

การใช้เหมืองข้อมูลช่วยในการตัดสินใจการให้สินเชื่อ  
กรณีศึกษา: บริษัท กรุงไทยคาร์เร้นท์ แอนด์ ลีส จำกัด (มหาชน)

ทิพย์ธิดา วงศ์พิพันธ์

งานค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตร  
มหาบัณฑิต สาขาวิชาเทคโนโลยีคอมพิวเตอร์และการสื่อสาร  
คณะวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์  
พ.ศ.2556

# **Lease Approval Using Data Mining Techniques**

**THIPTIDA VONGPIPAN**

**An Independent Study Submitted in Partial Fulfillment of the  
Requirements for the Degree of Master of Science (Computer and  
Communication Technology) Department of Computer and  
Communication Technology  
Faculty of Engineering, Dhurakij Pundit University  
2013**

หัวข้องานค้นคว้าอิสระ	การใช้เหมืองข้อมูลช่วยในการตัดสินใจการให้สินเชื่อ
ชื่อผู้เขียน	ทิพย์ธิดา วงศ์พิพันธ์
อาจารย์ที่ปรึกษางานค้นคว้าอิสระ	อาจารย์ ดร.วรพล พงษ์เพชร
สาขาวิชา	เทคโนโลยีคอมพิวเตอร์และการสื่อสาร
ปีการศึกษา	2555

### บทคัดย่อ

วัตถุประสงค์ของการวิจัยในครั้งนี้เพื่อนำความรู้ที่ได้จากการทำเหมืองข้อมูลช่วยในการตัดสินใจการให้สินเชื่อของ บริษัท กรุงไทยคาร์เร้นท์ แอนด์ ลีส จำกัด (มหาชน) มาเป็นแนวทางสนับสนุนการตัดสินใจการอนุมัติสินเชื่อของบริษัทได้อย่างมีประสิทธิภาพมากขึ้น ซึ่งอาจมีส่วนช่วยลดปริมาณหนี้สูญได้

โดยการศึกษาจะเริ่มตั้งแต่การศึกษารวบรวมข้อมูลประวัติลูกค้าที่ผ่านการพิจารณาให้สินเชื่อแล้ว การเตรียมและคัดเลือกข้อมูลให้สมบูรณ์เหมาะสำหรับการทำเหมืองข้อมูล ได้แก่ การโอนย้ายข้อมูล (Data Transfer) การลดขนาดของข้อมูล (Data Reduction) และการทำความสะอาดข้อมูล (Data Cleaning) หลังจากนั้นจึงได้ทำการทดสอบข้อมูล โดยใช้เทคนิคการจำแนกกลุ่มเพื่อหากฎที่ใช้ในการจำแนกลูกค้ากลุ่มดี และลูกค้ากลุ่มไม่ดี ซึ่งได้ทดลองจำแนกกลุ่มข้อมูลด้วยเทคนิคการจำแนกกลุ่ม 3 เทคนิค ซึ่งให้ผลลัพธ์เป็นกฎ ได้แก่ โมเดลการจำแนกกลุ่มแบบ Decision Tree : C4.5 หรือเรียกว่าโมเดล Classifies trees J48 โมเดลการจำแนกกลุ่มแบบ Decision Rule : Part หรือเรียกว่าโมเดล Classifies PART และ โมเดลการจำแนกกลุ่มแบบ Bayes : Naïve Bayes หรือเรียกว่าโมเดล Classifier Naïve Bayes

ผลลัพธ์ที่ได้คือกฎที่ใช้ในการจำแนก ลูกค้ากลุ่มดี และลูกค้ากลุ่มไม่ดี ซึ่งผลการทดลองพบว่า การจำแนกกลุ่มข้อมูลด้วยเทคนิค J48 ให้ผลลัพธ์ที่ดีที่สุด เนื่องจากมีความถูกต้องมากที่สุด และสามารถแบ่งกลุ่มได้ตามเงื่อนไขที่ชัดเจนมากที่สุด เมื่อเทียบกับอีกสองเทคนิค ซึ่งท้ายสุดจึงนำกฎที่ได้จากผลลัพธ์มาเป็นแนวทางและเงื่อนไข ในการนำมาเป็นตัวอย่างที่สำคัญสำหรับการเขียนโปรแกรมเป็นระบบการตัดสินใจอนุมัติสินเชื่อออนไลน์ได้อย่างมีประโยชน์

Independent Study Title	Lease Approval Using Data Mining Techniques
Author	Thiptida Vongpipan
Independent Study Advisor	Dr. Worapol Pongpech
Department	Computer and Communication Technology
Academic Year	2012

### **ABSTRACT**

This research investigates employing data mining techniques for extracting potential useful information from the Krungthai Car Rent and Lease Public Company Limited database. The obtained information can then be used to facilitate decision making for credit approval of the company.

The first part of the research focuses on organizing, analyzing, and cleaning approved customer's information through data transfer, data reduction, and data cleaning. The second part of the research focuses on identifies potential rejected customers and potential approved customers using data mining techniques such as Classifiers tree J48, Classifiers PART, and Naïve Bayes techniques.

The results are a set of potential approved customers and a set of potential rejected customers. The results also revealed that the J48 provides the highest correction rate and maximum number of association rules when compared to the other two techniques. Finally, a simple credit approval decision making web-based software using this obtained information as a guideline and conditions has been written as a concrete example of a useful application.

## กิตติกรรมประกาศ

งานค้นคว้าอิสระฉบับนี้สำเร็จลุล่วงได้ด้วยดี โดยได้รับความอนุเคราะห์อย่างยิ่งจาก อาจารย์ ดร.วรพล พงษ์เพชร ที่กรุณาเสียสละเวลาอันมีค่า และให้คำแนะนำ และชี้แนะแนวทางการค้นคว้าหาข้อมูลเพิ่มเติม อันเป็นประโยชน์ทำให้งานวิจัยครั้งนี้สมบูรณ์ ผู้วิจัยขอกราบขอบพระคุณและจารึกพระคุณนี้ไว้ในความทรงจำอย่างมิรู้ลืมเลื่อน นอกจากนั้นขอกราบขอบพระคุณคณะกรรมการอื่น ๆ อันประกอบด้วย ดร.ประศาสน์ จันทราทิพย์ และดร.ณัฐพัชร์ อาริรัชกุลกานต์ ที่กรุณาให้ความช่วยเหลือในการแก้ไข และให้คำแนะนำที่มีประโยชน์ที่มีส่วนทำให้งานวิจัยครั้งนี้มีคุณค่ามากยิ่งขึ้น

ขอขอบพระคุณ ผู้จัดการหน่วยงานต่างๆ ที่กรุณาให้ความเอื้อเฟื้อในการให้ข้อมูล และเสียสละเวลาให้สัมภาษณ์ เพื่อการพัฒนาระบบ

คุณค่าและประโยชน์ใดๆ ที่อาจมีจากงานค้นคว้าอิสระฉบับนี้ ผู้วิจัยขอมอบเป็นเครื่องบูชาพระคุณของบิดามารดาที่ให้กำเนิดและเลี้ยงดูให้การศึกษา ตลอดจนครูบาอาจารย์และผู้ที่มีพระคุณทุกท่านที่มีส่วนในการวางรากฐานการศึกษาให้แก่ผู้วิจัย

ผู้ศึกษาหวังเป็นอย่างยิ่งว่า งานค้นคว้าอิสระฉบับนี้ จะเป็นประโยชน์กับผู้ที่ต้องการศึกษาการใช้เหมืองข้อมูลช่วยในการตัดสินใจการให้สินเชื่อ และหากมีข้อผิดพลาดประการใดในงานค้นคว้าอิสระฉบับนี้ ผู้ศึกษาต้องกราบขออภัยเป็นอย่างสูงมา ณ ที่นี้ด้วย

ทิพย์ธิดา วงศ์พิพันธ์

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ฉ
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญตาราง.....	ช
สารบัญภาพ.....	ฉ
บทที่	
1. บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของสารนิพนธ์.....	2
1.3 ขอบเขตของการวิจัย.....	2
1.4 ผลที่คาดว่าจะได้รับ.....	3
1.5 นิยามศัพท์เฉพาะ.....	3
2. แนวคิด ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง.....	5
2.1 แนวคิดและทฤษฎีที่เกี่ยวข้อง.....	5
2.2 งานวิจัยที่เกี่ยวข้องกับประเด็นปัญหา.....	19
3. ระเบียบวิธีวิจัย.....	22
3.1 ขั้นตอนการดำเนินการวิจัย.....	22
3.2 เครื่องมือที่ใช้ในการวิจัย.....	23
4. ผลดำเนินการวิจัย.....	24
4.1 วิธีการเก็บและรวบรวมข้อมูล.....	24
4.2 การเตรียมข้อมูลและคัดเลือกข้อมูลที่จะนำมาใช้.....	26
4.3 การทดสอบข้อมูลและเลือกเทคนิคที่เหมาะสม.....	30
4.4 ผลการวิเคราะห์.....	35
4.5 การพัฒนาระบบวิเคราะห์สินเชื่อ.....	50
4.6 การนำระบบงานที่พัฒนาขึ้นมาใหม่มาใช้ในองค์กร.....	54

## สารบัญ (ต่อ)

บทที่	หน้า
5. สรุปผลการวิจัยและข้อเสนอแนะ.....	59
5.1 สรุปผลการวิจัย.....	59
5.2 ปัญหาและแนวทางแก้ไข.....	60
5.3 ข้อเสนอแนะสำหรับการพัฒนาต่อไป.....	60
บรรณานุกรม.....	62
ภาคผนวก.....	65
ก ผลการรันฉบับเต็ม.....	66
ข ขั้นตอนการลงโปรแกรม Weka.....	70
ประวัติผู้เขียน.....	77

สารบัญตาราง

ตารางที่	หน้า
4.1 การแปลงค่าของข้อมูลวัตถุประสงค์การใช้งาน.....	28
4.2 อธิบายความหมายแต่ละแอทริบิวต์.....	35
4.3 อธิบายความหมายแต่ละฟิลด์ Object Car typ และ Bus Typ.....	36
4.4 อธิบายความหมายแต่ละฟิลด์ CusTyp Long Run และ Capital.....	37
4.5 อธิบายความหมายแต่ละฟิลด์ NP Income และ OE.....	37
4.6 อธิบายความหมายแต่ละฟิลด์ Cur ratio Debt Ratio และ DE.....	37
4.7 อธิบายความหมายแต่ละฟิลด์ Growth Inc และ Growth NP.....	38
4.8 อธิบายความหมายแต่ละฟิลด์ Grade.....	38



สารบัญภาพ

ภาพที่	หน้า
2.1 การแทนต้นไม้การตัดสินใจ.....	8
2.2 เครื่องข่ายประสาท.....	10
2.3 แสดงหน้าจอโปรแกรม Weka.....	11
2.4 แสดงเมนูหลักของ Explorer.....	12
2.5 แสดงส่วนประกอบหลักของ Explorer.....	13
2.6 สร้างเพิ่มข้อมูลแบบ CSV.....	14
2.7 แสดงตัวอย่างเพิ่ม current.csv.....	14
4.1 แสดงแบบวิเคราะห์และพิจารณาสินเชื่อเพื่อเช่ารถยนต์.....	25
4.2 การโอนย้ายข้อมูลจากแบบฟอร์มนำมาอยู่ในรูปแบบเดียวกัน ในไฟล์ Excel ที่สร้างเป็นฐานข้อมูล.....	27
4.3 ผลการแปลงค่าของข้อมูลวัตถุประสงค์การใช้งาน.....	29
4.4 การบันทึกข้อมูลในรูปแบบนามสกุล .csv.....	29
4.5 แสดงไฟล์ข้อมูลในรูปแบบนามสกุล .csv.....	30
4.6 แสดงการเปิดโปรแกรม Weka 3.6.6.....	31
4.7 แสดงการเข้าหน้าโปรแกรม Weka 3.6.6.....	31
4.8 การนำเข้าข้อมูลในโปรแกรม Weka.....	32
4.9 หน้าจอการแสดงผลข้อมูลที่นำเข้า.....	32
4.10 หน้าจอแสดงการเลือกเทคนิค Classify.....	33
4.11 การเลือกเทคนิคที่ใช้ในการจำแนกกลุ่มข้อมูล แบบ trees J48.....	33
4.12 หน้าจอแสดงผลลัพธ์การจำแนกกลุ่มข้อมูลด้วยเทคนิค Decision Tree : J48....	34
4.13 หน้าจอแสดงผลลัพธ์ที่ได้ของโมเดล.....	34
4.14 หน้าจอแสดงผลลัพธ์การจำแนกกลุ่มข้อมูลด้วยเทคนิค Decision Tree : C4.5...	39
4.15 หน้าจอแสดงผลลัพธ์ที่ได้ของโมเดล.....	39
4.16 ผลลัพธ์บางส่วนจากการจำแนกกลุ่มแบบ Decision Tree : C4.5.....	40

สารบัญภาพ (ต่อ)

ภาพที่	หน้า
4.17 ผลลัพธ์บางส่วนจากการจำแนกกลุ่มแบบ Decision Rule : Part.....	41
4.18 ผลลัพธ์บางส่วนจากการจำแนกกลุ่มแบบ Bayes : Naive Bayes.....	42
4.19 หน้าจอบันทึกข้อมูลลูกค้าที่ต้องการวิเคราะห์ห้อนุมัติสินเชื่อ.....	51
4.20 หน้าจอแสดงผลการวิเคราะห์ข้อมูลลูกค้า.....	51
4.21 หน้าจอแสดงการนำกฎมาเขียนโปรแกรมส่วนแรก.....	52
4.22 หน้าจอแสดงการนำกฎมาเขียนโปรแกรมส่วนที่เป็นเงื่อนไข.....	52
4.23 หน้าจอแสดงการนำกฎมาเขียนโปรแกรมส่วนที่เป็นเงื่อนไขส่วนสุดท้าย.....	53
4.24 หน้าจอแสดงรายงานข้อมูลลูกค้า.....	53
4.25 แบบวิเคราะห์และพิจารณาสินเชื่อ (CREDIT SCORING)-Test A.....	54
4.26 แบบวิเคราะห์และพิจารณาสินเชื่อ (CREDIT SCORING)-Test B.....	56
4.27 หน้าจอแสดงรายงานข้อมูลลูกค้า.....	58

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

บริษัท กรุงไทยคาร์เร็นท์ แอนด์ ลีส จำกัด (มหาชน) ดำเนินธุรกิจหลักคือ ให้บริการรถเช่าระยะยาว (Operating Lease) สัญญาเช่าระยะเวลา 3-5 ปี และระยะสั้น (Rent) สัญญาเช่าระยะเวลาเป็นรายวัน รายสัปดาห์ และรายเดือน กลุ่มลูกค้าเป็นนิติบุคคลทั้งขนาดกลางและขนาดใหญ่ บริษัทข้ามชาติ หน่วยงานราชการและรัฐวิสาหกิจ โดยลูกค้าสามารถเลือก ยี่ห้อและรุ่นของรถยนต์ รวมถึงอุปกรณ์เสริมต่างๆ ได้ตามความต้องการ การทำสัญญาเช่าเพื่อการดำเนินงานเป็นการลดภาระทางการเงิน โดยที่ลูกค้าไม่ต้องลงทุนจ่ายเงินก้อนเพื่อซื้อรถยนต์ ไม่มีภาระในการจัดหางานและผู้และหลักประกันต่างๆ ดังนั้น หัวใจหลักของการพิจารณาอนุมัติให้เช่ารถ จำเป็นต้องมีการวิเคราะห์ข้อมูลของผู้เช่าอย่างเข้มงวดก่อนการให้เช่ารถยนต์

การวิเคราะห์ข้อมูลการให้สินเชื่อนี้ ถือเป็นสิ่งสำคัญอย่างมาก เพราะมีผลต่อรายได้และสภาพคล่องขององค์กร ซึ่งจากประวัติที่ผ่านมาบริษัทฯ ต้องสูญเสียรายได้จากลูกค้าที่ไม่ดี คิดเป็นหนี้สูญประมาณ 0.4 % ของรายได้รวม ซึ่งถือเป็นรายได้ที่ไม่ควรสูญเสีย

วิธีการตรวจเครดิตของลูกค้าแบบปัจจุบัน เป็นการตรวจสอบจากเอกสารผู้เช่า โดยพิจารณาจากคุณสมบัติของผู้เช่า เช่น ประเภทธุรกิจ ระยะเวลาดำเนินธุรกิจ รายได้ และกำไร เป็นต้น เพื่อบ่งบอกความสามารถในการชำระหนี้ และทำการวิเคราะห์ข้อมูลเหล่านี้ผ่าน excel ซึ่งทำให้เกิดความล่าช้าในการวิเคราะห์ และวิเคราะห์ได้ผลแบบไม่ละเอียด มีโอกาสผิดพลาดได้ ส่งผลให้การวิเคราะห์เครดิตไม่ดีพอ และก่อให้เกิดหนี้สูญ

จากที่กล่าวมา มีวิธีการหลายวิธีที่สามารถช่วยในการแก้ปัญหาดังกล่าวได้ เช่น การตรวจเครดิตบูโร การใช้โปรแกรมระบบการจัดการฐานข้อมูล DBMS (Data Base Management System) ซึ่งมีความสามารถทางการจัดการเพิ่มข้อมูลมากมาย เช่น การสร้างแฟ้ม การกำหนดความสัมพันธ์ของข้อมูลซึ่งความสัมพันธ์ข้อมูลนี้ สามารถใช้ประมวลผลข้อมูล หลาย ๆ แฟ้มรวมกันเสมือนเป็นแฟ้มใหญ่แฟ้มเดียวได้ มีวิธีการจัดระบบข้อมูลให้เป็นระเบียบเรียบร้อย สามารถที่จะค้นหา ข้อมูลที่ต้องการได้สะดวกรวดเร็ว และการทำเหมืองข้อมูล (Data Mining) วิธีเหล่านี้ที่กล่าวมาดี แต่ในการศึกษางานวิจัยครั้งนี้เลือกการทำเหมืองข้อมูล เพราะเป็นวิธีการที่เหมาะสมกับการ

เพื่อค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ที่มีอยู่ เพื่อนำไปวิเคราะห์ ปรากฏการณ์ต่าง ๆ โดยสามารถนำมาใช้ทำนายผลได้

การทำเหมืองข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้ ซึ่งข้อมูลที่ถูกเก็บไว้ในฐานข้อมูลหากเก็บไว้เฉย ๆ ก็จะไม่เกิดประโยชน์ ดังนั้นจึงต้องมีการสกัดสารสนเทศไปใช้ การสกัดสารสนเทศ หมายถึง การคัดเลือกข้อมูลออกมาใช้งานในส่วนที่เราต้องการ ในปัจจุบันการวิเคราะห์ข้อมูลจากฐานข้อมูลเดี่ยวอาจไม่ให้ความรู้เพียงพอและลึกซึ้งสำหรับการดำเนินงานภายใต้ภาวะที่มีการแข่งขันสูงและมีการเปลี่ยนแปลงที่รวดเร็วจึงจำเป็นต้องรวบรวมฐานข้อมูลหลาย ๆ ฐานข้อมูลเข้าด้วยกัน เรียกว่าคลังข้อมูล ( Data Warehouse) เป็นการเก็บรวบรวมข้อมูลจากหลายแหล่งมาเก็บไว้ในรูปแบบเดียวกันและรวบรวมไว้ในที่ ๆ เดียวกัน

งานวิจัยนี้จึงขอเสนอการประยุกต์ใช้เทคนิค การใช้เหมืองข้อมูล (Data Mining) มาช่วยในการตัดสินใจการให้สินเชื่อ ของลูกค้าแต่ละประเภทธุรกิจที่มาเช่ารถ ซึ่งเป็นกรณีศึกษาของบริษัท กรุงเทพ คาร์เร็นท์ แอนด์ ลีส จำกัด (มหาชน) โดยใช้ข้อมูลลูกค้า ของปี 2551-2553 ซึ่งการวิเคราะห์การให้สินเชื่อนี้ จะมีประโยชน์ต่อเจ้าหน้าที่ฝ่ายกฎหมายในส่วนของ การพิจารณาการให้สินเชื่อกับลูกค้า โดยการทำการเหมืองข้อมูลจะช่วยทำให้วิเคราะห์ข้อมูล ได้รวดเร็ว ทันสมัย ลดความเสี่ยงด้านหนี้สูญ ทำให้มีเกณฑ์ในการคัดเลือกลูกค้าชั้นดีที่มีประสิทธิภาพมากขึ้นให้กับองค์กรได้

## 1.2 วัตถุประสงค์ของสารนิพนธ์

1. เพื่อการวิเคราะห์ข้อมูลของลูกค้าที่มาเช่ารถ
2. เพื่อช่วยการตัดสินใจพิจารณาข้อมูลสินเชื่อของลูกค้า ให้มีประสิทธิภาพมากยิ่งขึ้น
3. เพื่อลดความเสี่ยงด้านหนี้สูญ ในการอนุมัติให้เช่ารถยนต์ ต่อลูกค้าที่ไม่เหมาะสม

## 1.3 ขอบเขตของการวิจัย

1. รวบรวมข้อมูลที่ต้องการนำมาใช้ ในการทำการเหมืองข้อมูล
2. ศึกษาวิธีการของเครื่องมือ โปรแกรม Weka ที่จะใช้ในการทำการเหมืองข้อมูล (Data Mining )
3. นำเหมืองข้อมูลที่ได้ (Data Mining ) มาพัฒนาใช้กับเครื่องมือ โปรแกรม Weka เพื่อใช้ในการตัดสินใจการให้สินเชื่อ
4. วิเคราะห์ผลการวิจัย
5. สรุปผลการวิจัย

#### 1.4 ผลที่คาดว่าจะได้รับ

1. ช่วยให้การวิเคราะห์สินเชื่อเป็นไปอย่างถูกต้องมากขึ้น ลดความผิดพลาดในการอนุมัติสินเชื่อ
2. ช่วยให้การตัดสินใจการให้สินเชื่อ ได้รวดเร็วขึ้น และทันสมัย
3. ทำให้มีเครื่องมือที่ช่วยในการตัดสินใจที่เสถียรขึ้น เป็นระบบมากขึ้น
4. ทำให้ได้ซอฟต์แวร์ที่ช่วยในการพัฒนา การตัดสินใจการให้สินเชื่อ
5. ทำให้ข้อมูลที่มี เก็บอย่างเป็นระบบ และได้ใช้ข้อมูลอย่างมีประสิทธิภาพมากที่สุด

#### 1.5 นิยามศัพท์เฉพาะ

การทำเหมืองข้อมูล (Data Mining) คือ กระบวนการที่กระทำกับข้อมูลจำนวนมาก เพื่อค้นหารูปแบบและความสัมพันธ์ ที่ซ่อนอยู่ในชุดข้อมูลนั้น ซึ่งการจัดระบบฐานข้อมูลก็มีความสำคัญในการจัดการกับข้อมูลจำนวนมาก และควรมีระบบฐานข้อมูลที่ดี ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท เช่นด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร

ระบบฐานข้อมูล (Database System) หมายถึง โครงสร้างสารสนเทศที่ประกอบด้วยรายละเอียดของข้อมูลที่เกี่ยวข้องกันที่จะนำมาใช้ในระบบต่าง ๆ ร่วมกัน โดยการจัดเก็บข้อมูลอย่างเป็นระบบ ซึ่งผู้ใช้สามารถจัดการกับข้อมูลได้ในลักษณะต่าง ๆ ทั้งการเพิ่ม การแก้ไข การลบ ตลอดจนการเรียกดูข้อมูลซึ่งส่วนใหญ่จะเป็นการประยุกต์นำเอาระบบคอมพิวเตอร์เข้ามาช่วยในการจัดการฐานข้อมูล

บิต (Bit) หมายถึง หน่วยของข้อมูลที่มีขนาดเล็กที่สุด

ไบท์ (Byte) หมายถึง หน่วยของข้อมูลที่เกิดจากการนำบิต มารวมกันเป็นตัวอักษร (Character)

เขตข้อมูล (Field) หมายถึง หน่วยของข้อมูลที่ประกอบขึ้นจากตัวอักษรตั้งแต่หนึ่งตัวขึ้นไปมารวมกันแล้วได้ความหมายของสิ่งใดสิ่งหนึ่ง

ระเบียน (Record) หมายถึง หน่วยของข้อมูลที่เกิดจากการเอาเขตข้อมูลหลาย ๆ เขตข้อมูลมารวมกัน เพื่อเกิดเป็นข้อมูลเรื่องใดเรื่องหนึ่ง

แฟ้มข้อมูล (File) หมายถึง หน่วยของข้อมูลที่เกิดจากการนำข้อมูลหลาย ๆ ระเบียนที่เป็นเรื่องเดียวกันมารวมกัน

เอสคิวแอล (SQL) ย่อมาจาก structured query language คือภาษาที่ใช้ในการเขียนโปรแกรม เพื่อจัดการกับฐานข้อมูลโดยเฉพาะ

ลูกค้าดี หมายถึง ลูกค้าที่มีประวัติการชำระเงินตรงตามกำหนดของเงื่อนไขที่ลูกค้าได้รับการอนุมัติ

ลูกค้าไม่ดี หมายถึง ลูกค้าที่มีประวัติการชำระเงินไม่ตรงตามกำหนดของเงื่อนไขที่ลูกค้าได้รับการอนุมัติ

DPU

## บทที่ 2

### แนวคิด ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง

จากการศึกษาและค้นคว้าเอกสาร และงานวิจัยที่เกี่ยวข้องกับการศึกษาเรื่อง การใช้เหมืองข้อมูลช่วยในการตัดสินใจการให้สินเชื่อ ของบริษัท กรุงไทย คาร์เร้นท์ แอนด์ ลีส จำกัด (มหาชน) ทำให้สามารถจำแนก และประมวลความรู้ที่ได้จากการทบทวนเอกสาร ออกเป็น 2 ส่วน ได้แก่ แนวคิด ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง ซึ่งมีรายละเอียด ดังนี้

#### 2.1 แนวคิดและทฤษฎีที่เกี่ยวข้อง

##### 2.1.1 การทำเหมืองข้อมูล

การทำเหมืองข้อมูล (Data Mining) คือการค้นหาคำความสัมพันธ์และรูปแบบทั้งหมด ซึ่งมีอยู่จริงในฐานข้อมูลแต่ได้ถูกซ่อนไว้ภายในข้อมูลจำนวนมาก โดยการทำเหมืองข้อมูลจะเหมาะกับการแก้ปัญหาบางชนิดเท่านั้น มีเทคนิคต่างๆ ที่ใช้ในการแก้ไขปัญหาอยู่หลายเทคนิค ซึ่งไม่มีเทคนิคใดสามารถแก้ปัญหาได้ทุกปัญหา ดังนั้นความหลากหลายของเทคนิคเป็นสิ่งที่จำเป็นที่จะนำไปสู่วิธีการแก้ปัญหาที่ดีที่สุดของการทำเหมืองข้อมูล

เหมืองข้อมูล วิทยา พรพัชรพงศ์(2547) กล่าวว่า เหมืองข้อมูล (Data Mining) หรือ การค้นหาคำความรู้ในฐานข้อมูล (Knowledge Discovery in Database - KDD) เป็นเทคนิคเพื่อค้นหา รูปแบบ (Pattern) ของความรู้จากข้อมูลจำนวนมากโดยอัตโนมัติ

##### 2.1.2 ขั้นตอนการทำงานของการทำงานการทำเหมืองข้อมูล

จากขั้นตอนการทำงานของการทำงานการทำเหมืองข้อมูล ประกอบไปด้วย 4 ขั้นตอนหลัก ๆ ดังนี้

1. การระบุปัญหาที่เกิดขึ้นกับธุรกิจเป็นการระบุขอบเขตของข้อมูลที่จะนำมาทำการวิเคราะห์เพื่อหาความได้เปรียบทางการตลาดหรือนำมาแก้ไขปัญหา
2. ส่วนของการทำเหมืองข้อมูล เป็นการนำเทคนิคของการทำเหมืองข้อมูลไปใช้ถ่ายทอดหรือทำการเปลี่ยนแปลงข้อมูลดิบให้อยู่ในรูปของข้อมูลจะนำไปใช้ได้จริงในทางธุรกิจ
3. การนำเอาข้อมูลที่เป็นผลลัพธ์ของส่วนการทำเหมืองข้อมูลมาลองปฏิบัติจริงกับธุรกิจ
4. การวัดประสิทธิภาพของเทคนิคการทำเหมืองข้อมูลที่จะนำมาใช้จากผลลัพธ์ เช่น วัดจากส่วนแบ่งของตลาด วัดจากปริมาณลูกค้า หรือ วัดจากกำไรสุทธิ เป็นต้น จากทั้ง 4 ขั้นตอนนี้

กล่าวมาข้างต้นคือการนำเอา Data Mining ไปใช้กับระบบทางธุรกิจ โดยแต่ละขั้นตอนจะพึ่งพาอาศัยกันผลลัพธ์จากขั้นตอนหนึ่งจะกลายมาเป็นอินพุตจากอีกขั้นตอนต่อไป ซึ่ง Data Mining จะเปลี่ยนข้อมูลดิบให้เป็นข้อมูลประยุกต์ ดังนั้นการระบุแหล่งข้อมูลที่ถูกดึงจึงเป็นสิ่งที่สำคัญอย่างยิ่งต่อผลลัพธ์ที่ได้จากการวิเคราะห์

### 2.1.3 งานของการทำเหมืองข้อมูล

ในทางปฏิบัติจริง การทำเหมืองข้อมูลจะประสบความสำเร็จกับงานบางกลุ่มเท่านั้น และต้องอยู่ภายใต้ ภาวะที่จำกัดปัญหาเหมาะสมกับการใช้เทคนิคการทำเหมืองข้อมูล จะเป็นปัญหาที่ต้องใช้เหตุผลในการแก้ เป็นปัญหาที่เกี่ยวข้องกับเศรษฐศาสตร์และการเงิน ซึ่งจะสามารถจัดรูปแบบของธุรกิจให้อยู่ในรูปแบบของงานทั้ง 6 งานได้ ดังนี้

1. การจัดหมวดหมู่ (Classification)
2. การประเมินค่า (Estimation)
3. การทำนายล่วงหน้า (Prediction)
4. การจัดกลุ่มโดยอาศัยความใกล้ชิด (Affinity Group)
5. การรวมตัว (Clustering)
6. การบรรยาย (Description)

ไม่มีเทคนิคหรือเครื่องมือเพียงชนิดเดียวของการทำเหมืองข้อมูล ที่เหมาะสมกับงานทุกชนิด งานในแต่ละชนิดก็จะมีเทคนิคของการทำเหมืองข้อมูล ที่แตกต่างกันไปขึ้นอยู่กับชนิดของงาน

1. การจัดหมวดหมู่ การจัดหมวดหมู่ถือว่าเป็นงานธรรมดาทั่วไปของ การทำเหมืองข้อมูล เพราะการทำความเข้าใจและการติดต่อสื่อสารต่างๆ ก็เกี่ยวข้องกับการแบ่งเป็นหมวดหมู่การจัดแยกประเภท และการแบ่งแยกชนิดโดยการจัดหมวดหมู่ประกอบด้วยการสำรวจจุดเด่นของวัตถุที่ปรากฏออกมา และทำการกำหนด จุดเด่นนั้นๆ เป็นตัวที่ใช้แบ่งหมวดหมู่ งานในการแบ่งหมวดหมู่คือการบ่งบอกลักษณะ โดยการอธิบายจุดเด่นที่เป็นที่รู้จักดีในหมวดหมู่นั้น และชุดข้อมูลเรียนรู้ (Training Set) ของตัวอย่างในแต่ละหมวดหมู่ ซึ่งมีภาระหน้าที่ในการสร้างโมเดลของบางชนิดที่ไม่สามารถจะจัดหมวดหมู่ของข้อมูลได้ ให้สามารถจัดเป็น หมวดหมู่ได้ ตัวอย่างของการจัดหมวดหมู่ เช่น การจัดหมวดหมู่ของผู้ยื่นขอเครดิต (Credits) เป็นระดับต่ำระดับกลางและระดับสูงของความเสถียรที่จะได้รับเป็นต้น

2. การประเมินค่า ทางธุรกิจอย่างต่อเนื่องจะก่อให้เกิดผลลัพธ์ที่มีประโยชน์กับธุรกิจ การป้อนข้อมูลที่เราเมื่ออยู่เข้าไป เพื่อใช้ในการประเมินสิ่งต่างๆ ที่จะก่อให้เกิดประโยชน์ หรือสำหรับตัวแปรที่เราไม่รู้ค่าแน่นอนเช่น รายได้จากการค้า จุดสูงสุดทางธุรกิจ หรือคุณภาพของบัตร



เครดิต ในทางปฏิบัติการประเมิน ค่าจะถูกใช้ในการทำงานการจัดหมวดหมู่ ตัวอย่างของการประเมินค่าเช่นการประเมินรายได้รวมของครอบครัวหรือการประเมินจำนวนบุตรในครอบครัว

