รายงานการวิจัย

การเปรียบเทียบประสิทธิภาพและทำนายผลสถานภาพการเป็นนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

An Efficiency Comparison and Prediction in Student Status, Faculty of Sciences, King Mongkut's Institute of Technology Ladkrabang

อัจฉรา แผ้วบาง

รศ.สายชล สินสมบูรณ์ทอง

b00273587

ได้รับทุนสนับสนุนงานวิจัยจากเงินรายได้ ประจำปีงบประมาณ 2558

คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อ	การเปรียบเทียบประสิทธิภาพและทำนายผลสถานภาพการเป็นนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ผู้วิจัย	อัจฉรา แผ้วบาง
ผู้ร่วมวิจัย	รศ.สายชล สินสมบูรณ์ทอง
สาขา	สถิติประยุกต์
พ.ศ.	2558

บทคัดย่อ

ในการศึกษาเรื่องการเปรียบเทียบประสิทธิภาพและทำนายผลสถานภาพการเป็นนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง มีวัตถุประสงค์เพื่อศึกษาทำความเข้าใจและเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่ม รวมทั้งเปรียบเทียบการทำนายผลสถานภาพการเป็นนักศึกษา วิธีการจำแนกกลุ่มที่นำมาเปรียบเทียบประสิทธิภาพและการทำนายผลคือ วิธีความใกล้เคียงกันมากที่สุดโดยใช้อัลกอริทึมชนิด IBk วิธีแผนภาพต้นไม้เพื่อการตัดสินใจโดยใช้อัลกอริทึมชนิด J48 วิธีโครงข่ายประสาทโดยใช้อัลกอริทึมชนิดเพอร์เซปตรอนแบบหลายชั้น วิธีซัพพอร์ตเวกเตอร์แมชชีนโดยใช้อัลกอริทึม SMO ชนิดโพลิโนเมียลเคอร์เนล และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม ในการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มทั้ง 5 วิธี จะใช้ค่าความถูกต้อง ค่าอัตราความถูกต้องเชิงบวก ค่าอัตราความถูกต้องเชิงลบ ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล โดยจะพิจารณาจากค่าเหล่านี้ที่มีค่ามากที่สุด ส่วนค่าอัตราความผิดพลาดเชิงบวกและค่าอัตราความผิดพลาดเชิงลบจะพิจารณาจากค่าน้อยที่สุด ในการทำนายผลของวิธีการจำแนกกลุ่มทั้ง 5 วิธี จะใช้ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ในการเปรียบเทียบโดยพิจารณาจากค่า MAE และ MSE ที่มีค่าน้อยที่สุด

ผลการศึกษาพบว่าวิธีโครงข่ายประสาทมีค่าความถูกต้อง ค่าอัตราความถูกต้องเชิงบวก ค่าความถูกต้องเชิงลบ ค่าความแม่นยำ ค่าอัตราความผิดพลาดเชิงบวก ค่าอัตราความผิดพลาดเชิงลบ ค่าความระลึกและค่าความถ่วงดุลดีที่สุดคือ 98.1481%, 1, 0.727, 0.981, 0.273, 0, 1 และ 0.99 ตามลำดับ ส่วนวิธีแผนภาพต้นไม้เพื่อการตัดสินใจและวิธีซัพพอร์ตเวกเตอร์แมชชีนมีค่าอัตราความถูกต้องเชิงบวก ค่าอัตราความผิดพลาดเชิงลบ และค่าความระลึกดีที่สุดคือ 1, 0 และ 1 ตามลำดับ ดังนั้นวิธีโครงข่ายประสาทมีประสิทธิภาพดีที่สุด ส่วนการเปรียบเทียบการทำนายผลของวิธีการจำแนกกลุ่มสถานภาพการเป็นนักศึกษา พบว่าวิธีโครงข่ายประสาทมีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุดคือ 0.0217 และ 0.0188 ตามลำดับ ดังนั้นวิธีโครงข่ายประสาทมีการทำนายผลดีที่สุด

คำสำคัญ : วิธีการจำแนกกลุ่ม วิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาท วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	An Efficiency Comparison and Prediction in Student Status, Faculty of Sciences, King Mongkut's Institute of Technology Ladkrabang
Researcher	Achara Phaeobang
Co-researcher	Assoc.Prof. Saichon Sinsomboonthong
Programme	Applied Statistics
Year	2015

Abstract

In this study, an efficiency comparison and prediction in student status, faculty of sciences, King mongkut's institute of technology Ladkrabang were investigated. The purposes are understand and compare efficiency the classification method and also compare prediction in student status, faculty of sciences, King mongkut's institute of technology Ladkrabang. The classification methods to efficiency comparison and prediction are k-nearest neighbor method by IBk algorithm, decision tree method by J48 algorithm, neural network method by multilayer perceptron algorithm, support vector machine method by polynomial kernel, SMO algorithm and binary logistic regression method. In efficiency comparison of five classification methods by using the maximum of accuracy, true positive rate, true negative rate, precision, recall, F-Measure and the minimum of false positive rate and false negative rate. In prediction of five classification methods by using the minimum of mean absolute error (MAE) and mean square error (MSE).

The result of the study demonstrated that neural network method has the best of accuracy, true positive rate, true negative rate, precision, false positive rate, false negative rate, recall, and F-Measure, 98.1481%, 1, 0.727, 0.981, 0.273, 0, 1 and 0.99 respectively. Nevertheless, decision tree and support vector machine method have the best of true positive rate, false negative rate and recall, 1, 0 and 1 respectively. Then, neural network method has the best efficiency. In addition, the prediction comparison of classification methods show that neural network method has the minimum MAE and MSE, 0.0217 and 0.0188 respectively. Then, neural network method has the best prediction.

Keywords : classification method, k-nearest neighbor method, decision tree method, neural network method, support vector machine method and binary logistic regression method

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

งานวิจัยเล่มนี้สำเร็จลุล่วงไปได้ด้วยดี เนื่องจากผู้จัดทำได้รับความช่วยเหลือจากบุคคลผู้มีพระคุณหลายท่าน ดังนี้

ขอขอบพระคุณโครงการวิจัยที่เอื้อเฟื้อทุนสนับสนุนในการวิจัยครั้งนี้ โดยใช้เงินรายได้คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ขอขอบคุณนางสาวอัมภาวรรณ ชัยประดิษฐ นักศึกษาปริญญาตรี ชั้นปีที่ 4 สาขาวิชาสถิติประยุกต์ ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่คอยให้ความช่วยเหลือด้านการวิเคราะห์ข้อมูลบางส่วน

ขอขอบคุณทุกท่านที่มิได้เอ่ยนามในที่นี้ที่ให้คำแนะนำที่เป็นประโยชน์ต่าง ๆ และคอยเป็นกำลังใจให้งานวิจัยเล่มนี้สำเร็จลุล่วงไปได้ด้วยดี

.....

นางอัจฉรา แผ้วบาง

(หัวหน้าโครงการวิจัย)

.....

รศ.สายชล สินสมบูรณ์ทอง

(ผู้ร่วมโครงการวิจัย)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	I
บทคัดย่อภาษาอังกฤษ	II
กิตติกรรมประกาศ	III
สารบัญ	IV
สารบัญตาราง	VII
สารบัญรูป	
บทที่ 1 บทนำ	**1**
1.1 ความสำคัญและที่มาของปัญหา	1
1.2 วัตถุประสงค์ของงานวิจัย	4
1.3 ขอบเขตของงานวิจัย	4
1.4 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย	5
1.5 นิยามคำศัพท์	5
บทที่ 2 ทฤษฎีและรายงานวิจัยที่เกี่ยวข้อง	**7**
2.1 ทฤษฎีที่เกี่ยวข้อง	7
2.1.1 วิธีความใกล้เคียงกันมากที่สุด	7
2.1.2 วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ	8
2.1.3 วิธีโครงข่ายประสาท	11
2.1.4 วิธีซัพพอร์ตเวกเตอร์แมชชีน	16
2.1.5 วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม	20
2.2 รายงานวิจัยที่เกี่ยวข้อง	21
บทที่ 3 วิธีดำเนินงานวิจัย	**23**
3.1 อุปกรณ์ในการวิจัย	23
3.1.1 อุปกรณ์ที่มีอยู่แล้ว	23
3.1.2 อุปกรณ์ที่ต้องการเพิ่มเติม	23
3.2 ขั้นตอนการดำเนินงานวิจัย	23

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

บทที่ 4 ผลการวิจัยและอภิปรายผล **27**

4.1 การวิเคราะห์ข้อมูล 27

4.1.1 วิธีความใกล้เคียงกันมากที่สุด 27

4.1.1.1 การสร้างตัวแบบสถานภาพการเป็นนักศึกษา 27

4.1.1.2 การทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา 29

4.1.1.3 การทำนายสถานภาพการเป็นนักศึกษา 30

4.1.2 วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 35

4.1.2.1 การสร้างตัวแบบสถานภาพการเป็นนักศึกษา 35

4.1.2.2 การทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา 37

4.1.2.3 การทำนายสถานภาพการเป็นนักศึกษา 38

4.1.3 วิธีโครงข่ายประสาท 43

4.1.3.1 การสร้างตัวแบบสถานภาพการเป็นนักศึกษา 43

4.1.3.2 การทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา 45

4.1.3.3 การทำนายสถานภาพการเป็นนักศึกษา 46

4.1.4 วิธีซัพพอร์ตเวกเตอร์แมชชีน 56

4.1.4.1 การสร้างตัวแบบสถานภาพการเป็นนักศึกษา 56

4.1.4.2 การทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา 58

4.1.4.3 การทำนายสถานภาพการเป็นนักศึกษา 59

4.1.5 วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 64

4.1.5.1 การสร้างตัวแบบสถานภาพการเป็นนักศึกษา 64

4.1.5.2 การทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา 66

4.1.5.3 การทำนายสถานภาพการเป็นนักศึกษา 67

4.2 ผลการเปรียบเทียบประสิทธิภาพและทำนายผลของวิธีการจำแนกกลุ่ม 72

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ **74**

5.1 สรุปผลการวิจัย 74

5.2 ข้อเสนอแนะ 75

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

		หน้า
ภาคผนวก ก	รายละเอียดและตัวอย่างข้อมูลที่นำมาวิเคราะห์	76
ภาคผนวก ข	การวิเคราะห์ข้อมูล	80
ภาคผนวก ค	ตัวอย่างการคำนวณ	91
เอกสารอ้างอิง		97

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

หน้า

ตารางที่ 4.1 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด 27
ตารางที่ 4.2 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด 28
ตารางที่ 4.3 ผลในส่วนของเมตริกซ์ความสับสนจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด 28
ตารางที่ 4.4 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด 29
ตารางที่ 4.5 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด 30
ตารางที่ 4.6 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด 30
ตารางที่ 4.7 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีความใกล้เคียงกันมากที่สุด 31
ตารางที่ 4.8 ผลในส่วนของการสรุปผลจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด 33
ตารางที่ 4.9 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด 34
ตารางที่ 4.10 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด 34
ตารางที่ 4.11 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 35
ตารางที่ 4.12 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 36
ตารางที่ 4.13 ผลในส่วนของเมตริกซ์ความสับสนจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 36
ตารางที่ 4.14 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 37

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

หน้า

ตารางที่ 4.15 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 38

ตารางที่ 4.16 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 38

ตารางที่ 4.17 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 39

ตารางที่ 4.18 ผลในส่วนของการสรุปผลจากการทำนายสถานภาพการเป็นนักศึกษาด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 41

ตารางที่ 4.19 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 42

ตารางที่ 4.20 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายสถานภาพการเป็นนักศึกษาด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 42

ตารางที่ 4.21 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษาด้วยวิธีโครงข่ายประสาท 43

ตารางที่ 4.22 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท 44

ตารางที่ 4.23 ผลในส่วนของเมตริกซ์ความสับสนจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท 44

ตารางที่ 4.24 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท 45

ตารางที่ 4.25 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท 45

ตารางที่ 4.26 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท 46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

หน้า

ตารางที่ 4.27 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท สำหรับอัตราการเรียนรู้ $\eta = 0.1$ และโมเมนตัม $\alpha = 0.5, 0.6, 0.7, 0.8$ และ 0.9 47

ตารางที่ 4.28 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท สำหรับอัตราการเรียนรู้ $\eta = 0.2$ และโมเมนตัม $\alpha = 0.5, 0.6, 0.7, 0.8$ และ 0.9 48

ตารางที่ 4.29 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท สำหรับอัตราการเรียนรู้ $\eta = 0.3$ และโมเมนตัม $\alpha = 0.5, 0.6, 0.7, 0.8$ และ 0.9 49

ตารางที่ 4.30 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท สำหรับอัตราการเรียนรู้ $\eta = 0.4$ และโมเมนตัม $\alpha = 0.5, 0.6, 0.7, 0.8$ และ 0.9 50

ตารางที่ 4.31 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท สำหรับอัตราการเรียนรู้ $\eta = 0.5$ และโมเมนตัม $\alpha = 0.5, 0.6, 0.7, 0.8$ และ 0.9 51

ตารางที่ 4.32 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีโครงข่ายประสาท 52

ตารางที่ 4.33 ผลในส่วนของการสรุปผลจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท 54

ตารางที่ 4.34 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนาย สถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท 55

ตารางที่ 4.35 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท 55

ตารางที่ 4.36 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบการพิจารณาสถานภาพ การเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 56

ตารางที่ 4.37 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบ สถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 57

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

หน้า

ตารางที่ 4.38 ผลในส่วนของเมตริกซ์ความสับสนจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 57

ตารางที่ 4.39 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 58

ตารางที่ 4.40 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 58

ตารางที่ 4.41 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 59

ตารางที่ 4.42 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 60

ตารางที่ 4.43 ผลในส่วนของการสรุปผลจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 62

ตารางที่ 4.44 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 63

ตารางที่ 4.45 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 63

ตารางที่ 4.46 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 64

ตารางที่ 4.47 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 65

ตารางที่ 4.48 ผลในส่วนของเมตริกซ์ความสับสนจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 65

ตารางที่ 4.49 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 66

ตารางที่ 4.50 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 66

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

หน้า

ตารางที่ 4.51 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 67

ตารางที่ 4.52 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 68

ตารางที่ 4.53 ผลในส่วนของการสรุปผลจากการทำนายสถานภาพการเป็นนักศึกษาด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 70

ตารางที่ 4.54 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 71

ตารางที่ 4.55 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายสถานภาพการเป็นนักศึกษาด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 71

ตารางที่ 4.56 ผลการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มสถานภาพการเป็นนักศึกษาทั้ง 5 วิธี 72

ตารางที่ 4.57 ผลการเปรียบเทียบการทำนายผลของวิธีการจำแนกกลุ่มสถานภาพการเป็นนักศึกษาทั้ง 5 วิธี 73

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

หน้า

รูปที่ 2.1 ตัวอย่างของวิธีความใกล้เคียงกันมากที่สุด 8
รูปที่ 2.2 ส่วนประกอบของแผนภาพต้นไม้เพื่อการตัดสินใจ 9
รูปที่ 2.3 โครงข่ายของเซลล์ประสาท 12
รูปที่ 2.4 ลักษณะโครงข่ายประสาทเทียมแบบส่งสัญญาณไปข้างหน้า 14
รูปที่ 2.5 ลักษณะโครงข่ายประสาทเทียมแบบมีการย้อนกลับ 14
รูปที่ 2.6 ลักษณะโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น 15
รูปที่ 2.7 การขยายตัวของเส้นขอบ 17
รูปที่ 2.8 เส้นขอบและเส้นแบ่งเมื่อแทนด้วยสมการเส้นตรง 18
รูปที่ 2.9 รูปแบบการวางตัวที่ไม่สามารถแบ่งด้วยเส้นตรงได้ 19

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มาของปัญหา

โดยส่วนใหญ่นักเรียนที่เข้ามาศึกษาต่อในมหาวิทยาลัย ชั้นปีที่ 1 นั้นจะมาจากหลายพื้นที่ทั้งในเขต กทม. และต่างจังหวัด ทำให้นักศึกษาจำเป็นต้องมีการปรับตัวในหลาย ๆ ด้าน ได้แก่ การศึกษา ที่อยู่อาศัย การเดินทาง เพื่อน สภาพแวดล้อม โดยเฉพาะด้านการศึกษาของนักศึกษาที่มีเกรดเฉลี่ยสะสมแตกต่างกัน ซึ่งอาจจะมีปัจจัยที่ส่งผลต่อการเรียนของนักศึกษา เช่น เพศ อายุ สาขาวิชา คะแนนเฉลี่ยสะสมในระดับมัธยมศึกษาตอนปลาย ประเภทของโรงเรียนที่นักศึกษาสำเร็จการศึกษา จังหวัดของโรงเรียนในระดับมัธยมศึกษาตอนปลาย วิธีผ่านการสอบคัดเลือก ที่พักอาศัย จำนวนหน่วยกิตที่ลงทะเบียนเรียน ความพร้อม แรงจูงใจ การเรียนรู้ ความจำ วิธีการสอนของอาจารย์ที่มีส่วนช่วยสนับสนุนให้นักศึกษาชั้นปีที่ 1 เกิดการพ้นสภาพการเป็นนักศึกษา ซึ่งมีค่าระดับคะแนนเฉลี่ยประจำภาคการศึกษา (GPS) ต่ำกว่า 2.00 เป็นเวลา 2 ภาคการศึกษาติดต่อกัน และค่าระดับคะแนนเฉลี่ยสะสม (GPA) ต่ำกว่า 2.00 ซึ่งการพ้นสภาพการเป็นนักศึกษาได้เกิดขึ้นในทุกสถานการศึกษา รวมทั้งสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง โดยเฉพาะนักศึกษาในคณะวิทยาศาสตร์ซึ่งเป็นหนึ่งในหลาย ๆ คณะที่นักศึกษาชั้นปีที่ 1 จะพ้นสภาพการเป็นนักศึกษามากกว่านักศึกษาชั้นปีอื่น ๆ ดังมีรายงานวิจัยของศุภมาส ไวกุลเพ็ชร และสมฤทัย รัศมีธรรม (2545) ได้ทำการศึกษานักศึกษาในคณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ซึ่งได้จัดการเรียนการสอนโดยแบ่งสาขาวิชาออกเป็น 5 สาขาวิชา ได้แก่ ฟิสิกส์ประยุกต์ ชีววิทยาประยุกต์ เคมี คณิตศาสตร์และวิทยาการคอมพิวเตอร์ และสถิติประยุกต์ มีหลักสูตรจำนวน 10 หลักสูตร ได้แก่ ฟิสิกส์ประยุกต์–เครื่องมือวิทยาศาสตร์และอุตสาหกรรม ฟิสิกส์ประยุกต์ จุลชีววิทยาอุตสาหกรรม เทคโนโลยีชีวภาพ เคมีอุตสาหกรรม–เครื่องมือวิเคราะห์ เคมีทรัพยากรสิ่งแวดล้อม เคมีอุตสาหกรรม วิทยาการคอมพิวเตอร์ คณิตศาสตร์ประยุกต์ และสถิติประยุกต์ จากการศึกษาปัจจัยที่มีผลต่อเกรดเฉลี่ยสะสมของนักศึกษาคณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง เกี่ยวกับสถานภาพส่วนตัว พฤติกรรมทางการเรียนและทัศนคติต่อการเรียนการสอนของอาจารย์ ข้อมูลที่รวบรวมได้นำมาวิเคราะห์เพื่อศึกษาถึงความสัมพันธ์ของปัจจัยที่มีผลต่อเกรดเฉลี่ยสะสม โดยใช้สถิติไค-สแควร์และค่าคราแมร์วี ผลการศึกษาพบว่าปัจจัยที่มีความสัมพันธ์กับเกรดเฉลี่ยสะสมของนักศึกษาที่ระดับนัยสำคัญทางสถิติ 0.05 ได้แก่ เพศ สาขาวิชา ชั้นปี วิธีผ่านการสอบคัดเลือก ค่าใช้จ่ายเฉลี่ยต่อเดือนเพื่อการศึกษา การตั้งใจเรียนในห้องเรียน การจดคำบรรยายระหว่างการเรียนในห้องเรียน จำนวนครั้งในการซักถามอาจารย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เกี่ยวกับสิ่งที่ข้องใจในวิชาที่เรียน ความพอใจในสาขาวิชาที่เรียนในปัจจุบัน ความเชี่ยวชาญในสาขาวิชาที่เรียน ความพอใจในการสอนของอาจารย์ ความชื่นชมในท่านอาจารย์ การเปิดโอกาสให้ซักถามเกี่ยวกับเนื้อหาวิชาที่เรียนในห้องเรียน และการเปิดโอกาสให้ซักถามเกี่ยวกับเนื้อหาวิชาที่เรียนนอกห้องเรียน

ณัฐณิชา และคณะ (2550) ทำการศึกษาถึงความแตกต่างระหว่างเพศในด้านพฤติกรรมที่มีผลต่อผลสัมฤทธิ์ทางการศึกษาของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง เกี่ยวกับสถานภาพทั่วไป พฤติกรรมการเรียน พฤติกรรมการใช้เวลาว่าง และพฤติกรรมการสังสรรค์ของนักศึกษา ข้อมูลที่รวบรวมได้นำมาวิเคราะห์เพื่อศึกษาถึงความแตกต่างระหว่างเพศในด้านพฤติกรรมที่มีผลต่อผลสัมฤทธิ์ทางการศึกษาโดยใช้สถิติการทดสอบแบบวิลคอกซัน แมนวิทนีย์ การทดสอบไค-สแควร์ และการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ พบว่านักศึกษาส่วนใหญ่มีระดับการศึกษาของบิดา มารดาในระดับปริญญาตรี นักศึกษาส่วนใหญ่มีภูมิลำเนาอยู่ต่างจังหวัด (ร้อยละ 61.3) อาศัยอยู่หอพัก (ร้อยละ 59.4) ใช้เวลาในการเดินทางมาเรียนน้อยกว่า 30 นาที (ร้อยละ 67.9) ค่าเฉลี่ยของเกรดเฉลี่ยสะสมของนักศึกษาเพศชายและนักศึกษาเพศหญิงคือ 2.44 และ 2.57 ตามลำดับ จากการทดสอบพบว่าเกรดเฉลี่ยสะสมมีความแตกต่างระหว่างเพศ ($p = 0.019$) ผลการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ พบว่า ในนักศึกษาเพศชาย พฤติกรรมการเรียน และพฤติกรรมการสังสรรค์มีอิทธิพลต่อผลสัมฤทธิ์ทางการศึกษา ส่วนนักศึกษาเพศหญิง พฤติกรรมการเรียนมีอิทธิพลต่อผลสัมฤทธิ์ทางการศึกษา

อนงค์ ผลสุขการ และสายชล สินสมบูรณ์ทอง (2553) ศึกษาเรื่องปัจจัยที่มีผลสัมฤทธิ์ทางด้านการเรียนของนักศึกษาชั้นปีที่ 1 คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง จากทุกสาขาวิชาและทุกหลักสูตร ในภาคเรียนที่ 1 ปีการศึกษา 2552 โดยมีตัวแปรตาม คือ ผลสัมฤทธิ์ทางด้านการเรียนของนักศึกษา ส่วนตัวแปรอิสระประกอบด้วยเพศ อายุ สาขาวิชา หลักสูตรที่เรียน วิธีผ่านการสอบคัดเลือก การศึกษาในโรงเรียนระดับมัธยมศึกษาตอนปลาย จังหวัดของโรงเรียน คะแนนเกรดเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลาย (GPAX) จำนวนหน่วยกิตที่ลงทะเบียนเรียน ระยะเวลาในการเดินทางมาเรียน ที่พักอาศัย พฤติกรรมด้านการเรียน และทัศนคติต่อการเรียนการสอนของอาจารย์ โดยวางแผนการสุ่มตัวอย่างแบบแบ่งชั้นภูมิชนิดสุ่มอย่างง่าย วิเคราะห์ข้อมูลโดยใช้การวิเคราะห์การถดถอยโลจิสติก ผลของการศึกษาพบว่าปัจจัยที่มีผลสัมฤทธิ์ทางด้านการเรียนของนักศึกษาชั้นปีที่ 1 คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ภาคเรียนที่ 1 ปีการศึกษา 2552 คือ อายุ คะแนน GPAX และจำนวนหน่วยกิตที่ลงทะเบียนเรียน โดยมีสมการเป็น $\hat{W} = -0.258\text{age} + 0.687\text{GPAX} + 0.318\text{credit}$ นอกจากนี้ ในการใช้สมการพยากรณ์ พบว่าจากข้อมูลจริงมีนักศึกษาชั้นปีที่ 1 คณะวิทยาศาสตร์ มีผลสัมฤทธิ์การด้านการเรียนผ่าน 321 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และเมื่อใช้สมการที่มีตัวแปรอายุ คะแนน GPAX และจำนวนหน่วยกิตที่ลงทะเบียนเรียนอยู่ในตัวแบบการถดถอยโลจิสติก จะพยากรณ์ว่านักศึกษามีผลสัมฤทธิ์การด้านการเรียนผ่าน 319 คน นั่นคือสมการพยากรณ์ถูกต้องถึง 99.4% ในขณะที่นักศึกษามีผลสัมฤทธิ์การด้านการเรียนไม่ผ่าน 58 คน แต่เมื่อใช้สมการที่มีตัวแปรอายุ คะแนน GPAX และจำนวนหน่วยกิตที่ลงทะเบียนเรียนอยู่ในตัวแบบการถดถอยโลจิสติก จะพยากรณ์ว่านักศึกษามีผลสัมฤทธิ์การด้านการเรียนไม่ผ่าน 3 คน นั่นคือสมการพยากรณ์ถูกต้องเพียง 5.20% ดังนั้นเปอร์เซ็นต์ทั้งหมดของการพยากรณ์ถูกต้องคือ 85.00%

สุจิตรา สุคนธมัต (2555) ศึกษาปัจจัยที่มีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนของนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง โดยศึกษาประชากรจากนักศึกษาระดับปริญญาตรีทุกชั้นปี จำนวน 3,402 คน สุ่มตัวอย่างขนาด 552 คน โดยใช้แผนการสุ่มตัวอย่างแบบแบ่งชั้นภูมิและเลือกกลุ่มตัวอย่างในแต่ละชั้นภูมิโดยวิธีการสุ่มแบบมีระบบ และใช้แบบสอบถามเป็นเครื่องมือในการเก็บรวบรวมข้อมูล วิธีการทางสถิติที่ใช้ในการวิเคราะห์ข้อมูล ได้แก่ ร้อยละ การทดสอบ Z การวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว การเปรียบเทียบรายคู่ และสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน ผลการสำรวจพบว่าเมื่อนักศึกษาประสบปัญหาวิตกกังวลในเรื่องการเรียน/เรื่องส่วนตัว ส่วนใหญ่จะขอคำปรึกษาจากบิดามารดา สาเหตุที่ทำให้นักศึกษามีผลการเรียนไม่เป็นไปตามที่คาดหวัง ได้แก่ เอาใจใส่การเรียนน้อย วิธีปรับปรุงตนเองของนักศึกษาเพื่อให้ผลการเรียนดีขึ้น ได้แก่ ตั้งใจเรียนมากขึ้น สาเหตุจากอาจารย์ผู้สอนที่ทำให้นักศึกษามีผลการเรียนไม่เป็นไปตามที่คาดหวัง ได้แก่ วิธีการสอนของอาจารย์และอาจารย์ออกข้อสอบยากเกินไป ผลจากการทดสอบสมมติฐานปัจจัยที่มีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนของนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ได้แก่ ปัจจัยทางด้านประเภทของการสอบคัดเลือกเข้าศึกษาไม่มีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนของนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ปัจจัยทางครอบครัว ได้แก่ สถานภาพครอบครัว ระดับการศึกษาของบิดามารดา อาชีพของบิดามารดา รายได้ของบิดามารดา ความสัมพันธ์ระหว่างบิดามารดากับนักศึกษาไม่มีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนของนักศึกษา ปัจจัยทางด้านสภาพแวดล้อมในคณะวิทยาศาสตร์ ได้แก่ สภาพแวดล้อมในมหาวิทยาลัยโดยทั่ว ๆ ไป และความสัมพันธ์ของกลุ่มเพื่อนกับนักศึกษามีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนของนักศึกษา ปัจจัยทางด้านตัวนักศึกษา ได้แก่ เพศ อายุ สาขาวิชา ชั้นปี ภูมิลำเนา บุคคลที่นักศึกษาพักอาศัย การทำงานพิเศษและทัศนคติในการเรียนของนักศึกษามีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนของนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้นคณะผู้วิจัยจึงได้นำวิธีการทำเหมืองข้อมูลมาใช้ในการเปรียบเทียบประสิทธิภาพและทำนายผลสถานภาพการเป็นนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง โดยวิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาท วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม เพื่อหาวิธีที่มีประสิทธิภาพและทำนายผลสถานภาพการเป็นนักศึกษาที่เหมาะสมต่อไป

1.2 วัตถุประสงค์ของงานวิจัย

1.2.1 เพื่อเปรียบเทียบประสิทธิภาพสถานภาพการเป็นนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ระหว่างวิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาท วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

1.2.2 เพื่อเปรียบเทียบการทำนายผลสถานภาพการเป็นนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ระหว่างวิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีความใกล้เคียงกันมากที่สุด วิธีโครงข่ายประสาท วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

1.3 ขอบเขตของงานวิจัย

โครงการวิจัยนี้ได้ทำการศึกษาเรื่องการเปรียบเทียบประสิทธิภาพและทำนายผลสถานสภาพการเป็นนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ระหว่างวิธีความใกล้เคียงกันมากที่สุด โดยใช้อัลกอริทึมชนิด IBk วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ โดยใช้อัลกอริทึมชนิด J48 วิธีโครงข่ายประสาท โดยใช้อัลกอริทึมชนิดเพอร์เซปตรอนแบบหลายชั้น โดยกำหนดค่าอัตราการเรียนรู้เป็น 0.1, 0.2, 0.3, 0.4 และ 0.5 ค่าโมเมนตัมเป็น 0.5, 0.6, 0.7, 0.8 และ 0.9 จำนวนรอบการสอน 30,000 รอบ และชั้นซ่อน 1 ชั้น วิธีซัพพอร์ตเวกเตอร์แมชชีน ใช้อัลกอริทึม SMO ชนิดโพลิโนเมียลเคอร์เนล และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม ในการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มทั้ง 5 วิธี ใช้ค่าความถูกต้อง (Accuracy) ค่าอัตราความถูกต้องเชิงบวก (TP Rate) ค่าอัตราความถูกต้องเชิงลบ (TN Rate) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-Measure) โดยจะพิจารณาจากค่าเหล่านี้ที่มีค่ามากที่สุด ส่วนค่าอัตราความผิดพลาดเชิงบวก (FP Rate) และค่าอัตราความผิดพลาดเชิงลบ (FN Rate) จะพิจารณาจากค่าน้อยที่สุด ในการทำนายผลของวิธีการจำแนกกลุ่มทั้ง 5 วิธี จะใช้ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error : MAE) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการเปรียบเทียบ โดยพิจารณาจากค่า MAE และ MSE ที่มีค่าน้อยที่สุด ทำการเก็บรวบรวมข้อมูลจากฐานข้อมูลและแฟ้มข้อมูลของสำนักทะเบียนและประมวลผล สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง เกี่ยวกับตัวแปรอิสระ ซึ่งประกอบด้วย เพศ สาขาวิชา วิธีผ่านการสอบคัดเลือก คะแนนเฉลี่ยประจำภาคการศึกษาที่ 1 ของปีการศึกษา 2556 คะแนนเฉลี่ยประจำภาคการศึกษาที่ 2 ของปีการศึกษา 2556 และตัวแปรตามประกอบด้วยสถานภาพการเป็นนักศึกษา โดยแบ่งเป็นคงสภาพการเป็นนักศึกษาและพ้นสภาพการเป็นนักศึกษา (เนื่องจากคะแนนเฉลี่ยสะสมและไม่ได้ลงทะเบียนเรียน) รวมระยะเวลาดำเนินโครงการ 1 ปี

1.4 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

1.4.1 ทำให้ทราบวิธีที่มีประสิทธิภาพดีที่สุดในการพิจารณาสถานภาพการเป็นนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

1.4.2 ทำให้ทราบวิธีที่มีการทำนายผลดีที่สุดในการพิจารณาสถานภาพการเป็นนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

1.5 นิยามคำศัพท์

สถานภาพการเป็นนักศึกษา เป็นการคงสภาพการเป็นนักศึกษาและพ้นสภาพการเป็นนักศึกษา (เนื่องจากคะแนนเฉลี่ยสะสมและไม่ได้ลงทะเบียนเรียน)

วิธีความใกล้เคียงกันมากที่สุด (K-Nearest Neighbor (KNN) Method) เป็นวิธีที่ไม่มีการสร้างตัวแบบจากข้อมูลฝึกหัดเก็บไว้ ทำนายข้อมูลใหม่โดยอาศัยการเปรียบเทียบกับข้อมูลฝึกหัดจำนวน k ตัว ที่อยู่ใกล้เคียงกันมากที่สุด ใช้คำตอบของข้อมูลฝึกหัดที่อยู่ใกล้เคียงกันมากที่สุด k ตัว ที่พบมากที่สุดเป็นคำตอบ วิธีนี้ทำนายได้เฉพาะข้อมูลเชิงกลุ่ม (nominal data) เท่านั้น

วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree Method) ปืนตัวแบบทางคณิตศาสตร์เพื่อหาทางเลือกที่ดีที่สุดโดยการนำข้อมูลมาสร้างตัวแบบการพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ซึ่งมีการเรียนรู้ข้อมูลแบบมีผู้สอน (supervised learning) สามารถสร้างตัวแบบการจำแนกกลุ่มได้จากกลุ่มตัวอย่างของข้อมูลฝึกหัดโดยอัตโนมัติและสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจำแนกกลุ่มได้อีกด้วย

วิธีโครงข่ายประสาท (Neural Network Method) เป็นวิธีที่ใช้หลักการเลียนแบบการทำงานของสมองมนุษย์ เส้นเชื่อมแต่ละเส้นจะมีน้ำหนักถ่วง (weight) เพื่อใช้กำหนดน้ำหนักถ่วงหรือความสำคัญของข้อมูลเข้า (input data) กำหนดค่าเริ่มต้นโดยการสุ่ม ในแต่ละโหนดทำการคำนวณค่าผลรวมเชิงเส้นแบบถ่วงน้ำหนักและผ่านฟังก์ชันกระตุ้น (activation function) คำนวณ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าความคลาดเคลื่อน (error) ระหว่างคำตอบที่ทำนายได้กับเฉลย ถ้ามีความคลาดเคลื่อนเกิดขึ้น ระบบจะทำการปรับปรุงค่าน้ำหนักถ่วงของแต่ละการเชื่อมต่อ (connection) ทำนายข้อมูลได้ทั้งข้อมูลเชิงกลุ่ม (nominal data) และข้อมูลเชิงตัวเลข (numeric data) วิธีโครงข่ายประสาทอยู่ในหมวดวิธีที่เป็น Functions ใช้สมการทางคณิตศาสตร์ในการสร้างตัวแบบ

วิธีซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine Method) เป็นวิธีที่ใช้กระบวนการสอนเครื่องแบบมีผู้สอน (supervised learning) เพื่อให้สามารถสร้างตัวจำแนกข้อมูล (classifier) ที่มีความทั่วไป (generalize) สูง นั่นคือสามารถทำงานได้ดีกับตัวอย่างที่ไม่รู้จัก (unknown database) ด้วยกระบวนการปรับรูปแบบข้อมูลจากข้อมูลที่มีมิติต่ำ (low dimension dataset) บนพื้นที่ข้อมูลนำเข้า (input space) ให้อยู่ในรูปแบบของข้อมูลที่มีมิติสูง (high dimension dataset) บนพื้นที่ข้อมูลคุณลักษณะ (feature space) โดยใช้ฟังก์ชันในการปรับรูปแบบข้อมูลเรียกว่า**ฟังก์ชันเคอร์เนล (kernel function)** ซึ่งความสามารถดังกล่าวช่วยให้การสร้างตัวจำแนกข้อมูลด้วยสมการกำลังสอง (quadratic equation) บนพื้นที่ข้อมูลคุณลักษณะเป็นไปได้ง่ายขึ้นและมีความชัดเจนในการจำแนกกลุ่มมากยิ่งขึ้นด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและรายงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 วิธีความใกล้เคียงกันมากที่สุด (K-Nearest Neighbor (KNN) Method)

