สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

รายงานการวิจัย

การเปรียบเทียบประสิทธิภาพและทำนายผลการอนุมัติบัตรเครดิต ของธนาคารพาณิชย์ไทย

An Efficiency Comparison and Prediction in Credit Card Approval of Thai Commercial Banks

รศ.สายชล สินสมบูรณ์ทอง

b 12779611
i

สรทบ.....
สขทะเบียน 142668
วัน.เดือน.ปี 23 พ.ค. 2559

ได้รับทุนสนับสนุนงานวิจัยจากเงินรายได้ ประจำปีงบประมาณ 2558
คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

งานวิจัยเล่มนี้สำเร็จลุล่วงไปได้ด้วยดี เนื่องจากผู้จัดทำได้รับความช่วยเหลือจากบุคคลผู้มีพระคุณหลายท่าน ดังนี้

ขอขอบพระคุณโครงการวิจัยที่เอื้อเฟื้อทุนสนับสนุนในการวิจัยครั้งนี้ โดยใช้เงินรายได้คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ขอขอบคุณนางสาวกวิสรา ทองธนบดีกุล นักศึกษาปริญญาตรี ชั้นปีที่ 4 สาขาวิชาสถิติประยุกต์ ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังที่คอยให้ความช่วยเหลือด้านการวิเคราะห์ข้อมูลบางส่วน

ขอขอบคุณทุกท่านที่มิได้เอ่ยนามในที่นี้ที่ให้คำแนะนำที่เป็นประโยชน์ต่าง ๆ และคอยเป็นกำลังใจให้งานวิจัยเล่มนี้สำเร็จลุล่วงไปได้ด้วยดี

.....

รศ.สายชล สินสมบูรณ์ทอง

(หัวหน้าโครงการวิจัย)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อ	การเปรียบเทียบประสิทธิภาพและทำนายผลการอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทย
ผู้วิจัย	รศ.สายชล สินสมบูรณ์ทอง
สาขา	สถิติประยุกต์
พ.ศ.	2558

บทคัดย่อ

ในการศึกษาเรื่องการเปรียบเทียบประสิทธิภาพและทำนายผลการอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทย มีวัตถุประสงค์เพื่อศึกษาทำความเข้าใจและเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่ม รวมทั้งเปรียบเทียบการทำนายผลการอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทย 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) ในการเปรียบเทียบ โดยพิจารณาจากค่า MAE และ MSE ที่มีค่าน้อยที่สุด

ผลการศึกษาพบว่าวิธีโครงข่ายประสาทมีค่าความถูกต้อง ค่าอัตราความถูกต้องเชิงบวก ค่าอัตราความผิดพลาดเชิงลบ ค่าความระลึกและค่าความถ่วงดุลดีที่สุดคือ 70.37%, 0.821, 0.179, 0.821 และ 0.742 ตามลำดับ ส่วนวิธีแผนภาพต้นไม้เพื่อการตัดสินใจและวิธีความใกล้เคียงกันมากที่สุดมีค่าอัตราความถูกต้องเชิงลบและค่าอัตราความผิดพลาดเชิงบวกดีที่สุดคือ 0.731 และ 0.269 ตามลำดับ และวิธีแผนภาพต้นไม้เพื่อการตัดสินใจมีค่าความแม่นยำดีที่สุดคือ 0.682 ดังนั้นวิธีโครงข่ายประสาทมีประสิทธิภาพดีที่สุด ส่วนการเปรียบเทียบการทำนายผลของวิธีการจำแนกกลุ่มในการพิจารณาอนุมัติบัตรเครดิตพบว่าวิธีโครงข่ายประสาทมีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุดคือ 0.3683 และ 0.1925 ตามลำดับ ดังนั้นวิธีโครงข่ายประสาทมีการทำนายผลดีที่สุด

คำสำคัญ : วิธีการจำแนกกลุ่ม วิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาท วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	An Efficiency Comparison and Prediction in Credit Card Approval of Thai Commercial Banks
Researcher	Assoc.Prof. Saichon Sinsomboonthong
Programme	Applied Statistics
Year	2015

Abstract

In this study, an efficiency comparison and prediction in credit card approval of Thai commercial banks were investigated. The purposes are understand and compare efficiency the classification method and also compare prediction in credit card approval of Thai commercial banks. 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, false negative rate, recall, and F-Measure, 70.37%, 0.821, 0.179, 0.821 and 0.742 respectively. Nevertheless, decision tree and k-nearest neighbor method have the best of true negative rate and false positive rate, 0.731, 0.269 respectively and decision tree method has the best of precision, 0.682. 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, 0.3683 and 0.1925 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

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VII

สารบัญรูป XII

บทที่ 1 บทนำ 1

1.1 ความสำคัญและที่มาของปัญหา 1

1.2 วัตถุประสงค์ของงานวิจัย 3

1.3 ขอบเขตของงานวิจัย 3

1.4 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย 4

1.5 นิยามคำศัพท์ 4

บทที่ 2 ทฤษฎีและรายงานวิจัยที่เกี่ยวข้อง 6

2.1 ทฤษฎีที่เกี่ยวข้อง 6

2.1.1 วิธีความใกล้เคียงกันมากที่สุด 6

2.1.2 วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 7

2.1.3 วิธีโครงข่ายประสาท 10

2.1.4 วิธีซัพพอร์ตเวกเตอร์แมชชีน 15

2.1.5 วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 19

2.2 รายงานวิจัยที่เกี่ยวข้อง 20

บทที่ 3 วิธีดำเนินงานวิจัย 22

3.1 อุปกรณ์ในการวิจัย 22

3.1.1 อุปกรณ์ที่มีอยู่แล้ว 22

3.1.2 อุปกรณ์ที่ต้องการเพิ่มเติม 22

3.2 ขั้นตอนการดำเนินงานวิจัย 23

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

บทที่ 4 ผลการวิจัยและอภิปรายผล **28**

4.1 การวิเคราะห์ข้อมูล 28

4.1.1 วิธีความใกล้เคียงกันมากที่สุด 28

4.1.1.1 การสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต 28

4.1.1.2 การทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต 30

4.1.1.3 การทำนายการพิจารณาอนุมัติบัตรเครดิต 32

4.1.2 วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 35

4.1.2.1 การสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต 35

4.1.2.2 การทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต 37

4.1.2.3 การทำนายการพิจารณาอนุมัติบัตรเครดิต 39

4.1.3 วิธีโครงข่ายประสาท 42

4.1.3.1 การสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต 42

4.1.3.2 การทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต 44

4.1.3.3 การทำนายการพิจารณาอนุมัติบัตรเครดิต 46

4.1.4 วิธีซัพพอร์ตเวกเตอร์แมชชีน 54

4.1.4.1 การสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต 54

4.1.4.2 การทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต 56

4.1.4.3 การทำนายการพิจารณาอนุมัติบัตรเครดิต 58

4.1.5 วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 61

4.1.5.1 การสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต 61

4.1.5.2 การทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต 63

4.1.5.3 การทำนายการพิจารณาอนุมัติบัตรเครดิต 65

4.2 ผลการเปรียบเทียบประสิทธิภาพและทำนายผลของวิธีการจำแนกกลุ่ม 68

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ **70**

5.1 สรุปผลการวิจัย 70

5.2 ข้อเสนอแนะ 71

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

		หน้า
ภาคผนวก ก	รายละเอียดและตัวอย่างข้อมูลที่นำมาวิเคราะห์	72
ภาคผนวก ข	การวิเคราะห์ข้อมูล	80
ภาคผนวก ค	ตัวอย่างการคำนวณ	91
เอกสารอ้างอิง		96

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

หน้า

ตารางที่ 4.1 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด 28
ตารางที่ 4.2 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด 29
ตารางที่ 4.3 ผลในส่วนของเมตริกซ์ความสับสนจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด 29
ตารางที่ 4.4 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด 30
ตารางที่ 4.5 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด 31
ตารางที่ 4.6 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด 31
ตารางที่ 4.7 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีความใกล้เคียงกันมากที่สุด 32
ตารางที่ 4.8 ผลในส่วนของการสรุปผลจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด 33
ตารางที่ 4.9 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด 34
ตารางที่ 4.10 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด 34
ตารางที่ 4.11 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 35
ตารางที่ 4.12 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 36
ตารางที่ 4.13 ผลในส่วนของเมตริกซ์ความสับสนจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 36
ตารางที่ 4.14 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 37

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

หน้า

ตารางที่ 4.15 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 37
ตารางที่ 4.16 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 38
ตารางที่ 4.17 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 39
ตารางที่ 4.18 ผลในส่วนของการสรุปผลจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 40
ตารางที่ 4.19 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 41
ตารางที่ 4.20 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ 41
ตารางที่ 4.21 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท 42
ตารางที่ 4.22 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท 43
ตารางที่ 4.23 ผลในส่วนของเมตริกซ์ความสับสนจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท 43
ตารางที่ 4.24 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท 44
ตารางที่ 4.25 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท 45
ตารางที่ 4.26 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท 45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

หน้า

ตารางที่ 4.27 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท สำหรับอัตราการเรียนรู้ $\eta = 0.1$ และโมเมนตัม $\alpha = 0.5, 0.6, 0.7, 0.8$ และ 0.9 46

ตารางที่ 4.28 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท สำหรับอัตราการเรียนรู้ $\eta = 0.2$ และโมเมนตัม $\alpha = 0.5, 0.6, 0.7, 0.8$ และ 0.9 47

ตารางที่ 4.29 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท สำหรับอัตราการเรียนรู้ $\eta = 0.3$ และโมเมนตัม $\alpha = 0.5, 0.6, 0.7, 0.8$ และ 0.9 48

ตารางที่ 4.30 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท สำหรับอัตราการเรียนรู้ $\eta = 0.4$ และโมเมนตัม $\alpha = 0.5, 0.6, 0.7, 0.8$ และ 0.9 49

ตารางที่ 4.31 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท สำหรับอัตราการเรียนรู้ $\eta = 0.5$ และโมเมนตัม $\alpha = 0.5, 0.6, 0.7, 0.8$ และ 0.9 50

ตารางที่ 4.32 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีโครงข่ายประสาท 51

ตารางที่ 4.33 ผลในส่วนของการสรุปผลจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท 52

ตารางที่ 4.34 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท 53

ตารางที่ 4.35 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท 53

ตารางที่ 4.36 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 54

ตารางที่ 4.37 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 55

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

หน้า

ตารางที่ 4.38 ผลในส่วนของเมตริกซ์ความสับสนจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 55

ตารางที่ 4.39 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 56

ตารางที่ 4.40 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 56

ตารางที่ 4.41 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 57

ตารางที่ 4.42 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 58

ตารางที่ 4.43 ผลในส่วนของการสรุปผลจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 59

ตารางที่ 4.44 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 60

ตารางที่ 4.45 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน 60

ตารางที่ 4.46 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 61

ตารางที่ 4.47 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 62

ตารางที่ 4.48 ผลในส่วนของเมตริกซ์ความสับสนจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 62

ตารางที่ 4.49 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 63

ตารางที่ 4.50 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 63

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

หน้า

ตารางที่ 4.51 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 64

ตารางที่ 4.52 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 65

ตารางที่ 4.53 ผลในส่วนของการสรุปผลจากการทำนายการพิจารณาอนุมัติบัตรเครดิตด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 66

ตารางที่ 4.54 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 67

ตารางที่ 4.55 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม 67

ตารางที่ 4.56 ผลการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มในการพิจารณาอนุมัติบัตรเครดิตทั้ง 5 วิธี 68

ตารางที่ 4.57 ผลการเปรียบเทียบการทำนายผลของวิธีการจำแนกกลุ่มในการพิจารณาอนุมัติบัตรเครดิตทั้ง 5 วิธี 69

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

หน้า

รูปที่ 2.1 ตัวอย่างของวิธีความใกล้เคียงกันมากที่สุด 7
รูปที่ 2.2 ส่วนประกอบของแผนภาพต้นไม้เพื่อการตัดสินใจ 8
รูปที่ 2.3 โครงข่ายของเซลล์ประสาท 11
รูปที่ 2.4 ลักษณะโครงข่ายประสาทเทียมแบบส่งสัญญาณไปข้างหน้า 13
รูปที่ 2.5 ลักษณะโครงข่ายประสาทเทียมแบบมีการย้อนกลับ 13
รูปที่ 2.6 ลักษณะโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น 14
รูปที่ 2.7 การขยายตัวของเส้นขอบ 16
รูปที่ 2.8 เส้นขอบและเส้นแบ่งเมื่อแทนด้วยสมการเส้นตรง 17
รูปที่ 2.9 รูปแบบการวางตัวที่ไม่สามารถแบ่งด้วยเส้นตรงได้ 18

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มาของปัญหา

จากแนวโน้มของการเจริญเติบโตของบัตรเครดิตในปัจจุบันทำให้บัตรเครดิตมีการเจริญเติบโตอย่างต่อเนื่อง เนื่องจากบัตรเครดิตมีบทบาทสำคัญเป็นที่ยอมรับในสังคมปัจจุบันและยังได้รับความนิยมเพิ่มมากขึ้น ถือได้ว่าบัตรเครดิตมีส่วนช่วยในการส่งเสริมพัฒนาสังคมในยุคปัจจุบันเป็นอย่างมาก สังเกตได้จากในสังคมปัจจุบันบัตรเครดิตได้เข้ามามีบทบาทในการดำรงชีวิตของประชาชนและเป็นที่ยอมรับกันอย่างกว้างขวาง โดยพิจารณาได้จากจำนวนผู้ถือบัตรเครดิตมีปริมาณที่เพิ่มมากขึ้นและปริมาณการใช้จ่ายผ่านบัตรเครดิตที่เพิ่มขึ้นอย่างรวดเร็ว (มัณฑิตา สกุลรัตนศักดิ์, 2556)

ในการศึกษาเพื่อเปรียบเทียบประสิทธิภาพและผลการทำนายของวิธีการจำแนกกลุ่มข้อมูลทางด้านการทำเหมืองข้อมูล ได้มีผู้ศึกษาจำนวนมาก เช่น กัมพล กลมรัตน์ธาดา (2553) กล่าวถึงวิกฤตการณ์ทางเศรษฐกิจของประเทศสหรัฐอเมริกาในปี พ.ศ. 2551 ก่อให้เกิดความไม่มั่นคงต่อทุกภาคธุรกิจ ซึ่งธุรกิจธนาคารเป็นหนึ่งในธุรกิจที่ได้รับผลกระทบอย่างมากตามสภาวะถดถอยทางเศรษฐกิจในครั้งนี้ งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาถึงอัตราส่วนทางการเงินต่าง ๆ ที่ส่งผลต่อปัญหาความไม่มั่นคงของธุรกิจธนาคาร เพื่อให้ภาครัฐและเอกชนสามารถใช้อัตราส่วนทางการเงินดังกล่าวเพื่อเป็นเครื่องมือในการตรวจสอบความเข้มแข็งทางการเงินและการดำเนินงานของธุรกิจธนาคาร งานวิจัยนี้จึงศึกษาถึงปัจจัยที่มีความสำคัญในการชี้วัดและเป็นสัญญาณเตือนภัยล่วงหน้าถึงความไม่มั่นคงทางการเงินของธนาคาร และได้ทำการเปรียบเทียบความแม่นยำและความถูกต้องของตัวแบบสัญญาณเตือนภัยล่วงหน้าในการทำนายความไม่มั่นคงทางการเงินของธนาคารโดยใช้วิธีการถดถอยโลจิสติกและโครงข่ายประสาท โดยทดสอบอัตราส่วนทางการเงิน 21 อัตราส่วนเพื่อการจำแนกกลุ่มความไม่มั่นคงของธนาคาร โดยผลลัพธ์ชี้ให้เห็นว่าระบบสัญญาณเตือนภัยล่วงหน้าทางการเงินที่ได้จากวิธีการทั้งสองดังกล่าวข้างต้นมีความสามารถสูงในการจำแนกกลุ่มโดยสามารถจำแนกกลุ่มของธนาคารที่มีความมั่นคงและกลุ่มของธนาคาร ที่มีความไม่มั่นคงได้เป็นอย่างดี และผลจากการเปรียบเทียบระหว่างวิธีการถดถอยโลจิสติกและโครงข่ายประสาทพบว่าวิธีโครงข่ายประสาทสามารถทำนายข้อมูลและจำแนกกลุ่มได้แม่นยำกว่าวิธีการถดถอยโลจิสติก ดังนั้นงานการวิจัยนี้จึงสามารถนำไปประยุกต์ใช้เพื่อเป็นแนวทางหนึ่งในการที่จะเป็นตัวแบบเพื่อเป็นระบบสัญญาณเตือนภัยล่วงหน้าทางการเงินในการบ่งชี้ถึงสถานะความไม่มั่นคงของธุรกิจธนาคารต่อไปได้ วาทินี นุ้ยเพียร และคณะ (2553) ได้เลือกใช้อัลกอริทึมโครงข่ายประสาทแบบมัลติเลเยอร์เปอร์เซ็ปตรอน ซัพพอร์ตเวกเตอร์แมชชีน นาอีฟเบย์ และความใกล้เคียงกันมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ที่สุด เพื่อประเมินประสิทธิภาพค่าความถูกต้อง (accuracy) ค่าความแม่นยำ (precision) ค่าความระลึก (recall) และค่าความถ่วงดุล (F-measure) ใช้ข้อมูลจาก UCI ประกอบด้วย Ozone Days และ Adult เลือกกลุ่มข้อมูลโดยมีจำนวนคำตอบ (class) เท่ากันในข้อมูลแต่ละชุดเป็นการทดลองแบบมีการเรียนรู้ จากผลการวิจัยอัลกอริทึมที่ดีที่สุดของข้อมูล Ozone Days คือซัพพอร์ตเวกเตอร์แมชชีน 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 วลัยลักษณ์ สุขสมบุญ และสมชาย ปราการเจริญ (2553) ได้นำเสนอการเปรียบเทียบวิธีการจำแนกกลุ่มของปัญหาสำหรับ Helpdesk ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน วิธีนาอีฟเบย์ และวิธีความใกล้เคียงกันมากที่สุด โดยนำข้อมูลมาตัดคำภาษาอังกฤษด้วยวิธีการอ้างอิงพจนานุกรมและคลังคำศัพท์เฉพาะที่ผู้วิจัยสร้างขึ้นใหม่ จากนั้นนำข้อมูลปัญหา Helpdesk แปลงให้อยู่ในรูปของเวกเตอร์ทางคณิตศาสตร์และทำการทดลองด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน วิธีนาอีฟเบย์และวิธีความใกล้เคียงกันมากที่สุดโดยใช้การตรวจสอบไขว้กับ K-Fold = 10 โดยทดลองปรับค่าพารามิเตอร์ของฟังก์ชันแกน Linear, Polynomial และ Radius Basic Function ในวิธีซัพพอร์ตเวกเตอร์แมชชีน วิธีนาอีฟเบย์และวิธีความใกล้เคียงกันมากที่สุด ผลการทดลองสามารถสรุปได้ว่าการทดลองเปรียบเทียบวิธีการจำแนกกลุ่มของปัญหาสำหรับ Helpdesk แสดงให้เห็นว่าเทคนิคที่นำเสนอด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความถูกต้องที่ดีที่สุดโดยใช้ฟังก์ชันแกน Polynomial

นอกจากนี้ยังมีงานวิจัยของณัฐวุฒิ ศิริกุลรุ่งโรจน์ และคณะ (2556) ได้ทำการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มและผลการทำนายของวิธีการจำแนกกลุ่ม โดยเลือกใช้วิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาท และวิธีซัพพอร์ตเวกเตอร์แมชชีน เพื่อประเมินประสิทธิภาพการจำแนกกลุ่ม และประเมินผลการทำนาย โดยใช้ข้อมูลการตลาดของธนาคาร ข้อมูลการสำรวจรายได้ ข้อมูลการมีชีวิตอยู่รอดของผู้ป่วยที่ได้เข้ารับการผ่าตัดมะเร็งเต้านม ข้อมูลการซื้อรถจักรยาน และข้อมูลโรคเบาหวานของชนพื้นเมืองชาวอเมริกัน โดยแบ่งข้อมูลเป็นชุดสร้างตัวแบบ ชุดทดสอบความถูกต้องของตัวแบบ และชุดทำนายตัวแบบ ในอัตราส่วน 70, 20 และ 10 ตามลำดับ จากการเปรียบเทียบข้อมูลทั้ง 5 ชุด วิธีการจำแนกกลุ่มที่มีประสิทธิภาพการจำแนกกลุ่มดีที่สุดสำหรับข้อมูลการตลาดของธนาคาร ข้อมูลการซื้อรถจักรยาน และข้อมูลโรคเบาหวานของชนพื้นเมืองชาว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อเมริกัน โดยเปรียบเทียบจากค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก ค่าความถ่วงดุล คือ วิธีโครงข่ายประสาท ส่วนวิธีการจำแนกกลุ่มที่มีผลการทำนายดีที่สุดสำหรับข้อมูลการตลาดของธนาคาร ข้อมูลการมีชีวิตอยู่รอดของผู้ป่วยที่ได้เข้ารับการผ่าตัดมะเร็งเต้านม ข้อมูลการซื้อจักรยาน และข้อมูลโรคเบาหวานของชนพื้นเมืองชาวอเมริกัน โดยเปรียบเทียบจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ยคือ วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

ดังนั้นผู้วิจัยจึงได้นำวิธีการทำเหมืองข้อมูลมาประยุกต์ใช้ในการเปรียบเทียบประสิทธิภาพการอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทยและเปรียบเทียบการทำนายผลการอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทยโดยวิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาท วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม เพื่อหาวิธีที่มีประสิทธิภาพดีที่สุดและการทำนายผลดีที่สุดในการอนุมัติบัตรเครดิตต่อไป

1.2 วัตถุประสงค์ของงานวิจัย

1.2.1 เพื่อเปรียบเทียบประสิทธิภาพการอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทยระหว่างวิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาท วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

1.2.2 เพื่อเปรียบเทียบการทำนายผลการอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทยระหว่างวิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาท วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

1.3 ขอบเขตของงานวิจัย

โครงการวิจัยนี้ได้ทำการศึกษาเรื่องการเปรียบเทียบประสิทธิภาพและทำนายผลการอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทย 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 20,000 รอบ และชั้นซ่อน 1 ชั้น วิธีซัพพอร์ตเวกเตอร์แมชชีน ใช้อัลกอริทึม SMO ชนิดโพลิโนเมียลเคอร์เนล และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม ในการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มทั้ง 5 วิธี จะใช้ค่าความถูกต้อง (Accuracy) ค่าอัตราความถูกต้องเชิงบวก (TP Rate) ค่าอัตราความถูกต้องเชิงลบ (TN Rate) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-Measure) โดยจะพิจารณาจากค่าเหล่านี้ที่มีค่ามากที่สุด ส่วนค่าอัตราความผิดพลาดเชิงบวก (FP Rate) และค่าอัตราความผิดพลาดเชิงลบ (FN Rate) จะพิจารณาจากค่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

น้อยที่สุด ในการทำนายผลของวิธีการจำแนกกลุ่มทั้ง 5 วิธี จะใช้ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error : MAE) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) ในการเปรียบเทียบ โดยพิจาณาจากค่า MAE และ MSE ที่มีค่าน้อยที่สุด ทำการเก็บรวบรวมข้อมูลลูกค้าธนาคารพาณิชย์ไทยจากแบบฟอร์มใบสมัครบัตรเครดิตจำนวน 519 ชุด โดยแบ่งเป็นข้อมูลชุดฝึกหัดจำนวน 365 ชุด ข้อมูลชุดทดสอบประสิทธิภาพจำนวน 100 ชุด และข้อมูลชุดทดสอบจำนวน 54 ชุด เกี่ยวกับเพศ อายุ ระดับการศึกษา สถานภาพสมรส จำนวนบุตร อาชีพ รายได้ต่อเดือน ประเภทธุรกิจในการประกอบการ ตำแหน่งหน้าที่ในการทำงาน อายุในการทำงาน ประเภทที่อยู่อาศัย สถานะการอยู่อาศัย การผ่อน/เช่าที่อยู่อาศัยต่อเดือน และการพิจารณาอนุมัติบัตรเครดิต ซึ่งตัวแปรอิสระประกอบด้วย เพศ อายุ ระดับการศึกษา สถานภาพสมรส จำนวนบุตร อาชีพ รายได้ต่อเดือน ประเภทธุรกิจในการประกอบการ ตำแหน่งหน้าที่ในการทำงาน อายุในการทำงาน ประเภทที่อยู่อาศัย สถานการอยู่อาศัย และการผ่อน/เช่าที่อยู่อาศัยต่อเดือน และ ตัวแปรตามประกอบด้วยการพิจารณาอนุมัติบัตรเครดิต โดยแบ่งเป็นการอนุมัติหรือไม่อนุมัติบัตรเครดิต รวมระยะเวลาดำเนินโครงการ 1 ปี

1.4 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

1.4.1 ทำให้ทราบวิธีที่มีประสิทธิภาพดีที่สุดในการอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทย

1.4.2 ทำให้ทราบวิธีที่มีการทำนายผลดีที่สุดในการอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทย

1.5 นิยามคำศัพท์

บัตรเครดิตหรือบัตรสินเชื่อ (Credit Card) เป็นบริการที่สถาบันทางการเงินต่าง ๆ ออกให้แก่ลูกค้าเพื่อใช้จ่ายแทนเงินสด บัตรเครดิตที่รู้จักกัน เช่น วีซ่า มาสเตอร์การ์ด เจซีบี อเมริกันเอกซ์เพรส ดิสคัฟเวอร์ และไดเนอร์สคลับ สามารถใช้ได้ตามจำนวนวงเงินบัตรที่อนุมัติ หักออกด้วยค่าสินค้าและบริการที่ใช้จ่ายผ่านบัตร ค่าธรรมเนียม ดอกเบี้ย และหนี้สินคงค้างที่ยังไม่ได้ชำระ

