รายงานวิจัยฉบับสมบูรณ์

การวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคเหมืองข้อมูล

Factors Analysis of Undergraduate Student's Retirement Using by Data Mining

โดย

นางขวัญฤทัย นกแก้ว

ได้รับทุนอุดหนุนจากงบประมาณบำรุงการศึกษาประจำปี 2555

มหาวิทยาลัยราชภัฏยะลา

ชื่อ : นางขวัญฤทัย นกแก้ว
ชื่องานวิจัย : การวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคเหมืองข้อมูล
สาขา: เทคโนโลยีสารสนเทศ ภาควิชาวิทยาศาสตร์ประยุกต์
คณะ: วิทยาศาสตร์เทคโนโลยีและการเกษตร
มหาวิทยาลัย: มหาวิทยาลัยราชภัฏยะลา
ปีการศึกษา : 2555

บทคัดย่อ

งานวิจัยนี้เป็นการนำเสนอการวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาระดับปริญญาตรี โดยใช้การค้นหากฎความสัมพันธ์ (Association Rule Discovery) ซึ่งเป็นเทคนิคหนึ่งของเหมืองข้อมูล(Data Mining) เทคนิคการจำแนกประเภทข้อมูล(Data Classification) เป็นเทคนิคหนึ่งในการจำแนกประเภทข้อมูล ให้ความแม่นยำในการทำนายสูง เนื่องจากมีการนำเทคนิคการหากฎความสัมพันธ์เข้ามารวมกับเทคนิคการจำแนกประเภทข้อมูลเพื่อเพิ่มประสิทธิภาพในการจำแนกประเภทข้อมูล ในงานวิจัยนี้ใช้ชุดข้อมูลของนักศึกษามหาวิทยาลัยราชภัฏยะลา ที่เข้าศึกษาระหว่างปี พ.ศ. 2550-2553 ซึ่งพ้นสภาพการเป็นนักศึกษา จำนวน 1804 คน มีจำนวน 14 ปัจจัย ผู้วิจัยได้ทำการวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษา ผลจากการวิเคราะห์พบว่ามีปัจจัยสำคัญ 2 ปัจจัยคือ สาขาที่นักศึกษาเลือกเรียนในระดับปริญญาตรี และวุฒิการศึกษาเดิมในระดับมัธยมปลาย

คำสำคัญ: เหมืองข้อมูล, การจำแนกประเภทข้อมูล

Research: Factors Analysis of Undergraduate Student's Retirement Using by Data mining
Author : Mrs. Kwanrutai Nokkaew
Major Program: Information Technology
Faculty: Faculty of Technology Science and Agriculture
Yala Rajabhat University

Abstract

This research presents an analysis of the factors that affect of undergraduates. Search using association rules. Which is the data mining techniques for classification. As one technique for data classification. The high precision of prediction. Because the technique is given to the relationship with the classification techniques to improve the classification. In this research, a series of Yala Rajabhat University. Admission during the year 2550-2553, which will no study be a student in 1804 has 13 inputs, the researcher analyzed the factors that affect a student's dismissal. The results of the analysis showed that the two major factors Field study in an undergraduate student. And education in the high school.

Keyword : Data mining, Associative Classification

กิตติกรรมประกาศ

งานวิจัยฉบับนี้ได้รับการสนับสนุนทุนวิจัยจากสถาบันวิจัยชายแดนใต้ ในส่วนของงบประมาณบำรุงการศึกษา ที่เปิดโอกาสให้อาจารย์สามารถสร้างองค์ความรู้ใหม่และพัฒนาผลงานวิจัยที่มีประโยชน์กับสถาบันวิจัย และมหาวิทยาลัยราชภัฏยะลา

ขอขอบพระคุณอาจารย์ซอและ เกปัน ผู้อำนวยการกองบริการการศึกษา มหาวิทยาลัยราชภัฏยะลา และนายซอพี บูแด ที่ให้ความอนุเคราะห์ข้อมูลนักศึกษาเพื่อนำมาประกอบการทำวิจัยในครั้งนี้

ขอบคุณนักศึกษาสาขาเทคโนโลยีสารสนเทศ ชั้นปีที่ 3 ปีการศึกษา 2555 ในการจัดการจำแนกข้อมูลใช้ในการวิเคราะห์

นางขวัญฤทัย นกแก้ว

ผู้วิจัย

สารบัญ

		หน้า
บทคัดย่อภาษาไทย		ก
บทคัดย่อภาษาอังกฤษ		ข
กิตติกรรมประกาศ		ค
สารบัญ		ง
สารบัญภาพ		ฉ
สารบัญตาราง		ช
บทที่ 1	บทนำ	1
	ความสำคัญและที่มาของปัญหา	1
	วัตถุประสงค์	2
	สมมติฐานการวิจัย	3
	ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2	เอกสารและงานวิจัยที่เกี่ยวข้อง	5
	การทำเหมืองข้อมูล (Data Mining)	5
	กระบวนการค้นหาความรู้ (Knowledge Discovery in Database (KDD))	6
	ขั้นตอนการทำงานของ เหมืองข้อมูล	8
	อัลกอริทึมที่ใช้การทำ เหมืองข้อมูล	12
	การวิเคราะห์ปัจจัย	24
	เทคโนโลยีที่นำมาพัฒนางานวิจัย	28
	งานวิจัยที่เกี่ยวข้อง	30
บทที่ 3	วิธีดำเนินการวิจัย	33
	กฎระเบียบการพ้นสภาพการเป็นนักศึกษา	33
	ขั้นตอนการวิเคราะห์โดยใช้ เหมืองข้อมูล	34
	ขั้นตอนการวิเคราะห์โดยใช้ เครื่องมือในงานวิจัย	37
บทที่ 4	ผลการดำเนินงาน	43
	การกลั่นกรองข้อมูล (Data Cleaning)	43
	การรวบรวมข้อมูล (Data Integration)	43
	การคัดเลือกข้อมูล (Data Selection)	44

การแปลงรูปแบบข้อมูล (Data Transformation) 45
การทำเหมืองข้อมูล (Data Mining) 45
การประเมินรูปแบบ (Pattern Evaluation) 48
บทที่ 5 สรุปผล อภิปรายผล และข้อเสนอแนะ 51
สรุปผลการวิจัย 53
บรรณานุกรม 55
ภาคผนวก ก บันทึกข้อความขออนุเคราะห์ข้อมูล 56
ภาคผนวก ข ผลการวิเคราะห์กฎความสัมพันธ์ (Association Rule) 58
ภาคผนวก ค ประวัติผู้วิจัย 66

สารบัญภาพ

		หน้า
ภาพที่ 2.1	กระบวนการค้นหาความรู้ (Knowledge Discovery Process)	7
ภาพที่ 2.2	กระบวนการจำแนกข้อมูล	14
ภาพที่ 2.3	Itemset lattice	18
ภาพที่ 2.4	Itemset lattice กรณีกำจัด Infrequent itemsets	19
ภาพที่ 2.5	ขั้นตอนการสร้าง Frequent itemsets ด้วยขั้นตอนวิธี Apriori	20
ภาพที่ 2.6	แสดงองค์ประกอบการวิเคราะห์ปัจจัย	24
ภาพที่ 2.7	โลโก้ของซอฟต์แวร์วีก้า (WEKA)	29
ภาพที่ 3.1	โปรแกรม WEKA	39
ภาพที่ 3.2	โมดูล Explorer ของโปรแกรม WEKA	40
ภาพที่ 3.3	ไฟล์ข้อมูลที่ต้องการวิเคราะห์	40
ภาพที่ 3.4	เลือกเทคนิค Associations	41
ภาพที่ 3.5	ผลจากการวิเคราะห์ เทคนิค Associations ขั้นตอนวิธี PredictiveApriori	41
ภาพที่ 4.1	ข้อมูลนักศึกษาพ้นสภาพในปีการศึกษา 2550 - 2553	44
ภาพที่ 4.2	การแปลงข้อมูลให้อยู่ในรูปแบบ CSV	45
ภาพที่ 4.3	หน้าจอโปรแกรม WEKA 3.6.8	46
ภาพที่ 4.4	การนำเข้าข้อมูลเพื่อใช้ทำเหมืองข้อมูล	47
ภาพที่ 4.5	เลือกเทคนิคที่นำมาใช้ในการวิเคราะห์ปัจจัยขั้นตอนวิธี Apriori	47
ภาพที่ 5.1	แสดงขั้นตอนการวิเคราะห์ปัจจัย	52

สารบัญตาราง

		หน้า
ตารางที่ 2.1	นิยามเทอมทั่วไป	16
ตารางที่ 3.1	ตัวอย่างข้อมูลประวัตินักศึกษาที่พ้นสภาพ	34
ตารางที่ 3.2	ตัวอย่างข้อมูลประวัตินักศึกษาที่ทำให้สมบูรณ์	36
ตารางที่ 3.3	ตัวอย่างตารางข้อมูลนักศึกษาขั้นต้น	37
ตารางที่ 4.1	แบบฟอร์มการจัดเก็บข้อมูล	43
ตารางที่ 4.2	สรุปสาขาที่มีจำนวนนักศึกษาพ้นสภาพ	48

บทที่ 1
บทนำ

ความสำคัญและที่มาของปัญหา

มหาวิทยาลัยราชภัฏยะลา เป็นสถาบันอุดมศึกษาเพื่อพัฒนาท้องถิ่นจังหวัดชายแดนภาคใต้ โดยมุ่งเน้นการพัฒนาทรัพยากรมนุษย์ให้มีคุณภาพชีวิตที่สูงขึ้น ด้วยกระบวนการพัฒนาองค์ความรู้บนพื้นฐานการบูรณาการศาสตร์สากล และภูมิปัญญาท้องถิ่น การผลิตบัณฑิตระดับปริญญาตรี และระดับปริญญาโท ในทุกปีการศึกษาจะประสบกับปัญหานักศึกษาพ้นสภาพการเป็นนักศึกษา เนื่องจากนักศึกษามีผลการเรียนไม่เข้าหลักเกณฑ์มาตรฐาน โดยทางมหาวิทยาลัยได้กำหนดเกณฑ์ตามเงื่อนไขการพิจารณา 6 เงื่อนไข ดังนี้ เงื่อนไขแรก นักศึกษาที่ได้รับค่าคะแนนเฉลี่ยสะสมต่ำกว่า 1.6 ในภาคการศึกษาปกติที่สองที่เข้าศึกษาในมหาวิทยาลัย(ไม่นับภาคเรียนที่ลาพักหรือถูกให้พัก) เงื่อนไขที่สอง นักศึกษาได้รับค่าระดับคะแนนเฉลี่ยสะสมต่ำกว่า 1.70 ในภาคการศึกษาปกติถัดไปหลังจากได้รับการรอพินิจครั้งที่ 1 เงื่อนไขที่สาม นักศึกษาได้รับค่าระดับคะแนนเฉลี่ยสะสมกว่า 1.80 ในภาคการศึกษาปกติ ถัดไปหลังจากได้รับการรอพินิจครั้งที่ 2 เงื่อนไขที่สี่ ใช้เวลาการศึกษาเกิน 8 ปีการศึกษาสำหรับการลงทะเบียนเรียนเต็มเวลา และเกิน 12 ปีการศึกษาสำหรับการลงทะเบียนเรียนไม่เต็มเวลา ในกรณีหลักสูตร 4 ปี ใช้เวลาการศึกษาเกิน 10 ปีการศึกษาสำหรับการลงทะเบียนเรียนเต็มเวลา และเกิน 15 ปีการศึกษาสำหรับการลงทะเบียนเรียนไม่เต็มเวลา ในกรณีหลักสูตร 5 ปี เงื่อนไขที่ห้า นักศึกษาลงทะเบียนครบตามที่หลักสูตรกำหนดแต่ยังได้รับคะแนนเฉลี่ยสะสม ต่ำกว่า 1.80
เงื่อนไขที่หก นักศึกษากระทำการทุจริต หรือมีความประพฤติอันเป็นการเสื่อมเสียแก่ มหาวิทยาลัย มหาวิทยาลัยเห็นควรให้ออกหรือไล่ออกตามข้อบังคับของมหาวิทยาลัยว่าด้วยวินัยนักศึกษา

การวิเคราะห์ข้อมูลโดยนำเทคนิคเหมืองข้อมูล ช่วยในเรื่องของการจัดการข้อมูล สามารถนำข้อมูลที่มีการบันทึกย้อนหลังหลายปี มาประกอบการตัดสินใจ เทคนิคเหมืองข้อมูล สามารถใช้ค้นข้อมูลสำคัญที่ปะปนกับข้อมูลอื่น ๆ ในฐานข้อมูลที่ใช้เทคนิคเหมืองข้อมูล บางครั้งเรียกว่าการค้นหาข้อมูลด้วยองค์ความรู้ KDD (Knowledge Discovery in Database)

ด้วยเหตุผลดังกล่าวผู้ศึกษาวิจัยจึงมีความสนใจที่จะนำฐานข้อมูลของกองบริการการศึกษา มหาวิทยาลัยราชภัฏยะลา ซึ่งประกอบไปด้วย กลุ่มข้อมูลนักศึกษา ข้อมูลการลงทะเบียนของนักศึกษาในแต่ละภาคการศึกษามาศึกษา โดยใช้กระบวนการทางเทคโนโลยีสารสนเทศ ที่เรียกว่าการทำเหมืองข้อมูล เพื่อนำมาหาความสัมพันธ์ของสาเหตุที่อาจจะมีผลกระทบต่อการพ้นสภาพของนักศึกษา เพื่อให้เกิดประโยชน์ต่อการพัฒนาด้านการวางแผนเพื่อแก้ปัญหาปรับลดจำนวน

นักศึกษาพ้นสภาพในปีต่อไป การทำเหมืองข้อมูล มีเทคนิคที่ใช้ในการวิเคราะห์หลายเทคนิค สำหรับงานวิจัยการวิเคราะห์ปัจจัยการพ้นสภาพการเป็นนักศึกษานี้จะใช้เทคนิคกฎความสัมพันธ์ (Association Rule) เพื่อค้นหาความสัมพันธ์ของข้อมูล สองชุดหรือมากกว่าสองชุดขึ้นไปไว้ด้วยกัน การหากฎความสัมพันธ์นั้นจะมีขั้นตอนวิธีการหาหลายวิธี แต่ขั้นตอนวิธีที่เป็นที่รู้จักและใช้อย่างแพร่หลายกฎการจำแนก(Classification Rules)ในการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ (Associative Classification) เป็นเทคนิคหนึ่งในเทคนิคในการจำแนกประเภทข้อมูล (Data Classification) ที่น่าสนใจ ซึ่งเป็นเทคนิคที่ให้ความแม่นยำในการทำนายสูง เนื่องจากมีการนำเทคนิคการหากฎความสัมพันธ์เข้ามาใช้ร่วมกับเทคนิคการจำแนกประเภทข้อมูล

วัตถุประสงค์

เพื่อวิเคราะห์ปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษา มหาวิทยาลัยราชภัฏยะลา

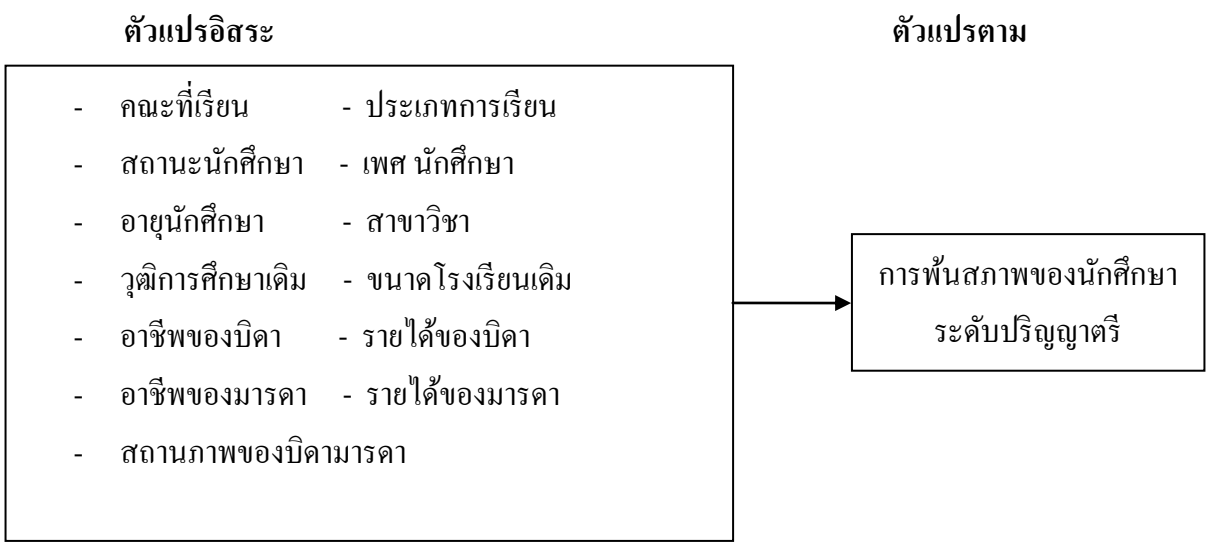
ขอบเขตการศึกษา

1. ใช้การจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ (Associative Rule) ของเทคนิคเหมืองข้อมูล (Data Mining) ในการวิเคราะห์ข้อมูลโดยการศึกษาผ่านโปรแกรม WEKA

2. ข้อมูลที่นำมาศึกษาคือข้อมูล นักศึกษาพ้นสภาพการเป็นนักศึกษา ระหว่างปีการศึกษา พ.ศ. 2550 -2553 จำนวน 1804 คน

การศึกษาวิจัยเรื่องวิเคราะห์ปัจจัยการพ้นสภาพของนักศึกษาระดับปริญญาตรี มีกรอบแนวคิดในการวิจัยตามความมุ่งหมายของการศึกษา

ตัวแปรอิสระ



ตัวแปรตาม จำนวน 1 ตัวแปร คือการพ้นสภาพของนักศึกษาระดับปริญญาตรี

ตัวแปรอิสระ จำนวน 13 ตัวแปร คือ คณะที่เรียน ประเภทการเรียน สถานะนักศึกษา เพศนักศึกษา อายุนักศึกษา สาขาวิชา วุฒิการศึกษาเดิม ขนาดโรงเรียนเดิม ทุนการศึกษา รายได้ของบิดา อาชีพของบิดา รายได้ของมารดา อาชีพของมารดา สถานภาพของบิดามารดา

กรอบแนวคิดของโครงการวิจัย

การศึกษาวิจัยเรื่องวิเคราะห์ปัจจัยการพ้นสภาพของนักศึกษาระดับปริญญาตรีมีกรอบแนวคิดในการวิจัยตามความมุ่งหมายของการศึกษา

สมมติฐานการวิจัย

ผลการวิเคราะห์การทำเหมืองข้อมูล สามารถสกัดปัจจัยที่มีผลกระทบของการพ้นสภาพของนักศึกษาได้

ประชากรและกลุ่มตัวอย่าง

ประชากร คือ นักศึกษามหาวิทยาลัยราชภัฏยะลา

กลุ่มตัวอย่าง คือ นักศึกษาอยู่ในเกณฑ์การพิจารณา

ประโยชน์ที่คาดว่าจะได้รับ

ขยายผลการใช้เทคนิคเหมืองข้อมูลผ่านโปรแกรม WEKA ในการวิเคราะห์ข้อมูลเพื่อหาสาเหตุการพ้นสภาพของนักศึกษา

นำสิ่งที่ได้จากเหมืองข้อมูลไปเป็นส่วนหนึ่งในการวางแผน ปรับลดจำนวนนักศึกษาที่พ้นสภาพการเป็นนักศึกษา

นิยามศัพท์เฉพาะ

เหมืองข้อมูล : การวิเคราะห์ข้อมูลจากฐานข้อมูลนักศึกษากองบริการการศึกษา มหาวิทยาลัยราชภัฏยะลา

การจำแนกประเภทข้อมูล : การคัดเลือกข้อมูลนักศึกษา เฉพาะสาขาที่นักศึกษาเลือกเข้าศึกษา กับหลักสูตรที่นักศึกษาได้ศึกษามาในระดับมัธยมศึกษาตอนปลายของนักศึกษาที่ศึกษา มหาวิทยาลัยราชภัฏยะลา

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การศึกษาทฤษฎีและผลงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคเหมืองข้อมูล ในบทนี้ผู้วิจัยได้ศึกษาเอกสาร แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง ประกอบด้วย

2.1. การทำเหมืองข้อมูล (Data Mining)

2.2. กระบวนการค้นหาความรู้ในฐานข้อมูล(Knowledge Discovery in Database KDD)

2.3. ขั้นตอนการทำเหมืองข้อมูล

2.4. ขั้นตอนวิธีที่ใช้ในการทำเหมืองข้อมูล

2.5. การวิเคราะห์ปัจจัย

2.6. เทคโนโลยีที่นำมาพัฒนางานวิจัย

2.7. งานวิจัยที่เกี่ยวข้อง

2.1. การทำเหมืองข้อมูล

เหมืองข้อมูล คือ การวิเคราะห์ข้อมูลที่ออกแบบมาเพื่อสนับสนุนการตัดสินใจ เป็นเครื่องมือที่ช่วยค้นหาและแยกแยะสาระที่เป็นประโยชน์ที่ถูกจัดเก็บในฐานข้อมูลขนาดใหญ่ออกมา เป็นการเพิ่มคุณค่าให้กับฐานข้อมูลที่มีอยู่ ลักษณะของสาระจะต้องมีความชัดเจน มีรูปแบบ มีเหตุผลใหม่ ใช้ประโยชน์ได้ และเข้าใจได้ Data Mining เป็นขั้นตอนหนึ่งในการทำกระบวนการค้นหาความรู้

ซึ่ง เหมืองข้อมูลยังเป็นกระบวนการของการกลั่นกรองสารสนเทศ ที่ซ่อนอยู่ในฐานข้อมูลใหญ่ จัดได้ว่าเป็นเทคโนโลยีใหม่ที่เข้ามาช่วยให้การวิเคราะห์ข้อมูลทำได้โดยอัตโนมัติและมีประสิทธิภาพสูง จึงได้รับความสนใจนำไปใช้อย่างแพร่หลายในทุกวงการ เพื่อทำนายแนวโน้มและพฤติกรรม โดยอาศัยข้อมูลในอดีต และเพื่อใช้สารสนเทศเหล่านี้ในการสนับสนุนการตัดสินใจทางธุรกิจ และการประมวลผลข้อมูลเกี่ยวกับลักษณะการใช้ ช่วงระยะเวลา และ เชื่อมโยงข้อมูล ที่มีผู้เข้าใช้บริการจากแหล่งข้อมูลต่าง ๆ เช่น ผู้ใช้บริการเว็บ เพื่อนำมาพิจารณาปรับปรุง การให้บริการ โดยอาจจะเพิ่มหรือลดการให้บริการบางชนิดให้เหมาะสมกับกลุ่มผู้ใช้ แต่ละสภาพแวดล้อม ซึ่งอาจจะมีความสนใจแตกต่างกันไป ซึ่งอาจจะมีหนึ่งเทคนิค หรือมากกว่ามาประมวลผลเพื่อดึงความรู้หรือสิ่งที่ต้องการจากฐานข้อมูลที่ได้ผ่านขั้นตอนการจัดเตรียมข้อมูล (Data Cleaning) เพื่อนำข้อมูลเข้าสู่กระบวนการประมวลผลโดยใช้เทคนิคในการทำงานเหมืองข้อมูลต่อไป

ประเภทข้อฐานข้อมูลที่สามรถทำ Data Mining

1. ฐานข้อมูลเชิงสัมพันธ์ (Relation Database) จะมีการจัดเก็บข้อมูลในลักษณะที่เป็นกลุ่มของข้อมูลที่มีความสัมพันธ์กัน ในฐานข้อมูลหนึ่งๆ สามารถที่จะมีตารางตั้งแต่ 1 ตารางเป็นต้นไป และในแต่ละตารางนั้นก็สามารถมีได้หลายคอลัมน์ (Column) หลายแถว (Row)

2. คลังข้อมูล (Data Warehouse) เป็นการเก็บรวบรวมข้อมูลจากหลายแหล่งมาเก็บไว้ในรูปแบบเดียวกันและรวบรวมไว้ในที่เดียวกัน

3. ฐานข้อมูลการเปลี่ยนแปลงรายการ (Transactional Database) เกิดจาก แฟ้มหลัก ที่มีการจัดเก็บข้อมูลที่ทำไว้ก่อนรูปแบบการเก็บแบบถาวร และมีข้อมูลเก็บไว้อย่างสมบูรณ์ และเมื่อมีการแก้ไขหรือปรับปรุงแฟ้มข้อมูลหลักนี้ ระบบจะใช้วิธีสร้างแฟ้มข้อมูลขึ้นมาใหม่ซึ่งจะเปลี่ยนแปลงข้อมูลเฉพาะบางรายการ เรียกว่า "ฐานข้อมูลการเปลี่ยนแปลงรายการ " (Transaction Database) แล้วเขียนโปรแกรมให้คอมพิวเตอร์อ่านแฟ้มข้อมูลใหม่นี้ เข้าไปจัดการแก้ไขปรับปรุงข้อมูลในแฟ้มหลักให้ ดู Transaction Database เปรียบเทียบ

4. ฐานข้อมูลขั้นสูง (Advanced Database) เป็นฐานข้อมูลที่จัดเก็บข้อมูลในรูปแบบอื่น ๆ เช่นข้อมูลแบบ Object-oriented ข้อมูลที่เป็น text file ข้อมูลมัลติมีเดีย ข้อมูลในรูปของ web งานของเหมืองข้อมูล(Task of data mining)

ในการวิเคราะห์เหมืองข้อมูลจะเลือกใช้เทคนิคการทำงานอย่างใดอย่างหนึ่งไม่ได้ เพราะในแต่ละขั้นตอนวิธี ก็จะมีเทคนิคของเหมืองข้อมูลที่แตกต่างกันไปขึ้นอยู่กับงานวิจัยและวัตถุประสงค์ของงาน รูปแบบของข้อมูลที่ผู้วิจัยมีอยู่ในฐานข้อมูล ซึ่งจะสามารถจัดรูปแบบของขั้นตอนวิธีของการทำเหมืองข้อมูลทั้งหมด 6 ขั้นตอนวิธีได้ ดังนี้

1. การจัดหมวดหมู่ (Classification) การจัดหมวดหมู่ของเหมืองข้อมูลเพราะการทำความเข้าใจและการติดต่อสื่อสารต่าง ๆ ก็เกี่ยวข้องกับการแบ่งเป็นหมวดหมู่การจัดแยกประเภทและการแบ่งแยกชนิดโดยการจัดหมวดหมู่ประกอบด้วยการสำรวจจุดเด่นของวัตถุที่ปรากฏออกมา และทำการกำหนด จุดเด่นนั้น ๆ เป็นตัวที่ใช้แบ่งหมวดหมู่งานในการแบ่งหมวดหมู่ และให้ข้อมูลเรียนรู้รูปแบบข้อมูล (Training Set) วิธีการนี้เป็นกระบวนการหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence) ของตัวอย่างในแต่ละหมวดหมู่ ซึ่งมีภาระหน้าที่ในการสร้างโมเดลของบางชนิดที่ไม่สามารถจะจัดหมวดหมู่ของข้อมูลได้ ให้สามารถจัดเป็น หมวดหมู่ได้ ตัวอย่างของการจัดหมวดหมู่ เช่น การจัดหมวดหมู่ของผู้ยื่นขอเครดิต (Credits) เป็นระดับต่ำระดับกลาง และระดับสูง ของความเสี่ยงที่จะได้รับ เป็นต้น

2. การประเมินค่า (Estimation) การประเมินค่าทางธุรกิจอย่างต่อเนื่องจะก่อให้เกิดผลลัพธ์

ที่มีประโยชน์กับธุรกิจ การป้อนข้อมูล ที่เรามีอยู่เข้าไป เพื่อใช้ในการประเมินสิ่งต่าง ๆ ที่จะก่อให้เกิดประโยชน์ หรือสำหรับตัวแปรที่เราไม่รู้ค่าแน่นอน เช่น รายได้จากการค้าขาย จุดสูงสุดทางธุรกิจ หรือดุลยภาพของบัตรเครดิต ในทางปฏิบัติการประเมินค่าจะถูกใช้ในการทำงานการจัดหมวดหมู่ ตัวอย่างของการประเมินค่า เช่น การประเมินรายได้รวมของครอบครัว หรือการประเมินจำนวนบุตรในครอบครัว

3. การทำนายล่วงหน้า (Prediction) การทำนายล่วงหน้าก็เป็นงานที่มีลักษณะคล้ายกับการจัดหมวดหมู่หรือการประเมินค่า จะใช้สถิติการบันทึกของการจัดหมวดหมู่ในการทำนายอนาคตของพฤติกรรมข้อมูลหรือการประเมินค่าที่จะเกิดขึ้นในอนาคต ตัวอย่างของงานการทำนายล่วงหน้า เช่น การทำนายการเปลี่ยนแปลงพฤติกรรมของตลาด การทำนายจำนวนลูกค้าที่จะออกจากธุรกิจของเราใน 6 เดือนข้างหน้า หรือ การทำนายอนาคตว่าจะเกิดโรคระบาดอะไรขึ้นและกระทรวงสาธารณสุขจะมีแผนการแก้ไขปัญหาอย่างไร เป็นต้น

