ระบบสนับสนุนการตัดสินใจเพื่อแนะนำการลงทะเบียนรายวิชาให้นักศึกษา สาขาวิชาวิทยาการคอมพิวเตอร์

ณรัฐศภรณ์ เหมรา

ปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชานวัตกรรมเทคโนโลยีดิจิทัล
มหาวิทยาลัยแม่โจ้
พ.ศ. 2564

ระบบสนับสนุนการตัดสินใจเพื่อแนะนำการลงทะเบียนรายวิชาให้นักศึกษา สาขาวิชาวิทยาการคอมพิวเตอร์

ณรัฐศภรณ์ เหมรา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของความสมบูรณ์ของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชานวัตกรรมเทคโนโลยีดิจิทัล
สำนักบริหารและพัฒนาวิชาการ มหาวิทยาลัยแม่โจ้
พ.ศ. 2564

ลิขสิทธิ์ของมหาวิทยาลัยแม่โจ้

ระบบสนับสนุนการตัดสินใจเพื่อแนะนำการลงทะเบียนรายวิชาให้นักศึกษา
สาขาวิชาวิทยาการคอมพิวเตอร์

ณรัฐศภรณ์ เหมรา

วิทยานิพนธ์นี้ได้รับการพิจารณาอนุมัติให้เป็นส่วนหนึ่งของความสมบูรณ์ของการศึกษา
ตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชานวัตกรรมเทคโนโลยีดิจิทัล

พิจารณาเห็นชอบโดย **อาจารย์ที่ปรึกษา**

อาจารย์ที่ปรึกษาหลัก .....
(ผู้ช่วยศาสตราจารย์ ดร.พาสน์ ปราโมกข์ชน)
วันที่.....เดือน.....พ.ศ. .....

อาจารย์ที่ปรึกษาร่วม .....
(อาจารย์ ดร.กิติศักดิ์ โอสถานันต์กุล)
วันที่.....เดือน.....พ.ศ. .....

อาจารย์ที่ปรึกษาร่วม .....
(ผู้ช่วยศาสตราจารย์ ดร.ปวีณ เขื่อนแก้ว)
วันที่.....เดือน.....พ.ศ. .....

ประธานอาจารย์ผู้รับผิดชอบหลักสูตร .....
(อาจารย์ ดร.กิตติกร หาญตระกูล)
วันที่.....เดือน.....พ.ศ. .....

สำนักบริหารและพัฒนาวิชาการรับรองแล้ว .....
(รองศาสตราจารย์ ดร.ญาณิน โอภาสพัฒนกิจ)
รองอธิการบดี ปฏิบัติการแทน
อธิการบดี มหาวิทยาลัยแม่โจ้
วันที่.....เดือน.....พ.ศ. .....

ชื่อเรื่อง ระบบสนับสนุนการตัดสินใจเพื่อแนะนำการลงทะเบียนรายวิชาให้นักศึกษา สาขาวิชาวิทยาการคอมพิวเตอร์
ชื่อผู้เขียน นายณรัฐศภรณ์ เหมรา
ชื่อปริญญา วิทยาศาสตรมหาบัณฑิต สาขาวิชานวัตกรรมเทคโนโลยีดิจิทัล
อาจารย์ที่ปรึกษาหลัก ผู้ช่วยศาสตราจารย์ ดร.พาสน์ ปราโมกข์ชน

บทคัดย่อ

ผลการเรียนเป็นสิ่งที่สำคัญอย่างหนึ่งในการศึกษาที่มีผลกับนักศึกษาทุกคน ในการสมัครเข้าทำงานในอนาคต และโอกาศได้รับทุนการศึกษา ปัจจุบันมีการพัฒนาระบบการศึกษาอย่างต่อเนื่อง มีการคิดค้นแนวคิดและทฤษฎี รวมไปถึงการนำเทคโนโลยีใหม่มาใช้ในการศึกษา แต่อย่างไรก็ตามยังมีปัญหาที่ยังไม่ได้รับการแก้ไข ได้แก่ ปัญหาผลการเรียนของนักศึกษาลดลง ปัญหานักศึกษาเลือกวิชาเรียนไม่สอดคล้องกับทักษะการเรียนของนักศึกษา เป็นต้น งานวิจัยนี้นำเสนอกระบวนการคิด วิเคราะห์ และออกแบบแนวทางแก้ปัญหาผลการเรียนลดลง ผู้วิจัยดำเนินการวิจัยโดยใช้เทคนิคต่างๆ ด้านวิทยาการข้อมูลเพื่อค้นหาวิชาที่ให้ผลการเรียนอยู่ในเกณฑ์ดี เพื่อนำไปใช้ในระบบคำแนะนำให้นักศึกษาเลือกรายวิชาเรียน ด้วยกระบวนการ CRISP-DM (กระบวนการมาตรฐานอุตสาหกรรมข้ามสำหรับการทำเหมืองข้อมูล) ผู้วิจัยสร้างโมเดลแนะนำการเลือกวิชาเรียนจากข้อมูลนักศึกษาปี 2555 มาสองโมเดลมาเพื่อแนะนำให้กับนักศึกษาเทียบกับข้อมูลลงทะเบียนรายวิขาของนักศึกษาปี 2556 โดยเมื่อทดลองโมเดลแรกที่ใช้เทคนิคทางสถิติพบว่านักศึกษาเลือกวิชาตามที่ระบบแนะนำจำนวน 15 วิชา มี 13 วิชาที่มีนักศึกษาส่วนใหญ่เรียนในเกณฑ์ดี กล่าวคือระบบสามารถแนะนำวิชาได้ถูกต้อง 86.67 % และโมเดลที่สองการใช้กฎความสัมพันธ์ร่วมกับสถิติพบว่านักศึกษาสามารถเรียนวิชาที่แนะนำนั้นได้ผลการเรียนในเกณฑ์ดีหรือไม่ โดยหากแนะนำจากกฎความสัมพันธ์จะถูกต้องที่ 93.56 มีร้อยละวิชาที่มีผลที่ 50 และหากแนะนำจากสถิติจะถูกต้องที่ 79.89 และมีร้อยละวิชาที่มีผลที่ 100 หากนำทั้งสองวิธีมารวมกันระบบสามารถแนะนำวิชาได้ถูกต้อง 79.74 % และมีร้อยละวิชาที่มีผลที่ 100 การเลือกใช้โมเดลจากวิธีการทางสถิติอาจเหมาะกับสามารถนำไปแนะนำในวงกว้างๆ ของนักศึกษา ในขณะที่โมเดลที่สองเหมาะสมที่จะนำไปสร้างโปรแกรมระบบสนับสนุนการตัดสินใจสามารถสร้างความเชื่อมันได้มากกว่าโมเดลจากวิธีทางสถิติ โดยเฉพาะการแนะนำรายวิชาที่เจาะจง เพราะพิจรณาจากรายวิชาที่นักศึกษาเคยเรียนผ่านมาแล้วร่วมด้วย

คำสำคัญ : เทคนิคกฎความสัมพันธ์, การทำเหมืองข้อมูล, การพัฒนาการศึกษา, CRISP-DM

มหาวิทยาลัยแม่โจ้
MAEJO UNIVERSITY
แม่โจ้ ๒๘๙๙

Title DECISION SUPPORT SYSTEM FOR SUGGESTING COURSE REGISTRATION TO COMPUTER SCIENCE STUDENT

Author Mr. Naratsaporn Heamra

Degree Master of Science in Digital Technology Innovation

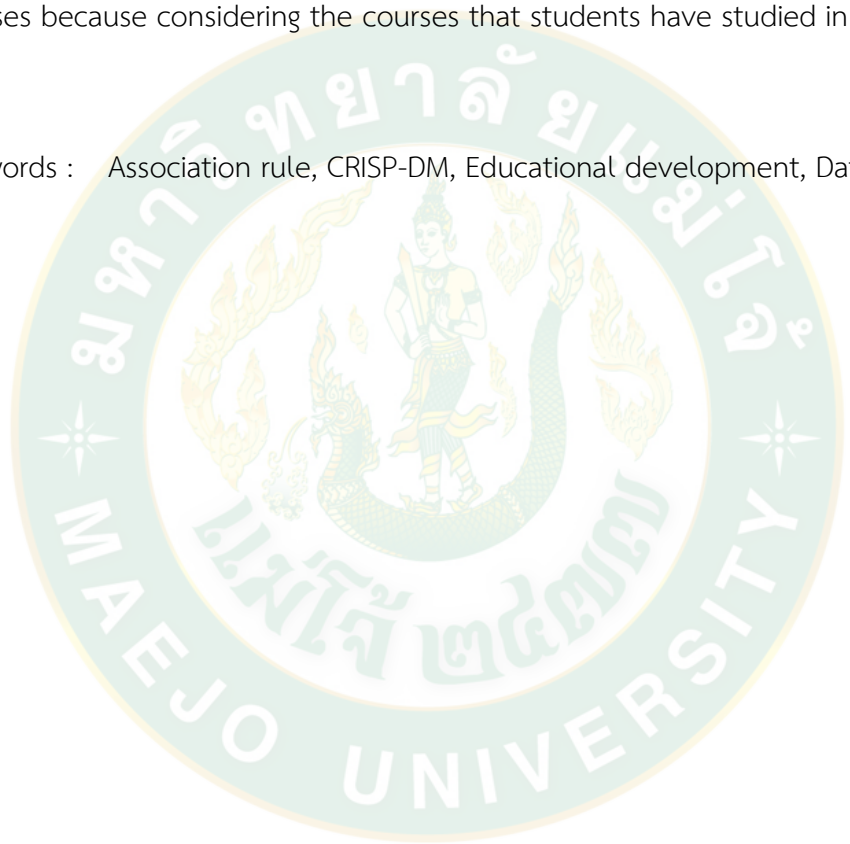
Advisory Committee Chairperson Assistant Professor Dr. Part Pramokchon

ABSTRACT

Academic study results are a significant affects all students for their career or the opportunity to receive scholarships in the future. Currently, there is the continuous development of the education system which applies many novel concepts and theories of innovation for education. However, there are still some unsolved problems, especially the issue of student's grades decreasing. The event of choosing courses to study is not consistent with the student's ability. This research focuses on the analytical processes to solve problems of academic study performance. This research uses various data science techniques to find out which courses are likely to perform well and recommend them to students. CRISP-DM (cross-industry standard process for data mining) is a standard framework of data science which is applied to construct a course recommender model. The researcher created two models obtained by using descriptive statistical techniques and Apriori rules mining. Corpus for constructing both models is course registration data collected in 2012-2013. This research uses 2012 data for training models and 2013 data for testing model performance. The first model created using a descriptive statistical technique recommends 15 courses for registration. These courses tend to get good grades. When we test the model against 2013 data, 13 courses also showed good study outcomes. This result presents that the system was able to recommend 86.67% of the courses correctly. The second model, the use of association-rules mining, found that students who enroll in the courses recommended by the model

also got good grades. The association-rule based model suggests correct courses at 93.56% with 50% of all. Meanwhile, the statistics model will be correct at 79.89% and with 100 % courses. When both models work together, it gets 79.74% accuracy and covers all courses. The statistical method is suitable for choosing roughly courses and covering all courses. The second model is suitable for creating the decision support system that could build more faith than the first model from the statistical technique. Especially the association-rule based model can introduce specific courses because considering the courses that students have studied in the past.

Keywords : Association rule, CRISP-DM, Educational development, Data Mining



กิตติกรรมประกาศ

งานวิจัยนี้บรรลุตามวัตถุประสงค์ได้ด้วยดี ผู้วิจัยต้องขอขอบคุณ อาจารย์ ดร. พาสน์ ปราโมกข์ชน อาจารย์ที่ปรึกษาวิทยานิพนธ์เป็นอย่างสูง ที่ได้ให้ความรู้ คำแนะนำ ในการค้นคว้าวิจัย จนงานวิจัยฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี ขอขอบคุณอาจารย์ที่ปรึกษาร่วม อาจารย์ ดร. กิติศักดิ์ โอสถานันต์กุล และ อาจารย์ ดร. ปวีณ เขื่อนแก้ว ที่ให้ความรู้ และขอบคุณ อาจารย์ประธานหลักสูตร อาจารย์ ดร. กิตติกร หาญตระกูล สำหรับโอกาสในครั้งนี้ ขอขอบคุณเจ้าหน้าที่ทุกท่านที่ให้ความช่วยเหลือประสานงาน และเทคนิคต่างๆ เป็นอย่างดี ขอขอบคุณสำนักบริหารและพัฒนาวิชาการสำหรับข้อมูลนักศึกษาที่นำมาใช้ในงานวิจัยนี้ ขอขอบคุณเพื่อนร่วมศึกษาทุกท่านที่ให้กำลังใจซึ่งกันเสมอมา สุดท้ายขอขอบคุณพระคุณบิดา มารดาและคนในครอบครัวที่เป็นแรงผลักดันจนมีวันนี้ สำหรับงานวิจัยนี้หากผิดพลาดประการใดผู้วิจัยขออภัยมา ณ ที่นี้ด้วย