ธนาคารพาณิชย์ไทย (Thai Commercial Bank) หมายถึง การประกอบธุรกิจประเภทรับฝากเงินที่ต้องจ่ายคืนเมื่อทวงถาม หรือเมื่อสิ้นระยะเวลาอันได้กำหนดไว้และใช้ประโยชน์จากเงินนั้นในทางหนึ่งหรือหลายทาง เช่น การให้กู้ยืม ซื้อขายหรือเก็บเงินตามตั๋วเงินหรือตราสารเปลี่ยนมืออื่นใด ซื้อหรือขายเงินตราต่างประเทศ ทั้งนี้จะประกอบธุรกิจประเภทอื่น ๆ อันเป็นประเพณีของธนาคารพาณิชย์ด้วยก็ได้ สำหรับธนาคารพาณิชย์ไทยคือธนาคารที่ได้รับอนุญาตให้ประกอบการธนาคารพาณิชย์และหมายความรวมถึงสาขาของธนาคารในประเทศไทยที่ได้รับอนุญาตให้ประกอบการธนาคารพาณิชย์ด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

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

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

วิธีโครงข่ายประสาท (Neural Network Method) เป็นวิธีที่ใช้หลักการเลียนแบบการทำงานของสมองมนุษย์ เส้นเชื่อมแต่ละเส้นจะมีน้ำหนักถ่วง (weight) เพื่อใช้กำหนดน้ำหนักถ่วงหรือความสำคัญของข้อมูลเข้า (input data) กำหนดค่าเริ่มต้นโดยการสุ่ม ในแต่ละโหนดทำการคำนวณค่าผลรวมเชิงเส้นแบบถ่วงน้ำหนักและผ่านฟังก์ชันกระตุ้น (activation function) คำนวณค่าความคลาดเคลื่อน (error) ระหว่างคำตอบที่ทำนายได้กับเฉลย ถ้ามีความคลาดเคลื่อนเกิดขึ้นระบบจะทำการปรับปรุงค่าน้ำหนักถ่วงของแต่ละการเชื่อมต่อ (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

วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม (Binary Logistic Regression Method) เป็นวิธีการวิเคราะห์การถดถอยแบบหนึ่งโดยที่ตัวแปรตามเป็นตัวแปรเชิงคุณภาพที่มีค่าได้เพียง 2 ค่า (dichotomous or binary variable) ส่วนตัวแปรอิสระอาจจะเป็นตัวแปรเชิงปริมาณหรือเชิงคุณภาพ หรืออาจจะมีทั้งตัวแปรเชิงปริมาณและตัวแปรเชิงคุณภาพก็ได้

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

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

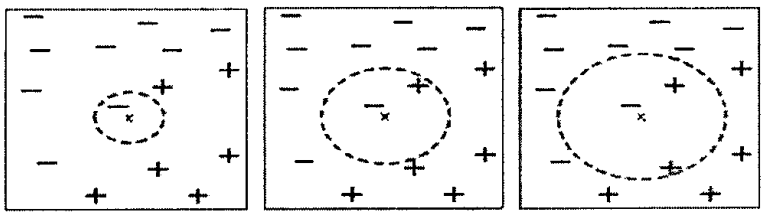
โดยที่ $dist(x_i, x_j)$ คือ ระยะห่างระหว่างตัวอย่าง x_i กับตัวอย่าง x_j

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

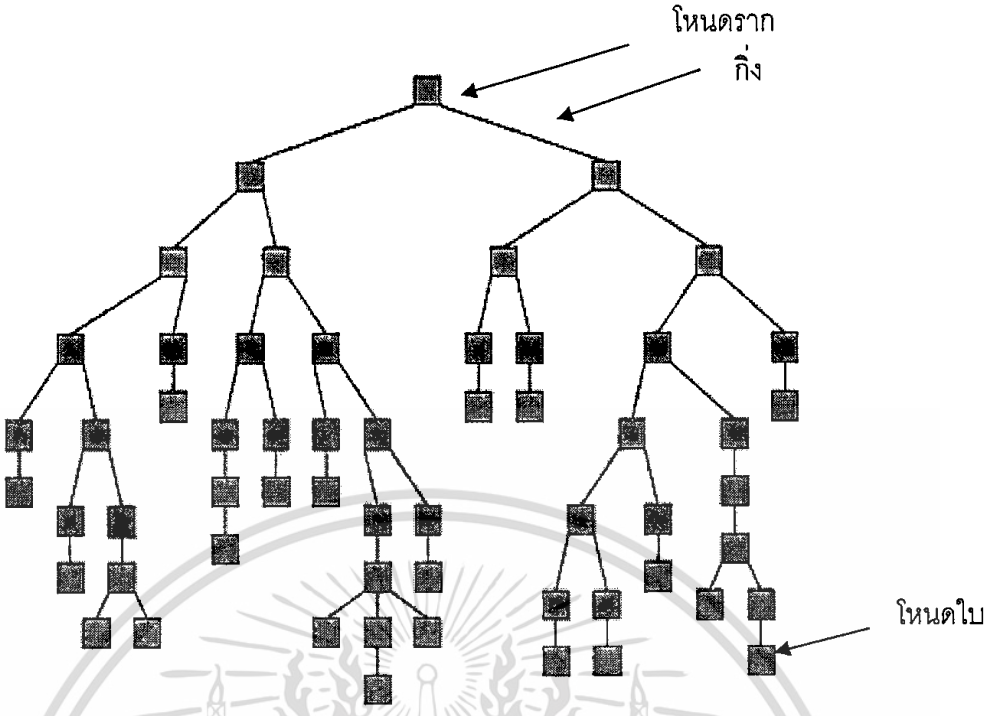
2.1.2.1 ส่วนประกอบของแผนภาพต้นไม้เพื่อการตัดสินใจ

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

ใช้ในการพิจารณาให้สินเชื่อแก่บุคคลต่าง ๆ

ใช้ในการทำนายว่าลูกค้าคนไหนที่มีโอกาสจะยกเลิกการใช้บริการและเหตุผลเพราะอะไร

2.1.3 วิธีโครงข่ายประสาท (Neural Network Method)

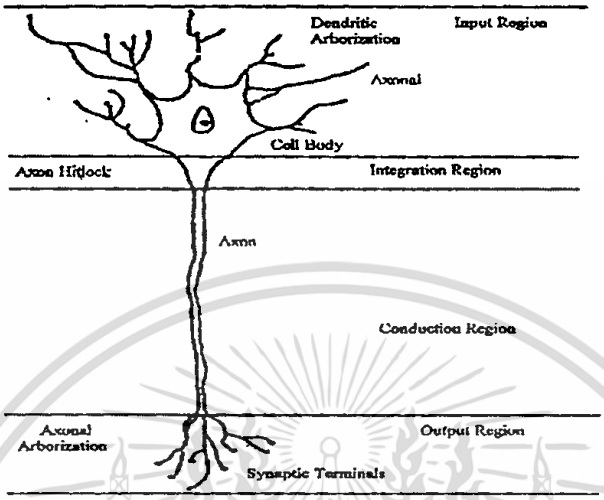
วิธีโครงข่ายประสาทใช้หลักการเลียนแบบการทำงานของสมองมนุษย์ เส้นเชื่อมแต่ละเส้น จะมีน้ำหนักถ่วง (weight) เพื่อใช้กำหนดน้ำหนักถ่วงหรือความสำคัญของข้อมูลเข้า (input data) กำหนดค่าเริ่มต้นโดยการสุ่ม ในแต่ละโหนดทำการคำนวณค่าผลรวมเชิงเส้นแบบถ่วงน้ำหนักและผ่านฟังก์ชันกระตุ้น (activation function) คำนวณค่าความคลาดเคลื่อน (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)

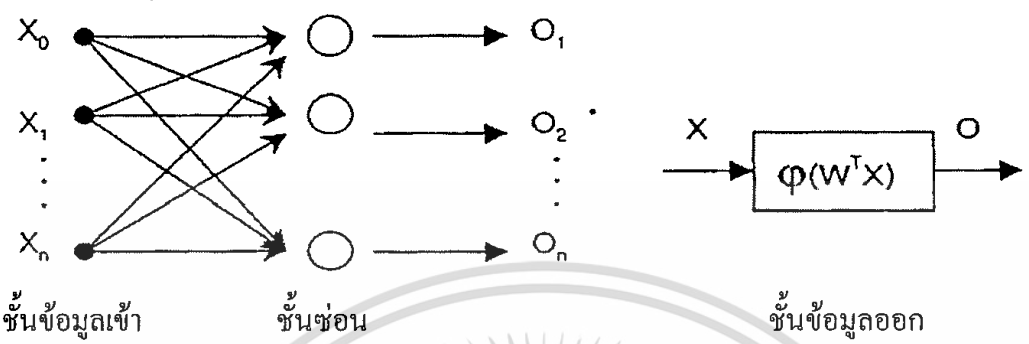
เพื่อให้โครงข่ายประสาทสามารถเรียนรู้ได้อย่างมีประสิทธิภาพ จำเป็นต้องมีการเชื่อมโยงกันระหว่างเซลล์ประสาท โดยทั่วไปสามารถแบ่งการเชื่อมโยงของโครงข่ายได้ 2 ลักษณะ คือ

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

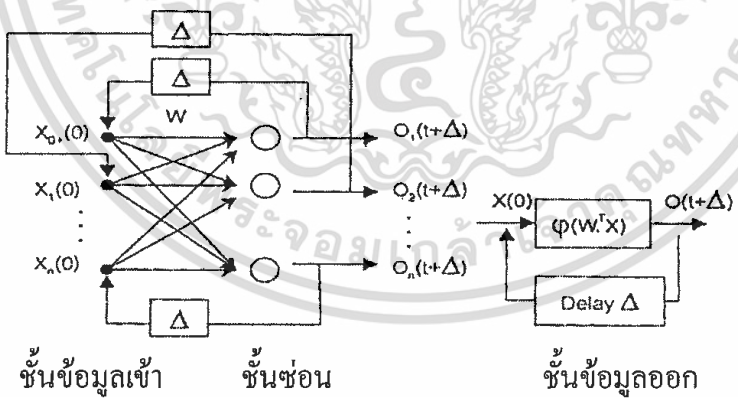
มากกว่าหนึ่งชั้นได้ การเชื่อมต่อระหว่างชั้นของโครงข่ายแบบส่งสัญญาณไปข้างหน้าจะมีค่าถ่วงน้ำหนัก (weight) เป็นตัวเชื่อมและสัญญาณนำเข้าที่เข้ามาจะถูกส่งไปตามทิศทางของลูกศรจนถึงชั้นข้อมูลออกโดยไม่มีการย้อนกลับ สามารถแสดงตัวแบบโครงข่ายแบบส่งสัญญาณไปข้างหน้าได้ดังรูปที่ 2.4



รูปที่ 2.4 ลักษณะโครงข่ายประสาทเทียมแบบส่งสัญญาณไปข้างหน้า

2) โครงข่ายแบบมีการย้อนกลับ (Feedback network)

โครงข่ายชนิดนี้มีชื่อเรียกอีกชื่อหนึ่งว่า โครงข่ายหันกลับ (recurrent network) เป็นโครงข่ายที่จะอาศัยทั้งข้อมูลในปัจจุบันและข้อมูลที่มีการประวิงเวลามาใช้ในการประมวลผลของโครงข่ายประสาท สามารถแสดงตัวแบบโครงข่ายแบบมีการย้อนกลับได้ดังรูปที่ 2.5



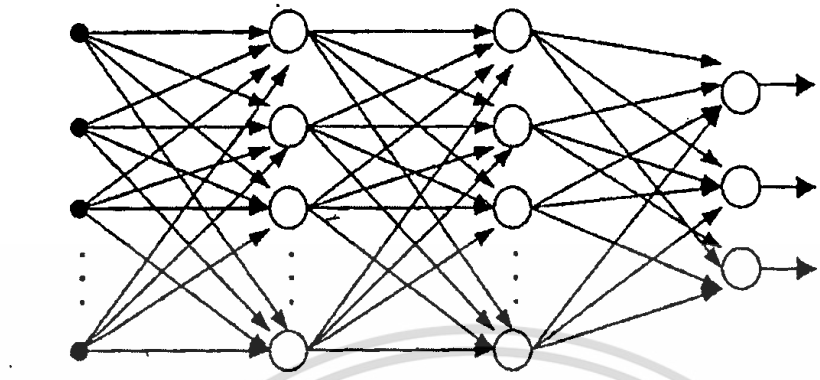
รูปที่ 2.5 ลักษณะโครงข่ายประสาทเทียมแบบมีการย้อนกลับ

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



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 แต่ถ้าอัตราการเรียนรู้ต่ำก็ควรจะมีค่าโมเมนตัมที่สูง

2.1.4 วิธีซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine Method)

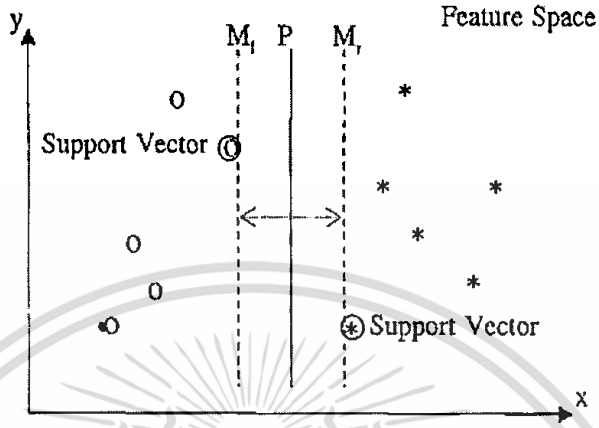
เป้าหมายของวิธีซัพพอร์ทเวกเตอร์แมชชีนคือกระบวนการสอนเครื่องแบบมีผู้สอน (supervised learning) เพื่อให้สามารถสร้างตัวจำแนกข้อมูล (classifier) ที่มีความทั่วไป (generalize) สูง นั่นคือสามารถทำงานได้ดีกับตัวอย่างที่ไม่รู้จัก (unknown database) ด้วยกระบวนการปรับรูปแบบข้อมูลจากข้อมูลที่มีมิติต่ำ (low dimension dataset) บนพื้นที่ข้อมูลนำเข้า (input space) ให้อยู่ในรูปแบบของข้อมูลที่มีมิติสูง (high dimension dataset) บนพื้นที่ข้อมูลคุณลักษณะ (feature space) โดยใช้ฟังก์ชันในการปรับรูปแบบข้อมูลเรียกว่าฟังก์ชันเคอร์เนล **(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 เส้นตรงใดที่เป็นเส้นที่ดีที่สุด โดยเส้นตรงนั้นจะเพิ่มเส้นขอบ (margin) ออกไปทั้งสองข้าง โดยเส้นขอบที่เพิ่มนั้น

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

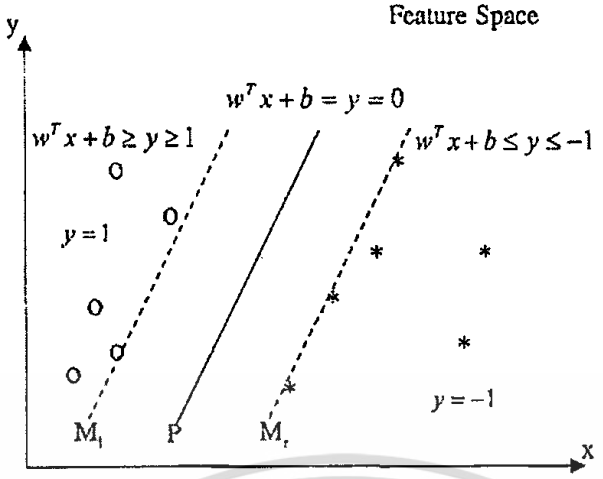
$$x = ((x_1, y_1),...,(x_i, y_i)) \quad (2.5)$$

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

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง



รูปที่ 2.8 เส้นขอบและเส้นแบ่งเมื่อแทนด้วยสมการเส้นตรง

จากรูปที่ 2.8 เส้นตรง M_1 แทนด้วยสมการ $w^T x + b \geq y \geq 1$ ซึ่งข้อมูล y ที่มากกว่า 1 ก็จะถูกกำหนดค่าใหม่โดยให้ y เท่ากับ 1 และพจน์ w ก็คือค่าความชัน เช่นเดียวกันกับเส้นตรง M_2 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

142668

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

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

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

โดยที่ M คือ ความกว้างของเส้นขอบ

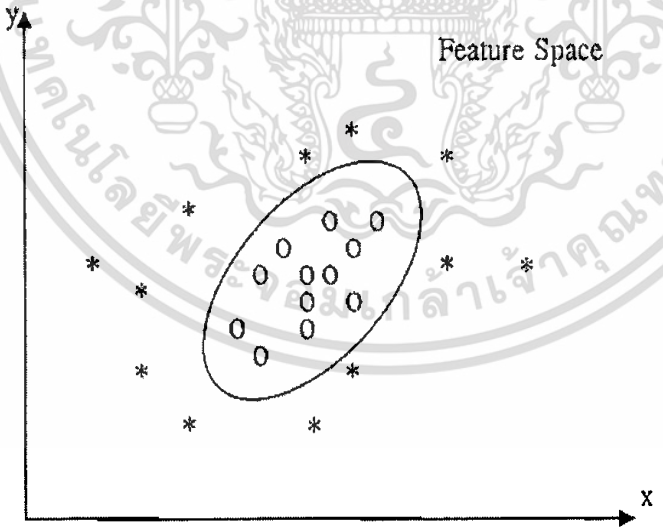
α คือ สัมประสิทธิ์คงที่

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

วิวัช ชลไชยะและกรุง สินอภิรมย์สราญ (2550) ธนาคารใช้การให้คะแนนสินเชื่อเพื่อจัดลำดับความเสี่ยงของลูกค้าที่ขอสินเชื่อตามศักยภาพของแต่ละคน คะแนนที่ได้ช่วยให้ธนาคารสามารถระบุลูกค้าที่มีความเสี่ยงสูงจากกลุ่มลูกค้าที่ขอสินเชื่อ ตัวแบบการให้คะแนนสินเชื่อได้จากข้อมูลที่สำคัญ ได้แก่ ข้อมูลส่วนตัวของลูกค้า ประวัติการชำระสินเชื่อและพฤติกรรมของลูกค้า งานวิจัยนี้เสนอตัวแบบการให้คะแนนแบบผสม โดยมีพื้นฐานมาจาก 2 เทคนิค การทำเหมืองข้อมูล คือการวิเคราะห์การเกาะกลุ่มและการจำแนกประเภท ซึ่งจะเรียกวิธีการนี้ว่า การวิเคราะห์การเกาะกลุ่มของตัวทำนายหลากหลาย (CLAMP) วิธีการนี้พัฒนาตัวแบบ 2 ส่วน ส่วนแรกใช้กระบวนการวิเคราะห์การเกาะกลุ่มที่ใช้ขั้นตอนวิธีของการวิเคราะห์การเกาะกลุ่มแบบค่าเฉลี่ยเอ็กซ์ กระบวนการนี้แบ่งกั้นข้อมูลออกเป็น k กลุ่ม โดยค่า k ได้จากการพิจารณาค่าเกณฑ์การวัดข้อมูล ส่วนที่ 2 เลือกวิธีการจำแนกประเภทข้อมูลจาก J48 (ตัวแบบต้นไม้การตัดสินใจ) วิธีการจำแนกแบบเบย์อย่างง่าย (ตัวแบบเชิงความน่าจะเป็น) สมการถดถอยแบบโลจิสติก (ตัวแบบเชิงสถิติ) และข่ายงานประสาท (ตัวแบบปัญญาประดิษฐ์) เกณฑ์ที่ใช้ในการเลือกตัวแบบจำแนกประเภทแบ่ง 40% เป็นข้อมูลพัฒนาตัวแบบ สำหรับสร้างตัวแบบ 30% เป็นข้อมูลประเมิน สำหรับเลือกวิธีการจำแนกประเภทที่ดีที่สุดในกลุ่ม และ 30% เป็นข้อมูลทดสอบ เพื่อป้องกันปัญหาตัวแบบเหมาะสมเฉพาะข้อมูลพัฒนาตัวแบบ วิธีการจำแนกประเภทที่ใช้ CLAMP แสดงความถูกต้องที่ดีกว่าการใช้วิธีการจำแนกประเภทเพียงอย่างเดียว

บุษกร สุคนธวงศาโรจน์ (2551) ได้นำเสนอการจำแนกบุคลากรในองค์กรสำหรับการสร้างแผนที่ความรู้โดยข้อมูลที่ใช้ในการทำลองได้จากหน่วยงานการไฟฟ้าฝ่ายผลิตแห่งประเทศไทย สายงานรองผู้ว่าการบัญชีและการเงิน สายงานรองผู้ว่าการผลิตไฟฟ้า และสายงานรองผู้ว่าการเชื้อเพลิง โดยทำการทดลองกับกลุ่มตัวอย่าง 28 หน่วยงาน โดยใช้วิธีการจำแนกประเภทความเชี่ยวชาญเชิงเปรียบเทียบด้วยอัลกอริทึม 3 แบบ ได้แก่ Decision Tree (C4.5), OneR, Naïve Bayes การทดลองเพื่อหาอัลกอริทึมที่มีความแม่นยำสูงสุดในการทำนายความเชี่ยวชาญโดยใช้ตัววัดประสิทธิภาพ 4 ตัว ได้แก่ Accuracy, Precision/Recall, F-measure และ Root Mean-Squared Error (RMSE) จากการทดลองพบว่าอัลกอริทึมที่ให้ค่าความถูกต้องในการทำนายคลาสของความเชี่ยวชาญสูงสุดคือ C4.5

กิตติพล วิแสงและคณะ (2552) ได้ศึกษาปัจจัยเสี่ยงของโรคเบาหวานที่มีความสำคัญอย่างยิ่งในการนำมาเป็นเครื่องมือประเมินการเกิดโรคเบาหวานแทนการตรวจเลือด บทความนี้ศึกษาและวิเคราะห์ปัจจัยเสี่ยงและลำดับของปัจจัยเสี่ยงในการเกิดโรคเบาหวาน โดยตัวแบบที่ใช้ศึกษาคือตัวแบบ Back-propagation Neural Networks, Radial Basis Function Network และตัวแบบ Naïve Bayes ผลวินิจฉัยทางการแพทย์ได้ถูกนำมาอ้างอิงเพื่อวัดความถูกต้องของตัวแบบ ผลการศึกษาแสดงให้เห็นว่าตัวแบบ Back-propagation Neural Networks ให้ผลลัพธ์ที่มีความ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ถูกต้องมากที่สุด ปัจจัยเสี่ยงและลำดับของปัจจัยเสี่ยงสามารถตัดสินได้จากตัวแบบที่ศึกษา ซึ่งข้อมูลเหล่านี้สามารถนำไปสร้างเป็นแบบประเมินความเสี่ยงโรคเบาหวานโดยไม่ต้องอาศัยการตรวจเลือดได้ รวมทั้งเป็นเครื่องมือในการประเมินตนเองและสามารถพัฒนาเป็นระบบผู้เชี่ยวชาญสำหรับการวินิจฉัยโรคเบาหวานต่อไปได้

พลอยพรรณ สอนสุวิทย์ และตรัสพงศ์ ไทยอุปถัมภ์ (2552) ได้นำเสนอการเปรียบเทียบประสิทธิภาพการตรวจจับสิ่งผิดปกติทางเครือข่าย โดยการสแกนหาจุดอ่อน ซึ่งเป็นการบุกรุกทางเครือข่ายที่มีความสำคัญ งานวิจัยนี้ได้ศึกษาเทคนิควิธีการจำแนก (Classification) มาตรวจจับการสแกน ได้แก่ SVM, C4.5, Naïve Bayes และ Neural Network จากการศึกษา ทำให้ทราบถึงประสิทธิภาพของแต่ละเทคนิค เช่น ร้อยละของความถูกต้อง (% Correct) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error) พบว่าวิธี C4.5 วิเคราะห์ได้แม่นยำที่สุดและมีค่า Error ต่ำที่สุด

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

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

3.1.2 อุปกรณ์ที่ต้องการเพิ่ม

1) แบบฟอร์มใบสมัครบัตรเครดิต

3.2 ขั้นตอนการดำเนินงานวิจัย

งานวิจัยนี้ได้ทำการศึกษาเรื่องการเปรียบเทียบประสิทธิภาพและทำนายผลการอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทยโดยใช้การวิเคราะห์จำแนกกลุ่ม 5 วิธี คือ วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีความใกล้เคียงกันมากที่สุด วิธีโครงข่ายประสาท วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม โดยมีขั้นตอนการดำเนินงานดังนี้

3.2.1 การเก็บรวบรวมข้อมูล

เก็บรวบรวมข้อมูลลูกค้าจากแบบฟอร์มใบสมัครบัตรเครดิตการอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทย จำนวน 519 ชุด โดยข้อมูลประกอบด้วยคุณลักษณะ (Attribute) ต่าง ๆ ดังนี้

1. เพศ
    1) เพศชาย
    2) เพศหญิง
2. อายุ หน่วยเป็นปี
3. ระดับการศึกษา
    1) ประถมศึกษา
    2) มัธยมศึกษาตอนต้น
    2) มัธยมศึกษาตอนปลาย/ปวช./เทียบเท่า
    3) อนุปริญญา/ปวส./ปวท./เทียบเท่า
    4) ปริญญาตรี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) ปริญญาโท

6) ปริญญาเอก

4. สถานภาพสมรส

1) โสด

2) สมรสจดทะเบียน

3) สมรสไม่จดทะเบียน

4) หม้าย

5) หย่าร้าง

6) แยกกันอยู่

5. จำนวนบุตร หน่วยเป็นคน

6. อาชีพ

1) รับราชการ	2) พนักงานรัฐวิสาหกิจ
3) พนักงานบริษัท	4) เจ้าของกิจการ
5) พ่อบ้าน/แม่บ้าน	6) อาชีพอิสระ
7) แพทย์/สัตวแพทย์/ทันตแพทย์	8) พยาบาล
9) นักกฎหมาย	10) ตำรวจ/ทหาร
11) นักการเมือง	12) นักศึกษา
13) ครู/อาจารย์	14) ศิลปิน/นักแสดง
15) นักบัญชี	16) นักคอมพิวเตอร์/นักเทคโนโลยีสารสนเทศ
17) ข้าราชการบำนาญ/เกษียณอายุ	18) รับจ้าง/พนักงานรายวัน/ชั่วคราว
19) เกษตร/ปศุสัตว์/ประมง	20) วิศวกร/สถาปนิก/มัณฑนากร
21) อื่น ๆ	