ไม่มีการสร้างตัวแบบจากข้อมูลฝึกหัดเก็บไว้ ทำนายข้อมูลใหม่โดยอาศัยการเปรียบเทียบกับข้อมูลฝึกหัดจำนวน k ตัว ที่อยู่ใกล้เคียงกันมากที่สุด ใช้คำตอบของข้อมูลฝึกหัดที่อยู่ใกล้เคียงกันมากที่สุด k ตัว ที่พบมากที่สุดเป็นคำตอบ วิธีนี้ทำนายได้เฉพาะข้อมูลเชิงกลุ่ม (nominal data) เท่านั้น

วิธีความใกล้เคียงกันมากที่สุดเป็นวิธีการที่ได้รับความนิยมในการใช้งานอย่างมาก เนื่องจากเป็นวิธีการที่ง่ายและมีประสิทธิภาพซึ่งสามารถนำไปประยุกต์ใช้กับงานได้หลายอย่าง เช่น งานทางด้านการจำแนกกลุ่ม รวมถึงงานทางด้านการแทนที่ข้อมูลสูญหาย ซึ่งมีขั้นตอนการดำเนินการดังนี้ (Tan, P-N. and et al. : 2006)

2.1.1.1 กำหนดค่า k เพื่อใช้พิจารณาสมาชิกที่อยู่ใกล้เคียงกันมากที่สุด เช่น k = 3 คือ จะพิจารณาเฉพาะข้อมูล 3 ตัวแรกที่อยู่ใกล้กับจุดที่ต้องการจะทำนาย

2.1.1.2 คำนวณหาระยะห่างระหว่างข้อมูลตัวอย่างที่สนใจกับข้อมูลอื่น ๆ ทุกตัว ด้วยวิธีระยะห่างยูคลิเดียน (euclidean distance) จากสมการดังนี้

$$\text{dist}(x_i, x_j) = \sqrt{\sum_{k=1}^{n}(x_{i,k} - x_{j,k})^2} \qquad (2.1)$$

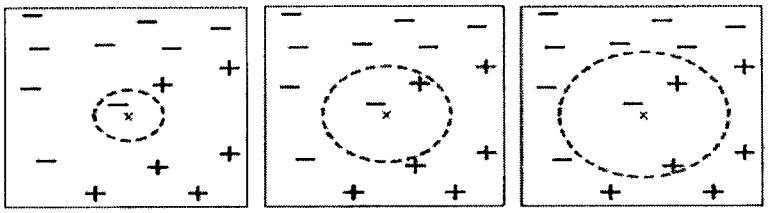
โดยที่ $\text{dist}(x_i, x_j)$ คือ ระยะห่างระหว่างตัวอย่าง x_i กับตัวอย่าง x_j

n คือ จำนวนคุณสมบัติทั้งหมดของตัวอย่าง

$x_{i,k}$ คือ คุณสมบัติที่ k ของตัวอย่าง x_i

2.1.1.3 เลือกค่าข้อมูลที่มีค่าระยะห่างน้อยที่สุด k ตัว เพื่อนำมาพิจารณาหาคำตอบ ดังรูปที่ 2.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



(a) ความใกล้เคียงกันมากที่สุดโดยพิจารณาจากข้อมูล 1 ตัว

(b) ความใกล้เคียงกันมากที่สุดโดยพิจารณาจากข้อมูล 2 ตัว

(c) ความใกล้เคียงกันมากที่สุดโดยพิจารณาจากข้อมูล 3 ตัว

รูปที่ 2.1 ตัวอย่างของวิธีความใกล้เคียงกันมากที่สุด

2.1.2 วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree Method)

วิธีแผนภาพต้นไม้เพื่อการตัดสินใจเป็นตัวแบบทางคณิตศาสตร์เพื่อหาทางเลือกที่ดีที่สุดโดยการนำข้อมูลมาสร้างตัวแบบการพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ซึ่งมีการเรียนรู้ข้อมูลแบบมีผู้สอน (supervised learning) สามารถสร้างตัวแบบการจำแนกกลุ่มได้จากกลุ่มตัวอย่างของข้อมูลฝึกหัดโดยอัตโนมัติ และสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจำแนกกลุ่มได้อีกด้วย

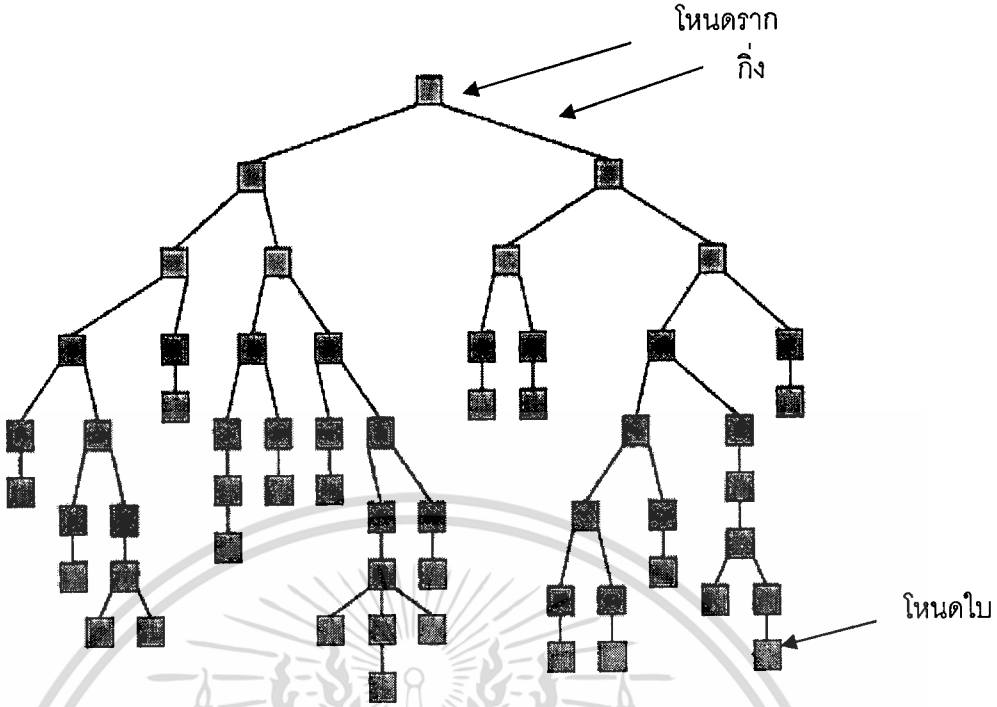
2.1.2.1 ส่วนประกอบของแผนภาพต้นไม้เพื่อการตัดสินใจ

1) โหนด (Node) คือคุณสมบัติต่าง ๆ เป็นจุดที่แยกข้อมูลว่าจะให้ไปในทิศทางใด ซึ่งโหนดที่อยู่สูงสุดเรียกว่า โหนดราก (root node)

2) กิ่ง (Branch) คือคุณสมบัติของโหนดที่แตกออกมาโดยจำนวนของกิ่งจะเท่ากับคุณสมบัติของโหนด

3) ใบ (Leaf) คือกลุ่มของผลลัพธ์ในการแยกแยะข้อมูล ซึ่งโหนดที่อยู่ล่างสุดเรียกว่า โหนดใบ (leaf node) โดยสามารถแสดงส่วนประกอบของแผนภาพต้นไม้เพื่อการตัดสินใจดังรูปที่ 2.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.2 ส่วนประกอบของแผนภาพต้นไม้เพื่อการตัดสินใจ

2.1.2.2 การสร้างแผนภาพต้นไม้เพื่อการตัดสินใจ

หลักการพื้นฐานของการสร้างแผนภาพต้นไม้เพื่อการตัดสินใจเป็นการสร้างจากบนลงล่าง คือ เริ่มจากการสร้างรากของต้นไม้ก่อน แล้วจึงแตกกิ่งไปจนถึงใบ โดยแสดงขั้นตอนการสร้างแผนภาพต้นไม้เพื่อการตัดสินใจได้ดังนี้

1) ต้นไม้เริ่มต้นโดยมีโหนดเพียงโหนดเดียวแสดงถึงชุดข้อมูลฝึกหัด (training data set)

2) ถ้าข้อมูลทั้งหมดอยู่ในกลุ่มเดียวกันแล้ว ให้โหนดนั้นเป็นใบและตั้งชื่อแยกตามกลุ่มของข้อมูลนั้น

3) ถ้าในโหนดมีข้อมูลหลายกลุ่มปะปนอยู่ จะต้องวัดค่าผลกำไร (gain) ของ แต่ละคุณลักษณะ (attribute) เพื่อที่จะใช้เป็นเกณฑ์ (criterion) ในการคัดเลือกคุณลักษณะที่มีความสามารถในการแบ่งข้อมูลออกเป็นกลุ่มต่าง ๆ ได้ดีที่สุด โดยคุณลักษณะที่มีผลกำไรมากที่สุดจะถูกเลือกให้เป็นตัวทดสอบหรือคุณลักษณะที่ใช้ในการตัดสินใจโดยแสดงในรูปของโหนดบนต้นไม้

4) กิ่งของต้นไม้ถูกสร้างขึ้นจากค่าต่าง ๆ ที่เป็นไปได้ของโหนดทดสอบ และข้อมูลจะถูกแบ่งออกตามกิ่งต่าง ๆ ที่สร้างขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) ทำการวนซ้ำเพื่อหาคุณลักษณะที่มีผลกำไรมากที่สุด สำหรับข้อมูลที่ถูกแบ่งแยกออกมาในแต่ละกิ่งเพื่อนำคุณลักษณะนี้มาสร้างเป็นโหนดตัดสินใจต่อไป โดยที่คุณลักษณะที่ถูกเลือกมาเป็นโหนดแล้วจะไม่ถูกเลือกมาอีกสำหรับโหนดในระดับต่อ ๆ ไป

6) ทำการวนซ้ำเพื่อแบ่งข้อมูลและแตกกิ่งของต้นไม้ไปเรื่อย ๆ โดยการวนซ้ำ จะสิ้นสุดก็ต่อเมื่อเงื่อนไขข้อใดข้อหนึ่งข้างบนนี้เป็นจริง

2.1.2.3 การคำนวณค่าผลกำไรสารสนเทศ (Information Gain)

แผนภาพต้นไม้เพื่อการตัดสินใจเป็นโครงสร้างที่ใช้แสดงกฎที่ได้จากเทคนิคการจำแนกกลุ่มของข้อมูล โดยแผนภาพต้นไม้เพื่อการตัดสินใจจะมีลักษณะคล้ายโครงสร้างต้นไม้ โดยที่แต่ละโหนดแสดงคุณลักษณะ ในการสร้างแผนภาพต้นไม้เพื่อการตัดสินใจ ปัญหาที่สำคัญที่ต้องพิจารณาคือควรจะตัดสินใจเลือกคุณลักษณะใดมาทำหน้าที่เป็นโหนดราก ในแต่ละขั้นตอนของการสร้างต้นไม้และต้นไม้ย่อย (subtree) ของแผนภาพต้นไม้เพื่อการตัดสินใจ เกณฑ์ที่ใช้ช่วยประกอบการเลือกคุณลักษณะเพื่อการคำนวณเกณฑ์ผลกำไร (gain criterion) ซึ่งเป็นค่าที่บ่งบอกว่าคุณลักษณะที่เป็นไปได้จากชุดข้อมูลมาทำหน้าที่เป็นโหนดราก ถ้าคุณลักษณะใดให้ผลกำไรสูงที่สุด แสดงว่าคุณลักษณะนั้นสามารถจำแนกกลุ่มของข้อมูลได้ดีที่สุด การใช้ผลกำไรสารสนเทศจะช่วยลดจำนวนครั้งของการทดสอบในการแยกแยะข้อมูล อีกทั้งยังรับประกันว่าแผนภาพต้นไม้เพื่อการตัดสินใจที่ได้จะไม่มีความซับซ้อนมากเกินไป (overfitting) ซึ่งผลกำไรสารสนเทศนั้นสามารถคำนวณได้ดังนี้

$$I(S_1,S_2,...,S_n) = -\sum_{i=1}^{n}\frac{S_i}{S}\log_2\frac{S_i}{S} \qquad (2.2)$$

เมื่อ S คือ เซตของข้อมูลซึ่งประกอบด้วยข้อมูล S ระเบียน (record)

n คือ จำนวนกลุ่มทั้งหมดที่ต่างกันของข้อมูลชุดนั้น

S_i คือ จำนวนข้อมูลที่เป็นสมาชิกของ S และอยู่ในกลุ่ม C_i

C_i คือ กลุ่มในลำดับที่ i โดยที่ i มีค่าระหว่าง 1 ถึง n

ค่าเอ็นโทรปี (Entropy) ของคุณลักษณะ A ซึ่งมีค่าของคุณลักษณะเป็น $(a_1,a_2,...,a_v)$ หาได้ดังนี้

$$E(A) = \sum_{j=1}^{v}\frac{S_{1j}+...+S_{nj}}{S} I(S_{1j}+...+S_{nj}) \qquad (2.3)$$

S_{ij} คือ จำนวนข้อมูลที่เป็นสมาชิกของ S และอยู่ในกลุ่ม C_i จากการแบ่งข้อมูลด้วยค่าที่เป็นไปได้ของคุณลักษณะ A

ดังนั้นจะสามารถพิจารณาเกณฑ์ผลกำไรได้ดังนี้

$$Gain(A) = I(S_{1j},S_{2j},...,S_{nj}) - E(A) \qquad (2.4)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อดี

1) เข้าใจได้ง่าย
2) สร้างกฎได้จากต้นไม้
3) เลือกเฉพาะคุณลักษณะ (attribute) ที่สำคัญในการสร้างตัวแบบ

ข้อเสีย

1) ใช้ได้กับคำตอบ (class) ที่เป็นข้อมูลเชิงกลุ่ม (nominal data) เท่านั้น
2) ความถูกต้องในการทำนายไม่สูง

การประยุกต์ใช้แผนภาพต้นไม้เพื่อการตัดสินใจ (Decision tree application)

แผนภาพต้นไม้เพื่อการตัดสินใจใช้ตอบคำถามที่ต้องการจำแนกประเภทข้อมูลที่ต้องการความเข้าใจประกอบ

ใช้ในการพิจารณาให้สินเชื่อแก่บุคลต่าง ๆ

ใช้ในการทำนายว่าลูกค้าคนไหนที่มีโอกาสจะยกเลิกการใช้บริการและเหตุผลเพราะอะไร

2.1.3 วิธีโครงข่ายประสาท (Neural Network Method)

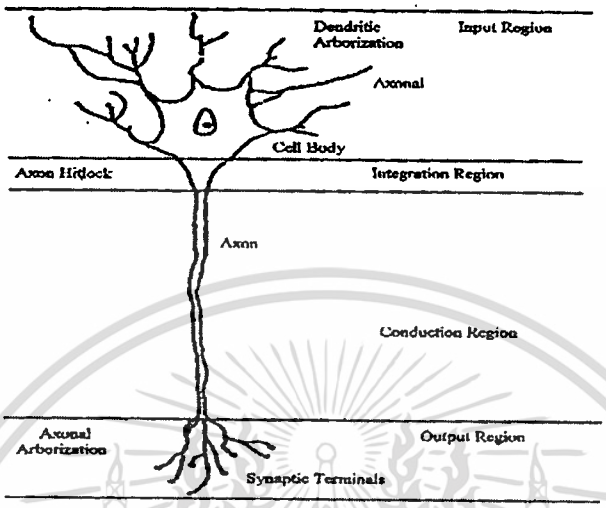
วิธีโครงข่ายประสาทใช้หลักการเลียนแบบการทำงานของสมองมนุษย์ เส้นเชื่อมแต่ละเส้น จะมีน้ำหนักถ่วง (weight) เพื่อใช้กำหนดน้ำหนักถ่วงหรือความสำคัญของข้อมูลเข้า (input data) กำหนดค่าเริ่มต้นโดยการสุ่ม ในแต่ละโหนดทำการคำนวณค่าผลรวมเชิงเส้นแบบถ่วงน้ำหนักและผ่านฟังก์ชันกระตุ้น (activation function) คำนวณค่าความคลาดเคลื่อน (error) ระหว่างคำตอบที่ทำนายได้กับเฉลย ถ้ามีความคลาดเคลื่อนเกิดขึ้น ระบบจะทำการปรับปรุงค่าน้ำหนักถ่วงของแต่ละการเชื่อมต่อ (connection) ทำนายข้อมูลได้ทั้งข้อมูลเชิงกลุ่ม (nominal data) และข้อมูลเชิงตัวเลข (numeric data) วิธีโครงข่ายประสาทอยู่ในหมวดวิธีที่เป็น Functions ใช้สมการทางคณิตศาสตร์ในการสร้างตัวแบบ

โครงข่ายประสาทเป็นศาสตร์ที่จำลองแบบความสามารถของมนุษย์ด้านการเรียนรู้ จดจำและจำแนกสิ่งต่าง ๆ ซึ่งใช้สมองเป็นส่วนสำคัญ ในการประมวลระบบของโครงข่ายประสาทนั้นจะเลียนแบบการทำงานของระบบสมองคือมีการส่งผ่านข้อมูลระหว่างกันโดยมีการเชื่อมต่อของเซลล์ประสาท (neuron) กันเป็นโครงข่ายร่างแหจำนวนมากและมีการประมวลผลในลักษณะขนาน (parallel processing) สาเหตุหลักที่โครงข่ายประสาทเป็นที่นิยมมากขึ้นเนื่องจากมีความยืดหยุ่นในการทำงานสูงและสามารถปรับตัวเองให้ทำงานในสภาพที่เปลี่ยนแปลงได้ดี อีกทั้งยังไม่จำเป็นต้องทราบตัวแบบทางคณิตศาสตร์ที่แน่นอนของกระบวนการ เพียงแต่ใช้ชุดข้อมูลที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประกอบด้วยข้อมูลเข้า (input data) และข้อมูลเป้าหมาย (target data) ของกระบวนการในจำนวนที่มากพอมาใช้ในการสอนโครงข่ายประสาท

2.1.3.1 ความรู้พื้นฐานของระบบประสาท (Neural system knowledge)



รูปที่ 2.3 โครงข่ายของเซลล์ประสาท

ภายในสมองมนุษย์ประกอบด้วยหน่วยประมวลผลขนาดเล็ก เรียกว่า เซลล์ประสาท (neuron) ซึ่งจะมีประมาณ 10 หน่วย ในเซลล์ประสาทแต่ละหน่วยดังแสดงในรูปที่ 2.3 ประกอบด้วย ใยประสาท (dendrites) ตัวเซลล์ (cell body) และเส้นใยประสาท (axon) ซึ่งแบ่งออกเป็น 4 บริเวณคือ

1) บริเวณนำกระแสประสาทเข้า (Input region) เป็นบริเวณที่จะมีการนำกระแสประสาท (nerve impulse) จากเซลล์ประสาทอื่นเข้ามาภายในตัวเซลล์โดยผ่านทางใยประสาทซึ่งมีลักษณะแตกเป็นกิ่งก้านคล้ายต้นไม้และมีจำนวนตั้งแต่ 1 ใยขึ้นไป

2) บริเวณการรวมกระแสประสาทเข้า (Integration region) เป็นบริเวณที่มีการรวมกระแสประสาทก่อนที่จะเข้าสู่บริเวณการนำแสประสาทรวมออกจากเซลล์

3) บริเวณการนำกระแสประสาทรวมออกจากเซลล์ (Conduction region) เป็นบริเวณที่จะนำกระแสประสาทรวมออกจากเซลล์โดยใช้เส้นใยประสาทเป็นทางผ่านซึ่งมีเพียง 1 เส้นใยต่อเซลล์เท่านั้น

4) บริเวณการนำกระแสประสาทรวมออก (Output region) เป็นบริเวณส่วนปลายของเส้นใยประสาทที่มีการแตกแขนงใช้ในการถ่ายทอดกระแสประสาทข้ามเซลล์ไปยังเซลล์ประสาทอื่นโดยผ่านทางใยประสาทของเซลล์ประสาทนั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.3.2 การเรียนรู้ของโครงข่ายประสาท (Neural network learning)

การเรียนรู้ของโครงข่ายประสาทจะมีประสิทธิภาพเพียงใดนั้นขึ้นอยู่กับค่าถ่วงน้ำหนัก (weight) ของโครงข่ายที่ทำการออกแบบ ซึ่งการฝึกหัดโครงข่ายประสาทคือการหาค่าถ่วงน้ำหนักที่เหมาะสมให้กับโครงข่ายประสาทนั้น ๆ โดยทั่วไปสามารถจำแนกวิธีการเรียนรู้ของโครงข่ายประสาทได้เป็น 2 ประเภท คือ การเรียนรู้แบบมีผู้สอนและการเรียนรู้แบบไม่มีผู้สอน

1) การเรียนรู้แบบมีผู้สอน (Supervised learning)

การเรียนรู้แบบมีผู้สอนจะกำหนดข้อมูลฝึกหัด (training data set) ให้กับโครงข่ายประสาท ซึ่งกลุ่มนี้ประกอบด้วยข้อมูลเข้า (Input data) และข้อมูลเป้าหมาย (target data) ที่ต้องการ จากนั้นโครงข่ายประสาทจะทำการคำนวณค่าถ่วงน้ำหนักที่เหมาะสมให้กับข้อมูลฝึกหัด โดยคำตอบที่ได้จากโครงข่ายประสาทจะถูกคำนวณค่าความผิดพลาด (error value) ว่ามีความห่างจากคำตอบที่ต้องการของข้อมูลนำเข้าในชุดเดียวกันมากน้อยเพียงใด ถ้ายังมีความผิดพลาดสูงอยู่ การฝึกหัดจะดำเนินต่อไปจนกว่าค่าความผิดพลาดจะลดลงต่ำกว่าค่าที่ยอมรับได้ (accept level) จึงจะหยุดฝึกหัด สุดท้ายค่าถ่วงน้ำหนักที่ได้จะเป็นเหมือนฟังก์ชันที่ใช้ในการแปลงข้อมูล

2) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning)

การเรียนรู้แบบไม่มีผู้สอนจะอาศัยชุดข้อมูลเข้าเพียงอย่างเดียวในการฝึกหัดโครงข่ายประสาทโดยไม่มีข้อมูลเป้าหมาย แต่จะใช้ข้อมูลออก (output data) จากโครงข่ายประสาทแทน เมื่อป้อนข้อมูลเข้าสู่โครงข่ายประสาท โครงข่ายประสาทจะคำนวณค่าความสัมพันธ์ที่มีอยู่ภายในกลุ่มข้อมูลเข้า โดยอาศัยค่าถ่วงน้ำหนักเป็นตัวแยกความแตกต่างของข้อมูลเข้าและนำไปเก็บไว้ในโหนดข้อมูลออกของโครงข่ายประสาท ซึ่งมีวัตถุประสงค์เพื่อใช้ในการจำแนกชุดข้อมูล (classification)

2.1.3.3 การเชื่อมโยงของโครงข่ายประสาทเทียม (Neural network linking)

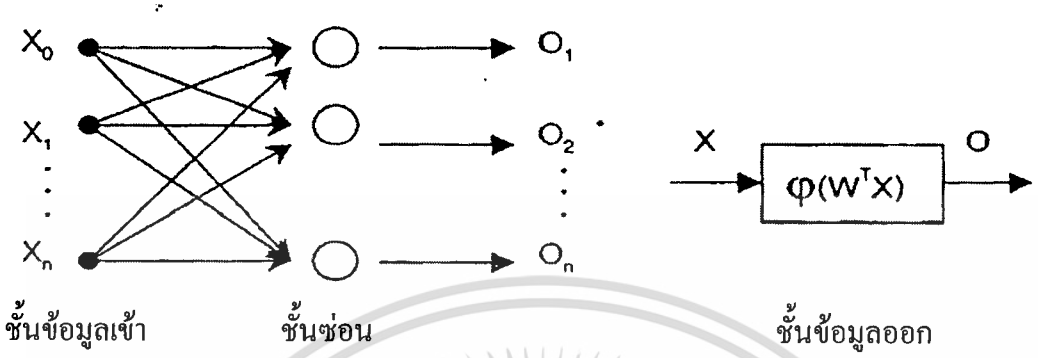
เพื่อให้โครงข่ายประสาทสามารถเรียนรู้ได้อย่างมีประสิทธิภาพ จำเป็นต้องมีการเชื่อมโยงกันระหว่างเซลล์ประสาท โดยทั่วไปสามารถแบ่งการเชื่อมโยงของโครงข่ายได้ 2 ลักษณะ คือ

1) โครงข่ายแบบส่งสัญญาณไปข้างหน้า (Feedforward network)

เป็นโครงข่ายที่การประมวลผลจะอาศัยชุดข้อมูลปัจจุบันและส่งค่าที่ประมวลผลได้ไปยังชั้นถัด ๆ ไป กล่าวคือ โครงข่ายชนิดนี้จะประกอบด้วยชั้นต่าง ๆ โดยชั้นแรกจะเป็นชั้นข้อมูลเข้า (input layer) และชั้นสุดท้ายเป็นชั้นข้อมูลออก (output layer) ส่วนระหว่างชั้นข้อมูลเข้ากับชั้นข้อมูลออกอาจจะมีหรือ ไม่มีชั้นซ่อน (hidden layer) อยู่ภายในก็ได้ ซึ่งขึ้นกับกฎการเรียนรู้ (learning rule) ที่ใช้ในการสอนโครงข่าย เช่น ถ้าเป็นโครงข่ายเพอร์เซปตรอนแบบหลายชั้น (multi-layer perceptron) จะมีชั้นซ่อนอยู่ระหว่างชั้นข้อมูลเข้ากับชั้นข้อมูลออก ซึ่งอาจมี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

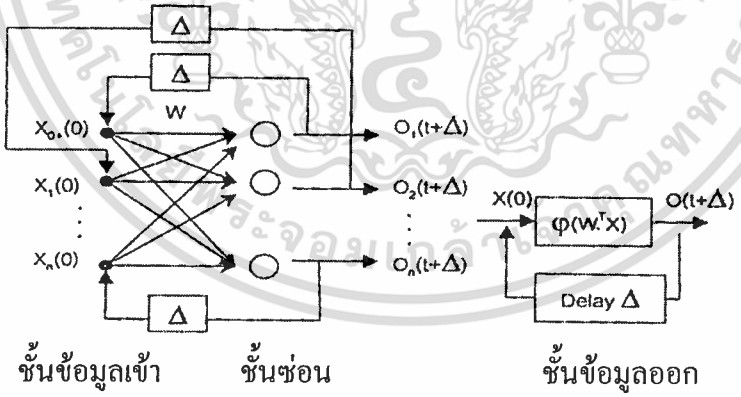
มากกว่าหนึ่งชั้นได้ การเชื่อมต่อระหว่างชั้นของโครงข่ายแบบส่งสัญญาณไปข้างหน้าจะมีค่าถ่วงน้ำหนัก (weight) เป็นตัวเชื่อมและสัญญาณนำเข้าที่เข้ามาจะถูกส่งไปตามทิศทางของลูกศรจนถึงชั้นข้อมูลออกโดยไม่มีการย้อนกลับ สามารถแสดงตัวแบบโครงข่ายแบบส่งสัญญาณไปข้างหน้าได้ดังรูปที่ 2.4



รูปที่ 2.4 ลักษณะโครงข่ายประสาทเทียมแบบส่งสัญญาณไปข้างหน้า

2) โครงข่ายแบบมีการย้อนกลับ (Feedback network)

โครงข่ายชนิดนี้มีชื่อเรียกอีกชื่อหนึ่งว่า โครงข่ายหันกลับ (recurrent network) เป็นโครงข่ายที่จะอาศัยทั้งข้อมูลในปัจจุบันและข้อมูลที่มีการประวิงเวลามาใช้ในการประมวลผลของโครงข่ายประสาท สามารถแสดงตัวแบบโครงข่ายแบบมีการย้อนกลับได้ดังรูปที่ 2.5



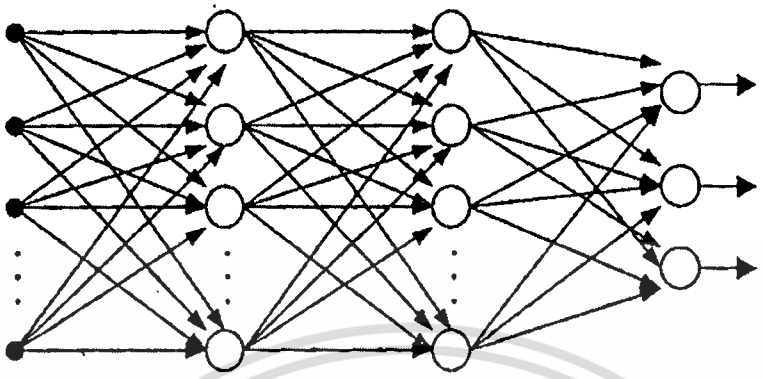
รูปที่ 2.5 ลักษณะโครงข่ายประสาทเทียมแบบมีการย้อนกลับ

2.1.3.4 การแพร่แบบย้อนกลับ (Back-propagation)

การแพร่แบบย้อนกลับเป็นขั้นตอนที่ใช้สอนโครงข่ายประสาทแบบเพอร์เซปตรอนหลายชั้น (multi-layer perceptron) ซึ่งตัวแบบโครงข่ายประสาทมีการเชื่อมโยงกันเป็นโครงข่ายแบบเป็นชั้น ๆ โครงข่ายชนิดนี้มีการเชื่อมโยงกัน 3 ชั้น ประกอบด้วยชั้นข้อมูลเข้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(input layer) ถัดมาเป็นชั้นซ่อน (hidden layer) และชั้นสุดท้ายคือชั้นข้อมูลออก (output layer) สามารถแสดงโครงข่ายประสาทแบบเพอร์เซปตรอนหลายชั้นที่มีชั้นซ่อน 2 ชั้น ได้ดังรูปที่ 2.6



ชั้นข้อมูลเข้า ชั้นซ่อนที่ 1 ชั้นซ่อนที่ 2 ชั้นข้อมูลออก

(Input layer) (First hidden layer) (Second hidden layer) (Output layer)

รูปที่ 2.6 ลักษณะโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น

ที่มาของชื่อการแพร่แบบย้อนกลับนั้นมาจากจุดที่ว่า วิธีการปรับค่าถ่วงน้ำหนักเพื่อให้ได้ค่าที่เหมาะสมนั้นจะใช้วิธีสอนว่าค่าเป้าหมาย (target) ของแต่ละข้อมูลเข้านั้นคืออะไร และใช้ค่าความผิดพลาด (error) ของข้อมูลออกมาใช้เป็นตัวชี้นำในการปรับค่าถ่วงน้ำหนัก ดังนั้นการแพร่แบบย้อนกลับ จึงเป็นกระบวนการเรียนรู้แบบมีผู้สอน แต่ปัญหาที่เกิดขึ้นคือไม่มีค่าเป้าหมายของสัญญาณที่ออกมาจาก แต่ละเซลล์ประสาทในชั้นซ่อน ดังนั้นจึงต้องอาศัยการแพร่ความผิดพลาดจากชั้นข้อมูลออกกลับมายังชั้นซ่อนนั่นเอง

2.1.3.5 ปัจจัยที่ส่งผลต่อการเรียนรู้การแพร่แบบย้อนกลับ

1) การกำหนดค่าเริ่มต้นของค่าถ่วงน้ำหนัก

ก่อนที่จะทำการสอนโครงข่ายประสาทแบบหลายชั้น จำเป็นต้องกำหนดค่าเริ่มต้นให้กับ ค่าถ่วงน้ำหนักที่เชื่อมโยงระหว่างชั้นทุกชั้น โดยค่านี้จะเป็นเลขจำนวนจริงที่มีค่าน้อย ๆ ที่ได้มาจากการสุ่ม ค่าเริ่มต้น (randomness)

2) การกำหนดเกณฑ์การหยุดฝึกหัด

เกณฑ์ในการหยุดฝึกหัดนั้นขึ้นกับผู้ที่ทำการออกแบบโครงข่ายประสาทว่าต้องการที่จะให้โครงข่ายประสาทมีความแม่นยำเพียงใด โดยทั่วไปนิยมใช้ค่าดัชนีที่ชี้ถึงค่าความผิดพลาดของระบบได้ ในงานวิจัยส่วนใหญ่ใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error, MSE)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) อัตราการเรียนรู้ (Learning rate, η)

อัตราการเรียนรู้เป็นค่าสัมประสิทธิ์ที่แสดงถึงการเรียนรู้ของโครงข่าย โดยทั่วไปค่าที่เหมาะสมจะอยู่ในช่วง 0.05 ถึง 0.5 ถ้าอัตราการเรียนรู้มีค่าสูง แสดงว่ากำหนดให้โครงข่ายมีการเปลี่ยนแปลงค่าถ่วงน้ำหนักที่มาก ในทางตรงกันข้ามถ้ามีอัตราการเรียนรู้ต่ำ แสดงว่ากำหนดให้โครงข่ายมีการเปลี่ยนแปลงค่าถ่วงน้ำหนักที่น้อย ซึ่งจำเป็นต้องใช้เวลาในการเรียนรู้ที่มากขึ้น แต่จะมีข้อดีคือโครงข่ายจะมีเสถียรภาพและไม่เกิดการแกว่ง (oscillation) ขณะที่ทำการเรียนรู้

4) ค่าคงที่โมเมนตัม (Momentum constant, α)

ค่าคงที่โมเมนตัมเป็นค่าสัมประสิทธิ์ที่ช่วยหน่วงไม่ให้การเปลี่ยนแปลงค่าถ่วงน้ำหนักนั้นมีค่ามากเกินไป เป็นการเพิ่มเสถียรภาพให้กับโครงข่ายประสาทได้อีกทางหนึ่ง ซึ่งค่าโมเมนตัมที่เหมาะสมจะมีค่าเข้าใกล้ 1.0 และควรจะกำหนดให้สอดคล้องกับอัตราการเรียนรู้ด้วย เช่น ถ้าอัตราการเรียนรู้สูงก็ควรที่จะมีค่าโมเมนตัมที่ต่ำ ทำให้การเปลี่ยนแปลงค่าถ่วงน้ำหนักนั้นไม่มากจนเกินไป

2.1.4 วิธีซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine Method)

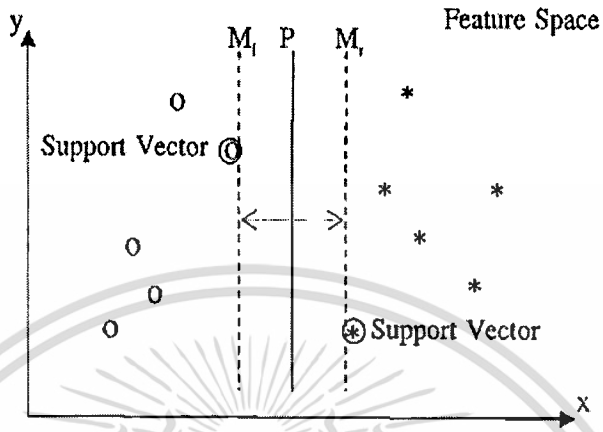
เป้าหมายของวิธีซัพพอร์ทเวกเตอร์แมชชีนคือกระบวนการสอนเครื่องแบบมีผู้สอน (supervised learning) เพื่อให้สามารถสร้างตัวจำแนกข้อมูล (classifier) ที่มีความทั่วไป (generalize) สูง นั่นคือสามารถทำงานได้ดีกับตัวอย่างที่ไม่รู้จัก (unknown database) ด้วยกระบวนการปรับรูปแบบข้อมูลจากข้อมูลที่มีมิติต่ำ (low dimension dataset) บนพื้นที่ข้อมูลนำเข้า (input space) ให้อยู่ในรูปแบบของข้อมูลที่มีมิติสูง (high dimension dataset) บนพื้นที่ข้อมูลคุณลักษณะ (feature space) โดยใช้ฟังก์ชันในการปรับรูปแบบข้อมูลเรียกว่า**ฟังก์ชันเคอร์เนล (kernel function)** ซึ่งความสามารถดังกล่าวช่วยให้การสร้างตัวจำแนกข้อมูลด้วยสมการกำลังสอง (quadratic equation) บนพื้นที่ข้อมูลคุณลักษณะเป็นไปได้ง่ายขึ้นและมีความชัดเจนในการจำแนกกลุ่มมากยิ่งขึ้นด้วย นอกจากนี้ตัวจำแนกข้อมูลที่ดีควรมีโครงสร้างแบบเส้นตรง (linear classifier) และสามารถสร้างพื้นที่ระยะห่างระหว่างตัวจำแนกข้อมูลกับค่าที่ใกล้ที่สุดของแต่ละกลุ่มข้อมูลได้มากที่สุดเพื่อประสิทธิภาพในการแยกแยะประเภทของชุดข้อมูลแต่ละประเภทออกจากกันอย่างชัดเจน ซึ่งเส้นที่เหมาะสมดังกล่าวเรียกว่า ระนาบแบ่งเขตข้อมูลที่เหมาะสม (optimal separating hyperplane)