ณรัฐศภรณ์ เหมรา

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ค

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ช

สารบัญ ..... ซ

สารบัญตาราง ..... ฎ

สารบัญภาพ ..... ฐ

บทที่ 1 บทนำ 1

1.1 ที่มาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์การวิจัย 1

1.3 ขอบเขตของการวิจัย 2

1.4 กรอบแนวคิดการวิจัย 2

1.5 ขั้นตอนในการสร้างและพัฒนาระบบ 3

1.6 เครื่องมือและอุปกรณ์ 4

1.7 ผลที่คาดว่าจะได้รับ 4

บทที่ 2 ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง 5

2.1 การทำเหมืองข้อมูล 5

2.2 ขั้นตอนการทำเหมืองข้อมูล 7

2.3 ประเภทของข้อมูลที่สามารถทำ Data Mining 7

2.4 ลักษณะเฉพาะของข้อมูลที่ใช้ทำเหมืองข้อมูล 7

2.5 ประโยชน์ของการใช้การทำเหมืองข้อมูล 8

2.6 แนวคิดในการพัฒนาการศึกษา 8

2.7 วรรณกรรมที่เกี่ยวข้องกับการพัฒนาการศึกษา 9
2.8 CRISP-DM Framework 10
2.9 วรรณกรรมที่เกี่ยวข้องกับ CRISP-DM 12
2.10 เทคนิคการทำเหมืองกฎความสัมพันธ์ (Association Rules Mining) 13
2.11 เทคนิคการใช้กฎความสัมพันธ์แบบอพริโอริ 18
2.12 วรรณกรรทที่เกี่ยวข้องกับเทคนิค Association Rules 20
2.13 สถิติกับการวิจัย 23
2.14 Weka (Waikato Environment for Knowledge Analysis) 25
2.15 หลักสูตรสาขาวิชาวิทยาการคอมพิวเตอร์ 26
บทที่ 3 วิธีการดำเนินงานวิจัย 35
3.1 การทำความเข้าใจปัญหา (Business Understanding) 35
3.2 การเตรียมข้อมูล (Data Understanding) 36
3.3 การแก้ไข และดัดแปลงข้อมูล (Data Preparation) 37
3.4 การสร้างและการทดสอบตัวแบบ (Modeling) 39
3.4.1 การคัดกรองข้อมูลตามสถิติ 39
3.5 การสร้างต้นแบบจากสถิติเป็นโมเดลลำดับแรก สร้างจากการใช้สถิติ 40
3.6 การสร้างโมเดลที่สอง โดยใช้การทำเหมืองกฎความสัมพันธ์ 44
3.6.1 กฎความสัมพันธ์ (Association Rules) 44
3.6.2 การวัดประสิทธิภาพของกฎความสัมพันธ์ (Support & Confidence) 44
3.6.3 การสร้างกฎความสัมพันธ์ด้วย Apriori Algorithm 44
3.6.4 การหากฎความสัมพันธ์ Apriori โดยโปรแกรม Weka 44
3.7 การออกแบบการทำงานของโปรแกรม 54
3.7.1 การนำกฎความสัมพันธ์จำนวนมากเข้าสู่ฐานข้อมูล 54
3.7.2 การออกแบบข้อมูลนำเข้า 55

3.8 วิธีการวัดประสิทธิภาพข้อมูล 60
3.9 ผลการวัดประสิทธิภาพ 60
3.10 สรุปผลการวิจัยและจัดทำรายงานวิทยานิพนธ์ 62
บทที่ 4 ผลการดำเนินการวิจัย 63
4.4 ผลการทดสอบความแม่นยำของตัวต้นแบบ (Evaluation) 64
4.6 การนำไปใช้งาน (Deployment) 65
4.6.1 การออกแบบข้อมูลนำเข้า 66
4.6.2 การแสดงผล 67
บทที่ 5 สรุปและข้อเสนอแนะ 72
5.1 สรุปผลการวิจัย 72
5.2 ปัญหาและอุปสรรค 73
5.3 ข้อเสนอแนะ 73
ภาคพนวก ก 75
ภาคพนวก ข 80
บรรณานุกรม 84
ประวัติผู้วิจัย 88

สารบัญตาราง

หน้า

ตารางที่ 1 แสดงข้อมูลไอเทมเซตของนักศึกษาแต่ละคนจำนวน 6 คน 15

ตารางที่ 2 แสดงค่า Transaction ของ k-itemsets 15

ตารางที่ 3 แสดงตัวอย่างการคำนวณค่า support ของวิชา 2 วิชา 16

ตารางที่ 4 แสดงตัวอย่างการคำนวณค่า support ของวิชา 3 วิชา 16

ตารางที่ 5 กลุ่มวิชาสังคมศาสตร์ หน่วยกิตต่ำสุด : 6 27

ตารางที่ 6 กลุ่มวิชามนุษยศาสตร์ หน่วยกิตต่ำสุด : 9 27

ตารางที่ 7 กลุ่มวิชาภาษาหน่วยกิตต่ำสุด : 9 28

ตารางที่ 8 กลุ่มวิชาวิทยาศาสตร์และคณิตศาสตร์ หน่วยกิตต่ำสุด : 6 28

ตารางที่ 9 กลุ่มวิชาแกน หน่วยกิตต่ำสุด : 30 29

ตารางที่ 10 กลุ่มวิชาเอกบังคับ หน่วยกิตต่ำสุด : 58 30

ตารางที่ 11 กลุ่มวิชาเอกเลือก หน่วยกิตต่ำสุด : 12 33

ตารางที่ 12 ข้อมูลดิบของนักศึกษาที่นำมาใช้ในการวิจัยครั้งนี้ 37

ตารางที่ 13 ตัวอย่างข้อมูลที่มีการดัดแปลงแล้ว 38

ตารางที่ 14 เกณฑ์การวัดผลมาตรฐานการศึกษาของประเทศไทย 38

ตารางที่ 15 ตัวอย่างข้อมูลที่มีการแก้ไขแล้ว 39

ตารางที่ 16 ตัวอย่างการหาสัดส่วน (Proportion) จากข้อมูลการลงทะเบียนของนักศึกษาปี 2555 41

ตารางที่ 17 จำนวนกฎที่ได้จากการค้นหากฎความสัมพันธ์ ด้วยค่า Support และ Confidence ต่างกัน ในการหมวดวิชาภายนอก 49

ตารางที่ 18 ตัวอย่างกฎความสัมพันธ์ 10 อันดับเรียงจากค่า Confidence จากมากไปน้อย 50

ตารางที่ 19 ตัวอย่างกฎความสัมพันธ์ 10 อันดับเรียงจากค่า lift จากมากไปน้อย 52

ตารางที่ 20 การแยกข้อมูลเพื่อนำไปประมวลผล 56

ตารางที่ 21 การแจกแจงรอบการทำงานจากตัวอย่างขั้นตอนที่ 2 59
ตารางที่ 22 การเปรียบเทียบจำนวนเปอร์เซ็นต์ถูกของแต่ละวิธี 61
ตารางที่ 23 ผลการทดสอบกับข้อมูลชุดเดิม 64
ตารางที่ 24 ผลการทดสอบกับข้อมูลอีกชุด 65
ตารางที่ 25 ความพึงพอใจของนักศึกษาที่ทดลองใช้งานระบบ 71
ตารางที่ 25 แสดงรายละเอียดวิชาที่ปรากฏในงานวิจัย 81

สารบัญภาพ

หน้า

ภาพที่ 1 กรอบแนวคิดในการวิจัย 3

ภาพที่ 2 แสดงกระบวนการค้นหาความรู้จากคลังข้อมูล 5

ภาพที่ 3 แสดงเทคนิคต่างๆของการค้นหาความรู้จากคลังข้อมูล 6

ภาพที่ 4 กราฟแสดงปัญหาผลการเรียนลดลงของนักศึกษาชั้นปีที่ 3 8

ภาพที่ 5 กระบวนการทำงานของ CRISP-DM 12

ภาพที่ 6 โครงสร้างแลตทิซแบบปกติ 19

ภาพที่ 7 โครงสร้างแลตทิซที่มีการใช้เทคนิคการพรุน 19

ภาพที่ 8 กราฟแสดงผลร้อยละของเกรดเฉลี่ยของนักศึกษาชั้นปีที่ 3 35

ภาพที่ 9 ฮีสโตแกรมแสดงความถี่ของรายวิชาและจำนวนนักศึกษาที่ลงทะเบียน 40

ภาพที่ 10 แสดงเปอร์เซ็นต์รายวิชาตามการคำนวณสถิติ 41

ภาพที่ 11 กราฟแสดงผู้ให้ข้อมูลแบบสอบถามของนักศึกษาตามชั้นปี 42

ภาพที่ 12 แบบสอบถามความพึงพอใจของผู้ทดลองใช้งาน 42

ภาพที่ 13 โมเดลสถิติจากแบบสอบถามความพึงพอใจของผู้ทดลองใช้งาน 43

ภาพที่ 14 ข้อมูลที่นำมาทดสอบกับโมเดลที่หนึ่ง 43

ภาพที่ 15 หน้าจอเริ่มใช้งานของโปรแกรม Weka 45

ภาพที่ 16 หน้าจอเลือกใช้งาน 45

ภาพที่ 17 เมนูเลือกไฟล์ข้อมูลที่นำเข้าไปประมวลผล 46

ภาพที่ 18 ไฟล์ข้อมูลที่นำเข้าไปประมวลผล 46

ภาพที่ 19 โปรแกรมโดยข้อมูลจะปรากฏขึ้นหลังจากเลือกไฟล์ 47

ภาพที่ 20 หน้าจอการใช้งาน Associate 47

ภาพที่ 21 หน้าจอการตั้งค่า Apriori algorithm 48

ภาพที่ 22 ตัวอย่างกฎความสัมพันธ์ที่ได้จากการใช้โปรแกรม..... 49
ภาพที่ 23 ผังงาน (Flowchart) แสดงการทำงานของโปรแกรมแนะนำการศึกษา * 54
ภาพที่ 24 ผังงาน (Flowchart) แสดงเงื่อนไขการทำงานภายในโปรแกรม 57
ภาพที่ 25 กราฟเปรียบเทียบจำนวนวิชาที่แนะนำ และจำนวนที่แนะนำถูกต้องของแต่ละวิธี..... 60
ภาพที่ 26 แผนผังสรุปการวิจัย..... 62
ภาพที่ 27 การเปรียบเทียบค่าความถูกต้องการทดสอบโมเดลกับข้อมูลชุดเดิม และชุดใหม่..... 65
ภาพที่ 28 หน้าจอการกรอกข้อมูลที่ต้องการนำไปค้นหาวิชาที่แนะนำ 66
ภาพที่ 29 หน้าจอแสดงรายงานผลการค้นหารายวิชาที่ควรเลือก 67
ภาพที่ 30 หน้าจอแสดงรายงานผลการค้นหารายวิชาที่ควรเลือกแบบกฎความสัมพันธ์ 68
ภาพที่ 31 หากไม่ค้นพบรายวิชาจะปรากฏปุ่มเพื่อค้นหาแบบสถิติ..... 68
ภาพที่ 32 หน้าจอแสดงผลการค้นหารายวิชาจากสถิติ 69
ภาพที่ 33 หน้าจอแสดงผลการค้นหารายวิชาจากสถิติ 70

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

ในปัจจุบันข้อมูลมีปริมาณมากขึ้นตามเทคโนโลยีสารสนเทศที่พัฒนาขึ้นโดยเฉพาะข้อมูลในสถานศึกษา การวิเคราะห์ข้อมูลการศึกษามาเป็นเรื่องที่นิยมมากในปัจจุบัน หรือเรียกกันว่าดาต้าไมนิง หรือ การทำเหมืองข้อมูล (Data Mining) การทำเหมืองข้อมูลเป็นกระบวนการหารูปแบบและความสัมพันธ์ในชุดข้อมูล การทำเหมืองข้อมูลกำลังได้รับความนิยมอย่างสูงในปัจจุบัน และมีการใช้งานประยุกต์ใช้ในหลายรูปแบบอย่างกว้างขวางโดยเฉพาะเกี่ยวข้องกับการศึกษา เช่น ปัจจัยข้อมูลนักศึกษาที่มีผลต่อระดับผลการเรียน การค้นหาภาควิชาที่เหมาะสมที่สุดให้กับนักเรียน เป็นต้น ในงานวิจัยนี้มีการนำเสนอเทคนิคการต่างๆมาประยุกต์ใช้ในการช่วยแก้ปัญหาผลการเรียนลดลงของนักศึกษา เนื่องจากผลการศึกษาเป็นสิ่งสำคัญ เป็นโอกาสในการสำเร็จการศึกษา การรับทุนการศึกษาและเลือกหางานในอนาคต อีกทั้งในปัจจุบันมีงานวิจัยที่เกี่ยวข้องกับการพัฒนาการศึกษามากมาย แต่การนำไปใช้งานกับกลุ่มเป้าหมาย หรือตัวนักศึกษาโดยตรงยังมีน้อย ในงานวิจัยนี้จึงจะมีการนำเสนอในส่วนนี้ด้วย

เกณฑ์มาตรฐานหลักสูตรสาขาวิทยาการคอมพิวเตอร์กำหนดให้นักศึกษาแต่ละคนต้องลงทะเบียนศึกษาไม่น้อยกว่า 40 วิชา โดยจากข้อมูลของนักศึกษาวิทยาการคอมพิวเตอร์ ระหว่างปี 2555 – 2556 พบว่านักศึกษาปีชั้นปีที่ 2 ถึง 3 คือช่วงที่มีลงทะเบียนมากที่สุดคือ นอกจากนั้นพบว่ามีนักศึกษาจำนวนไม่ต่ำกว่า 35% ที่เกรดเฉลี่ยลดลง ผลลัพธ์ของการสืบค้นความรู้นี้จะช่วยให้นักศึกษาเลือกเรียนวิชาที่ทำให้ผลการเรียนของตนเองอยู่ในเกณฑ์ดี โดยการเลือกวิชานั้นจะต้องเหมาะสมกับความรู้และความสามารถของนักศึกษาแต่ละบุคคล