7. รายได้ต่อเดือน หน่วยเป็นบาท

8. ประเภทธุรกิจในการประกอบการ

1) ข้าราชการ	2) สายการบิน
3) ขนส่ง	4) ธุรกิจโรงแรม
5) ประกันภัย	6) ที่ปรึกษา
7) นำเข้า/ส่งออก	8) อุตสาหกรรม/โรงงาน
9) การสื่อสาร/โทรคมนาคม	10) โรงพิมพ์/สิ่งพิมพ์/บรรจุภัณฑ์
11) การค้าปลีก/ค้าส่ง/ห้างสรรพสินค้า	12) อสังหาริมทรัพย์/พัฒนาที่ดิน/ก่อสร้าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

13) ปศุสัตว์/ประมง/เกษตรกรรม/โรงสีข้าว 14) สาธารณูปโภค/ไฟฟ้า/ประปา

15) สถาบันการศึกษา/สถาบันกวดวิชา 16) สถานพยาบาล/สถานีอนามัย/โรงพยาบาล

17) ธนาคาร/สถาบันการเงิน/บริษัทหลักทรัพย์/โรงจำนำ

18) กองทัพบก/เรือ/อากาศ/ตำรวจ

19) บริการ/บันเทิง/ท่องเที่ยว/ร้านอาหาร/ภัตตาคาร 20) ยานพาหนะ/อะไหล่

21) เคมีภัณฑ์ 22) ร้านสะดวกซื้อ

23) อุปกรณ์อิเล็กทรอนิกส์/เครื่องใช้ไฟฟ้า 24) ปิโตรเคมี/ผลิตภัณฑ์

25) พลาสติก 26) อัญมณี/เครื่องประดับ

27) เครื่องจักร/ผลิตภัณฑ์โลหะ 28) วัสดุก่อสร้าง

29) รับเหมาก่อสร้าง 30) สิ่งทอ

31) เฟอร์นิเจอร์/โรงเลื่อย 32) เวชภัณฑ์/โรงพยาบาล/คลินิก

33) อื่น ๆ

9. ตำแหน่งหน้าที่ในการทำงาน

10. อายุในการทำงาน หน่วยเป็นปี

11. ประเภทที่อยู่อาศัย

1) บ้านเดี่ยว

2) ทาวน์เฮาส์/อาคารพาณิชย์

3) คอนโดมิเนี่ยม/อพาร์ตเมนท์/แมนชั่น/หอพัก

12. สถานะการอยู่อาศัย

1) บ้านของตนเอง/คู่สมรสปลอดภาระ

2) บ้านของตนเอง/คู่สมรสไม่ปลอดภาระ

3) บ้านบิดามารดา

4) บ้านญาติ/พี่น้อง/บุคคลอื่น

5) บ้านพักสวัสดิการ

13. การผ่อน/เช่าที่อยู่อาศัยต่อเดือน หน่วยเป็นบาท

14. การพิจารณาอนุมัติบัตรเครดิต

1) อนุมัติ

2) ไม่อนุมัติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2 บันทึกข้อมูลการอนุมัติบัตรเครดิตจำนวน 519 ชุด ลงในโปรแกรม Microsoft Excel โดยให้แนวคอลัมน์เป็นเพศ อายุ ระดับการศึกษา สถานภาพสมรส จำนวนบุตร อาชีพ รายได้ต่อเดือน ประเภทธุรกิจในการประกอบการ ตำแหน่งหน้าที่ในการทำงาน อายุในการทำงาน ประเภทที่อยู่อาศัย สถานะการอยู่อาศัย การผ่อน/เช่าที่อยู่อาศัยต่อเดือน และการพิจารณาอนุมัติบัตรเครดิต ส่วนแนวแถวเป็นลำดับที่ของลูกค้า

3.2.3 การแบ่งข้อมูล

นำข้อมูลทั้งหมดมาแบ่งออกเป็น 3 ส่วน จากข้อมูลการอนุมัติบัตรเครดิต จำนวน 519 ชุด เพื่อใช้สำหรับการวิเคราะห์ข้อมูล โดยแบ่งข้อมูลออกเป็นสัดส่วนดังนี้

ส่วนที่ 1 : ข้อมูลชุดฝึกหัด (Training Data Set) เพื่อนำไปสร้างตัวแบบ

มีข้อมูล 70% ของข้อมูลทั้งหมด ซึ่งจะได้ข้อมูลในส่วนที่ 1 จำนวน 365 ชุด

ส่วนที่ 2 : ข้อมูลชุดทดสอบประสิทธิภาพ (Evaluation Data Set) เพื่อนำไปทดสอบความถูกต้องของตัวแบบ

มีข้อมูล 20% ของข้อมูลทั้งหมด ซึ่งจะได้ข้อมูลในส่วนที่ 2 จำนวน 100 ชุด

ส่วนที่ 3 : ข้อมูลชุดทดสอบ (Testing Data Set) เพื่อนำไปทำนายตัวแบบ

มีข้อมูล 10% ของข้อมูลทั้งหมด ซึ่งจะได้ข้อมูลในส่วนที่ 3 จำนวน 54 ชุด

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

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

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) 20,000 รอบ การวิจัยครั้งนี้ใช้อัลกอริทึมของวิธีโครงข่ายประสาทชนิดเพอร์เซปตรอนหลายชั้นที่มีชั้นซ่อน (Hidden Layer) 1 ชั้น แม้ว่าโครงสร้างโครงข่ายประสาทที่ซับซ้อนสามารถมีชั้นซ่อนมากกว่า 1 ชั้น แต่ในทางปฏิบัติ การกำหนดชั้นซ่อน 1 ชั้น ก็เพียงพอต่อการวิเคราะห์ข้อมูล (Berson, A. and Stephen, J. Smith, 1997)

4) วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) ใช้อัลกอริทึม SMO ชนิดโพลิโนเมียลเคอร์เนล (Polynomial Kernel) เนื่องจากงานวิจัยที่อ้างอิงจากวาทินี นุ้ยเพียร และคณะ (2553) ได้ผลว่าวิธีซัพพอร์ตเวกเตอร์แมชชีนที่ใช้อัลกอริทึมชนิดโพลิโนเมียลเคอร์เนล ดีที่สุด

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

การนำผลการวิเคราะห์ข้อมูลมาประเมินผลเพื่อเปรียบเทียบประสิทธิภาพและทำนายผลการอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทยของวิธีการจำแนกกลุ่มโดยนำข้อมูลที่ได้จากการวิเคราะห์มาเปรียบเทียบว่าวิธีในการจำแนกกลุ่มข้อมูลวิธีใดระหว่างวิธีความใกล้เคียงกันมากที่สุด 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 Weka 3.7 แสดงผลลัพธ์ของการทำนายการพิจารณาอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทย ด้วยวิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาท วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม ได้ผลดังนี้

4.1.1 วิธีความใกล้เคียงกันมากที่สุด (K-Nearest Neighbor Method)

วิธีความใกล้เคียงกันมากที่สุดนั้นมีอัลกอริทึมหลายอัลกอริทึม แต่ในงานวิจัยนี้ใช้อัลกอริทึม IBk ในการทำนายการพิจารณาอนุมัติบัตรเครดิต ได้ผลดังนี้

4.1.1.1 การสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต

สร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิตจากข้อมูลฝึกหัด (Training data set) ด้วยไฟล์ CreditApproval.csv จำนวน 365 คน ได้ผลดังนี้

ตารางที่ 4.1 ผลในส่วนของการสรุปผล (Summary) จากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด

Correctly Classified Instances	233	63.8356 %
Incorrectly Classified Instances	132	36.1644 %
Kappa statistic	0.2737	
Mean absolute error	0.3625	
Root mean squared error	0.5996	
Relative absolute error	72.6503 %	
Root relative squared error	120.0365 %	
Coverage of cases (0.95 level)	63.8356 %	
Mean rel. region size (0.95 level)	50 %	
Total Number of Instances	365	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.1 จากข้อมูลจำนวน 365 คน ทำนายข้อมูลถูกต้องจำนวน 233 คน คิดเป็น 63.8356% และทำนายข้อมูลไม่ถูกต้องจำนวน 132 คน คิดเป็น 36.1644% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.2737 แสดงว่าข้อมูลมีความสอดคล้องกันค่อนข้างน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.3625 ซึ่งมีค่าค่อนข้างน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.5996)^2 = 0.3595$ ซึ่งมีค่าค่อนข้างน้อย แสดงว่าตัวแบบมีความถูกต้องพอสมควร

ตารางที่ 4.2 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบ (Detailed Accuracy By Class) จากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.598	0.325	0.627	0.598	0.612	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.675	0.402	0.648	0.675	0.662	no
Weighted Avg.	0.638	0.365	0.638	0.638	0.638	

จากตารางที่ 4.2 สำหรับคำตอบอนุมัติบัตรเครดิต (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก (TP Rate) = 0.598 ค่าอัตราความผิดพลาดเชิงบวก (FP Rate) = 0.325 ค่าความแม่นยำ (Precision) = 0.627 ค่าความระลึก (Recall) = 0.598 และค่าความถ่วงดุล (F-Measure) = 0.612 ส่วนคำตอบไม่อนุมัติบัตรเครดิต (Class = no) มีค่าอัตราความถูกต้องเชิงลบ (TN Rate) = 0.675 ค่าอัตราความผิดพลาดเชิงลบ (FN Rate) = 0.402 ค่าความแม่นยำ = 0.648 ค่าความระลึก = 0.675 และค่าความถ่วงดุล = 0.662

ตารางที่ 4.3 ผลในส่วนของเมตริกซ์ความสับสน (Confusion Matrix) จากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด

		ผลการจำแนกอนุมัติบัตรเครดิต	
		อนุมัติ	ไม่อนุมัติ
ค่าที่แท้จริง	อนุมัติ	104	70
	ไม่อนุมัติ	62	129

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.3 มีข้อมูล 365 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 233 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าอนุมัติบัตรเครดิต 104 คน และจำนวนข้อมูลที่จำแนกถูกว่าไม่อนุมัติบัตรเครดิต 129 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 132 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วไม่อนุมัติบัตรเครดิต 62 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่อนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วอนุมัติบัตรเครดิต 70 คน

4.1.1.2 การทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต

ทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิตจากชุดข้อมูลทดสอบประสิทธิภาพ (Evaluation data set) ด้วยไฟล์ CreditApproval-eval.csv จำนวน 100 คน ได้ผลดังนี้

ตารางที่ 4.4 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด

Correctly Classified Instances	65	65 %
Incorrectly Classified Instances	35	35 %
Kappa statistic	0.2978	
Mean absolute error	0.3558	
Root mean squared error	0.5921	
Coverage of cases (0.95 level)	65 %	
Total Number of Instances	100	

จากตารางที่ 4.4 จากข้อมูลจำนวน 100 คน ทำนายข้อมูลถูกต้องจำนวน 65 คน คิดเป็น 65% และทำนายข้อมูลไม่ถูกต้องจำนวน 35 คน คิดเป็น 35% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.2978 แสดงว่าข้อมูลมีความสอดคล้องกันค่อนข้างน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.3558 ซึ่งมีค่าค่อนข้างน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.5921)^2 = 0.3506$ ซึ่งมีค่าค่อนข้างน้อย แสดงว่าตัวแบบมีความถูกต้องพอสมควร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.571	0.275	0.667	0.571	0.615	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.725	0.429	0.638	0.725	0.679	no
Weighted Avg.	0.65	0.353	0.652	0.65	0.648	

จากตารางที่ 4.5 สำหรับคำตอบอนุมัติบัตรเครดิต (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.571 , ค่าอัตราความผิดพลาดเชิงบวก = 0.275, ค่าความแม่นยำ = 0.667, ค่าความระลึก = 0.571 และค่าความถ่วงดุล = 0.615 ส่วนคำตอบไม่อนุมัติบัตรเครดิต (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.725, ค่าอัตราความผิดพลาดเชิงลบ = 0.429, ค่าความแม่นยำ = 0.638, ค่าความระลึก = 0.725 และค่าความถ่วงดุล = 0.679

ตารางที่ 4.6 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด

		ผลการจำแนกอนุมัติบัตรเครดิต	
		อนุมัติ	ไม่อนุมัติ
ค่าที่แท้จริง	อนุมัติ	28	21
	ไม่อนุมัติ	14	37

จากตารางที่ 4.6 มีข้อมูล 100 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 65 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าอนุมัติบัตรเครดิต 28 คน และจำนวนข้อมูลที่จำแนกถูกว่าไม่อนุมัติบัตรเครดิต 37 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 35 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วไม่อนุมัติบัตรเครดิต 14 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่อนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วอนุมัติบัตรเครดิต 21 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.1.3 การทำนายการพิจารณาอนุมัติบัตรเครดิต

การทำนายการพิจารณาอนุมัติบัตรเครดิตจากชุดข้อมูลทดสอบ (Testing data set) ด้วยไฟล์ CreditApproval-test.csv จำนวน 54 คน ได้ผลดังนี้

ตารางที่ 4.7 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ (Predictions on test set) ด้วยวิธีความใกล้เคียงกันมากที่สุด

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
1	1:yes	2:no	+	0.997
2	2:no	2:no		0.997
3	2:no	2:no		0.997
4	2:no	2:no		0.997
5	2:no	2:no		0.997
6	1:yes	1:yes		0.997
7	1:yes	1:yes		0.997
8	1:yes	1:yes		0.997
9	2:no	1:yes	+	0.997
10	2:no	1:yes	+	0.997
11	1:yes	2:no	+	0.997
12	2:no	2:no		0.997
13	2:no	2:no		0.997
14	1:yes	1:yes		0.997
15	1:yes	2:no	+	0.997
16	1:yes	1:yes		0.997
17	1:yes	1:yes		0.997
18	1:yes	1:yes		0.997
19	1:yes	2:no	+	0.997
20	1:yes	2:no	+	0.997
21	2:no	2:no		0.997
22	2:no	1:yes	+	0.997

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
28	1:yes	1:yes		0.997
29	1:yes	2:no	+	0.997
30	1:yes	1:yes		0.997
31	1:yes	1:yes		0.997
32	1:yes	1:yes		0.997
33	2:no	1:yes	+	0.997
34	2:no	2:no		0.997
35	1:yes	2:no	+	0.997
36	2:no	2:no		0.997
37	1:yes	2:no	+	0.997
38	2:no	2:no		0.997
39	2:no	2:no		0.997
40	2:no	2:no		0.997
41	2:no	2:no		0.997
42	2:no	1:yes	+	0.997
43	2:no	2:no		0.997
44	2:no	2:no		0.997
45	2:no	2:no		0.997
46	2:no	1:yes	+	0.997
47	1:yes	1:yes		0.997
48	1:yes	2:no	+	0.997
49	2:no	2:no		0.997

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)	ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
23	1:yes	1:yes		0.997	50	2:no	2:no		0.997
24	1:yes	2:no	+	0.997	51	1:yes	2:no	+	0.997
25	1:yes	2:no	+	0.997	52	1:yes	1:yes		0.997
26	1:yes	2:no	+	0.997	53	2:no	1:yes	+	0.997
27	2:no	2:no		0.997	54	1:yes	2:no	+	0.997

จากตารางที่ 4.7 ค่าทำนายที่ทายผิดในช่อง error จะมีเครื่องหมาย + อยู่จำนวน 21 ค่า คือค่าที่ 1, 9, 10, 11, 15, 19, 20, 22, 24, 25, 26, 29, 33, 35, 37, 42, 46, 48, 51, 53 และ 54 โดยที่ค่าที่ 9, 10, 22, 33, 42, 46 และ 53 ไม่อนุมัติบัตรเครดิต (class 2 : no) แต่ทำนายว่าอนุมัติบัตรเครดิต (class 1 : yes) และค่าที่ 1, 11, 15, 19, 20, 24, 25, 26, 29, 35, 37, 48, 51 และ 54 อนุมัติบัตรเครดิต แต่ทำนายว่าไม่อนุมัติบัตรเครดิต

ตารางที่ 4.8 ผลในส่วนของการสรุปผลจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด

Correctly Classified Instances	33	61.1111 %
Incorrectly Classified Instances	21	38.8889 %
Kappa statistic	0.2286	
Mean absolute error	0.3895	
Root mean squared error	0.6219	
Coverage of cases (0.95 level)	61.1111 %	
Total Number of Instances	54	

จากตารางที่ 4.8 จากข้อมูลจำนวน 54 คน ทำนายข้อมูลถูกต้องจำนวน 33 คน คิดเป็น 61.1111% และทำนายข้อมูลไม่ถูกต้องจำนวน 21 คน คิดเป็น 38.8889% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.2286 แสดงว่าข้อมูลมีความสอดคล้องกันค่อนข้างน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.3895 ซึ่งมีค่าค่อนข้างน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.6219)^2 = 0.3868$ ซึ่งมีค่าค่อนข้างน้อย แสดงว่าตัวแบบมีความถูกต้องพอสมควร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.9 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.5	0.269	0.667	0.5	0.571	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.731	0.5	0.576	0.731	0.644	no
Weighted Avg.	0.611	0.38	0.623	0.611	0.606	

จากตารางที่ 4.9 สำหรับคำตอบอนุมัติบัตรเครดิต (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.5 ค่าอัตราความผิดพลาดเชิงบวก = 0.269 ค่าความแม่นยำ = 0.667 ค่าความระลึก = 0.5 และค่าความถ่วงดุล = 0.571 ส่วนคำตอบไม่อนุมัติบัตรเครดิต (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.731 ค่าอัตราความผิดพลาดเชิงลบ = 0.5 ค่าความแม่นยำ = 0.576 ค่าความระลึก = 0.731 และค่าความถ่วงดุล = 0.644

ตารางที่ 4.10 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีความใกล้เคียงกันมากที่สุด

		ผลการจำแนกอนุมัติบัตรเครดิต	
		อนุมัติ	ไม่อนุมัติ
ค่าที่แท้จริง	อนุมัติ	14	14
	ไม่อนุมัติ	7	19

จากตารางที่ 4.10 มีข้อมูล 54 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 33 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าอนุมัติบัตรเครดิต 14 คน และจำนวนข้อมูลที่จำแนกถูกว่าไม่อนุมัติบัตรเครดิต 19 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 21 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วไม่อนุมัติบัตรเครดิต 7 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่อนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วอนุมัติบัตรเครดิต 14 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.2 วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree Method)

วิธีแผนภาพต้นไม้เพื่อการตัดสินใจนั้นมีอัลกอริทึมหลายอัลกอริทึม แต่ในงานวิจัยนี้ใช้อัลกอริทึม J48 ในการทำนายการพิจารณาอนุมัติบัตรเครดิต ได้ผลดังนี้

4.1.2.1 การสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต

สร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิตจากชุดข้อมูลฝึกหัด (Training data set) ด้วยไฟล์ CreditApproval.csv จำนวน 365 คน ได้ผลดังนี้

ตารางที่ 4.11 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

Correctly Classified Instances	220	60.274 %
Incorrectly Classified Instances	145	39.726 %
Kappa statistic	0.1945	
Mean absolute error	0.455	
Root mean squared error	0.5009	
Relative absolute error	91.1974 %	
Root relative squared error	100.2801 %	
Coverage of cases (0.95 level)	98.0822 %	
Mean rel. region size (0.95 level)	96.4384 %	
Total Number of Instances	365	

จากตารางที่ 4.11 จากข้อมูลจำนวน 365 คน ทำนายข้อมูลถูกต้องจำนวน 220 คน คิดเป็น 60.274% และทำนายข้อมูลไม่ถูกต้องจำนวน 145 คน คิดเป็น 39.726% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.1945 แสดงว่าข้อมูลมีความสอดคล้องกันน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.455 ซึ่งมีค่าปานกลาง แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.5009)^2 = 0.2509$ ซึ่งมีค่าค่อนข้างน้อย แสดงว่าตัวแบบมีความถูกต้องพอสมควร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.12 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.454	0.262	0.612	0.454	0.521	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.738	0.546	0.597	0.738	0.66	no
Weighted Avg.	0.603	0.41	0.605	0.603	0.594	

จากตารางที่ 4.12 สำหรับคำตอบอนุมัติบัตรเครดิต (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.454 ค่าอัตราความผิดพลาดเชิงบวก = 0.262 ค่าความแม่นยำ = 0.612 ค่าความระลึก = 0.454 และค่าความถ่วงดุล = 0.521 ส่วนคำตอบไม่อนุมัติบัตรเครดิต (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.738 ค่าอัตราความผิดพลาดเชิงลบ = 0.546 ค่าความแม่นยำ = 0.597 ค่าความระลึก = 0.738 และค่าความถ่วงดุล = 0.66

ตารางที่ 4.13 ผลในส่วนของเมตริกซ์ความสับสนจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

		ผลการจำแนกอนุมัติบัตรเครดิต	
		อนุมัติ	ไม่อนุมัติ
ค่าที่แท้จริง	อนุมัติ	79	95
	ไม่อนุมัติ	50	141

จากตารางที่ 4.13 มีข้อมูล 365 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 220 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าอนุมัติบัตรเครดิต 79 คน และจำนวนข้อมูลที่จำแนกถูกว่าไม่อนุมัติบัตรเครดิต 141 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 145 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วไม่อนุมัติบัตรเครดิต 50 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่อนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วอนุมัติบัตรเครดิต 95 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.2.2 การทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต

ทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิตจากชุดข้อมูลทดสอบประสิทธิภาพด้วยไฟล์ CreditApproval-eval.csv จำนวน 100 คน ได้ผลดังนี้

ตารางที่ 4.14 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

Correctly Classified Instances	71	71 %
Incorrectly Classified Instances	29	29 %
Kappa statistic	0.4172	
Mean absolute error	0.3929	
Root mean squared error	0.4689	
Coverage of cases (0.95 level)	96 %	
Total Number of Instances	100	

จากตารางที่ 4.14 จากข้อมูลจำนวน 100 คน ทำนายข้อมูลถูกต้องจำนวน 71 คน คิดเป็น 71% และทำนายข้อมูลไม่ถูกต้องจำนวน 29 คน คิดเป็น 29% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.4172 แสดงว่าข้อมูลมีความสอดคล้องกันปานกลาง มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.3929 ซึ่งมีค่าค่อนข้างน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.4689)^2 = 0.2199$ ซึ่งมีค่าค่อนข้างน้อย แสดงว่าตัวแบบมีความถูกต้องพอสมควร

ตารางที่ 4.15 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.529	0.176	0.763	0.592	0.667	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.824	0.408	0.677	0.824	0.743	no
Weighted Avg.	0.71	0.295	0.719	0.71	0.706	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.15 สำหรับคำตอบอนุมัติบัตรเครดิต (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.529 ค่าอัตราความผิดพลาดเชิงบวก = 0.176 ค่าความแม่นยำ = 0.763 ค่าความระลึก = 0.592 และค่าความถ่วงดุล = 0.667 ส่วนคำตอบไม่อนุมัติบัตรเครดิต (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.824 ค่าอัตราความผิดพลาดเชิงลบ = 0.408 ค่าความแม่นยำ = 0.677 ค่าความระลึก = 0.824 และค่าความถ่วงดุล = 0.743

ตารางที่ 4.16 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

		ผลการจำแนกอนุมัติบัตรเครดิต	
		อนุมัติ	ไม่อนุมัติ
ค่าที่แท้จริง	อนุมัติ	29	20
	ไม่อนุมัติ	9	42

จากตารางที่ 4.16 มีข้อมูล 100 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 71 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าอนุมัติบัตรเครดิต 29 คน และจำนวนข้อมูลที่จำแนกถูกว่าไม่อนุมัติบัตรเครดิต 42 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 29 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วไม่อนุมัติบัตรเครดิต 9 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่อนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วอนุมัติบัตรเครดิต 20 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.2.3 การทำนายการพิจารณาอนุมัติบัตรเครดิต

การทำนายการพิจารณาอนุมัติบัตรเครดิตจากข้อมูลทดสอบด้วยไฟล์ CensusIncome-test.csv จำนวน 54 คน ได้ผลดังนี้

ตารางที่ 4.17 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	กาทำนาย (prediction)	ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	กาทำนาย (prediction)
1	1:yes	2:no	+	0.8	28	1:yes	1:yes		0.75
2	2:no	2:no		0.8	29	1:yes	1:yes		0.833
3	2:no	1:yes	+	0.632	30	1:yes	1:yes		1
4	2:no	2:no		0.8	31	1:yes	1:yes		0.846
5	2:no	1:yes	+	1	32	1:yes	1:yes		0.643
6	1:yes	1:yes		0.714	33	2:no	1:yes	+	0.833
7	1:yes	2:no	+	0.8	34	2:no	2:no		0.67
8	1:yes	1:yes		0.75	35	1:yes	2:no	+	0.67
9	2:no	2:no		0.8	36	2:no	2:no		0.67
10	2:no	1:yes	+	0.846	37	1:yes	2:no	+	0.67
11	1:yes	1:yes		0.75	38	2:no	2:no		0.67
12	2:no	2:no		0.8	39	2:no	2:no		0.67
13	2:no	1:yes	+	0.75	40	2:no	2:no		0.67
14	1:yes	1:yes		1	41	2:no	2:no		0.67
15	1:yes	1:yes		0.75	42	2:no	2:no		0.67
16	1:yes	1:yes		0.714	43	2:no	2:no		0.67
17	1:yes	2:no	+	1	44	2:no	2:no		0.67
18	1:yes	1:yes		0.714	45	2:no	2:no		0.67
19	1:yes	2:no	+	0.8	46	2:no	2:no		0.67
20	1:yes	2:no	+	1	47	1:yes	2:no	+	0.67
21	2:no	1:yes	+	0.633	48	1:yes	2:no	+	0.67
22	2:no	1:yes	+	0.75	49	2:no	2:no		0.67

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.17 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ (ต่อ)

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	กาทำนาย (prediction)
23	1:yes	1:yes		0.714
24	1:yes	1:yes		0.833
25	1:yes	2:no	+	1
26	1:yes	1:yes		0.875
27	2:no	2:no		1

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	กาทำนาย (prediction)
50	2:no	2:no		0.67
51	1:yes	2:no	+	0.67
52	1:yes	2:no	+	0.67
53	2:no	2:no		0.67
54	1:yes	2:no	+	0.67

จากตารางที่ 4.17 ค่าทำนายที่ทายผิดในช่อง error จะมีเครื่องหมาย + อยู่จำนวน 20 ค่า คือค่าที่ 1, 3, 5, 7, 10, 13, 17, 19, 20, 21, 22, 25, 33, 35, 37, 47, 48, 51, 52 และ 54 โดยที่ค่าที่ 3, 5, 10, 13, 21, 22 และ 33 ไม่อนุมัติบัตรเครดิต (class 2 : no) แต่ทำนายว่าอนุมัติบัตรเครดิต (class 1 : yes) และค่าที่ 1, 7, 17, 19, 20, 25, 35, 37, 47, 48, 51, 52 และ 54 อนุมัติบัตรเครดิต แต่ทำนายว่าไม่อนุมัติบัตรเครดิต