3. การทำนายล่วงหน้า การทำนายล่วงหน้าก็เป็นงานที่มีลักษณะคล้ายกับการจัดหมวดหมู่หรือการประเมินค่า ยกเว้น เพียงแต่จะใช้สถิติการบันทึกของการจัดหมวดหมู่ในการทำอนาคตของพฤติกรรมหรือการประเมิน ค่าที่จะเกิดขึ้นในอนาคต ตัวอย่างของงานการทำนายล่วงหน้า เช่น การทำนายการเปลี่ยนแปลงพฤติกรรม ของตลาด หรือการทำนายจำนวนลูกค้าที่จะออกจากธุรกิจของเราใน 6 เดือนข้างหน้า เป็นต้น

4. การจัดกลุ่มโดยอาศัยความใกล้ชิดกันหรือการวิเคราะห์ของตลาด งานในการจัดกลุ่มหรือการวิเคราะห์ตลาด คือการตัดสินใจรวมสิ่งที่สามารถไปด้วยกันเข้าไว้ในกลุ่มเดียวกัน ตัวอย่างของการจัดกลุ่มโดยอาศัยความใกล้ชิดกันหรือการวิเคราะห์ตลาด เช่น การตัดสินใจว่าลูกค้าคนใดจัดอยู่ในกลุ่มค้าประเภทใด

5. การรวมตัว การรวมตัวคืองานที่ทำการรวมส่วนต่างๆ ในแต่ละส่วนที่ต่างชนิดกันให้อยู่ในรวมกันเป็นกลุ่มย่อย หรือคลัสเตอร์ (Clusters) โดยในแต่ละกลุ่มย่อย อาจจะประกอบด้วยส่วนต่างๆที่ต่างชนิดกัน ซึ่งความแตกต่างของการรวมตัวจากการจัดหมวดหมู่คือ การรวมตัวจะไม่พึ่งพาอาศัยการกำหนดหมวดหมู่ล่วงหน้า และไม่ใช้ตัวอย่างข้อมูลจะรวมตัวกันบนพื้นฐานของความคล้ายในตัวเอง

6. การบรรยาย ในบางครั้งวัตถุประสงค์ของการทำเหมืองข้อมูล คือต้องการอธิบายความสัมพันธ์ของฐานข้อมูลในทางที่จะเพิ่มความเข้าใจในส่วนของประชากร ผลิตภัณฑ์ หรือขบวนการให้มากขึ้น เทคนิคการทำเหมืองข้อมูล ส่วนใหญ่ต้องการเทรนนิ่งข้อมูลจำนวนมากที่ประกอบด้วยหลายๆ ตัวอย่างเพื่อจะสร้างกฎที่ใช้ในการจัดหมวดหมู่ กฎของความสัมพันธ์ กลุ่มย่อย การทำนายล่วงหน้า ดังนั้นชุดของข้อมูลขนาดเล็ก จะนำไปสู่ความไม่แน่ใจของผลสรุปที่ได้ไม่มีเทคนิคใดเลย ที่จะสามารถแก้ปัญหาของการทำเหมืองข้อมูลได้ทุกปัญหา ดังนั้นความหลากหลายของเทคนิคจึงเป็นสิ่งจำเป็นในการไปสู่วิธีการแก้ปัญหาของData Mining ได้ดีที่สุด

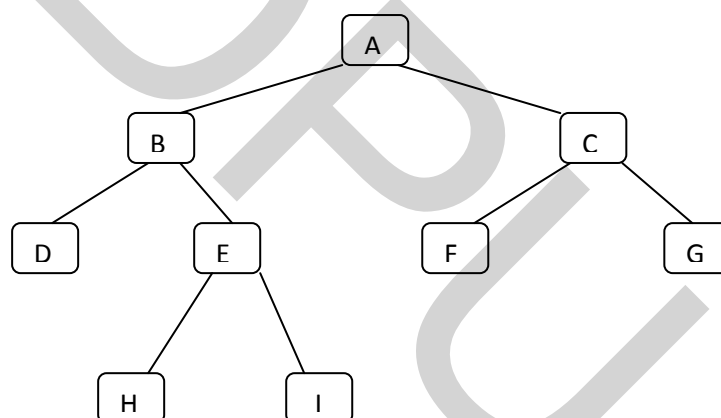
#### 2.1.4 เทคนิคการทำเหมืองข้อมูล

2.1.4.1 การจำแนกประเภทข้อมูล (Data Classification) กฤษณะ ไวยมัย และคณะ (2544, น.135) กล่าวว่า การจำแนกประเภทข้อมูล เป็นกระบวนการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้ โดยการสร้างกฎเพื่อช่วยในการตัดสินใจจากข้อมูลที่มีอยู่ เพื่อใช้ทำนายแนวโน้มการเกิดขึ้นของข้อมูลที่ยังไม่เกิดขึ้น โดยการนำเสนอกฎที่ได้จากเทคนิคการจำแนกประเภทข้อมูล นิยมนำเสนอในรูปแบบของแผนภูมิต้นไม้ซึ่งเรียกว่าต้นไม้ช่วยในการตัดสินใจ

2.1.4.2 ต้นไม้ช่วยการตัดสินใจ (decision tree)(C4.5) เป็นโครงสร้างที่ใช้แสดงกฎที่ได้จากเทคนิคการจำแนกประเภทข้อมูล โดยต้นไม้ช่วยการตัดสินใจจะมีลักษณะคล้ายโครงสร้างต้นไม้ที่แต่ละโหนดแสดงคุณลักษณะ (attribute), แต่ละกิ่งแสดงเงื่อนไขในการทดสอบ และโหนดปลาย (leaf node) แสดงกลุ่มที่กำหนดไว้

การแทนต้นไม้การตัดสินใจ (Decision Tree Representation)

1. โหนดภายใน (internal node) คือ คุณสมบัติต่าง ๆ ของข้อมูล ซึ่งเมื่อข้อมูลใด ๆ ตกลงมาที่โหนด จะใช้คุณสมบัตินี้เป็นตัวตัดสินใจว่าข้อมูลจะไปในทิศทางใด โดยโหนดภายในที่เป็นจุดเริ่มต้นของต้นไม้เรียกว่า โหนดราก
2. กิ่ง (Branch, link) เป็นค่าคุณสมบัติของคุณสมบัติในโหนดภายในที่แตกกิ่งนี้ออกมา ซึ่งโหนดภายในจะแตกกิ่งเป็นจำนวนเท่ากับจำนวนค่าคุณสมบัติของโหนดภายในนั้น
3. โหนดใบ (leaf node) คือ กลุ่มต่าง ๆ ซึ่งเป็นผลลัพธ์ในการแยกแยะข้อมูล



ภาพที่ 2.1 การแทนต้นไม้การตัดสินใจ

2.1.4.3 ตัวจำแนกประเภทเบย์อย่างง่าย (Naïve Bayes Classifier) เป็นตัวจำแนกประเภทแบบหนึ่งที่ใช้งานได้ดี เหมาะกับกรณีของเซตตัวอย่างมีจำนวนมากและคุณสมบัติ (Attribute) ของตัวอย่างไม่ขึ้นต่อกัน มีการนำตัวจำแนกประเภทเบย์อย่างง่ายไปประยุกต์ใช้งานในด้านการจำแนกประเภทข้อความ (text classification) การวินิจฉัย (diagnosis) และพบว่าใช้งานได้ดีไม่ต่างจากการจำแนกประเภทวิธีการอื่น

Naïve Bayes เป็นเทคนิคที่ถูกตั้งชื่อตาม Thomas Bayes นำเสนอโดย Kohavi ในปี 1996 ได้ใช้ทฤษฎี Bayes Theorem ในการแก้ปัญหาเรื่องการจำแนกประเภทข้อมูลที่ใช้แบบจำลองของการทำนาย โดยจะทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระแต่ละตัว กับตัวแปรตาม ทั้งนี้ เพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ทางทฤษฎี การทำนาย

ผลของ Naïve-Bayes จะมีความถูกต้อง เมื่อตัวแปรอิสระทั้งหมดมีความเป็นอิสระต่อกัน ซึ่งจะไม่ขึ้นกับตัวแปรอิสระตัวใดตัวหนึ่ง

ข้อจำกัดของอัลกอริทึม Naïve Bayes

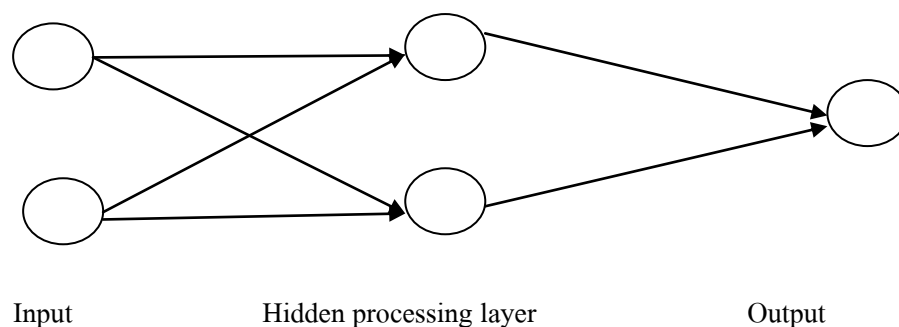
เทคนิคแบบ Naïve-Bayes จะไม่รองรับข้อมูลที่เป็นค่าต่อเนื่อง (Continuoues Data) ดังนั้น ตัวแปรอิสระ หรือตัวแปรตามที่มีค่าต่อเนื่อง จะต้องมีการแบ่งเป็นข้อมูลเป็นช่วง ๆ เช่น อายุงาน เป็นต้น จะเห็นได้ว่าถ้ามีการแบ่งช่วงของข้อมูลไม่เหมาะสม จะมีผลต่อคุณภาพของแบบจำลองที่สร้างขึ้นได้

2.1.4.4 กฎการตัดสินใจ (Decision rule) สามารถสร้างขึ้นจากต้นไม้การตัดสินใจได้ง่าย โดยจะอยู่ในรูปของกฎ IF...THEN โดยที่ IF {ส่วนของการกำหนดเงื่อนไข} THEN {ส่วนของการกระทำตามเงื่อนไข}

กฎการตัดสินใจ จะนำเสนอรูปของ Production Rules ในลักษณะของ IF...Then ที่ใช้ในการแก้ปัญหา สามารถสร้างกฎได้ง่ายและเป็นที่เข้าใจของผู้นำกฎนี้ไปใช้งาน ส่วนประกอบของกฎ มีด้วยกัน 2 ส่วน ได้แก่ IF ที่จะเรียกว่า Antecedent เป็นส่วนของการกระทำตามเงื่อนไขที่กำหนด หรืออาจเป็นส่วนของการแสดงบทสรุปว่าจะเป็นอะไร หรือจะได้อะไร เป็นต้น

เราสามารถเชื่อมกฎที่สร้างขึ้นด้วยคำเชื่อม AND , OR ซึ่งจะเรียกว่า Production rules

2.1.4.5 เครือข่ายประสาท (Neural Network) เป็นเทคโนโลยีที่มาจากงานวิจัยด้านปัญญาประดิษฐ์ (Artificial Intelligence: AI) เพื่อใช้ในการคำนวณค่าฟังก์ชันจากกลุ่มข้อมูล วิธีการของเครือข่ายประสาทเป็นวิธีการที่ให้เครื่องเรียนรู้จากตัวอย่างต้นแบบแล้วฝึกให้ระบบได้รู้จักคิดแก้ปัญหาที่กว้างขึ้นได้ใน โครงสร้างของเครือข่ายประสาทประกอบด้วยโหนด Node สำหรับ Input-output และการประมวลผลกระจายอยู่ในโครงสร้างเป็นชั้น ๆ ได้แก่ Input layer, output layer และ Hidden layers การประมวลผลของเครือข่ายประสาทจะอาศัยการส่งการทำงานผ่านโหนดต่าง ๆ ใน Layers เหล่านี้ ตัวอย่าง เครือข่ายประสาท กฤษณะ ไวยมัย และคณะ (2544, น.138)



ภาพที่ 2.2 เครือข่ายประสาท

2.1.4.6 การวิเคราะห์กลุ่ม (Cluster Analysis) เป็นเทคนิคการแบ่งกลุ่มหน่วยข้อมูล หรือเป็นการแบ่งคน สัตว์ สิ่งของ องค์กร ฯลฯ ออกเป็นกลุ่มย่อยอย่างน้อย 2 กลุ่ม โดยมี หลักเกณฑ์ในการแบ่งดังนี้ “ให้หน่วยที่อยู่ในกลุ่มเดียวกันมีลักษณะที่สนใจเหมือนกันหรือ คล้ายกัน แต่หน่วยที่อยู่ต่างกลุ่มกันจะมีลักษณะที่สนใจต่างกัน”

คำว่าลักษณะที่สนใจอาจจะมีหลาย ๆ ตัวแปร เช่น ถ้าสนใจความคิดเห็นทางการเมือง จะมี คำถามหลาย ๆ คำถามด้านการเมือง และจะนำคำตอบเหล่านั้นมาแบ่งกลุ่ม (กัลยา วานิชย์บัญชา. 2552, น.286)

การจัด Case (หมายถึง คน สัตว์ สิ่งของ หรือ องค์กร ฯลฯ) หรือเป็นการจัดตัวแปร ออกเป็นกลุ่มย่อย ๆ ตั้งแต่ 2 กลุ่มขึ้นไป Case ที่อยู่ในกลุ่มเดียวกันจะมีลักษณะที่เหมือนกันหรือ คล้ายกัน ส่วน Case ที่อยู่ต่างกลุ่มกันจะมีลักษณะที่แตกต่างกัน

ตัวแปรที่อยู่ในกลุ่มเดียวกันจะมีความสัมพันธ์กันมากกว่าตัวแปรที่อยู่ต่างกลุ่มกัน

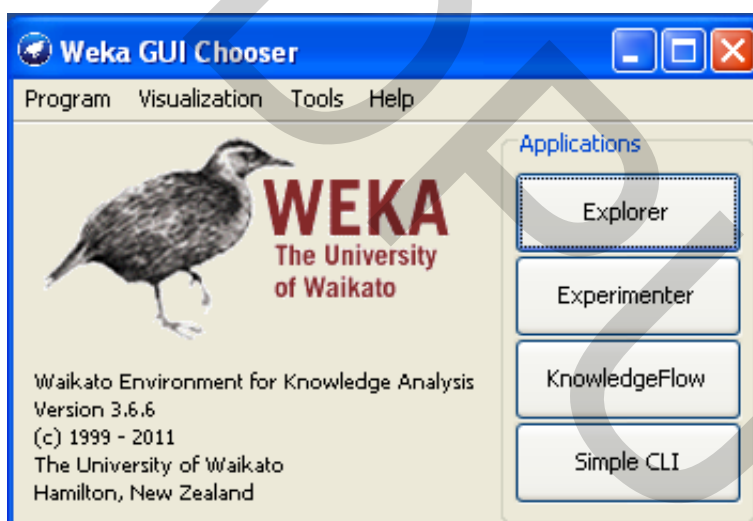
ตัวแปรที่อยู่ต่างกลุ่มกันจะมีความสัมพันธ์กันน้อยหรือไม่มีความสัมพันธ์กันเลย

2.1.4.7 การค้นหากฎความสัมพันธ์ (Association rule discovery) การค้นหา ความสัมพันธ์ของข้อมูลจากฐานข้อมูลขนาดใหญ่ที่มีอยู่เพื่อนำไปใช้ในการวิเคราะห์ หรือทำนาย ปรากฏการณ์ต่าง ๆ เป็นการประเมินจากข้อมูลในฐานข้อมูลที่รวบรวมไว้ โดยกฎความสัมพันธ์ที่ ได้มีส่วนช่วยในกระบวนการตัดสินใจทางด้านธุรกิจ ตัวอย่างของการค้นหากฎความสัมพันธ์ คือ การวิเคราะห์พฤติกรรมของผู้บริโภค โดยการหาความสัมพันธ์ระหว่างสินค้าที่ต่างชนิดกัน แต่ ลูกค้าจะเลือกซื้อไปพร้อมกัน

2.1.4.8 K-Nearest Neighbour (K-NN) คือ วิธีการในการจัดแบ่งคลาส เทคนิคนี้จะ ตัดสินใจ ว่าคลาสใดที่จะแทนเงื่อนไขหรือกรณีใหม่ๆ ได้บ้าง โดยการตรวจสอบจำนวนบางจำนวน (“K” ใน K-nearest neighbor) ของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด โดยจะ

หาผลรวม (Count Up)ของจำนวนเงื่อนไข หรือกรณีต่างๆสำหรับแต่ละคลาส และกำหนดเงื่อนไขใหม่ๆ ให้คลาสที่เหมือนกันกับคลาสที่ใกล้เคียงกันมากที่สุด

2.1.5 โปรแกรม WEKA ย่อมาจาก Waikato Environment for Knowledge Analysis เริ่มพัฒนา มาตั้งแต่ปี 1997 โดยมหาวิทยาลัย Waikato ประเทศนิวซีแลนด์ เป็นซอฟต์แวร์เสรีที่อยู่ภายใต้ ข้อตกลงของ GNU General Public License เขียนโดยใช้ภาษาจาวา ซอฟต์แวร์วิก้านอกจากจะ สนับสนุนเกี่ยวกับเทคนิคการเรียนรู้เรื่องเครื่องจักร(Machine Learning)แล้วยังมีอัลกอริทึมที่ สนับสนุนการทำงานของการทำงานเหมือนข้อมูลด้วยกิจกรรมที่เกี่ยวข้องกับการทำเหมืองข้อมูล ประกอบด้วยการเตรียมข้อมูล(Data Preprocessing)การทำเหมืองข้อมูลด้วยเทคนิคการจัดกลุ่ม (Clustering)การทำเหมืองข้อมูลด้วยเทคนิคการวิเคราะห์ความสัมพันธ์(Associating)เทคนิคการ คัดเลือกข้อมูล(Selecting Attributes)และเทคนิคการทำเหมืองข้อมูลด้วยรูปภาพ(Visualization)

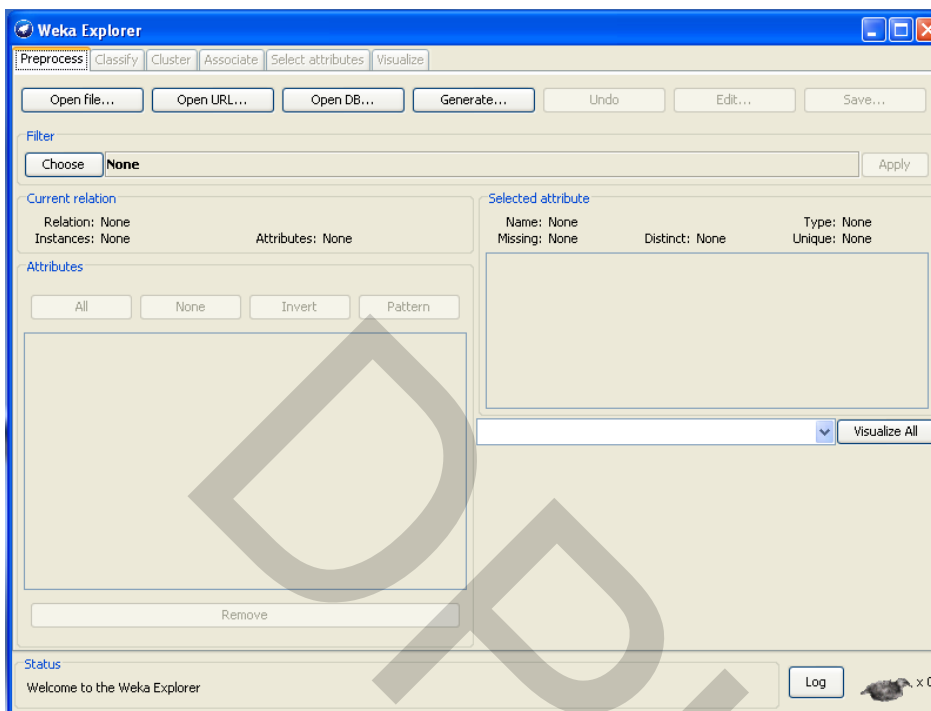


ภาพที่ 2.3 แสดงหน้าจอโปรแกรม Weka

### โปรแกรมหลักของซอฟต์แวร์ Weka

- 1) Simple CLI (Command Line Interface) เป็นโปรแกรมรับคำสั่งการทำงานผ่านการพิมพ์
- 2) Explorer เป็นโปรแกรมที่ออกแบบในลักษณะ GUI
- 3) Experimenter เป็นโปรแกรมที่ออกแบบการทดลองและการทดสอบผล
- 4) KnowledgeFlow เป็นโปรแกรมออกแบบผังการไหลของความรู้
- 5) ArffViewer เป็นโปรแกรมที่ใช้ สำหรับแก้ไขแฟ้มประเภท Arff

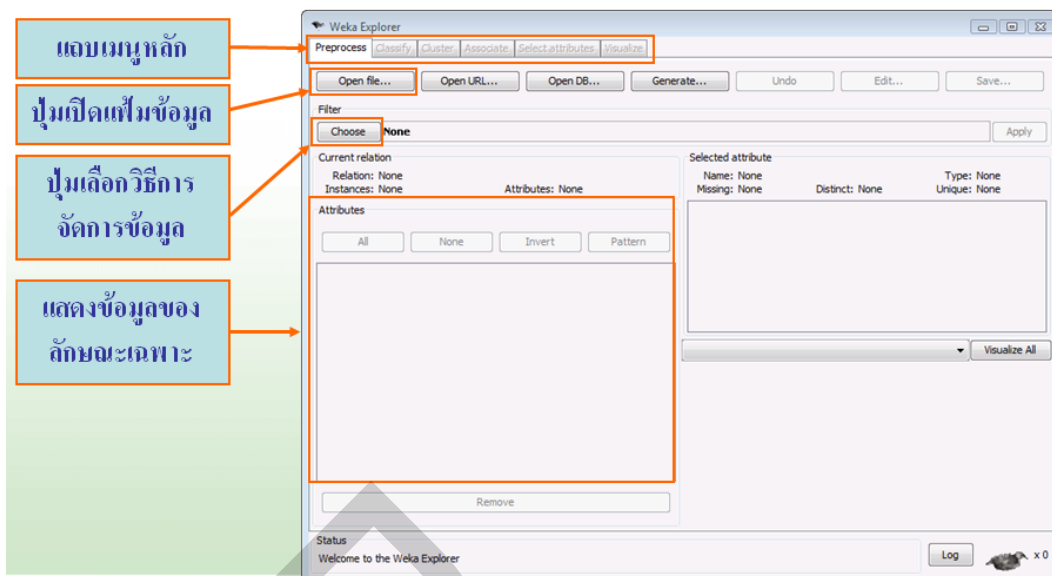
- 6) Log เป็นโปรแกรมที่ใช้ อ่านข้อความบันทึกเก็บระหว่างการทำงาน  
ในการวิจัยนี้จะใช้เฉพาะ โปรแกรม Explorer ในการศึกษาวิเคราะห์และทำเหมืองข้อมูล



ภาพที่ 2.4 แสดงเมนูหลักของ Explorer

### Preprocess การเตรียมข้อมูล

- 1) Classify รวมโมดูลการทำเหมืองข้อมูลแบบจัดแบ่งประเภท
- 2) Cluster รวมโมดูลการทำเหมืองข้อมูลแบบการเกาะกลุ่ม
- 3) Associate รวมโมดูลการทำเหมืองข้อมูลแบบกฎที่เชื่อมโยง
- 4) Select attributes รวมโมดูลสำหรับกาสรวิเคราะห์ ความเกี่ยวพันของลักษณะประจำ
- 5) Visualize นำเสนอข้อมูลด้วยภาพนามธรรมสองมิติ



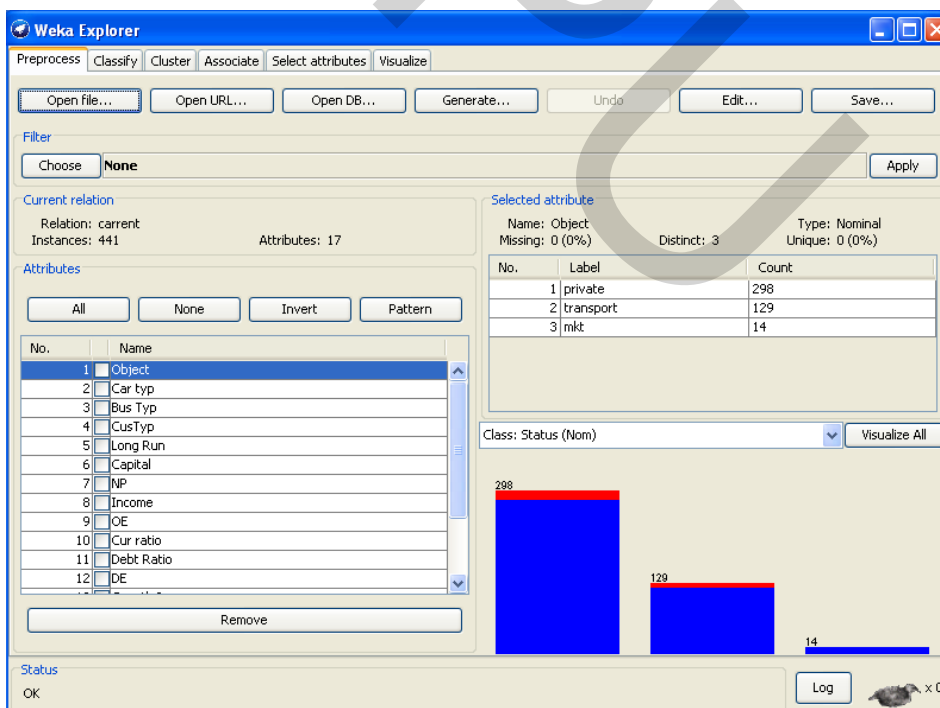
ภาพที่ 2.5 แสดงส่วนประกอบหลักของ Explorer

### ประเภทของเพิ่มข้อมูลที่รับได้

- 1) เพิ่มข้อมูลที่รับต้องเป็น ARFF หรือ CSV
- 2) ในกรณีที่เพิ่มข้อมูลอยู่ในเครือข่ายสามารถเรียกใช้ผ่าน URL ได้
- 3) สามารถเรียกใช้ข้อมูลจากฐานข้อมูลได้ โดยเชื่อมโยงผ่าน JDBC

Object	Car type	Bus Typ	CusTyp	Long Run	Capital	NP	Income	OE	Cur ratio	Debt Ratio	DE	Growth	Int	Grow
private	Large	four	CCC		15	10	15	15	10	5	5	5		3
transport	Van	two	CCC		0	8	0	0	0	0	0	0		0
transport	Van	four	CCC		13	8	15	13	10	5	4	0		5
private	MPV	five	CCC		0	8	0	0	0	0	0	0		0
transport	Pickup	five	CCC		14	10	15	13	10	5	5	5		5
transport	Pickup	five	CCC		13	0	0	0	0	4	3	0		0
private	Medium	five	CCC		13	8	15	15	8	5	5	5		5
transport	Pickup	four	CCC		15	9	15	15	10	5	5	5		5
transport	Small	five	CCC		14	8	13	0	8	5	5	5		0
transport	Pickup	five	CCC		15	8	15	15	10	5	4	4		5
transport	MPV	two	CCC		0	8	0	0	0	0	0	0		0
private	Large	two	CCC		15	10	13	15	8	4	4	0		4
private	Large	four	CCC		14	8	13	13	0	5	4	4		5
private	Large	four	CCC		14	8	13	15	8	5	4	4		4
private	Small	two	CCC		0	0	15	0	0	0	0	0		0
private	Large	five	CCC		15	0	15	13	8	5	5	5		4
private	Large	four	EEE		14	8	15	13	8	5	4	4		3
mkt	Pickup	five	CCC		15	8	15	13	10	5	4	0		4
private	Medium	four	CCC		13	8	13	13	8	5	5	4		0
transport	Small	two	CCC		14	8	15	15	10	5	4	0		5
private	Small	five	CCC		14	8	15	15	10	5	4	4		5
private	Small	four	CCC		15	8	15	15	10	5	4	3		5
transport	Pickup	five	CCC		14	0	15	15	10	5	4	0		5
private	Large	four	CCC		0	0	0	0	0	0	0	0		0
private	Large	four	CCC		15	8	14	15	10	5	5	5		4
private	Medium	four	CCC		15	8	15	13	10	5	5	5		0

ภาพที่ 2.6 สร้างเพิ่มข้อมูลแบบ CSV



ภาพที่ 2.7 แสดงตัวอย่างเพิ่ม current.csv



### 2.1.6 ความรู้ทั่วไปเกี่ยวกับระบบฐานข้อมูล

ฐานข้อมูลเป็นการจัดเก็บข้อมูลอย่างเป็นระบบ ทำให้ผู้ใช้สามารถใช้ข้อมูลที่เกี่ยวข้องในระบบงานต่าง ๆ ร่วมกันได้ โดยที่จะไม่เกิดความซ้ำซ้อนของข้อมูล และยังสามารถหลีกเลี่ยงความขัดแย้งของข้อมูลด้วย อีกทั้งข้อมูลในระบบก็จะถูกต้องเชื่อถือได้ และเป็นมาตรฐานเดียวกัน โดยจะมีการกำหนดระบบความปลอดภัยของข้อมูลขึ้น

นับได้ว่าปัจจุบันเป็นยุคของสารสนเทศ เป็นที่ยอมรับกันว่า สารสนเทศเป็นข้อมูลที่ผ่านการกลั่นกรองอย่างเหมาะสม สามารถนำมาใช้ประโยชน์อย่างมากมาย ไม่ว่าจะเป็นการนำมาใช้งานด้านธุรกิจ การบริหาร และกิจการอื่น ๆ องค์กรที่มีข้อมูลปริมาณมาก ๆ จะพบความยุ่งยากลำบากในการจัดเก็บข้อมูล ตลอดจนการนำข้อมูลที่ต้องการออกมาใช้ให้ทันต่อเหตุการณ์ ดังนั้นคอมพิวเตอร์จึงถูกนำมาใช้เป็นเครื่องมือช่วยในการจัดเก็บข้อมูล การประมวลผลข้อมูล ซึ่งทำให้ระบบการจัดเก็บข้อมูลเป็นไปได้สะดวก ทั้งนี้โปรแกรมแต่ละโปรแกรมจะต้องสร้างวิธควบคุมและจัดการกับข้อมูลขึ้นเอง ฐานข้อมูลจึงเข้ามามีบทบาทสำคัญอย่างมาก โดยเฉพาะระบบงานต่าง ๆ ที่ใช้คอมพิวเตอร์ การออกแบบและพัฒนาระบบฐานข้อมูล จึงต้องคำนึงถึงการควบคุมและการจัดการความถูกต้อง ตลอดจนประสิทธิภาพในการเรียกใช้ข้อมูลด้วย จากการจัดเก็บข้อมูลรวมเป็น ฐานข้อมูลจะก่อให้เกิดประโยชน์ดังนี้

#### 1. สามารถลดความซ้ำซ้อนของข้อมูลได้

การเก็บข้อมูลชนิดเดียวกันไว้หลาย ๆ ที่ ทำให้เกิดความซ้ำซ้อน (Redundancy) ดังนั้นการนำข้อมูลมารวมเก็บไว้ในฐานข้อมูล จะช่วยลดปัญหาการเกิดความซ้ำซ้อนของข้อมูลได้ โดยระบบจัดการฐานข้อมูล (Database Management System : DBMS) จะช่วยควบคุมความซ้ำซ้อนได้ เนื่องจากระบบจัดการฐานข้อมูลจะทราบได้ตลอดเวลาว่ามีข้อมูลซ้ำซ้อนกันอยู่ที่ใดบ้าง

#### 2. หลีกเลี่ยงความขัดแย้งของข้อมูลได้

หากมีการเก็บข้อมูลชนิดเดียวกันไว้หลาย ๆ ที่ และมีการปรับปรุงข้อมูลเดียวกันนี้ แต่ปรับปรุงไม่ครบทุกที่ที่มีข้อมูลเก็บอยู่ก็จะทำให้เกิดปัญหาข้อมูลชนิดเดียวกัน อาจมีค่าไม่เหมือนกันในแต่ละที่ที่เก็บข้อมูลอยู่ จึงก่อให้เกิดความขัดแย้งของข้อมูลขึ้น (Inconsistency)

#### 3. สามารถใช้ข้อมูลร่วมกันได้

ฐานข้อมูลจะเป็นการจัดเก็บข้อมูลรวมไว้ด้วยกัน ดังนั้นหากผู้ใช้ต้องการใช้ข้อมูลในฐานข้อมูลที่มาจากแฟ้มข้อมูลต่างๆ ก็จะทำให้ทำได้โดยง่าย

#### 4. สามารถรักษาความถูกต้องเชื่อถือได้ของข้อมูล

บางครั้งพบว่าการจัดเก็บข้อมูลในฐานข้อมูลอาจมีข้อผิดพลาดเกิดขึ้น เช่น จากการที่ผู้ป้อนข้อมูลป้อนข้อมูลผิดพลาดคือป้อนจากตัวเลขหนึ่งไปเป็นอีกตัวเลขหนึ่ง โดยเฉพาะกรณีมี