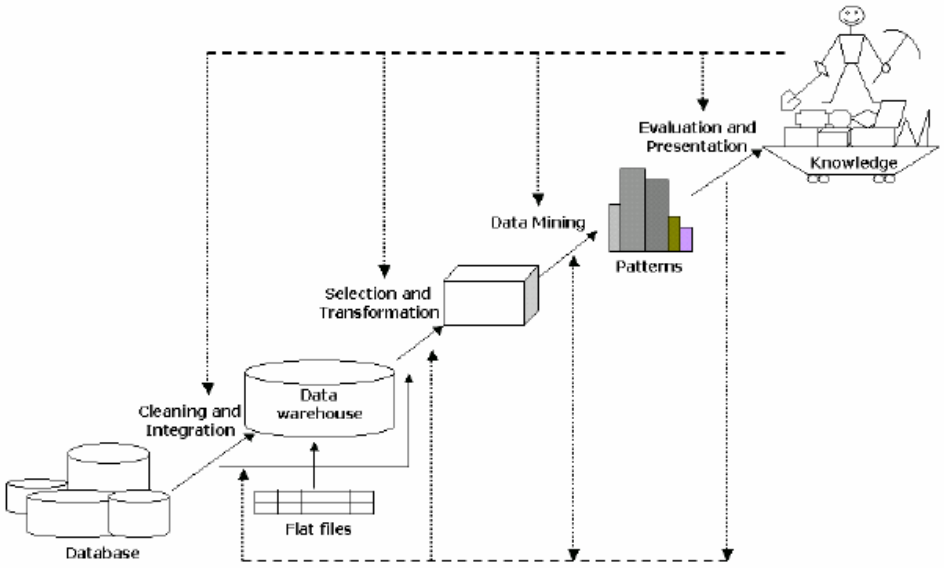
4. การจัดกลุ่มโดยอาศัยความใกล้ชิดกัน หรือการวิเคราะห์ของตลาด (Affinity Group) งานในการจัดกลุ่มหรือการวิเคราะห์ตลาด คือการตัดสินใจรวมสิ่งที่สามารถไปด้วยกันเข้าไว้ในกลุ่มเดียวกันตัวอย่างของการจัดกลุ่มโดยอาศัยความใกล้ชิดกันหรือการวิเคราะห์ตลาด เช่น การตัดสินใจว่าสิ่งใดบ้างที่จะไปอยู่ด้วยกันอย่างสม่ำเสมอในรถเข็นในซุปเปอร์มาร์เกต

5. การรวมตัว (Clustering) การรวมตัวคืองานที่ทำการรวมส่วนต่าง ๆ ในแต่ละส่วนที่ต่างชนิดกันให้อยู่รวมกันเป็นกลุ่มย่อย หรือคลัสเตอร์ (Clusters) โดยในแต่ละคลัสเตอร์อาจจะประกอบด้วยส่วนต่าง ๆ ที่ต่างชนิดกัน ซึ่งความแตกต่างของการรวมตัวจากการจัดหมวดหมู่ คือ การรวมตัวจะไม่พึ่งพาอาศัยการกำหนดหมวดหมู่ล่วงหน้า และไม่ใช้ตัวอย่าง ข้อมูลจะรวมตัวกันบนพื้นฐานของความคล้ายในตัวเอง

6. การบรรยาย (Description) ในบางครั้งวัตถุประสงค์ของเหมืองข้อมูลคือต้องการอธิบายความสับสนของฐานข้อมูลในทางที่จะเพิ่มความเข้าใจในส่วนของประชากร ผลิตภัณฑ์ หรือขบวนการให้มากขึ้นเทคนิคเหมืองข้อมูลส่วนใหญ่ต้องการ ข้อมูลเรียนรู้รูปแบบข้อมูล (Training Set) ข้อมูลจำนวนมากที่ประกอบด้วยการวิเคราะห์

2.2. กระบวนการค้นหาองค์ความรู้ในฐานข้อมูล (Knowledge Discovery in Database : KDD)

กระบวนการค้นหาองค์ความรู้ในฐานข้อมูล หมายถึง กระบวนการในการค้นหาลักษณะแฝงของข้อมูลที่อยู่ในกลุ่มข้อมูลจำนวนมาก ซึ่งมีขั้นตอนการทำเหมืองข้อมูล(Data Mining) เป็นกระบวนการที่สำคัญในการค้นหาลักษณะที่น่าสนใจของข้อมูลเหล่านี้ เช่น รูปแบบ ความสัมพันธ์ การเปลี่ยนแปลง โครงสร้างที่เด่นชัด หรือลักษณะที่ผิดปกติของข้อมูลจากข้อมูลจำนวนมาก ๆ ที่เก็บอยู่ในฐานข้อมูลแสดงดังภาพที่ 2.1



ภาพที่ 2.1 กระบวนการค้นหาองค์ความรู้ในฐานข้อมูล (Knowledge Discovery in Database: KDD**)**

จากที่ได้กล่าวแล้วว่าการทำเหมืองข้อมูล (Data Mining) เป็นขั้นตอนหนึ่งที่สำคัญในกระบวนการค้นหาลักษณะแฝงของข้อมูล ที่มีประโยชน์ในกระบวนการค้นหาองค์ความรู้ในฐานข้อมูล (Knowledge Discovery in Database : KDD) ซึ่งโดยทั่วไปกระบวนการจะประกอบด้วยขั้นตอนต่าง ๆ ดังนี้

1. การคัดเลือกข้อมูล (Data Selection) เป็นการระบุถึงแหล่งข้อมูลที่จะนำมาใช้ในการทำเหมืองข้อมูล รวมถึง การนำข้อมูลที่ต้องการ ออกมาจากฐานข้อมูลเพื่อทำการพิจารณาในเบื้องต้นต่อไป

2. การกรองข้อมูล (Data Cleaning) เป็นกระบวนการที่ทำให้เกิดความมั่นใจในคุณภาพของข้อมูลที่จะนำมาใช้วิเคราะห์ว่าข้อมูลที่มีนั้นมีความสมบูนณ์และถูกต้องหรือไม่ หากข้อมูลที่มีอยู่ยังไม่สมบรูณ์ถูกต้องจะต้องนำข้อมูลที่ไม่ถูกต้องออกจกาฐานข้อมูล

3. การแปลงรูปแบบข้อมูล (Data Transformation) เป็นการแปลงข้อมูลที่เลือกมาให้อยู่ในรูปแบบที่เหมาะสม สำหรับการนำไปใช้วิเคราะห์ตามขั้นตอนวิธี (Algorithm) และแบบจำลองที่ใช้ในการทำเหมืองข้อมูลต่อไป

4. การวิเคราะห์เหมืองข้อมูล มีรูปแบบของงานตามลักษณะของแบบจำลองที่ใช้ในการทำเหมืองข้อมูล (Data Mining) นั้นสามารถแบ่งกลุ่มได้เป็น 2 ประเภทใหญ่ๆ คือ

4.1 Predictive Data Mining เป็นการคาดคะเนลักษณะหรือประมาณค่าที่ชัดเจนของข้อมูลที่จะเกิดขึ้น โดยใช้พื้นฐานจากข้อมูลที่ผ่านมาในอดีต

4.2 Descriptive Data Mining เป็นการหาแบบจำลองเพื่ออธิบายลักษณะบางอย่างของข้อมูลที่มีอยู่ ซึ่งโดยส่วนมากจะเป็นลักษณะการแบ่งกลุ่มให้กับข้อมูล

5. การวิเคราะห์และประเมินผลลัพธ์ที่ได้ (Result Analysis and Evaluation) เป็นขั้นตอนการแปลความหมาย และการประเมินผลลัพธ์ที่ได้ว่ามีความเหมาะสมหรือตรงกับวัตถุประสงค์ที่ต้องการหรือไม่ โดยทั่วไปควรมีการแสดงผลในรูปแบบที่สามารถเข้าใจได้โดยง่าย

2.3 ขั้นตอนการทำเหมืองข้อมูล (Data Mining)

2.3.1. ทำความเข้าใจปัญหา **(Problem formulation)** การกำหนดวัตถุประสงค์ของงานวิจัย หรือ ทางธุรกิจ คือจะต้องเข้าใจปัญหาและความต้องการทำ การกำหนดวัตถุประสงค์ จะเป็นส่วนที่กำหนดว่าเมื่อไหร่ที่จะใช้ Data Mining ในการแก้ปัญหาซึ่งในส่วนนี้จะประกอบด้วยการวิเคราะห์การวิเคราะห์เบื้องต้นว่าเรามีข้อมูลใดอยู่บ้าง และต้องการอะไรจากข้อมูลซึ่งขั้นตอนนี้จะสามารถมองถึง ขั้นตอนวิธี และฐานข้อมูลที่สัมพันธ์กับวัตถุประสงค์ทางธุรกิจได้

การใช้งาน Data Mining ให้ได้ประโยชน์สูงสุดจำเป็นต้องมีการกำหนดวัตถุประสงค์ที่ชัดเจน เช่น ต้องการ เพิ่มยอดการตอบรับการขายทางจดหมาย ขึ้นอยู่กับการระบุเป้าหมายว่า จะเพิ่มอัตราการตอบรับหรือเพิ่มมูลค่าการตอบรับซึ่ง จำเป็นที่จะต้องสร้าง Model ที่แตกต่างกัน วัตถุประสงค์ที่กำหนดขึ้นมาจะต้องมีการระบุวิธีการในการวัดผลลัพธ์ที่ได้จาก โครงการ รวมถึงต้นทุนที่สมเหตุสมผลด้วย

2.3.2. การเตรียมข้อมูล (Data Preparation)เป็นหัวใจของขั้นตอนในการทำทั้งหมด เป็นช่วงที่ใช้เวลามากที่สุดในขั้นตอน โดยปกติแล้วต้องการเวลาประมาณ 60% ของเวลาทั้งหมดในการเตรียมข้อมูล ในขั้นตอนนี้อาจสามารถแบ่ง ออกได้เป็นขั้นตอน 5 ขั้นตอน ย่อยดังต่อไปนี้

2.3.2.1 การเลือกข้อมูล (Data Selection) จุดประสงค์ คือการระบุแหล่งของข้อมูลที่มี และทำการดึงเอาข้อมูลออกมาใช้สำหรับการวิเคราะห์เบื้องต้นในการ เตรียมตัวสำหรับการที่จะทำการวิเคราะห์เหมืองข้อมูล ในขั้นต่อ ๆ ไป การเลือกข้อมูลนั้นจะแตกต่างไปตามวัตถุประสงค์ของแต่ละงานออกไป ที่ได้กำหนดไว้ตั้งแต่ต้น และการเลือกข้อมูลก็ยังถูกกำหนดโดยลักษณะงานที่จะถูกนำมาใช้อีกด้วย ตัวแปรที่ถูกเลือกมาแต่ละตัวนั้นจะต้องถูกทำความเข้าใจว่าตัวแปรแต่ละตัวหมายความว่าอะไร ประกอบด้วยอะไร ไม่เพียงแต่คำจำกัดความทางการวิจัยเท่านั้น แต่จะต้องมีคำอธิบายอย่างชัดเจนเกี่ยวกับชนิดของข้อมูล ค่าที่เป็นไปได้ แหล่งกำเนิดของข้อมูล รูปแบบของข้อมูล และลักษณะอื่น ๆ จะมีตัวแปร 2 ชนิด คือ

ก. ตัวแปรแบบเชิงคุณภาพ (Qualitative Variable)

ตัวแปรที่กำหนด (Nominal Variable) กล่าวถึงชนิดนี้ของข้อมูล ที่มันอ้างถึง

แต่ไม่มีลำดับ ในค่าที่เป็นไปได้ (Possible Value) ตัวอย่างเช่น สถานภาพ แต่งงาน (โสด แต่งงาน หย่า หม้าย ไม่ทราบ) เพศ (ชาย หญิง) ระดับการศึกษา (ปริญญาโท ปริญญาตรี ม. ปลาย ปวช)

ตัวแปรลำดับ (Ordinal Variable) มีลำดับสำหรับค่าที่เป็นไปได้ ตัวอย่าง เช่น ลำดับของ ลูกค้า (ดี ปานกลาง ไม่ดี)

ข. ตัวแปรแบบเชิงปริมาณ (Quantitative Variable) ซึ่งมีการวัดความแตกต่าง ระหว่างค่าที่เป็นไปได้ ค่าที่ต่อเนื่อง(Continuous) เช่นรายได้ เฉลี่จำนวนครั้งที่ซื้อ รายได้ ค่าเป็นจำนวนเต็ม (Discrete) เช่นจำนวนพนักงาน เวลาปี เดือน ฤดู ไตรมาส

ตัวแปรของข้อมูลมีหลายตัว แต่ตัวแปรที่ถูกเลือกสำหรับทำเหมืองข้อมูล (Data Mining) นั้นถูกเรียกว่า ตัวแปรใช้งานล่าสุด (Active Variable) เพราะว่ามันจะถูกใช้สร้างความแตกต่างของกลุ่มย่อยต่าง ๆ และสามารถถูกนำมาทำนายผลได้ เมื่อผู้วิจัยทำการเลือกข้อมูลจะต้องพิจารณาอายุของข้อมูลด้วย เพราะว่าสถานการณ์ภายนอกเปลี่ยนแปลงตลอดเวลาซึ่งจะทำให้ประสิทธิภาพของการวิเคราะห์เหมืองข้อมูลลดลง

2.3.2.2 การกลั่นกรองข้อมูล (Screening Information**)** จุดประสงค์ก็เพื่อทำให้มั่นใจว่าคุณภาพของข้อมูลที่ถูกเลือกนั้นเหมาะสม ข้อมูลที่สมบูรณ์เป็นเครื่องประกัน ว่าการทำเหมืองข้อมูล (Data Mining) จะสำเร็จ ในขั้นตอนนี้เป็นขั้นตอนที่มีปัญหามากกว่า ในขั้นตอนของการเตรียมข้อมูล เพราะข้อมูลส่วนใหญ่ที่มีในองค์กร ไม่ได้ถูกเตรียมมาเพื่อทำเหมืองข้อมูล (Data Mining) โดยเฉพาะ ข้อมูลจะถูกนำมาจากแหล่งต่าง ๆ ถูกจัดเก็บไม่ดี ข้อมูลที่ถูกนำมาจาก ภายนอกแล้วนำมาเพื่อให้เข้ากับข้อมูลภายในที่มีอยู่ ปัญหาหลักของข้อมูล คือ คุณภาพและ ความสมบูรณ์ของข้อมูล ในขั้นตอนนี้ก่อนอื่นจะต้องทำการทบทวนโครงสร้างของข้อมูลใหม่ และวัดคุณภาพของข้อมูล โดยวิธีทางสถิติ หรือสุ่มตัวอย่าง เครื่องมือที่ใช้ในการทำการกลั่นกรองข้อมูลมีดังต่อไปนี้

ระหว่างการทำขั้นตอนการกลั่นกรองข้อมูลจะมีปัญหาบ่อย ๆ ที่มักพบได้ 1. Noisy Data คือตัวแปรตัวหนึ่งหรือมากกว่ามีค่าซึ่งเกินกว่าค่าที่เราคาดไว้ ซึ่งอาจจะหมายถึงแง่ดีหรือแง่ร้ายก็ได้ ในแง่ดีก็คือข้อมูลจะแสดงอย่างชัดเจน ในแง่ร้าย คือข้อมูลอาจจะเป็นข้อมูลที่ไม่สมบูรณ์ สาเหตุ ที่เกิดขึ้นได้อาจจะมาจากความเลินเล่อของมนุษย์ ตัวอย่างเช่น Operator ใส่อายุให้คนเป็น 300 ปี หรือใส่ค่าของรายได้ เป็นติดลบ ค่าเหล่านี้ควรจะถูกแก้ไข หรือเอาออกจากการวิเคราะห์ ควรมีขั้นตอนการตรวจข้อมูลก่อนนำมาใช้ 2. ค่าที่หายไป (Missing Value) คือค่าที่ไม่ได้แสดงในข้อมูลที่เราได้เลือกแล้ว หรือค่าที่ไม่สมบูรณ์ที่เราลบออกไป ระหว่างการทำตรวจข้อมูลผิดพลาด (Noise Detection) ค่าอาจจะหายไปเพราะเกิดจากความเลินเล่อของมนุษย์ เพราะว่าไม่มีข้อมูลนั้นระหว่างการทำนำเข้าข้อมูล การจัดการกับค่าที่หายไป นั้นสามารถจัดการได้ด้วยเทคนิคที่ต่าง ๆ กัน

2.3.2.3 การสำรวจและตรวจสอบข้อมูล (Data Monitoring) เมื่อทำการเก็บข้อมูลเรียบร้อยแล้ว ขั้นตอนต่อไปที่ควรกระทำก็ คือการตรวจสอบข้อมูล เหตุที่ต้องทำการตรวจสอบข้อมูลมี 2 ข้อ 1. นักวิเคราะห์ควรมีความคุ้นเคยกับตัวข้อมูล ไม่ใช่รู้แต่ชื่อ ลักษณะและความหมายของมันเท่านั้น แต่ต้องรู้ถึงเนื้อหา หรือความมุ่งหมายที่แท้จริงของข้อมูลด้วย 2. อาจมีความผิดพลาดของการเก็บสะสมข้อมูล เกิดขึ้นในขณะที่ทำการรวบรวมข้อมูลจากฐานข้อมูลหลาย ๆ แหล่งเข้ามาเป็นหนึ่งเดียวเพื่อใช้ในการวิเคราะห์ ซึ่งนักวิเคราะห์ ที่ดีจะต้องทำการตรวจสอบข้อมูลเหล่านี้ให้ถูกต้อง ตัวอย่างของความผิดพลาดที่เกิดขึ้น ได้แก่ ความผิดพลาดในการเก็บข้อมูล จากลักษณะ ที่ไม่ต้องการ ซึ่งเกิดจากความสับสนในการตั้งชื่อ ลักษณะ (Attribute) นั้น เราต้องการเก็บค่าของระดับการศึกษาของผู้สมัครเข้าศึกษาต่อ ซึ่งในความเป็นจริงถูกเก็บไว้ในลักษณะ (Attribute) ที่ชื่อ “LEVEL_EDU” แต่ในฐานข้อมูลนั้นบังเอิญมีลักษณะ (Attribute) อีกตัวหนึ่งชื่อ “EDUCATION” ซึ่งเก็บระดับการศึกษาที่ผู้สมัครต้องการเข้าศึกษา ซึ่งถ้าเราไม่ได้ตรวจสอบความสัมพันธ์และความมุ่งหมายที่แท้จริงของแต่ละลักษณะ (Attribute) แล้ว ก็อาจเกิดการสับสน โดยเก็บข้อมูลของลักษณะ (Attribute) “EDUCATION” ไปแทนก็ได้ ซึ่งเมื่อนำข้อมูลที่ได้ไปทำเหมืองข้อมูล(Data Mining) ผลลัพธ์ที่ได้ ก็จะผิดพลาดด้วย

2.3.2.4 การแปลงข้อมูล (Data Transformation) ระหว่างขั้นตอนของการแปลงข้อมูลที่ได้กลั่นกรองแล้วจะถูกแปลงให้เป็นรูปแบบของข้อมูลที่พร้อมจะถูก วิเคราะห์ รูปแบบของข้อมูลที่พร้อมจะถูกวิเคราะห์ คือรูปแบบของข้อมูลที่ไม่มีความขัดแย้ง ถูกจัดระเบียบมาอย่างเรียบร้อย กลั่นกรองมาจากแหล่งข้อมูลภายนอกและภายใน ขั้นตอนนี้เป็นขั้นตอนที่สำคัญมากเนื่องจากความถูกต้อง และสมบูรณ์ของผลลัพธ์สุดท้ายซึ่งขึ้นอยู่กับว่า นักวิจัย จะนำข้อมูลนั้นตัดสินใจ กำหนดโครงสร้างและเสนอลักษณะของนำเข้าข้อมูลอย่างไร

2.3.2.5 การปรับแต่งข้อมูล (Data Engineering) ขั้นตอนก่อนหน้านี้เป็นขั้นตอนของการสร้าง และการตรวจสอบความถูกต้องของข้อมูลที่จะนำมาใช้ แต่ในขั้นตอนนี้ที่เราต้องทำ คือการปรับแต่งฐานข้อมูล ซึ่งในขั้นตอนนี้จะมีปัญหาหลัก ๆ อยู่ 3 ข้อ คือ 1. ฐานข้อมูลที่ได้อาจมีลักษณะ(Attributes) จำนวนมากที่สามารถใช้ประโยชน์ได้แต่ถูกละเลย การเลือกกลุ่มของ attributes ที่จะใช้เป็นปัญหาที่สำคัญปัญหาหนึ่ง 2. ฐานข้อมูลที่ได้อาจมีจำนวนระเบียน (record) มากเกินไปกว่าที่จะสามารถทำการวิเคราะห์ให้เสร็จลงได้ในเวลาที่เหมาะสม ซึ่งในกรณีนี้เราต้องทำการสุ่มข้อมูลตัวอย่างขึ้นมาใช้แทน 3. ข้อมูลบางอย่างอาจใช้ให้เกิดประโยชน์ได้ โดยการนำเสนอในรูปแบบของการวิเคราะห์แบบเฉพาะเจาะจง การทำเหมืองข้อมูลนั้นจะมีการทำซ้ำขึ้นมาหลาย ๆ ครั้ง เพื่อทดสอบ การใช้ลักษณะ (Attribute) ที่แตกต่าง ขนาดของกลุ่มตัวอย่างที่ต่างกัน เช่น เราจะ

ทำนายอนาคตเมื่อเวลาผ่านไป 1 , 2 , 3 , หรือ 4 เดือน เราอาจทำนายได้โดยใช้เพียงลักษณะ (Attribute) เป็นตัวทำนายหรืออาจใช้ข้อมูลทุกอย่างที่เรามีเป็นตัวทำนายก็ได้ เป็นต้น

2.3.3. การนำเสนอข้อมูล (Visualization)

เป็นการนำเสนอข้อมูลในรูปแบบกราฟิก การนำเสนอจะสามารถทำได้มากกว่า 2 มิติ ซึ่งจะสร้างความละเอียดของการนำเสนอ และสร้างความเข้าใจให้มากขึ้น

2.3.4. การวิเคราะห์ (Analysis)

หลังจากเลือกขั้นตอนวิธี (Algorithm) ที่เหมาะสมกับลักษณะของปัญหาแล้ว เราก็จะนำ algorithm นั้นมาทำการวิเคราะห์ ข้อมูลในฐานข้อมูลที่เตรียมไว้ ซึ่งในบางครั้งขั้นตอนนี้จะถูกเรียกว่าเหมืองข้อมูล (Data Mining) ในขณะที่จะเรียกกระบวนการทั้งหมดว่า กระบวนการค้นหาองค์ความรู้ในฐานข้อมูล (Knowledge Discovery in Database: KDD)ผลลัพธ์ที่ได้จากขั้นตอนนี้จะเป็นรูปแบบของความสัมพันธ์ของ ข้อมูลที่จะนำมาใช้ ในการพยากรณ์ (Prediction) หรือวิเคราะห์ต่อไป นำข้อมูลที่จัดเตรียมไว้มาทำเหมืองข้อมูล (Data Mining) ซึ่งมีการทำงานอยู่ 4 ชนิดด้วยกัน คือ

การจัดแบ่งข้อมูล (Data Segmentation) เป็นกระบวนการแบ่งข้อมูลภายในฐานข้อมูลออกเป็นกลุ่มเพื่อให้ง่ายต่อการวิเคราะห์ เช่น การแบ่งลูกค้าออก ตามอายุ เพศ รายได้ เป็นต้น

การสร้างแบบจำลองพยากรณ์ (Predictive Modeling) แบ่งเป็น 2 ลักษณะ คือ การจัดหมวดหมู่ (Classification) เป็นการจัดกลุ่มให้กับแต่ละข้อมูลในฐานข้อมูล โดยมีการระบุค่า หรือลักษณะที่เป็นไปได้ของข้อมูล ภายในแต่ละกลุ่ม เช่น การจัดกลุ่มของผู้ป่วยตามผลของการใช้ยารักษา เพื่อระบุรูปแบบการรักษาให้กับผู้ป่วยใหม่ ที่เข้ารับการรักษา เป็นต้น

การพยากรณ์ค่าที่เป็นไปได้ (Value Prediction) การกระจายของค่าที่เป็นไปได้ของตัวแปรใด ๆ ในกลุ่มข้อมูล การทำนายค่าที่เป็นตัวเลข เช่น การทำนายภาษีที่จะเก็บได้ในปี

การหาความสัมพันธ์ของข้อมูล (Associations) ภายในกลุ่มข้อมูล เพื่อใช้ลักษณะของข้อมูลหนึ่ง ๆในการบอกลักษณะที่จะเกิดขึ้นกับข้อมูลอีกตัวหนึ่ง ซึ่งอาจจะเป็นการหาความสัมพันธ์ของข้อมูลในกลุ่มเดียวกัน เช่น การระบุว่าในกลุ่มของลูกค้าที่ซื้อนมนั้น จะมีลูกค้า 64% ที่ซื้อขนมปังด้วย หรืออาจจะเป็นการหาความสัมพันธ์ของ ตัวแปรระหว่างกลุ่มข้อมูลก็ได้ ตัวอย่าง เช่น ในทุก ๆ ครั้งที่ดัชนีของตลาดหุ้นหนึ่งลดลง 5% ดัชนีของตลาดหุ้นอื่นจะเพิ่มขึ้น 13% ภายในช่วง 2-6 เดือนหลังจากนั้น เป็นต้น ซึ่งลักษณะของการหาความสัมพันธ์นั้นอาจแบ่งได้เป็น 3 กลุ่ม คือ การหาความสัมพันธ์ระหว่างข้อมูล (Association discovery) การหาความสัมพันธ์ในลักษณะที่เป็นลำดับของข้อมูล (Sequential Pattern discovery) และ การหาความสัมพันธ์ของข้อมูลกับช่วงเวลาใด ๆ (Similar Time Sequence discovery)

การแสดงลักษณะของข้อมูลที่ผิดปกติ (Deviation Detection) ข้อมูลผิดไปจากที่คาดไว้ โดยมีการแสดงผลอยู่ในลักษณะที่สามารถทำความเข้าใจและแปลความหมายได้ง่าย เช่น การใช้กราฟ เป็นต้น

2.3.5 Interpreted หลังจากที่การสร้าง Model แล้วจำเป็นต้องตรวจสอบผลลัพธ์และตีความหมาย ความถูกต้องที่ตรวจออกมาได้นั้น เป็นชุดตัวอย่างที่ส่งเข้าไป ดังนั้นผลลัพธ์ที่ได้อาจมีความปรวนแปรหากมีการนำไปใช้กับข้อมูลอื่น ๆ

2.3.6 Presentation เป็นการแสดงผลการวิเคราะห์โดยอาศัยเครื่องมือที่มีความสามารถและเข้าใจง่าย การแสดงผลอาจจะอยู่ในรูปแบบของ รายงาน ตาราง กราฟ แผนที่หลายมิติ เป็นต้น

2.4 ขั้นตอนวิธีที่ใช้การทำเหมืองข้อมูล (Data Mining)

เทคนิคของเหมืองข้อมูลการแก้ปัญหาของงานชนิดต่างๆ โดยใช้วิธีเหมืองข้อมูลในแต่ละงานก็จะมีเทคนิคของเหมืองข้อมูลที่จะนำมาใช้ได้อย่างเหมาะสม โดยเทคนิคของเหมืองข้อมูลนั้นมีมากมาย ส่วนใหญ่มาจากศาสตร์ทางปัญญาประดิษฐ์ AI (Artificial Intelligence) หรือศาสตร์อื่นๆ ซึ่งจะขอยกตัวอย่างของเทคนิคที่ถูกใช้กันค่อนข้างแพร่หลาย

2.4.1 ดีซีชันทรี (Decision Tree) เป็นแบบจำลองที่มีลักษณะคล้ายกับต้นไม้ จะมีการสร้างกฎต่างๆ ขึ้นเพื่อใช้ในการตัดสินใจ ดีซีชันทรีเป็นวิธีที่ได้รับความนิยม เนื่องจากความไม่ซับซ้อนของขั้นตอนวิธี ทำให้เครื่องมือที่ใช้ในการทำที่วางขายกันอยู่ในท้องตลาด ต่างก็ใช้วิธีนี้ข้อดีของวิธีนี้คือสามารถตีความและเข้าใจลักษณะของรูปแบบข้อมูล (Pattern) ได้ง่าย เพราะ มีการแยกออกเป็นกฎหรือข้อกำหนดต่าง ๆ แต่ก็ยังคงมีปัญหาในเรื่องของการให้น้ำหนักความน่าเชื่อถือหรือการให้ค่าน้ำหนักในแต่ละโหนด (Node) ซึ่งถ้าให้น้ำหนักผิดไป อาจจะทำให้การตีความผิดไปได้