ตารางที่ 4.18 ผลในส่วนของการสรุปผลจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

Correctly Classified Instances	34	62.963 %
Incorrectly Classified Instances	20	37.037 %
Kappa statistic	0.2643	
Mean absolute error	0.4456	
Root mean squared error	0.5249	
Coverage of cases (0.95 level)	92.5926 %	
Total Number of Instances	54	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.18 โดยจากข้อมูลจำนวน 54 คน ทำนายข้อมูลถูกต้องจำนวน 34 คน คิดเป็น 62.963% และทำนายข้อมูลไม่ถูกต้องจำนวน 20 คน คิดเป็น 37.037% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.2643 แสดงว่าข้อมูลมีความสอดคล้องกันค่อนข้างน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.4456 ซึ่งมีค่าปานกลาง แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.5249)^2 = 0.1446$ ซึ่งมีค่าค่อนข้างน้อย แสดงว่าตัวแบบมีความถูกต้องพอสมควร

ตารางที่ 4.19 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.536	0.269	0.682	0.536	0.6	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.731	0.464	0.594	0.731	0.655	no
Weighted Avg.	0.63	0.363	0.639	0.63	0.627	

จากตารางที่ 4. 19 สำหรับคำตอบอนุมัติบัตรเครดิต (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.536 ค่าอัตราความผิดพลาดเชิงบวก = 0.269 ค่าความแม่นยำ = 0.682 ค่าความระลึก = 0.536 และค่าความถ่วงดุล = 0.6 ส่วนคำตอบไม่อนุมัติบัตรเครดิต (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.731 ค่าอัตราความผิดพลาดเชิงลบ = 0.464 ค่าความแม่นยำ = 0.594 ค่าความระลึก = 0.731 และค่าความถ่วงดุล = 0.655

ตารางที่ 4.20 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

		ผลการจำแนกอนุมัติบัตรเครดิต	
		อนุมัติ	ไม่อนุมัติ
ค่าที่แท้จริง	อนุมัติ	15	13
	ไม่อนุมัติ	7	19

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.20 มีข้อมูล 54 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 34 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าอนุมัติบัตรเครดิต 15 คน และจำนวนข้อมูลที่จำแนกถูกว่าไม่อนุมัติบัตรเครดิต 19 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 20 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วไม่อนุมัติบัตรเครดิต 7 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่อนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วอนุมัติบัตรเครดิต 13 คน

4.1.3 วิธีโครงข่ายประสาท (Neural Network Method)

วิธีโครงข่ายประสาทใช้อัลกอริทึมชนิดเพอร์เซปตรอนแบบหลายชั้น (Multilayer Perception) ได้ผลการทำนายการพิจารณาอนุมัติบัตรเครดิตดังนี้

4.1.3.1 การสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต

สร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิตจากชุดข้อมูลฝึกหัด ด้วยไฟล์ CreditApproval.csv จำนวน 365 คน ได้ผลดังนี้

ตารางที่ 4.21 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท

Correctly Classified Instances	214	58.6301 %
Incorrectly Classified Instances	151	41.3699 %
Kappa statistic	0.1676	
Mean absolute error	0.4293	
Root mean squared error	0.565	
Relative absolute error	86.0392 %	
Root relative squared error	113.1115 %	
Coverage of cases (0.95 level)	91.2329 %	
Mean rel. region size (0.95 level)	85.7534 %	
Total Number of Instances	365	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตางรางที่ 4.21 จากข้อมูลจำนวน 365 คน ทำนายข้อมูลถูกต้องจำนวน 214 คน คิดเป็น 58.6301% และทำนายข้อมูลไม่ถูกต้องจำนวน 151 คน คิดเป็น 41.3699% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.1676 แสดงว่าข้อมูลมีความสอดคล้องค่อนข้างน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.4293 ซึ่งมีค่าปานกลาง แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.565)^2 = 0.3192$ ซึ่งมีค่าค่อนข้างน้อย แสดงว่าตัวแบบมีความถูกต้องพอสมควร

ตารางที่ 4.22 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.523	0.356	0.572	0.523	0.547	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.644	0.477	0.597	0.644	0.62	no
Weighted Avg.	0.586	0.419	0.585	0.586	0.585	

จากตารางที่ 4.22 สำหรับคำตอบอนุมัติบัตรเครดิต (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.523 , ค่าอัตราความผิดพลาดเชิงบวก = 0.356, ค่าความแม่นยำ = 0.572, ค่าความระลึก = 0.523 และค่าความถ่วงดุล = 0.547 ส่วนคำตอบไม่อนุมัติบัตรเครดิต (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.644, ค่าอัตราความผิดพลาดเชิงลบ = 0.477, ค่าความแม่นยำ = 0.597, ค่าความระลึก = 0.644 และค่าความถ่วงดุล = 0.62

ตารางที่ 4.23 ผลในส่วนของเมตริกซ์ความสับสนจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท

		ผลการจำแนกอนุมัติบัตรเครดิต	
		อนุมัติ	ไม่อนุมัติ
ค่าที่แท้จริง	อนุมัติ	91	83
	ไม่อนุมัติ	68	123

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.23 มีข้อมูล 365 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 214 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าอนุมัติบัตรเครดิต 91 คน และจำนวนข้อมูลที่จำแนกถูกว่าไม่อนุมัติบัตรเครดิต 123 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 151 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วไม่อนุมัติบัตรเครดิต 68 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่อนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วอนุมัติบัตรเครดิต 83 คน

4.1.3.2 การทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต

ทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิตจากชุดข้อมูลทดสอบประสิทธิภาพจากไฟล์ CreditApproval-eval.csv จำนวน 100 คน ได้ผลดังนี้

ตารางที่ 4.24 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท

Correctly Classified Instances	65	65 %
Incorrectly Classified Instances	35	35 %
Kappa statistic	0.2955	
Mean absolute error	0.3517	
Root mean squared error	0.5396	
Coverage of cases (0.95 level)	83 %	
Total Number of Instances	100	

จากตารางที่ 4.24 จากข้อมูลจำนวน 100 คน ทำนายข้อมูลถูกต้องจำนวน 65 คน คิดเป็น 65% และทำนายข้อมูลไม่ถูกต้องจำนวน 35 คน คิดเป็น 35% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.2955 แสดงว่าข้อมูลมีความสอดคล้องกันค่อนข้างน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.3517 ซึ่งมีค่าค่อนข้างน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีความค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.5396)^2 = 0.2912$ ซึ่งมีค่าค่อนข้างน้อย แสดงว่าตัวแบบมีความถูกต้องพอสมควร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.25 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท

	TP-Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.49	0.196	0.706	0.490	0.578	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.804	0.510	0.621	0.804	0.701	no
Weighted Avg.	0.65	0.356	0.663	0.65	0.641	

จากตารางที่ 4.25 สำหรับคำตอบอนุมัติบัตรเครดิต (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.49 ค่าอัตราความผิดพลาดเชิงบวก = 0.196 ค่าความแม่นยำ = 0.706 ค่าความระลึก = 0.490 และค่าความถ่วงดุล = 0.578 ส่วนคำตอบไม่อนุมัติบัตรเครดิต (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.804 ค่าอัตราความผิดพลาดเชิงลบ = 0.510 ค่าความแม่นยำ = 0.621 ค่าความระลึก = 0.804 และค่าความถ่วงดุล = 0.701

ตารางที่ 4.26 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท

		ผลการจำแนกอนุมัติบัตรเครดิต	
		อนุมัติ	ไม่อนุมัติ
ค่าที่แท้จริง	อนุมัติ	24	25
	ไม่อนุมัติ	10	41

จากตารางที่ 4.26 มีข้อมูล 100 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 65 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าอนุมัติบัตรเครดิต 24 คน และจำนวนข้อมูลที่จำแนกถูกว่าไม่อนุมัติบัตรเครดิต 41 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 35 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วไม่อนุมัติบัตรเครดิต 10 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่อนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วอนุมัติบัตรเครดิต 25 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.3.3 การทำนายการพิจารณาอนุมัติบัตรเครดิต

การทำนายการพิจารณาอนุมัติบัตรเครดิตจากชุดข้อมูลทดสอบ ด้วยไฟล์ CreditApproval-test.csv จำนวน 54 คน ได้ผลดังนี้

$$n_{hidden} \le n_{h\,max} = \frac{n_{data\ set} \times n_{input}}{n_{input} + n_{output}}$$
$$= \frac{54 \times 15}{15+2}$$
$$= 47.647 \approx 48$$

จะได้จำนวนโหนดมากที่สุดในชั้นซ่อนประมาณ 48 โหนด โดยในที่นี้จะกำหนดจำนวนโหนดในชั้นซ่อนโดยแบ่งเป็นช่วงละ 5 คือ 5, 10, 15, 20, 25, 30, 35, 40, 45 และ 50 กำหนดค่าอัตราการเรียนรู้เป็น 0.1, 0.2, 0.3, 0.4 และ 0.5 ค่าโมเมนตัมเป็น 0.5, 0.6, 0.7, 0.8 และ 0.9 และใช้จำนวนรอบการฝึกหัดคือ 20,000 รอบ

ตารางที่ 4.27 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท 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
15-5-2	0.4017	0.3442	0.2941	0.3139	0.2504
15-10-2	0.2630	0.2874	0.3077	0.2872	0.2802
15-15-2	0.2586	0.3146	0.3028	0.2833	0.2778
15-20-2	0.3785	0.3010	0.2836	0.2783	0.2783
15-25-2	0.4093	0.3142	0.3783	0.3106	0.3025
15-30-2	0.3464	0.3636	0.4012	0.3130	0.2474
15-35-2	0.2614	0.3284	0.3163	0.3196	0.5170
15-40-2	0.4721	0.3279	0.2742	0.2931	0.2424
15-45-2	0.3505	0.4619	0.2696	0.2829	0.4815
15-50-2	0.2798	0.3505	0.3626	0.3476	0.4812

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.27 เมื่อกำหนดค่าอัตราการเรียนรู้เท่ากับ 0.1 และค่าโมเมนตัมเท่ากับ 0.9 จะพบว่าโครงข่ายประสาทที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (MSE = 0.2424) คือโครงข่ายประสาท 15-40-2 ซึ่งเป็นโครงข่ายประสาทที่มีโหนดข้อมูลเข้าเท่ากับ 15 โหนด โหนดซ่อนเท่ากับ 40 โหนด และโหนดข้อมูลออกเท่ากับ 2 โหนด

ตารางที่ 4.28 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท 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
15-5-2	0.3259	0.3175	0.2518	0.2622	0.2784
15-10-2	0.3046	0.2852	0.2535	0.2496	0.268
15-15-2	0.3191	0.2911	0.2995	0.2516	0.4989
15-20-2	0.2557	0.2668	0.2264	0.2515	0.2500
15-25-2	0.2924	0.2579	0.2873	0.2764	0.2500
15-30-2	0.2550	0.2907	0.2366	0.4749	0.4811
15-35-2	0.2288	0.2616	0.3038	0.2494	0.5185
15-40-2	0.2822	0.3336	0.3582	0.2332	0.4815
15-45-2	0.2964	0.2757	0.2362	0.2369	0.4815
15-50-2	0.3259	0.3086	0.3645	0.2377	0.2500

จากตารางที่ 4.28 เมื่อกำหนดค่าอัตราการเรียนรู้เท่ากับ 0.2 และค่าโมเมนตัมเท่ากับ 0.7 จะพบว่าโครงข่ายประสาทที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (MSE = 0.2264) คือโครงข่ายประสาท 15-20-2 ซึ่งเป็นโครงข่ายประสาทที่มีโหนดข้อมูลเข้าเท่ากับ 15 โหนด โหนดซ่อนเท่ากับ 20 โหนด และโหนดข้อมูลออกเท่ากับ 2 โหนด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.29 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท 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
15-5-2	0.2819	0.2617	0.2550	0.2561	0.2738
15-10-2	0.2390	0.1925	0.2123	0.2632	0.3453
15-15-2	0.2861	0.2139	0.2417	0.2596	0.2500
15-20-2	0.3037	0.2354	0.4230	0.4834	0.2500
15-25-2	0.2437	0.2521	0.2377	0.4816	0.2500
15-30-2	0.2357	0.3007	0.4555	0.2729	0.2500
15-35-2	0.2585	0.2268	0.4606	0.3824	0.2844
15-40-2	0.2712	0.3193	0.3915	0.2499	0.4815
15-45-2	0.2834	0.3104	0.4639	0.4876	0.4815
15-50-2	0.2855	0.2864	0.2449	0.2500	0.2500

จากตารางที่ 4.29 เมื่อกำหนดค่าอัตราการเรียนรู้เท่ากับ 0.3 และค่าโมเมนตัมเท่ากับ 0.6 จะพบว่าโครงข่ายประสาทที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (MSE = 0.1925) คือ โครงข่ายประสาท 15-10-2 ซึ่งเป็นโครงข่ายประสาทที่มีโหนดข้อมูลเข้าเท่ากับ 15 โหนด โหนดซ่อนเท่ากับ 10 โหนด และโหนดข้อมูลออกเท่ากับ 2 โหนด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.30 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท 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
15-5-2	0.2503	0.2239	0.2535	0.2635	0.2784
15-10-2	0.2570	0.2044	0.2429	0.2645	0.3216
15-15-2	0.2588	0.2530	0.2472	0.2641	0.4679
15-20-2	0.3028	0.2769	0.2411	0.2611	0.4145
15-25-2	0.2122	0.2524	0.4853	0.4962	0.4282
15-30-2	0.3430	0.2638	0.4464	0.3105	0.5185
15-35-2	0.3121	0.2379	0.2984	0.4740	0.2500
15-40-2	0.2792	0.4187	0.4875	0.4963	0.5185
15-45-2	0.2787	0.3284	0.3909	0.5184	0.2376
15-50-2	0.3703	0.4160	0.3947	0.2497	0.2500

จากตารางที่ 4.30 เมื่อกำหนดค่าอัตราการเรียนรู้เท่ากับ 0.4 และค่าโมเมนตัมเท่ากับ 0.6 จะพบว่าโครงข่ายประสาทที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (MSE = 0.2044) คือโครงข่ายประสาท 15-10-2 ซึ่งเป็นโครงข่ายประสาทที่มีโหนดข้อมูลเข้าเท่ากับ 15 โหนด โหนดซ่อนเท่ากับ 10 โหนด และโหนดข้อมูลออกเท่ากับ 2 โหนด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.31 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท 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
15-5-2	0.2444	0.2451	0.2532	0.2619	0.4815
15-10-2	0.2521	0.2210	0.2483	0.2622	0.4967
15-15-2	0.2627	0.2470	0.2527	0.262	0.4815
15-20-2	0.2724	0.2889	0.2483	0.2636	0.25
15-25-2	0.3022	0.2558	0.4744	0.3325	0.4815
15-30-2	0.2952	0.2991	0.2516	0.2499	0.4815
15-35-2	0.3315	0.3692	0.4456	0.5185	0.2500
15-40-2	0.3192	0.3931	0.2497	0.4815	0.4815
15-45-2	0.2630	0.4086	0.2492	0.5185	0.5185
15-50-2	0.3551	0.3512	0.3531	0.2500	0.4815

จากตารางที่ 4.31 เมื่อกำหนดค่าอัตราการเรียนรู้เท่ากับ 0.5 และค่าโมเมนตัมเท่ากับ 0.6 จะพบว่าโครงข่ายประสาทที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (MSE = 0.2210) คือโครงข่ายประสาท 15-10-2 ซึ่งเป็นโครงข่ายประสาทที่มีโหนดข้อมูลเข้าเท่ากับ 15 โหนด โหนดซ่อนเท่ากับ 10 โหนด และโหนดข้อมูลออกเท่ากับ 2 โหนด

จากการเปรียบเทียบทั้งหมด เมื่อกำหนดค่าอัตราการเรียนรู้เท่ากับ 0.3 และค่าโมเมนตัมเท่ากับ 0.6 จะพบว่าโครงข่ายประสาทที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (MSE = 0.1925) คือโครงข่ายประสาท 15-10-2 ซึ่งเป็นโครงข่ายประสาทที่มีโหนดข้อมูลเข้าเท่ากับ 15 โหนด โหนดซ่อนเท่ากับ 10 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	2:no	+	0.983
2	2:no	2:no		0.96
3	2:no	1:yes	+	0.511
4	2:no	2:no		0.531
5	2:no	1:yes	+	0.513
6	1:yes	1:yes		0.512
7	1:yes	1:yes		0.501
8	1:yes	1:yes		0.511
9	2:no	1:yes	+	0.511
10	2:no	1:yes	+	0.511
11	1:yes	2:no	+	0.506
12	2:no	2:no		0.96
13	2:no	1:yes	+	0.511
14	1:yes	1:yes		0.997
15	1:yes	1:yes		0.511
16	1:yes	1:yes		0.517
17	1:yes	1:yes		0.511
18	1:yes	1:yes		0.599
19	1:yes	2:no	+	0.532
20	1:yes	1:yes		0.511
21	2:no	1:yes	+	0.511
22	2:no	1:yes	+	0.511
23	1:yes	1:yes		0.519
24	1:yes	1:yes		0.955
25	1:yes	1:yes		0.514
26	1:yes	1:yes		0.512
27	2:no	1:yes	+	0.508

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
28	1:yes	1:yes		0.987
29	1:yes	1:yes		0.998
30	1:yes	1:yes		0.512
31	1:yes	1:yes		0.966
32	1:yes	1:yes		0.511
33	2:no	1:yes	+	0.511
34	2:no	2:no		1
35	1:yes	1:yes		0.501
36	2:no	2:no		0.927
37	1:yes	1:yes		0.511
38	2:no	2:no		1
39	2:no	2:no		0.997
40	2:no	2:no		0.911
41	2:no	2:no		0.517
42	2:no	2:no		0.573
43	2:no	2:no		0.859
44	2:no	1:yes	+	0.511
45	2:no	2:no		0.976
46	2:no	1:yes	+	0.511
47	1:yes	1:yes		0.511
48	1:yes	1:yes		0.511
49	2:no	2:no		0.985
50	2:no	2:no		0.799
51	1:yes	2:no	+	0.509
52	1:yes	1:yes		0.511
53	2:no	2:no		0.983
54	1:yes	2:no	+	0.903

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.32 ค่าทำนายที่ทายผิดในช่อง error จะมีเครื่องหมาย + อยู่จำนวน 16 ค่า คือค่าที่ 1, 3, 5, 9, 10, 11, 13, 19, 21, 22, 27, 33, 44, 46, 51 และ 54 โดยที่ค่าที่ 3, 5, 9, 10, 13, 21, 22, 27, 33, 44 และ 46 ไม่อนุมัติบัตรเครดิต (class 2 : no) แต่ทำนายว่าอนุมัติบัตรเครดิต (class 1 : yes) และค่าที่ 1, 11, 19, 51 และ 54 อนุมัติบัตรเครดิต แต่ทำนายว่าไม่อนุมัติบัตรเครดิต

ตารางที่ 4.33 ผลในส่วนของการสรุปผลจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท

Correctly Classified Instances	38	70.3704 %
Incorrectly Classified Instances	16	29.6296 %
Kappa statistic	0.4017	
Mean absolute error	0.3683	
Root mean squared error	0.4387	
Relative absolute error	73.5309 %	
Root relative squared error	87.5061 %	
Coverage of cases (0.95 level)	98.1481 %	
Mean rel. region size (0.95 level)	87.037 %	
Total Number of Instances	54	

จากตารางที่ 4.33 จากข้อมูลจำนวน 54 คน ทำนายข้อมูลถูกต้องจำนวน 38 คน คิดเป็น 70.3704% และทำนายข้อมูลไม่ถูกต้องจำนวน 16 คน คิดเป็น 29.6296% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.4017 แสดงว่าข้อมูลมีความสอดคล้องกันปานกลาง มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.3683 ซึ่งมีค่าค่อนข้างน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.4387)^2 = 0.1925$ ซึ่งมีค่าค่อนข้างน้อย แสดงว่าตัวแบบมีความถูกต้องพอสมควร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.34 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.821	0.423	0.676	0.821	0.742	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.577	0.179	0.75	0.577	0.652	no
Weighted Avg.	0.704	0.305	0.712	0.704	0.699	

จากตารางที่ 4.34 สำหรับคำตอบอนุมัติบัตรเครดิต (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.821 ค่าอัตราความผิดพลาดเชิงบวก = 0.423 ค่าความแม่นยำ = 0.676 ค่าความระลึก = 0.821 และค่าความถ่วงดุล = 0.742 ส่วนคำตอบไม่อนุมัติบัตรเครดิต (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.577 ค่าอัตราความผิดพลาดเชิงลบ = 0.179 ค่าความแม่นยำ = 0.75 ค่าความระลึก = 0.577 และค่าความถ่วงดุล = 0.652

ตารางที่ 4.35 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีโครงข่ายประสาท

		ผลการจำแนกอนุมัติบัตรเครดิต	
		อนุมัติ	ไม่อนุมัติ
ค่าที่แท้จริง	อนุมัติ	23	5
	ไม่อนุมัติ	11	15

จากตารางที่ 4.35 มีข้อมูล 54 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 38 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าอนุมัติบัตรเครดิต 23 คน และจำนวนข้อมูลที่จำแนกถูกว่าไม่อนุมัติบัตรเครดิต 15 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 16 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วไม่อนุมัติบัตรเครดิต 11 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่อนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วอนุมัติบัตรเครดิต 5 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.4 วิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine Method)

วิธีซัพพอร์ตเวกเตอร์แมชชีนนั้นมีฟังก์ชันหลายฟังก์ชัน แต่ในงานวิจัยนี้ใช้อัลกอริทึม SMO ชนิดโพลิโนเมียลเคอร์เนล (Polynomial Kernel) ในการทำนายการพิจารณาอนุมัติบัตรเครดิต ได้ผลดังนี้

4.1.4.1 การสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต

สร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิตจากชุดข้อมูลฝึกหัด ด้วยไฟล์ CreditApproval.csv จำนวน 365 คน ได้ผลดังนี้

ตารางที่ 4.36 ผลในส่วนของการสรุปผล จากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

Correctly Classified Instances	247	67.6712 %
Incorrectly Classified Instances	118	32.3288 %
Kappa statistic	0.3507	
Mean absolute error	0.3233	
Root mean squared error	0.5686	
Relative absolute error	64.795 %	
Root relative squared error	113.8361 %	
Coverage of cases (0.95 level)	67.6712 %	
Mean rel. region size (0.95 level)	50 %	
Total Number of Instances	365	

จากตารางที่ 4.36 จากข้อมูลจำนวน 365 คน ทำนายข้อมูลถูกต้องจำนวน 247 คน คิดเป็น 67.6712% และทำนายข้อมูลไม่ถูกต้องจำนวน 118 คน คิดเป็น 32.3288% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.3507 แสดงว่าข้อมูลมีความสอดคล้องกันค่อนข้างน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.3233 ซึ่งมีค่าค่อนข้างน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.5686)^2 = 0.3233$ ซึ่งมีค่าค่อนข้างน้อย แสดงว่าตัวแบบมีความถูกต้องพอสมควร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.37 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.638	0.288	0.669	0.638	0.653	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.712	0.362	0.683	0.712	0.697	no
Weighted Avg.	0.677	0.327	0.676	0.677	0.676	

จากตารางที่ 4. 37 สำหรับคำตอบอนุมัติบัตรเครดิต (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.638 ค่าอัตราความผิดพลาดเชิงบวก = 0.288 ค่าความแม่นยำ = 0.669 ค่าความระลึก = 0.638 และค่าความถ่วงดุล = 0.653 ส่วนคำตอบไม่อนุมัติบัตรเครดิต (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.712 ค่าอัตราความผิดพลาดเชิงลบ = 0.362 ค่าความแม่นยำ = 0.683 ค่าความระลึก = 0.712 และค่าความถ่วงดุล = 0.697

ตารางที่ 4.38 ผลในส่วนของเมตริกซ์ความสับสนจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

		ผลการจำแนกอนุมัติบัตรเครดิต	
		อนุมัติ	ไม่อนุมัติ
ค่าที่แท้จริง	อนุมัติ	111	63
	ไม่อนุมัติ	55	136

จากตารางที่ 4.38 มีข้อมูล 365 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 247 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าอนุมัติบัตรเครดิต 111 คน และจำนวนข้อมูลที่จำแนกถูกว่าไม่อนุมัติบัตรเครดิต 136 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 118 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วไม่อนุมัติบัตรเครดิต 55 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่อนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วอนุมัติบัตรเครดิต 63 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.4.2 การทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต

ทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิตจากชุดข้อมูลทดสอบประสิทธิภาพ ด้วยไฟล์ CreditApproval-eval.csv จำนวน 100 คน ได้ผลดังนี้

ตารางที่ 4.39 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

Correctly Classified Instances	65	65 %
Incorrectly Classified Instances	35	35 %
Kappa statistic	0.2972	
Mean absolute error	0.35	
Root mean squared error	0.5916	
Coverage of cases (0.95 level)	65 %	
Total Number of Instances	100	

จากตารางที่ 4.39 จากข้อมูลจำนวน 100 คน ทำนายข้อมูลถูกต้องจำนวน 65 คน คิดเป็น 65% และทำนายข้อมูลไม่ถูกต้องจำนวน 35 คน คิดเป็น 35% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.2972 แสดงว่าข้อมูลมีความสอดคล้องกันค่อนข้างน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.35 ซึ่งมีค่าค่อนข้างน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.5916)^2 = 0.35$ ซึ่งมีค่าค่อนข้างน้อย แสดงว่าตัวแบบมีความถูกต้องพอสมควร

ตารางที่ 4.40 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.551	0.255	0.675	0.551	0.607	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.745	0.449	0.633	0.745	0.685	no
Weighted Avg.	0.650	0.354	0.654	0.650	0.646	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4. 40 สำหรับคำตอบอนุมัติบัตรเครดิต (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.551 ค่าอัตราความผิดพลาดเชิงบวก = 0.255 ค่าความแม่นยำ = 0.675 ค่าความระลึก = 0.551 และค่าความถ่วงดุล = 0.607 ส่วนคำตอบไม่อนุมัติบัตรเครดิต (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.745 ค่าอัตราความผิดพลาดเชิงลบ = 0.449 ค่าความแม่นยำ = 0.633 ค่าความระลึก = 0.745 และค่าความถ่วงดุล = 0.685