ผู้ใช้หลายคนต้องใช้ข้อมูลจากฐานข้อมูลร่วมกัน หากผู้ใช้คนใดคนหนึ่งแก้ไขข้อมูลผิดพลาดก็ทำให้ผู้อื่นได้รับผลกระทบตามไปด้วย ในระบบจัดการฐานข้อมูล (DBMS) จะสามารถใส่กฎเกณฑ์เพื่อควบคุมความผิดพลาดที่เกิดขึ้น

#### 5. สามารถกำหนดความป็นมาตรฐานเดียวกันของข้อมูลได้

การเก็บข้อมูลร่วมกันไว้ในฐานข้อมูลจะทำให้สามารถกำหนดมาตรฐานของข้อมูลได้ รวมทั้งมาตรฐานต่าง ๆ ในการจัดเก็บข้อมูลให้เป็นไปในลักษณะเดียวกันได้ เช่นการกำหนดรูปแบบการเขียนวันที่ ในลักษณะ วัน/เดือน/ปี หรือ ปี/เดือน/วัน ทั้งนี้จะมีผู้ที่คอยบริหารฐานข้อมูลที่เราเรียกว่า ผู้บริหารฐานข้อมูล (Database Administrator : DBA) เป็นผู้กำหนดมาตรฐานต่างๆ

#### 6. สามารถกำหนดระบบความปลอดภัยของข้อมูลได้

ระบบความปลอดภัยในที่นี้ เป็นการป้องกันไม่ให้ผู้ใช้ที่ไม่มีสิทธิมาใช้ หรือมาเห็นข้อมูลบางอย่างในระบบ ผู้บริหารฐานข้อมูลจะสามารถกำหนดระดับการเรียกใช้ข้อมูลของผู้ใช้แต่ละคนได้ตามความเหมาะสม

#### 7. เกิดความเป็นอิสระของข้อมูล

ในระบบฐานข้อมูลจะมีตัวจัดการฐานข้อมูลที่ทำหน้าที่เป็นตัวเชื่อมโยงกับฐานข้อมูลโปรแกรมต่าง ๆ อาจไม่จำเป็นต้องมีโครงสร้างข้อมูลทุกครั้ง ดังนั้นการแก้ไขข้อมูลบางครั้ง จึงอาจกระทำเฉพาะกับโปรแกรมที่เรียกใช้ข้อมูลที่เปลี่ยนแปลงเท่านั้น ส่วนโปรแกรมที่ไม่ได้เรียกใช้ข้อมูลดังกล่าว ก็จะเป็นอิสระจากการเปลี่ยนแปลง

#### 2.1.7 รูปแบบของระบบฐานข้อมูล รูปแบบของระบบฐานข้อมูล มีอยู่ด้วยกัน 3 ประเภท คือ

1. ฐานข้อมูลเชิงสัมพันธ์ (Relational Database) เป็นการเก็บข้อมูลในรูปแบบที่เป็นตาราง (Table) หรือเรียกว่า รีเลชัน (Relation) มีลักษณะเป็น 2 มิติ คือเป็นแถว (row) และเป็นคอลัมน์ (column) การเชื่อมโยงข้อมูลระหว่างตาราง จะเชื่อมโยงโดยใช้แอททริบิวต์ (attribute) หรือคอลัมน์ที่เหมือนกันทั้งสองตารางเป็นตัวเชื่อมโยงข้อมูล ฐานข้อมูลเชิงสัมพันธ์นี้จะเป็นรูปแบบของฐานข้อมูลที่นิยมใช้ในปัจจุบัน

2. ฐานข้อมูลแบบเครือข่าย (Network Database) ฐานข้อมูลแบบเครือข่ายจะเป็นการรวมระเบียบต่าง ๆ และความสัมพันธ์ระหว่างระเบียบแต่จะต่างกับฐานข้อมูลเชิงสัมพันธ์ คือ ในฐานข้อมูลเชิงสัมพันธ์จะแฝงความสัมพันธ์เอาไว้ โดยระเบียบที่มีความสัมพันธ์กันจะต้องมีค่าของข้อมูลในแอททริบิวต์ใดแอททริบิวต์หนึ่งเหมือนกัน แต่ฐานข้อมูลแบบเครือข่าย จะแสดงความสัมพันธ์อย่างชัดเจน

3. ฐานข้อมูลแบบลำดับชั้น (Hierarchical Database) ฐานข้อมูลแบบลำดับชั้น เป็นโครงสร้างที่จัดเก็บข้อมูลในลักษณะความสัมพันธ์แบบพ่อ-ลูก (Parent-Child Relationship Type :

PCR Type) หรือเป็นโครงสร้างรูปแบบต้นไม้ (Tree) ข้อมูลที่จัดเก็บในที่นี้ คือ ระเบียบ (Record) ซึ่งประกอบด้วยค่าของเขตข้อมูล (Field) ของเอนทิตีหนึ่ง ๆ ฐานข้อมูลแบบลำดับชั้นนี้คล้ายคลึงกับฐานข้อมูลแบบเครือข่าย แต่ต่างกันที่ฐานข้อมูลแบบลำดับชั้น มีกฎเพิ่มขึ้นมาอีกหนึ่งประการ คือในแต่ละกรอบจะมีลูกศรวิ่งเข้าหาได้ไม่เกิน 1 หัวลูกศร

### 2.1.8 โปรแกรมฐานข้อมูล

โปรแกรมฐานข้อมูล เป็นโปรแกรมหรือซอฟต์แวร์ที่ช่วยจัดการข้อมูลหรือรายการต่าง ๆ ที่อยู่ในฐานข้อมูล ไม่ว่าจะเป็นการจัดเก็บ การเรียกใช้ การปรับปรุงข้อมูล จะช่วยให้ผู้ใช้สามารถค้นหาข้อมูลได้อย่างรวดเร็ว ซึ่งโปรแกรมฐานข้อมูลที่นิยมใช้มีอยู่ด้วยกันหลายตัว แต่ในที่นี้ของยกตัวอย่างเช่น SQL เป็นโปรแกรมฐานข้อมูลที่มีโครงสร้างของภาษาที่เข้าใจง่าย ไม่ซับซ้อน มีประสิทธิภาพการทำงานสูง สามารถทำงานที่ซับซ้อนได้โดยใช้คำสั่งเพียงไม่กี่คำสั่ง โปรแกรม SQL จึงเหมาะที่จะใช้กับระบบฐานข้อมูลเชิงสัมพันธ์ และเป็นภาษาหนึ่งที่มีผู้นิยมใช้กันมาก โดยทั่วไปโปรแกรมฐานข้อมูลของบริษัทต่าง ๆ ที่มีใช้อยู่ในปัจจุบัน เช่น Oracle, DB2 ก็มักจะมีคำสั่ง SQL ที่ต่างจากมาตรฐานไปบ้างเพื่อให้เป็นจุดเด่นของแต่ละโปรแกรมไป

### 2.1.9 เอสคิวแอล (SQL)

SQL ย่อมาจาก structured query language คือภาษาที่ใช้ในการเขียนโปรแกรมเพื่อจัดการกับฐานข้อมูลโดยเฉพาะ เป็นภาษามาตรฐานบนระบบฐานข้อมูลเชิงสัมพันธ์และเป็นระบบเปิด (open system) หมายถึงเราสามารถใส่คำสั่ง sql กับฐานข้อมูลชนิดใดก็ได้ และ คำสั่งงานเดียวกันเมื่อสั่งงานผ่าน ระบบฐานข้อมูลที่แตกต่างกันจะได้ ผลลัพธ์เหมือนกัน ทำให้เราสามารถเลือกใช้ฐานข้อมูล ชนิดใดก็ได้โดยไม่ติดขัดกับฐานข้อมูลใดฐานข้อมูลหนึ่ง นอกจากนี้แล้ว SQL ยังเป็นชื่อโปรแกรมฐานข้อมูล ซึ่งโปรแกรม SQL เป็น โปรแกรมฐานข้อมูลที่มีโครงสร้างของภาษาที่เข้าใจง่าย ไม่ซับซ้อน มีประสิทธิภาพการทำงานสูง สามารถทำงานที่ซับซ้อนได้โดยใช้คำสั่งเพียงไม่กี่คำสั่ง โปรแกรม SQL จึงเหมาะที่จะใช้กับระบบฐานข้อมูลเชิงสัมพันธ์

เอสคิวแอล (SQL) คือ ภาษาสอบถามข้อมูล หรือภาษาจัดการข้อมูลอย่างมีโครงสร้าง มีการพัฒนาภาษาคอมพิวเตอร์ และ โปรแกรมฐานข้อมูลที่รองรับมากมาย เพราะจัดการข้อมูลได้ง่าย เช่น MySQL ซึ่งเป็น Open Source ที่ใช้งานได้ทั้งใน Linux และ Windows

SQL เป็นภาษาที่ใช้ในการเขียนโปรแกรม เพื่อจัดการกับฐานข้อมูลโดยเฉพาะ เราสามารถแบ่งการทำงานได้เป็น 4 ประเภท ดังนี้

## 1) Select query

ใช้ในการดึงข้อมูลในฐานข้อมูล จะมีการค้นหารายการจากตารางในฐานข้อมูล ตั้งแต่หนึ่งตารางขึ้นไป ตามเงื่อนไขที่ตั้ง ผลลัพธ์ที่ได้จะเป็นเซตของข้อมูลที่สามารถสร้างเป็นตารางใหม่

## 2) Update query

ใช้สำหรับการแก้ไขข้อมูลในตาราง โดยแก้ไขในคอลัมน์ที่มีค่าตรงตามเงื่อนไขมีรูปแบบดังนี้ Update ชื่อตาราง Set [ชื่อคอลัมน์=ค่าที่จะใส่เข้าไปในคอลัมน์นั้น ๆ] Where เงื่อนไข

## 3) Insert query

ใช้ในการเพิ่มเติมข้อมูลใหม่ ๆ เข้าไปในฐานข้อมูล มีรูปแบบดังนี้ Insert Into ชื่อตาราง [=ชื่อคอลัมน์1,2,..] Values [ค่าที่จะใส่ลงในคอลัมน์ 1,2,..]

## 4) Delete query

ใช้ลบข้อมูลออกจากตาราง มีรูปแบบดังนี้ Delete From ชื่อตาราง Where เงื่อนไข

## 2.1.10 MySQL (มายเอสคิวแอล)

เป็นระบบจัดการฐานข้อมูลโดยใช้ภาษา SQL. แม้ว่า MySQL เป็นซอฟต์แวร์โอเพนซอร์ส แต่แตกต่างจากซอฟต์แวร์โอเพนซอร์สทั่วไป โดยมีการพัฒนาภายใต้บริษัท MySQL AB ในประเทศสวีเดน โดยจัดการ MySQL ทั้งในรูปแบบที่ให้ใช้ฟรี และแบบที่ใช้ในเชิงธุรกิจ นิยมใช้งานร่วมกับภาษาโปรแกรม PHP นอกจากนี้ หลายภาษาโปรแกรมที่สามารถทำงานร่วมกับฐานข้อมูล MySQL ซึ่งรวมถึง ภาษาซี ซีพลัสพลัส ปาสคาล ซีชาร์ป ภาษาจาวา ภาษาเพิร์ล พีเอชพี ไพทอน รูบี และภาษาอื่น ใช้งานผ่าน API สำหรับโปรแกรมที่ติดต่อผ่าน ODBC หรือ ส่วนเชื่อมต่อกับภาษาอื่น (database connector) เช่น เอเอสพี สามารถเรียกใช้ MySQL ผ่านทาง MyODBC,ADO,ADO.NET เป็นต้น

## 2.2 ผลงานวิจัยที่เกี่ยวข้อง

งานวิจัยของปรีชา ยามันสะบีตัน และคณะ (2549) ศึกษาเรื่อง การทำเหมืองข้อมูลมาใช้ในการบริหารลูกค้าสัมพันธ์สำหรับนักเรียนระดับอุดมศึกษา ซึ่งพบว่าใช้ข้อมูลของนักเรียนระดับปริญญาตรี มหาวิทยาลัยสุโขทัยธรรมาธิราช จากฐานข้อมูลทั้งหมด 94,244 คน ได้ขนาดตัวอย่างจำนวน 94,244 คน เป็นการเลือกกลุ่มตัวอย่างด้วยวิธีสุ่มเท่ากันอย่างเป็นระบบ โดยนำข้อมูลส่วนตัวของนักศึกษาและข้อมูลการลงทะเบียนเรียนของนักศึกษามาใช้ในการทำนายว่าในอนาคต นักศึกษาจะมีโอกาสถูกรื้อหรือไม่ โดยใช้เทคนิคโครงสร้างต้นไม้ตัดสินใจ (Decision Tree) ในการสร้างกฎได้ 1,082 กฎ และได้กฎที่สำคัญ 2 ข้อ คือ กฎข้อที่หนึ่ง ถ้านักศึกษามีอัตราส่วนการสอบผ่านน้อยกว่า หรือ เท่ากับ 0.333 และมีภาคการศึกษาน้อยกว่าหรือเท่ากับ 10 ภาคการศึกษาและมีอัตราสอบไม่ผ่านมากกว่า 0.654 นักศึกษาคณนั้นจะถูกตัดชื่อออกกลางคัน ส่วนกฎข้อที่สองนั้นคือ ถ้านักศึกษามีอัตราส่วนการสอบผ่านมากกว่าหรือเท่ากับ 0.704 และมีจำนวนชุดวิชาที่ลงทะเบียนเรียนมากกว่า 4 ชุดวิชา และจำนวนภาคการศึกษามากกว่า 6 ภาคการศึกษา และนักศึกษามีอายุมากกว่า 27 ปีแล้ว นักศึกษาคณนั้นจะสำเร็จการศึกษา

นอกจากนี้ นางสาวบุญมา เฟ่งชวน (2548: 102) ได้เสนอขั้นตอนวิธีการศึกษาตัวแบบและพัฒนาตัวแบบ เพื่อนำตัวแบบที่ได้นำไปประยุกต์ใช้ในการสนับสนุนการตัดสินใจด้านการผลิตบัณฑิตระดับปริญญาตรี โดยขั้นตอนวิธีที่นำเสนอแบ่งเป็น 5 ส่วน ส่วนแรกเป็นการเตรียมข้อมูลสถานะการปฏิบัติงานทำของบัณฑิต โดยนำข้อมูลสถานะการปฏิบัติงานทำของบัณฑิตมาแปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมสำหรับการนำไปศึกษา ส่วนที่ 2 เป็นการศึกษาตัวแบบเพื่อใช้ในการทำนายแนวโน้มการเลือกอาชีพแรกหลังสำเร็จการศึกษาของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคการค้นหากฎความสัมพันธ์ ผู้วิจัยได้ศึกษาตัวแปรที่จะนำมาใช้สร้างเป็นเงื่อนไขและผลลัพธ์ ซึ่งในการวิจัยนี้ได้ใช้สาขาวิชาและเพศเป็นเงื่อนไขส่วนผลลัพธ์คืออาชีพ ส่วนที่ 3 เป็นการพัฒนาตัวแบบโดยนำข้อมูลสถานะการปฏิบัติงานทำของบัณฑิตที่ผ่านการเตรียมข้อมูลเรียบร้อยแล้วมาสร้างตัวแบบตามที่ได้ศึกษา มา โดยแบ่งข้อมูลออกเป็น 2 ชุด คือ ข้อมูลเรียนรู้ และข้อมูลตรวจสอบ ส่วนที่ 4 เป็นการทดสอบตัวแบบ เพื่อหาความถูกต้องของตัวแบบ และคำนวณหาร้อยละความถูกต้องของตัวแบบ เพื่อพิจารณาว่าตัวแบบที่ศึกษาและพัฒนาขึ้นสามารถนำไปใช้ในการสนับสนุนการตัดสินใจด้านการผลิตบัณฑิตระดับปริญญาตรีได้หรือไม่ ส่วนที่ 5 เป็นการพัฒนาระบบสนับสนุนการตัดสินใจหลังจากที่ทดสอบตัวแบบแล้วว่าตัวแบบที่ศึกษาและสร้างขึ้นนั้น สามารถนำไปใช้ได้ จากค่าความเชื่อมั่นสูงสุด และค่าสนับสนุนสูงสุดของแต่ละเงื่อนไข ต่อจากนั้นจึงทำการแปลความหมายโดยนำข้อมูลรายบุคคลนักศึกษาที่ผ่านการแปลงข้อมูลแล้วกับตัวแบบที่ผ่านการจำแนกกฎความสัมพันธ์แล้วมาประมวลผล โดยการตรวจสอบสาขาวิชาและเพศของนักศึกษาและนำมาเปรียบเทียบกับตัว

แบบก็จะได้ว่านักศึกษาจะประกอบอาชีพอะไรตามตัวแบบที่ได้ เมื่อแปลความหมายเสร็จเรียบร้อยแล้ว สามารถเรียกดูรายงานจำนวนผู้ประกอบอาชีพต่าง ๆ โดยระบุเงื่อนไขต่าง ๆ เช่น ปี หรือ อาชีพ ซึ่งรายงานก็สามารถเลือกแบบรายงานที่ออกรายงานได้จากผลการวิจัยแสดงให้เห็นว่าสามารถสร้างตัวแบบในรูปแบบกฎความสัมพันธ์ได้จริงและทดสอบความถูกต้องของตัวแบบตามสัดส่วนที่แบ่งข้อมูลเป็น 3 ช่วง คือ 60 : 40, 70 : 30 และ 80 : 20 ช่วงละ 4 ครั้ง และคิดค่าเฉลี่ยร้อยละความถูกต้องได้ผล (Train : Validation) 60 : 40, 70 : 30, 80 : 20 ความถูกต้องของตัวแบบ (ค่าเฉลี่ย) 65.39%, 74.72%, 81.89% ตามลำดับ

ในส่วนของงานวิจัย ญัฐริน เจริญเกียรติบุตรและปานใจ ธารทัศน์วงศ์ (2549) ได้ใช้การทำเหมืองข้อมูล โดยใช้เทคนิคการค้นหากฎความสัมพันธ์ (Association Rule Discovery) เพื่อช่วยในการแนะนำการศึกษาต่อระดับอุดมศึกษาจากผลการศึกษาในระดับมัธยมปลาย 7 รายวิชา คือ วิชาภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ ฟิสิกส์ เคมี และ ชีววิทยา เป็นองค์ประกอบในการคัดเลือกโดยแบ่ง Training Set : Validation เป็น 60 : 40, 70 : 30 และ 75 : 25 พบว่า ความถูกต้องเป็น 82.32 %, 81.87%, 97.78% และ 92.37% ตามลำดับ ซึ่งโมเดลนี้มีความถูกต้องเฉลี่ยเป็น 87.06% สังเกตได้ว่าสัดส่วน Training Set มากขึ้น ทำให้มีความถูกต้องของผลลัพธ์มากขึ้น เมื่อสำรวจความพึงพอใจในการใช้โปรแกรมของนักเรียนและครูพบว่ามีความพึงพอใจในระดับมากที่สุด และข้อมูลสามารถนำไปใช้ประโยชน์ในการแนะนำการศึกษาคิดเป็น 95.12%

จิราพร ยิ่งกว่าชาติและคณะ (2549) ได้นำการทำเหมืองข้อมูลมาใช้ในการทำนายผลสำเร็จ การศึกษาของนักศึกษาระดับอุดมศึกษาด้วยการใช้เทคนิคการเรียนรู้แบบเบย์ โดยใช้ข้อมูลนักศึกษา ของมหาวิทยาลัยราชภัฏอุบลราชธานี ปีการศึกษา 2534-2545 ซึ่งมีข้อมูลนักศึกษา 37,859 คน และได้คัดเลือกข้อมูลที่มีความครบถ้วนสามารถนำมาสร้างแบบจำลองได้ทั้งหมด 5,140 คน นอกจากนี้ ยังมีข้อมูลนักศึกษาของมหาวิทยาลัยศรีปทุม 800 คน และได้คัดเลือกข้อมูลที่มีความครบถ้วน สามารถนำมาสร้างแบบจำลองได้ทั้งหมด 145 คน จากผลการวิเคราะห์เบื้องต้นพบว่าตัวแปรที่มีผล ต่อการสำเร็จการศึกษาคงนี้ คือ เกรดเฉลี่ยในชั้นปีแรก อาชีพของมารดา และรายได้ของครอบครัว เมื่อนำตัวแปรดังกล่าวมาสร้างแบบจำลองด้วยเทคนิคข่ายงานเบย์จะ ได้ความแม่นยำในการทำนาย ผลถึง 91.26% อย่างไรก็ตาม ผลการวิจัยยังมีอีกตัวแปรทั้งสามว่ามีความสอดคล้องกับผลการวิเคราะห์ ความถดถอยเชิงพหุคูณ (Multiple Regression Analysis) นั่นคือ มีผลต่อการอธิบายค่าของตัวแปร ตาม ดังนั้น จึงน่าจะเชื่อว่าตัวแปรทั้งสามอัน ได้แก่ เกรดเฉลี่ยในชั้นปีแรก อาชีพของมารดา และ รายได้ของครอบครัว นี้มีความเหมาะสมในการนำมาสร้างแบบจำลองเพื่อทำนายผลสำเร็จ การศึกษาของนักศึกษา

ชลนิศา สาระ (2550) เสนอวิธีการจำแนกกลุ่มสถานภาพการสำเร็จการศึกษาโดยแบบจำลองต้นไม้ตัดสินใจ ซึ่งใช้อัลกอริทึม C4.5 ทำให้เข้าใจถึงกระบวนการ พารามิเตอร์ที่ใช้ และการวัดประสิทธิภาพในการสร้างแบบจำลองต้นไม้สำหรับอัลกอริทึม C4.5

วิโรจน์ ภาคคีรี, กาญจนา วิริยะพันธ์ (2550) ได้ทำการวิเคราะห์รูปแบบการบูรณูกข้อมูลบนเครือข่าย โดยใช้เทคนิคดาต้าไมนนิ่ง ซึ่งได้ใช้เทคนิคการคัดเลือกกลุ่มข้อมูล 3 ชนิด คือ การใช้แผนผังต้นไม้เพื่อการตัดสินใจ โดยใช้อัลกอริทึม C4.5 ซึ่งใช้กฎความสัมพันธ์ของริเปเปอร์ และการใช้นิรอลเน็ทเวิร์ค ซึ่งการใช้แผนผังต้นไม้จะมีร้อยละของความถูกต้องมากที่สุด

ยิ่งลักษณ์ ยิ่งสวัสดิ์ (2553) การพัฒนาเครื่องมือจัดกลุ่มลูกค้าโดยใช้เทคนิคดาต้าไมนนิ่ง โดยใช้ต้นไม้การตัดสินใจในการเรียนรู้ข้อมูลของลูกค้าในอดีต เพื่อทำการสร้างตัวแบบจำลองสำหรับการตัดสินใจในการจัดกลุ่มลูกค้าร้านค้าช่วง เพื่อให้สามารถพัฒนาความสัมพันธ์ และเพิ่มประสิทธิภาพในการดูแลรักษาลูกค้าให้ดียิ่งขึ้น

เมื่อผู้วิจัยได้ศึกษาแนวคิด ทฤษฎี และผลงานวิจัยที่เกี่ยวข้องดังที่ได้นำเสนอแล้วนั้น ทำให้ผู้วิจัยมีความเข้าใจในหลักการการใช้เหมืองข้อมูล ขั้นตอนการจัดทำ การจัดเตรียมข้อมูลให้สมบูรณ์ ประโยชน์ที่ได้จากการทำเหมืองข้อมูลมากขึ้น ซึ่งเป็นส่วนสำคัญที่ทำให้ผู้วิจัยนำความรู้ที่ได้เหล่านี้ไปประยุกต์ใช้ในการทำเหมืองข้อมูลเพื่อนำมาช่วยในการตัดสินใจการให้สินเชื่อ ภายในองค์กร ซึ่งจะนำเสนอในบทที่ 3 ระเบียบวิธีวิจัย และ บทที่ 4 ผลดำเนินการวิจัย ต่อไป

## บทที่ 3

### ระเบียบวิธีวิจัย

#### 3.1 ขั้นตอนการดำเนินการวิจัย

งานวิจัยนี้ จะกล่าวถึงขั้นตอนการดำเนินการวิจัย และได้้นำเทคนิคการใช้เหมืองข้อมูล มาช่วยในการตัดสินใจการให้สินเชื่อการเช่ารถยนต์ของบริษัท กรุงไทย คาร์เร็นท์ แอนด์ ลีส จำกัด (มหาชน) โดยมีขั้นตอนการดำเนินการวิจัย ดังนี้

##### 3.1.1 วิธีการเก็บ และรวบรวมข้อมูล

จากการที่ผู้วิจัยได้ทำการสัมภาษณ์ ผู้จัดการแผนกกฎหมาย ของบริษัท กรุงไทย คาร์เร็นท์ แอนด์ ลีส จำกัด (มหาชน) ผู้วิจัยพบว่า บริษัทฯ มีการเก็บข้อมูลคือประวัติของลูกค้า โดย ทำการวิเคราะห์ ข้อมูลลูกค้าจากแบบแบบวิเคราะห์และพิจารณาสินเชื่อ (CREDIT SCORING) และมีการเก็บข้อมูล ของลูกค้าแต่ละรายอยู่ในลักษณะของไฟล์ Excel จึงทำการเก็บและรวบรวมข้อมูลประวัติของลูกค้า ที่พิจารณาให้สินเชื่อแล้ว ในปี 2551–2553 จำนวน 441 ราย มาใช้ในการวิจัยครั้งนี้

##### 3.1.2 การเตรียมข้อมูล และคัดเลือกข้อมูลที่จะนำมาใช้

หลังจากที่ผู้วิจัยได้รวบรวมข้อมูลแล้ว ในขั้นตอนแรก ต้องมีการเตรียมข้อมูลและคัดเลือกข้อมูลที่จะนำมาใช้ เนื่องจากข้อมูลมีรูปแบบที่ยังไม่สามารถนำไปใช้ ผ่านกระบวนการทดสอบและวิเคราะห์ด้วยเครื่องมือการทำเหมืองข้อมูลได้ จึงต้องมีการเตรียมข้อมูลให้เป็นไปตามรูปแบบมาตรฐานเดียวกัน และคัดเลือกเฉพาะข้อมูลที่สมบูรณ์และรูปแบบถูกต้อง ครบถ้วน และลดความหลากหลายของข้อมูล เพื่อให้ได้ผลการทดสอบที่มีประสิทธิภาพสูงสุด ได้แก่ขั้นตอนการโอนย้ายข้อมูล (Data Transfer) การลดขนาดของข้อมูล (Data Reduction) และการทำความสะอาดข้อมูล (Data Cleaning)

##### 3.1.3 การทดสอบข้อมูล และเลือกเทคนิคที่เหมาะสม

เป็นการนำข้อมูลที่ได้ออกแล้วจากขั้นตอนที่ผ่านมา จึงนำข้อมูลดังกล่าว มาทดสอบกับโปรแกรม Weka ซึ่งต้องเลือกเทคนิคการจำแนกกลุ่มข้อมูลที่เหมาะสม โดยแต่ละเทคนิคจะให้ผลลัพธ์ออกมาในรูปแบบที่แตกต่างกัน ซึ่งในการศึกษาวิจัยครั้งนี้ต้องการนำโมเดลที่ได้จากการเลือกเทคนิคที่เหมาะสม มาใช้ในการวิเคราะห์เพื่ออนุมัติให้สินเชื่อลูกค้า และนำไปพัฒนาระบบสนับสนุนการตัดสินใจในการให้สินเชื่อลูกค้า ซึ่งได้มีการทดสอบจำนวน 3 เทคนิค โดย



รายละเอียดแต่ละเทคนิคได้อธิบายไว้ในบทที่ 2 และจะแสดงรายละเอียดของผลลัพธ์ในแต่ละเทคนิคในบทที่ 4

#### 3.1.4 ผลการวิเคราะห์

เป็นการวิเคราะห์ผลลัพธ์ที่ได้จากการทดสอบของเทคนิคที่นำมาทดสอบ จำนวน 3 เทคนิค โดยนำผลลัพธ์ที่ได้ของแต่ละเทคนิคมาวิเคราะห์ความถูกต้องจากข้อมูลที่นำมาทดสอบ เพื่อทำการเลือกกฎ ที่มีความน่าเชื่อถือมากที่สุด จากกฎที่มีค่าความเป็นจริงมากที่สุด เพื่อใช้เป็นเงื่อนไขในการพัฒนาระบบสนับสนุนการตัดสินใจเพื่อการอนุมัติสินเชื่อ ในขั้นตอนต่อไป

#### 3.1.5 พัฒนาระบบสนับสนุนการตัดสินใจอนุมัติสินเชื่อ

เป็นการนำกฎที่ได้จากการทำเหมืองข้อมูลมาเป็นเงื่อนไขในการพัฒนาโปรแกรมให้เป็นเครื่องมือช่วยในการตัดสินใจเพื่อการอนุมัติสินเชื่อ

#### 3.1.6 สรุปผลการวิจัยและข้อเสนอแนะ

เป็นการกล่าวถึงผลการวิจัย การใช้เหมืองข้อมูลช่วยในการตัดสินใจการให้สินเชื่อ นอกจากสามารถนำมาพัฒนาเป็นระบบอนุมัติสินเชื่อแล้ว ยังสามารถนำการใช้เหมืองข้อมูลมาพัฒนาเพื่อช่วยในการตัดสินใจสำหรับผู้บริหาร เพื่อช่วยในการบริหารด้านอื่น ๆ อีกด้วย เช่น การหาความสัมพันธ์ของข้อมูลด้านความพึงพอใจของลูกค้า การทำนายว่าลูกค้าจะเช่ารถต่อหรือไม่ เป็นต้น

#### 3.1.7 เรียบเรียงงานค้นคว้าอิสระ

### 3.2 เครื่องมือที่ใช้ในการวิจัย

3.2.1 ระบบปฏิบัติการ Windows XP Professional Version 2002 pack 3

3.2.2 โปรแกรม Microsoft Excel ใช้ในการคัดเลือกข้อมูลที่จะนำมาใช้ เตรียมข้อมูล สำหรับนำเข้าทดสอบในโปรแกรม Weka

3.2.3 โปรแกรม CS 4, PHP, Apache และฐานข้อมูล My SQL พัฒนาระบบสนับสนุนการตัดสินใจอนุมัติสินเชื่อ

3.2.4 โปรแกรม Microsoft Word ใช้ในการจัดทำเอกสารงานค้นคว้าอิสระ

## บทที่ 4

### ผลดำเนินการวิจัย



จากวิธีการดำเนินงานที่ผู้ศึกษาได้ทำการนำเสนอไปในบทที่ 3 นั้นได้ผลการดำเนินวิจัยในแต่ละขั้นตอนดังนี้

1. วิธีการเก็บ และรวบรวมข้อมูล
2. การเตรียมข้อมูล และคัดเลือกข้อมูลที่จะนำมาใช้
3. การทดสอบข้อมูล และเลือกเทคนิคที่เหมาะสม
4. ผลการวิเคราะห์
5. พัฒนาระบบสนับสนุนการตัดสินใจอนุมัติสินเชื่อ

#### 4.1 วิธีการเก็บ และรวบรวมข้อมูล

งานวิจัยครั้งนี้ได้รวบรวมข้อมูลประวัติของลูกค้าที่พิจารณาให้สินเชื่อ จากแบบวิเคราะห์และพิจารณาสินเชื่อเพื่อเช่ารถยนต์ ( CREDIT SCORING ) จัดทำโดยแผนกกฎหมายของบริษัท กรุงไทย คาร์เร็นท์ แอนด์ ลีส จำกัด (มหาชน) ตามภาพที่ 4.1

ข้อมูลที่น่ามาใช้ในการศึกษานี้ ได้เก็บรวบรวมข้อมูลลูกค้าผ่านการอนุมัติแล้ว ตั้งแต่ปี พ.ศ. 2551-2553 มีจำนวนลูกค้ารวม 441 ราย โดยข้อมูลที่จัดเก็บมีทั้งหมด 30 แอทริบิวต์ แต่ข้อมูลที่มีความจำเป็นต้องใช้ในการประกอบการพิจารณาให้สินเชื่อมีจำนวน 17 แอทริบิวต์ ประกอบด้วย แอทริบิวต์ ดังนี้คือ วัตถุประสงค์การใช้งานรถยนต์, ประเภทรถยนต์, ประเภทธุรกิจ, ประเภทลูกค้า, ระยะเวลาดำเนินธุรกิจ, ทุนจดทะเบียน, กำไรสุทธิ รายได้หลัก, ส่วนผู้ถือหุ้น, อัตราสินทรัพย์หมุน/หนี้สินหมุนเวียน, อัตราหนี้สินรวม/สินทรัพย์รวม, อัตราหนี้สินรวม/ส่วนผู้ถือหุ้น, อัตราการเพิ่มของยอดขาย, อัตราการเพิ่มของกำไรสุทธิ, วงเงินอนุมัติ, เกรดลูกค้า และสถานะลูกค้า โดยรวบรวมข้อมูลลูกค้าทั้งหมดให้อยู่ในรูปแบบฐานข้อมูลเดียวกัน