2.4.2 คลัสเตอร์ลิ่ง (Clustering) วิธีคลัสเตอร์ลิ่งนี้เป็นวิธีที่อาจจะเรียกว่าเป็นการทำเหมืองข้อมูลแบบอ้อม ๆ เนื่องจากการหาผลลัพธ์ในแต่ละครั้งนั้นผู้วิเคราะห์ยังไม่อาจจะทราบว่าสิ่งที่ต้องการจะหานั้นคืออะไร จำเป็นต้องรอจนกว่าการค้นหาจะทำเสร็จสมบูรณ์จึงจะทราบข้อมูลที่ซ่อนอยู่ เปรียบเสมือนกับการมีข้อมูลจำนวนมาก อยู่ในตะกร้า แล้วจากนั้นก็จัดเรียงข้อมูลหน่วยนั้นให้อยู่เป็นกลุ่มก้อนซึ่งทำให้สังเกตลักษณะเด่นที่ซ่อนเร้นอยู่ภายในข้อมูลจำนวนมากหน่วยนั้น

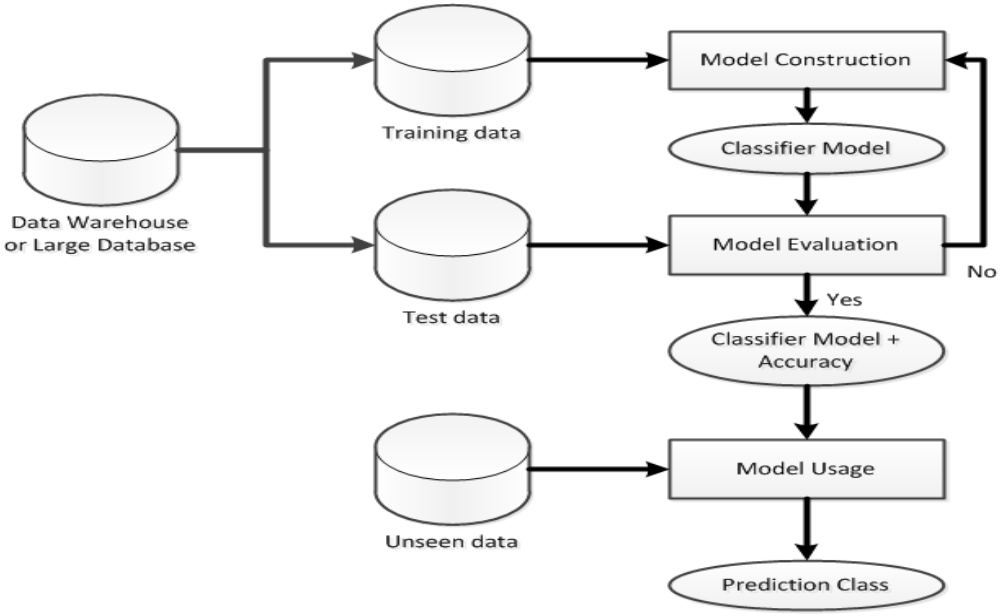
2.4.3 นิวรอนเน็ตเวิร์ก (Neuron Network)นิวรอนเน็ตเวิร์ก คือระบบที่มีการประมวลผลข้อมูลซึ่งรวมคุณสมบัติของไบโอลอจิกคอล นิวรอนเน็ตเวิร์ก ถูกพัฒนาขึ้นโดยโมเดลทางคณิตศาสตร์ของกระบวนการเรียนรู้ของมนุษย์ เลียนแบบการทำงานของสมอง และจะเรียนรู้จากชุดข้อมูลของชุดข้อมูลของชุดความรู้ นิวรอนเน็ตเวิร์ก ประกอบด้วยหน่วยความจำจำนวนมากเรียกว่านิวรอน (Neurons) เซลล์ (Cells) หรือโหนด (Nodes) แต่ละนิวรอนต่อกันโดยคอนเน็กชั้นลิงค์ (Connection Link) ที่มีค่าน้ำหนักของมันอยู่ ในแต่ละการเชื่อมต่อ โดยค่าน้ำหนักจะแสดง

รายละเอียดที่เน็ตเวิร์กใช้ในการแก้ปัญหา โดยนิวรอนเน็ตเวิร์กถูกใช้ในการแก้ปัญหาอย่างกว้างขวาง เช่น การเก็บและการเรียกข้อมูล การแยกประเภทของข้อมูล การเปลี่ยนจากรูปแบบของอินพุท (Input) ให้อยู่ในรูปแบบของเอาท์พุท (Output) ความสามารถในการ ตรวจสอบรูปแบบของข้อมูลที่คล้ายคลึงกับความคิดของมนุษย์ เป็นต้น ถึงแม้ว่านิวรอนเน็ตเวิร์ก สามารถนำไปประยุกต์ใช้กับงานหลายๆ ชนิดได้อย่างมีประสิทธิภาพ

2.4.4 จีเนติก อัลกอริทึ่ม (Genetic Algorithms : GA) จีเนติก อัลกอริทึ่ม เป็นทฤษฎีที่จำลองกระบวนการวิวัฒนาการทางธรรมชาติ คือการคัดเลือกทางธรรมชาติ และอาศัยพื้นฐานความคิดทางพันธุกรรมในการถ่ายทอดลักษณะต่าง ๆ ไปยังรุ่นถัดไป ที่สามารถนำมาพัฒนาใช้ในการหาคำตอบที่เหมาะสมที่สุดของแต่ละปัญหา จีเนติกอัลกอริทึ่มเป็นวิธีการหาคำตอบโดยการพิจารณา และดำเนินการจากกลุ่มของคำตอบของ ปัญหาที่ถูกสร้างขึ้นมาโดยการเข้ารหัส คือการแปลงค่าตัวแปรหรือพารามิเตอร์ (Parameters) ของปัญหา ให้อยู่ในรูปโครงสร้างของโครโมโซม (Cromosomes) ที่กำหนด เพื่อคัดเลือกโครโมโซมคำตอบที่ เหมาะสมสำหรับสร้างวิวัฒนาการของคำตอบให้ดีขึ้นตามกระบวนการทางพันธุศาสตร์ โดยการแลกเปลี่ยนค่าพารามิเตอร์ต่าง ๆ ระหว่างโครโมโซมที่ถูกคัดเลือกอันจะทำให้คำตอบของปัญหาถูกปรับปรุงให้ดีขึ้น

2.4.5 การจำแนกประเภทข้อมูล (Data Classification) การจำแนกประเภทข้อมูลคือกระบวนการสร้างโมเดลจำแนกประเภทข้อมูล (Data Classification Model) เพื่อทำนายกลุ่มของข้อมูลใหม่ (Unseen data) ตัวอย่างของกลุ่ม เช่น กลุ่มของลูกค้าที่ซื้อคอมพิวเตอร์ไม่ซื้อคอมพิวเตอร์ กลุ่มของลูกค้าที่ฐานะดี ปานกลาง แย่ กลุ่มของการผลิตสินค้า ผ่านเกณฑ์ ไม่ผ่านเกณฑ์ ในที่นี้คำว่ากลุ่มจะเรียกว่า ชั้นของข้อมูล (Class) ซึ่งใน ชั้นข้อมูลเดียวกันนั้นจะต้องมีข้อมูลที่มีความเหมือนหรือคล้ายคลึงกันมากกว่าข้อมูลที่อยู่ในชั้นข้อมูลที่แตกต่างกัน การสร้างโมเดลจำแนกประเภทข้อมูล จะเกิดขึ้นมาจากการหาความสัมพันธ์ของข้อมูลในฐานข้อมูลขนาดใหญ่ โดยข้อมูลทั้งหมดจะมีการแบ่งออกเป็น 2 กลุ่ม คือชุดข้อมูลของชุดความรู้(Training Set) เป็นชุดข้อมูลที่มีบทบาทในการสร้างโมเดลจำแนกประเภทข้อมูลขึ้นมา และมีกลุ่มข้อมูลทดสอบ (Test Set) เป็นชุดข้อมูลประเมินความถูกต้องของโมเดลจำแนกประเภทข้อมูล โมเดลจำแนกประเภทข้อมูลได้ถูกนำมาประยุกต์ใช้งานหลาย ๆ ด้าน ไม่ว่าจะเป็นการวิเคราะห์หุ้น เพื่อหาว่าหุ้นแต่ละบริษัทมีคุณภาพเป็นอย่างไร เมื่อมีปัจจัยที่เกี่ยวข้อง ไม่ว่าจะเป็น การเติบโตของรายได้ ความสามารถในการควบคุมต้นทุน ความผันผวนของรายได้และกำไร และผู้บริหาร หรือจะเป็นการพยากรณ์อากาศ การจัดสรรกฎหมายที่เหมาะสมในการพิจารณาคดีความ การจัดการความสัมพันธ์ของลูกค้า (Customer Relationship Management) และอื่นๆ

กระบวนการสร้างตัวโมเดลจำแนกประเภทข้อมูล แบ่งออกเป็น 3 ขั้นตอน ซึ่งภาพรวมของกระบวนการสร้างโมเดลจำแนกประเภทข้อมูลแสดงได้ดังรูปด้านล่าง



ภาพที่ 2.2 กระบวนการจำแนกข้อมูล

ที่มา : http://th.wikipedia.org.th

กระบวนการของแต่ละขั้นตอนมีดังนี้

2.4.5.1 การสร้างโมเดลจำแนกประเภท โดยอาศัยการเรียนรู้ (Model Construction Learning) เป็นขั้นตอนการสร้างโมเดลจำแนกประเภท โดยอาศัยการเรียนรู้จากข้อมูลที่ได้กำหนดชั้นข้อมูล (Class) ไว้เรียบร้อยแล้วหรือเรียกว่าชุดข้อมูลเรียนรู้ (Training data) ซึ่งโมเดลจำแนกประเภทที่ได้จะแสดงด้วยวิธีการพื้นฐานทางเหมืองข้อมูล (Data Mining) ยกตัวอย่างเช่น ต้นไม้ตัดสินใจ (Decision Tree) โมเดลจำแนกประเภทที่ได้จะมีลักษณะคล้ายต้นไม้จริงกลับหัวที่มีโหนดรากอยู่ด้านบนสุดและโหนดใบอยู่ล่างสุดของต้นไม้ แต่ละโหนดบนต้นไม้จะมีคุณลักษณะ (Attribute) เป็นตัวเลือกทดสอบ มีกิ่งซึ่งเป็นค่าที่เป็นไปได้ของคุณลักษณะ (Attribute value) ที่ถูกเลือกทดสอบไว้ และมีโหนดใบแสดง ชั้นข้อมูล (class) ที่กำหนดไว้

2.4.5.2 ขั้นตอนตรวจสอบความ (Model Evaluation Accuracy) โดยอาศัยข้อมูลที่ใช้สำหรับทดสอบเรียกว่าข้อมูลทดสอบ (Testing Data) ซึ่งกลุ่มที่แท้จริงของข้อมูลที่ใช้ทดสอบจะถูกนำมาเปรียบเทียบกับกลุ่มที่หามาได้จากโมเดลจำแนกประเภท เพื่อทดสอบว่าโมเดลจำแนกประเภทนี้สามารถจัดกลุ่มประเภทข้อมูลได้อย่างถูกต้องมากน้อยเพียงใด และมีการปรับปรุงโมเดลจำแนกประเภทจนกว่าจะได้ค่าความถูกต้องในระดับที่ยอมรับได้

2.4.5.3 ขั้นตอนการนำโมเดลจำแนกประเภทที่สร้างขึ้นมาใช้(Model Usage Classification) ข้อมูลที่ไม่เคยเห็นมาก่อน (Unseen Data) เพื่อทำนายและกำหนดกลุ่มให้กับข้อมูลนั้น

2.4.6 กฎความสัมพันธ์ (Association Rule) เป็นการค้นหากฎความสัมพันธ์ของข้อมูล โดยค้นหาความสัมพันธ์ของข้อมูลทั้งสองชุดหรือมากกว่าสองชุดขึ้นไปไว้ด้วยกัน ความสำคัญของกฎการวัดโดยใช้ข้อมูลสองตัวด้วยกันคือค่าสนับสนุน (Support) ซึ่งเป็นเปอร์เซ็นต์ของการดำเนินการที่กฎสามารถนำไปใช้ หรือเป็นเปอร์เซ็นต์ของการดำเนินการที่กฎที่ใช้มีความถูกต้อง และข้อมูลตัวที่สองที่นำมาใช้วัดคือค่าความมั่นใจ (Confidence) ซึ่งเป็นจำนวนของกรณีที่กฎถูกต้องโดยสัมพันธ์กับจำนวนของกรณีที่กฎสามารถนำไปใช้ได้ ในการหากฎความสัมพันธ์นั้นจะมีขั้นตอนวิธีการหาหลายวิธีด้วยกัน แต่ขั้นตอนวิธีที่เป็นที่รู้จักและใช้อย่างแพร่หลาย กฎการจำแนก (Classification Rules) ในการจำแนกประเภทข้อมูลโดยใช้กฎความสัมพันธ์ (Associative Classification) เป็นเทคนิคหนึ่งในเทคนิคในการจำแนกประเภทข้อมูล (Data Classification) ที่น่าสนใจ ซึ่งเป็นเทคนิคที่ให้ความแม่นยำในการทำนายสูง เนื่องจากมีการนำเทคนิคการหากฎความสัมพันธ์เข้ามาใช้ร่วมกับเทคนิคการจำแนกประเภทข้อมูล

การวิเคราะห์หรือการค้นพบกฎความสัมพันธ์ (Association Rule Mining/analysis) หมายถึงข้อมูลกลุ่มหนึ่งเกิดขึ้นกับข้อมูลกลุ่มอื่น ๆ ด้วยค่าความเชื่อมันระดับหนึ่ง ตัวอย่างการวิเคราะห์กฎความสัมพันธ์มักถูกนำมาประยุกต์กับในธุรกิจการขาย เช่น ในศูนย์การค้าเพื่อวิเคราะห์ว่าลูกค้ามักซื้อกลุ่มสินค้าหนึ่ง ๆ กับสินค้ากลุ่มอื่น ๆ ใดบ้าง เพื่อประโยชน์ในหลาย ๆ ด้าน ทั้งในเรื่องการจัดการสินค้าคงคลัง การจัดสินค้าในชั้นวางขาย การจัดโปรโมชัน รวมถึงการสร้างความพึงพอใจให้ลูกค้า และการลดต้นทุนของธุรกิจ โดยการประยุกต์ในธุรกิจเช่นนี้ ชุดข้อมูลที่ใช้ในการวิเคราะห์ความสัมพันธ์จึงมีชื่อเรียกหลากหลายชื่อ อาทิเช่น Market Basket Dataset/transactions และ Transactional Dataset เป็นต้น

ค่าสนับสนุนและค่าความเชื่อมั่น

ค่าสนับสนุน (Support) หรือค่าความถี่ของ Item sets ใดๆ ค่าสนับสนุนของ Item sets X เขียนได้เป็น sup(X) หมายถึงสัดส่วนของจำนวนทรานแซคชันที่มี Item sets X ปรากฏอยู่ (σ) ต่อจำนวน ทรานแซคชันทั้งหมดในชุดข้อมูล (ntrans)

ค่าความเชื่อมั่น (Confidence) ของกฎ เขียนในรูป conf(X => Y) หรือX => Y) (c%) โดยที่ X และ Y เป็น Item sets คือ อัตราส่วนของจำนวนทรานแซคชันที่มีทั้ง Item sets X และ Y ต่อจำนวนทรานแซคชันที่มี Item sets X โดยที่ $X \subset I$, $Y \subset I$ และ $X \cap Y = \varnothing$

ตารางที่ 2.1 นิยามเทอมทั่วไป

เทอม	คำอธิบาย
$I = \{i_1, i_2, ..., i_n\}$	เป็นเซตของตัวอักษรหรือตัวเลข เรียก item
$D = \{ t_1 , t_2 , \dots , t_n \}$	เป็นเซตของทรานแซคชัน (Transactions) ซึ่งเรียกว่า D โดยที่ D เป็นชุดข้อมูล
$T_i = \{ i_{i1} , i_{i2} , \dots , i_{ik} \}$	ทรานแซคชันหนึ่งประกอบด้วยเซตของ items โดยที่ $T \subseteq I$
TID	ดัชนีของแต่ละทรานแซคชัน
Item sets	เซตของ items
k-Item sets	เป็นเซตของ items ที่มีความยาว k ตัว เช่น {AB, BC, CD, CE,…} เป็นเซตของ 2- Item sets {ABC, BCD, } เป็นเซตของ 3- Item sets
Candidate k-Item sets	Item sets ขนาดความยาว k ที่เราสนใจและต้องการตรวจสอบว่าเป็น Frequent k-Item sets หรือไม่ หรือกล่าวอีกนัยหนึ่งว่า k- Item sets เหล่านั้นจะถูกตรวจนับค่าความสนับสนุน ขณะที่อ่านชุดข้อมูล (Dataset)
Ntrans หรือ $\|D\|$	จำนวนทรานแซคชันทั้งหมดในชุดข้อมูล
Nitems หรือ $\|I\|$	จำนวน Items ทั้งหมดในชุดข้อมูล

ค่าสนับสนุนถูกนำมาใช้ เพื่อการกำจัดกฎความสัมพันธ์ที่ไม่น่าสนใจ ทั้งนี้เนื่องจากกฎที่ประกอบด้วย Item sets ซึ่งมีค่าสนับสนุนน้อย นั่นหมายถึง Item sets นั้น ๆ ไม่เป็นที่น่าสนใจของลูกค้า หรือถูกซื้อด้วยจำนวนที่ไม่มากนัก ดังนั้นกฎความสัมพันธ์ที่มี Item sets เหล่านั้นปรากฏอยู่จึงไม่ถูกสนใจไปด้วย ในขณะที่ค่าความเชื่อมั่นใช้เพื่อแสดงถึงความน่าเชื่อถือของกฎ ยิ่งมีค่าความเชื่อมั่นสูงนั่นก็หมายความว่า Item sets หนึ่งถูกซื้อร่วมกับ Item sets อีกกลุ่มหนึ่งมากขึ้นเท่านั้น

การคำนวณค่าสนับสนุนและค่าความเชื่อมั่นที่เกี่ยวข้องกับ Item sets X และ Y เขียนในรูปแบบสมการ (2.1) และสมการ (2.2)

สมการ (2.1)

สมการ (2.2)

นิยามกฎความสัมพันธ์และขั้นตอนการวิเคราะห์ความสัมพันธ์ การค้นพบกฎความสัมพันธ์ กำหนดให้มีชุดข้อมูลเชิงรายการ (Input Data Set) และค่า Threshold จำนวนสองค่าคือความสนับสนุนขั้นต่ำ (Minimum Support หรือ Minsup) และค่าความเชื่อมั่นขั้นต่ำ (Minimum confidence หรือ minconf) ทำการหากฎความสัมพันธ์ซึ่งมีค่าสนับสนุนของ Item sets ที่ปรากฏในกฎมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำ และมีค่าความเชื่อมั่นของกฎมากกว่าหรือเท่ากับค่าความเชื่อมั่นขั้นต่ำ

กระบวนการในการหากฎความสัมพันธ์สามารถแบ่งออกเป็นสองขั้นตอนหลัก ๆ ได้แก่

ขั้นตอนที่ 1 การหาข้อมูลที่เกิดขึ้นบ่อย (Frequent Item sets) เป็นการเลือกเฉพาะ Item sets ซึ่งมีค่าความสนับสนุนมากกว่าหรือเท่ากับความสนับสนุนขั้นต่ำ

ขั้นตอนที่ 2 การหากฎความสัมพันธ์ เป็นการพิจารณาว่ากฎความสัมพันธ์ใด ที่มีค่าความเชื่อมั่นมากกว่าหรือเท่ากับความเชื่อมั่นขั้นต่ำ

การหากฎความสัมพันธ์ทำได้หลายวิธี เช่นโดยใช้วิธี Brute-force หมายถึง การหากฎที่เป็นไปได้ทั้งหมดจากชุดข้อมูลที่กำหนดให้ จำนวนกฎที่เป็นไปได้ทั้งหมดคำนวณได้จากสมการ (2.3)

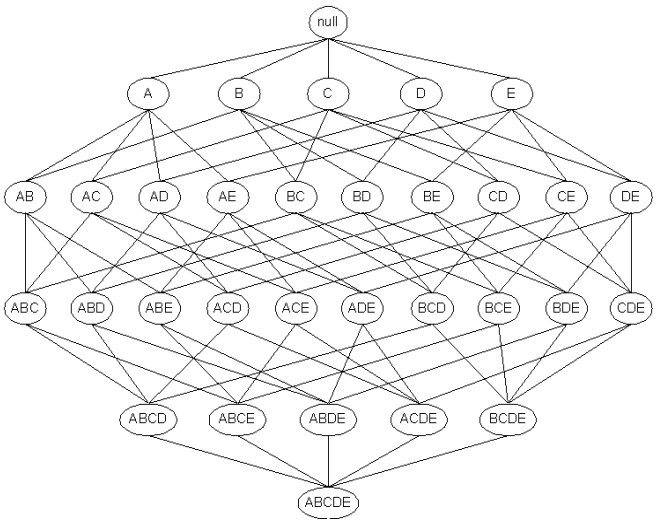
สมการ (2.3)

โดยที่ *nrules* หมายถึงจำนวนกฎความสัมพันธ์ และ *nitems* คือจำนวน items ทั้งหมดที่ปรากฏในชุดข้อมูลเชิงรายการ ซึ่งจะเห็นว่าการหากฎในลักษณะนี้ ต้องใช้ทรัพยากรในระบบคอมพิวเตอร์มหาศาล ทั้งที่เป็นการเปรียบเทียบและการคำนวณ เรียกโดยรวมว่า เวลาในการคำนวณ (Computation time) สูงมาก ทั้งๆ ที่บางส่วนหรือบางกฎอาจไม่จำเป็นต้องทำ จึงได้มีการพัฒนาขั้นตอนวิธีต่างๆ เพื่อที่จะลด (Computation time) ลง ขั้นตอนในการหาข้อมูลที่เกิดขึ้นบ่อย (Frequent Item sets) ใช้เวลาและใช้ทรัพยากรคอมพิวเตอร์มากกว่าขั้นตอนในการหากฎความสัมพันธ์ ซึ่งรายละเอียดจะอธิบายในส่วนถัดไป ทั้งนี้ได้มีการนำเสนอเทคนิควิธีการในการหา Frequent Item sets จากนักวิจัยจำนวนมากเพื่อแก้ปัญหานี้ โดยในส่วนถัดๆ ไปจะนำเสนอรายละเอียดของขั้นตอนวิธีที่สำคัญ ๆ

การสร้างข้อมูลที่เกิดขึ้นบ่อย (Frequent Item sets)

เราสามารถใช้โครงสร้างแลตทิซ (Lattice structure) ในการแจกแจง Item sets ทั้งหมดที่เป็นไปได้จากจำนวน Items ที่มีอยู่ เช่นตัวอย่างโครงสร้างแลตทิซของ 5 Items คือ I = {A, B, C, D, E} แสดง

ได้ดังภาพที่ 2.3 ซึ่งมีความยาว Item set จากระดับชั้น (Level) ที่ 1 (Item set) ถึงระดับชั้นที่ 5 (5- Item set)



ภาพที่ 2.3 Item sets lattice

ถ้าในชุดข้อมูลมีจำนวน Item set เท่ากับ k items ดังนั้นจำนวน Item sets ที่มีโอกาสเป็นFrequent Item sets ทั้งหมดคำนวณได้จาก $2^k - 1$ (ไม่รวมเซตว่าง) ซึ่งจะเห็นได้ว่าเมื่อมีการใช้งานจริงในภาคธุรกิจ หรือลักษณะงานอื่นๆ ค่าของ k จะสูงมาก ผลที่ตามมาคือการค้นหาและเปรียบเทียบรวมถึงการคำนวณจะมีจำนวนมากขึ้นเป็นทวีคูณ โดยใช้แนววิธี Brute-force ในการหา Frequent Item sets ขั้นตอนวิธีจะต้องแจง Item sets es ทั้งหมดที่เป็นไปได้ตามโครงสร้างแลตทิซ เราเรียก Item sets ชุดนี้ว่า Candidate Item sets และจากนั้นจะต้องนำ Item sets เหล่านี้ไปเปรียบเทียบกับ Items ต่างๆ ในแต่ละทรานแซคชัน ถ้าตรงกันก็เพิ่มค่าถี่ในกับ Item sets นั้น ๆ ซึ่งค่าซับซ้อน (Complexity) ของการค้นหานี้ จะเท่ากับ $O(NMw)$ โดยที่ N คือจำนวนทรานแซคชันทั้งหมด M คือจำนวน Candidate Item sets ทั้งหมดที่เป็นไปได้ ซึ่งเท่ากับ $2^k - 1$ และ w คือความยาวของทรานแซคชันที่ยาวสูงสุด

หลักการขั้นตอนวิธีของ Apriori ในส่วนนี้จะอธิบายถึงหลักการเพื่อช่วยกำจัดหรือลดจำนวนข้อมูล แทนที่จะแจกแจงทั้งหมดเหมือนที่แสดงด้วยโครงสร้างแลตทิซ เราจะใช้หลกัการที่เรียกว่า Apriori ซึ่งมีนิยามดังนี้

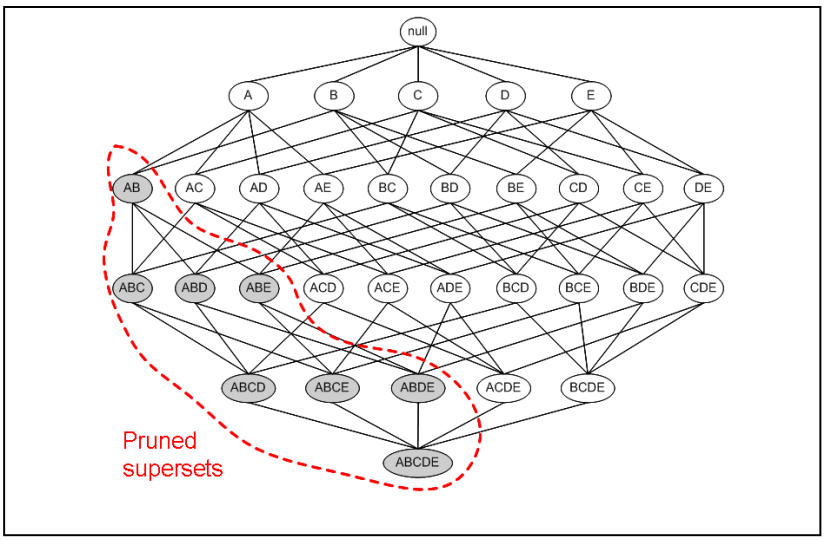
ถ้า Item sets หนึ่งๆ เป็น Frequent แล้ว ทุกๆ สับเซตของ Item sets นั้นจะต้องเป็น Frequent ด้วย

เพื่อให้มองเห็นภาพ จะใช้ตัวอย่างจากข้อมูลรูปโครงสร้างแลตทิซ ในภาพที่ 2.4 เพื่ออธิบายว่าแนวคิด Apriori จะลดจำนวน Candidate Item sets ได้อย่างไร สมมติให้ทรานแซคชันหนึ่งๆ ประกอบด้วย Items สามตัวคือ {C, D, E} ดังนั้นทรานแซคชันดังกล่าวจะประกอบด้วยสับเซตดังนี้ {C D}, {C, E}, {D, E}, {C}, {D}, {E} ซึ่ง 3-Item sets ={CDE} เป็น Frequent แล้ว ดังนั้นสับเซตขนาด 2 และขนาด 1 ของ Item sets ดังกล่าวต้องเป็น Frequent ด้วย ดังนี้

เซตของ Frequent 2-Item sets = {CD, CE, DE}

เซตของ Frequent 1-Item sets = {C, D, E}

ในทำนองตรงกันข้าม ถ้า {A, B} ไม่เป็น Frequent Item sets หรือเรียกว่าเป็น Infrequent Item sets ดังนั้นทุกๆ Superset ของ Item sets ดังกล่าวก็จะเป็น Infrequent Item sets ด้วย ดังแสดงได้ด้วยโครงสร้างแลตทิซในภาพที่ 2.4 ซึ่งหลักการในการกำจัด Infrequent Item sets นี้เรียกว่า support-based pruning กล่าวคือในการตรวจนับความถี่ของ Item sets ใดๆ อยู่ แล้วพบว่าค่าสนับสนุนหรือค่าความถี่น้อยกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำ เราสามารถกำจัด Item sets นั้นๆ ออกไป รวมถึงไม่จำเป็นต้องพิจารณาตรวจนับทุกๆ Superset ของ Item sets นั้นๆ ด้วยเช่นกัน ด้วยวิธีการเช่นนี้จะทำให้จำนวน Candidate Item sets ลดขนาดลง คุณสมบัตินี้มีชื่อเรียกว่า anti-monotone สำหรับการนับจำนวนความถี่ ซึ่งมีดังนี้