ตารางที่ 4.41 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

		ผลการจำแนกอนุมัติบัตรเครดิต	
		อนุมัติ	ไม่อนุมัติ
ค่าที่แท้จริง	อนุมัติ	27	22
	ไม่อนุมัติ	13	38

จากตารางที่ 4.41 มีข้อมูล 100 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 65คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าอนุมัติบัตรเครดิต 27 คน และจำนวนข้อมูลที่จำแนกถูกว่าไม่อนุมัติบัตรเครดิต 38 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 35 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วไม่อนุมัติบัตรเครดิต 13 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่อนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วอนุมัติบัตรเครดิต 22 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.4.3 การทำนายการพิจารณาอนุมัติบัตรเครดิต

การทำนายการพิจารณาอนุมัติบัตรเครดิตจากชุดข้อมูลทดสอบ ด้วยไฟล์ CreditApproval-test.csv จำนวน 54 คน ได้ผลดังนี้

ตารางที่ 4.42 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
1	1:yes	2:no	+	1
2	2:no	2:no		1
3	2:no	2:no		1
4	2:no	2:no		1
5	2:no	1:yes	+	1
6	1:yes	1:yes		1
7	1:yes	2:no	+	1
8	1:yes	1:yes		1
9	2:no	2:no		1
10	2:no	1:yes	+	1
11	1:yes	2:no	+	1
12	2:no	2:no		1
13	2:no	1:yes	+	1
14	1:yes	1:yes		1
15	1:yes	2:no	+	1
16	1:yes	2:no	+	1
17	1:yes	2:no	+	1
18	1:yes	1:yes		1
19	1:yes	2:no	+	1
20	1:yes	1:yes		1
21	2:no	1:yes	+	1
22	2:no	1:yes	+	1
23	1:yes	1:yes		1

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
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	2:no	1:yes	+	1
34	2:no	2:no		1
35	1:yes	2:no	+	1
36	2:no	2:no		1
37	1:yes	1:yes		1
38	2:no	2:no		1
39	2:no	2:no		1
40	2:no	2:no		1
41	2:no	1:yes	+	1
42	2:no	2:no		1
43	2:no	2:no		1
44	2:no	2:no		1
45	2:no	2:no		1
46	2:no	1:yes	+	1
47	1:yes	1:yes		1
48	1:yes	1:yes		1
49	2:no	2:no		1
50	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)
24	1:yes	1:yes		1
25	1:yes	2:no	+	1
26	1:yes	2:no	+	1
27	2:no	2:no		1

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
51	1:yes	2:no	+	1
52	1:yes	2:no	+	1
53	2:no	2:no		1
54	1:yes	2:no	+	1

จากตารางที่ 4.42 ค่าทำนายที่ทายผิดในช่อง error จะมีเครื่องหมาย + อยู่จำนวน 21 ค่า คือค่าที่ 1, 5, 7, 10, 11, 13, 15, 16, 17, 19, 21, 22, 25, 26, 33, 35, 41, 46, 51 และ 52 โดยที่ค่าที่ 5, 10, 13, 21, 22, 33, 41 และ 46 ไม่อนุมัติบัตรเครดิต (class 2 : no) แต่ทำนายว่าอนุมัติบัตรเครดิต (class 1 : yes) และค่าที่ 1, 7, 11, 15, 16, 17, 19, 25, 26, 35, 51, 52 และ 54 อนุมัติบัตรเครดิต แต่ทำนายว่าไม่อนุมัติบัตรเครดิต

ตารางที่ 4.43 ผลในส่วนของการสรุปผลจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

Correctly Classified Instances	33	61.1111 %
Incorrectly Classified Instances	21	38.8889 %
Kappa statistic	0.2265	
Mean absolute error	0.3889	
Root mean squared error	0.6236	
Coverage of cases (0.95 level)	61.1111 %	
Total Number of Instances	54	

จากตารางที่ 4.43 โดยจากข้อมูลจำนวน 54 คน ทำนายข้อมูลถูกต้องจำนวน 33 คน คิดเป็น 61.1111% และทำนายข้อมูลไม่ถูกต้องจำนวน 21 คน คิดเป็น 38.8889% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.2265 แสดงว่าข้อมูลมีความสอดคล้องกันค่อนข้างน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.3889 ซึ่งมีค่าค่อนข้างน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.6236)^2 = 0.3889$ ซึ่งมีค่าค่อนข้างน้อย แสดงว่าตัวแบบมีความถูกต้องพอสมควร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.44 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.536	0.308	0.652	0.536	0.588	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.692	0.464	0.581	0.692	0.632	no
Weighted Avg.	0.611	0.383	0.618	0.611	0.609	

จากตารางที่ 4.44 สำหรับคำตอบอนุมัติบัตรเครดิต (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.536 ค่าอัตราความผิดพลาดเชิงบวก = 0.308 ค่าความแม่นยำ = 0.652 ค่าความระลึก = 0.536 และค่าความถ่วงดุล = 0.614 ส่วนคำตอบไม่อนุมัติบัตรเครดิต (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.692 ค่าอัตราความผิดพลาดเชิงลบ = 0.464 ค่าความแม่นยำ = 0.581 ค่าความระลึก = 0.692 และค่าความถ่วงดุล = 0.632

ตารางที่ 4.45 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน

		ผลการจำแนกอนุมัติบัตรเครดิต	
		อนุมัติ	ไม่อนุมัติ
ค่าที่แท้จริง	อนุมัติ	15	13
	ไม่อนุมัติ	8	18

จากตารางที่ 4.45 มีข้อมูล 54 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 33 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าอนุมัติบัตรเครดิต 15 คน และจำนวนข้อมูลที่จำแนกถูกว่าไม่อนุมัติบัตรเครดิต 18 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 21 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วไม่อนุมัติบัตรเครดิต 8 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่อนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วอนุมัติบัตรเครดิต 13 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.5 วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม (Binary Logistic Regression Method)

เนื่องจากข้อมูลการพิจารณาอนุมัติบัตรเครดิตมีตัวแปรอิสระหลายตัวและมีตัวแปรตามเพียงตัวเดียว เป็นตัวแปรเชิงคุณภาพที่มีค่าได้เพียง 2 ค่า คือ อนุมัติและไม่อนุมัติบัตรเครดิต จึงเลือกใช้วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม ได้ผลการทำนายการพิจารณาอนุมัติบัตรเครดิตดังนี้

4.1.5.1 การสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต

สร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิตจากชุดข้อมูลฝึกหัด ด้วยไฟล์ CreditApproval.csv จำนวน 365 คน ได้ผลดังนี้

ตารางที่ 4.46 ผลในส่วนของการสรุปผลจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

Correctly Classified Instances	225	61.6438 %
Incorrectly Classified Instances	140	38.3562 %
Kappa statistic	0.2308	
Mean absolute error	0.3935	
Root mean squared error	0.5951	
Relative absolute error	78.8575 %	
Root relative squared error	119.1393 %	
Coverage of cases (0.95 level)	70.6849 %	
Mean rel. region size (0.95 level)	61.0959 %	
Total Number of Instances	365	

จากตางรางที่ 4.46 จากข้อมูลจำนวน 365 คน ทำนายข้อมูลถูกต้องจำนวน 225 คน คิดเป็น 61.6438% และทำนายข้อมูลไม่ถูกต้องจำนวน 140 คน คิดเป็น 38.3562% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.2308 แสดงว่าข้อมูลมีความสอดคล้องค่อนข้างน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.3935 ซึ่งมีค่าปานกลาง แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.5951)^2 = 0.3541$ ซึ่งมีค่าค่อนข้างน้อย แสดงว่าตัวแบบมีความถูกต้องพอสมควร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.47 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.529	0.361	0.599	0.592	0.595	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.639	0.408	0.632	0.639	0.635	no
Weighted Avg.	0.616	0.386	0.616	0.616	0.616	

จากตารางที่ 4.47 สำหรับคำตอบอนุมัติบัตรเครดิต (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.529 ค่าอัตราความผิดพลาดเชิงบวก = 0.361 ค่าความแม่นยำ = 0.599 ค่าความระลึก = 0.592 และค่าความถ่วงดุล = 0.595 ส่วนคำตอบไม่อนุมัติบัตรเครดิต (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.639 ค่าอัตราความผิดพลาดเชิงลบ = 0.408 ค่าความแม่นยำ = 0.632 ค่าความระลึก = 0.639 และค่าความถ่วงดุล = 0.635

ตารางที่ 4.48 ผลในส่วนของเมตริกซ์ความสับสน จากการสร้างตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

		ผลการจำแนกอนุมัติบัตรเครดิต	
		อนุมัติ	ไม่อนุมัติ
ค่าที่แท้จริง	อนุมัติ	103	71
	ไม่อนุมัติ	69	122

จากตารางที่ 4.48 มีข้อมูล 365 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 225 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าอนุมัติบัตรเครดิต 103 คน และจำนวนข้อมูลที่จำแนกถูกว่าไม่อนุมัติบัตรเครดิต 122 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 140 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วไม่อนุมัติบัตรเครดิต 69 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่อนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วอนุมัติบัตรเครดิต 71 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.5.2 การทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต

ทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิตจากชุดข้อมูลทดสอบประสิทธิภาพด้วยไฟล์ CreditApproval-eval.csv จำนวน 100 คน ได้ผลดังนี้

ตารางที่ 4.49 ผลในส่วนของการสรุปผลจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

Correctly Classified Instances	63	63 %
Incorrectly Classified Instances	37	37 %
Kappa statistic	0.2564	
Mean absolute error	0.3675	
Root mean squared error	0.5692	
Coverage of cases (0.95 level)	75 %	
Total Number of Instances	100	

จากตารางที่ 4.49 จากข้อมูลจำนวน 100 คน ทำนายข้อมูลถูกต้องจำนวน 63 คน คิดเป็น 63% และทำนายข้อมูลไม่ถูกต้องจำนวน 37 คน คิดเป็น 37% มีค่าสถิติแคปปา (Kappa statistic) คือ 0.2564 แสดงว่าข้อมูลมีความสอดคล้องกันค่อนข้างน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.3675 ซึ่งมีค่าค่อนข้างน้อย แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.5692)^2 = 0.3239$ ซึ่งมีค่าค่อนข้างน้อย แสดงว่าตัวแบบมีความถูกต้องพอสมควร

ตารางที่ 4.50 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.51	0.255	0.658	0.51	0.575	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.745	0.49	0.613	0.745	0.673	no
Weighted Avg.	0.63	0.375	0.635	0.63	0.625	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.50 สำหรับคำตอบอนุมัติบัตรเครดิต (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.51 ค่าอัตราความผิดพลาดเชิงบวก = 0.255 ค่าความแม่นยำ = 0.658 ค่าความระลึก = 0.51 และค่าความถ่วงดุล = 0.575 ส่วนคำตอบไม่อนุมัติบัตรเครดิต (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.745 ค่าอัตราความผิดพลาดเชิงลบ = 0.49 ค่าความแม่นยำ = 0.613 ค่าความระลึก = 0.745 และค่าความถ่วงดุล = 0.673

ตารางที่ 4.51 ผลในส่วนของเมตริกซ์ความสับสนจากการทดสอบความถูกต้องของตัวแบบการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

		ผลการจำแนกอนุมัติบัตรเครดิต	
		อนุมัติ	ไม่อนุมัติ
ค่าที่แท้จริง	อนุมัติ	25	24
	ไม่อนุมัติ	13	38

จากตารางที่ 4.51 มีข้อมูล 100 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 63 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าอนุมัติบัตรเครดิต 25 คน และจำนวนข้อมูลที่จำแนกถูกว่าไม่อนุมัติบัตรเครดิต 38 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 37 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วไม่อนุมัติบัตรเครดิต 13 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่อนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วอนุมัติบัตรเครดิต 24 คน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.5.3 การทำนายการพิจารณาอนุมัติบัตรเครดิต

การทำนายการพิจารณาอนุมัติบัตรเครดิตจากชุดข้อมูลทดสอบ ด้วยไฟล์ CreditApproval-test.csv จำนวน 54 คน ได้ผลดังนี้

ตารางที่ 4.52 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
1	1:yes	2:no	+	1
2	2:no	2:no		1
3	2:no	2:no		0.925
4	2:no	2:no		1
5	2:no	1:yes	+	0.99
6	1:yes	1:yes		0.97
7	1:yes	2:no	+	1
8	1:yes	2:no	+	0.816
9	2:no	1:yes	+	1
10	2:no	2:no		1
11	1:yes	2:no	+	0.961
12	2:no	2:no		1
13	2:no	2:no		0.93
14	1:yes	1:yes		1
15	1:yes	1:yes		0.99
16	1:yes	2:no	+	1
17	1:yes	2:no	+	1
18	1:yes	1:yes		0.974
19	1:yes	2:no	+	1
20	1:yes	1:yes		0.985
21	2:no	1:yes	+	1
22	2:no	2:no		0.981
23	1:yes	1:yes		0.953

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
28	1:yes	1:yes		0.63
29	1:yes	1:yes		1
30	1:yes	2:no	+	0.849
31	1:yes	1:yes		1
32	1:yes	1:yes		0.805
33	2:no	1:yes	+	0.994
34	2:no	2:no		0.965
35	1:yes	2:no	+	0.771
36	2:no	2:no		1
37	1:yes	1:yes		0.562
38	2:no	2:no		1
39	2:no	2:no		1
40	2:no	1:yes	+	0.999
41	2:no	1:yes	+	0.564
42	2:no	1:yes	+	0.983
43	2:no	1:yes	+	0.999
44	2:no	1:yes	+	1
45	2:no	1:yes	+	0.997
46	2:no	1:yes	+	0.994
47	1:yes	1:yes		0.733
48	1:yes	1:yes		0.545
49	2:no	2:no		0.699
50	2:no	2:no		0.982

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.52 ผลในส่วนของการทำนายชุดข้อมูลทดสอบ ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม (ต่อ)

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน (error)	การทำนาย (prediction)
24	1:yes	1:yes		1
25	1:yes	2:no	+	1
26	1:yes	2:no	+	0.981
27	2:no	2:no		1

ตัวอย่าง (instance)	ค่าที่แท้จริง (actual)	ค่าที่ทำนายได้ (predicted)	ความคลาดเคลื่อน	การทำนาย (prediction)
51	1:yes	2:no	+	0.937
52	1:yes	2:no	+	0.882
53	2:no	2:no		0.838
54	1:yes	2:no	+	0.83

จากตารางที่ 4.52 ค่าทำนายที่ทายผิดในช่อง error จะมีเครื่องหมาย + อยู่จำนวน 25 ค่า คือค่าที่ 1, 5, 7, 8, 9, 11, 16, 17, 19, 21, 25, 26, 30, 33, 35, 40, 41, 42, 43, 44, 45, 46, 51,52 และ 54 โดยที่ค่าที่ 5, 9, 21, 33, 40, 41, 42, 43, 44, 45 และ 46 ไม่อนุมัติบัตรเครดิต (class 2 : no) แต่ทำนายว่าอนุมัติบัตรเครดิต (class 1 : yes) และค่าที่ 1, 7, 8, 11, 16, 17, 19, 25, 26, 30, 35, 51, 52 และ 54 อนุมัติบัตรเครดิต แต่ทำนายว่าไม่อนุมัติบัตรเครดิต

ตารางที่ 4.53 ผลในส่วนของการสรุปผลจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

Correctly Classified Instances	29	53.7037 %
Incorrectly Classified Instances	25	46.2963 %
Kappa statistic	0.0766	
Mean absolute error	0.4830	
Root mean squared error	0.6558	
Coverage of cases (0.95 level)	66.6667%	
Total Number of Instances	54	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.53 โดยจากข้อมูลจำนวน 54 คน ทำนายข้อมูลถูกต้องจำนวน 29 คน คิดเป็น 53.7037% และทำนายข้อมูลไม่ถูกต้องจำนวน 25 คน คิดเป็น 46.2963% มีค่าสถิติแคปป้า (Kappa statistic) คือ 0.0766 แสดงว่าข้อมูลมีความสอดคล้องกันน้อย มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE) คือ 0.4830 ซึ่งมีค่าปานกลาง แสดงว่าค่าที่ทำนายได้ใกล้เคียงกับค่าที่แท้จริงพอสมควร และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) คือ $(0.6558)^2 = 0.4301$ ซึ่งมีค่าค่อนข้างน้อย แสดงว่าตัวแบบมีความถูกต้องพอสมควร

ตารางที่ 4.54 ผลในส่วนของความถูกต้องของรายละเอียดในแต่ละคำตอบจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
	0.5	0.423	0.56	0.5	0.528	yes
	TN Rate	FN Rate	Precision	Recall	F-Measure	Class
	0.577	0.5	0.517	0.577	0.545	no
Weighted Avg.	0.537	0.46	0.539	0.537	0.537	

จากตารางที่ 4.54 สำหรับคำตอบอนุมัติบัตรเครดิต (Class = yes) มีค่าอัตราความถูกต้องเชิงบวก = 0.5 ค่าอัตราความผิดพลาดเชิงบวก = 0.423 ค่าความแม่นยำ = 0.56 ค่าความระลึก = 0.5 และค่าความถ่วงดุล = 0.528 ส่วนคำตอบไม่อนุมัติบัตรเครดิต (Class = no) มีค่าอัตราความถูกต้องเชิงลบ = 0.577 ค่าอัตราความผิดพลาดเชิงลบ = 0.5 ค่าความแม่นยำ = 0.517 ค่าความระลึก = 0.577 และค่าความถ่วงดุล = 0.545

ตารางที่ 4.55 ผลในส่วนของเมตริกซ์ความสับสนจากการทำนายการพิจารณาอนุมัติบัตรเครดิต ด้วยวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม

		ผลการจำแนกอนุมัติบัตรเครดิต	
		อนุมัติ	ไม่อนุมัติ
ค่าที่แท้จริง	อนุมัติ	14	14
	ไม่อนุมัติ	11	15

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.55 มีข้อมูล 54 คน ตัวแบบสามารถทำนายข้อมูลได้ถูกต้อง 29 คน โดยมีจำนวนข้อมูลที่จำแนกถูกว่าอนุมัติบัตรเครดิต 14 คน และจำนวนข้อมูลที่จำแนกถูกว่าไม่อนุมัติบัตรเครดิต 15 คน ตัวแบบทำนายข้อมูลไม่ถูกต้อง 25 คน โดยมีจำนวนข้อมูลที่จำแนกผิดว่าอนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วไม่อนุมัติบัตรเครดิต 11 คน และจำนวนข้อมูลที่จำแนกผิดว่าไม่อนุมัติบัตรเครดิต ซึ่งค่าที่แท้จริงแล้วอนุมัติบัตรเครดิต 14 คน

4.2 ผลการเปรียบเทียบประสิทธิภาพและทำนายผลของวิธีการจำแนกกลุ่ม

จากการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มได้ถูกต้องและความแม่นยำในการทำนายการพิจารณาอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทย ระหว่างวิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาท วิธีซัพพอร์ตเวกเตอร์แมชชีน และวิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม ได้ผลดังตารางที่ 4.56 และ 4.57

ตารางที่ 4.56 ผลการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มในการพิจารณาอนุมัติบัตรเครดิตทั้ง 5 วิธี

การเปรียบเทียบประสิทธิภาพ	ค่าความถูกต้อง (Accuracy)	ค่าอัตราความถูกต้องเชิงบวก (TP Rate)	ค่าอัตราความถูกต้องเชิงลบ (TN Rate)	ค่าความแม่นยำ (Precision)
วิธีความใกล้เคียงกันมากที่สุด	61.11%	0.5	**0.731**	0.667
วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ	62.96%	0.536	**0.731**	**0.682**
วิธีโครงข่ายประสาท	**70.37%**	**0.821**	0.577	0.676
วิธีซัพพอร์ตเวกเตอร์แมชชีน	61.11%	0.536	0.692	0.652
วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม	53.70%	0.5	0.577	0.56

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.56 ผลการเปรียบเทียบประสิทธิภาพของวิธีการจำแนกกลุ่มในการพิจารณาอนุมัติบัตรเครดิต ทั้ง 5 วิธี (ต่อ)

การเปรียบเทียบประสิทธิภาพ	ค่าอัตราความผิดพลาดเชิงบวก (FP Rate)	ค่าอัตราความผิดพลาดเชิงลบ (FN Rate)	ค่าความระลึก (Recall)	ค่าความถ่วงดุล (F-Measure)
วิธีความใกล้เคียงกันมากที่สุด	**0.269**	0.5	0.5	0.571
วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ	**0.269**	0.464	0.536	0.6
วิธีโครงข่ายประสาท	0.423	**0.179**	**0.821**	**0.742**
วิธีซัพพอร์ตเวกเตอร์แมชชีน	0.308	0.464	0.536	0.588
วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม	0.423	0.5	0.5	0.528

ตารางที่ 4.57 ผลการเปรียบเทียบการทำนายผลของวิธีการจำแนกกลุ่มในการพิจารณาอนุมัติบัตรเครดิต ทั้ง 5 วิธี

การเปรียบเทียบการทำนายผล	ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE)	ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE)
วิธีความใกล้เคียงกันมากที่สุด	0.3895	0.3868
วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ	0.4456	0.2755
วิธีโครงข่ายประสาท	**0.3683**	**0.1925**
วิธีซัพพอร์ตเวกเตอร์แมชชีน	0.3889	0.3889
วิธีการถดถอยโลจิสติกแบบ 2 กลุ่ม	0.4830	0.4301

จากตารางที่ 4.56 พบว่าวิธีโครงข่ายประสาทมีค่าความถูกต้อง ค่าอัตราความถูกต้องเชิงบวก ค่าอัตราความผิดพลาดเชิงลบ ค่าความระลึกและค่าความถ่วงดุลดีที่สุดคือ 70.37%, 0.821, 0.179, 0.821 และ 0.742 ตามลำดับ ส่วนวิธีแผนภาพต้นไม้เพื่อการตัดสินใจและวิธีความใกล้เคียงกันมากที่สุดมีค่าอัตราความถูกต้องเชิงลบและค่าอัตราความผิดพลาดเชิงบวกดีที่สุดคือ 0.731 และ 0.269 ตามลำดับ และวิธีแผนภาพต้นไม้เพื่อการตัดสินใจมีค่าความแม่นยำที่ดีที่สุด โดยมีค่าเท่ากับ 0.682 ดังนั้นวิธีโครงข่ายประสาทมีประสิทธิภาพดีที่สุด

จากตารางที่ 4.57 พบว่าวิธีโครงข่ายประสาทมีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุดคือ 0.3683 และ 0.1925 ตามลำดับ แสดงว่ามีความคลาดเคลื่อนน้อยที่สุด ดังนั้นวิธีโครงข่ายประสาทมีการทำนายผลดีที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

ในการทำวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาทำความเข้าใจและเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่ม รวมทั้งเปรียบเทียบการทำนายผลการอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทย วิธีการจำแนกกลุ่มที่นำมาเปรียบเทียบประสิทธิภาพและการทำนายผลคือ วิธีความใกล้เคียงกันมากที่สุด โดยใช้อัลกอริทึมชนิด IBk วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ โดยใช้อัลกอริทึมชนิด J48 วิธีโครงข่ายประสาท โดยใช้อัลกอริทึมชนิดเพอร์เซปตรอนแบบหลายชั้น 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 20,000 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 ค่าอัตราความผิดพลาดเชิงลบ ค่าความระลึกและค่าความถ่วงดุลดีที่สุดคือ 70.37%, 0.821, 0.179, 0.821 และ 0.742 ตามลำดับ ส่วนวิธีแผนภาพต้นไม้เพื่อการตัดสินใจและวิธีความใกล้เคียงกันมากที่สุดมีค่าอัตราความถูกต้องเชิงลบและค่าอัตราความผิดพลาดเชิงบวกดีที่สุดคือ 0.731 และ 0.269 ตามลำดับ และวิธีแผนภาพต้นไม้เพื่อการตัดสินใจมีค่าความแม่นยำดีที่สุดคือ 0.682 ดังนั้นวิธีโครงข่ายประสาทมีประสิทธิภาพดีที่สุด

ส่วนการเปรียบเทียบการทำนายผลของวิธีการจำแนกกลุ่มในการพิจารณาอนุมัติบัตรเครดิต พบว่าวิธีโครงข่ายประสาทมีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยและความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุดคือ 0.3683 และ 0.1925 ตามลำดับ ดังนั้นวิธีโครงข่ายประสาทมีการทำนายผลดีที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 ข้อเสนอแนะ

1. ตัวแปรที่นำมาใช้ในการงานวิจัยนี้เป็นเพียงส่วนหนึ่งของการพิจารณาอนุมัติบัตรเครดิตของธนาคารพาณิชย์ไทยเท่านั้น เพื่อให้การทำนายมีประสิทธิภาพมากขึ้น ควรเพิ่มตัวแปรที่เกี่ยวข้องอื่น ๆ อีก