การทำเหมืองข้อมูล ได้ถูกนำมาประยุกต์ใช้ในหลายประเภท เช่น ด้านธุรกิจ ด้านการแพทย์ ด้านการศึกษา เป็นต้น ในการวิจัยนี้นำเสนอใช้เทคนิคกฎความสัมพันธ์ร่วมกับสถิติเป็นหนึ่งในขั้นตอนการทำเหมืองข้อมูลเพื่อใช้ในการพัฒนาการศึกษา การสนับสนุนการศึกษาโดยการแนะนำวิชาที่ทำให้นักศึกษาคาดการณ์ว่าเมื่อเรียนแล้วจะได้ผลการเรียนที่อยู่ในเกณฑ์ดี เกณฑ์ดีหมายถึงเกรด B ขึ้นไปตามการวัดผลระดับผลการเรียนหลักสูตรการศึกษาขั้นพื้นฐานของประเทศไทย

1.2 วัตถุประสงค์การวิจัย

1. สร้างโมเดลโดยวิธีการทำเหมืองข้อมูลจากผลการเรียนของนักศึกษาโดยใช้เทคนิคต่างๆ เพื่อเปรียบเทียบและประยุกต์ใช้ในการพัฒนาระบบแนะนำการลงทะเบียนของนักศึกษา และงานอื่น ๆ ที่

เกี่ยวข้องเพื่อช่วยลดโอกาสการมีผลการเรียนไม่ถึงเกณฑ์ดีของนักศึกษาให้ลดลงตัวอย่างเช่น รูปแบบกฎความสัมพันธ์ของรายวิชาที่มีผลต่อผลการเรียนของนักศึกษา

2. สร้างแอพพลิเคชั่นระบบแนะนำการลงทะเบียนโดยใช้โมเดลจากเหมืองข้อมูล เพื่อเป็นการนำผลลัพธ์จากการศึกษาโมเดลในวัตถุประสงค์ แรกไปใช้งานจริง

1.3 ขอบเขตของการวิจัย

1. พัฒนาโมเดลต้นแบบที่สามารถบอกวิชาที่เหมาะสมกับนักศึกษาแต่ละบุคคล
2. นักศึกษาสามารถพิจารณาวิชาที่เหมาะสมกับตัวเองจากการแนะนำของระบบ
3. ลดระยะเวลาการวางแผนการเรียน จากการแนะนำการเลือกลงทะเบียนเรียนจากระบบ
4. ลดความเสี่ยงที่นักศึกษาเลือกลงทะเบียนเรียนแล้วไม่ได้ผลการเรียนที่ดี

1.4 กรอบแนวคิดการวิจัย

ในกระบวนการทดลองการวิจัยได้แบ่งกระบวนการทำงานแสดงดังภาพที่ 1 โดย

1. ศึกษาข้อมูลหลักสูตร และวิธีการเลือกลงทะเบียนของนักศึกษา เป็นการศึกษาข้อมูลพื้นฐานของหลักสูตรที่เกี่ยวข้องกับงานวิจัย เช่น โครงสร้างหลักสูตร วิชาที่เกี่ยวข้อง เป็นต้น

2. เตรียม ตรวจสอบความถูกต้อง และแก้ไขข้อมูล เป็นการศึกษารูปแบบของข้อมูลที่เกี่ยวข้องกับงานวิจัย พร้อมกับตรวจสอบข้อมูลเพื่อหาความผิดปกติของข้อมูล และแก้ไขข้อมูลก่อนทำการวิเคราะห์ข้อมูลดังนั้นข้อมูลใดๆไม่มีความชัดเจนก็จะถูกคัดออกไป

3. ใช้เทคนิคเหมืองข้อมูล (Data Mining) โดยอาศัยหลักสถิติและเทคนิคเหมืองข้อมูลเป็นกระบวนการในการค้นหารูปแบบ (pattern) หรือความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ในชุดข้อมูล

4. การประเมินผล (Evaluation) เป็นกระบวนการในการตัดสินผลลัพธ์การวิจัย คือเทคนิคเหมืองข้อมูลที่ใช้ โดยนำผลที่ได้จากการการประเมินผลนำมาเปรียบเทียบ เพื่อนำไปใช้งานแอพพลิเคชั่นระบบแนะนำการลงทะเบียน

ภาพที่ 1 กรอบแนวคิดในการวิจัย

1.5 ขั้นตอนในการสร้างและพัฒนาระบบ

1. ศึกษาข้อมูลนักศึกษา และหลักสูตรการศึกษาวิทยาการคอมพิวเตอร์ตั้งแต่ปีการศึกษา 2555 -2556 คณะวิทยาศาสตร์ มหาวิทยาลัยแม่โจ้

2. ศึกษาทฤษฎี และกระบวนการมาตรฐานการทำเหมืองข้อมูลโดยใช้กระบวนการ CRISP-DM

3. ศึกษาวิธีการทำเหมืองข้อมูลโดยใช้เทคนิคต่างๆ

4. คัดกรองข้อมูลผลการเรียนของนักศึกษา โดยเฉพาะตรวจสอบความสมบูรณ์ของข้อมูล

5. นำเทคนิคต่างๆ สร้างโมเดลการเลือกลงทะเบียนเรียนของนักศึกษาสาขาวิทยาการคอมพิวเตอร์

6. ทดสอบประสิทธิภาพโมเดลโดยพิจารณาความถูกต้องแม่นยำ

7. สรุปผลการวิจัย และจัดทำรายงาน

1.6 เครื่องมือและอุปกรณ์

1. ฮาร์ดแวร์

    1.1 คอมพิวเตอร์แบบพกพา

2. ซอฟต์แวร์

    2.1 Weka 3.4

    2.2 PhpMyAdmin

    2.3 Subline 3

1.7 ผลที่คาดว่าจะได้รับ

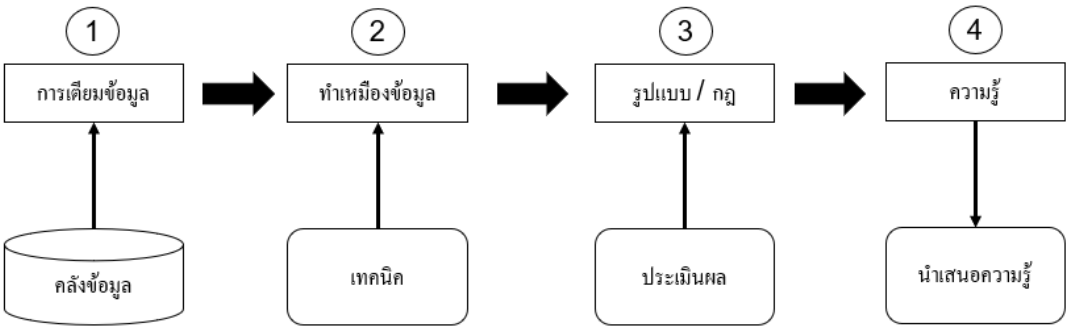
1. พัฒนาโมเดลต้นแบบที่สามารถบอกวิชาที่เหมาะสมกับนักศึกษาแต่ละบุคคล

2. นักศึกษาสามารถพิจารณาวิชาที่เหมาะสมกับตัวเองจากการแนะนำของระบบ เพื่อลดระยะเวลาการวางแผนการเรียน จากการแนะนำการเลือกลงทะเบียนเรียนจากระบบ และลดความเสี่ยงที่นักศึกษาเลือกลงทะเบียนเรียนแล้วไม่ได้ผลการเรียนที่ดี (ตั้งแต่ B ขึ้นไป)

บทที่ 2
ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง

2.1 การทำเหมืองข้อมูล

การทำเหมืองข้อมูล (Data Mining) คือ การรวบรวมข้อมูลจากฐานข้อมูลขนาดใหญ่เพื่อค้นหาความรู้ ในรูปแบบความสัมพันธ์ต่างๆและรูปแบบของกฎความสัมพันธ์ (สารานุกรมเสรี, 2562) ด้วยการสำรวจและวิเคราะห์ให้อยู่ในรูปแบบความหมายของรูปแบบความสัมพันธ์ต่างๆ และรูปแบบของกฎความสัมพันธ์ กระบวนการที่จะค้นหาความรู้จากคลังข้อมูลนั้นมีขั้นตอนการค้นหาหลัก 4 ขั้นตอน คือ การเตรียมข้อมูลส่วนใหญ่มักมาจากฐานข้อมูล การสำรวจแล้วบันทึกผลการสำรวจ การสร้างแบบสอบถาม โดยบางครั้งอาจมีการดัดแปลง และแก้ไข ขั้นต่อมาคือการทำเหมืองข้อมูลนั้นคือการเลือกใช้เทคนิคที่เหมาะสมกับข้อมูลที่เตรียม จากนั้นจะได้รูปแบบ หรือกฎความสัมพันธ์ และประเมินความแม่นยำของรูปแบบ หรือกฎความสัมพันธ์นั้นๆ เพื่อนำความรู้ที่ได้จากรูปแบบ หรือกฎความสัมพันธ์นั้นนำเสนอต่อผู้ใช้งานต่อไป



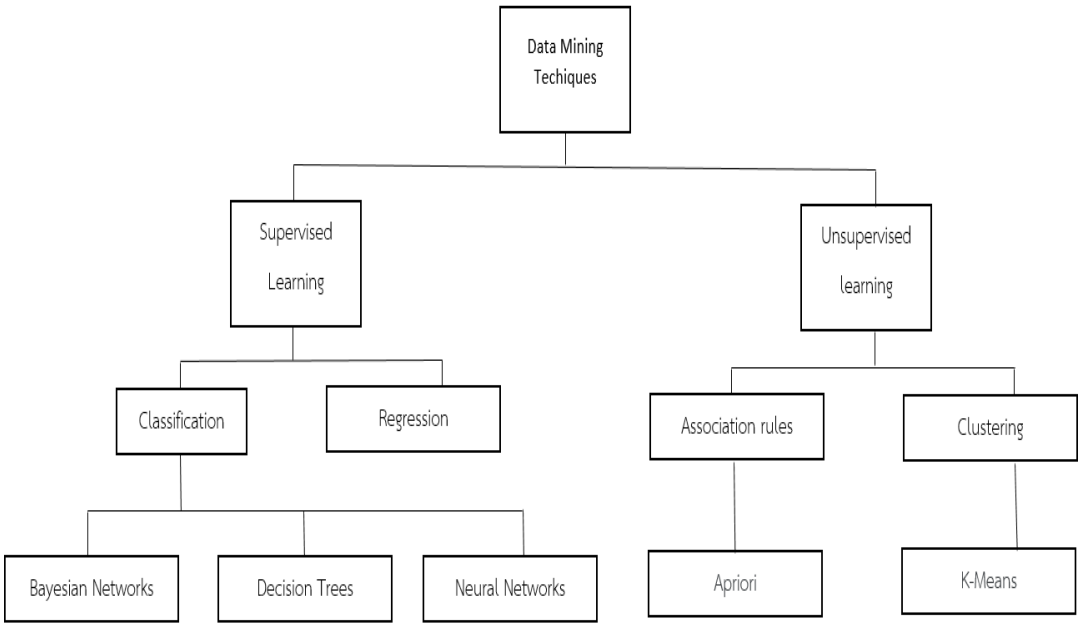
ภาพที่ 2 แสดงกระบวนการค้นหาความรู้จากคลังข้อมูล

Supervised Learning จะมีต้นแบบที่เป็นเป้าหมาย (Target) หรือ ผลเฉลย (Answer) ในขณะที่ Unsupervised Learning จะไม่มีเป้าหมาย (Target) หรือ ผลเฉลย (Answer) ในการบอกว่าโมเดลที่ได้สร้างนั้นถูกหรือผิด เช่น การทำนายต่างๆ อาจเป็นยอดขายสินค้า จะใช้ข้อมูลในอดีต ที่รู้ว่ายอดขายเกิดขึ้นจริงเท่าไหร่ เป็น Target หรือ Answer และใช้การเปรียบเทียบปัจจัยต่างๆ อาจเป็นตัวแปร หรือ Input กรณีนี้จะเป็น Supervised Learning ประเภทการถดถอย (Regression) ซึ่งจะมีข้อตกลงเบื้องต้น (Assumption) การวิเคราะห์กลุ่มลูกค้าจากพฤติกรรมการซื้อสินค้าในระบบ ซึ่ง

ถ้าหากก่อนหน้านี้ไม่มีการแบ่งกลุ่มมาก่อนเป็นการใช้ Unsupervised Learning ประเภท Clustering เป็นต้น

Unsupervised learning เป็นกลุ่ม algorithm ที่ไม่มีเป้าหมาย (Target) หรือ ผลเฉลย (Answer) อย่างชัดเจนว่าถ้าทำงานแล้วได้ผลลัพธ์แล้วถูกหรือผิด ตัวอย่างเช่นถ้ามีนักเรียน 100 คนในห้องแล้วให้เราทำการแบ่งกลุ่ม 100 คนนั้นเป็น 2 กลุ่ม ในกรณีนี้จะไม่มีกฎชัดเจนว่าควรแบ่งตามเพศ , อายุ , สีเสื้อหรืออื่นๆ แค่แบ่งออกมาให้ได้ 2 กลุ่มซึ่งอาจจะแบ่งว่า นักเรียนคนไหนอยู่กลุ่มห้องส่วนหน้าคือกลุ่ม 1 ส่วนใครอยู่ครึ่งห้องหลังคือกลุ่ม 2 เป็นต้น

Association Rules เป็นกฎการหาความสัมพันธ์ในรูปแบบความน่าจะเป็นในรูปแบบหนึ่ง ที่ใช้สำหรับการวิเคราะห์โอกาสที่เกิดร่วมกันในเหตุการณ์ต่างๆหลายเหตุการณ์ ตัวอย่างเช่น การวิเคราะห์การขายสินค้าของร้านขายสินค้าแห่งหนึ่ง โดยอาศัยกฎการหาความสัมพันธ์ของสินค้าที่มักจะขายพร้อมกัน ข้อมูลการขายสินค้าสามารถนำมาช่วยในการวางแผนทางการตลาด เช่น การจัดโปรโมชั่นให้กับสินค้า การนำสินค้ามาจัดวางใกล้กัน เป็นต้น กฎการหาความสัมพันธ์ถือเป็น Unsupervised Learning รูปแบบหนึ่ง ดังภาพที่ 3



