = 0.4122

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5.1.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทดสอบ 71 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

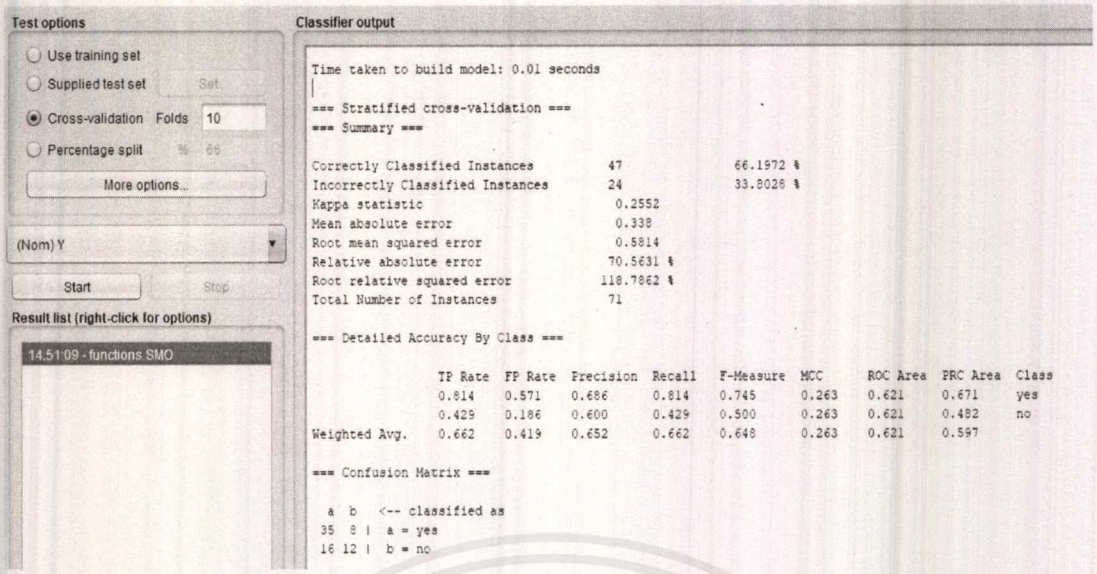


รูปที่ 4.160 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10



รูปที่ 4.161 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.162 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูปที่ 4.160 ถึง 4.162 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยการปรับความไม่สมดุลด้วยการวิธีสุ่มเกิน (Over Sampling) ยอดคงเหลือในบัตรเครดิตของลูกค้า โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

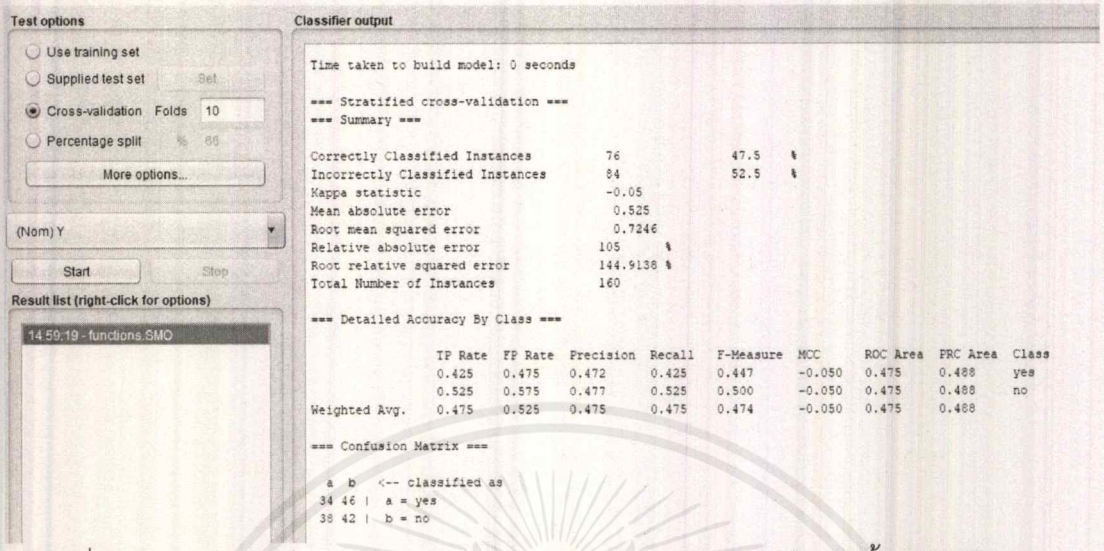
การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 71 คน มีค่าทำนายความถูกต้องจำนวน 45 คน คิดเป็นร้อยละ 63.3803 โดยมีค่าความไว = 0.5000 ค่าความจำเพาะ = 0.7714 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6051)^2$ = 0.3661

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 71 คน มีค่าทำนายความถูกต้องจำนวน 43 คน คิดเป็นร้อยละ 60.5634 โดยมีค่าความไว = 0.4000 ค่าความจำเพาะ = 0.7561 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.628)^2$ = 0.3944

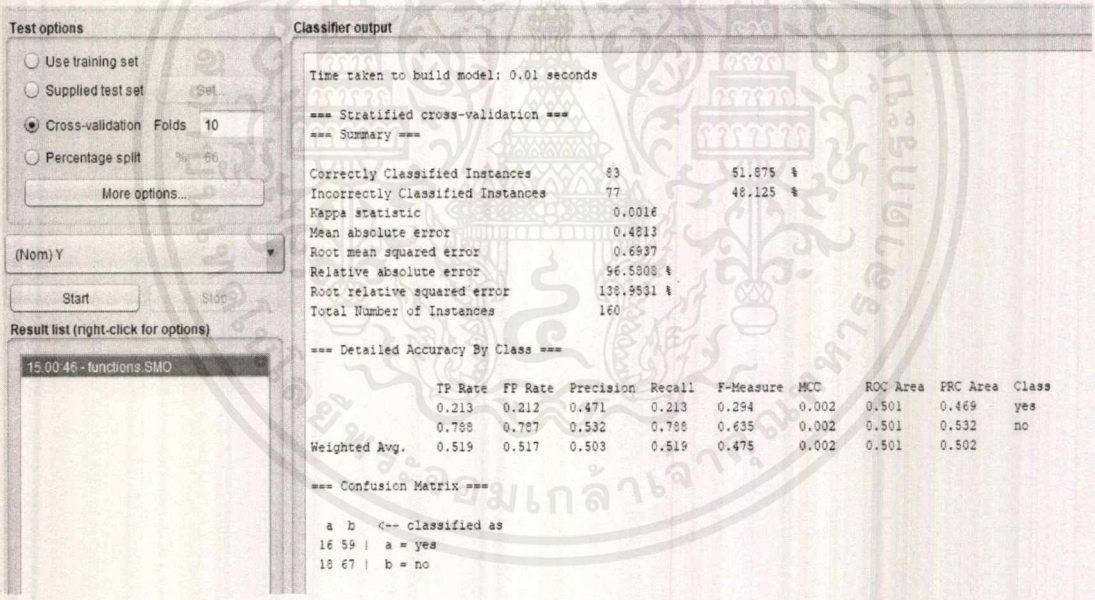
การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 71 คน มีค่าทำนายความถูกต้องจำนวน 47 คน คิดเป็นร้อยละ 66.1927 โดยมีค่าความไว = 0.8140 ค่าความจำเพาะ = 0.4286 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5814)^2$ = 0.3380

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5.1.3 คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทดสอบ 160 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

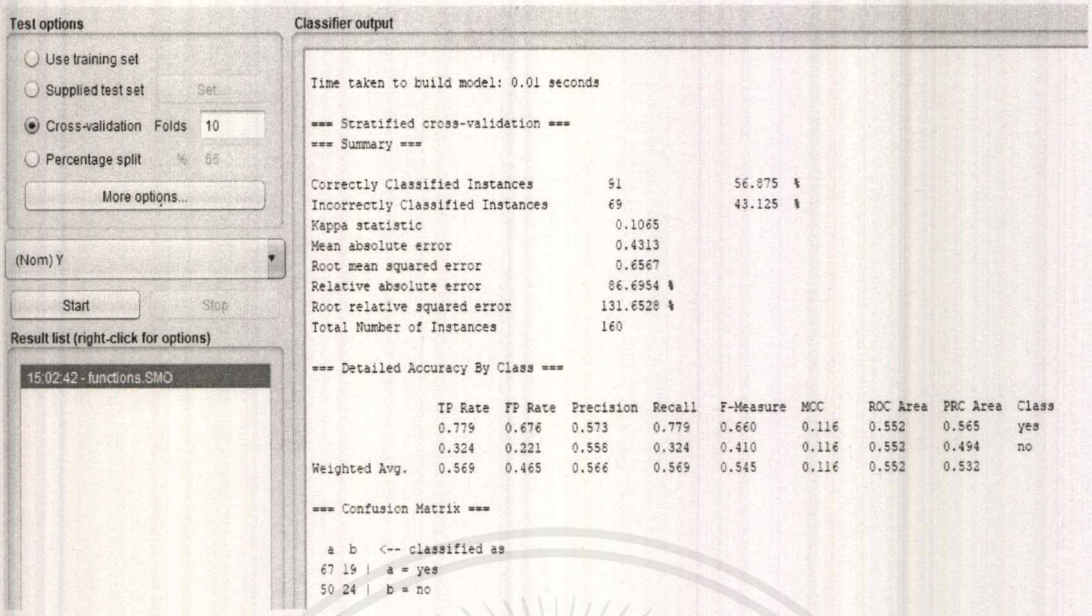


รูปที่ 4.163 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 10



รูปที่ 4.164 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.165 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูปที่ 4.163 ถึง 4.165 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยการปรับความไม่สมดุลด้วยการวิธีสุ่มเกิน (Over Sampling) คุณภาพไวน์แดง โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 160 ขวด มีค่าทำนายความถูกต้องจำนวน 76 ขวด คิดเป็นร้อยละ 47.5 โดยมีค่าความไว = 0.4250 ค่าความจำเพาะ = 0.5250 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.7246)^2$ = 0.5250

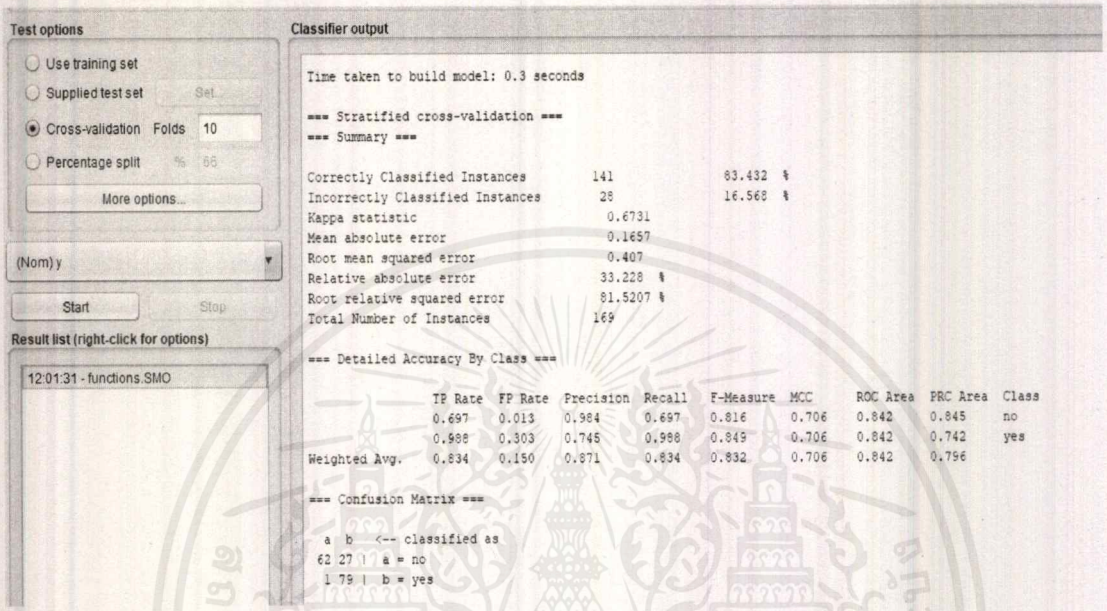
การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 160 ขวด มีค่าทำนายความถูกต้องจำนวน 83 ขวด คิดเป็นร้อยละ 51.875 โดยมีค่าความไว = 0.2133 ค่าความจำเพาะ = 0.7882 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6937)^2$ = 0.4812

การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 160 ขวด มีค่าทำนายความถูกต้องจำนวน 91 ขวด คิดเป็นร้อยละ 56.875 โดยมีค่าความไว = 0.7791 ค่าความจำเพาะ = 0.3243 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6567)^2$ = 0.4313

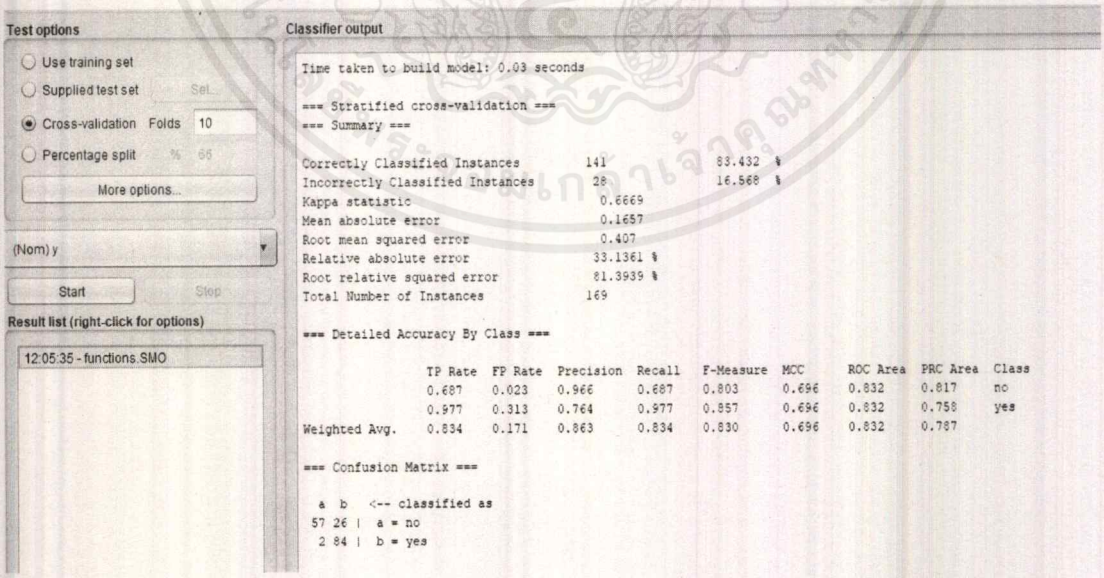
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5.2 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-Sampling)

4.5.2.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทดสอบ 169 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

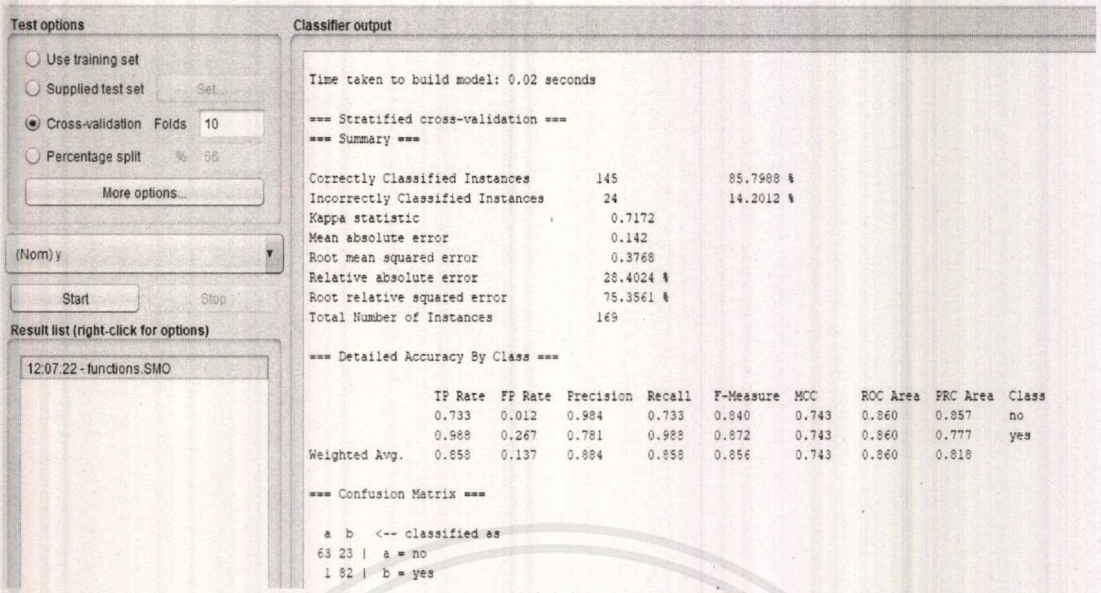


รูปที่ 4.166 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10



รูปที่ 4.167 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.168 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูปที่ 4.166 ถึง 4.168 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชีน โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-Sampling) การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

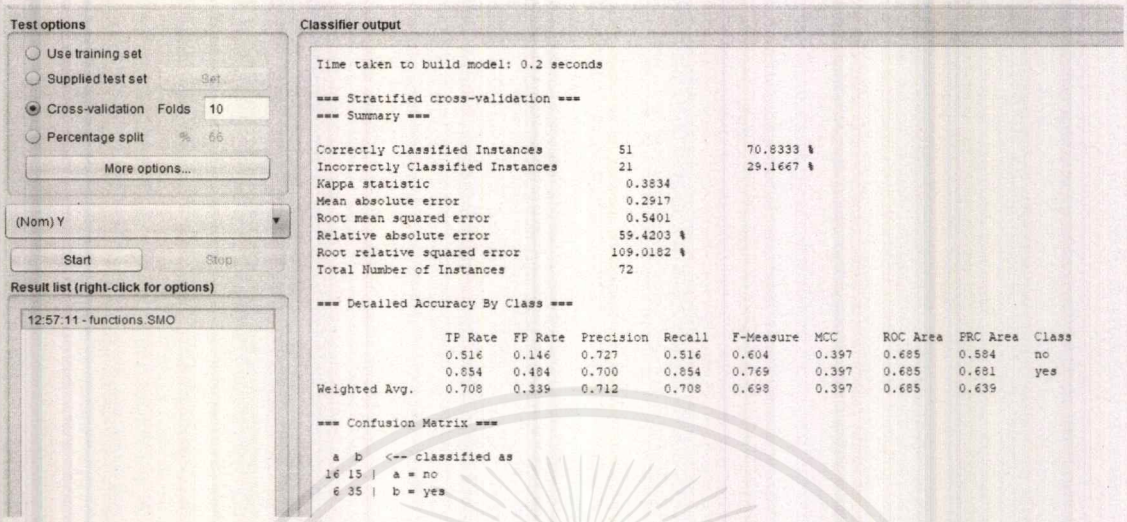
การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 169 คน มีค่าทำนายความถูกต้องจำนวน 141 คน คิดเป็นร้อยละ 83.432 โดยมีค่าความไว = 0.6966 ค่าความจำเพาะ = 0.9875 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.407)^2$ = 0.1656

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 169 คน มีค่าทำนายความถูกต้องจำนวน 141 คน คิดเป็นร้อยละ 83.432 โดยมีค่าความไว = 0.6867 ค่าความจำเพาะ = 0.9767 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.407)^2$ = 0.1656

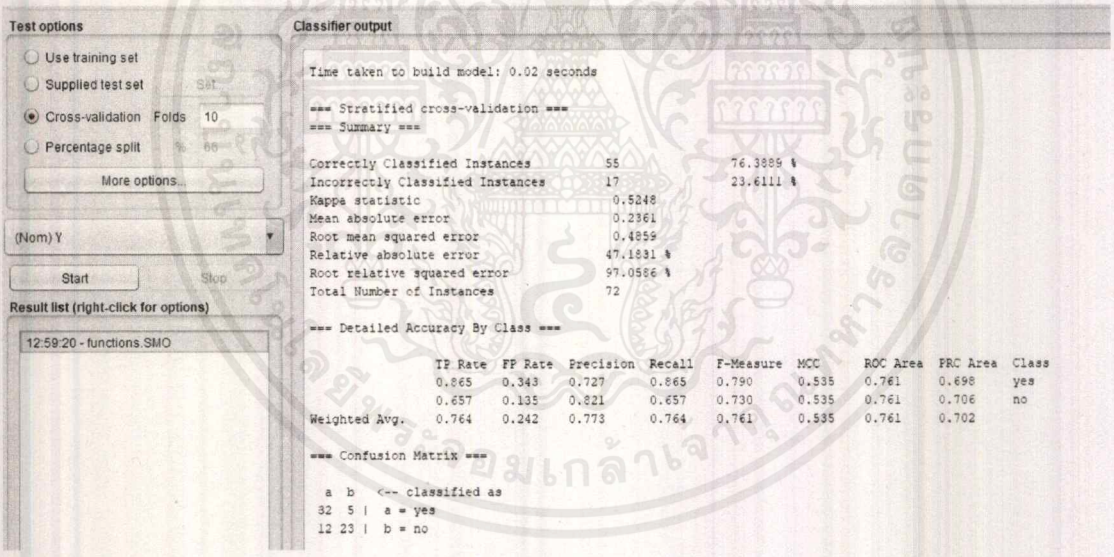
การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 169 คน มีค่าทำนายความถูกต้องจำนวน 145 คน คิดเป็นร้อยละ 85.7988 โดยมีค่าความไว = 0.7326 ค่าความจำเพาะ = 0.9880 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.3768)^2$ = 0.1420

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5.2.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทดสอบ 72 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

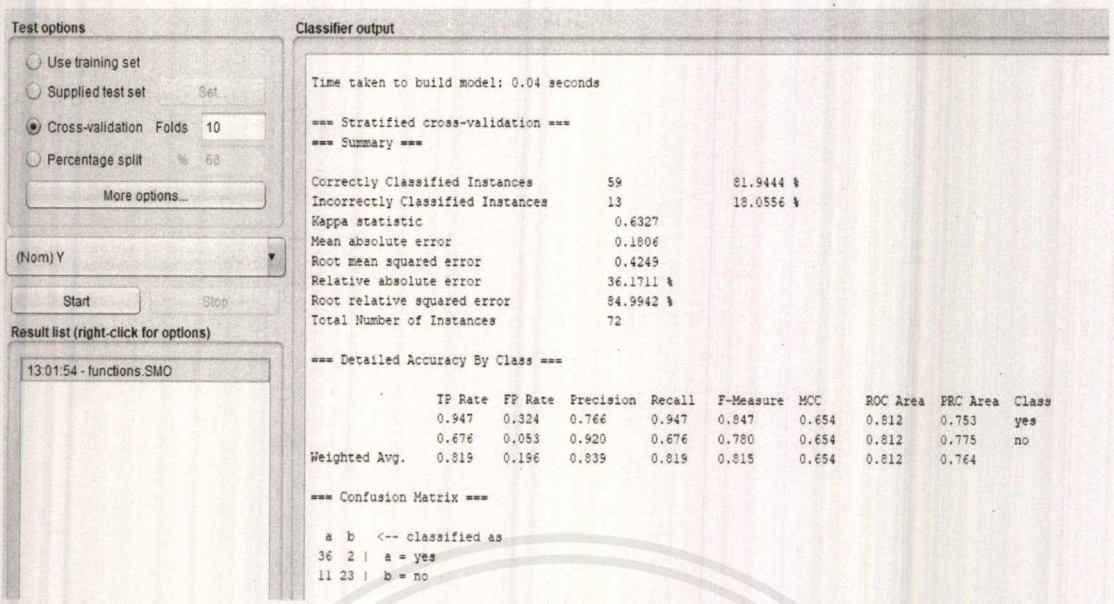


รูปที่ 4.169 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10



รูปที่ 4.170 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.171 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูปที่ 4.169 ถึง 4.171 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชีน โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-Sampling) ยอดคงเหลือในบัตรเครดิตของลูกค้าโดยทำการสุ่ม 3 ครั้ง Random Seed 10 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

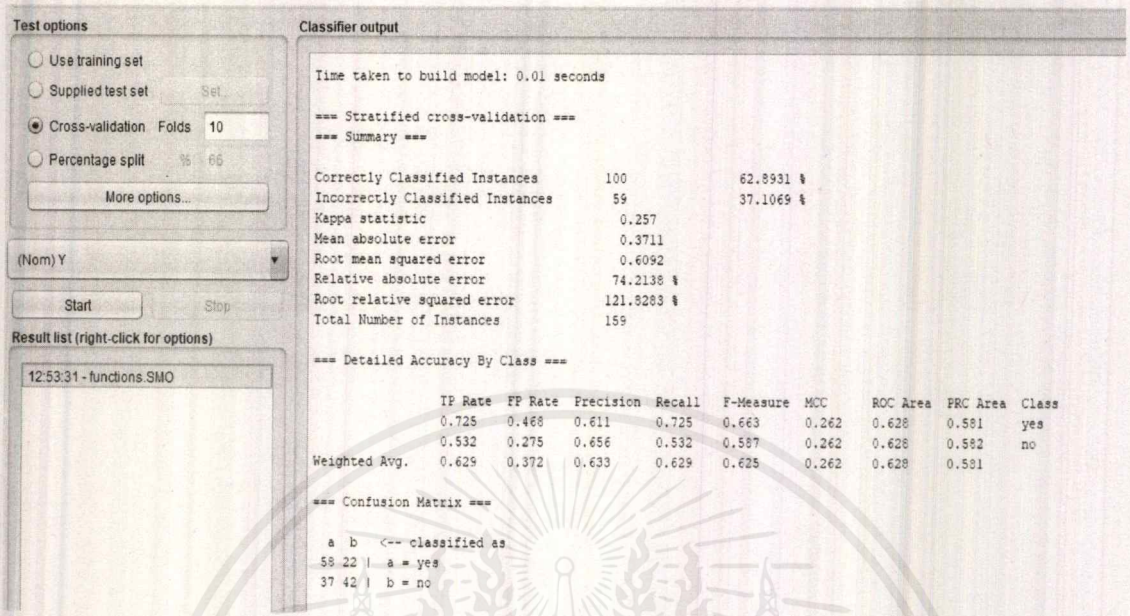
การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 72 คน มีค่าทำนายความถูกต้องจำนวน 51 คน คิดเป็นร้อยละ 70.8333 โดยมีค่าความไว = 0.5161 ค่าความจำเพาะ = 0.8537 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5401)^2$ = 0.2917

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 72 คน มีค่าทำนายความถูกต้องจำนวน 55 คน คิดเป็นร้อยละ 76.3889 โดยมีค่าความไว = 0.8649 ค่าความจำเพาะ = 0.6571 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4859)^2$ = 0.2361

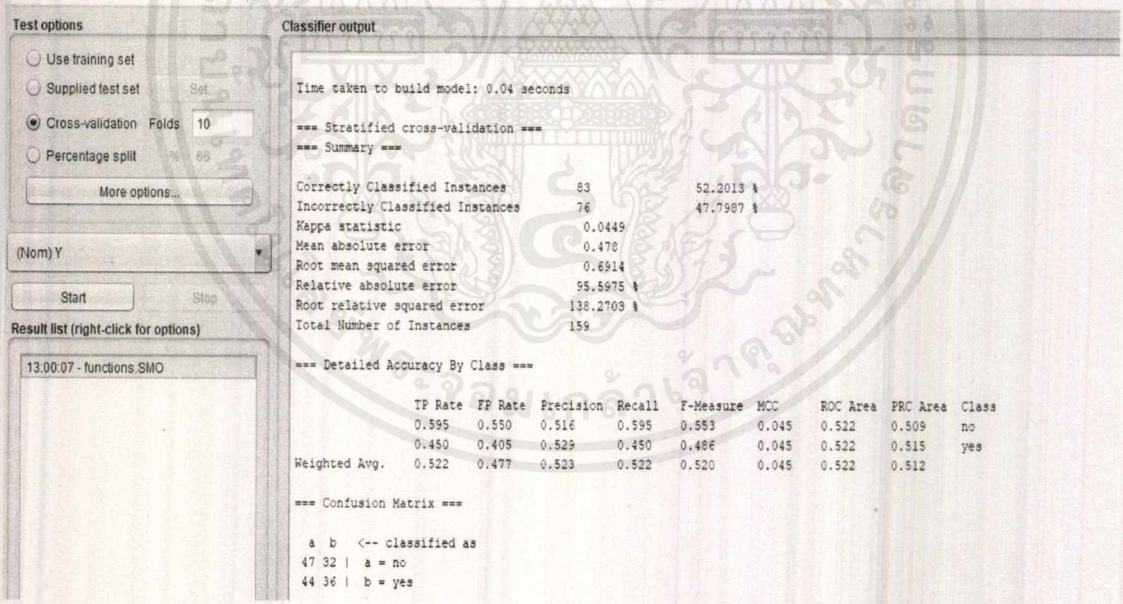
การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 72 คน มีค่าทำนายความถูกต้องจำนวน 59 คน คิดเป็นร้อยละ 81.9444 โดยมีค่าความไว = 0.9474 ค่าความจำเพาะ = 0.6765 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.4249)^2$ = 0.1805

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5.2.3 คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทดสอบ 159 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

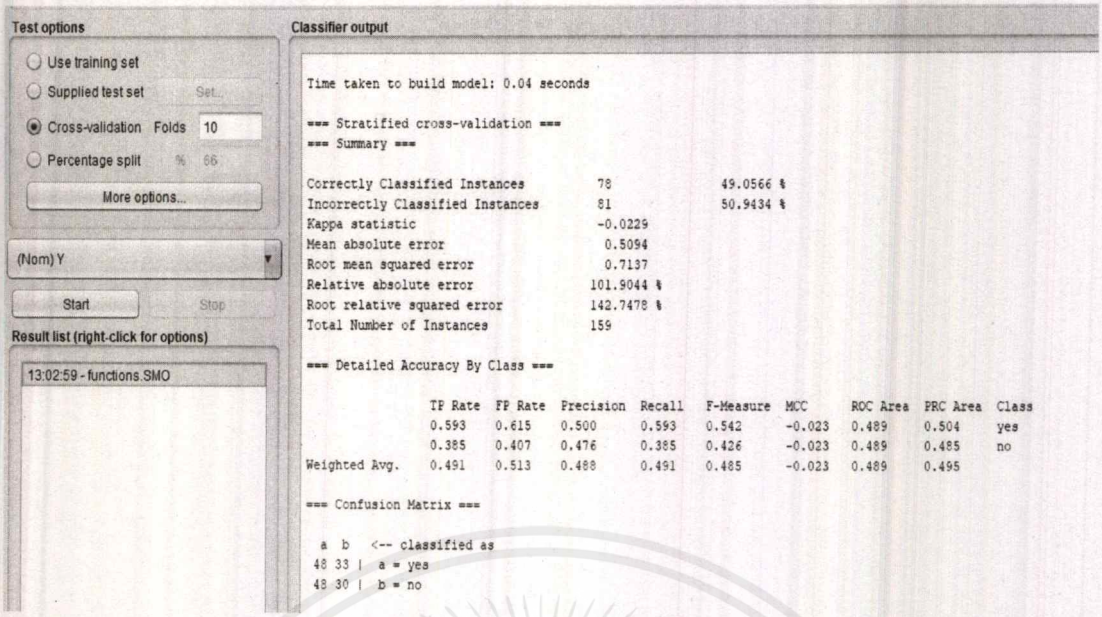


รูปที่ 4.172 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 10



รูปที่ 4.173 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.174 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูปที่ 4.172 ถึง 4.174 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชีน โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-Sampling) คุณภาพไวน์แดง โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 159 ขวด มีค่าทำนายความถูกต้องจำนวน 100 ขวด คิดเป็นร้อยละ 62.8931 โดยมีค่าความไว = 0.725 ค่าความจำเพาะ = 0.5316 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6092)^2$ = 0.3711

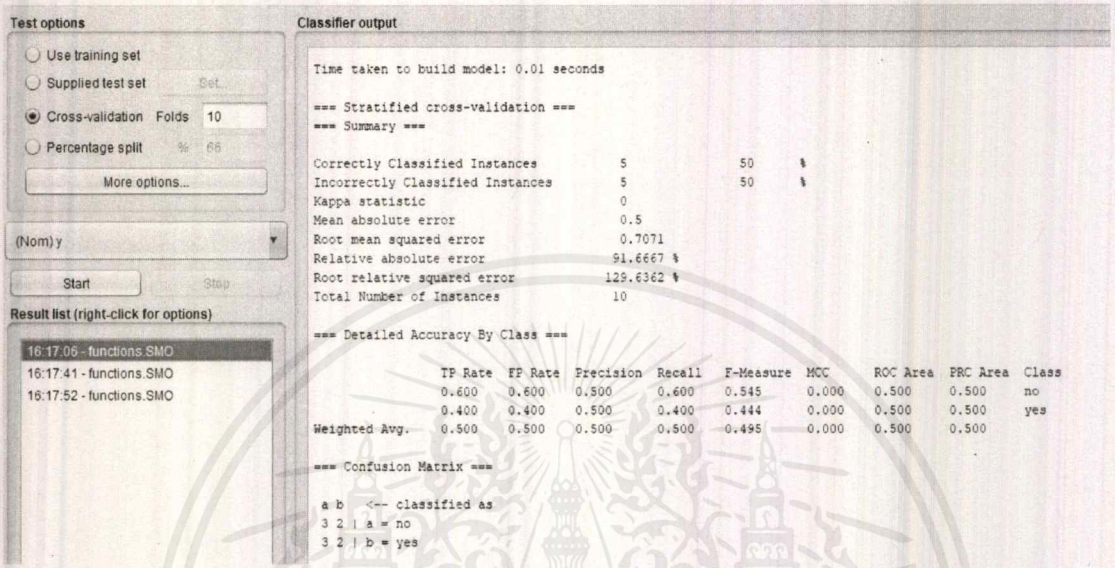
การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 159 ขวด มีค่าทำนายความถูกต้องจำนวน 83 ขวด คิดเป็นร้อยละ 52.2013 โดยมีค่าความไว = 0.5950 ค่าความจำเพาะ = 0.4500 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6914)^2$ = 0.4780

การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 159 ขวด มีค่าทำนายความถูกต้องจำนวน 78 ขวด คิดเป็นร้อยละ 49.0566 โดยมีค่าความไว = 0.5926 ค่าความจำเพาะ = 0.3846 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.7137)^2$ = 0.5094

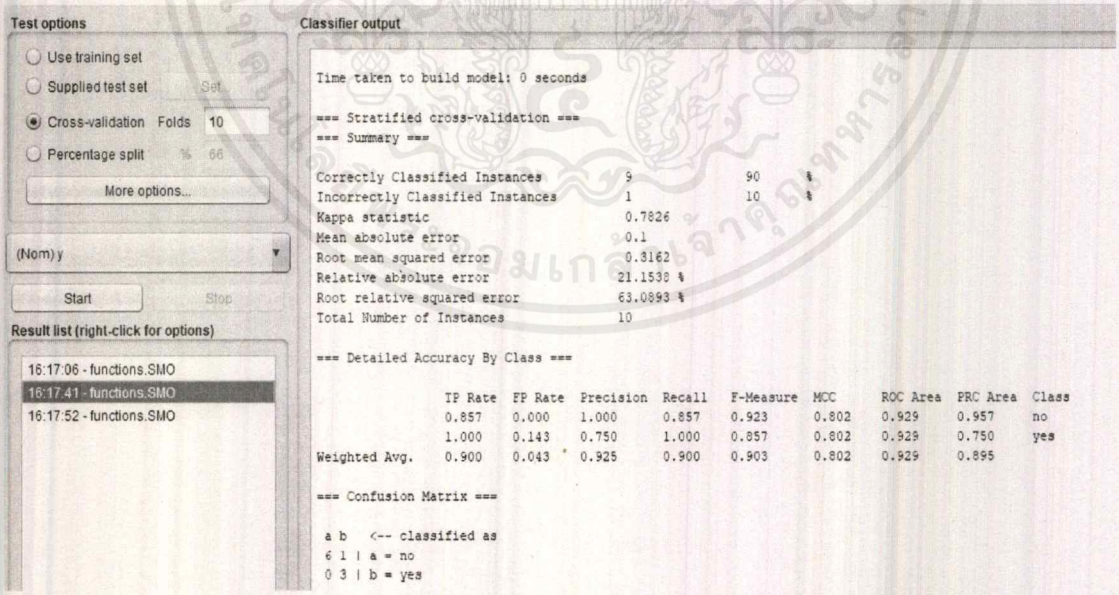
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5.3 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด (Under Sampling)

4.5.3.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทดสอบ 10 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

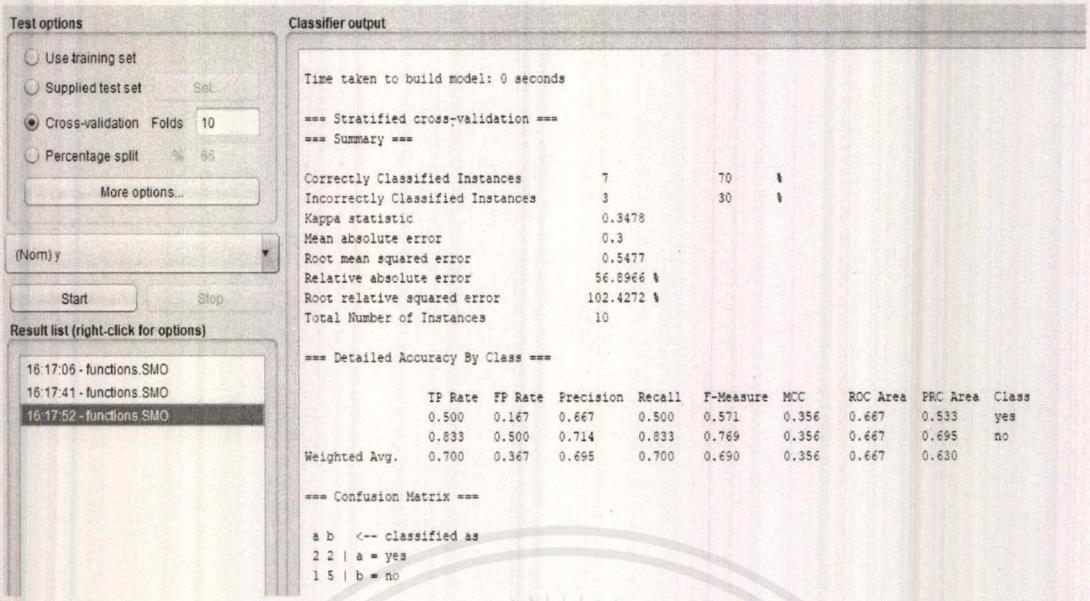


รูปที่ 4.175 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10



รูปที่ 4.176 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.177 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูปที่ 4.175 ถึง 4.177 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด (Under Sampling) การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

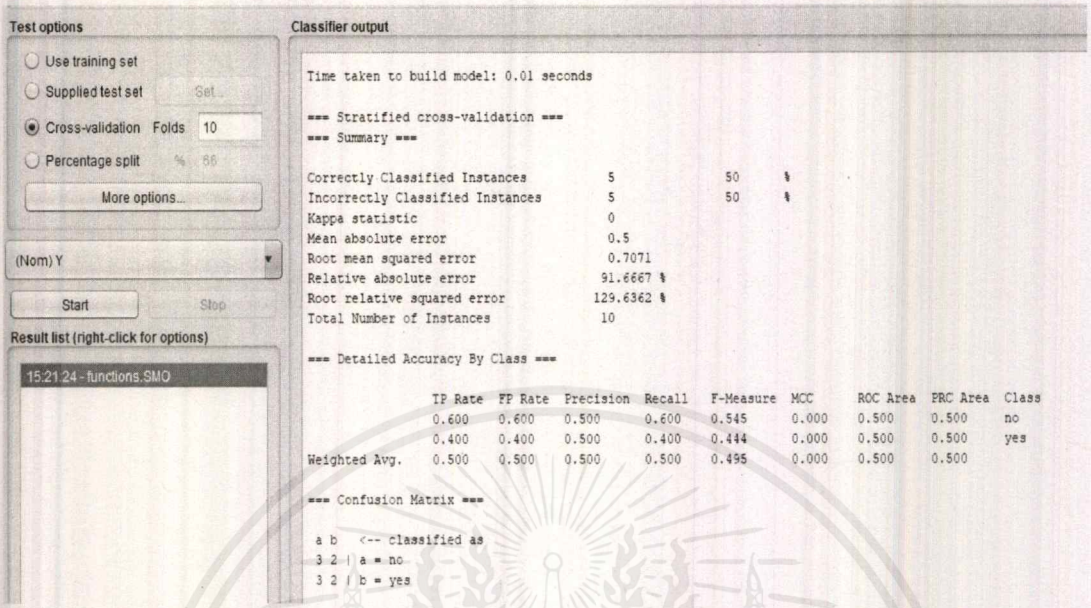
การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 5 คน คิดเป็นร้อยละ 50 โดยมีค่าความไว = 0.6000 ค่าความจำเพาะ = 0.4000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.7071)^2$ = 0.4999

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 9 คน คิดเป็นร้อยละ 90 โดยมีค่าความไว = 0.8570 ค่าความจำเพาะ = 1 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.3162)^2$ = 0.1000

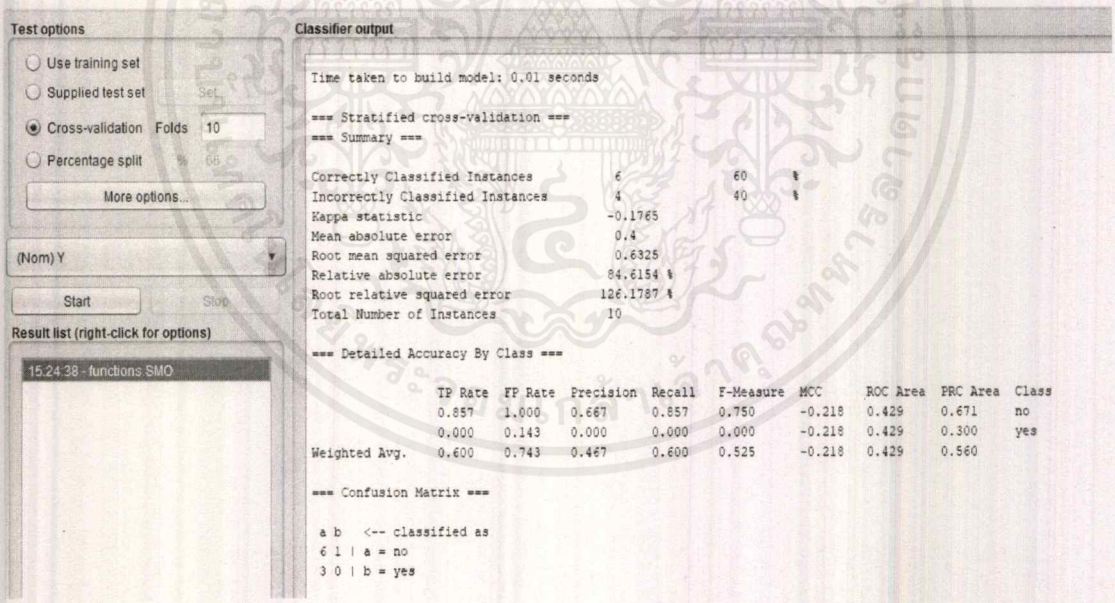
การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 7 คน คิดเป็นร้อยละ 70 โดยมีค่าความไว = 0.5000 ค่าความจำเพาะ = 0.8333 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5477)^2$ = 0.3000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5.3.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทดสอบ 10 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

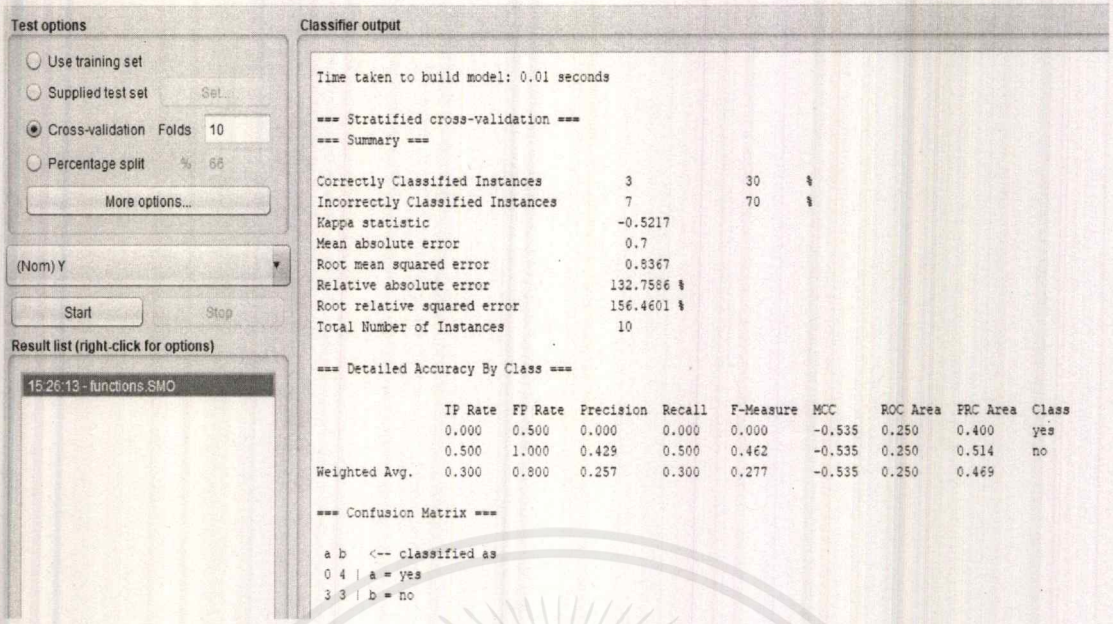


รูปที่ 4.178 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10



รูปที่ 4.179 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.180 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูปที่ 4.178 ถึง 4.180 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด (Under Sampling) ยอดคงเหลือในบัตรเครดิตของลูกค้า โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