ภาพที่ 2.4 Item sets lattice กรณีกำจัด Infrequent Item sets

คุณสมบัติ Monotonicity กำหนดให้ $J = 2^I$ เป็น Power set ของ I และ I หมายถึงจำนวน Items ในชุดข้อมูล f จะเรียกว่าเป็น Monotone ถ้า

สมการ (2.4)

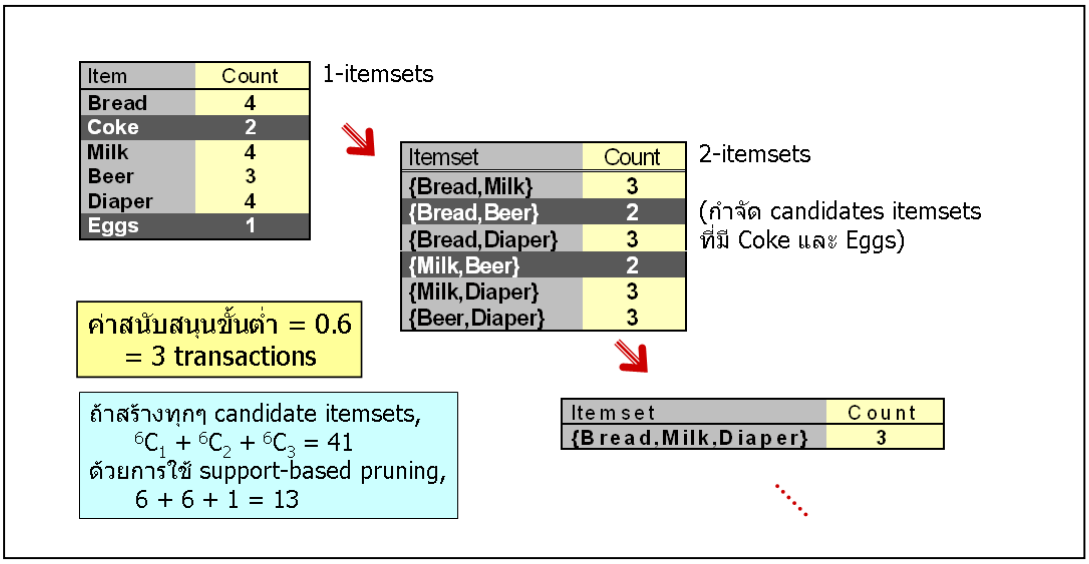
จาก (2.4) ซึ่งหมายความว่า ถ้า X เป็นสับเซตของ Y แล้ว ***f(X)* จะต้องไม่มากกว่า *f(Y)*** ในทางตรงกันข้าม f จะเรียกว่าเป็น Anti-monotone ถ้า

$$\forall X,Y \in I : (X \subseteq Y) \Rightarrow f(X) \leq f(Y) \qquad \text{สมการ (2.5)}$$

นั่นคือจาก (2.5) ถ้า X เป็นสับเซตของ Y แล้ว ***f(Y)* จะต้องไม่มากกว่า *f(X)***

จากคุณสมบัติ Anti-monotone ในสมการ (2.5) เราสามารถนำมาประยุกต์ให้ f เป็นค่าสนับสนุน ก็จะสอดคล้องกับการกำจัด Infrequent Item sets ที่เรียกว่า Support-based pruning ที่กล่าวแล้วข้างต้นนั่นเอง

การสร้าง Frequent Item sets ด้วยขั้นตอนวิธี Apriori ขั้นตอนวิธี Apriori เป็นขั้นตอนวิธีแรกในการใช้เทคนิค Support-based pruning เพื่อลดจำนวนครั้งในการสร้าง Candidate Item sets ภาพที่ 2.5 แสดงขั้นตอนในการหา Frequent Item sets โดยการกำจัด Candidate Item sets ที่ไม่จำเป็นต้องสร้างเมื่อเปรียบเทียบกับแนววิธี Brute-force กำหนดค่าสนับสนุนขั้นต่ำเท่ากับ 0.6 หรือ 3 ทรานแซคชัน



ภาพที่ 2.5 ขั้นตอนการสร้าง Frequent Item sets ด้วยขั้นตอนวิธี Apriori

ขั้นตอนวิธีในการสร้าง Frequent Item sets แสดงด้วยรหัสเทียม (Pseudocode) เมื่อกำหนดให้ C_k = Candidate k-Item sets และ L_k = Frequent k-Item sets ส่วนสัญลักษณ์อื่นๆ เป็นไปตามนิยามเทอมทั่วไป

ขั้นตอนวิธี Apirori ในการหา Frequent Item sets

Input: ชุดข้อมูล D, และค่า Threshold *minsup*

Output: Frequent Item sets ทุกระดับ

ขั้นตอนที่ 1 L_1 = {Frequent 1-Item sets };

ขั้นตอนที่ 2 For (k =2 ; $L_{k-1} \neq \square$; k++)

ขั้นตอนที่ 3 C_k = apriori-gen(L_{k-1}); //generate candidates

ขั้นตอนที่ 4 For all transactions $t \square D$

4.1 C_t = subset (C_k, t); //Candidates contained in t

4.2 For all candidates $c \square C_t$ do

4.2.1 c.count++;

4.3 End for

ขั้นตอนที่ 5 End for

ขั้นตอนที่ 6 $L_k = \{c \square C_k \mid c.\text{count} \geq \text{minsup}\}$

ขั้นตอนที่ 7 Answer = $\bigcup_1^k L_k$;

ขั้นตอนที่ 8 End for

การทำงานของขั้นตอนวิธี Apriori แบ่งเป็นส่วนหลัก ๆ ดังนี้

- ขั้นตอนวิธีเริ่มต้นทำงาน โดยการอ่านชุดข้อมูลนำเข้า D รอบแรก เพื่อหา Frequent 1-Item sets และเก็บไว้ในเซต L_1 (ขั้นตอนที่ 1)
- จากนั้นทำขั้นตอนซ้ำ จากระดับสอง (k = 2) เพื่อสร้าง Candidate k-Item sets จาก Frequent (k-1)-Item sets ซึ่งเป็นสมาชิกของ L_{k-1} โดยการสร้าง Candidate (k-1)-Item sets ในขั้นตอนนี้จะทำโดยขั้นตอนวิธี apriori_gen ซึ่งจะกล่าวถึงรายละเอียดในส่วนถัดไป (ขั้นตอนที่ 2 - 8)
- ในระหว่างการสร้าง Candidate Item sets แต่ละระดับ ขั้นตอนวิธีจะทำการนับ Item sets เหล่านั้น เพื่อพิจารณาหา Frequent Item sets โดยการอ่านชุดข้อมูล (ขั้นตอนที่ 4 - 5) ซึ่งรายละเอียดจะอธิบายในส่วนถัดๆ ไป
- หลังจากมีการนับค่าความถี่ของ Candidate Item sets แต่ละตัว แล้ว ขั้นตอนวิธีจะกำจัด Item sets ที่มีค่าความสนับสนุไม่ถึงค่าสนับสนุนขั้นต่ำ (ขั้นตอนที่ 6)
- ขั้นตอนวิธีจะหยุดทำงานเมื่อไม่สามารถสร้าง Candidate Item sets ในระดับที่สูงขึ้นได้อีกต่อไป

ซึ่งจะเห็นว่าขั้นตอนวิธี Apriori ทำงานในลักษณะ Level-wise กล่าวคือจะทำงานทีละระดับจากโครงสร้างแลตทิซ เริ่มจากระดับที่ 1 จนถึงระดับที่ k ซึ่งเป็นระดับสูงสุดที่มี Frequent Item sets นอกจากนั้นขั้นตอนวิธีดังกล่าวยังมีอีกหนึ่งคุณลักษณะที่สำคัญ ได้แก่ การใช้กลยุทธ์ที่เรียกว่า generate-and-test ในการหา Frequent Item sets กล่าวคือในแต่ละระดับ ขั้นตอนวิธีจะทำการสร้าง Item sets ที่เป็น Candidate ก่อน แล้วจึงนำเอา Item sets เหล่านั้นไปทดสอบ โดยการนับจำนวนกับ Items ต่างๆ ในชุดข้อมูลทรานแซคชัน และทำการเปรียบเทียบกับค่าสนับสนุนขั้นต่ำ

ปัจจัยที่มีผลกระทบต่อประสิทธิภาพของขั้นตอนวิธี Apriori

ประสิทธิภาพของขั้นตอนวิธี Apriori ในการสร้าง Frequent Item sets ขึ้นอยู่กับปัจจัยดังต่อไปนี้

1) ค่าสนับสนุนขั้นต่ำ (Minimum support threshold) ยิ่งกำหนดให้มีค่าน้อย จำนวน Frequent Item sets ก็จะยิ่งมากขึ้น ซึ่งก็หมายความว่าขั้นตอนวิธีจะต้องสร้าง Candidate Item sets จำนวนมากเช่นกัน
2) จำนวนของ Items (nitems) ซึ่งหมายถึงมิติ (Dimensionality) ยิ่งมีค่ามาก นั่นก็หมายถึงต้องใช้พื้นที่ในการจัดเก็บข้อมูลดังกล่าวสูงขึ้น ซึ่งก็คือขนาดของแฟ้มใหญ่ขึ้น ดังนั้นถ้าจำนวน Frequent Item sets มากขึ้นตามจำนวน Items ที่มากขึ้น การทำ I/O และการคำนวณเปรียบเทียบ ก็จะมากขึ้นไปด้วย
3) จำนวนของทรานแซคชัน (ntrans) เหตุผลเช่นเดียวกับข้อ 2) และเนื่องจากขั้นตอนวิธี Apriori ต้องอ่านชุดข้อมูลซ้ำๆ หลายรอบ ถ้าแต่ละรอบต้องใช้เวลานานขึ้น เวลาโดยรวมก็จะนานขึ้นเช่นเดียวกัน
4) ความยาวเฉลี่ยของทรานแซคชัน กล่าวคือแต่ละทรานแซคชันประกอบด้วยหลายๆ Items จะทำให้ค่าความยาวเฉลี่ยของทรานแซคชันมีค่าสูงขึ้นด้วย ส่งผลให้ระดับ Item sets lattice มีค่าสูง ซึ่งมีผลกระทบที่ตามมาสองประเด็น คือจะได้ Frequent Item sets ในระดับสูงขึ้น ซึ่งก็คือจะมีจำนวน Candidate Item sets มากขึ้นด้วย รวมถึงการทำ Hashing ในแต่ละทรานแซคชันก็จะใช้เวลามากขึ้น เพราะมีจำนวน Items ในแต่ละทรานแซคชันมากขึ้นนั่นเอง

การสร้างกฎความสัมพันธ์ ในส่วนนี้จะเป็นการอธิบายการสร้างกฎความสัมพันธ์จาก Frequent Item sets ที่หามาได้ในขั้นตอนที่ 1 โดยแต่ละ Frequent k-Item sets สมมติว่าชื่อ Item sets Y สามารถสร้างกฎได้สูงสุดตามสมการ (2.6)

$$\text{nrules} = 2^k - 2 \qquad \text{สมการ (2.6)}$$

โดยไม่รวม Y และ $Y \Rightarrow \varnothing$ ซึ่งกฎความสัมพันธ์นี้สร้างได้จากการแบ่ง Item sets Y ออกเป็นสองสับเซต ได้แก่ X และ $Y - X$ ซึ่งเขียนในรูปแบบ $X \Rightarrow Y$ - X และสอดคล้องกับค่าความเชื่อมั่นขั้นต่ำ จะเห็นว่าในการสร้างกฎไม่จำเป็นต้องอ่านข้อมูลจากชุดข้อมูลทรานแซคชันเหมือนกับขั้นตอนการหา Frequent Item sets ดังนั้นเวลาที่ใช้ในการประมวลส่วนนี้จะน้อยกว่าเวลาที่ใช้ในการหาหรือสร้าง Frequent Item sets

การกำจัดกฎที่ไม่ต้องการ หลังจากสร้างกฎความสัมพันธ์ได้จาก Frequent Item sets Y ใดๆ ได้แล้ว ในทำนองเดียวกับการหา Candidate Item sets ที่มีการกำจัด Item sets บางตัวทิ้งได้ อันเนื่องมาจากคุณสมบัติ Monotone แต่ในการสร้างกฎ จะสามารถใช้ทฤษฎีของมาตรวัดความเชื่อมั่น (Confidence measure) ในการกำจัดกฎบางกฎได้

ทฤษฎีความเชื่อมั่น

ถ้ากฎ $X \Rightarrow Y - X$ ไม่สอดคล้องกับค่าความเชื่อมั่นขั้นต่ำแล้ว ดังนั้นกฎ $X' \Rightarrow Y - X'$ โดยที่ X' เป็นสับเซตของ X ก็จะไม่สอดคล้องกับค่าความเชื่อมั่นขั้นต่ำเช่นเดียวกัน

แนวทางในการพิสูจน์ทฤษฎีทำได้โดย พิจารณากฎ

$X' \Rightarrow Y - X'$ ซึ่งมีค่าความเชื่อมั่นเท่ากับ $(Y) / \sigma(X')$

และ

$X \Rightarrow Y - X$ ซึ่งค่าความเชื่อมั่น $(Y) / \sigma(X)$

เนื่องจาก X' เป็นสับเซตของ X ดังนั้น $(X') >= \sigma(X)$ นั่นก็คือค่าความเชื่อมั่นของกฎ (1) ไม่มีทางมีค่ามากกว่าค่าความเชื่อมั่นของกฎ (2)

การสร้างกฎความสัมพันธ์โดยใช้ขั้นตอนวิธี Apriori ขั้นตอนวิธี Apriori ใช้แนววิธีทำทีละระดับ (Level-wise) ในการสร้างกฎความสัมพันธ์ โดยที่แต่ละระดับหมายถึงจำนวน Items ซึ่งปรากฏในส่วนขวามือของกฎ ($\Rightarrow$) ในตอนเริ่มต้น กฎซึ่งมีค่าความเชื่อมั่นมากกว่าหรือเท่ากับค่าความเชื่อมั่นขั้นต่ำและมี Item ทางขวามือของกฎเพียงตัวเดียว จะถูกสร้างขึ้นมาก่อน จากนั้นจึงสร้างกฎที่เป็น Candidate ซึ่งเกิดจาก Items ในกฎนั้น ๆ เป็นลำดับถัดไป

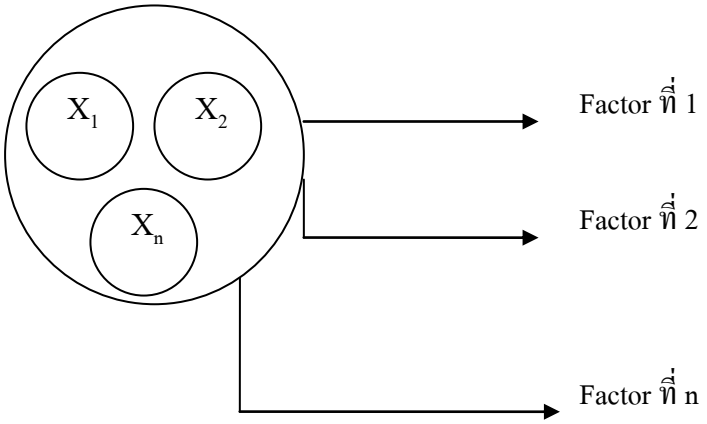
2.5 **การวิเคราะห์ปัจจัย**

การวิเคราะห์องค์ประกอบเป็นเทคนิคทางสถิติ สาหรับวิเคราะห์ตัวแปรหลายตัว (Multivariate Analysis Techniques)

กัลยา วานิชบัญชา (2551) สรุปว่า เป็นการวิเคราะห์หลายตัวแปรเทคนิคหนึ่งเพื่อการสรุปรายละเอียดของตัวแปรหลายตัว หรือเรียกว่าเป็นเทคนิคที่ใช้ในการลดจำนวนตัวแปรเทคนิคหนึ่งโดยการศึกษาถึงโครงสร้างความสัมพันธ์ของตัวแปร และสร้างตัวแปรใหม่เรียกว่า องค์ประกอบ โดยองค์ประกอบที่สร้างขึ้นจะเป็นการนำตัวแปรที่มีความสัมพันธ์กันหรือมีความร่วมกันสูงมารวมกันเป็นองค์ประกอบเดียวกัน ส่วนตัวแปรที่อยู่คนละองค์ประกอบมีความร่วมกันน้อย หรือไม่มีความสัมพันธ์กันเลย

ถ้านำตัวแปรใด ๆ ที่สามารถวัดได้หรือสังเกตได้มาหาค่าสหสัมพันธ์กัน ก็จะพบว่าบางตัวแปรมีสหสัมพันธ์กันสูงและบางตัวแปรไม่มีสหสัมพันธ์กัน การที่กลุ่มตัวแปรที่มีสหสัมพันธ์กันแสดงว่ามีค่าที่ร่วมกันอยู่ ซึ่งเรียกว่าค่าสามัญ ซึ่งความสัมพันธ์ที่มีค่าร่วมกันอยู่นี้ สามารถหาได้โดยใช้วิธีลดจำนวนตัวแปรให้น้อยลงแล้วจัดตัวแปรเหล่านั้นเป็นกลุ่ม ๆ เฉพาะบางตัวแปรที่มีความสัมพันธ์กันเพื่อนำไปอธิบายหรือพยากรณ์สิ่งต่าง ๆ ได้ วิธีการนี้เรียกว่า การวิเคราะห์องค์ประกอบ (Factor Analysis)

การวิเคราะห์องค์ประกอบมีเงื่อนไขและข้อแตกต่างจากการวิเคราะห์สหสัมพันธ์คาโนนิคัล ซึ่งนำตัวแปรที่แปรรูปคะแนนแล้วมาจัดเป็นคู่ ๆ (Canonical Pair)โดยเรียงตามลำดับค่าสหสัมพันธ์จากมากไปหาน้อย และแต่ละคู่จะไม่มีความสัมพันธ์ข้ามคู่กัน ส่วนการวิเคราะห์องค์ประกอบ จะนำตัวแปรทั้งหมดมาวิเคราะห์พร้อมกันแล้วจำแนกออกมาได้หลายองค์ประกอบที่มีความสัมพันธ์ร่วมกันอยู่โดยเรียงองค์ประกอบตามลำดับค่าความสัมพันธ์จากมากไปหาน้อยเช่นกันดังภาพที่ 2.6



ภาพที่ 2.6 แสดงองค์ประกอบการวิเคราะห์ปัจจัย

โดยที่องค์ประกอบที่ 1 มีค่าสหสัมพันธ์สูงสุดและเรียงลงมาตามลำดับและไม่มีความสัมพันธ์กันระหว่างองค์ประกอบ แต่ตัวแปรภายในแต่ละองค์ประกอบจะมีความสัมพันธ์กันหรือมีค่าสามัญร่วมกันอยู่

การวิเคราะห์องค์ประกอบ หรือ Factor Analysis เป็นวิธีการวิเคราะห์องค์ประกอบของกลุ่มที่สนใจ ซึ่งประกอบด้วยตัวแปรต่าง ๆ จำนวนหลายตัว ว่าตัวแปรแต่ละตัวเกิดจากองค์ประกอบใด ซึ่งโดยปกติแล้วจำนวนองค์ประกอบ (Factor) จะมีจำนวนน้อยกว่าตัวแปร (Variable) ซึ่งในที่นี้องค์ประกอบ หมายถึง สิ่งที่ต้องการศึกษาซึ่งมักเป็นสิ่งที่มองไม่เห็น แต่ส่งผลต่อสิ่งที่มองเห็น หรือสิ่งที่ต้องการวัดตัวแปร หมายถึง สิ่งที่เราต้องการวัด ต้องการศึกษา

การวิเคราะห์องค์ประกอบ (Factor Analysis) หรือบางครั้งเรียกว่า การวิเคราะห์ปัจจัยเป็นเทคนิคที่จะจับกลุ่มหรือรวมกลุ่มหรือรวมตัวแปรที่มีความสัมพันธ์กันไว้ในกลุ่มหรือปัจจัยเดียวกัน ตัวแปรที่อยู่ในปัจจัยเดียวกันจะมีความสัมพันธ์กันมาก โดยความสัมพันธ์นั้นอาจจะเป็นในทิศทางบวก (ไปในทิศทางเดียวกัน) หรือทิศทางลบ (ไปในทางตรงกันข้าม) ก็ได้ ส่วนตัวแปรที่คนละปัจจัยจะไม่มีความสัมพันธ์กันหรือมีความสัมพันธ์กันน้อย หรือในอีกความหมายหนึ่งของการวิเคราะห์องค์ประกอบ หรือเรียกว่า การวิเคราะห์ตัวประกอบ เป็นเทคนิคทางสถิติที่ใช้วิเคราะห์ผลการวัด โดยใช้เครื่องมือหรือเทคนิคหลายชุดหรือหลายด้านอาจใช้แบบทดสอบแบบวัด แบบสำรวจ ฯลฯ อาจใช้ชุดเดียวแต่มีการวัดแยกเป็นรายด้านหรือหลายชุดก็ได้ ผลการวิเคราะห์จะช่วยให้ทราบว่าเครื่องมือหรือเทคนิคเหล่านั้นวัดแต่ละองค์ประกอบมากน้อยเพียงใด สำหรับการพิจารณาผลจากการวิเคราะห์จะใช้หลักเหตุผล องค์ประกอบที่วัดนั้นผลจากการวิเคราะห์องค์ประกอบจะปรากฏค่าต่าง ๆ ที่สำคัญ คือค่า Communality ซึ่งเขียนด้วย h2 เป็นค่าความแปรปรวนที่แต่ละฉบับ (ด้าน) แบ่งให้กับแต่ละองค์ประกอบ เป็นส่วนที่ชี้ถึงว่าแต่ละฉบับ (ด้าน) วัดองค์ประกอบนั้นร่วมกับตัวแปรอื่นมากน้อยเพียงใด ค่า Eigenvalues เป็นผลรวมกำลังสองของสัมประสิทธิ์ขององค์ประกอบร่วมในแต่ละองค์ประกอบ ซึ่งต้องมีค่าไม่ต่ำกว่า 1 จึงจะถือว่าเป็นองค์ประกอบหนึ่ง ๆ ที่แท้จริง ส่วน Factor Loading เป็นค่าน้ำหนัก องค์ประกอบที่แต่ละฉบับ (ด้าน) วัดในองค์ประกอบนั้นการวิเคราะห์องค์ประกอบจะยึดหลักที่ว่าตัวแปรหรือข้อมูลต่าง ๆ มีความสัมพันธ์กันมากนั้นเนื่องมาจากตัวแปรเหล่านี้มีองค์ประกอบร่วมกัน (Common Factor) สังเกตได้จากการจัดกลุ่มของตัวแปรหรือค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร ดังนั้น สามารถใช้องค์ประกอบร่วมแทนตัวแปรกลุ่มนั้นได้ ทำให้ทราบถึงโครงสร้างและแบบแผนของข้อมูล ทำให้หาองค์ประกอบร่วมของตัวแปรได้ และสามารถหาน้ำหนักองค์ประกอบ (Factor Loading) ของตัวแปรแต่ละตัวได้ ซึ่งค่าน้ำหนักองค์ประกอบนี้สามารถอธิบายได้ถึง ความแปรปรวนร่วมระหว่างกันของตัวแปร ทำให้ทราบถึงโครงสร้างและรูปแบบของข้อมูล ทำให้หาองค์ประกอบร่วมของแต่ละตัวได้ ซึ่งค่าน้ำหนัก

องค์ประกอบนี้ สามารถอธิบายได้ถึงความแปรปรวนร่วมระหว่างตัวแปรกับองค์ประกอบนั้นอันแสดงถึงขนาด (Magnitude) ของความสัมพันธ์ระหว่างตัวแปรกับองค์ประกอบ

2.5.1 ประเภทของเทคนิคการวิเคราะห์องค์ประกอบ

เทคนิคของการวิเคราะห์องค์ประกอบ แบ่งออกเป็น 2 ประเภทคือ

2.5.1.1 การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis) การวิเคราะห์องค์ประกอบเชิงสำรวจจะใช้ในกรณีที่ผู้ศึกษาไม่มีความรู้ หรือมีความรู้น้อยมากเกี่ยวกับโครงสร้างความสัมพันธ์ของตัวแปรเพื่อศึกษาโครงสร้างของตัวแปร และลดจำนวนตัวแปรที่มีอยู่เดิมให้มีการรวมกันได้

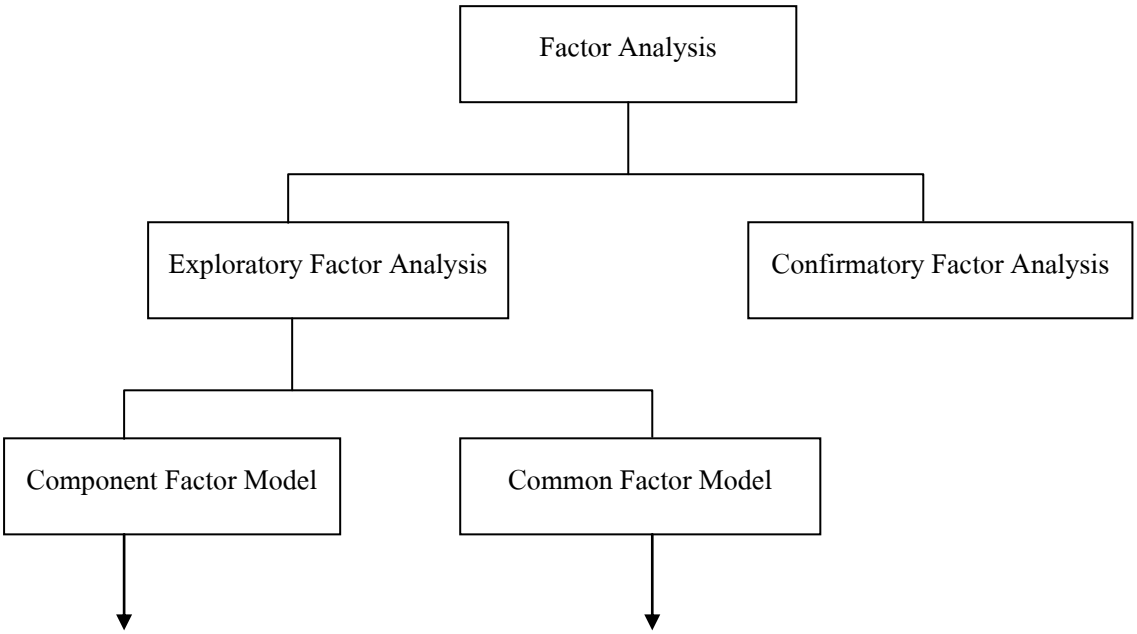
2.5.1.2 การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis) การวิเคราะห์องค์ประกอบเชิงยืนยันจะใช้ กรณีที่ผู้ศึกษาทราบโครงสร้างความสัมพันธ์ของตัวแปร หรือคาดว่าโครงสร้างความสัมพันธ์ของตัวแปรควรจะเป็นรูปแบบใด หรือคาดว่าตัวแปรใดบ้างที่มีความสัมพันธ์กันมากและควรอยู่ในองค์ประกอบเดียวกัน หรือคาดว่ามีตัวแปรใดที่ไม่มีความสัมพันธ์กัน ควรจะอยู่ต่างองค์ประกอบกัน หรือกล่าวได้ว่า ผู้ศึกษาทราบโครงสร้างความสัมพันธ์ของตัวแปร หรือคาดไว้ว่าโครงสร้างความสัมพันธ์ของตัวแปรเป็นอย่างไรและจะใช้เทคนิคการวิเคราะห์องค์ประกอบเชิงยืนยันมาตรวจสอบหรือยืนยันความสัมพันธ์ว่าเป็นอย่างที่คาดไว้หรือไม่ โดยการวิเคราะห์หาความตรงเชิงโครงสร้างนั่นเอง

2.5.2 วัตถุประสงค์ของเทคนิค Factor Analysis

2.5.2.1 เพื่อศึกษาว่าองค์ประกอบร่วมที่จะสามารถอธิบายความสัมพันธ์ร่วมกันระหว่างตัวแปรต่าง ๆ โดยที่จำนวนองค์ประกอบร่วมที่หาได้จะมีจำนวนน้อยกว่าจำนวนตัวแปรนั้น จึงทำให้ทราบว่ามีองค์ประกอบร่วมอะไรบ้าง โมเดลนี้ เรียกว่า การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis Model : EFA)