 <b>แบบวิเคราะห์และพิจารณาสินเชื่อเพื่อเช่ารถยนต์ (CREDIT SCORING)</b>					เลขที่ CD-09024 วันที่ 14/2/51	
บริษัท ปลอดภัย จำกัด					ลูกค้านี้เกรด <input type="checkbox"/> A <input type="checkbox"/> B <input type="checkbox"/> C <input checked="" type="checkbox"/> D	
ประเภทธุรกิจ ขายเป็นจักรยานยนต์/อะไหล่และบริการซ่อม						
ประเภทรถที่ใช้ รถกระบะ					วัตถุประสงค์การใช้ขนส่งสินค้า	
					เจ้าหน้าที่การตลาด คุณอริสรา	
กลุ่มธุรกิจ	A	B	C	คะแนน	D	
1. ประเภทธุรกิจ	1.ผลงานและสารบัญช	13.ธุรกิจการเกษตร	25.บริการขนส่ง,ซ่อมบำรุง,กำจัดขยะ		33.ธุรกิจศึกษา	
	2. สื่อสารและเทคโนโลยี	14. อาหารและเครื่องดื่ม	26. การท่องเที่ยวและโรงแรม		34. ธุรกิจรถเช่า	
	3. ธนาคารและเงินทุนหลักทรัพย์	15. ของใช้ครัวเรือน	27. การขนส่งและโลจิสติกส์		35. รับเหมานาคเล็กทุนต่ำกว่าล้านบาท	
	4. ประกันภัยและประกันชีวิต	16. ของใช้ส่วนตัวและเวชภัณฑ์	28. ธุรกิจดูแลความสะอาด			
	5. อุตสาหกรรมยานยนต์	17. แฟชั่น	29. ที่ปรึกษาด้านธุรกิจ			
	6. เครื่องมือ-เครื่องจักร	18. บรรจุภัณฑ์	30. การศึกษา			
	7. ปีโตรเคมีและเคมีภัณฑ์	19. กระดาษและวัสดุการพิมพ์	31. มูลนิธิ			
	8. ชิ้นส่วนอิเล็กทรอนิกส์	20. วัสดุก่อสร้าง	32. บุคคล			
	9. บริการพาณิชย์	21. กิจานอสังหาริมทรัพย์ขนาดใหญ่				
	10. การแพทย์	22. เหมืองแร่				
	11. ธุรกิจสื่อและสิ่งพิมพ์	23. เครื่องใช้ไฟฟ้า				
	12. ราชการและรัฐวิสาหกิจ	24. คอมพิวเตอร์				
	<input checked="" type="checkbox"/> 5 คะแนน	<input type="checkbox"/> 4 คะแนน	<input type="checkbox"/> 2 คะแนน	5	<input type="checkbox"/> 0 คะแนน	
2. ประเภทลูกค้า	<input type="checkbox"/> ลูกค้าเก่าชำระดี=>ปี=( 5 )	<input type="checkbox"/> ลูกค้าเก่าชำระดี=( 4 )	<input checked="" type="checkbox"/> ลูกค้าใหม่=( 3 )	3	<input type="checkbox"/> ลูกค้าเก่าชำระไม่ดี(0)	
3. ดำเนินธุรกิจมานาน	<input type="checkbox"/> >10ปี= ( 15 )	<input type="checkbox"/> 6-10ปี= ( 14 )	<input checked="" type="checkbox"/> 3-5ปี= ( 13 )	13	<input type="checkbox"/> < 3ปี= ( 0 )	
4. ทุนจดทะเบียน (ชำระแล้ว)	<input type="checkbox"/> >100ล้านบาท= ( 10 )	<input type="checkbox"/> 61-100ล้านบาท= ( 9 )	<input type="checkbox"/> 3-60ล้านบาท= ( 8 )	0	<input checked="" type="checkbox"/> <3ล้านบาท=( 0 )	
5. ค่าไรสุทธิ	<input type="checkbox"/> ค่าไร 3 ปี PPP=(15)	<input type="checkbox"/> ค่าไร 2 ปีล่าสุด PPL=(14)	<input type="checkbox"/> สลิปกัน LPP, LPL, PLP=(13)	0	<input checked="" type="checkbox"/> ขาดทุนต่อเนื่องLLL(0)	
6. รายได้หลัก	<input type="checkbox"/> เพิ่มขึ้นต่อเนื่อง 2 ปี= ( 15 )	<input type="checkbox"/> เพิ่มขึ้นต่อเนื่อง 1 ปี= ( 14 )	<input checked="" type="checkbox"/> เพิ่ม-ลดสลับกัน=( 13 )	13	<input type="checkbox"/> ลดลงต่อเนื่อง= ( 0 )	
7. ส่วนผู้ถือหุ้น	<input type="checkbox"/> เพิ่มขึ้นต่อเนื่อง 2 ปี= ( 10 )	<input type="checkbox"/> เพิ่มขึ้นต่อเนื่อง 1 ปี= ( 9 )	<input type="checkbox"/> เพิ่ม-ลดสลับกัน= ( 8 )	0	<input checked="" type="checkbox"/> ลดลงต่อเนื่องปี (0)	
8. สินทรัพย์หมุนเวียน/หนี้สินหมุนเวียน	<input type="checkbox"/> ค่าเฉลี่ย 3 ปี > 1 เท่า=( 5 )	<input type="checkbox"/> ค่าเฉลี่ย 0.76-1 เท่า=( 4 )	<input checked="" type="checkbox"/> ค่าเฉลี่ย 0-0.75 เท่า=( 3 )	3	<input type="checkbox"/> ค่าติดลบ=( 0 )	
9. หนี้สินรวม/สินทรัพย์รวม (%)	<input type="checkbox"/> ค่าเฉลี่ย< 50%=( 5 )	<input checked="" type="checkbox"/> ค่าเฉลี่ย50-150%=( 4 )	<input type="checkbox"/> ค่าเฉลี่ย 151-250%=( 3 )	4	<input type="checkbox"/> >250%=( 0 )	
10. หนี้สินรวม/ส่วนของผู้ถือหุ้น (%)	<input type="checkbox"/> ค่าเฉลี่ย< 100%=( 5 )	<input type="checkbox"/> ค่าเฉลี่ย100-200%=( 4 )	<input type="checkbox"/> ค่าเฉลี่ย 201-300%=( 3 )	0	<input checked="" type="checkbox"/> >300%=( 0 )	
 <b>แบบวิเคราะห์และพิจารณาสินเชื่อเพื่อเช่ารถยนต์ (CREDIT SCORING)</b>						
กลุ่มธุรกิจ	A	B	C	คะแนน	D	
11. อัตราการเพิ่มของยอดขาย (%)	<input checked="" type="checkbox"/> ค่าเฉลี่ย> 10%=( 5 )	<input type="checkbox"/> ค่าเฉลี่ย 6-10%=( 4 )	<input type="checkbox"/> ค่าเฉลี่ย 1-5%=( 3 )	5	<input type="checkbox"/> ค่าเฉลี่ย ติดลบ=(0)	
12. อัตราการเพิ่มของกำไรสุทธิ (%)	<input type="checkbox"/> ค่าเฉลี่ย> 10%=( 5 )	<input type="checkbox"/> ค่าเฉลี่ย 6-10%=( 4 )	<input type="checkbox"/> ค่าเฉลี่ย 1-5%=( 3 )	0	<input checked="" type="checkbox"/> ค่าเฉลี่ย ติดลบ=(0)	
				รวมคะแนน	46	

ภาพที่ 4.1 แสดงแบบวิเคราะห์และพิจารณาสินเชื่อเพื่อเช่ารถยนต์ (CREDIT SCORING)

เกรดลูกค้า	A = คะแนนรวมตั้งแต่ 89-100 คะแนน	<input type="checkbox"/>			
	B = คะแนนรวมตั้งแต่ 76 - 88 คะแนน	<input type="checkbox"/>			
	C = คะแนนรวมตั้งแต่ 60 - 75 คะแนน	<input type="checkbox"/>			
	D = คะแนนรวมต่ำกว่า 60 คะแนน	<input checked="" type="checkbox"/>	→	พิจารณาครั้งที่ 2	
เอกสารประกอบการพิจารณา	<input checked="" type="checkbox"/> สำเนาหนังสือรับรอง	<input checked="" type="checkbox"/> ภพ. 01 และ ภพ.20	<input checked="" type="checkbox"/> BOL		
4. สัญญาว่าจ้างหรือโครงการในอนาคต.....					
5. ค่าประกันโดยธนาคาร.....					
6. ค่าประกันโดยผู้ถือหุ้นใหญ่.....					
7. ประเภทรถที่เช่า.....					
ผลการพิจารณาครั้งที่ 2.....					
.....					
.....					
วิธีพิจารณาเงิน	<input type="checkbox"/> 1	<input type="checkbox"/> 2	อำนาจอนุมัติ		
			มูลค่าสัญญา < 3 ล้านบาท	.....	DM.MKT.
ผลการพิจารณา	<input checked="" type="checkbox"/> อนุมัติ	<input type="checkbox"/> ไม่อนุมัติ			DM.CD.
วงเงินคงเหลือ	0	บาท	มูลค่าสัญญา 3-6 ล้านบาท	.....	AMD.
			มูลค่าสัญญา 50-100 ล้านบาท	.....	EX.COM.
			มูลค่าสัญญา >100 ล้านบาท	.....	BOD.
มูลค่าสัญญา	234,000	บาท	มูลค่าสัญญา 20-50 ล้านบาท	.....	MD.

ภาพที่ 4.1 (ต่อ)

## 4.2 การเตรียมข้อมูล และคัดเลือกข้อมูลที่จะนำมาใช้

เนื่องจากการเก็บข้อมูลของบริษัทฯ ได้จัดเก็บไว้ในรูปแบบที่ไม่สามารถนำมาใช้ในการทดสอบและวิเคราะห์ด้วยเครื่องมือการทำเหมืองข้อมูลได้ จึงต้องมีการเตรียมข้อมูล และคัดเลือกข้อมูลที่จะนำมาใช้ในการทำเหมืองข้อมูล เพื่อให้ได้ผลการทดสอบที่มีประสิทธิภาพสูงสุด โดยประกอบด้วย 3 ขั้นตอน คือ การโอนย้ายข้อมูล (Data Transfer) การลดขนาดของข้อมูล (Data Reduction) และการทำความสะอาดข้อมูล (Data Cleaning)

### 4.2.1 การโอนย้ายข้อมูล (Data Transfer)

การโอนย้ายข้อมูลเป็นการโอนข้อมูลเดิม ที่ถูกจัดเก็บอยู่ในรูปแบบไฟล์ Excel ที่มีรูปแบบการกรอกข้อมูลลงในแบบฟอร์มที่กำหนดไว้ล่วงหน้า ของลูกค้าแต่ละราย ทำให้ข้อมูลของลูกค้า ไม่ถูกจัดเก็บ หรือรวบรวมไว้ในฐานข้อมูลเดียวกัน จึงจำเป็นต้องมีการโอนย้ายข้อมูลของลูกค้าแต่ละราย ให้อยู่ในรูปแบบฐานข้อมูลเดียวกัน ซึ่งใช้ excel ในการสร้างเป็นฐานข้อมูล และบันทึกข้อมูลให้อยู่ในรูปแบบ .CSV เพื่อให้สามารถนำไปใช้งานได้ และสะดวกต่อการนำเข้าข้อมูลในโปรแกรม Weka ตามภาพที่ 4.1 ข้อมูลของลูกค้า ไม่ถูกจัดเก็บ หรือรวบรวมไว้ในฐานข้อมูล

เดียวกัน จึงจำเป็นต้องมีการโอนย้ายข้อมูลของลูกค้าย่อยราย ให้อยู่ในรูปแบบฐานข้อมูลเดียวกัน และตามภาพที่ 4.2

	A	B	C	D	E	F	J	L	O
1	เกรด	เลขที่	วัน	เดือน	ปี	รหัส	ประเภทธุรกิจ	วัตถุประสงค์การใช้	ประเภทธุรกิจ
2	D	CD-08024	14	2	2008	H1808	ขายรถจักรยานยนต์/อะไหล่และบริการซ่อม	ขนส่งสินค้า	5
3	A	CD-08025	15	2	2008	H5874	ผลิตชิ้นส่วนรถยนต์	รถสวนกลาง	5
4	C	CD-08026	15	2	2008	H5871	ผลิตและส่งออกอุปกรณ์รถยนต์ประเภทไฟในเตาแก๊ส	รถผู้บริหาร	5
5	A	CD-08027		2	2008	H5886	ดำเนินการบริการซ่อมบำรุงปรังและทดสอบมอเตอร์อุตสาหกรรม		5
6	D	CD-08028	22	2	2008	A7715*	บริการรถเช่าและท่องเที่ยว		0
7	C	CD-08029	27	2	2008	H5432	ที่ปรึกษาเกี่ยวกับการบริหารงานพาณิชย์	รถผู้บริหาร	2
8	D	CD-08030	27	2	2008	H5893	อุปกรณ์ไหลอมและทำความร้อนสำหรับเครื่องแปลงกระแสไฟฟ้า	รถผู้บริหาร	5
9	B	CD-08031	26	2	2008	xH5788	ขายเครื่องใช้ไฟฟ้าภายในประเทศ	สำหรับผู้บริหาร	4
10	A	CD-08032	27	2	2008	H5897	ผลิตและจำหน่ายตั้งไม้โอ๊ค	ใช้เป็นรถสวนกลาง	4
11	B	CD-08033	28	2	2008	C-H5895	จำหน่ายกาแฟ	ผู้จัดการสาขา	4
12	A	CD-08034	28	2	2008	H5906	ผลิตน้ำเข้าและจำหน่ายสารเคมี	สำหรับผู้บริหารระดับล่าง	5
13	A	CD-08035	28	2	2008	H5905	ผลิตเพื่อขายและส่งออกชิ้นส่วนเครื่องรับโทรทัศน์	สำหรับผู้บริหาร	5
14	A	CD-08036	29	2	2008	H5904	รับจ้างตรวจเช็คสารตรวจแ	...รับจ้างงานตามเหมือง	4
15	B	CD-08037	3	3	2008	R2253	ผลิตเสื้อขึ้นใน	ให้ผู้จัดการ	4
16	A	CD-08038	3	3	2008	C-H5915	ผลิตยา	รถผู้บริหาร	4
17	A	CD-08039	4	3	2008	H5896	ขายคอมพิวเตอร์	การตลาด	4
18	B	CD-08041	6	3	2008	H5913	ชิ้นส่วนยานยนต์	รถผู้จัดการ	5
19	B	CD-08042	12	3	2008	H5922	ให้บริการโครงการผลิตกระแสไฟฟ้า	ระดับผู้บริหาร	5
20	B	CD-08043	14	3	2008	C-H5933	ธนาคาร	สำหรับผู้จัดการสาขา	5
21	A	CD-08044	14	3	2008	H5845	อุตสาหกรรมปูนสำเร็จรูป	สวนกลาง	4
22	B	CD-08045	14	3	2008	H5923	ประกันภัย	วังในกรุงเทพ บริเวณทล	5
23	B	CD-08046	19	3	2008	H5930	จำหน่ายยาเวชภัณฑ์	วังในกทม	4

ภาพที่ 4.2 การโอนย้ายข้อมูลจากแบบฟอร์มนำมาอยู่ในรูปแบบเดียวกันในไฟล์ Excel ที่สร้างเป็นฐานข้อมูล

#### 4.2.2 การลดขนาดของข้อมูล (Data Reduction)

การลดขนาดของข้อมูล เพื่อให้ได้ข้อมูลที่สมบูรณ์ และมีรูปแบบที่สามารถนำไปทดลองการทำเหมืองข้อมูลกับโปรแกรม Weka ได้ จึงต้องทำการลดขนาดของข้อมูล 3 ส่วน คือ ส่วนที่ 1. ลดจำนวนแอทริบิวต์ โดยข้อมูลที่รวบรวมมาจากแบบวิเคราะห์และพิจารณาสินเชื่อเพื่อเช่ารถยนต์ (CREDIT SCORING) มีทั้งหมดจำนวน 30 แอทริบิวต์ แต่ข้อมูลที่มีความจำเป็นต้องใช้ในการประกอบการพิจารณาให้สินเชื่อมีจำนวน 17 แอทริบิวต์ จึงลดจำนวนแอทริบิวต์ส่วนที่ไม่จำเป็นออก คงเหลือแอทริบิวต์ ดังนี้คือ วัตถุประสงค์การใช้งานรถยนต์ ประเภทรถยนต์ ประเภทธุรกิจ ประเภทลูกค้า ระยะเวลาดำเนินธุรกิจ ทุนจดทะเบียน กำไรสุทธิ รายได้หลัก ส่วนผู้ถือหุ้น อัตราสินทรัพย์หมุน : หนี้สินหมุนเวียน อัตราหนี้สินรวม : สินทรัพย์รวม อัตราหนี้สินรวม : ส่วนผู้ถือหุ้น อัตราการเพิ่มของยอดขาย อัตราการเพิ่มของกำไรสุทธิ วงเงินอนุมัติ เกรดลูกค้า และสถานะลูกค้า ส่วนที่ 2. ลดข้อมูลที่มีแอทริบิวต์ไม่ครบถ้วน ซึ่งหลังจากลดจำนวนแอทริบิวต์แล้ว ต้องตรวจสอบความครบถ้วนของข้อมูล ซึ่งข้อมูลลูกค้าบางราย มีข้อมูลไม่ครบทุกแอทริบิวต์ จึงต้องคัดเลือกข้อมูลเฉพาะรายที่ระบุเบียบมีความสมบูรณ์ครบถ้วน โดยหลังจากคัดเลือกแล้วคงเหลือข้อมูลที่สมบูรณ์จำนวน 441 ราย และส่วนที่ 3. ลดความหลากหลายของแอทริบิวต์ ได้แก่ วงเงินอนุมัติ ซึ่ง

ลูกค้าแต่ละรายจะได้รับอนุมัติในวงเงินที่ไม่เท่ากัน จึงต้องจัดกลุ่มของวงเงินให้คงเหลือเพียง 4 กลุ่ม  
ได้แก่

กลุ่มที่ 1 1-10,000,000

กลุ่มที่ 2 10,000,001-50,000,000

กลุ่มที่ 3 50,000,001-100,000,000

กลุ่มที่ 4 100,000,001-1,000,000,000

#### 4.2.3 การทำความสะอาดข้อมูล (Data Cleaning)

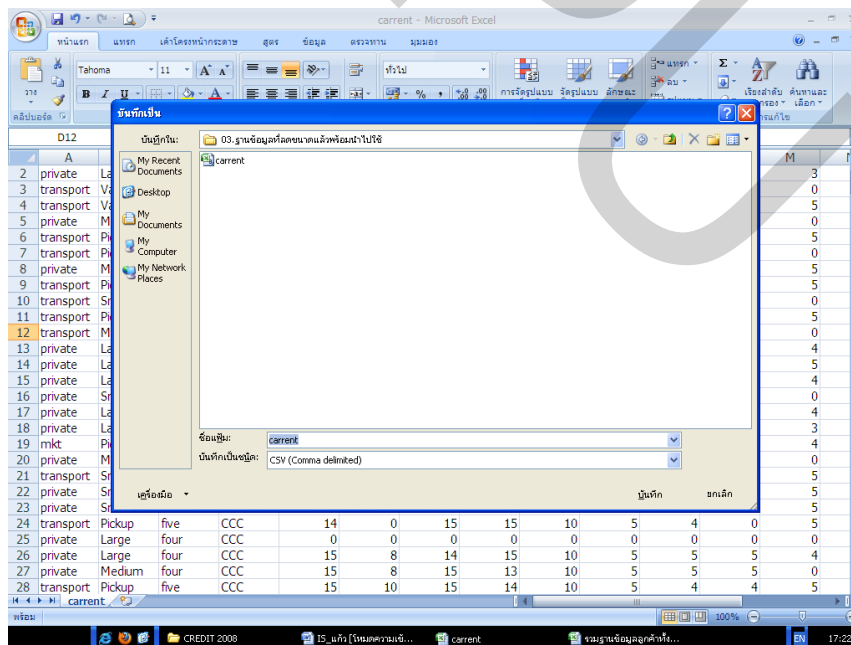
แก้ค่าของข้อมูลที่มีความหมายเหมือนกันให้เป็นมาตรฐานเดียวกัน ตามตารางภาพที่ 4.1

#### ตารางที่ 4.1 การแปลงค่าของข้อมูลวัตถุประสงค์การใช้งาน

ข้อมูลเดิม	ข้อมูลใหม่
ขนส่งสินค้า	transport
งานบริการ-ขนส่งสินค้า	
ใช้บรรทุกของ	
พนักงาน	private
ผู้จัดการสาขา	
ประจำตำแหน่ง	
ผู้บริหาร	
การตลาด	mkt.
ติดต่อลูกค้า	

	A	B	C	D	E	F	G	H	I	J	K	L	M
2	private	Large	four	CCC	15	10	15	15	10	5	5	5	3
3	transport	Van	two	CCC	0	8	0	0	0	0	0	0	0
4	transport	Van	four	CCC	13	8	15	13	10	5	4	0	5
5	private	MPV	five	CCC	0	8	0	0	0	0	0	0	0
6	transport	Pickup	five	CCC	14	10	15	13	10	5	5	5	5
7	transport	Pickup	five	CCC	13	0	0	0	0	4	3	0	0
8	private	Medium	five	CCC	13	8	15	15	8	5	5	5	5
9	transport	Pickup	four	CCC	15	9	15	15	10	5	5	5	5
10	transport	Small	five	CCC	14	8	13	0	8	5	5	5	0
11	transport	Pickup	five	CCC	15	8	15	15	10	5	4	4	5
12	transport	MPV	two	CCC	0	8	0	0	0	0	0	0	0
13	private	Large	two	CCC	15	10	13	15	8	4	4	0	4
14	private	Large	four	CCC	14	8	13	13	0	5	4	4	5
15	private	Large	four	CCC	14	8	13	15	8	5	4	4	4
16	private	Small	two	CCC	0	0	15	0	0	0	0	0	0
17	private	Large	five	CCC	15	0	15	13	8	5	5	5	4
18	private	Large	four	EEE	14	8	15	13	8	5	4	4	3
19	mkt	Pickup	five	CCC	15	8	15	13	10	5	4	0	4
20	private	Medium	four	CCC	13	8	13	13	8	5	5	4	0
21	transport	Small	two	CCC	14	8	15	15	10	5	4	0	5
22	private	Small	five	CCC	14	8	15	15	10	5	4	4	5
23	private	Small	four	CCC	15	8	15	15	10	5	4	3	5
24	transport	Pickup	five	CCC	14	0	15	15	10	5	4	0	5
25	private	Large	four	CCC	0	0	0	0	0	0	0	0	0
26	private	Large	four	CCC	15	8	14	15	10	5	5	5	4
27	private	Medium	four	CCC	15	8	15	13	10	5	5	5	0
28	transport	Pickup	five	CCC	15	10	15	14	10	5	4	4	5

ภาพที่ 4.3 ผลการแปลงค่าของข้อมูลวัตถุประสงค์การใช้งาน



ภาพที่ 4.4 การบันทึกข้อมูลในรูปแบบนามสกุล .csv

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Object	Car typ	Bus Typ	CusTyp	Long Run	Capital	NP	Income	OE	Cur ratio	Debt Ratio	DE	Growth Inc Grow
2	private	Large	four	CCC	15	10	15	15	10	5	5	5	3
3	transport	Van	two	CCC	0	8	0	0	0	0	0	0	0
4	transport	Van	four	CCC	13	8	15	13	10	5	4	0	5
5	private	MPV	five	CCC	0	8	0	0	0	0	0	0	0
6	transport	Pickup	five	CCC	14	10	15	13	10	5	5	5	5
7	transport	Pickup	five	CCC	13	0	0	0	0	4	3	0	0
8	private	Medium	five	CCC	13	8	15	15	8	5	5	5	5
9	transport	Pickup	five	CCC	15	9	15	15	10	5	5	5	5
10	transport	Small	five	CCC	14	8	13	0	8	5	5	5	0
11	transport	Pickup	five	CCC	15	8	15	15	10	5	4	4	5
12	transport	MPV	two	CCC	0	8	0	0	0	0	0	0	0
13	private	Large	two	CCC	15	10	13	15	8	4	4	0	4
14	private	Large	four	CCC	14	8	13	13	0	5	4	4	5
15	private	Large	four	CCC	14	8	13	15	8	5	4	4	4
16	private	Small	two	CCC	0	0	15	0	0	0	0	0	0
17	private	Large	five	CCC	15	0	15	13	8	5	5	5	4
18	private	Large	four	EEE	14	8	15	13	8	5	4	4	3
19	mkt	Pickup	five	CCC	15	8	15	13	10	5	4	0	4
20	private	Medium	four	CCC	13	8	13	13	8	5	5	4	0
21	transport	Small	two	CCC	14	8	15	15	10	5	4	0	5
22	private	Small	five	CCC	14	8	15	15	10	5	4	4	5
23	private	Small	four	CCC	15	8	15	15	10	5	4	3	5
24	transport	Pickup	five	CCC	14	0	15	15	10	5	4	0	5
25	private	Large	four	CCC	0	0	0	0	0	0	0	0	0
26	private	Large	four	CCC	15	8	14	15	10	5	5	5	4
27	private	Medium	four	CCC	15	8	15	13	10	5	5	5	0

ภาพที่ 4.5 แสดงไฟล์ข้อมูลในรูปแบบนามสกุล .csv

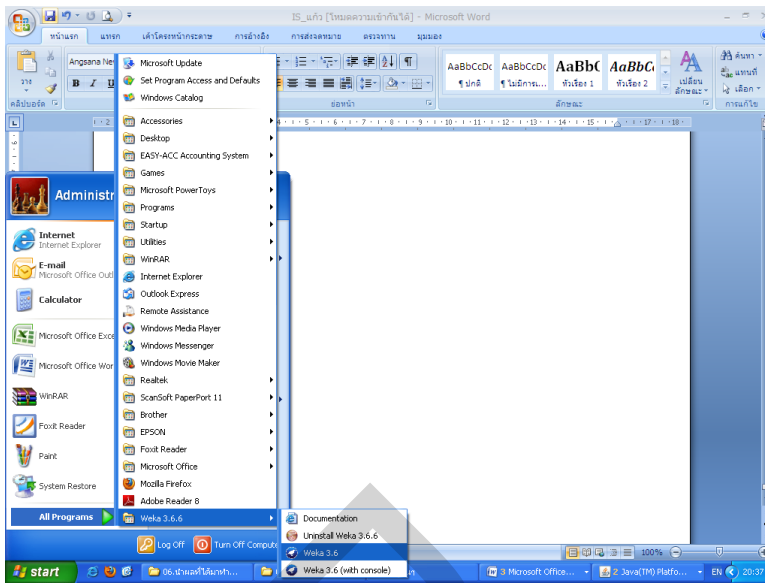
### 4.3 การทดสอบข้อมูล และเลือกเทคนิคที่เหมาะสม

#### 4.3.1 การทดสอบข้อมูล

หลังจากเตรียมข้อมูลที่มีความสมบูรณ์ ครบถ้วน และลดความหลากหลายของข้อมูลแล้ว จึงนำข้อมูลดังกล่าว มาทดสอบกับโปรแกรม Weka ซึ่งมีขั้นตอนดังนี้

ขั้นตอนที่ 1. เปิดโปรแกรม ตามภาพที่ 4.6 และแสดงหน้าโปรแกรมตามภาพที่ 4.7



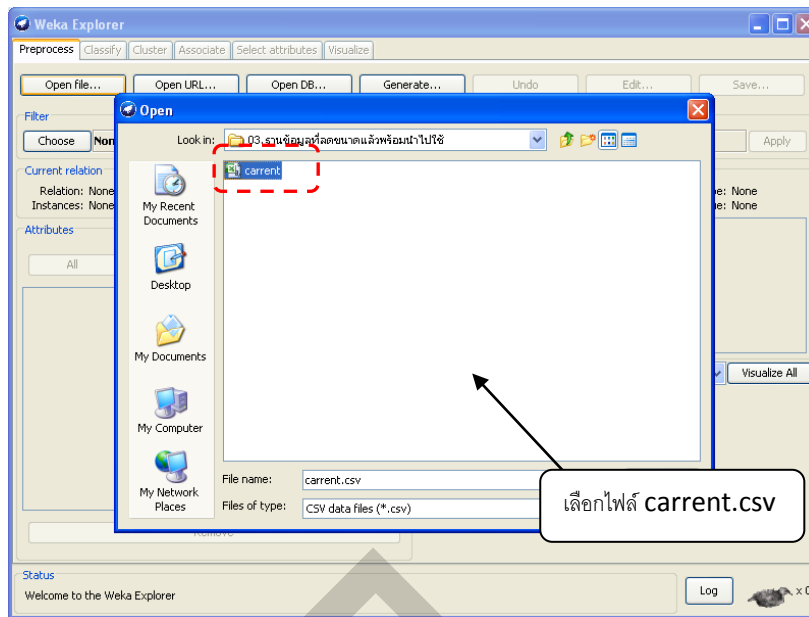


ภาพที่ 4.6 แสดงการเปิดโปรแกรม Weka 3.6.6

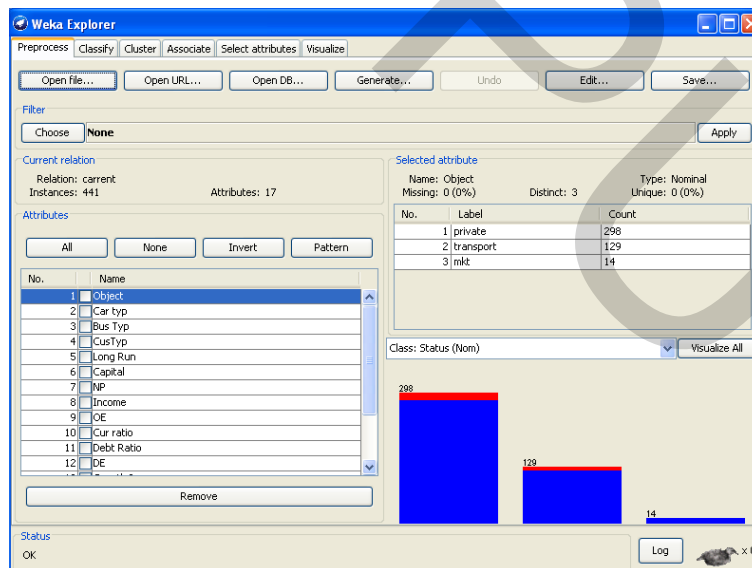


ภาพที่ 4.7 แสดงการเข้าหน้าโปรแกรม Weka 3.6.6

ขั้นตอนที่ 2. นำเข้าข้อมูลที่จัดเตรียมไว้ โดยการ เลือก Application >> Explorer >> Open file เลือกไฟล์ข้อมูลที่ต้องการนำเข้า ตามภาพที่ 4.8 และ หลังจากนั้น จะแสดงหน้าจอข้อมูลตามภาพที่ 4.9 ซึ่งหน้าจอนี้จะทำการวิเคราะห์หาจำนวนของข้อมูลที่ใช้ในการวิจัย เพื่อให้ทราบจำนวนเปอร์เซ็นต์ ต่อข้อมูลทั้งหมด และค่าเฉลี่ยของข้อมูลในแต่ละแอทริบิวต์