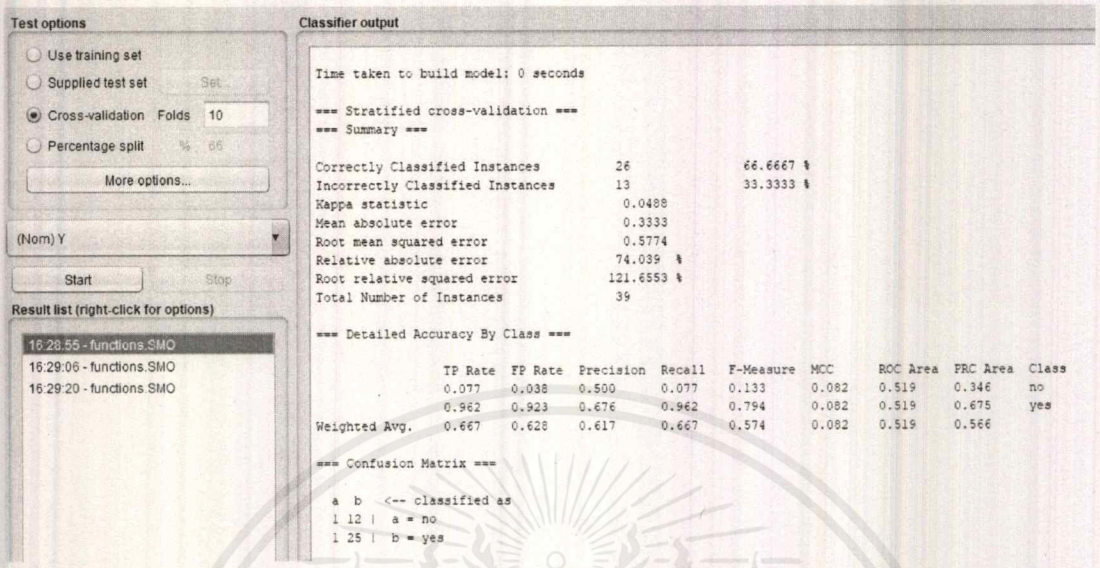
การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 5 คน คิดเป็นร้อยละ 50 โดยมีค่าความไว = 0.6000 ค่าความจำเพาะ = 0.4000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.7071)^2$ = 0.4999

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 6 คน คิดเป็นร้อยละ 60 โดยมีค่าความไว = 0.8571 ค่าความจำเพาะ = 0 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6325)^2$ = 0.4001

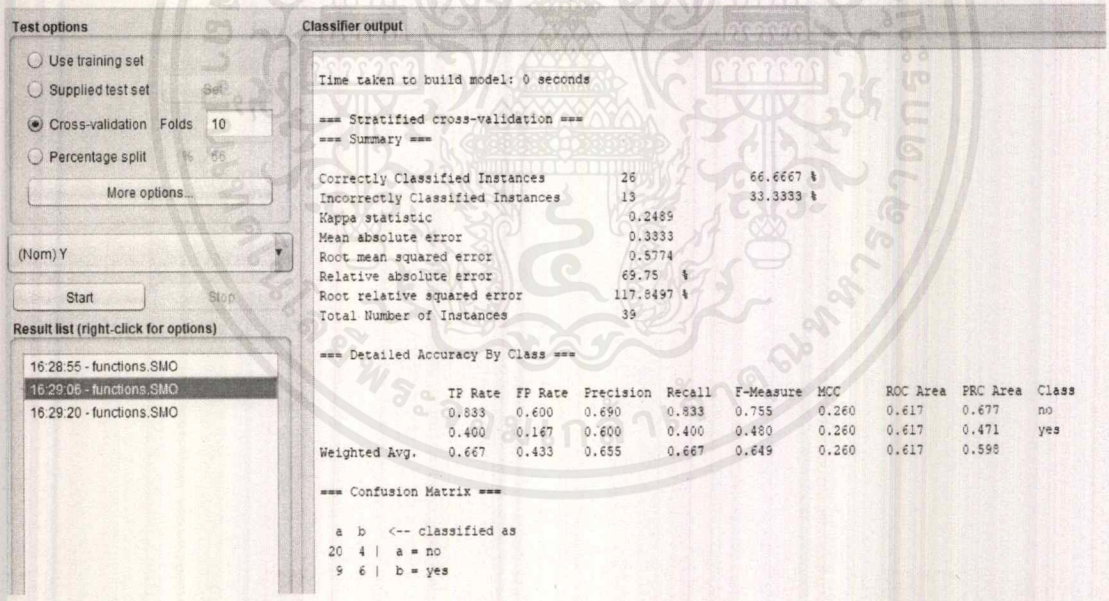
การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 คน มีค่าทำนายความถูกต้องจำนวน 3 คน คิดเป็นร้อยละ 30 โดยมีค่าความไว = 0.0000 ค่าความจำเพาะ = 0.5000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.8367)^2$ = 0.7000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5.3.3 คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทดสอบ 39 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

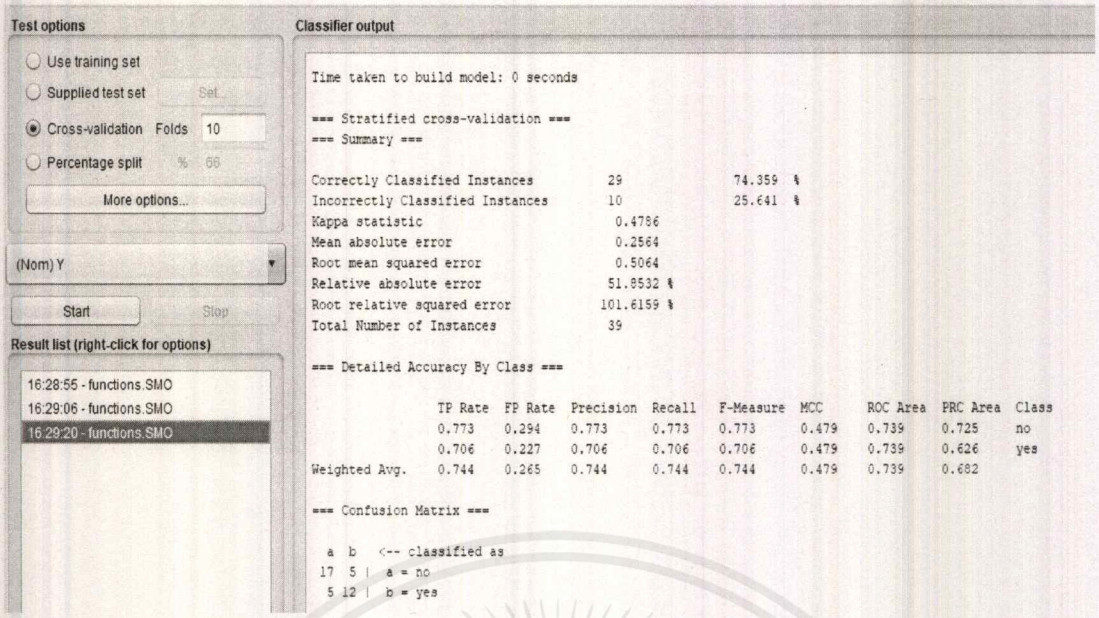


รูปที่ 4.181 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 10



รูปที่ 4.182 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.183 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูปที่ 4.181 ถึง 4.183 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด (Under Sampling) คุณภาพไวน์แดง โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 39 ขวด มีค่าทำนายความถูกต้องจำนวน 26 ขวด คิดเป็นร้อยละ 66.6667 โดยมีค่าความไว = 0.0769 ค่าความจำเพาะ = 0.9615 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5774)^2$ = 0.3334

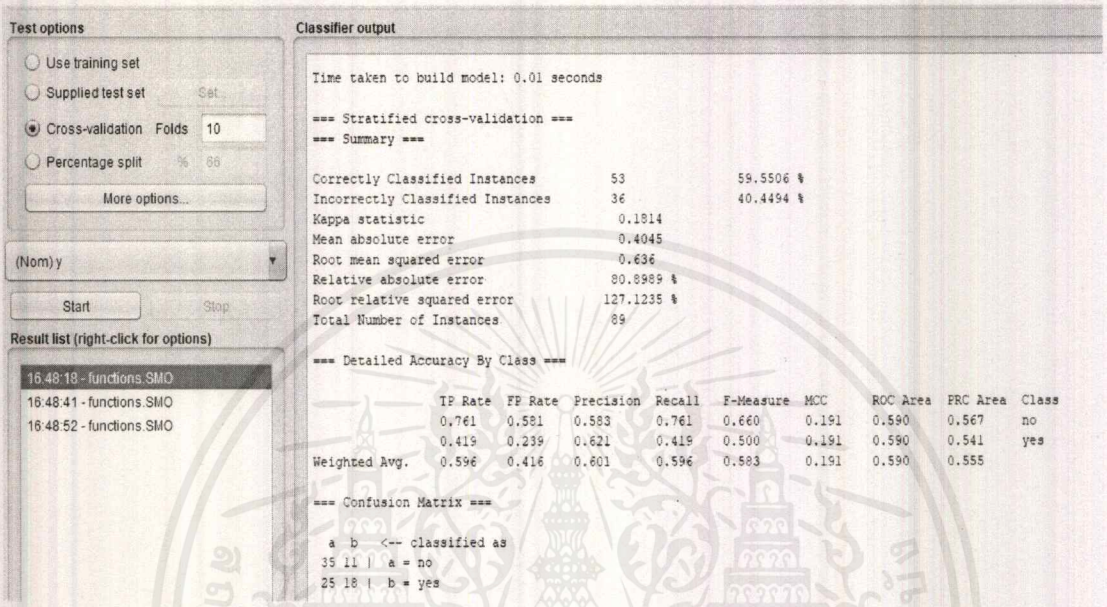
การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 39 ขวด มีค่าทำนายความถูกต้องจำนวน 26 ขวด คิดเป็นร้อยละ 66.6667 โดยมีค่าความไว = 0.8333 ค่าความจำเพาะ = 0.4000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5774)^2$ = 0.3334

การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 10 ขวด มีค่าทำนายความถูกต้องจำนวน 29 ขวด คิดเป็นร้อยละ 74.3590 โดยมีค่าความไว = 0.7727 ค่าความจำเพาะ = 0.7059 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5064)^2$ = 0.2564

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5.4 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยการปรับความไม่สมดุลด้วยวิธีการผสมผสาน (Hybrid Methods)

4.5.4.1 การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทดสอบ 89 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) y

Start Stop

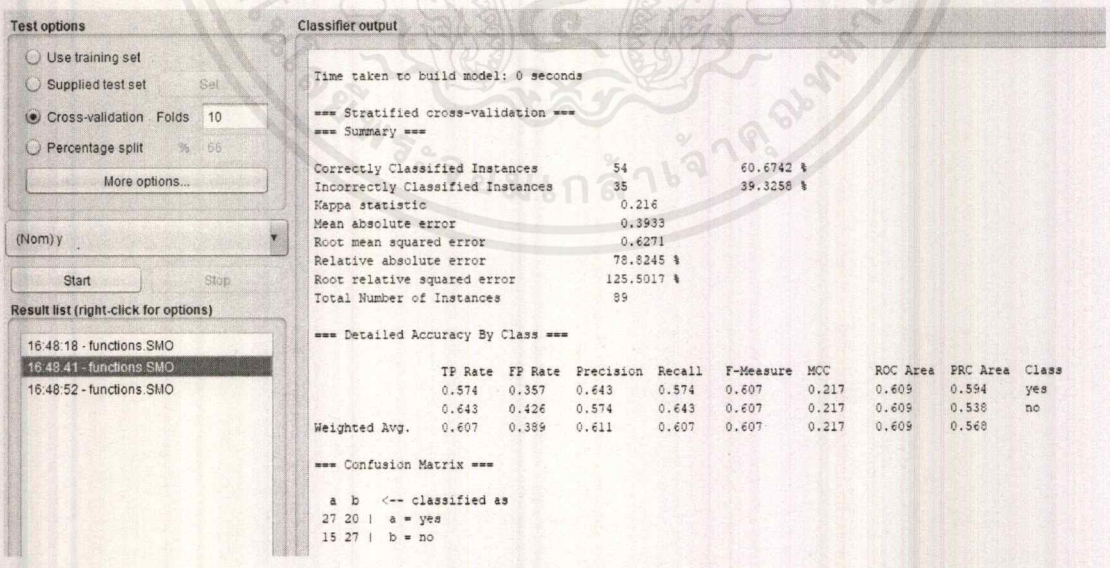
Result list (right-click for options)

16:48:18 - functions.SMO
16:48:41 - functions.SMO
16:48:52 - functions.SMO

Classifier output

```
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          53               59.5506 %
Incorrectly Classified Instances        36               40.4494 %
Kappa statistic                          0.1814
Mean absolute error                      0.4045
Root mean squared error                  0.636
Relative absolute error                 80.8989 %
Root relative squared error            127.1235 %
Total Number of Instances               89

=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
                 0.761     0.581     0.583     0.761     0.660     0.191     0.590     0.567     no
                 0.419     0.239     0.621     0.419     0.500     0.191     0.590     0.541     yes
Weighted Avg.    0.596     0.416     0.601     0.596     0.583     0.191     0.590     0.555

=== Confusion Matrix ===

  a  b   <-- classified as
 35 11 |  a = no
 25 18 |  b = yes
```

รูปที่ 4.184 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) y

Start Stop

Result list (right-click for options)

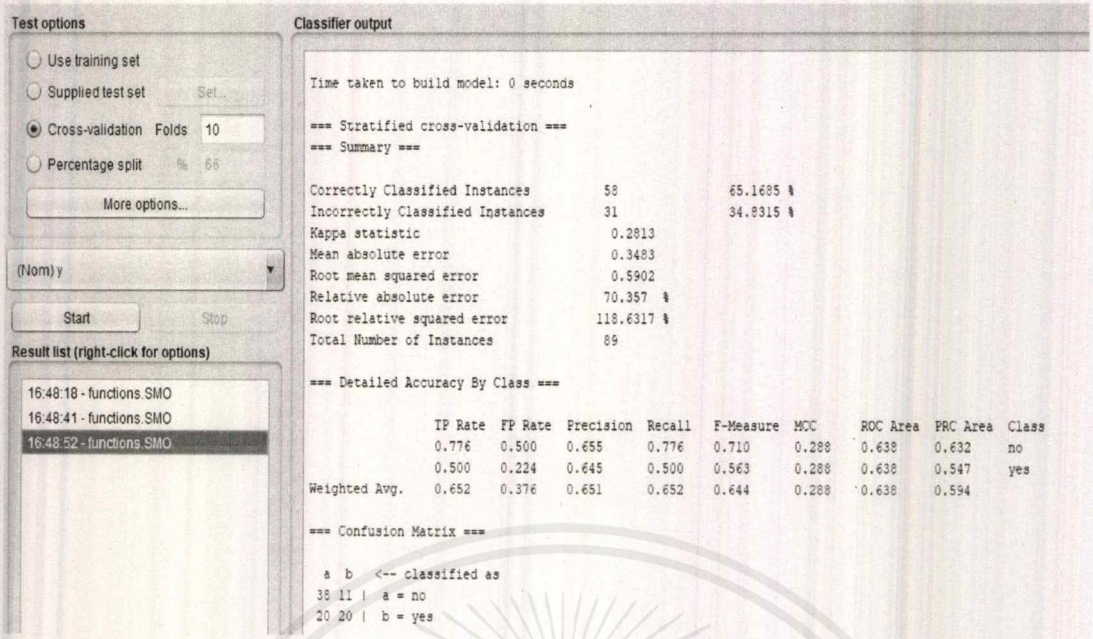
16:48:18 - functions.SMO
16:48:41 - functions.SMO
16:48:52 - functions.SMO

Classifier output

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          54               60.6742 %
Incorrectly Classified Instances        35               39.3258 %
Kappa statistic                          0.216
Mean absolute error                      0.3933
Root mean squared error                  0.6271
Relative absolute error                 78.8245 %
Root relative squared error            125.5017 %
Total Number of Instances               89

=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
                 0.574     0.357     0.643     0.574     0.607     0.217     0.609     0.594     yes
                 0.643     0.426     0.574     0.643     0.607     0.217     0.609     0.538     no
Weighted Avg.    0.607     0.389     0.611     0.607     0.607     0.217     0.609     0.568