ภาพที่ 3 แสดงเทคนิคต่างๆของการค้นหาความรู้จากคลังข้อมูล

2.2 ขั้นตอนการทำเหมืองข้อมูล

- Data Cleaning เป็นขั้นตอนสำหรับการคัดข้อมูลที่ไม่เกี่ยวข้องออกไป
- Data Integration เป็นขั้นตอนการรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกัน
- Data Selection เป็นขั้นตอนการดึงข้อมูลสำหรับการวิเคราะห์จากแหล่งที่บันทึกไว้
- Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน
- Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่
- Pattern Evaluation เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล
- Knowledge Representation เป็นขั้นตอนการนำเสนอความรู้ที่ค้นพบโดยใช้ เทคนิคในการนำเสนอเพื่อให้เข้าใจ

2.3 ประเภทของข้อมูลที่สามารถทำ Data Mining

- Relational databases ฐานข้อมูลที่จัดเก็บอยู่ในรูปแบบของตาราง โดยในแต่ละตารางจะประกอบไปด้วย แถวและคอลัมน์ ความสัมพันธ์ของข้อมูลทั้งหมดสามารถแสดงได้โดย Entity Relationship Model
- Data Warehouses เป็นการเก็บรวบรวมข้อมูลจากหลายแหล่งมาเก็บไว้ในรูปแบบเดียวกันและรวบรวมไว้ ในที่เดียวกัน
- Transactional databases ประกอบด้วยข้อมูลที่แต่ละทรานแซกชันแทนด้วยเหตุการณ์ในขณะใดขณะหนึ่ง เช่น ใบเสร็จรับเงิน จะเก็บข้อมูลในรูปชื่อลูกค้าและรายการสินค้าที่ลูกค้ารายซื้อ
- Advanced database เป็นฐานข้อมูลที่จัดเก็บในรูปแบบอื่นๆ เช่น ข้อมูลแบบ Object oriented ข้อมูลที่เป็น Text file ข้อมูลมัลติมีเดีย ข้อมูลในรูปของ Web Site

2.4 ลักษณะเฉพาะของข้อมูลที่ใช้ทำเหมืองข้อมูล

- ข้อมูลขนาดใหญ่ เกินกว่าจะพิจารณาความสัมพันธ์ที่ซ่อนอยู่ภายในข้อมูลได้ด้วยตาเปล่าหรือโดยการใช้ Database Management System (DBMS) ในการจัดการฐานข้อมูล
- ข้อมูลที่มาจากหลายแหล่งโดยอาจรวบรวมมาจากหลายระบบปฏิบัติการหรือหลาย DBMS เช่น Oracle , DB2 , MS SQL , MS Access เป็นต้น
- ข้อมูลที่มีโครงสร้างซับซ้อน เช่น ข้อมูลรูปภาพ ข้อมูลมัลติมีเดีย ข้อมูลเหล่านี้ สามารถนำ มาทำ Mining ได้เช่นกันแต่ต้องใช้เทคนิคการทำ Data Mining ขั้นสูง
- ข้อมูลที่ไม่มีการเปลี่ยนแปลงตลอดช่วงเวลาที่ทำการ Mining หากข้อมูลที่มีอยู่นั้น เป็นข้อมูลที่เปลี่ยนแปลงตลอดเวลาจะต้องแก้ปัญหานี้ก่อน โดยบันทึกฐานข้อมูลนั้น ไว้และนำฐานข้อมูลที่บันทึกไว้มาทำ Mining แต่เนื่องจากข้อมูลนั้นมีการ เปลี่ยนแปลงอยู่ตลอดเวลา

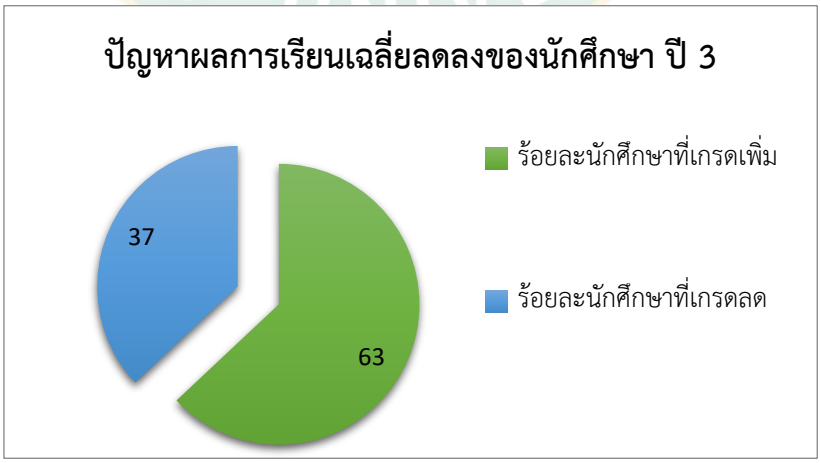
จึงทำให้ผลลัพธ์ที่ได้จาการทำ Mining สมเหตุสมผล ในช่วงเวลาหนึ่งเท่านั้น ดังนั้นเพื่อให้ได้ผลลัพธ์ที่มีความถูกต้องเหมาะสมอยู่ ตลอดเวลาจึงต้องทำ Mining ใหม่ทุกครั้งในช่วงเวลาที่เหมาะสม

2.5 ประโยชน์ของการใช้การทำเหมืองข้อมูล

ข้อมูลถูกจัดเก็บเพื่อนำไปสร้างระบบการสนับสนุนการตัดสินใจ (Decision Support System) เพื่อเป็นการง่ายต่อการนำข้อมูลมาใช้ในการวิเคราะห์เพื่อการตัดสินใจ ส่วนมากข้อมูลจะถูกจัดเก็บแยกมาจากระบบปฏิบัติการ (Operational System) โดยจัดอยู่ในรูปของคลังหรือเหมืองข้อมูล(Data Warehouse) เป็นการง่ายต่อการนำเอาไปใช้ในการสืบค้นความรู้

2.6 แนวคิดในการพัฒนาการศึกษา

เนื่องจากผู้วิจัยในขณะที่เป็นนักศึกษามักมีปัญหาเรื่องเกี่ยวเลือกลงทะเบียนเรียนร่วมกับเพื่อนๆในสาขา ซึ่งจะเลี่ยงไม่ได้เลยว่าผลการเรียนนั้นมีผลสำคัญในการเลือกลงทะเบียนเพราะนักศึกษาส่วนใหญ่บางครั้งเลือกลงทะเบียนเรียนตามความชอบ บางครั้งอาจไม่ได้คำนึงถึงแต่มักจำคำนึงถึงผลลัพธ์มากกว่าความชอบเพราะต้องการให้เกรดเฉลี่ยออกมาดี ผู้วิจัยจึงมีแนวคิดที่จะพัฒนาระบบช่วยแนะนำผลการเรียนของนักศึกษาโดยคำนึงถึงผลลัพธ์หรือเกรดแต่ละรายวิชาให้ออกมาดีที่สุด โดยจากภาพที่ 3 แสดงข้อมูลนักศึกษาจากปี 2555 เปรียบเทียบกับปี 2556 โดยเลือกเฉพาะชั้นปีที่ 3 เนื่องจากเป็นชั้นปีที่มีการลงเรียนในกลุ่มวิชาทั่วไป และวิชาอิสระมากที่สุด ผลปรากฏว่ามีนักศึกษาชั้นปีที่ 3 เกรดเฉลี่ยลดลงจากชั้นปีที่ 2 ร้อยละ 37 ซึ่งงานวิจัยนี้ต้องการที่จะช่วยลดจำนวนนี้



ภาพที่ 4 กราฟแสดงปัญหาผลการเรียนลดลงของนักศึกษาชั้นปีที่ 3

2.7 วรรณกรรมที่เกี่ยวข้องกับการพัฒนาการศึกษา

บทความที่มีชื่อว่า “ DATA MINING APPROACH FOR PREDICTING STUDENT PERFORMANCE ” (Osmanbegović และSuljić, 2555) กล่าวว่าที่ผ่านมาได้มีงานวิจัยที่ใช้เทคนิคการทำเหมืองข้อมูลในการวิเคราะห์ข้อมูลเกี่ยวกับสถานศึกษา เช่น การวิเคราะห์ความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาที่มีผลต่อระดับผลการเรียนนำเสนอการวิเคราะห์กระบวนการตัดสินใจด้วยเทคนิคต้นไม้ตัดสินใจเป็นเทคนิคอย่างหนึ่งในการทำเหมืองข้อมูลซึ่งอาศัยวามสัมพันธ์ของปัจจัยข้อมูลนักศึกษาเป็นการบ่งชี้ถึงระดับผลการเรียนของนักศึกษา

บทความที่ชื่อว่า “ Self-regulated learning with the Second Life environment: An empirical Study ” (Al-HatemMasood และ Al-Samarraie, 2561) กล่าวว่าการใช้กฎความสัมพันธ์เพื่อคาดการณ์ผลการเรียนของนักศึกษา เพื่อสามารถแจ้งเตือนนักศึกษาให้ปรับวิธีการเรียนในวิชาเรียนที่เหลือ เพื่อให้มีผลการเรียนเฉลี่ยรวมที่สูงขึ้น และสามารถผ่านเกณฑ์ได้ เมื่อเรียนจนครบหลักสูตร

บทความที่ชื่อว่า “ Decision Support System of Scholarship Grantee Selection using Data Mining ” (Sugiyarti และคณะ, 2561) การใช้เทคนิคต่างๆเพื่อค้นหาภาควิชาที่เหมาะสมที่สุดให้กับนิสิต เพื่อที่เป็นทางเลือกให้กับนักศึกษาเลือกเรียนโดยพิจารณาจากทักษะความสามารถของตน เพื่อสร้างตัวแบบสำหรับปัจจัยที่มีผลต่อการตัดสินใจเลือกสาขาวิชาเรียนของนักศึกษา โดยการใช้เทคนิคการทำเหมืองข้อมูลด้วยวิธี K-fold cross validation พบว่าสามารถบอกตัวแปรสำคัญที่มีผลต่อการตัดสินใจในการเลือกสาขาวิชาเรียนของนักศึกษาในระดับอุดมศึกษาได้

บทความที่ชื่อว่า “ Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining ” (BansalSharma และGoel, 2560) การใช้เทคนิคเหมืองข้อมูลในการหาความถนัด เพื่อพัฒนาศักยภาพของนักศึกษา ใช้เทคนิคการจัดกลุ่มด้วยอัลกอริธึม K-mean กำหนดการจัดกลุ่มนักศึกษาได้ เป็น 2 กลุ่ม กลุ่มนักศึกษาที่ด้อย กลุ่มนักศึกษาที่เด่น เพื่อเป็นข้อมูลให้อาจารย์ที่ปรึกษาหรืออาจารย์ผู้สอนได้ใช้ประโยชน์ในการดูแลนักศึกษาได้ตรงกลุ่มเป้าหมายมากขึ้น เพื่อลดความเสี่ยงของนักศึกษาที่จะมีผลการเรียนที่ไม่ผ่านเกณฑ์ของแต่ละวิชา

2.8 CRISP-DM Framework

CRISP-DM เป็นเหมือน blueprint ที่ใช้กันอย่างกว้างขวาง เช่นเดียวกันกับกระบวนการ ISO ในโรงงานอุตสาหกรรม หรือกระบวนการ CMMI ซึ่งเป็นมาตรฐานในการพัฒนาซอฟต์แวร์ กระบวนการมาตรฐานในการวิเคราะห์ข้อมูลด้านเหมืองข้อมูล (Kanwaseth, 2019) พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือกันของ 3 บริษัท คือ DaimlerChrysler SPSS และNCR กระบวนการทำงานนี้เรียกว่า “ Cross-Industry Standard Process for Data Mining ” หรือเรียกย่อ ว่า “ CRISP-DM ” ในกระบวนการนี้ประกอบด้วย 6 ขั้นตอน ดังภาพที่ 5

ขั้นตอนที่ 1 Business Understanding เป็นขั้นตอนแรกในกระบวนการ CRISP-DM ซึ่งเน้นไปที่การเข้าใจปัญหาและแปลงปัญหาที่ได้ให้อยู่ในรูปโจทย์ของการวิเคราะห์ข้อมูลทางการทำเหมืองข้อมูล (Data Mining) พร้อมทั้งวางแผนในการดำเนินการคร่าวๆ ตัวอย่างการนำเทคนิคการทำเหมืองข้อมูล (Data Mining) ไปใช้ในการวิเคราะห์ด้านอื่น

ขั้นตอนที่ 2 Data Understanding ขั้นตอนนี้เริ่มจากการเก็บรวบรวมข้อมูล หลังจากนั้นจะเป็นการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมาได้เพื่อดูความถูกต้องของข้อมูล และพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์

ขั้นตอนที่ 3 Data Preparation ขั้นตอนนี้เป็นขั้นตอนที่ทำการแปลงข้อมูลที่ได้ทำการเก็บรวบรวมมา (raw data) ให้กลายเป็นข้อมูลที่สามารถนำไปวิเคราะห์ในขั้นถัดไปได้ โดยการแปลงข้อมูลนี้อาจจะต้องมีการทำข้อมูลให้ถูกต้อง (data cleaning) เช่น การแปลงข้อมูลให้อยู่ในช่วง (scale) เดียวกัน หรือการเติมข้อมูลที่ขาดหายไป เป็นต้น โดยขั้นตอนนี้จะเป็นขั้นตอนที่ใช้เวลามากที่สุดของกระบวนการ CRISP-DM