2.1.4.1 แนวความคิดของซัพพอร์ทเวกเตอร์แมชชีน

ซัพพอร์ทเวกเตอร์แมชชีนเป็นสมการที่ใช้ในการจำแนกค่าคุณลักษณะของ 2 กลุ่ม ที่วางตัวอยู่ในพื้นที่คุณลักษณะ (feature space) ออกจากกันโดยจะสร้างเส้นแบ่ง (plane) ที่เป็นเส้นตรงขึ้นมาและเพื่อให้ทราบว่าเส้นตรงที่แบ่ง 2 กลุ่มออกจากกันนั้น เส้นตรงใดที่เป็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เส้นที่ดีที่สุด โดยเส้นตรงนั้นจะเพิ่มเส้นขอบ (margin) ออกไปทั้งสองข้าง โดยเส้นขอบที่เพิ่มนั้นจะขนานกับเส้นเดิมเสมอ เส้นขอบที่เพิ่มขึ้นมานี้จะขยายออกไปจนกว่าจะสัมผัสกับค่าของกลุ่มตัวอย่างที่ใกล้ที่สุด ดังรูปที่ 2.7



รูปที่ 2.7 การขยายตัวของเส้นขอบ

จากรูปที่ 2.7 เส้น M_l และ M_r คือเส้นขอบที่ขยายออกไปด้านซ้ายและขวา ตามลำดับ และ P คือเส้นแบ่งข้อมูลทั้ง 2 กลุ่ม เมื่อเส้น M_l และ M_r ขยายออกจนไปสัมผัสค่าข้อมูลที่ใกล้ที่สุด ซึ่งข้อมูลที่อยู่บนเส้นขอบของทั้งสองฝั่งนั้นเรียกว่า **ซัพพอร์ทเวกเตอร์ (support vector)** จะวัดค่าระยะความห่างของเส้นขอบ โดยเส้น P จะเปลี่ยนความชันไปเรื่อย ๆ เพื่อที่จะหาความกว้างสูงสุดของเส้นขอบ

กระบวนการโดยรวมของซัพพอร์ทเวกเตอร์แมชชีนนั้นเป็นการหาค่าความชันของเส้น P ที่มีขนาดของเส้นขอบสูงสุด

2.1.4.2 สมการพื้นฐานของซัพพอร์ทเวกเตอร์แมชชีน

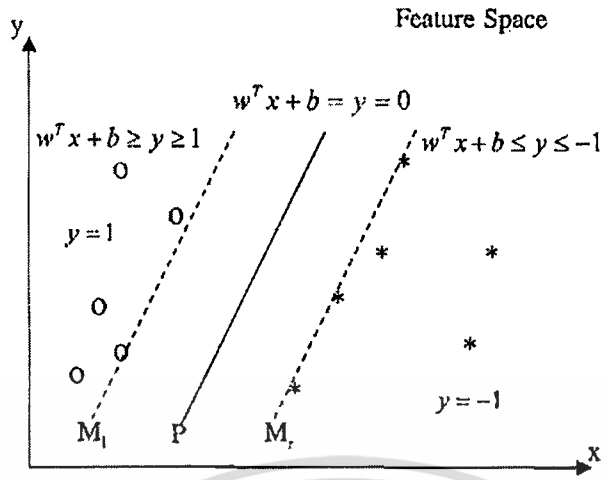
ถ้านำแนวคิดของซัพพอร์ทเวกเตอร์แมชชีนที่กล่าวไปแล้วในข้อ 2.1.4.1 มาเขียนเป็นสมการเพื่อใช้ในการแก้ปัญหา โดยข้อมูลที่นำมาวางลงในพื้นที่คุณลักษณะนั้นเป็นกลุ่มข้อมูลที่อยู่ในรูปของเวกเตอร์

$$x = ((x_1, y_1),...,(x_i, y_i)) \quad (2.5)$$

เมื่อ x คือ ชุดค่าคุณลักษณะ

ค่าคุณลักษณะที่วางตัวอยู่ในพื้นที่คุณลักษณะจะถูกแบ่งด้วยเส้นตรงดังรูปที่ 2.8 และเมื่อนำเส้นตรงมาแทนค่าด้วยสมการเส้นตรง $y = mx + b$ โดยมีการกำหนดกลุ่มของข้อมูลทั้งสองฝั่งเป็นเพียง 2 ค่า ที่ซึ่งแทนด้วยค่า y เพื่อให้ข้อมูลที่อยู่ในกลุ่มเดียวกันที่มาจากหลายค่ากลายเป็นค่าเดียว ดังสมการในรูปที่ 2.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.8 เส้นขอบและเส้นแบ่งเมื่อแทนด้วยสมการเส้นตรง

จากรูปที่ 2.8 เส้นตรง M_l แทนด้วยสมการ $w^T x + b \geq y \geq 1$ ซึ่งข้อมูล y ที่มากกว่า 1 ก็จะถูกกำหนดค่าใหม่โดยให้ y เท่ากับ 1 และพจน์ w ก็คือค่าความชัน เช่นเดียวกันกับเส้นตรง M_r ที่ค่าของ y จะถูกกำหนดค่าใหม่เมื่อ y ที่น้อยกว่า –1 ให้เท่ากับ –1 ดังนั้นสมการที่เกิดขึ้นใหม่ จากสมการเส้นขอบ 2.6 และ 2.7 สามารถกำหนดได้ดังสมการที่ 2.8

เมื่อ $w^T x + b \geq y$ กำหนด $y = 1$ (2.6)

$w^T x + b \leq y$ กำหนด $y = -1$ (2.7)

$y(w^T x + b) - 1 \geq 0$ (2.8)

โดย y คือ ค่ากลุ่มข้อมูล (1, –1)

w คือ ค่าความชัน

x คือ ค่าคุณลักษณะ

b คือ ค่าคงที่ (ค่าตัดแกน y)

2.1.4.3 ค่าความกว้างเส้นขอบ (Margin)

การคำนวณความกว้างของเส้นขอบต้องทำการคำนวณพจน์ w ให้อยู่ในรูปปกติมาตรฐาน (normalization) โดยคำนวณจากสมการที่ 2.6 และ 2.7 เมื่อแทนค่า y ลงไปแล้ว

$$w^T x^- + b = 1$$

$$w^T x^- + b = -1$$

$$w^T (x^+ - x^-) = 2$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$M = \left(\frac{w}{\|w\|}\right)^T \left(x^+ - x^-\right)$$

$$= \frac{2}{\|w\|} \quad (2.9)$$

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i \quad (2.10)$$

โดยที่ M คือ ความกว้างของเส้นขอบ

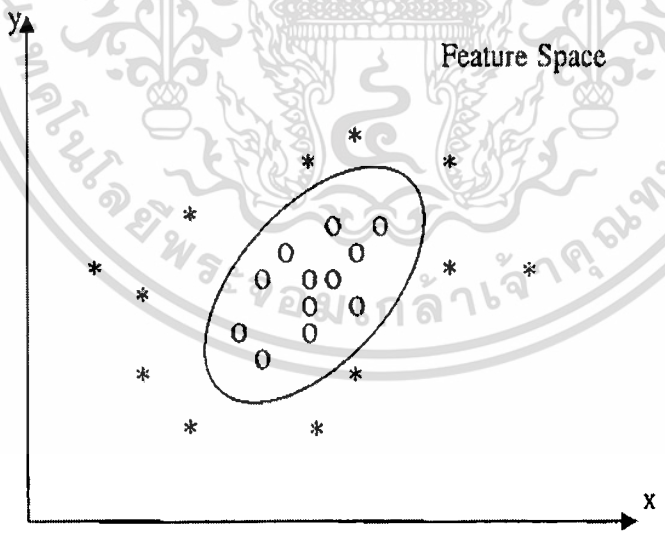
α คือ สัมประสิทธิ์คงที่

เมื่อนำค่า w ไปใส่ในสมการที่ 2.8 ซึ่งเป็นสมการในการหาเส้นแบ่ง จะได้

$$y_i\left(\sum_{i=1}^{N} \alpha_i y_i \left(x_i^T, x_j\right) + b\right) - 1 \geq 0 \quad (2.11)$$

2.1.4.4 เคอร์เนล (Kernel)

ในความเป็นจริงนั้นข้อมูล 2 กลุ่ม ไม่ได้วางตัวในพื้นที่คุณลักษณะ และไม่สามารถแบ่งได้โดยเส้นตรง แต่ข้อมูลอาจจะจับกลุ่มกันในตำแหน่งต่าง ๆ ดังนั้นจึงเป็นปัญหาทำให้ไม่สามารถที่จะใช้สมการซัพพอร์ทเวกเตอร์แมชชีนแบบเชิงเส้นได้ ดังนั้นจะต้องมีเครื่องมือมาช่วยให้ข้อมูลเหล่านั้นเรียงตัวใหม่ในพื้นที่ เรียกว่า **พื้นที่หลายมิติ (higher dimensional space)**



รูปที่ 2.9 รูปแบบการวางตัวที่ไม่สามารถแบ่งด้วยเส้นตรงได้

ในเคอร์เนลนั้นคือการคูณกันของชุดเวกเตอร์ของ x ใด ๆ

$$K\left(x_i, x_j\right) = x_i^T x_j \quad (2.12)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เคอร์เนลที่นิยมใช้มีอยู่ 3 ชนิด คือ

1) โพลิโนเมียล (Polynomial)

$$K(x_i, x_j) = (< x_i, x_j > +1)^d \quad (2.13)$$

เมื่อ d คือ ค่าเลขยกกำลัง

2) ฟังก์ชันเบสิสเรเดียล (Radial Basis Function : RBF)

$$K(x_i, x_j) = \exp\left(-\frac{x_i - x_j^2}{2\sigma^2}\right) \quad (2.14)$$

เมื่อ σ คือ ค่าพารามิเตอร์

3) ซิกมอยด์ (Sigmoid)

$$K(x_i, x_j) = \tanh(k < x_i, x_j > +\mu) \quad (2.15)$$

เมื่อ k, μ คือค่าพารามิเตอร์

ดังนั้นจากสมการของเคอร์เนลนั้นสามารถที่จะแทนลงไปในตำแหน่งของ x_i^T, x_j ในสมการที่ 2.11 จึงเขียนเป็นสมการใหม่ดังนี้

$$y_i \sum_{i=1}^{N} (\alpha_i y_i K(x_i, x_j) + b) - 1 \geq 0 \quad (2.16)$$

สมการที่ 2.16 เป็นสมการที่ใช้ในขั้นตอนที่จะเรียนรู้ว่าจะวางตำแหน่งเส้นแบ่งไว้ที่ตำแหน่งใด โดยทำงานร่วมกับเคอร์เนล เพื่อแปลงให้ข้อมูลที่ยากต่อการแบ่งแบบเชิงเส้นสามารถแบ่งได้เมื่อทำให้เป็นข้อมูลแบบหลายมิติ (higher dimension) ดังนั้นจึงมีอีกสมการหนึ่งที่ใช้ค่า w และ b เดิมมาจัดตำแหน่งของข้อมูลเพื่อที่ให้ทราบว่าข้อมูลนั้นเป็นกลุ่มใด กำหนดได้ดังสมการที่ 2.17

$$f(x) = \text{sgn}\left(\sum_{i=1}^{\infty} (\alpha_i y_i K(x_i, x_j) + b)\right) \quad (2.17)$$

เมื่อ $f(x)$ คือค่า y หาในรูปของ x

2.1.5 วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม (Binary Logistic Regression Method) เป็นการวิเคราะห์การถดถอยแบบหนึ่งโดยที่ตัวแปรตามเป็นตัวแปรเชิงคุณภาพที่มีค่าได้เพียง 2 ค่า (dichotomous or binary variable) ส่วนตัวแปรอิสระอาจจะเป็นตัวแปรเชิงปริมาณหรือเชิงคุณภาพ หรืออาจจะมีทั้งตัวแปรเชิงปริมาณและตัวแปรเชิงคุณภาพก็ได้ (กัลยา วานิชย์บัญชา, 2552)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 รายงานวิจัยที่เกี่ยวข้อง

ชลนิศา สาระ (2550) เสนอวิธีการจำแนกกลุ่มสถานภาพการสำเร็จการศึกษาโดยแบบจำลอง (model) ต้นไม้เพื่อการตัดสินใจซึ่งใช้อัลกอริทึม C4.5 ทำให้เข้าใจถึงกระบวนการพารามิเตอร์ที่ใช้และการวัดประสิทธิภาพในการสร้างแบบจำลองต้นไม้สำหรับอัลกอริทึม C4.5

ภัทรพงศ์ พงศ์ภัทรกานต์ (2552) ได้นำเสนอการเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลของแบบจำลอง C5.0, CART, SVM และ SVM ร่วมกับ C5.0 ภายใต้หลักการทำงานของเหมืองข้อมูลโดยใช้ชุดข้อมูลจำนวน 9 ชุด ทำการเลือกชุดข้อมูลที่เป็นไบนารีคลาสและมีคุณลักษณะ (attribute) กับจำนวนตัวอย่างที่หลากหลายรูปแบบ SVM จะทำการคัดแยกชุดข้อมูลออกเป็นไบนารีคลาสได้ดีมาเปรียบเทียบประสิทธิภาพในการจำแนกกลุ่มข้อมูลโดยใช้ C5.0, CART, SVM และ SVM ผสมผสานกับ C5.0 โดยทำการทดลองวัดประสิทธิภาพความถูกต้องเปรียบเทียบกัน ซึ่งผลการทดลองพบว่าแบบจำลอง SVM ผสมผสานกับ C5.0 ที่ผู้วิจัยนำเสนอมีประสิทธิภาพสูงที่สุดทุกชุดข้อมูลจากจำนวน 9 ชุด ที่ได้ทำกาทดลอง ซึ่งสรุปผลได้ว่าการใช้ตัวแบบผสมผสานกันสามารถจำแนกประเภทข้อมูลประเภทไบนารีคลาสได้อย่างมีประสิทธิภาพสูงกว่าการใช้ตัวแบบ C5.0, CART, และ SVM เพียงอย่างเดียว

ภัทรพงศ์ พงศ์ภัทรกานต์ (2553) ได้นำเสนอการวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาระดับปริญญาตรี คอมมิตตีแมชชีนเป็นการทำงานระหว่าง SVM ร่วมกับ C5.0 โดยได้ทำการทดลองวัดประสิทธิภาพความถูกต้องเปรียบเทียบกับโครงข่ายประสาทและ C5.0 ซึ่งแบบจำลองคอมมิตตีแมชชีนมีประสิทธิภาพในการจำแนกข้อมูลสูงที่สุด มีค่าเท่ากับ 75.32

วลัยลักษณ์ สุขสมบุญ และสมชาย ปราการเจริญ (2553) ได้นำเสนอการเปรียบเทียบวิธีการจำแนกประเภทของปัญหาสำหรับ Helpdesk ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน วิธีนาอีฟเบย์และวิธีความใกล้เคียงกันมากที่สุด โดยนำข้อมูลมาตัดคำภาษาอังกฤษด้วยวิธีการอ้างอิงพจนานุกรมและคลังคำศัพท์เฉพาะที่ผู้วิจัยสร้างขึ้นใหม่ จากนั้นนำข้อมูลปัญหา Helpdesk แปลงให้อยู่ในรูปของเวกเตอร์ทางคณิตศาสตร์และทำการทดลองด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน วิธีนาอีฟเบย์และวิธีความใกล้เคียงกันมากที่สุดโดยใช้การตรวจสอบไขว้กับ K-Fold = 10 โดยทดลองปรับค่าพารามิเตอร์ของฟังก์ชันแกน Linear, Polynomial และ Radius Basic Function ในวิธีซัพพอร์ตเวกเตอร์แมชชีน วิธีนาอีฟเบย์และวิธีความใกล้เคียงกันมากที่สุด ผลการทดลองสามารถสรุปได้ว่าการทดลองเปรียบเทียบวิธีการจำแนกประเภทของปัญหาสำหรับ Helpdesk แสดงให้เห็นว่าเทคนิคที่นำเสนอด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความถูกต้องที่ดีที่สุดโดยใช้ฟังก์ชันแกน Polynomial

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วาทินี นุ้ยเพียร และคณะ (2553) ได้ทำการศึกษาเพื่อเปรียบเทียบวิธีการจำแนกข้อมูลโดยเลือกใช้อัลกอริทึมโครงข่ายประสาทแบบมัลติเลเยอร์เปอร์เซ็ปตรอน ซัพพอร์ตเวกเตอร์แมชชีน นาอีฟเบย์และความใกล้เคียงกันมากที่สุดเพื่อประเมินประสิทธิภาพค่าความถูกต้อง (accuracy) ค่าความแม่นยำ (precision) ค่าความระลึก (recall) และค่าความถ่วงดุล (F-measure) ใช้ข้อมูลจาก UCI ประกอบด้วย Ozone Days และ Adult เลือกกลุ่มข้อมูลโดยมีจำนวนคำตอบ (class) เท่ากันในข้อมูลแต่ละชุด เป็นการทดลองแบบมีการเรียนรู้ จากผลการวิจัยอัลกอริทึมที่ดีที่สุดของข้อมูล Ozone Days คือซัพพอร์ตเวกเตอร์แม็ชชีน ฟังก์ชันเคอร์เนลแบบ Rbf มีค่าความถูกต้อง 94.83% ค่าความแม่นยำ 96% ค่าความระลึก 96% และค่าความถ่วงดุล 96% ส่วนข้อมูล Adult คือซัพพอร์ตเวกเตอร์แมชชีน ฟังก์ชันเคอร์เนลแบบโพลิโนเมียล มีค่าความถูกต้อง 79.66% ค่าความแม่นยำ 80% ค่าความระลึก 80% และค่าความถ่วงดุล 80% อัลกอริทึมที่สามารถเลือกใช้ได้ดีคือ ซัพพอร์ตเวกเตอร์แมชชีน สามารถใช้กับลักษณะข้อมูลที่เป็นตัวเลขหรือข้อมูลเชิงกลุ่มแบบข้อความ ซึ่งเทคนิคเหล่านี้สามารถประยุกต์ใช้กับการสร้างเทคโนโลยีการจัดเก็บและนำเสนอเนื้อหาแบบมีโครงสร้าง โดยสามารถวิเคราะห์ จำแนกหรือจัดแบ่งข้อมูลที่มีความสัมพันธ์กับข้อมูลอื่น ๆ แบบเชิงความหมายได้ต่อไป

เดช ธรรมศิริ และพยุง มีสัจ (2554) ได้นำเสนอการจำแนกประเภท โดยทำการทดสอบประสิทธิภาพด้วยข้อมูล Austrian Credit และ Bankrupt Data โดยนำเอาวิธีซัพพอร์ตเวกเตอร์แมชชีนและหาค่าพารามิเตอร์ที่เหมาะสมร่วมด้วยการเลือกคุณลักษณะที่เหมาะสมโดยใช้ขั้นตอนวิธีเชิงพันธุกรรมเปรียบเทียบผลการวิจัยกับวิธีต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาท วิธีซัพพอร์ตเวกเตอร์แมชชีนกับวิธีเชิงพันธุกรรม พบว่าวิธีซัพพอร์ตเวกเตอร์แมชชีนที่ใช้ขั้นตอนวิธีเชิงพันธุกรรมจะให้ค่าความแม่นยำสูงที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
วิธีการดำเนินงาน

3.1 อุปกรณ์ในการวิจัย

3.1.1 อุปกรณ์ที่มีอยู่แล้ว

1) เครื่องคอมพิวเตอร์

2) เครื่องพิมพ์เลเซอร์

3) โปรแกรมสำเร็จรูป Weka version 3.7

4) โปรแกรมสำเร็จรูป Microsoft Excel 2007

3.2 ขั้นตอนการดำเนินงานวิจัย

งานวิจัยนี้ได้ทำการศึกษาเรื่องการเปรียบเทียบประสิทธิภาพและทำนายผลสถานภาพการเป็นนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ระหว่างวิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาท วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม โดยมีขั้นตอนการดำเนินงานดังนี้

3.2.1 การเก็บรวบรวมข้อมูล

เก็บรวบรวมข้อมูลสถานภาพการเป็นนักศึกษาชั้นปีที่ 1 ในภาคการศึกษาที่ 2 ปีการศึกษา 2556 คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง จำนวน 1,625 คน (www.reg.kmitl.ac.th/index/index.php) โดยข้อมูลประกอบด้วยคุณลักษณะ (Attribute) ต่าง ๆ ดังนี้ (คู่มือนักศึกษาระดับปริญญาตรี : 2556)

1. เพศ
    1) เพศชาย
    2) เพศหญิง
2. สาขาวิชา
    1) สาขาคณิตศาสตร์
    2) สาขาวิทยาการคอมพิวเตอร์
    3) สาขาวิชาเคมีอุตสาหกรรม
    4) สาขาวิชาเคมีสิ่งแวดล้อม
    5) สาขาวิชาจุลชีววิทยาอุตสาหกรรม
    6) สาขาเทคโนโลยีชีวภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7) สาขาฟิสิกส์ประยุกต์

8) สาขาสถิติประยุกต์

3. วิธีผ่านการสอบคัดเลือก

1) ระบบแอดมิดชั่น

2) ระบบโควตา

3) ระบบรับตรง

4. คะแนนเฉลี่ยประจำภาคการศึกษาที่ 1 ของปีการศึกษา 2556

5. คะแนนเฉลี่ยประจำภาคการศึกษาที่ 2 ของปีการศึกษา 2556

6. สถานภาพการเป็นนักศึกษา

1) คงสภาพการเป็นนักศึกษา

2) พ้นสภาพการเป็นนักศึกษา เนื่องจาก คะแนนเฉลี่ยสะสม และไม่ได้ลงทะเบียนเรียน

3.2.2 บันทึกข้อมูลสถานภาพการเป็นนักศึกษาจำนวน 1,625 คน ลงในโปรแกรม Microsoft Excel โดยให้แนวคอลัมน์เป็นเพศ สาขาวิชา วิธีผ่านการสอบคัดเลือก คะแนนเฉลี่ยประจำภาคการศึกษาที่ 1 ของปีการศึกษา 2556 คะแนนเฉลี่ยประจำภาคการศึกษาที่ 2 ของปีการศึกษา 2556 และสถานภาพการเป็นนักศึกษา ส่วนแนวแถวเป็นลำดับที่ของนักศึกษา

3.2.3 การแบ่งข้อมูล

นำข้อมูลทั้งหมดมาแบ่งออกเป็น 3 ส่วน จากข้อมูลสถานภาพการเป็นนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง จำนวน 1,625 คน เพื่อใช้สำหรับการวิเคราะห์ข้อมูล โดยแบ่งข้อมูลออกเป็นสัดส่วนดังนี้

ส่วนที่ 1 : ข้อมูลชุดฝึกหัด (Training Data Set) เพื่อนำไปสร้างตัวแบบ

มีข้อมูล 70% ของข้อมูลทั้งหมด ซึ่งจะได้ข้อมูลในส่วนที่ 1 จำนวน 1,138 คน

ส่วนที่ 2 : ข้อมูลชุดทดสอบประสิทธิภาพ (Evaluation Data Set) เพื่อนำไปทดสอบความถูกต้องของตัวแบบ

มีข้อมูล 20% ของข้อมูลทั้งหมด ซึ่งจะได้ข้อมูลในส่วนที่ 2 จำนวน 325 คน

ส่วนที่ 3 : ข้อมูลชุดทดสอบ (Testing Data Set) เพื่อนำไปทำนายตัวแบบ

มีข้อมูล 10% ของข้อมูลทั้งหมด ซึ่งจะได้ข้อมูลในส่วนที่ 3 จำนวน 162 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.4 เครื่องมือที่ใช้ในการวิจัย

ในการวิเคราะห์ข้อมูลครั้งนี้ใช้โปรแกรม WEKA (Waikato Environment for Knowledge Analysis) เวอร์ชั่น 3.7 ซึ่งเป็นโปรแกรมที่สามารถดาวน์โหลดได้จากเว็บไซต์ ซึ่งอยู่ภายใต้การควบคุมของ GPL License ซึ่งโปรแกรม WEKA ได้ถูกพัฒนามาจากภาษาจาวาทั้งหมด ซึ่งเป็นที่นิยมในการใช้งานด้านการทำเหมืองข้อมูล

3.2.5 วิธีการวิเคราะห์ข้อมูล

ในการทำวิจัยครั้งนี้ คณะผู้วิจัยได้จัดข้อมูลแต่ละชุดออกเป็น 3 ส่วน โดยส่วนที่ 1 ใช้ข้อมูล 70 เปอร์เซ็นต์ของข้อมูลทั้งหมดในการสร้างตัวแบบ ส่วนที่ 2 ใช้ข้อมูล 20 เปอร์เซ็นต์ของข้อมูลทั้งหมดในการทดสอบความถูกต้องของตัวแบบ ส่วนที่ 3 ใช้ข้อมูล 10 เปอร์เซ็นต์ของข้อมูลทั้งหมดในการทำนายตัวแบบ แปลงไฟล์ข้อมูลให้เป็นนามสกุล *.csv เพื่อใช้วิเคราะห์ประสิทธิภาพการจำแนกกลุ่มข้อมูลในโปรแกรม WEKA ซึ่งเป็นโปรแกรมที่สามารถนำมาทดสอบอัลกอริทึมของวิธีการจำแนกกลุ่มได้เนื่องจากมีอัลกอริทึมที่ได้ระบุไว้ให้เลือกใช้ในโปรแกรมครบตามที่กำหนด คณะผู้วิจัยได้กำหนดวิธีการจำแนกกลุ่มเพื่อนำมาทดสอบดังนี้

1) วิธีความใกล้เคียงกันมากที่สุด (K-Nearest Neighbor : KNN) ใช้อัลกอริทึมชนิด IBk เนื่องจากเป็นฟังก์ชันหลักที่สนใจ ซึ่งเป็นพื้นฐานของอัลกอริทึม 8.1 อัลกอริทึม IBk ยังสามารถกำหนดน้ำหนักระยะห่างและทางเลือก (option) เพื่อกำหนดค่า k โดยใช้ cross-validation (Kumar, V. and Wu, X., 2009)

2) วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree) ใช้อัลกอริทึมชนิด J48 ซึ่งพัฒนามาจาก ID3 สามารถใช้ได้กับข้อมูลแบบไม่ต่อเนื่องและแบบต่อเนื่อง ต่างจาก ID3 ที่ใช้ได้เพียงข้อมูลแบบไม่ต่อเนื่องเท่านั้น (รุจิรา ธรรมสมบัติ, 2554)

3) วิธีโครงข่ายประสาท (Neural Network) ใช้อัลกอริทึมชนิดเพอร์เซปตรอนหลายชั้น (Multilayer Perceptron) โดยกำหนดค่าอัตราการเรียนรู้ (Learning Rate) เป็น 0.1, 0.2, 0.3, 0.4 และ 0.5 ค่าโมเมนตัม (Momentum) เป็น 0.5, 0.6, 0.7, 0.8 และ 0.9 จำนวนรอบการสอน (Training Time) 30,000 รอบ การวิจัยครั้งนี้ใช้อัลกอริทึมของวิธีโครงข่ายประสาทชนิดเพอร์เซปตรอนหลายชั้นที่มีชั้นซ่อน (Hidden Layer) 1 ชั้น แม้ว่าโครงสร้างโครงข่ายประสาทที่ซับซ้อนสามารถมีชั้นซ่อนมากกว่า 1 ชั้น แต่ในทางปฏิบัติ การกำหนดชั้นซ่อน 1 ชั้น ก็เพียงพอต่อการวิเคราะห์ข้อมูล (Berson, A. and Stephen, J. S., 1997)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) ใช้อัลกอริทึมชนิดโพลิโนเมียลเคอร์เนล (Polynomial Kernel) เนื่องจากงานวิจัยที่อ้างอิงจากวาทินี นุ้ยเพียร และคณะ (2553) ได้ผลว่าวิธีซัพพอร์ตเวกเตอร์แมชชีนที่ใช้อัลกอริทึมชนิดโพลิโนเมียลเคอร์เนลดีที่สุด

5) วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม (Binary Logistic Regression Method) เป็นการวิเคราะห์การถดถอยแบบหนึ่งโดยที่ตัวแปรตามเป็นตัวแปรเชิงคุณภาพที่มีค่าได้เพียง 2 ค่า (dichotomous or binary variable) ส่วนตัวแปรอิสระอาจจะเป็นตัวแปรเชิงปริมาณหรือเชิงคุณภาพ หรืออาจจะมีทั้งตัวแปรเชิงปริมาณและตัวแปรเชิงคุณภาพก็ได้ (กัลยา วานิชย์บัญชา, 2552)

การนำผลการวิเคราะห์ข้อมูลมาประเมินผลเพื่อเปรียบเทียบประสิทธิภาพและทำนายผลสถานภาพการเป็นนักศึกษาของวิธีการจำแนกกลุ่มโดยนำข้อมูลที่ได้จากการวิเคราะห์มาเปรียบเทียบว่าวิธีในการจำแนกกลุ่มข้อมูลวิธีใดระหว่างวิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาท วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม จะมีค่าความถูกต้อง (Accuracy) ค่าอัตราความถูกต้องเชิงบวก (TP Rate) ค่าอัตราความถูกต้องเชิงลบ (TN Rate) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-Measure) มากที่สุด ส่วนค่าอัตราความผิดพลาดเชิงบวก (FP Rate) และค่าอัตราความผิดพลาดเชิงลบ (FN Rate) จะพิจารณาจากค่าน้อยที่สุด จะทำให้มีประสิทธิภาพดีที่สุด ในการทำนายผลของวิธีการจำแนกกลุ่มทั้ง 5 วิธี จะใช้ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error : MAE) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) ในการเปรียบเทียบ โดยพิจาณาจากค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและค่าความคลาดเคลื่อนกำลังสองเฉลี่ยที่มีค่าน้อยที่สุด จะทำให้การทำนายผลดีที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการวิเคราะห์ข้อมูล

4.1 การวิเคราะห์ข้อมูล

จากโปรแกรม Weka3.7 แสดงผลลัพธ์ของการทำนายสถานภาพการเป็นนักศึกษาด้วยวิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาท วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม ได้ผลดังนี้

4.1.1 วิธีความใกล้เคียงกันมากที่สุด (K-Nearest Neighbor Method)

วิธีความใกล้เคียงกันมากที่สุดนั้นมีอัลกอริทึมหลายอัลกอริทึม แต่ในงานวิจัยนี้ใช้อัลกอริทึม IBk ในการทำนายสถานภาพการเป็นนักศึกษา ได้ผลดังนี้

4.1.1.1 การสร้างตัวแบบสถานภาพการเป็นนักศึกษา

สร้างตัวแบบสถานภาพการเป็นนักศึกษาจากข้อมูลฝึกหัด (Training data set) ด้วยไฟล์ StudentStatus.csv จำนวน 1,138 คน ได้ผลดังนี้

ตารางที่ 4.1 ผลในส่วนของการสรุปผล (Summary) จากการสร้างตัวแบบสถานภาพการเป็นนักศึกษาด้วยวิธีความใกล้เคียงกันมากที่สุด

Correctly Classified Instances	1016	89.2794%
Incorrectly Classified Instances	122	10.7206%
Kappa statistic	0.3869	
Mean absolute error	0.108	
Root mean squared error	0.3271	
Relative absolute error	56.2234 %	
Root relative squared error	105.7296 %	
Coverage of cases (0.95 level)	89.2794 %	
Mean rel. region size (0.95 level)	50%	
Total Number of Instances	1138	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.1 จากข้อมูลจำนวน 1,138 คน ทำนายข้อมูลถูกต้องจำนวน 1,016 คน คิดเป็น 89.2794% และทำนายข้อมูลไม่ถูกต้องจำนวน 122 คน คิดเป็น 10.7206% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.3869 แสดงว่าข้อมูลมีความสอดคล้องกันค่อนข้างน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.1080 ซึ่งมีค่าน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงมาก และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.3271)^2 = 0.107$ ซึ่งมีค่าน้อย แสดงว่าตัวแบบมีความถูกต้องมาก

ตารางที่ 4.2 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบ (Detailed Accuracy By Class) จากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.952	0.598	0.93	0.952	0.941	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.402	0.048	0.5	0.402	0.445	no
Weighted Avg.	0.893	0.539	0.884	0.893	0.888	

จากตารางที่ 4.2 สำหรับคำตอบคงสภาพการเป็นนักศึกษา (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก (TP Rate) = 0.952 ค่าอัตราความผิดพลาดเชิงบวก (FP Rate) = 0.598 ค่าความแม่นยำ (Precision) = 0.93 ค่าความระลึก (Recall) = 0.952 และค่าความถ่วงดุล (F-Measure) = 0.941 ส่วนคำตอบพ้นสภาพการเป็นนักศึกษา (Class = no) มีค่าอัตราความถูกต้องเชิงลบ (TN Rate) = 0.402 ค่าอัตราความผิดพลาดเชิงลบ (FN Rate) = 0.048 ค่าความแม่นยำ = 0.5 ค่าความระลึก = 0.402 และค่าความถ่วงดุล = 0.445

ตารางที่ 4.3 ผลในส่วนของเมตริกซ์ความสับสน (Confusion Matrix) จากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด

		ผลการจำแนกสถานภาพการเป็นนักศึกษา	
		คงสภาพการเป็นนักศึกษา	พ้นสภาพการเป็นนักศึกษา
ค่าที่แท้จริง	คงสภาพการเป็นนักศึกษา	967	49
	พ้นสภาพการเป็นนักศึกษา	73	49

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.3 มีข้อมูล 1,138 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 1,016 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าคงสภาพการเป็นนักศึกษา 967 คน และจำนวนข้อมูลที่จำแนกถูกว่าพ้นสภาพการเป็นนักศึกษา 49 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 122 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าคงสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วพ้นสภาพการเป็นนักศึกษา 73 คน และจำนวนข้อมูลที่จำแนกผิดว่าพ้นสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วคงสภาพการเป็นนักศึกษา 49 คน

4.1.1.2 การทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา

ทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษาจากชุดข้อมูลทดสอบประสิทธิภาพ (Evaluation data set) ด้วยไฟล์ StudentStatus-eval.csv จำนวน 325 คน ได้ผลดังนี้

ตารางที่ 4.4 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด

Correctly Classified Instances	298	91.6923%
Incorrectly Classified Instances	27	8.3077%
Kappa statistic	0.4455	
Mean absolute error	0.0853	
Root mean squared error	0.2893	
Coverage of cases (0.95 level)	91.6923 %	
Total Number of Instances	325	