=== Confusion Matrix ===

  a  b   <-- classified as
 27 20 |  a = yes
 15 27 |  b = no
```

รูปที่ 4.185 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.186 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 30

จากรูปที่ 4.184 ถึง 4.186 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชีน โดยการปรับความไม่สมดุลด้วยวิธีการผสมผสาน (Hybrid Methods) การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

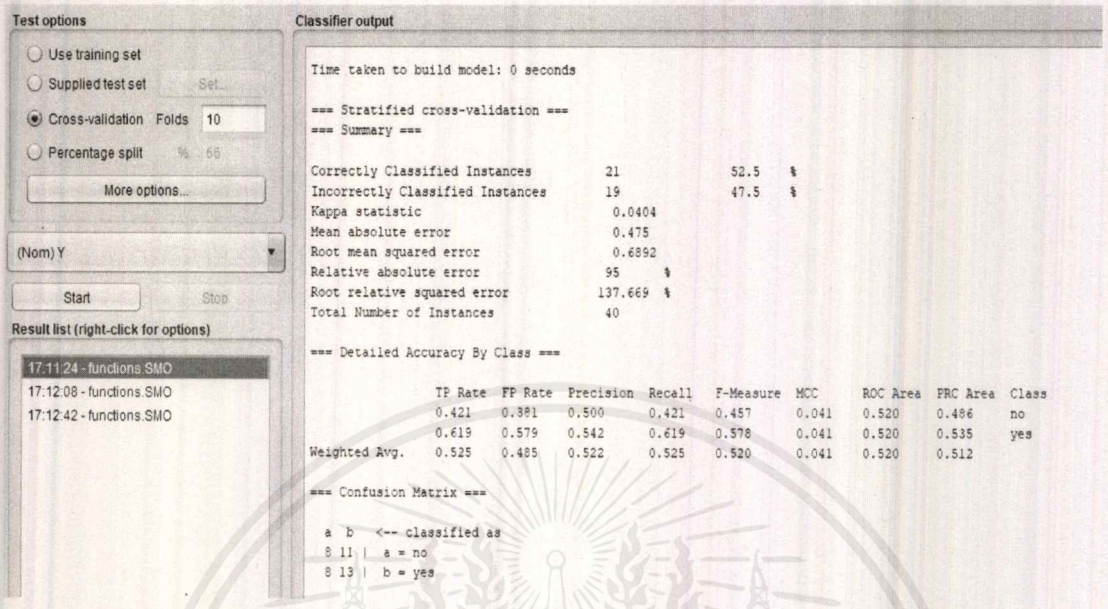
การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 89 คน มีค่าทำนายความถูกต้องจำนวน 53 คน คิดเป็นร้อยละ 59.5506 โดยมีค่าความไว = 0.7609 ค่าความจำเพาะ = 0.4186 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6360)^2$ = 0.4045

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 89 คน มีค่าทำนายความถูกต้องจำนวน 54 คน คิดเป็นร้อยละ 60.6742 โดยมีค่าความไว = 0.5745 ค่าความจำเพาะ = 0.6429 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6271)^2$ = 0.3933

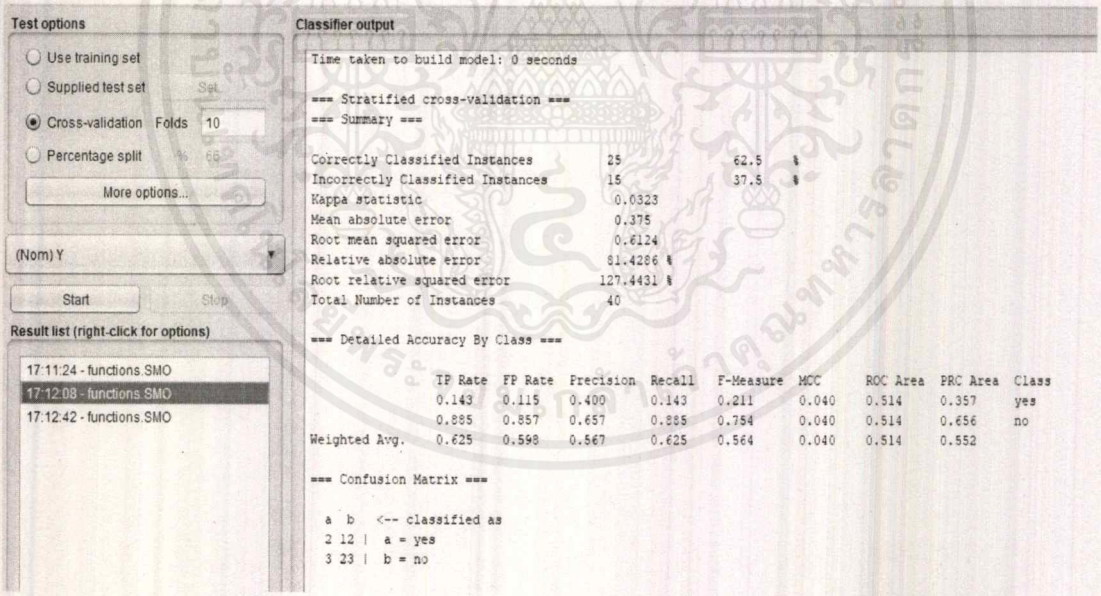
การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มคนไข้จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 89 คน มีค่าทำนายความถูกต้องจำนวน 58 คน คิดเป็นร้อยละ 65.1685 โดยมีค่าความไว = 0.7755 ค่าความจำเพาะ = 0.5000 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.5902)^2$ = 0.3483

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5.4.2 ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance) จำนวนข้อมูลทดสอบ 40 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

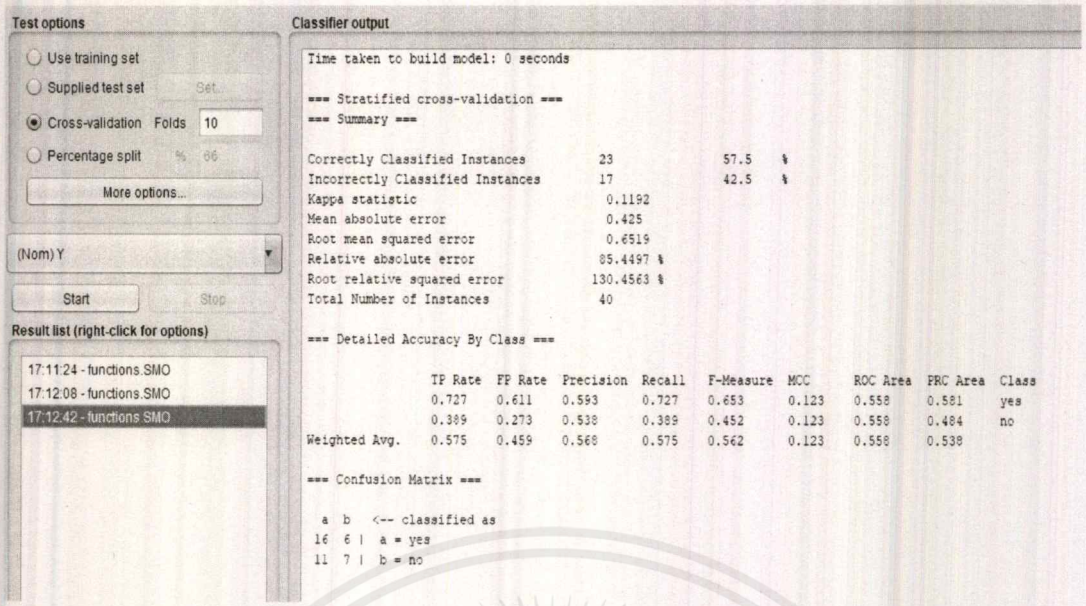


รูปที่ 4.187 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10



รูปที่ 4.188 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.189 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO ยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 30

จากรูปที่ 4.187 ถึง 4.189 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยการปรับความไม่สมดุลด้วยวิธีการผสมผสาน (Hybrid Methods) ยอดคงเหลือในบัตรเครดิตของลูกค้า โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

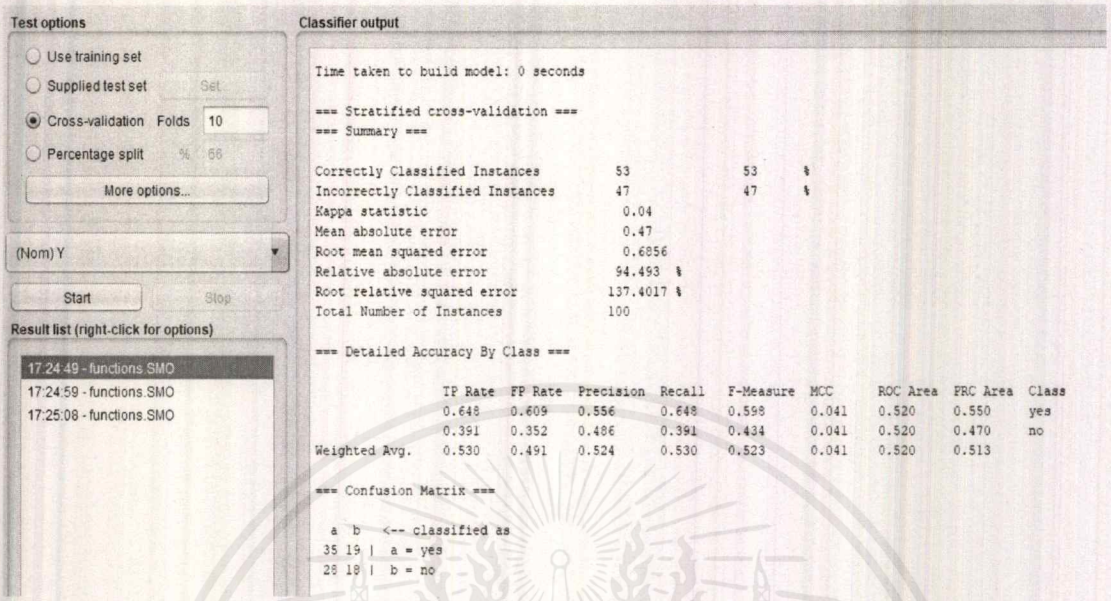
การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 40 คน มีค่าทำนายความถูกต้องจำนวน 21 คน คิดเป็นร้อยละ 52.5 โดยมีค่าความไว = 0.4211 ค่าความจำเพาะ = 0.6190 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6892)^2$ = 0.4750

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 40 คน มีค่าทำนายความถูกต้องจำนวน 25 คน คิดเป็นร้อยละ 62.5 โดยมีค่าความไว = 0.1429 ค่าความจำเพาะ = 0.8846 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6124)^2$ = 0.3750

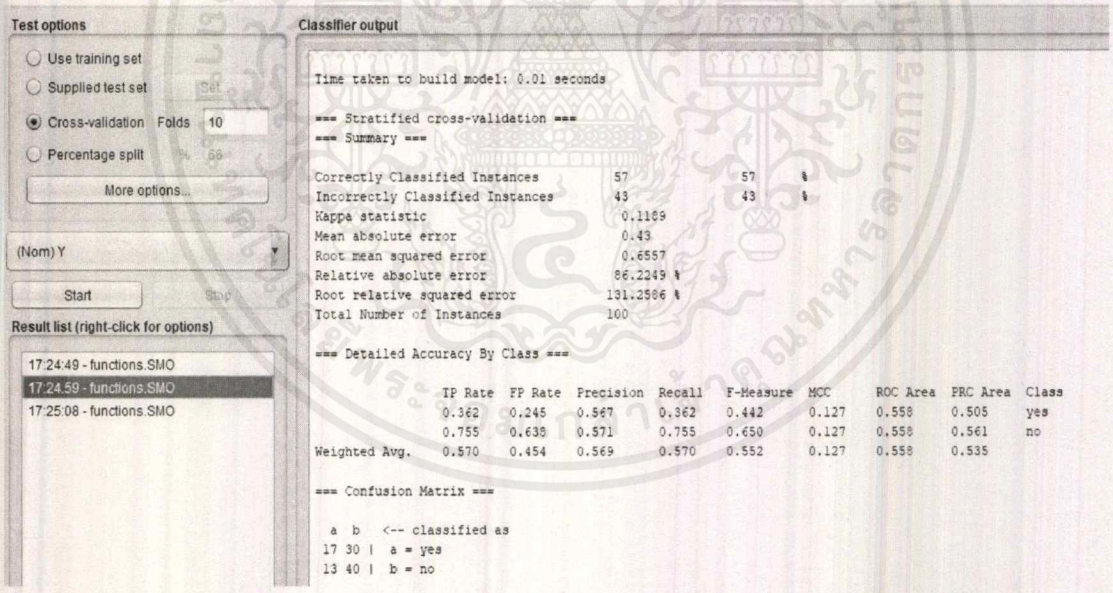
การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มลูกค้าจากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 40 คน มีค่าทำนายความถูกต้องจำนวน 23 คน คิดเป็นร้อยละ 57.5 โดยมีค่าความไว = 0.7273 ค่าความจำเพาะ = 0.3889 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6519)^2$ = 0.4250

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.5.4.3 คุณภาพไวน์แดง (Red Wine Quality) จำนวนข้อมูลทดสอบ 100 ค่า เมื่อกำหนดค่า Random Seed 10, 20 และ 30 ตามลำดับ

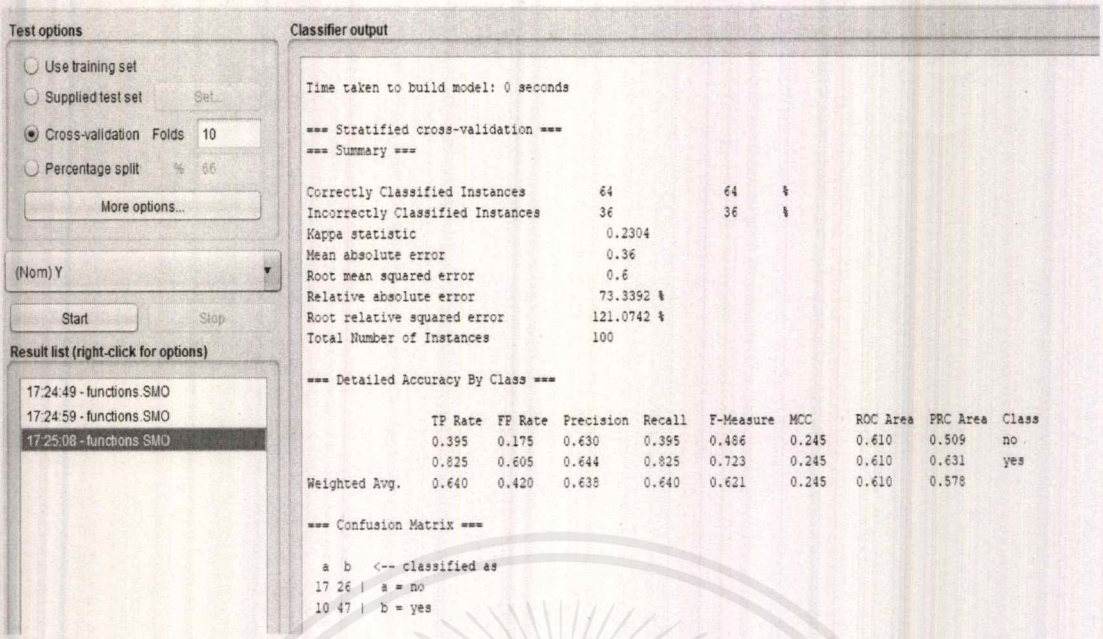


รูปที่ 4.190 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 10



รูปที่ 4.191 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.192 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ขั้นตอนวิธี SMO คุณภาพไวน์แดง เมื่อ Random Seed 30

จากรูปที่ 4.190 ถึง 4.192 ผลการวิเคราะห์ข้อมูลด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยการปรับความไม่สมดุลด้วยวิธีการผสมผสาน (Hybrid Methods) คุณภาพไวน์แดง โดยทำการสุ่ม 3 ครั้ง Random Seed 10, 20 และ 30 ตามลำดับ จากผลลัพธ์ที่ได้พบว่า

การสุ่มครั้งที่ 1 Random Seed = 10 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 100 ขวด มีค่าทำนายความถูกต้องจำนวน 53 ขวด คิดเป็นร้อยละ 53 โดยมีค่าความไว = 0.6481 ค่าความจำเพาะ = 0.3913 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6856)^2$ = 0.4700

การสุ่มครั้งที่ 2 Random Seed = 20 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 100 ขวด มีค่าทำนายความถูกต้องจำนวน 57 ขวด คิดเป็นร้อยละ 57 โดยมีค่าความไว = 0.3617 ค่าความจำเพาะ = 0.5747 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6557)^2$ = 0.4299

การสุ่มครั้งที่ 3 Random Seed = 30 กลุ่มไวน์จากชุดข้อมูลทดสอบร้อยละ 10 จำนวนทั้งหมด 100 ขวด มีค่าทำนายความถูกต้องจำนวน 64 ขวด คิดเป็นร้อยละ 64 โดยมีค่าความไว = 0.3953 ค่าความจำเพาะ = 0.8246 และค่าคลาดเคลื่อนกำลังสองเฉลี่ย = $(0.6000)^2$ = 0.3600

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.6 ผลการเปรียบเทียบประสิทธิภาพวิธีการจำแนกทั้ง 4 วิธี และวิธีการปรับความไม่สมดุลทั้ง 4 วิธี

4.6.1 ผลการเปรียบเทียบประสิทธิภาพของวิธีการปรับความไม่สมดุลของข้อมูลทั้ง 4 วิธี โดยการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว วิธีต้นไม้ตัดสินใจ วิธีโครงข่ายประสาทเทียม และวิธีซัพพอร์ตเวกเตอร์แมชชีน

4.6.1.1 ชุดข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก

ตารางที่ 4.2 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยวิธีการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน เมื่อ Random Seed 10, 20 และ 30

วิธีการปรับความไม่สมดุลของข้อมูล	วิธีการจำแนก	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
วิธีการสุ่มเกิน	วิธีเพื่อนบ้านใกล้สุด k ตัว	**76.9697**	**0.8919**	**0.6703**	0.2229
	วิธีต้นไม้ตัดสินใจ	66.6667	0.6757	0.6593	0.2395
	วิธีโครงข่ายประสาทเทียม	64.8485	0.8649	0.4725	**0.2228**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	64.8485	0.7838	0.5385	0.3515
Random Seed 20					
วิธีการสุ่มเกิน	วิธีเพื่อนบ้านใกล้สุด k ตัว	**74.5455**	0.7303	**0.7632**	**0.1887**
	วิธีต้นไม้ตัดสินใจ	67.8788	0.6629	0.6974	0.2410
	วิธีโครงข่ายประสาทเทียม	61.8182	0.5843	0.6579	0.2430
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	67.2727	**0.8202**	0.5000	0.3273
Random Seed 30					
วิธีการสุ่มเกิน	วิธีเพื่อนบ้านใกล้สุด k ตัว	**80.6061**	**0.7582**	0.8649	**0.1676**
	วิธีต้นไม้ตัดสินใจ	73.9394	0.7143	0.7703	0.2008
	วิธีโครงข่ายประสาทเทียม	72.1212	0.5604	**0.9189**	0.1934
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	58.7879	0.7253	0.4189	0.4122
ค่าเฉลี่ย					
วิธีการสุ่มเกิน	วิธีเพื่อนบ้านใกล้สุด k ตัว	**77.3738**	**0.7935**	**0.7661**	**0.1931**
	วิธีต้นไม้ตัดสินใจ	69.4950	0.6843	0.7090	0.2271
	วิธีโครงข่ายประสาทเทียม	66.2626	0.6699	0.6831	0.2197
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	63.6364	0.7764	0.4858	0.3637

จากตารางที่ 4.2 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว ให้ค่าความถูกต้องสูงสุดคือร้อยละ 77.3738 ให้ค่าความไวสูงสุดคือ 0.7935 ให้ค่าความจำเพาะสูงสุดคือ 0.7661 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.1931

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.3 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยวิธีการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE เมื่อ Random Seed 10, 20 และ 30

วิธีการปรับความไม่สมดุลของข้อมูล	วิธีการจำแนก	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
วิธีการสุ่มเกินเทคนิค SMOTE	วิธีเพื่อนบ้านใกล้สุด k ตัว	86.3905	0.8651	0.8625	0.1316
	วิธีต้นไม้ตัดสินใจ	**88.1657**	**0.8989**	0.8625	**0.1046**
	วิธีโครงข่ายประสาทเทียม	84.6154	0.7303	0.9750	0.1223
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	83.432	0.6966	**0.9875**	0.1656
Random Seed 20					
วิธีการสุ่มเกินเทคนิค SMOTE	วิธีเพื่อนบ้านใกล้สุด k ตัว	**86.3905**	**0.8765**	0.8488	0.1322
	วิธีต้นไม้ตัดสินใจ	79.2899	0.7470	0.8372	0.1427
	วิธีโครงข่ายประสาทเทียม	85.2071	0.7229	**0.9767**	**0.1176**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	83.432	0.6867	**0.9767**	0.1656
Random Seed 30					
วิธีการสุ่มเกินเทคนิค SMOTE	วิธีเพื่อนบ้านใกล้สุด k ตัว	80.4734	**0.8023**	0.8072	0.1930
	วิธีต้นไม้ตัดสินใจ	82.8402	0.7674	0.8916	**0.1220**
	วิธีโครงข่ายประสาทเทียม	81.6568	0.7442	0.8916	0.1422
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	**85.7988**	0.7326	**0.9880**	0.1420
ค่าเฉลี่ย					
วิธีการสุ่มเกินเทคนิค SMOTE	วิธีเพื่อนบ้านใกล้สุด k ตัว	**84.4181**	**0.8480**	0.8395	0.1523
	วิธีต้นไม้ตัดสินใจ	83.4319	0.8044	0.8638	**0.1231**
	วิธีโครงข่ายประสาทเทียม	83.8264	0.7325	0.9478	0.1274
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	84.2209	0.7053	**0.9841**	0.1577

จากตารางที่ 4.3 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว ให้ค่าความถูกต้องสูงสุดคือร้อยละ 84.4181 และให้ค่าความไวสูงสุดคือ 0.8480 ส่วนวิธีซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความจำเพาะสูงสุดคือ 0.9841 และวิธีต้นไม้ตัดสินใจให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.1931

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.4 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยวิธีการปรับความไม่สมดุลด้วยวิธีการสุ่มลด เมื่อ Random Seed 10, 20 และ 30

วิธีการปรับความไม่สมดุลของข้อมูล	วิธีการจำแนก	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
วิธีการสุ่มลด	วิธีเพื่อนบ้านใกล้สุด k ตัว	80	**0.8571**	0.6667	0.1719
	วิธีต้นไม้ตัดสินใจ	50	0.7143	0	0.4605
	วิธีโครงข่ายประสาทเทียม	**90**	**0.8571**	**1**	**0.0991**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	**90**	0.8570	**1**	0.1000
Random Seed 20					
วิธีการสุ่มลด	วิธีเพื่อนบ้านใกล้สุด k ตัว	50	**0.6000**	0.4000	0.4173
	วิธีต้นไม้ตัดสินใจ	50	**0.6000**	0.4000	0.4028
	วิธีโครงข่ายประสาทเทียม	**80**	**0.6000**	**1**	**0.1944**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	50	**0.6000**	0.4000	0.4999
Random Seed 30					
วิธีการสุ่มลด	วิธีเพื่อนบ้านใกล้สุด k ตัว	70	0.2500	**1**	0.2537
	วิธีต้นไม้ตัดสินใจ	30	0	0.5000	0.5236
	วิธีโครงข่ายประสาทเทียม	**80**	**0.5000**	**1**	**0.1987**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	70	**0.5000**	0.8333	0.3000
ค่าเฉลี่ย					
วิธีการสุ่มลด	วิธีเพื่อนบ้านใกล้สุด k ตัว	66.6667	0.5690	0.6889	0.2810
	วิธีต้นไม้ตัดสินใจ	43.3333	0.4381	0.3000	0.4623
	วิธีโครงข่ายประสาทเทียม	**83.3333**	**0.6524**	**1**	**0.1641**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	70.0000	0.6523	0.7444	0.3000

จากตารางที่ 4.4 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการจำแนกด้วยวิธีโครงข่ายประสาทเทียม ให้ค่าความถูกต้องสูงสุดคือร้อยละ 83.3333 ให้ค่าความไวสูงสุดคือ 0.6524 ให้ค่าความจำเพาะสูงสุดคือ 1 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.1641

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยวิธีการปรับความไม่สมดุลด้วยวิธีการผสมผสาน เมื่อ Random Seed 10, 20 และ 30

วิธีการปรับความไม่สมดุลของข้อมูล	วิธีการจำแนก	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
วิธีการผสมผสาน	วิธีเพื่อนบ้านใกล้สุด k ตัว	**70.7865**	0.6522	0.7674	0.2803
	วิธีต้นไม้ตัดสินใจ	65.1685	0.6522	0.6511	0.2606
	วิธีโครงข่ายประสาทเทียม	66.2921	0.5217	**0.8140**	**0.2272**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	59.5506	**0.7609**	0.4186	0.4045
Random Seed 20					
วิธีการผสมผสาน	วิธีเพื่อนบ้านใกล้สุด k ตัว	68.5393	**0.7872**	0.5714	2
	วิธีต้นไม้ตัดสินใจ	**73.0337**	0.7021	**0.7619**	**0.2124**
	วิธีโครงข่ายประสาทเทียม	60.6742	0.6170	0.5952	0.2642
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	60.6742	0.6429	0.6271	0.3933
Random Seed 30					
วิธีการผสมผสาน	วิธีเพื่อนบ้านใกล้สุด k ตัว	68.5393	0.7143	0.6500	0.2855
	วิธีต้นไม้ตัดสินใจ	**78.6517**	**0.8163**	**0.7500**	**0.1997**
	วิธีโครงข่ายประสาทเทียม	59.5506	0.7551	0.4000	0.2698
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	65.1685	0.7755	0.5000	0.3483
ค่าเฉลี่ย					
วิธีการผสมผสาน	วิธีเพื่อนบ้านใกล้สุด k ตัว	69.2884	0.7179	0.6629	0.2823
	วิธีต้นไม้ตัดสินใจ	**72.2846**	0.7235	**0.7210**	**0.2242**
	วิธีโครงข่ายประสาทเทียม	62.1723	0.6313	0.6031	0.2537
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	61.7978	**0.7264**	0.5152	0.3820

จากตารางที่ 4.5 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกโดยการปรับความไม่สมดุลด้วยวิธีการวิธีการผสมผสานเมื่อ Random Seed 10, 20 และ 30 จะพบว่าการจำแนกด้วยวิธีต้นไม้ตัดสินใจ ให้ค่าความถูกต้องสูงสุดคือร้อยละ 72.2846 ให้ค่าความจำเพาะสูงสุดคือ 0.7210 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.2242 ส่วนวิธีซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความไวสูงสุดคือ 0.7264

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.6.1.2 ข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้า

ตารางที่ 4.6 ผลการวิเคราะห์ข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้าโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน เมื่อ Random Seed 10, 20 และ 30

วิธีการปรับความไม่สมดุลของข้อมูล	วิธีการจำแนก	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
วิธีการสุ่มเกิน	วิธีเพื่อนบ้านใกล้สุด k ตัว	73.2394	0.7778	0.6856	0.2604
	วิธีต้นไม้ตัดสินใจ	**78.8732**	**0.8056**	**0.7714**	**0.1860**
	วิธีโครงข่ายประสาทเทียม	66.1972	0.6111	0.7143	0.2735
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	63.3803	0.5000	**0.7714**	0.3661
Random Seed 20					
วิธีการสุ่มเกิน	วิธีเพื่อนบ้านใกล้สุด k ตัว	**71.8310**	**0.8333**	0.6342	0.2757
	วิธีต้นไม้ตัดสินใจ	67.6056	0.7333	0.6341	0.2910
	วิธีโครงข่ายประสาทเทียม	66.1972	0.7000	0.6341	**0.2536**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	60.5634	0.4000	**0.7561**	0.3944
Random Seed 30					
วิธีการสุ่มเกิน	วิธีเพื่อนบ้านใกล้สุด k ตัว	70.4225	**0.8372**	0.5000	0.2875
	วิธีต้นไม้ตัดสินใจ	**71.831**	0.8140	**0.5714**	0.2572
	วิธีโครงข่ายประสาทเทียม	69.0141	0.7674	**0.5714**	**0.2428**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	66.1927	0.8140	0.4286	0.3380
ค่าเฉลี่ย					
วิธีการสุ่มเกิน	วิธีเพื่อนบ้านใกล้สุด k ตัว	71.8310	**0.8161**	0.6066	0.2745
	วิธีต้นไม้ตัดสินใจ	**72.7699**	0.7843	**0.6590**	**0.2447**
	วิธีโครงข่ายประสาทเทียม	67.1362	0.6928	0.6399	0.2566
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	63.3788	0.5713	0.6520	0.3662

จากตารางที่ 4.6 ผลการวิเคราะห์ข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้าโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการจำแนกด้วยวิธีต้นไม้ตัดสินใจ ให้ค่าความถูกต้องสูงสุดคือร้อยละ 72.7699 ให้ค่าความจำเพาะสูงสุดคือ 0.6590 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.2447 ส่วนวิธีเพื่อนบ้านใกล้สุด k ตัว ให้ค่าความไวสูงสุดคือ 0.8161

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 ผลการวิเคราะห์ข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้าโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE เมื่อ Random Seed 10, 20 และ 30

วิธีการปรับความไม่สมดุลของข้อมูล	วิธีการจำแนก	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
วิธีการสุ่มเกินเทคนิค SMOTE	วิธีเพื่อนบ้านใกล้สุด k ตัว	**76.3889**	**0.6774**	0.8293	**0.2292**
	วิธีต้นไม้ตัดสินใจ	69.4444	0.5484	0.8049	0.2735
	วิธีโครงข่ายประสาทเทียม	72.2222	0.5806	0.8293	0.2408
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	70.8333	0.5161	**0.8537**	0.2917
Random Seed 20					
วิธีการสุ่มเกินเทคนิค SMOTE	วิธีเพื่อนบ้านใกล้สุด k ตัว	70.8333	0.7838	0.6286	0.2831
	วิธีต้นไม้ตัดสินใจ	72.2222	0.7838	0.6571	0.2375
	วิธีโครงข่ายประสาทเทียม	68.0556	0.7568	**0.6000**	0.2748
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	**76.3889**	**0.8649**	0.6571	**0.2361**
Random Seed 30					
วิธีการสุ่มเกินเทคนิค SMOTE	วิธีเพื่อนบ้านใกล้สุด k ตัว	**83.3333**	0.9231	0.7273	**0.1627**
	วิธีต้นไม้ตัดสินใจ	76.3889	0.7865	**0.7353**	0.2175
	วิธีโครงข่ายประสาทเทียม	77.7778	0.8684	0.6765	0.1920
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	81.9444	**0.9474**	0.6765	0.1805
ค่าเฉลี่ย					
วิธีการสุ่มเกินเทคนิค SMOTE	วิธีเพื่อนบ้านใกล้สุด k ตัว	**76.8518**	**0.7948**	0.7284	**0.2250**
	วิธีต้นไม้ตัดสินใจ	72.6852	0.7062	**0.7324**	0.2428
	วิธีโครงข่ายประสาทเทียม	72.6852	0.7353	0.7019	0.2359
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	76.3889	0.7761	0.7291	0.2361

จากตารางที่ 4.7 ผลการวิเคราะห์ข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้าโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว ให้ค่าความถูกต้องสูงสุดคือร้อยละ 76.5818 ให้ค่าความไวสูงสุดคือ 0.7948 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.2250 ส่วนวิธีต้นไม้ตัดสินใจ ให้ค่าความจำเพาะสูงสุดคือ 0.7324

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.8 ผลการวิเคราะห์ข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้าโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด เมื่อ Random Seed 10, 20 และ 30

วิธีการปรับความไม่สมดุลของข้อมูล	วิธีการจำแนก	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
วิธีการสุ่มลด	วิธีเพื่อนบ้านใกล้สุด k ตัว	50	0.6000	0.4000	0.4173
	วิธีต้นไม้ตัดสินใจ	**60**	**0.8000**	**0.4000**	**0.3549**
	วิธีโครงข่ายประสาทเทียม	30	0.2000	0.4000	0.6374
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	50	0.6000	0.4000	0.4999
Random Seed 20					
วิธีการสุ่มลด	วิธีเพื่อนบ้านใกล้สุด k ตัว	**60**	0.7142	**0.3333**	0.3356
	วิธีต้นไม้ตัดสินใจ	50	0.7142	0	**0.3298**
	วิธีโครงข่ายประสาทเทียม	40	0.4286	**0.3333**	0.5196
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	**60**	**0.8571**	0	0.4001
Random Seed 30					
วิธีการสุ่มลด	วิธีเพื่อนบ้านใกล้สุด k ตัว	50	**0.5000**	0.5000	**0.4173**
	วิธีต้นไม้ตัดสินใจ	50	0.2500	**0.6667**	0.4445
	วิธีโครงข่ายประสาทเทียม	50	**0.5000**	0.5000	0.5054
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	30	0	0.5000	0.7000
ค่าเฉลี่ย					
วิธีการสุ่มลด	วิธีเพื่อนบ้านใกล้สุด k ตัว	**53.3333**	**0.6047**	**0.4111**	0.3901
	วิธีต้นไม้ตัดสินใจ	**53.3333**	0.5881	0.3556	**0.3764**
	วิธีโครงข่ายประสาทเทียม	40.0000	0.3762	**0.4111**	0.5541
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	46.6667	0.4857	0.3000	0.5333

จากตารางที่ 4.8 ผลการวิเคราะห์ข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้าโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว ให้ค่าความถูกต้องสูงสุดคือร้อยละ 53.3333 ให้ค่าความไวสูงสุดคือ 0.6047 และให้ค่าความจำเพาะสูงสุดคือ 0.7324 ส่วนวิธีต้นไม้ตัดสินใจ ให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.3764

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.9 ผลการวิเคราะห์ข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้าโดยการปรับความไม่สมดุลด้วยวิธีการผสมผสาน เมื่อ Random Seed 10, 20 และ 30

วิธีการปรับความไม่สมดุลของข้อมูล	วิธีการจำแนก	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
วิธีการผสมผสาน	เพื่อนบ้านใกล้สุด k ตัว	40	0.3158	0.4762	0.5691
	ต้นไม้ตัดสินใจ	42.5	0.4210	0.4286	0.4970
	โครงข่ายประสาทเทียม	50	**0.5789**	0.4286	**0.3750**
	ซัพพอร์ตเวกเตอร์แมชชีน	**52.5**	0.4211	**0.6190**	0.4750
Random Seed 20					
วิธีการผสมผสาน	เพื่อนบ้านใกล้สุด k ตัว	60	0.5714	0.6154	0.3808
	ต้นไม้ตัดสินใจ	75	**0.8462**	0.5714	**0.2183**
	โครงข่ายประสาทเทียม	57.5	0.4231	0.8571	0.2517
	ซัพพอร์ตเวกเตอร์แมชชีน	62.5	0.1429	**0.8846**	0.3750
Random Seed 30					
วิธีการผสมผสาน	เพื่อนบ้านใกล้สุด k ตัว	45	0.5909	0.2778	0.5236
	ต้นไม้ตัดสินใจ	42.5	0.5454	0.2778	0.5010
	โครงข่ายประสาทเทียม	52.5	0.6818	0.3333	**0.3448**
	ซัพพอร์ตเวกเตอร์แมชชีน	**57.5**	**0.7273**	**0.3889**	0.4250
ค่าเฉลี่ย					
วิธีการผสมผสาน	เพื่อนบ้านใกล้สุด k ตัว	48.3333	0.4927	0.4565	0.4912
	ต้นไม้ตัดสินใจ	53.3333	**0.6042**	0.4259	0.4054
	โครงข่ายประสาทเทียม	53.3333	0.5613	0.5397	**0.3238**
	ซัพพอร์ตเวกเตอร์แมชชีน	**57.5000**	0.4304	**0.6308**	0.4250

จากตารางที่ 4.9 ผลการวิเคราะห์ข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้าโดยการปรับความไม่สมดุลด้วยวิธีการผสมผสาน เมื่อ Random Seed 10 20 และ 30 จะพบว่าการจำแนกด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน ให้ค่าความถูกต้องสูงสุดคือร้อยละ 57.5000 และให้ค่าความจำเพาะสูงสุดคือ 0.6308 ส่วนวิธีต้นไม้ตัดสินใจ ให้ค่าความไวสูงสุดคือ 0.6042 และวิธีโครงข่ายประสาทเทียม ให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.3238

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.6.1.3 ชุดข้อมูลคุณภาพไวน์

ตารางที่ 4.10 ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยวิธีการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน เมื่อ Random Seed 10, 20 และ 30

วิธีการปรับความไม่สมดุลของข้อมูล	วิธีการจำแนก	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
วิธีการสุ่มเกิน	วิธีเพื่อนบ้านใกล้สุด k ตัว	53.75	**0.6250**	0.4500	0.4596
	วิธีต้นไม้ตัดสินใจ	53.75	0.5250	**0.5500**	0.3390
	วิธีโครงข่ายประสาทเทียม	**54.375**	0.5875	0.5000	**0.2924**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	47.5	0.4250	0.5316	0.5250
Random Seed 20					
วิธีการสุ่มเกิน	วิธีเพื่อนบ้านใกล้สุด k ตัว	**61.875**	**0.6400**	0.6000	0.3759
	วิธีต้นไม้ตัดสินใจ	50.625	0.3067	0.6824	0.3232
	วิธีโครงข่ายประสาทเทียม	56.25	0.2400	**0.8471**	**0.2814**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	51.875	0.2133	0.7882	0.4812
Random Seed 30					
วิธีการสุ่มเกิน	วิธีเพื่อนบ้านใกล้สุด k ตัว	**65.625**	0.7093	0.5946	0.3392
	วิธีต้นไม้ตัดสินใจ	60	0.6279	0.5676	0.3390
	วิธีโครงข่ายประสาทเทียม	58.125	0.5465	**0.6216**	**0.2834**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	56.875	**0.7791**	0.3243	0.4313
ค่าเฉลี่ย					
วิธีการสุ่มเกิน	วิธีเพื่อนบ้านใกล้สุด k ตัว	**60.4167**	**0.6581**	0.5482	0.3916
	วิธีต้นไม้ตัดสินใจ	54.7917	0.4865	0.6000	0.3337
	วิธีโครงข่ายประสาทเทียม	56.2500	0.4580	**0.6562**	**0.2857**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	52.0833	0.4725	0.5480	0.4792

จากตารางที่ 4.10 ผลการวิเคราะห์ข้อมูลคุณภาพไวน์โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว ให้ค่าความถูกต้องสูงสุดคือร้อยละ 60.4167 และให้ค่าความไวสูงสุดคือ 0.6581 ส่วนวิธีโครงข่ายประสาทเทียม ให้ค่าความจำเพาะสูงสุดคือ 0.6562 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.2857

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.11 ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยวิธีการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE เมื่อ Random Seed 10, 20 และ 30

วิธีการปรับความไม่สมดุลของข้อมูล	วิธีการจำแนก	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
วิธีการสุ่มเกินเทคนิค SMOTE	วิธีเพื่อนบ้านใกล้สุด k ตัว	62.2642	0.6625	**0.5822**	0.3722
	วิธีต้นไม้ตัดสินใจ	60.3774	0.6875	0.5190	0.3145
	วิธีโครงข่ายประสาทเทียม	59.7484	**0.7750**	0.4177	**0.2724**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	**62.8931**	0.725	0.5316	0.3711
Random Seed 20					
วิธีการสุ่มเกินเทคนิค SMOTE	วิธีเพื่อนบ้านใกล้สุด k ตัว	57.2327	0.5443	**0.6000**	0.4243
	วิธีต้นไม้ตัดสินใจ	56.6038	0.7848	0.3500	0.2739
	วิธีโครงข่ายประสาทเทียม	**58.4906**	**0.8228**	0.3500	**0.2678**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	52.2013	0.5950	0.4500	0.4780
Random Seed 30					
วิธีการสุ่มเกินเทคนิค SMOTE	วิธีเพื่อนบ้านใกล้สุด k ตัว	**49.6855**	0.4815	0.5128	0.4994
	วิธีต้นไม้ตัดสินใจ	47.7987	**0.7284**	0.2180	0.3169
	วิธีโครงข่ายประสาทเทียม	49.0566	0.2346	**0.7564**	**0.3010**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	49.0566	0.5926	0.3846	0.5094
ค่าเฉลี่ย					
วิธีการสุ่มเกินเทคนิค SMOTE	วิธีเพื่อนบ้านใกล้สุด k ตัว	**56.3941**	0.5628	**0.5650**	0.4320
	วิธีต้นไม้ตัดสินใจ	54.9266	**0.7336**	0.3623	0.3018
	วิธีโครงข่ายประสาทเทียม	55.7652	0.6108	0.5080	**0.2804**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	54.7170	0.6375	0.4554	0.4528

จากตารางที่ 4.11 ผลการวิเคราะห์ข้อมูลคุณภาพไวน์โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว ให้ค่าความถูกต้องสูงสุดคือร้อยละ 56.3941 และให้ค่าความจำเพาะสูงสุดคือ 0.5650 ส่วนวิธีต้นไม้ตัดสินใจ ให้ค่าความไวสูงสุดคือ 0.7336 และวิธีโครงข่ายประสาทเทียม ให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.2804

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.12 ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยวิธีการปรับความไม่สมดุลด้วยวิธีการสุ่มลด เมื่อ Random Seed 10, 20 และ 30

วิธีการปรับความไม่สมดุลของข้อมูล	วิธีการจำแนก	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
วิธีการสุ่มลด	วิธีเพื่อนบ้านใกล้สุด k ตัว	56.4103	0.3846	0.6538	0.4132
	วิธีต้นไม้ตัดสินใจ	46.1538	0.3077	0.5384	0.4718
	วิธีโครงข่ายประสาทเทียม	56.4103	**0.5385**	0.5769	0.3716
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	**66.6667**	0.0769	**0.9615**	**0.3334**
Random Seed 20					
วิธีการสุ่มลด	วิธีเพื่อนบ้านใกล้สุด k ตัว	58.9744	0.7083	**0.4000**	0.3895
	วิธีต้นไม้ตัดสินใจ	51.2821	0.6250	0.3333	0.4196
	วิธีโครงข่ายประสาทเทียม	61.5385	0.7500	**0.4000**	**0.3328**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	**66.6667**	**0.8333**	**0.4000**	0.3334
Random Seed 30					
วิธีการสุ่มลด	วิธีเพื่อนบ้านใกล้สุด k ตัว	71.7949	**0.7727**	0.6471	0.2675
	วิธีต้นไม้ตัดสินใจ	61.5385	**0.7727**	0.4118	0.3446
	วิธีโครงข่ายประสาทเทียม	69.2308	**0.7727**	0.5882	0.2929
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	**74.3590**	**0.7727**	**0.7059**	**0.2564**
ค่าเฉลี่ย					
วิธีการสุ่มลด	วิธีเพื่อนบ้านใกล้สุด k ตัว	62.3932	0.6219	0.5670	0.3567
	วิธีต้นไม้ตัดสินใจ	52.9915	0.5685	0.4278	0.4120
	วิธีโครงข่ายประสาทเทียม	62.3932	**0.6871**	0.5217	0.3324
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	**69.2308**	0.5610	**0.6891**	**0.3077**

จากตารางที่ 4.12 ผลการวิเคราะห์ข้อมูลคุณภาพไวน์โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการจำแนกด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความถูกต้องสูงสุดคือร้อยละ 69.2308 ให้ค่าความจำเพาะสูงสุดคือ 0.6891 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.3077 ส่วนวิธีโครงข่ายประสาทเทียม ให้ค่าความไวสูงสุดคือ 0.6871

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.13 ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยวิธีการปรับความไม่สมดุลด้วยวิธีการผสมผสาน เมื่อ Random Seed 10, 20 และ 30

วิธีการปรับความไม่สมดุลของข้อมูล	วิธีการจำแนก	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
วิธีการผสมผสาน	วิธีเพื่อนบ้านใกล้สุด k ตัว	56	0.5370	**0.5870**	0.4306
	วิธีต้นไม้ตัดสินใจ	**57**	0.5926	0.5435	0.3741
	วิธีโครงข่ายประสาทเทียม	52	0.5185	0.5217	**0.3191**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	53	**0.6481**	0.3913	0.4700
Random Seed 20					
วิธีการผสมผสาน	วิธีเพื่อนบ้านใกล้สุด k ตัว	62	**0.5957**	0.6415	0.3725
	วิธีต้นไม้ตัดสินใจ	54	0.4681	0.6038	0.2916
	วิธีโครงข่ายประสาทเทียม	70	0.4680	**0.9057**	**0.2192**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	57	0.3617	0.5747	0.4299
Random Seed 30					
วิธีการผสมผสาน	วิธีเพื่อนบ้านใกล้สุด k ตัว	**67**	**0.5349**	0.7719	0.3260
	วิธีต้นไม้ตัดสินใจ	65	0.5116	0.7544	0.3190
	วิธีโครงข่ายประสาทเทียม	52	**0.5349**	0.5087	**0.3185**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	64	0.3953	**0.8246**	0.3600
ค่าเฉลี่ย					
วิธีการผสมผสาน	วิธีเพื่อนบ้านใกล้สุด k ตัว	**61.6667**	**0.5559**	**0.6668**	0.3764
	วิธีต้นไม้ตัดสินใจ	58.6667	0.5241	0.6339	0.3282
	วิธีโครงข่ายประสาทเทียม	58.0000	0.5071	0.6454	**0.2856**
	วิธีซัพพอร์ตเวกเตอร์แมชชีน	58.0000	0.4684	0.5969	0.4200

จาตารางที่ 4.13 ผลการวิเคราะห์ข้อมูลคุณภาพไวน์โดยการปรับความไม่สมดุลด้วยวิธีการผสมผสานเมื่อ Random Seed 10, 20 และ 30 จะพบว่าการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว ให้ค่าความถูกต้องสูงสุดคือร้อยละ 61.6667 ให้ค่าความไวสูงสุดคือ 0.5559 และให้ค่าความจำเพาะสูงสุดคือ 0.6668 ส่วนวิธีโครงข่ายประสาทเทียม ให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.2856

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.6.2 ผลการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกทั้ง 4 วิธี โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกิน วิธีการสุ่มเกินโดยเทคนิค SMOTE วิธีการสุ่มลด และวิธีการผสมผสาน

4.6.2.1 ชุดข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก

ตารางที่ 4.14 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว เมื่อ Random Seed 10, 20 และ 30

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	76.9697	**0.8919**	0.6703	0.2229
	วิธีการสุ่มเกินเทคนิค SMOTE	**86.3905**	0.8651	**0.8625**	**0.1316**
	วิธีการสุ่มลด	50	0.6000	0.4000	0.4173
	วิธีการผสมผสาน	70.7865	0.6522	0.7674	0.2803
Random Seed 20					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	74.5455	0.7303	0.7632	0.1887
	วิธีการสุ่มเกินเทคนิค SMOTE	**86.3905**	**0.8795**	**0.8488**	**0.1322**
	วิธีการสุ่มลด	80	0.8571	0.6667	0.1719
	วิธีการผสมผสาน	68.5393	0.7872	0.5714	0.2811
Random Seed 30					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	**80.6061**	0.7582	0.8649	**0.1676**
	วิธีการสุ่มเกินเทคนิค SMOTE	80.4734	**0.8023**	0.8072	0.1930
	วิธีการสุ่มลด	70	0.2500	**1**	0.2537
	วิธีการผสมผสาน	68.5393	0.7143	0.6500	0.2855
ค่าเฉลี่ย					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	77.3737	0.7934	0.7661	0.1930
	วิธีการสุ่มเกินเทคนิค SMOTE	**84.4181**	**0.8489**	**0.8395**	**0.1522**
	วิธีการสุ่มลด	66.6666	0.5690	0.6889	0.2809
	วิธีการผสมผสาน	69.2883	0.7179	0.6629	0.2823

จากตารางที่ 4.14 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกโดยการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความถูกต้องสูงสุดคือร้อยละ 84.4181 ให้ค่าความไวสูงสุดคือ 0.8489 ให้ค่าความจำเพาะสูงสุดคือ 0.8395 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.1522

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.15 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยการจำแนกด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) เมื่อ Random Seed 10, 20 และ 30

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	66.6667	0.6757	0.6593	0.2395
	วิธีการสุ่มเกินเทคนิค SMOTE	**88.1657**	**0.8989**	**0.8625**	**0.1046**
	วิธีการสุ่มลด	50	0.6000	0.4000	0.4028
	วิธีการผสมผสาน	65.1685	0.6522	0.6511	0.2606
Random Seed 20					
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	67.8788	0.6629	0.6974	0.2410
	วิธีการสุ่มเกินเทคนิค SMOTE	**79.2899**	**0.7470**	**0.8372**	**0.1427**
	วิธีการสุ่มลด	50	0.7143	0	0.4605
	วิธีการผสมผสาน	73.0337	0.7021	0.7619	0.2124
Random Seed 30					
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	73.9394	0.7143	0.7703	0.2008
	วิธีการสุ่มเกินเทคนิค SMOTE	**82.8402**	0.7674	**0.8916**	**0.1220**
	วิธีการสุ่มลด	30	0	0.5000	0.5236
	วิธีการผสมผสาน	78.6517	**0.8163**	0.7500	0.1997
ค่าเฉลี่ย					
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	69.4949	0.6843	0.7090	0.2271
	วิธีการสุ่มเกินเทคนิค SMOTE	**83.4319**	**0.8044**	**0.8637**	**0.1231**
	วิธีการสุ่มลด	43.3333	0.4381	0.3000	0.4623
	วิธีการผสมผสาน	72.2846	0.7235	0.7210	0.2242

จากตารางที่ 4.15 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกโดยการจำแนกด้วยวิธีต้นไม้ตัดสินใจ เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความถูกต้องสูงสุดคือร้อยละ 83.4319 ให้ค่าความไวสูงสุดคือ 0.8044 ให้ค่าความจำเพาะสูงสุดคือ 0.8637 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.1231

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.16 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยการจำแนกด้วยวิธีโครงข่ายประสาทเทียม เมื่อ Random Seed 10, 20 และ 30

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	64.8485	**0.8649**	0.4725	0.2228
	วิธีการสุ่มเกินเทคนิค SMOTE	**84.6154**	0.7303	0.9750	**0.1223**
	วิธีการสุ่มลด	80	0.6000	**1**	0.1944
	วิธีการผสมผสาน	66.2921	0.5217	0.8140	0.2272
Random Seed 20					
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	61.8182	0.5843	0.6579	0.2430
	วิธีการสุ่มเกินเทคนิค SMOTE	85.2071	0.7229	0.9767	0.1176
	วิธีการสุ่มลด	**90**	**0.8571**	1	**0.0991**
	วิธีการผสมผสาน	60.6742	0.6170	0.5952	0.2642
Random Seed 30					
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	72.1212	0.5604	0.9189	0.1934
	วิธีการสุ่มเกินเทคนิค SMOTE	**81.6568**	0.7442	0.8916	**0.1422**
	วิธีการสุ่มลด	80	0.5000	**1**	0.1987
	วิธีการผสมผสาน	59.5506	**0.7551**	0.4000	0.2698
ค่าเฉลี่ย					
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	66.2626	0.6698	0.6831	0.2197
	วิธีการสุ่มเกินเทคนิค SMOTE	**83.8264**	**0.7324**	0.9477	**0.1273**
	วิธีการสุ่มลด	83.3333	0.6523	1	0.1640
	วิธีการผสมผสาน	62.1723	0.6312	0.603	0.2537

จากตารางที่ 4.16 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกโดยการจำแนกด้วยวิธีโครงข่ายประสาทเทียม เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความถูกต้องสูงสุดคือร้อยละ 83.8264 ให้ค่าความไวสูงสุดคือ 0.7324 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.1273 ส่วนการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลด ให้ค่าความจำเพาะสูงสุดคือ 0.9477

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.17 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก โดยการจำแนกด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน เมื่อ Random Seed 10, 20 และ 30

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	64.8485	**0.7838**	0.5385	0.3515
	วิธีการสุ่มเกินเทคนิค SMOTE	**83.432**	0.6966	**0.9875**	**0.1656**
	วิธีการสุ่มลด	50	0.6000	0.4000	0.4999
	วิธีการผสมผสาน	59.5506	0.7609	0.4186	0.4045
Random Seed 20					
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	67.2727	0.8202	0.5000	0.3273
	วิธีการสุ่มเกินเทคนิค SMOTE	83.432	0.6867	0.9767	0.1656
	วิธีการสุ่มลด	**90**	**0.8570**	**1**	**0.1000**
	วิธีการผสมผสาน	60.6742	0.6429	0.6271	0.3933
Random Seed 30					
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	58.7879	0.7253	0.4189	0.4122
	วิธีการสุ่มเกินเทคนิค SMOTE	**85.7988**	0.7326	**0.9880**	**0.1420**
	วิธีการสุ่มลด	70	0.5000	0.8333	0.3000
	วิธีการผสมผสาน	65.1685	**0.7755**	0.5000	0.3483
ค่าเฉลี่ย					
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	63.6363	**0.7764**	0.4858	0.3636
	วิธีการสุ่มเกินเทคนิค SMOTE	**84.2209**	0.7053	**0.9840**	**0.1577**
	วิธีการสุ่มลด	70	0.6523	0.7444	0.2999
	วิธีการผสมผสาน	61.7977	0.7264	0.5152	0.3820

จากตารางที่ 4.17 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกโดยการจำแนกด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความถูกต้องสูงสุดคือร้อยละ 84.2209 ให้ค่าความจำเพาะสูงสุดคือ 0.9840 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.1577 ส่วนการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกิน ให้ค่าความไวสูงสุดคือ 0.7764

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.6.2.2 ชุดข้อมูลยอดเงินคงเหลือในบัตรเครดิตของลูกค้า

ตารางที่ 4.18 ผลการวิเคราะห์ข้อมูลยอดเงินคงเหลือในบัตรเครดิตของลูกค้าโดยการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว เมื่อ Random Seed 10, 20 และ 30

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	73.2394	**0.7778**	0.6856	0.2604
	วิธีการสุ่มเกินเทคนิค SMOTE	**76.3889**	0.6774	**0.8293**	**0.2292**
	วิธีการสุ่มลด	50	0.6000	0.4000	0.4173
	วิธีการผสมผสาน	40	0.3158	0.4762	0.5691
Random Seed 20					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	**71.8310**	**0.8333**	**0.6342**	**0.2757**
	วิธีการสุ่มเกินเทคนิค SMOTE	70.8333	0.7838	0.6286	0.2831
	วิธีการสุ่มลด	60	0.7142	0.3333	0.3356
	วิธีการผสมผสาน	60	0.5714	0.6154	0.3808
Random Seed 30					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	70.4225	0.8372	0.5000	0.2875
	วิธีการสุ่มเกินเทคนิค SMOTE	**83.3333**	**0.9231**	**0.7273**	**0.1627**
	วิธีการสุ่มลด	50	0.5000	0.5000	0.4173
	วิธีการผสมผสาน	45	0.5909	0.2778	0.5236
ค่าเฉลี่ย					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	71.8309	**0.8161**	0.6066	0.2745
	วิธีการสุ่มเกินเทคนิค SMOTE	**76.8518**	0.7947	**0.7284**	**0.225**
	วิธีการสุ่มลด	53.3333	0.6047	0.4111	0.3900
	วิธีการผสมผสาน	48.3333	0.4927	0.4564	0.4911

จากตารางที่ 4.18 ผลการวิเคราะห์ข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้าโดยการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความถูกต้องสูงสุดคือร้อยละ 76.8518 ให้ค่าความจำเพาะสูงสุดคือ 0.7284 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.225 ส่วนวิธีการสุ่มเกิน ให้ค่าความไวสูงสุดคือ 0.8161

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.19 ผลการวิเคราะห์ข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้าโดยการจำแนกด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) เมื่อ Random Seed 10, 20 และ 30

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	**78.8732**	**0.8056**	0.7714	**0.1860**
	วิธีการสุ่มเกินเทคนิค SMOTE	69.4444	0.5484	**0.8049**	0.2735
	วิธีการสุ่มลด	60	0.8000	0.4000	0.3549
	วิธีการผสมผสาน	42.5	0.4210	0.4286	0.4970
Random Seed 20					
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	67.6056	0.7333	0.6341	0.2910
	วิธีการสุ่มเกินเทคนิค SMOTE	**72.2222**	0.7838	**0.6571**	0.2375
	วิธีการสุ่มลด	50	0.7142	0	0.3298
	วิธีการผสมผสาน	75	**0.8462**	0.5714	**0.2183**
Random Seed 30					
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	71.831	**0.8140**	0.5714	0.2572
	วิธีการสุ่มเกินเทคนิค SMOTE	**76.3889**	0.7865	**0.7353**	**0.2175**
	วิธีการสุ่มลด	50	0.2500	0.6667	0.4445
	วิธีการผสมผสาน	42.5	0.5454	0.2778	0.5010
ค่าเฉลี่ย					
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	**72.7699**	**0.7843**	0.6589	0.2447
	วิธีการสุ่มเกินเทคนิค SMOTE	72.6851	0.7062	**0.7324**	**0.2428**
	วิธีการสุ่มลด	53.3333	0.5880	0.355	0.3764
	วิธีการผสมผสาน	53.3333	0.6042	0.4259	0.4054

จากตารางที่ 4.19 ผลการวิเคราะห์ข้อมูลยอดเงินคงเหลือในบัตรเครดิตของลูกค้าโดยการจำแนกด้วยวิธีต้นไม้ตัดสินใจ เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกิน ให้ค่าความถูกต้องสูงสุดคือร้อยละ 72.7699 ให้ค่าความไวสูงสุดคือ 0.7843 ส่วนการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความจำเพาะสูงสุดคือ 0.7324 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.2428

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.20 ผลการวิเคราะห์ข้อมูลยอดเงินคงเหลือในบัตรเครดิตของลูกค้าโดยการจำแนกด้วยวิธีโครงข่ายประสาทเทียม เมื่อ Random Seed 10, 20 และ 30

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	66.1972	**0.6111**	0.7143	0.2735
	วิธีการสุ่มเกินเทคนิค SMOTE	**72.2222**	0.5806	**0.8293**	**0.2408**
	วิธีการสุ่มลด	30	0.2000	0.4000	0.6374
	วิธีการผสมผสาน	50	0.5789	0.4286	0.3750
Random Seed 20					
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	66.1972	0.7000	0.6341	0.2536
	วิธีการสุ่มเกินเทคนิค SMOTE	**68.0556**	**0.7568**	0.6000	0.2748
	วิธีการสุ่มลด	40	0.4286	0.3333	0.5196
	วิธีการผสมผสาน	57.5	0.4231	**0.8571**	**0.2517**
Random Seed 30					
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	69.0141	0.7674	0.5714	0.2428
	วิธีการสุ่มเกินเทคนิค SMOTE	**77.7778**	**0.8684**	**0.6765**	**0.1920**
	วิธีการสุ่มลด	50	0.5000	0.5000	0.5054
	วิธีการผสมผสาน	52.5	0.6818	0.3333	0.3448
ค่าเฉลี่ย					
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	67.1361	0.6928	0.6399	0.2566
	วิธีการสุ่มเกินเทคนิค SMOTE	**72.6852**	**0.7352**	**0.7019**	**0.2358**
	วิธีการสุ่มลด	40	0.3762	0.4111	0.5541
	วิธีการผสมผสาน	53.3333	0.5612	0.5396	0.3238

จากตารางที่ 4.20 ผลการวิเคราะห์ข้อมูลยอดเงินคงเหลือในบัตรเครดิตของลูกค้าโดยการจำแนกด้วยวิธีโครงข่ายประสาทเทียม เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความถูกต้องสูงสุดคือร้อยละ 72.6852 ให้ค่าความไวสูงสุดคือ 0,7352 ให้ค่าความจำเพาะสูงสุดคือ 0.7019 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.2358

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.21 ผลการวิเคราะห์ข้อมูลยอดเงินคงเหลือในบัตรเครดิตของลูกค้าโดยการจำแนกด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน เมื่อ Random Seed 10, 20 และ 30

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	63.3803	0.5000	0.7714	0.3661
	วิธีการสุ่มเกินเทคนิค SMOTE	**70.8333**	0.5161	**0.8537**	**0.2917**
	วิธีการสุ่มลด	50	**0.6000**	0.4000	0.4999
	วิธีการผสมผสาน	52.5	0.4211	0.6190	0.4750
Random Seed 20					
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	60.5634	0.4000	0.7561	0.3944
	วิธีการสุ่มเกินเทคนิค SMOTE	**76.3889**	**0.8649**	0.6571	**0.2361**
	วิธีการสุ่มลด	60	0.8571	0	0.4001
	วิธีการผสมผสาน	62.5	0.1429	**0.8846**	0.3750
Random Seed 30					
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	66.1927	0.8140	0.4286	0.3380
	วิธีการสุ่มเกินเทคนิค SMOTE	**81.9444**	**0.9474**	**0.6765**	**0.1805**
	วิธีการสุ่มลด	30	0	0.5000	0.7000
	วิธีการผสมผสาน	57.5	0.7273	0.3889	0.4250
ค่าเฉลี่ย					
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	63.3788	0.5713	0.6520	0.3661
	วิธีการสุ่มเกินเทคนิค SMOTE	**76.3888**	**0.7761**	**0.7291**	**0.2361**
	วิธีการสุ่มลด	46.6666	0.4857	0.3	0.5333
	วิธีการผสมผสาน	57.5	0.4304	0.6308	0.425

จากตารางที่ 4.21 ผลการวิเคราะห์ข้อมูลยอดเงินคงเหลือในบัตรเครดิตของลูกค้าโดยการจำแนกด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความถูกต้องสูงสุดคือร้อยละ 72.3888 ให้ค่าความไวสูงสุดคือ 0.7761 ให้ค่าความจำเพาะสูงสุดคือ 0.7291 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.2361

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.6.2.3 ชุดข้อมูลคุณภาพไวน์แดง

ตารางที่ 4.22 ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว เมื่อ Random Seed 10, 20 และ 30

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	53.75	0.6250	0.4500	0.4596
	วิธีการสุ่มเกินเทคนิค SMOTE	**62.2642**	**0.6625**	0.5822	**0.3722**
	วิธีการสุ่มลด	56.4103	0.3846	**0.6538**	0.4132
	วิธีการผสมผสาน	56	0.5370	0.5870	0.4306
Random Seed 20					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	61.875	0.6400	0.6000	0.3759
	วิธีการสุ่มเกินเทคนิค SMOTE	57.2327	0.5443	0.6000	0.4243
	วิธีการสุ่มลด	58.9744	**0.7083**	0.4000	0.3895
	วิธีการผสมผสาน	**62**	0.5957	**0.6415**	**0.3725**
Random Seed 30					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	65.625	0.7093	0.5946	0.3392
	วิธีการสุ่มเกินเทคนิค SMOTE	49.6855	**0.4815**	0.5128	0.4994
	วิธีการสุ่มลด	**71.7949**	0.7727	0.6471	**0.2675**
	วิธีการผสมผสาน	67	0.5349	**0.7719**	0.3260
ค่าเฉลี่ย					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	60.4166	**0.6581**	0.5482	0.3915
	วิธีการสุ่มเกินเทคนิค SMOTE	56.3941	0.5627	**0.565**	0.4319
	วิธีการสุ่มลด	**62.3932**	0.621	0.566	**0.3567**
	วิธีการผสมผสาน	61.6666	0.5558	0.6668	0.3763

จากตารางที่ 4.22 ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลด ให้ค่าความถูกต้องสูงสุดคือร้อยละ 62.3932 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.3567 ส่วนการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกิน ให้ค่าความไวสูงสุดคือ 0.6581 และการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความจำเพาะสูงสุดคือ 0.565

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.23 ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยการจำแนกด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) เมื่อ Random Seed 10, 20 และ 30

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	53.75	0.5250	0.5500	0.3390
	วิธีการสุ่มเกินเทคนิค SMOTE	**60.3774**	**0.6875**	**0.5190**	**0.3145**
	วิธีการสุ่มลด	46.1538	0.3077	0.5384	0.4718
	วิธีการผสมผสาน	57	0.5926	0.5435	0.3741
Random Seed 20					
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	50.625	0.3067	**0.6824**	0.3232
	วิธีการสุ่มเกินเทคนิค SMOTE	**56.6038**	**0.7848**	0.3500	**0.2739**
	วิธีการสุ่มลด	51.2821	0.6250	0.3333	0.4196
	วิธีการผสมผสาน	54	0.4681	0.6038	0.2916
Random Seed 30					
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	60	0.6279	0.5676	0.3390
	วิธีการสุ่มเกินเทคนิค SMOTE	47.7987	0.7284	0.2180	**0.3169**
	วิธีการสุ่มลด	61.5385	**0.7727**	0.4118	0.3446
	วิธีการผสมผสาน	**65**	0.5116	**0.7544**	0.3190
ค่าเฉลี่ย					
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	54.7916	0.4865	0.6	0.3337
	วิธีการสุ่มเกินเทคนิค SMOTE	54.9263	**0.7335**	0.3623	**0.3016**
	วิธีการสุ่มลด	52.9914	0.5684	0.4278	0.412
	วิธีการผสมผสาน	**58.6666**	0.5241	**0.6339**	0.3282

จากตารางที่ 4.23 ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยการจำแนกด้วยวิธีต้นไม้ตัดสินใจ เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการปรับความไม่สมดุลของข้อมูลด้วยวิธีการผสมผสาน ให้ค่าความถูกต้องสูงสุดคือร้อยละ 58.6666 และให้ค่าความจำเพาะสูงสุดคือ 0:6339 ส่วนการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความไวสูงสุดคือ 0.7335 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.3016

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.24 ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยการจำแนกด้วยวิธีโครงข่ายประสาทเทียม เมื่อ Random Seed 10, 20 และ 30

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	54.375	0.5875	0.5000	0.2924
	วิธีการสุ่มเกินเทคนิค SMOTE	**59.7484**	**0.7750**	0.4177	**0.2724**
	วิธีการสุ่มลด	56.4103	0.5385	**0.5769**	0.3716
	วิธีการผสมผสาน	52	0.5185	0.5217	0.3191
Random Seed 20					
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	56.25	0.2400	0.8471	0.2814
	วิธีการสุ่มเกินเทคนิค SMOTE	58.4906	**0.8228**	0.3500	0.2678
	วิธีการสุ่มลด	61.5385	0.7500	0.4000	0.3328
	วิธีการผสมผสาน	70	0.4680	**0.9057**	**0.2192**
Random Seed 30					
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	58.125	0.5465	0.6216	**0.2834**
	วิธีการสุ่มเกินเทคนิค SMOTE	49.0566	0.2346	**0.7564**	0.3010
	วิธีการสุ่มลด	**69.2308**	**0.7727**	0.5882	0.2929
	วิธีการผสมผสาน	52	0.5349	0.5087	0.3185
ค่าเฉลี่ย					
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	56.25	0.458	0.6562	0.2857
	วิธีการสุ่มเกินเทคนิค SMOTE	55.7652	0.6108	**0.5080**	**0.2804**
	วิธีการสุ่มลด	**62.3932**	**0.6870**	0.5217	0.3324
	วิธีการผสมผสาน	58	0.5071	0.6453	0.2856

จากตารางที่ 4.24 ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยการจำแนกด้วยวิธีโครงข่ายประสาทเทียม เมื่อ Random Seed 10, 20 และ 30 จะพบว่าการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลด ให้ค่าความถูกต้องสูงสุดคือร้อยละ 62.3932 และให้ค่าความไวสูงสุดคือ 0.6870 ส่วนการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความจำเพาะสูงสุดคือ 0.5080 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.2804

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.25 ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยการจำแนกด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน เมื่อ Random Seed 10, 20 และ 30

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	47.5	0.4250	0.5250	0.5250
	วิธีการสุ่มเกินเทคนิค SMOTE	62.8931	**0.725**	0.5316	0.3711
	วิธีการสุ่มลด	**66.6667**	0.0769	**0.9615**	**0.3334**
	วิธีการผสมผสาน	53	0.6481	0.3913	0.4700
Random Seed 20					
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	51.875	0.2133	**0.7882**	0.4812
	วิธีการสุ่มเกินเทคนิค SMOTE	52.2013	0.5950	0.4500	0.4780
	วิธีการสุ่มลด	**66.6667**	**0.8333**	0.4000	**0.3334**
	วิธีการผสมผสาน	57	0.3617	0.5747	0.4299
Random Seed 30					
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	56.875	**0.7791**	0.3243	0.4313
	วิธีการสุ่มเกินเทคนิค SMOTE	49.0566	0.5926	0.3846	0.5094
	วิธีการสุ่มลด	**74.3590**	0.7727	0.7059	**0.2564**
	วิธีการผสมผสาน	64	0.3953	**0.8246**	0.3600
ค่าเฉลี่ย					
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	52.0833	0.4724	0.5458	0.4791
	วิธีการสุ่มเกินเทคนิค SMOTE	54.717	**0.6375**	0.4554	0.4528
	วิธีการสุ่มลด	**69.2308**	0.5609	**0.6891**	**0.3077**
	วิธีการผสมผสาน	58	0.4683	0.5968	0.4199

จากตารางที่ 4.25 ผลการวิเคราะห์ข้อมูลคุณภาพไวน์แดงโดยการจำแนกด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน เมื่อ Random Seed 10, 20 และ 30 จะการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลด ให้ค่าความถูกต้องสูงสุดคือร้อยละ 69.2308 ให้ค่าความจำเพาะสูงสุดคือ 0.6891 และให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.3077 ส่วนการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความไวสูงสุดคือ 0.6375

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.6.3 ผลการเปรียบเทียบประสิทธิภาพในภาพรวมของวิธีการจำแนกทั้ง 4 วิธี และการปรับความไม่สมดุลด้วย 4 วิธี

4.6.3.1 ชุดข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก

ตารางที่ 4.26 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10, 20 และ 30

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	76.9697	0.8919	0.6703	0.2229
	วิธีการสุ่มเกินเทคนิค SMOTE	86.3905	0.8651	0.8625	0.1316
	วิธีการสุ่มลด	50	0.6000	0.4000	0.4173
	วิธีการผสมผสาน	70.7865	0.6522	0.7674	0.2803
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	66.6667	0.6757	0.6593	0.2395
	วิธีการสุ่มเกินเทคนิค SMOTE	**88.1657**	**0.8989**	0.8625	**0.1046**
	วิธีการสุ่มลด	50	0.6000	0.4000	0.4028
	วิธีการผสมผสาน	65.1685	0.6522	0.6511	0.2606
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	64.8485	0.8649	0.4725	0.2228
	วิธีการสุ่มเกินเทคนิค SMOTE	84.6154	0.7303	0.9750	0.1223
	วิธีการสุ่มลด	80	0.6000	**1**	0.1944
	วิธีการผสมผสาน	66.2921	0.5217	0.8140	0.2272
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	64.8485	0.7838	0.5385	0.3515
	วิธีการสุ่มเกินเทคนิค SMOTE	83.432	0.6966	0.9875	0.1656
	วิธีการสุ่มลด	50	0.6000	0.4000	0.4999
	วิธีการผสมผสาน	59.5506	0.7609	0.4186	0.4045
Random Seed 20					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	74.5455	0.7303	0.7632	0.1887
	วิธีการสุ่มเกินเทคนิค SMOTE	86.3905	**0.8795**	0.8488	0.1322
	วิธีการสุ่มลด	80	0.8571	0.6667	0.1719
	วิธีการผสมผสาน	68.5393	0.7872	0.5714	0.2811
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	67.8788	0.6629	0.6974	0.2410
	วิธีการสุ่มเกินเทคนิค SMOTE	79.2899	0.7470	0.8372	0.1427
	วิธีการสุ่มลด	50	0.7143	0	0.4605
	วิธีการผสมผสาน	73.0337	0.7021	0.7619	0.2124
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	61.8182	0.5843	0.6579	0.2430
	วิธีการสุ่มเกินเทคนิค SMOTE	85.2071	0.7229	0.9767	0.1176
	วิธีการสุ่มลด	**90**	0.8571	**1**	**0.0991**
	วิธีการผสมผสาน	60.6742	0.6170	0.5952	0.2642
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	67.2727	0.8202	0.5000	0.3273
	วิธีการสุ่มเกินเทคนิค SMOTE	83.432	0.6867	0.9767	0.1656
	วิธีการสุ่มลด	**90**	0.8570	**1**	0.1000
	วิธีการผสมผสาน	60.6742	0.6429	0.6271	0.3933

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.26 ผลการวิเคราะห์ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อ Random Seed 10, 20 และ 30 (ต่อ)

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 30					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	80.6061	0.7582	0.8649	0.1676
	วิธีการสุ่มเกินเทคนิค SMOTE	80.4734	0.8023	0.8072	0.1930
	วิธีการสุ่มลด	70	0.2500	1	0.2537
	วิธีการผสมผสาน	68.5393	0.7143	0.6500	0.2855
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	73.9394	0.7143	0.7703	0.2008
	วิธีการสุ่มเกินเทคนิค SMOTE	82.8402	0.7674	0.8916	0.1220
	วิธีการสุ่มลด	30	0	0.5000	0.5236
	วิธีการผสมผสาน	78.6517	0.8163	0.7500	0.1997
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	72.1212	0.5604	0.9189	0.1934
	วิธีการสุ่มเกินเทคนิค SMOTE	81.6568	0.7442	0.8916	0.1422
	วิธีการสุ่มลด	80	0.5000	1	0.1987
	วิธีการผสมผสาน	59.5506	0.7551	0.4000	0.2698
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	58.7879	0.7253	0.4189	0.4122
	วิธีการสุ่มเกินเทคนิค SMOTE	85.7988	0.7326	0.9880	0.1420
	วิธีการสุ่มลด	70	0.5000	0.8333	0.3000
	วิธีการผสมผสาน	65.1685	0.7755	0.5000	0.3483
ค่าเฉลี่ย					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	77.3738	0.7935	0.7661	0.1931
	วิธีการสุ่มเกินเทคนิค SMOTE	84.4181*	0.8490*	0.8395	0.1523
	วิธีการสุ่มลด	66.6667	0.5690	0.6889	0.2810
	วิธีการผสมผสาน	69.2884	0.7179	0.6629	0.2823
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	69.4950	0.6843	0.7090	0.2271
	วิธีการสุ่มเกินเทคนิค SMOTE	83.4319	0.8044	0.8638	0.1231*
	วิธีการสุ่มลด	43.3333	0.4381	0.3000	0.4623
	วิธีการผสมผสาน	72.2846	0.7235	0.7210	0.2242
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	66.2626	0.6699	0.6831	0.2197
	วิธีการสุ่มเกินเทคนิค SMOTE	83.8264	0.7325	0.9478	0.1274
	วิธีการสุ่มลด	83.3333	0.6524	1*	0.1641
	วิธีการผสมผสาน	62.1723	0.6313	0.6031	0.2537
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	63.6364	0.7764	0.4858	0.3637
	วิธีการสุ่มเกินเทคนิค SMOTE	84.2209	0.7053	0.9841	0.1577
	วิธีการสุ่มลด	70	0.6523	0.7444	0.3000
	วิธีการผสมผสาน	61.7978	0.7264	0.5152	0.3820

* หมายถึง ค่าความถูกต้อง ค่าความไว ค่าความจำเพาะเฉลี่ยสูงสุด และค่าคลาดเคลื่อนกำสองเฉลี่ยต่ำสุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.26 ผลการวิเคราะห์ชุดข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกโดยการจำแนกข้อมูลและการปรับความไม่สมดุลของข้อมูล เมื่อ Random Seed 10, 20 และ 30 พบว่า

วิธีเพื่อนบ้านใกล้สุด k ตัวโดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความถูกต้องเฉลี่ยสูงสุดคือร้อยละ 84.4181 ให้ค่าความไวเฉลี่ยสูงสุดคือ 0.8490 ให้ค่าความจำเพาะเฉลี่ยสูงสุดคือ 0.8395 และให้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.1523

วิธีต้นไม้ตัดสินใจโดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความถูกต้องเฉลี่ยสูงสุดคือร้อยละ 83.4319 ให้ค่าความไวเฉลี่ยสูงสุดคือ 0.8044
ให้ค่าความจำเพาะเฉลี่ยสูงสุดคือ 0.8638 และให้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.1231

วิธีโครงข่ายประสาทเทียมโดยการปรับความไม่สมดุลของข้อมูลด้วยด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความถูกต้องเฉลี่ยสูงสุดคือร้อยละ 83.8264 ให้ค่าความไวเฉลี่ยสูงสุดคือ 0.7325 และให้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.1274 ส่วนการปรับความไม่สมดุลด้วยวิธีการสุ่มลดให้ค่าความจำเพาะเฉลี่ยสูงสุดคือ 1

วิธีซัพพอร์ตเวกเตอร์แมชชีนโดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินด้วยเทคนิค SMOTE ให้ค่าความถูกต้องเฉลี่ยสูงสุดคือร้อยละ 84.2209 ให้ค่าความจำเพาะเฉลี่ยสูงสุดคือ 0.9841 และให้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.1577 ส่วนการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินให้ค่าความไวเฉลี่ยสูงสุดคือ 0.7764

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.6.3.2 ชุดข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้า

ตารางที่ 4.27 ผลการวิเคราะห์ชุดข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10, 20 และ 30

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	73.2394	0.7778	0.6856	0.2604
	วิธีการสุ่มเกินเทคนิค SMOTE	76.3889	0.6774	0.8293	0.2292
	วิธีการสุ่มลด	50	0.6000	0.4000	0.4173
	วิธีการผสมผสาน	40	0.3158	0.4762	0.5691
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	**78.8732**	**0.8056**	0.7714	**0.1860**
	วิธีการสุ่มเกินเทคนิค SMOTE	69.4444	0.5484	0.8049	0.2735
	วิธีการสุ่มลด	60	0.8000	0.4000	0.3549
	วิธีการผสมผสาน	42.5	0.4210	0.4286	0.4970
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	66.1972	0.6111	0.7143	0.2735
	วิธีการสุ่มเกินเทคนิค SMOTE	72.2222	0.5806	0.8293	0.2408
	วิธีการสุ่มลด	30	0.2000	0.4000	0.6374
	วิธีการผสมผสาน	50	0.5789	0.4286	0.3750
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	63.3803	0.5000	0.7714	0.3661
	วิธีการสุ่มเกินเทคนิค SMOTE	70.8333	0.5161	**0.8537**	0.2917
	วิธีการสุ่มลด	50	0.6000	0.4000	0.4999
	วิธีการผสมผสาน	52.5	0.4211	0.6190	0.4750
Random Seed 20					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	71.8310	0.8333	0.6342	0.2757
	วิธีการสุ่มเกินเทคนิค SMOTE	70.8333	0.7838	0.6286	0.2831
	วิธีการสุ่มลด	60	0.7142	0.3333	0.3356
	วิธีการผสมผสาน	60	0.5714	0.6154	0.3808
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	67.6056	0.7333	0.6341	0.2910
	วิธีการสุ่มเกินเทคนิค SMOTE	72.2222	0.7838	0.6571	0.2375
	วิธีการสุ่มลด	50	0.7142	0	0.3298
	วิธีการผสมผสาน	75	0.8462	0.5714	**0.2183**
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	66.1972	0.7000	0.6341	0.2536
	วิธีการสุ่มเกินเทคนิค SMOTE	68.0556	0.7568	0.6000	0.2748
	วิธีการสุ่มลด	40	0.4286	0.3333	0.5196
	วิธีการผสมผสาน	57.5	0.4231	0.8571	0.2517
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	60.5634	0.4000	0.7561	0.3944