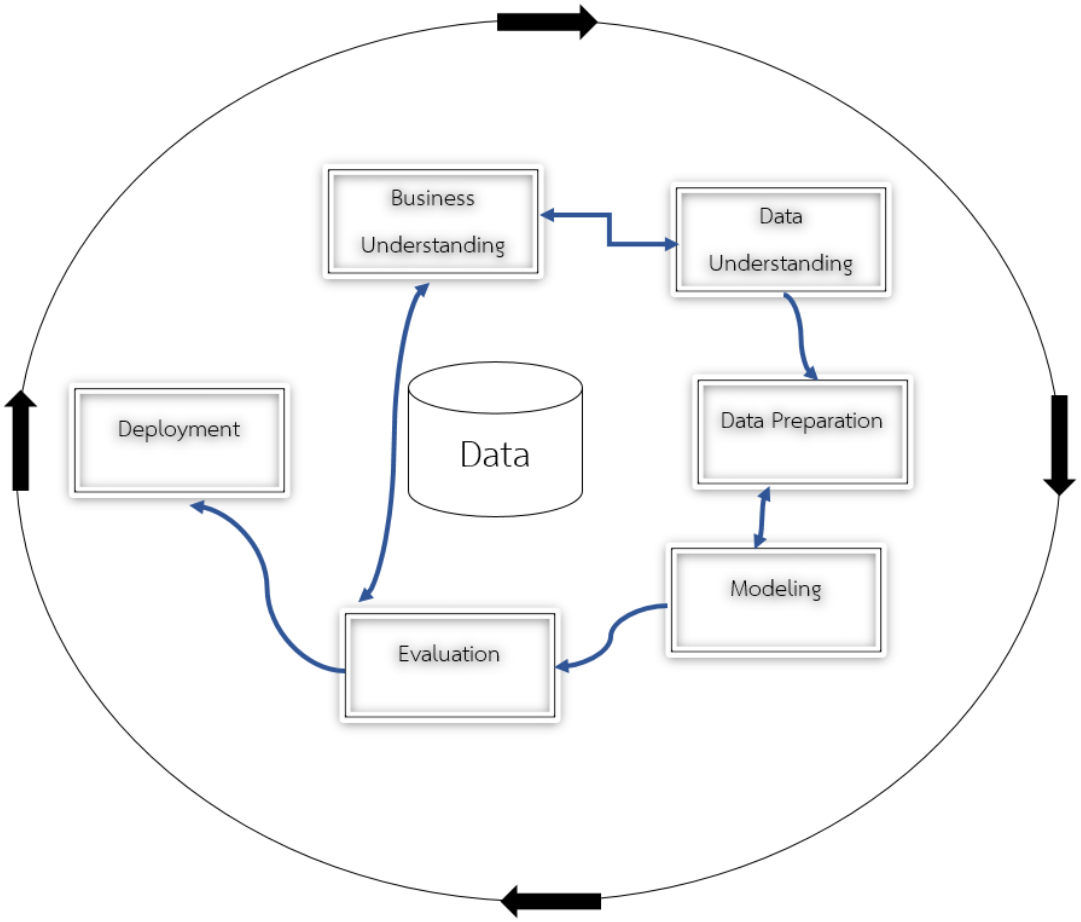
ขั้นตอนที่ 4 Modeling ขั้นตอนนี้จะเป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคทางการทำเหมืองข้อมูล (Data Mining) ที่ได้แนะนำไปแล้ว เช่น การจำแนกประเภทข้อมูล หรือ การแบ่งกลุ่มข้อมูล ซึ่งในขั้นตอนนี้หลายเทคนิคจะถูกนำมาใช้เพื่อให้ได้คำตอบที่ดีที่สุด ดังนั้นในบางครั้งอาจจะต้องมีการย้อนกลับไปที่ขั้นตอนที่ 3 Data Preparationเพื่อแปลงข้อมูลบางส่วนให้เหมาะสมกับแต่ละเทคนิคด้วย ตัวอย่างเทคนิคในการวิเคราะห์ข้อมูล เช่น

- การแบ่งกลุ่มข้อมูล (Clustering)
- การหากฎความสัมพันธ์ (Association Rules)
- การจำแนกประเภทข้อมูล (Classification)

ขั้นตอนที่ 5 Evaluation ในขั้นตอนนี้เราจะได้ผลการวิเคราะห์ข้อมูลด้วยเทคนิคทางการทำเหมืองข้อมูล (Data Mining) แล้วแต่ก่อนที่จะนำผลลัพธ์ที่ได้ไปใช้งานต่อไปก็จะต้องมีการวัดประสิทธิภาพของผลลัพธ์ที่ได้ว่าตรงกับวัตถุประสงค์ที่ได้ตั้งไว้ในขั้นตอนแรก หรือ มีความน่าเชื่อถือมากน้อยเพียงใด ซึ่งอาจจะย้อนกลับไปยังขั้นตอนก่อนหน้าเพื่อเปลี่ยนแปลงแก้ไขเพื่อให้ได้ผลลัพธ์ตามที่ต้องการได้ หากเป็นการสร้างโมเดลด้วยเทคนิค Association Rules การทดสอบประสิทธิภาพของโมเดลจะอยู่ในรูปแบบกฎพร้อมกับค่าความเชื่อมั่น หากเป็นการสร้างโมเดลด้วยเทคนิค Classification มีการทดสอบประสิทธิภาพของโมเดลอยู่ 3 แบบใหญ่ คือ

- Self-consitency test เป็นการแบ่งข้อมูลที่ใช้ในการสร้างโมเดล (model) และข้อมูลที่ใช้ในการทดสอบโมเดลเป็นข้อมูลชุดเดียวกัน
- Split test เป็นการแบ่งข้อมูลด้วยการสุ่มออกเป็น 2 ส่วน เช่น 70% ต่อ 30% หรือ 80% ต่อ 20% โดยข้อมูลส่วนที่หนึ่ง (70% หรือ 80%) ใช้ในการสร้างโมเดลและข้อมูลส่วนที่สอง (30% หรือ 20%) ใช้ใน การทดสอบประสิทธิภาพของโมเดล
- Cross-validation test เป็นการแบ่งข้อมูลออกเป็นหลายส่วน เช่น 5-fold cross-validation คือ ทำการแบ่งข้อมูลออกเป็น 5 ส่วน

ขั้นตอนที่ 6 Deployment ในกระบวนการทำงานของ CRISP-DM นั้นไม่ได้หยุดเพียงแค่ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคทางการทำเหมืองข้อมูล (Data Mining) เท่านั้น แม้ว่าผลลัพธ์ที่ได้จะแสดงถึงองค์ความรู้ที่มีประโยชน์ แต่จะต้องนำองค์ความรู้ที่ได้เหล่านี้ไปใช้ได้จริงในองค์กรหรือบริษัท ตัวอย่างเช่น การสร้างรายงานเพื่อให้ผู้บริหารหรือนักการตลาดเข้าใจได้ง่ายและสามารถนำไปออกโปรโมชั่นได้ เป็นต้น



ภาพที่ 5 กระบวนการทำงานของ CRISP-DM

2.9 วรรณกรรมที่เกี่ยวข้องกับ CRISP-DM

บทความ “ An introduction to data mining and other techniques for advanced analytics ” (Leventhal, 2553) CRISP-DM เป็นเหมือน blueprint ที่ใช้กันอย่างกว้างขวาง เช่นเดียวกันกับกระบวนการ ISO ในโรงงานอุตสาหกรรม หรือกระบวนการ CMMI ซึ่งเป็นมาตรฐานในการพัฒนาซอฟต์แวร์ กระบวนการมาตรฐานในการวิเคราะห์ข้อมูลด้านเหมืองข้อมูล พัฒนาขึ้นในปี ค.ศ. 1996 โดยความร่วมมือกันของ 3 บริษัท คือ DaimlerChrysler SPSS และNCR กระบวนการทำงานนี้เรียกว่า “ Cross-Industry Standard Process for Data Mining ” หรือเรียกย่อ ว่า “ CRISP-DM ” ในกระบวนการนี้ประกอบด้วย 6 ขั้นตอน ดังภาพที่ 3

บทความ “ Predicted Increase Enrollment in Higher Education Using Neural Networks and Data Mining Techniques ” (Nakhkob และKhademi, 2558) ปัจจุบันมีงานวิจัยที่นิยมใช้ CRISP-DM เพิ่มขึ้นโดยเฉพาะกับการศึกษา เช่น การประยุกต์ใช้เทคโนโลยีการทำเหมืองข้อมูลเพื่อทำนายความสำเร็จของนักเรียนและกรณีล้มเหลว นำเสนอการใช้ CRISP-DM กับชุดข้อมูล

เพื่อแยกรูปแบบที่ซ่อนอยู่จากข้อมูลของนักเรียน รูปแบบเหล่านี้สามารถเห็นได้ในความสัมพันธ์กับตัวแปรของนักเรียน หลังจากนั้นแบบจำลองถูกสร้างและทดสอบโดยใช้ชุดข้อมูลตัวอย่างของนักศึกษาระดับปริญญาตรี

บทความ “ Analyzing undergraduate students performance using educational data mining ” (AsifMerceronAli และHaider, 2560) กล่าวถึงการวิเคราะห์เชิงทำนายของผลการเรียนของนักเรียนและเลือก CRISP-DM สำหรับขั้นตอนในการออกแบบพื้นฐานของเหมืองข้อมูล เพื่อทำนายผลลัพธ์ข้อมูลของนักเรียนเมื่อสิ้นสุดรอบการเรียนหนึ่งปี มีการแสดงสถิติเชิงพรรณนาการวิเคราะห์เพื่อให้ได้ข้อมูลเชิงลึก จากนั้นนำไปเปรียบเทียบเพื่อทำนายผลการเรียนกับโมเดลอื่นๆ

บทความ “ A hybrid data envelopment analysis approach to analyse college graduation rate at higher education institutions ” (ChenChen และ Oztekin, 2559) วิเคราะห์อัตราการสำเร็จการศึกษาของวิทยาลัยในสถาบันอุดมศึกษา กลายเป็นจุดสนใจหลักในการวัดประสิทธิภาพของสถาบันการศึกษา บทความนี้วิธีการวิเคราะห์ DEA ภายใต้กระบวกการของ CRISP-DM เพื่อประเมินประสิทธิภาพของนักศึกษาระดับปริญญาตรีวิทยาลัย การวิเคราะห์เหล่านี้ให้ข้อมูลที่เป็นประโยชน์และการสนับสนุนนโยบายสำหรับผู้บริหารมหาวิทยาลัย

การสร้างแบบจำลองการขายผลิตภัณฑ์และพยากรณ์ยอดขายประกันชีวิต โดยเทคนิคการทำเหมืองข้อมูล กรณีศึกษาบริษัทประกันชีวิตแห่งหนึ่งงานการวิจัยนี้มี (พลอาสา, 2558) การพัฒนาแบบจำลองสำหรับการขายผลิตภัณฑ์ประกันชีวิต ใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) ตามกรอบ CRISP-DM โดยการสร้างแบบจำลองการแบ่งกลุ่ม (Clustering) ด้วยวิธี Simple K-Means เพื่อใช้ในการจัดกลุ่มลูกค้าที่ซื้อกรมธรรม์ประกันชีวิต จากนั้นจึงสร้างแบบจำลองการหาความสัมพันธ์ (Association Rule) ด้วยวิธี Apriori สร้างตัวแบบ (Model) เพื่อตอบสนองความต้องการของลูกค้า

2.10 เทคนิคการทำเหมืองกฎความสัมพันธ์ (Association Rules Mining)

กฎความสัมพันธ์ (Association rule) การค้นหากฎความสัมพันธ์มีขั้นตอนหลัก 2 ขั้นตอน คือ การนับความถี่ (Frequent Items Generation) จากตารางที่ 1 แสดงข้อมูลไอเทมเซตของนักศึกษาแต่ละคนเรียกว่า ไอเทมเซต ซึ่งนิยามศัพท์ที่เกี่ยวข้องกับกฎความสัมพันธ์ มีดังนี้

- ไอเทมเซต (Itemset) คือ ความสัมพันธ์ของข้อมูลที่หาได้ในฐานข้อมูลโดยไอเทมเซตประกอบด้วย ไอเทมที่ k-itemsets ตัวอย่างไอเทมที่แสดงความสัมพันธ์ของข้อมูล ดังตัวอย่างในตารางที่ 1 มีผลการเรียนเป็น k-itemsets คือ 6 ประกอบไปด้วย KM100high หมายถึงรหัสวิชา KM100 มีผลการเรียน B ขึ้นไป KM100low หมายถึงรหัสวิชา KM100 มีผลการเรียนต่ำกว่า B

เช่นเดียวกับรหัสวิชา SP241 และ SP242 เช่น นักศึกษารหัส 55XXXXXX01 มีผลการเรียน KM100 B ขึ้นไป วิชา SP241 มีผลการเรียนน้อยกว่า B วิชา SP242 มีผลการเรียนน้อยกว่า B

- Frequent Itemset คือ ไอเทมเซตที่ผ่านค่าสนับสนุนขั้นต่ำ (Minimum Support) การสร้างกฎความสัมพันธ์จากฟรีเคว้นไอเทมเซตที่หาได้ จะพิจารณาจากฟรีเคว้นไอเทมเซต ที่มีความยาวมากกว่า 2 รายการขึ้นไป โดยกฎความสัมพันธ์ทั่วไปเขียนได้ดังนี้ X => Y

- X (Left Hand Side) แสดงรูปแบบของไอเทมเซตด้านซ้ายของกฎความสัมพันธ์ และ Y (Right Hand Side) แสดงรูปแบบของไอเทมเซตด้านขวาของกฎความสัมพันธ์