จากตารางที่ 4.4 จากข้อมูลจำนวน 325 คน ทำนายข้อมูลถูกต้องจำนวน 298 คน คิดเป็น 91.6923% และทำนายข้อมูลไม่ถูกต้องจำนวน 27 คน คิดเป็น 8.3077% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.4455 แสดงว่าข้อมูลมีความสอดคล้องกันค่อนข้างน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.0853 ซึ่งมีค่าน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงมากและมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.2893)^2 = 0.084$ ซึ่งมีค่าน้อย แสดงว่าตัวแบบมีความถูกต้องมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.96	0.536	0.95	0.96	0.955	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.464	0.04	0.52	0.464	0.491	no
Weighted Avg.	0.917	0.493	0.913	0.917	0.915	

จากตารางที่ 4.5 สำหรับคำตอบคงสภาพการเป็นนักศึกษา (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.96 ค่าอัตราความผิดพลาดเชิงบวก = 0.536 ค่าความแม่นยำ = 0.95 ค่าความระลึก = 0.96 และค่าความถ่วงดุล = 0.955 ส่วนคำตอบพ้นสภาพการเป็นนักศึกษา (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.464 ค่าอัตราความผิดพลาดเชิงลบ = 0.04 ค่าความแม่นยำ = 0.52 ค่าความระลึก = 0.464 และค่าความถ่วงดุล = 0.491

ตารางที่ 4.6 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด

		ผลการจำแนกสถานภาพการเป็นนักศึกษา	
		คงสภาพการเป็นนักศึกษา	พ้นสภาพการเป็นนักศึกษา
ค่าที่แท้จริง	คงสภาพการเป็นนักศึกษา	285	12
	พ้นสภาพการเป็นนักศึกษา	15	13

จากตารางที่ 4.6 มีข้อมูล 325 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 298 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าคงสภาพการเป็นนักศึกษา 285 คน และจำนวนข้อมูลที่จำแนกถูกว่าพ้นสภาพการเป็นนักศึกษา 13 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 27 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าคงสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วพ้นสภาพการเป็นนักศึกษา 15 คน และจำนวนข้อมูลที่จำแนกผิดว่าพ้นสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วคงสภาพการเป็นนักศึกษา 12 คน

4.1.1.3 การทำนายสถานภาพการเป็นนักศึกษา

การทำนายสถานภาพการเป็นนักศึกษาจากชุดข้อมูลทดสอบ (Testing data set) ด้วยไฟล์ StudentStatus-test.csv จำนวน 162 คน ได้ผลดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ (Predictions on test set) ด้วยวิธีความใกล้เคียงกันมากที่สุด

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
1	1:Yes	1:Yes		0.999
2	1:Yes	1:Yes		0.999
3	1:Yes	1:Yes		0.999
4	1:Yes	1:Yes		0.999
5	1:Yes	1:Yes		0.999
6	1:Yes	1:Yes		0.999
7	1:Yes	1:Yes		0.999
8	1:Yes	1:Yes		0.999
9	1:Yes	1:Yes		0.999
10	1:Yes	1:Yes		0.999
11	1:Yes	1:Yes		0.999
12	1:Yes	1:Yes		0.999
13	1:Yes	1:Yes		0.999
14	2:No	1:Yes	+	0.999
15	1:Yes	1:Yes		0.999
16	1:Yes	1:Yes		0.999
17	1:Yes	1:Yes		0.999
18	1:Yes	1:Yes		0.999
19	1:Yes	1:Yes		0.999
20	1:Yes	1:Yes		0.999
21	1:Yes	1:Yes		0.999
22	1:Yes	1:Yes		0.999
23	1:Yes	1:Yes		0.999
24	1:Yes	1:Yes		0.999
25	1:Yes	1:Yes		0.999
26	1:Yes	1:Yes		0.999
27	1:Yes	1:Yes		0.999
28	1:Yes	1:Yes		0.999
29	1:Yes	1:Yes		0.999
30	1:Yes	1:Yes		0.999
31	1:Yes	1:Yes		0.999
32	1:Yes	1:Yes		0.999
33	1:Yes	1:Yes		0.999
34	1:Yes	1:Yes		0.999
35	1:Yes	1:Yes		0.999

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
36	1:Yes	1:Yes		0.999
37	1:Yes	1:Yes		0.999
38	1:Yes	1:Yes		0.999
39	1:Yes	1:Yes		0.999
40	1:Yes	1:Yes		0.999
41	1:Yes	1:Yes		0.999
42	1:Yes	1:Yes		0.999
43	1:Yes	1:Yes		0.999
44	1:Yes	1:Yes		0.999
45	1:Yes	1:Yes		0.999
46	1:Yes	2:No	+	0.999
47	1:Yes	1:Yes		0.999
48	1:Yes	1:Yes		0.999
49	1:Yes	1:Yes		0.999
50	1:Yes	1:Yes		0.999
51	1:Yes	1:Yes		0.999
52	1:Yes	1:Yes		0.999
53	1:Yes	1:Yes		0.999
54	1:Yes	1:Yes		0.999
55	1:Yes	1:Yes		0.999
56	1:Yes	1:Yes		0.999
57	1:Yes	1:Yes		0.999
58	1:Yes	1:Yes		0.999
59	1:Yes	1:Yes		0.999
60	1:Yes	1:Yes		0.999
61	1:Yes	1:Yes		0.999
62	1:Yes	1:Yes		0.999
63	1:Yes	1:Yes		0.999
64	1:Yes	1:Yes		0.999
65	1:Yes	1:Yes		0.999
66	1:Yes	1:Yes		0.999
67	1:Yes	1:Yes		0.999
68	1:Yes	1:Yes		0.999
69	1:Yes	1:Yes		0.999
70	1:Yes	1:Yes		0.999

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ (Predictions on test set) ด้วยวิธีความใกล้เคียงกันมากที่สุด (ต่อ)

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
71	1:Yes	1:Yes		0.999
72	1:Yes	1:Yes		0.999
73	1:Yes	1:Yes		0.999
74	1:Yes	1:Yes		0.999
75	1:Yes	1:Yes		0.999
76	1:Yes	1:Yes		0.999
77	1:Yes	1:Yes		0.999
78	1:Yes	1:Yes		0.999
79	1:Yes	1:Yes		0.999
80	1:Yes	1:Yes		0.999
81	2:No	2:No		0.999
82	1:Yes	1:Yes		0.999
83	1:Yes	1:Yes		0.999
84	1:Yes	1:Yes		0.999
85	1:Yes	1:Yes		0.999
86	1:Yes	1:Yes		0.999
87	1:Yes	1:Yes		0.999
88	2:No	1:Yes	+	0.999
89	1:Yes	1:Yes		0.999
90	1:Yes	1:Yes		0.999
91	1:Yes	1:Yes		0.999
92	1:Yes	1:Yes		0.999
93	1:Yes	1:Yes		0.999
94	1:Yes	1:Yes		0.999
95	1:Yes	1:Yes		0.999
96	1:Yes	1:Yes		0.999
97	1:Yes	1:Yes		0.999
98	1:Yes	1:Yes		0.999
99	1:Yes	1:Yes		0.999
100	2:No	2:No		0.999
101	1:Yes	1:Yes		0.999
102	1:Yes	1:Yes		0.999
103	1:Yes	1:Yes		0.999
104	1:Yes	1:Yes		0.999
105	1:Yes	1:Yes		0.999

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
106	1:Yes	1:Yes		0.999
107	2:No	1:Yes	+	0.999
108	1:Yes	1:Yes		0.999
109	1:Yes	1:Yes		0.999
110	1:Yes	1:Yes		0.999
111	1:Yes	1:Yes		0.999
112	1:Yes	1:Yes		0.999
113	2:No	2:No		0.999
114	1:Yes	1:Yes		0.999
115	1:Yes	1:Yes		0.999
116	1:Yes	1:Yes		0.999
117	1:Yes	1:Yes		0.999
118	1:Yes	1:Yes		0.999
119	1:Yes	1:Yes		0.999
120	1:Yes	1:Yes		0.999
121	1:Yes	1:Yes		0.999
122	1:Yes	1:Yes		0.999
123	1:Yes	1:Yes		0.999
124	1:Yes	1:Yes		0.999
125	1:Yes	1:Yes		0.999
126	1:Yes	1:Yes		0.999
127	1:Yes	1:Yes		0.999
128	1:Yes	1:Yes		0.999
129	1:Yes	1:Yes		0.999
130	2:No	1:Yes	+	0.999
131	1:Yes	1:Yes		0.999
132	1:Yes	1:Yes		0.999
133	1:Yes	1:Yes		0.999
134	1:Yes	1:Yes		0.999
135	2:No	2:No		0.999
136	1:Yes	1:Yes		0.999
137	1:Yes	1:Yes		0.999
138	1:Yes	1:Yes		0.999
139	1:Yes	1:Yes		0.999
140	1:Yes	1:Yes		0.999

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ (Predictions on test set) ด้วยวิธีความใกล้เคียงกันมากที่สุด (ต่อ)

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
141	1:Yes	1:Yes		0.999
142	2:No	2:No		0.999
143	1:Yes	1:Yes		0.999
144	2:No	1:Yes	+	0.999
145	1:Yes	1:Yes		0.999
146	1:Yes	1:Yes		0.999
147	1:Yes	1:Yes		0.999
148	1:Yes	1:Yes		0.999
149	1:Yes	1:Yes		0.999
150	2:No	1:Yes	+	0.999
151	1:Yes	1:Yes		0.999

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
152	1:Yes	1:Yes		0.999
153	1:Yes	1:Yes		0.999
154	1:Yes	1:Yes		1
155	1:Yes	1:Yes		0.999
156	1:Yes	1:Yes		0.999
157	1:Yes	1:Yes		0.999
158	1:Yes	1:Yes		0.999
159	1:Yes	2:No	+	0.999
160	1:Yes	1:Yes		0.999
161	1:Yes	1:Yes		0.999
162	1:Yes	1:Yes		0.999

จากตารางที่ 4.7 ค่าทำนายที่ทายผิดในช่อง error จะมีเครื่องหมาย + อยู่จำนวน 8 ค่า คือค่าที่ 14, 46, 88, 107, 130, 144, 150 และ 159 โดยที่ค่าที่ 14, 88, 107, 130, 144 และ 150 พ้นสภาพการเป็นนักศึกษาแต่ทำนายว่าคงสภาพการเป็นนักศึกษา และค่าที่ 46 และ 159 คงสภาพการเป็นนักศึกษาแต่ทำนายว่าพ้นสภาพการเป็นนักศึกษา

ตารางที่ 4.8 ผลในส่วนของการสรุปผลจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด

Correctly Classified Instances	154	95.0617%
Incorrectly Classified Instances	8	4.9383%
Kappa statistic	0. 5308	
Mean absolute error	0.0502	
Root mean squared error	0.222	
Coverage of cases (0.95 level)	95.0617 %	
Total Number of Instances	162	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.8 จากข้อมูลจำนวน 162 คน ทำนายข้อมูลถูกต้องจำนวน 154 คน คิดเป็น 95.0617% และทำนายข้อมูลไม่ถูกต้องจำนวน 8 คน คิดเป็น 4.9383% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.5308 แสดงว่าข้อมูลมีความสอดคล้องกันปานกลาง มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.0502 ซึ่งมีค่าน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงมาก และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.222)^2 = 0.049$ ซึ่งมีค่าน้อย แสดงว่าตัวแบบมีความถูกต้องมาก

ตารางที่ 4.9 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.987	0.545	0.961	0.987	0.974	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.455	0.013	0.714	0.455	0.556	no
Weighted Avg.	0.951	0.509	0.945	0.951	0.945	

จากตารางที่ 4.9 สำหรับคำตอบคงสภาพการเป็นนักศึกษา (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.987 ค่าอัตราความผิดพลาดเชิงบวก = 0.545 ค่าความแม่นยำ = 0.961 ค่าความระลึก = 0.987 และค่าความถ่วงดุล = 0.974 ส่วนคำตอบพ้นสภาพการเป็นนักศึกษา (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.455 ค่าอัตราความผิดพลาดเชิงลบ = 0.013 ค่าความแม่นยำ = 0.714 ค่าความระลึก = 0.455 และค่าความถ่วงดุล = 0.556

ตารางที่ 4.10 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีความใกล้เคียงกันมากที่สุด

		ผลการจำแนกสถานภาพการเป็นนักศึกษา	
		คงสภาพการเป็นนักศึกษา	พ้นสภาพการเป็นนักศึกษา
ค่าที่แท้จริง	คงสภาพการเป็นนักศึกษา	149	2
	พ้นสภาพการเป็นนักศึกษา	6	5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.10 มีข้อมูล 162 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 154 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าคงสภาพการเป็นนักศึกษา 149 คน และจำนวนข้อมูลที่จำแนกถูกว่าพ้นสภาพการเป็นนักศึกษา 5 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 8 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าคงสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วพ้นสภาพการเป็นนักศึกษา 6 คน และจำนวนข้อมูลที่จำแนกผิดว่าพ้นสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วคงสภาพการเป็นนักศึกษา 2 คน

4.1.2 วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree Method)

วิธีแผนภาพต้นไม้เพื่อการตัดสินใจนั้นมีอัลกอริทึมหลายอัลกอริทึม แต่ในงานวิจัยนี้ใช้อัลกอริทึม J48 ในการทำนายสถานภาพการเป็นนักศึกษา ได้ผลดังนี้

4.1.2.1 การสร้างตัวแบบสถานภาพการเป็นนักศึกษา

สร้างตัวแบบสถานภาพการเป็นนักศึกษาจากชุดข้อมูลฝึกหัด (Training data set) ด้วยไฟล์ StudentStatus.csv จำนวน 1,138 คน ได้ผลดังนี้

ตารางที่ 4.11 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

Correctly Classified Instances	1083	95.167 %
Incorrectly Classified Instances	55	4.833 %
Kappa statistic	0.718	
Mean absolute error	0.0825	
Root mean squared error	0.2074	
Relative absolute error	42.95 %	
Root relative squared error	67.0523%	
Coverage of cases (0.95 level)	96.7487%	
Mean rel. region size (0.95 level)	54.0422 %	
Total Number of Instances	1138	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.11 จากข้อมูลจำนวน 1,138 คน ทำนายข้อมูลถูกต้องจำนวน 1,083 คน คิดเป็น 95.167% และทำนายข้อมูลไม่ถูกต้องจำนวน 55 คน คิดเป็น 4.833% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.718 แสดงว่าข้อมูลมีความสอดคล้องกันค่อนข้างมาก มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.0825 ซึ่งมีค่าน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงมาก และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.2074)^2 = 0.0430$ ซึ่งมีค่าน้อย แสดงว่าตัวแบบมีความถูกต้องมาก

ตารางที่ 4.12 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.987	0.344	0.96	0.987	0.973	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.656	0.013	0.86	0.656	0.744	no
Weighted Avg.	0.952	0.309	0.949	0.952	0.949	

จากตารางที่ 4.12 สำหรับคำตอบคงสภาพการเป็นนักศึกษา (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.987 ค่าอัตราความผิดพลาดเชิงบวก = 0.344 ค่าความแม่นยำ = 0.96 ค่าความระลึก = 0.987 และค่าความถ่วงดุล = 0.973 ส่วนคำตอบพ้นสภาพการเป็นนักศึกษา (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.656 ค่าอัตราความผิดพลาดเชิงลบ = 0.013 ค่าความแม่นยำ = 0.86 ค่าความระลึก = 0.656 และค่าความถ่วงดุล = 0.744

ตารางที่ 4.13 ผลในส่วนของเมตริกซ์ความสับสนจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

		ผลการจำแนกสถานภาพการเป็นนักศึกษา	
		คงสภาพการเป็นนักศึกษา	พ้นสภาพการเป็นนักศึกษา
ค่าที่แท้จริง	คงสภาพการเป็นนักศึกษา	1,003	13
	พ้นสภาพการเป็นนักศึกษา	42	80

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.13 มีข้อมูล 1,138 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 1,083 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าคงสภาพการเป็นนักศึกษา 1,003 คน และจำนวนข้อมูลที่จำแนกถูกว่าพ้นสภาพการเป็นนักศึกษา 80 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 55 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าคงสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วพ้นสภาพการเป็นนักศึกษา 42 คน และจำนวนข้อมูลที่จำแนกผิดว่าพ้นสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วคงสภาพการเป็นนักศึกษา 13 คน

4.1.2.2 การทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา

ทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษาจากชุดข้อมูลทดสอบประสิทธิภาพด้วยไฟล์ StudentStatus-eval.csv จำนวน 325 คน ได้ผลดังนี้

ตารางที่ 4.14 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

Correctly Classified Instances	312	96%
Incorrectly Classified Instances	13	4%
Kappa statistic	0.7418	
Mean absolute error	0.0769	
Root mean squared error	0.1904	
Coverage of cases (0.95 level)	96.9231 %	
Total Number of Instances	325	

จากตารางที่ 4.14 จากข้อมูลจำนวน 325 คน ทำนายข้อมูลถูกต้องจำนวน 312 คน คิดเป็น 96% และทำนายข้อมูลไม่ถูกต้องจำนวน 13 คน คิดเป็น 4% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.7418 แสดงว่าข้อมูลมีความสอดคล้องกันค่อนข้างมาก มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.0769 ซึ่งมีค่าน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงมาก และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.1904)^2 = 0.0363$ ซึ่งมีค่าน้อย แสดงว่าตัวแบบมีความถูกต้องมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.15 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.98	0.25	0.977	0.98	0.978	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.75	0.02	0.778	0.75	0.764	no
Weighted Avg.	0.96	0.23	0.959	0.96	0.96	

จากตารางที่ 4.15 สำหรับคำตอบคงสภาพการเป็นนักศึกษา (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.98 ค่าอัตราความผิดพลาดเชิงบวก = 0.25 ค่าความแม่นยำ = 0.977 ค่าความระลึก = 0.98 และค่าความถ่วงดุล = 0.978 ส่วนคำตอบพ้นสภาพการเป็นนักศึกษา (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.75 ค่าอัตราความผิดพลาดเชิงลบ = 0.02 ค่าความแม่นยำ = 0.778 ค่าความระลึก = 0.75 และค่าความถ่วงดุล = 0.764

ตารางที่ 4.16 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

		ผลการจำแนกสถานภาพการเป็นนักศึกษา	
		คงสภาพการเป็นนักศึกษา	พ้นสภาพการเป็นนักศึกษา
ค่าที่แท้จริง	คงสภาพการเป็นนักศึกษา	291	6
	พ้นสภาพการเป็นนักศึกษา	7	21

จากตารางที่ 4.16 มีข้อมูล 325 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 312 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าคงสภาพการเป็นนักศึกษา 291 คน และจำนวนข้อมูลที่จำแนกถูกว่าพ้นสภาพการเป็นนักศึกษา 21 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 13 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าคงสภาพการเป็นนักศึกษาซึ่งค่าที่แท้จริงแล้วพ้นสภาพการเป็นนักศึกษา 7 คน และจำนวนข้อมูลที่จำแนกผิดว่าพ้นสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วคงสภาพการเป็นนักศึกษา 6 คน

4.1.2.3 การทำนายสถานภาพการเป็นนักศึกษา

การทำนายสถานภาพการเป็นนักศึกษาจากข้อมูลทดสอบด้วยไฟล์ StudentStatus-test.csv จำนวน 162 คน ได้ผลดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.17 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
1	1:Yes	1:Yes		0.965
2	1:Yes	1:Yes		0.965
3	1:Yes	1:Yes		0.965
4	1:Yes	1:Yes		0.965
5	1:Yes	1:Yes		0.965
6	1:Yes	1:Yes		0.965
7	1:Yes	1:Yes		0.965
8	1:Yes	1:Yes		0.965
9	1:Yes	1:Yes		0.965
10	1:Yes	1:Yes		0.965
11	1:Yes	1:Yes		0.965
12	1:Yes	1:Yes		0.965
13	1:Yes	1:Yes		0.939
14	2:No	1:Yes	+	0.965
15	1:Yes	1:Yes		0.965
16	1:Yes	1:Yes		0.965
17	1:Yes	1:Yes		0.965
18	1:Yes	1:Yes		0.965
19	1:Yes	1:Yes		0.965
20	1:Yes	1:Yes		0.939
21	1:Yes	1:Yes		0.965
22	1:Yes	1:Yes		0.965
23	1:Yes	1:Yes		0.939
24	1:Yes	1:Yes		0.965
25	1:Yes	1:Yes		0.965
26	1:Yes	1:Yes		0.965
27	1:Yes	1:Yes		0.965
28	1:Yes	1:Yes		0.965
29	1:Yes	1:Yes		0.965
30	1:Yes	1:Yes		0.965
31	1:Yes	1:Yes		0.965
32	1:Yes	1:Yes		0.965
33	1:Yes	1:Yes		0.965
34	1:Yes	1:Yes		0.965
35	1:Yes	1:Yes		0.965
36	1:Yes	1:Yes		0.965

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
37	1:Yes	1:Yes		0.965
38	1:Yes	1:Yes		0.965
39	1:Yes	1:Yes		0.965
40	1:Yes	1:Yes		0.965
41	1:Yes	1:Yes		0.965
42	1:Yes	1:Yes		0.965
43	1:Yes	1:Yes		0.965
44	1:Yes	1:Yes		0.965
45	1:Yes	1:Yes		0.965
46	1:Yes	1:Yes		0.965
47	1:Yes	1:Yes		0.965
48	1:Yes	1:Yes		0.965
49	1:Yes	1:Yes		0.965
50	1:Yes	1:Yes		0.965
51	1:Yes	1:Yes		0.965
52	1:Yes	1:Yes		0.965
53	1:Yes	1:Yes		0.965
54	1:Yes	1:Yes		0.965
55	1:Yes	1:Yes		0.965
56	1:Yes	1:Yes		0.965
57	1:Yes	1:Yes		0.965
58	1:Yes	1:Yes		0.965
59	1:Yes	1:Yes		0.939
60	1:Yes	1:Yes		0.965
61	1:Yes	1:Yes		0.965
62	1:Yes	1:Yes		0.965
63	1:Yes	1:Yes		0.965
64	1:Yes	1:Yes		0.939
65	1:Yes	1:Yes		0.965
66	1:Yes	1:Yes		0.965
67	1:Yes	1:Yes		0.965
68	1:Yes	1:Yes		0.965
69	1:Yes	1:Yes		0.965
70	1:Yes	1:Yes		0.965
71	1:Yes	1:Yes		0.965
72	1:Yes	1:Yes		0.965

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.17 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ (ต่อ)

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
73	1:Yes	1:Yes		0.965
74	1:Yes	1:Yes		0.965
75	1:Yes	1:Yes		0.965
76	1:Yes	1:Yes		0.965
77	1:Yes	1:Yes		0.965
78	1:Yes	1:Yes		0.965
79	1:Yes	1:Yes		0.965
80	1:Yes	1:Yes		0.965
81	2:No	2:No		0.97
82	1:Yes	1:Yes		0.965
83	1:Yes	1:Yes		0.965
84	1:Yes	1:Yes		0.965
85	1:Yes	1:Yes		0.965
86	1:Yes	1:Yes		0.965
87	1:Yes	1:Yes		0.965
88	2:No	1:Yes	+	0.965
89	1:Yes	1:Yes		0.965
90	1:Yes	1:Yes		0.965
91	1:Yes	1:Yes		0.965
92	1:Yes	1:Yes		0.965
93	1:Yes	1:Yes		0.965
94	1:Yes	1:Yes		0.965
95	1:Yes	1:Yes		0.965
96	1:Yes	1:Yes		0.965
97	1:Yes	1:Yes		0.965
98	1:Yes	1:Yes		0.965
99	1:Yes	1:Yes		0.939
100	2:No	2:No		0.97
101	1:Yes	1:Yes		0.965
102	1:Yes	1:Yes		0.965
103	1:Yes	1:Yes		0.965
104	1:Yes	1:Yes		0.939
105	1:Yes	1:Yes		0.965
106	1:Yes	1:Yes		0.965
107	2:No	1:Yes	+	0.965
108	1:Yes	1:Yes		0.965

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
109	1:Yes	1:Yes		0.965
110	1:Yes	1:Yes		0.965
111	1:Yes	1:Yes		0.939
112	1:Yes	1:Yes		0.965
113	2:No	2:No		0.97
114	1:Yes	1:Yes		0.965
115	1:Yes	1:Yes		0.965
116	1:Yes	1:Yes		0.965
117	1:Yes	1:Yes		0.965
118	1:Yes	1:Yes		0.965
119	1:Yes	1:Yes		0.965
120	1:Yes	1:Yes		0.965
121	1:Yes	1:Yes		0.965
122	1:Yes	1:Yes		0.965
123	1:Yes	1:Yes		0.965
124	1:Yes	1:Yes		0.965
125	1:Yes	1:Yes		0.965
126	1:Yes	1:Yes		0.965
127	1:Yes	1:Yes		0.965
128	1:Yes	1:Yes		0.965
129	1:Yes	1:Yes		0.965
130	2:No	1:Yes	+	0.965
131	1:Yes	1:Yes		0.965
132	1:Yes	1:Yes		0.965
133	1:Yes	1:Yes		0.965
134	1:Yes	1:Yes		0.965
135	2:No	2:No		0.97
136	1:Yes	1:Yes		0.939
137	1:Yes	1:Yes		0.965
138	1:Yes	1:Yes		0.965
139	1:Yes	1:Yes		0.965
140	1:Yes	1:Yes		0.965
141	1:Yes	1:Yes		0.965
142	2:No	1:Yes	+	0.965
143	1:Yes	1:Yes		0.965
144	2:No	2:No		0.97

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.17 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ (ต่อ)

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
145	1:Yes	1:Yes		0.965
146	1:Yes	1:Yes		0.965
147	1:Yes	1:Yes		0.965
148	1:Yes	1:Yes		0.965
149	1:Yes	1:Yes		0.965
150	2:No	2:No		0.97
151	1:Yes	1:Yes		0.965
152	1:Yes	1:Yes		0.965
153	1:Yes	1:Yes		0.965

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาด. เคลื่อน (error)	การทำนาย (prediction)
154	1:Yes	1:Yes		0.965
155	1:Yes	1:Yes		0.965
156	1:Yes	1:Yes		0.965
157	1:Yes	1:Yes		0.965
158	1:Yes	1:Yes		0.965
159	1:Yes	1:Yes		0.939
160	1:Yes	1:Yes		0.965
161	1:Yes	1:Yes		0.965
162	1:Yes	1:Yes		0.939

จากตารางที่ 4.17 ค่าทำนายที่ทายผิดในช่อง error จะมีเครื่องหมาย + อยู่จำนวน 5 ค่า คือค่าที่ 14, 88, 107, 130 และ 142 โดยที่ค่าที่ 14, 88, 107, 130 และ 142 พ้นสภาพการเป็นนักศึกษา แต่ทำนายว่าคงสภาพการเป็นนักศึกษา และไม่มีค่าใดคงสภาพการเป็นนักศึกษา แต่ทำนายว่าพ้นสภาพการเป็นนักศึกษา

ตารางที่ 4.18 ผลในส่วนของการสรุปผลจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

Correctly Classified Instances	157	96.9136 %
Incorrectly Classified Instances	5	3.0864 %
Kappa statistic	0.6911	
Mean absolute error	0.0651	
Root mean squared error	0.1734	
Coverage of cases (0.95 level)	96.9136 %	
Total Number of Instances	162	

จากตารางที่ 4.18 โดยจากข้อมูลจำนวน 162 คน ทำนายข้อมูลถูกต้องจำนวน 157 คน คิดเป็น 96.9136% และทำนายข้อมูลไม่ถูกต้องจำนวน 5 คน คิดเป็น 3.0864% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.6911 แสดงว่าข้อมูลมีความสอดคล้องกันปานกลาง มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.0651 ซึ่งมีค่าน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงมาก และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.1734)^2 = 0.0301$ ซึ่งมีค่าน้อย แสดงว่าตัวแบบมีความถูกต้องมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.19 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	1	0.455	0.968	1	0.984	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.545	0	1	0.545	0.706	no
Weighted Avg.	0.969	0.424	0.97	0.969	0.965	

จากตารางที่ 4.19 สำหรับคำตอบคงสภาพการเป็นนักศึกษา (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 1 ค่าอัตราความผิดพลาดเชิงบวก = 0.455 ค่าความแม่นยำ = 0.968 ค่าความระลึก = 1 และค่าความถ่วงดุล = 0.984 ส่วนคำตอบพ้นสภาพการเป็นนักศึกษา (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.545 ค่าอัตราความผิดพลาดเชิงลบ = 0 ค่าความแม่นยำ = 1 ค่าความระลึก = 0.545 และค่าความถ่วงดุล = 0.706

ตารางที่ 4.20 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

		ผลการจำแนกสถานภาพการเป็นนักศึกษา	
		คงสภาพการเป็นนักศึกษา	พ้นสภาพการเป็นนักศึกษา
ค่าที่แท้จริง	คงสภาพการเป็นนักศึกษา	151	0
	พ้นสภาพการเป็นนักศึกษา	5	6

จากตารางที่ 4.20 มีข้อมูล 162 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 157 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าคงสภาพการเป็นนักศึกษา 151 คน และจำนวนข้อมูลที่จำแนกถูกว่าพ้นสภาพการเป็นนักศึกษา 6 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 5 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าคงสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วพ้นสภาพการเป็นนักศึกษา 5 คน และจำนวนข้อมูลที่จำแนกผิดว่าพ้นสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วคงสภาพการเป็นนักศึกษา 0 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.3 วิธีโครงข่ายประสาท (Neural Network Method)

วิธีโครงข่ายประสาทใช้อัลกอริทึมชนิดเพอร์เซปตรอนแบบหลายชั้น (Multilayer Perception) ได้ผลการทำนายสถานภาพการเป็นนักศึกษาดังนี้

4.1.3.1 การสร้างตัวแบบสถานภาพการเป็นนักศึกษา

สร้างตัวแบบสถานภาพการเป็นนักศึกษาจากชุดข้อมูลฝึกหัด ด้วยไฟล์ StudentStatus.csv จำนวน 1,138 คน ได้ผลดังนี้

ตารางที่ 4.21 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท

Correctly Classified Instances	1053	92.5308 %
Incorrectly Classified Instances	85	7.4692 %
Kappa statistic	0.5745	
Mean absolute error	0.0792	
Root mean squared error	0.2533	
Relative absolute error	41.2339 %	
Root relative squared error	81.8631%	
Coverage of cases (0.95 level)	95.4306%	
Mean rel. region size (0.95 level)	53.9543 %	
Total Number of Instances	1138	

จากตางรางที่ 4.21 จากข้อมูลจำนวน 1,138 คน ทำนายข้อมูลถูกต้องจำนวน 1,053 คน คิดเป็น 92.5308% และทำนายข้อมูลไม่ถูกต้องจำนวน 85 คน คิดเป็น 7.4692% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.5745 แสดงว่าข้อมูลมีความสอดคล้องปานกลาง มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.0792 ซึ่งมีค่าน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงมาก และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.2533)^2 = 0.0642$ ซึ่งมีค่าน้อย แสดงว่าตัวแบบมีความถูกต้องมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.22 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.969	0.443	0.948	0.969	0.959	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.557	0.031	0.687	0.557	0.615	no
Weighted Avg.	0.925	0.398	0.92	0.925	0.922	

จากตารางที่ 4.22 สำหรับคำตอบคงสภาพการเป็นนักศึกษา (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.969 ค่าอัตราความผิดพลาดเชิงบวก = 0.443 ค่าความแม่นยำ = 0.948 ค่าความระลึก = 0.969 และค่าความถ่วงดุล = 0.959 ส่วนคำตอบพ้นสภาพการเป็นนักศึกษา (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.557 ค่าอัตราความผิดพลาดเชิงลบ = 0.031 ค่าความแม่นยำ = 0.687 ค่าความระลึก = 0.557 และค่าความถ่วงดุล = 0.615

ตารางที่4.23 ผลในส่วนของเมตริกซ์ความสับสนจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท

		ผลการจำแนกสถานภาพการเป็นนักศึกษา	
		คงสภาพการเป็นนักศึกษา	พ้นสภาพการเป็นนักศึกษา
ค่าที่แท้จริง	คงสภาพการเป็นนักศึกษา	985	31
	พ้นสภาพการเป็นนักศึกษา	54	68

จากตารางที่ 4.23 มีข้อมูล 1,138 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 1,053 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าคงสภาพการเป็นนักศึกษา 985 คน และจำนวนข้อมูลที่จำแนกถูกว่าพ้นสภาพการเป็นนักศึกษา 68 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 85 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าคงสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วพ้นสภาพการเป็นนักศึกษา 54 คน และจำนวนข้อมูลที่จำแนกผิดว่าพ้นสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วคงสภาพการเป็นนักศึกษา 31 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.3.2 การทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา

ทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษาจากชุดข้อมูลทดสอบประสิทธิภาพจากไฟล์ StudentStatus-eval.csv จำนวน 325 คน ได้ผลดังนี้

ตารางที่ 4.24 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท

Correctly Classified Instances	307	94.4615 %
Incorrectly Classified Instances	18	5.5385%
Kappa statistic	0.6365	
Mean absolute error	0.0602	
Root mean squared error	0.2312	
Coverage of cases (0.95 level)	95.0769 %	
Total Number of Instances	325	

จากตารางที่ 4.24 จากข้อมูลจำนวน 325 คน ทำนายข้อมูลถูกต้องจำนวน 307 คน คิดเป็น 94.4615% และทำนายข้อมูลไม่ถูกต้องจำนวน 18 คน คิดเป็น 5.5385% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.6365 แสดงว่าข้อมูลมีความสอดคล้องกันปานกลาง มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.0602 ซึ่งมีค่าน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงมาก และมีความค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.2312)^2 = 0.0535$ ซึ่งมีค่าน้อย แสดงว่าตัวแบบมีความถูกต้องมาก

ตารางที่ 4.25 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.973	0.357	0.967	0.973	0.97	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.643	0.027	0.692	0.643	0.667	no
Weighted Avg.	0.945	0.329	0.943	0.945	0.944	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.25 สำหรับคำตอบคงสภาพการเป็นนักศึกษา (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.973 ค่าอัตราความผิดพลาดเชิงบวก = 0.357 ค่าความแม่นยำ = 0.967 ค่าความระลึก = 0.973 และค่าความถ่วงดุล = 0.97 ส่วนคำตอบพ้นสภาพการเป็นนักศึกษา (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.643 ค่าอัตราความผิดพลาดเชิงลบ = 0.027 ค่าความแม่นยำ = 0.692 ค่าความระลึก = 0.643 และค่าความถ่วงดุล = 0.667

ตารางที่ 4.26 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท

		ผลการจำแนกสถานภาพการเป็นนักศึกษา	
		คงสภาพการเป็นนักศึกษา	พ้นสภาพการเป็นนักศึกษา
ค่าที่แท้จริง	คงสภาพการเป็นนักศึกษา	289	8
	พ้นสภาพการเป็นนักศึกษา	10	18

จากตารางที่ 4.26 มีข้อมูล 325 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 307 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าคงสภาพการเป็นนักศึกษา 289 คน และจำนวนข้อมูลที่จำแนกถูกว่าพ้นสภาพการเป็นนักศึกษา 18 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 18 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าคงสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วพ้นสภาพการเป็นนักศึกษา 10 คน และจำนวนข้อมูลที่จำแนกผิดว่าพ้นสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วคงสภาพการเป็นนักศึกษา 8 คน

4.1.3.3 การทำนายสถานภาพการเป็นนักศึกษา

การทำนายสถานภาพการเป็นนักศึกษาจากชุดข้อมูลทดสอบด้วยไฟล์ StudentStatus-test.csv จำนวน 162 คน ได้ผลดังนี้

$$n_{hidden} \le n_{h\,max} = \frac{n_{data\ set} \times n_{input}}{n_{input} + n_{output}}$$
$$= \frac{162 \times 6}{6 + 2}$$
$$= 121.5 \approx 122$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จะได้จำนวนโหนดมากที่สุดในชั้นซ่อนประมาณ 122 โหนด โดยในที่นี้จะกำหนดจำนวนโหนดในชั้นซ่อนโดยแบ่งเป็นช่วงละ 15 คือ 15, 30, 45, 60, 75, 90, 105, 120, 135 และ 150 กำหนดค่าอัตราการเรียนรู้เป็น 0.1, 0.2, 0.3, 0.4 และ 0.5 กำหนดค่าโมเมนตัมเป็น 0.5, 0.6, 0.7, 0.8 และ 0.9 และใช้จำนวนรอบการฝึกหัด คือ 30,000 รอบ