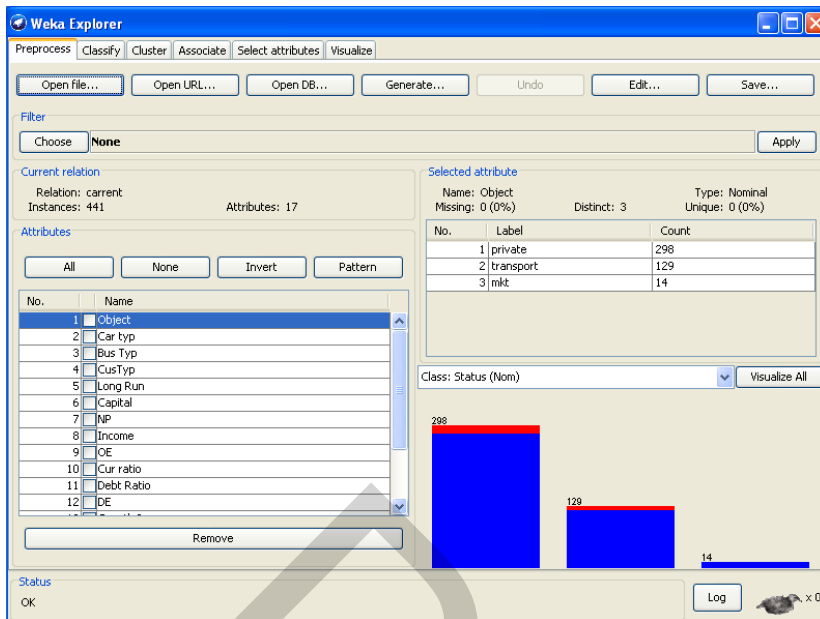


ภาพที่ 4.8 การนำเข้าข้อมูลในโปรแกรม Weka

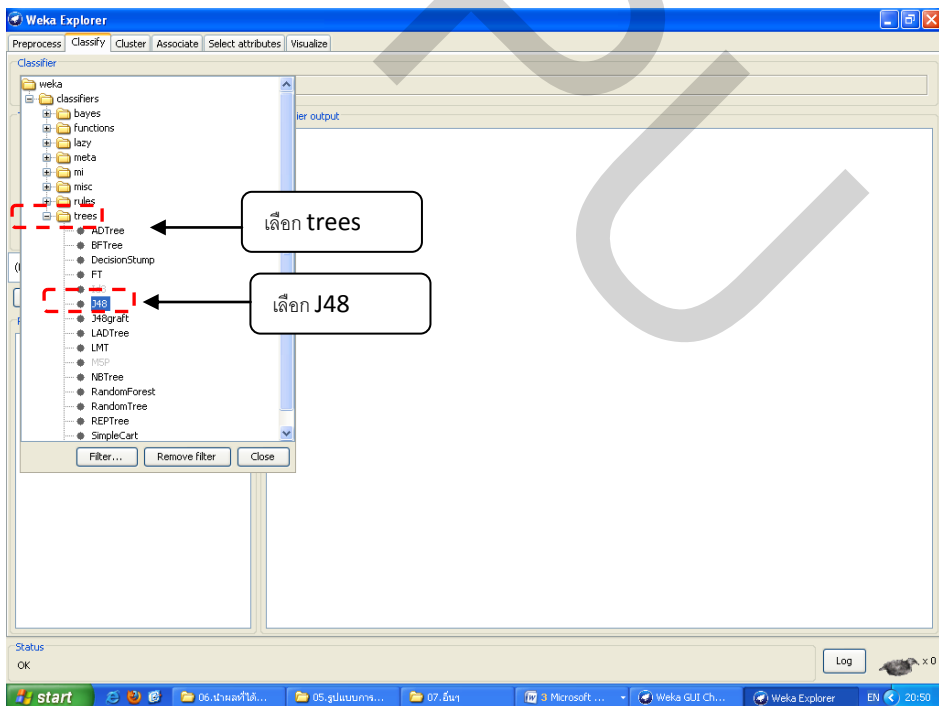


ภาพที่ 4.9 หน้าจอการแสดงผลข้อมูลที่นำเข้า

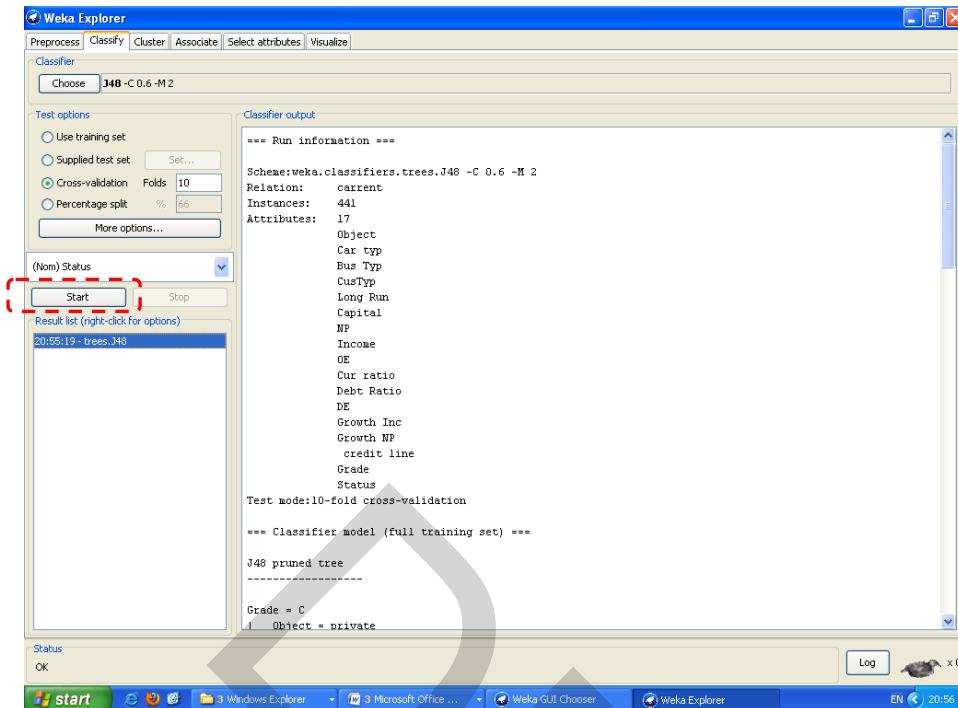
ขั้นตอนที่ 3. เลือกเทคนิคที่ต้องการ ได้แก่ การจัดกลุ่มข้อมูลโดยการเลือก Classify >> Choose (ตามภาพ 4.10) >> trees >> เลือกเทคนิคการจำแนกข้อมูลที่ต้องการ เช่น J48 (ตามภาพ 4.11) กดปุ่ม Start (ตามภาพ 4.12) และจะแสดงผลลัพธ์ที่ได้ตามภาพ 4.13



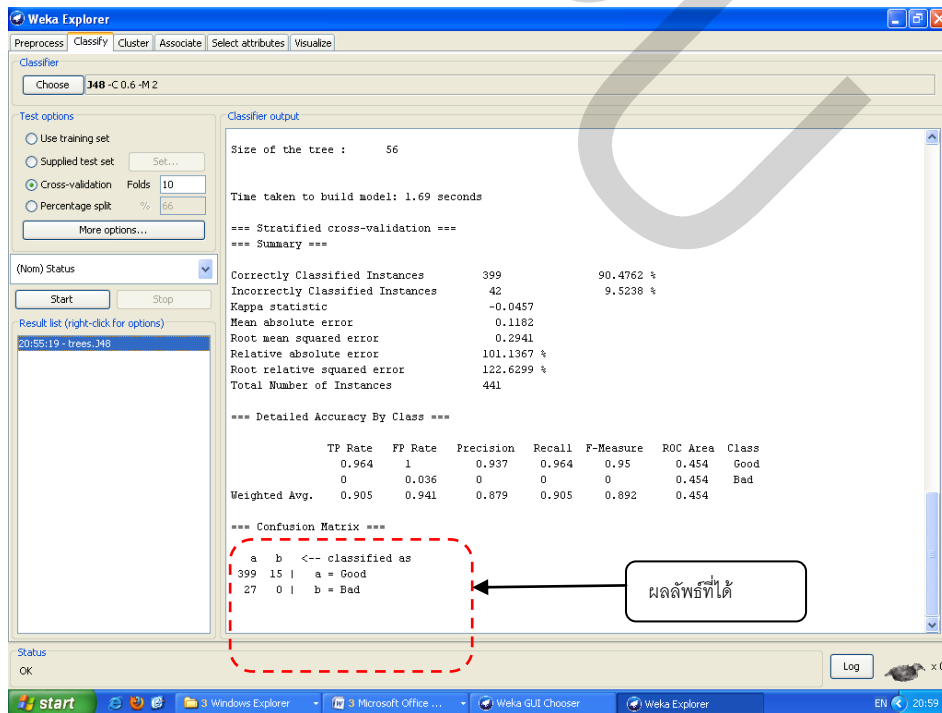
ภาพที่ 4.10 หน้าจอแสดงการเลือกเทคนิค Classify



ภาพที่ 4.11 การเลือกเทคนิคที่ใช้ในการจำแนกกลุ่มข้อมูล แบบ trees J48



ภาพที่ 4.12 หน้าจอแสดงผลการทำงานของโปรแกรมจำแนกกลุ่มข้อมูลด้วยเทคนิค Decision Tree: J48



ภาพที่ 4.13 หน้าจอแสดงผลที่ได้ของโมเดล

### 4.3.2 การเลือกเทคนิค

เทคนิคการจำแนกกลุ่มข้อมูลมีหลายเทคนิค แต่ละเทคนิคก็ให้ผลลัพธ์ออกมาในรูปแบบที่แตกต่างกัน ในการศึกษาวิจัยครั้งนี้ต้องการนำโมเดลที่ได้จากการเลือกเทคนิคที่เหมาะสมมาใช้ในการวิเคราะห์เพื่ออนุมัติให้สินเชื่อลูกค้า และนำไปพัฒนาระบบสนับสนุนการตัดสินใจในการให้สินเชื่อลูกค้า ดังนั้นเทคนิคการจำแนกกลุ่มที่นำมาใช้ในการศึกษาครั้งนี้ ซึ่งรายละเอียดแต่ละเทคนิคได้อธิบายไว้ในบทที่ 2

1. โมเดลการจำแนกกลุ่มแบบ Decision Tree : C4.5 หรือเรียกว่าโมเดล Classifires trees J48 ในโปรแกรม Weka
2. โมเดลการจำแนกกลุ่มแบบ Decision Rule : Part หรือเรียกว่าโมเดล Classifires PART ในโปรแกรม Weka
3. โมเดลการจำแนกกลุ่มแบบ Bayes : Naïve Bayes หรือเรียกว่าโมเดล Classifier Naïve Bayes ในโปรแกรม Weka

### 4.4 ผลการวิเคราะห์

ในการศึกษาครั้งนี้ทำการทดสอบ โมเดลโดยใช้โปรแกรมประยุกต์ Weka เวอร์ชัน 3.6.6 ข้อมูลที่ใช้ในการทดสอบมีทั้งหมด 441 แถว และ 17 แอททริบิวต์ ในการทดสอบได้ใช้การทดสอบแบบ 10-fold cross validation ซึ่งได้แก่ การแบ่งกลุ่มข้อมูลออกเป็น 10 กลุ่ม ในแต่ละรอบจะนำกลุ่มข้อมูลจำนวน 9 กลุ่ม เป็นกลุ่มศึกษา (Training Set) และกลุ่มข้อมูลที่เหลือเป็นกลุ่มทดสอบ (Testing Set) โดยทำซ้ำเป็นจำนวน 10 รอบ เพื่อเปลี่ยนกลุ่มทดสอบให้ครบทุกกลุ่ม และนำผลลัพธ์ที่ได้มาหาค่าเฉลี่ย โดยอธิบายความหมายแต่ละแอททริบิวต์ ซึ่งมี 17 แอททริบิวต์ ดังนี้

#### ตารางที่ 4.2 อธิบายความหมายแต่ละแอททริบิวต์

Object	วัตถุประสงค์การใช้งานรถยนต์
Car typ	ประเภทรถยนต์
Bus Typ	ประเภทธุรกิจ
CusTyp	ประเภทลูกค้า
Long Run	ระยะเวลาดำเนินธุรกิจ

ตารางที่ 4.2 (ต่อ)

Capital	ทุนจดทะเบียน
NP	กำไรสุทธิ
Income	รายได้หลัก
OE	ส่วนผู้ถือหุ้น
Cur ratio	อัตราสินทรัพย์หมุน/หนี้สินหมุนเวียน
Debt Ratio	อัตราหนี้สินรวม/สินทรัพย์รวม
DE	อัตราหนี้สินรวม/ส่วนผู้ถือหุ้น
Growth Inc	อัตราการเพิ่มของยอดขาย
Growth NP	อัตราการเพิ่มของกำไรสุทธิ
Credit line	วงเงินอนุมัติ
Grade	เกรดลูกค้า
Status	สถานะลูกค้า

ตารางที่ 4.3 อธิบายความหมายแต่ละฟิลด์

Object	ความหมาย	Car typ	ความหมาย	Bus Typ	ความหมาย
mkt	รถเพื่อใช้เดินทางพบ ลูกค้า	Large	รถยนต์เก๋งขนาดใหญ่	five	5 คัน
private	รถเพื่อการเดินทาง ส่วนตัว	Lux	รถยนต์Luxury	four	4 คัน
transport	รถเพื่อการขนส่ง	Medium	รถยนต์เก๋งขนาดกลาง	two	2 คัน
		MPV	รถยนต์เอนกประสงค์	zero	0 คัน
		Pickup	รถกระบะ		
		Small	รถยนต์เก๋งขนาดเล็ก		
		Van	รถยนต์ตู้		

ตารางที่ 4.4 อธิบายความหมายแต่ละฟิลด์

CusTyp	ความหมาย	Long Run	ความหมาย	Capital	ความหมาย
CCC	ลูกค้าเก่าชำระดี>3ปี	0	< 3ปี	0	< 3ล้านบาท
DDD	ลูกค้าเก่าชำระดี	13	3-5ปี	8	3-60ล้านบาท
EEE	ลูกค้าใหม่	14	6-10ปี	9	61-100ล้านบาท
ZZZ	ลูกค้าเก่าชำระไม่ดี	15	> 10ปี	10	> 100ล้านบาท

ตารางที่ 4.5 อธิบายความหมายแต่ละฟิลด์

NP	ความหมาย	Income	ความหมาย	OE	ความหมาย
0	ขาดทุนต่อเนื่องLLL	0	ลดลงต่อเนื่อง	0	ลดลงต่อเนื่อง 2 ปี
13	สลับกัน LPP, LPL, PLP	13	เพิ่ม-ลดสลับกัน	8	เพิ่ม-ลดสลับกัน
14	กำไร 2 ปีล่าสุด PPL	14	เพิ่มขึ้นต่อเนื่อง 1 ปี	9	เพิ่มขึ้นต่อเนื่อง 1 ปี
15	กำไร 3 ปี PPP	15	เพิ่มขึ้นต่อเนื่อง 2 ปี	10	เพิ่มขึ้นต่อเนื่อง 2 ปี

ตารางที่ 4.6 อธิบายความหมายแต่ละฟิลด์

Cur ratio	ความหมาย	Debt Ratio	ความหมาย	DE	ความหมาย
0	ค่าติดลบ	0	> 250%	0	> 300%
3	ค่าเฉลี่ย 0-0.75 เท่า	3	ค่าเฉลี่ย 151-250%	3	ค่าเฉลี่ย 201-300%
4	ค่าเฉลี่ย 0.76-1 เท่า	4	ค่าเฉลี่ย 50-150%	4	ค่าเฉลี่ย 100-200%
5	ค่าเฉลี่ย 3 ปี > 1 เท่า	5	ค่าเฉลี่ย < 50%	5	ค่าเฉลี่ย < 100%

ตารางที่ 4.7 อธิบายความหมายแต่ละฟิลด์

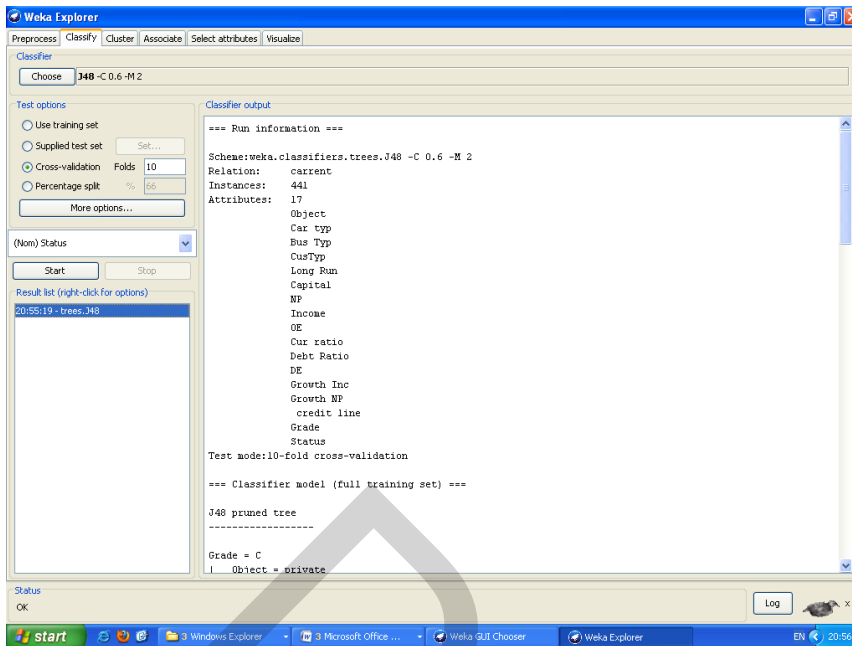
Growth Inc	ความหมาย	Growth NP	ความหมาย
0	ค่าเฉลี่ย ต่ำ	0	ค่าเฉลี่ย ต่ำ
3	ค่าเฉลี่ย 1-5%	3	ค่าเฉลี่ย 1-5%
4	ค่าเฉลี่ย 6-10%	4	ค่าเฉลี่ย 6-10%
5	ค่าเฉลี่ย > 10%	5	ค่าเฉลี่ย > 10%

ตารางที่ 4.8 อธิบายความหมายแต่ละฟิลด์

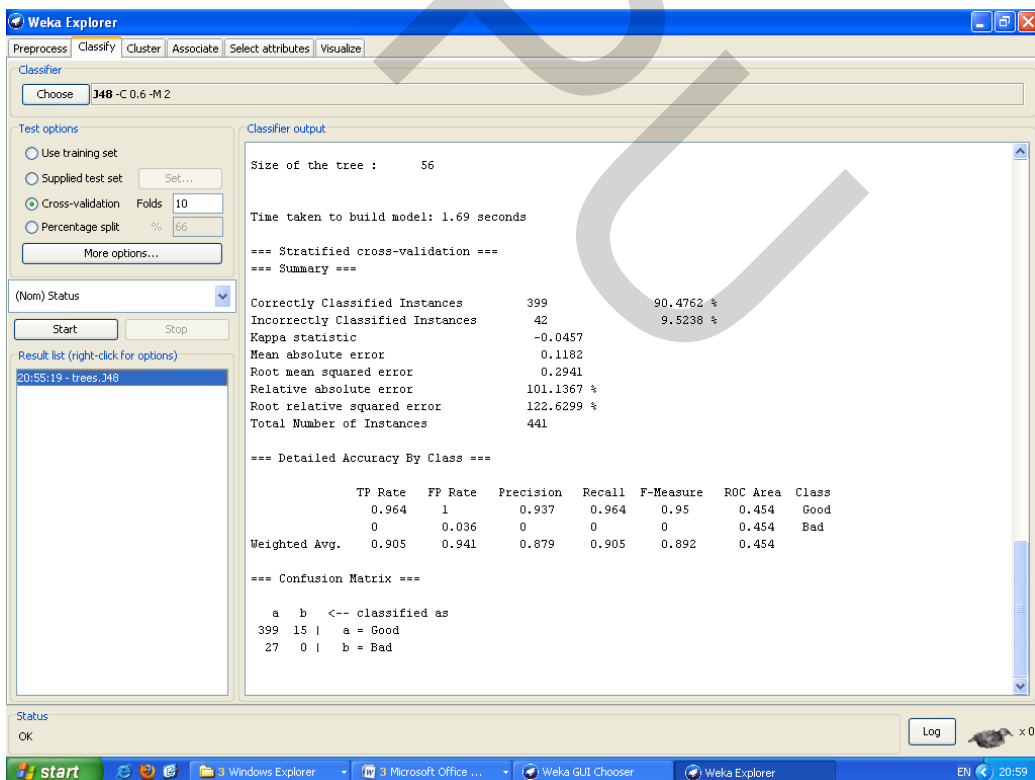
Grade	ความหมาย
A	เกรดลูกค้า A = คะแนนรวมตั้งแต่ 89-100 คะแนน
B	B = คะแนนรวมตั้งแต่ 76 - 88 คะแนน
C	C = คะแนนรวมตั้งแต่ 60 - 75 คะแนน
D	D = คะแนนรวมต่ำกว่า 60 คะแนน
Z	

จากการทดสอบ 3 เทคนิคที่ได้กล่าวไว้ในบทที่ 3 ได้รายละเอียดผลลัพธ์ ดังนี้  
**ผลลัพธ์เทคนิคที่ 1 Decision Tree : C4.5** หรือเรียกว่าโมเดล Classifires trees J48  
 แสดงตามภาพที่ 4.14 4.15 และ 4.16

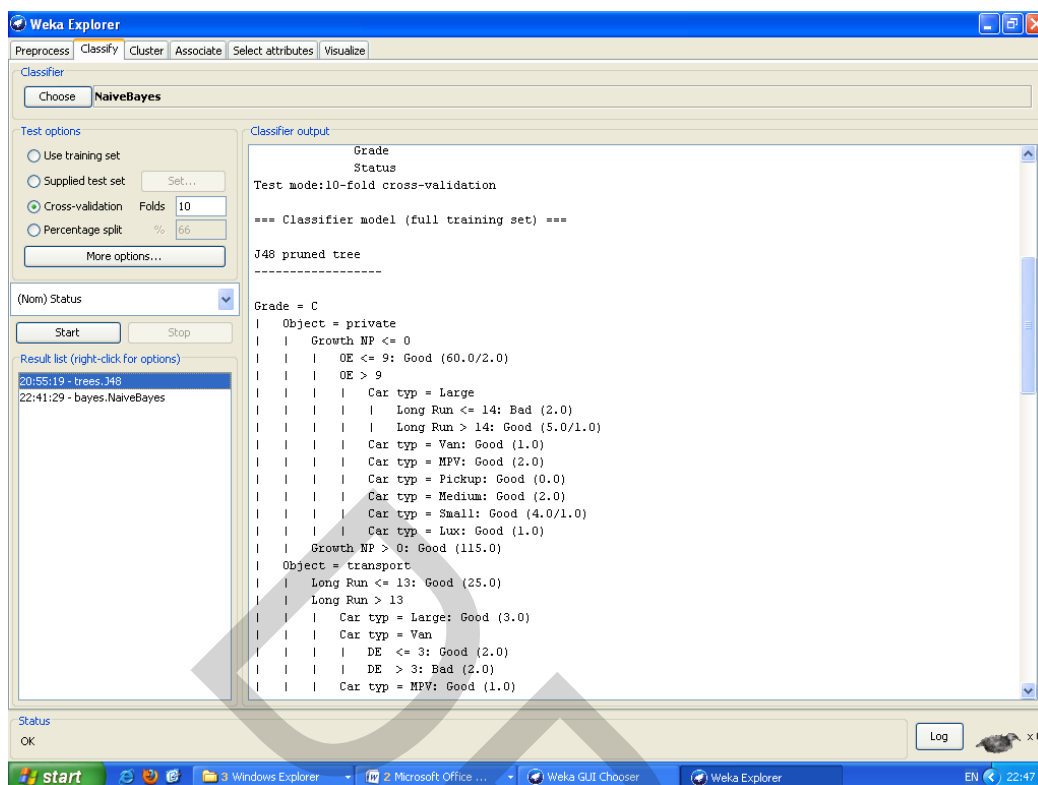




ภาพที่ 4.14 หน้าจอแสดงผลฟังก์ชันการจำแนกกลุ่มข้อมูลด้วยเทคนิค Decision Tree : C4.5



ภาพที่ 4.15 หน้าจอแสดงผลฟังก์ชันที่ได้ของโมเดล



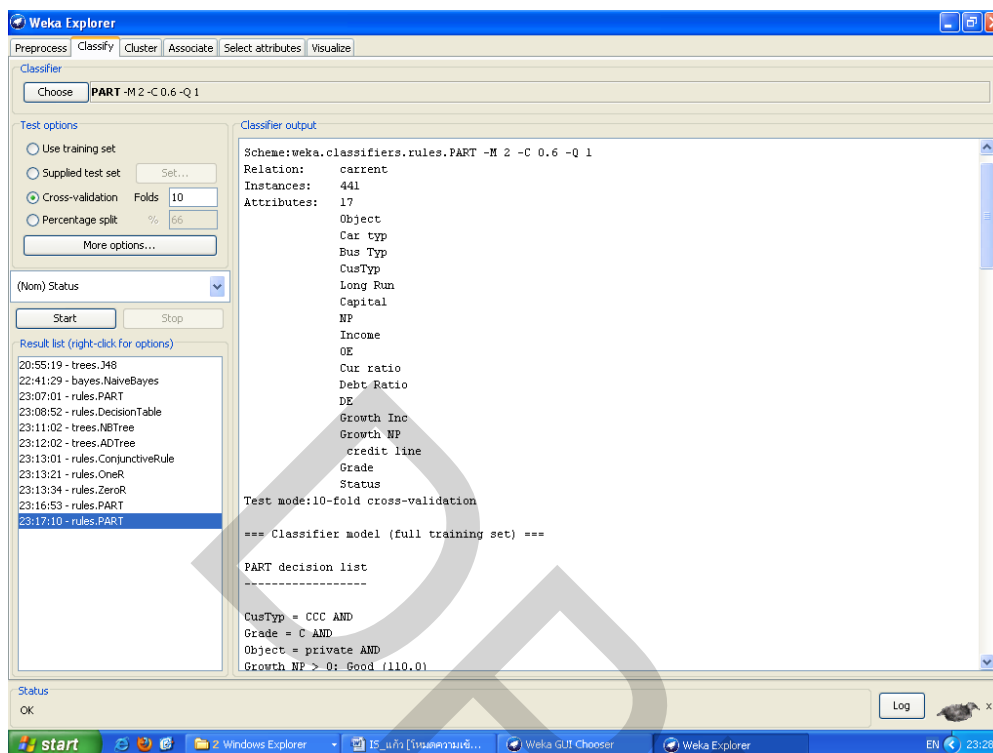
ภาพที่ 4.16 ผลลัพธ์บางส่วนจากการจำแนกกลุ่มแบบ Decision Tree : C4.5

จากการทดสอบเทคนิคนี้ พบว่า ให้ผลลัพธ์การจำแนกกลุ่มลูกค้าดี และไม่ดี ซึ่ง ลูกค้าดี หมายถึง ลูกค้าที่มีประวัติการชำระเงินตรงตามกำหนดของเงื่อนไขที่ลูกค้าได้รับการอนุมัติ และ ลูกค้าไม่ดี หมายถึง ลูกค้าที่มีประวัติการชำระเงินไม่ตรงตามกำหนดของเงื่อนไขที่ลูกค้าได้รับการอนุมัติ มีความถูกต้องเป็น 90.47% และได้จำนวนกฎจำนวน 40 กฎ ซึ่งจะแสดงผลเป็นกฎในลักษณะของต้นไม้การตัดสินใจ ที่มีกิ่งต้นไม้จำนวนมาก

โดยผลลัพธ์ที่ได้จะแสดงเงื่อนไขของกฎ โดยค่าของข้อมูลของแอทริบิวต์ที่มีค่าต่อเนื่องจะใช้เครื่องหมาย “<” “<=” และ “>” “>=” และในการพิจารณาส่วนข้อมูลที่เป็นค่าไม่ต่อเนื่องจะใช้เครื่องหมาย “=” และใช้เงื่อนไข “AND” ในการเชื่อมโยงระหว่างแอทริบิวต์ โดยจะเชื่อมโยงไปจนถึงแอทริบิวต์ที่จะแบ่งกลุ่มได้ และจะมีข้อมูลความถูกต้องของกฎ เช่น

Car typ = Pickup: Good (6.0/1.0) หมายถึง ถ้าประเภทรถ เป็นรถ Pickup จะเป็นลูกค้าดี และมีข้อมูลที่สอดคล้องหรือให้ผลลัพธ์ตรงกันจำนวน 6 ตัวอย่าง และให้ผลลัพธ์ที่ไม่ตรงกันจำนวน 1 ตัวอย่าง และได้แสดงตัวอย่างผลการรันไว้ที่ภาคผนวก

## ผลลัพธ์เทคนิคที่ 2 การจำแนกกลุ่มแบบ Decision Rule : Part แสดงตามภาพที่ 4.17

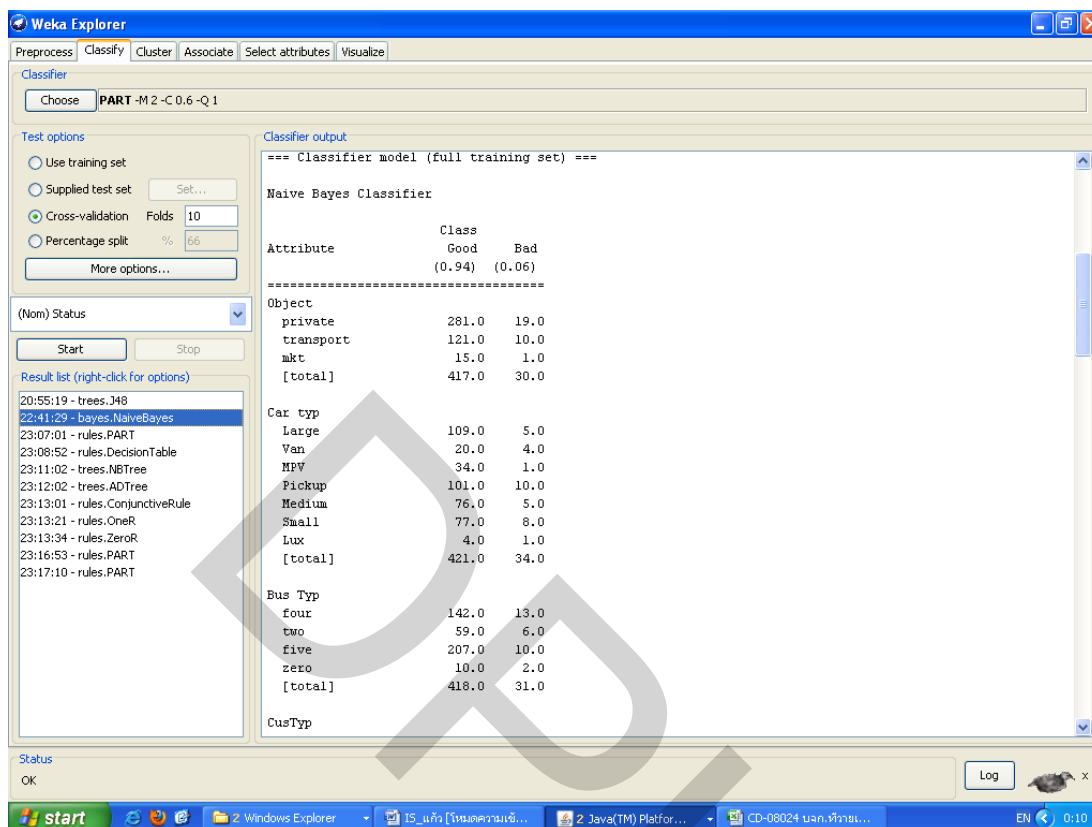


ภาพที่ 4.17 ผลลัพธ์บางส่วนจากการจำแนกกลุ่มแบบ Decision Rule: Part

จากการทดสอบเทคนิคนี้ พบว่า ให้ผลลัพธ์การจำแนกกลุ่มลูกค้าดี และไม่ดี มีความถูกต้องเป็น 90.24% และได้จำนวนกฎจำนวน 26 กฎ ซึ่งแสดงผลลัพธ์เป็นกฎที่เข้าใจง่ายให้จำนวนกฎน้อย ตามที่แสดงดังภาพที่ 4.17

เช่น CusTyp = CCC AND Grade = C AND Object = private AND Growth NP > 0: Good(110.0) หมายถึง ถ้าประเภทลูกค้าเป็นลูกค้าใหม่ และเกรดเท่ากับผลคะแนนรวมตั้งแต่ 60-75 คะแนน และวัตถุประสงค์การใช้งานคือใช้งานส่วนตัว และอัตราการเพิ่มของกำไรสุทธิมากกว่า 0 จะเป็นลูกค้าดี และมีข้อมูลที่สอดคล้องหรือให้ผลลัพธ์ตรงกันจำนวน 110 ตัวอย่าง

### ผลลัพธ์เทคนิคที่ 3 การจำแนกกลุ่มแบบ Bayes : Naive Bayes แสดงตามภาพที่ 4.18



ภาพที่ 4.18 ผลลัพธ์บางส่วนจากการจำแนกกลุ่มแบบ Bayes : Naive Bayes

จากการทดสอบเทคนิคนี้ พบว่า ให้ผลลัพธ์การจำแนกกลุ่มลูกค้าดี และไม่ดี มีความถูกต้องเป็น 87.52%

วิธีนี้จะไม่รองรับข้อมูลที่เป็นค่าต่อเนื่อง (Continuous Data) ดังนั้น ผลลัพธ์จะแสดงการแบ่งกลุ่มข้อมูลเป็นช่วง ๆ ได้แก่ แบ่งกลุ่มวัตถุประสงค์การใช้งานรถยนต์ (Object) มี 3 กลุ่ม คือ มีวัตถุประสงค์ใช้งานส่วนตัว (private) ใช้ทำการขนส่ง (transport) และ ใช้สำหรับเดินทางพบลูกค้า (mkt.) และแสดงให้เห็นว่าลูกค้าดี และลูกค้าไม่ดี มีจำนวนเท่าไร เช่น ถ้าแบ่งกลุ่มตามวัตถุประสงค์การใช้งานรถยนต์แบบส่วนตัว จะได้ผลลัพธ์เป็นลูกค้าดีจำนวน 281 ราย และลูกค้าไม่ดีจำนวน 19 ราย ถ้าแบ่งกลุ่มตามวัตถุประสงค์การใช้งานรถยนต์เพื่อขนส่ง จะได้ผลลัพธ์เป็นลูกค้าดีจำนวน 121 ราย และลูกค้าไม่ดีจำนวน 10 ราย ถ้าแบ่งกลุ่มตามวัตถุประสงค์การใช้งานรถยนต์สำหรับเดินทางพบลูกค้า จะได้ผลลัพธ์เป็นลูกค้าดีจำนวน 15 ราย และลูกค้าไม่ดีจำนวน 1 ราย