- ค่าสนับสนุน (Support) คือ ค่าความน่าจะเป็นของจำนวนไอเทมเซตที่พบใน ฐานข้อมูลต่อจำนวนรายการทั้งหมด เช่น คำนวณหาค่าสนับสนุนของไอเทมเซต ตารางที่ 2 แสดงตัวอย่างการคำนวณค่า support ของแต่ละวิชาโดยคำนวณค่าสนับสนุนจากไอเทมรายวิชา ตัวอย่างเช่น KM100high คือ ในรายวิชา KM100 ที่มีเกรด B ขึ้นไป ปรากฏ Transaction ID ที่ 1,4 ดังนั้นค่าสนับสนุนไอเทมรายวิชานี้จึงเท่ากับ 33% ตารางที่ 3 แสดงตัวอย่างการคำนวณค่า support ของวิชา 2 วิชา และตารางที่ 4 แสดงตัวอย่างการคำนวณค่า support ของวิชา 3 วิชา ค่า support ของ โดยค่าสนับสนุนหาได้จาก

$$Support = \frac{frq(X,Y)}{N}$$

frq (X) คือ ค่าสนับสนุนของไอเทมเซต X

frq (Y) คือ จำนวนรายการข้อมูลของไอเทมเซต Y

N หรือ Number of all Transactions คือ จำนวนรายการข้อมูลทั้งหมด

ตารางที่ 1 แสดงข้อมูลไอเทมเซตของนักศึกษาแต่ละคนจำนวน 6 คน

Student id	Itemset
55XXXXXX01	KM100high, SP241low, SP242low
55XXXXXX02	KM100low, SP241high, SP242high
55XXXXXX03	KM100low, SP241low, SP242low
55XXXXXX04	KM100high , SP241low, SP242low
55XXXXXX05	KM100low , SP241low, SP242low
55XXXXXX06	KM100low, SP241low, SP242low

ตารางที่ 2 แสดงค่า Transaction ของ k-itemsets

Items	Transaction ID						Support
	1	2	3	4	5	6	
KM100high	1	0	0	1	0	0	2/6 = 33%
KM100low	0	1	1	0	1	1	4/6 = 67%
SP241high	0	1	0	0	0	1	1/6 = 17%
SP241low	1	0	1	1	1	0	5/6 = 83%
SP242high	0	1	0	0	0	0	1/6 = 17%
SP242low	1	0	1	1	1	1	5/6 = 83%

ตารางที่ 3 แสดงตัวอย่างการคำนวณค่า support ของวิชา 2 วิชา

Items	Transaction ID						Support
	1	2	3	4	5	6	
{KM100high, SP241high}	0	0	0	0	0	0	2/6 = 33%
{KM100high, SP241low }	1	0	0	1	0	0	2/6 = 33%
{KM100high, SP242high}	0	0	0	0	0	0	0/6 = 0%
{KM100high, SP242low}	1	0	0	1	0	0	1/6 = 17%
{SP241high, SP242 high}	0	1	0	0	0	0	0/6 = 0%
{ SP241high, SP242low}	0	0	0	0	0	0	0/6 = 0%
{ SP241low, SP242low}	1	0	1	1	1	1	5/6 = 83%

ตารางที่ 4 แสดงตัวอย่างการคำนวณค่า support ของวิชา 3 วิชา

Items	Transaction ID						Support
	1	2	3	4	5	6	
{KM100high, SP241high, SP242high }	0	0	0	0	0	0	0/6 = 0%
{KM100high, SP241high, SP242low }	0	0	0	0	0	0	2/6 = 33%
{KM100high,	1	0	0	1	0	0	2/6 = 33%

SP241low, SP242low }							
{KM100low, SP241low, SP242low }	0	0	1	0	1	1	3/6 = 50%
{KM100low, SP241high, SP242high }	0	1	0	0	0	0	1/6 = 17%

- ค่าความเชื่อมั่น (Confidence) คือ การแสดงค่าความเชื่อมั่นของกฎความสัมพันธ์ เมื่อรูปแบบLHS ที่อยู่ทางด้านซ้ายของกฎเกิดขึ้นแล้วมีโอกาสเกิดรูปแบบ RHS ที่อยู่ทางด้านขวามากน้อยเท่าใด ซึ่งจะมีค่าอยู่ระหว่าง 0-1 ถ้าใกล้เคียง 1 หมายถึงมีความเชื่อมั่นในการหาความสัมพันธ์มาก อาจคำนวณค่าที่จะ เกิดขึ้นในรูปแบบเปอร์เซ็นต์ก็ได้ วิธีการคำนวณค่าความเชื่อมั่นได้จาก

$$Confidence = \frac{frq(X,Y)}{frq(X)}$$

frq (X, Y) คือค่าสนับสนุนที่รูปแบบ X และ Y ของกฎความสัมพันธ์เกิดขึ้นพร้อมกัน

frq (X) คือค่าสนับสนุนรูปแบบที่อยู่ด้านซ้ายของกฎความสัมพันธ์

- ค่าสหสัมพันธ์ หรือเรียกว่าค่าลิฟต์ (Lift) คือค่าเป็นมาตรวัดทีบ่งบอกว่าการเกิดรูปแบบ Itemset X และ Itemset Y มีความสัมพันธ์กันแค่ไหน โดยถ้าค่าลิฟท์ของ X และ Y มีค่าใกล้ 1 แสดงว่าทั้งสอง Itemset มีความเป็นอิสระต่อกัน (Independent) แต่ถ้าค่าลิฟท์มีค่าเกิน 1 มาก แสดงว่ารูปแบบทั้งสองมีความสัมพันธ์กันโดยไม่เกิดขึ้นจากการสุ่ม ค่าลิฟต์คำนวณได้จาก

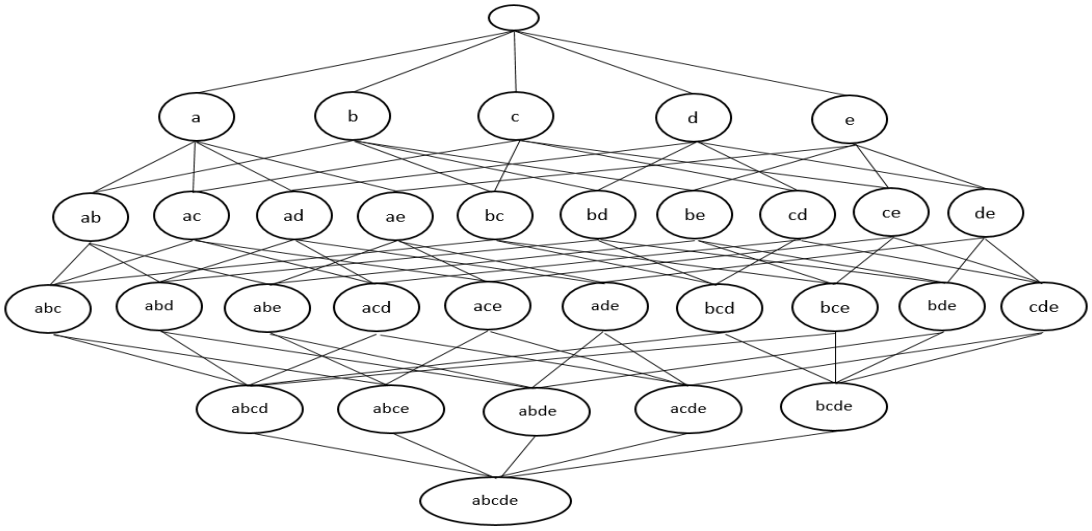
$$Lift = \frac{Support}{Supp(X) * Supp(Y)}$$

Support คือค่าสนับสนุนของกฎความสัมพันธ์

Supp (X, Y) คือค่าสนับสนุนที่รูปแบบ X และ Y ของกฎความสัมพันธ์เกิดขึ้นพร้อมกัน

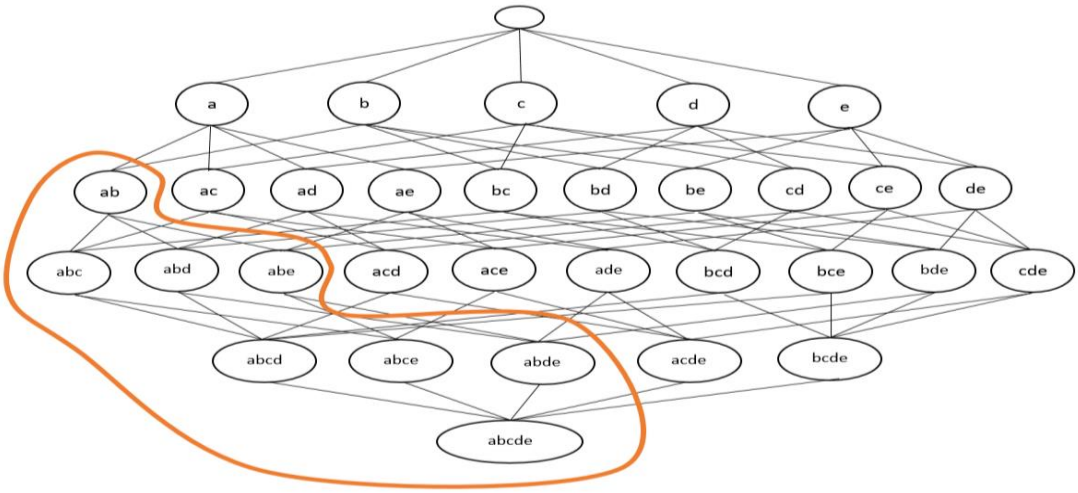
2.11 เทคนิคการใช้กฎความสัมพันธ์แบบอพริโอริ

อพริโอริ (Apriori) เป็นอัลกอริทึมพื้นฐานที่ใช้ในการหาความสัมพันธ์ของข้อมูลโดยใช้หลักการค้นหาแบบวงกว้างก่อนนับทรานแซคชัน ซึ่งจะทำการสร้าง และตรวจสอบเซตไอเท็มที่เกิดขึ้นบ่อยทีละชั้น โดยเริ่มจากเซ็ตไอเท็มที่มี จำนวนสมาชิกเท่ากับหนึ่งถ้าเซ็ตไอเท็มใดมีค่าสนับสนุนน้อยกว่าค่าสนับสนุนที่กำหนดก็จะตัดเซตไอเท็มนั้นออก ไม่นำไปสร้างเซ็ตไอเท็มในชั้นต่อไป การทำงานของอัลกอริทึมจะวนไปเรื่อย จนกระทั่งไล่ทุกระดับชั้น หรือไม่เหลือเซ็ตไอเท็มในชั้นต่อไป ในการนับจำนวนทรานแซคชันอัลกอลิทึมอพริโอริ จะทำการไล่ทรานแซคชันครั้งเดียวในแต่ละระดับชั้นในการตรวจดูว่าทรานแซคชันนั้นบรรจุเซตไอเท็มใดบ้าง เพื่อความรวดเร็วจะเก็บ เซตไอเท็มในแต่ละระดับชั้นทั้งหมดไว้ในโครงสร้างโครงสร้างแลตทิซแบบปกติ(ไพชยนต์ คงไชย และคณะ, 2557) จุดเด่นของอัลกอริทึมนี้อยู่ที่ความสามารถในความเร็วของการค้นหาไอเท็มเซตที่ปรากฏบ่อยด้วยการละเว้นการพิจารณาไอเท็มเซตที่ปรากฏซ้ำด้วยความถี่ ที่ต่ำกว่าเกณฑ์มีการกำหนดค่าสนับสนุนขั้นต่ำ (Minimum Support) มีการกำหนดค่าความเชื่อมั่นขั้นต่ำ (Minimum Confidence) ในการกำหนดขั้นต่ำทั้งสองค่านี้ จะขึ้นอยู่กับผู้ใช้ระบบเป็นผู้กำหนดเอง หรือจะใช้เชี่ยวชาญ (Expert user) เป็นผู้กำหนดให้ก็ได้ โดยกฎความสัมพันธ์ที่ได้นั้นจะต้องมีค่าสนับสนุน (Support) และค่าความ เชื่อมั่น (Confidence) ไม่น้อยกว่าค่าขั้นต่ำที่ได้กำหนดเอาไว้ข้างต้น



ภาพที่ 6 โครงสร้างแลตทิซแบบปกติ

จากภาพที่ 6 หาแต่ละโหนดคือ Item set โดยเริ่มจากไอเท็มเซต {a},{b},{c},{d},{e} ในแถวแรก แถวที่สองเป็นไอเท็มที่มีสองไอเท็มโดยเกิดจากไอเท็มเซ็ตแถวแรก เป็นเช่นนี้จนแถวของโหนดสุดท้าย จะประกอบไปด้วยไอเท็มเซต {a,b,c,d,e} ที่มีห้าไอเท็มมีที่มาจากห้าโหนดในแถวแรก



ภาพที่ 7 โครงสร้างแลตทิซที่มีการใช้เทคนิคการพรุน

จากภาพที่ 7 จะเห็นได้ว่ามีการทำเส้นตั้งแต่ไอเท็มเซต {a, b} ที่มีสองไอเท็ม จนมาถึงไอเท็มเซต {A, B, C, D, E} ที่มีห้าไอเท็มซึ่งมีความหมายว่า ไม่สนใจไอเท็มที่มี {a, b} เป็นซับเซต เนื่องจากค่า Support ของ a และ b ไม่ถึงเกณฑ์ Minimum Support เทคนิคนี้จึงได้ตัดเซตที่มี a และ b เป็นสมาชิกออกเพื่อช่วยลดเวลาในการสร้าง Candidate Itemsets และการทำงานของ อัลกอริทึม Apriori มีการทำงานเป็นแบบลูป