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

ข้อมูลของการพิจารณาอนุมัติบัตรเครดิต

ตารางที่ ก-1 คุณลักษณะและรายละเอียดของการพิจารณาอนุมัติบัตรเครดิต

คุณลักษณะ	รายละเอียด
1) เพศ (SEX)	ชาย (Male) หญิง (Female)
2) อายุ (AGE)	เป็นตัวเลข (Numeric) หน่วย : ปี
3) ระดับการศึกษา (EDUCATION)	ประถมศึกษา (Primary) มัธยมศึกษาตอนต้น (Secondary) มัธยมศึกษาตอนปลาย/ปวช./เทียบเท่า (High School) อนุปริญญา/ปวส./ปวท./เทียบเท่า (Diploma) ปริญญาตรี (Bachelor) ปริญญาโท (Master) ปริญญาเอก (Doctorate)
4) สถานภาพสมรส (MARRIAGE)	โสด (Single) สมรสจดทะเบียน (Register) สมรสไม่จดทะเบียน (Not Register) หม้าย (Widow) หย่าร้าง (Divorce) แยกกันอยู่ (Separate)
5) จำนวนบุตร (CHILD)	เป็นตัวเลข (Numeric) หน่วย : คน
6) อาชีพ (OCCUPATION)	รับราชการ (Serve) พนักงานรัฐวิสาหกิจ (Enterprise) พนักงานบริษัท (Employees) เจ้าของกิจการ (Owner) พ่อบ้าน/แม่บ้าน (Steward/Housewife) อาชีพอิสระ (Freelance) แพทย์/สัตวแพทย์/ทันตแพทย์ (Physician/Veterinary/Dentist) พยาบาล (Nurse) นักกฎหมาย (Lawyer) ตำรวจ/ทหาร (Police/Soldier)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คุณลักษณะ	รายละเอียด
6) อาชีพ (OCCUPATION) (ต่อ)	นักการเมือง (Politician) นักศึกษา (Student) ครู/อาจารย์ (Teacher) ศิลปิน/นักแสดง (Artist/Player) นักบัญชี (Accountant) นักคอมพิวเตอร์/นักเทคโนโลยีสารสนเทศ (Computer/Technology) ข้าราชการบำนาญ/เกษียณอายุ (Pensioner/Retire) รับจ้าง/พนักงานรายวัน/ชั่วคราว (Hire/Staff/Temporary) เกษตรกรรม/ปศุสัตว์/ประมง (Farm/Livestock/Fishing) วิศวกร/สถาปนิก/มัณฑนากร (Engineer/Architect/Decorate) อื่น ๆ (Other)
7) รายได้ต่อเดือน (INCOME)	เป็นตัวเลข (Numeric) หน่วย : บาท
8) ประเภทธุรกิจในการประกอบการ (BUSINESS)	ข้าราชการ (Serve) สายการบิน (Airline) ขนส่ง (Transport) ธุรกิจโรงแรม (Hotel) ประกันภัย (Insurance) ที่ปรึกษา (Advisor) นำเข้า/ส่งออก (Import/Export) อุตสาหกรรม/โรงงาน (Industry/Plant) การสื่อสาร/โทรคมนาคม/เครื่องมือสื่อสาร (Communication/Telecommunication/Mobile) โรงพิมพ์/สิ่งพิมพ์/บรรจุภัณฑ์ (Printer/Publication/Package) การค้าปลีก/ค้าส่ง/ห้างสรรพสินค้า (Retail/Wholesale/Shopping Mall) อสังหาริมทรัพย์/พัฒนาที่ดิน/ก่อสร้าง (Property/Development/Construction) ปศุสัตว์/ประมง/เกษตรกรรม/โรงสีข้าว (Livestock/Fishing/Farm/Rice Mill) สาธารณูปโภค/ไฟฟ้า/ประปา (Utility/Electricity/Water)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คุณลักษณะ	รายละเอียด
8) ประเภทธุรกิจในการประกอบการ (BUSINESS) (ต่อ)	สถาบันการศึกษา/สถาบันกวดวิชา (Educational Institution/Tutor) สถานีอนามัย/โรงพยาบาล (Health/Hospital) ธนาคาร/สถาบันการเงิน/บริษัทหลักทรัพย์/โรงจำนำ (Bank/Finance/Company/Pawn) กองทัพบก/เรือ/อากาศ/ตำรวจ (Army) บริการ/บันเทิง/ท่องเที่ยว/ร้านอาหาร/ภัตตาคาร (Service/Entertainment/Travel/Restaurant) ยานพาหนะ/อะไหล่ (Vehicle/Parts) เคมีภัณฑ์ (Chemical) ร้านสะดวกซื้อ (Minimart) อุปกรณ์อิเล็กทรอนิกส์/เครื่องใช้ไฟฟ้า (Electronic) ปิโตรเคมี/ผลิตภัณฑ์ (Petrochemical/Products) พลาสติก (Plastic) อัญมณี/เครื่องประดับ (Jewel/Ornament) เครื่องจักร/ผลิตภัณฑ์โลหะ (Machine/Metal Products) วัสดุก่อสร้าง (Material) รับเหมาก่อสร้าง (Construction) สิ่งทอ (Textiles) เฟอร์นิเจอร์/โรงเลื่อย (Furniture/Sawmill) เวชภัณฑ์/คลินิก (Medical Supplies/Clinic) อื่น ๆ (Other)
9) ตำแหน่งหน้าที่ในการทำงาน (POSITION)	พนักงานบัญชี (Accounting) เจ้าของกิจการ (Owner) ครู / อาจารย์ (Teacher) พนักงานสำรองตั๋ว (Booking) เจ้าหน้าที่ธุรการ/ บริหาร (Administration) ผู้ช่วยพยาบาล (Assistant nurse) การตลาด (Marketing) ผู้จัดการ (Manager) ผู้ช่วยผู้จัดการ (Assistant Manager)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คุณลักษณะ	รายละเอียด
9) ตำแหน่งหน้าที่ในการทำงาน (POSITION) (ต่อ)	เลขานุการ (Secretary) พนักงานขาย (Saleman) พนักงานคอมพิวเตอร์ (Computer) พนักงานจัดซื้อ (Purchase) เจ้าหน้าที่บุคคล (Person) หัวหน้างาน (Chief) มัคคุเทศน์/ ล่าม (Guide) แม่บ้าน (Housekeeper) หัวหน้าพ่อครัว (Chef) นักแสดง (Actor) เจ้าหน้าที่วิเคราะห์นโยบาย (Analyst) เจ้าหน้าที่การเงิน (Cashier) เจ้าหน้าที่คุมสินค้า (Stock Product) เจ้าหน้าที่วิจัยทดลอง (Research) นักร้อง / นักดนตรี (Singer) พยาบาล (Nurse) พนักงานตรวจสอบคุณภาพ (Q.C.) พนักงานออกแบบ (Design) พนักงานรับโทรศัพท์ / ประชาสัมพันธ์ (Operator) เจ้าหน้าสินเชื่อ (Loan) สารวัตร (Inspector) นักวิชาการตรวจสอบ/ เจ้าหน้าที่ตรวจสอบ (Audit) พนักงานช่างเครื่อง /ช่างเทคนิค (Auto Machine) นักวิชาการ (Academician) นักสังคมสงเคราะห์ (Social Worker) เจ้าหน้าที่ฝ่ายฝึกอบรม (Training) เจ้าหน้าที่ป้องกันและสาธารณภัย (Public Hazard) เจ้าหน้าที่บริการร้านอาหาร (Steward) สัตวแพทย์ (Veterinarian) มัณฑนากร (Decorate) เภสัชกร (Pharmacist)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คุณลักษณะ	รายละเอียด
10) อายุในการทำงาน (WORKING AGE)	เป็นตัวเลข (Numeric) หน่วย : ปี
11) ประเภทที่อยู่อาศัย (TYPE OF HOUSING)	บ้านเดี่ยว (Detached) ทาวน์เฮาส์/อาคารพาณิชย์ (Townhouse) คอนโดมิเนี่ยม/อพาร์ตเมนท์/แมนชั่น/หอพัก (Condominium/Apartment/Mansion/Dorm)
12) สถานะการอยู่อาศัย (RESIDENCY STATUS)	บ้านของตนเอง/คู่สมรสปลอดภาระ (Non Load Own) บ้านของตนเอง/คู่สมรสไม่ปลอดภาระ (Load Own) บ้านบิดามารดา (Parents) บ้านญาติ/พี่น้อง/บุคคลอื่น (Relative) บ้านพักสวัสดิการ (Welfare) บ้านเช่า (Rent)
13) ประเภทบัตรเครดิตที่ต้องการ (CREDIT TYPE)	บัตรมาสเตอร์ (Master) บัตรวีซ่า (Visa)
14) วิธีการรับรายได้ (GET MONEY)	รับเป็นเงินสด/เช็ค (Cash) เข้าบัญชีธนาคาร (Bank Account)
15) การพิจารณาอนุมัติบัตรเครดิต (APPROVAL)	อนุมัติ (yes) ไม่อนุมัติ (no)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ ก-2 ตัวอย่างข้อมูลการพิจารณาอนุมัติบัตรเครดิต

No	SEX	AGE	EDUCATION	MARRIAGE	CHILD	OCCUPATION	INCOME	BUSINESS
1	female	27	bachelor	single	0	accountant	16,300	other
2	male	26	bachelor	single	0	accountant	15,300	other
3	female	43	diploma	register	0	accountant	16,400	other
4	female	39	bachelor	single	0	owner	150,000	other
5	female	30	master	register	0	serve	15,220	finance
6	female	46	master	register	2	teacher	25,447	institution
7	female	29	diploma	divorce	1	accountant	15,200	travel
8	female	26	bachelor	single	0	employee	15,200	travel
9	female	42	bachelor	register	1	serve	15,585	institution
10	female	50	bachelor	register	1	serve	15,585	institution
11	female	58	bachelor	widow	0	serve	23,393	institution
12	female	40	bachelor	register	1	serve	16,060	institution
13	male	52	bachelor	non register	2	serve	27,000	institution
14	female	24	bachelor	single	0	employee	16,500	shopping
15	female	32	bachelor	single	0	employee	22,500	other
16	female	31	bachelor	single	0	employee	15,000	other
17	female	45	bachelor	single	0	employee	15,000	institution
18	female	36	bachelor	register	1	employee	15,000	clinic
19	male	53	doctorate	divorce	2	serve	30,730	institution
20	female	26	bachelor	single	0	employee	15,300	other
21	female	26	bachelor	single	0	employee	15,300	other
22	female	31	bachelor	single	0	employee	15,000	other
23	male	34	bachelor	single	0	employee	18,400	electronic
24	female	26	bachelor	single	0	accountant	15,500	other
25	female	37	bachelor	register	1	computer	15,300	other
26	female	27	bachelor	single	0	employee	20,000	travel
27	female	28	bachelor	single	0	employee	17,500	other
28	female	23	bachelor	single	0	employee	15,000	other

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

No	POSITION	WORKING AGE	TYPE OF HOUSING	RESIDENCY STATUS	CREDIT TYPE	GET MONEY	APPROVAL
1	account	10	detached	parent	visa	cash	yes
2	account	5	townhouse	rent	visa	cash	yes
3	account	20	detached	parent	visa	cash	no
4	owner	12	detached	non load	visa	cash	yes
5	account	7	detached	non load	visa	bank	yes
6	teacher	13	detached	non load	visa	bank	no
7	account	2	detached	non load	visa	bank	yes
8	booking	2	detached	parent	visa	bank	yes
9	admin	20	detached	welfare	visa	bank	yes
10	admin	20	detached	welfare	visa	bank	yes
11	assistant	39	detached	non load	visa	bank	yes
12	admin	13	detached	non load	visa	bank	yes
13	admin	11	detached	non load	visa	bank	yes
14	marketing	2	apartment	rent	visa	bank	no
15	manager	3.8	detached	parent	visa	bank	no
16	manager	2	detached	relative	visa	bank	yes
17	assistant ma	6	detached	relative	visa	bank	yes
18	secretary	1	detached	parent	visa	bank	yes
19	teacher	30	detached	welfare	visa	bank	yes
20	saleman	2	apartment	welfare	visa	bank	no
21	saleman	2	apartment	welfare	visa	bank	no
22	saleman	2	apartment	welfare	visa	bank	no
23	manager	4	townhouse	non load	visa	bank	no
24	account	7	detached	welfare	visa	bank	yes
25	computer	8	townhouse	relative	visa	bank	yes
26	booking	8	dorm	rent	visa	bank	no
27	manager	4	detached	parent	visa	bank	no
28	account	2	detached	parent	visa	bank	no

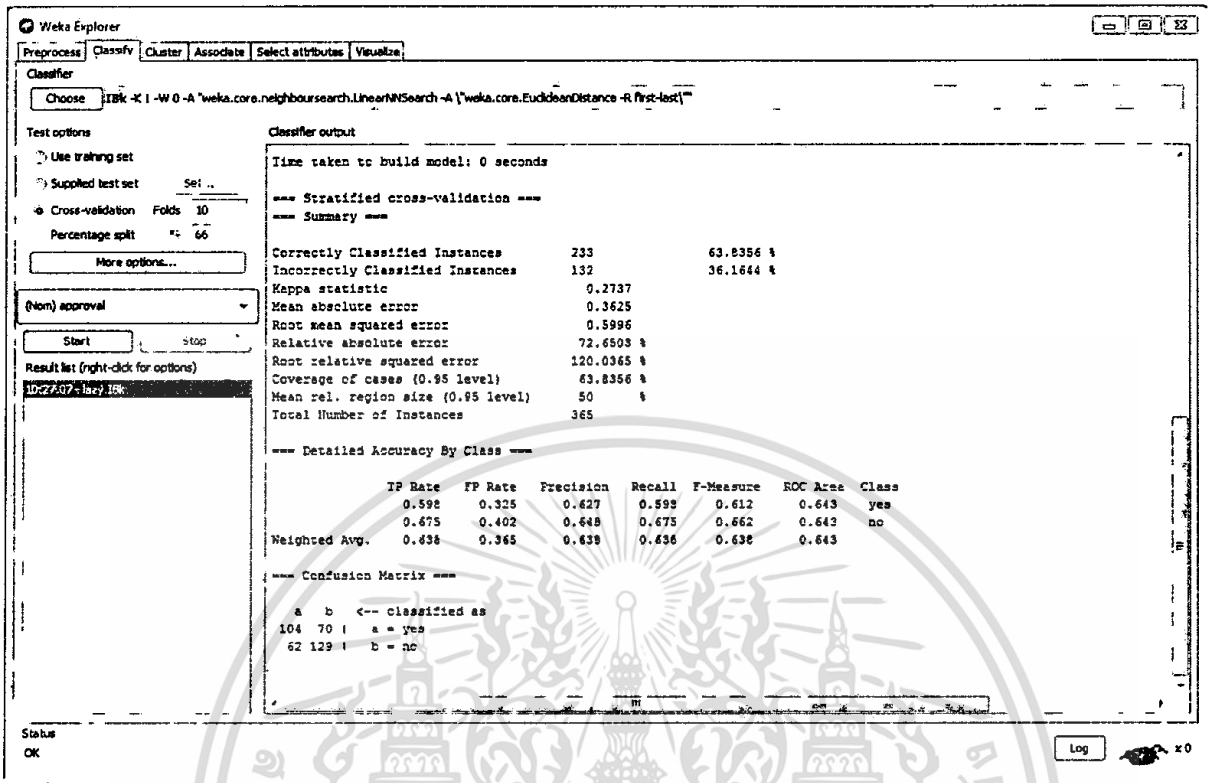
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

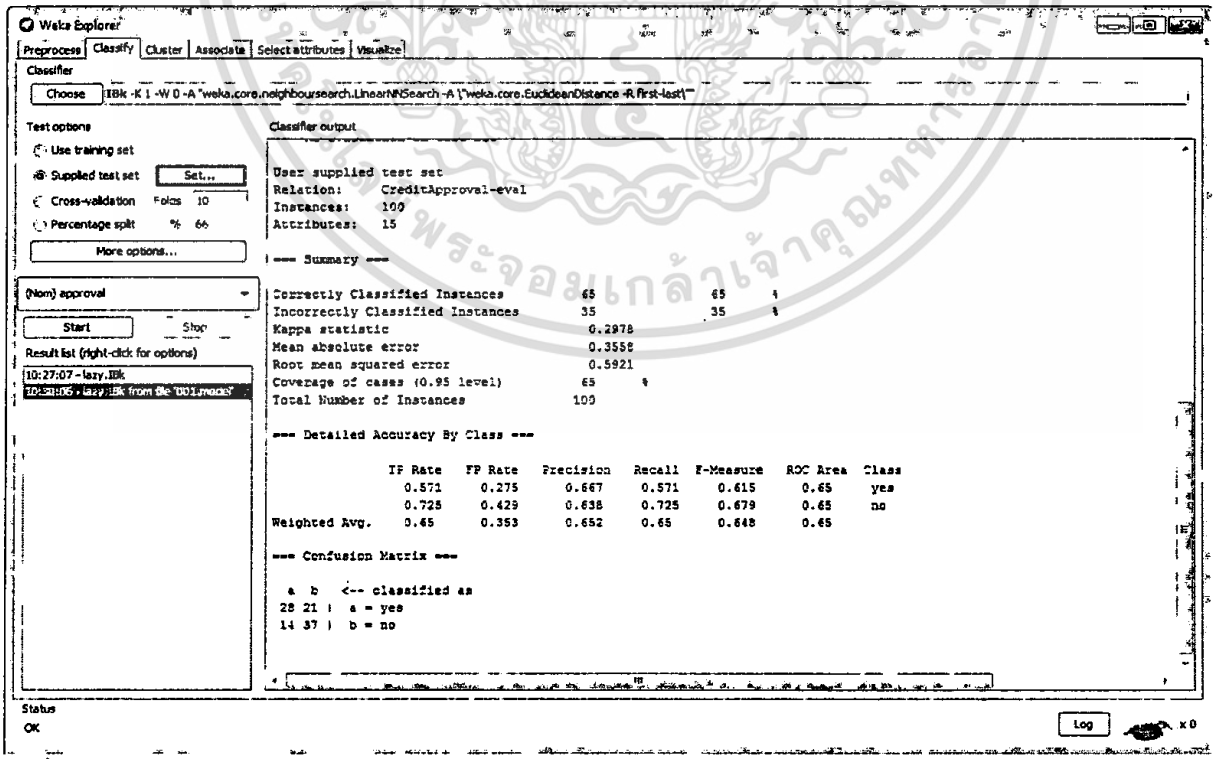
การวิเคราะห์ข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. วิธีความใกล้เคียงกันมากที่สุด (K-Nearest Neighbor : KNN)

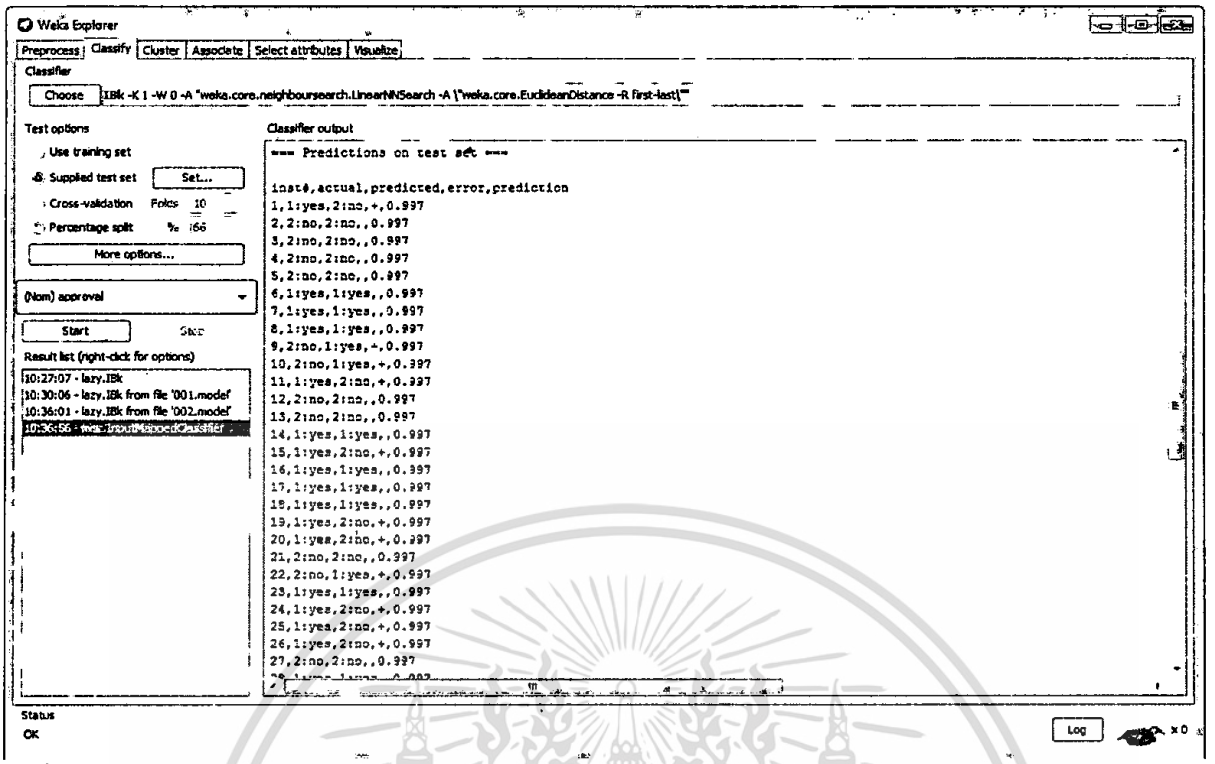


รูปที่ ข-1 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการสร้างตัวแบบโดยวิธีความใกล้เคียงกันมากที่สุด

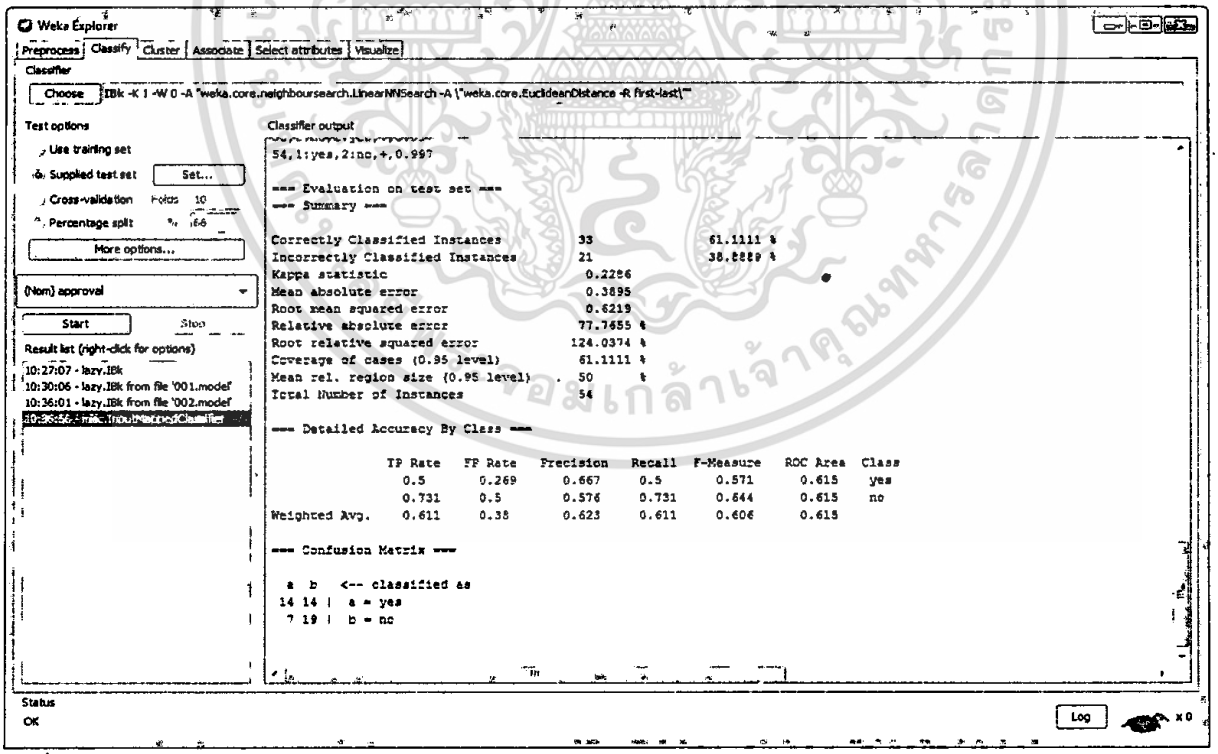


รูปที่ ข-2 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการทดสอบตัวแบบโดยวิธีความใกล้เคียงกันมากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



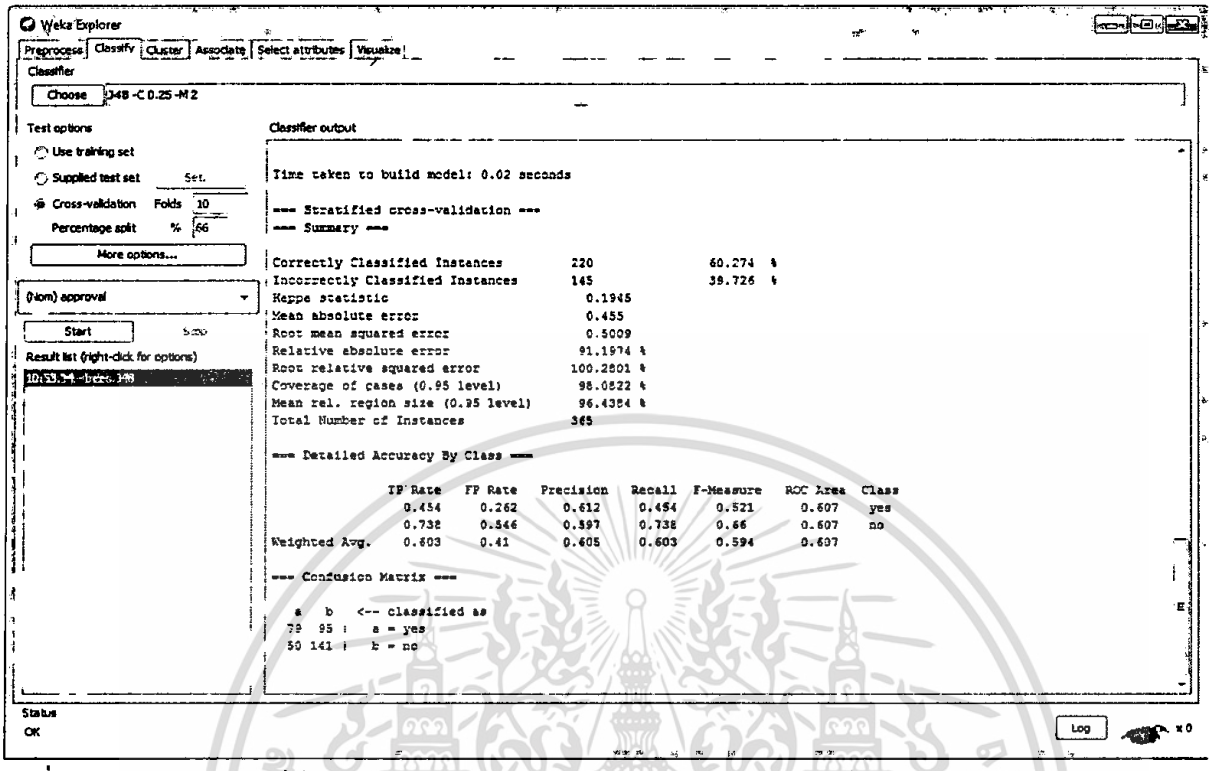
รูปที่ ข-3 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการทำนายตัวแบบโดยวิธีความใกล้เคียงกันมากที่สุด