2.5.2.2 เพื่อต้องการทดสอบสมมุติฐานเกี่ยวกับโครงสร้างขององค์ประกอบว่าองค์ประกอบแต่ละองค์ประกอบด้วยตัวแปรอะไรบ้าง และตัวแปรแต่ละตัวควรมีน้ำหนักหรืออัตราความสัมพันธ์กับองค์ประกอบมากน้อยเพียงใด ตรงกับที่คาดคะเนไว้หรือไม่ หรือสรุปได้ว่าเพื่อต้องการทดสอบว่าตัวประกอบอย่างนี้ตรงกับโมเดลหรือตรงกับทฤษฎีที่มีอยู่หรือไม่ โมเดลนี้เรียกว่า การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis Model: CFA) ซึ่งเทคนิคของ Factor Analysis สามารถสรุปได้เป็นรูปแบบดังนี้

สรุปรูปแบบการวิเคราะห์ตัวประกอบ



2.5.3 ประโยชน์ของเทคนิควิเคราะห์ (Factor Analysis)

2.5.3.1 ลดจำนวนตัวแปร โดยการรวมตัวแปรหลาย ๆ ตัวให้อยู่ในองค์ประกอบเดียวกัน องค์ประกอบที่ได้ถือเป็นตัวแปรใหม่ ที่สามารถหาค่าข้อมูลขององค์ประกอบที่สร้างขึ้นได้ เรียกว่า น้ำหนักของปัจจัย จึงสามารถนำองค์ประกอบดังกล่าวไปเป็นตัวแปรสำหรับการวิเคราะห์ทางสถิติต่อไป เช่น การวิเคราะห์ความถดถอยและสหสัมพันธ์ (Regression and Correlation Analysis) การวิเคราะห์ความแปรปรวน (ANOVA) การทดสอบสมมุติฐาน T – test Z – test และการวิเคราะห์จำแนกกลุ่ม (Discriminant Analysis) เป็นต้น

2.5.3.2 ใช้ในการแก้ปัญหาอันเนื่องมาจากการที่ตัวแปรอิสระของเทคนิคการวิเคราะห์สมการความถดถอยมีความสัมพันธ์กัน (Multicollinearity) ซึ่งวิธีการอย่างหนึ่งในการแก้ปัญหานี้ คือ การรวมตัวแปรอิสระที่มีความสัมพันธ์ไว้ด้วยกัน โดยการสร้างเป็นตัวแปรใหม่หรือเรียกว่า

องค์ประกอบ โดยใช้เทคนิค Factor Analysis แล้วนำองค์ประกอบดังกล่าวไปเป็นตัวแปรอิสระในการวิเคราะห์ความถดถอยต่อไป

2.5.3.3 ทำให้เห็นโครงสร้างความสัมพันธ์ของตัวแปรที่ศึกษา เนื่องจากเทคนิค Factor Analysis จะหาค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ของตัวแปรที่ละคู่ แล้วรวมตัวแปรที่สัมพันธ์กันมากไว้ในองค์ประกอบเดียวกัน จึงสามารถวิเคราะห์โครงสร้างที่แสดงความสัมพันธ์ของตัวแปรต่าง ๆ ที่อยู่ในองค์ประกอบเดียวกันได้ ทำให้สามารถอธิบายความหมายของแต่ละองค์ประกอบได้ ตามความหมายของตัวแปรต่าง ๆ ที่อยู่ในองค์ประกอบนั้น ทำให้สามารถนำไปใช้ในด้านการวางแผนได้ เช่น การพัฒนาพหุปัญญาสำหรับนักเรียนชั้นมัธยมศึกษาตอนต้นตามทฤษฎีพหุปัญญาของการ์ดเนอร์ (2546)

2.5.4 ปัญหาการวิเคราะห์องค์ประกอบ

2.5.4.1 การวิเคราะห์องค์ประกอบไม่มีตัวแปรตาม ซึ่งแตกต่างกับการทดสอบสถิติการวิเคราะห์ถดถอยเชิงพหุแบบปกติ สถิติการวิเคราะห์ถดถอยโลจิสติค สถิติการวิเคราะห์จำแนกประเภท และการวิเคราะห์เส้นทาง ดังนั้น สถิติการวิเคราะห์องค์ประกอบ จึงไม่สามารถใช้แก้ปัญหาการวิจัยที่ต้องการหาตัวทำนายได้

2.5.4.2 ขั้นตอนการสกัดองค์ประกอบไม่สามารถระบุจำนวนรอบของการสกัดได้ ดังนั้นหลังจากขั้นตอนการสกัดองค์ประกอบนักวิจัยจึงไม่สามารถระบุจำนวนรอบของการสกัดองค์ประกอบได้ว่ามีกี่รอบจึงจะพอดี

2.5.4.3 ในปัจจุบันการวิจัยที่ต้องการทดสอบเพื่อลดจำนวนตัวแปร มีเพียงสถิติการวิเคราะห์องค์ประกอบเท่านั้น เนื่องจากสถิตินี้สามารถรวมตัวแปรหลาย ๆ ตัวให้อยู่ในองค์ประกอบเดียวกัน และทำให้เห็นโครงสร้างความสัมพันธ์ของตัวแปรที่ศึกษา โดยการหาค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ของตัวแปรทีละคู่ แล้วรวมตัวแปรที่สัมพันธ์กันมากไว้ในองค์ประกอบเดียวกัน หลังจากนี้จึงสามารถวิเคราะห์ถึงโครงสร้างที่แสดงความสัมพันธ์ของตัวแปรต่าง ๆ ที่อยู่ในองค์ประกอบเดียวกันได้ ดังนั้นเมื่อนักวิจัยต้องการวิเคราะห์ให้ได้ผลการวิเคราะห์ดังกล่าวข้างต้น จึงมีสถิติให้เลือกใช้เฉพาะสถิติการวิเคราะห์องค์ประกอบเพียงตัวเดียว แต่ยังไม่มีวิธีการทางสถิติวิธีอื่น ๆ จึงทำให้นักวิจัยต้องเลือกใช้วิธีการวิเคราะห์องค์ประกอบทั้ง ๆ ที่วิธีนี้มีข้อจำกัดดังกล่าวข้างต้น

2.6. เทคโนโลยีที่นำมาพัฒนางานวิจัย

โปรแกรม WEKA (Waikato Environment for Knowledge Analysis) : เริ่มพัฒนามาตั้งแต่ปี 1997 โดยมหาวิทยาลัย Waikato ประเทศนิวซีแลนด์ เป็นซอฟต์แวร์สำเร็จ รูปประกอบประเภทฟรีแวร์ อยู่ภายใต้การควบคุมของ GPL License ซึ่งโปรแกรม WEKA ได้ถูกพัฒนามาจากภาษาจาวา

ทั้งหมด ซึ่งเขียนมาโดยเน้นกับงานทางด้านการเรียนรู้ด้วยเครื่อง (Machine Learning) และ การทำเหมืองข้อมูล (Data Mining) โปรแกรมจะประกอบไปด้วยโมดูลย่อย ๆ สำหรับใช้ในการจัดการข้อมูล และเป็นโปรแกรมที่สามารถใช้ Graphic User Interface (GUI) และ ใช้คาสั่งในการให้ซอฟต์แวร์ประมวลผล และสามารถรัน (Run) ได้หลายระบบปฏิบัติการ และสามารถพัฒนาต่อยอดโปรแกรมได้ เป็นเครื่องมือที่ใช้ทำงานในด้านการทำเหมืองข้อมูลที่รวบรวมแนวคิดขั้นตอนวิธีมากมาย ซึ่งขั้นตอนวิธีสามารถเลือกใช้งานโดยตรงได้จาก 2 ทางคือจากชุดเครื่องมือที่มีขั้นตอนวิธีมาให้ หรือเลือกใช้จากขั้นตอนวิธีที่ได้เขียนเป็นโปรแกรมลงไปเป็นชุดเครื่องมือเพิ่มเติม และชุดเครื่องมือมีฟังก์ชันสำหรับการทำงานร่วมกับข้อมูล ได้แก่ Pre-Processing, Classification, Regression, Clustering, Association rules, Selection และ Visualization ภาพแสดง โลโก้โปรแกรม ดังภาพที่ 2.7

ภาพที่ 2.7 โลโก้ของซอฟต์แวร์วีก้าเป็นรูปนกเฉพาะถิ่นของประเทศนิวซีแลนด์

ข้อดีของซอฟต์แวร์วีก้า

1. เป็นซอฟต์แวร์เสรีที่สามารถดาวน์โหลดได้ฟรี
2. สามารถทำงานได้ทุกระบบปฏิบัติการ
3. เชื่อมต่อ SQL Database โดยใช้ Java Database Connectivity
4. มีการเตรียมข้อมูลและเทคนิคในการสร้างแบบจำลองที่ครอบคลุม
5. มีลักษณะที่ง่ายต่อการใช้งาน

ความสามารถของซอฟต์แวร์วีก้า

1. สนับสนุนเกี่ยวกับการทำเหมืองข้อมูล (Data Mining)
2. การเตรียมข้อมูล (Data Preprocessing)
3. การทำเหมืองข้อมูลด้วยเทคนิคการจำแนกข้อมูล (Classification)

4. การทำเหมืองข้อมูลด้วยเทคนิคการจัดกลุ่ม (Clustering)
5. การทำเหมืองข้อมูลด้วยเทคนิคการวิเคราะห์ความสัมพันธ์ (Associating)
6. เทคนิคการคัดเลือกข้อมูล (Selecting Attributes)
7. เทคนิคการนำเสนอข้อมูลด้วยรูปภาพ (Visualization)

สรุปข้อดีของโปรแกรม WEKA คือ มีขั้นตอนวิธีที่รู้จักกันดีของการทำเหมืองข้อมูลให้เลือกใช้อย่างครบถ้วน และสามารถเขียนฟังก์ชันเพิ่มเข้าไปในโปรแกรมเองได้

2.7 งานวิจัยที่เกี่ยวข้อง

นภดล สายคติกรณ์ (2553) งานวิจัยเพื่อศึกษาหาปัจจัยและโครงสร้างความสัมพันธ์ของปัจจัยที่มีผลต่อความปลอดภัยในการใช้บริการระบบเครือข่ายคอมพิวเตอร์ของนักศึกษา จากนั้นทำการพัฒนาแบบจำลองสมการโครงสร้างผู้วิจัย รวบรวมตัวชีวัดจากการศึกษาวรรณกรรมอ้างอิงและทฤษฎีได้จำนวน 39 ตัวชี้วัด และทำการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างจำนวน 250 ตัวอย่างทำการวิเคราะห์ปัจจัยสามารถสกัดปัจจัยได้ 6 ปัจจัย จากนั้นทำการหาสมการแสดงโครงสร้างความสัมพันธ์ของปัจจัยที่ได้ด้วยการวิเคราะห์สมการโครงสร้างหลายกลุ่ม จากแบบจำลองที่พัฒนาขึ้นจำแนกได้เป็น ความพอใจในความปลอดภัยของระบบเครือข่ายของเฉพาะเพศชาย เฉพาะเพศหญิง และรวมทั้งเพศชายและเพศหญิง จากนั้นนำแบบจำลองทั้งสามมาทดสอบความแม่นยำของสมการ ตามประเภทต่าง ๆ จากการวิจัยพบว่าความแม่นยำของสมการโครงสร้างของการพยากรณ์เฉพาะเพศชาย การพยากรณ์เฉพาะเพศหญิง และการพยากรณ์สมการทั้งสองแบบรวมกันมีค่าเท่ากับ 21.09% 27.75% และ 10.14% ตามลำดับ

ประกาศิต ช่างสุพรรณ,(2553) งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบสำหรับจำแนกระดับความมั่นใจของผู้เรียน โดยใช้เทคนิคต้นไม้ตัดสินใจอาศัยข้อมูลการทำแบบฝึกหัดวิชาฟิสิกส์ โดยรวมตัวแปรทั้งส่วนคะแนนการทำแบบฝึกหัดแต่ละคนและส่วนของการวิเคราะห์แบบฝึกหัดแต่ละข้อ รวมทั้งสิ้น 11 ตัวแปร แยกชุด ข้อมูลตามรายวิชา คือวิชาฟิสิกส์ 1, วิชาฟิสิกส์ 3 และรวมกันทั้งฟิสิกส์ 1 และฟิสิกส์ 3 จากนั้นแบ่งชุดข้อมูลตามระดับความมั่นใจของผู้เรียนเป็น 2 แบบ คือ แบบ 2 ระดับ (มั่นใจและไม่มั่นใจ) และแบบ 3 ระดับ (มั่นใจสูง มั่นใจปานกลาง และมั่นใจต่ำ) รวมทั้งสิ้น 6 ชุดข้อมูล ทำการทดสอบตัวแบบที่ ได้ด้วยวิธี 10-fold Cross Validation ผลการทดลองพบว่า ชุดข้อมูลของวิชา ฟิสิกส์ 1 แบบ 2 ระดับ มีความถูกต้องสูงสุดเป็น 79.05% และเมื่อใช้ชุดข้อมูลคนละวิชา ที่มีระดับความมั่นใจเท่ากัน ตรวจสอบเพื่อหา

ค่าตัวแบบที่ดีที่สุด พบว่า ชุดข้อมูลที่รวมทั้งฟิสิกส์ 1 และฟิสิกส์ 3 แบบ 2 ระดับ มีค่าเฉลี่ยความถูกต้องสูงที่สุด คือ 78.28% ซึ่งอยู่ในระดับดี และสามารถที่จะใช้ตัวแบบดังกล่าวทั้งสองวิชาได้

ภัทรพงศ์ พงศ์ภัทรกานต์,(2553) งานวิจัยนี้เป็นการนำเสนอการวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาระดับปริญญาตรีโดยใช้คอมมิตตีแมชชีน ภายใต้หลักการทำงานของเหมืองข้อมูล โดยใช้ชุด ข้อมูลนักศึกษาที่เข้าศึกษาระหว่างปี พ.ศ. 2546-2549 ของมหาวิทยาลัยราชภัฏเลย มีจำนวน 11 แอททิรบิวต์ และ 12,865 ชุดข้อมูล ซึ่งคอมมิตตีแมชชีนเป็นการทำงานระหว่าง SVM ร่วมกับ C5.0 โดยได้ทำการทดลองวัดประสิทธิภาพความถูกต้องเปรียบเทียบกับนิวรอลเน็ตเว็ร์กและ C5.0 ซึ่งแบบจำลองแบบคอมมิตตีแมชชีน มีประสิทธิภาพในการจำแนกข้อมูลสูงที่สุด มีค่าเท่ากับ 75.32 และได้ทำการวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาระดับปริญญาตรี ผลการทดลองพบว่ามีปัจจัยที่มีความสำคัญ จำนวน 3 ปัจจัย คือ ขนาดโรงเรียนเดิม, อาชีพของบิดา และอาชีพของมารดา และสามารถนำมาเขียนกฎการจำแนกสร้างความสัมพันธ์ของแอททิรบิวต์ ซึ่งเทคนิคนี้สามารถ ประยุกต์ใช้กับการวิเคราะห์จำแนกข้อมูลที่มีความสัมพันธ์กับข้อมูลอื่นๆ ได้เป็นอย่างดี

รินทร์ลภัส จินานุศิลประสาท.(2553)ในการดำเนินธุรกิจขององค์กรไม่ว่าเป็นธุรกิจใดก็ตามจุดหมายสูงสุดขององค์กรคือ การได้เป็นผู้นำของธุรกิจซึ่งปัจจุบันนี้ธุรกิจต่าง ๆ ล่วงอยู่ในยุดของการเติบโตด้านเทคโนโลยีและมีสภาวะการแข่งขันสูงอีกปัจจัยหนึ่งที่ทำให้ธุรกิจบรรลุเป้าหมายคือ สามารถวิเคราะห์พฤติกรรมของลูกค้ากลุ่มเป้าหมายได้ถูกต้องแม่นยำ ก็จะมีข้อได้เปรียบเหนือกว่าคู่แข่ง สำหรับงานวิจัยนี้ใช้เทคนิคการวิเคราะห์หลักเกณฑ์การเชื่อมโยงโดยใช้ขั้นตอนวิธี Apriori ช่วยในการวิเคราะห์ความสัมพันธ์และพฤติกรรมที่ซ่อนอยู่ซึ่งสรุป พบว่าลูกค้าที่ซื้อประกันภัย TA, 10/7 , 80/20 , 5/5 ,Vol3, Fire และ 20/20 มีพฤติกรรมซื้อประกันภัยแบบชำระเบี้ยครั้งเดียว โดยมีช่วงความเชื่อมั่นมีค่าเท่ากับ 0.9 จากการวิเคราะห์สามารถนำมาเป็นข้อมูลสนับสนุนการตัดสินใจในการวางแผนการตลาด และวางกลยุทธ์ลูกค้าสัมพันธ์

วีระ จิรกิจอนุสรณ์ (2553) ภาษีมูลค่าเพิ่มในแหล่งรายได้ของภาครัฐที่นำมาจัดทำงบประมาณในการพัฒนาประเทศ ประสิทธิผลของการพยากรณ์รายได้ภาษีมูลค่าเพิ่มจะช่วยให้รัฐสามารถวางแผนงบประมาณอย่างมีประสิทธิภาพเทคนิคของเหมืองข้อมูลได้ถูกนำมาใช้งานประยุกต์หลายแขนงรวมทั้งการศึกษาวิจัยคาดการณ์ภาษีมูลค่าเพิ่ม อย่างไรก็ตามปัจจัยหรือตัวแปรหรือตัวแปรที่ใช้ในการศึกษายังไม่ครอบคลุมทั้งหมดงานวิจัยฉบับนี้ศึกษาเพิ่มเติมปัจจัย อื่น

ๆ ที่อาจจะส่งผลต่อการคาดการณ์ และงานวิจัยนี้สร้างแบบจำลองด้วยเงื่อนไขที่หลากหลาย จากการศึกษาวิจัยพบว่าแบบจำลอง Multilayer perceptron ที่ใช้ตัวแปร ค่าดัชนีราคาผู้บริโภค ดัชนีสินค้านำเข้า ดัชนีการอุปโภคบริโภคภาคเอกชน ดัชนีการลงทุนภาคเอกชน รายจ่ายภาครัฐ จะให้ค่าการคาดการณ์ผลการจัดเก็บภาษีมูลค่าเพิ่ม ที่มีความผิดพลาดในการคาดการณ์เฉลี่ยคิดเป็น 98% ในขั้นตอนการเตรียมข้อมูล ข้อมูลที่ใช้ในงานวิจัยนี้จัดเก็บเป็นรายเดือนตั้งแต่ปี 2544 ถึงปี 2552 จัดเก็บในรูปแบบไฟล์ csv ตัวแปรของงานวิจัยมาจากหลายแหล่งข้อมูล ในศาสตร์ของการทำเหมืองข้อมูลการหาความสัมพันธ์ของตัวแปรแต่ละตัวแปรมีผลต่อการจัดเก็บข้อมูลโดยใช้ information gain และ gain ratio เพื่อจะคัดเลือกตัวแปรไปทดลองแบบจำลอง โดยแบ่งข้อมูลเป็นช่วงข้อมูลของแต่ละตัวแปร เป็นจำนวน 10 ช่วง แล้วนำไปเข้าคำนวณโปรแกรม WEKA เพื่อหาความสัมพันธ์ของข้อมูล

วรากุล กาญจนกัญโห (2553) งานวิจัยนี้ใช้ข้อมูล Data Set ด้านสังคมและเศรษฐกิจ เมื่อ ปี ค.ศ.1990 ข้อมูลด้านสถิติที่เกี่ยวข้องกับการบังคับใช้ กฎหมายจากการสำรวจของ LEMAS เมื่อปี ค.ศ.1990 และ จากรายงานข้อมูลด้านอาชญากรรมของ FBI UCR เมื่อปี ค.ศ.1995 เพื่อใช้ในการวิเคราะห์และคัดเลือกปัจจัยสำคัญที่เป็นสาเหตุของการเกิดอาชญากรรม โดยใช้หลักการของ Generalized Regression Neural Network (GRNN) และหาประสิทธิภาพของการพยากรณ์โดยแบบจำลองโครงข่าย ประสาทเทียมแบบแพร่กระจายย้อนกลับ (Back Propagation Algorithm) ผลการวิจัยพบว่าวิธีเชิงพันธุกรรม (Genetic Algorithm) มีประสิทธิภาพในการคัดเลือกปัจจัยโดยให้ค่าความเหมาะสมดีที่สุด (Best Fitness) และมีปัจจัยทั้งหมด 8ปัจจัยสำคัญที่มีอิทธิพลต่อการเกิดอาชญากรรม ซึ่งสามารถอธิบายการเกิดอาชญากรรมได้ประมาณ 80.68% (R Square=0.8060) และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAE)เท่ากับ 0.0908

บทที่ 3
วิธีดำเนินการวิจัย

ในรายงานวิจัยในครั้งนี้มีจุดมุ่งหมาย เพื่อศึกษาลักษณะทั่วไปของนักศึกษาที่พ้นสภาพการเป็นนักศึกษาและศึกษาว่าปัจจัยด้านเศรษฐกิจ ด้านการเรียน ด้านพื้นฐานความรู้จากแผนการเรียนรู้เดิมที่นักศึกษาได้ศึกษามา และการเลือกสาขาวิชาในระดับปริญญาตรี มีผลต่อการพ้นสภาพนักศึกษา ในการวิจัยครั้งนี้ กลุ่มตัวอย่างเป็นนักศึกษาที่พ้นสภาพเป็นนักศึกษาในปีการศึกษา พ.ศ. 2550 -2553

3.1 กฎระเบียบการพ้นสภาพการเป็นนักศึกษา

3.2 ขั้นตอนการวิเคราะห์โดยใช้เหมืองข้อมูล (Data mining)

3.3 ขั้นตอนการวิเคราะห์โดยใช้ เครื่องมือในงานวิจัย

3.1 กฎระเบียบการพ้นสภาพการเป็นนักศึกษา

ตามระเบียบการพ้นสภาพการเป็นนักศึกษาได้ต้องเป็นไปตามเกณฑ์ที่มหาวิทยาลัยกำหนด คือ

1. นักศึกษาที่ได้รับค่าคะแนนเฉลี่ยสะสมต่ำกว่า 1.60 ในภาคการศึกษาปกติที่สองที่เข้าศึกษาในมหาวิทยาลัย(ไม่นับภาคเรียนที่ลาพักหรือถูกให้พัก)

2. นักศึกษาที่ได้รับค่าระดับคะแนนเฉลี่ยสะสมต่ำกว่า 1.70 ในภาคการศึกษาปกติถัดไปหลังจากได้รับการรอพินิจครั้งที่ 1

3. นักศึกษาที่ได้รับค่าระดับคะแนนเฉลี่ยสะสมต่ำกว่า 1.80 ในภาคเรียนปกติถัดไปหลักจากได้รับการรอพินิจครั้งที่ 2

4. ใช้เวลาการศึกษาเกิน 8 ปีการศึกษา สำหรับการลงทะเบียนเรียนเต็มเวลาและเกิน 12 ปีการศึกษา สำหรับการลงทะเบียนเรียนไม่เต็มเวลา ในหลักสูตร 4 ปี ใช้เวลาการศึกษาเกิน 10 ปีการศึกษาสำหรับการลงทะเบียนเรียนเต็มเวลา และเกิน 15 ปีการศึกษา สำหรับการลงทะเบียนเรียนไม่เต็มเวลาในกรณีหลักสูตร 5 ปี

5. นักศึกษาลงทะเบียนครบตามหลักสูตรกำหนดแต่ยังได้รับคะแนนเฉลี่ยสะสม ต่ำกว่า 1.80

6. กระทำการทุจริต หรือมีความประพฤติอันเป็นการเสื่อมเสียแก่มหาวิทยาลัยและมหาวิทยาลัยเห็นสมควรให้ออกหรือไล่ออกตามข้อบังคับของมหาวิทยาลัยว่าด้วยวินัยนักศึกษา

3.2 ขั้นตอนการวิเคราะห์โดยใช้เหมืองข้อมูล (Data mining)

งานวิจัยนี้นี้มุ้งเน้นในส่วนของนักศึกษาที่พ้นสภาพการเป็นนักศึกษา แต่ในกระบวนการประมวลผลนำข้อมูลนักศึกษาที่ได้เข้าศึกษาในปีการศึกษาเดียวกันมาเปรียบเทียบ ปัจจัย โดยจำนวนนักศึกษาที่เข้าศึกษา ในปีการศึกษา 2549-2552 จำนวนทั้งสิ้น 4,528 คน นำข้อมูลทั้งหมดโดยการคัดแยกอย่างง่าย ใช้ข้อมูลนักศึกษาพ้นสภาพในระดับปริญญาตรี ระหว่างปีการศึกษา พ.ศ. 2550 - 2553 นักศึกษาพ้นสภาพจำนวน 1,804 คน และนักศึกษาที่ยังศึกษาต่อ จำนวน 2,724 คน โดยมีกระบวนการทำงานดังนี้

3.2.1 การเตรียมข้อมูล ขั้นตอนการเตรียมการข้อมูลจะมีการระบุชุดข้อมูลหลักที่ใช้โดยเหมืองข้อมูล และจัดเรียงให้เรียบร้อย เนื่องจากข้อมูลในคลังข้อมูลงานทะเบียนนักศึกษาจากกองบริการการศึกษา ได้ถูกคัดกรองและเปลี่ยนรูปแบบของข้อมูลเพื่อดำเนินการของเหมืองข้อมูล การจำแนกและวิเคราะห์ข้อมูลจะศึกษาข้อมูลเพื่อระบุลักษณะหรือรูปแบบของข้อมูลทั่วไปในขั้นตอนนี้โดยงานวิจัยจะสนใจ ปัจจัย คณะที่เรียน สาขาที่เรียน อายุนักศึกษา วุฒิการศึกษาเดิม เพศของนักศึกษา อาชีพของบิดา-มารดา สถานภาพของบิดา-มารดา ได้ข้อมูลนักศึกษาพ้นสภาพในช่วงปีการศึกษา 2550 – 2553 จำนวนนักศึกษา 1804 คน

ตารางที่ 3.1 ตัวอย่างข้อมูลประวัตินักศึกษาที่พ้นสภาพ

STUDENT CODE	PREFIX NAME	STUDENT NAME	STUDENT SURNAME	PROGRAMNAME	FINISHDATE	SCHOOLID	SCHOOLNAME	courses	Grade
404865001	นางสาว	อุไร	สุดจิตร	เทคโนโลยีสารสนเทศ	21/6/2007	1133	ดีบุกพังงาวิทยายน	Science Curriculum - Mathematics.	2.25
404865036	นางสาว	ศิรินันท์	เหมนะ	เทคโนโลยีสารสนเทศ	21/6/2007	1826	วิทยาลัยเทคนิคพิจิตร	Science Curriculum - Mathematics.	2.1
405052050	นางสาว	บีบัส	สาและ	ชีววิทยาประยุกต์	3/7/2007			Science Curriculum - Mathematics.	2
404655009	นางสาว	ชื่นจิตต์	สงวนวงศ์ภักดี	การบริหารธุรกิจ (แขนงวิชาการตลาด)	10/7/2007			Arts curriculum	2.32
404903038	นาย	ดารซี	หะมะ	สังคมศึกษา	15/7/2007			Arts curriculum	2.25
404906009	นางสาว	นูรีมัง	วาเต๊ะ	ภาษาอังกฤษ	16/7/2007	990000528	โรงเรียนธรรมศาสน์วิทยา	Science Curriculum - Mathematics.	2.33
404906038	นางสาว	นูรฮาซูรา	อาแด	ภาษาอังกฤษ	16/7/2007			Science Curriculum - Mathematics.	2.1

จากตารางที่ 3.1 เป็นตัวอย่างข้อมูลประวัติ ของนักศึกษา เช่น รหัสประจำตัว ชื่อ เพศ สัญชาติ ที่อยู่ หลักสูตรที่เรียนระดับ ผลการเรียนระดับมัธยม สาขาวิชาที่นักศึกษา ศึกษาอยู่ ฯลฯ เมื่อเราได้ข้อมูลทั้งหมดแล้ว ขั้นต่อมาก็คือ การเตรียมข้อมูลเพื่อให้พร้อมที่จะนำไปทำเหมืองข้อมูล ซึ่งแบ่งเป็นขั้นต่างๆ ได้ดังนี้

3.2.1.1 การกลั่นกรองข้อมูล เป็นการนำข้อมูลที่ไม่มีค่า ข้อมูลที่ขาดหาย และข้อมูลที่ไม่แน่นอนออกไป ข้อมูลที่ได้มานั้น เป็นข้อมูลที่ยังไม่สมบูรณ์ ที่จะสามารถนำไปใช้ผ่านกระบวนการเหมืองข้อมูลได้ จึงต้องมีการจัดการข้อมูล การเตรียมข้อมูลเบื้องต้น มีวิธีการดังนี้

เลือกเฉพาะคอลัมน์สำคัญที่คาดว่าจะสามารถนำมาใช้ประโยชน์ได้ และเป็นคอลัมน์ที่มีข้อมูลค่อนข้างครบถ้วน เช่น จากใน ตารางที่ 1 คอลัมน์สำคัญที่มีข้อมูลค่อนข้างมาก ได้แก่ ข้อมูลรหัสนักศึกษา ที่อยู่ อายุ เพศ โรงเรียนเดิม หลักสูตรโรงเรียนเดิม เกรดเฉลี่ยที่จบการศึกษาในโรงเรียนเดิม