ข้อดีของอพริโอริ คือช่วยให้ทราบพฤติกรรมของเป้าหมายได้โดยการใช้อัลกอริทึมจัดการเชื่อมความสัมพันธ์ของเหตุการณ์ต่างๆ ที่เราต้องการหาความสัมพันธ์ของเป้าหมายคัดกรองข้อมูลออกมาตามความสัมพันธ์วิเคราะห์ข้อมูลมาจนมีความน่าเชื่อถือและนำไปใช้ได้จริง

อพริโอริ ถือเป็นอัลกอริทึมที่นิยมใช้ในการหากฎความสัมพันธ์ของข้อมูล ข้อเสียของอพริโอริถ้าฐานข้อมูลมีการเพิ่มข้อมูลเข้ามา หรือเกิดมีการเปลี่ยนแปลงข้อมูลอัลกอริทึมอพริโอริ จะต้องนำข้อมูลทั้งหมดมารวมกันก่อน แล้วจึงจะสามารถนำข้อมูล ทั้งหมดไปค้นหากฎความสัมพันธ์ใหม่ทั้งหมด โดยไม่สามารถนำกฎความสัมพันธ์ที่หาได้จากกลุ่มข้อมูลเก่าก่อนหน้ามาใช้ให้เกิดประโยชน์ได้ ทำให้เสียเวลาในการทำงานเพื่อค้นหากฎความสัมพันธ์ใหม่ทั้งหมด

2.12 วรรณกรรทที่เกี่ยวข้องกับเทคนิค Association Rules

การสร้างแบบจำลองกฎความสัมพันธ์สำหรับฐานข้อมูลการสั่งซื้อสินค้าโดยใช้เทคนิค เอฟพี-กโรธ (ชื่นมัจฉา, 2559) สร้างแบบจำลองความสัมพันธ์ของการซื้อสินค้าโดยใช้กฎความสัมพันธ์ (Association Rules) ด้วยเทคนิค เอฟพีกโรธ (FP - Growth) โดยวิเคราะห์จากการซื้อสินค้าของลูกค้าในแต่ละรายบุคคล ซึ่งแยกความต้องการของลูกค้าออกมาเป็นกฎความสัมพันธ์ได้ 4 รูปแบบ ประกอบด้วย 1. ค่าพารามิเตอร์ Support ไว้ที่ 0.1 และค่าพารามิเตอร์ Confidence ไว้ที่ 0.7 กฎความสัมพันธ์ ของสิ้นค้าทั้งหมด 15 กฎ 2. ค่าพารามิเตอร์ Support ไว้ที่ 0.1 และค่าพารามิเตอร์ Confidence ไว้ที่ 0.8 กฎความสัมพันธ์ ของสิ้นค้าทั้งหมด 11 กฎ 3. ค่าพารามิเตอร์ Support ไว้ที่ 0.2 และค่าพารามิเตอร์ Confidence ไว้ที่ 0.7 กฎความสัมพันธ์ ของสิ้นค้าทั้งหมด 6 กฎ 4. ค่าพารามิเตอร์ Support ไว้ที่ 0.2 และค่าพารามิเตอร์ Confidence ไว้ที่ 0.8 กฎความสัมพันธ์ของสิ้นค้าทั้งหมด 6 กฎ ซึ่งสามารถนาไปใช้ในการสนับสนุนการตัดสินใจ

การประยุกต์ใช้เอฟพี-กโรธกับงานแนะแนวการศึกษาต่อในระดับอุดมศึกษา (คงทิม และภูบุญอบ, 2552) ใช้กฎความสัมพันธ์ซึ่งเป็นหนึ่งในเทคนิคของเหมืองข้อมูลเพื่อช่วยแนะแนวการศึกษาให้กับ นักเรียนที่จะศึกษาต่อในระดับอุดมศึกษา ใช้ผลการเรียนเฉลี่ยของเจ็ดรายวิชาหลัก ได้แก่ วิชา คณิตศาสตร์ เคมี ชีววิทยา ฟิสิกส์ ภาษาไทย สังคมศึกษา และวิชา ภาษาอังกฤษ ในระดับชั้นมัธยมศึกษาตอนปลาย สาย วิทยาศาสตร์-คณิตศาสตร์ มาค้นหาความสัมพันธ์กับการศึกษาต่อของนักเรียน โดยใช้อัลกอริทึมเอฟพี-กโรธ (FPGrowth Algorithm) เพื่อนำรูปแบบความสัมพันธ์เหล่านั้นมา แนะแนวการศึกษาให้กับนักเรียนที่จะศึกษาต่อในระดับอุดมศึกษา ซึ่งมีค่าเฉลี่ยความถูกต้องรวมเป็น 89.87%

กฎความสัมพันธ์ของรายวิชาที่มีผลต่อการพ้นสภาพนักศึกษาโดยใช้อัลกอริทึมอพริโอริ(มหัทธนชัยมาลัยวงศ์สมหอม และตันตรานนท์, 2559) หากฎความสัมพันธ์ของรายวิชาที่มีผลต่อการพ้นสภาพนักศึกษา โดยใช้ อัลกอริทึมอพริโอริโดยการศึกษากลุ่มตัวอย่างจาก นักศึกษาของภาควิชาคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเชียงใหม่ ที่เข้าศึกษา ปีการศึกษา 2553-2558 คัดกรองนักศึกษาเฉพาะที่พ้นสภาพเนื่องจากมีผลการศึกษาต่ำกว่าเกณฑ์การประเมินผลการศึกษา นำข้อมูลนักศึกษากลุ่มนี้ไปตรวจสอบรายวิชาที่มีผล การเรียนต่ำกว่า C ที่ส่งผลทำให้พ้นสภาพการเป็นนักศึกษา จากนั้นแบ่งรายวิชาที่ส่งผลต่อ การพ้นสภาพนักศึกษาเป็น 4 กลุ่มตามชั้นปีของนักศึกษาที่พ้นสภาพ ผลจากวิจัยพบว่ากฎความสัมพันธ์ของรายวิชาที่มีผลต่อการพ้นสภาพนักศึกษาชั้นปีที่ 1, 2, 3 มีจำนวนเท่ากับ 27, 15, 14 กฎ ตามลำดับ ความสัมพันธ์ของรายวิชาในชั้นปีที่ 1 ส่งผลต่อรายวิชาในชั้นปีที่ 2 และชั้นปีที่ 3 เมื่อทดสอบความถูกต้องของกฎ ความสัมพันธ์ที่สร้างขึ้น พบว่าความถูกต้องของกฎความสัมพันธ์ของรายวิชาในชั้นปีที่ 2 และ 3 มีค่าสูงกว่าชั้นปีที่ 1 สาเหตุเนื่องมาจากรายวิชาในชั้นปีที่ 1 มีความหลากหลายซึ่งต้องเรียนในกลุ่มวิชาการศึกษาทั่วไป กลุ่มวิชาพื้นฐาน วิชาชีพ และกลุ่มวิชาชีพ แต่รายวิชาในชั้นปีที่ 2 เน้นเรียนในกลุ่มวิชาชีพ จากนั้นพัฒนาระบบแนะนำการลดความเสี่ยงการพ้นสภาพนักศึกษา ปรับปรุงหลักสูตร กระบวนการเรียนการสอน การวางแผนในการลงทะเบียนของนักศึกษา และงานอื่น ๆ ที่เกี่ยวข้องเพื่อคุณภาพการศึกษาให้ดีขึ้น

การใช้เทคนิคเหมืองข้อมูลเพื่อการวิเคราะห์ความสัมพันธ์รายวิชา (โพธิ์ผลิ, 2552) ใช้เทคนิคการจำแนกข้อมูลด้วยกฎความสัมพันธ์ และเทคนิคการจำแนกข้อมูลด้วยต้นไม้ตัดสินใจ มาเปรียบเทียบค่าความแม่นยำในการจำแนกข้อมูลของนักศึกษา รวมถึงศึกษาปัจจัยอื่นๆ ที่จะสามารถนำมาใช้เป็นตัวแบ่งข้อมูลในการค้นหาความสัมพันธ์ของรายวิชา เช่น เพศ สถานศึกษาตอนปลาย ผลการเรียนระดับดี หรือพอใช้ เพื่อให้ได้ความสัมพันธ์รายวิชาที่มีความถูกต้องและแม่นยำ ซึ่งการทดลองทำการแบ่งกลุ่มด้วยปัจจัยต่างๆพบว่าสามารถสร้างกฎความสัมพันธ์ โดยมรค่าความแม่นยำเฉลี่ยสูงกว่าร้อยละ 70 โดยปัจจัยที่ส่งผลมากที่สุด คือ เพศของนักศึกษา โดยอภิบายไว้ว่าเพศหญิงเรียนในวิชาหมวดเฉพาะได้ดีกว่าเพศชาย

การใช้กฎความสัมพันธ์ ร่วมกับฟัซซี่กฎความสัมพันธ์ เพื่อคาดการณ์ผลการเรียนของนักศึกษาใช้ (รักผกาวงศ์ และรักผกาวงศ์, 2558) เทคนิคกฎความสัมพันธ์ (association rule mining) โดยหาความสัมพันธ์ของผลการเรียนของแต่ละวิชา และใช้ เทคนิคฟัซซี่กฎความสัมพันธ์ (Fuzzy association rule mining) เพื่อคาดการณ์ผลการเรียนรวมโดยเฉลี่ยว่าอยู่ใน ระดับสูง ปานกลาง

หรือต่ำ เพื่อน่าผลไปคัดกรองนักศึกษาที่เรียนวิชาและได้เกรดตามกฎที่ค้นพบ ซึ่งคาดการณ์ว่า จะได้ผลการเรียนรวมโดยเฉลี่ยในระดับต่ำ ให้ปรับวิธีการเรียนตั้งแต่ต้น เพื่อที่จะมีผลการเรียนผ่านเกณฑ์ เมื่อเรียน จนครบหลักสูตร

การใช้เทคนิค Data Mining เพื่อค้นหาภาควิชาที่เหมาะสมที่สุดให้กับนิสิต (ส่งศิริรักธรรมานนท์ และไวยมัย, 2547) ปัญหานิสิตเลือกเรียนในสาขาวิชาที่ไม่ตรงกับความสามารถที่แท้จริง เนื่องด้วยนักศึกษาขาดประสบการณ์ และไม่รู้จักแต่ละภาควิชามากพอ งานวิจัยนี้นำเสนอการใช้เทคนิค classification ในการชี้แนะแนวทางการเลือกภาควิชาที่เหมาะสมที่สุดกับลักษณะและความสามารถของนิสิตแต่ละคน ใช้การหาลักษณะของนิสิตที่จัดอยู่ในกลุ่มที่เหมาะสมในแต่ละภาควิชา เพื่อให้นักศึกษาสามารถเลือกภาควิชาที่เหมาะสมกับตนเองได้มีความถูกต้องมากกว่า 80 เปอร์เซ็นต์

การเปรียบเทียบประสิทธิภาพของเทคนิค Apriori และ FP-Growth ในการสร้างกฎความสัมพันธ์ของโรคมะเร็งต่อมลูกหมาก (ทองคำ, 2561) เปรียบเทียบประสิทธิภาพของเทคนิค Apriori และ FP-Growth ในการสร้างกฎความ สัมพันธ์ของโรคมะเร็งต่อมลูกหมาก รวบรวมข้อมูลจากฐานข้อมูล SEER ระหว่างเดือน มกราคม พ.ศ. 2547 ถึง พ.ศ. 2557 จำนวน 2,308 ระเบียน ข้อมูลทั้งหมดได้ถูกนำ มาสร้างกฎความสัมพันธ์ด้วยเทคนิค Apriori และเทคนิค FP-Growth ผลการศึกษา พบว่า เทคนิค FP-Growth มีความสามารถในการสร้างกฎความสัมพันธ์ได้มากกว่าเล็กน้อย ทั้งนี้อาจเนื่องจากข้อมูล แต่หากตัวแปร และจำนวนข้อมูลน้อย จึงทำ ให้กฎที่ได้มีจำนวนจำกัด และเท่ากันจากทั้งสองเทคนิค

การหาปัจจัยที่ส่งผลต่อความเสี่ยงของนักศึกษาเรียนอ่อนด้วยเทคนิคกฎความสัมพันธ์ (สุวรรณโณ และสิงห์เอี่ยม, 2554) จากการวิจัยพบว่าสามารถใช้เทคนิคกฎ ความสัมพันธ์ซึ่งเป็นเทคนิคในการทำเหมืองข้อมูล มาประยุกต์ใช้ เพื่อค้นหารูปแบบของข้อมูลนักศึกษาที่มีความเสี่ยงที่จะประสบปัญหาการเรียนอ่อนได้พบว่าปัจจัยส่วนใหญ่ที่ ส่งผลให้นักศึกษามีความเสี่ยงที่จะเรียนอ่อนคือ วิธี การเข้าศึกษา คะแนนการสอบเข้าศึกษาของวิชา ต่างๆ และเพศของนักศึกษา ดังนั้นหากผู้ที่เกี่ยวข้องนำผลจากการวิจัยชิ้นนี้ไปใช้เพื่อป้องกัน ไม่ให้เกิดการเรียนอ่อนด้วยการหาความเสี่ยงของ นักศึกษาชั้นปีที่ 1 แล้วให้คำแนะนำและดูแล นักศึกษากลุ่มที่มีค่าความเสี่ยงที่จะเรียนอ่อนสูงก็ จะช่วยให้นักศึกษาวางแผนการศึกษาและลดความ เสี่ยงที่จะเรียนได้ผลการเรียนต่ำลงได้