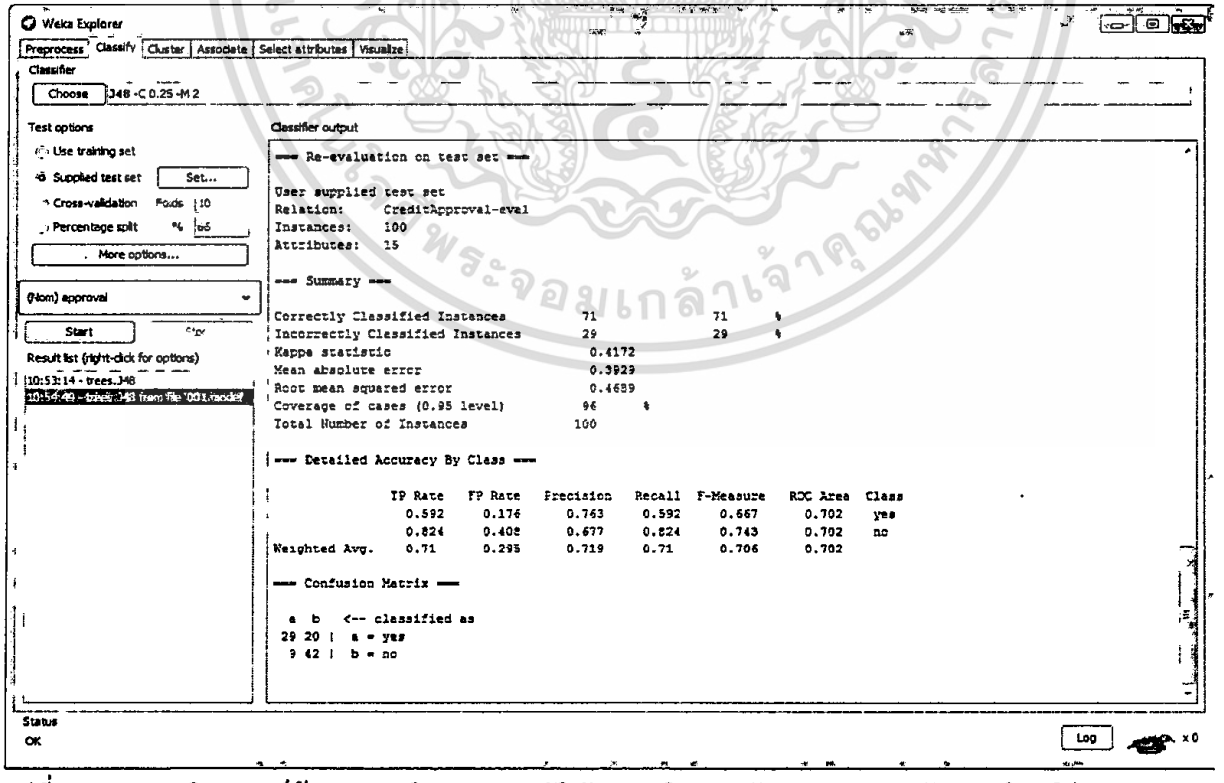
รูปที่ ข-4 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการทำนายตัวแบบโดยวิธีความใกล้เคียงกันมากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ (Decision Tree)

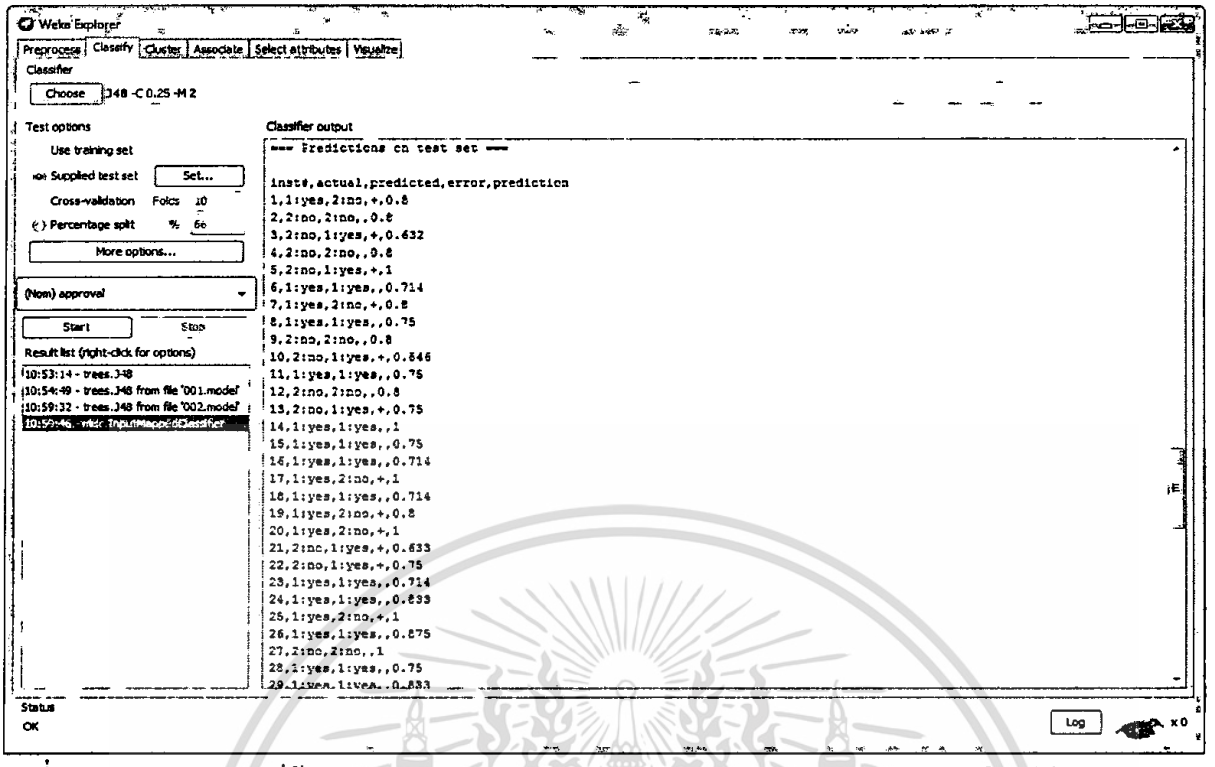


รูปที่ ข-5 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการสร้างตัวแบบโดยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

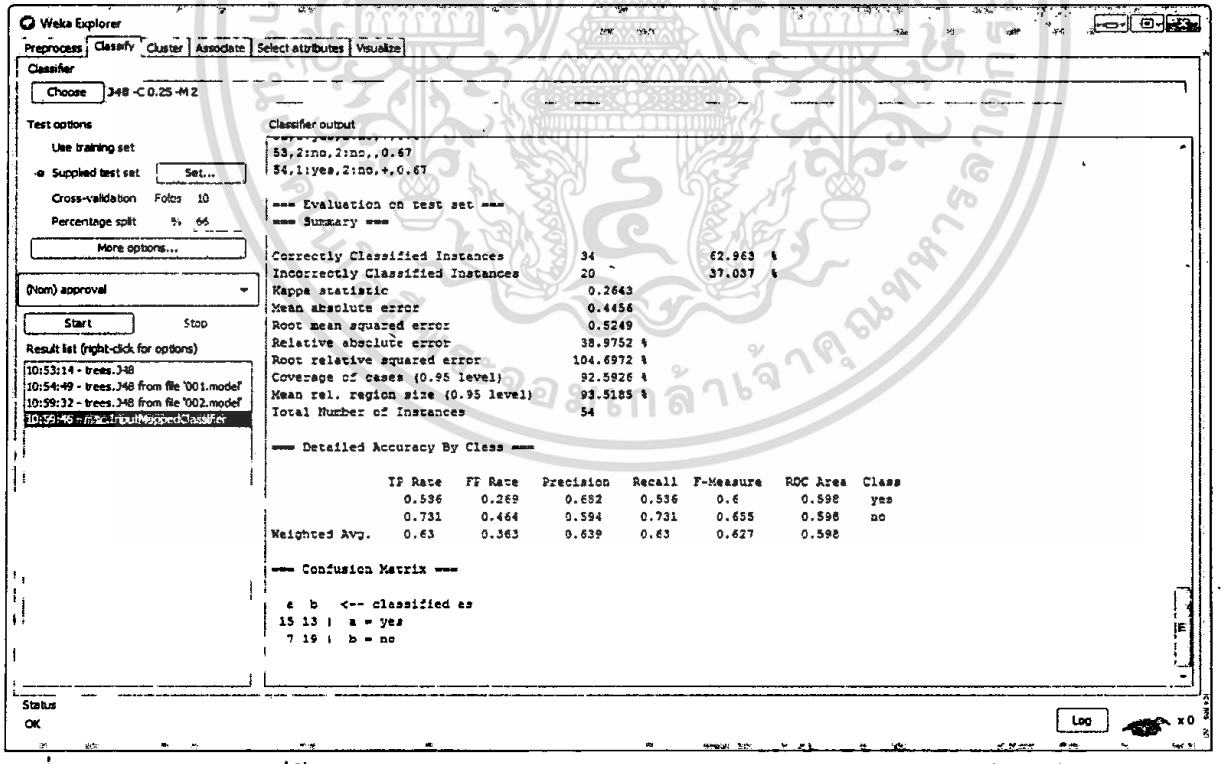


รูปที่ ข-6 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการทดสอบตัวแบบโดยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



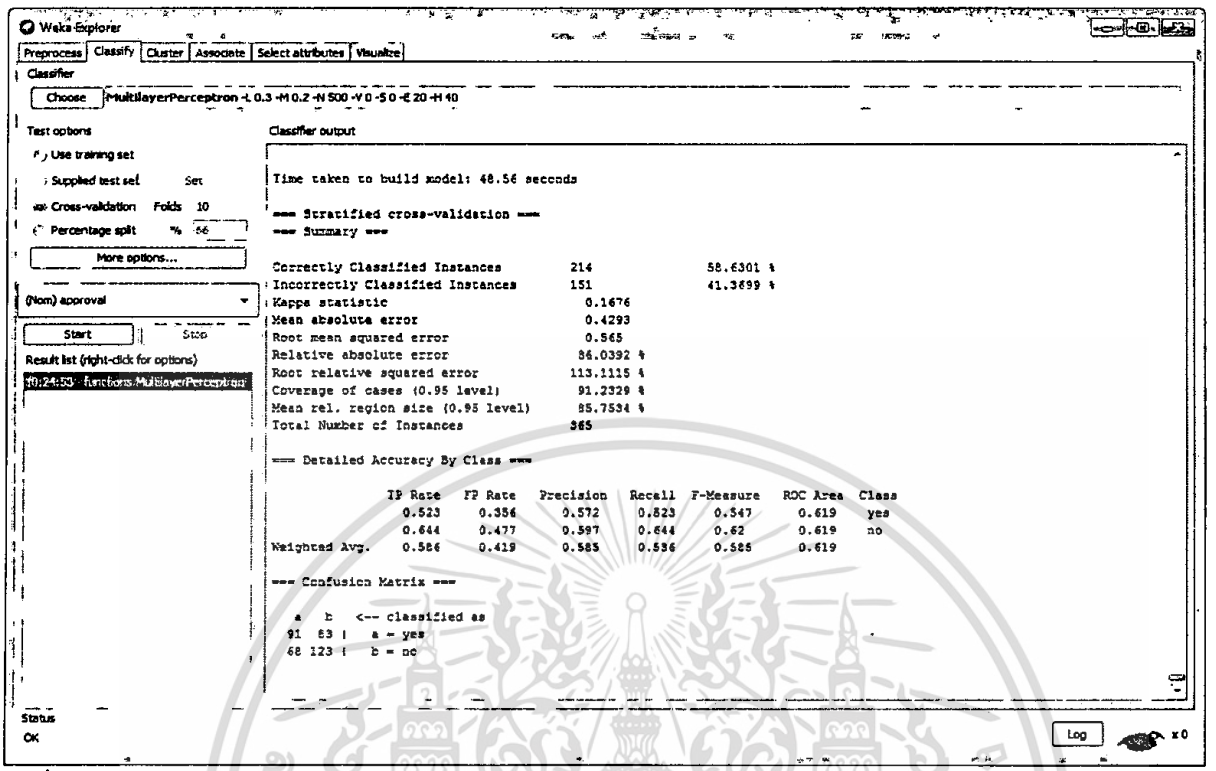
รูปที่ ข-7 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการทำนายตัวแบบโดยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ



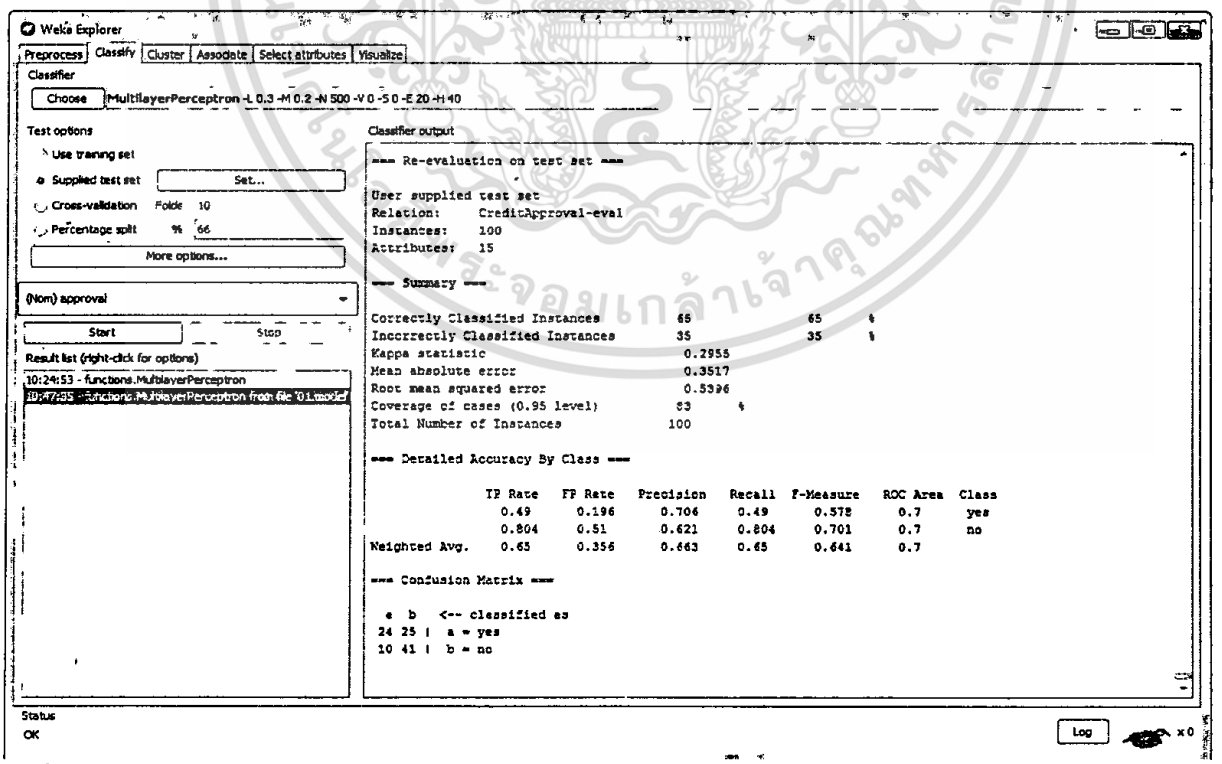
รูปที่ ข-8 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการทำนายตัวแบบโดยวิธีแผนภาพต้นไม้เพื่อการตัดสินใจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. วิธีโครงข่ายประสาท (Neural Network)

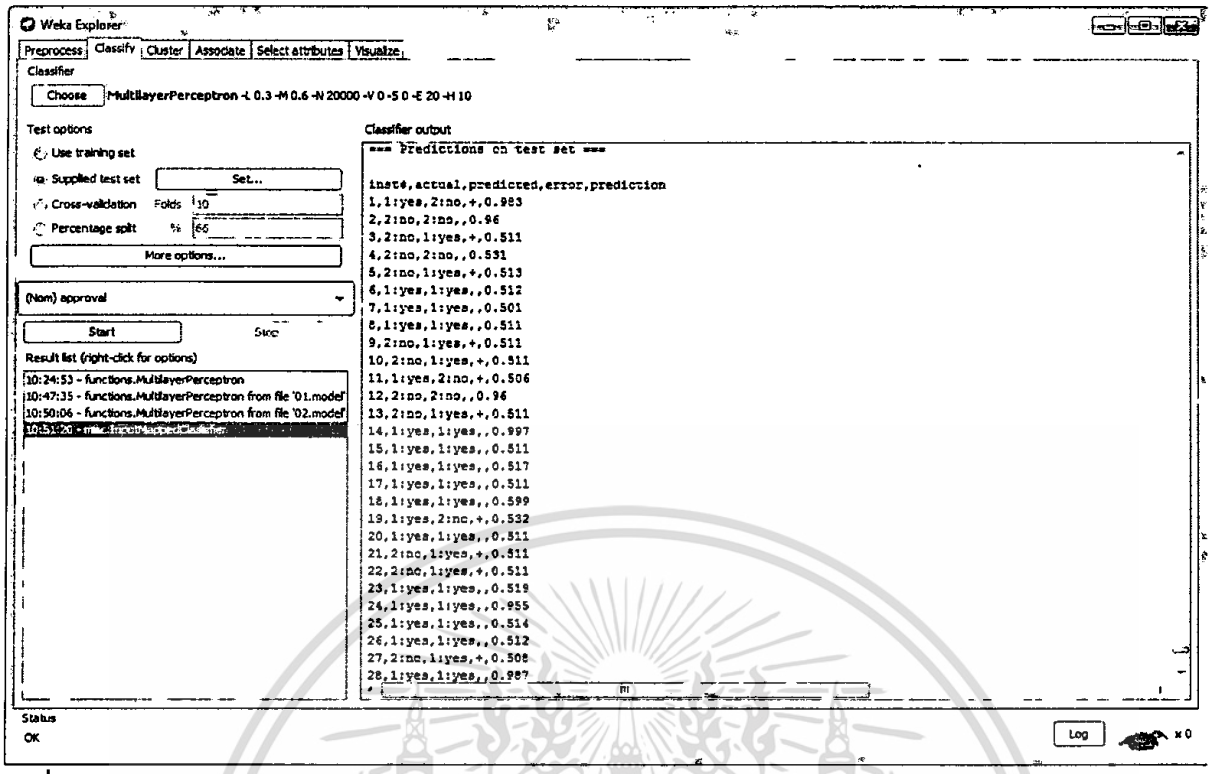


รูปที่ ข-9 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการสร้างตัวแบบโดยวิธีโครงข่ายประสาท

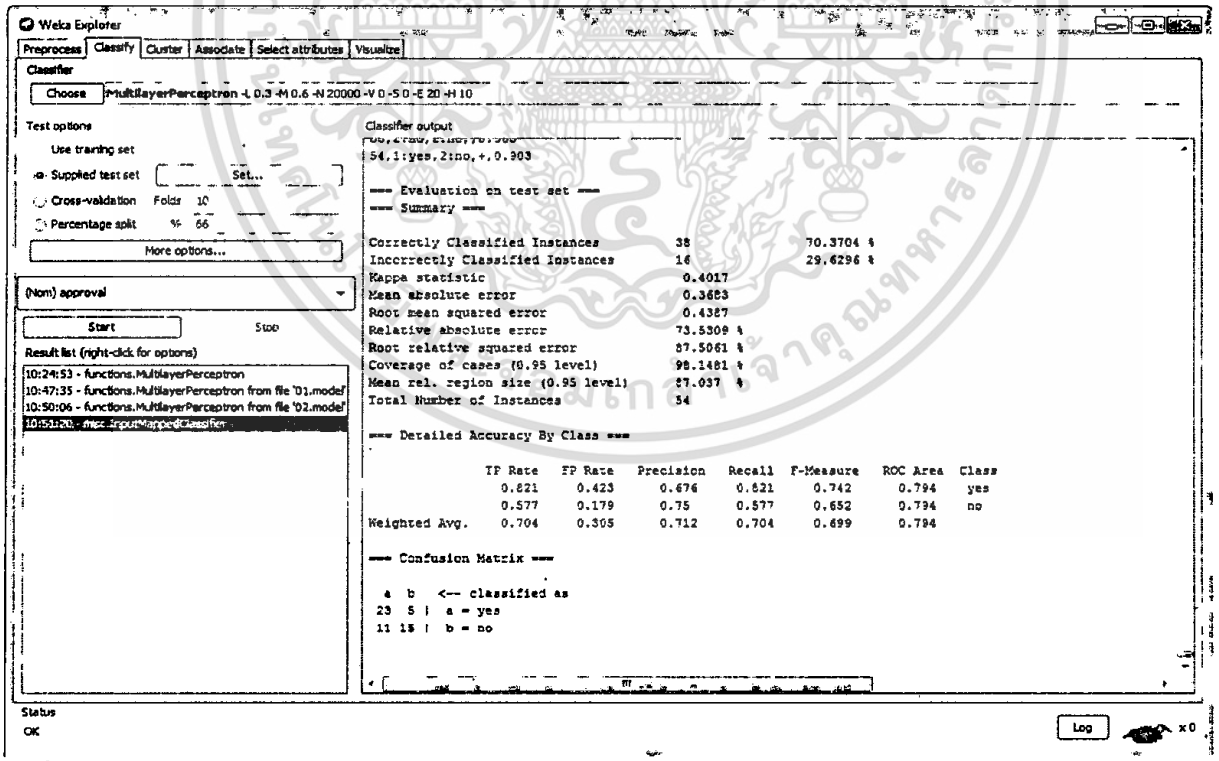


รูปที่ ข-10 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการทดสอบตัวแบบโดยวิธีโครงข่ายประสาท

เอกสารนี้มีไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



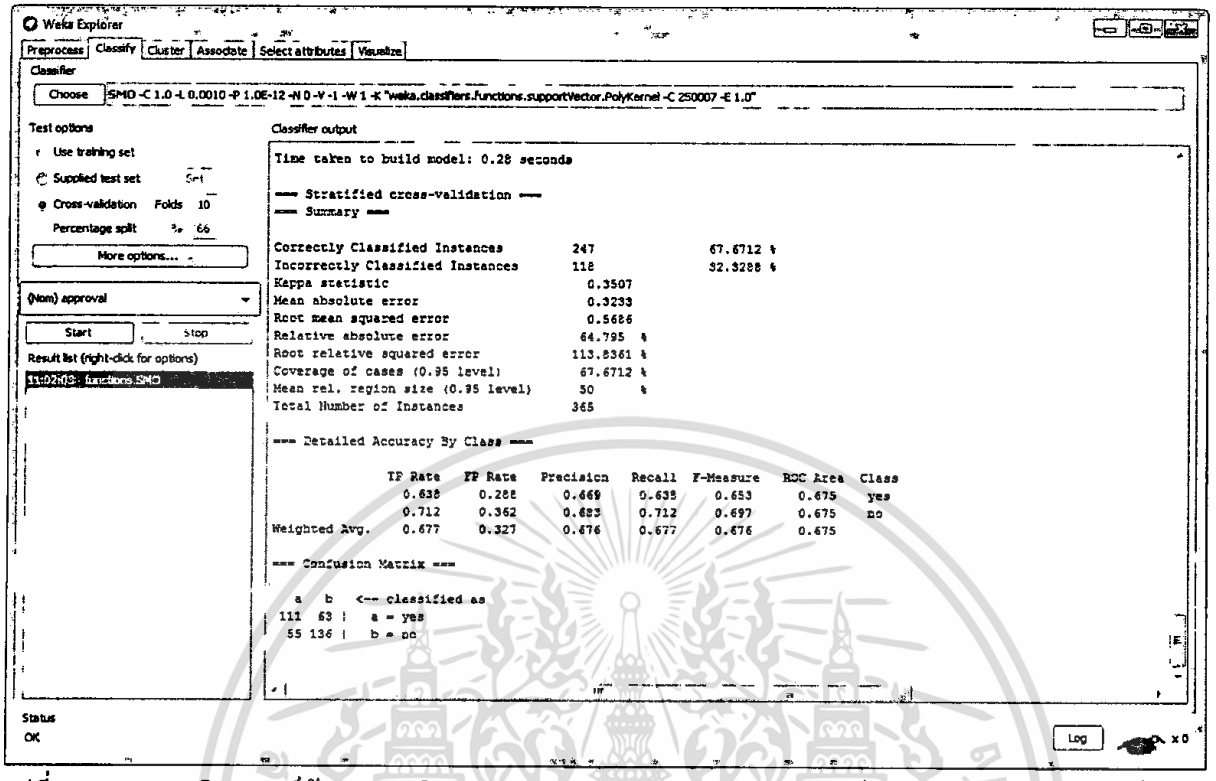
รูปที่ ข-11 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการทำนายตัวแบบโดยวิธีโครงข่ายประสาท



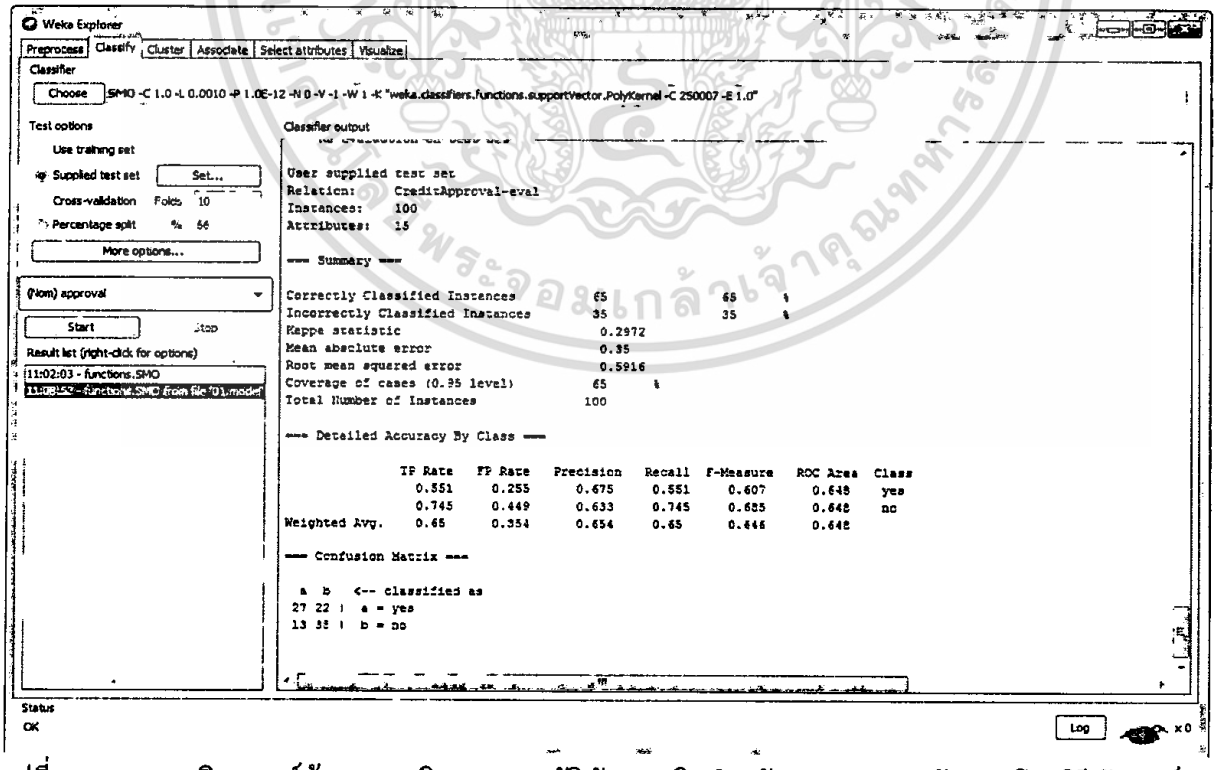
รูปที่ ข-12 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการทำนายตัวแบบโดยวิธีโครงข่ายประสาท