ตารางที่ 4.27 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของการทำนายสถานภาพการเป็นนักศึกษาด้วยวิธีโครงข่ายประสาท สำหรับอัตราการเรียนรู้ $\eta = 0.1$ และโมเมนตัม $\alpha = 0.5, 0.6, 0.7, 0.8$ และ 0.9

โครงข่าย	โมเมนตัม				
	0.5	0.6	0.7	0.8	0.9
6-15-2	0.0590	0.0556	0.0401	0.0452	0.0512
6-30-2	0.0771	0.0282	0.0448	0.0484	0.0432
6-45-2	0.0637	0.0576	0.0635	0.0489	0.0431
6-60-2	0.0312	0.0518	0.0425	0.0558	0.0436
6-75-2	0.0417	0.0389	0.0428	0.0432	0.0337
6-90-2	0.0574	0.0503	0.0563	0.0461	0.0505
6-105-2	0.0398	0.0363	0.0381	0.0659	0.0576
6-120-2	0.0489	0.0501	0.0284	0.0376	0.0496
6-135-2	0.0605	0.0590	0.0605	0.0665	0.0904
6-150-2	0.0434	0.0473	0.0551	0.0719	0.0573

จากตารางที่ 4.27 เมื่อกำหนดค่าอัตราการเรียนรู้เท่ากับ 0.1 และค่าโมเมนตัมเท่ากับ 0.6 จะพบว่าโครงข่ายประสาทที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (MSE = 0.0282) คือโครงข่ายประสาท 6-30-2 ซึ่งเป็นโครงข่ายประสาทที่มีโหนดข้อมูลเข้าเท่ากับ 6 โหนด โหนดซ่อนเท่ากับ 30 โหนด และโหนดข้อมูลออกเท่ากับ 2 โหนด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.28 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของการทำนายสถานภาพการเป็นนักศึกษาด้วยวิธีโครงข่ายประสาท สำหรับอัตราการเรียนรู้ $\eta = 0.2$ และโมเมนตัม $\alpha = 0.5, 0.6, 0.7, 0.8$ และ 0.9

โครงข่าย	โมเมนตัม				
	0.5	0.6	0.7	0.8	0.9
6-15-2	0.0188	0.0429	0.0321	0.0557	0.0548
6-30-2	0.0475	0.0489	0.0412	0.0485	0.0476
6-45-2	0.0565	0.0500	0.0405	0.0506	0.0484
6-60-2	0.0249	0.0441	0.0432	0.0589	0.0604
6-75-2	0.0512	0.0357	0.0299	0.0498	0.0375
6-90-2	0.0515	0.0608	0.0469	0.0436	0.0544
6-105-2	0.0493	0.0536	0.0456	0.0433	0.0489
6-120-2	0.0529	0.0506	0.0549	0.0325	0.0565
6-135-2	0.0654	0.0745	0.0656	0.0483	0.0518
6-150-2	0.0507	0.0419	0.0464	0.0487	0.0447

จากตารางที่ 4.28 เมื่อกำหนดค่าอัตราการเรียนรู้เท่ากับ 0.2 และค่าโมเมนตัมเท่ากับ 0.5 จะพบว่าโครงข่ายประสาทที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (MSE = 0.0188) คือโครงข่ายประสาท 6-15-2 ซึ่งเป็นโครงข่ายประสาทที่มีโหนดข้อมูลเข้าเท่ากับ 6 โหนด โหนดซ่อนเท่ากับ 15 โหนด และโหนดข้อมูลออกเท่ากับ 2 โหนด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.29 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของการทำนายสถานภาพการเป็นนักศึกษาด้วยวิธีโครงข่ายประสาท สำหรับอัตราการเรียนรู้ $\eta = 0.3$ และโมเมนตัม $\alpha = 0.5, 0.6, 0.7, 0.8$ และ 0.9

โครงข่าย	โมเมนตัม				
	0.5	0.6	0.7	0.8	0.9
6-15-2	0.0448	0.0519	0.0530	0.0342	0.0555
6-30-2	0.0686	0.0467	0.0459	0.0475	0.0675
6-45-2	0.0373	0.0610	0.0428	0.0404	0.0556
6-60-2	0.0477	0.0536	0.0366	0.0496	0.0432
6-75-2	0.0430	0.0540	0.0452	0.0454	0.0494
6-90-2	0.0660	0.0607	0.0393	0.0575	0.0571
6-105-2	0.0590	0.0614	0.0367	0.0566	0.0679
6-120-2	0.0668	0.0469	0.0528	0.0592	0.0679
6-135-2	0.0643	0.0546	0.0525	0.0528	0.0679
6-150-2	0.0681	0.0471	0.0540	0.0452	0.0679

จากตารางที่ 4.29 เมื่อกำหนดค่าอัตราการเรียนรู้เท่ากับ 0.3 และค่าโมเมนตัมเท่ากับ 0.8 จะพบว่าโครงข่ายประสาทที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (MSE = 0.0342) คือโครงข่ายประสาท 6-15-2 ซึ่งเป็นโครงข่ายประสาทที่มีโหนดข้อมูลเข้าเท่ากับ 6 โหนด โหนดซ่อนเท่ากับ 15 โหนด และโหนดข้อมูลออกเท่ากับ 2 โหนด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.30 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของการทำนายสถานภาพการเป็นนักศึกษาด้วยวิธีโครงข่ายประสาท สำหรับอัตราการเรียนรู้ $\eta = 0.4$ และโมเมนตัม $\alpha = 0.5, 0.6, 0.7, 0.8$ และ 0.9

โครงข่าย	โมเมนตัม				
	0.5	0.6	0.7	0.8	0.9
6-15-2	0.0394	0.0553	0.0442	0.0723	0.0617
6-30-2	0.0424	0.0548	0.0541	0.0612	0.0738
6-45-2	0.0355	0.0285	0.0478	0.0605	0.0618
6-60-2	0.0353	0.0495	0.0335	0.0555	0.0556
6-75-2	0.0511	0.0338	0.0448	0.0439	0.0617
6-90-2	0.0495	0.0538	0.0428	0.0494	0.0679
6-105-2	0.0564	0.0412	0.0605	0.0404	0.0679
6-120-2	0.0518	0.0431	0.0474	0.0475	0.0679
6-135-2	0.0499	0.0507	0.0427	0.0555	0.0679
6-150-2	0.0573	0.0612	0.0374	0.0556	0.0679

จากตารางที่ 4.30 เมื่อกำหนดค่าอัตราการเรียนรู้เท่ากับ 0.4 และค่าโมเมนตัมเท่ากับ 0.6 จะพบว่าโครงข่ายประสาทที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (MSE = 0.0285) คือโครงข่ายประสาท 6-45-2 ซึ่งเป็นโครงข่ายประสาทที่มีโหนดข้อมูลเข้าเท่ากับ 6 โหนด โหนดซ่อนเท่ากับ 45 โหนด และโหนดข้อมูลออกเท่ากับ 2 โหนด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.31 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของการทำนายสถานภาพการเป็นนักศึกษาด้วยวิธีโครงข่ายประสาท สำหรับอัตราการเรียนรู้ $\eta = 0.5$ และโมเมนตัม $\alpha = 0.5, 0.6, 0.7, 0.8$ และ 0.9

โครงข่าย	โมเมนตัม				
	0.5	0.6	0.7	0.8	0.9
6-15-2	0.0443	0.0486	0.0593	0.0632	0.0679
6-30-2	0.0431	0.0457	0.0419	0.0620	0.0556
6-45-2	0.0424	0.0413	0.0357	0.0695	0.0493
6-60-2	0.0454	0.0329	0.0389	0.0386	0.0610
6-75-2	0.0349	0.0457	0.0370	0.0812	0.0679
6-90-2	0.0474	0.0548	0.0497	0.0556	0.0679
6-105-2	0.0475	0.0364	0.0450	0.0622	0.0679
6-120-2	0.0582	0.0521	0.0585	0.0602	0.0679
6-135-2	0.0552	0.0450	0.0556	0.0371	0.0679
6-150-2	0.0475	0.0607	0.0490	0.0709	0.0680

จากตารางที่ 4.31 เมื่อกำหนดค่าอัตราการเรียนรู้เท่ากับ 0.5 และค่าโมเมนตัมเท่ากับ 0.6 จะพบว่าโครงข่ายประสาทที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (MSE = 0.0329) คือโครงข่ายประสาท 6-60-2 ซึ่งเป็นโครงข่ายประสาทที่มีโหนดข้อมูลเข้าเท่ากับ 6 โหนด โหนดซ่อนเท่ากับ 60 โหนด และโหนดข้อมูลออกเท่ากับ 2 โหนด

จากการเปรียบเทียบทั้งหมด เมื่อกำหนดค่าอัตราการเรียนรู้เท่ากับ 0.2 และค่าโมเมนตัมเท่ากับ 0.5 จะพบว่าโครงข่ายประสาทที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (MSE = 0.0188) คือโครงข่ายประสาท 6-15-2 ซึ่งเป็นโครงข่ายประสาทที่มีโหนดข้อมูลเข้าเท่ากับ 6 โหนด โหนดซ่อนเท่ากับ 15 โหนด และโหนดข้อมูลออกเท่ากับ 2 โหนด ได้ผลดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.32 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีโครงข่ายประสาท

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
1	1:Yes	1:Yes		0.999
2	1:Yes	1:Yes		1
3	1:Yes	1:Yes		1
4	1:Yes	1:Yes		1
5	1:Yes	1:Yes		1
6	1:Yes	1:Yes		1
7	1:Yes	1:Yes		1
8	1:Yes	1:Yes		1
9	1:Yes	1:Yes		1
10	1:Yes	1:Yes		1
11	1:Yes	1:Yes		1
12	1:Yes	1:Yes		1
13	1:Yes	1:Yes		1
14	2:No	1:Yes	+	1
15	1:Yes	1:Yes		1
16	1:Yes	1:Yes		1
17	1:Yes	1:Yes		1
18	1:Yes	1:Yes		1
19	1:Yes	1:Yes		1
20	1:Yes	1:Yes		1
21	1:Yes	1:Yes		1
22	1:Yes	1:Yes		1
23	1:Yes	1:Yes		1
24	1:Yes	1:Yes		1
25	1:Yes	1:Yes		1
26	1:Yes	1:Yes		0.969
27	1:Yes	1:Yes		1
28	1:Yes	1:Yes		1
29	1:Yes	1:Yes		1
30	1:Yes	1:Yes		1
31	1:Yes	1:Yes		1
32	1:Yes	1:Yes		1
33	1:Yes	1:Yes		1
34	1:Yes	1:Yes		1
35	1:Yes	1:Yes		1
36	1:Yes	1:Yes		1

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
37	1:Yes	1:Yes		1
38	1:Yes	1:Yes		1
39	1:Yes	1:Yes		1
40	1:Yes	1:Yes		1
41	1:Yes	1:Yes		1
42	1:Yes	1:Yes		1
43	1:Yes	1:Yes		1
44	1:Yes	1:Yes		1
45	1:Yes	1:Yes		1
46	1:Yes	1:Yes		0.887
47	1:Yes	1:Yes		1
48	1:Yes	1:Yes		1
49	1:Yes	1:Yes		1
50	1:Yes	1:Yes		1
51	1:Yes	1:Yes		1
52	1:Yes	1:Yes		1
53	1:Yes	1:Yes		1
54	1:Yes	1:Yes		1
55	1:Yes	1:Yes		0.974
56	1:Yes	1:Yes		1
57	1:Yes	1:Yes		1
58	1:Yes	1:Yes		1
59	1:Yes	1:Yes		1
60	1:Yes	1:Yes		1
61	1:Yes	1:Yes		1
62	1:Yes	1:Yes		1
63	1:Yes	1:Yes		1
64	1:Yes	1:Yes		0.903
65	1:Yes	1:Yes		1
66	1:Yes	1:Yes		1
67	1:Yes	1:Yes		1
68	1:Yes	1:Yes		1
69	1:Yes	1:Yes		1
70	1:Yes	1:Yes		1
71	1:Yes	1:Yes		1
72	1:Yes	1:Yes		1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.32 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีโครงข่ายประสาท (ต่อ)

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
73	1:Yes	1:Yes		1
74	1:Yes	1:Yes		1
75	1:Yes	1:Yes		1
76	1:Yes	1:Yes		1
77	1:Yes	1:Yes		1
78	1:Yes	1:Yes		1
79	1:Yes	1:Yes		1
80	1:Yes	1:Yes		1
81	2:No	2:No		1
82	1:Yes	1:Yes		1
83	1:Yes	1:Yes		1
84	1:Yes	1:Yes		1
85	1:Yes	1:Yes		1
86	1:Yes	1:Yes		1
87	1:Yes	1:Yes		1
88	2:No	2:No		0.948
89	1:Yes	1:Yes		1
90	1:Yes	1:Yes		1
91	1:Yes	1:Yes		1
92	1:Yes	1:Yes		1
93	1:Yes	1:Yes		1
94	1:Yes	1:Yes		1
95	1:Yes	1:Yes		1
96	1:Yes	1:Yes		1
97	1:Yes	1:Yes		1
98	1:Yes	1:Yes		1
99	1:Yes	1:Yes		0.999
100	2:No	2:No		1
101	1:Yes	1:Yes		1
102	1:Yes	1:Yes		1
103	1:Yes	1:Yes		0.957
104	1:Yes	1:Yes		1
105	1:Yes	1:Yes		0.996
106	1:Yes	1:Yes		1
107	2:No	1:Yes	+	1
108	1:Yes	1:Yes		1

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
109	1:Yes	1:Yes		1
110	1:Yes	1:Yes		1
111	1:Yes	1:Yes		1
112	1:Yes	1:Yes		1
113	2:No	2:No		1
114	1:Yes	1:Yes		1
115	1:Yes	1:Yes		1
116	1:Yes	1:Yes		1
117	1:Yes	1:Yes		1
118	1:Yes	1:Yes		1
119	1:Yes	1:Yes		1
120	1:Yes	1:Yes		1
121	1:Yes	1:Yes		1
122	1:Yes	1:Yes		1
123	1:Yes	1:Yes		1
124	1:Yes	1:Yes		1
125	1:Yes	1:Yes		1
126	1:Yes	1:Yes		1
127	1:Yes	1:Yes		1
128	1:Yes	1:Yes		1
129	1:Yes	1:Yes		1
130	2:No	1:Yes	+	1
131	1:Yes	1:Yes		1
132	1:Yes	1:Yes		1
133	1:Yes	1:Yes		1
134	1:Yes	1:Yes		1
135	2:No	2:No		0.997
136	1:Yes	1:Yes		1
137	1:Yes	1:Yes		1
138	1:Yes	1:Yes		1
139	1:Yes	1:Yes		1
140	1:Yes	1:Yes		1
141	1:Yes	1:Yes		1
142	2:No	2:No		0.975
143	1:Yes	1:Yes		1
144	2:No	2:No		1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.32 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีโครงข่ายประสาท (ต่อ)

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
145	1:Yes	1:Yes		1
146	1:Yes	1:Yes		1
147	1:Yes	1:Yes		1
148	1:Yes	1:Yes		1
149	1:Yes	1:Yes		1
150	2:No	2:No		0.979
151	1:Yes	1:Yes		1
152	1:Yes	1:Yes		1
153	1:Yes	1:Yes		1

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
154	1:Yes	1:Yes		1
155	1:Yes	1:Yes		1
156	1:Yes	1:Yes		1
157	1:Yes	1:Yes		1
158	1:Yes	1:Yes		1
159	1:Yes	1:Yes		0.898
160	1:Yes	1:Yes		1
161	1:Yes	1:Yes		1
162	1:Yes	1:Yes		1

จากตารางที่ 4.32 ค่าทำนายที่ทายผิดในช่อง error จะมีเครื่องหมาย + อยู่จำนวน 3 ค่า คือค่าที่ 14, 107 และ 130 โดยที่ค่าที่ 14, 107 และ 130 พ้นสภาพการเป็นนักศึกษา แต่ทำนายว่าคงสภาพการเป็นนักศึกษาและไม่มีค่าใดคงสภาพการเป็นนักศึกษา แต่ทำนายว่าพ้นสภาพการเป็นนักศึกษา

ตารางที่ 4.33 ผลในส่วนของการสรุปผลจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท

Correctly Classified Instances	159	98.1481 %
Incorrectly Classified Instances	3	1.8519 %
Kappa statistic	0.8325	
Mean absolute error	0.0217	
Root mean squared error	0.137	
Coverage of cases (0.95 level)	98.1481 %	
Total Number of Instances	162	

จากตารางที่ 4.33 จากข้อมูลจำนวน 162 คน ทำนายข้อมูลถูกต้องจำนวน 159 คน คิดเป็น 98.1481% และทำนายข้อมูลไม่ถูกต้องจำนวน 3 คน คิดเป็น 1.8519% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.8325 แสดงว่าข้อมูลมีความสอดคล้องกันมาก มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.0217 ซึ่งมีค่าค่อนข้างน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.137)^2 = 0.0188$ ซึ่งมีค่าน้อย แสดงว่าตัวแบบมีความถูกต้องมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.34 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	1	0.273	0.981	1	0.99	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.727	0	1	0.727	0.842	no
Weighted Avg.	0.981	0.254	0.982	0.981	0.98	

จากตารางที่ 4.34 สำหรับคำตอบคงสภาพการเป็นนักศึกษา (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 1 ค่าอัตราความผิดพลาดเชิงบวก = 0.273 ค่าความแม่นยำ = 0.981 ค่าความระลึก = 1 และค่าความถ่วงดุล = 0.99 ส่วนคำตอบพ้นสภาพการเป็นนักศึกษา (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.727 ค่าอัตราความผิดพลาดเชิงลบ = 0 ค่าความแม่นยำ = 1 ค่าความระลึก = 0.727 และค่าความถ่วงดุล = 0.842

ตารางที่ 4.35 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีโครงข่ายประสาท

		ผลการจำแนกสถานภาพการเป็นนักศึกษา	
		คงสภาพการเป็นนักศึกษา	พ้นสภาพการเป็นนักศึกษา
ค่าที่แท้จริง	คงสภาพการเป็นนักศึกษา	151	0
	พ้นสภาพการเป็นนักศึกษา	3	8

จากตารางที่ 4.35 มีข้อมูล 162 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 159 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าคงสภาพการเป็นนักศึกษา 151 คน และจำนวนข้อมูลที่จำแนกถูกว่าพ้นสภาพการเป็นนักศึกษา 8 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 3 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าคงสภาพการเป็นนักศึกษาซึ่งค่าที่แท้จริงแล้วพ้นสภาพการเป็นนักศึกษา 3 คน และจำนวนข้อมูลที่จำแนกผิดว่าพ้นสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วคงสภาพการเป็นนักศึกษา 0 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.4 วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine Method)

วิธีซัพพอร์ตเวกเตอร์แมชชีนนั้นมีฟังก์ชันหลายฟังก์ชัน แต่ในงานวิจัยนี้ใช้อัลกอริทึม SMO ชนิดโพลิโนเมียลเคอร์เนล (Polynomial Kernel) ในการทำนายสถานภาพการเป็นนักศึกษา ได้ผลดังนี้

4.1.4.1 การสร้างตัวแบบสถานภาพการเป็นนักศึกษา

สร้างตัวแบบสถานภาพการเป็นนักศึกษาจากชุดข้อมูลฝึกหัดด้วยไฟล์ StudentStatus.csv จำนวน 1,138 คน ได้ผลดังนี้

ตารางที่ 4.36 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

Correctly Classified Instances	1044	91.7399%
Incorrectly Classified Instances	94	8.2601 %
Kappa statistic	0.4114	
Mean absolute error	0.0826	
Root mean squared error	0.2874	
Relative absolute error	43.0149 %	
Root relative squared error	92.8968%	
Coverage of cases (0.95 level)	91.7399 %	
Mean rel. region size (0.95 level)	50 %	
Total Number of Instances	1138	

จากตารางที่ 4.36 จากข้อมูลจำนวน 1,138 คน ทำนายข้อมูลถูกต้องจำนวน 1,044 คน คิดเป็น 91.7399% และทำนายข้อมูลไม่ถูกต้องจำนวน 94 คน คิดเป็น 8.2601% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.4114 แสดงว่าข้อมูลมีความสอดคล้องกันปานกลาง มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.0826 ซึ่งมีค่าน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงมากและมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.2874)^2 = 0.0826$ ซึ่งมีค่าน้อย แสดงว่าตัวแบบมีความถูกต้องมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.37 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.99	0.689	0.923	0.99	0.955	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.311	0.010	0.792	0.311	0.447	no
Weighted Avg.	0.917	0.616	0.909	0.917	0.901	

จากตารางที่ 4.37 สำหรับคำตอบคงสภาพการเป็นนักศึกษา (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.99 ค่าอัตราความผิดพลาดเชิงบวก = 0.689 ค่าความแม่นยำ = 0.923 ค่าความระลึก = 0.99 และค่าความถ่วงดุล = 0.955 ส่วนคำตอบพ้นสภาพการเป็นนักศึกษา (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.311 ค่าอัตราความผิดพลาดเชิงลบ = 0.010 ค่าความแม่นยำ = 0.792 ค่าความระลึก = 0.311 และค่าความถ่วงดุล = 0.447

ตารางที่ 4.38 ผลในส่วนของเมตริกซ์ความสับสนจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

		ผลการจำแนกสถานภาพการเป็นนักศึกษา	
		คงสภาพการเป็นนักศึกษา	พ้นสภาพการเป็นนักศึกษา
ค่าที่แท้จริง	คงสภาพการเป็นนักศึกษา	1,006	10
	พ้นสภาพการเป็นนักศึกษา	84	38

จากตารางที่ 4.38 มีข้อมูล 1,138 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 1,044 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าคงสภาพการเป็นนักศึกษา 1,006 คน และจำนวนข้อมูลที่จำแนกถูกว่าพ้นสภาพการเป็นนักศึกษา 38 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 94 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าคงสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วพ้นสภาพการเป็นนักศึกษา 84 คน และจำนวนข้อมูลที่จำแนกผิดว่าพ้นสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วคงสภาพการเป็นนักศึกษา 10 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.4.2 การทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา

ทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษาจากชุดข้อมูลทดสอบประสิทธิภาพด้วยไฟล์ StudentStatus-eval.csv จำนวน 325 คน ได้ผลดังนี้

ตารางที่ 4.39 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

Correctly Classified Instances	305	93.8462%
Incorrectly Classified Instances	20	6.1538%
Kappa statistic	0.5151	
Mean absolute error	0.0615	
Root mean squared error	0.2481	
Coverage of cases (0.95 level)	93.8462%	
Total Number of Instances	325	

จากตารางที่ 4.39 จากข้อมูลจำนวน 325 คน ทำนายข้อมูลถูกต้องจำนวน 305 คน คิดเป็น 93.8462% และทำนายข้อมูลไม่ถูกต้องจำนวน 20 คน คิดเป็น 6.1538% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.5151 แสดงว่าข้อมูลมีความสอดคล้องกันปานกลาง มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.0615 ซึ่งมีค่าน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงมาก และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.2481)^2 = 0.0616$ ซึ่งมีค่าน้อย แสดงว่าตัวแบบมีความถูกต้องมาก

ตารางที่ 4.40 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.987	0.571	0.948	0.987	0.967	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.429	0.013	0.75	0.429	0.545	no
Weighted Avg.	0.938	0.523	0.931	0.938	0.931	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.40 สำหรับคำตอบคงสภาพการเป็นนักศึกษา (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.987 ค่าอัตราความผิดพลาดเชิงบวก = 0.571 ค่าความแม่นยำ = 0.948 ค่าความระลึก = 0.987 และค่าความถ่วงดุล = 0.967 ส่วนคำตอบพ้นสภาพการเป็นนักศึกษา (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.429 ค่าอัตราความผิดพลาดเชิงลบ = 0.013 ค่าความแม่นยำ = 0.75 ค่าความระลึก = 0.429 และค่าความถ่วงดุล = 0.545

ตารางที่ 4.41 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

		ผลการจำแนกสถานภาพการเป็นนักศึกษา	
		คงสภาพการเป็นนักศึกษา	พ้นสภาพการเป็นนักศึกษา
ค่าที่แท้จริง	คงสภาพการเป็นนักศึกษา	293	4
	พ้นสภาพการเป็นนักศึกษา	16	12

จากตารางที่ 4.41 มีข้อมูล 325 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 305 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าคงสภาพการเป็นนักศึกษา 293 คน และจำนวนข้อมูลที่จำแนกถูกว่าพ้นสภาพการเป็นนักศึกษา 12 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 20 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าคงสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วพ้นสภาพการเป็นนักศึกษา 16 คน และจำนวนข้อมูลที่จำแนกผิดว่าพ้นสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วคงสภาพการเป็นนักศึกษา 4 คน

4.1.4.3 การทำนายสถานภาพการเป็นนักศึกษา

การทำนายสถานภาพการเป็นนักศึกษาจากชุดข้อมูลทดสอบด้วยไฟล์ StudentStatus-test.csv จำนวน 162 คน ได้ผลดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.42 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
1	1:Yes	1:Yes		1
2	1:Yes	1:Yes		1
3	1:Yes	1:Yes		1
4	1:Yes	1:Yes		1
5	1:Yes	1:Yes		1
6	1:Yes	1:Yes		1
7	1:Yes	1:Yes		1
8	1:Yes	1:Yes		1
9	1:Yes	1:Yes		1
10	1:Yes	1:Yes		1
11	1:Yes	1:Yes		1
12	1:Yes	1:Yes		1
13	1:Yes	1:Yes		1
14	2:No	1:Yes	+	1
15	1:Yes	1:Yes		1
16	1:Yes	1:Yes		1
17	1:Yes	1:Yes		1
18	1:Yes	1:Yes		1
19	1:Yes	1:Yes		1
20	1:Yes	1:Yes		1
21	1:Yes	1:Yes		1
22	1:Yes	1:Yes		1
23	1:Yes	1:Yes		1
24	1:Yes	1:Yes		1
25	1:Yes	1:Yes		1
26	1:Yes	1:Yes		1
27	1:Yes	1:Yes		1
28	1:Yes	1:Yes		1
29	1:Yes	1:Yes		1
30	1:Yes	1:Yes		1
31	1:Yes	1:Yes		1
32	1:Yes	1:Yes		1
33	1:Yes	1:Yes		1
34	1:Yes	1:Yes		1
35	1:Yes	1:Yes		1
36	1:Yes	1:Yes		1

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
37	1:Yes	1:Yes		1
38	1:Yes	1:Yes		1
39	1:Yes	1:Yes		1
40	1:Yes	1:Yes		1
41	1:Yes	1:Yes		1
42	1:Yes	1:Yes		1
43	1:Yes	1:Yes		1
44	1:Yes	1:Yes		1
45	1:Yes	1:Yes		1
46	1:Yes	1:Yes		1
47	1:Yes	1:Yes		1
48	1:Yes	1:Yes		1
49	1:Yes	1:Yes		1
50	1:Yes	1:Yes		1
51	1:Yes	1:Yes		1
52	1:Yes	1:Yes		1
53	1:Yes	1:Yes		1
54	1:Yes	1:Yes		1
55	1:Yes	1:Yes		1
56	1:Yes	1:Yes		1
57	1:Yes	1:Yes		1
58	1:Yes	1:Yes		1
59	1:Yes	1:Yes		1
60	1:Yes	1:Yes		1
61	1:Yes	1:Yes		1
62	1:Yes	1:Yes		1
63	1:Yes	1:Yes		1
64	1:Yes	1:Yes		1
65	1:Yes	1:Yes		1
66	1:Yes	1:Yes		1
67	1:Yes	1:Yes		1
68	1:Yes	1:Yes		1
69	1:Yes	1:Yes		1
70	1:Yes	1:Yes		1
71	1:Yes	1:Yes		1
72	1:Yes	1:Yes		1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.42 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน (ต่อ)

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
73	1:Yes	1:Yes		1
74	1:Yes	1:Yes		1
75	1:Yes	1:Yes		1
76	1:Yes	1:Yes		1
77	1:Yes	1:Yes		1
78	1:Yes	1:Yes		1
79	1:Yes	1:Yes		1
80	1:Yes	1:Yes		1
81	2:No	2:No		1
82	1:Yes	1:Yes		1
83	1:Yes	1:Yes		1
84	1:Yes	1:Yes		1
85	1:Yes	1:Yes		1
86	1:Yes	1:Yes		1
87	1:Yes	1:Yes		1
88	2:No	1:Yes	+	1
89	1:Yes	1:Yes		1
90	1:Yes	1:Yes		1
91	1:Yes	1:Yes		1
92	1:Yes	1:Yes		1
93	1:Yes	1:Yes		1
94	1:Yes	1:Yes		1
95	1:Yes	1:Yes		1
96	1:Yes	1:Yes		1
97	1:Yes	1:Yes		1
98	1:Yes	1:Yes		1
99	1:Yes	1:Yes		1
100	2:No	1:Yes	+	1
101	1:Yes	1:Yes		1
102	1:Yes	1:Yes		1
103	1:Yes	1:Yes		1
104	1:Yes	1:Yes		1
105	1:Yes	1:Yes		1
106	1:Yes	1:Yes		1
107	2:No	1:Yes	+	1
108	1:Yes	1:Yes		1

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
109	1:Yes	1:Yes		1
110	1:Yes	1:Yes		1
111	1:Yes	1:Yes		1
112	1:Yes	1:Yes		1
113	2:No	1:Yes	+	1
114	1:Yes	1:Yes		1
115	1:Yes	1:Yes		1
116	1:Yes	1:Yes		1
117	1:Yes	1:Yes		1
118	1:Yes	1:Yes		1
119	1:Yes	1:Yes		1
120	1:Yes	1:Yes		1
121	1:Yes	1:Yes		1
122	1:Yes	1:Yes		1
123	1:Yes	1:Yes		1
124	1:Yes	1:Yes		1
125	1:Yes	1:Yes		1
126	1:Yes	1:Yes		1
127	1:Yes	1:Yes		1
128	1:Yes	1:Yes		1
129	1:Yes	1:Yes		1
130	2:No	1:Yes	+	1
131	1:Yes	1:Yes		1
132	1:Yes	1:Yes		1
133	1:Yes	1:Yes		1
134	1:Yes	1:Yes		1
135	2:No	1:Yes	+	1
136	1:Yes	1:Yes		1
137	1:Yes	1:Yes		1
138	1:Yes	1:Yes		1
139	1:Yes	1:Yes		1
140	1:Yes	1:Yes		1
141	1:Yes	1:Yes		1
142	2:No	1:Yes	+	1
143	1:Yes	1:Yes		1
144	2:No	2:No		1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.42 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน (ต่อ)

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
145	1:Yes	1:Yes		1
146	1:Yes	1:Yes		1
147	1:Yes	1:Yes		1
148	1:Yes	1:Yes		1
149	1:Yes	1:Yes		1
150	2:No	1:Yes	+	1
151	1:Yes	1:Yes		1
152	1:Yes	1:Yes		1
153	1:Yes	1:Yes		1

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
154	1:Yes	1:Yes		1
155	1:Yes	1:Yes		1
156	1:Yes	1:Yes		1
157	1:Yes	1:Yes		1
158	1:Yes	1:Yes		1
159	1:Yes	1:Yes		1
160	1:Yes	1:Yes		1
161	1:Yes	1:Yes		1
162	1:Yes	1:Yes		1

จากตารางที่ 4.42 ค่าทำนายที่ทายผิดในช่อง error จะมีเครื่องหมาย + อยู่จำนวน 9 ค่า คือค่าที่ 14, 88, 100, 107, 113, 130, 135, 142 และ150 โดยที่ค่าที่ 14, 88, 100, 107, 113, 130, 135, 142 และ 150 พ้นสภาพการเป็นนักศึกษา แต่ทำนายว่าคงสภาพการเป็นนักศึกษา และไม่มีค่าใดคงสภาพการเป็นนักศึกษา แต่ทำนายว่าพ้นสภาพการเป็นนักศึกษา

ตารางที่ 4.43 ผลในส่วนของการสรุปผลจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

Correctly Classified Instances	153	94.4444 %
Incorrectly Classified Instances	9	5.5556%
Kappa statistic	0.2929	
Mean absolute error	0.0556	
Root mean squared error	0.2357	
Coverage of cases (0.95 level)	94.4444%	
Total Number of Instances	162	

จากตารางที่ 4.43 โดยจากข้อมูลจำนวน 162 คน ทำนายข้อมูลถูกต้องจำนวน 153 คน คิดเป็น 94.4444% และทำนายข้อมูลไม่ถูกต้องจำนวน 9 คน คิดเป็น 5.5556% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.2929 แสดงว่าข้อมูลมีความสอดคล้องกันค่อนข้างน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.0556 ซึ่งมีค่าน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงมาก และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.2357)^2 = 0.0556$ ซึ่งมีค่าน้อย แสดงว่าตัวแบบมีความถูกต้องมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.44 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	1	0.818	0.944	1	0.971	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.182	0	1	0.182	0.308	no
Weighted Avg.	0.944	0.763	0.948	0.944	0.926	

จากตารางที่ 4.44 สำหรับคำตอบคงสภาพการเป็นนักศึกษา (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 1 ค่าอัตราความผิดพลาดเชิงบวก = 0.818 ค่าความแม่นยำ = 0.944 ค่าความระลึก = 1 และค่าความถ่วงดุล = 0.971 ส่วนคำตอบพ้นสภาพการเป็นนักศึกษา (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.182 ค่าอัตราความผิดพลาดเชิงลบ = 0 ค่าความแม่นยำ = 1 ค่าความระลึก = 0.182 และค่าความถ่วงดุล = 0.308

ตารางที่ 4.45 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

		ผลการจำแนกสถานภาพการเป็นนักศึกษา	
		คงสภาพการเป็นนักศึกษา	พ้นสภาพการเป็นนักศึกษา
ค่าที่แท้จริง	คงสภาพการเป็นนักศึกษา	151	0
	พ้นสภาพการเป็นนักศึกษา	9	2

จากตารางที่ 4.45 มีข้อมูล 162 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 153 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าคงสภาพการเป็นนักศึกษา 151 คน และจำนวนข้อมูลที่จำแนกถูกว่าพ้นสภาพการเป็นนักศึกษา 2 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 9 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าคงสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วพ้นสภาพการเป็นนักศึกษา 9 คน และจำนวนข้อมูลที่จำแนกผิดว่าพ้นสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วคงสภาพการเป็นนักศึกษา 0 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.5 วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม (Binary Logistic Regression Method)

เนื่องจากข้อมูลสถานภาพการเป็นนักศึกษามีตัวแปรอิสระหลายตัวและมีตัวแปรตามเพียงตัวเดียว เป็นตัวแปรเชิงคุณภาพที่มีค่าได้เพียง 2 ค่า จึงเลือกใช้วิธีการถดถอยโลจิสติกแบบ 2 กลุ่มได้ผลการทำนายสถานภาพการเป็นนักศึกษา ได้ผลดังนี้

4.1.5.1 การสร้างตัวแบบสถานภาพการเป็นนักศึกษา

สร้างตัวแบบสถานภาพการเป็นนักศึกษาจากชุดข้อมูลฝึกหัด ด้วยไฟล์ StudentStatus.csv จำนวน 1,138 คน ได้ผลดังนี้

ตารางที่ 4.46 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

Correctly Classified Instances	1056	92.7944 %
Incorrectly Classified Instances	82	7.2056 %
Kappa statistic	0.524	
Mean absolute error	0.1155	
Root mean squared error	0.2384	
Relative absolute error	60.1474 %	
Root relative squared error	77.0462 %	
Coverage of cases (0.95 level)	98.7698 %	
Mean rel. region size (0.95 level)	70.7821 %	
Total Number of Instances	1138	