	วิธีการสุ่มเกินเทคนิค SMOTE	**76.3889**	**0.8649**	0.6571	0.2361
	วิธีการสุ่มลด	60	0.8571	0	0.4001
	วิธีการผสมผสาน	62.5	0.1429	**0.8846**	0.3750

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.27 ผลการวิเคราะห์ชุดข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อ Random Seed 10, 20 และ 30 (ต่อ)

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำสองเฉลี่ย
Random Seed 30					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	70.4225	0.8372	0.5000	0.2875
	วิธีการสุ่มเกินเทคนิค SMOTE	**83.3333**	0.9231	0.7273	**0.1627**
	วิธีการสุ่มลด	50	0.5000	0.5000	0.4173
	วิธีการผสมผสาน	45	0.5909	0.2778	0.5236
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	71.831	0.8140	0.5714	0.2572
	วิธีการสุ่มเกินเทคนิค SMOTE	76.3889	0.7865	**0.7353**	0.2175
	วิธีการสุ่มลด	50	0.2500	0.6667	0.4445
	วิธีการผสมผสาน	42.5	0.5454	0.2778	0.5010
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	69.0141	0.7674	0.5714	0.2428
	วิธีการสุ่มเกินเทคนิค SMOTE	77.7778	0.8684	0.6765	0.1920
	วิธีการสุ่มลด	50	0.5000	0.5000	0.5054
	วิธีการผสมผสาน	52.5	0.6818	0.3333	0.3448
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	66.1927	0.8140	0.4286	0.3380
	วิธีการสุ่มเกินเทคนิค SMOTE	81.9444	**0.9474**	0.6765	0.1805
	วิธีการสุ่มลด	30	0	0.5000	0.7000
	วิธีการผสมผสาน	57.5	0.7273	0.3889	0.4250
ค่าเฉลี่ย					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	71.8309	**0.8161***	0.6066	0.2745
	วิธีการสุ่มเกินเทคนิค SMOTE	**76.8518***	0.7947	**0.7284**	**0.2250***
	วิธีการสุ่มลด	53.3333	0.6047	0.4111	0.3900
	วิธีการผสมผสาน	48.3333	0.4927	0.4564	0.4911
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	**72.7699**	**0.7843**	0.6589	0.2447
	วิธีการสุ่มเกินเทคนิค SMOTE	72.6851	0.7062	**0.7324***	**0.2428**
	วิธีการสุ่มลด	53.3333	0.5880	0.3555	0.3764
	วิธีการผสมผสาน	53.3333	0.6042	0.4259	0.4054
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	67.1361	0.6928	0.6399	0.2566
	วิธีการสุ่มเกินเทคนิค SMOTE	**72.6852**	**0.7352**	**0.7019**	**0.2358**
	วิธีการสุ่มลด	40	0.3762	0.4111	0.5541
	วิธีการผสมผสาน	53.3333	0.5612	0.5396	0.3238
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	63.3788	0.5713	0.6520	0.3661
	วิธีการสุ่มเกินเทคนิค SMOTE	**76.3888**	**0.7761**	**0.7291**	**0.2361**
	วิธีการสุ่มลด	46.6666	0.4857	0.3000	0.5333
	วิธีการผสมผสาน	57.5000	0.4304	0.6308	0.4250

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.27 ผลการวิเคราะห์ชุดข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้าโดยการจำแนกข้อมูลและการปรับความไม่สมดุลของข้อมูล เมื่อ Random Seed 10, 20 และ 30 พบว่า

วิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความถูกต้องเฉลี่ยสูงสุดคือร้อยละ 76.8518 ให้ค่าความจำเพาะเฉลี่ยสูงสุดคือ 0.7284 และให้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.2250 ส่วนการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินให้ค่าความไวเฉลี่ยสูงสุดคือ 0.8161

วิธีต้นไม้ตัดสินใจโดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินให้ค่าความถูกต้องเฉลี่ยสูงสุดคือ ร้อยละ 72.7699 และให้ค่าความไวเฉลี่ยสูงสุดคือ 0.7843 ส่วนการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความจำเพาะเฉลี่ยสูงสุดคือ 0.7324 และให้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.2428

วิธีโครงข่ายประสาทเทียมโดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความถูกต้องเฉลี่ยสูงสุดคือร้อยละ 72.6852 ให้ค่าความไวเฉลี่ยสูงสุคือ 0.7352 ให้ค่าความจำเพาะเฉลี่ยสูงสุดคือ 0.7019 และให้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.2358

วิธีซัพพอร์ตเวกเตอร์แมชชีนโดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความถูกต้องเฉลี่ยสูงสุดคือร้อยละ 76.3888 ให้ค่าความไวเฉลี่ยสูงสุดคือ 0.7761 ให้ค่าความจำเพาะเฉลี่ยสูงสุดคือ 0.7291 และให้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.2361

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.6.3.3 ชุดข้อมูลคุณภาพไวน์แดง

ตารางที่ 4.28 ผลการวิเคราะห์ชุดข้อมูลคุณภาพไวน์แดง เมื่อ Random Seed 10, 20 และ 30

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 10					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	53.75	0.6250	0.4500	0.4596
	วิธีการสุ่มเกินเทคนิค SMOTE	62.2642	0.6625	0.5822	0.3722
	วิธีการสุ่มลด	56.4103	0.3846	0.6538	0.4132
	วิธีการผสมผสาน	56	0.5370	0.5870	0.4306
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	53.75	0.5250	0.5500	0.3390
	วิธีการสุ่มเกินเทคนิค SMOTE	60.3774	0.6875	0.5190	0.3145
	วิธีการสุ่มลด	46.1538	0.3077	0.5384	0.4718
	วิธีการผสมผสาน	57	0.5926	0.5435	0.3741
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	54.375	0.5875	0.5000	0.2924
	วิธีการสุ่มเกินเทคนิค SMOTE	59.7484	**0.7750**	0.4177	**0.2724**
	วิธีการสุ่มลด	56.4103	0.5385	0.5769	0.3716
	วิธีการผสมผสาน	52	0.5185	0.5217	0.3191
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	47.5	0.4250	0.5250	0.5250
	วิธีการสุ่มเกินเทคนิค SMOTE	62.8931	0.725	0.5316	0.3711
	วิธีการสุ่มลด	**66.6667**	0.0769	**0.9615**	0.3334
	วิธีการผสมผสาน	53	0.6481	0.3913	0.4700
Random Seed 20					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	61.875	0.6400	0.6000	0.3759
	วิธีการสุ่มเกินเทคนิค SMOTE	57.2327	0.5443	0.6000	0.4243
	วิธีการสุ่มลด	58.9744	0.7083	0.4000	0.3895
	วิธีการผสมผสาน	62	0.5957	0.6415	0.3725
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	50.625	0.3067	0.6824	0.3232
	วิธีการสุ่มเกินเทคนิค SMOTE	56.6038	0.7848	0.3500	0.2739
	วิธีการสุ่มลด	51.2821	0.6250	0.3333	0.4196
	วิธีการผสมผสาน	54	0.4681	0.6038	0.2916
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	56.25	0.2400	0.8471	0.2814
	วิธีการสุ่มเกินเทคนิค SMOTE	58.4906	0.8228	0.3500	0.2678
	วิธีการสุ่มลด	61.5385	0.7500	0.4000	0.3328
	วิธีการผสมผสาน	**70**	0.4680	**0.9057**	**0.2192**
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	51.875	0.2133	0.7882	0.4812
	วิธีการสุ่มเกินเทคนิค SMOTE	52.2013	0.5950	0.4500	0.4780
	วิธีการสุ่มลด	66.6667	**0.8333**	0.4000	0.3334
	วิธีการผสมผสาน	57	0.3617	0.5747	0.4299

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.28 ผลการวิเคราะห์ชุดข้อมูลคุณภาพไวน์แดง เมื่อ Random Seed 10, 20 และ 30 (ต่อ)

วิธีการจำแนก	วิธีการปรับความไม่สมดุลของข้อมูล	ค่าความถูกต้อง	ค่าความไว	ค่าความจำเพาะ	ค่าคลาดเคลื่อนกำลังสองเฉลี่ย
Random Seed 30					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	65.625	0.7093	0.5946	0.3392
	วิธีการสุ่มเกินเทคนิค SMOTE	49.6855	0.4815	0.5128	0.4994
	วิธีการสุ่มลด	71.7949	0.7727	0.6471	0.2675
	วิธีการผสมผสาน	67	0.5349	**0.7719**	0.3260
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	60	0.6279	0.5676	0.3390
	วิธีการสุ่มเกินเทคนิค SMOTE	47.7987	0.7284	0.2180	0.3169
	วิธีการสุ่มลด	61.5385	0.7727	0.4118	0.3446
	วิธีการผสมผสาน	65	0.5116	0.7544	0.3190
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	58.125	0.5465	0.6216	0.2834
	วิธีการสุ่มเกินเทคนิค SMOTE	49.0566	0.2346	0.7564	0.3010
	วิธีการสุ่มลด	69.2308	0.7727	0.5882	0.2929
	วิธีการผสมผสาน	52	0.5349	0.5087	0.3185
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	56.875	**0.7791**	0.3243	0.4313
	วิธีการสุ่มเกินเทคนิค SMOTE	49.0566	0.5926	0.3846	0.5094
	วิธีการสุ่มลด	**74.3590**	0.7727	0.7059	**0.2564**
	วิธีการผสมผสาน	64	0.3953	0.8246	0.3600
ค่าเฉลี่ย					
เพื่อนบ้านใกล้สุด k ตัว	วิธีการสุ่มเกิน	60.4167	**0.6581**	0.5482	0.3916
	วิธีการสุ่มเกินเทคนิค SMOTE	56.3941	0.5628	0.5650	0.4320
	วิธีการสุ่มลด	**62.3932**	0.6219	0.5670	**0.3567**
	วิธีการผสมผสาน	61.6667	0.5559	**0.6668**	0.3764
ต้นไม้ตัดสินใจ	วิธีการสุ่มเกิน	54.7917	0.4865	0.6000	0.3337
	วิธีการสุ่มเกินเทคนิค SMOTE	54.9266	**0.7336***	0.3623	**0.3018**
	วิธีการสุ่มลด	52.9915	0.5685	0.4278	0.4120
	วิธีการผสมผสาน	**58.6667**	0.5241	**0.6339**	0.3282
โครงข่ายประสาทเทียม	วิธีการสุ่มเกิน	56.2500	0.4580	**0.6562**	0.2857
	วิธีการสุ่มเกินเทคนิค SMOTE	55.7652	0.6108	0.5080	**0.2804***
	วิธีการสุ่มลด	**62.3932**	**0.6871**	0.5217	0.3324
	วิธีการผสมผสาน	58.0000	0.5071	0.6454	0.2856
ซัพพอร์ตเวกเตอร์แมชชีน	วิธีการสุ่มเกิน	52.0833	0.4725	0.5458	0.4792
	วิธีการสุ่มเกินเทคนิค SMOTE	54.7170	**0.6375**	0.4554	0.4528
	วิธีการสุ่มลด	**69.2308***	0.5610	**0.6891***	**0.3077**
	วิธีการผสมผสาน	58.0000	0.4684	0.5969	0.4200

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.28 ผลการวิเคราะห์ชุดข้อมูลคุณภาพไวน์แดงโดยการจำแนกข้อมูลและการปรับความไม่สมดุลของข้อมูล เมื่อ Random Seed 10, 20 และ 30 พบว่า

วิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลดให้ค่าความถูกต้องเฉลี่ยสูงสุดคือร้อยละ 62.3932 และให้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.3567 ส่วนการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินให้ค่าความไวเฉลี่ยสูงสุดคือ 0.6581 ส่วนการปรับความไม่สมดุลของข้อมูลด้วยวิธีการผสมผสานให้ค่าความจำเพาะเฉลี่ยสูงสุดคือ 0.6668

วิธีต้นไม้ตัดสินใจโดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มผสมผสานให้ค่าความถูกต้องเฉลี่ยสูงสุดคือร้อยละ 58.6667 และให้ค่าความจำเพาะเฉลี่ยสูงสุดคือ 0.6339 ส่วนการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความไวเฉลี่ยสูงสุดคือ 0.7336 และให้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.3018

วิธีโครงข่ายประสาทเทียมโดยการปรับความไม่สมดุลของข้อมูลด้วยด้วยวิธีการสุ่มลดให้ค่าความถูกต้องเฉลี่ยสูงสุดคือร้อยละ 62.3932 และให้ค่าความไวเฉลี่ยสูงสุดคือ 0.6871 ส่วนการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินให้ค่าความจำเพาะเฉลี่ยสูงสุดคือ 0.6562 ส่วนการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.2804

วิธีซัพพอร์ตเวกเตอร์แมชชีนโดยการปรับความไม่สมดุลของข้อมูลด้วยด้วยวิธีการสุ่มลดให้ค่าความถูกต้องเฉลี่ยสูงสุดคือร้อยละ 69.2308 ให้ค่าความจำเพาะเฉลี่ยสูงสุดคือ 0.6891 และให้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดคือ 0.3077 ส่วนการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ให้ค่าความไวเฉลี่ยสูงสุดคือ 0.6375

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

5.1 สรุปผลการวิจัย

ในการจัดทำปัญหาพิเศษครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการปรับข้อมูลที่ไม่สมดุล 4 วิธี เมื่อมีการจำแนก 4 วิธีคือ วิธีเพื่อนบ้านใกล้สุด k ตัว วิธีต้นไม้ตัดสินใจ วิธีโครงข่ายประสาทเทียม และวิธีซัพพอร์ตเวกเตอร์แมชชีน ว่าวิธีการจำแนก 4 วิธี ในการปรับความไม่สมดุลของข้อมูลวิธีใดมีประสิทธิภาพที่ดีที่สุด เพื่อเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูล 4 วิธี เมื่อมีการปรับข้อมูลที่ไม่สมดุลของข้อมูล 4 วิธี คือ วิธีการสุ่มเกิน วิธีการสุ่มเกินโดยเทคนิค SMOTE วิธีการสุ่มลด และวิธีการสุ่มผสมผสาน ว่าวิธีการปรับความไม่สมดุลของข้อมูล 4 วิธี ในการจำแนกวิธีใดมีประสิทธิภาพที่ดีที่สุด และเพื่อเปรียบเทียบประสิทธิภาพในภาพรวมของการจำแนกข้อมูล 4 วิธี และการปรับความไม่สมดุลของข้อมูล 4 วิธี ว่าวิธีใดมีประสิทธิภาพในการจำแนกและการปรับความไม่สมดุลของข้อมูลที่ดีที่สุด โดยพิจารณาจากค่าความถูกต้อง ค่าความไว ค่าความจำเพาะ และค่าคลาดเคลื่อนกำสองเฉลี่ย เนื่องจากวิธีดังกล่าวข้างต้นเป็นวิธีที่นิยมนำมาใช้ในการจำแนกข้อมูลที่ทราบผลลัพธ์แน่นอน โดยใช้หลักการของการทำเหมืองข้อมูลมาใช้ในการจำแนกข้อมูล

จากการค้นคว้าและศึกษาในการหาข้อมูลที่ไม่สมดุลได้ข้อมูล 3 ชุด คือ ชุดข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก มีจำนวนกลุ่มส่วนน้อยร้อยละ 4.49 ชุดข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้า มีจำนวนกลุ่มส่วนน้อยร้อยละ 10 และชุดข้อมูลคุณภาพไวน์แดง มีจำนวนกลุ่มส่วนน้อยร้อยละ 19.52

ชุดข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก เมื่อพิจารณาวิธีการปรับความไม่สมดุลของข้อมูลแต่ละวิธีเพียงอย่างเดียวโดยใช้ค่าเฉลี่ย เมื่อ Random Seed 10, 20 และ 30 จะพบว่าวิธีการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินโดยการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว มีประสิทธิภาพดีที่สุด วิธีการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE โดยการจำแนกด้วยเพื่อนบ้านใกล้สุด k ตัว มีประสิทธิภาพดีที่สุด วิธีการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลดโดยการจำแนกด้วยวิธีโครงข่ายประสาทเทียมมีประสิทธิภาพดีที่สุด และวิธีการปรับความไม่สมดุลของข้อมูลด้วยวิธีการผสมผสานโดยการจำแนกด้วยวิธีต้นไม้ตัดสินใจมีประสิทธิภาพดีที่สุด เมื่อพิจารณาจากวิธีการจำแนกเพียงอย่างเดียวโดยใช้ค่าเฉลี่ย เมื่อ Random Seed 10, 20 และ 30 จะพบว่าวิธีการจำแนกเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE มีประสิทธิภาพดีที่สุด วิธีการจำแนกต้นไม้ตัดสินใจโดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE มีประสิทธิภาพดีที่สุด วิธีการจำแนกโครงข่ายประสาทเทียมโดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE มีประสิทธิภาพดีที่สุด และวิธีการจำแนกซัพพอร์ตเวกเตอร์แมชชีนโดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE มีประสิทธิภาพดีที่สุด และเมื่อพิจารณาในภาพรวมของการจำแนกและการปรับความไม่สมดุลโดยใช้ค่าเฉลี่ย เมื่อ Random Seed 10, 20 และ 30 วิธีที่มีประสิทธิภาพสูงสุดคือวิธีเพื่อนบ้านใกล้สุด k ตัวโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชุดข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้า เมื่อพิจารณาวิธีการปรับความไม่สมดุลของข้อมูลแต่ละวิธีเพียงอย่างเดียวโดยใช้ค่าเฉลี่ย เมื่อ Random Seed 10, 20 และ 30 จะพบว่าวิธีการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินโดยการจำแนกด้วยวิธีต้นไม้ตัดสินใจมีประสิทธิภาพดีที่สุด วิธีการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE โดยการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว มีประสิทธิภาพดีที่สุด วิธีการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลด โดยการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว มีประสิทธิภาพดีที่สุด และวิธีการผสมผสานโดยการจำแนกด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพดีที่สุด เมื่อพิจารณาจากวิธีการจำแนกเพียงอย่างเดียวโดยใช้ค่าเฉลี่ย เมื่อ Random Seed 10, 20 และ 30 จะพบว่าวิธีการจำแนกเพื่อนบ้านใกล้สุด k ตัว โดยวิธีการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE มีประสิทธิภาพดีที่สุด วิธีการจำแนกต้นไม้ตัดสินใจโดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE และวิธีการสุ่มลด มีประสิทธิภาพดีที่สุด วิธีการจำแนกโครงข่ายประสาทเทียม โดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE มีประสิทธิภาพดีที่สุด และวิธีการจำแนกด้วยซัพพอร์ตเวกเตอร์แมชชีนโดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE มีประสิทธิภาพดีที่สุด และเมื่อพิจารณาในภาพรวมของการจำแนกและการปรับความไม่สมดุลโดยใช้ค่าเฉลี่ย เมื่อ Random Seed 10, 20 และ 30 วิธีที่มีประสิทธิภาพสูงสุดคือวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE

ชุดข้อมูลคุณภาพไวน์แดงเมื่อพิจารณาวิธีการปรับความไม่สมดุลของข้อมูลแต่ละวิธีเพียงอย่างเดียวโดยใช้ค่าเฉลี่ย เมื่อ Random Seed 10, 20 และ 30 จะพบว่าวิธีการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินโดยการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว มีประสิทธิภาพดีที่สุด วิธีการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE โดยการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว มีประสิทธิภาพดีที่สุด วิธีการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลดโดยการจำแนกด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพดีที่สุด และวิธีการปรับความไม่สมดุลของข้อมูลด้วยวิธีการผสมผสานโดยการจำแนกด้วยวิธีเพื่อนบ้านใกล้สุด k ตัว มีประสิทธิภาพดีที่สุด เมื่อพิจารณาจากวิธีการจำแนกเพียงอย่างเดียวเมื่อ Random Seed 10, 20 และ 30 จะพบว่าวิธีการจำแนกเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลด มีประสิทธิภาพดีที่สุด วิธีการจำแนกต้นไม้ตัดสินใจโดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE และวิธีการผสมผสานมีประสิทธิภาพดีที่สุด วิธีการจำแนกโครงข่ายประสาทเทียมโดยการปรับความไม่สมดุลของข้อมูลด้วยด้วยวิธีการสุ่มเกินเทคนิค SMOTE และวิธีการสุ่มลดมีประสิทธิภาพดีที่สุด และวิธีการจำแนกซัพพอร์ตเวกเตอร์แมชชีนโดยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลดมีประสิทธิภาพดีที่สุด และเมื่อพิจารณาในภาพรวมของการจำแนกและการปรับความไม่สมดุลโดยใช้ค่าเฉลี่ย เมื่อ Random Seed 10, 20 และ 30 วิธีที่มีประสิทธิภาพสูงสุดคือวิธีซัพพอร์ตเวกเตอร์แมชชีนโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด

ดังนั้นชุดข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก วิธีที่มีประสิทธิภาพสูงสุดคือวิธีเพื่อนบ้านใกล้สุด k ตัว โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ชุดข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้า วิธีที่มีประสิทธิภาพสูงสุดคือวิธีเพื่อนบ้านใกล้สุด k ตัวโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE ส่วนชุดข้อมูลคุณภาพไวน์แดง วิธีที่มีประสิทธิภาพสูงสุดคือวิธีซัพพอร์ตเวกเตอร์แมชชีนโดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นั่นคือถ้าพิจารณาวิธีการจำแนกเพียงอย่างเดียวจะพบว่าวิธีการจำแนกที่ดีที่สุดคือ วิธีเพื่อนบ้านใกล้สุด k ตัวรองลงมาคือวิธีต้นไม้ตัดสินใจ ส่วนถ้าพิจารณาวิธีการปรับความไม่สมดุลของข้อมูลเพียงอย่างเดียวจะพบว่าวิธีการปรับความไม่สมดุลของข้อมูลดีที่สุดคือวิธีการสุ่มเกินโดยเทคนิค SMOTE รองลงมาคือวิธีการสุ่มลด

5.2 อภิปรายผล

จากการสรุปผลปัญหาพิเศษครั้งนี้ ชุดข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก วิธีเพื่อนบ้านใกล้สุด k ตัวใช้ขั้นตอนวิธี IBk โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินเทคนิค SMOTE มีประสิทธิภาพในการจำแนกดีที่สุด เพราะว่ามีค่าความถูกต้องสูงสุด แต่การจำแนกข้อมูลให้ผลไม่สอดคล้องกับเชาวนันท์ และคณะ (2556) เรื่องแบบจำลองการทำนายผลการรักษาผู้ป่วยมะเร็งปากมดลูกวิธีโครงข่ายประสาทเทียมโดยการปรับด้วยวิธีการสุ่มเกินโดยเทคนิค SMOTE มีประสิทธิภาพในการจำแนกดีที่สุด เนื่องจากข้อมูลมีลักษณะแตกต่างกัน โดยชุดข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวกมีกลุ่มส่วนน้อยร้อยละ 4.49 ส่วนชุดข้อมูลผลการรักษาผู้ป่วยมะเร็งปากมดลูกอาจมีกลุ่มส่วนน้อยมากกว่านี้ ชุดข้อมูลยอดคงเหลือในบัตรเครดิต วิธีเพื่อนบ้านใกล้สุด k ตัวโดยใช้ขั้นวิธี IBk โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มเกินโดยเทคนิค SMOTE มีประสิทธิภาพในการจำแนกดีที่สุด เพราะว่ามีค่าความถูกต้องสูงสุด แต่การจำแนกข้อมูลให้ผลไม่สอดคล้องกับเชาวนันท์ และคณะ (2556) เรื่องแบบจำลองการทำนายผลการรักษาผู้ป่วยมะเร็งปากมดลูก วิธีโครงข่ายประสาทเทียมโดยการปรับด้วยวิธีการสุ่มเกินโดยเทคนิค SMOTE มีประสิทธิภาพในการจำแนกดีที่สุด เนื่องจากข้อมูลมีลักษณะแตกต่างกัน โดยชุดข้อมูลยอดคงเหลือในบัตรเครดิตมีกลุ่มส่วนน้อยร้อยละ 10 ส่วนชุดข้อมูลผลการรักษาผู้ป่วยมะเร็งปากมดลูกอาจมีกลุ่มส่วนน้อยน้อยกว่าหรือมากกว่านี้ และชุดข้อมูลคุณภาพไวน์แดง วิธีซัพพอร์ตเวกเตอร์แมชชีนโดยใช้ขั้นตอนวิธี SMO โดยการปรับความไม่สมดุลด้วยวิธีการสุ่มลดมีประสิทธิภาพในการจำแนกดีที่สุด เพราะว่ามีค่าความถูกต้องสูงสุด ให้ผลไม่สอดคล้องกับ He and Ghodsi (2010) เรื่องการศึกษาเกี่ยวกับการจำแนกข้อมูลที่ค้นพบได้ยาก วิธีซัพพอร์ตเวกเตอร์แมชชีนโดยการปรับด้วยวิธีการสุ่มเกินมีประสิทธิภาพในการจำแนกที่ดีที่สุด เนื่องจากข้อมูลมีลักษณะแตกต่างกัน โดยชุดข้อมูลคุณภาพไวน์แดงมีกลุ่มส่วนน้อยร้อยละ 19.52 ส่วนชุดข้อมูลการศึกษาเกี่ยวกับการจำแนกข้อมูลที่ค้นพบได้ยากมีกลุ่มส่วนน้อยน้อยกว่าร้อยละ 19.52

5.3 ข้อเสนอแนะ

1. ชุดข้อมูลที่นำมาปรับความไม่สมดุลนั้นควรจะมีจำนวนชุดข้อมูลที่มากขึ้น เพื่อให้มีประสิทธิภาพในการจำแนกดีขึ้น

2. เพื่อให้ได้ข้อสรุปของผลการวิเคราะห์ข้อมูลที่มีความสมบูรณ์มากขึ้น ดังนั้นผู้วิจัยอาจวิเคราะห์ข้อมูลด้วยวิธีการจำแนกวิธีอื่นๆ เช่น นาอีฟเบส์ เบส์เนท การถดถอยลอจิสติกทวิภาค และ ฐานกฎ เป็นต้น

3. ควรศึกษาวิธีการปรับความไม่สมดุลของข้อมูลวิธีอื่นๆ เช่น Cost Sensitive Learning เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

กีระชาติ สุขสุทธิ์. 2559. “การจำแนกข้อมูลไม่สมดุลโดยใช้การปรับปรุงข้อมูลร่วมกับการหาค่าพารามิเตอร์ที่เหมาะสมด้วยขั้นตอนวิธีทางพันธุกรรมที่มีการเริ่มต้นใหม่.” วิทยานิพนธ์ปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์, มหาวิทยาลัยเทคโนโลยีสุรนารี.

เกร็ดความรู้. 2556. **บัตรเครดิต.** [Online]. เข้าถึงได้จาก : https://www.เกร็ดความรู้.net/บัตเครดิต/.

เชาวนันท์ โสโท, พุธษดี ศิริแสงตระกูล และวรชัย ตั้งวรพงศ์ชัย. 2556. “แบบจำลองการทำนายผลการรักษาผู้ป่วยมะเร็งปากมดลูกด้วยโครงข่ายประสาทเทียม.” วารสารวิจัย มข. 13(1) : 39-50.

ทิพย์ธิดา วงศ์พิพันธ์. 2555. “การใช้เหมืองข้อมูลช่วยในการตัดสินใจการให้สินเชื่อ.” งานค้นคว้าอิสระตามหลักสูตรวิทยาศาสตร์มหาบัณฑิต สาขาวิชาเทคโนโลยีคอมพิวเตอร์และการสื่อสาร คณะวิศวกรรมศาสตร์, มหาวิทยาลัยธุรกิจบัณฑิต.

นฤพนต์ ว่องประชานุกูล. 2548. “วิธีที่เหมาะสมสำหรับการตัดกิ่งต้นไม้ตัดสินใจของการทำเหมืองข้อมูลทางด้านวิทยาศาสตร์.” วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์, มหาวิทยาลัยเทคโนโลยีสุรนารี.

เบญจภรณ์ จันทรกองกุล, สุวรรณา รัศมีขวัญ, สุนิสา ริมเจริญ, ภูสิต กุลเกษม, กฤษณะ ชินสาร, อัณณ์นุพันธ์ รอดทุกข์, ปิยนุช วรบุตร และจรรยา อ้นปันส์. 2557. **วิธีการที่เหมาะสมสำหรับการแบ่งกลุ่มข้อมูลที่ไม่สมดุลสูง.** [Online]. เข้าถึงได้จาก : http://digital_collect.lib.buu.ac.th/dcms/files/2559_047.pdf.

พนิดา สมบัติมาก, ภัสสร จันทร์หอม, ศุภกร รัศมี และโอฬาร รุ่งมณีธรรมคุณ. 2560. “การเปรียบเทียบประสิทธิภาพในการจำแนกกลุ่มเมื่อข้อมูล.” ปัญหาพิเศษระดับปริญญาตรี ภาควิชาสถิติ คณะวิทยาศาสตร์, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

ภรันยา ปาลวิสุทธิ์. 2559. “การเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุลโดยวิธีการการสุ่มเพิ่มตัวอย่างกลุ่มน้อยสำหรับสำหรับข้อมูลการเป็นโรคอินเตอร์เน็ต.” วารสารเทคโนโลยีสารสนเทศ. 12(1) : 54-63.

รติพร จันทรกลั่น. 2557. “การปรับปรุงอลักอริทึมซัพพอร์ตเวกเตอร์แมชชีนสำหรับการจำแนกข้อมูลภาพไบโอเมตริกซ์.” วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์, มหาวิทยาลัยเทคโนโลยีสุรนารี.

วีระยุทธ มายุศิริ, จารี ทองคำ และวาทินี สุขมาก. 2557. “การพัฒนาแบบจำลองเพื่อการพยากรณ์การรักษาซ้ำของผู้ป่วยโรคจิตเภทโดยเทคนิคเหมืองข้อมูล.” วารสารวิทยาศาสตร์และเทคโนโลยี. 10(0) : 144-153.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วราภรณ์ ม้วนทอง, วรุณศักดิ์ เลี่ยมแหลม และเสฏฐา ศาสนนันทน์. 2558. "แบบจำลองโครงข่ายประสาทเทียมสำหรับ Clarifier กรณีศึกษาโรงงานผลิตน้ำประปาบางเขน." วารสารวิศวกรรมศาสตร์. 10(1) : 32-44.

สุรวัชร ศรีเปารยะ และสายชล สินสมบูรณ์ทอง. 2560. "การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรัง: กรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดีย." วารสารวิทยาศาสตร์และเทคโนโลยี. 25(5) : 839-853.

อังศุมาลี สุทธภักติ 2553. "การเปรียบเทียบประสิทธิภาพของการจำแนกหมวดหมู่ของข้อคิดเห็นในแบบสอบถามปลายเปิด โดยวิธีเนอีฟเบย์และซัพพอร์ตเวคเตอร์แมชชีน." วิทยาลัยราชพฤกษ์.

Cortez, P. 2009. **Wine Quality Data Set.** [Online]. เข้าถึงได้จาก : http://archive.ics.uci.edu/ml/datasets/Wine+Quality.

He, H. and Ghodsi, A. 2010. "Rare Class Classification by Support Vector Machine." 548-551. in **20th International Conference on Pattern Recognition.** Turkey : ICPR.

Hogan, S. 2013. **Tests of Auditory Perception in Children with OME.** [Online]. เข้าถึงได้จาก : https://vincentarelbundock.gitub.io/Rdatasets/doc/MASS/OME.html.

Hu, S-b and Shao, P. 2012. "Improved nearest neighbor interpolators based on confidence region in medical image registration." Biomedical Signal Processing and Control. 7: 525-536.

James, G., Witten, D., Hastie, T. and Tibshirani, R. 2013. **Credit Card Balance Data.** [Online]. เข้าถึงได้จาก : https://vincentarelbundock.github.io/Rdataset/doc/ISLR/Credit.html.

McCarthy, K., Zabar, B. and Weiss, G. 2005. "Cost-Sensitive Learning vs. Sampling: Which is Best for Handling Unbalanced Classes with Unequal Error Costs?." Proceedings of the 2007 International Conference on Data. 7(1) : 35-41.

Medthai. 2552. **ภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก.** [Online]. เข้าถึงได้จาก: https://medthai.com/.

Rahman, M. M. and Davis, D. N. 2013. "Addressing the Class Imbalance Problem in Medical Datasets." International Journal of Machine Learning and Computing. 3(2) : 224-228.

Thaicardonline. 2554. **วงเงินบัตรเครดิต.** [Online]. เข้าถึงได้จาก : https://www.thaicardonline.com/วงเงินบัตรเครดิตคืออะไร/.

Wordpress. 2558. **ไวน์แดง.** [Online]. เข้าถึงได้จาก : https://ferneuf.wordpress. com/278-2/ไวน์แดง.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก
รายละเอียด ตัวอย่างข้อมูลที่นำมาวิเคราะห์ และตัวอย่างการแบ่งข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception inChildren with OME)

ตารางที่ ก-1 คุณลักษณะและรายละเอียดการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก

คุณลักษณะ	รายละเอียด
X_1 อายุ	หน่วย : เดือน
X_2 อัตราความเสี่ยงการเป็นโรคหูชั้นกลางอักเสบ	low คือ ความเสี่ยงน้อย high คือ ความเสี่ยงมาก N/A คือ ไม่มีความเสี่ยง
X_3 ความดังที่ได้รับ	หน่วย : เดซิเบล
X_4 สัญญาณในการกระตุ้น	coherent คือ สัญญาณต่อเนื่อง incoherent คือ สัญญาณไม่ต่อเนื่อ
X_5 จำนวนคำตอบที่ถูกต้องจากการทดลองฟัง	หน่วย : ครั้ง
Y การรับรู้ทางหู	No คือ ได้ยินเสียงไม่ปกติ Yes คือ ได้ยินเสียงปกติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก-2 ตัวอย่างข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก

No	X_1	X_2	X_3	X_4	X_5	Y
1	30	low	35	coherent	1	no
2	30	low	35	incoherent	4	yes
3	30	low	40	coherent	0	no
4	30	low	40	incoherent	1	no
5	30	low	45	coherent	2	no
6	30	low	45	incoherent	2	yes
7	30	low	50	coherent	3	no
8	30	low	50	incoherent	4	no
9	30	low	55	coherent	3	no
10	30	low	55	incoherent	2	no
11	60	low	35	coherent	2	yes
12	60	low	35	incoherent	3	no
13	60	low	40	coherent	1	no
14	60	low	40	incoherent	1	no
15	60	low	45	coherent	1	no
16	60	low	45	incoherent	5	no
17	60	low	50	coherent	4	no
18	60	low	50	incoherent	2	no
19	60	low	55	coherent	3	no
20	60	low	55	incoherent	4	yes
21	60	low	35	coherent	4	no
22	60	low	35	incoherent	4	no
23	60	low	40	coherent	4	no
24	60	low	40	incoherent	5	no
25	60	low	45	coherent	5	no
26	60	low	45	incoherent	6	no
27	60	low	50	coherent	9	no
28	60	low	50	incoherent	8	no
29	60	low	55	coherent	6	no
30	60	low	55	incoherent	5	no

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ยอดคงเหลือในบัตรเครดิตของลูกค้า (Credit Card Balance)

ตารางที่ ก-3 คุณลักษณะและรายละเอียดยอดคงเหลือในบัตรเครดิตของลูกค้า

คุณลักษณะ	รายละเอียด
X_1 รายได้ต่อเดือน	หน่วย : สกุลเงินดอลลาร์สหรัฐอเมริกา (USD)
X_2 วงเงินในบัตรเครดิต	หน่วย : สกุลเงินดอลลาร์สหรัฐอเมริกา (USD)
X_3 จำนวนครั้งในการใช้บัตรเครดิตต่อปี	หน่วย : ครั้ง
X_4 จำนวนบัตรเครดิตที่มี	หน่วย : บัตร
X_5 อายุ	หน่วย : ปี
X_6 จำนวนปีในการศึกษา	หน่วย : ปี
X_7 เพศ	Male คือ เพศชาย Female คือ เพศหญิง
X_8 การศึกษา	Yes คือ กำลังศึกษา No คือ สำเร็จการศึกษา
X_9 สถานภาพสมรส	Yes คือ แต่งงานแล้ว No คือ โสด
X_{10} ยอดเงินคงเหลือในบัตรเครดิต	หน่วย : สกุลเงินดอลลาร์สหรัฐอเมริกา (USD)
Y การชำระเงินในเดือนถัดไป	No คือ ไม่ชำระเงิน Yes คือ ชำระเงิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก-4 ตัวอย่างข้อมูลยอดคงเหลือในบัตรเครดิตของลูกค้า

No	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Y
1	14.891	3606	283	2	34	11	Male	no	yes	333	no
2	106.025	6645	483	3	82	15	Female	yes	yes	903	yes
3	104.593	7075	514	4	71	11	Male	no	no	580	no
4	148.924	9504	681	3	36	11	Female	no	no	964	no
5	55.882	4897	357	2	68	16	Male	no	yes	331	no
6	80.18	8047	569	4	77	10	Male	no	no	1151	no
7	20.996	3388	259	2	37	12	Female	no	no	203	no
8	71.408	7114	512	2	87	9	Male	no	no	872	no
9	15.125	3300	266	5	66	13	Female	no	no	279	no
10	71.061	6819	491	3	41	19	Female	yes	yes	1350	no
11	63.095	8117	589	4	30	14	Male	no	yes	1407	no
12	15.045	1311	138	3	64	16	Male	no	no	0	no
13	80.616	5308	394	1	57	7	Female	no	yes	204	no
14	43.682	6922	511	1	49	9	Male	no	yes	1081	no
15	19.144	3291	269	2	75	13	Female	no	no	148	no
16	20.089	2525	200	3	57	15	Female	no	yes	0	no
17	53.598	3714	286	3	73	17	Female	no	yes	0	no
18	36.496	4378	339	3	69	15	Female	no	yes	368	no
19	49.57	6384	448	1	28	9	Female	no	yes	891	no
20	42.079	6626	479	2	44	9	Male	no	no	1048	no
21	17.7	2860	235	4	63	16	Female	no	no	89	yes
22	37.348	6378	458	1	72	17	Female	no	no	968	no
23	20.103	2631	213	3	61	10	Male	no	yes	0	no
24	64.027	5179	398	5	48	8	Male	no	yes	411	no
25	10.742	1757	156	3	57	15	Female	no	no	0	no
26	14.09	4323	326	5	25	16	Female	no	yes	671	no
27	42.471	3625	289	6	44	12	Female	yes	no	654	no
28	32.793	4534	333	2	44	16	Male	no	no	467	no
29	186.634	13414	949	2	41	14	Female	no	yes	1809	no