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. วิธีซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine)

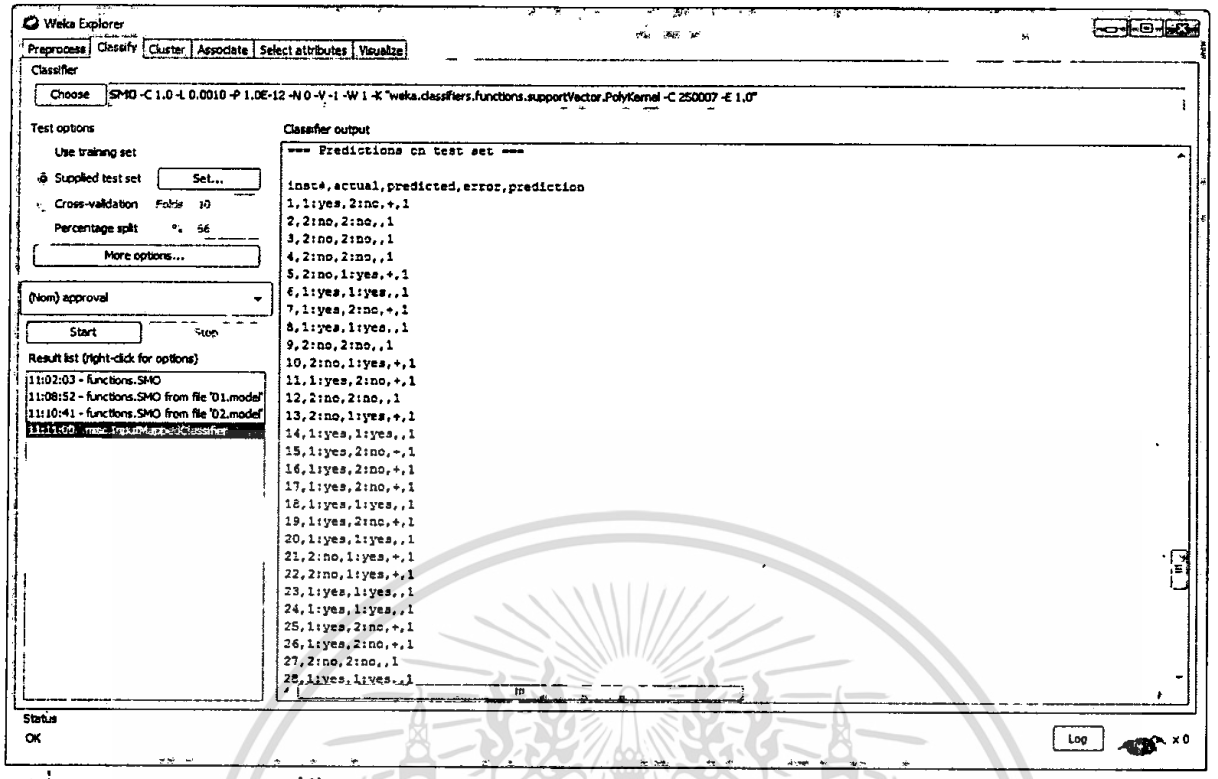


รูปที่ ข-13 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการสร้างตัวแบบโดยวิธีซัพพอร์ทเวกเตอร์แมชชีน

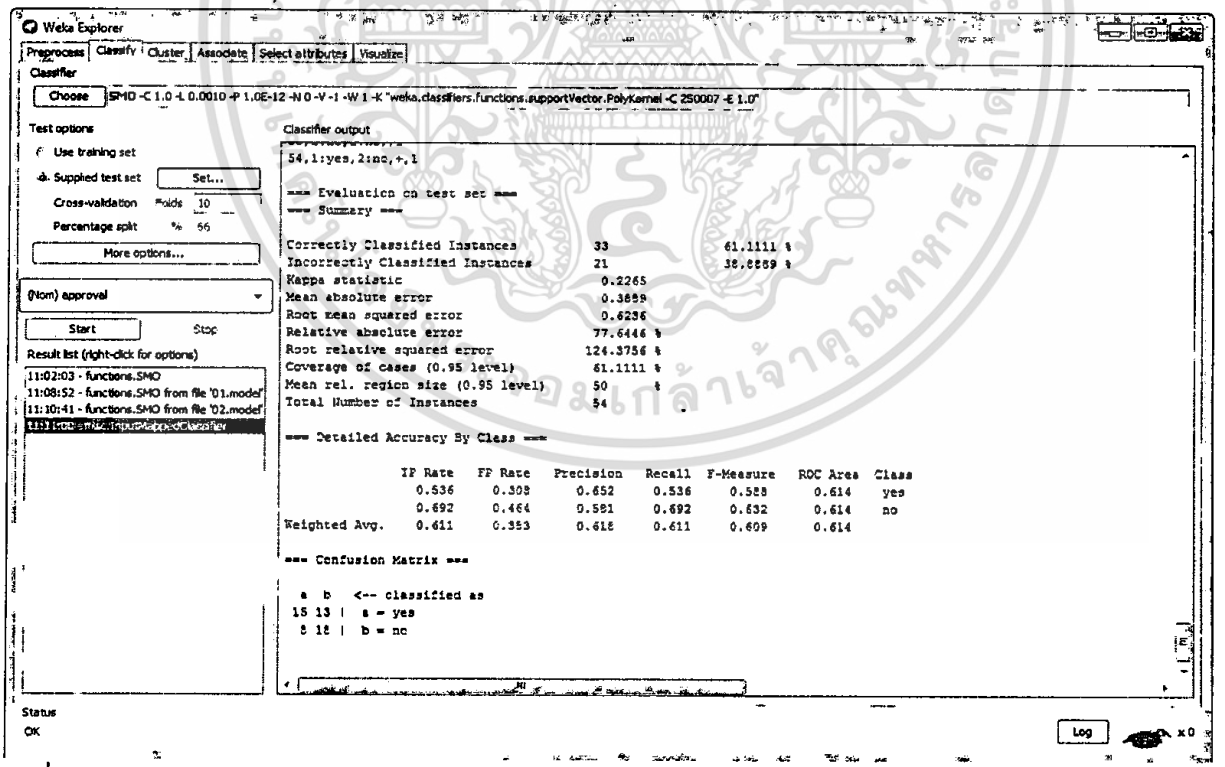


รูปที่ ข-14 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการทดสอบตัวแบบโดยวิธีซัพพอร์ทเวกเตอร์แมชชีน

เอกสารนี้สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



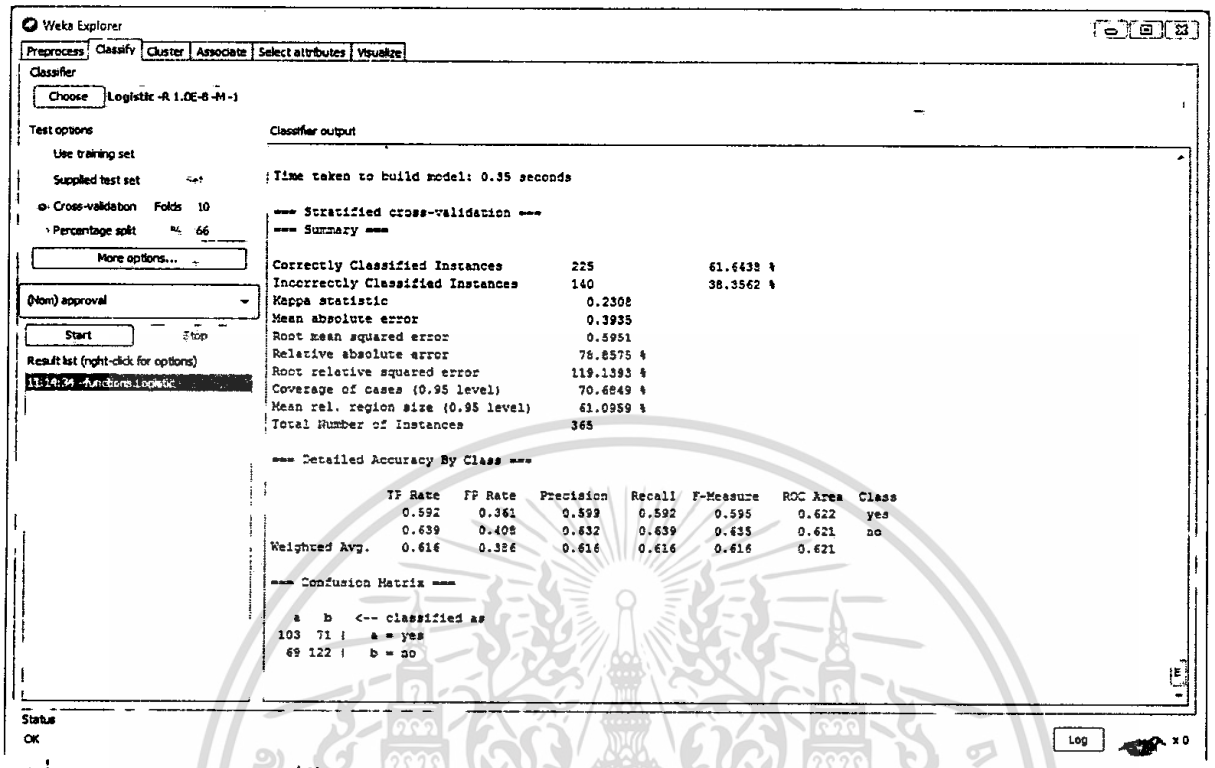
รูปที่ ข-15 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการทำนายตัวแบบโดยวิธีซัพพอร์ทเวกเตอร์แมชชีน



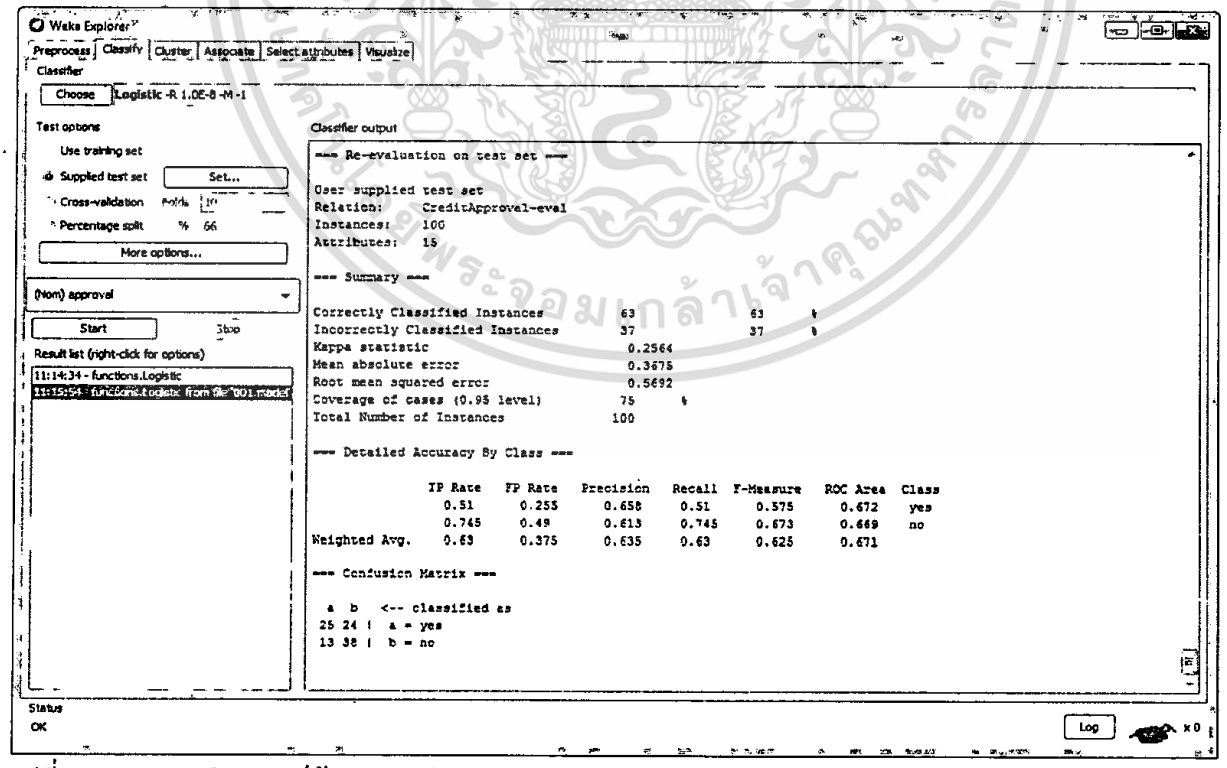
รูปที่ ข-16 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการทำนายตัวแบบโดยวิธีซัพพอร์ทเวกเตอร์แมชชีน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. วิธีการถดถอยโลจิสติก (Logistic Regression)

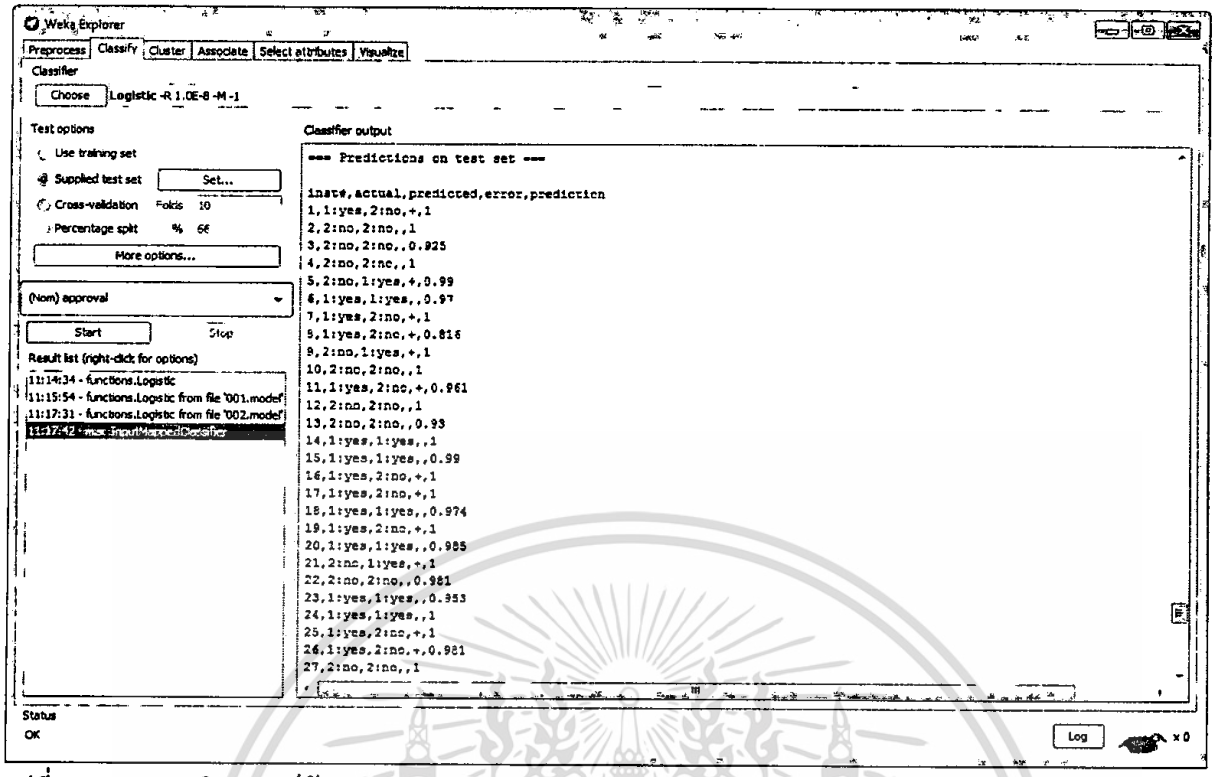


รูปที่ ข-17 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการสร้างตัวแบบโดยวิธีการถดถอยโลจิสติก

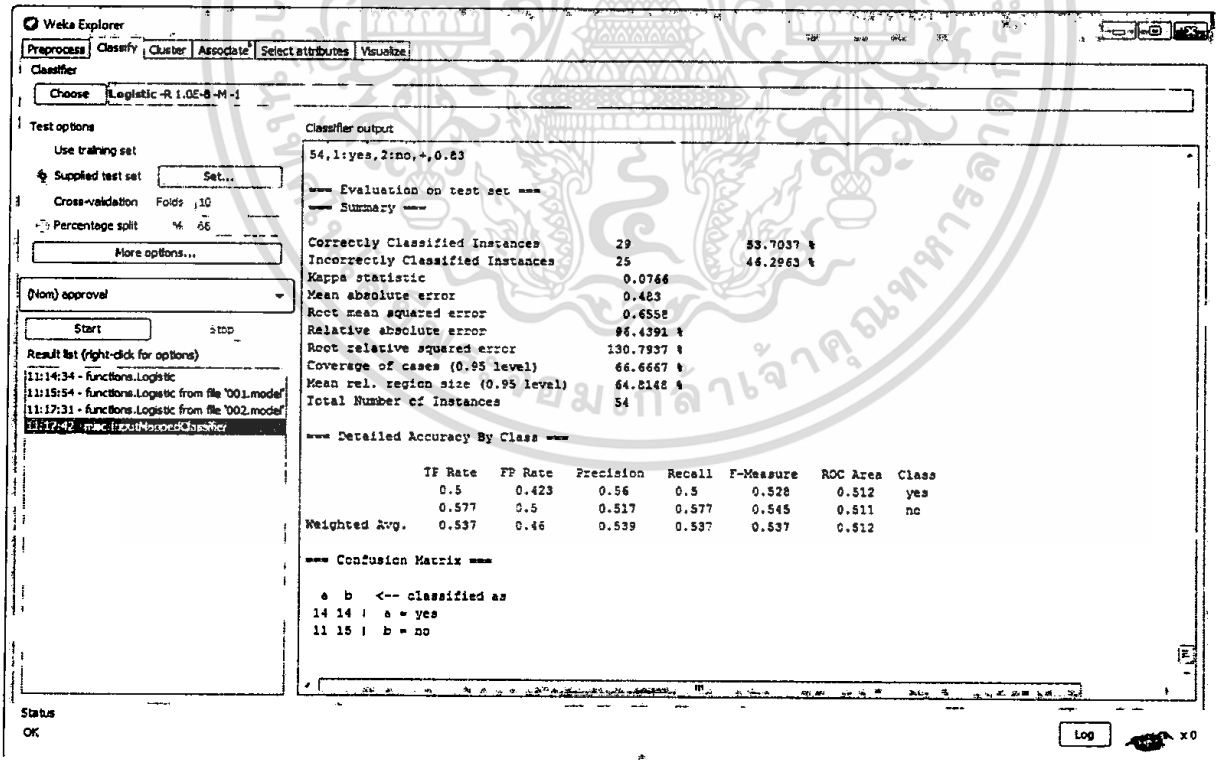


รูปที่ ข-18 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการทดสอบตัวแบบโดยวิธีการถดถอยโลจิสติก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ข-19 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการทำนายตัวแบบโดยวิธีการถดถอยโลจิสติก



รูปที่ ข-20 ผลการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการทำนายตัวแบบโดยวิธีการถดถอยโลจิสติก

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) ของการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิต โดยวิธีความใกล้เคียงกันมากที่สุด

จากรูปที่ ข-8

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

$= \frac{14+19}{14+19+7+14}$

$= 0.6111$ หรือ 61.11%

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

$= \frac{14}{14+7}$

$= 0.6667$ หรือ 66.67%

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

$= \frac{14}{14+14}$

$= 0.5$ หรือ 50%

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

•

$= \frac{2\times(0.5\times 0.6667)}{0.5+0.6667}$

$= 0.5714$ หรือ 57.14%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

<u>ตัวอย่างที่ 2</u> การคำนวณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของการวิเคราะห์ข้อมูลการพิจารณาอนุมัติบัตรเครดิตสำหรับการทำนายตัวแบบโดยวิธีความใกล้เคียงกันมากที่สุด

จากรูปที่ ข-7

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

เช่น ระเบียนที่ 2 ค่าจริง (actual) = 2:No ค่าทำนาย (predicted) = 2:No จะได้ $y_i = y_2 = 1$

ระเบียนที่ 6 ค่าจริง (actual) = 1:Yes ค่าทำนาย (predicted) = 1:Yes จะได้ $y_i = y_6 = 1$

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

เช่น ระเบียนที่ 1 ค่าจริง (actual) = 1:Yes ค่าทำนาย (predicted) = 2:No จะได้ $y_i = y_1 = 0$

ระเบียนที่ 9 ค่าจริง (actual) = 2:No ค่าทำนาย (predicted) = 1:Yes จะได้ $y_i = y_9 = 0$

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

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
1	0	0.997	0.994
2	1	0.997	0.000
3	1	0.997	0.000
4	1	0.997	0.000
5	1	0.997	0.000
6	1	0.997	0.000
7	1	0.997	0.000
8	1	0.997	0.000
9	0	0.997	0.994
10	0	0.997	0.994
11	0	0.997	0.994
12	1	0.997	0.000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
13	1	0.997	0.000
14	1	0.997	0.000
15	0	0.997	0.994
16	1	0.997	0.000
17	1	0.997	0.000
18	1	0.997	0.000
19	0	0.997	0.994
20	0	0.997	0.994
21	1	0.997	0.000
22	0	0.997	0.994
23	1	0.997	0.000
24	0	0.997	0.994
25	0	0.997	0.994
26	0	0.997	0.994
27	1	0.997	0.000
28	1	0.997	0.000
29	0	0.997	0.994
30	1	0.997	0.000
31	1	0.997	0.000
32	1	0.997	0.000
33	0	0.997	0.994
34	1	0.997	0.000
35	0	0.997	0.994
36	1	0.997	0.000
37	0	0.997	0.994
38	1	0.997	0.000
39	1	0.997	0.000
40	1	0.997	0.000
41	1	0.997	0.000
42	0	0.997	0.994

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	y_i	$\hat{y}_i$	e_i^2
43	1	0.997	0.000
44	1	0.997	0.000
45	1	0.997	0.000
46	0	0.997	0.994
47	1	0.997	0.000
48	0	0.997	0.994
49	1	0.997	0.000
50	1	0.997	0.000
51	0	0.997	0.994
52	1	0.997	0.000
53	0	0.997	0.994
54	0	0.997	0.994
รวม			20.874

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

$$= \frac{20.874}{54}$$

$$= 0.38656$$

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

$$= \sqrt{0.38656}$$

$$= 0.6219$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

1) มัณฑิตา สกุลรัตนศักดิ์. 2556. พฤติกรรมการใช้บริการของผู้ถือบัตรเครดิต บริษัทบัตรกรุงไทย จำกัด (มหาชน) ในเขตกรุงเทพมหานคร มหาวิทยาลัยศรีปทุม.

2) กัมพล กลมรัตน์ธาดา. 2553. ระบบสัญญาณเตือนภัยล่วงหน้าทางการเงิน : การเปรียบเทียบระหว่างวิธีการถดถอยโลจิสติกส์และโครงข่ายประสาทเทียม. วิทยานิพนธ์ วิทยศาสตรมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย.

3) วาทินี นุ้ยเพียร และคณะ. 2553. การเปรียบเทียบประสิทธิภาพและวิเคราะห์การจำแนกข้อมูลโดยใช้โครงข่ายประสาทเทียม ซัพพอร์ตเวกเตอร์แมชชีน นาอีฟเบย์ และเคเนียรเรสต์เนเบอร์. ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

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

5) ณัฐวุฒิ ศิริกุลรุ่งโรจน์ รัฐธชัย บุญวิเศษ สหัสนัยน์ ไพสานต์ และสุรวัชร ศรีเปารยะ. 2556. การเปรียบเทียบประสิทธิภาพของการจำแนกประเภทโดยใช้วิธีความใกล้เคียงกันมากที่สุด วิธีแผนภาพต้นไม้เพื่อการตัดสินใจ วิธีโครงข่ายประสาทเทียมและวิธีซัพพอร์ตเวกเตอร์แมชชีน. ปัญหาพิเศษปริญญาตรี สาขาวิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

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

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

8) วิวัช ชลไชยะ และกรุง สินอภิรมย์สราญ. 2550. การปรับปรุงตัวแบบการให้คะแนนสินเชื่อโดยใช้การวิเคราะห์การเกาะกลุ่มของตัวทำนายหลากหลาย. วิทยานิพนธ์ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

9) บุษกร สุคนธวงศาโรจน์. 2551. อัลกอริทึมในการจำแนกบุคลากรในองค์กรสำหรับการสร้างแผนที่ความรู้. ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

10) กิตติพล วิแสง สิรภัทร เชี่ยวชาญวัฒนา และคำรณ สุนัติ. 2552. การวิเคราะห์ปัจจัยเสี่ยงของโรคเบาหวาน. ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น.

11) พลอยพรรณ สอนสุวิทย์ และตรัสพงศ์ ไทยอุปถัมภ์. 2552. การเปรียบเทียบประสิทธิภาพการตรวจจับสิ่งผิดปกติทางเครือข่ายชนิด Probing Detection Efficiency Comparison of Probing-Type Network Anomaly. ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเชียงใหม่.

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

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

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

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