จากตางรางที่ 4.46 จากข้อมูลจำนวน 1,138 คน ทำนายข้อมูลถูกต้องจำนวน 1,056 คน คิดเป็น 92.7944% และทำนายข้อมูลไม่ถูกต้องจำนวน 82 คน คิดเป็น 7.2056% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.524 แสดงว่าข้อมูลมีความสอดคล้องปานกลาง มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.1155 ซึ่งมีค่าน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงมาก และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.2384)^2 = 0.0568$ ซึ่งมีค่าน้อย แสดงว่าตัวแบบมีความถูกต้องมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.47 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.988	0.574	0.935	0.988	0.961	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.426	0.012	0.813	0.426	0.559	no
Weighted Avg.	0.928	0.514	0.922	0.928	0.918	

จากตารางที่ 4.47 สำหรับคำตอบคงสภาพการเป็นนักศึกษา (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.988 ค่าอัตราความผิดพลาดเชิงบวก = 0.574 ค่าความแม่นยำ = 0.935 ค่าความระลึก = 0.988 และค่าความถ่วงดุล = 0.961 ส่วนคำตอบพ้นสภาพการเป็นนักศึกษา (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.426 ค่าอัตราความผิดพลาดเชิงลบ = 0.012 ค่าความแม่นยำ = 0.813 ค่าความระลึก = 0.426 และค่าความถ่วงดุล = 0.559

ตารางที่ 4.48 ผลในส่วนของเมตริกซ์ความสับสน จากการสร้างตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

		ผลการจำแนกสถานภาพการเป็นนักศึกษา	
		คงสภาพการเป็นนักศึกษา	พ้นสภาพการเป็นนักศึกษา
ค่าที่แท้จริง	คงสภาพการเป็นนักศึกษา	1,004	12
	พ้นสภาพการเป็นนักศึกษา	70	52

จากตารางที่ 4.48 มีข้อมูล 1,138 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 1,056 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าคงสภาพการเป็นนักศึกษา 1,004 คน และจำนวนข้อมูลที่จำแนกถูกว่าพ้นสภาพการเป็นนักศึกษา 52 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 82 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าคงสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วพ้นสภาพการเป็นนักศึกษา 70 คน และจำนวนข้อมูลที่จำแนกผิดว่าพ้นสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วคงสภาพการเป็นนักศึกษา 12 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.5.2 การทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา

ทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษาจากชุดข้อมูลทดสอบประสิทธิภาพด้วยไฟล์ StudentStatus-eval.csv จำนวน 100 คน ได้ผลดังนี้

ตารางที่ 4.49 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

Correctly Classified Instances	306	94.1538%
Incorrectly Classified Instances	19	5.8462%
Kappa statistic	0.5484	
Mean absolute error	0.1015	
Root mean squared error	0.2223	
Coverage of cases (0.95 level)	98.7692%	
Total Number of Instances	325	

จากตารางที่ 4.49 จากข้อมูลจำนวน 325 คน ทำนายข้อมูลถูกต้องจำนวน 306 คน คิดเป็น 94.1538% และทำนายข้อมูลไม่ถูกต้องจำนวน 19 คน คิดเป็น 5.8462% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.5484 แสดงว่าข้อมูลมีความสอดคล้องกันปานกลาง มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.1015 ซึ่งมีค่าน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงมาก และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.2223)^2 = 0.0494$ ซึ่งมีค่าน้อย แสดงว่าตัวแบบมีความถูกต้องมาก

ตารางที่ 4.50 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.987	0.536	0.951	0.987	0.969	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.464	0.013	0.765	0.464	0.578	no
Weighted Avg.	0.942	0.491	0.935	0.942	0.935	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.50 สำหรับคำตอบคงสภาพการเป็นนักศึกษา (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.987 ค่าอัตราความผิดพลาดเชิงบวก = 0.536 ค่าความแม่นยำ = 0.951 ค่าความระลึก = 0.987 และค่าความถ่วงดุล = 0.969 ส่วนคำตอบพ้นสภาพการเป็นนักศึกษา (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.464 ค่าอัตราความผิดพลาดเชิงลบ = 0.013 ค่าความแม่นยำ = 0.765 ค่าความระลึก = 0.464 และค่าความถ่วงดุล = 0.578

ตารางที่ 4.51 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบสถานภาพการเป็นนักศึกษา ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

		ผลการจำแนกสถานภาพการเป็นนักศึกษา	
		คงสภาพการเป็นนักศึกษา	พ้นสภาพการเป็นนักศึกษา
ค่าที่แท้จริง	คงสภาพการเป็นนักศึกษา	293	4
	พ้นสภาพการเป็นนักศึกษา	15	13

จากตารางที่ 4.51 มีข้อมูล 325 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 306 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าคงสภาพการเป็นนักศึกษา 293 คน และจำนวนข้อมูลที่จำแนกถูกว่าพ้นสภาพการเป็นนักศึกษา 13 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 19 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าคงสภาพการเป็นนักศึกษาซึ่งค่าที่แท้จริงแล้วพ้นสภาพการเป็นนักศึกษา 15 คน และจำนวนข้อมูลที่จำแนกผิดว่าพ้นสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วคงสภาพการเป็นนักศึกษา 4 คน

4.1.5.3 การทำนายสถานภาพการเป็นนักศึกษา

การทำนายสถานภาพการเป็นนักศึกษาจากชุดข้อมูลทดสอบด้วยไฟล์ StudentStatus-test.csv จำนวน 162 คน ได้ผลดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.52 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
1	1:Yes	1:Yes		0.915
2	1:Yes	1:Yes		0.974
3	1:Yes	1:Yes		0.958
4	1:Yes	1:Yes		0.915
5	1:Yes	1:Yes		0.989
6	1:Yes	1:Yes		0.986
7	1:Yes	1:Yes		0.978
8	1:Yes	1:Yes		0.999
9	1:Yes	1:Yes		0.998
10	1:Yes	1:Yes		0.989
11	1:Yes	1:Yes		0.959
12	1:Yes	1:Yes		0.978
13	1:Yes	1:Yes		0.944
14	2:No	1:Yes	+	0.971
15	1:Yes	1:Yes		0.984
16	1:Yes	1:Yes		0.975
17	1:Yes	1:Yes		0.997
18	1:Yes	1:Yes		0.984
19	1:Yes	1:Yes		0.992
20	1:Yes	1:Yes		0.952
21	1:Yes	1:Yes		0.697
22	1:Yes	1:Yes		0.986
23	1:Yes	1:Yes		0.845
24	1:Yes	1:Yes		0.988
25	1:Yes	1:Yes		0.898
26	1:Yes	1:Yes		0.809
27	1:Yes	1:Yes		0.987
28	1:Yes	1:Yes		0.982
29	1:Yes	1:Yes		0.892
30	1:Yes	1:Yes		0.981
31	1:Yes	1:Yes		0.992
32	1:Yes	1:Yes		0.972
33	1:Yes	1:Yes		0.998
34	1:Yes	1:Yes		0.999
35	1:Yes	1:Yes		0.984
36	1:Yes	1:Yes		0.999

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
37	1:Yes	1:Yes		0.957
38	1:Yes	1:Yes		0.989
39	1:Yes	1:Yes		0.975
40	1:Yes	1:Yes		1
41	1:Yes	1:Yes		0.985
42	1:Yes	1:Yes		0.909
43	1:Yes	1:Yes		0.884
44	1:Yes	1:Yes		0.999
45	1:Yes	1:Yes		0.986
46	1:Yes	1:Yes		0.68
47	1:Yes	1:Yes		0.875
48	1:Yes	1:Yes		0.999
49	1:Yes	1:Yes		0.996
50	1:Yes	1:Yes		0.999
51	1:Yes	1:Yes		0.999
52	1:Yes	1:Yes		0.992
53	1:Yes	1:Yes		0.999
54	1:Yes	1:Yes		0.949
55	1:Yes	1:Yes		0.801
56	1:Yes	1:Yes		0.952
57	1:Yes	1:Yes		0.997
58	1:Yes	1:Yes		0.927
59	1:Yes	1:Yes		0.82
60	1:Yes	1:Yes		1
61	1:Yes	1:Yes		0.754
62	1:Yes	1:Yes		0.898
63	1:Yes	1:Yes		0.9
64	1:Yes	2:No	+	0.609
65	1:Yes	1:Yes		0.997
66	1:Yes	1:Yes		0.918
67	1:Yes	1:Yes		0.97
68	1:Yes	1:Yes		0.98
69	1:Yes	1:Yes		0.83
70	1:Yes	1:Yes		0.995
71	1:Yes	1:Yes		0.95
72	1:Yes	1:Yes		0.858

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.52 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม (ต่อ)

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
73	1:Yes	1:Yes		0.992
74	1:Yes	1:Yes		0.95
75	1:Yes	1:Yes		0.982
76	1:Yes	1:Yes		0.966
77	1:Yes	1:Yes		0.943
78	1:Yes	1:Yes		0.976
79	1:Yes	1:Yes		0.969
80	1:Yes	1:Yes		0.972
81	2:No	2:No		0.891
82	1:Yes	1:Yes		0.978
83	1:Yes	1:Yes		0.937
84	1:Yes	1:Yes		0.994
85	1:Yes	1:Yes		0.979
86	1:Yes	1:Yes		0.961
87	1:Yes	1:Yes		0.999
88	2:No	1:Yes	+	0.903
89	1:Yes	1:Yes		0.924
90	1:Yes	1:Yes		0.963
91	1:Yes	1:Yes		0.992
92	1:Yes	1:Yes		0.974
93	1:Yes	1:Yes		0.955
94	1:Yes	1:Yes		0.735
95	1:Yes	1:Yes		0.982
96	1:Yes	1:Yes		0.987
97	1:Yes	1:Yes		0.917
98	1:Yes	1:Yes		0.894
99	1:Yes	1:Yes		0.843
100	2:No	1:Yes	+	0.731
101	1:Yes	1:Yes		0.992
102	1:Yes	1:Yes		0.79
103	1:Yes	1:Yes		0.872
104	1:Yes	1:Yes		0.882
105	1:Yes	1:Yes		0.856
106	1:Yes	1:Yes		0.996
107	2:No	1:Yes	+	0.979
108	1:Yes	1:Yes		0.995
109	1:Yes	1:Yes		0.999
110	1:Yes	1:Yes		0.984
111	1:Yes	1:Yes		0.865
112	1:Yes	1:Yes		0.99
113	2:No	1:Yes	+	0.761
114	1:Yes	1:Yes		0.95
115	1:Yes	1:Yes		0.998
116	1:Yes	1:Yes		0.883
117	1:Yes	1:Yes		0.971
118	1:Yes	1:Yes		0.799
119	1:Yes	1:Yes		0.903
120	1:Yes	1:Yes		0.997
121	1:Yes	1:Yes		0.981
122	1:Yes	1:Yes		0.999
123	1:Yes	1:Yes		0.983
124	1:Yes	1:Yes		0.977
125	1:Yes	1:Yes		0.988
126	1:Yes	1:Yes		0.999
127	1:Yes	1:Yes		0.97
128	1:Yes	1:Yes		0.997
129	1:Yes	1:Yes		0.963
130	2:No	1:Yes	+	0.931
131	1:Yes	1:Yes		0.878
132	1:Yes	1:Yes		0.952
133	1:Yes	1:Yes		0.959
134	1:Yes	1:Yes		0.94
135	2:No	1:Yes	+	0.703
136	1:Yes	1:Yes		0.918
137	1:Yes	1:Yes		0.931
138	1:Yes	1:Yes		0.953
139	1:Yes	1:Yes		0.968
140	1:Yes	1:Yes		0.928
141	1:Yes	1:Yes		0.997
142	2:No	1:Yes	+	0.75
143	1:Yes	1:Yes		0.967
144	2:No	2:No		0.664

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.52 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม (ต่อ)

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
145	1:Yes	1:Yes		0.98
146	1:Yes	1:Yes		0.911
147	1:Yes	1:Yes		0.969
148	1:Yes	1:Yes		0.97
149	1:Yes	1:Yes		0.934
150	2:No	1:Yes	+	0.503
151	1:Yes	1:Yes		0.801
152	1:Yes	1:Yes		0.973
153	1:Yes	1:Yes		0.881

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
154	1:Yes	1:Yes		0.996
155	1:Yes	1:Yes		0.917
156	1:Yes	1:Yes		0.968
157	1:Yes	1:Yes		0.94
158	1:Yes	1:Yes		0.999
159	1:Yes	1:Yes		0.688
160	1:Yes	1:Yes		0.99
161	1:Yes	1:Yes		0.947
162	1:Yes	1:Yes		0.722

จากตารางที่ 4.52 ค่าทำนายที่ทายผิดในช่อง error จะมีเครื่องหมาย + อยู่จำนวน 10 ค่า คือค่าที่ 14, 64, 88, 100, 107, 113, 130, 135, 142 และ 150 โดยที่ค่าที่ 14, 88, 100, 107, 113, 130, 135, 142 และ 150 พ้นสภาพการเป็นนักศึกษาแต่ทำนายว่าคงสภาพการเป็นนักศึกษา และค่าที่ 64 คงสภาพการเป็นนักศึกษาแต่ทำนายว่าพ้นสภาพการเป็นนักศึกษา

ตารางที่ 4.53 ผลในส่วนของการสรุปผลจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

Correctly Classified Instances	152	93.8272 %
Incorrectly Classified Instances	10	6.1728%
Kappa statistic	0.2643	
Mean absolute error	0.1035	
Root mean squared error	0.218	
Coverage of cases (0.95 level)	98.7654%	
Total Number of Instances	162	

จากตารางที่ 4.53 โดยจากข้อมูลจำนวน 162 คน ทำนายข้อมูลถูกต้องจำนวน 152 คน คิดเป็น 93.8272% และทำนายข้อมูลไม่ถูกต้องจำนวน 10 คน คิดเป็น 6.1728% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.2643 แสดงว่าข้อมูลมีความสอดคล้องกันค่อนข้างน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.1035 ซึ่งมีค่าน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงมาก และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.218)^2 = 0.0475$ ซึ่งมีค่าน้อย แสดงว่าตัวแบบมีความถูกต้องมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.54 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.993	0.818	0.943	0.993	0.968	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.182	0.007	0.667	0.182	0.286	no
Weighted Avg.	0.938	0.763	0.925	0.938	0.921	

จากตารางที่ 4.54 สำหรับคำตอบคงสภาพการเป็นนักศึกษา (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.993 ค่าอัตราความผิดพลาดเชิงบวก = 0.818 ค่าความแม่นยำ = 0.943 ค่าความระลึก = 0.993 และค่าความถ่วงดุล = 0.968 ส่วนคำตอบพ้นสภาพการเป็นนักศึกษา (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.182 ค่าอัตราความผิดพลาดเชิงลบ = 0.007 ค่าความแม่นยำ = 0.667 ค่าความระลึก = 0.182 และค่าความถ่วงดุล = 0.286

ตารางที่ 4.55 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายสถานภาพการเป็นนักศึกษา ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

		ผลการจำแนกสถานภาพการเป็นนักศึกษา	
		คงสภาพการเป็นนักศึกษา	พ้นสภาพการเป็นนักศึกษา
ค่าที่แท้จริง	คงสภาพการเป็นนักศึกษา	150	1
	พ้นสภาพการเป็นนักศึกษา	9	2

จากตารางที่ 4.55 มีข้อมูล 162 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 152 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าคงสภาพการเป็นนักศึกษา 150 คน และจำนวนข้อมูลที่จำแนกถูกว่าพ้นสภาพการเป็นนักศึกษา 2 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 10 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าคงสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วพ้นสภาพการเป็นนักศึกษา 9 คน และจำนวนข้อมูลที่จำแนกผิดว่าพ้นสภาพการเป็นนักศึกษา ซึ่งค่าที่แท้จริงแล้วคงสภาพการเป็นนักศึกษา 1 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 ผลการเปรียบเทียบประสิทธิภาพและทำนายผลของวิธีการจำแนกกลุ่ม

จากการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มได้ถูกต้องและความแม่นยำในการทำนายสถานภาพการเป็นนักศึกษา ระหว่างวิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาท วิธีซัพพอร์ตเวกเตอร์แมชชีนและวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม ได้ผลดังตารางที่ 4.56 และ 4.57

ตารางที่ 4.56 ผลการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มในสถานภาพการเป็นนักศึกษา

ทั้ง 5 วิธี

การเปรียบเทียบประสิทธิภาพ	ค่าความถูกต้อง (Accuracy)	ค่าอัตราความถูกต้องเชิงบวก (TP Rate)	ค่าอัตราความถูกต้องเชิงลบ (TN Rate)	ค่าความแม่นยำ (Precision)
วิธีความใกล้เคียงกันมากที่สุด	95.0617%	0.987	0.455	0.961
วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ	96.9136%	**1**	0.545	0.968
วิธีโครงข่ายประสาท	**98.1481%**	**1**	**0.727**	**0.981**
วิธีซัพพอร์ตทเวกเตอร์แมชชีน	94.4444%	**1**	0.182	0.944
วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม	93.8272%	0.993	0.182	0.943

ตารางที่ 4.56 ผลการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มในสถานภาพการเป็นนักศึกษา

ทั้ง 5 วิธี (ต่อ)

การเปรียบเทียบประสิทธิภาพ	ค่าอัตราความผิดพลาดเชิงบวก (FP Rate)	ค่าอัตราความผิดพลาดเชิงลบ (FN Rate)	ค่าความระลึก (Recall)	ค่าความถ่วงดุล (F-Measure)
วิธีความใกล้เคียงกันมากที่สุด	0.545	0.013	0.987	0.974
วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ	0.455	**0**	**1**	0.984
วิธีโครงข่ายประสาท	**0.273**	**0**	**1**	**0.99**
วิธีซัพพอร์ตเวกเตอร์แมชชีน	0.818	**0**	**1**	0.971
วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม	0.818	0.007	0.993	0.968

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.57 ผลการเปรียบเทียบการทำนายผลของวิธีการจำแนกกลุ่มสถานภาพการเป็นนักศึกษาทั้ง 5 วิธี

การเปรียบเทียบการทำนายผล	ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE)	ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE)
วิธีความใกล้เคียงกันมากที่สุด	0.0502	0.0493
วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ	0.0651	0.0301
วิธีโครงข่ายประสาท	**0.0217**	**0.0188**
วิธีซัพพอร์ตเวกเตอร์แมชชีน	0.0556	0.0556
วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม	0.1035	0.0475

จากตารางที่ 4.56 พบว่าวิธีโครงข่ายประสาทมีทุกค่า (ค่าความถูกต้อง ค่าอัตราความถูกต้องเชิงบวก ค่าอัตราความถูกต้องเชิงลบ ค่าความแม่นยำ ค่าอัตราความผิดพลาดเชิงบวก ค่าอัตราความผิดพลาดเชิงลบ ค่าความระลึก และค่าความถ่วงดุล) ดีที่สุด คือ 98.1481%, 1, 0.727, 0.981, 0.273, 0, 1 และ 0.99 ตามลำดับ ส่วนวิธีแผนภาพต้นไม้เพื่อการตัดสินใจและวิธีซัพพอร์ตเวกเตอร์แมชชีนมีค่าอัตราความถูกต้องเชิงบวก ค่าอัตราความผิดพลาดเชิงลบ และค่าความระลึกดีที่สุดเช่นเดียวกับวิธีโครงข่ายประสาท คือ 1, 0 และ 1 ตามลำดับ ดังนั้นจากการพิจารณาโดยภาพรวมพบว่าวิธีโครงข่ายประสาทมีประสิทธิภาพดีที่สุด

จากตารางที่ 4.57 พบว่าวิธีโครงข่ายประสาทมีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด คือ 0.0217 และ 0.0188 ตามลำดับ แสดงว่ามีความคลาดเคลื่อนน้อยที่สุด ดังนั้นวิธีโครงข่ายประสาทมีการทำนายผลดีที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

ในการทำวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาทำความเข้าใจและเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่ม รวมทั้งเปรียบเทียบการทำนายผลสถานภาพการเป็นนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง วิธีการจำแนกกลุ่มที่นำมาเปรียบเทียบประสิทธิภาพและการทำนายผลคือ วิธีความใกล้เคียงกันมากที่สุดโดยใช้อัลกอริทึมชนิด IBk วิธีแผนภาพต้นไม้เพื่อการตัดสินใจโดยใช้อัลกอริทึมชนิด J48 วิธีโครงข่ายประสาทโดยใช้อัลกอริทึมชนิดเพอร์เซปตรอนแบบหลายชั้น กำหนดค่าอัตราการเรียนรู้เป็น 0.1, 0.2, 0.3, 0.4 และ 0.5 ค่าโมเมนตัมเป็น 0.5, 0.6, 0.7, 0.8 และ 0.9 จำนวนรอบการสอน 30,000 รอบ และชั้นซ่อน 1 ชั้น วิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้อัลกอริทึม SMO ชนิดโพลิโนเมียลเคอร์เนล และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม ในการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มทั้ง 5 วิธี จะใช้ค่าความถูกต้อง ค่าอัตราความถูกต้องเชิงบวก ค่าอัตราความถูกต้องเชิงลบ ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล โดยจะพิจารณาจากค่าเหล่านี้ที่มีค่ามากที่สุด ส่วนค่าอัตราความผิดพลาดเชิงบวกและค่าอัตราความผิดพลาดเชิงลบจะพิจารณาจากค่าน้อยที่สุด ในการทำนายผลของวิธีการจำแนกกลุ่มทั้ง 5 วิธี จะใช้ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ในการเปรียบเทียบ โดยพิจารณาจากค่า MAE และ MSE ที่มีค่าน้อยที่สุด พบว่าวิธีโครงข่ายประสาทมีค่าความถูกต้อง ค่าอัตราความถูกต้องเชิงบวก ค่าอัตราความถูกต้องเชิงลบ ค่าความแม่นยำ ค่าอัตราความผิดพลาดเชิงบวก ค่าอัตราความผิดพลาดเชิงลบ ค่าความระลึกและค่าความถ่วงดุลดีที่สุดคือ 98.1481%, 1, 0.727, 0.981, 0.273, 0, 1 และ 0.99 ตามลำดับ ส่วนวิธีแผนภาพต้นไม้เพื่อการตัดสินใจและวิธีซัพพอร์ตเวกเตอร์แมชชีนมีค่าอัตราความถูกต้องเชิงบวก ค่าอัตราความผิดพลาดเชิงลบและค่าความระลึกดีที่สุดคือ 1, 0 และ 1 ตามลำดับ ดังนั้นวิธีโครงข่ายประสาทมีประสิทธิภาพดีที่สุด

ส่วนการเปรียบเทียบการทำนายผลของวิธีการจำแนกกลุ่มสถานภาพการเป็นนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง พบว่าวิธีโครงข่ายประสาทมีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุดคือ 0.0217 และ 0.0188 ตามลำดับ ดังนั้นวิธีโครงข่ายประสาทมีการทำนายผลดีที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 ข้อเสนอแนะ

1. ตัวแปรที่นำมาใช้ในการงานวิจัยนี้เป็นเพียงส่วนหนึ่งของการพิจารณาสถานภาพการเป็นนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง เท่านั้น เพื่อให้การทำนายมีประสิทธิภาพมากขึ้น ควรเพิ่มตัวแปรที่เกี่ยวข้องอื่น ๆ อีก เช่น อายุ ภาควิชา ที่พักอาศัย จังหวัดของโรงเรียนในระดับมัธยมศึกษาตอนปลาย คะแนนเฉลี่ยสะสมในระดับมัธยมศึกษาตอนปลาย ประเภทของโรงเรียนที่นักศึกษาสำเร็จการศึกษา การเข้าเป็นนักศึกษาของสถาบันฯ การศึกษาในมหาวิทยาลัย อันดับที่เลือกเข้าเป็นนักศึกษาในสถาบันฯ คะแนน GAT/PAT จำนวนหน่วยกิตที่ลงทะเทียนเรียน จำนวนหน่วยกิตที่ผ่าน และคะแนนเฉลี่ยทุกรายวิชา เป็นต้น

2. เพื่อให้ผลสรุปครอบคลุมกว้างขวางเพิ่มขึ้น ควรจะทำการศึกษาวิธีอื่น ๆ ที่เป็นเทคนิคการทำเหมืองข้อมูลในการจำแนกกลุ่มเหมือนกัน เช่น วิธีนาอีฟเบย์ (Naïve Bayes Method) และโครงข่ายความเชื่อของเบย์เซียน (Bayesian Belief Network)

3. เพื่อให้ได้ข้อสรุปของผลการวิเคราะห์ข้อมูลที่มีความสมบูรณ์มากขึ้น เราอาจจะใช้อัลกอริทึมประเภทอื่น ๆ โดยวิธีความใกล้เคียงกันมากที่สุดยังมีอัลกอริทึม KStar และ LWL วิธีแผนภาพต้นไม้เพื่อการตัดสินใจมีอัลกอริทึม Decision Stump, LMT, Random Forest, Random Tree และ REP Tree วิธีโครงข่ายประสาทสามารถกำหนดค่าอัตราการเรียนรู้และค่าโมเมนตัมที่ละเอียดมากขึ้นกว่าเดิม อาจกำหนดจำนวนชั้นซ่อนมากกว่า 1 ชั้นได้ และอาจเพิ่มจำนวนรอบการสอนให้มากขึ้น เช่น 100,000 รอบ และวิธีซัพพอร์ตเวกเตอร์แมชชีนมีอัลกอริทึม Normalized Poly Kernel, REF Kernel และ Puk

4. นอกจากที่กล่าวมาแล้วในข้อที่ 3 เราอาจจะวิเคราะห์ข้อมูลด้วยอัลกอริทึมประเภทอื่น ๆ อีก วิธีซัพพอร์ตเวกเตอร์แมชชีนมี Kernel Function อีก 2 แบบ คือ Radial Basis Function Kernel และ Sigmoid Kernel และวิธีแผนภาพต้นไม้เพื่อการตัดสินใจมีอัลกอริทึมที่ใช้ในการจำแนกกลุ่มอื่น ๆ อีก เช่น ID3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก
รายละเอียดและตัวอย่างข้อมูลที่นำมาวิเคราะห์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อมูลของสถานภาพการเป็นนักศึกษา

ตารางที่ ก-1 คุณลักษณะและรายละเอียดของสถานภาพการเป็นนักศึกษา

คุณลักษณะ	รายละเอียด
1) สาขาวิชา (DEPARTMENT)	สาขาวิชาคณิตศาสตร์ (MATH) สาขาวิชาวิทยาการคอมพิวเตอร์ (COM) สาขาวิชาเคมีอุตสาหกรรม (IDC) สาขาวิชาเคมีสิ่งแวดล้อม (ERC) สาขาวิชาเทคโนโลยีชีวภาพ (BIO) สาขาวิชาจุลชีววิทยาอุตสาหกรรม (MICRO) สาขาวิชาฟิสิกส์ประยุกต์ (PHY) สาขาวิชาสถิติประยุกต์ (STAT)
2) เพศ (SEX)	ชาย (Male) หญิง (Female)
3) คะแนนเฉลี่ยประจำภาคการศึกษาที่ 1 ปีการศึกษา 2556 (GPS1)	เป็นตัวเลข (Numeric)
4) คะแนนเฉลี่ยประจำภาคการศึกษาที่ 2 ปีการศึกษา 2556 (GPS2)	เป็นตัวเลข (Numeric)
5) วิธีผ่านการสอบคัดเลือก (TYPE)	ระบบแอดมิดชั่น (Admission) ระบบโควตา (Quota) ระบบรับตรง (Test)
6) สถานภาพการเป็นนักศึกษา (STATUS)	คงสภาพการเป็นนักศึกษา (Yes) พ้นสภาพการเป็นนักศึกษา เนื่องจากคะแนนเฉลี่ยสะสม และไม่ได้ลงทะเบียนเรียน (No)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก-2 ตัวอย่างข้อมูลสถานภาพการเป็นนักศึกษา

No	DEPARTMENT	SEX	GPS1	GPS2	TYPE	STATUS
1	MATH	F	1.64	2.29	Quota	Yes
2	MATH	F	2.25	2.35	Quota	Yes
3	MATH	M	1.4	2.4	Test	Yes
4	MATH	M	1.78	2.12	Quota	Yes
5	MATH	F	2.5	2.21	Admission	Yes
6	MATH	F	2.07	2.41	Test	Yes
7	MATH	M	2.3	2	Admission	Yes
8	MATH	F	2.85	3	Test	Yes
9	MATH	F	2.57	2.92	Admission	Yes
10	MATH	F	2.33	2.71	Quota	Yes
11	MATH	M	1.5	2.33	Test	Yes
12	MATH	F	2.25	2	Test	Yes
13	MATH	F	2.07	1.7	Test	Yes
14	MATH	F	2.08	2.1	Admission	No
15	MATH	F	2.08	2.33	Test	Yes
16	MATH	F	1.91	2.25	Test	Yes
17	MATH	F	2.5	2.8	Admission	Yes
18	MATH	M	2.4	2.08	Admission	Yes
19	COM	F	2.92	2.85	Quota	Yes
20	COM	M	2.5	1.85	Test	Yes
21	COM	M	1.16	2	Test	Yes
22	COM	M	2.71	2.71	Quota	Yes
23	COM	F	2.08	1.64	Test	Yes
24	COM	F	2.78	2.42	Admission	Yes
25	COM	M	2	1.9	Test	Yes
26	COM	F	1.5	2.12	Admission	Yes
27	COM	F	2.5	2.64	Admission	Yes
28	COM	M	2.6	2.25	Test	Yes

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

No	DEPARTMENT	SEX	GPS1	GPS2	TYPE	STATUS
29	COM	F	2	2	Admission	Yes
30	COM	M	1.62	3.1	Test	Yes
31	COM	F	2.75	2.58	Test	Yes
32	COM	F	2.5	2.25	Admission	Yes
33	COM	F	3.21	2.92	Test	Yes
34	COM	M	3.28	3.42	Quota	Yes
35	COM	M	2.5	2.5	Admission	Yes
36	COM	F	3.57	3.21	Admission	Yes
37	COM	M	2.64	2.21	Quota	Yes
38	COM	M	2.92	2.2	Test	Yes
39	COM	F	2.42	2.29	Test	Yes
40	COM	M	3.57	3.35	Admission	Yes
41	COM	F	2.78	2.64	Quota	Yes
42	COM	M	2.14	1.92	Admission	Yes
43	COM	F	1.78	2.5	Quota	Yes
44	COM	M	3.35	3.07	Test	Yes
45	COM	F	2.85	2.64	Quota	Yes
46	COM	F	1.64	2	Quota	Yes
47	COM	F	1.92	2.33	Quota	Yes
48	COM	M	3.35	3.42	Admission	Yes
49	COM	F	2.85	2.78	Test	Yes
50	COM	F	3	3.21	Test	Yes
51	COM	F	3.35	3.28	Test	Yes
52	COM	M	2.57	2.75	Admission	Yes
53	COM	M	3.08	3.5	Quota	Yes
54	COM	M	2.71	2.06	Quota	Yes
55	COM	M	1.66	2.25	Quota	Yes
56	COM	M	2.57	2.21	Quota	Yes
57	COM	F	2.78	3.08	Admission	Yes
58	COM	M	2	2.08	Test	Yes

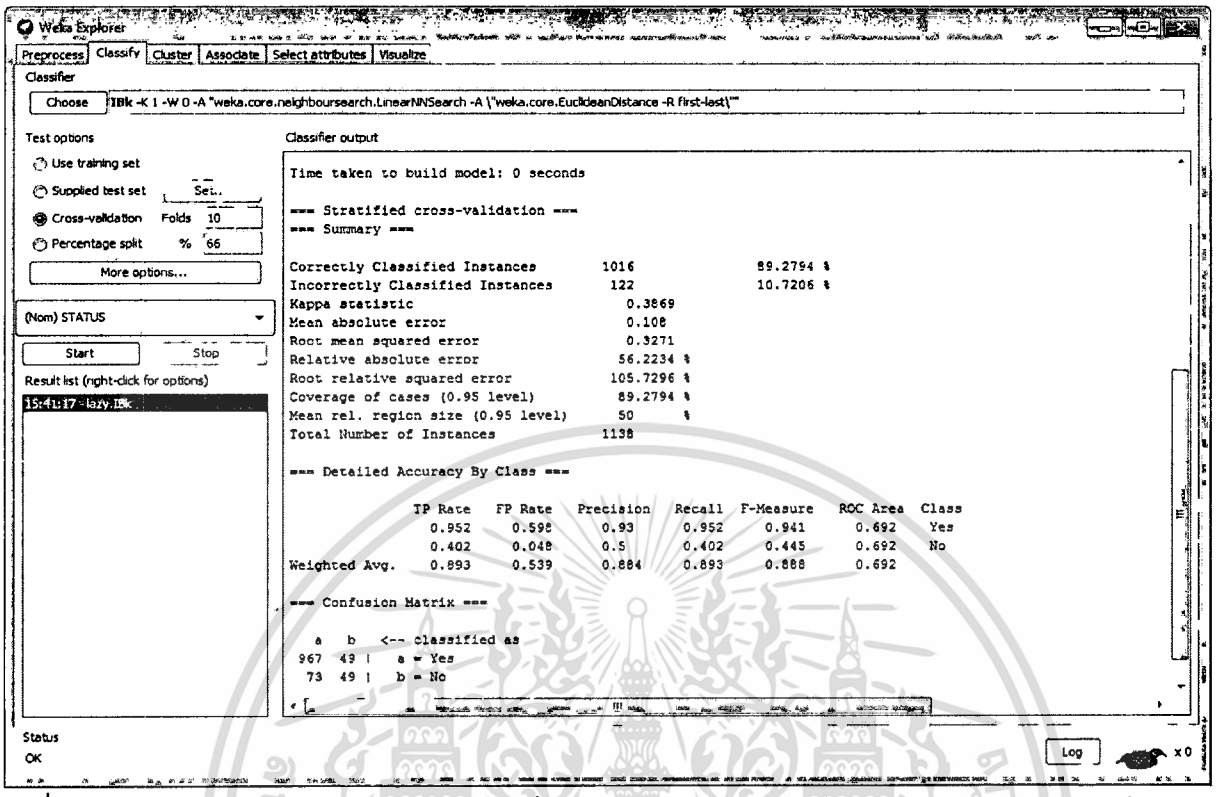
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

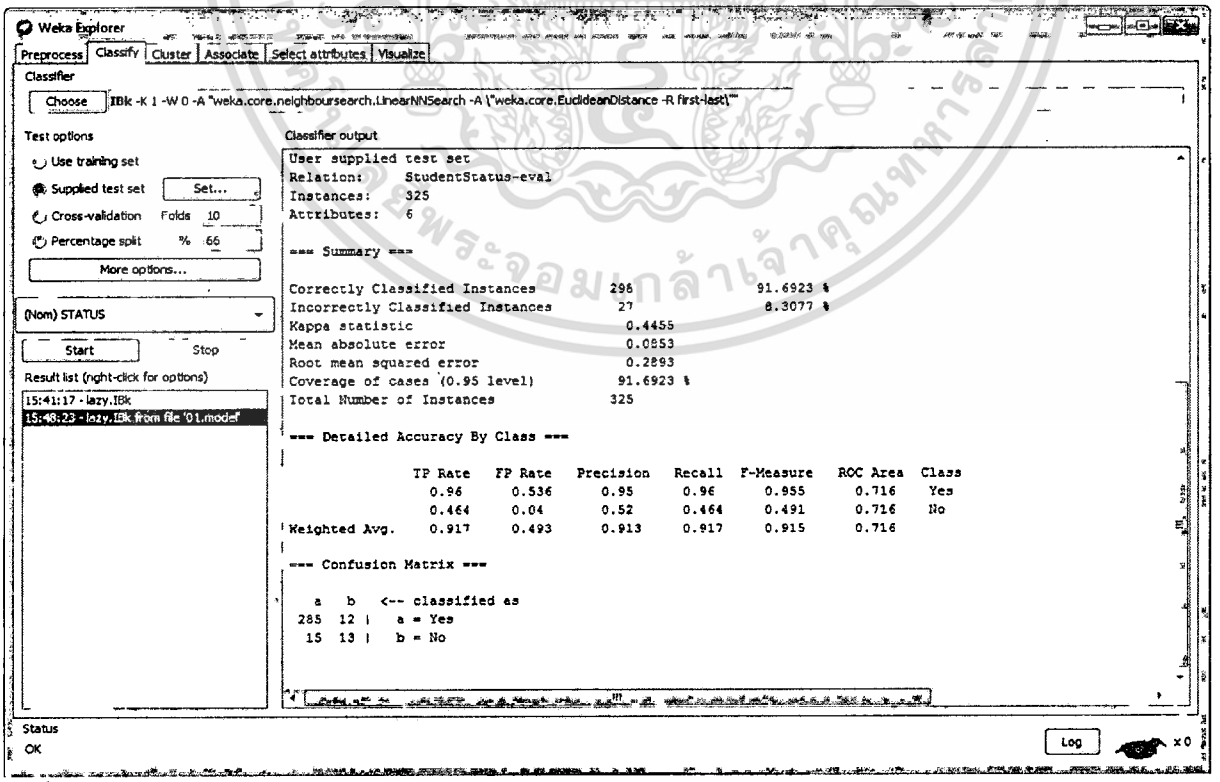
การวิเคราะห์ข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. วิธีความใกล้เคียงกันมากที่สุด (K-Nearest Neighbor : KNN)