30	26.813	5611	411	4	55	16	Female	no	no	915	yes

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. คุณภาพไวน์แดง (Red Wine Quality)

ตารางที่ ก-5 คุณลักษณะและรายละเอียดคุณภาพไวน์แดง

คุณลักษณะ	รายละเอียด
X_1 ค่าความเป็นกรด	หน่วย : pH
X_2 ค่าความเป็นกรดระเหย	หน่วย : pH
X_3 ค่ากรดซิตริก	หน่วย : pH
X_4 ค่าน้ำตาลรีดิวซ์	หน่วย : กรัม
X_5 ค่าคลอไรด์	หน่วย : กรัม
X_6 ค่าซัลเฟอร์ไดออกไซด์	หน่วย : กรัม
X_7 ค่าซัลเฟอร์ไดออกไซด์ทั้งหมด	หน่วย : กรัม
X_8 ค่าความหนาแน่น	หน่วย : กิโลกรัมต่อลูกบาศก์เมตร
X_9 ค่าความเป็นกรดหรือด่าง	หน่วย : pH
X_{10} ค่าเกลือของกรดซัลฟิวริก	หน่วย : pH
X_{11} ค่าแอลกอฮอล์	หน่วย : กรัม
X_{12} คะแนนคุณภาพ	หน่วย : คะแนน
Y คุณภาพไวน์แดง	No คือ คุณภาพไม่ดี Yes คือ คุณภาพดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก-6 ตัวอย่างข้อมูลคุณภาพไวน์แดง

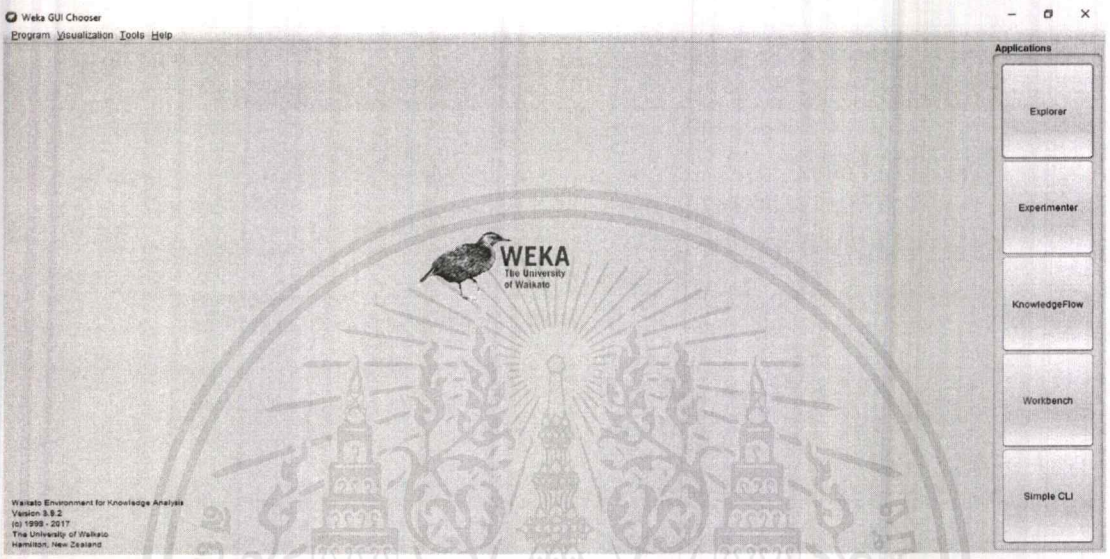
No	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	Y
1	7.4	0.7	0	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	5	no
2	7.8	0.76	0.04	2.3	0.092	15	54	0.997	3.26	0.65	9.8	4	yes
3	11.2	0.28	0.56	1.9	0.075	17	60	0.998	3.16	0.58	9.8	6	yes
4	7.4	0.7	0	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	4	no
5	7.4	0.66	0	1.8	0.075	13	40	0.9978	3.51	0.56	9.4	4	no
6	7.9	0.6	0.06	1.6	0.069	15	59	0.9964	3.3	0.46	9.4	4	yes
7	7.3	0.65	0	1.2	0.065	15	21	0.9946	3.39	0.47	10	7	yes
8	7.8	0.58	0.02	2	0.073	9	18	0.9968	3.36	0.57	9.5	7	no
9	7.5	0.5	0.36	6.1	0.071	17	102	0.9978	3.35	0.8	10.5	5	no
10	6.7	0.58	0.08	1.8	0.097	15	65	0.9959	3.28	0.54	9.2	4	no
11	7.5	0.5	0.36	6.1	0.071	17	102	0.9978	3.35	0.8	10.5	5	yes
12	5.6	0.615	0	1.6	0.089	16	59	0.9943	3.58	0.52	9.9	4	yes
13	7.8	0.61	0.29	1.6	0.114	9	29	0.9974	3.26	1.56	9.1	5	no
14	8.9	0.62	0.18	3.8	0.176	52	145	0.9986	3.16	0.88	9.2	5	no
15	8.9	0.62	0.19	3.9	0.17	51	148	0.9986	3.17	0.93	9.2	4	no
16	8.5	0.28	0.56	1.8	0.092	35	103	0.9969	3.3	0.75	10.5	7	no
17	8.1	0.56	0.28	1.7	0.368	16	56	0.9968	3.11	1.28	9.3	5	yes
18	7.4	0.59	0.08	4.4	0.086	6	29	0.9974	3.38	0.5	9	4	yes
19	7.9	0.32	0.51	1.8	0.341	17	56	0.9969	3.04	1.08	9.2	6	no
20	8.9	0.22	0.48	1.8	0.077	29	60	0.9968	3.39	0.53	9.4	6	no
21	7.6	0.39	0.31	2.3	0.082	23	71	0.9982	3.52	0.65	9.7	5	yes
22	7.9	0.43	0.21	1.6	0.106	10	37	0.9966	3.17	0.91	9.5	5	yes
23	8.5	0.49	0.11	2.3	0.084	9	67	0.9968	3.17	0.53	9.4	5	no
24	6.9	0.4	0.14	2.4	0.085	21	40	0.9968	3.43	0.63	9.7	6	no
25	6.3	0.39	0.16	1.4	0.08	11	23	0.9955	3.34	0.56	9.3	5	no
26	7.6	0.41	0.24	1.8	0.08	4	11	0.9962	3.28	0.59	9.5	4	yes
27	7.9	0.43	0.21	1.6	0.106	10	37	0.9966	3.17	0.91	9.5	5	no
28	7.1	0.71	0	1.9	0.08	14	35	0.9972	3.47	0.55	9.4	4	yes
29	7.8	0.645	0	2	0.082	8	16	0.9964	3.38	0.59	9.8	6	yes
30	6.7	0.675	0.07	2.4	0.089	17	82	0.9958	3.35	0.54	10.1	4	no

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างการแบ่งข้อมูล

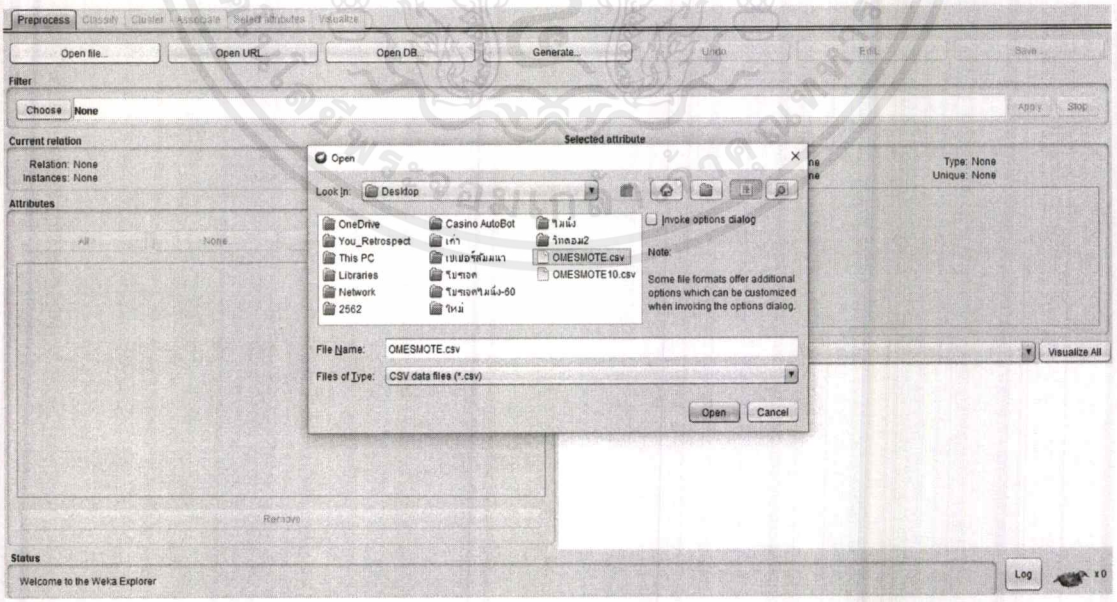
การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME) จำนวนข้อมูลทั้งหมด 1690 ค่า ทำการแบ่งข้อมูล โดยกำหนด Random Seed 10, 20 และ 30

1. เปิดโปรแกรม WEKA เวอร์ชั่น 3.9.2 กดปุ่ม Explorer



รูปที่ ก-1 หน้าต่างโปรแกรม WEKA

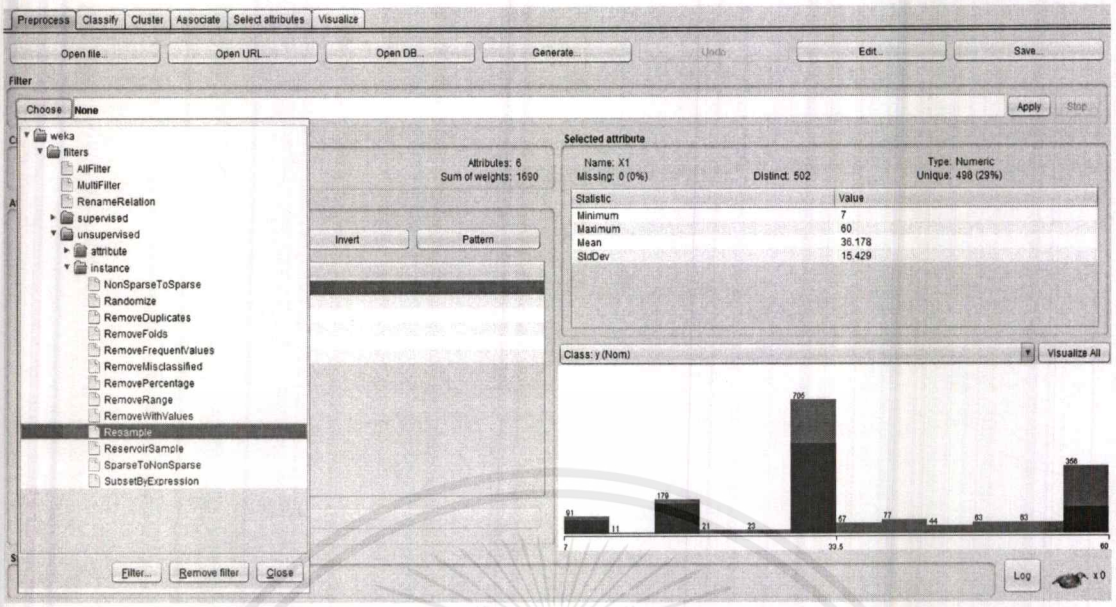
2. กดปุ่ม Open File (นามสกุล .csv) จะได้ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก จำนวนข้อมูลทั้งหมด 1690 ค่า



รูปที่ ก-2 แสดงการเปิดไฟล์ การรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก

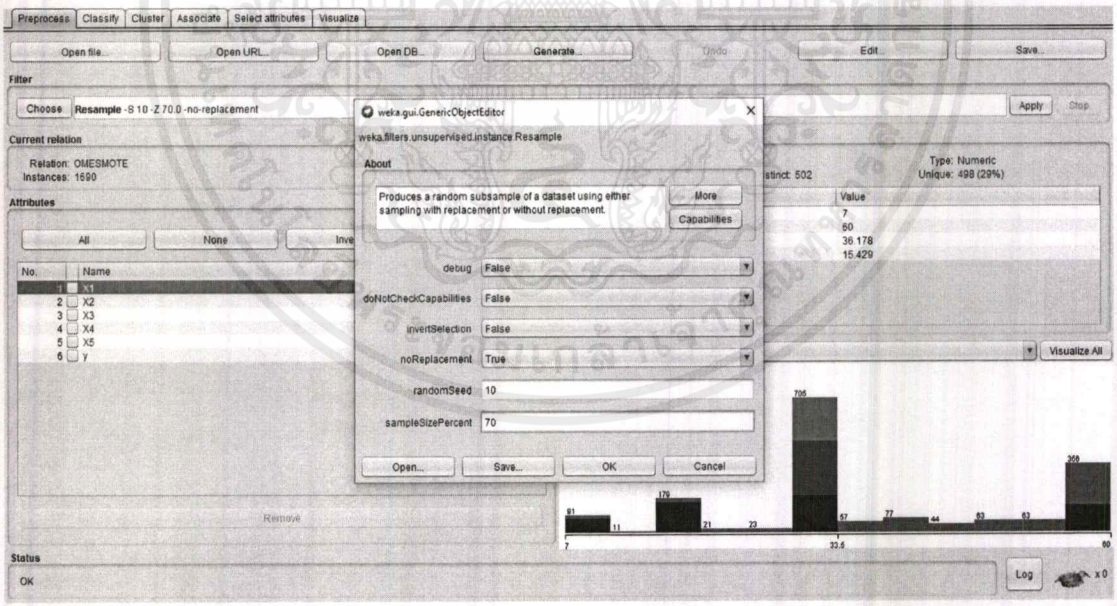
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. กดปุ่ม Choose ⟶ filters ⟶ unsupervised ⟶ instance ⟶ Resample



รูปที่ ก-3 ขั้นตอนการแบ่งข้อมูล การเลือกการสุ่ม

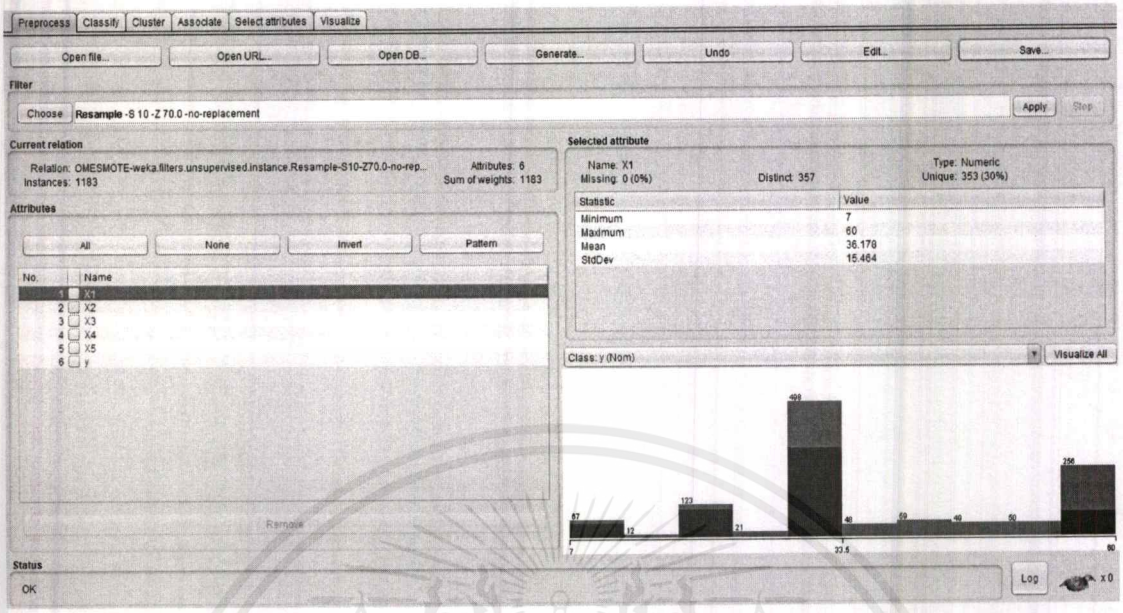
4. กดช่อง Resample - S 1 -Z 100.0 ⟶ ในช่อง invertSelection เป็น False ⟶ ในช่อง noReplacement เป็น True ⟶กำหนดค่า randomSeed เท่ากับ 10 ⟶ในช่อง sampleSize Percent กำหนดให้เป็น 70 คือ 70%



รูปที่ ก-4 ขั้นตอนการแบ่งข้อมูล 70%

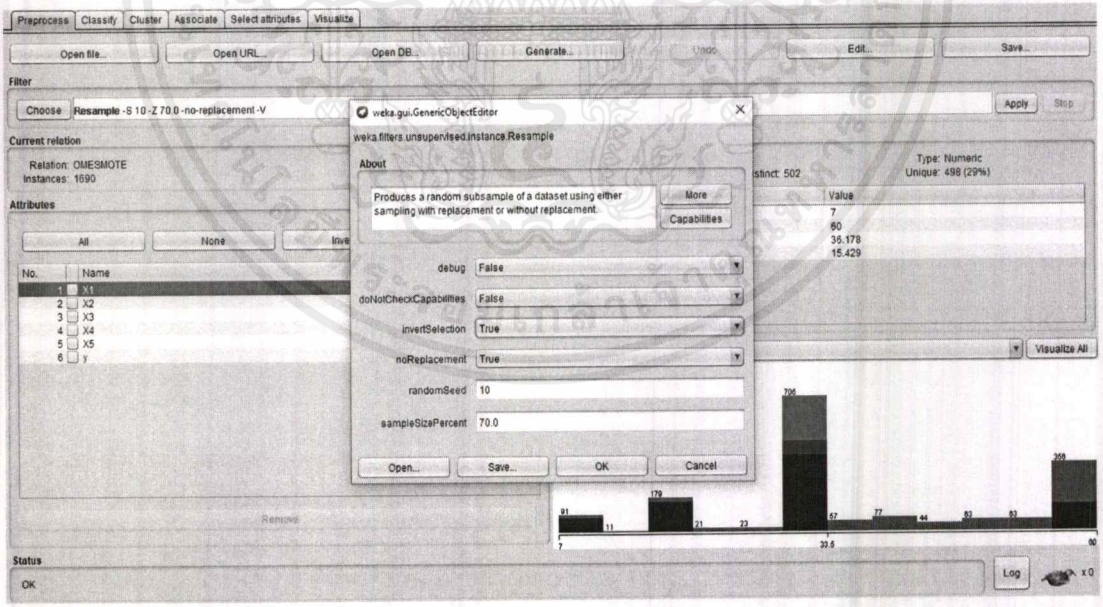
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. กด Apply จะแสดงค่าข้อมูล 70% จำนวน 1183 ค่า ⟶ กด Save กำหนดชื่อไฟล์เป็น OME SMOTE 70 ⟶ กด Undo



รูปที่ ก-5 แสดงผลลัพธ์ข้อมูล 70%

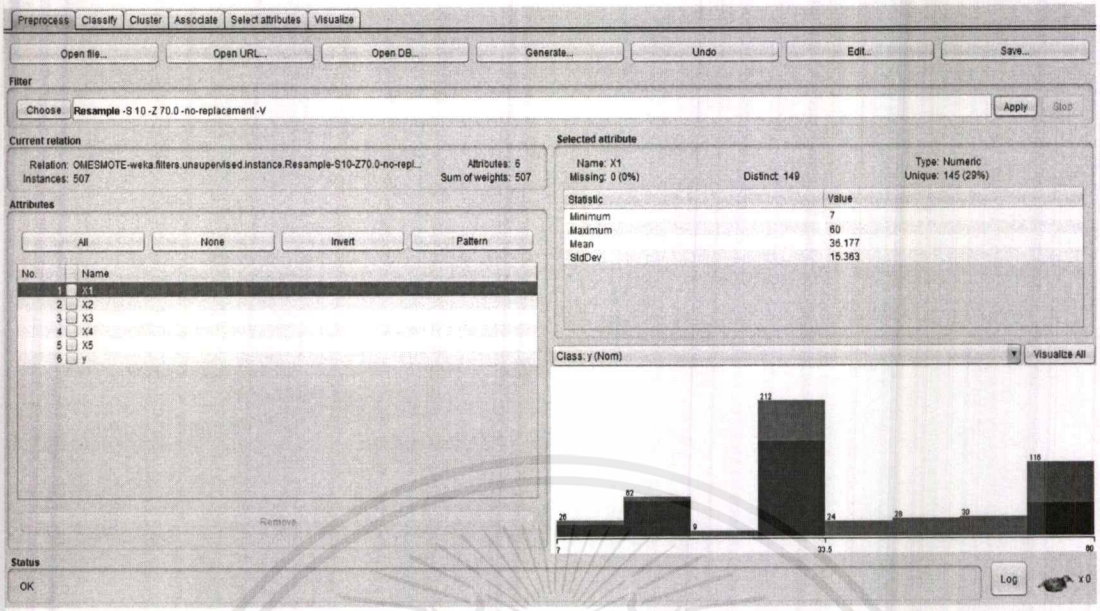
6. กดช่อง Resample - S 10 ⟶ ในช่อง invertSelection เป็น True ⟶ ในช่อง noReplacement เป็น True ⟶ กำหนดค่า randomSeed เท่ากับ 10 ⟶ ในช่อง sampleSize Percent กำหนดให้เป็น 70 เพื่อต้องการแบ่งข้อมูล 30% จาก 70%



รูปที่ ก-6 ขั้นตอนการแบ่งข้อมูล 30%

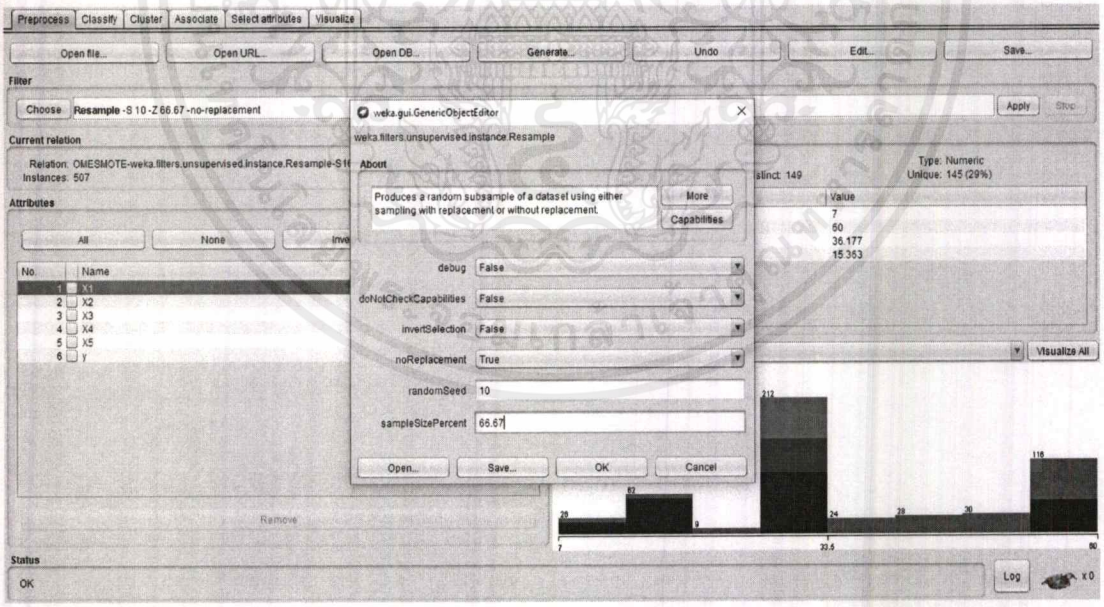
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7. กด Apply จะได้ข้อมูล 30% จำนวน 507 ค่า



รูปที่ ก-7 แสดงผลลัพธ์ข้อมูล 30%

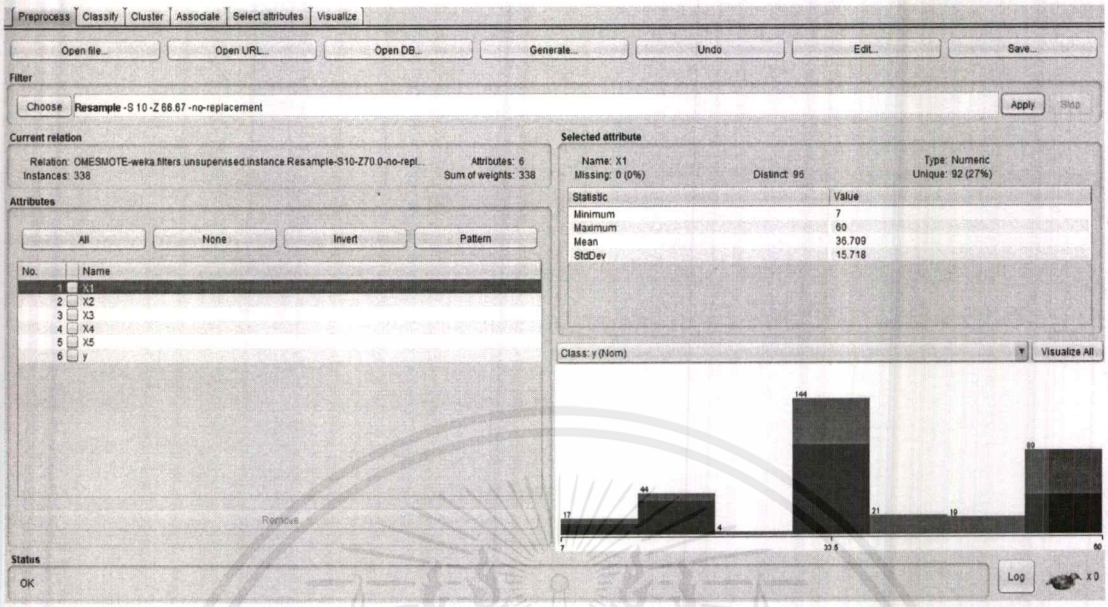
8. กดช่อง Resample - S 10 ⟶ในช่อง invertSelection เป็น False ⟶ ในช่อง noReplacement เป็น True ⟶ กำหนดค่า randomSeed เท่ากับ 10 ⟶ในช่อง sampleSize Percent กำหนดให้เป็น 66.67 เพื่อต้องการแบ่งข้อมูล 20% จาก 30%



รูปที่ ก-8 ขั้นตอนการแบ่งข้อมูล 20%

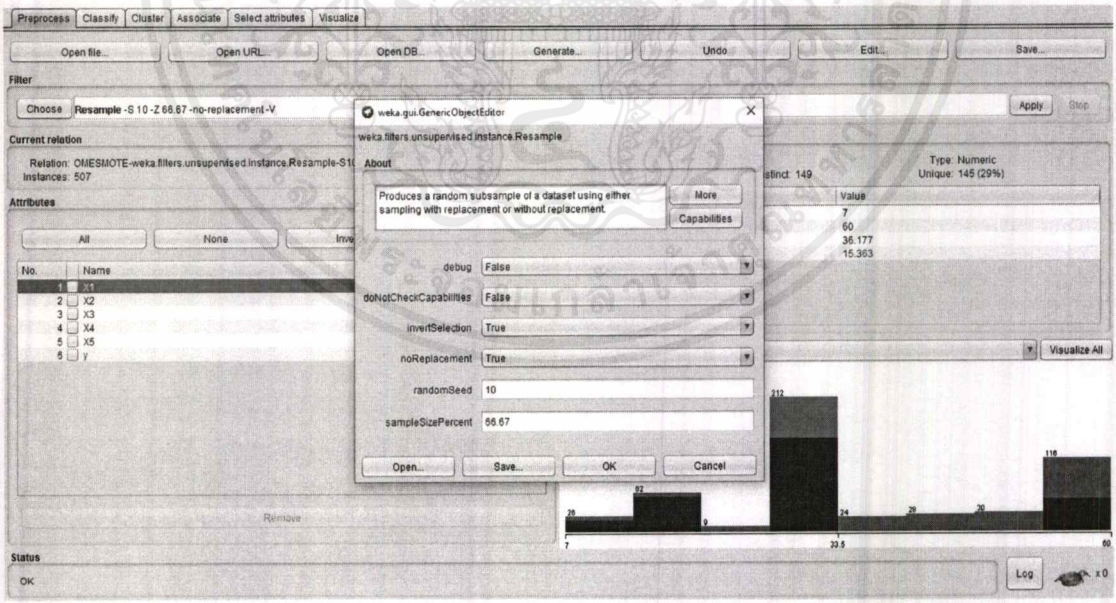
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

9. กด Apply จะได้ข้อมูล 20% จำนวน 338 ค่า ⟶ กด Save กำหนดชื่อไฟล์เป็น OME SMOTE 20 ⟶ กด Undo



รูปที่ ก-9 แสดงผลลัพธ์ข้อมูล 20%

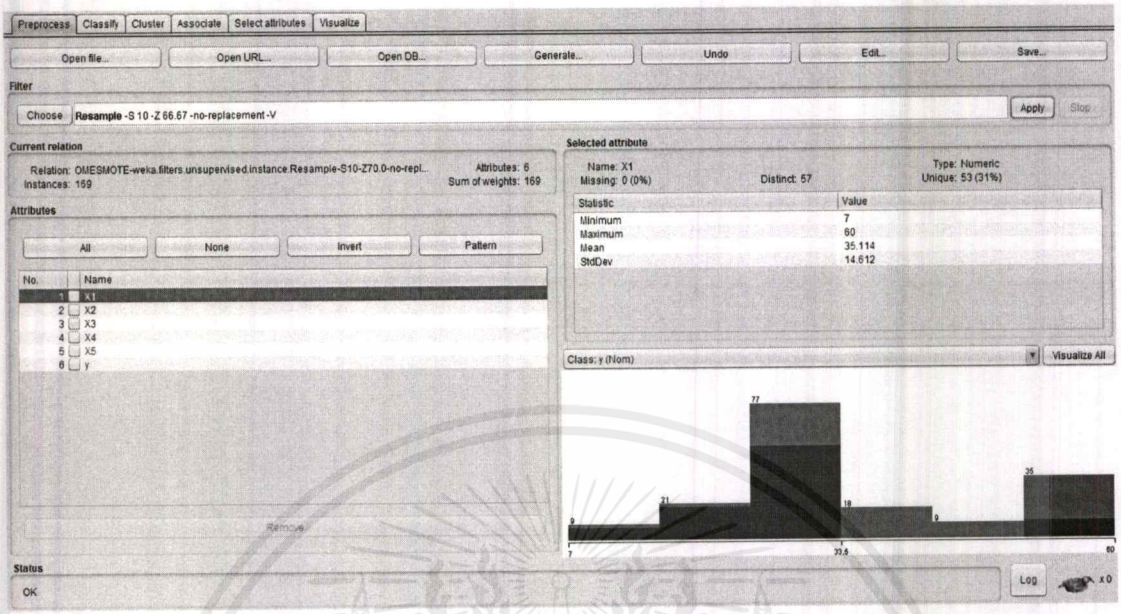
10. กดช่อง Resample - S 10 ⟶ ในช่อง invertSelection เป็น True ⟶ ในช่อง noReplacement เป็น True ⟶ กำหนดค่า randomSeed เท่ากับ 10 ⟶ ในช่อง sampleSizePercent กำหนดให้เป็น 66.67 เพื่อต้องการแบ่งข้อมูล 10% จาก 30%



รูปที่ ก-10 ขั้นตอนการแบ่งข้อมูล 10%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

11. กด Apply จะได้ข้อมูล 10% จำนวน 169 ค่า ⟶ กด Save กำหนดชื่อไฟล์เป็น OME SMOTE 10



รูปที่ ก-11 แสดงผลลัพธ์ข้อมูล 10%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข
การปรับความไม่สมดุลของข้อมูล

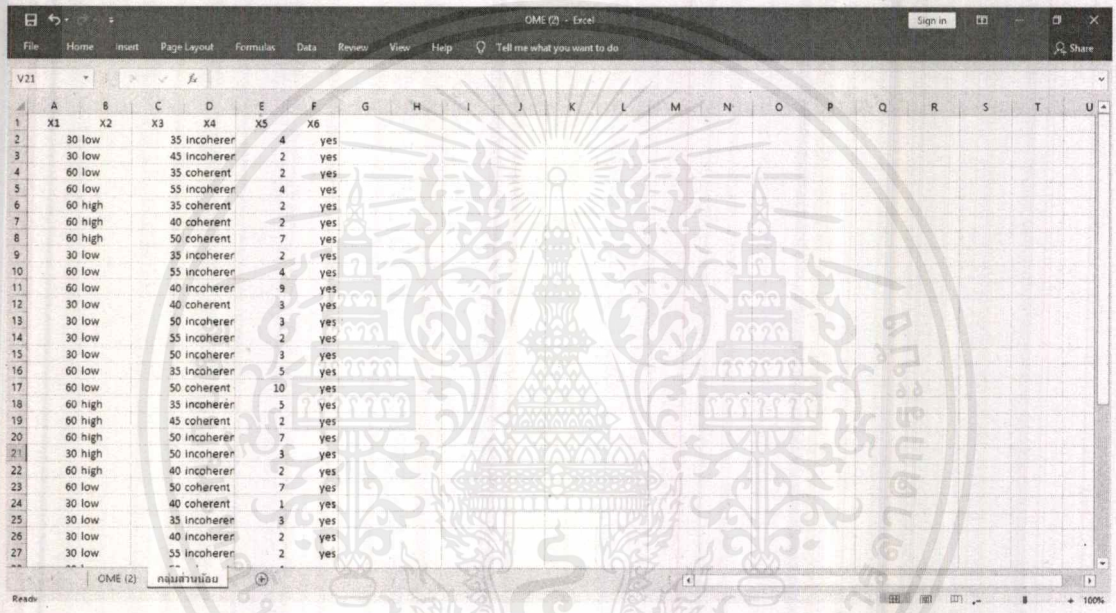
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชุดข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก (Auditory Perception in Children with OME)

จำนวนข้อมูลทั้งหมด 892 ค่า พบค่าในกลุ่มส่วนมาก (กลุ่ม No) 850 ค่า คิดเป็นร้อยละ 95.51 และค่าในกลุ่มส่วนน้อย (กลุ่ม Yes) 40 ค่า คิดเป็นร้อยละ 4.49

1. วิธีการสุ่มเกิน (Over Sampling)

ทำการสุ่มข้อมูลในกลุ่มส่วนน้อยเพื่อสร้างข้อมูลใหม่ ของกลุ่มส่วนน้อยให้มีจำนวนเพิ่มมากขึ้นให้ใกล้เคียงหรือเท่ากับจำนวนในกลุ่มส่วนมาก โดยจะใช้วิธีการสุ่มกลุ่มตัวอย่างอย่างง่าย (Simple random sampling) แบบวิธีการสุ่มกลุ่มตัวอย่างโดยวิธีใช้ตารางเลขสุ่ม (Table of Random Numbers) (สายชล, 2560)



	X1	X2	X3	X4	X5	X6
2	30	low	35	incoherer	4	yes
3	30	low	45	incoherer	2	yes
4	60	low	35	coherent	2	yes
5	60	low	55	incoherer	4	yes
6	60	high	35	coherent	2	yes
7	60	high	40	coherent	2	yes
8	60	high	50	coherent	7	yes
9	30	low	35	incoherer	2	yes
10	60	low	55	incoherer	4	yes
11	60	low	40	incoherer	9	yes
12	30	low	40	coherent	3	yes
13	30	low	50	incoherer	3	yes
14	30	low	55	incoherer	2	yes
15	30	low	50	incoherer	3	yes
16	60	low	35	incoherer	5	yes
17	60	low	50	coherent	10	yes
18	60	high	35	incoherer	5	yes
19	60	high	45	coherent	2	yes
20	60	high	50	incoherer	7	yes
21	30	high	50	incoherer	3	yes
22	60	high	40	incoherer	2	yes
23	60	low	50	coherent	7	yes
24	30	low	40	coherent	1	yes
25	30	low	35	incoherer	3	yes
26	30	low	40	incoherer	2	yes
27	30	low	55	incoherer	2	yes

รูปที่ ข-1 ข้อมูลกลุ่มส่วนน้อย 40 ค่า ที่ต้องนำไปทำการสุ่มเพิ่ม

1. กำหนดจุดเริ่มต้นในตารางเลขสุ่ม โดยหลับตาและปลายปากกาจิ้มตำแหน่งใดๆ บนตารางเลขสุ่ม สมมติว่า จุดเริ่มต้นคือ แถวที่1 คอลัมน์ที่ 25 ในที่นี้ คือ เลข 4 ดังแสดงในรูปที่ 23

52285	53301	71193	18991	34854	61701	10262	41876	19487	06996
70189	45193	46899	90746	97060	46547	64523	16987	60706	51116
83441	17072	50243	83300	63817	07510	05828	95271	07689	29757
16254	51933	02155	93543	20033	88132	16695	58878	39877	32928
84056	70489	14252	94132	04605	38293	60501	20415	82886	04396
01398	76800	59510	70789	59184	72725	81987	60820	67407	06777
46083	69602	70703	54693	85747	69453	40158	84787	65193	07982
74331	57826	36074	65258	56350	67475	10856	05061	54175	45490
25271	57349	62441	93647	49612	26541	48268	02745	50788	47621
39202	76311	95471	81348	27869	51539	78557	98949	37103	52406

รูปที่ ข-2 ตารางเลขสุ่ม

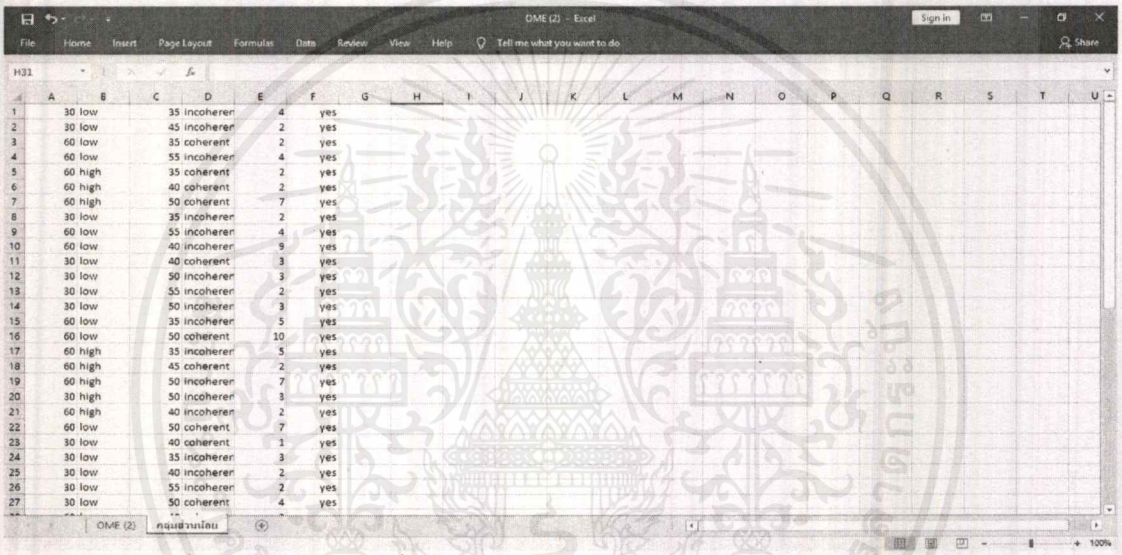
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ใช้จุดสุ่มเริ่มต้นที่ได้ในข้อที่ 1 อ่านเลขสุ่ม 2 หลักแรก จากซ้ายไปขวา เป็นเช่นนี้ไปเรื่อยๆ จนถึงจำนวนข้อมูลที่ต้องการ

3. ค่าเลขสุ่ม 2 หลักที่อ่าน เอาเฉพาะค่า 01 ถึง 40 (เนื่องจากข้อมูลในกลุ่มส่วนน้อยที่นำมาทำการสุ่มอย่างง่ายมี 40 ค่า)

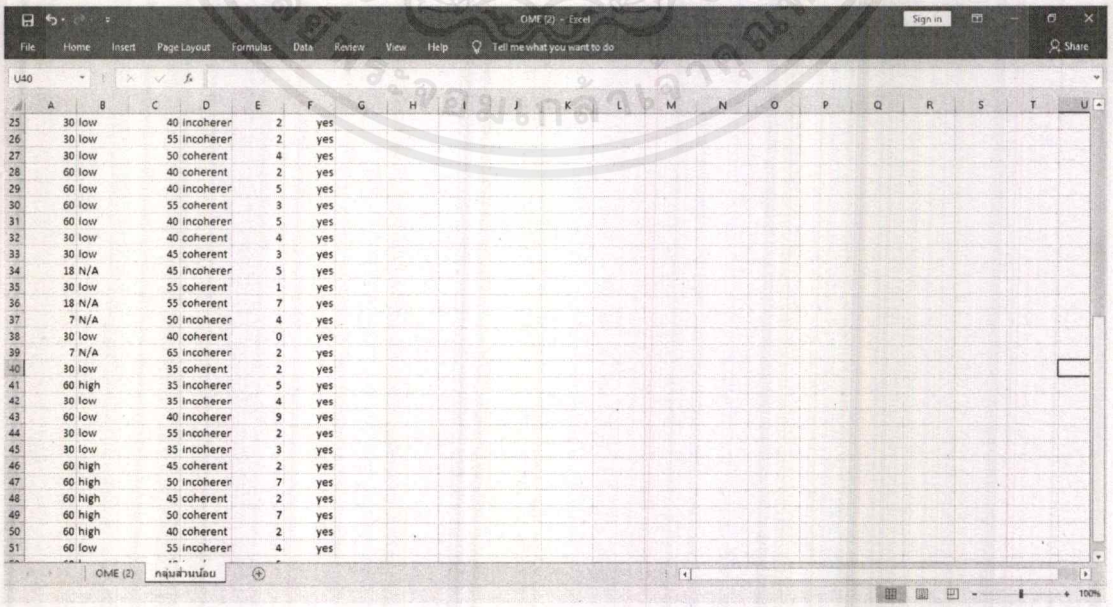
4. จากรูปที่ 23 เริ่มอ่านค่าจาก 46, 17, 01, 10, 26, 24, ... เนื่องจาก 46 เป็นค่าที่เกินจากจำนวนข้อมูลของกลุ่มส่วนน้อย 40 ค่า จึงไม่เอาและอ่านตัวถัดไปเรื่อยๆ จนถึง n = 800

5. นำค่าที่สุ่มได้มาเทียบกับข้อมูลในโปรแกรม Excel ยกตัวอย่างเช่น สุ่มได้เลขที่ 17 ซึ่งในโปรแกรม Excel คือแถวลำดับที่ 17 นำแถวลำดับนี้ไปต่อในชุดข้อมูลเดิม ในที่นี้คือต่อจากลำดับที่ 40 ซึ่งข้อมูลทีชสุ่มได้จะอยู่ในลำดับที่ 41 ทำแบบนี้ไปเรื่อย ๆ จนถึง n = 800



	A	B	C	D	E	F
1	30	low	35	incoherer	4	yes
2	30	low	45	incoherer	2	yes
3	60	low	35	coherent	2	yes
4	60	low	55	incoherer	4	yes
5	60	high	35	coherent	2	yes
6	60	high	40	coherent	2	yes
7	60	high	50	coherent	7	yes
8	30	low	35	incoherer	2	yes
9	60	low	55	incoherer	4	yes
10	60	low	40	incoherer	9	yes
11	30	low	40	coherent	3	yes
12	30	low	50	incoherer	3	yes
13	30	low	55	incoherer	2	yes
14	30	low	50	incoherer	3	yes
15	60	low	35	incoherer	5	yes
16	60	low	50	coherent	10	yes
17	60	high	35	incoherer	5	yes
18	60	high	45	coherent	2	yes
19	60	high	50	incoherer	7	yes
20	30	high	50	incoherer	3	yes
21	60	high	40	incoherer	2	yes
22	60	low	50	coherent	7	yes
23	30	low	40	coherent	1	yes
24	30	low	35	incoherer	3	yes
25	30	low	40	incoherer	2	yes
26	30	low	55	incoherer	2	yes
27	30	low	50	coherent	4	yes

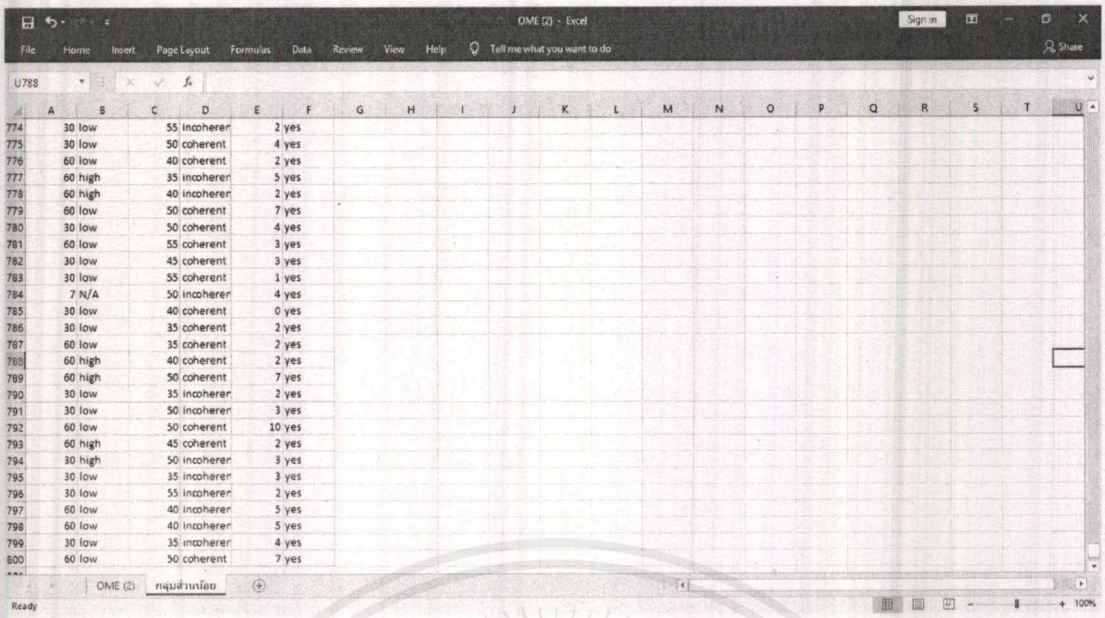
รูปที่ ข-3 ข้อมูลกลุ่มส่วนน้อย 40 ค่า



	A	B	C	D	E	F
25	30	low	40	incoherer	2	yes
26	30	low	55	incoherer	2	yes
27	30	low	50	coherent	4	yes
28	60	low	40	coherent	2	yes
29	60	low	40	incoherer	5	yes
30	60	low	55	coherent	3	yes
31	60	low	40	incoherer	5	yes
32	30	low	40	coherent	4	yes
33	30	low	45	coherent	3	yes
34	18	N/A	45	incoherer	5	yes
35	30	low	55	coherent	1	yes
36	18	N/A	55	coherent	7	yes
37	7	N/A	50	incoherer	4	yes
38	30	low	40	coherent	0	yes
39	7	N/A	65	incoherer	2	yes
40	30	low	35	coherent	2	yes
41	60	high	35	incoherer	5	yes
42	30	low	35	incoherer	4	yes
43	60	low	40	incoherer	9	yes
44	30	low	55	incoherer	2	yes
45	30	low	35	incoherer	3	yes
46	60	high	45	coherent	2	yes
47	60	high	50	incoherer	7	yes
48	60	high	45	coherent	2	yes
49	60	high	50	coherent	7	yes
50	60	high	40	coherent	2	yes
51	60	low	55	incoherer	4	yes

รูปที่ ข-4 ที่ได้จากการสุ่ม 800 ค่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



OME (2) - Excel

	A	B	C	D	E	F
774	30	low	55	incoherer	2	yes
775	30	low	50	coherent	4	yes
776	60	low	40	coherent	2	yes
777	60	high	35	incoherer	5	yes
778	60	high	40	incoherer	2	yes
779	60	low	50	coherent	7	yes
780	30	low	50	coherent	4	yes
781	60	low	55	coherent	3	yes
782	30	low	45	coherent	3	yes
783	30	low	55	coherent	1	yes
784	7	N/A	50	incoherer	4	yes
785	30	low	40	coherent	0	yes
786	30	low	35	coherent	2	yes
787	60	low	35	coherent	2	yes
788	60	high	40	coherent	2	yes
789	60	high	50	coherent	7	yes
790	30	low	35	incoherer	2	yes
791	30	low	50	incoherer	3	yes
792	60	low	50	coherent	10	yes
793	60	high	45	coherent	2	yes
794	30	high	50	incoherer	3	yes
795	30	low	35	incoherer	3	yes
796	30	low	55	incoherer	2	yes
797	60	low	40	incoherer	5	yes
798	60	low	40	incoherer	5	yes
799	30	low	35	incoherer	4	yes
800	60	low	50	coherent	7	yes

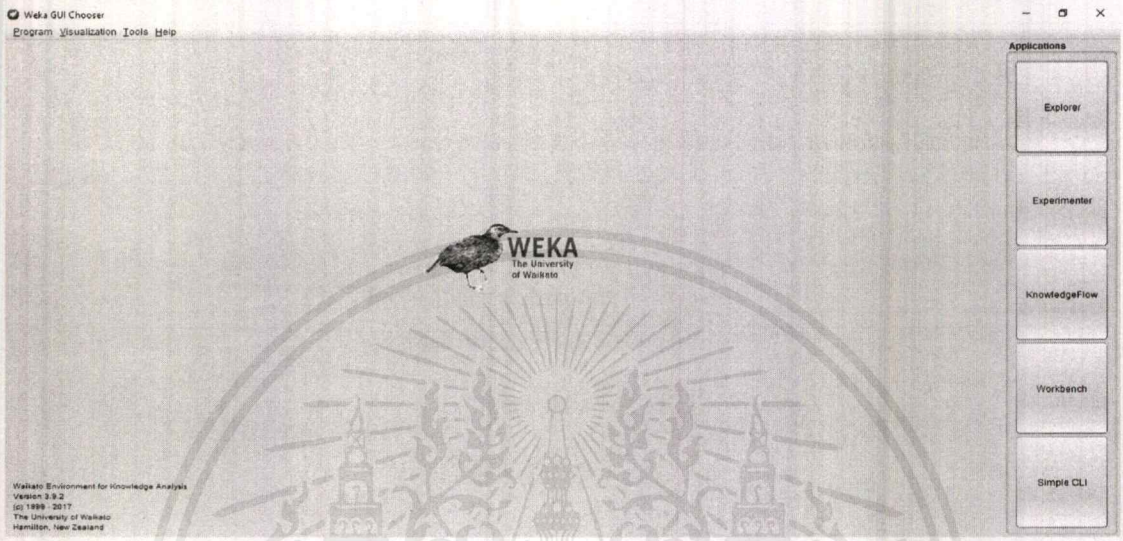
รูปที่ ข-4 ที่ได้จากการสุ่ม 800 ค่า (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. วิธีการสุ่มเกินโดยเทคนิค SMOTE (Synthetic Minority Over-Sampling Technique)

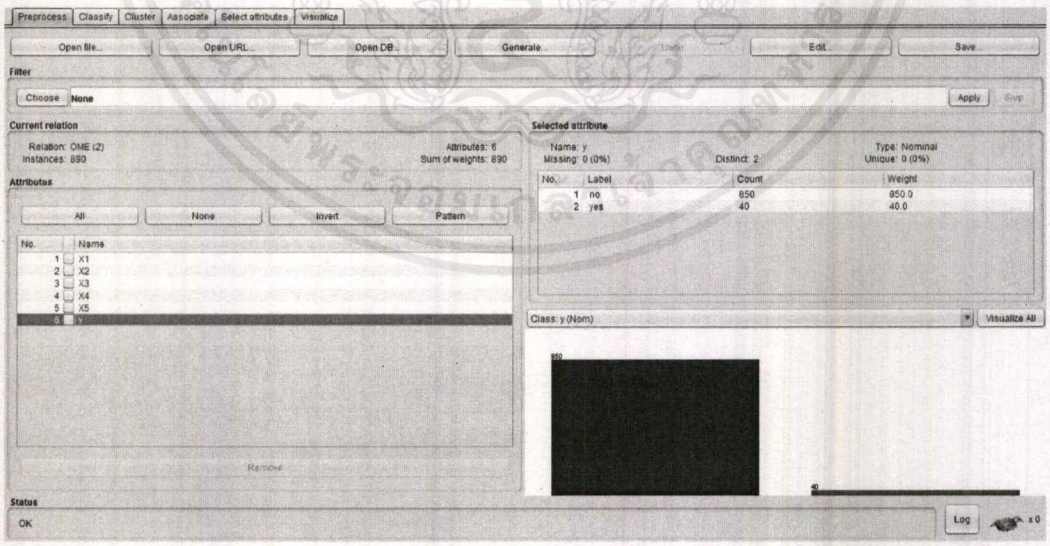
จะทำการสุ่มสร้างข้อมูลจากกลุ่มส่วนน้อยตามจำนวนที่กำหนด โดยการวัดระยะห่างจากจุดข้อมูลตัวอย่างไปยังจุดข้อมูลใกล้เคียง แล้วสุ่มสร้างข้อมูลสังเคราะห์ขึ้นให้ใกล้เคียงกับคลาสมาก

1. เปิดโปรแกรม WEKA แล้วกดปุ่ม Explorer



รูปที่ ข-5 หน้าต่างโปรแกรม WEKA

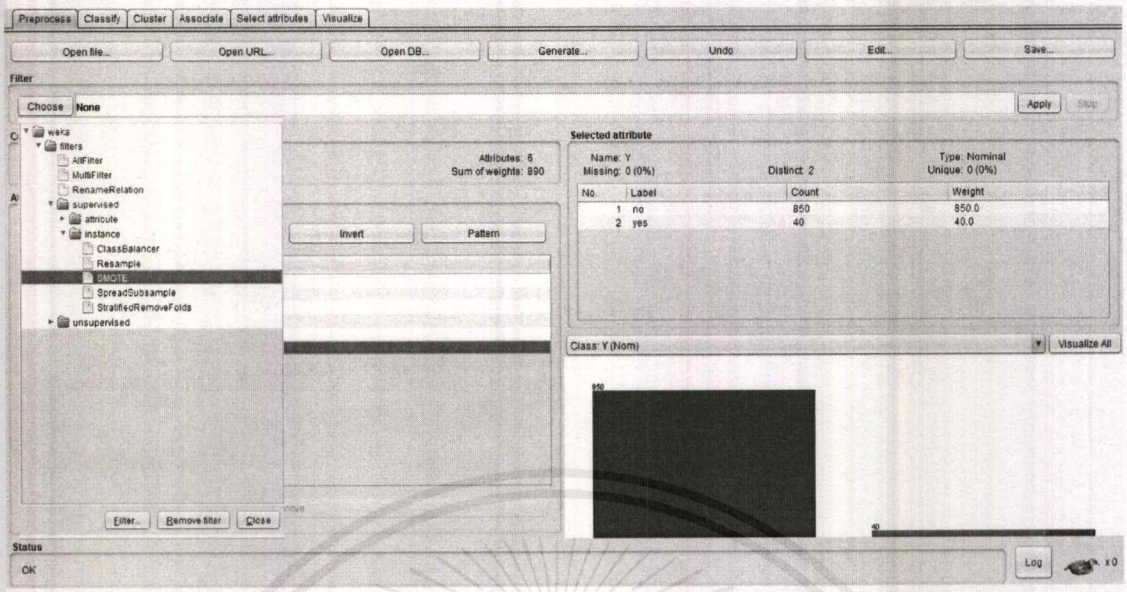
2. กดปุ่ม Open File (นามสกุล .csv) ไฟล์ข้อมูล จะได้ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก จำนวนข้อมูลทั้งหมด 890 ค่า พบค่าในกลุ่มส่วนมาก (Majority) 850 ค่า คิดเป็นร้อยละ 95.51 และค่าในกลุ่มส่วนน้อย (Minority) 40 ค่า คิดเป็นร้อยละ 4.49



รูปที่ ข-6 จำนวนข้อมูลไม่สมดุลในตัวแปรตาม Y

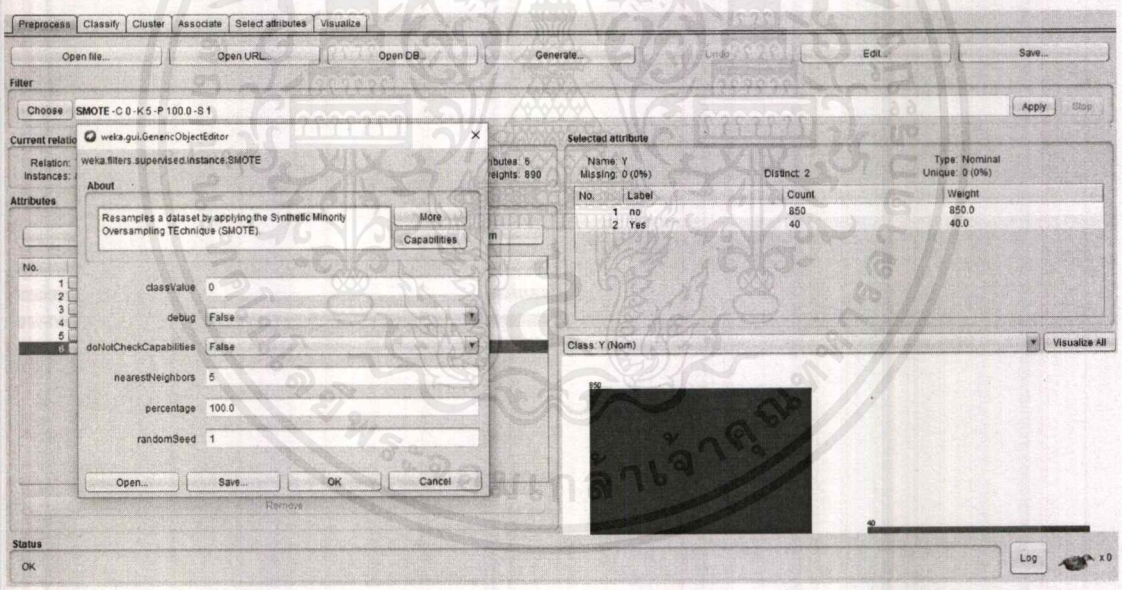
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. กดปุ่ม Choose ⟶ filters ⟶ supervised ⟶ instance ⟶ SMOTE



รูปที่ ข-7 ขั้นตอนการปรับข้อมูลไม่สมดุล การเลือกวิธีสุ่มเกินเทคนิค SMOTE

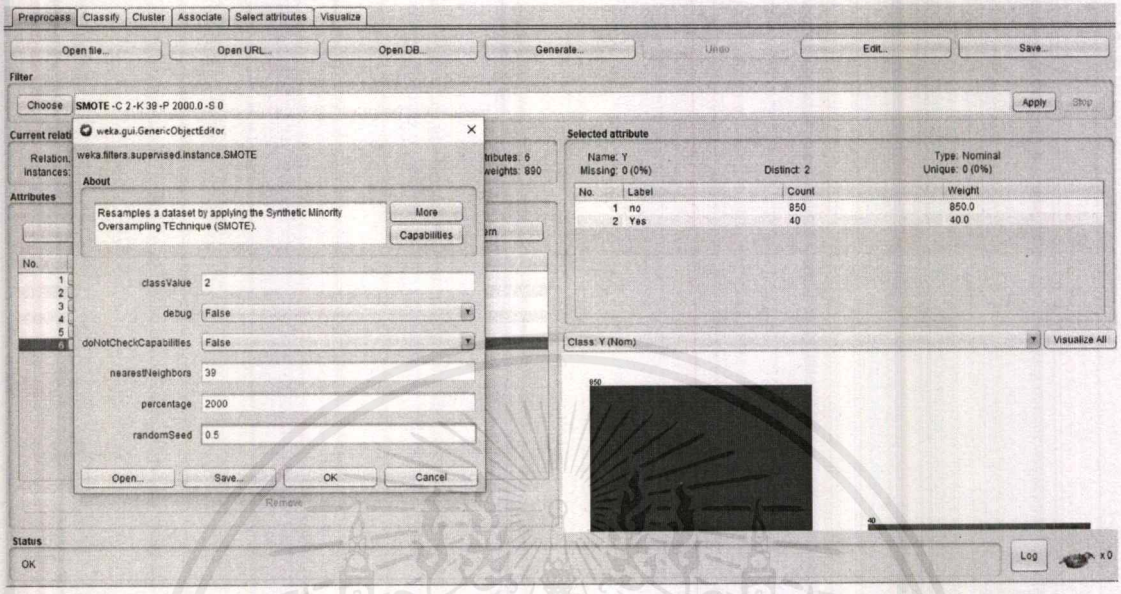
4. กดช่อง SMOTE -C 0 -K 5 -P 100.0 -S 1 เพื่อกำหนดค่าต่าง ๆ



รูปที่ ข-8 ขั้นตอนการปรับข้อมูลไม่สมดุล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. กำหนดค่าในช่อง classValue กำหนดค่าให้เป็น 2 ⟶ ในช่อง debug เป็น False ⟶ ในช่อง doNotCheckCapabilities เป็น False ⟶ กำหนดค่า nearestNeighbors เท่ากับ 39 ⟶ กำหนดค่า percentage เท่ากับ 2000 คือ 2000% ⟶ กำหนดค่า randomSeed เท่ากับ 0.5



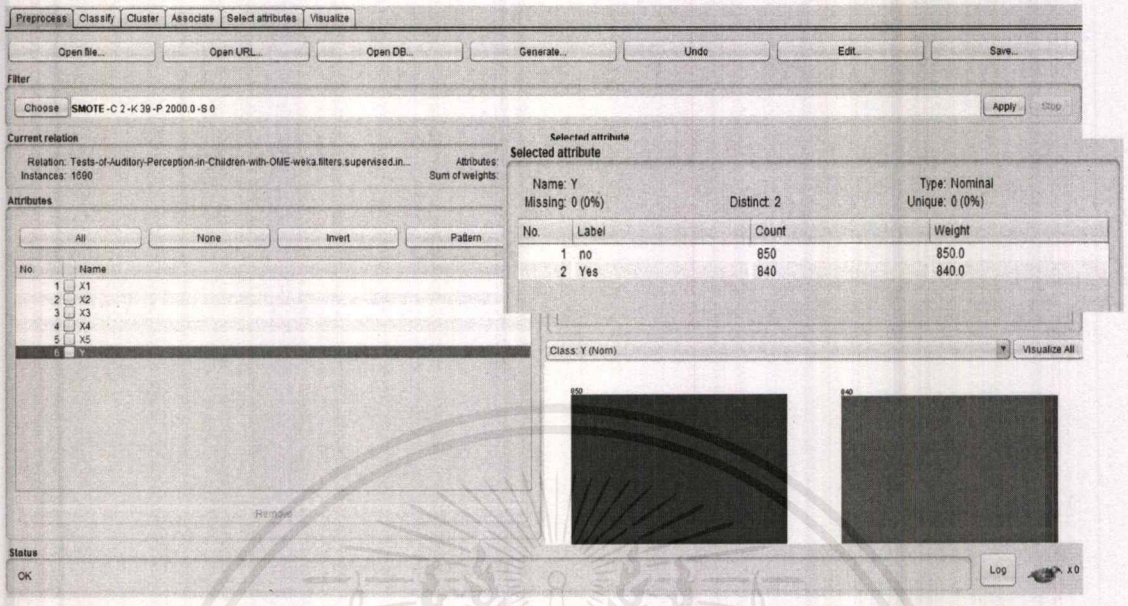
รูปที่ ข-9 กำหนดค่าต่าง ๆ ในขั้นตอนการปรับข้อมูลไม่สมดุล

<u>หมายเหตุ</u>

- ในช่อง classValue กำหนดค่าให้เป็น 2 เพราะกลุ่มส่วนน้อย (Minority) ที่ต้องการเพิ่มข้อมูลอยู่ใน No ที่ 2 Label ที่ Yes
- ในช่อง nearestNeighbors กำหนดค่าให้เป็น 39 เพราะจำนวนสมาชิกที่อยู่ใกล้กับข้อมูลของกลุ่มส่วนน้อย (Minority) ที่มี 40 ค่า คือ 39 ค่า
- ในช่อง percentage กำหนดค่าให้เป็น 2000 เพราะต้องการให้จำนวนข้อมูลกลุ่มส่วนน้อย (Minority) เพิ่มขึ้นให้เท่า ๆ กันกับกลุ่มส่วนมาก (Majority) ในการกด Apply ครั้งเดียว (จะกำหนดให้น้อยกว่านี้ก็ได้แต่ต้องกด Apply หลาย ๆ ครั้ง)
- ในช่อง randomSeed กำหนดค่าให้เป็น 0.5 เพราะเป็นการระบุค่าเริ่มต้นให้กับตัวแปรสมมติให้ค่าสุ่มตัวเลขระหว่าง 0 ถึง 1 ในที่นี้สมมติให้เป็น 1

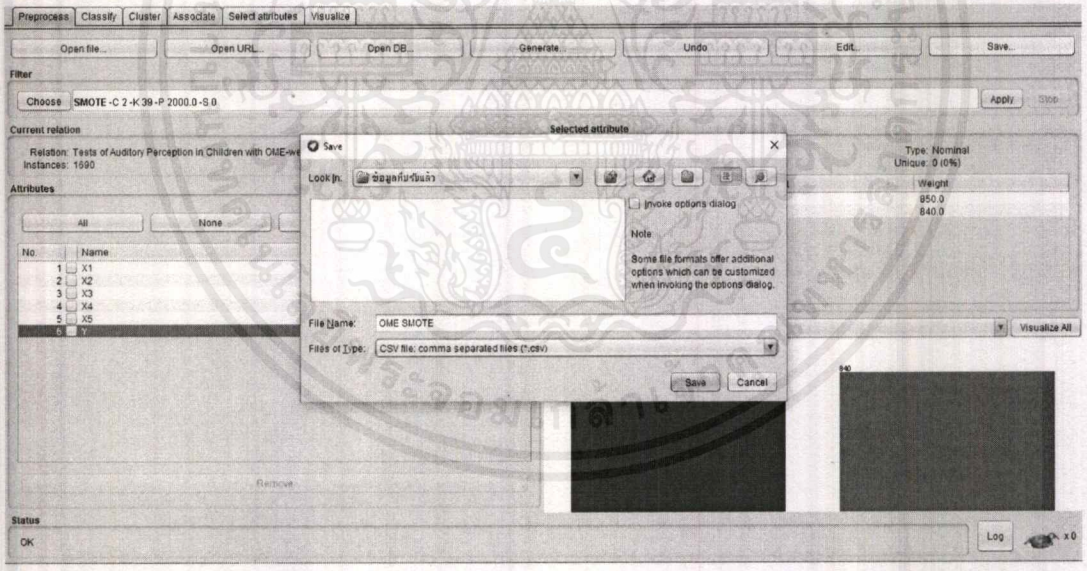
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6. กดปุ่ม Apply จะแสดงค่าข้อมูลที่ปรับแล้ว จำนวน 1690 ค่า พบค่าในกลุ่มส่วนมาก (Majority) 850 ค่า และค่าในกลุ่มส่วนน้อย (Minority) 840 ค่า



รูปที่ ข-10 ผลลัพธ์ข้อมูลที่ทำการปรับแล้ว

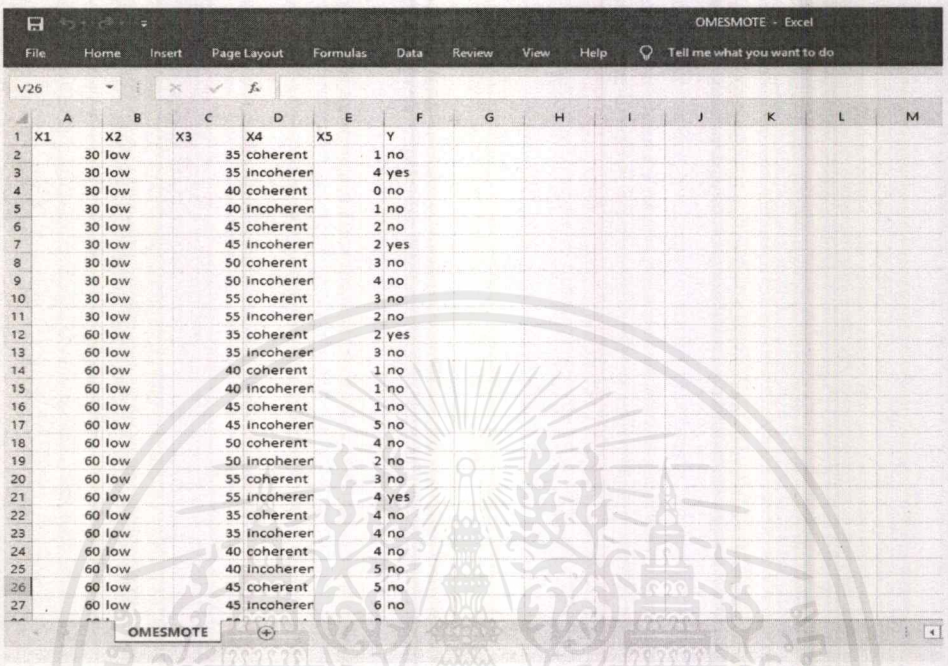
7. กดปุ่ม Save กำหนดชื่อไฟล์เป็น OME SMOTE



รูปที่ ข-11 ขั้นตอนการบันทึกไฟล์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

8. จากข้อมูลเดิม 890 ค่า หลังการทำการปรับโดยวิธี SMOTE ได้ข้อมูลเพิ่มอีก 800 ค่า ในกลุ่มส่วนน้อย (Minority) จาก 40 ค่า รวมเป็น 840 ค่า ทำให้จำนวนข้อมูลทั้งหมดเป็น 1690 ค่า แบ่งเป็นกลุ่มส่วนมาก (Majority) 850 ค่า คิดเป็นร้อยละ 50.30 กลุ่มส่วนน้อย (Minority) 840 ค่า คิดเป็นร้อยละ 49.70 ข้อมูลที่ได้มาใหม่เป็นข้อมูลที่ได้จากการปรับข้อมูลด้วยวิธี SMOTE ทำให้ชุดข้อมูลสมดุล



	X1	X2	X3	X4	X5	Y
2	30	low	35	coherent	1	no
3	30	low	35	incoherer	4	yes
4	30	low	40	coherent	0	no
5	30	low	40	incoherer	1	no
6	30	low	45	coherent	2	no
7	30	low	45	incoherer	2	yes
8	30	low	50	coherent	3	no
9	30	low	50	incoherer	4	no
10	30	low	55	coherent	3	no
11	30	low	55	incoherer	2	no
12	60	low	35	coherent	2	yes
13	60	low	35	incoherer	3	no
14	60	low	40	coherent	1	no
15	60	low	40	incoherer	1	no
16	60	low	45	coherent	1	no
17	60	low	45	incoherer	5	no
18	60	low	50	coherent	4	no
19	60	low	50	incoherer	2	no
20	60	low	55	coherent	3	no
21	60	low	55	incoherer	4	yes
22	60	low	35	coherent	4	no
23	60	low	35	incoherer	4	no
24	60	low	40	coherent	4	no
25	60	low	40	incoherer	5	no
26	60	low	45	coherent	5	no
27	60	low	45	incoherer	6	no

รูปที่ ข-12 ผลลัพธ์ที่ได้จากการปรับข้อมูลแล้ว

	A	B	C	D	E	F
1665	18.80467	low	57.02613	incoheren	2.856309	yes
1666	7	low	56.8904	incoheren	2.823065	yes
1667	9.801326	low	50.834	incoheren	2.624805	yes
1668	31.50147	low	64.74462	incoheren	2.82702	yes
1669	49.10514	low	56.29286	incoheren	2.562123	yes
1670	53.0307	low	45.95709	incoheren	3.005427	yes
1671	44.33098	low	56.37691	incoheren	3.181446	yes
1672	32.45989	low	44.70488	incoheren	3.884514	yes
1673	30	low	38.61566	incoheren	2.574976	yes
1674	30	low	51.19864	incoheren	2	yes
1675	30	low	35.3673	incoheren	2.40846	yes
1676	30	low	35	incoheren	2	yes
1677	9.548201	low	63.59716	incoheren	2	yes
1678	30	low	35	incoheren	3.017956	yes
1679	31.66449	low	48.22101	incoheren	2.721421	yes
1680	30	low	53.2615	incoheren	1.755723	yes
1681	30	low	40.56545	incoheren	2.752436	yes
1682	30	low	44.49703	incoheren	2.192829	yes
1683	30	low	53.73761	incoheren	1.127526	yes
1684	30	low	37.36315	incoheren	1.612352	yes
1685	30	low	44.21376	incoheren	2	yes
1686	52.69558	low	37.21894	incoheren	4.540861	yes
1687	40.7911	low	47.10152	incoheren	4.955426	yes
1688	59.34458	low	35	incoheren	2.484732	yes
1689	30	low	37.88253	incoheren	2.57348	yes
1690	30	low	43.28447	incoheren	2.261595	yes
1691	30	low	40.57319	incoheren	2.799426	yes

รูปที่ ข-12 ผลลัพธ์ที่ได้จากการปรับข้อมูลแล้ว (ต่อ)

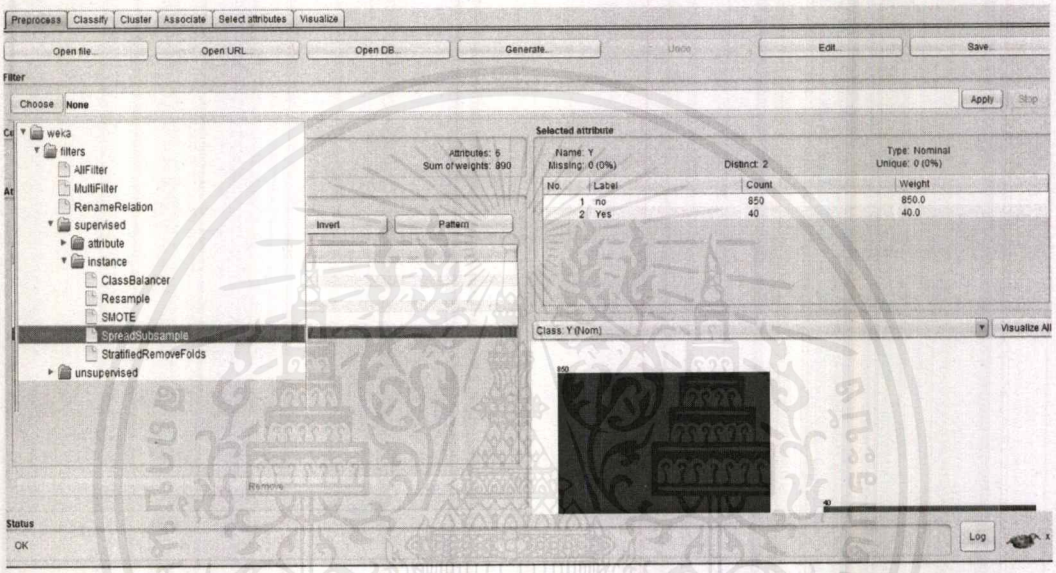
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. วิธีการสุ่มลด (Under Sampling)

การปรับข้อมูลให้มีความสมดุลด้วยวิธีการสุ่มลดจำนวนข้อมูลจากกลุ่มส่วนมากลงเพื่อทำให้จำนวนข้อมูลระหว่างกลุ่มส่วนมากและกลุ่มส่วนน้อยมีจำนวนใกล้เคียงกันมากขึ้น

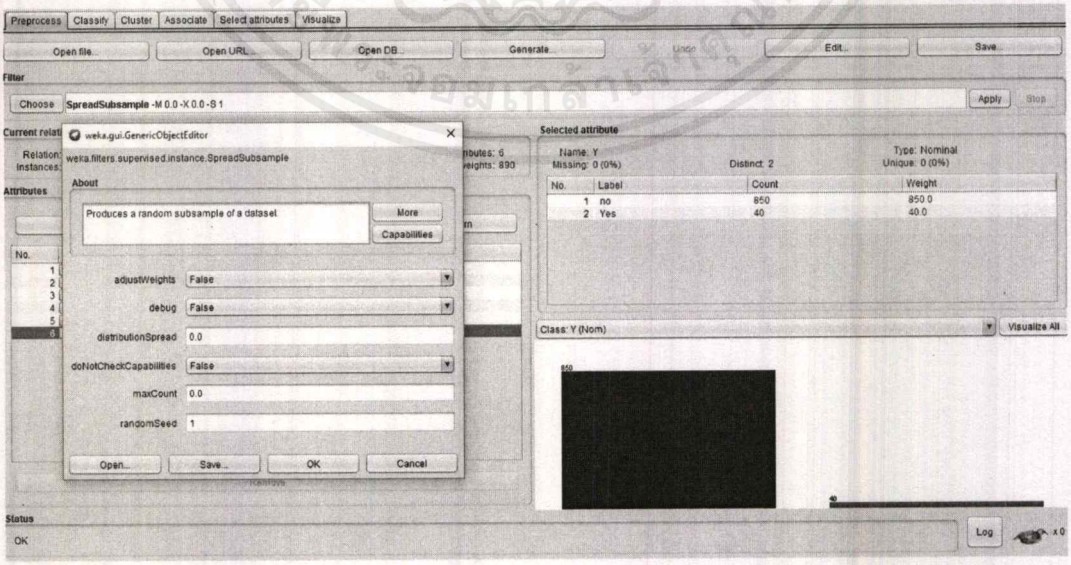
1. เปิด WEKA แล้วกดปุ่ม Explorer และกดปุ่ม Open File (นามสกุล .csv) ไฟล์ข้อมูล จะได้ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก จำนวนข้อมูลทั้งหมด 890 ค่า พบค่าในกลุ่มส่วนมาก (Majority) 850 ค่า คิดเป็นร้อยละ 95.51 และค่าในกลุ่มส่วนน้อย (Minority) 40 ค่า คิดเป็นร้อยละ 4.49 ดังรูปที่ ข-6

2. กดปุ่ม Choose ⟶ filters ⟶ supervised ⟶ instance ⟶ SpreadSubsample



รูปที่ ข-13 ขั้นตอนการปรับข้อมูลไม่สมดุล การเลือกวิธีสุ่มลด

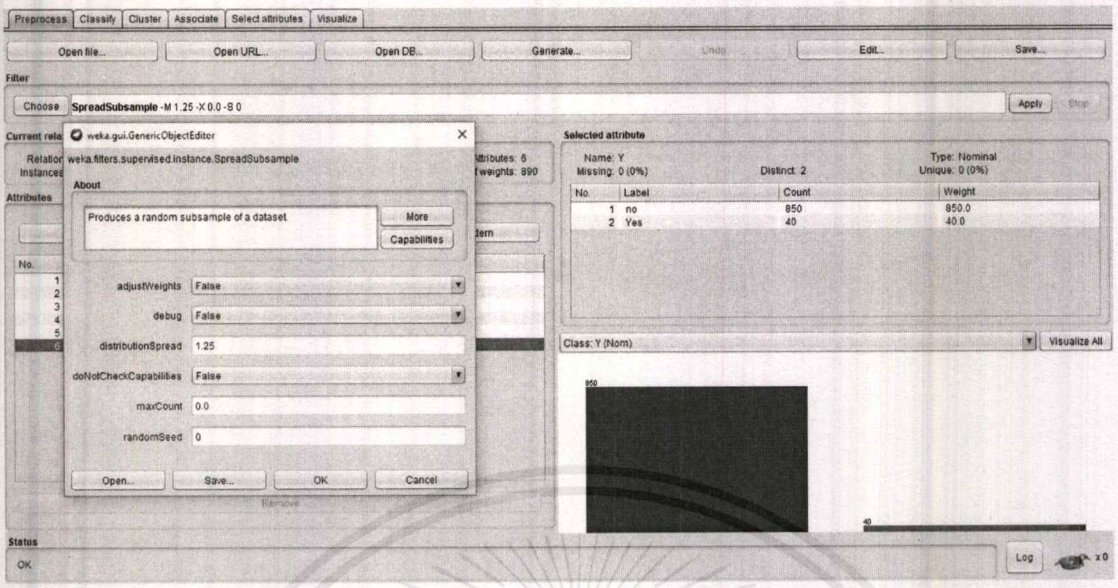
3. กดช่อง SpreadSubsample -M 0.0 -X 0.0 -S 1



รูปที่ ข-14 ขั้นตอนการปรับข้อมูลไม่สมดุล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. กำหนดค่าในช่อง distributionSpread เท่ากับ 1.25 ⟶ กำหนดค่า randomSeed เท่ากับ 0



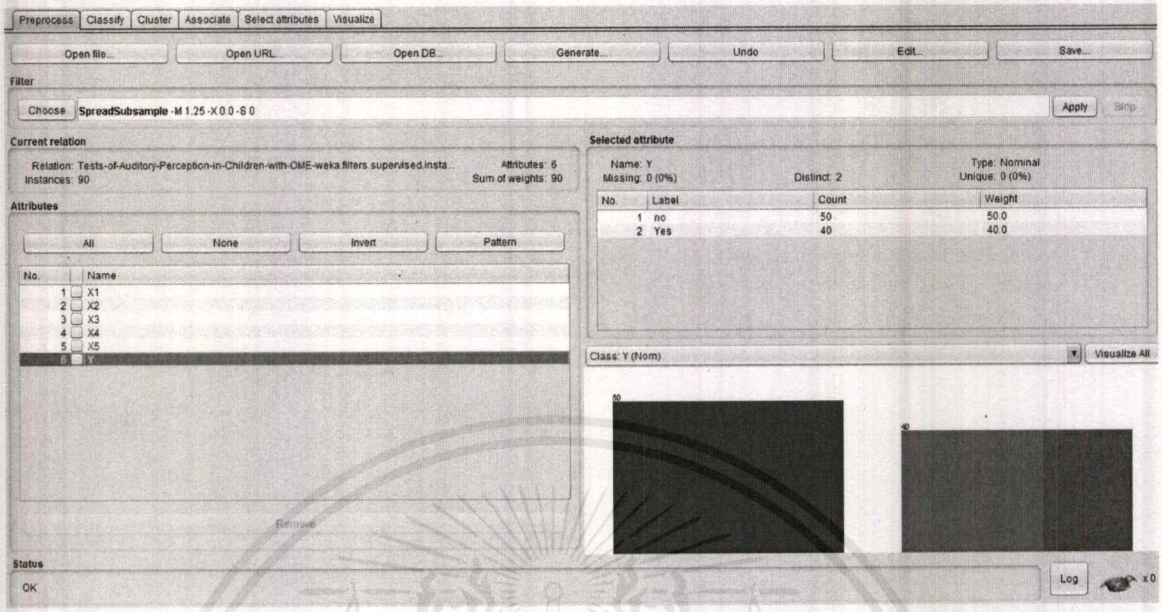
รูปที่ ข-15 ขั้นตอนการกำหนดค่าปรับข้อมูลไม่สมดุล

<u>หมายเหตุ</u>

- ในช่อง distributionSpread กำหนดค่าเป็น 1.25 เพราะเป็นการกระจายตัวของข้อมูลให้เกิดความสมดุลกัน หากกำหนดเป็น 0 ข้อมูลจะไม่มีการกระจายตัวและหากกำหนดเป็น 1ข้อมูลจะกระจายตัวเท่ากัน) จึงกำหนดให้เป็น 1.25 เพื่อให้เกิดการกระจายตัวที่ใกล้เคียงกัน
- ในช่อง randomSeed กำหนดค่าเป็น 0 เพราะเป็นการระบุค่าเริ่มต้นให้กับตัวแปร โดยให้ค่าอยู่ระหว่าง 0-1

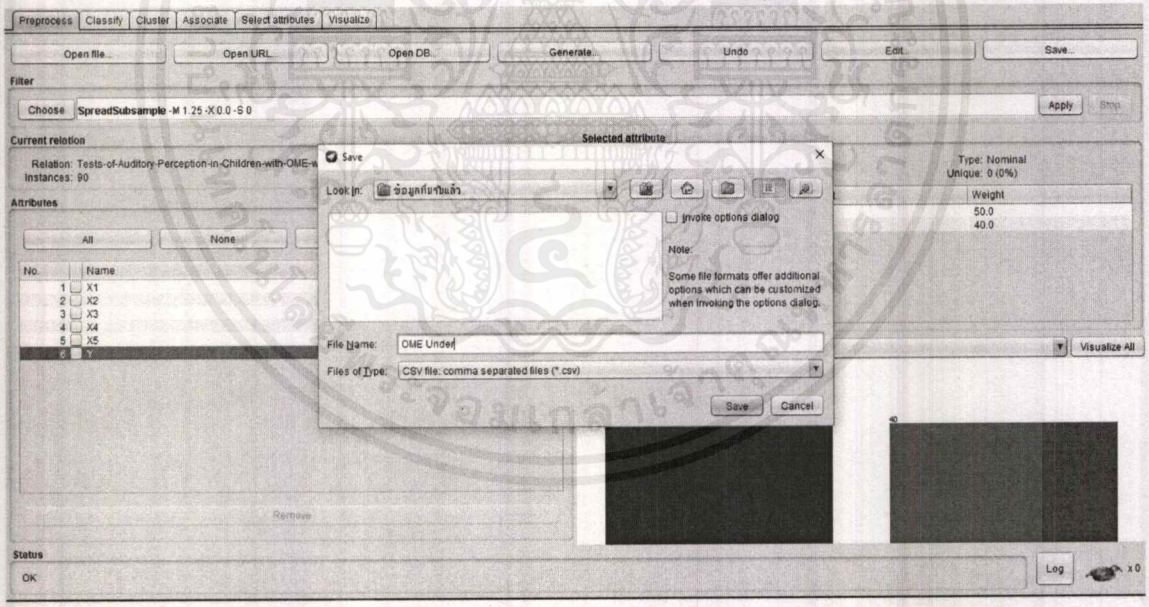
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. กดปุ่ม Apply จะแสดงค่าข้อมูลที่ปรับแล้ว จำนวนข้อมูลทั้งหมด 90 พบค่าในกลุ่มส่วนมาก (Majority) 50 ค่า และค่าในกลุ่มส่วนน้อย (Minority) 40 ค่า



รูปที่ ข-16 ผลลัพธ์ข้อมูลที่ทำการปรับแล้ว

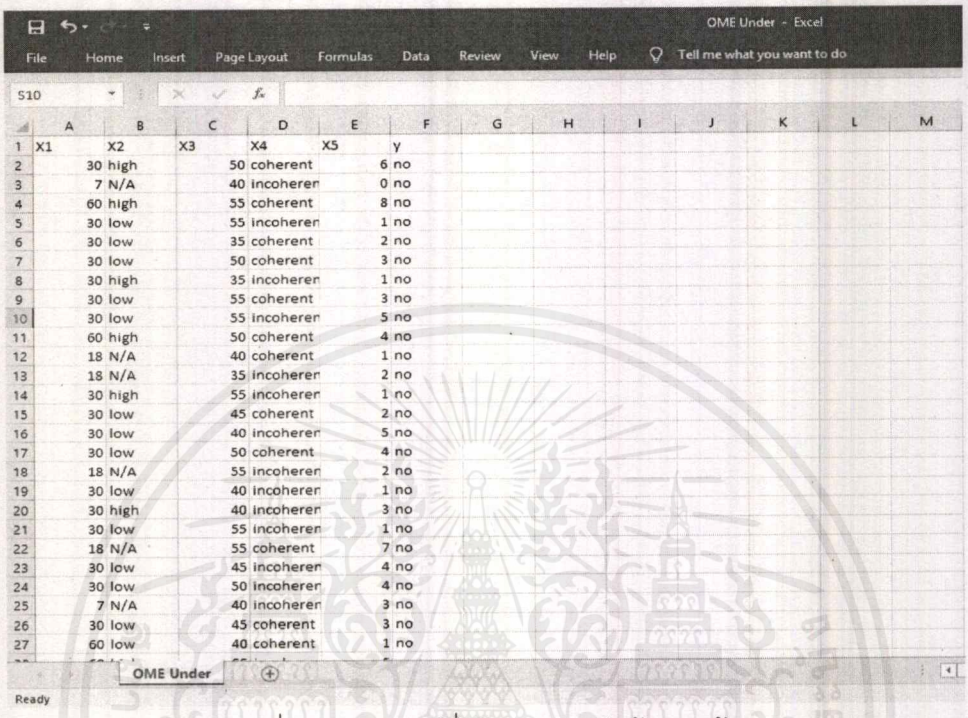
6. กดปุ่ม Save กำหนดชื่อไฟล์เป็น OME Under



รูปที่ ข-17 ขั้นตอนการบันทึกไฟล์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7. จากข้อมูลเดิม 890 ค่า หลังการทำการปรับโดยวิธี Under ข้อมูลในกลุ่มส่วนมาก (Majority) ลดลงเหลือ 50 ค่า จาก 850 ค่า ทำให้จำนวนข้อมูลทั้งหมดเป็น 90 ค่า แบ่งเป็นกลุ่มส่วนมาก (Majority) 50 ค่า คิดเป็นร้อยละ 55.56 กลุ่มส่วนน้อย (Minority) 40 ค่า คิดเป็นร้อยละ 44.44 ข้อมูลที่ได้มาใหม่เป็นข้อมูลที่ได้จากการปรับข้อมูลด้วยวิธี Under ทำให้ชุดข้อมูลสมดุล



OME Under - Excel

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do

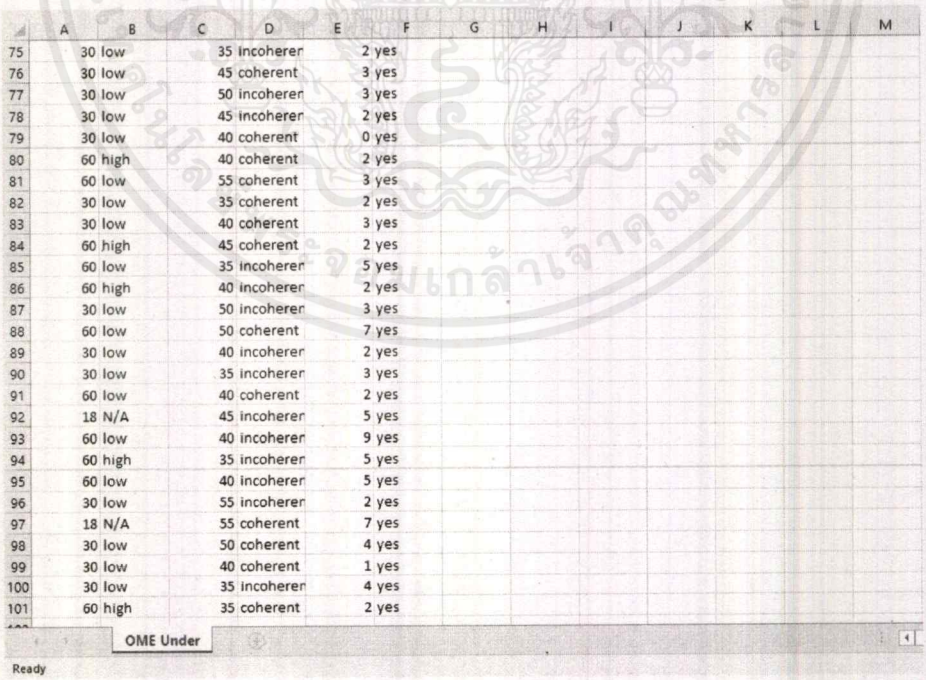
S10

	X1	X2	X3	X4	X5	y
2		30 high		50 coherent		6 no
3		7 N/A		40 incoheren		0 no
4		60 high		55 coherent		8 no
5		30 low		55 incoheren		1 no
6		30 low		35 coherent		2 no
7		30 low		50 coherent		3 no
8		30 high		35 incoheren		1 no
9		30 low		55 coherent		3 no
10		30 low		55 incoheren		5 no
11		60 high		50 coherent		4 no
12		18 N/A		40 coherent		1 no
13		18 N/A		35 incoheren		2 no
14		30 high		55 incoheren		1 no
15		30 low		45 coherent		2 no
16		30 low		40 incoheren		5 no
17		30 low		50 coherent		4 no
18		18 N/A		55 incoheren		2 no
19		30 low		40 incoheren		1 no
20		30 high		40 incoheren		3 no
21		30 low		55 incoheren		1 no
22		18 N/A		55 coherent		7 no
23		30 low		45 incoheren		4 no
24		30 low		50 incoheren		4 no
25		7 N/A		40 incoheren		3 no
26		30 low		45 coherent		3 no
27		60 low		40 coherent		1 no

OME Under

Ready

รูปที่ ข-18 ผลลัพธ์ที่ได้จากการปรับข้อมูลแล้ว



75	30 low	35 incoheren	2 yes		
76	30 low	45 coherent	3 yes		
77	30 low	50 incoheren	3 yes		
78	30 low	45 incoheren	2 yes		
79	30 low	40 coherent	0 yes		
80	60 high	40 coherent	2 yes		
81	60 low	55 coherent	3 yes		
82	30 low	35 coherent	2 yes		
83	30 low	40 coherent	3 yes		
84	60 high	45 coherent	2 yes		
85	60 low	35 incoheren	5 yes		
86	60 high	40 incoheren	2 yes		
87	30 low	50 incoheren	3 yes		
88	60 low	50 coherent	7 yes		
89	30 low	40 incoheren	2 yes		
90	30 low	35 incoheren	3 yes		
91	60 low	40 coherent	2 yes		
92	18 N/A	45 incoheren	5 yes		
93	60 low	40 incoheren	9 yes		
94	60 high	35 incoheren	5 yes		
95	60 low	40 incoheren	5 yes		
96	30 low	55 incoheren	2 yes		
97	18 N/A	55 coherent	7 yes		
98	30 low	50 coherent	4 yes		
99	30 low	40 coherent	1 yes		
100	30 low	35 incoheren	4 yes		
101	60 high	35 coherent	2 yes		

OME Under

Ready

รูปที่ ข-18 ผลลัพธ์ที่ได้จากการปรับข้อมูลแล้ว (ต่อ)

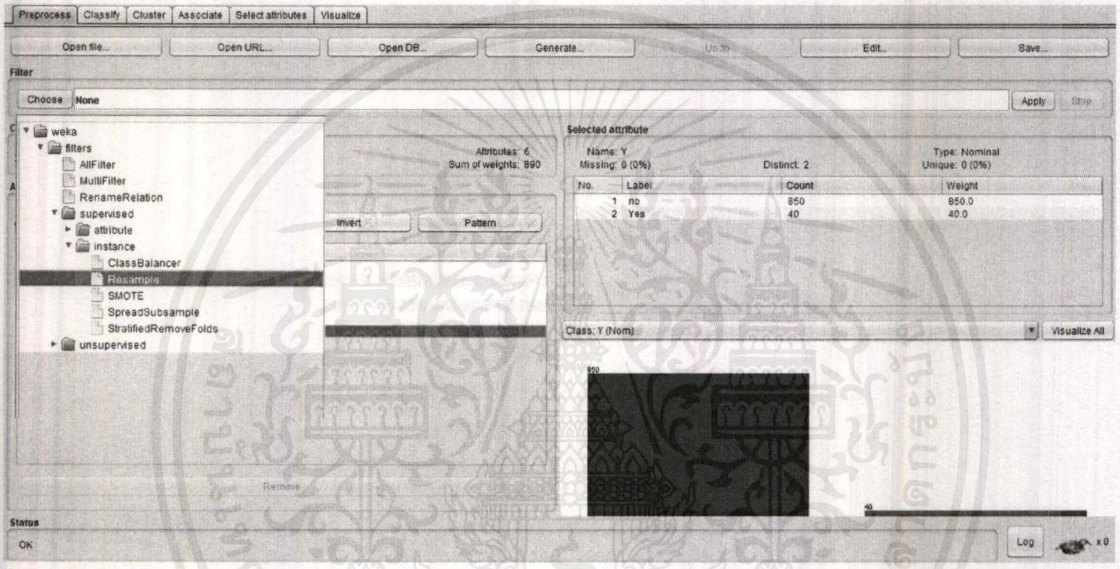
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. วิธีสุ่มผสมผสาน (Hybrid Methods)

เป็นวิธีการที่นำวิธีสุ่มเกินและวิธีสุ่มลดมาทำงานร่วมกันโดยการใช้วิธีนี้จะเป็นการสุ่มลดจำนวนข้อมูลจากกลุ่มส่วนมาก และทำการสุ่มเพิ่มข้อมูลในกลุ่มส่วนน้อย ให้จำนวนข้อมูลจากทั้งสองคลาสมีจำนวนใกล้เคียงกันหรือเท่ากัน

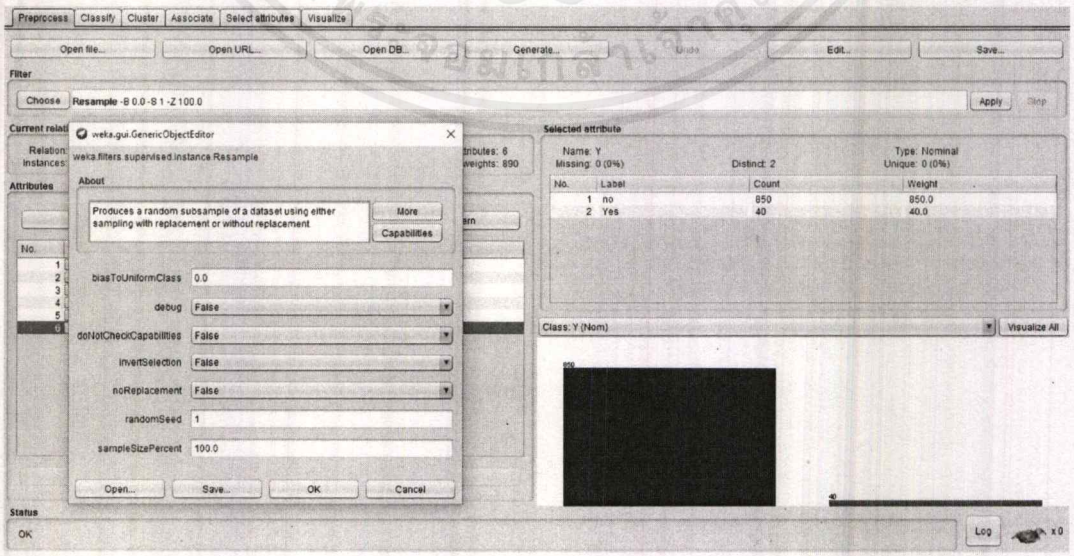
1. เปิด WEKA แล้วกดปุ่ม Explorer และกดปุ่ม Open File (นามสกุล .csv) ไฟล์ข้อมูล จะได้ข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก จำนวนข้อมูลทั้งหมด 890 ค่า พบค่าในกลุ่มส่วนมาก (Majority) 850 ค่า คิดเป็นร้อยละ 95.51 และค่าในกลุ่มส่วนน้อย (Minority) 40 ค่า คิดเป็นร้อยละ 4.49 ดังรูปที่ ข-6

2. กดปุ่ม Choose ⟶ filters ⟶ supervised ⟶ instance ⟶ Resample



รูปที่ ข-19 ขั้นตอนการปรับข้อมูลไม่สมดุล การเลือกวิธีสุ่มผสมผสาน

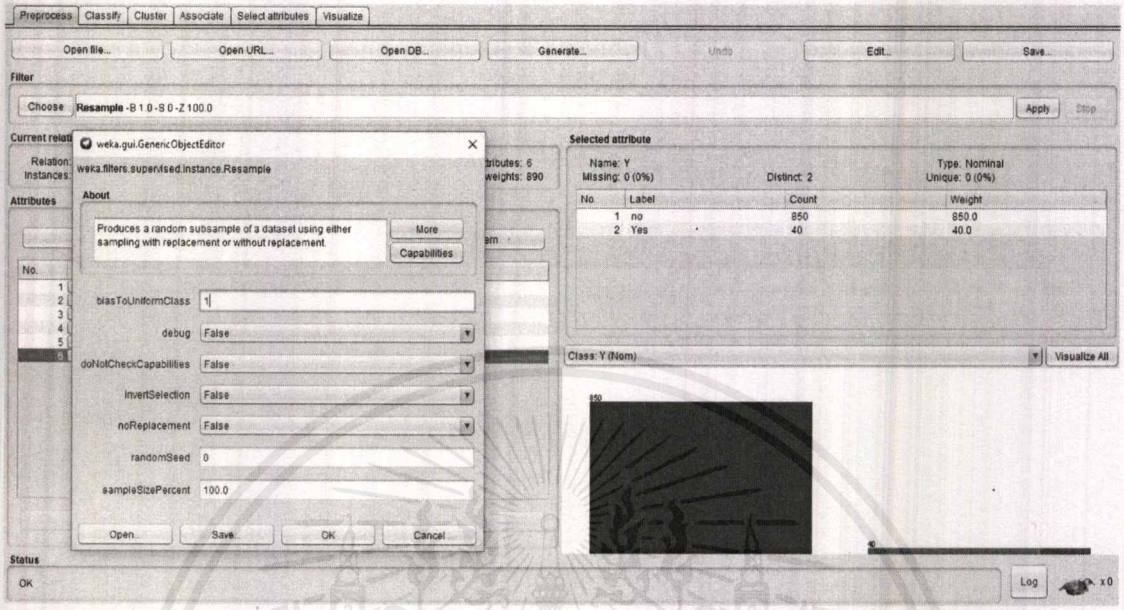
3. กดช่อง Resample - B 0.0 - S 1 -Z 100.0



รูปที่ ข-20 ขั้นตอนการปรับข้อมูลไม่สมดุล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. กำหนดค่าในช่อง invertSelection เป็น False ⟶ ในช่อง noReplacement เป็น False กำหนดค่า biasToUniformClass เท่ากับ 1 ⟶ กำหนดค่า randomSeed เท่ากับ 0 ⟶ ในช่อง sampleSizePercent กำหนดให้เป็น 100 คือ 100%



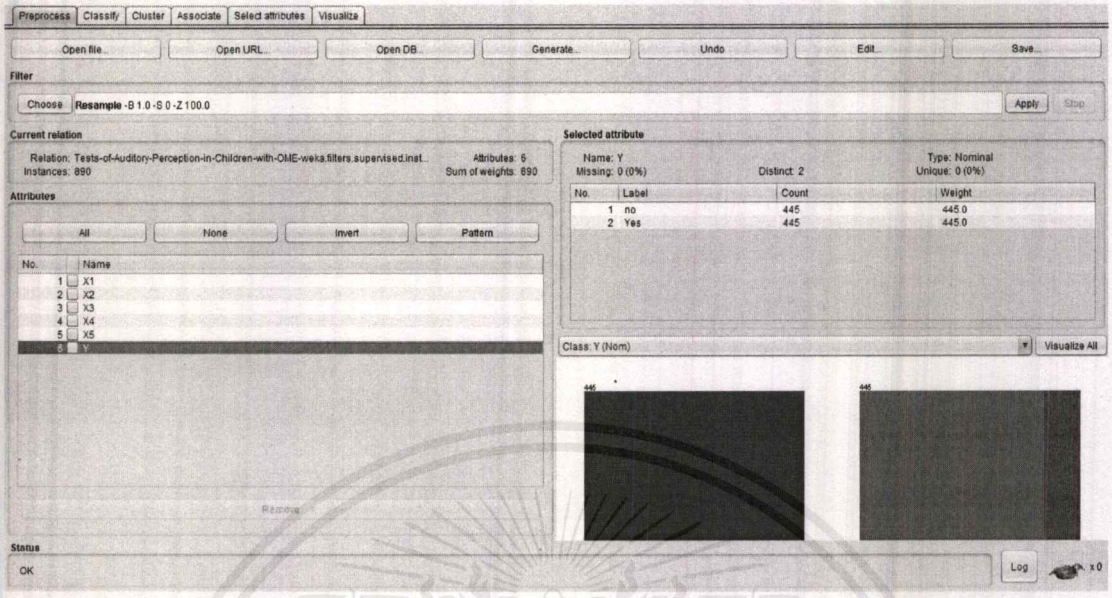
รูปที่ ข-21 ขั้นตอนการกำหนดค่าปรับข้อมูลไม่สมดุล

<u>หมายเหตุ</u>

- ในช่อง biasToUniformClass กำหนดค่าเป็น 1 เพราะเป็นการกำหนดความเอนเอียงของข้อมูล กำหนดเป็น 1 เพื่อให้ข้อมูลเท่ากัน
- ในช่อง randomSeed กำหนดค่าเป็น 0 เพราะเป็นการระบุค่าเริ่มต้นให้กับตัวแปร โดยให้ค่าอยู่ระหว่าง 0-1
- ในช่อง sampleSizePercent กำหนดค่าเป็น 100 เพราะเป็นการกำหนดขนาดของข้อมูล กำหนดเป็น 100 เพราะข้อมูลจะได้ไม่แตกต่างไปจากข้อมูลก่อนปรับมากนัก

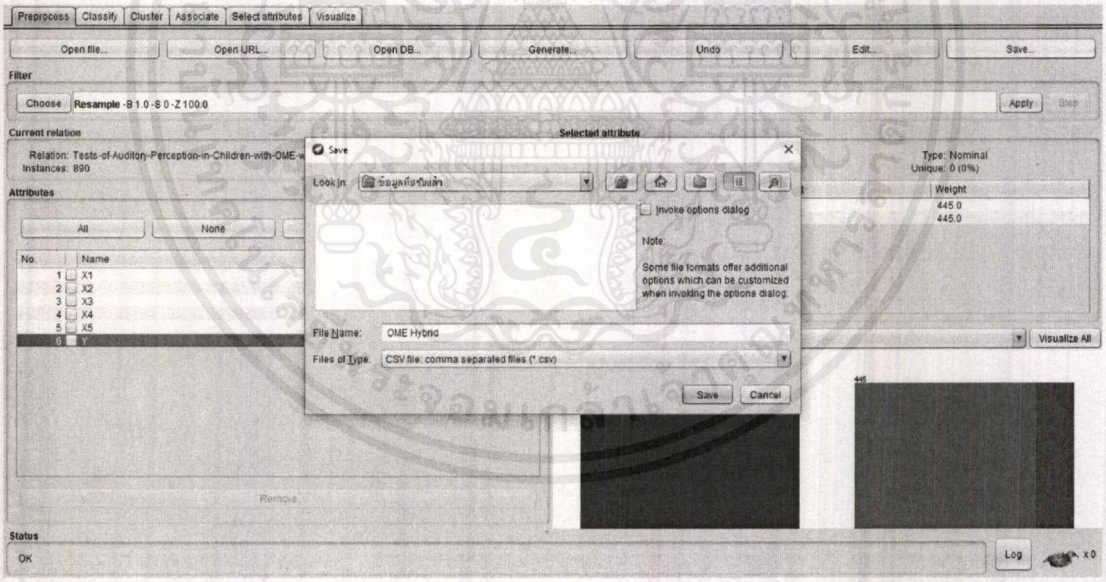
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. กดปุ่ม Apply จะแสดงค่าข้อมูลที่ปรับแล้ว จำนวน 890 ค่า พบค่าในกลุ่มส่วนมาก (Majority) 445 ค่า และค่าในกลุ่มส่วนน้อย (Minority) 445 ค่า



รูปที่ ข-22 ผลลัพธ์ข้อมูลที่ทำการปรับแล้ว

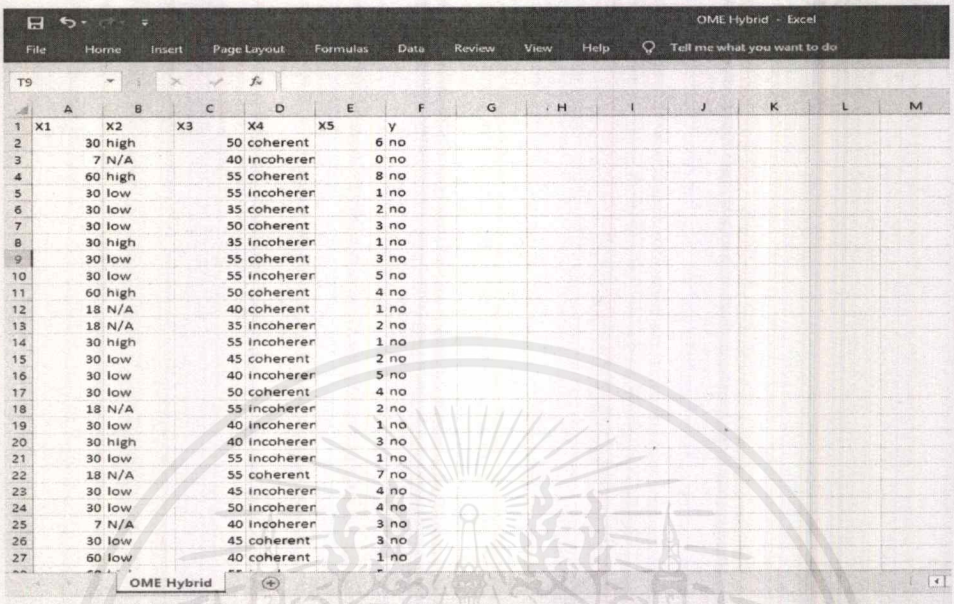
6. กดปุ่ม Save กำหนดชื่อไฟล์เป็น OME Hybrid



รูปที่ ข-23 ขั้นตอนการบันทึกไฟล์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7. จากข้อมูลเดิม 890 ค่า หลังการทำการปรับโดยวิธี Hybrid ข้อมูลในกลุ่มส่วนมาก (Majority) ลดลงเหลือ 445 ค่า จาก 850 ค่า ส่วนข้อมูลในกลุ่มส่วนน้อย (Minority) เพิ่มขึ้นเป็น 445 ค่า จาก 40 ค่า ข้อมูลที่ได้มาใหม่เป็นข้อมูลที่ได้จากการปรับข้อมูลด้วยวิธี Hybrid ทำให้ชุดข้อมูลสมดุล



	X1	X2	X3	X4	X5	y
2	30	high	50	coherent	6	no
3	7	N/A	40	incoheren	0	no
4	60	high	55	coherent	8	no
5	30	low	55	incoheren	1	no
6	30	low	35	coherent	2	no
7	30	low	50	coherent	3	no
8	30	high	35	incoheren	1	no
9	30	low	55	coherent	3	no
10	30	low	55	incoheren	5	no
11	60	high	50	coherent	4	no
12	18	N/A	40	coherent	1	no
13	18	N/A	35	incoheren	2	no
14	30	high	55	incoheren	1	no
15	30	low	45	coherent	2	no
16	30	low	40	incoheren	5	no
17	30	low	50	coherent	4	no
18	18	N/A	55	incoheren	2	no
19	30	low	40	incoheren	1	no
20	30	high	40	incoheren	3	no
21	30	low	55	incoheren	1	no
22	18	N/A	55	coherent	7	no
23	30	low	45	incoheren	4	no
24	30	low	50	incoheren	4	no
25	7	N/A	40	incoheren	3	no
26	30	low	45	coherent	3	no
27	60	low	40	coherent	1	no

รูปที่ ข-24 ผลลัพธ์ที่ได้จากการปรับข้อมูลแล้ว

	A	B	C	D	E	F
865	60	low	40	coherent	2	yes
866	7	N/A	50	incoheren	4	yes
867	60	high	50	coherent	7	yes
868	30	low	55	incoheren	2	yes
869	30	low	45	coherent	3	yes
870	60	low	55	incoheren	4	yes
871	60	low	50	coherent	7	yes
872	60	high	40	coherent	2	yes
873	60	high	35	incoheren	5	yes
874	60	high	35	incoheren	5	yes
875	60	low	50	coherent	10	yes
876	60	high	45	coherent	2	yes
877	60	high	50	coherent	7	yes
878	60	high	50	incoheren	7	yes
879	60	low	50	coherent	7	yes
880	30	low	40	coherent	0	yes
881	30	low	45	coherent	3	yes
882	30	low	55	coherent	1	yes
883	60	low	55	incoheren	4	yes
884	60	high	40	incoheren	2	yes
885	60	low	50	coherent	7	yes
886	30	high	50	incoheren	3	yes
887	7	N/A	65	incoheren	2	yes
888	60	low	55	coherent	3	yes
889	30	low	35	incoheren	3	yes
890	30	low	45	incoheren	2	yes
891	30	low	35	incoheren	3	yes

รูปที่ ข-24 ผลลัพธ์ที่ได้จากการปรับข้อมูลแล้ว (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

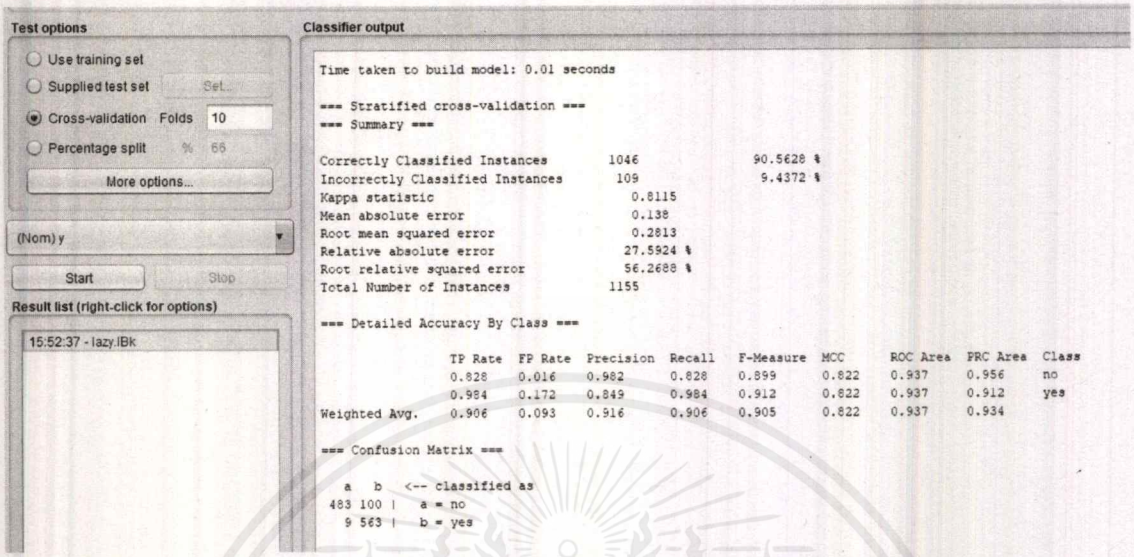
ภาคผนวก ค

การวิเคราะห์ข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

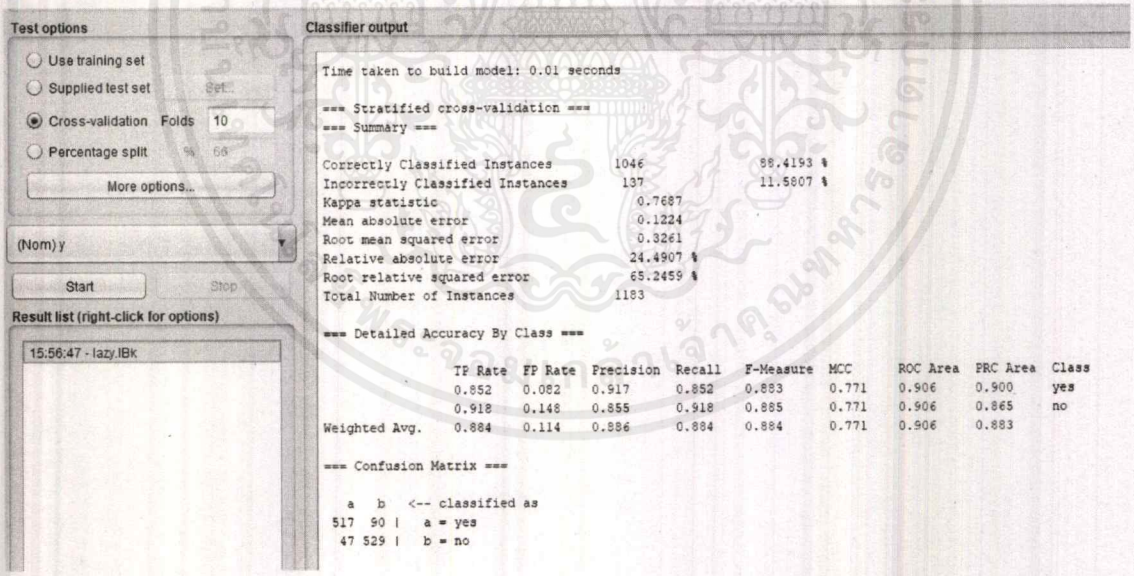
1. วิธีเพื่อนบ้านใกล้สุด k ตัว

1.1 การปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกิน



รูปที่ ค-1 ผลการวิเคราะห์ข้อมูลสำหรับการสร้างตัวแบบวิธีเพื่อนบ้านใกล้สุด k ตัว ด้วยการการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกิน

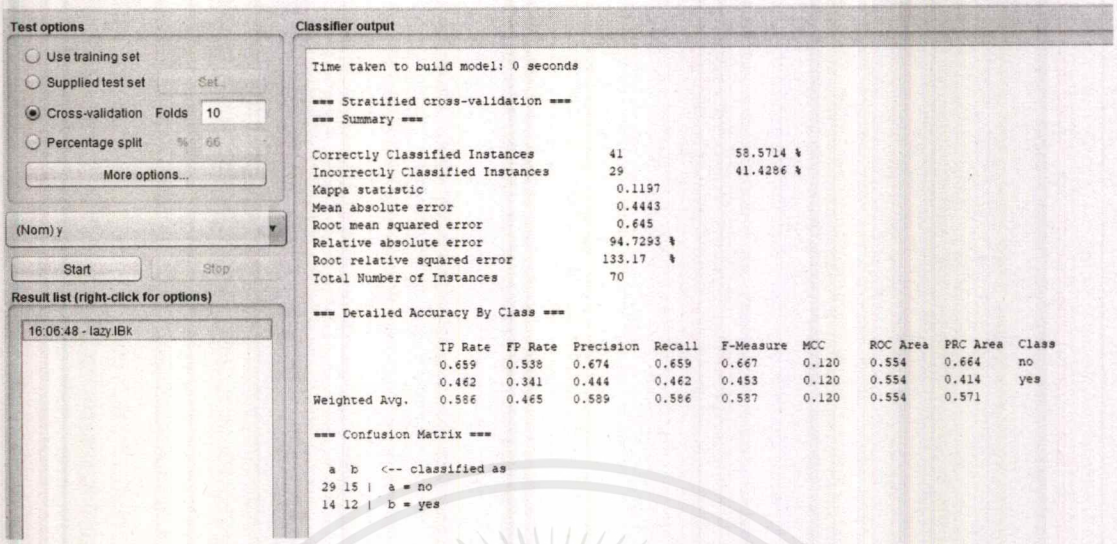
1.2 การปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE



รูปที่ ค-2 ผลการวิเคราะห์ข้อมูลสำหรับการสร้างตัวแบบวิธีเพื่อนบ้านใกล้สุด k ตัว ด้วยการการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3 การปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลด



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) y

Start Stop

Result list (right-click for options)

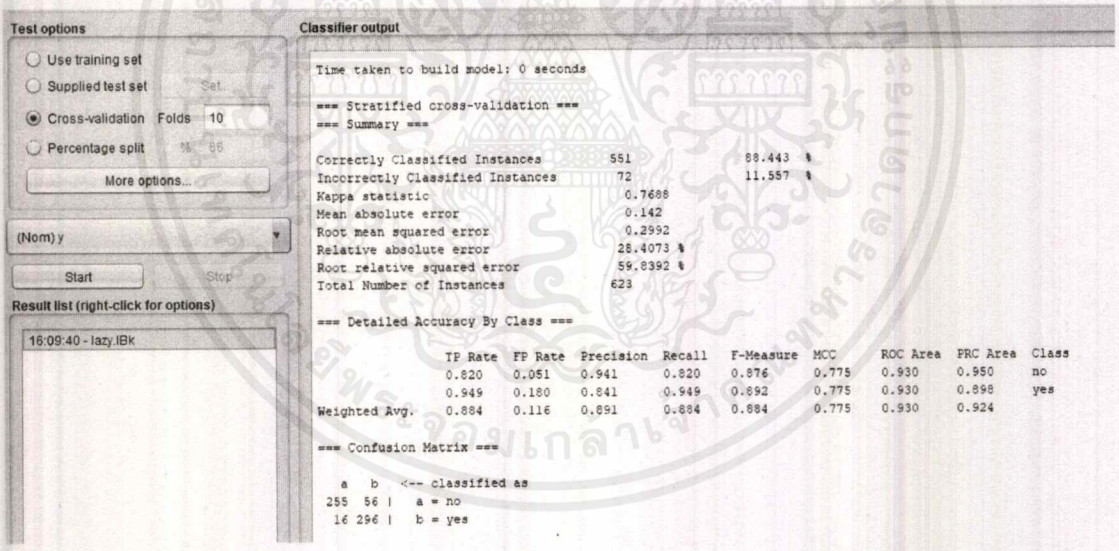
16:06:48 - lazy.IBk

Classifier output

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          41               58.5714 %
Incorrectly Classified Instances        29               41.4286 %
Kappa statistic                          0.1197
Mean absolute error                      0.4443
Root mean squared error                  0.645
Relative absolute error                 94.7293 %
Root relative squared error            133.17   %
Total Number of Instances               70

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.659    0.538    0.674      0.659   0.667      0.120    0.554     0.664     no
                 0.462    0.341    0.444      0.462   0.453      0.120    0.554     0.414     yes
Weighted Avg.    0.586    0.465    0.589      0.586   0.587      0.120    0.554     0.571

=== Confusion Matrix ===

  a  b   <-- classified as
 29 15 |  a = no
 14 12 |  b = yes
```