สำหรับคอลัมน์ที่มีค่าสำหรับทุกแถวเป็นค่าเดียวกัน เช่น วันที่จบการศึกษา จะเป็นข้อมูลที่ไม่สามารถแยกความแตกต่างของแต่ละแถวได้เลย ดังนั้นในการทำเหมืองข้อมูลจะไม่สามรถใช้ประโยชน์จากคอลัมน์นี้ ดังนั้น จึงไม่นำคอลัมน์นี้มาพิจารณา

แก้ไขข้อมูลให้ถูกต้องสมบูรณ์ ได้แก่ การแก้ไขค่าว่างของข้อมูล ซึ่งสามารถแก้ไขได้หลายวิธี เช่น แก้ไขโดยจำกัดข้อมูลที่ในแถวเป็นค่าว่าง (NULL) ข้อมูลบางแถวค่าในคอลัมน์ ชื่อโรงเรียนเดิมหายไป ซึ่งจะเห็นได้ว่าถ้ามีแต่รหัสนักศึกษา และวิชาที่นักศึกษาศึกษา โดยที่ไม่มีข้อมูลโรงเรียนเดิม แล้ว เราก็ไม่สามารถจะนำแถวนั้นพิจารณาเพื่อหาความสัมพันธ์ที่น่าสนใจได้

จากตารางที่ 3.1 ที่เป็นข้อมูลประวัตินักศึกษา เราได้นำมาปรับเปลี่ยนข้อมูลบางส่วนเพื่อให้สมบูรณ์ขึ้นได้แก่

- การตัดคอลัมน์ที่ไม่จำเป็นในการทำเหมืองข้อมูลออก เช่น คอลัมน์ชื่อ-สกุล นักศึกษา เพราะ ชื่อนิสิตแต่ละคนไม่สามารถนำมาทำเหมืองข้อมูลได้
- คัดเลือกเฉพาะคอลัมน์ที่คาดว่าจะสามารถนำมาทำเหมืองข้อมูลได้ เช่น คัดเลือกคอลัมน์โรงเรียนเดิม หลักสูตรที่นักศึกษาเรียนในระดับมัธยม สาขาที่นักศึกษาเลือกศึกษา แต่เนื่องจากชื่อโรงเรียนของนักศึกษาจะเป็นคอลัมน์ที่ไม่สมบรูณ์จะต้องกำจัดตามหลักการเหมืองข้อมูล ผลที่ได้จากการทำข้อมูลจากตารางที่ 3.1 มีการกำจัดข้อมูลที่ไม่สมบรูณ์ออกไปดังแสดงตารางที่ 3.2

ตารางที่ 3.2 ตัวอย่างข้อมูลประวัตินักศึกษาที่ทำให้สมบูรณ์

PROGRAMNAME	SCHOOLNAME	Course	Grade
IT	one	Science Curriculum - Mathematics.	2.25
IT	two	Science Curriculum - Mathematics.	2.1
Biological applications	three	Science Curriculum - Mathematics.	2
Field Marketing	three	Arts curriculum	2.32
Social studies	three	Arts curriculum	2.25
English language	four	Science Curriculum - Mathematics.	2.33
English language	three	Science Curriculum - Mathematics.	2.1
Science of administration	three	Arts curriculum	2.22

จากตารางที่ 3.2 ที่เป็นตารางข้อมูลการลงทะเบียนเรียนของนิสิต เราได้ปรับข้อมูลบางส่วนให้สมบูรณ์ขึ้นได้แก่

- การตัดบางคอลัมน์ที่ไม่น่าสนใจที่จะนำมาทำเหมืองข้อมูลออก เช่น คอลัมน์โรงเรียนเดิม

3.2.1.2 การคัดเลือกข้อมูล เป็นการเลือกเฉพาะข้อมูลที่ต้องการนำมาวิเคราะห์ในการทำเหมืองข้อมูล เราจำเป็นต้องคัดเลือกเฉพาะข้อมูลนักศึกษาที่สามารถนำมาใช้ประโยชน์ได้ คัดเลือกข้อมูลนักศึกษา เฉพาะสาขาที่นักศึกษาเลือกเข้าศึกษา กับ หลักสูตรที่นักศึกษาได้ศึกษามาในระดับมัธยมศึกษาตอนปลาย เนื่องมาจากถ้าข้อมูลที่เราได้มานั้นย้อนหลังไปถึง 3 ปี จากข้อมูลดังกล่าวมีความสัมพันธ์กันเบื้องต้น หลังจากที่ทำตามขั้นตอนข้างต้นทั้งหมดแล้วจะได้ข้อมูลที่มีความสมบูรณ์มากขึ้น

3.2.1.3 การแปลงรูปแบบข้อมูล เป็นการแปลงข้อมูลที่เลือกมาให้อยู่ในรูปแบบที่เหมาะสมสำหรับการนำไปใช้วิเคราะห์ตามขั้นตอนวิธีที่ใช้ในการทำเหมืองข้อมูลต่อไป จากตารางที่ 3.2 จะเห็นได้ว่าข้อมูลที่เลือกสามารถตอบเป้าหมายที่ต้องการจะศึกษา คือปัจจัยที่ส่งผลให้นักศึกษาพ้นสภาพการเป็นนักศึกษา เพื่อที่จะได้นำตารางนี้ไปปรับเปลี่ยนเพื่อให้เหมาะสมกับเทคนิคต่าง ๆ ของเหมืองข้อมูลต่อไป ผลลัพธ์ที่ได้ทั้งหมดแสดงได้ดังตารางที่ 3.3

ตารางที่ 3.3 ตัวอย่างตารางข้อมูลนักศึกษาขั้นต้น

PROGRAMNAME	Course
IT	Science Curriculum - Mathematics.
IT	Science Curriculum - Mathematics.
Biological applications	Science Curriculum - Mathematics.
Field Marketing	Arts curriculum
Social studies	Arts curriculum
English language	Science Curriculum - Mathematics.
English language	Science Curriculum - Mathematics.

จากข้อมูลในตารางที่ 3.3 นี้ถือได้ว่าเป็นข้อมูลเบื้องต้นในรูปแบบสมบูรณ์ที่พร้อมจะนำไปทำเหมืองข้อมูลแล้ว แต่เราอาจต้องปรับเปลี่ยนรูปแบบของข้อมูลเพื่อให้เหมาะสมกับแต่ละเทคนิคของเหมืองข้อมูลที่จะเลือกใช้

3.2.1.4. การทำเหมืองข้อมูลเป็นการใช้เทคนิคภายในการทำเหมืองข้อมูล โดยทั่วไปประเภทของงานตามลักษณะของแบบจำลองที่ใช้ในการทำเหมืองข้อมูล คือ แบบจำลองเชิงทำนาย และแบบจำลองเชิงพรรณนา สำหรับงานวิจัยนี้จะนำแบบจำลองเชิงทำนาย มาวิเคราะห์ ผู้วิจัยใช้โปรแกรม WEKA โดยเลือกเทคนิค Data Classification และ Association Rule โดยที่ ผู้วิจัยทำการจำแนกประเภทข้อมูล เพื่อการสร้างโมเดลจำแนกประเภทข้อมูล เพื่อทำนายกลุ่มของข้อมูลใหม่เป็นขั้นตอนการนำโมเดลจำแนกประเภทที่สร้างขึ้นมาใช้กับข้อมูลที่ไม่เคยเห็นมาก่อน เพื่อทำนายและกำหนดกลุ่มให้กับข้อมูลนั้น

5. การประเมินรูปแบบ เป็นขั้นตอนการเลือกรูปแบบที่ยืนยันสมมติฐานที่มีเหตุผลว่ามีความเหมาะสมหรือตรงกับวัตถุประสงค์ที่ต้องการหรือไม่ การแปลความหมายและการประเมินผลลัพธ์ที่ได้ โดยทั่วไปควรมีการแสดงผลในรูปแบบที่สามารถเข้าใจได้โดยง่าย

3.3 ขั้นตอนการวิเคราะห์โดยใช้ เครื่องมือในงานวิจัย

ขั้นตอนการหาปัจจัยในการพ้นสภาพของนักศึกษา มาวิเคราะห์ความสัมพันธ์กับนักศึกษาที่สำเร็จการศึกษา

@attribute Programname

@attribute Course

วิธีการหาปัจจัย

1. นำข้อมูลที่บันทึกในโปรแกรมชนิด Excel (.xls) แปลงเป็นโปรแกรมชนิด CSV(*.csv)

2. นำข้อมูลที่อยู่ในโปรแกรมชนิด CSV (*.csv) มาเข้าโปรแกรม WEKA จากนั้นบันทึกเป็นโปรแกรมชนิด ARFF (.arff)

ในการวิเคราะห์จำเป็นจะต้องเปลี่ยนไฟล์ข้อมูลให้เป็น CSV(comma delimited) เพื่อที่จะนำข้อมูลที่ได้มาประมวลผลสำหรับงานวิจัยนี้ใช้ WEKA เป็นซอฟต์แวร์ด้านการทำเหมืองข้อมูลที่ได้รับการยอมรับและแพร่หลายในต่างประเทศรวมทั้งประเทศไทยด้วย ซึ่งเห็นได้จากงานวิจัยหลายเรื่องที่นำเอาโปรแกรม WEKA มาใช้ในการพัฒนารูปแบบจำลองต่างๆ สาเหตุอีกประการหนึ่งที่ทำให้ได้รับความนิยมก็คือเป็นซอฟต์แวร์เปิด (Open Source Software) ที่อยู่ภายใต้การควบคุมของ GPL License เป็นโปรแกรมที่ได้พัฒนาขึ้นบนพื้นฐานของภาษาจาวา สามารถรันได้หลายระบบปฏิบัติการ และสามารถพัฒนาต่อยอดโปรแกรมได้ เป็นเครื่องมือที่ใช้ทำงานในด้านการทำเหมืองข้อมูลที่รวบรวมแนวคิดขั้นตอนวิธีมากมาย ซึ่งขั้นตอนวิธีสามารถเลือกใช้งานโดยตรงได้จาก 2 ทางคือจากชุดเครื่องมือที่มีขั้นตอนวิธีมาให้ หรือเลือกใช้จากขั้นตอนวิธีที่ได้เขียนเป็นโปรแกรมลงไปเป็นชุดเครื่องมือเพิ่มเติม และชุดเครื่องมือมีฟังก์ชันสำหรับการทำงานร่วมกับข้อมูล

ข้อดีของโปรแกรม WEKA คือ มีขั้นตอนวิธีที่รู้จักกันดีของการทำเหมืองข้อมูลให้เลือกใช้อย่างครบถ้วน สามารถเขียนฟังก์ชันเพิ่มเข้าไปในโปรแกรมได้เอง และสามารถนำมาปรับใช้กับข้อมูลที่ผู้วิจัยนำมาใช้ในการทำวิจัยในครั้งนี้คือ ระบบฐานจากกองบริการการศึกษา ส่วนของข้อมูลนักศึกษา มหาวิทยาลัยราชภัฏยะลา

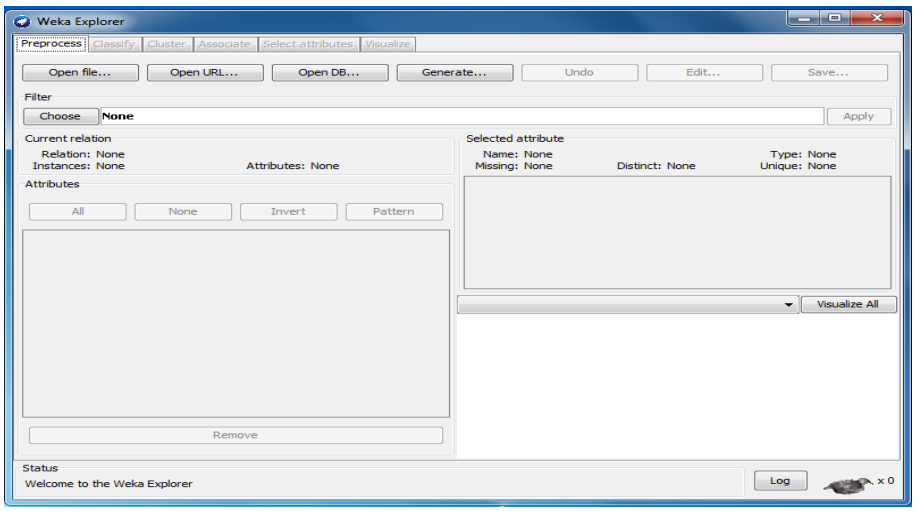
3.3.1 ขั้นตอนการทำเหมืองข้อมูล (Data Mining)

ขั้นตอนการทำเหมืองข้อมูลโดยแบ่งกลุ่มมีดังนี้

1. เปิดโปรแกรม WEKA

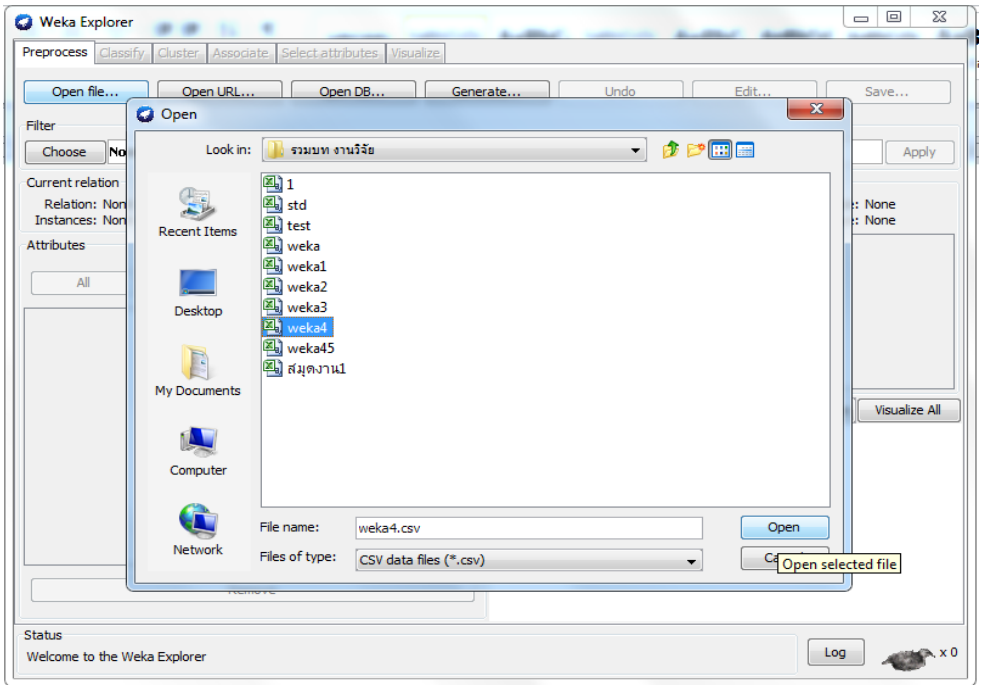
ภาพที่ 3.1 โปรแกรม WEKA

2. เปิดโมดูล Explorer ของโปรแกรม WEKA



ภาพที่ 3.2 โมดูล Explorer ของโปรแกรม WEKA

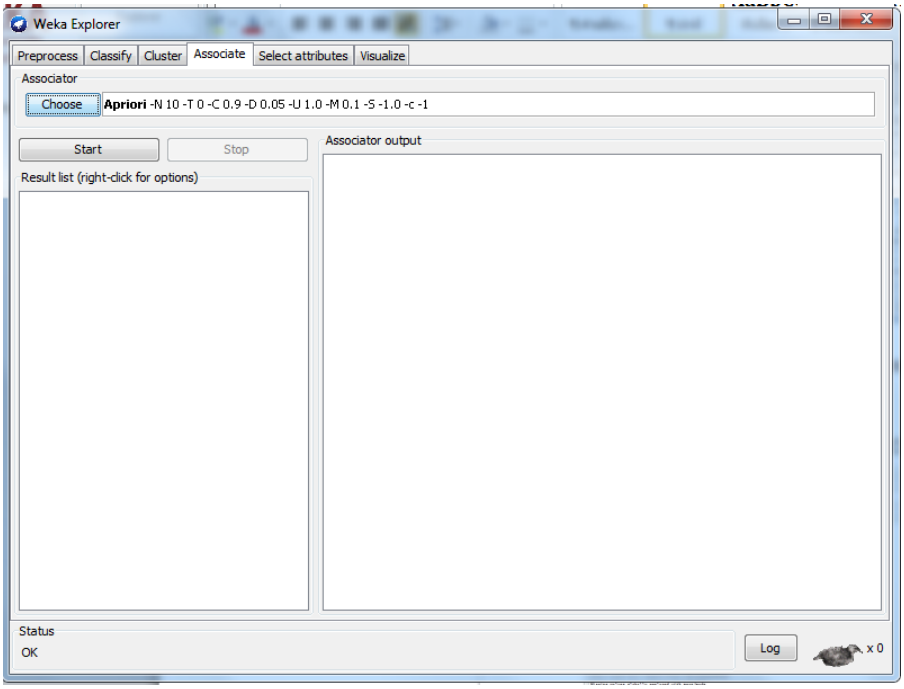
3. เปิดไฟล์ข้อมูลที่ต้องการวิเคราะห์



ภาพที่ 3.3 ไฟล์ข้อมูลที่ต้องการวิเคราะห์

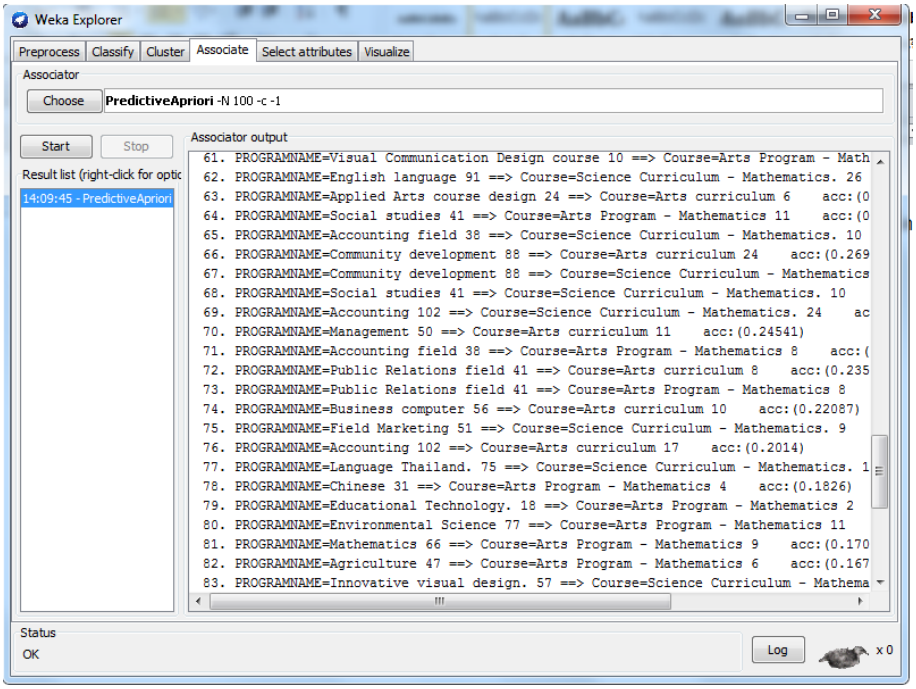
ขั้นตอนการทำเหมืองข้อมูลโดยการสร้างกฎความสัมพันธ์มีดังนี้

1. นำชุดข้อมูลปัจจัยที่ผลต่อการวิเคราะห์ และสำหรับสอนเพื่อสร้างกฎความสัมพันธ์โดยที่สูตรที่ศึกษา
2. เลือกเทคนิค associations rule เลือก ขั้นตอนวิธี PredictiveApriori



ภาพที่ 3.4 เลือกเทคนิค associations

4. เลือกเทคนิค associations rule เลือก ขั้นตอนวิธี PredictiveApriori



ภาพที่ 3.5 ผลจากการวิเคราะห์ เทคนิค associations ขั้นตอนวิธี PredictiveApriori

เครื่องมือในงานวิจัย คือโปรแกรม SPSS และ โปรแกรม WEKA

โปรแกรม SPSS เป็นโปรแกรมสำเร็จรูปทางสถิติใช้ในการวิเคราะห์ข้อมูล ในการทำวิจัยในครั้งนี้ เพื่อดูระดับค่าข้อมูลที่ได้จะการวิเคราะห์ความถูกต้อง

โปรแกรม WEKA เป็น Open Source สามารถที่จะ ดาวน์โหลดได้จากอินเทอร์เน็ต ไม่มีปัญหาเรื่องของลิขสิทธิ์ ใช้วิเคราะห์ผลได้ถูกต้องตามกฎหมาย มีเทคนิคที่เหมาะกับงานวิจัยนี้ Data classification และ Association Rules และใช้เทคนิคต้นไม้แห่งการตัดสินใจ จะวิเคราะห์ข้อมูลที่ได้เป็นกฎ หรือโมเดลที่สามารถนำไปใช้ในการตัดสินใจ ที่ใช้ในงานวิจัย

ทำการวัดประสิทธิภาพตัวแบบที่ได้ด้วยความถูกต้อง ค่าความแม่นยำ ค่าความระลึก ค่า F-Measurement ของการจัดกลุ่มของข้อมูล

สถิติที่ใช้ในการวิเคราะห์ข้อมูล

ค่าสัมบรูณ์ของค่าความคลาดเคลื่อนเฉลี่ย (Mean Adsolute Error : MAE) การวัดความคลาดเคลื่อนของค่าจริงและค่าที่พยากรณ์ได้โดยใช้ค่าสัมประสิทธิ์ต่าง ๆ หรือจำนวนข้อมูลต่าง ๆ จะพิจารณาจากการที่ค่าจริงใกล้เคียงค่าพยากรณ์ที่สุด หรือทำให้เกิดความคลาดเคลื่อนน้อยที่สุด ย่อมเป็นค่าที่เหมาะสมกับการใช้พยากรณ์ให้ได้ผลลัพธ์ที่แม่นยำ การวัดความความคลาดเคลื่อนสามารถวัดได้จากค่าต่าง ๆดังนี้ไปนี้

บทที่ 4
ผลการดำเนินงาน

งานวิจัยการวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษา ระดับปริญญาตรี มีกระบวนการในการทำเหมืองข้อมูล ทั้งหมด 6 ขั้นตอน ต่อไปนี้

4.1 การกลั่นกรองข้อมูล (Data Cleaning)

4.2 การรวบรวมข้อมูล (Data Integration)

4.3 การคัดเลือกข้อมูล (Data Selection)

4.4 การแปลงรูปแบบข้อมูล (Data Transformation)

4.5 การทำเหมือนข้อมูล (Data Mining)

4.6 การประเมินรูปแบบ (Pattern Evaluation)

4.1 การกลั่นกรองข้อมูล

เป็นการนำข้อมูลที่ไม่สมบรูณ์ หรือข้อมูลที่ไม่ทราบค่า ออกไปจากการประมวลผล เก็บรวบรวมข้อมูลนักศึกษาพ้นสภาพในระดับปริญญาตรี ระหว่างปีการศึกษา พ.ศ. 2550 - 2553 จำนวนนักศึกษา 1804 คน โดยแบ่งรายละเอียดการเก็บข้อมูล ดังนี้

ตารางที่ 4.1 แบบฟอร์มการจัดเก็บข้อมูล

STUDENTCODE	PREFIXNAME	STUDENTNAME	STUDENTSURNAME	PROGRAMNAME	SCHOOLNAME	COURSE	GRADE

4.2 การรวบรวมข้อมูล

ผู้วิจัยจำเป็นต้องคัดเลือกเฉพาะข้อมูลนักศึกษาที่พ้นสภาพการเป็นนักศึกษา โดยการคัดเลือกข้อมูลจะมีการคัดเลือกข้อมูลนักศึกษาที่เข้าศึกษาในระดับปริญญาตรี ในช่วง ปีการศึกษา 2549-2552 จำนวนทั้งสิ้น 4528 คน นำข้อมูลทั้งหมดโดยการคัดแยกอย่างง่าย ใช้ข้อมูลนักศึกษาพ้นสภาพในระดับปริญญาตรี ระหว่างปีการศึกษา พ.ศ. 2550 -2553 จำนวนนักศึกษา 1804 คน โดยมีกระบวนการทำงาน ดังนี้

	B	C	D	E	L	M	N	O	P
1	PREFIXNAME	STUDENTNAME	STUDENTSURNAME	PROGRAMNAME	FINISHDATE	SCHOOLID	SCHOOLNAME	courses	Grade
2	นางสาว	อุไร	สุดจิตร	เทคโนโลยีสารสนเทศ	21/6/2007	1133	ศึกษาพิงวิทยายน	Science Curriculum - Mathematics.	2.25
3	นางสาว	ศิริพันท์	เหมมะ	เทคโนโลยีสารสนเทศ	21/6/2007	1826	วิทยาลัยเทคนิคพิจิตร	Science Curriculum - Mathematics.	2.1
4	นางสาว	บิบัส	สาและ	ชั่ววิทยาประยุกต์	3/7/2007			Arts Program - Mathematics	2
5	นางสาว	ชื่นจิตต์	สงวนวงศ์ภัคดี	การบริหารธุรกิจ(แขนงวิชาการตลาด)	10/7/2007			Arts curriculum	2.32
6	นาย	ดารชี	ทะมะ	สังคมศึกษา	15/7/2007			Science Curriculum - Mathematics.	2.25
7	นางสาว	นูริมัง	วาเต๊ะ	ภาษาอังกฤษ	16/7/2007	990000528	โรงเรียนธรรมศาสน์วิทยา	Science Curriculum - Mathematics.	2.33
8	นางสาว	นูรฮาธูรา	อาแด	ภาษาอังกฤษ	16/7/2007			Science Curriculum - Mathematics.	2.1
9	นาย	ศราวุธ	หนูประกอบ	รัฐประศาสนศาสตร์	16/7/2007			Science Curriculum - Mathematics.	2.22
10	นาย	อับดุลอาซิ	คดเฉาะ	รัฐประศาสนศาสตร์	16/7/2007			Science Curriculum - Mathematics.	2.23
11	นาย	อาฟานดี	มะเก๊ะ	รัฐประศาสนศาสตร์	16/7/2007			Science Curriculum - Mathematics.	2.29
12	นาย	พัลลภ	เจะอามะ	การพัฒนาชุมชน	16/7/2007			Science Curriculum - Mathematics.	2.45
13	นาย	อันวา	สาและ	การพัฒนาชุมชน	16/7/2007			Arts Program - Mathematics	2.55
14	นาย	โมฮำหมัดสุไลมาน	มะแซ	นิเทศศาสตร์(แขนงวิชาการประชาสัมพันธ์)	16/7/2007	0	มหาวิทยาลัยราชภัฏยะลา	Arts Program - Mathematics	2.56
15	นาย	สุกรี	สะแลแม	นิเทศศาสตร์(แขนงวิชาการประชาสัมพันธ์)	16/7/2007			Arts Program - Mathematics	2.51
16	นางสาว	นาธีนี๊ะ	นิแห	นิเทศศาสตร์(แขนงวิชาการประชาสัมพันธ์)	16/7/2007			Arts Program - Mathematics	2.25
17	นางสาว	ฟาอีซ๊ะ	ยะพา	นิเทศศาสตร์(แขนงวิชาการประชาสัมพันธ์)	16/7/2007			Arts Program - Mathematics	2.1
18	นาย	แวอุเซ็ง	วาเลาะ	ศิลปกรรม(แขนงออกแบบนิเทศศิลป์)	16/7/2007			Arts Program - Mathematics	2
19	นาย	ดุรณีชิ	มะลาแง	ศิลปกรรม(แขนงออกแบบประยุกต์ศิลป์)	16/7/2007			Arts Program - Mathematics	2.32
20	นาย	รอยาลี	ภาหมีเ๊ะ	สุขศึกษา	16/7/2007	990000528	โรงเรียนธรรมศาสน์วิทยา	Arts Program - Mathematics	2.25
21	นาย	นูร์ดิน	มามะ	สุขศึกษา	16/7/2007	990000528	โรงเรียนธรรมศาสน์วิทยา	Arts Program - Mathematics	2.33
22	นาย	สมศักดิ์	อาด๊ำ	สุขศึกษา	16/7/2007	990000528	โรงเรียนธรรมศาสน์วิทยา	Arts Program - Mathematics	2.1
23	นาย	อดุลย์	แมเราะ	เกษตรศาสตร์	16/7/2007	990000528	โรงเรียนธรรมศาสน์วิทยา	Arts Program - Mathematics	2.22
24	นาย	อาดิส	มะลิ	เกษตรศาสตร์	16/7/2007			Arts Program - Mathematics	2.23
25	นาย	มะอ๊ารี	สะตะระ	เกษตรศาสตร์	16/7/2007			Arts Program - Mathematics	2.29
26	นางสาว	สะกีนะ	เจะอุมา	ฟิสิกส์	16/7/2007			Arts Program - Mathematics	2.45
27	นาย	มะยูดิน	มะแอะ	การบริหารธุรกิจ(แขนงวิชาการบัญชี)	19/7/2007			Arts Program - Mathematics	2.55
28	นางสาว	ดารีหา	แหมล	ภาษาไทย	22/7/2007			Arts Program - Mathematics	2.56