#### 4.4.1 การวิเคราะห์และการคัดเลือกกฎ

จากผลการทดลอง การจำแนกกลุ่มทั้ง 3 แบบ คือ 1) แบบ Decision Tree: C4.5 2) แบบ Decision Rule: Part และ 3) แบบ Bayes: Naive Bayes พบว่าแต่ละแบบให้ผลลัพธ์ที่มีความถูกต้องคือ 90.47 เปอร์เซ็นต์ 90.24 เปอร์เซ็นต์ และ 87.52 เปอร์เซ็นต์ ตามลำดับ และให้กฎจำนวน 40 กฎ 26 กฎ และ ไม่มีกฎ ตามลำดับ ซึ่งการจำแนกกลุ่มแบบ Decision Tree: C4.5 ให้ผลลัพธ์ที่มีความถูกต้องมากที่สุด สามารถแบ่งกลุ่มได้ตามเงื่อนไขที่ชัดเจนมากที่สุด และสามารถแสดงกฎให้สามารถนำไปวิเคราะห์กฎต่อไปได้ (ผลการรันฉบับเต็มอยู่ที่ภาคผนวก) โดยแบ่งเป็น 4 กลุ่มหลัก ๆ ดังนี้

1. กฎที่จำแนกกลุ่มลูกค้าดี (Good) ได้ถูกต้อง 100 เปอร์เซ็นต์ มีจำนวน 23 กฎ ซึ่งกฎที่อยู่ในกลุ่มนี้เป็นกฎที่มีความน่าเชื่อถือมากที่สุด เพราะสามารถจำแนกกลุ่มตัวอย่างได้ถูกต้องทั้งหมด

ตัวอย่างกฎในกลุ่มนี้ เช่น Grade = C and Object = private and Growth NP > 0 : Good (115.0) หมายถึง ลูกค้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการเดินทางส่วนตัว และ ไม่มีผลขาดทุนต่อเนื่อง จะเป็นกลุ่มลูกค้าดี โดยมีข้อมูลที่สอดคล้องกับกฎนี้ จำนวน 115 ตัวอย่าง

2. กฎที่จำแนกกลุ่มลูกค้าไม่ดี (Bad) ได้ถูกต้อง 100 เปอร์เซ็นต์ มีจำนวน 4 กฎ ซึ่งกฎที่อยู่ในกลุ่มนี้เป็นกฎที่มีความน่าเชื่อถือมากที่สุด เพราะสามารถจำแนกกลุ่มตัวอย่างได้ถูกต้องทั้งหมด

ตัวอย่างกฎในกลุ่มนี้ เช่น Grade = C and Object = transport and LongRun >13 Car typ = Pickup and Bus typ = four and OE <=0 : Bad (2.0) หมายถึง ลูกค้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการขนส่ง และมีระยะเวลาการดำเนินธุรกิจมากกว่า 5 ปี ประเภทรถ เป็นรถกระบะ และประเภทธุรกิจ เป็นกลุ่ม Four และ ส่วนของผู้ถือหุ้นลดลงต่อเนื่อง 2 ปี เป็นกลุ่มลูกค้าไม่ดี โดยมีข้อมูลที่สอดคล้องกับกฎนี้ จำนวน 2 ตัวอย่าง

3. กฎที่จำแนกกลุ่มลูกค้าดี (Good) ได้ถูกต้องน้อยกว่า 100 เปอร์เซ็นต์ มีจำนวน 12 กฎ

ตัวอย่างกฎในกลุ่มนี้ เช่น Grade = C and Object = private and Growth NP <= 0 and OE <=9 : Good (59.63/2.0)

หมายถึง ลูกค้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการเดินทางส่วนตัว และ มีผลขาดทุนต่อเนื่อง และ ส่วนของผู้ถือหุ้น ไม่มีการเพิ่มขึ้นต่อเนื่อง 2 ปี จะเป็นกลุ่มลูกค้าดี โดยมีข้อมูลที่สอดคล้องกับกฎนี้ จำนวน 58 ตัวอย่าง และมีข้อมูลที่สอดคล้องกับกฎนี้ จำนวน 2 ตัวอย่าง

4. กฎที่จำแนกกลุ่มลูกค้าไม่ดี (Bad) ได้ถูกต้องน้อยกว่า 100 เปอร์เซ็นต์ มีจำนวน 1 กฎ

ตัวอย่างกฎในกลุ่มนี้ เช่น Grade = B and Car typ = Small and Income <= 13 and LongRun <=14 : BAD (3.0/1.0)

หมายถึง เป็นลูก้าเกรด B รถที่เช่าเป็นรถยนต์เก๋งขนาดเล็ก รายได้หลัก ไม่มีการเพิ่มขึ้นอย่างต่อเนื่อง ระยะเวลาการดำเนินธุรกิจต่ำกว่า 15 ปี เป็นกลุ่มลูก้าไม่ดี โดยมีข้อมูลที่สอดคล้องกับกฎนี้ จำนวน 2 ตัวอย่าง และมีข้อมูลที่ไม่สอดคล้องกับกฎนี้ จำนวน 1 ตัวอย่าง

แต่เนื่องจากกฎที่ได้ มีบางกฎที่มีความถูกต้อง น้อยกว่า 90% หรือ บางกฎไม่มีตัวอย่างสนับสนุนกับกฎดังกล่าว หรือ บางกฎที่ไม่สมบูรณ์

ดังนั้น จึงทำการลบกฎดังกล่าวออกไป จำนวน 12 กฎ ดังต่อไปนี้

1. กฎที่ไม่มีกลุ่มตัวอย่างสนับสนุน มีจำนวน 3 กฎ ได้แก่

Grade = B and CarTyp = Lux : Good (0.0) หมายถึง ถ้าลูก้าได้เกรด B และประเภทรถที่ใช้เป็นกลุ่ม Luxury เป็นกลุ่มลูก้าดี

Grade = C and Object = transport and LongRun>13 and CarTyp = Lux : Good (0.0) หมายถึง ถ้าลูก้าได้เกรด C และวัตถุประสงค์การใช้รถเพื่อการขนส่ง และ ระยะเวลาการดำเนินธุรกิจ มากกว่า 6 ปี และประเภทรถที่ใช้เป็นกลุ่ม Luxury เป็นกลุ่มลูก้าดี นอกจากนี้ไม่มีกลุ่มตัวอย่างสนับสนุนแล้ว เงื่อนไขยังไม่เหมาะสมด้วย เนื่องจาก ถ้าใช้รถเพื่อการขนส่ง ไม่สอดคล้องกับประเภทที่เป็นกลุ่ม Luxury

Grade = C and Object = private and Growth NP <= 0 and OE >9 and CarTyp = Pickup : Good (0.0) หมายถึง ถ้าลูก้าได้เกรด C และวัตถุประสงค์การใช้รถเพื่อการเดินทางส่วนตัว และมีการผลขาดทุนต่อเนื่อง และส่วนของผู้ถือหุ้น เพิ่มขึ้นต่อเนื่องอย่างน้อย 1 ปี และประเภทรถเป็นรถกระบะ เป็นกลุ่มลูก้าดี

2. กฎที่มีความถูกต้องน้อยกว่า 90% มีจำนวน 8 กฎ ได้แก่

Grade = B and Car typ = Small and Income <= 13 and Long Run >14 : Good (3.0/1.0) หมายถึง ถ้าลูก้าได้เกรด B และประเภท เป็นรถยนต์เก๋งขนาดเล็ก (เครื่องยนต์ 1.5-1.6 ซีซี) และรายได้หลัก ลดลงอย่างต่อเนื่อง หรือ เพิ่มลดสลับกัน และ ระยะเวลาการดำเนินธุรกิจนานกว่า 10 ปี เป็นกลุ่มลูก้าดี โดยมีข้อมูลที่สอดคล้องกับกฎนี้ จำนวน 2 ตัวอย่าง และมีข้อมูลที่ไม่สอดคล้องกับกฎนี้ จำนวน 1 ตัวอย่าง มีความถูกต้องของตัวอย่างที่สนับสนุนคิดเป็น 67%

Grade = B and Car typ = Small and Income <= 13 LongRun <=14 : Bad (3.0/1.0) หมายถึง ถ้าลูก้าได้เกรด B และประเภท เป็นรถยนต์เก๋งขนาดเล็ก (เครื่องยนต์ 1.5-1.6 ซีซี) และรายได้หลัก ลดลงอย่างต่อเนื่อง หรือ เพิ่มลดสลับกัน และ ระยะเวลาการดำเนินธุรกิจน้อยกว่า 10 ปี

เป็นกลุ่มลูกค้าไมดี โดยมีข้อมูลที่สอดคล้องกับกฎนี้ จำนวน 2 ตัวอย่าง และมีข้อมูลที่ไม่สอดคล้องกับกฎนี้ จำนวน 1 ตัวอย่าง มีความถูกต้องของตัวอย่างที่สนับสนุนคิดเป็น 67%

Grade = C and Object = private and Growth NP  $\leq 0$  and OE  $> 9$  and Car typ = Small : Good (4.0/1.0) หมายถึง ถ้าลูกค้าได้เกรด C และมีวัตถุประสงค์การใช้รถยนต์เพื่อการเดินทางส่วนตัว และมีขาดทุนต่อเนื่อง และ ส่วนของผู้ถือหุ้น เพิ่มขึ้นต่อเนื่อง 2 ปี และ ประเภทรถเป็นรถยนต์เก๋งขนาดเล็ก (เครื่องยนต์ 1.5-1.6 ซีซี) เป็นกลุ่มลูกค้าดี โดยมีข้อมูลที่สอดคล้องกับกฎนี้ จำนวน 3 ตัวอย่าง และมีข้อมูลที่ไม่สอดคล้องกับกฎนี้ จำนวน 1 ตัวอย่าง มีความถูกต้องของตัวอย่างที่สนับสนุนคิดเป็น 75%

Grade = D and Bus Typ = Four and OE  $\leq 0$  : Good (7.0/2.0) หมายถึง ถ้าลูกค้าได้เกรด D และประเภทธุรกิจอยู่ในกลุ่ม D ได้แก่ ธุรกิจจัดหางาน ธุรกิจรถเช่า และธุรกิจรับเหมาก่อสร้างขนาดเล็กทุนต่ำกว่า 3 ล้านบาท และส่วนของผู้ถือหุ้น ลดลงต่อเนื่อง 2 ปี เป็นกลุ่มลูกค้าดี โดยมีข้อมูลที่สอดคล้องกับกฎนี้ จำนวน 5 ตัวอย่าง และมีข้อมูลที่ไม่สอดคล้องกับกฎนี้ จำนวน 2 ตัวอย่าง มีความถูกต้องของตัวอย่างที่สนับสนุนคิดเป็น 71%

Grade = B and Car typ = Pickup : Good (6.0/1.0) หมายถึง ถ้าลูกค้าได้เกรด B และประเภทรถที่ใช้เป็นรถกระบะ เป็นกลุ่มลูกค้าดี โดยมีข้อมูลที่สอดคล้องกับกฎนี้ จำนวน 5 ตัวอย่าง และมีข้อมูลที่ไม่สอดคล้องกับกฎนี้ จำนวน 1 ตัวอย่าง มีความถูกต้องของตัวอย่างที่สนับสนุนคิดเป็น 83%

Grade = C and Object = private and Growth NP  $\leq 0$  and OE  $> 9$  and Car typ = Large and LongRun  $> 14$  : Good (5.0/1.0) หมายถึง ถ้าลูกค้าได้เกรด C และมีวัตถุประสงค์การใช้รถเพื่อการเดินทางส่วนตัว และมีผลขาดทุนต่อเนื่อง และ ส่วนของผู้ถือหุ้น เพิ่มขึ้นต่อเนื่อง 2 ปี และประเภทรถที่ใช้เป็นรถยนต์เก๋งขนาดใหญ่ (เครื่องยนต์ มากกว่า 2.0 ซีซี) และมีระยะเวลาการดำเนินธุรกิจ นานกว่า 10 ปี เป็นกลุ่มลูกค้าดี โดยมีข้อมูลที่สอดคล้องกับกฎนี้ จำนวน 4 ตัวอย่าง และมีข้อมูลที่ไม่สอดคล้องกับกฎนี้ จำนวน 1 ตัวอย่าง มีความถูกต้องของตัวอย่างที่สนับสนุนคิดเป็น 80%

Grade = D and Bus Typ = Zero : Good (6.0/1.0) หมายถึง ถ้าลูกค้าได้เกรด D และประเภทธุรกิจอยู่ในกลุ่ม Zero ได้แก่ ธุรกิจจัดหางาน ธุรกิจรถเช่า และธุรกิจรับเหมาก่อสร้างขนาดเล็กทุนต่ำกว่า 3 ล้านบาท เป็นกลุ่มลูกค้าดี โดยมีข้อมูลที่สอดคล้องกับกฎนี้ จำนวน 5 ตัวอย่าง และมีข้อมูลที่ไม่สอดคล้องกับกฎนี้ จำนวน 1 ตัวอย่าง มีความถูกต้องของตัวอย่างที่สนับสนุนคิดเป็น 83%

Grade = C and Object = transport and LongRun > 13 and CarTyp = Pickup and Bus typ = Four and OE > 0 : Good (13.2/2.0) หมายถึง ถ้าลูกค้าได้เกรด C และ วัตถุประสงค์การใช้รถเพื่อการขนส่ง และ ระยะเวลาการดำเนินธุรกิจ มากกว่า 6 ปี และประเภทรถที่ใช้เป็นรถกระบะ และประเภทธุรกิจ อยู่ในกลุ่ม Four และ ส่วนของผู้ถือหุ้น ไม่มีการลดลงอย่างต่อเนื่อง เป็นกลุ่มลูกค้าดี โดยมีข้อมูลที่สอดคล้องกับกฎนี้ จำนวน 11 ตัวอย่าง และมีข้อมูลที่ไม่สอดคล้องกับกฎนี้ จำนวน 2 ตัวอย่าง มีความถูกต้องของตัวอย่างที่สนับสนุนคิดเป็น 85%

3. กฎที่ไม่สมบูรณ์ มีจำนวน 1 กฎ ได้แก่

Grade = Z หมายถึง ลูกค้าที่ไม่มีเกรด จึงทำให้กฎไม่สมบูรณ์

**กฎที่อยู่ในกลุ่มลูกค้าดี (Good) มีจำนวน 24 กฎ ได้แก่**

1. ลูกค้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการเดินทางส่วนตัว และ ไม่มีผลขาดทุนต่อเนื่อง จะเป็นกลุ่มลูกค้าดี

กฎที่อ้างอิง : Grade = C and Object = private and Growth NP > 0 : Good (115.0)

2. ลูกค้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการเดินทางส่วนตัว และ มีผลขาดทุนต่อเนื่อง และ ส่วนของผู้ถือหุ้น ไม่มีการเพิ่มขึ้นต่อเนื่อง 2 ปี จะเป็นกลุ่มลูกค้าดี

กฎที่อ้างอิง : Grade = C and Object = private and Growth NP <= 0 and OE <= 9 : Good (59.63/2.0)

3. ลูกค้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการเดินทางส่วนตัว และ มีผลขาดทุนต่อเนื่อง และ ส่วนของผู้ถือหุ้น มีการเพิ่มขึ้นต่อเนื่อง 2 ปี ประเภทรถที่ใช้เป็นกลุ่ม Luxury จะเป็นกลุ่มลูกค้าดี

กฎที่อ้างอิง : Grade = C and Object = private and Growth NP <= 0 and OE > 9 and Car typ = Lux : Good (1.0)

4. ลูกค้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการเดินทางส่วนตัว และ มีผลขาดทุนต่อเนื่อง และ ส่วนของผู้ถือหุ้น มีการเพิ่มขึ้นต่อเนื่อง 2 ปี ประเภทรถที่ใช้เป็นรถยนต์เก๋งขนาดกลาง (เครื่องยนต์ขนาด 1.6-1.8 ซีซี) จะเป็นกลุ่มลูกค้าดี

กฎที่อ้างอิง : Grade = C and Object = private and Growth NP <= 0 and OE > 9 and Car typ = Medium : Good (2.0)

5. ลูกค้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการเดินทางส่วนตัว และ มีผลขาดทุนต่อเนื่อง และ ส่วนของผู้ถือหุ้น มีการเพิ่มขึ้นต่อเนื่อง 2 ปี ประเภทรถที่ใช้เป็นรถยนต์เอนกประสงค์ จะเป็นกลุ่มลูกค้าดี



กฎที่อ้างอิง : Grade = C and Object = private and Growth NP  $\leq$  0 and OE  $>$  9 and Car typ = MPV : Good (2.0)

6. ลูก้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการเดินทางส่วนตัว และมีผลขาดทุนต่อเนื่อง และส่วนของผู้ถือหุ้น มีการเพิ่มขึ้นต่อเนื่อง 2 ปี ประเภทรถที่ใช้เป็นรถยนต์ตู้ จะเป็นกลุ่มลูก้าดี

กฎที่อ้างอิง : Grade = C and Object = private and Growth NP  $\leq$  0 and OE  $>$  9 and Car typ = Van : Good (1.0)

7. ลูก้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการขนส่ง และมีระยะเวลาการดำเนินงานธุรกิจ น้อยกว่าหรือเท่ากับ 5 ปี จะเป็นกลุ่มลูก้าดี

กฎที่อ้างอิง : Grade = C and Object = transport and LongRun  $\leq$  13 : Good (25.34)

8. ลูก้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการขนส่ง และมีระยะเวลาการดำเนินงานธุรกิจ มากกว่า 5 ปี ประเภทรถ เป็นรถยนต์เก๋งขนาดเล็ก (เครื่องยนต์ 1.5-1.6 ซีซี) จะเป็นกลุ่มลูก้าดี

กฎที่อ้างอิง : Grade = C and Object = transport and LongRun  $>$  13 Car typ = Small : Good (20.0/1.0)

9. ลูก้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการขนส่ง และมีระยะเวลาการดำเนินงานธุรกิจ มากกว่า 5 ปี ประเภทรถ เป็นรถยนต์เก๋งขนาดกลาง (เครื่องยนต์ 1.6-1.8 ซีซี) จะเป็นกลุ่มลูก้าดี

กฎที่อ้างอิง : Grade = C and Object = transport and LongRun  $>$  13 Car typ = Medium : Good (3.0)

10. ลูก้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการขนส่ง และมีระยะเวลาการดำเนินงานธุรกิจ มากกว่า 5 ปี ประเภทรถ เป็นรถเอนกประสงค์ จะเป็นกลุ่มลูก้าดี

กฎที่อ้างอิง : Grade = C and Object = transport and LongRun  $>$  13 Car typ = MPV : Good (1.0)

11. ลูก้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการขนส่ง และมีระยะเวลาการดำเนินงานธุรกิจ มากกว่า 5 ปี ประเภทรถ เป็นรถยนต์เก๋งขนาดใหญ่ (เครื่องยนต์ มากกว่า 2.0 ซีซี) จะเป็นกลุ่มลูก้าดี

กฎที่อ้างอิง : Grade = C and Object = transport and LongRun  $>$  13 Car typ = Large : Good (3.0)

12. ลูกค้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการขนส่ง และมีระยะเวลาการดำเนินงานธุรกิจ มากกว่า 5 ปี ประเภทรถ เป็นรถกระบะ และประเภทธุรกิจ อยู่ในกลุ่ม Zero จะเป็นกลุ่มลูกค้าดี

กฎที่อ้างอิง : Grade = C and Object = transport and LongRun >13 Car typ = Pickup and Bus typ = Zero : Good (1.0)

13. ลูกค้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการขนส่ง และมีระยะเวลาการดำเนินงานธุรกิจ มากกว่า 5 ปี ประเภทรถ เป็นรถกระบะ และประเภทธุรกิจ อยู่ในกลุ่ม Five จะเป็นกลุ่มลูกค้าดี

กฎที่อ้างอิง : Grade = C and Object = transport and LongRun >13 Car typ = Pickup and Bus typ = Five : Good (28.0/1.0)

14. ลูกค้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการขนส่ง และมีระยะเวลาการดำเนินงานธุรกิจ มากกว่า 5 ปี ประเภทรถ เป็นรถกระบะ และประเภทธุรกิจ อยู่ในกลุ่ม Two จะเป็นกลุ่มลูกค้าดี

กฎที่อ้างอิง : Grade = C and Object = transport and LongRun >13 Car typ = Pickup and Bus typ = Two : Good (4.0)

15. ลูกค้าเกรด C และวัตถุประสงค์การใช้รถเพื่อการขนส่ง และมีระยะเวลาการดำเนินงานธุรกิจ มากกว่า 5 ปี ประเภทรถ เป็นรถตู้ และอัตราหนี้สินรวมต่อสินทรัพย์รวม มีค่าเฉลี่ยมากกว่า 151% จะเป็นกลุ่มลูกค้าดี

กฎที่อ้างอิง : Grade = C and Object = transport and LongRun >13 Car typ = Van and DE <= 3 : Good (2.0)

16. ลูกค้าเกรด C และวัตถุประสงค์การใช้รถเพื่อใช้เดินทางพบลูกค้า จะเป็นกลุ่มลูกค้าดี

กฎที่อ้างอิง : Grade = C and Object = mkt : Good (9.03)

17. ลูกค้าเกรด D และประเภทธุรกิจ อยู่ในกลุ่ม Two จะเป็นกลุ่มลูกค้าดี

กฎที่อ้างอิง : Grade = D and Bus typ = Two : Good (8.0)

18. ลูกค้าเกรด D และประเภทธุรกิจ อยู่ในกลุ่ม Five จะเป็นกลุ่มลูกค้าดี

กฎที่อ้างอิง : Grade = D and Bus typ = Five : Good (10.0)

19. ลูกค้าเกรด B และประเภทรถ เป็นรถยนต์เก๋งขนาดใหญ่ (เครื่องยนต์ มากกว่า 2.0 ซีซี) จะเป็นกลุ่มลูกค้าดี

กฎที่อ้างอิง : Grade = B and Car typ = Large : Good (17.0)

20. ลูกค้ายกรด B และประเภทรถ เป็นรถตู้ จะเป็นกลุ่มลูกค้ายดี

กฎที่อ้างอิง : Grade = B and Car typ = VAN : Good (3.0)

21. ลูกค้ายกรด B และประเภทรถ เป็นรถเอนกประสงค์ จะเป็นกลุ่มลูกค้ายดี

กฎที่อ้างอิง : Grade = B and Car typ = MPV : Good (2.0)

22. ลูกค้ายกรด B และประเภทรถ เป็นรถยนต์เก๋งขนาดกลาง (เครื่องยนต์ 1.6-1.8 ซีซี)

จะเป็นกลุ่มลูกค้ายดี

กฎที่อ้างอิง : Grade = B and Car typ = Mudium : Good (12.0/1.0)

23. ลูกค้ายกรด B และประเภทรถ เป็นรถยนต์เก๋งขนาดเล็ก (เครื่องยนต์ 1.5 ซีซี) และมีรายได้หลัก เพิ่มขึ้นต่อเนื่องอย่างน้อย 1 ปี จะเป็นกลุ่มลูกค้ายดี

กฎที่อ้างอิง : Grade = B and Car typ = Small and Income > 13 : Good (6.0)

24. ลูกค้ายกรด A จะเป็นกลุ่มลูกค้ายดี

กฎที่อ้างอิง : Grade = A : Good (48.0/3.0)

**กฎที่อยู่ในกลุ่มลูกค้ายไม่ดี (Bad) มีจำนวน 4 กฎ ได้แก่**

1. ลูกค้ายกรด C และวัตถุประสงค์การใช้รถเพื่อการขนส่ง และมีระยะเวลาการดำเนินงานธุรกิจ มากกว่า 5 ปี ประเภทรถ เป็นรถตู้ และอัตราหนี้สินรวมต่อสินทรัพย์รวม มีค่าเฉลี่ยต่ำกว่า 151% จะเป็นกลุ่มลูกค้ายไม่ดี

กฎที่อ้างอิง : Grade = C and Object = transport and LongRun >13 Car typ = Van and DE > 3 : Bad (2.0)

2. ลูกค้ายกรด C และวัตถุประสงค์การใช้รถเพื่อการขนส่ง และมีระยะเวลาการดำเนินงานธุรกิจ มากกว่า 5 ปี ประเภทรถ เป็นรถกระบะ และประเภทธุรกิจ เป็นกลุ่ม Four และ ส่วนของผู้ถือหุ้น ลดลงต่อเนื่อง 2 ปี จะเป็นกลุ่มลูกค้ายไม่ดี

กฎที่อ้างอิง : Grade = C and Object = transport and LongRun >13 Car typ = Pickup and Bus typ = four and OE <=0 : Bad (2.0)

3. ลูกค้ายกรด D และประเภทธุรกิจ เป็นกลุ่ม Four และ ส่วนของผู้ถือหุ้น ไม่มีการลดลงต่อเนื่อง 2 ปี จะเป็นกลุ่มลูกค้ายไม่ดี

กฎที่อ้างอิง : Grade = D and Bus typ = four and OE > 0 : Bad (2.0)

4. ลูกค้ายกรด C และวัตถุประสงค์การใช้รถเพื่อการเดินทางส่วนตัว และมีผลขาดทุนต่อเนื่อง และ ส่วนของผู้ถือหุ้น เพิ่มขึ้นต่อเนื่องอย่างน้อย 2 ปี และประเภทรถยนต์ที่ใช้เป็นรถยนต์เก๋งขนาดใหญ่ (เครื่องยนต์มากกว่า 2.0 ซีซี) และมีระยะเวลาการดำเนินงานธุรกิจ ไม่เกิน 10 ปี จะเป็นกลุ่มลูกค้ายไม่ดี

กฎที่อ้างอิง : Grade = C and Object = private and Growth NP  $\leq 0$  and OE  $> 9$  and Car typ = Large and LongRun  $\leq 14$  : Bad (2.0)

สรุปกฎจากผลการทดสอบการจำแนกกลุ่มข้อมูลแบบ Decision Tree : C4.5 เป็นการจัดกลุ่มที่เหมาะสมที่สุดในการทดสอบเนื่องจากมีกฎจำนวนมาก และมีเงื่อนไขที่เหมาะสม แต่มีบางกฎที่มีเงื่อนไขไม่เหมาะสม จึงทำการลบกฎดังกล่าวออก จำนวน กฎ 12 กฎ จากทั้งหมด 40 กฎ เหลือ 28 กฎ

เมื่อได้ผลของการทำเหมืองข้อมูลแล้วจึงทำการเลือกกฎที่มีความน่าเชื่อถือมากที่สุดเพื่อใช้เป็นเงื่อนไขในการเขียนโปรแกรมระบบวิเคราะห์ห้สินเชื่อต่อไป

#### 4.5 การพัฒนาระบบวิเคราะห์ห้สินเชื่อ

เป็นการนำกฎที่ได้จากการทำเหมืองข้อมูล มาใช้เป็นเงื่อนไขในการเขียน โปรแกรมระบบวิเคราะห์ห้สินเชื่อออนไลน์ เพื่อเป็นเครื่องมือช่วยในการตัดสินใจการอนุมัติสินเชื่อ ของ บริษัทกรุงไทยคาร์เร้นท์ แอนด์ ลีส จำกัด (มหาชน) โดยใช้โปรแกรม PHP และใช้ฐานข้อมูล My SQL ในการพัฒนาโปรแกรม

##### 4.5.1 การออกแบบหน้าจอ

โปรแกรม CS 4 ใช้ในการออกแบบหน้าจอระบบวิเคราะห์ห้สินเชื่อ เพื่อบันทึกข้อมูลลูกค้าที่ต้องการวิเคราะห์ห้อนุมัติสินเชื่อ ซึ่งหน้าจอนี้จะบันทึกรายละเอียดต่าง ๆ ที่ใช้ในการวิเคราะห์ ดังนี้คือ ชื่อ นามสกุลลูกค้า วัตถุประสงค์การใช้งานรถยนต์ ประเภทรถยนต์ ประเภทธุรกิจ ประเภทลูกค้า ระยะเวลาดำเนินธุรกิจ ทุนจดทะเบียน กำไรสุทธิ รายได้หลัก ส่วนผู้ถือหุ้น อัตราสินทรัพย์หมุน : หนี้สินหมุนเวียน อัตราหนี้สินรวม : สินทรัพย์รวม อัตราหนี้สินรวม : ส่วนผู้ถือหุ้น อัตราการเพิ่มของยอดขาย : อัตราการเพิ่มของกำไรสุทธิ วงเงินอนุมัติ เกรดลูกค้า ซึ่งเมื่อทำการป้อนข้อมูลรายละเอียดของลูกค้าที่ต้องการขอสินเชื่อเสร็จ ก็ทำการบันทึกและประมวลผล โปรแกรมก็จะประมวลผลและแสดงผลห้ลูกค้านี้ว่าเป็น ลูกค้าดี (Good) หรือ ลูกค้าไม่ดี (Bad)

ระบบวิเคราะห์สินเชื่อออนไลน์ | ONLINE ANALYSIS FUND

หน้าหลัก | ออกรายงาน

>> กรณำบันทึกข้อมูล

ชื่อ :  นามสกุล :

วัตถุประสงค์การใช้จ่ายรถยนต์ :  ประเภทรถยนต์ :

ประเภทธุรกิจ :  ประเภทลูกค้า :

ระยะเวลาผ่อนชำระ :  ทุนจดทะเบียน :

ทำบัญชี :  รายได้หลัก :

ส่วนผู้ถือหุ้น :  มีตราหุ้นที่มอบหมายให้สิ่งของ :

มีจำนวนหุ้นรวม :  มีตราหุ้นที่มอบหมายให้สิ่งของ :

มีจำนวนหุ้นที่มอบหมายให้สิ่งของ :  มีตราหุ้นที่มอบหมายให้สิ่งของ :

มีจำนวนหุ้นที่มอบหมายให้สิ่งของ :  มีตราหุ้นที่มอบหมายให้สิ่งของ :

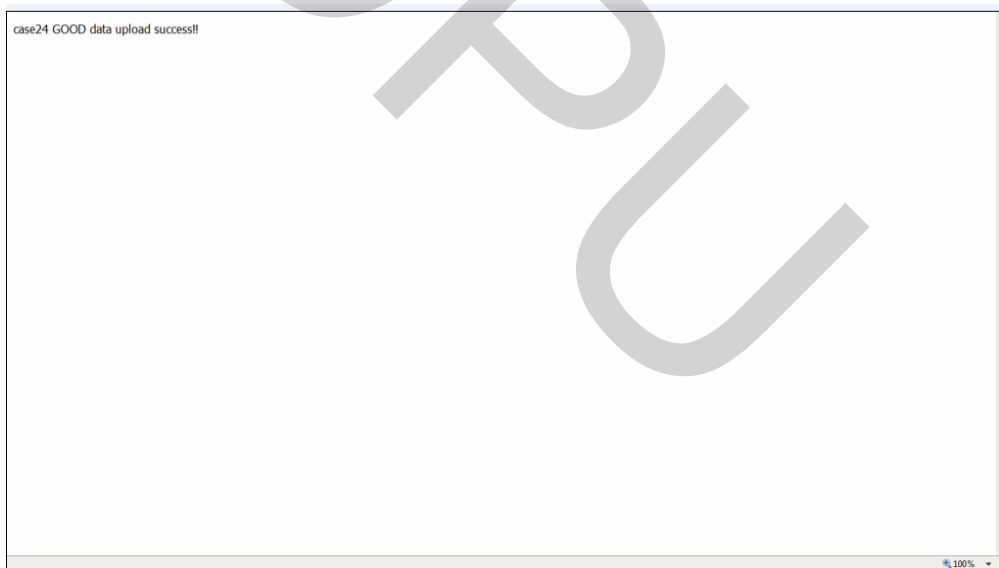
วงเงินอนุมัติ : 0

เหตุผล :