รูปที่ ค-3 ผลการวิเคราะห์ข้อมูลสำหรับการสร้างตัวแบบวิธีเพื่อนบ้านใกล้สุด k ตัว ด้วยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลด

1.4 การปรับความไม่สมดุลของข้อมูลด้วยวิธีการผสมผสาน



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) y

Start Stop

Result list (right-click for options)

16:09:40 - lazy.IBk

Classifier output

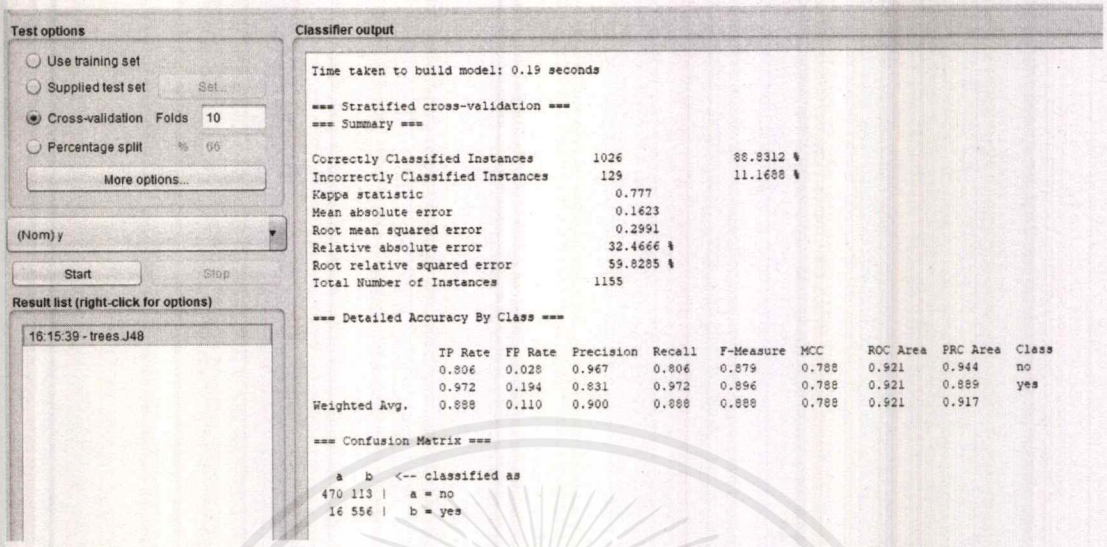
```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         551               88.443  %
Incorrectly Classified Instances        72               11.557  %
Kappa statistic                          0.7688
Mean absolute error                      0.142
Root mean squared error                  0.2992
Relative absolute error                 28.4073 %
Root relative squared error             59.8392 %
Total Number of Instances               623

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.820    0.051    0.941      0.820   0.876      0.775    0.930     0.950     no
                 0.949    0.180    0.841      0.949   0.892      0.775    0.930     0.898     yes
Weighted Avg.    0.884    0.116    0.891      0.884   0.884      0.775    0.930     0.924

=== Confusion Matrix ===

   a   b   <-- classified as
 255  56 |   a = no
  16 296 |   b = yes
```

รูปที่ ข-4 ผลการวิเคราะห์ข้อมูลสำหรับการสร้างตัวแบบวิธีเพื่อนบ้านใกล้สุด k ตัว ด้วยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการผสมผสาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. วิธีต้นไม้ตัดสินใจ

2.1 การปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกิน



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) y

Start Stop

Result list (right-click for options)

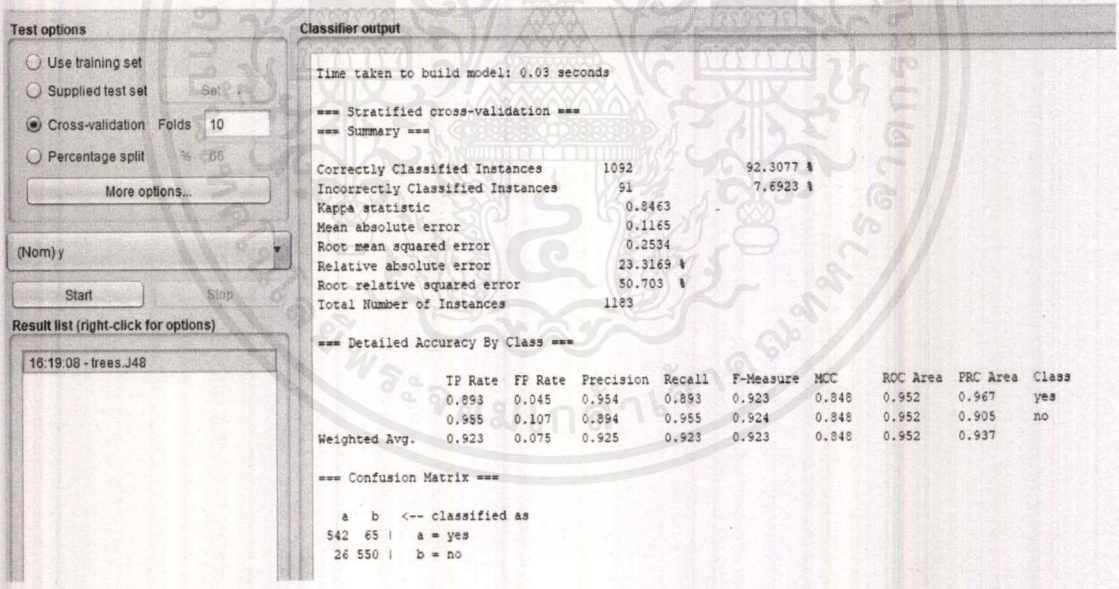
16:15:39 - trees.J48

Classifier output

```
Time taken to build model: 0.19 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1026               88.8312 %
Incorrectly Classified Instances       129               11.1688 %
Kappa statistic                          0.777
Mean absolute error                      0.1623
Root mean squared error                  0.2991
Relative absolute error                 32.4666 %
Root relative squared error             59.8285 %
Total Number of Instances             1155

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.806    0.028    0.967      0.806   0.879      0.788    0.921     0.944     no
                 0.972    0.194    0.831      0.972   0.896      0.788    0.921     0.889     yes
Weighted Avg.    0.888    0.110    0.900      0.888   0.888      0.788    0.921     0.917

=== Confusion Matrix ===

   a   b   <-- classified as
 470 113 |   a = no
  16 556 |   b = yes
```

รูปที่ ค-5 ผลการวิเคราะห์ข้อมูลสำหรับการสร้างตัวแบบวิธีต้นไม้ตัดสินใจ
ด้วยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกิน

2.2 การปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) y

Start Stop

Result list (right-click for options)

16:19:08 - trees.J48

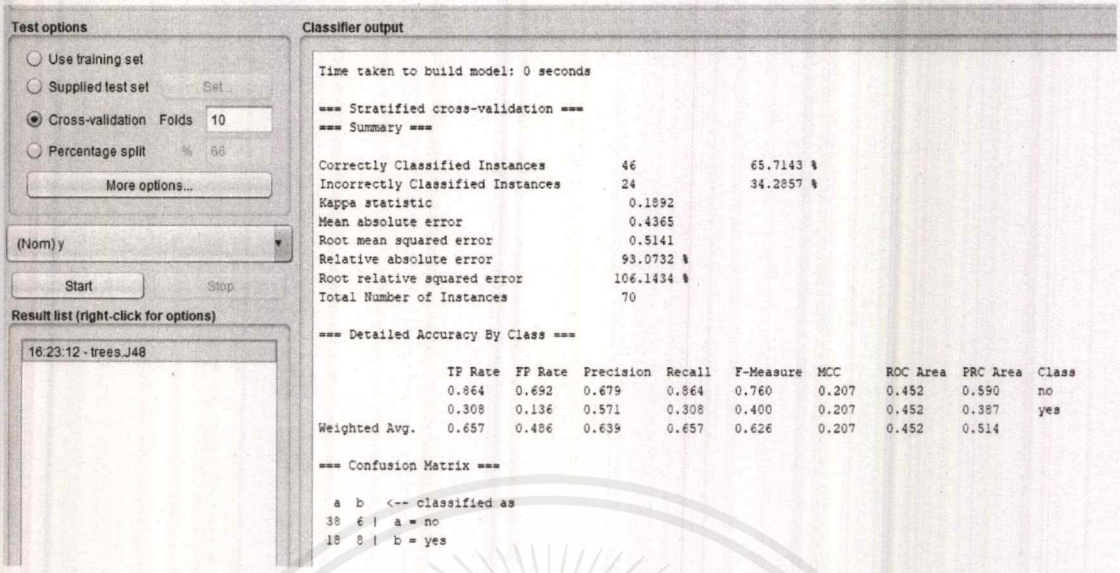
Classifier output

```
Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1092               92.3077 %
Incorrectly Classified Instances        91                7.6923 %
Kappa statistic                          0.8463
Mean absolute error                      0.1165
Root mean squared error                  0.2534
Relative absolute error                 23.3169 %
Root relative squared error             50.703  %
Total Number of Instances             1183

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.893    0.045    0.954      0.893   0.923      0.848    0.952     0.967     yes
                 0.955    0.107    0.894      0.955   0.924      0.848    0.952     0.905     no
Weighted Avg.    0.923    0.075    0.925      0.923   0.923      0.848    0.952     0.937

=== Confusion Matrix ===

   a   b   <-- classified as
 542  65 |   a = yes
  26 550 |   b = no
```

รูปที่ ค-6 ผลการวิเคราะห์ข้อมูลสำหรับการสร้างตัวแบบวิธีต้นไม้ตัดสินใจ
ด้วยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE

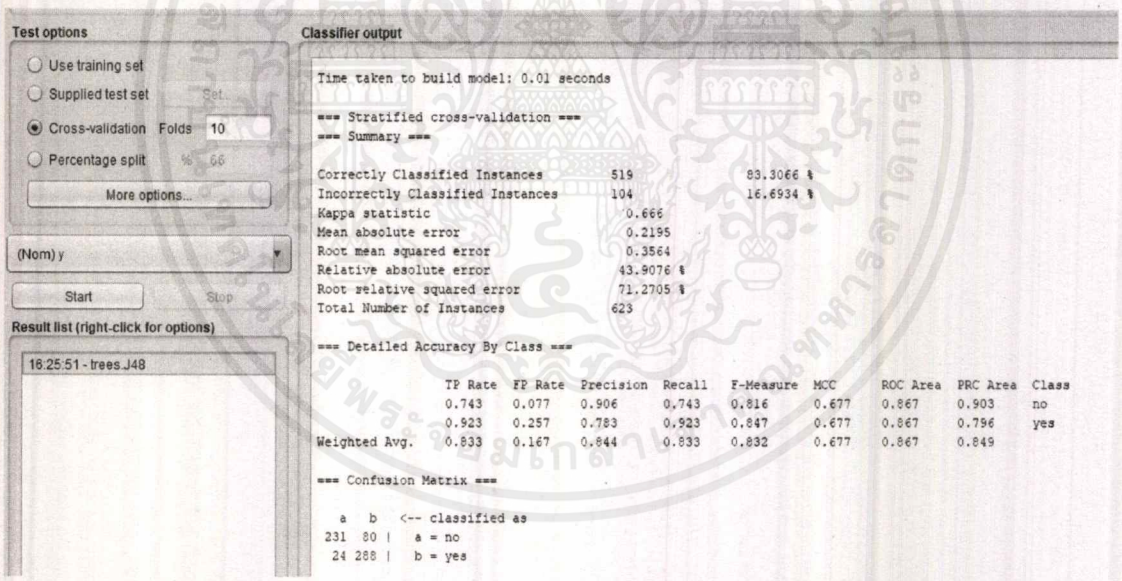
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 การปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลด



รูปที่ ค-7 ผลการวิเคราะห์ข้อมูลสำหรับการสร้างตัวแบบวิธีต้นไม้ตัดสินใจ
ด้วยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลด

2.4 การปรับความไม่สมดุลของข้อมูลด้วยวิธีการผสมผสาน

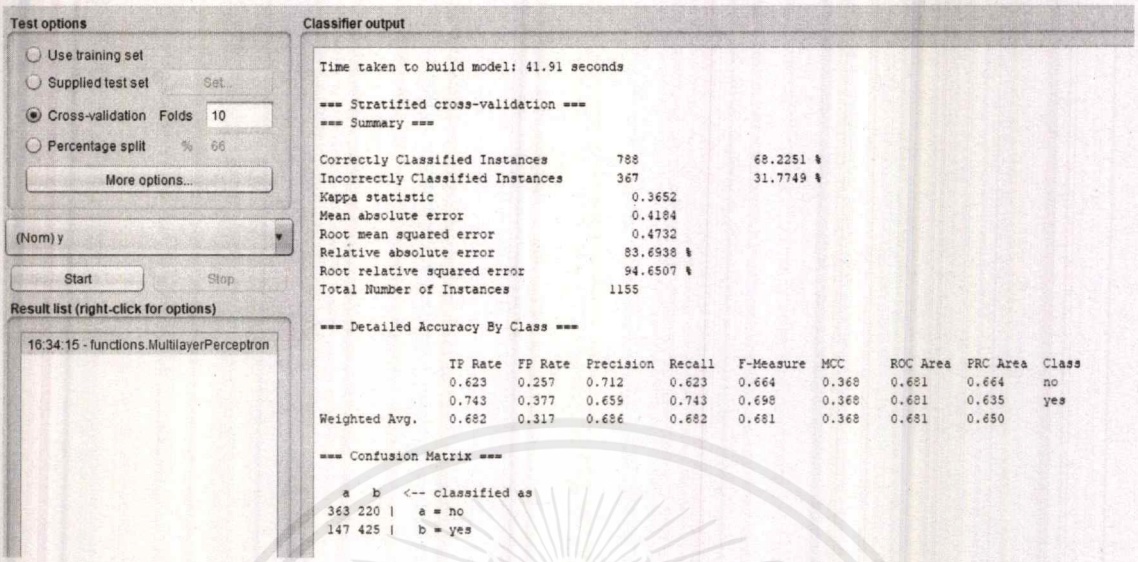


รูปที่ ค-8 ผลการวิเคราะห์ข้อมูลสำหรับการสร้างตัวแบบวิธีต้นไม้ตัดสินใจ
ด้วยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการผสมผสาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. วิธีโครงข่ายประสาทเทียม

3.1 การปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกิน



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) y

Start Stop

Result list (right-click for options)

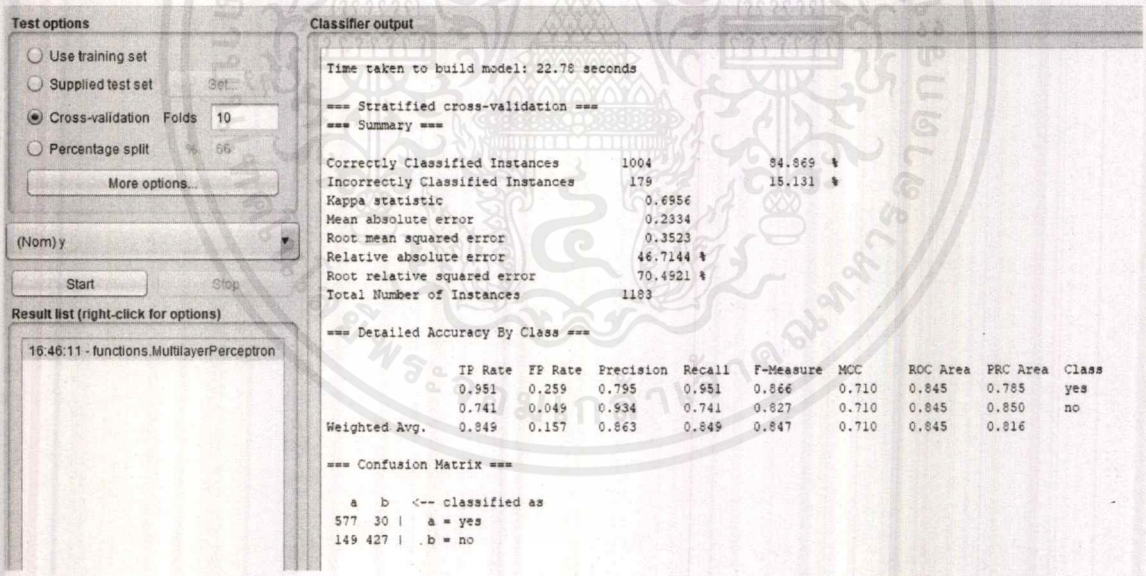
16:34:15 - functions.MultilayerPerceptron

Classifier output

```
Time taken to build model: 41.91 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         788               68.2251 %
Incorrectly Classified Instances       367               31.7749 %
Kappa statistic                          0.3652
Mean absolute error                      0.4184
Root mean squared error                  0.4732
Relative absolute error                 83.6938 %
Root relative squared error             94.6507 %
Total Number of Instances             1155

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.623     0.257     0.712       0.623   0.664      0.368    0.681     0.664     no
               0.743     0.377     0.659       0.743   0.698      0.368    0.681     0.635     yes
Weighted Avg.  0.682     0.317     0.686       0.682   0.681      0.368    0.681     0.650

=== Confusion Matrix ===

   a   b   <-- classified as
 363 220 |   a = no
 147 425 |   b = yes
```

รูปที่ ค-9 ผลการวิเคราะห์ข้อมูลสำหรับการสร้างตัวแบบวิธีโครงข่ายประสาทเทียม
ด้วยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกิน

3.2 การปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) y

Start Stop

Result list (right-click for options)

16:46:11 - functions.MultilayerPerceptron

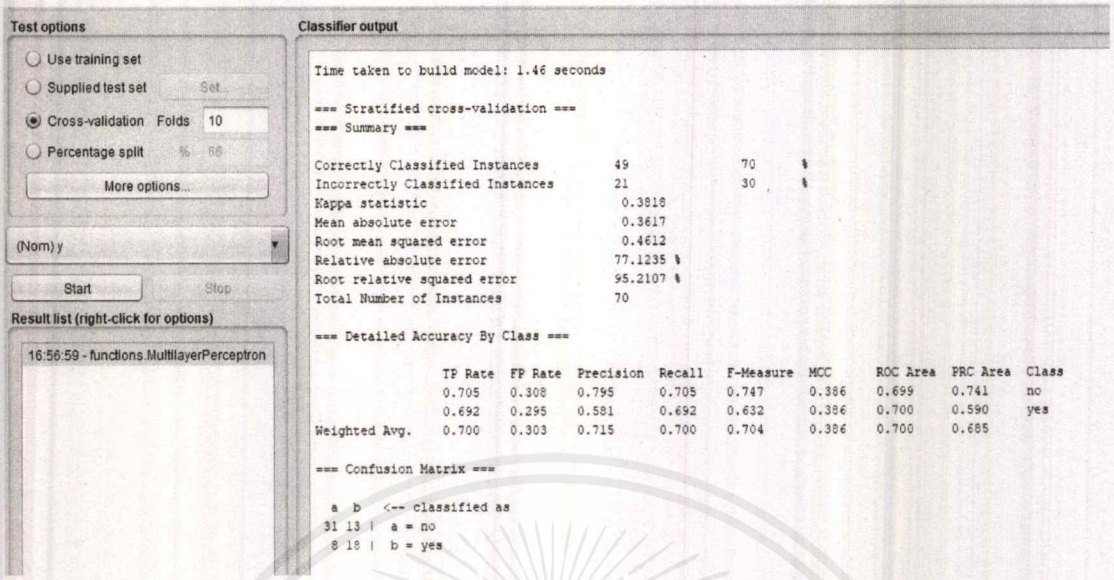
Classifier output

```
Time taken to build model: 22.78 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1004               84.869  %
Incorrectly Classified Instances       179               15.131  %
Kappa statistic                          0.6956
Mean absolute error                      0.2334
Root mean squared error                  0.3523
Relative absolute error                 46.7144 %
Root relative squared error             70.4921 %
Total Number of Instances              1183

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.951     0.259     0.795       0.951   0.866      0.710    0.845     0.785     yes
               0.741     0.049     0.934       0.741   0.827      0.710    0.845     0.850     no
Weighted Avg.  0.849     0.157     0.863       0.849   0.847      0.710    0.845     0.816

=== Confusion Matrix ===

   a   b   <-- classified as
 577  30 |   a = yes
 149 427 |   b = no
```

รูปที่ ค-10 ผลการวิเคราะห์ข้อมูลสำหรับการสร้างตัวแบบวิธีโครงข่ายประสาทเทียม
ด้วยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 การปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลด



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) y

Start Stop

Result list (right-click for options)

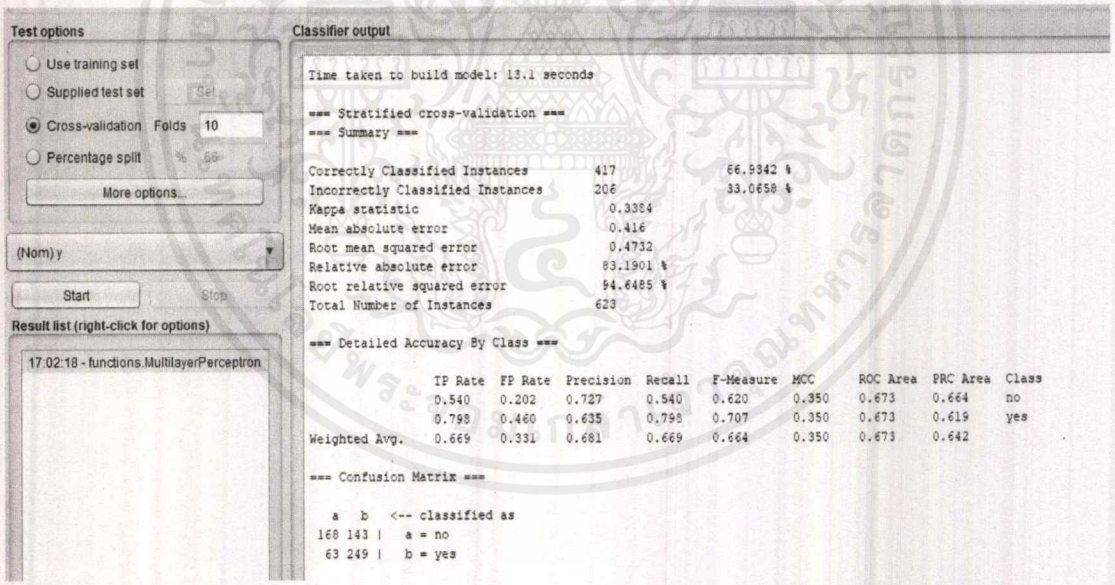
16:56:59 - functions.MultilayerPerceptron

Classifier output

```
Time taken to build model: 1.46 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          49               70      %
Incorrectly Classified Instances        21               30      %
Kappa statistic                          0.3818
Mean absolute error                      0.3617
Root mean squared error                  0.4612
Relative absolute error                 77.1235 %
Root relative squared error             95.2107 %
Total Number of Instances               70

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
               0.705     0.308     0.795     0.705     0.747     0.386     0.699     0.741     no
               0.692     0.295     0.581     0.692     0.632     0.386     0.700     0.590     yes
Weighted Avg.  0.700     0.303     0.715     0.700     0.704     0.386     0.700     0.685

=== Confusion Matrix ===

  a  b   <-- classified as
 31 13 |  a = no
  8 18 |  b = yes
```

รูปที่ ค-11 ผลการวิเคราะห์ข้อมูลสำหรับการสร้างตัวแบบวิธีโครงข่ายประสาทเทียม ด้วยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลด

3.4 การปรับความไม่สมดุลของข้อมูลด้วยวิธีการผสมผสาน



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) y

Start Stop

Result list (right-click for options)