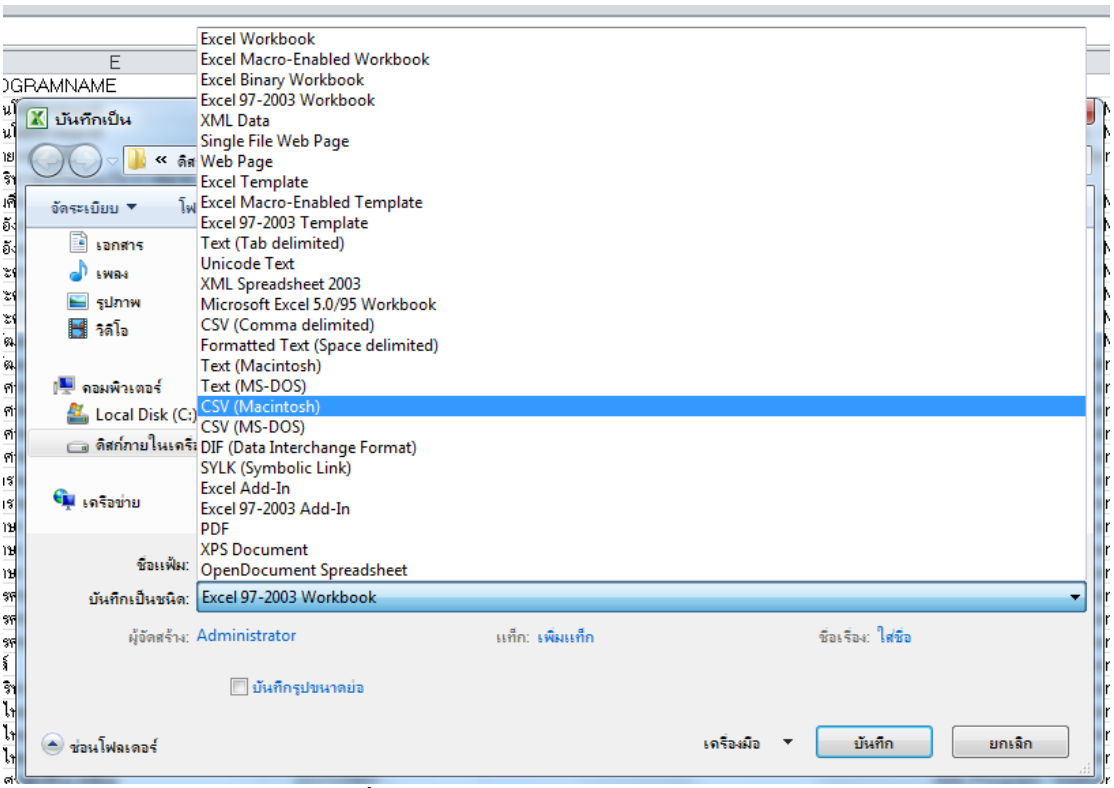
ภาพที่ 4.1 ข้อมูลนักศึกษาพ้นสภาพในปีการศึกษา 2550 - 2553

4.3 การคัดเลือกข้อมูล

จากการวิเคราะห์ข้อมูลที่ได้รับ สามารถคัดเลือกข้อมูล เพื่อนำมาวิเคราะห์ เทคนิคการจำแนกประเภทข้อมูล เป็นเทคนิคหนึ่งที่สำคัญของการสืบค้นความรู้บนฐานข้อมูลขนาดใหญ่ หรือเหมืองข้อมูล เป็นกระบวนการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนด เพื่อให้เกิดความแตกต่างระหว่าง กลุ่มของข้อมูล เพื่อนำไปใช้ในการทำนายต่อไปว่า ข้อมูลที่ได้มาควรจะจัดไปอยู่ในกลุ่มใด จำแนกข้อมูลออกเป็นกลุ่มตามที่ได้กำหนดไว้ จะขึ้นอยู่กับ การวิเคราะห์เซตของชุดข้อมูล (Dataset) โดยนำชุดข้อมูลมาสอนให้ระบบเรียนรู้ว่ามีข้อมูลใดอยู่ในกลุ่มเดียวกันบ้าง ซึ่งผลลัพธ์ที่ได้จากการเรียนรู้ คือ โมเดลจัดประเภทข้อมูล (Classifier Model) และโมเดลนี้ ผู้วิจัยได้รับการอนุเคราะห์ข้อมูล จากกองบริการการศึกษา มหาวิทยาลัยราชภัฏยะลา เพื่อนำไฟล์ข้อมูลมาพิจารณาดึงฟิลด์ข้อมูลที่เกี่ยวข้อง จากการศึกษา ฟิลด์ที่เกี่ยวข้องมีจำนวน 5 ฟิลด์ คือ คณะและสาขาวิชาที่นักศึกษาเลือกเรียน โรงเรียนเดิมที่นักศึกษาจบการศึกษา แผนการศึกษาในระดับมัธยมศึกษาที่นักศึกษาเลือกเรียน และเกรดเฉลี่ยที่นักศึกษาจบการศึกษา ผลปรากฏว่าปัจจัยที่มีความสัมพันธ์และสามารถวิเคราะห์ได้จากข้อมูลที่ซ้อนเร้นว่ามีเพียง 1 คู่คือ โปรแกรมวิชาที่นักศึกษาเลือกเรียนในระดับปริญญาตรี กับ แผนการศึกษาในระดับมัธยมศึกษา

4.4 การแปลงรูปแบบข้อมูล(Data Transformation)

นำไฟล์ข้อมูลที่ได้ จากกองบริการการศึกษามาเปลี่ยนสภาพของข้อมูลให้เหมาะสมกับเครื่องมือที่ผู้วิจัยใช้ในการวิเคราะห์ ในที่นี้ จะต้องแปลงข้อมูลที่ได้มาเป็น ไฟล์ *.csv เพื่อส่งข้อมูลเข้าประมวลผล ด้วยโปรแกรม weka-3-6-8jre



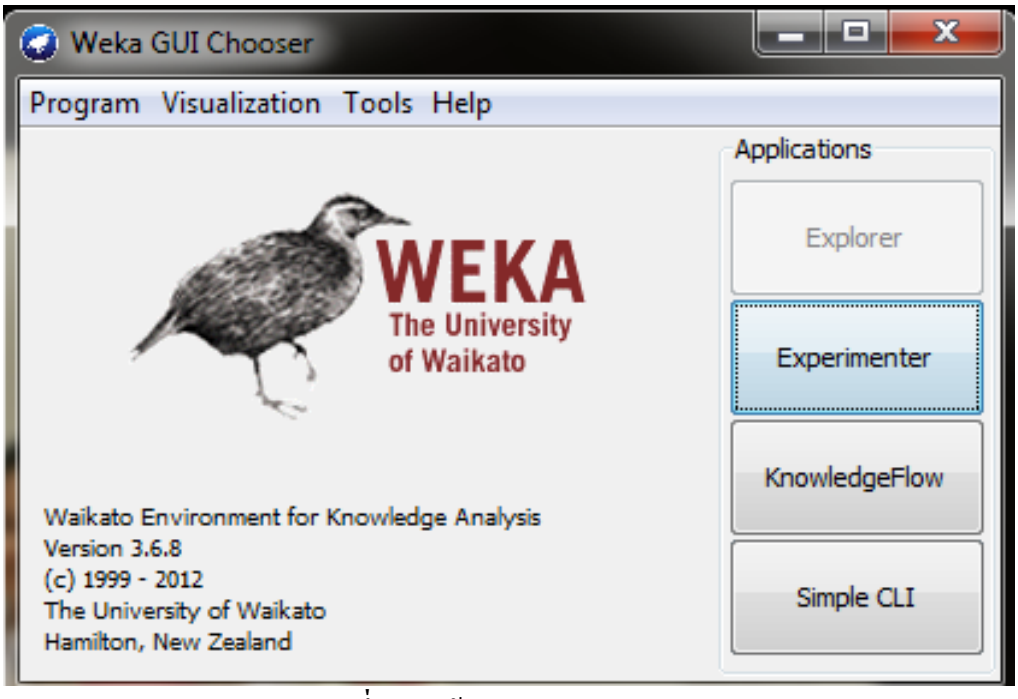
ภาพที่ 4.2 การแปลงข้อมูลให้อยู่ในรูปแบบ CSV

4.5 การทำเหมืองข้อมูล

เป็นการใช้เทคนิคภายในการทำเหมืองข้อมูล โดยทั่วไปประเภทของงานตามลักษณะของแบบจำลองที่ใช้ในการทำเหมืองข้อมูลจากงานวิจัยการวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษา ระดับปริญญาตรี โดยใช้เทคนิคเหมืองข้อมูล

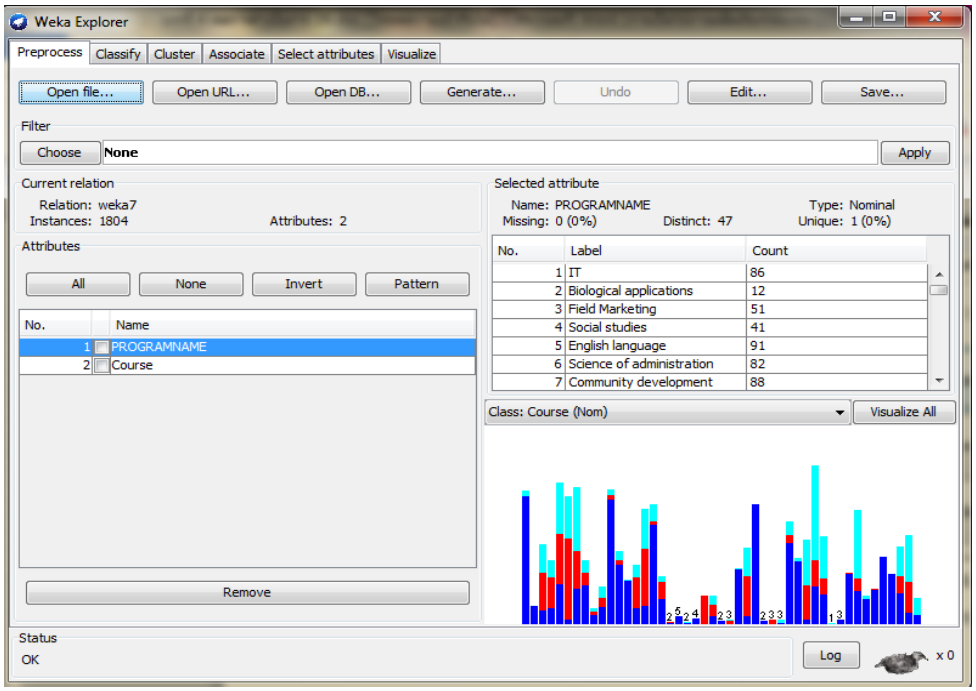
ซึ่งการค้นหากฎความสัมพันธ์ (Association Rule Discovery) ก็ถือเป็นเทคนิคหนึ่งของ การทำเหมืองข้อมูล ที่นักวิจัยต่างให้ความสนใจในการนำ เทคนิคนี้ไปช่วยในการค้นหากฎความสัมพันธ์ของข้อมูลในกลุ่มข้อมูลขนาดใหญ่ โดยจะนำไปวิเคราะห์ข้อมูลพฤติกรรมการเลือกสาขาที่นักศึกษาได้เลือกเรียนเพื่อหาความสัมพันธ์ต่าง ๆ และสามารถนำมาสรุปเป็นการค้นหาความสัมพันธ์ระหว่างรายการในแต่ละรายการหรือกลุ่มของรายการที่ปรากฏขึ้นในฐานข้อมูล

รูปแบบของกฎความสัมพันธ์ที่ได้จะอยู่ในรูปแบบ “IF X Then Y” โดยที่ X และ Y จะต้องเกิดขึ้นพร้อมกันภายในทรานแซคชั่นเดียวกัน โดยใช้สัญลักษณ์ X-->Y ซึ่งมีตัวแปรที่ใช้ในการวัดค่าอยู่ 2 ตัวแปร คือ ค่าสนับสนุน (Support Factor) และค่าความเชื่อมั่น (Confidence Factor) อัลกอริทึม Predictive Apriori ถือเป็นอัลกอริทึมที่นิยมใช้ในการหากฎความสัมพันธ์ของข้อมูล แต่ถ้าฐานข้อมูลมีการเพิ่มข้อมูลเข้ามา หรือเกิดมีการเปลี่ยนแปลงข้อมูล อัลกอริทึม Predictive Apriori Apriori จะต้องนำข้อมูลทั้งหมดมารวมกันก่อน แล้วจึงจะสามารถนำข้อมูลทั้งหมดไปค้นหากฎความสัมพันธ์ใหม่ทั้งหมด โดยไม่สามารถนำกฎความสัมพันธ์ที่หาได้จากกลุ่ม

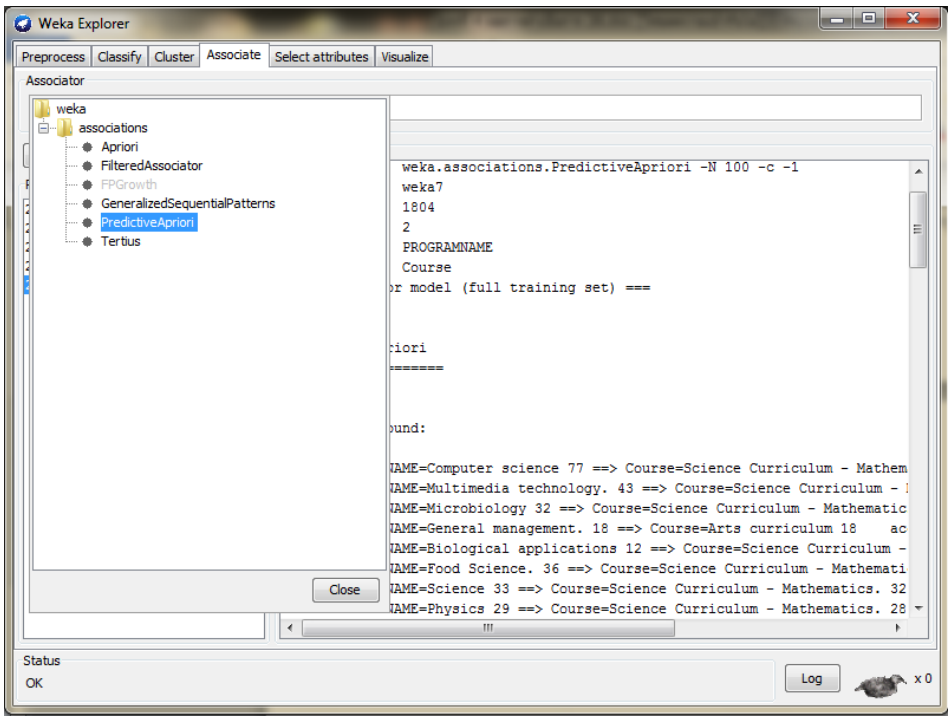


ภาพที่ 4.3 หน้าจอโปรแกรม Weka 3.6.8

การนำข้อมูลเพื่อใช้ในทำเหมืองข้อมูล ผ่านเมนู Application เลือก Explorer เข้าสู่เมนูในการนำเข้าข้อมูลเพื่อใช้ทำ เหมืองข้อมูล



ภาพที่ 4.4 การนำเข้าข้อมูลเพื่อใช้ทำ เหมืองข้อมูล



ภาพที่ 4.5 เลือกเทคนิคที่นำมาใช้ในการวิเคราะห์ปัจจัย อัลกอริทึม Apriori

4.6 การประเมินรูปแบบ

จากการวิเคราะห์ปัจจัยโดยใช้อัลกอริทึม Apriori โดยใช้โปรแกรม Weka 3.6.8 โดยปรากฏผลการวิเคราะห์ดังนี้ ข้อมูลกองบริการจำนวนนักศึกษาที่พ้นสภาพการเป็นนักศึกษา ในปีการศึกษา 2550-2553 จำนวน 1804 คน ข้อมูลดังกล่าวเมื่อนำเข้ามาวิเคราะห์กับโปรแกรม Weka 3.6.8 โดยวิเคราะห์จากความสัมพันธ์ระหว่าง หลักสูตรที่นักศึกษาได้ศึกษาผ่านมาในระดับมัธยมศึกษาตอนปลาย กับ หลักสูตรที่นักศึกษาเลือกเรียนในระดับปริญญาตรี สามารถสกัดเพื่อหาความสัมพันธ์แฝง ได้โดยสรุป หลักสูตรที่นักศึกษาได้ศึกษาผ่านมาในระดับมัธยมศึกษาตอนปลาย ร้อยละ 51.72 ที่ผ่านการศึกษาในหลักสูตร วิทย์-คณิต และเลือกเรียนในหลักสูตรที่นักศึกษาเลือกเรียนในระดับปริญญาตรี คือ หลักสูตรวิทยาศาสตร์บัณฑิต สาขา เทคโนโลยีสารสนเทศ หลักสูตรที่นักศึกษาได้ศึกษาผ่านมาในระดับมัธยมศึกษาตอนปลาย ร้อยละ 26.33 ที่ผ่านการศึกษาในหลักสูตร ศิลป์-คณิต และเลือกเรียนในหลักสูตรที่นักศึกษาเลือกเรียนในระดับปริญญาตรี คือ หลักสูตรบริหารธุรกิจบัณฑิต สาขาบัญชี และ หลักสูตรที่นักศึกษาได้ศึกษาผ่านมาในระดับมัธยมศึกษาตอนปลาย ร้อยละ 21.95 ที่ผ่านการศึกษาในหลักสูตร ศิลป์ และเลือกเรียนในหลักสูตรที่นักศึกษาเลือกเรียนในระดับปริญญาตรี คือ หลักสูตรครุศาสตร์บัณฑิต สาขา ภาษาไทย

จากการหาความสัมพันธ์ระดับหลักสูตรระดับปริญญาตรี ที่นักศึกษามีจำนวนพ้นสภาพการเป็นนักศึกษามากที่สุด ใน 3 ลำดับ

ตารางที่ 4.2 สรุปสาขาที่มีจำนวนนักศึกษาพ้นสภาพ

สาขา	หลักสูตร				
	วิทย์-คณิต	ศิลป์-คณิต	ศิลป์	รวม	ร้อยละ
การบริหารธุรกิจ(แขนงวิชาการบัญชี)	34	81	25	140	7.76
การบริหารธุรกิจ(แขนงวิชาการตลาด)	12	47	46	105	5.82
ภาษาอังกฤษ	26	33	32	91	5.04
พัฒนาชุมชน	23	41	24	88	4.88
สุขศึกษา	80	4	3	87	4.82
เทคโนโลยีสารสนเทศ	82	4	0	86	4.77
รัฐประศาสนศาสตร์	3	27	52	82	4.55
วิทยาศาสตร์สิ่งแวดล้อม	64	11	2	77	4.27

สาขา	หลักสูตร				
	วิทย์-คณิต	ศิลป์-คณิต	ศิลป์	รวม	ร้อยละ
วิทยาการคอมพิวเตอร์	77	0	0	77	4.27
ภาษาไทย	12	26	37	75	4.16
การบริหารธุรกิจ(แขนงวิชาคอมพิวเตอร์ธุรกิจ)	24	33	17	74	4.10
นิเทศศาสตร์	21	44	8	73	4.05
การจัดการทั่วไป	20	19	29	68	3.77
คณิตศาสตร์	52	9	5	66	3.66
ออกแบบนวัตกรรมทัศนศิลป์	7	30	20	57	3.16
การศึกษาปฐมวัย	5	18	26	49	2.72
เกษตรศาสตร์	38	6	3	47	2.61
เทคโนโลยีมัลติมีเดีย	43	0	0	43	2.38
เคมี	40	2	0	42	2.33
นิเทศศาสตร์(แขนงวิชาการประชาสัมพันธ์)	25	8	8	41	2.27
สังคมศึกษา	10	11	20	41	2.27
วิทยาศาสตร์และเทคโนโลยีการอาหาร	35	1	0	36	2.00
ชีววิทยา	34	1	0	35	1.94
วิทยาศาสตร์	33	1	0	34	1.88
จุลชีววิทยา	32	0	0	32	1.77
ภาษาจีน	16	4	11	31	1.72
ฟิสิกส์	28	1	0	29	1.61
ศิลปกรรม(แขนงออกแบบประยุกต์ศิลป์)	11	8	8	27	1.50
เทคโนโลยีการศึกษา	16	2	0	18	1.00
ภาษามลายู	6	11	0	17	0.94
ศิลปกรรม(แขนงออกแบบนิเทศศิลป์)	6	2	2	10	0.55
พลศึกษา	5	0	0	5	0.28

สาขา	หลักสูตร				
	วิทย์-คณิต	ศิลป์-คณิต	ศิลป์	รวม	ร้อยละ
วิทยาศาสตร์การกีฬา	5	0	0	5	0.28
วัฒนธรรมศึกษา	0	2	3	5	0.28
เทคโนโลยีและนวัตกรรมการศึกษา	3	1	0	4	0.22
เทคโนโลยีการเกษตร	4	0	0	4	0.22
คอมพิวเตอร์ศึกษา	0	2	1	3	0.17
รวม	886	362	311	1804	100.00

จากตารางสามารถสรุปสาขาที่นักศึกษา พ้นสภาพ ภายในปีการศึกษา 2550-2553 มี 3 ลำดับ คือ 1. การบริหารธุรกิจ(แขนงวิชาการบัญชี) จำนวน 140 คน 2. การบริหารธุรกิจ(แขนงวิชาการตลาด) จำนวน 105 คนและสาขาวิชาภาษาอังกฤษ จำนวน 91 คน

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

ในรายงานวิจัยในครั้งนี้มีจุดมุ่งหมาย เพื่อศึกษาปัจจัยของนักศึกษาที่พ้นสภาพการเป็นนักศึกษาและศึกษาว่าปัจจัยด้านเศรษฐกิจ ด้านการเรียน ด้านพื้นฐานความรู้จากแผนการเรียนรู้เดิมที่นักศึกษาได้ศึกษามา ด้านปัญหาส่วนตัว และด้านระบบการจัดการศึกษาของมหาวิทยาลัยมีผลต่อการพ้นสภาพนักศึกษา ในการวิจัยครั้งนี้ กลุ่มตัวอย่างเป็นนักศึกษาที่พ้นสภาพเป็นนักศึกษาในปีการศึกษา พ.ศ. 2550 -2553 ซึ่งสรุปสาระสำคัญและผลการศึกษาได้ดังนี้

วัตถุประสงค์

เพื่อวิเคราะห์ปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษา มหาวิทยาลัยราชภัฏยะลา

สมมติฐานการวิจัย

ผลการวิเคราะห์การทำเหมืองข้อมูล สามารถสกัดปัจจัยที่มีผลกระทบของการพ้นสภาพของนักศึกษาได้

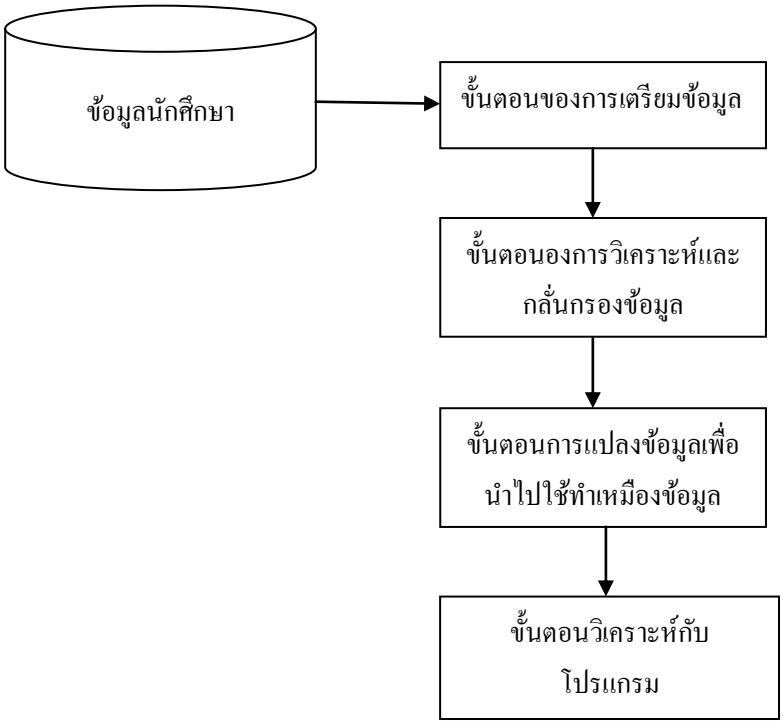
วิธีดำเนินการวิจัย

ประชากรที่ใช้ในการวิจัย

จำนวนนักศึกษาที่เข้าศึกษา ในปีการศึกษา 2549-2552 จำนวนทั้งสิ้น 4,528 คน และจำนวนข้อมูลนักศึกษาพ้นสภาพในระดับปริญญาตรี ระหว่างปีการศึกษา พ.ศ. 2550 -2553 นักศึกษาพ้นสภาพจำนวน 1,804 คน และนักศึกษาที่ยังศึกษาต่อ จำนวน 2,724 คน

การวิเคราะห์ข้อมูล

วิจัยนี้นี้มุ้งเน้นในส่วนของนักศึกษาที่พ้นสภาพการเป็นนักศึกษา แต่ในกระบวนการประมวลผลนำข้อมูลนักศึกษาที่ได้เข้าศึกษาในปีการศึกษาเดียวกันมาเปรียบเทียบปัจจัย โดยจำนวนนักศึกษาที่เข้าศึกษา ในปีการศึกษา 2549-2552 จำนวนทั้งสิ้น 4,528 คน นำข้อมูลทั้งหมดโดยการคัดแยกอย่างง่าย ใช้ข้อมูลนักศึกษาพ้นสภาพในระดับปริญญาตรี ระหว่างปีการศึกษา พ.ศ. 2550 -2553 นักศึกษาพ้นสภาพจำนวน 1,804 คน และนักศึกษาที่ยังศึกษาต่อจำนวน 2,724 คน โดยมีกระบวนการทำงานดังนี้



ภาพที่ 5.1 แสดงขั้นตอนการวิเคราะห์ปัจจัย

ขั้นตอนการเตรียมการข้อมูล (Data Preparation Phase) จะมีการระบุชุดข้อมูลหลักที่ใช้โดยเหมืองข้อมูล และจัดเรียงให้เรียบร้อย ข้อมูลทั่วไปในขั้นตอนนี้โดยงานวิจัยจะสนใจ ปัจจัย คณะที่เรียน สาขาที่เรียน อายุนักศึกษา วุฒิการศึกษาเดิม เพศของนักศึกษา วุฒิการศึกษาเดิม อาชีพของบิดา-มารดา สถานภาพของบิดา-มารดา ได้ข้อมูลนักศึกษาพ้นสภาพในช่วงปีการศึกษา 2550 – 2553 จำนวนนักศึกษา 1804 คน

ขั้นตอนองการวิเคราะห์และกลั่นกรองข้อมูล (Data Analysis and Classification Phase) เป็นการนำข้อมูลที่ไม่มีค่า ข้อมูลที่ขาดหาย และข้อมูลที่ไม่แน่นอนออกไป ข้อมูลที่ได้มานั้น เป็นข้อมูลที่ยังไม่สมบูรณ์ ที่จะสามารถนำไปใช้ผ่านกระบวนการเหมืองข้อมูลได้ จึงต้องมีการจัดการข้อมูล การเตรียมข้อมูลเบื้องต้น

ขั้นตอนการแปลงข้อมูลเพื่อนำไปใช้ทำเหมืองข้อมูล (Knowledge Acquisition Phase) การแปลงข้อมูลที่เลือกมาให้อยู่ในรูปแบบที่เหมาะสมสำหรับการนำไปใช้วิเคราะห์ตามขั้นตอนวิธีที่ใช้ในการทำเหมืองข้อมูล ขั้นตอนวิธี Predictive Apriori ถือเป็นขั้นตอนวิธีที่นิยมใช้ในการหากฎความสัมพันธ์ของข้อมูล แต่ถ้าฐานข้อมูลมีการเพิ่มข้อมูลเข้ามา หรือเกิดมีการเปลี่ยนแปลงข้อมูล ขั้นตอนวิธี Predictive Apriori Apriori จะต้องนำข้อมูลทั้งหมดมารวมกันก่อน แล้วจึงจะสามารถ

นำข้อมูลทั้งหมดไปค้นหากฎความสัมพันธ์ใหม่ทั้งหมด โดยไม่สามารถนำกฎความสัมพันธ์ที่หาได้จากกลุ่ม

จะเห็นได้ว่าข้อมูลที่เลือกสามารถตอบเป้าหมายที่ต้องการจะศึกษา คือปัจจัยที่ส่งผลให้นักศึกษาพ้นสภาพการเป็นนักศึกษา เพื่อที่เราจะได้สามารถนำตารางนี้ไปปรับเปลี่ยนเพื่อให้เหมาะสมกับเทคนิคต่าง ๆ