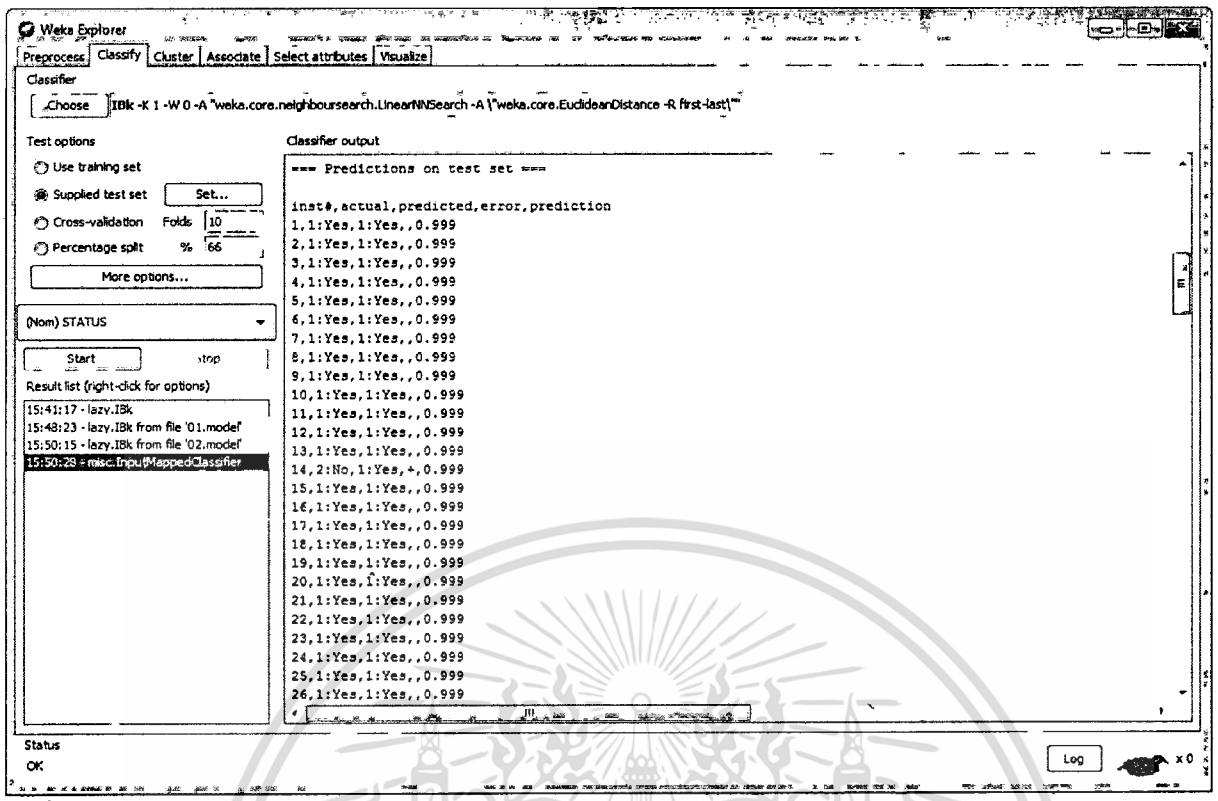


รูปที่ ข-1 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการสร้างตัวแบบโดยวิธีความใกล้เคียงกันมากที่สุด

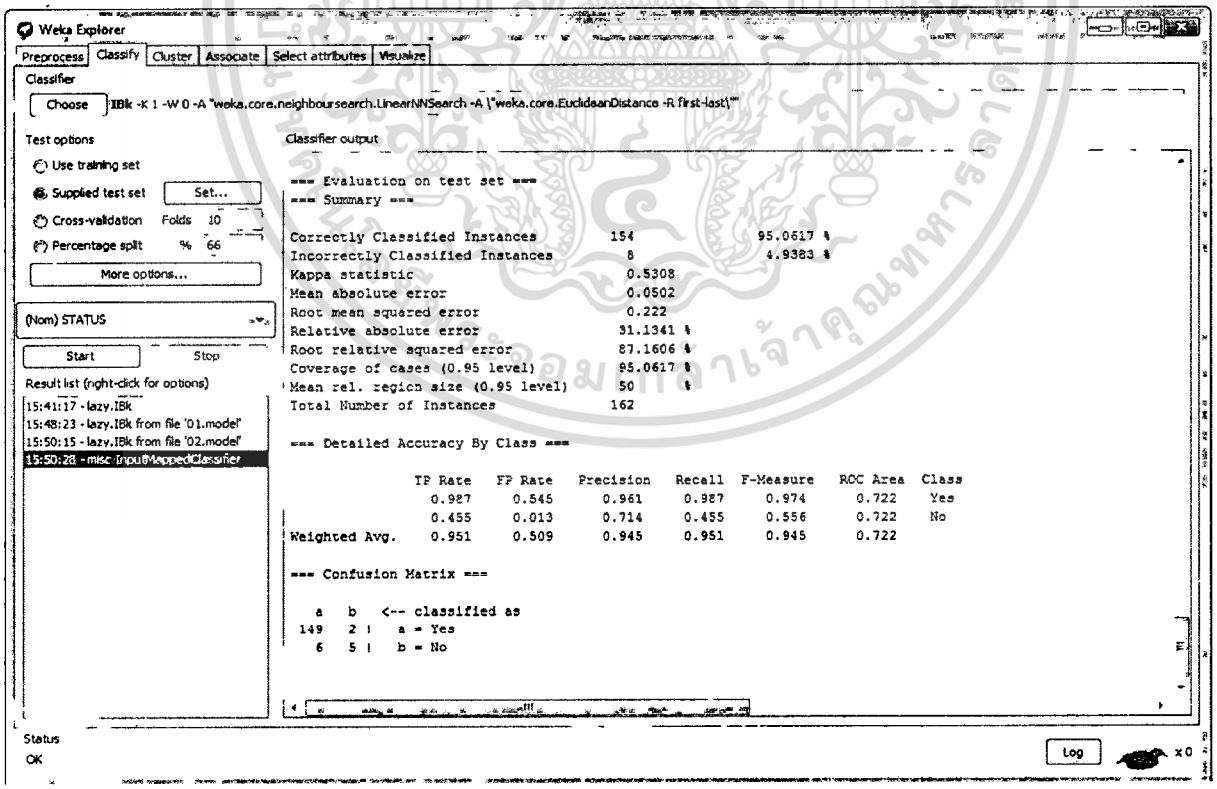


รูปที่ ข-2 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการทดสอบตัวแบบโดยวิธีความใกล้เคียงกันมากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



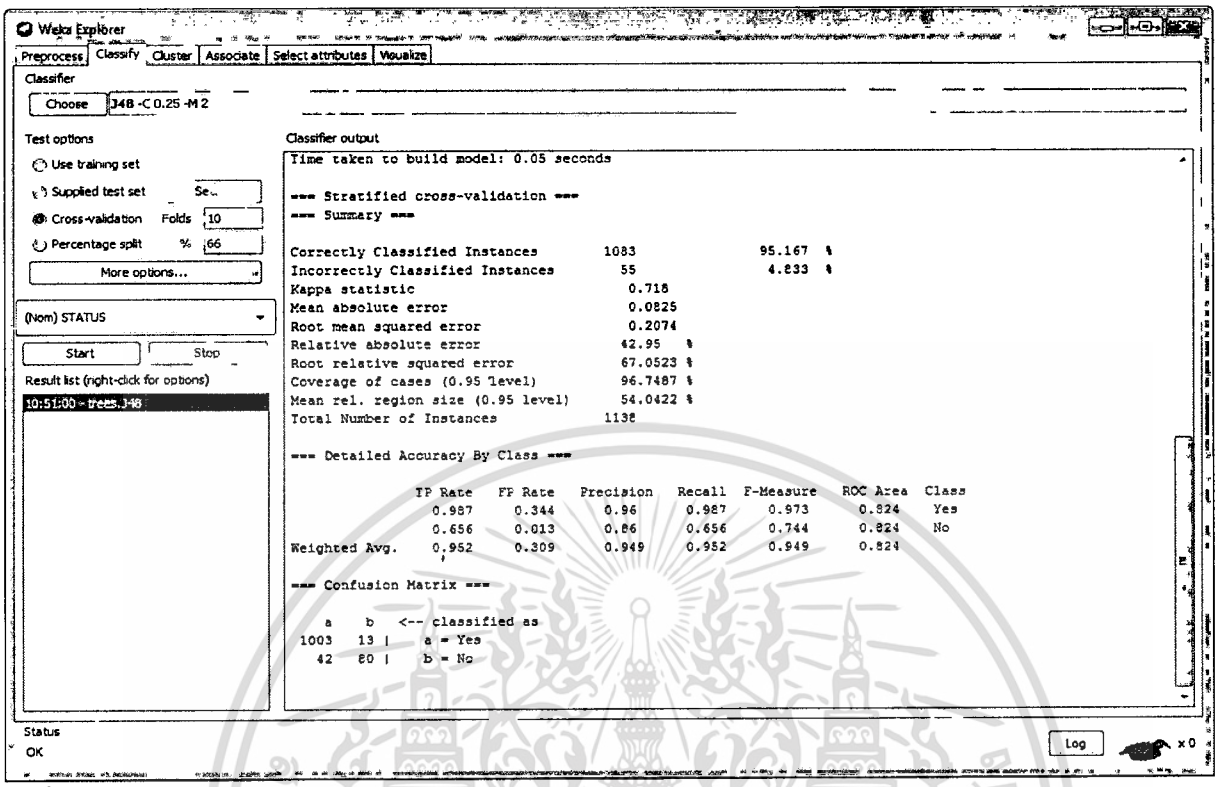
รูปที่ ข-3 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการทำนายตัวแบบโดยวิธีความใกล้เคียงกันมากที่สุด



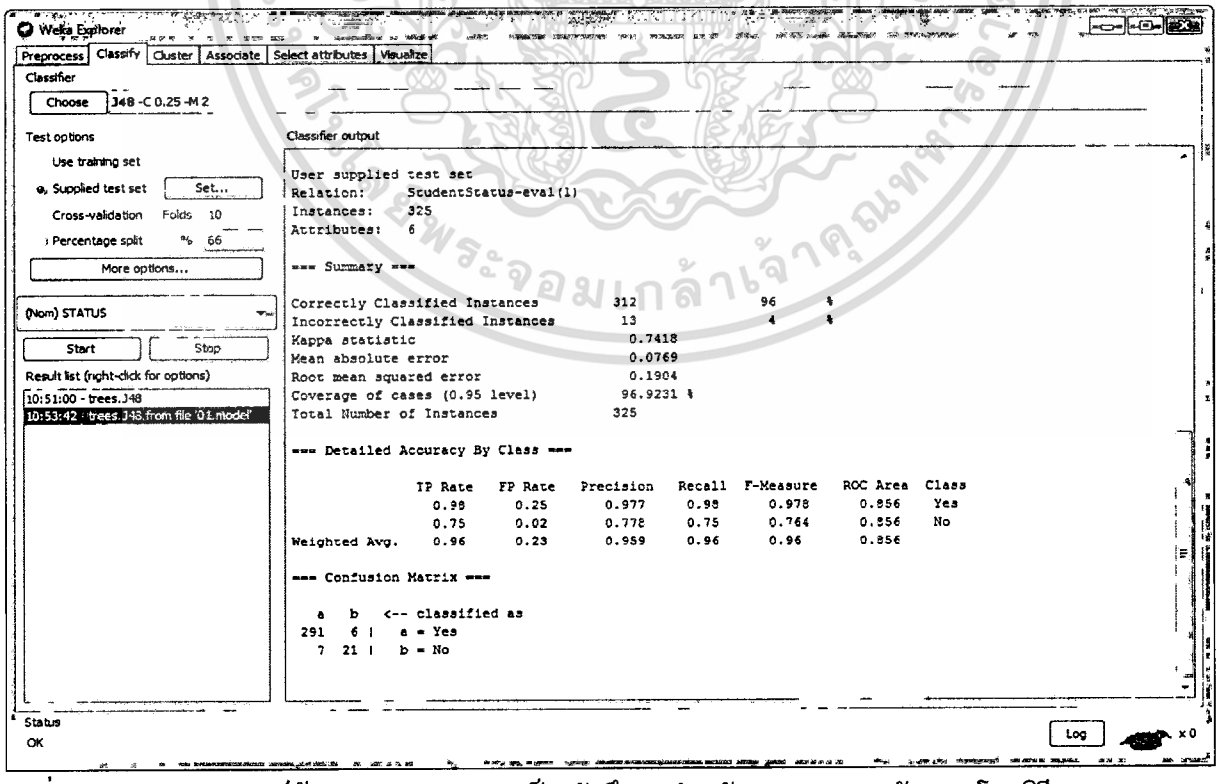
รูปที่ ข-4 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการทำนายตัวแบบโดยวิธีความใกล้เคียงกันมากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree)

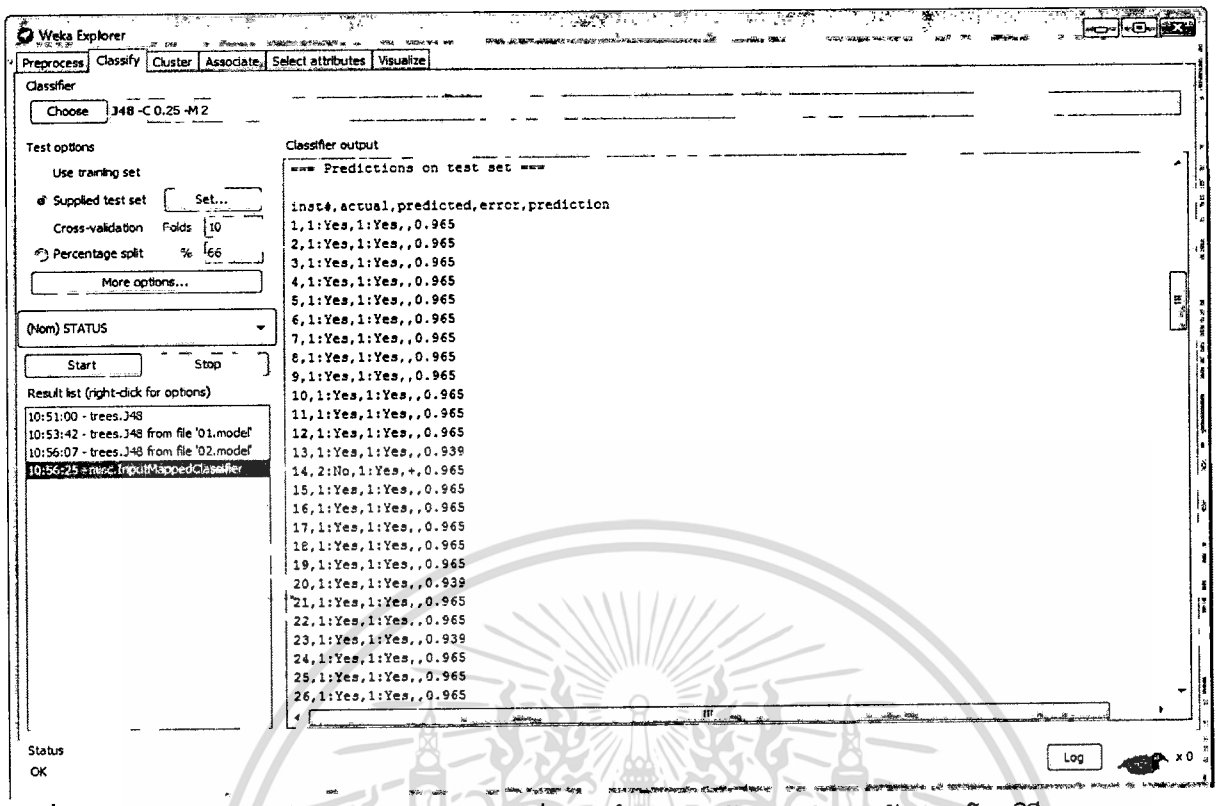


รูปที่ ข-5 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการสร้างตัวแบบโดยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

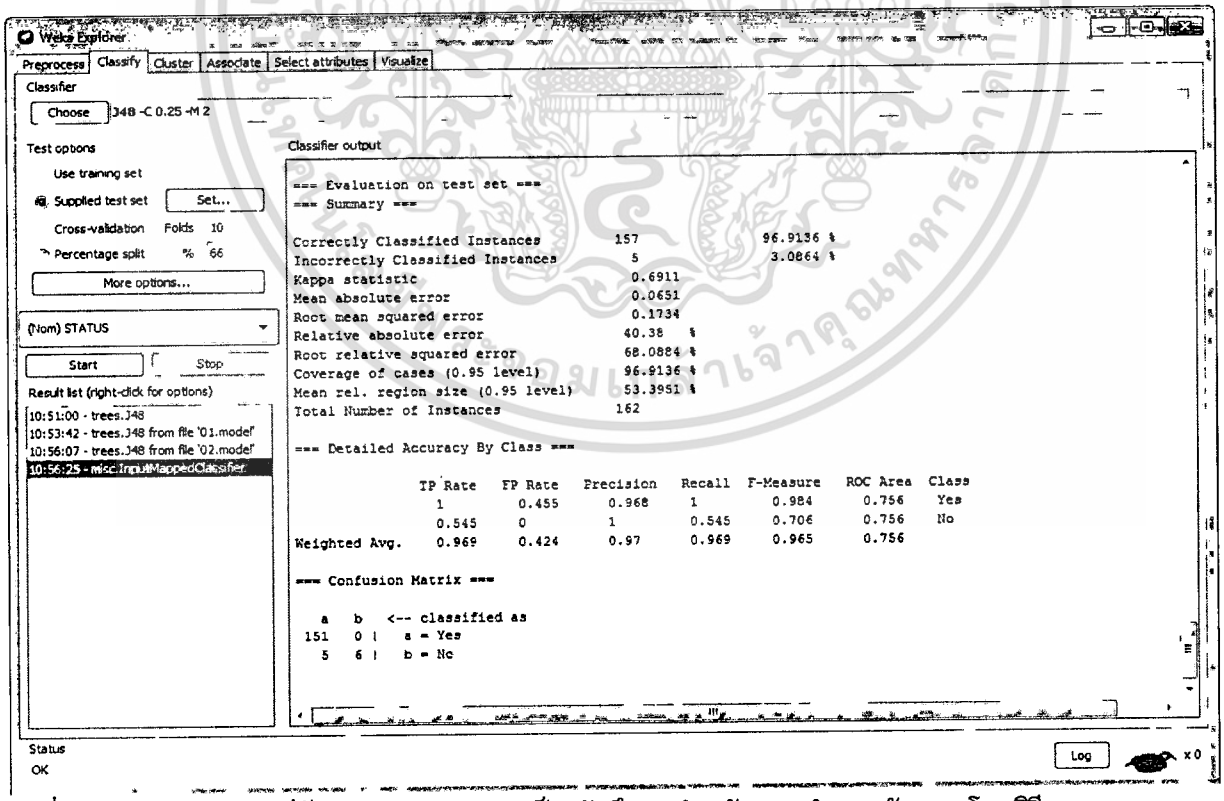


รูปที่ ข-6 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการทดสอบตัวแบบโดยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

เอกสารนี้สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



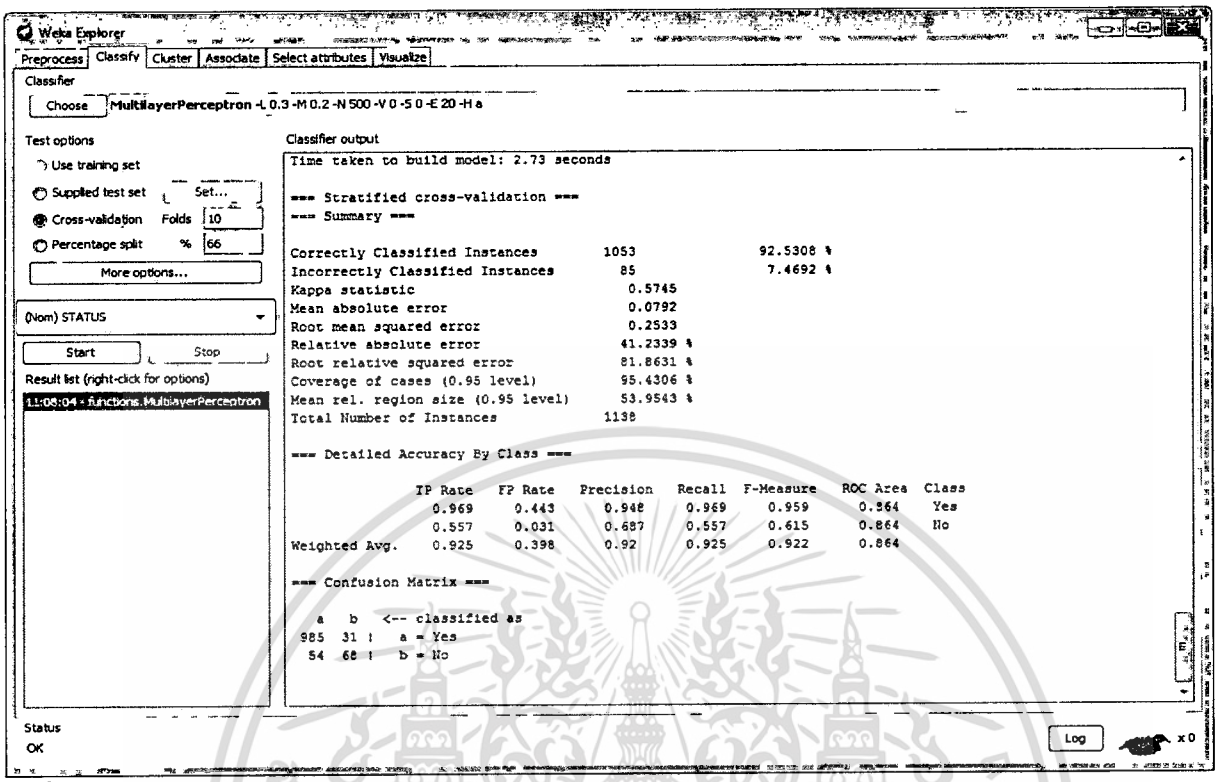
รูปที่ ข-7 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการทำนายตัวแบบโดยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ



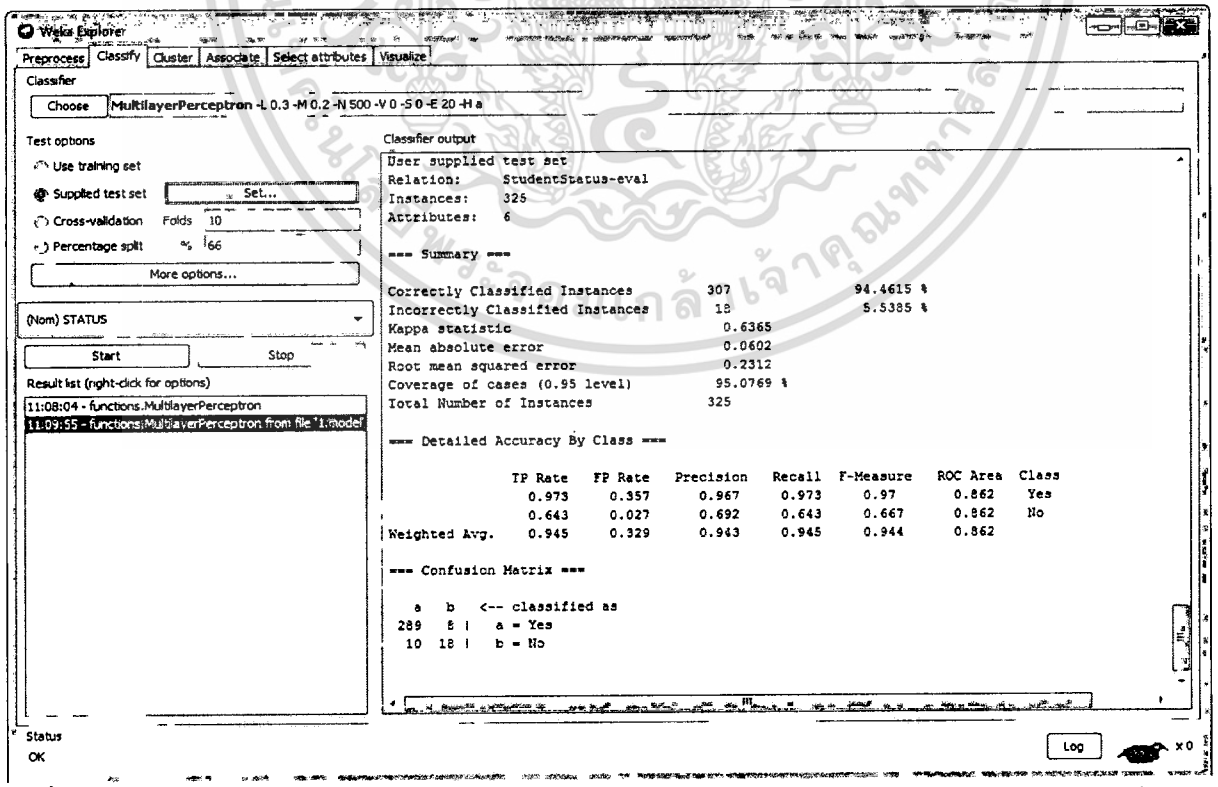
รูปที่ ข-8 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการทำนายตัวแบบโดยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. วิธีโครงข่ายประสาท (Neural Network)

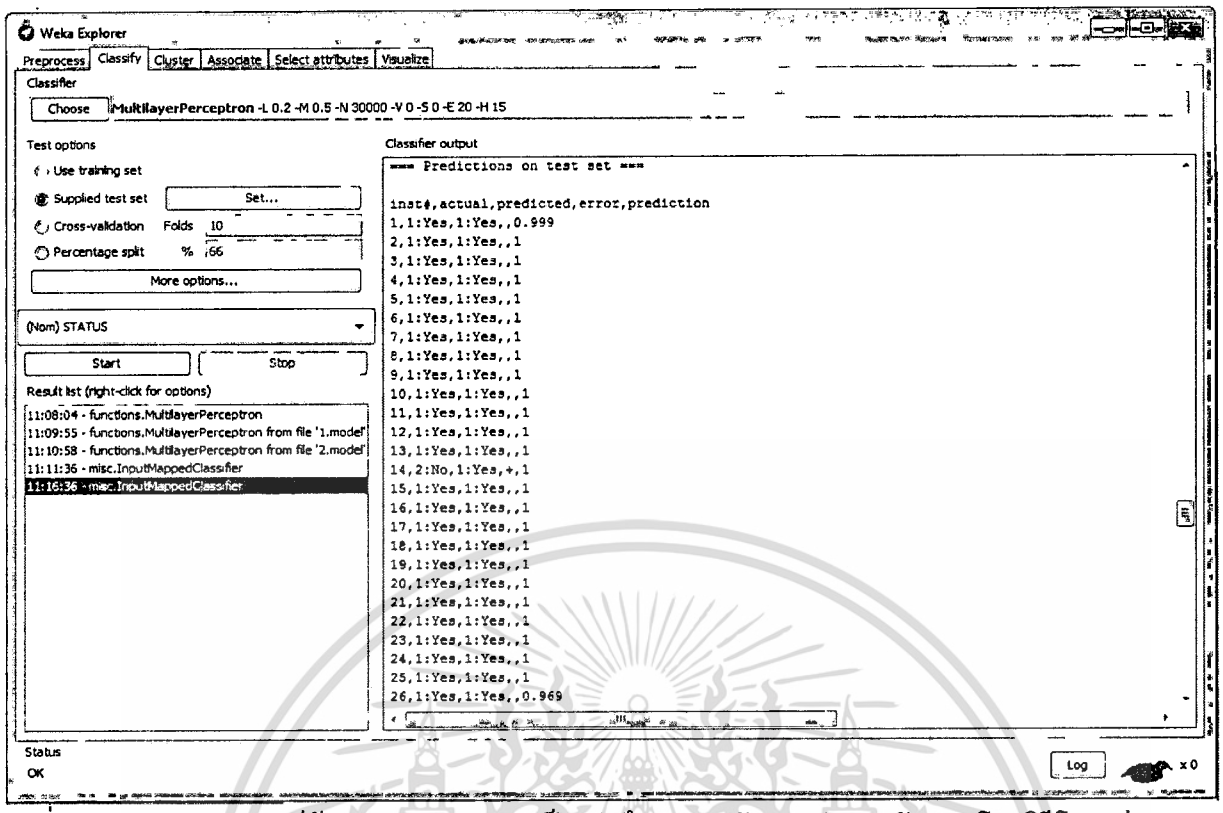


รูปที่ ข-9 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการสร้างตัวแบบโดยวิธีโครงข่ายประสาท

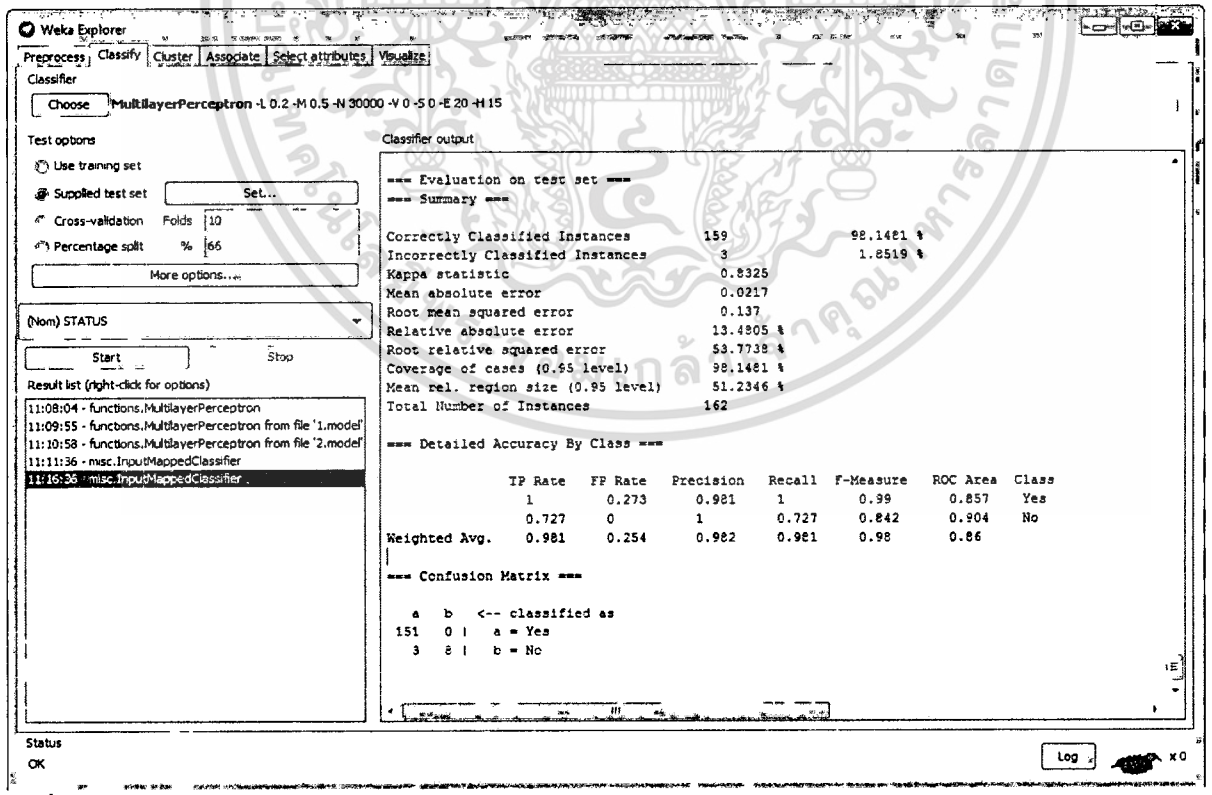


รูปที่ ข-10 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการทดสอบตัวแบบโดยวิธีโครงข่ายประสาท

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



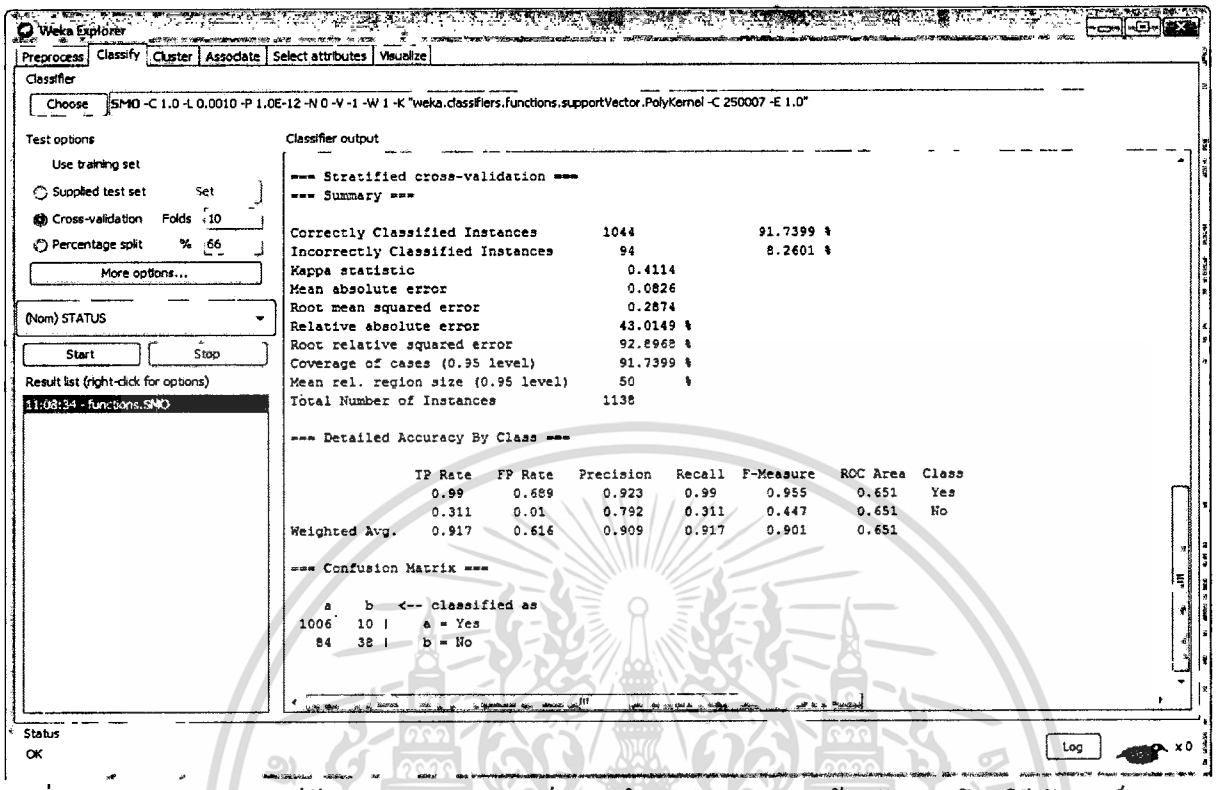
รูปที่ ข-11 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการทำนายตัวแบบโดยวิธีโครงข่ายประสาท