17:02:18 - functions.MultilayerPerceptron

Classifier output

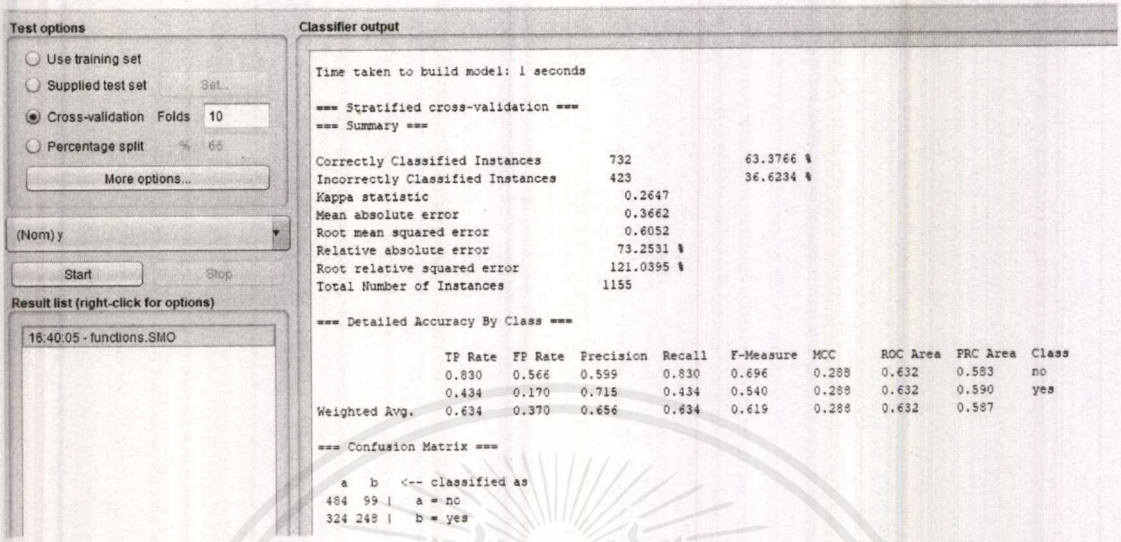
```
Time taken to build model: 13.1 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         417               66.9342 %
Incorrectly Classified Instances       206               33.0658 %
Kappa statistic                          0.3384
Mean absolute error                      0.416
Root mean squared error                  0.4732
Relative absolute error                 83.1901 %
Root relative squared error             94.6485 %
Total Number of Instances               623

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
               0.540     0.202     0.727     0.540     0.620     0.350     0.673     0.664     no
               0.798     0.460     0.635     0.798     0.707     0.350     0.673     0.619     yes
Weighted Avg.  0.669     0.331     0.681     0.669     0.664     0.350     0.673     0.642

=== Confusion Matrix ===

   a   b   <-- classified as
 168 143 |   a = no
  63 249 |   b = yes
```

รูปที่ ค-12 ผลการวิเคราะห์ข้อมูลสำหรับการสร้างตัวแบบวิธีโครงข่ายประสาทเทียม ด้วยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการผสมผสาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

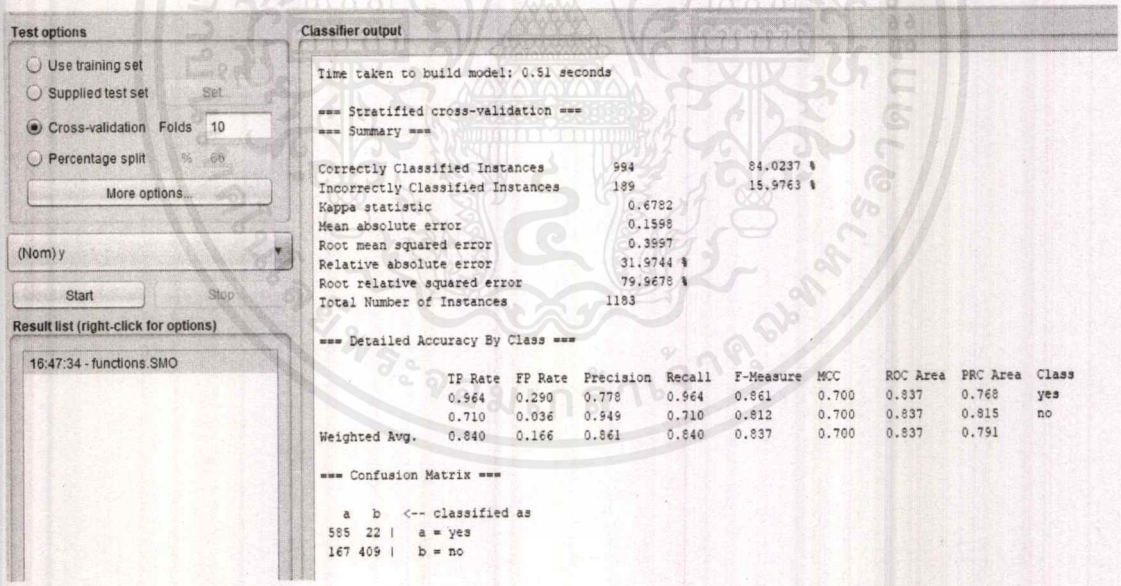
4. วิธีซัพพอร์ทเวกเตอร์แมชชีน

4.1 การปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกิน



รูปที่ ค-13 ผลการวิเคราะห์ข้อมูลสำหรับการสร้างตัวแบบวิธีซัพพอร์ทเวกเตอร์แมชชีน
ด้วยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกิน

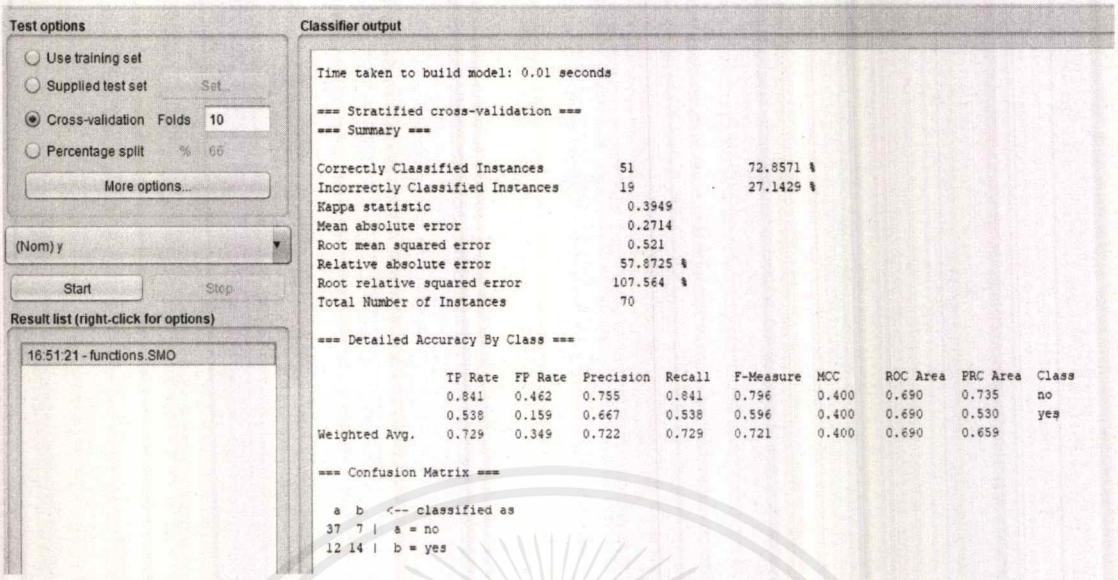
4.2 การปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE



รูปที่ ค-14 ผลการวิเคราะห์ข้อมูลสำหรับการสร้างตัวแบบวิธีซัพพอร์ทเวกเตอร์แมชชีน
ด้วยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเกินเทคนิค SMOTE

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 การปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลด



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) y

Start Stop

Result list (right-click for options)

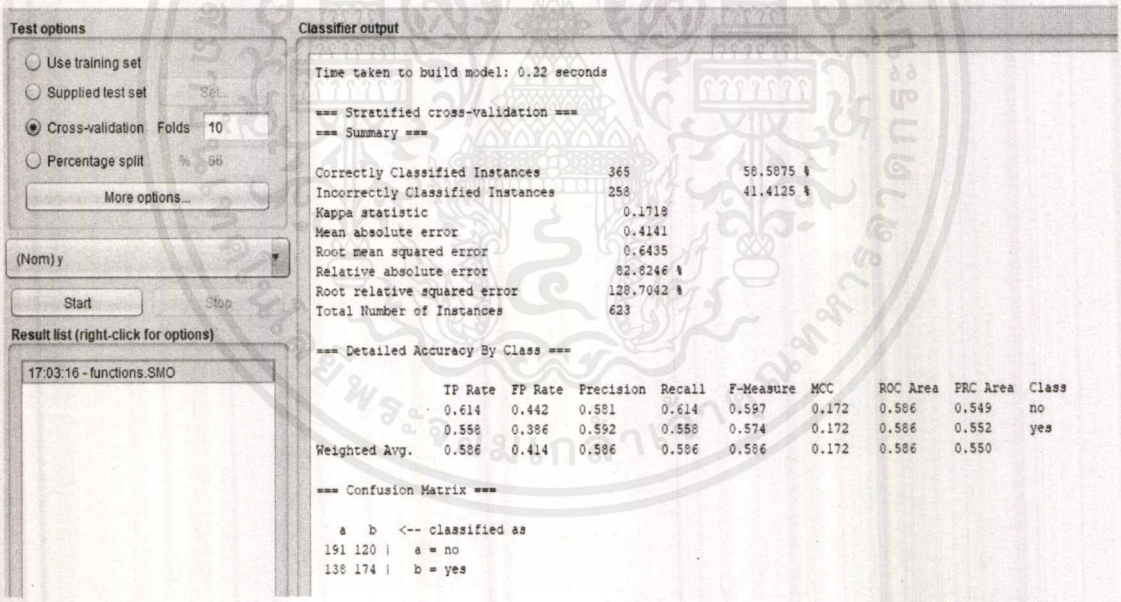
16:51:21 - functions.SMO

Classifier output

```
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          51               72.8571 %
Incorrectly Classified Instances        19               27.1429 %
Kappa statistic                          0.3949
Mean absolute error                      0.2714
Root mean squared error                  0.521
Relative absolute error                 57.8725 %
Root relative squared error            107.564  %
Total Number of Instances               70

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
                 0.841     0.462     0.755     0.841     0.796     0.400     0.690     0.735     no
                 0.538     0.159     0.667     0.538     0.596     0.400     0.690     0.530     yes
Weighted Avg.    0.729     0.349     0.722     0.729     0.721     0.400     0.690     0.659

=== Confusion Matrix ===

  a  b   <-- classified as
 37  7 |  a = no
 12 14 |  b = yes
```

รูปที่ ค-15 ผลการวิเคราะห์ข้อมูลสำหรับการสร้างตัวแบบวิธีซัพพอร์ทเวกเตอร์แมชชีน ด้วยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มลด

4.4 การปรับความไม่สมดุลของข้อมูลด้วยวิธีการผสมผสาน



Test options

Use training set
Supplied test set Set...
Cross-validation Folds 10
Percentage split % 66
More options...

(Nom) y

Start Stop

Result list (right-click for options)

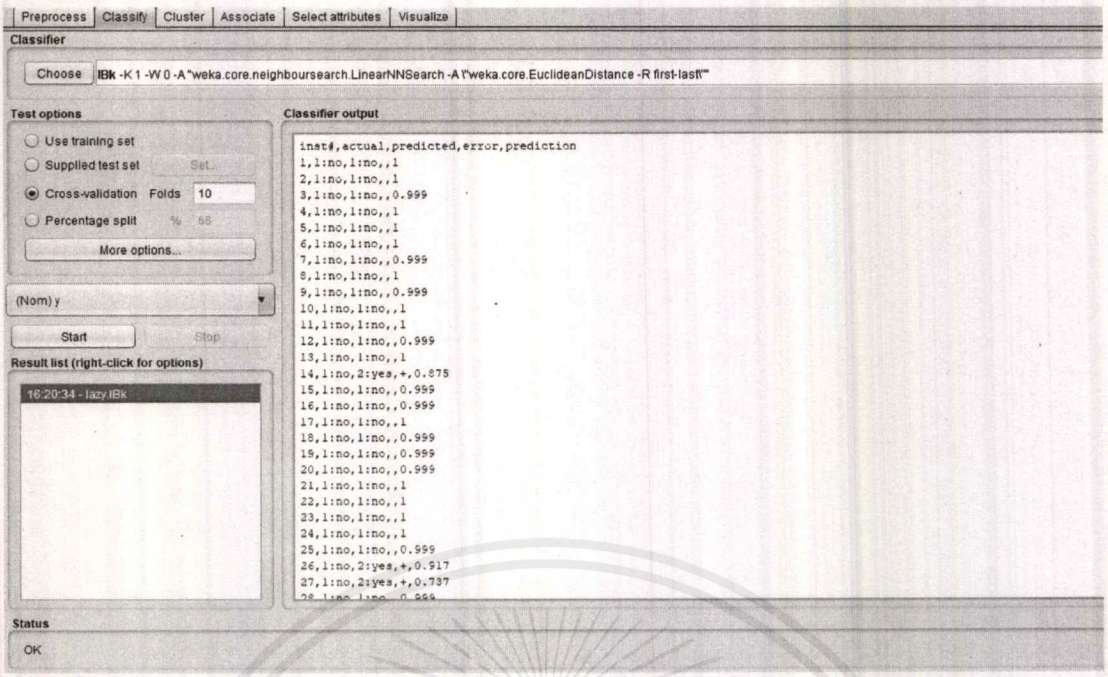
17:03:16 - functions.SMO

Classifier output

```
Time taken to build model: 0.22 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         365               58.5875 %
Incorrectly Classified Instances       258               41.4125 %
Kappa statistic                          0.1718
Mean absolute error                      0.4141
Root mean squared error                  0.6435
Relative absolute error                 82.8246 %
Root relative squared error            128.7042 %
Total Number of Instances              623

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
                 0.614     0.442     0.581     0.614     0.597     0.172     0.586     0.549     no
                 0.558     0.386     0.592     0.558     0.574     0.172     0.586     0.552     yes
Weighted Avg.    0.586     0.414     0.586     0.586     0.586     0.172     0.586     0.550

=== Confusion Matrix ===

   a   b   <-- classified as
 191 120 |   a = no
 138 174 |   b = yes
```

รูปที่ ค-16 ผลการวิเคราะห์ข้อมูลสำหรับการสร้างตัวแบบวิธีซัพพอร์ทเวกเตอร์แมชชีน ด้วยการปรับความไม่สมดุลของข้อมูลด้วยวิธีการผสมผสาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ตัวอย่างที่1 การคำนวณค่าความถูกต้อง (Accuracy) ค่าความไว (Sensitivity) ค่าความจำเพาะ (Specificity) และค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของการวิเคราะห์ข้อมูลข้อมูลโรคมะเร็งเต้านมของรัฐวิสคอนซิน สำหรับการทำนายตัวแบบด้วยวิธีต้นไม้ตัดสินใจ

จากรูปที่ ค-1

ค่าความถูกต้อง (Accuracy) $= \frac{TP+TN}{TP+TN+FP+FN} \times 100$

$= \frac{483+563}{483+563+9+100} \times 100$

$= 90.5628\%$

ค่าความไว (Sensitivity) $= \frac{TP}{TP+FN}$

$= \frac{483}{483+100}$

$= 0.828$

ค่าความจำเพาะ (Specificity) $= \frac{TN}{TN+FP}$

$= \frac{563}{563+9}$

$= 0.984$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE)

1) กำหนดให้ค่า $y_i = 1$ ได้จากในกรณีที่ค่าจริง (actual) ใน class attribute ของข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก ในระเบียนนั้นตรงกับค่าทำนาย (predicted)

เช่น ระเบียนที่ 1 ค่าจริง (actual) = 1 : ได้ยินเสียงไม่ปกติ

ค่าทำนาย (predicted) = 1 : ได้ยินเสียงไม่ปกติ

จะได้ $y_i = y_1 = 1$

2) กำหนดให้ค่า $y_i = 0$ ได้จากในกรณีที่ค่าจริง (actual) ใน class attribute ของของข้อมูลการรับรู้ทางหูของเด็กที่มีภาวะน้ำคั่งในหูชั้นกลางหรือหูน้ำหนวก ในระเบียนนั้นไม่ตรงกับค่าทำนาย (predicted)

เช่น ระเบียนที่ 14 ค่าจริง (actual) = 1 : ได้ยินเสียงไม่ปกติ

ค่าทำนาย (predicted) = 2 : ได้ยินเสียงปกติ

จะได้ $y_i = y_{14} = 0$

กำหนดให้ค่า $\hat{y}_i$ ได้จาก ค่าทำนาย (prediction) ซึ่งอยู่ที่คอลัมน์ขวาสุดของระเบียนนั้นในช่อง Classifier output

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
1	1	1	0
2	1	1	0
3	1	0.999	0.000001
4	1	1	0
5	1	1	0
6	1	1	0
7	1	0.999	0.000001
8	1	1	0
9	1	0.999	0.000001
10	1	1	0
11	1	1	0
12	1	0.999	0.000001
13	1	1	0
14	0	0.875	0.765625
15	1	0.999	0.000001
16	1	0.999	0.000001
17	1	1	0
18	1	0.999	0.000001
19	1	0.999	0.000001
20	1	0.999	0.000001
21	1	1	0
22	1	1	0
23	1	1	0
24	1	1	0
25	1	0.999	0.000001
26	0	0.917	0.840889
27	0	0.737	0.543169
28	1	0.999	0.000001
29	1	0.999	0.000001
30	1	1	0
31	1	0.999	0.000001
32	1	1	0
33	1	0.999	0.000001
34	1	1	0
35	1	1	0
36	0	0.643	0.413449
37	1	1	0
38	1	1	0
39	0	1	1
40	0	1	1
41	1	1	0
42	1	1	0
43	1	1	0
44	1	1	0

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
45	1	0.999	0.000001
46	1	1	0
47	1	1	0
48	1	1	0
49	1	0.999	0.000001
50	1	0.999	0.000001
51	1	0.999	0.000001
52	1	1	0
53	0	1	1
54	1	0.999	0.000001
55	1	0.533	0.218089
56	1	1	0
57	1	1	0
58	0	0.737	0.543169
59	1	1	0
60	1	0.909	0.008281
61	1	1	0
62	1	0.694	0.093636
63	1	1	0
64	1	0.733	0.071289
65	1	1	0
66	1	0.8	0.04
67	1	1	0
68	1	0.833	0.027889
69	1	1	0
70	1	0.818	0.033124
71	1	0.636	0.132496
72	1	0.929	0.005041
73	1	0.556	0.197136
74	1	0.875	0.015625
75	1	0.636	0.132496
76	1	0.929	0.005041
77	1	1	0
78	1	1	0
79	1	0.8	0.04
80	1	0.737	0.069169
81	1	0.875	0.015625
82	1	0.636	0.132496
83	1	1	0
84	1	1	0
85	1	1	0
86	1	0.875	0.015625
87	1	0.636	0.132496
88	1	1	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
89	1	0.929	0.005041
90	1	0.917	0.006889
91	1	0.875	0.015625
92	1	1	0
93	1	0.9	0.01
94	1	0.875	0.015625
95	1	0.694	0.093636
96	1	1	0
97	1	0.917	0.006889
98	1	1	0
99	1	1	0
100	1	0.789	0.044521
101	1	0.737	0.069169
102	1	0.875	0.015625
103	1	0.9	0.01
104	1	0.789	0.044521
105	1	0.694	0.093636
106	1	1	0
107	1	0.917	0.006889
108	1	0.875	0.015625
109	1	0.833	0.027889
110	1	0.694	0.093636
111	1	0.909	0.008281
112	1	0.9	0.01
113	1	0.9	0.01
114	1	0.818	0.033124
115	1	0.643	0.127449
116	1	1	0
117	1	1	0
118	1	0.999	0.000001
119	0	1	1
120	1	1	0
121	1	1	0
122	1	0.5	0.25
123	1	1	0
124	1	1	0
125	1	1	0
126	1	1	0
127	1	1	0
128	1	0.999	0.000001
129	1	0.999	0.000001
130	1	1	0
131	1	1	0
132	1	1	0

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
133	0	0.765	0.585225
134	1	1	0
135	1	0.999	0.000001
136	1	0.999	0.000001
137	1	1	0
138	1	1	0
139	1	0.999	0.000001
140	1	0.999	0.000001
141	0	0.63	0.3969
142	1	1	0
143	1	1	0
144	0	1	1
145	1	1	0
146	1	1	0
147	1	1	0
148	0	0.882	0.777924
149	1	0.5	0.25
150	0	0.767	0.588289
151	0	1	1
152	1	1	0
153	1	0.999	0.000001
154	0	0.833	0.693889
155	1	1	0
156	0	0.933	0.870489
157	1	1	0
158	1	0.999	0.000001
159	1	1	0
160	1	1	0
161	1	1	0
162	1	1	0
163	1	0.999	0.000001
164	1	0.999	0.000001
165	1	1	0
166	1	1	0
167	1	1	0
168	0	0.909	0.826281
169	1	1	0
170	1	1	0
171	1	0.999	0.000001
172	1	1	0
173	1	0.999	0.000001
174	1	1	0
175	1	0.999	0.000001
176	1	1	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
177	1	0.703	0.088209
178	1	1	0
179	1	0.846	0.023716
180	1	1	0
181	1	1	0
182	1	0.8	0.04
183	1	0.882	0.013924
184	1	0.8	0.04
185	1	1	0
186	1	1	0
187	1	0.933	0.004489
188	1	0.703	0.088209
189	1	1	0
190	1	0.714	0.081796
191	1	1	0
192	1	0.793	0.042849
193	1	0.703	0.088209
194	1	0.909	0.008281
195	1	1	0
196	1	0.846	0.023716
197	1	0.909	0.008281
198	1	0.933	0.004489
199	1	1	0
200	1	1	0
201	1	1	0
202	1	1	0
203	1	0.882	0.013924
204	1	1	0
205	1	0.895	0.011025
206	1	0.846	0.023716
207	1	1	0
208	1	0.846	0.023716
209	1	1	0
210	1	1	0
211	1	1	0
212	1	0.885	0.013225
213	1	0.63	0.1369
214	1	0.909	0.008281
215	1	1	0
216	1	0.765	0.055225
217	1	1	0
218	1	0.885	0.013225
219	1	1	0
220	1	1	0

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
221	1	0.63	0.1369
222	1	0.793	0.042849
223	1	0.895	0.011025
224	1	1	0
225	1	0.793	0.042849
226	1	1	0
227	1	1	0
228	1	1	0
229	1	0.765	0.055225
230	1	1	0
231	1	1	0
232	1	1	0
233	1	1	0
234	1	0.999	0.000001
235	1	1	0
236	1	0.999	0.000001
237	1	0.999	0.000001
238	1	1	0
239	0	1	1
240	0	1	1
241	1	0.999	0.000001
242	1	0.999	0.000001
243	1	0.999	0.000001
244	1	0.999	0.000001
245	0	0.68	0.4624
246	0	0.85	0.7225
247	1	1	0
248	1	0.999	0.000001
249	1	1	0
250	1	1	0
251	0	1	1
252	1	0.999	0.000001
253	1	1	0
254	1	0.999	0.000001
255	1	0.999	0.000001
256	1	0.999	0.000001
257	1	1	0
258	1	0.999	0.000001
259	1	0.999	0.000001
260	1	0.999	0.000001
261	1	1	0
262	1	1	0
263	1	0.999	0.000001
264	0	0.73	0.5329

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
265	1	1	0
266	1	0.999	0.000001
267	1	1	0
268	1	0.999	0.000001
269	1	1	0
270	1	0.999	0.000001
271	1	1	0
272	0	0.741	0.549081
273	1	1	0
274	0	0.846	0.715716
275	1	0.999	0.000001
276	1	1	0
277	1	1	0
278	1	1	0
279	1	0.999	0.000001
280	1	0.999	0.000001
281	0	0.68	0.4624
282	1	1	0
283	1	1	0
284	1	1	0
285	1	1	0
286	1	1	0
287	0	0.828	0.685584
288	0	0.68	0.4624
289	1	0.999	0.000001
290	1	1	0
291	0	0.95	0.9025
292	1	0.828	0.029584
293	0	0.75	0.5625
294	0	0.75	0.5625
295	1	1	0
296	1	1	0
297	1	0.769	0.053361
298	1	0.68	0.1024
299	1	1	0
300	1	0.769	0.053361
301	1	1	0
302	1	0.88	0.0144
303	0	0.5	0.25
304	1	0.73	0.0729
305	1	0.9	0.01
306	1	1	0
307	1	0.937	0.003969
308	1	1	0

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
309	1	0.929	0.005041
310	1	1	0
311	1	1	0
312	1	1	0
313	1	0.741	0.067081
314	1	0.828	0.029584
315	1	1	0
316	1	0.833	0.027889
317	1	1	0
318	1	0.741	0.067081
319	1	0.917	0.006889
320	1	0.929	0.005041
321	1	1	0
322	1	0.737	0.069169
323	1	0.88	0.0144
324	1	1	0
325	1	1	0
326	1	0.73	0.0729
327	0	0.5	0.25
328	1	0.769	0.053361
329	1	0.929	0.005041
330	1	1	0
331	1	0.875	0.015625
332	1	0.917	0.006889
333	1	0.692	0.094864
334	1	1	0
335	0	0.75	0.5625
336	1	1	0
337	1	1	0
338	1	1	0
339	1	1	0
340	1	0.741	0.067081
341	1	1	0
342	1	1	0
343	1	0.68	0.1024
344	0	0.75	0.5625
345	1	0.88	0.0144
346	1	0.692	0.094864
347	1	1	0
348	1	0.667	0.110889
349	0	0.933	0.870489
350	1	1	0
351	1	0.999	0.000001
352	0	0.737	0.543169

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
353	1	1	0
354	1	1	0
355	1	1	0
356	1	1	0
357	1	1	0
358	0	0.933	0.870489
359	1	1	0
360	1	1	0
361	1	1	0
362	0	0.722	0.521284
363	1	1	0
364	1	1	0
365	1	0.999	0.000001
366	1	1	0
367	1	1	0
368	1	0.999	0.000001
369	1	1	0
370	1	1	0
371	1	1	0
372	1	1	0
373	1	1	0
374	1	1	0
375	1	0.999	0.000001
376	1	0.999	0.000001
377	1	0.999	0.000001
378	1	1	0
379	1	1	0
380	1	1	0
381	1	0.999	0.000001
382	1	1	0
383	1	0.999	0.000001
384	0	0.786	0.617796
385	1	0.999	0.000001
386	0	0.786	0.617796
387	1	0.999	0.000001
388	1	1	0
389	1	0.999	0.000001
390	1	0.999	0.000001
391	1	1	0
392	1	0.999	0.000001
393	1	0.999	0.000001
394	0	0.9	0.81
395	0	0.737	0.543169
396	1	0.999	0.000001

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
397	1	1	0
398	0	0.571	0.326041
399	1	1	0
400	1	0.999	0.000001
401	0	1	1
402	1	1	0
403	1	0.999	0.000001
404	1	1	0
405	1	0.999	0.000001
406	1	0.999	0.000001
407	1	0.909	0.008281
408	1	1	0
409	1	0.895	0.011025
410	1	1	0
411	1	0.929	0.005041
412	1	0.577	0.178929
413	1	0.88	0.0144
414	1	1	0
415	1	0.909	0.008281
416	1	0.833	0.027889
417	1	0.577	0.178929
418	1	0.722	0.077284
419	1	0.778	0.049284
420	1	1	0
421	1	1	0
422	1	1	0
423	1	0.895	0.011025
424	1	0.577	0.178929
425	1	0.875	0.015625
426	1	0.8	0.04
427	1	0.778	0.049284
428	1	0.929	0.005041
429	1	0.722	0.077284
430	1	1	0
431	1	1	0
432	1	1	0
433	1	0.875	0.015625
434	1	0.577	0.178929
435	1	0.833	0.027889
436	1	0.833	0.027889
437	1	0.778	0.049284
438	1	0.722	0.077284
439	1	0.88	0.0144
440	1	0.88	0.0144

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
441	1	0.714	0.081796
442	1	0.571	0.184041
443	1	1	0
444	1	1	0
445	1	0.786	0.045796
446	1	1	0
447	1	1	0
448	1	1	0
449	1	0.571	0.184041
450	1	0.737	0.069169
451	1	1	0
452	1	0.9	0.01
453	1	0.806	0.037636
454	1	1	0
455	1	0.917	0.006889
456	1	0.937	0.003969
457	1	0.933	0.004489
458	1	0.929	0.005041
459	1	1	0
460	1	0.722	0.077284
461	1	0.722	0.077284
462	0	0.6	0.36
463	1	1	0
464	1	1	0
465	0	1	1
466	1	1	0
467	1	0.583	0.173889
468	1	1	0
469	0	0.862	0.743044
470	1	1	0
471	0	0.889	0.790321
472	0	0.583	0.339889
473	0	0.857	0.734449
474	1	0.999	0.000001
475	1	1	0
476	1	0.999	0.000001
477	0	0.933	0.870489
478	1	0.999	0.000001
479	1	0.999	0.000001
480	1	1	0
481	1	0.999	0.000001
482	1	1	0
483	1	0.583	0.173889
484	0	0.833	0.693889

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
485	1	1	0
486	1	1	0
487	1	1	0
488	1	1	0
489	0	0.92	0.8464
490	1	1	0
491	0	0.862	0.743044
492	1	0.999	0.000001
493	1	1	0
494	1	1	0
495	0	0.833	0.693889
496	1	1	0
497	1	1	0
498	1	1	0
499	1	0.999	0.000001
500	1	1	0
501	1	1	0
502	1	1	0
503	1	0.999	0.000001
504	1	1	0
505	1	1	0
506	1	0.999	0.000001
507	1	1	0
508	0	0.708	0.501264
509	1	1	0
510	1	1	0
511	1	0.999	0.000001
512	1	0.999	0.000001
513	1	1	0
514	1	1	0
515	0	0.889	0.790321
516	1	1	0
517	1	0.999	0.000001
518	0	0.706	0.498436
519	1	0.999	0.000001
520	1	0.999	0.000001
521	0	0.944	0.891136
522	1	1	0
523	1	1	0
524	1	1	0
525	1	0.857	0.020449
526	1	0.708	0.085264
527	1	0.929	0.005041
528	1	0.708	0.085264

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
529	1	1	0
530	1	0.917	0.006889
531	1	0.708	0.085264
532	1	0.722	0.077284
533	1	0.583	0.173889
534	1	0.875	0.015625
535	1	1	0
536	1	0.862	0.019044
537	1	0.706	0.086436
538	1	0.944	0.003136
539	1	0.583	0.173889
540	1	0.889	0.012321
541	1	1	0
542	1	0.583	0.173889
543	1	0.632	0.135424
544	1	0.929	0.005041
545	1	1	0
546	1	1	0
547	1	0.944	0.003136
548	1	0.875	0.015625
549	1	0.937	0.003969
550	1	0.706	0.086436
551	1	1	0
552	1	1	0
553	1	0.833	0.027889
554	1	1	0
555	1	1	0
556	1	0.583	0.173889
557	1	1	0
558	0	0.583	0.339889
559	1	0.875	0.015625
560	1	0.583	0.173889
561	1	0.933	0.004489
562	1	0.706	0.086436
563	1	1	0
564	1	0.708	0.085264
565	1	0.632	0.135424
566	1	0.722	0.077284
567	1	0.929	0.005041
568	1	0.706	0.086436
569	1	0.833	0.027889
570	1	0.708	0.085264
571	1	0.833	0.027889
572	1	1	0

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
573	1	0.875	0.015625
574	1	1	0
575	1	0.833	0.027889
576	1	0.706	0.086436
577	0	0.583	0.339889
578	1	0.632	0.135424
579	1	0.9	0.01
580	1	0.708	0.085264
581	1	1	0
582	0	0.9	0.81
583	1	1	0
584	1	1	0
585	1	0.999	0.000001
586	0	0.913	0.833569
587	1	1	0
588	1	1	0
589	1	0.999	0.000001
590	1	0.533	0.218089
591	1	0.999	0.000001
592	1	0.999	0.000001
593	1	0.999	0.000001
594	1	0.999	0.000001
595	1	1	0
596	1	0.999	0.000001
597	0	1	1
598	0	1	1
599	1	1	0
600	1	1	0
601	1	1	0
602	1	0.999	0.000001
603	1	1	0
604	0	0.737	0.543169
605	0	0.769	0.591361
606	1	0.999	0.000001
607	1	1	0
608	1	1	0
609	1	0.999	0.000001
610	1	1	0
611	0	0.857	0.734449
612	1	0.999	0.000001
613	1	1	0
614	1	0.999	0.000001
615	1	1	0
616	1	0.999	0.000001

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
617	1	0.999	0.000001
618	1	1	0
619	0	0.833	0.693889
620	1	1	0
621	1	1	0
622	1	1	0
623	1	1	0
624	0	0.667	0.444889
625	0	0.833	0.693889
626	1	1	0
627	1	1	0
628	1	1	0
629	1	1	0
630	1	1	0
631	1	0.999	0.000001
632	1	0.999	0.000001
633	1	1	0
634	1	1	0
635	1	1	0
636	1	0.999	0.000001
637	1	1	0
638	1	1	0
639	1	0.9	0.01
640	1	0.9	0.01
641	1	1	0
642	1	0.765	0.055225
643	1	0.823	0.031329
644	1	1	0
645	1	1	0
646	1	1	0
647	1	0.937	0.003969
648	1	1	0
649	1	0.913	0.007569
650	1	1	0
651	1	1	0
652	1	1	0
653	1	0.737	0.069169
654	1	1	0
655	1	1	0
656	1	0.8	0.04
657	1	0.823	0.031329
658	1	0.632	0.135424
659	1	1	0
660	1	0.823	0.031329

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
661	1	1	0
662	1	1	0
663	1	0.765	0.055225
664	1	0.632	0.135424
665	1	0.8	0.04
666	1	0.913	0.007569
667	1	0.9	0.01
668	1	0.724	0.076176
669	1	0.621	0.143641
670	1	0.769	0.053361
671	1	1	0
672	1	1	0
673	1	1	0
674	1	1	0
675	1	0.913	0.007569
676	1	0.9	0.01
677	1	0.632	0.135424
678	1	0.875	0.015625
679	1	1	0
680	1	0.765	0.055225
681	1	0.913	0.007569
682	1	1	0
683	1	0.9	0.01
684	1	1	0
685	1	1	0
686	1	0.724	0.076176
687	1	1	0
688	1	0.9	0.01
689	1	0.9	0.01
690	1	0.857	0.020449
691	1	0.765	0.055225
692	1	0.8	0.04
693	1	1	0
694	1	1	0
695	1	0.917	0.006889
696	1	1	0
697	1	1	0
698	1	1	0
699	1	0.999	0.000001
700	1	1	0
701	1	1	0
702	1	0.999	0.000001
703	1	1	0
704	0	0.933	0.870489

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
705	1	0.999	0.000001
706	1	0.999	0.000001
707	1	0.999	0.000001
708	0	0.933	0.870489
709	1	0.999	0.000001
710	1	1	0
711	1	0.999	0.000001
712	1	1	0
713	1	1	0
714	1	1	0
715	1	0.999	0.000001
716	1	0.999	0.000001
717	1	0.999	0.000001
718	1	1	0
719	0	0.722	0.521284
720	1	1	0
721	1	1	0
722	1	1	0
723	1	0.999	0.000001
724	1	1	0
725	0	0.667	0.444889
726	1	1	0
727	1	1	0
728	1	1	0
729	1	1	0
730	1	1	0
731	1	1	0
732	1	0.999	0.000001
733	1	1	0
734	1	0.999	0.000001
735	1	1	0
736	1	1	0
737	1	1	0
738	1	0.999	0.000001
739	1	1	0
740	1	1	0
741	1	0.999	0.000001
742	1	0.999	0.000001
743	1	0.999	0.000001
744	0	0.655	0.429025
745	1	0.533	0.218089
746	1	0.999	0.000001
747	1	1	0
748	1	1	0

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
749	1	1	0
750	1	1	0
751	1	1	0
752	1	1	0
753	1	1	0
754	1	0.875	0.015625
755	1	1	0
756	1	0.88	0.0144
757	1	0.885	0.013225
758	1	0.722	0.077284
759	1	0.76	0.0576
760	1	0.65	0.1225
761	1	1	0
762	1	0.76	0.0576
763	1	1	0
764	1	1	0
765	1	1	0
766	1	0.818	0.033124
767	1	0.875	0.015625
768	1	1	0
769	1	0.76	0.0576
770	1	0.722	0.077284
771	1	0.76	0.0576
772	1	0.875	0.015625
773	1	1	0
774	1	0.933	0.004489
775	1	1	0
776	1	0.786	0.045796
777	1	1	0
778	1	0.9	0.01
779	1	1	0
780	1	1	0
781	1	1	0
782	1	1	0
783	1	1	0
784	1	0.76	0.0576
785	1	0.722	0.077284
786	1	0.722	0.077284
787	1	0.75	0.0625
788	1	0.917	0.006889
789	1	0.933	0.004489
790	1	0.724	0.076176
791	1	0.75	0.0625
792	1	0.937	0.003969

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
793	1	0.933	0.004489
794	1	0.65	0.1225
795	1	1	0
796	1	1	0
797	1	1	0
798	1	0.885	0.013225
799	1	0.724	0.076176
800	1	0.786	0.045796
801	1	1	0
802	1	0.722	0.077284
803	1	1	0
804	1	0.76	0.0576
805	1	0.917	0.006889
806	1	1	0
807	1	0.76	0.0576
808	1	0.8	0.04
809	1	0.818	0.033124
810	1	1	0
811	0	0.808	0.652864
812	0	0.63	0.3969
813	1	1	0
814	1	1	0
815	0	0.8	0.64
816	1	0.533	0.218089
817	1	0.999	0.000001
818	0	0.808	0.652864
819	1	1	0
820	1	0.999	0.000001
821	1	0.999	0.000001
822	1	0.999	0.000001
823	0	0.833	0.693889
824	1	1	0
825	1	1	0
826	1	0.999	0.000001
827	1	1	0
828	0	0.8	0.64
829	1	1	0
830	1	1	0
831	1	0.999	0.000001
832	1	1	0
833	1	1	0
834	0	0.684	0.467856
835	1	1	0
836	1	0.999	0.000001

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
837	1	0.999	0.000001
838	1	1	0
839	1	1	0
840	1	0.999	0.000001
841	1	1	0
842	0	0.857	0.734449
843	1	1	0
844	1	1	0
845	0	0.75	0.5625
846	0	0.8	0.64
847	0	0.8	0.64
848	1	1	0
849	1	1	0
850	1	1	0
851	1	1	0
852	1	1	0
853	1	1	0
854	1	1	0
855	1	1	0
856	1	1	0
857	1	0.999	0.000001
858	1	1	0
859	0	0.667	0.444889
860	0	0.808	0.652864
861	1	1	0
862	0	0.857	0.734449
863	1	1	0
864	1	1	0
865	1	1	0
866	1	1	0
867	1	0.999	0.000001
868	1	1	0
869	1	0.917	0.006889
870	1	0.808	0.036864
871	1	1	0
872	1	1	0
873	1	1	0
874	1	0.864	0.018496
875	1	0.75	0.0625
876	1	1	0
877	1	0.857	0.020449
878	1	1	0
879	1	1	0
880	1	0.933	0.004489

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
881	1	1	0
882	1	1	0
883	1	0.864	0.018496
884	1	1	0
885	1	1	0
886	1	0.8	0.04
887	1	0.786	0.045796
888	1	1	0
889	1	0.818	0.033124
890	1	0.864	0.018496
891	1	0.818	0.033124
892	1	1	0
893	1	0.63	0.1369
894	1	1	0
895	1	0.75	0.0625
896	1	0.684	0.099856
897	1	1	0
898	1	0.917	0.006889
899	1	1	0
900	1	0.808	0.036864
901	1	1	0
902	1	1	0
903	1	1	0
904	1	0.895	0.011025
905	1	1	0
906	1	1	0
907	1	0.75	0.0625
908	1	0.895	0.011025
909	1	0.786	0.045796
910	1	0.875	0.015625
911	1	1	0
912	1	1	0
913	1	0.833	0.027889
914	1	0.857	0.020449
915	1	0.933	0.004489
916	1	0.864	0.018496
917	1	0.933	0.004489
918	1	0.63	0.1369
919	1	1	0
920	1	0.684	0.099856
921	1	1	0
922	1	0.833	0.027889
923	1	1	0
924	1	1	0

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
925	1	0.933	0.004489
926	1	1	0
927	1	0.999	0.000001
928	1	1	0
929	1	1	0
930	1	0.999	0.000001
931	0	1	1
932	1	1	0
933	0	0.765	0.585225
934	0	0.909	0.826281
935	0	0.722	0.521284
936	1	0.999	0.000001
937	0	0.75	0.5625
938	0	0.815	0.664225
939	1	1	0
940	1	1	0
941	1	1	0
942	1	0.999	0.000001
943	0	0.655	0.429025
944	1	1	0
945	1	1	0
946	1	1	0
947	1	0.999	0.000001
948	1	1	0
949	1	1	0
950	1	0.999	0.000001
951	1	0.999	0.000001
952	1	0.999	0.000001
953	1	1	0
954	1	1	0
955	1	1	0
956	0	0.75	0.5625
957	1	0.999	0.000001
958	1	0.999	0.000001
959	1	1	0
960	1	1	0
961	1	0.999	0.000001
962	1	0.999	0.000001
963	1	0.999	0.000001
964	0	0.722	0.521284
965	1	1	0
966	1	0.999	0.000001
967	1	0.999	0.000001
968	0	0.846	0.715716

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
969	1	1	0
970	1	0.999	0.000001
971	1	1	0
972	1	0.999	0.000001
973	1	1	0
974	1	0.999	0.000001
975	1	1	0
976	1	0.999	0.000001
977	0	0.917	0.840889
978	1	1	0
979	1	0.999	0.000001
980	1	0.999	0.000001
981	1	0.999	0.000001
982	1	1	0
983	1	0.999	0.000001
984	1	0.722	0.077284
985	1	1	0
986	1	0.722	0.077284
987	1	1	0
988	1	1	0
989	1	0.87	0.0169
990	1	1	0
991	1	0.846	0.023716
992	1	0.778	0.049284
993	1	0.778	0.049284
994	1	1	0
995	1	1	0
996	1	0.815	0.034225
997	1	0.889	0.012321
998	1	1	0
999	1	0.765	0.055225
1000	1	0.778	0.049284
1001	1	1	0
1002	1	0.815	0.034225
1003	1	0.846	0.023716
1004	1	0.889	0.012321
1005	1	1	0
1006	1	1	0
1007	1	1	0
1008	1	0.889	0.012321
1009	1	1	0
1010	1	1	0
1011	1	0.937	0.003969
1012	1	0.917	0.006889

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
1013	1	0.846	0.023716
1014	1	0.815	0.034225
1015	1	1	0
1016	1	0.917	0.006889
1017	1	0.846	0.023716
1018	1	0.75	0.0625
1019	1	0.846	0.023716
1020	1	1	0
1021	1	0.909	0.008281
1022	1	0.815	0.034225
1023	1	0.724	0.076176
1024	1	0.87	0.0169
1025	1	0.765	0.055225
1026	1	0.846	0.023716
1027	1	0.917	0.006889
1028	1	0.909	0.008281
1029	1	1	0
1030	1	1	0
1031	1	0.724	0.076176
1032	1	0.833	0.027889
1033	1	1	0
1034	1	0.909	0.008281
1035	1	0.833	0.027889
1036	1	0.87	0.0169
1037	1	1	0
1038	1	0.75	0.0625
1039	1	0.937	0.003969
1040	1	1	0
1041	1	0.999	0.000001
1042	1	0.999	0.000001
1043	1	1	0
1044	0	0.769	0.591361
1045	1	1	0
1046	1	1	0
1047	1	1	0
1048	1	0.999	0.000001
1049	1	0.999	0.000001
1050	1	0.999	0.000001
1051	1	0.999	0.000001
1052	1	1	0
1053	1	0.999	0.000001
1054	1	0.999	0.000001
1055	1	1	0
1056	1	0.999	0.000001

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
1057	1	1	0
1058	0	0.909	0.826281
1059	1	1	0
1060	1	1	0
1061	1	0.999	0.000001
1062	1	0.999	0.000001
1063	1	0.999	0.000001
1064	1	0.999	0.000001
1065	1	1	0
1066	1	1	0
1067	0	0.654	0.427716
1068	1	0.533	0.218089
1069	1	1	0
1070	1	1	0
1071	1	1	0
1072	1	1	0
1073	1	0.999	0.000001
1074	1	1	0
1075	1	0.999	0.000001
1076	1	0.999	0.000001
1077	1	1	0
1078	1	0.999	0.000001
1079	1	1	0
1080	0	0.87	0.7569
1081	1	0.999	0.000001
1082	0	1	1
1083	1	1	0
1084	1	1	0
1085	1	1	0
1086	1	0.999	0.000001
1087	1	1	0
1088	1	1	0
1089	1	1	0
1090	1	0.999	0.000001
1091	1	0.999	0.000001
1092	1	0.999	0.000001
1093	0	0.654	0.427716
1094	1	1	0
1095	1	1	0
1096	1	1	0
1097	1	1	0
1098	1	1	0
1099	1	0.909	0.008281
1100	1	0.9	0.01

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
1101	1	1	0
1102	1	0.87	0.0169
1103	1	0.8	0.04
1104	1	1	0
1105	1	1	0
1106	1	1	0
1107	1	0.654	0.119716
1108	1	0.704	0.087616
1109	1	1	0
1110	1	0.889	0.012321
1111	1	0.842	0.024964
1112	1	0.909	0.008281
1113	1	0.933	0.004489
1114	1	1	0
1115	1	0.909	0.008281
1116	1	0.704	0.087616
1117	1	0.704	0.087616
1118	1	0.686	0.098596
1119	1	0.654	0.119716
1120	1	1	0
1121	1	1	0
1122	1	0.686	0.098596
1123	1	0.686	0.098596
1124	1	0.686	0.098596
1125	1	1	0
1126	1	0.556	0.197136
1127	1	1	0
1128	1	1	0
1129	1	0.933	0.004489
1130	1	0.8	0.04
1131	1	0.8	0.04
1132	1	0.937	0.003969
1133	1	1	0
1134	1	1	0
1135	1	1	0
1136	1	1	0
1137	1	1	0
1138	1	1	0
1139	1	0.9	0.01
1140	1	0.909	0.008281
1141	1	1	0
1142	1	0.909	0.008281
1143	1	0.686	0.098596
1144	1	1	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
1145	1	0.867	0.017689
1146	1	1	0
1147	1	1	0
1148	1	1	0
1149	1	0.87	0.0169
1150	1	0.867	0.017689
1151	1	0.704	0.087616

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
1152	1	1	0
1153	1	0.769	0.053361
1154	1	0.8	0.04
1155	1	0.87	0.0169
รวม			91.417846

ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) = $\dfrac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}$

$= \dfrac{91.417846}{1155} = 0.07914965$

รากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) = $\sqrt{MSE} = \sqrt{0.07914965} = 0.2813$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้