สถานะลูกค้า : -กรุณาสั่ง-

บันทึกและประมวลผล

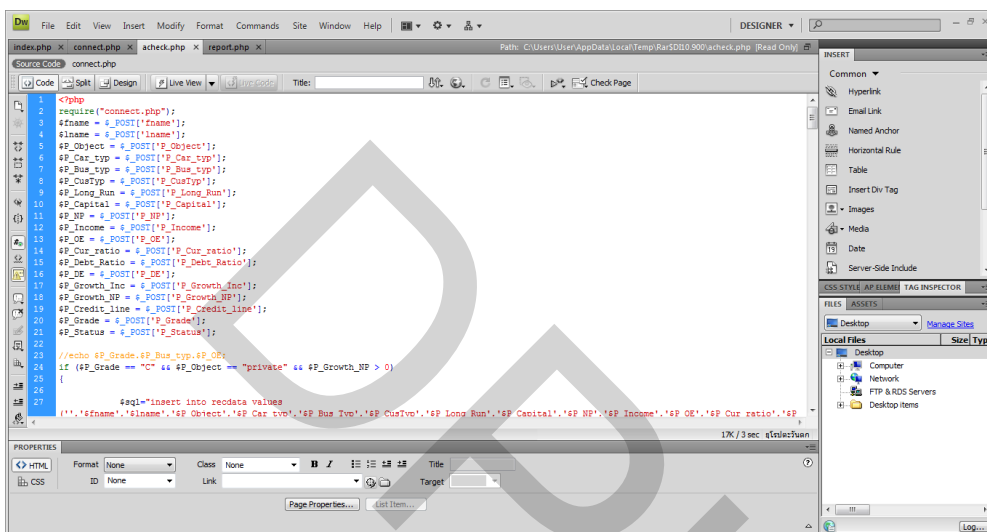
ภาพที่ 4.19 หน้าจอบันทึกข้อมูลลูกค้าที่ต้องการวิเคราะห์ห่อนุมัติสินเชื่อ



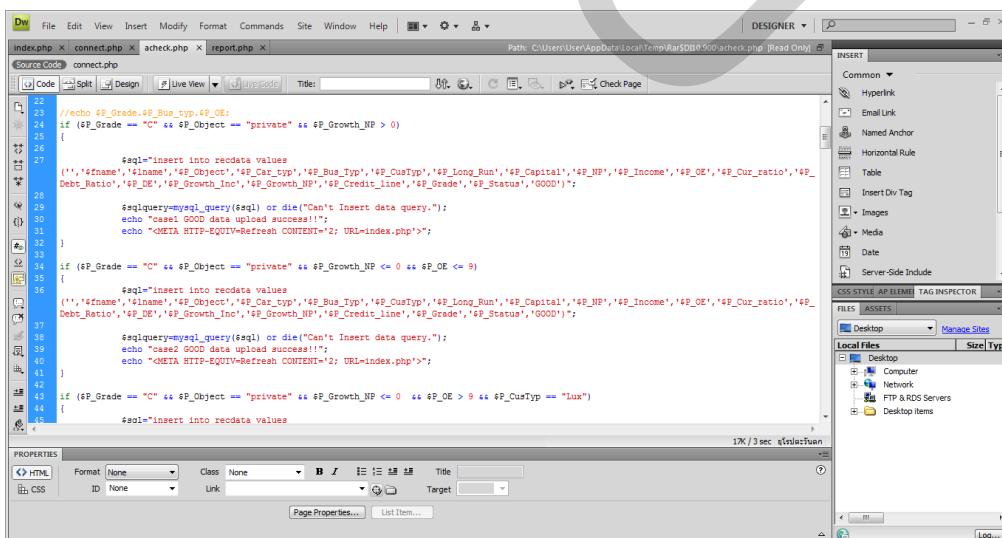
ภาพที่ 4.20 หน้าจอแสดงผลการวิเคราะห์ข้อมูลลูกค้า

### 4.5.2 การเขียนโปรแกรม

นำกฎที่ได้จากการทำเหมืองข้อมูลมาเป็นเงื่อนไขในการเขียนโปรแกรม ผลการสรุปได้กฎทั้งหมด 28 กฎ โดยทำการแปลงกฎที่ได้มาให้เป็น โค้ด ของตัวโปรแกรมใน 1 กฎที่ได้จะเท่ากับ 1 กรณี ในโค้ด ของโปรแกรมจะได้ทั้งหมด 28 กรณี



ภาพที่ 4.21 หน้าจอแสดงการนำกฎมาเขียน โปรแกรมส่วนแรก



ภาพที่ 4.22 หน้าจอแสดงการนำกฎมาเขียน โปรแกรมส่วนที่เป็นเงื่อนไข

```

272 }
273 }
274 }
275 if ($P_Grade == "D" && $P_Bus_Typ == "four" && $P_OE > 0)
276 {
277     $sql="insert into recodata values
278     ('','$name','$name','$P_Object','$P_Car_Typ','$P_Bus_Typ','$P_CusTyp','$P_Long_Run','$P_Capital','$P_NP','$P_Income','$P_OE','$P_Cur_ratio','$P_
279     Debt_Ratio','$P_DE','$P_Growth_Inc','$P_Growth_NP','$P_Credit_Line','$P_Grade','$P_Status','$Bad')";
280
281     $sqlquery=mysql_query($sql) or die("Can't Insert data query.");
282     echo "case3 Bad data upload success!!";
283     echo "<META HTTP-EQUIV=Refresh CONTENT=2; URL=index.php>";
284 }
285 }
286 if ($P_Grade == "C" && $P_Object == "private" && $P_Growth_NP <= 0 && $P_OE > 9 && $P_Car_Typ == "Large" && $P_Long_Run <= 14)
287 {
288     $sql="insert into recodata values
289     ('','$name','$name','$P_Object','$P_Car_Typ','$P_Bus_Typ','$P_CusTyp','$P_Long_Run','$P_Capital','$P_NP','$P_Income','$P_OE','$P_Cur_ratio','$P_
290     Debt_Ratio','$P_DE','$P_Growth_Inc','$P_Growth_NP','$P_Credit_Line','$P_Grade','$P_Status','$Bad')";
291
292     $sqlquery=mysql_query($sql) or die("Can't Insert data query.");
293     echo "case4 Bad data upload success!!";
294     echo "<META HTTP-EQUIV=Refresh CONTENT=2; URL=index.php>";
295 }

```

ภาพที่ 4.23 หน้าจอแสดงการนำกฎหมายเงินโปรแกรมส่วนที่เป็นเงื่อนไขส่วนสุดท้าย

#### 4.5.3 การออกแบบรายงาน

การนำผลที่ได้จากการประมวลผลของโปรแกรม มาออกรายงานเพื่อให้ทราบข้อมูลต่าง ๆ ของลูกค้าที่มาขอสินเชื่อกับทางบริษัทฯ ซึ่งจะมีข้อมูลแสดงสถานะของลูกค้าที่ดี และลูกค้าที่ไม่ดี

ระบบวิเคราะห์สินเชื่อออนไลน์ | ONLINE ANALYSIS FUND

หน้าหลัก | ออกรายงาน

>> รายงาน

ชื่อ	นามสกุล	วัตถุประสงค์การใช้งาน	ประเภทรถยนต์	ประเภทขนส่ง	ประเภทธุรกิจ	ประเภทลูกค้า	ระยะเวลาดำเนินการเสร็จ	หมวดหมู่ประเมิน	ค่าใช้จ่ายสุทธิ	รายได้สุทธิ	ส่วนผู้ถือหุ้น	อัตราส่วนเงินปันผลต่อมูลค่าหุ้น	อัตราเงินปันผลรวม	อัตราส่วนเงินปันผลต่อหุ้น	อัตราการเพิ่มขึ้นของยอดขาย	อัตราการเพิ่มขึ้นของกำไรสุทธิ	วงเงินอนุมัติ	เกรดลูกค้า	สถานะลูกค้า
คุณนาย	สามารถ	ไป	transport	Van		14							4			0	C	Bad	
คุณนาย	วิมล															0	A	GOOD	

หน้า 1 จาก 1

ภาพที่ 4.24 หน้าจอแสดงรายงานข้อมูลลูกค้า

#### 4.6 การนำระบบงานที่พัฒนาขึ้นมาใหม่มาใช้ในองค์กร

ข้อมูลลูกค้าที่นำมาวิเคราะห์เพื่อให้เกรดลูกค้าตามภาพที่ 4.25 และ 4.26



แบบวิเคราะห์และพิจารณาสินเชื่อเพื่อเช่ารถยนต์ (CREDIT SCORING)

เลขที่ CD-080240

วันที่ 25/12/55

บริษัท Test A	จำกัด	ลูกค้าเกรด	<input type="checkbox"/> A	<input type="checkbox"/> B	<input checked="" type="checkbox"/> C	<input type="checkbox"/> D
ประเภทธุรกิจ ขายรถจักรยานยนต์/อะไหล่และบริการซ่อม						
ประเภทรถที่ใช้ รถยนต์เก๋งขนาดใหญ่						
วัตถุประสงค์การใช้ รถคู่ประสงค์การใช้ รถประจำตำแหน่ง						
จนท.การตลาด คุณอริสรา						

กลุ่มธุรกิจ	A	B	C	คะแนน	D
1. ประเภทธุรกิจ	1. พนักงานและสาธารณูปโภค	13. ธุรกิจการเกษตร	25. บริการเฉพาะ, ซ่อมบำรุง, กำจัดขยะ		33. ธุรกิจจัดหางาน
	2. สื่อสารและเทคโนโลยี	14. อาหารและเครื่องดื่ม	26. การท่องเที่ยวและโรงแรม		34. ธุรกิจรถเช่า
	3. ธนาคารและเงินทุนหลักทรัพย์	15. ของใช้ครัวเรือน	27. การขนส่งและโลจิสติกส์		35. รับแทนขนาดเล็กทุนต่ำกว่า 13 ล้านบาท
	4. ประกันภัยและประกันชีวิต	16. ของใช้ส่วนตัวและเวชภัณฑ์	28. ธุรกิจดูแลความสะอาด		
	5. อุตสาหกรรมยานยนต์	17. แฟชั่น	29. ที่ปรึกษาด้านธุรกิจ		
	6. เครื่องมือ-เครื่องจักร	18. บรรจุภัณฑ์	30. การศึกษา		
	7. ปีโตรเคมีและเคมีภัณฑ์	19. กระดาษและวัสดุการพิมพ์	31. มูลนิธิ		
	8. ชิ้นส่วนอิเล็กทรอนิกส์	20. วัสดุก่อสร้าง	32. บุคคล		
	9. บริการพาณิชย์	21. พัฒนาอสังหาริมทรัพย์ขนาดใหญ่			
	10. การแพทย์	22. เหมืองแร่			
	11. ธุรกิจสื่อและสิ่งพิมพ์	23. เครื่องใช้ไฟฟ้า			
	12. ราชการและรัฐวิสาหกิจ	24. คอมพิวเตอร์			
	<input checked="" type="checkbox"/> 5 คะแนน	<input type="checkbox"/> 4 คะแนน	<input type="checkbox"/> 2 คะแนน	5	<input type="checkbox"/> 0 คะแนน
2. ประเภทลูกค้า	<input checked="" type="checkbox"/> ลูกค้าเก่าชำระดี > 3 ปี = ( 5 )	<input type="checkbox"/> ลูกค้าเก่าชำระดี = ( 4 )	<input type="checkbox"/> ลูกค้าใหม่ = ( 3 )	3	<input type="checkbox"/> ลูกค้าเก่าชำระไม่ดี ( 0 )
3. ดำเนินธุรกิจมานาน	<input type="checkbox"/> > 10 ปี = ( 15 )	<input type="checkbox"/> 6-10 ปี = ( 14 )	<input checked="" type="checkbox"/> 3-5 ปี = ( 13 )	13	<input type="checkbox"/> < 3 ปี = ( 0 )
4. ทุนจดทะเบียน (ชำระแล้ว)	<input type="checkbox"/> > 100 ล้านบาท = ( 10 )	<input type="checkbox"/> 61-100 ล้านบาท = ( 9 )	<input checked="" type="checkbox"/> 3-60 ล้านบาท = ( 8 )	0	<input type="checkbox"/> < 3 ล้านบาท = ( 0 )
5. ก้าวโรสุทธิ	<input type="checkbox"/> ถ้าไร 3 ปี PPP = ( 15 )	<input checked="" type="checkbox"/> ถ้าไร 2 ปีล่าสุด PPL = ( 14 )	<input type="checkbox"/> สลับกัน LPP, LPL, PLP = ( 13 )	0	<input type="checkbox"/> ขาดทุนต่อเนื่อง LLL ( 0 )
6. รายได้หลัก	<input type="checkbox"/> เพิ่มขึ้นต่อเนื่อง 2 ปี = ( 15 )	<input type="checkbox"/> เพิ่มขึ้นต่อเนื่อง 1 ปี = ( 14 )	<input checked="" type="checkbox"/> เพิ่ม-ลดสลับกัน = ( 13 )	13	<input type="checkbox"/> ลดลงต่อเนื่อง = ( 0 )
7. ส่วนผู้ถือหุ้น	<input type="checkbox"/> เพิ่มขึ้นต่อเนื่อง 2 ปี = ( 10 )	<input type="checkbox"/> เพิ่มขึ้นต่อเนื่อง 1 ปี = ( 9 )	<input checked="" type="checkbox"/> เพิ่ม-ลดสลับกัน = ( 8 )	0	<input type="checkbox"/> ลดลงต่อเนื่อง 2 ปี ( 0 )
8. สินทรัพย์หมุนเวียน/หนี้สินหมุนเวียน	<input type="checkbox"/> ค่าเฉลี่ย 3 ปี > 1 เท่า = ( 5 )	<input type="checkbox"/> ค่าเฉลี่ย 0.76-1 เท่า = ( 4 )	<input checked="" type="checkbox"/> ค่าเฉลี่ย 0-0.75 เท่า = ( 3 )	3	<input type="checkbox"/> ค่าติดลบ = ( 0 )
9. หนี้สินรวม/สินทรัพย์รวม (%)	<input type="checkbox"/> ค่าเฉลี่ย < 50% = ( 5 )	<input type="checkbox"/> ค่าเฉลี่ย 50-150% = ( 4 )	<input checked="" type="checkbox"/> ค่าเฉลี่ย 151-250% = ( 3 )	4	<input type="checkbox"/> > 250% = ( 0 )
10. หนี้สินรวม/ส่วนของผู้ถือหุ้น (%)	<input type="checkbox"/> ค่าเฉลี่ย < 100% = ( 5 )	<input type="checkbox"/> ค่าเฉลี่ย 100-200% = ( 4 )	<input checked="" type="checkbox"/> ค่าเฉลี่ย 201-300% = ( 3 )	0	<input type="checkbox"/> > 300% = ( 0 )

ภาพที่ 4.25 แบบวิเคราะห์และพิจารณาสินเชื่อ (CREDIT SCORING)-Test A



แบบวิเคราะห์และพิจารณาสินเชื่อเพื่อเช่ารถยนต์ (CREDIT SCORING)

กลุ่มธุรกิจ	A	B	C	คะแนน	D
11. อัตราการเพิ่มของยอดขาย (%)	<input type="checkbox"/> ค่าเฉลี่ย > 10% = (5)	<input type="checkbox"/> ค่าเฉลี่ย 6-10% = (4)	<input type="checkbox"/> ค่าเฉลี่ย 1-5% = (3)	5	<input checked="" type="checkbox"/> ค่าเฉลี่ย ติดลบ = (0)
12. อัตราการเพิ่มของกำไรสุทธิ (%)	<input type="checkbox"/> ค่าเฉลี่ย > 10% = (5)	<input type="checkbox"/> ค่าเฉลี่ย 6-10% = (4)	<input checked="" type="checkbox"/> ค่าเฉลี่ย 1-5% = (3)	0	<input type="checkbox"/> ค่าเฉลี่ย ติดลบ = (0)
			รวมคะแนน	46	
กรณลูกค้า A = คะแนนรวมตั้งแต่ 89-100 คะแนน <input type="checkbox"/> B = คะแนนรวมตั้งแต่ 76 - 88 คะแนน <input type="checkbox"/> C = คะแนนรวมตั้งแต่ 60 - 75 คะแนน <input checked="" type="checkbox"/> D = คะแนนรวมต่ำกว่า 60 คะแนน <input type="checkbox"/> → พิจารณาครั้งที่ 2					
เอกสารประกอบการพิจารณา <input checked="" type="checkbox"/> สำนวนหนังสือรับรอง <input type="checkbox"/> ภพ. 01 และ ภพ. 20 <input type="checkbox"/> BOL					
ความคิดเห็นเพิ่มเติม..... ..... .....			..... ..... .....		..... ..... .....
..... ..... .....			..... ..... .....		..... ..... .....
การพิจารณาครั้งที่ 2					
1. ความน่าเชื่อถือของบริษัทในเครือ..... 2. ความน่าเชื่อถือของสินค้าและซัพพลายเออร์..... 3. ความน่าเชื่อถือของบุคคล..... 4. สัญญาจ้างหรือโครงการในอนาคต..... 5. คำประกันโดยธนาคาร..... 6. คำประกันโดยผู้ถือหุ้นใหญ่..... 7. ประเภทที่เช่า.....					
ผลการพิจารณาครั้งที่ 2..... ..... .....					
วิธีพิจารณาเงิน <input type="checkbox"/> 1 <input type="checkbox"/> 2			อำนาจอนุมัติ		
ผลการพิจารณา <input checked="" type="checkbox"/> อนุมัติ <input type="checkbox"/> ไม่อนุมัติ			มูลค่าสัญญา < 3 ล้านบาท .....DM.MKT. .....DM.CD.		
วงเงินอนุมัติ ปี 2555 จำนวน 0 บาท			มูลค่าสัญญา 3-6 ล้านบาท .....AMD.		
มูลค่าสัญญา 0 บาท			มูลค่าสัญญา 6-20 ล้านบาท .....DMD.		
วงเงินคงเหลือ 0 บาท			มูลค่าสัญญา 20-50 ล้านบาท .....MD.		
			มูลค่าสัญญา 50- 100 ล้านบาท .....EX.COM.		
			มูลค่าสัญญา >100 ล้านบาท .....BOD.		

ภาพที่ 4.25 (ต่อ)



แบบวิเคราะห์และพิจารณาสินเชื่อเพื่อเช่ารถยนต์ (CREDIT SCORING)

เลขที่ CD-080241

วันที่ 25/12/55

บริษัท Test B จำกัด	ถูกคัดเกรด <input type="checkbox"/> A <input type="checkbox"/> B <input checked="" type="checkbox"/> C <input type="checkbox"/> D
ประเภทธุรกิจ ขายรถจักรยานยนต์/อะไหล่และบริการซ่อม	
ประเภทรถที่ใช้ รถกระบะ	วัตถุประสงค์การใช้รถขนส่ง
จนท.การตลาด คุณสุชพันธ์	

กลุ่มธุรกิจ	A	B	C	คะแนน	D
1. ประเภทธุรกิจ	1.พลังงานและสาธารณูปโภค	13.ธุรกิจการเกษตร	25.บริการเฉพาะ,ซ่อมบำรุง,กำจัดขยะ		33.ธุรกิจจัดหางาน
	2.สื่อสารและเทคโนโลยี	14.อาหารและเครื่องดื่ม	26.การท่องเที่ยวและโรงแรม		34.ธุรกิจรถเช่า
	3.ธนาคารและเงินทุนหลักทรัพย์	15.ของใช้ครัวเรือน	27.การขนส่งและโลจิสติกส์		35.รับเหมามาตรับเลือกต่ำกว่า3ล้านบาท
	4.ประกันภัยและประกันชีวิต	16.ของใช้ส่วนตัวและเวชภัณฑ์	28.ธุรกิจดูแลความสะอาด		
	5.อุตสาหกรรมยานยนต์	17.แฟชั่น	29.ที่ปรึกษาด้านธุรกิจ		
	6.เครื่องมือ-เครื่องจักร	18.บรรจุภัณฑ์	30.การศึกษา		
	7.ปิโตรเคมีและเคมีภัณฑ์	19.กระดาษและวัสดุการพิมพ์	31.มูลนิธิ		
	8. ชิ้นส่วนอิเล็กทรอนิกส์	20.วัสดุก่อสร้าง	32.บุคคล		
	9.บริการพาณิชย์	21.พัฒนาอสังหาฯ,รับเหมามาตรับใหญ่			
	10.การแพทย์	22.เหมืองแร่			
	11.ธุรกิจสื่อและสิ่งพิมพ์	23.เครื่องใช้ไฟฟ้า			
	12.ราชการและรัฐวิสาหกิจ	24.คอมพิวเตอร์			
	<input type="checkbox"/> 5 คะแนน	<input checked="" type="checkbox"/> 4 คะแนน	<input type="checkbox"/> 2 คะแนน	5	<input type="checkbox"/> 0 คะแนน
2.ประเภทลูกค้า	<input checked="" type="checkbox"/> ลูกค้าชำระดี>3ปี-(5)	<input type="checkbox"/> ลูกค้าชำระดี-(4)	<input type="checkbox"/> ลูกค้าใหม่-(3)	3	<input type="checkbox"/> ลูกค้าชำระไม่ดี(0)
3.ดำเนินธุรกิจมานาน	<input checked="" type="checkbox"/> >10ปี-(15)	<input type="checkbox"/> 6-10ปี-(14)	<input type="checkbox"/> 3-5ปี-(13)	13	<input type="checkbox"/> <3ปี-(0)
4.ทุนจดทะเบียน (ชำระแล้ว)	<input type="checkbox"/> >100ล้านบาท-(10)	<input type="checkbox"/> 61-100ล้านบาท-(9)	<input checked="" type="checkbox"/> 3-60ล้านบาท-(8)	0	<input type="checkbox"/> <3ล้านบาท-(0)
5.กำไรสุทธิ	<input type="checkbox"/> กำไร 3 ปี PPP=(15)	<input type="checkbox"/> กำไร 2 ปีล่าสุด PPL=(14)	<input checked="" type="checkbox"/> สลับกัน LPP, LPL, PLP=(13)	0	<input checked="" type="checkbox"/> ขาดทุนต่อเนื่องLLL(0)
6.รายได้หลัก	<input type="checkbox"/> เพิ่มขึ้นต่อเนื่อง 2 ปี=(15)	<input type="checkbox"/> เพิ่มขึ้นต่อเนื่อง 1 ปี=(14)	<input checked="" type="checkbox"/> เพิ่ม-ลดสลับกัน=(13)	13	<input type="checkbox"/> ลดลงต่อเนื่อง=(0)
7.ส่วนผู้ถือหุ้น	<input type="checkbox"/> เพิ่มขึ้นต่อเนื่อง 2 ปี=(10)	<input type="checkbox"/> เพิ่มขึ้นต่อเนื่อง 1 ปี=(9)	<input type="checkbox"/> เพิ่ม-ลดสลับกัน=(8)	0	<input checked="" type="checkbox"/> ลดลงต่อเนื่อง2ปี(0)
8.สินทรัพย์หมุนเวียน/หนี้สินหมุนเวียน	<input type="checkbox"/> ค่าเฉลี่ย 3 ปี > 1 เท่า=(5)	<input checked="" type="checkbox"/> ค่าเฉลี่ย 0.76-1 เท่า=(4)	<input type="checkbox"/> ค่าเฉลี่ย 0-0.75 เท่า=(3)	3	<input type="checkbox"/> ค่าติดลบ=(0)
9.หนี้สินรวม/สินทรัพย์รวม (%)	<input type="checkbox"/> ค่าเฉลี่ย< 50%=(5)	<input type="checkbox"/> ค่าเฉลี่ย50-150%=(4)	<input type="checkbox"/> ค่าเฉลี่ย 151-250%=(3)	4	<input checked="" type="checkbox"/> >250%=(0)
10.หนี้สินรวม/ส่วนของผู้ถือหุ้น (%)	<input type="checkbox"/> ค่าเฉลี่ย< 100%=(5)	<input type="checkbox"/> ค่าเฉลี่ย100-200%=(4)	<input type="checkbox"/> ค่าเฉลี่ย 201-300%=(3)	0	<input checked="" type="checkbox"/> >300%=(0)

ภาพที่ 4.26 แบบวิเคราะห์และพิจารณาสินเชื่อ (CREDIT SCORING) -Test B

## แบบวิเคราะห์และพิจารณาสินเชื่อเพื่อเช่ารถยนต์ (CREDIT SCORING)

กลุ่มธุรกิจ	A	B	C	คะแนน	D
11. อัตราการเพิ่มของยอดขาย (%)	<input type="checkbox"/> ค่าเฉลี่ย > 10% = (5)	<input type="checkbox"/> ค่าเฉลี่ย 6-10% = (4)	<input type="checkbox"/> ค่าเฉลี่ย 1-5% = (3)	5	<input checked="" type="checkbox"/> ค่าเฉลี่ย คิดลบ = (0)
12. อัตราการเพิ่มของกำไรสุทธิ (%)	<input type="checkbox"/> ค่าเฉลี่ย > 10% = (5)	<input type="checkbox"/> ค่าเฉลี่ย 6-10% = (4)	<input checked="" type="checkbox"/> ค่าเฉลี่ย 1-5% = (3)	0	<input type="checkbox"/> ค่าเฉลี่ย คิดลบ = (0)
			รวมคะแนน	46	
เกณฑ์การให้คะแนน A = คะแนนรวมทั้งสิ้นตั้งแต่ 89-100 คะแนน <input type="checkbox"/> B = คะแนนรวมทั้งสิ้นตั้งแต่ 76-88 คะแนน <input type="checkbox"/> C = คะแนนรวมทั้งสิ้นตั้งแต่ 60-75 คะแนน <input checked="" type="checkbox"/> D = คะแนนรวมทั้งสิ้นต่ำกว่า 60 คะแนน <input type="checkbox"/>					
เอกสารประกอบการพิจารณา <input checked="" type="checkbox"/> สิ้นหนังกี่รอบ <input checked="" type="checkbox"/> กพ. 01 และ กพ. 20 <input type="checkbox"/> BOL					
ความคิดเห็นเพิ่มเติม..... ..... ..... .....					
การพิจารณาครั้งที่ 2 ..... ..... ..... ..... ..... ..... .....					
ผลการพิจารณาครั้งที่ 2 ..... ..... .....					
วิธีพิจารณาเงิน <input type="checkbox"/> 1 <input type="checkbox"/> 2					
<b>อำนาจอนุมัติ</b> มูลค่าสัญญา < 3 ล้านบาท .....DM.MKT. .....DM.CD. มูลค่าสัญญา 3-6 ล้านบาท .....AMD. มูลค่าสัญญา 6-20 ล้านบาท .....DMD. มูลค่าสัญญา 20-50 ล้านบาท .....MD. มูลค่าสัญญา 50-100 ล้านบาท .....EX.COM.					
ผลการพิจารณา <input checked="" type="checkbox"/> อนุมัติ <input type="checkbox"/> ไม่อนุมัติ					
วงเงินอนุมัติ ปี 2555 จำนวน 0 บาท					
มูลค่าสัญญา 0 บาท					
วงเงินคงเหลือ 0 บาท					

ภาพที่ 4.26 (ต่อ)

นำเกรดที่ได้และข้อมูลลูกค้ามาบันทึกในระบบที่พัฒนาใหม่ เพื่อให้ทราบผลการอนุมัติ ตามภาพที่ 4.27

ชื่อ	นามสกุล	วัตถุประสงค์การใช้งานรถยนต์	ประเภทรถยนต์	ประเภทธุรกิจ	ประเภทลูกค้า	ระยะเวลาดำเนินการธุรกิจ	หมวดทะเบียน	กำไรสุทธิ	รายได้หลัก	ส่วนผู้ถือหุ้น	อัตราสินทรัพย์หมุนเวียน	อัตราหนี้สินรวม/สินทรัพย์รวม :	อัตราหนี้สินรวม/ส่วนผู้ถือหุ้น :	อัตราการเพิ่มของยอดขาย	อัตราการเพิ่มของกำไรสุทธิ	วงเงินอนุมัติ	เกรดลูกค้า	สถานะลูกค้า
Test A		private	Large		CCC	13	8	14	13	8	3	3	3	0	3	0	C	GOOD
Test B		transport	Pickup		CCC	15	8	13	13	0	4	0	0	0	3	0	C	Bad

หน้าที่ :  
1)

ภาพที่ 4.27 หน้าจอแสดงรายงานข้อมูลลูกค้า

จากการนำระบบที่พัฒนาไปใช้ในการอนุมัติให้สินเชื่อลูกค้า จะเห็นได้ว่า กรณีลูกค้าชื่อ Test B (นามสมมติ) หากอนุมัติจากผลคะแนนของลูกค้าตามระบบงานเดิมเพียงอย่างเดียว จะทำให้มีการอนุมัติสินเชื่อต่อลูกค้ารายนั้น แต่หากนำเกรดของลูกค้าที่ได้และข้อมูลของลูกค้ามาบันทึกในระบบที่พัฒนาใหม่ ผลที่ได้คือ Bad หมายถึง ลูกค้ารายดังกล่าวไม่ผ่านการอนุมัติสินเชื่อ ช่วยให้บริษัทมีการพิจารณาสินเชื่อของลูกค้าได้อย่างมีประสิทธิภาพมากยิ่งขึ้น อีกทั้งยังอาจช่วยลดความเสี่ยงด้านหนี้สูญ ในการอนุมัติให้เช่ารถยนต์ต่อลูกค้าที่ไม่เหมาะสม

## บทที่ 5

### สรุปผลการวิจัย และข้อเสนอแนะ

#### 5.1 สรุปผลการวิจัย

การวิจัยเรื่อง การใช้เหมืองข้อมูลช่วยในการตัดสินใจการให้สินเชื่อ ของบริษัท กรุงไทย คาร์เร็นท์ แอนด์ ลีส จำกัด (มหาชน) มีวัตถุประสงค์เพื่อวิเคราะห์ข้อมูลลูกค้าเช่ารถ เพื่อพิจารณาสินเชื่อได้อย่างมีประสิทธิภาพ และลดปริมาณหนี้สูญ โดยมีขั้นตอนหลักๆ 5 ส่วน ได้แก่

ส่วนที่ 1 ศึกษาและเก็บรวบรวมข้อมูลประวัติลูกค้าที่พิจารณาให้สินเชื่อแล้ว ในปี 2551–2553 จำนวน 441 ราย อยู่ในรูปแบบไฟล์ Excel

ส่วนที่ 2 เตรียมข้อมูลให้เป็นไปตามรูปแบบมาตรฐานเดียวกัน และคัดเลือกเฉพาะข้อมูลที่สมบูรณ์และรูปแบบถูกต้อง ครบถ้วน และลดความหลากหลายของข้อมูล เพื่อนำมาใช้กับโปรแกรม Weka 3.5.7 ได้ โดยมี 3 ขั้นตอน ได้แก่ ขั้นตอนการโอนย้ายข้อมูล (Data Transfer) การลดขนาดของข้อมูล (Data Reduction) และการทำความสะอาดข้อมูล (Data Cleaning)

ส่วนที่ 3 การทดสอบข้อมูล และเลือกเทคนิคที่เหมาะสม ซึ่งในการศึกษาวิจัยครั้งนี้ ได้ทดสอบเทคนิคการจำแนกกลุ่ม จำนวน 3 เทคนิค ให้ผลลัพธ์เป็นกฎ ได้แก่ โมเดลการจำแนกกลุ่มแบบ Decision Tree: C4.5 หรือเรียกว่าโมเดล Classifiers trees J48 โมเดลการจำแนกกลุ่มแบบ Decision Rule: Part หรือเรียกว่าโมเดล Classifiers PART และ โมเดลการจำแนกกลุ่มแบบ Bayes: Naïve Bayes หรือเรียกว่าโมเดล Classifier Naïve Bayes ในโปรแกรม Weka จากการทดสอบการจำแนกกลุ่มดังกล่าวข้างต้น พบว่า การจำแนกกลุ่มแบบ Decision Tree: C4.5 ให้ผลลัพธ์ที่ดีที่สุด เนื่องจากมีความถูกต้องมากที่สุด สามารถแบ่งกลุ่มได้ตามเงื่อนไขที่ชัดเจนมากที่สุด และสามารถแสดงกฎให้สามารถนำไปวิเคราะห์กฎต่อไปได้ โดยผลลัพธ์มีความถูกต้อง 90.47 เปอร์เซ็นต์ และได้กฎจำนวน 40 กฎ

ส่วนที่ 4 การวิเคราะห์ข้อมูล จากผลการทดสอบการจำแนกกลุ่มข้อมูลแบบ Decision Tree: C4.5 เป็นการจับกลุ่มที่เหมาะสมที่สุดในการทดสอบเนื่องจากมีกฎจำนวนมาก และมีเงื่อนไขที่เหมาะสม แต่มีบางกฎที่มีเงื่อนไขไม่เหมาะสม กฎที่มีความถูกต้องน้อยกว่า 90% บางกฎไม่มีตัวอย่างสนับสนุนกับกฎดังกล่าว หรือ บางกฎที่ไม่สมบูรณ์ จึงทำการลบกฎดังกล่าวออก จำนวน

กฎ 12 กฎ จากทั้งหมด 40 กฎ เหลือ 28 กฎ เพื่อใช้เป็นเงื่อนไขในการเขียนโปรแกรมระบบวิเคราะห์สินเชื่อต่อไป

ส่วนที่ 5 การพัฒนาระบบวิเคราะห์สินเชื่อ เป็นการนำกฎที่ได้จากการทำเหมืองข้อมูล มาใช้เป็นเงื่อนไขในการเขียนโปรแกรมระบบวิเคราะห์สินเชื่อออนไลน์ เพื่อเป็นเครื่องมือช่วยในการตัดสินใจการอนุมัติสินเชื่อ ของ บริษัท กรุงไทยคาร์เร้นท์ แอนด์ ลีส จำกัด (มหาชน) โดยใช้โปรแกรม PHP และใช้ฐานข้อมูล My SQL ในการพัฒนาโปรแกรม โดยโปรแกรมที่ได้จะเป็นเครื่องมือช่วยวิเคราะห์สินเชื่อให้มีประสิทธิภาพมากยิ่งขึ้นเพื่อลดปริมาณหนี้สูญให้น้อยลง ซึ่งควรทำควบคู่กับการพิจารณา สภาพการณ์ปัจจุบันประกอบด้วย เช่น การแข่งขัน ภาวะเศรษฐกิจ และ ตลาดอุตสาหกรรมรถเช่า เป็นต้น