ขั้นตอนวิเคราะห์กับโปรแกรม Weka 3.6.8 (Prognosis Phase) โดยวิเคราะห์จากความสัมพันธ์ระหว่าง หลักสูตรที่นักศึกษาได้ศึกษาผ่านมาในระดับมัธยมศึกษาตอนปลาย กับหลักสูตรที่นักศึกษาเลือกเรียนในระดับปริญญาตรี สามารถสกัดเพื่อหาความสัมพันธ์แฝง ได้โดยสรุป หลักสูตรที่นักศึกษาได้ศึกษาผ่านมาในระดับมัธยมศึกษาตอนปลาย ร้อยละ 51.72 ที่ผ่านการศึกษาในหลักสูตร วิทย์-คณิต และเลือกเรียนในหลักสูตรที่นักศึกษาเลือกเรียนในระดับปริญญาตรี คือ หลักสูตรวิทยาศาสตร์บัณฑิต สาขาเทคโนโลยีสารสนเทศ หลักสูตรที่นักศึกษาได้ศึกษาผ่านมาในระดับมัธยมศึกษาตอนปลาย ร้อยละ 26.33 ที่ผ่านการศึกษาในหลักสูตร ศิลป์-คณิต และเลือกเรียนในหลักสูตรที่นักศึกษาเลือกเรียนในระดับปริญญาตรี คือ หลักสูตรบริหารธุรกิจบัณฑิต สาขาบัญชี และ หลักสูตรที่นักศึกษาได้ศึกษาผ่านมาในระดับมัธยมศึกษาตอนปลาย ร้อยละ 21.95 ที่ผ่านการศึกษาในหลักสูตร ศิลป์ และเลือกเรียนในหลักสูตรที่นักศึกษาเลือกเรียนในระดับปริญญาตรี คือ หลักสูตรครุศาสตร์บัณฑิต สาขา ภาษาไทย

สรุปผลการวิจัย

การประยุกต์ใช้เทคนิคการวิเคราะห์ค้นหากฎความสัมพันธ์ (Association Rule Discovery) ก็ถือเป็นเทคนิคหนึ่งของเหมืองข้อมูล (Data Mining) ที่นักวิจัยต่างให้ความสนใจในการนำ เทคนิคนี้ไปช่วยในการค้นหากฎความสัมพันธ์ของข้อมูลในกลุ่มข้อมูลขนาดใหญ่ (Data Set) โดยจะนำไปวิเคราะห์ข้อมูลพฤติกรรมการเลือกสาขาที่นักศึกษาได้เลือกเรียนเพื่อหาความสัมพันธ์ต่าง ๆ และสามารถนำมาสรุปเป็น เป็นการค้นหาความสัมพันธ์ระหว่างรายการในแต่ละรายการหรือกลุ่มของรายการที่ปรากฏขึ้นในฐานข้อมูลรูปแบบของกฎความสัมพันธ์ที่ได้จะอยู่ในรูปแบบ “IF X Then Y” โดยที่ X และ Y จะต้องความสัมพันธ์ สามารถได้สมการการทำนายโอกาสการพ้นสภาพของนักศึกษาได้เป็นการเบื้องต้น และปัจจัยที่มีผลต่อการพ้นสภาพนักศึกษา คือ ปัจจัยส่งผลต่อการพ้นสภาพนักศึกษาจากข้อมูลที่ผู้วิจัยได้รับจากกองบริการการศึกษาสามารถวิเคราะห์ได้ 2 ปัจจัย คือ สาขาที่นักศึกษาเลือกเรียนในระดับปริญญาตรี และ วุฒิการศึกษาเดิม ซึ่งรูปแบบในการทำนายนี้สามารถนำไปเป็นต้นแบบที่จะนำไปปรับเปลี่ยนตัวแปรอื่น ๆ ที่เกี่ยวข้องได้ตามความเหมาะสม เพราะสาเหตุที่แท้จริงไม่มีใครทราบได้นอกจากตัวของนักศึกษาที่พ้นสภาพเอง ซึ่งอาจมีปัจจัยหลาย

อย่างเข้ามาประกอบด้วย เช่น ค่าใช้จ่ายต่าง ๆ การต้องออกไปทำงานช่วยเหลือครอบครัว การลาออกไปแต่งงาน ความเกียจคร้านในการเรียน การเปลี่ยนสถาบันเรียน หรือความไม่มีเหตุผลใด ๆ เลย เป็นต้น เนื่องจากสภาพพื้นฐานครอบครัวของนักศึกษาส่วนใหญ่ มีอาชีพเกษตรกรและรายได้น้อย และไม่ได้รับทุนการศึกษา ทำให้เป็นสาเหตุหลักของความไม่พร้อมกับการศึกษา

อภิปรายผล

มีนักเรียนจำนวนมากในระดับมัธยมศึกษาตอนต้น ที่ยังลังเล และไม่รู้ว่าตนเองควรจะเลือกเรียนต่อระดับมัธยมปลายในสาขาใด อาจจะด้วยเหตุผลหลายประการ ไม่ว่าจะเป็นเรื่องของผลการเรียน ความชอบ ความสามารถเฉพาะด้าน และส่วนหนึ่งอาจจะมาจากคำแนะนำหรือข้อบังคับของผู้ปกครอง ซึ่งส่งผลต่อการเลือกสาขาในระดับมัธยมปลาย ทำให้นักเรียนที่ถูกบังคับ หรือเลือกสาขาผิด อาจจะเรียนได้ไม่ดี ไม่มีความสุขในการเรียน เกิดการเปลี่ยนหรือย้ายสาขาระหว่างเรียน ทำให้เสียเวลา และส่งผลต่อเนื่องในการเลือกคณะ สาขาในการเรียนต่อระดับอุดมศึกษา ในการวิจัยครั้งควรจะปรับปรุงรู้แบบการเก็บข้อมูล โดยพิจารณาจากความชอบทางการเรียน และใช้คำถามเชิงจิตวิทยามาผสมผสานในการสร้างข้อคำถามเกี่ยวกับการวัดทักษะทางด้านปฏิบัติและวิชาการ เพื่อให้นักเรียนได้ค้นหาตัวเองและแนวทางการศึกษาต่อ ซึ่งผลที่ได้จะแบ่งเป็นคณะ และสาขาตามทักษะดังกล่าว เป็นการวัดตามความชอบ นักเรียนที่มีคะแนน หรือผลการเรียนน้อย เพื่อไม่ให้นักเรียนประสบกับปัญหาการพ้นสภาพในการเป็นนักศึกษา

ข้อเสนอแนะ

ในงานวิจัยนี้มีประสบกับปัญหาทางด้านข้อมูลเนื่องจากข้อมูลที่ได้รับจาก กองบริการการศึกษามีรายละเอียดที่ไม่ตรงตามความต้องการ ซึ่งส่งผลให้การผลการวิเคราะห์ออกมาไม่ชัดเจน ในการเก็บข้อมูลเพื่อตอบวัตถุประสงค์ในงานวิจัยชิ้นนี้ ผู้วิจัยจะต้องทำการเก็บแบบสอบถามสำหรับนักศึกษาเข้าศึกษาในปีการศึกษา และจะต้องเป็นข้อมูลเป็นปัจจุบัน เพื่อให้ได้ผลการวิเคราะห์ที่สมบรูณ์

บรรณานุกรม

ภาษาไทย

กฤษณะ ไวยมัย และ ธีระวัฒน์ พงษ์ศิริปรีดา, ”การใช้เทคนิค Association Rule Discovery เพื่อการจัดสรรกฎหมายในการพิจารณาคดีความ” NECTEC Technical Journal Vol. III, No. 11-143

ณ.ชนม์ ประยูรวงศ์ , แสงสุรีย์ วสุพงศ์อัยยะ, ”ศึกษาเบื้องต้นกับปัจจัยที่ส่งผลให้นักศึกษามีผลการเรียนอยู่ในภาวะรอพินิจ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์” ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์

นภดล สายคติกรณ์, “การพัฒนาแบบจำลองสมการโครสร้างแบบหลายกลุ่ม เพื่อศึกษาและวิเคราะห์ปัจจัยเชิงสาเหตุที่มีผลต่อความปลอดภัยในการใช้บริการระบบเครือข่ายคอมพิวเตอร์ของนักศึกษา,” NCCIT2010-179, มหาวิทยาลัยพระจอมเกล้าพระนครเหนือ ,2553

ประกาศิต ช่างสุพรรณ,”การจำแนกระดับความมั่นใจของผู้เรียน โดยใช้เทคนิคต้นไม้ตัดสินใจ,” มหาวิทยาลัยพระจอมเกล้าพระนครเหนือ,2553

ภัทรพงศ์ พงศ์ภัทรกานต์,การวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษา ระดับปริญญาตรี โดยใช้คอมมิตตีแมชชีน,มหาวิทยาลัยราชภัฏเลย2553

รินทร์ลภัส จินานุศิลปะสารท, “การทำเหมืองข้อมูลเพื่อค้นหารูปแบบความสัมพันธ์ที่ซ่อนอยู่โดยใช้ การวิเคราะห์หลักเกณฑ์การเชื่อมโยงของข้อมูลลูกค้าประกันภัย,” NCCIT2010-183, มหาวิทยาลัยพระจอมเกล้าพระนครเหนือ,2553

วีระ จิรกิจอนุสรณ์,“การคาดการณ์ภาษีมูลค่าเพิ่มด้วยเทคนิคของเหมืองข้อมูล,” มหาวิทยาลัยเกษตรศาสตร์,2553

วรากุล กาญจนกัญโห,“การวิเคราะห์ปัจจัยที่มีความสัมพันธ์ต่อการเกิดอาชญากรรมโดยใช้หลักการของ GRNN,”มหาวิทยาลัยขอนแก่น,2553

ภาษาอังกฤษ

Pang-Ning Tan, Michael Steinbach, Vipin Kumar.” Introduction to Data Mining”. Pearson International Edition

ภาคผนวก ก

บันทึกข้อความขออนุเคราะห์ข้อมูล

บันทึกข้อความ

ส่วนราชการ สาขาคอมพิวเตอร์ ภาควิชาวิทยาศาสตร์ประยุกต์ โทร.๐-๗๓๒๒-๗๑๕๑ ต่อ ๙๔๑๐

ที่ **วันที่** 30 ธันวาคม 2555

เรื่อง ขอความอนุเคราะห์เก็บข้อมูล

เรียนอธิการบดี มหาวิทยาลัยราชภัฏยะลา

ด้วย นางขวัญฤทัย นกแก้ว อาจารย์สาขาเทคโนโลยีสารสนเทศ ภาควิชาวิทยาศาสตร์ประยุกต์ คณะวิทยาศาสตร์เทคโนโลยีและการเกษตร มหาวิทยาลัยราชภัฏยะลา ได้รับทุนอุดหนุนการวิจัยจากงบประมาณบำรุงการศึกษา ประจำปีงบประมาณ พ.ศ. 2555 ในหัวข้อวิจัย เรื่อง "การวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษา ระดับปริญญาตรี โดยใช้เทคนิคเหมืองข้อมูล" ซึ่งในการศึกษาครั้งนี้ได้กำหนดกลุ่มตัวอย่างในการเก็บรวบรวมข้อมูลคือ รายละเอียด ข้อมูลประวัตินักศึกษาของมหาวิทยาลัยราชภัฏยะลา ตั้งแต่ปี พ.ศ. 2550 ถึง พ.ศ. 2554 ที่พ้นสภาพการเป็นนักศึกษา เพื่อให้การศึกษาครั้งนี้สำเร็จลุล่วงด้วยดี ผู้วิจัย จึงใคร่ขอความอนุเคราะห์ข้อมูล ดังกล่าวดำเนินการเก็บรวบรวมข้อมูลในการทำวิจัยตามประสงค์

จึงเรียนมาเพื่อโปรดพิจารณา

(.....)

ความเห็นที่ 1	
มอบ ซอพีอนุเคราะห์ข้อมูลโดยประสานกับ อ.ขวัญฤทัย (นายซอและ เกปัน) ผู้อำนวยการกองบริการการศึกษา 30 ธ.ค. 54 เวลา 14:04:58 , Non-PKI Simple Sign	

ภาคผนวก ข

ผลการวิเคราะห์ Association rule

=== Run information ===

Scheme: weka.associations.PredictiveApriori -N 100 -c -1

Relation: weka7

Instances: 1804

Attributes: 2

PROGRAMNAME

Course

=== Associator model (full training set) ===

PredictiveApriori

==========

Best rules found:

1. PROGRAMNAME=Computer science 77 ==> Course=Science Curriculum - Mathematics. 77 acc:(0.99336)

2. PROGRAMNAME=Multimedia technology. 43 ==> Course=Science Curriculum - Mathematics. 43 acc:(0.99102)

3. PROGRAMNAME=Microbiology 32 ==> Course=Science Curriculum - Mathematics. 32 acc:(0.98938)

4. PROGRAMNAME=General management. 18 ==> Course=Arts curriculum 18 acc:(0.985)

5. PROGRAMNAME=Biological applications 12 ==> Course=Science Curriculum - Mathematics. 12 acc:(0.98019)

6. PROGRAMNAME=Food Science. 36 ==> Course=Science Curriculum - Mathematics. 35 acc:(0.97407)

7. PROGRAMNAME=Science 33 ==> Course=Science Curriculum - Mathematics. 32 acc:(0.97195)

8. PROGRAMNAME=Physics 29 ==> Course=Science Curriculum - Mathematics. 28 acc:(0.96842)

9. PROGRAMNAME=Biology 23 ==> Course=Science Curriculum - Mathematics. 22 acc:(0.96066)

10. PROGRAMNAME=IT 86 ==> Course=Science Curriculum - Mathematics. 82 acc:(0.95619)

11. PROGRAMNAME=Chemical 42 ==> Course=Science Curriculum - Mathematics. 40 acc:(0.95582)

12. PROGRAMNAME=Gymnastics 5 ==> Course=Science Curriculum - Mathematics. 5 acc:(0.94377)

13. PROGRAMNAME=Health education 87 ==> Course=Science Curriculum - Mathematics. 80 acc:(0.92326)

14. PROGRAMNAME=Sports Science. 4 ==> Course=Science Curriculum - Mathematics. 4 acc:(0.92167)

15. PROGRAMNAME=Educational Technology. 18 ==> Course=Science Curriculum - Mathematics. 16 acc:(0.8872)

16. PROGRAMNAME=Cultural studies 3 ==> Course=Arts curriculum 3 acc:(0.88468)

17. PROGRAMNAME=Technology and innovation studies. 3 ==> Course=Science Curriculum - Mathematics. 3 acc:(0.88468)

18. PROGRAMNAME=Environmental Science 77 ==> Course=Science Curriculum - Mathematics. 64 acc:(0.8257)

19. PROGRAMNAME=Agricultural technology. 2 ==> Course=Science Curriculum - Mathematics. 2 acc:(0.82305)

20. PROGRAMNAME=The aquaculture 2 ==> Course=Science Curriculum - Mathematics. 2 acc:(0.82305)

21. PROGRAMNAME=Agriculture 47 ==> Course=Science Curriculum - Mathematics. 38 acc:(0.80993)

22. PROGRAMNAME=Mathematics 66 ==> Course=Science Curriculum - Mathematics. 52 acc:(0.80016)

23. PROGRAMNAME=Science of administration 82 ==> Course=Arts curriculum 52 acc:(0.60078)

24. PROGRAMNAME=Accounting 102 ==> Course=Arts Program - Mathematics 61 acc:(0.56918)

25. PROGRAMNAME=Communication. 73 ==> Course=Arts Program - Mathematics 44 acc:(0.56566)

26. PROGRAMNAME=Malay 17 ==> Course=Arts Program - Mathematics 11 acc:(0.56202)

27. PROGRAMNAME=Public Relations field 41 ==> Course=Science Curriculum - Mathematics. 25 acc:(0.55811)

28. PROGRAMNAME=Design Department of Applied Arts 3 ==> Course=Arts curriculum 2 acc:(0.53637)

29. PROGRAMNAME=Computer Studies. 3 ==> Course=Arts Program - Mathematics 2 acc:(0.53637)

30. PROGRAMNAME=Visual Communication Design course 10 ==> Course=Science Curriculum - Mathematics. 6 acc:(0.5276)

31. PROGRAMNAME=Marketing 54 ==> Course=Arts Program - Mathematics 29 acc:(0.52303)

32. PROGRAMNAME=Early Childhood Education 49 ==> Course=Arts curriculum 26 acc:(0.52054)

33. PROGRAMNAME=Innovative visual design. 57 ==> Course=Arts Program - Mathematics 30 acc:(0.51885)

34. PROGRAMNAME=Accounting field 38 ==> Course=Arts curriculum 20 acc:(0.51832)

35. PROGRAMNAME=Chinese 31 ==> Course=Science Curriculum - Mathematics. 16 acc:(0.51283)

36. PROGRAMNAME=Language Thailand. 75 ==> Course=Arts curriculum 37 acc:(0.50709)

37. PROGRAMNAME=Social studies 41 ==> Course=Arts curriculum 20 acc:(0.5025)

38. PROGRAMNAME=Business computer 56 ==> Course=Arts Program - Mathematics 27 acc:(0.50176)

39. PROGRAMNAME=Community development 88 ==> Course=Arts Program - Mathematics 41 acc:(0.4957)

40. PROGRAMNAME=Field Marketing 51 ==> Course=Arts curriculum 24 acc:(0.49533)

41. PROGRAMNAME=Applied Arts course design 24 ==> Course=Science Curriculum - Mathematics. 11 acc:(0.47745)

42. PROGRAMNAME=Marketing 54 ==> Course=Arts curriculum 22 acc:(0.43305)

43. PROGRAMNAME=Management 50 ==> Course=Science Curriculum - Mathematics. 20 acc:(0.42253)

44. PROGRAMNAME=Department computers 18 ==> Course=Arts curriculum 7 acc:(0.41283)

45. PROGRAMNAME=Management 50 ==> Course=Arts Program - Mathematics 19 acc:(0.39067)

46. PROGRAMNAME=Malay 17 ==> Course=Science Curriculum - Mathematics. 6 acc:(0.37783)

47. PROGRAMNAME=Early Childhood Education 49 ==> Course=Arts Program - Mathematics 18 acc:(0.37104)

48. PROGRAMNAME=Chinese 31 ==> Course=Arts curriculum 11 acc:(0.36415)

49. PROGRAMNAME=Department computers 18 ==> Course=Arts Program - Mathematics 6 acc:(0.35657)

50. PROGRAMNAME=English language 91 ==> Course=Arts Program - Mathematics 33 acc:(0.35425)

51. PROGRAMNAME=Field Marketing 51 ==> Course=Arts Program - Mathematics 18 acc:(0.34887)

52. PROGRAMNAME=Innovative visual design. 57 ==> Course=Arts curriculum 20 acc:(0.34396)

53. PROGRAMNAME=English language 91 ==> Course=Arts curriculum 32 acc:(0.34178)

54. PROGRAMNAME=Language Thailand. 75 ==> Course=Arts Program - Mathematics 26 acc:(0.33605)

55. PROGRAMNAME=Business computer 56 ==> Course=Science Curriculum - Mathematics. 19 acc:(0.32956)

56. PROGRAMNAME=Science of administration 82 ==> Course=Arts Program - Mathematics 27 acc:(0.31751)

57. PROGRAMNAME=Department computers 18 ==> Course=Science Curriculum - Mathematics. 5 acc:(0.30671)

58. PROGRAMNAME=Applied Arts course design 24 ==> Course=Arts Program - Mathematics 7 acc:(0.30512)

59. PROGRAMNAME=Communication. 73 ==> Course=Science Curriculum - Mathematics. 21 acc:(0.28052)

60. PROGRAMNAME=Visual Communication Design course 10 ==> Course=Arts curriculum 2 acc:(0.28044)

61. PROGRAMNAME=Visual Communication Design course 10 ==> Course=Arts Program - Mathematics 2 acc:(0.28044)

62. PROGRAMNAME=English language 91 ==> Course=Science Curriculum - Mathematics. 26 acc:(0.27833)

63. PROGRAMNAME=Applied Arts course design 24 ==> Course=Arts curriculum 6 acc:(0.27517)

64. PROGRAMNAME=Social studies 41 ==> Course=Arts Program - Mathematics 11 acc:(0.27316)

65. PROGRAMNAME=Accounting field 38 ==> Course=Science Curriculum - Mathematics. 10 acc:(0.2714)

66. PROGRAMNAME=Community development 88 ==> Course=Arts curriculum 24 acc:(0.26919)

67. PROGRAMNAME=Community development 88 ==> Course=Science Curriculum - Mathematics. 23 acc:(0.26196)

68. PROGRAMNAME=Social studies 41 ==> Course=Science Curriculum - Mathematics. 10 acc:(0.25975)

69. PROGRAMNAME=Accounting 102 ==> Course=Science Curriculum - Mathematics. 24 acc:(0.24619)

70. PROGRAMNAME=Management 50 ==> Course=Arts curriculum 11 acc:(0.24541)

71. PROGRAMNAME=Accounting field 38 ==> Course=Arts Program - Mathematics 8 acc:(0.24436)

72. PROGRAMNAME=Public Relations field 41 ==> Course=Arts curriculum 8 acc:(0.23529)

73. PROGRAMNAME=Public Relations field 41 ==> Course=Arts Program - Mathematics 8 acc:(0.23529)

74. PROGRAMNAME=Business computer 56 ==> Course=Arts curriculum 10 acc:(0.22087)

75. PROGRAMNAME=Field Marketing 51 ==> Course=Science Curriculum - Mathematics. 9 acc:(0.22073)

76. PROGRAMNAME=Accounting 102 ==> Course=Arts curriculum 17 acc:(0.2014)

77. PROGRAMNAME=Language Thailand. 75 ==> Course=Science Curriculum - Mathematics. 12 acc:(0.19967)

78. PROGRAMNAME=Chinese 31 ==> Course=Arts Program - Mathematics 4 acc:(0.1826)

79. PROGRAMNAME=Educational Technology. 18 ==> Course=Arts Program - Mathematics 2 acc:(0.17662)

80. PROGRAMNAME=Environmental Science 77 ==> Course=Arts Program - Mathematics 11 acc:(0.17611)

81. PROGRAMNAME=Mathematics 66 ==> Course=Arts Program - Mathematics 9 acc:(0.17068)

82. PROGRAMNAME=Agriculture 47 ==> Course=Arts Program - Mathematics 6 acc:(0.16781)

83. PROGRAMNAME=Innovative visual design. 57 ==> Course=Science Curriculum - Mathematics. 7 acc:(0.15287)

84. PROGRAMNAME=Communication. 73 ==> Course=Arts curriculum 8 acc:(0.12136)

85. PROGRAMNAME=Early Childhood Education 49 ==> Course=Science Curriculum - Mathematics. 5 acc:(0.12075)

86. PROGRAMNAME=Mathematics 66 ==> Course=Arts curriculum 5 acc:(0.07479)

87. PROGRAMNAME=Agriculture 47 ==> Course=Arts curriculum 3 acc:(0.06655)

88. PROGRAMNAME=Marketing 54 ==> Course=Science Curriculum - Mathematics. 3 acc:(0.05709)

89. PROGRAMNAME=Chemical 42 ==> Course=Arts Program - Mathematics 2 acc:(0.05331)

90. Course=Arts curriculum 396 ==> PROGRAMNAME=Science of administration 52 acc:(0.04871)

91. PROGRAMNAME=IT 86 ==> Course=Arts Program - Mathematics 4 acc:(0.04791)

92. Course=Arts Program - Mathematics 475 ==> PROGRAMNAME=Accounting 61 acc:(0.04769)

93. PROGRAMNAME=Health education 87 ==> Course=Arts Program - Mathematics 4 acc:(0.04751)

94. PROGRAMNAME=Science of administration 82 ==> Course=Science Curriculum - Mathematics. 3 acc:(0.04109)

95. PROGRAMNAME=Health education 87 ==> Course=Arts curriculum 3 acc:(0.03943)

96. Course=Arts curriculum 396 ==> PROGRAMNAME=Language Thailand. 37 acc:(0.03849)

97. Course=Arts Program - Mathematics 475 ==> PROGRAMNAME=Communication. 44 acc:(0.03801)

98. Course=Arts Program - Mathematics 475 ==> PROGRAMNAME=Community development 41 acc:(0.03641)

99. Course=Science Curriculum - Mathematics. 933 ==> PROGRAMNAME=IT 82 acc:(0.0361)

=== Run information ===

Scheme: weka.associations.Tertius -K 10 -F 0.0 -N 1.0 -L 4 -G 0 -c 0 -I 0 -P 0

Relation: weka7

Instances: 1804

Attributes: 2

PROGRAMNAME

Course

=== Associator model (full training set) ===

Tertius

=======

1. /* 0.167617 0.000000 */ PROGRAMNAME = Computer science ==> Course = Science Curriculum - Mathematics.
2. /* 0.161617 0.002217 */ PROGRAMNAME = IT ==> Course = Science Curriculum - Mathematics.
3. /* 0.150061 0.003880 */ PROGRAMNAME = Health education ==> Course = Science Curriculum - Mathematics.
4. /* 0.123283 0.016630 */ PROGRAMNAME = Science of administration ==> Course = Arts curriculum
5. /* 0.120168 0.000000 */ PROGRAMNAME = Multimedia technology. ==> Course = Science Curriculum - Mathematics.
6. /* 0.116514 0.022727 */ PROGRAMNAME = Accounting ==> Course = Arts Program - Mathematics
7. /* 0.109005 0.007206 */ PROGRAMNAME = Environmental Science ==> Course = Science Curriculum - Mathematics.
8. /* 0.106899 0.001109 */ PROGRAMNAME = Chemical ==> Course = Science Curriculum - Mathematics.
9. /* 0.102579 0.000554 */ PROGRAMNAME = Food Science. ==> Course = Science Curriculum - Mathematics.
10. /* 0.101982 0.000000 */ PROGRAMNAME = Microbiology ==> Course = Science Curriculum - Mathematics.

Number of hypotheses considered: 382

Number of hypotheses explored: 344

ภาคผนวก ค

ประวัติผู้วิจัย

ประวัติคณะผู้วิจัย

1. ชื่อ (ภาษาไทย) ขวัญฤทัย นกแก้ว

(ภาษาอังกฤษ) Kwanrutai Nokkeaw

2. เลขหมายบัตรประจำตัวประชาชน 3-9-402-004-27-489

3. ตำแหน่งปัจจุบัน อาจารย์สัญญาจ้าง

4. หน่วยงานและสถานที่อยู่ที่ติดต่อได้สะดวก พร้อมหมายเลขโทรศัพท์ โทรสาร

อาจารย์ประจำหลักสูตรเทคโนโลยีสารสนเทศ ภาควิชาวิทยาศาสตร์ประยุกต์

คณะวิทยาศาสตร์เทคโนโลยีและการเกษตร

สถานที่/ห้องพัก/อาคาร : อาคาร 4 ห้อง 405 สาขาวิชาคอมพิวเตอร์

โทรศัพท์ : 081-0928038 อีเมล์ : a_kwanrutai@hotmail.com

5. ประวัติการศึกษา

ชื่อ - ชื่อสกุล : นางขวัญฤทัย นกแก้ว

สัญชาติ : ไทย **เชื้อชาติ :** ไทย **ศาสนา** : พุทธ **เกิดวันที่** : 11 **เดือน** : ธันวาคม **พ.ศ.** 2523

ภูมิลำเนาอยู่บ้านเลขที่ 4/126 **หมู่ที่** 6 **ถนน** ปุณณกัณฑ์ **ซอย** ชุมชนปาล์มซิตี้ **ตำบล** คอหงส์

อำเภอ / กิ่งอำเภอ หาดใหญ่ **จังหวัด** สงขลา **รหัสไปรษณีย์** 90110 **โทรศัพท์** 081-092-8038

ระดับการศึกษา:

สถาบันการศึกษา: ม.เทคโนโลยีพระจอมเกล้าพระนครเหนือ **ระดับ**: ปริญญาโท

วุฒิการศึกษา: วท.ม.เทคโนโลยีสารสนเทศ **สาขาวิชา:** เทคโนโลยีสารสนเทศ

ปีการศึกษาที่จบ: 2551

สถาบันการศึกษา: สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ **ระดับ:** ปริญญาตรี

วุฒิการศึกษา: วท.บ วิทยาศาสตร์บัณฑิต **สาขาวิชา:** สถิติประยุกต์

ปีการศึกษาที่จบ: 2545

6. สาขาวิชาการที่มีความชำนาญพิเศษ Data mining

7. ประสบการณ์ที่เกี่ยวข้องกับการวิจัย

Saelim K. and Utakrit N., **A Design and Development of Blood Requirement Forecast system using Neural Network**, Proceedings of The 5st National Computer and Information Technology Conference (NCCIT 2009),King Mongkut's University of Technology North Bangkok,May.22-23, 2009, Bangkok, Thailand