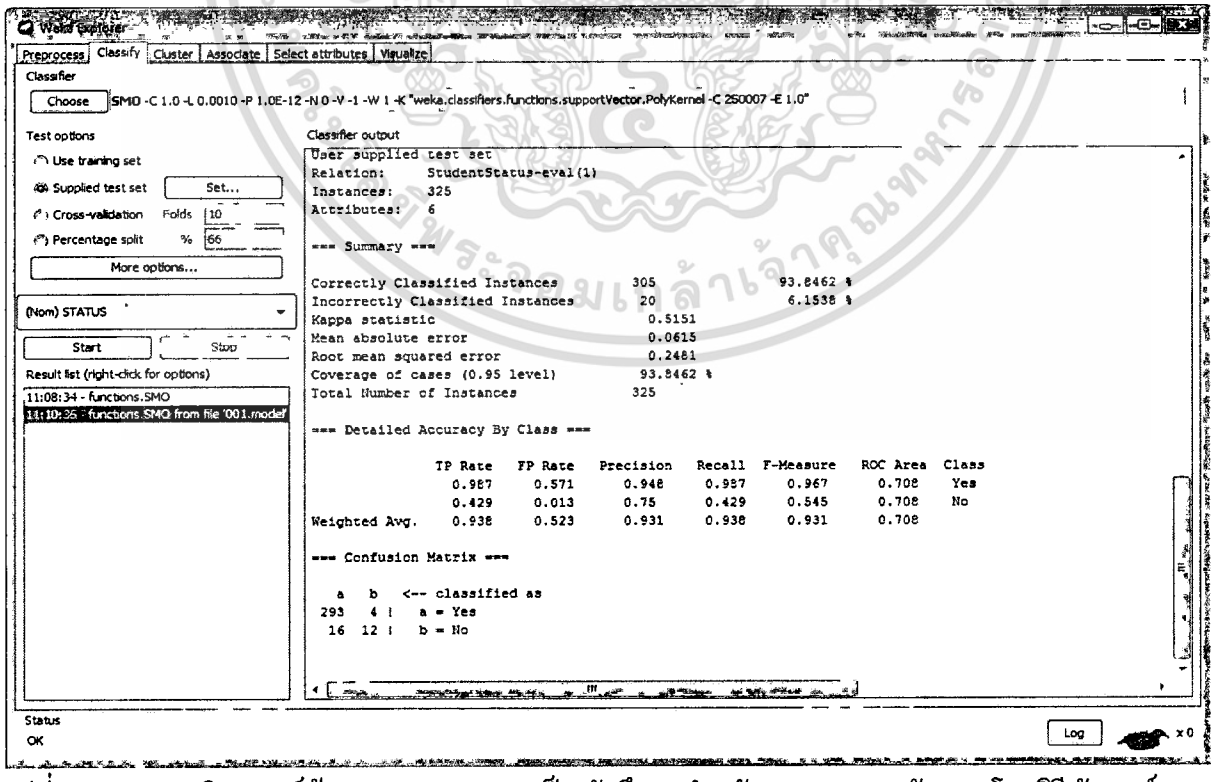
รูปที่ ข-12 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการทำนายตัวแบบโดยวิธีโครงข่ายประสาท

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. วิธีซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine)

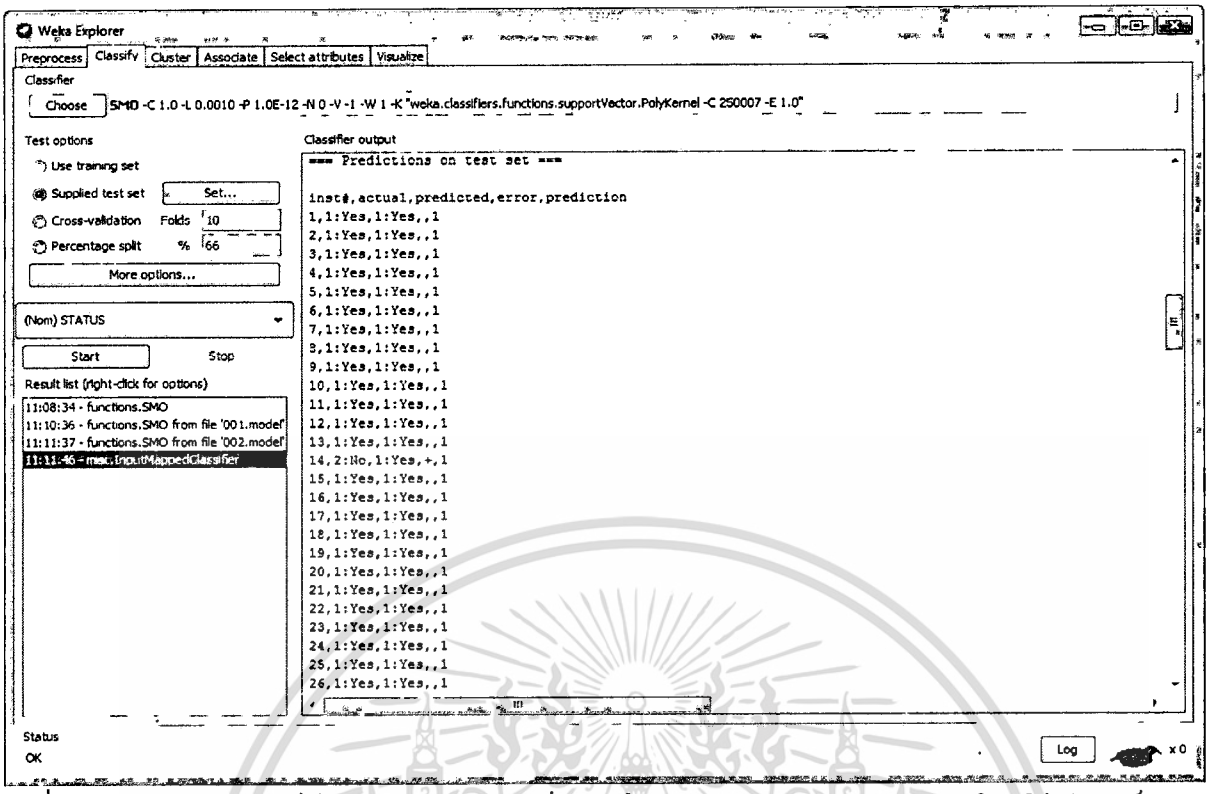


รูปที่ ข-13 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการสร้างตัวแบบโดยวิธีซัพพอร์ทเวกเตอร์แมชชีน

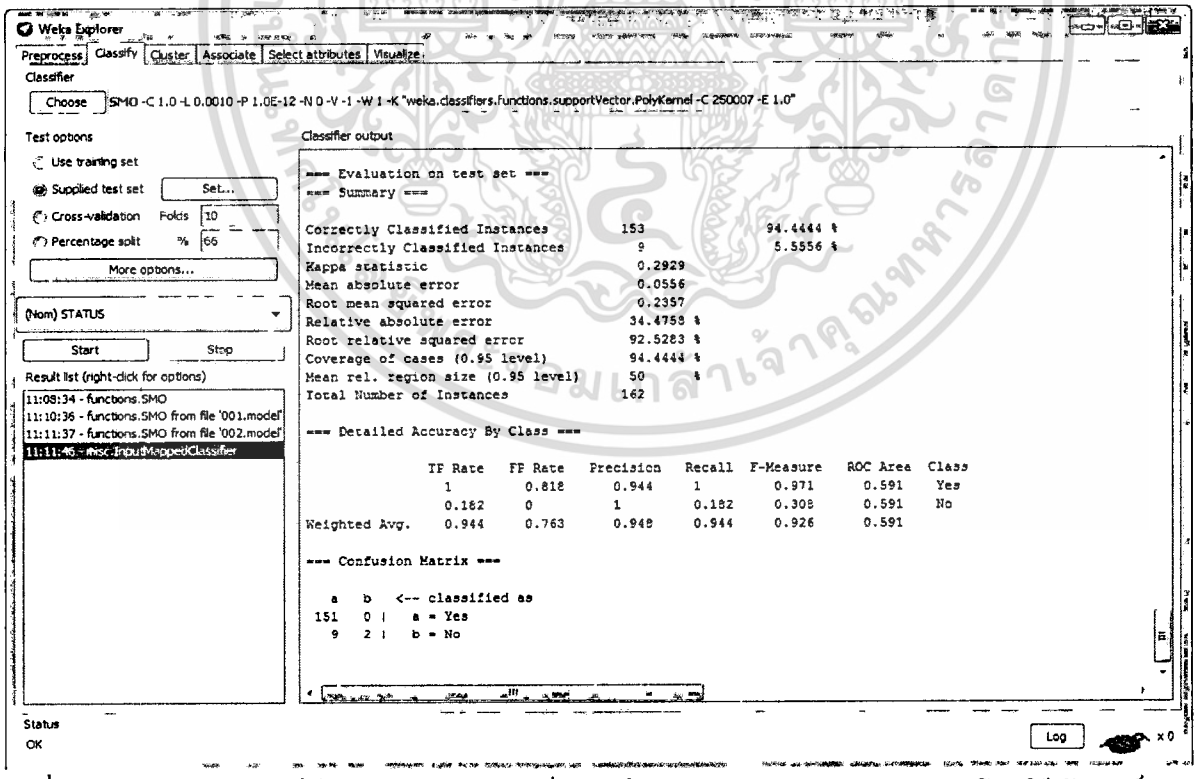


รูปที่ ข-14 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการทดสอบตัวแบบโดยวิธีซัพพอร์ทเวกเตอร์แมชชีน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



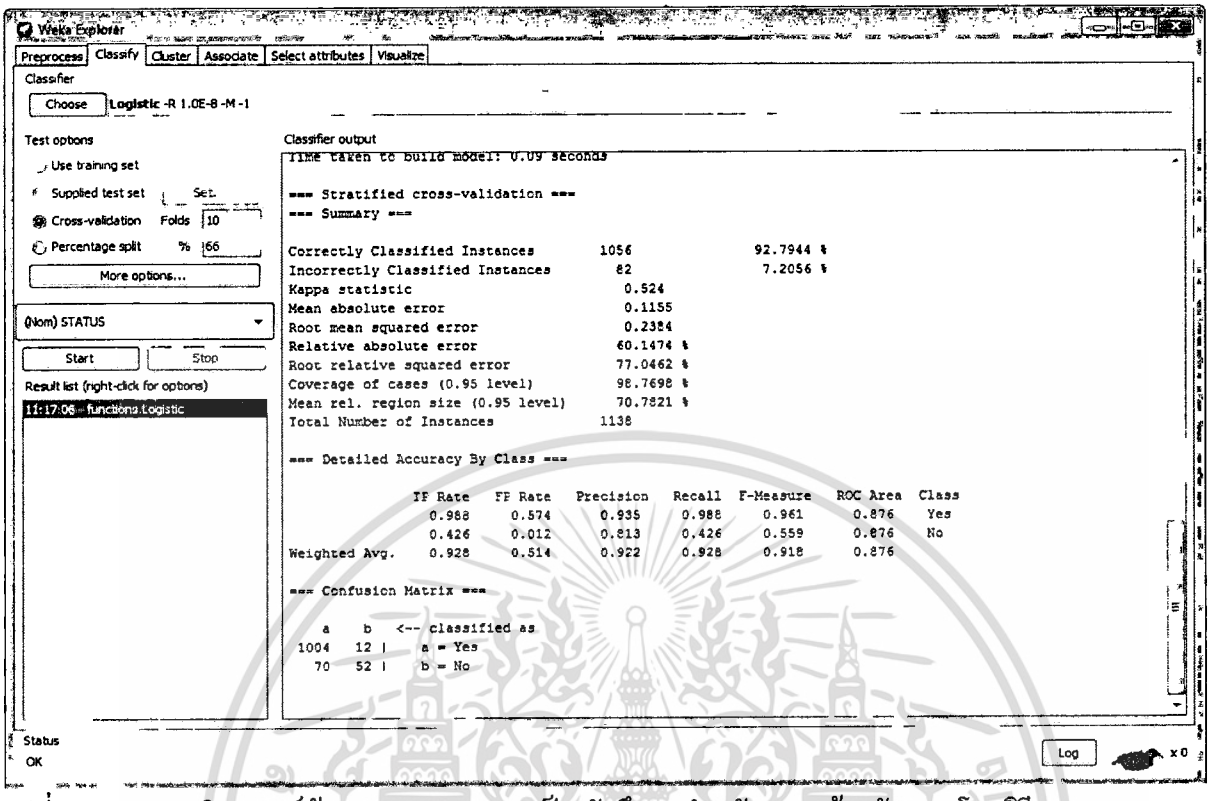
รูปที่ ข-15 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการทำนายตัวแบบโดยวิธีซัพพอร์ท เวกเตอร์แมชชีน



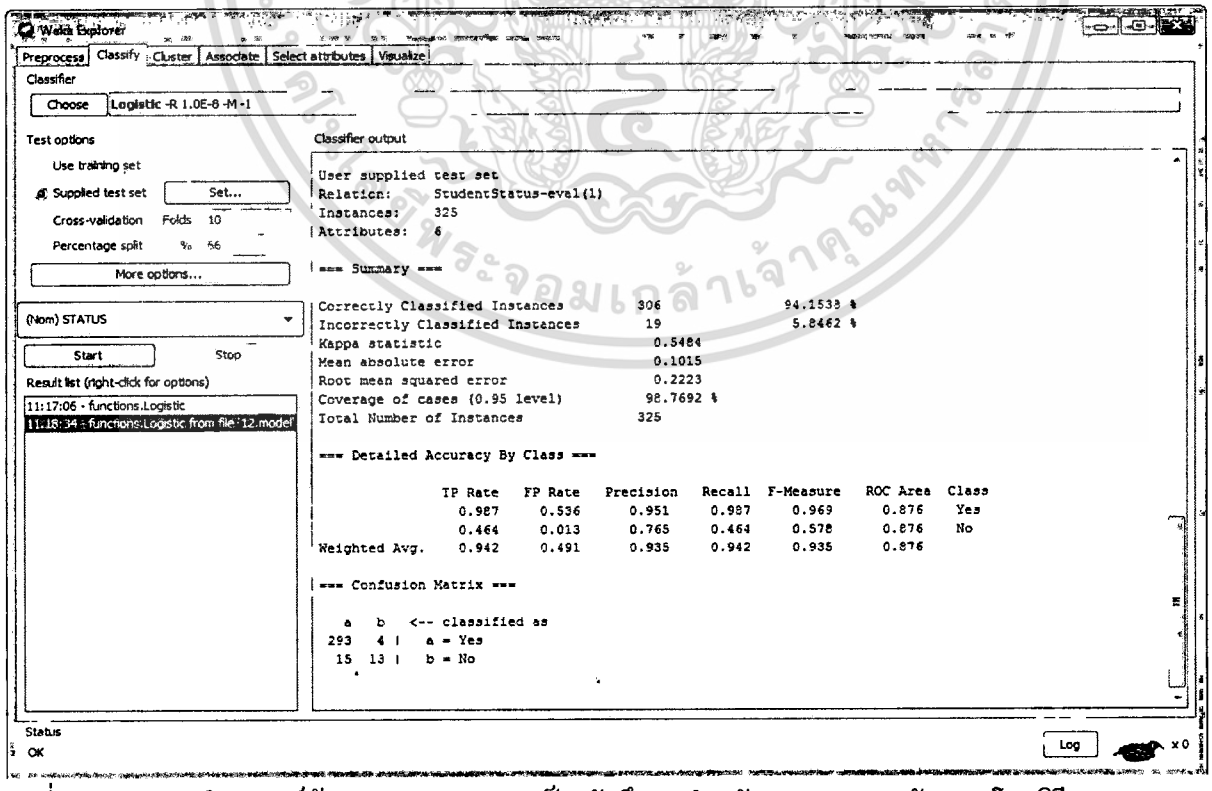
รูปที่ ข-16 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการทำนายตัวแบบโดยวิธีซัพพอร์ท เวกเตอร์แมชชีน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม (Binary Logistic Regression Method)

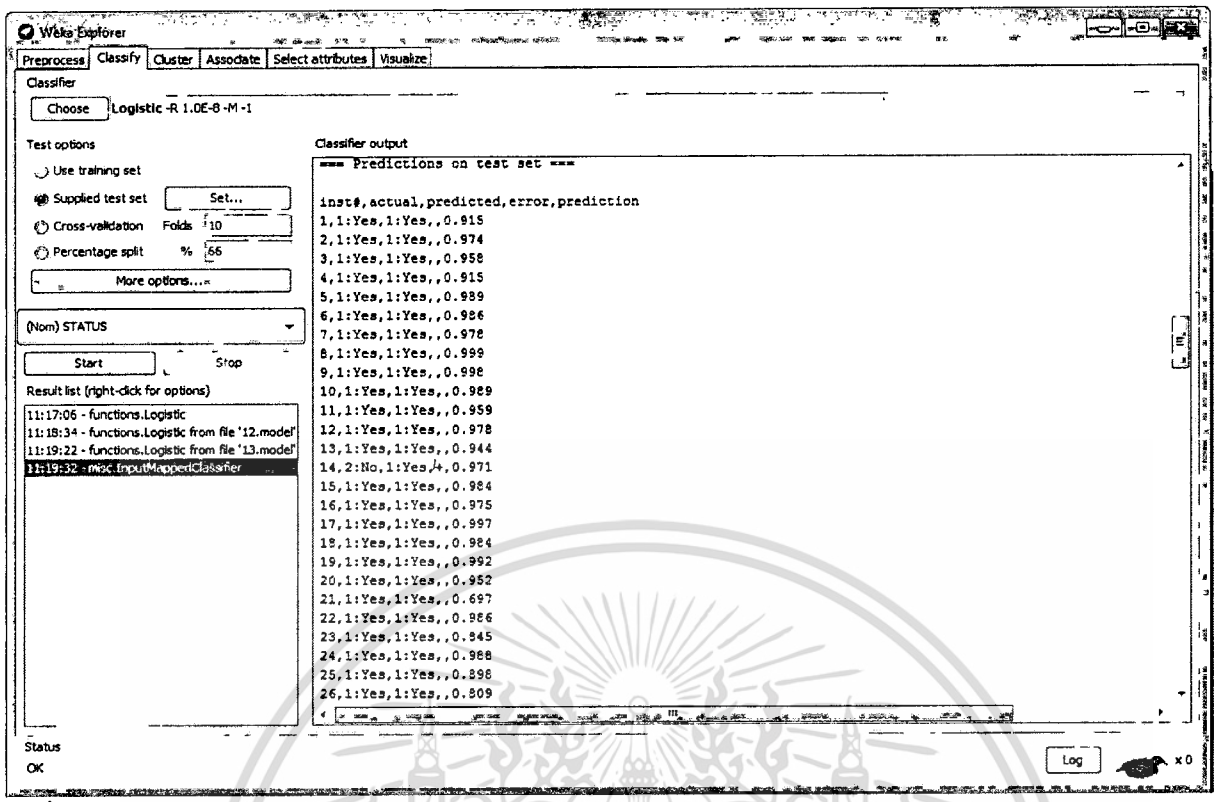


รูปที่ ข-17 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการสร้างตัวแบบโดยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

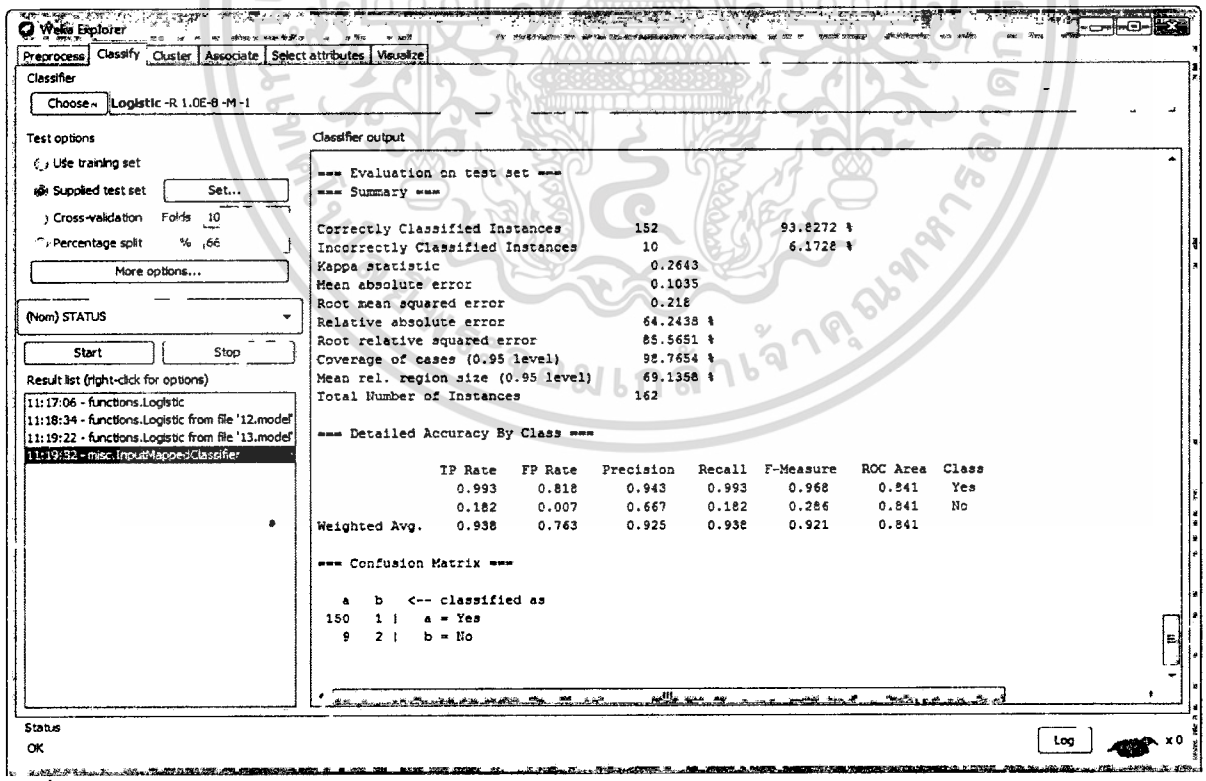


รูปที่ ข-18 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการทดสอบตัวแบบโดยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ข-19 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการทำนายตัวแบบโดยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม



รูปที่ ข-20 ผลการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการทำนายตัวแบบโดยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ค

ตัวอย่างการคำนวณ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

<u>ตัวอย่างที่ 1</u> การคำนวณค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) ค่าความถ่วงดุล (F-Measure) ของการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษา โดยวิธีความใกล้เคียงกันมากที่สุด

จากรูปที่ ข-4

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

$= \frac{149 + 5}{149 + 5 + 6 + 2}$

$= 0.9506$ หรือ 95.06%

ค่าความแม่นยำ (Precision) $= \frac{TP}{TP + FP}$

$= \frac{149}{149 + 6}$

$= 0.9613$ หรือ 96.13%

ค่าความระลึก (Recall) $= \frac{TP}{TP + FN}$

$= \frac{149}{149 + 2}$

$= 0.9868$ หรือ 98.68%

ค่าความถ่วงดุล (F-Measure) $= \frac{2 \times (Recall \times Precision)}{Recall + Precision}$

$= \frac{2 \times (0.9868 \times 0.9613)}{0.9868 + 0.9613}$

$= 0.9739$ หรือ 97.39%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

<u>ตัวอย่างที่ 2</u> การคำนวณค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของการวิเคราะห์ข้อมูลสถานภาพการเป็นนักศึกษาสำหรับการทำนายตัวแบบโดยวิธีความใกล้เคียงกันมากที่สุด

จากรูปที่ ข-3

กำหนดให้ค่า $y_i = 1$ ได้จากกรณีที่ค่าจริง (actual) ใน class attribute ของข้อมูลสถานภาพการเป็นนักศึกษาในระเบียนนั้นตรงกันกับค่าทำนาย (predicted)

เช่น ระเบียนที่ 1 ค่าจริง (actual) = 1:Yes ค่าทำนาย (predicted) = 1:Yes จะได้ $y_i = y_1 = 1$

ระเบียนที่ 81 ค่าจริง (actual) = 2:No ค่าทำนาย (predicted) = 2:No จะได้ $y_i = y_{81} = 1$

กำหนดให้ค่า $y_i = 0$ ได้จากกรณีที่ค่าจริง (actual) ใน class attribute ของข้อมูลสถานภาพการเป็นนักศึกษาในระเบียนนั้นไม่ตรงกันกับค่าทำนาย (predicted)

เช่น ระเบียนที่ 14 ค่าจริง (actual) = 2:No ค่าทำนาย (predicted) = 1:Yes จะได้ $y_i = y_{14} = 0$

ระเบียนที่ 46 ค่าจริง (actual) = 1:Yes ค่าทำนาย (predicted) = 2:No จะได้ $y_i = y_{46} = 0$

กำหนดให้ค่า $\hat{y}_i$ ได้จากค่าการทำนาย (predicted) ซึ่งอยู่ที่คอลัมน์ขวาสุดของระเบียนนั้นในช่อง Classifier output

ลำดับที่	y_i	$\hat{y}_i$	e_i^2	$\lvert e_i \rvert$
1	1	0.999	0.000001	0.001
2	1	0.999	0.000001	0.001
3	1	0.999	0.000001	0.001
4	1	0.999	0.000001	0.001
5	1	0.999	0.000001	0.001
6	1	0.999	0.000001	0.001
7	1	0.999	0.000001	0.001
8	1	0.999	0.000001	0.001
9	1	0.999	0.000001	0.001
10	1	0.999	0.000001	0.001
11	1	0.999	0.000001	0.001

ลำดับที่	y_i	$\hat{y}_i$	e_i^2	$\lvert e_i \rvert$
12	1	0.999	0.000001	0.001
13	1	0.999	0.000001	0.001
14	0	0.999	0.998001	0.999
15	1	0.999	0.000001	0.001
16	1	0.999	0.000001	0.001
17	1	0.999	0.000001	0.001
18	1	0.999	0.000001	0.001
19	1	0.999	0.000001	0.001
20	1	0.999	0.000001	0.001
21	1	0.999	0.000001	0.001
22	1	0.999	0.000001	0.001

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2	$\|e_i\|$
23	1	0.999	0.000001	0.001
24	1	0.999	0.000001	0.001
25	1	0.999	0.000001	0.001
26	1	0.999	0.000001	0.001
27	1	0.999	0.000001	0.001
28	1	0.999	0.000001	0.001
29	1	0.999	0.000001	0.001
30	1	0.999	0.000001	0.001
31	1	0.999	0.000001	0.001
32	1	0.999	0.000001	0.001
33	1	0.999	0.000001	0.001
34	1	0.999	0.000001	0.001
35	1	0.999	0.000001	0.001
36	1	0.999	0.000001	0.001
37	1	0.999	0.000001	0.001
38	1	0.999	0.000001	0.001
39	1	0.999	0.000001	0.001
40	1	0.999	0.000001	0.001
41	1	0.999	0.000001	0.001
42	1	0.999	0.000001	0.001
43	1	0.999	0.000001	0.001
44	1	0.999	0.000001	0.001
45	1	0.999	0.000001	0.001
46	0	0.999	0.998001	0.999
47	1	0.999	0.000001	0.001
48	1	0.999	0.000001	0.001
49	1	0.999	0.000001	0.001
50	1	0.999	0.000001	0.001
51	1	0.999	0.000001	0.001
52	1	0.999	0.000001	0.001

ลำดับที่	y_i	$\hat{y}_i$	e_i^2	$\|e_i\|$
53	1	0.999	0.000001	0.001
54	1	0.999	0.000001	0.001
55	1	0.999	0.000001	0.001
56	1	0.999	0.000001	0.001
57	1	0.999	0.000001	0.001
58	1	0.999	0.000001	0.001
59	1	0.999	0.000001	0.001
60	1	0.999	0.000001	0.001
61	1	0.999	0.000001	0.001
62	1	0.999	0.000001	0.001
63	1	0.999	0.000001	0.001
64	1	0.999	0.000001	0.001
65	1	0.999	0.000001	0.001
66	1	0.999	0.000001	0.001
67	1	0.999	0.000001	0.001
68	1	0.999	0.000001	0.001
69	1	0.999	0.000001	0.001
70	1	0.999	0.000001	0.001
71	1	0.999	0.000001	0.001
72	1	0.999	0.000001	0.001
73	1	0.999	0.000001	0.001
74	1	0.999	0.000001	0.001
75	1	0.999	0.000001	0.001
76	1	0.999	0.000001	0.001
77	1	0.999	0.000001	0.001
78	1	0.999	0.000001	0.001
79	1	0.999	0.000001	0.001
80	1	0.999	0.000001	0.001
81	1	0.999	0.000001	0.001
82	1	0.999	0.000001	0.001

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2	$\lvert e_i \rvert$
83	1	0.999	0.000001	0.001
84	1	0.999	0.000001	0.001
85	1	0.999	0.000001	0.001
86	1	0.999	0.000001	0.001
87	1	0.999	0.000001	0.001
88	0	0.999	0.998001	0.999
89	1	0.999	0.000001	0.001
90	1	0.999	0.000001	0.001
91	1	0.999	0.000001	0.001
92	1	0.999	0.000001	0.001
93	1	0.999	0.000001	0.001
94	1	0.999	0.000001	0.001
95	1	0.999	0.000001	0.001
96	1	0.999	0.000001	0.001
97	1	0.999	0.000001	0.001
98	1	0.999	0.000001	0.001
99	1	0.999	0.000001	0.001
100	1	0.999	0.000001	0.001
101	1	0.999	0.000001	0.001
102	1	0.999	0.000001	0.001
103	1	0.999	0.000001	0.001
104	1	0.999	0.000001	0.001
105	1	0.999	0.000001	0.001
106	1	0.999	0.000001	0.001
107	0	0.999	0.998001	0.999
108	1	0.999	0.000001	0.001
109	1	0.999	0.000001	0.001
110	1	0.999	0.000001	0.001
111	1	0.999	0.000001	0.001
112	1	0.999	0.000001	0.001

ลำดับที่	y_i	$\hat{y}_i$	e_i^2	$\lvert e_i \rvert$
113	1	0.999	0.000001	0.001
114	1	0.999	0.000001	0.001
115	1	0.999	0.000001	0.001
116	1	0.999	0.000001	0.001
117	1	0.999	0.000001	0.001
118	1	0.999	0.000001	0.001
119	1	0.999	0.000001	0.001
120	1	0.999	0.000001	0.001
121	1	0.999	0.000001	0.001
122	1	0.999	0.000001	0.001
123	1	0.999	0.000001	0.001
124	1	0.999	0.000001	0.001
125	1	0.999	0.000001	0.001
126	1	0.999	0.000001	0.001
127	1	0.999	0.000001	0.001
128	1	0.999	0.000001	0.001
129	1	0.999	0.000001	0.001
130	0	0.999	0.998001	0.999
131	1	0.999	0.000001	0.001
132	1	0.999	0.000001	0.001
133	1	0.999	0.000001	0.001
134	1	0.999	0.000001	0.001
135	1	0.999	0.000001	0.001
136	1	0.999	0.000001	0.001
137	1	0.999	0.000001	0.001
138	1	0.999	0.000001	0.001
139	1	0.999	0.000001	0.001
140	1	0.999	0.000001	0.001
141	1	0.999	0.000001	0.001
142	1	0.999	0.000001	0.001

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2	$\lvert e_i \rvert$
143	1	0.999	0.000001	0.001
144	0	0.999	0.998001	0.999
145	1	0.999	0.000001	0.001
146	1	0.999	0.000001	0.001
147	1	0.999	0.000001	0.001
148	1	0.999	0.000001	0.001
149	1	0.999	0.000001	0.001
150	0	0.999	0.998001	0.999
151	1	0.999	0.000001	0.001
152	1	0.999	0.000001	0.001

ลำดับที่	y_i	$\hat{y}_i$	e_i^2	$\lvert e_i \rvert$
153	1	0.999	0.000001	0.001
154	1	1	0	0
155	1	0.999	0.000001	0.001
156	1	0.999	0.000001	0.001
157	1	0.999	0.000001	0.001
158	1	0.999	0.000001	0.001
159	0	0.999	0.998001	0.999
160	1	0.999	0.000001	0.001
161	1	0.999	0.000001	0.001
162	1	0.999	0.000001	0.001
รวม			7.984161	8.145

ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) $= \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n}$

$= \frac{8.145}{162}$

$= 0.0502$

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

$= \frac{7.984161}{162}$

$= 0.04928$

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

$= \sqrt{0.04928}$

$= 0.222$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

1) ศุภมาส ไวกุลเพ็ชร และสมฤทัย รัศมีธรรม. 2545. การศึกษาปัจจัยที่มีผลต่อเกรดเฉลี่ยสะสมของนักศึกษาคณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

2) ณัฐณิชา และคณะ. 2550. การศึกษาความแตกต่างระหว่างเพศในด้านพฤติกรรมที่มีผลต่อผลสัมฤทธิ์ทางการศึกษาของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

3) อนงค์ ผลสุขการ และสายชล สินสมบูรณ์ทอง. 2553. ปัจจัยที่มีผลสัมฤทธิ์ทางด้านการเรียนของนักศึกษาชั้นปีที่ 1 คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. วารสารวิทยาศาสตร์ลาดกระบัง 19 : 2 (กรกฎาคม-ธันวาคม 2553) : 1-22.

4) สุจิตรา สุคนธมัต 2555. ปัจจัยที่มีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนของนักศึกษา คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. วารสารวิทยาศาสตร์ลาดกระบัง 21 : 2 (กรกฎาคม-ธันวาคม 2555) : 16-35.

5) Tan, P-N, Steinbach, M. and Kumar, V. 2006. Introduction to Data Mining Boston : Pearson Addison Wesley.

6) กัลยา วานิชย์บัญชา. 2552. การวิเคราะห์ข้อมูลหลายตัวแปร. บริษัทธรรมสาร จำกัด : กรุงเทพฯ

7) ชลนิศา สาระ 2550. การจำแนกกลุ่มสถานภาพการสำเร็จการศึกษาโดยแบบจำลองต้นไม้ตัดสินใจ. สารนิพนธ์ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

8) ภัทรพงศ์ พงศ์ภัทรกานต์. 2552. การเปรียบเทียบการจำแนกข้อมูลของแบบจำลอง CART, SVM, C5.0 และแบบผสมผสานกัน. ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเลย.

9) ภัทรพงศ์ พงศ์ภัทรกานต์. 2553. การวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาระดับปริญญาตรีโดยใช้คอมมิตตีแมชชีน. สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเลย.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

10) วลัยลักษณ์ สุขสมบุญ และสมชาย ปราการเจริญ. 2553. การเปรียบเทียบประสิทธิภาพการจำแนกประเภทปัญหาสำหรับระบบถามตอบโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน นาอีฟเบย์ และความใกล้เคียงกันมากที่สุด. ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

11) วาทินี นุ้ยเพียร และคณะ. 2553. การเปรียบเทียบประสิทธิภาพและวิเคราะห์การจำแนกข้อมูลโดยใช้โครงข่ายประสาทเทียม ซัพพอร์ตเวกเตอร์แมชชีน นาอีฟเบย์ และเคเนียรเร่สต์เนเบอร์. ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

12) เดช ธรรมศิริ และพยุง มีสัจ. 2554. การจำแนกข้อมูลด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีนโดยการปรับพารามิเตอร์และเลือกคุณลักษณะที่เหมาะสมด้วยขั้นตอนวิธีเชิงพันธุกรรม. ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

13) www.reg.kmitl.ac.th/index/index.php

14) คู่มือนักศึกษาระดับปริญญาตรี. 2556. สำนักทะเบียนและประมวลผล. สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

15) Kumar, V. and Wu, X. 2009. The Top Ten Algorithms in Data Mining. University of Minnesota Department of Computer Science and Engineering, Minneapolis, Minnesota : CRC Press.

16) รุจิรา ธรรมสมบัติ. 2554. ระบบสนับสนุนการตัดสินใจในการเลือกใช้แพคเกจอินเทอร์เน็ตมือถือโดยใช้ต้นไม้ตัดสินใจ. สาขาคอมพิวเตอร์ธุรกิจ คณะบริหารธุรกิจ วิทยาลัยราชพฤกษ์.

17) Berson, A. and Stephen, J. S. 1997. Data Warehousing, Data Mining, and OLAP.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้