## 5.2 ปัญหาและแนวทางแก้ไข

1. ข้อมูลประวัติลูกค้าที่ไม่ผ่านการพิจารณาสินเชื่อ บริษัทไม่ได้มีการจัดเก็บไว้ จึงไม่ได้นำข้อมูลมาทดสอบ จึงได้กฏลูกค้าไม่ดี (BAD) จำนวนน้อย และหากมีการเก็บข้อมูลดังกล่าวและนำมาทดสอบ ผลที่ได้อาจเป็นลูกค้าดี (GOOD) ทำให้บริษัทมีโอกาสในการขยายฐานลูกค้าเพิ่มขึ้น
2. การเก็บข้อมูลของบริษัท มีการบันทึกไม่ครบถ้วน หรือ ไม่ถูกต้อง เช่น ไม่มีระบุวัตถุประสงค์การใช้รถยนต์ ไม่ระบุเกรดลูกค้า รวมยอดคะแนนผิดพลาด หากมีระบบการเตือนให้การบันทึกข้อมูลให้สมบูรณ์ ถูกต้อง ก่อนการบันทึก จะทำให้ได้ข้อมูลที่สมบูรณ์และผลการวิจัยที่ได้มีความน่าเชื่อถือมากยิ่งขึ้น

## 5.3 ข้อเสนอแนะสำหรับการพัฒนาต่อไป

ในหัวข้อของการเสนอแนะในการดำเนินงานสามารถสรุปเป็นข้อคิดเห็นภายหลังการดำเนินการวิจัยได้ดังต่อไปนี้

1. ข้อมูลเป็นปัจจัยสำคัญที่นำมาทำเหมืองข้อมูล ดังนั้น ควรมีการนำระบบคลังข้อมูล (Data Warehouse) มาพัฒนาต่อเพื่อให้บริษัทสามารถนำข้อมูลมาวิเคราะห์ให้เกิดประโยชน์อย่างมีประสิทธิภาพ
2. การทำเหมืองข้อมูล สามารถนำมาเป็นต้นแบบในการหาความสัมพันธ์ของการวิเคราะห์การตัดสินใจการเช่ารถของลูกค้าอย่างต่อเนื่อง เพื่อการรักษาฐานลูกค้าเก่าในการตัดสินใจเช่าซ้ำ หรือนำมาหาความสัมพันธ์ของปัจจัยที่มีผลต่อการตัดสินใจเช่ารถของลูกค้า เพื่อขยายฐานลูกค้าใหม่ เป็นต้น

DU  
P  
U

บรรณานุกรม

## บรรณานุกรม

### ภาษาไทย

#### หนังสือ

- จิราพร ยิ่งกว่าชาติ และคณะ. (2549). การทำนายผลสำเร็จการศึกษาของนักศึกษาระดับอุดมศึกษา ด้วยการเรียนรู้แบบเบย์และการทำเหมืองข้อมูล (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). กรุงเทพฯ: มหาวิทยาลัยศรีปทุม.
- ชลนิตา สาระ. (2550). การจำแนกกลุ่มสถานภาพการสำเร็จการศึกษาโดยแบบจำลองต้นไม้ตัดสินใจ (สารนิพนธ์ปริญญาโทบริหารธุรกิจ). กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- ณัฐริน เจริญเกียรติบวร และปานใจ ชารัทสนวงศ์. (2549). การใช้เทคนิคเหมืองข้อมูลเพื่อช่วยในการแนะแนวการศึกษาต่อระดับอุดมศึกษา (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). กรุงเทพฯ: มหาวิทยาลัยศิลปากร.
- บุญมา เฟ่งชวน. (2548). การศึกษาตัวแบบและพัฒนาตัวแบบ เพื่อนำตัวแบบที่ได้นำไปประยุกต์ใช้ในการสนับสนุนการตัดสินใจด้านการผลิตบัณฑิต (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). กรุงเทพฯ: มหาวิทยาลัยศิลปากร.
- ปรีชา ยามันสะบีตัน และคณะ. (2549). การทำเหมืองข้อมูลมาใช้ในการบริหารลูกค้าสัมพันธ์ (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- ยิ่งลักษณ์ ยิ่งสวัสดิ์. (2553). การพัฒนาเครื่องมือจัดกลุ่มลูกค้าโดยใช้เทคนิคค้ำไม้หนึ่ง (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- วิโรจน์ ภาคคีรี, กาญจนา วิริยะพันธ์. (2550, กรกฎาคม – ธันวาคม). การวิเคราะห์รูปแบบการบุกรุกข้อมูลบนเครือข่าย โดยใช้เทคนิคค้ำไม้หนึ่ง. วารสารเทคโนโลยีสารสนเทศ, 3(6). หน้า 40-46.



กฤษณะ ไวยมัย, ชิดชนก ส่งศิริ และธนาวินท์ รักธรรมานนท์. (2544). การใช้เทคนิคด้าไมน์นิ่ง เพื่อพัฒนาคุณภาพการศึกษานิติตคณะวิศวกรรมศาสตร์. The NECTEC Technical journal, หน้า 135.

DPU

DU  
P  
U  
ภาคผนวก

## ภาคผนวก

ผลการรันฉบับเต็ม (วิธีการอ่าน ดูได้จากกฎที่คัดเลือกในบทที่ 4)

=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.6 -M 2

Relation: current

Instances: 441

Attributes: 17

Object

Car typ

Bus Typ

CusTyp

Long Run

Capital

NP

Income

OE

Cur ratio

Debt Ratio

DE

Growth Inc

Growth NP

credit line

Grade

Status

Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

-----

Grade = C

```

| Object = private
| | Growth NP <= 0
| | | OE <= 9: Good (59.63/2.0)
| | | OE > 9
| | | | Car typ = Large
| | | | | Long Run <= 14: Bad (2.0)
| | | | | Long Run > 14: Good (5.0/1.0)
| | | | Car typ = Van: Good (1.0)
| | | | Car typ = MPV: Good (2.0)
| | | | Car typ = Pickup: Good (0.0)
| | | | Car typ = Medium: Good (2.0)
| | | | Car typ = Small: Good (4.0/1.0)
| | | | Car typ = Lux: Good (1.0)
| | Growth NP > 0: Good (115.0)
| Object = transport
| | Long Run <= 13: Good (25.34)
| | Long Run > 13
| | | Car typ = Large: Good (3.0)
| | | Car typ = Van
| | | | DE <= 3: Good (2.0)
| | | | DE > 3: Bad (2.0)
| | | Car typ = MPV: Good (1.0)
| | | Car typ = Pickup
| | | | Bus Typ = four

```

| | | | | OE <= 0: Bad (2.0)  
 | | | | | OE > 0: Good (13.0/2.0)  
 | | | | Bus Typ = two: Good (4.0)  
 | | | | Bus Typ = five: Good (28.0/1.0)  
 | | | | Bus Typ = zero: Good (1.0)  
 | | | Car typ = Medium: Good (3.0)  
 | | | Car typ = Small: Good (20.0/1.0)  
 | | | Car typ = Lux: Good (0.0)  
 | Object = mkt: Good (9.03)  
 Grade = D  
 | Bus Typ = four  
 | | OE <= 0: Good (7.0/2.0)  
 | | OE > 0: Bad (2.0)  
 | Bus Typ = two: Good (8.0)  
 | Bus Typ = five: Good (10.0)  
 | Bus Typ = zero: Good (6.0/1.0)  
 Grade = A: Good (48.0/3.0)  
 Grade = B  
 | Car typ = Large: Good (17.0)  
 | Car typ = Van: Good (3.0)  
 | Car typ = MPV: Good (2.0)  
 | Car typ = Pickup: Good (6.0/1.0)  
 | Car typ = Medium: Good (12.0/1.0)  
 | Car typ = Small  
 | | Income <= 13  
 | | | Long Run <= 14: Bad (3.0/1.0)  
 | | | Long Run > 14: Good (3.0/1.0)  
 | | Income > 13: Good (6.0)  
 | Car typ = Lux: Good (0.0)

Grade = z: Good (3.0)

Number of Leaves : 40

Size of the tree : 56

Time taken to build model: 1.63 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	399	90.4762 %
Incorrectly Classified Instances	42	9.5238 %
Kappa statistic	-0.0457	
Mean absolute error	0.1182	
Root mean squared error	0.2941	
Relative absolute error	101.1442 %	
Root relative squared error	122.629 %	
Total Number of Instances	441	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.964	1	0.937	0.964	0.95	0.454	Good
	0	0.036	0	0	0.454		Bad
Weighted Avg.	0.905	0.941	0.879	0.905	0.892	0.454	

=== Confusion Matrix ===

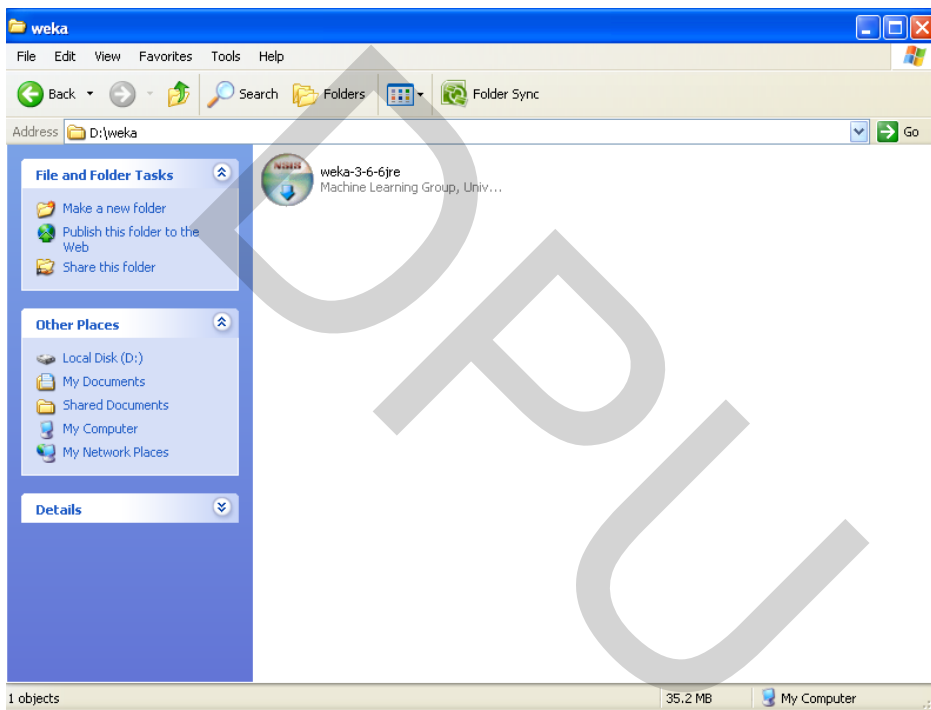
a b <-- classified as

399 15 | a = Good

27 0 | b = Bad

### ขั้นตอนการลงโปรแกรม Weka

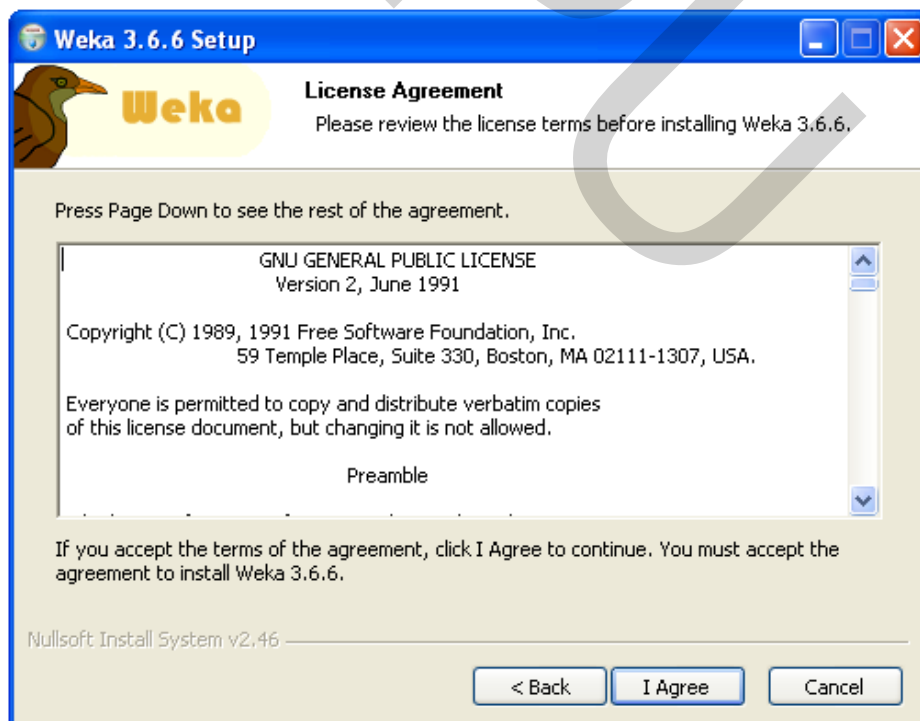
หลังจากเปิด D:\ แล้ว ให้คลิกที่ Folder Weka จากนั้นคลิกที่โปรแกรม weka-3-6-6jre



คลิกปุ่ม Next เพื่อทำการติดตั้ง Weka

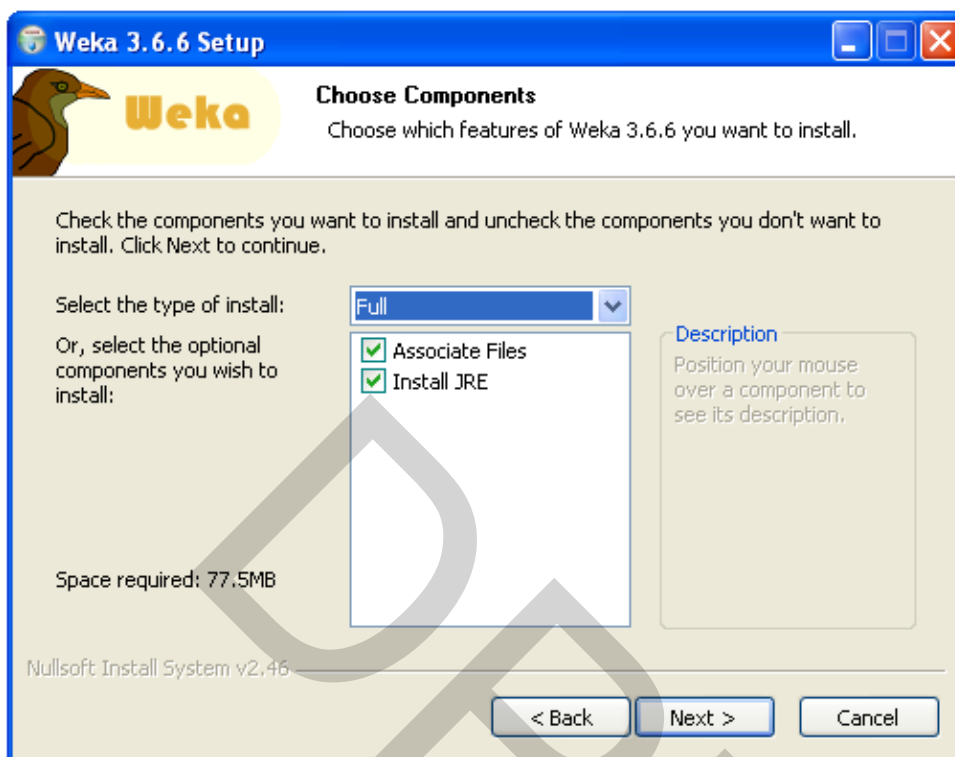


คลิกปุ่ม I Agree เพื่อทำการตกลง และทำการติดตั้งต่อไป

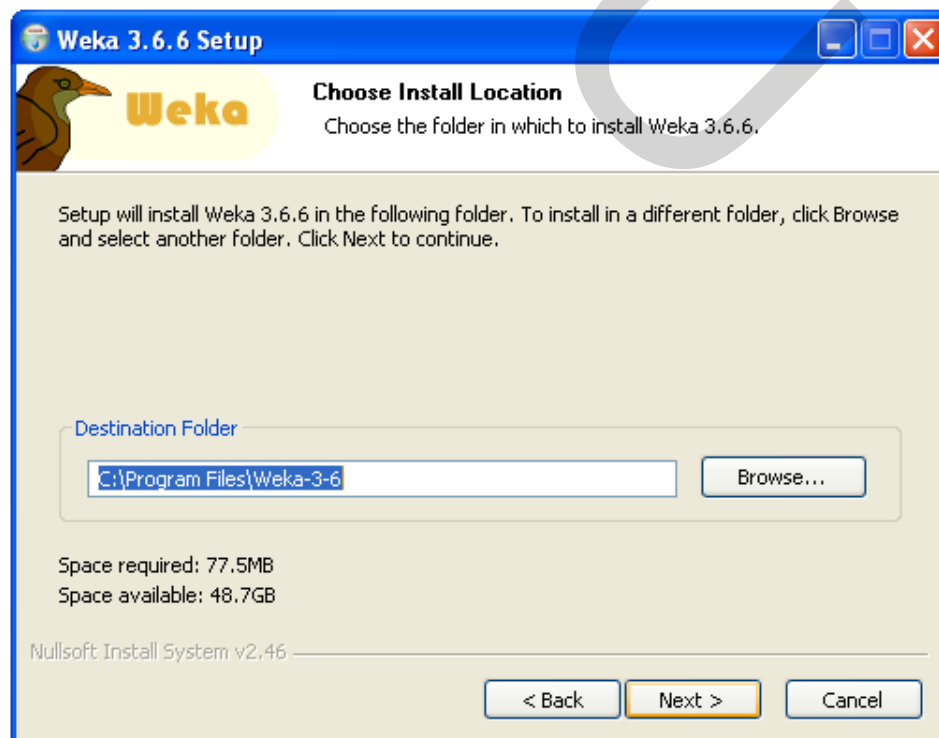


คลิกปุ่ม Next เพื่อทำการ Install

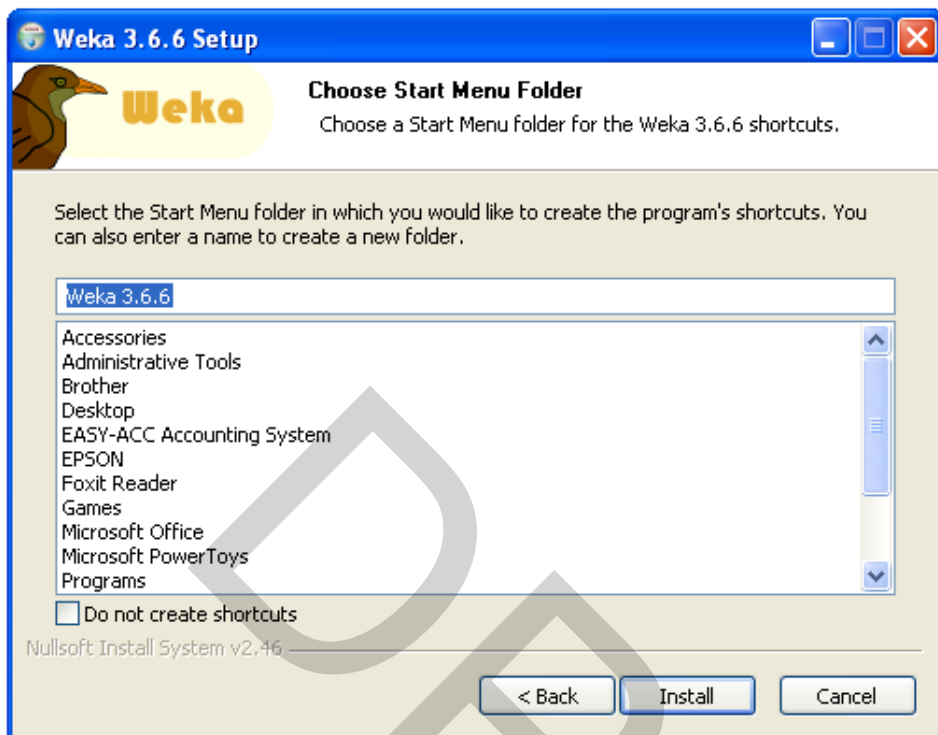




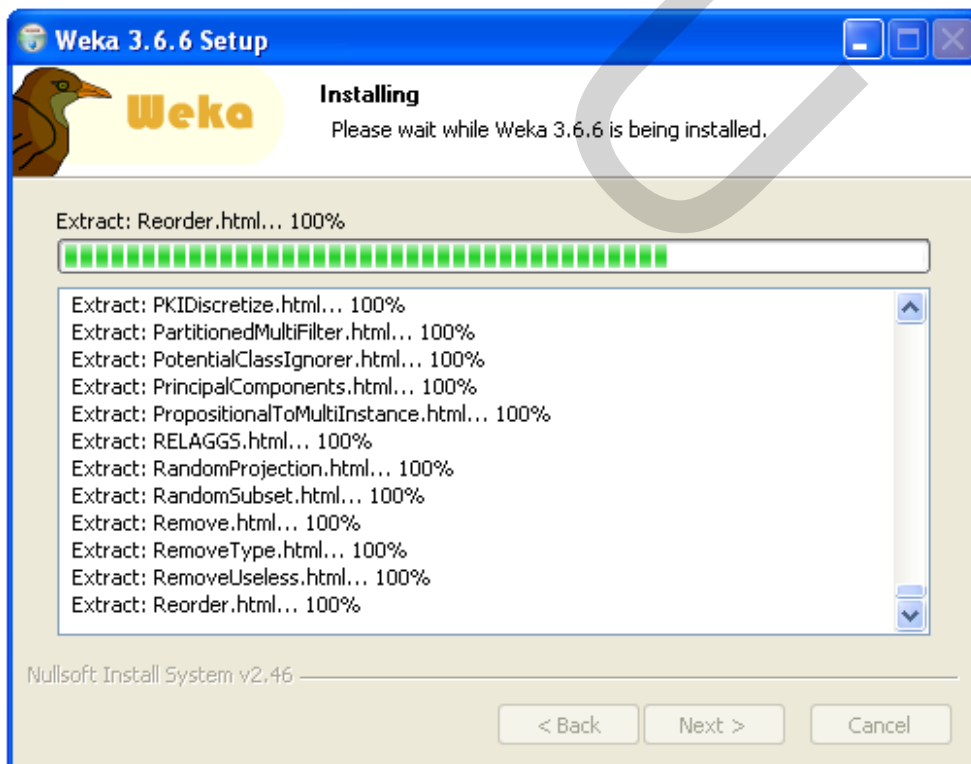
คลิกปุ่ม Browse เพื่อเลือกเพิ่มเก็บ Weka เมื่อเลือกเพิ่มเสร็จแล้ว คลิกปุ่ม Next เพื่อทำการติดตั้งต่อไป



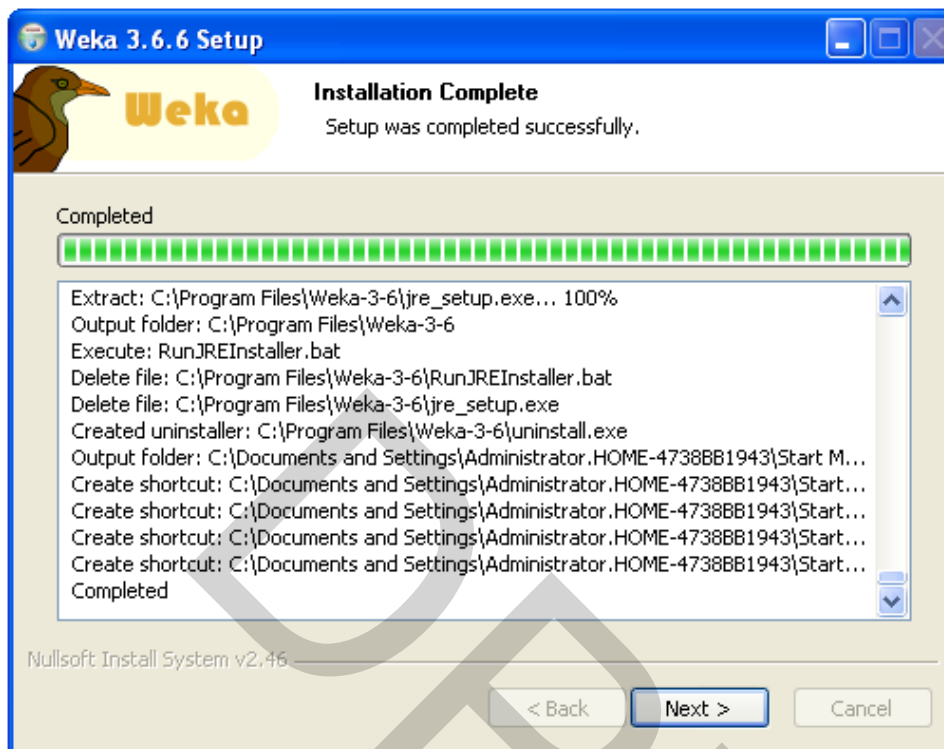
คลิก Install เพื่อเริ่มต้นการติดตั้งโปรแกรม



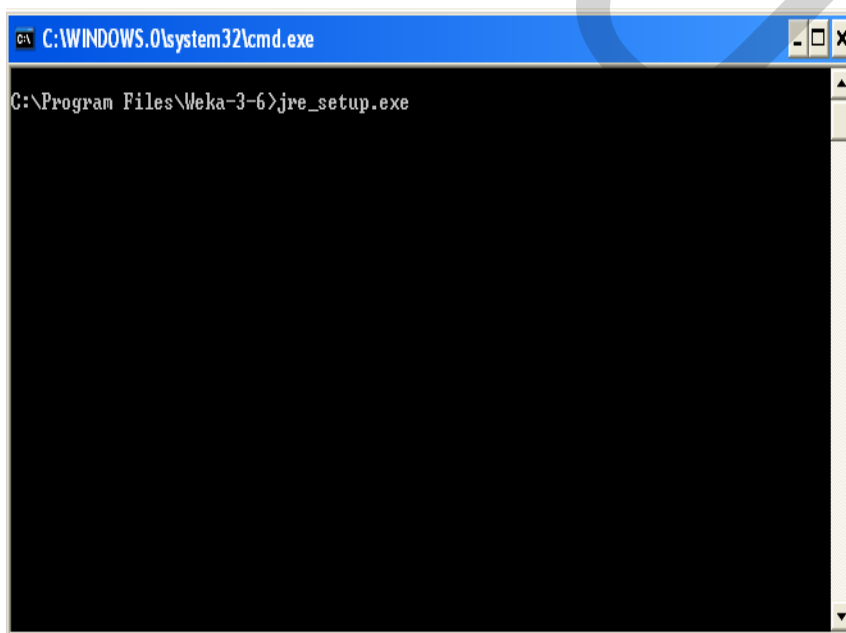
จะพบหน้าจอ Installing แสดงการประมวลผลการติดตั้งโปรแกรม



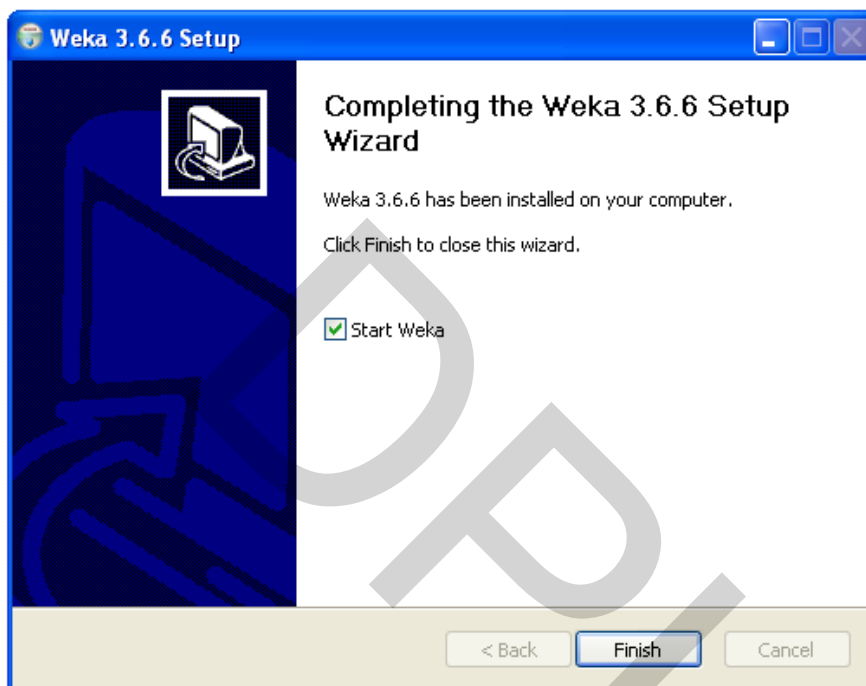
รอนจนกว่าแถบสีเขียวจะขึ้นเต็ม



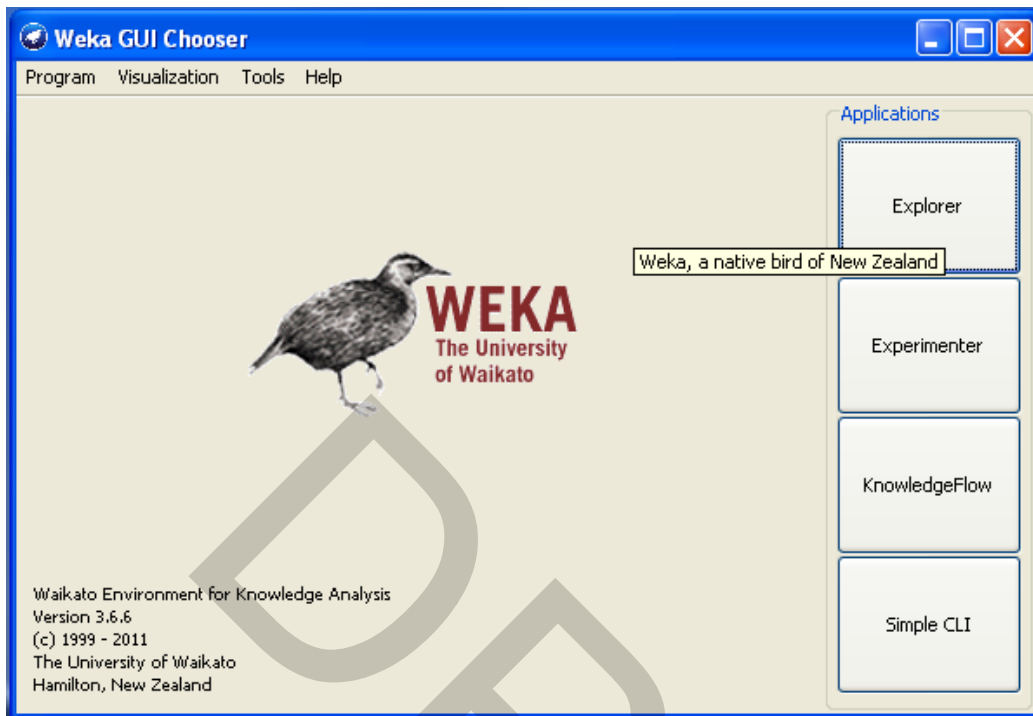
เมื่อประมวลผลเสร็จแล้ว จะแสดงหน้าจอขึ้นมา เราไม่ต้องคลิกปิด เพราะหน้าจอจะถูกปิดโดยอัตโนมัติหลังจากการติดตั้งเสร็จเรียบร้อยแล้ว



แสดงการติดตั้งโปรแกรม Waka ที่เสร็จสมบูรณ์แล้ว ตรง Start Waka ที่ถูกเช็คไว้ เพื่อให้รันหน้าจอโปรแกรม Waka 3.6.6 ถ้าไม่ต้องการให้รันโปรแกรมในขณะนี้ ให้คลิกเพื่อให้เครื่องหมายนั้นหายไป สุดท้ายให้คลิกที่ Finish



แสดงหน้าจอการทำงานของโปรแกรม Weka 3.6.6 เมื่อติดตั้งเสร็จสมบูรณ์จะพบหน้าจอดังกล่าว



### ประวัติผู้เขียน

ชื่อ-นามสกุล                      นางสาวทิพย์ธิดา วงศ์พิพันธ์  
ประวัติการศึกษา                  ปริญญาบัญชีบัณฑิต คณะบัญชี มหาวิทยาลัยธุรกิจบัณฑิต  
ปีที่สำเร็จการศึกษา 2548  
ตำแหน่งงานปัจจุบัน            พนักงานบัญชี  
สถานที่ทำงานปัจจุบัน          บริษัท ไทยคาร์เร็นท์ แอนด์ ลีส จำกัด (มหาชน)