การประยุกต์ใช้กฎความสัมพันธ์เพื่อวิเคราะห์ความเสี่ยงการออกกลางคันของนักศึกษาสาขาเทคโนโลยีสารสนเทศ (ปฤชานนท์ และศรีอุไร, 2561) ผลการสร้างกฎความสัมพันธ์พบว่ากฎความสัมพันธ์ที่ได้จากอัลกอริทึม Apriori มีค่าความเชื่อมั่นมากกว่า 90 % สรุปผลการวิจัยทำให้

ทราบว่าผล การเรียนในหมวดวิชาศึกษาทั่วไปหรือผลการเรียนในหมวดวิชาเฉพาะรายวิชาใดที่มีผลทำให้นักศึกษาออกกลางคัน และเป็นแนวทางให้ทางหลักสูตรได้จัดกิจกรรมส่งเสริมให้นักศึกษามีทักษะในรายวิชานั้นเพิ่มขึ้น เพื่อลดการลาออกของนักศึกษาในสาขาได้

การพัฒนากฎการทำนายผลการเรียนของนักศึกษาชั้นปีที่ 1 โดยใช้เทคนิคการทำเหมืองข้อมูล (ชัยวุฒิศักดิ์ และกล่อมวิเศษ, 2562) ใช้กฎ ความสัมพันธ์ (Association Rules) ด้วยอัลกอรทึมอพริโอริ (Apriori algorithm) และกฎการตัดสินใจสำหรับจำแนกข้อมูล (Data Classification) ด้วยเทคนิค J48 เพื่อจะได้นำมาวางแผนการเรียนของนักศึกษา จากการศึกษาพบว่ากฎ ที่ใช้ในการจำแนกผลการเรียนของนักศึกษาชั้นปีที่ 1 กลุ่มทีเกรดเฉลี่ยสะสมต่ำกว่า 2.00 และ กลุ่มที่ได้เกรดเฉลี่ยสูง กว่า 2.00 ด้วยเทคนิค J48 ให้ค่าความถูกต้องสูงถึง 91% และจำนวนกฎความสัมพันธ์ของรายวิชาที่มีผลต่อเกรดเฉลี่ยสะสมต่ำกว่า 2.00 ของนักศึกษาชั้นปีที่ 1 มีจำนวนเท่ากับ 5 ด้วยความเชื่อมั่นที่ 1.00 และ ค่าสหสัมพันธ์มากกว่า 1.00

2.13 สถิติกับการวิจัย

การวิเคราะห์ข้อมูลและสถิติเข้ามาเกี่ยวข้องกับกระบวนการวิจัยหลายขั้นตอน ตัวอย่างเช่น การคำนวณขนาด ตัวอย่าง การสุ่มตัวอย่าง การสร้างเครื่องมือและหาคุณภาพของเครื่องมือในการเก็บรวบรวมข้อมูล การวิเคราะห์ ข้อมูลเพื่อบรรยายลักษณะกลุ่มที่ศึกษา การวิเคราะห์ข้อมูลเพื่อตอบคำถามการวิจัยหรือสรุปผลการศึกษา ตามวัตถุประสงค์ การทดสอบสมมติฐานการวิจัย การนำเสนอและสรุปผลการวิจัย ใช้เป็นเครื่องมือในการ ตัดสินและสรุปผลการวิจัย เพื่อให้เข้าใจสถิติมากยิ่งขึ้นและนำความรู้สถิติไปใช้ในงานวิจัยได้อย่างเหมาะสม (คูณทวีแก้วตาใจเที่ยง และสุทธธง, 2563)

การแจกแจงความถี่ (Frequency) เป็นการนำข้อมูลที่เป็นค่าของตัวแปรที่สนใจมาจัดเรียงตามลำดับความมากน้อย และแบ่งเป็นช่วงของข้อมูล จำนวนข้อมูลในแต่ละช่วงเรียกว่า ความถี่ ในกรณีที่ความแตกต่างระหว่างจำนวนข้อมูลสูงสุดกับจำนวนข้อมูลต่ำสุดไม่มาก ไม่จำเป็นต้องแบ่งช่วงจำนวนข้อมูลเป็นกลุ่ม ในแต่ละช่วงมีจำนวน 1 ข้อมูลก็ได้ การแจกแจงความถี่มีจุดมุ่งหมายเพื่อให้ทราบภาพรวมของการแจกแจงข้อมูลทั้งหมดอย่างเป็นระบบ การจัดระบบและนำเสนอข้อมูลในเบื้องต้น สามารถนำเสนอข้อมูลในรูปของตารางและแผนภูมิ ในที่นี้จะขอแยกเป็น 2 ส่วนในการนำเสนอ คือ ตารางแจกแจงความถี่ และกราฟและแผนภูมิแบบต่างๆ

สัดส่วน (Proportion) คือ ความสัมพันธ์ของจำนวนย่อยกับจำนวนรวมทั้งหมด กล่าวคือ ให้ถือจำนวนรวมทั้งหมดเป็น 1 ส่วน เช่น ในการสำรวจคนในหมู่บ้านหนึ่งจำนวน 800 คน เป็นหญิง 300

คน ดังนั้นสัดส่วนของผู้หญิงในหมู่บ้านคือ $\frac{300}{800}$ = 0.37 และสัดส่วนของผู้ชายคือ $\frac{500}{800}$ = 0.63 เป็นต้น

การวัดแนวโน้มเข้าสู่ส่วนกลาง (Measures of Central Tendency) การวัดแนวโน้มเข้าสู่ส่วนกลางเป็นระเบียบวิธีทางสถิติในการหาค่าเพียงค่าเดียวที่จะใช้เป็นตัวแทนของข้อมูลทั้งชุด ค่าที่หาได้นี้จะทำให้สามารถทราบถึงลักษณะของข้อมูลทั้งหมดที่เก็บรวบรวมมาได้ ค่าที่หาได้นี้จะเป็นค่ากลาง ๆ เรียกว่า ค่ากลาง ประเภทของการวัดแนวโน้มเข้าสู่ส่วนกลาง การวัดแนวโน้มเข้าสู่ส่วนกลางมีอยู่หลายวิธีด้วยกัน ที่นิยมกันมาก ได้แก่

- ค่าเฉลี่ยเลขคณิต (Arithmetic Mean) หมายถึง การหารผลรวมของข้อมูลทั้งหมดด้วยจำนวนข้อมูลทั้งหมดสามารถคำนวณได้จากสูตร

$$\bar{x} = \frac{\sum x}{n}$$

$\bar{x}$ คือ ค่าเฉลี่ยเลขคณิต

$\sum x$ คือ ผลบวกของข้อมูลทุกค่า

n คือ จำนวนข้อมูลทั้งหมด

- มัธยฐาน (Median) หมายถึง ค่ากึ่งกลางของข้อมูลชุดนั้น หรือค่าที่อยู่ในตำแหน่งกึ่งกลางของข้อมูลชุดนั้น เมื่อได้จัดเรียงค่าของข้อมูลจากน้อยที่สุด ไปหามากที่สุดหรือจาหมากที่สุกไปหาน้อยที่สุด ค่ากึ่งกลางจะเป็นตัวแทนที่แสดงว่ามีข้อมูลที่มากกว่าและน้อยกว่านี้อยู่ 50 %
- ฐานนิยม (Mode) หมายถึง ค่าของคะแนนที่ซ้ำกันมากที่สุดหรือ ค่าคะแนนที่มีความถี่สูงที่สุดในข้อมูลชุดนั้น

2.14 Weka (Waikato Environment for Knowledge Analysis)

โปรแกรม Weka (Waikato Environment for Knowledge Analysis) เริ่มพัฒนามาตั้งแต่ปี 1997 โดยมหาวิทยาลัย Waikato ประเทศนิวซีแลนด์ เป็นซอฟต์แวร์สำเร็จรูป (เปรมปรีดิ์, 2560) อยู่ภายใต้การควบคุมของ GPL License โปรแกรม Weka ได้ถูกพัฒนามาจากภาษาจาวาทั้งหมด ซึ่งเขียนมาโดยเน้นกับงานทางด้านการเรียนรู้ด้วยเครื่อง (Machine Learning) และ การทำเหมืองข้อมูล (Data Mining) โปรแกรมจะประกอบไปด้วยฟังก์ชั่นสำหรับใช้ในการจัดการข้อมูล และเป็นโปรแกรมที่มี Graphic User Interface (GUI) ใช้ในการตั้งค่าในการส่งข้อมูลให้ซอฟต์แวร์ประมวลผลและสามารถพัฒนาต่อยอดโปรแกรมได้ เป็นเครื่องมือที่ใช้ทางานในด้านการทำเหมืองข้อมูลรวบรวมแนวคิดอัลกอริทึมมากมาย ซึ่งอัลกอริทึมสามารถเลือกใช้งานโดยตรงได้จาก 2 ทางคือจากชุดเครื่องมือที่มีอัลกอริทึมมาให้ หรือเลือกใช้จากอัลกอริทึมที่ได้เขียนเป็นโปรแกรมลงไปเป็นชุดเครื่องมือเพิ่มเติม และชุดเครื่องมือมีฟังก์ชั่นสำหรับการทำงานร่วมกับข้อมูล

ข้อดีโปรแกรม Weka

1.เป็นซอฟต์แวร์ที่เป็นฟรีแวร์

2.สามารถทางานได้ทุก OS

3. เชื่อมต่อ SQL Database โดยใช้ Java Database Connectivity

4. มีลักษณะที่ง่ายต่อการใช้งานเนื่องจากใช้

5.สนับสนุนเกี่ยวกับการทาเหมืองข้อมูล (Data Mining)

ข้อเสียโปรแกรม Weka

1.หาโหลดใช้งานได้ยาก

2.ฟังก์ชั่นอาจจะยังไม่ทันสมัย

การทำเหมืองข้อมูล คือ กระบวนการที่กระทำกับข้อมูล(โดยส่วนใหญ่จะมีจำนวนมาก) เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์

2.15 หลักสูตรสาขาวิชาวิทยาการคอมพิวเตอร์

หลักสูตร : วิทยาศาสตร์บัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์

ชื่อย่อ : วท.บ. (วิทยาการคอมพิวเตอร์)

คณะ : วิทยาศาสตร์

ระดับการศึกษา : ปริญญาตรี ปกติ

หลักสูตรวิทยาศาสตร์บัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ (หลักสูตรปรับปรุง พ.ศ.2555) มีชื่อภาษาอังกฤษคือ Bachelor of Science Program in Computer Science ใช้ระบบการศึกษาแบบทวิภาค โดยหนึ่งปีการศึกษาแบ่งออกเป็น 2 ภาคการศึกษาปกติ หนึ่งภาคการศึกษาปกติ มีระยะเวลาการศึกษาไม่น้อยกว่า 15 สัปดาห์ สำหรับการศึกษาภาคฤดูร้อน ให้มีระยะเวลาไม่น้อยกว่า 6 สัปดาห์ และมีชั่วโมงการศึกษาเท่ากับภาคการศึกษาปกติ การคิดหน่วยกิต ภาคทฤษฎีที่ใช้เวลาบรรยายหรืออภิปรายปัญหา 1 ชั่วโมงต่อสัปดาห์ ตลอดหนึ่งภาคการศึกษาปกติให้มีค่าเท่ากับ 1 หน่วยกิต ภาคปกติที่ใช้เวลาฝึกหรือทดลอง 2 ถึง 3 ชั่วโมงต่อสัปดาห์ ตลอดหนึ่งภาคการศึกษาปกติให้มีค่าเท่ากับ 1 หน่วยกิต การฝึกงานหรือการฝึกภาคสนาม ที่ใช้เวลาฝึก 90 ชั่งโมง ให้มีค่าเท่ากับ 1 หน่วยกิต การศึกษาตลอดหลักสูตร 4 ปี หรือ สำเร็จการศึกษาได้ไม่ก่อน 6 ภาคเรียนการศึกษาปกติ (ไม่เกิน 8 ปีการศึกษา) นับตั้งแต่วันลงทะเบียน ให้ลงทะเบียนเรียนได้ไม่น้อยกว่า 9 หน่วยกิต และ ไม่เกิน 22 หน่วยกิตในแต่ละภาคการศึกษา และไม่เกิน 9 หน่วยกิต ในภาคการศึกษาฤดูร้อน ต้องเรียนครบตามจำนวนหน่วยกิตที่กำหนดไว้ในหลักสูตรและจะต้องมีค่าระดับคะแนนเฉลี่ยสะสมทั้งหมดไม่ต่ำกว่า 2.00

ตารางที่ 5 กลุ่มวิชาสังคมศาสตร์ หน่วยกิตต่ำสุด : 6

รหัสวิชา	**รายวิชา**	**หน่วยกิต**
ศท021	สังคมศาสตร์ในชีวิตประจำวัน Social Sciences in Everyday Life	3 (3-0-6)
ศท022	อารยธรรมโลก World Civilization	3 (3-0-6)

ตารางที่ 6 กลุ่มวิชามนุษยศาสตร์ หน่วยกิตต่ำสุด : 9

รหัสวิชา	**รายวิชา**	**หน่วยกิต**
ศท011	มนุษย์กับความงามทางศิลปะ Man and Arts Appreciation	3 (3-0-6)
ศท012	จิตวิทยากับพฤติกรรมมนุษย์ Psychology and Human Behavior	3 (3-0-6)
ศท013	สุขภาพเพื่อการดำรงชีวิต Health for Life	3 (1-4-4)
ศท013	สุขภาพเพื่อการดำรงชีวิต Health for Life	3 (2-2-5)

ตารางที่ 7 กลุ่มวิชาภาษาหน่วยกิตต่ำสุด : 9

รหัสวิชา	รายวิชา	หน่วยกิต
ศท031	การใช้ภาษาไทย Thai Language Usage	3 (1-4-4)
ศท141	ภาษาอังกฤษพื้นฐาน 1 Fundamental English 1	3 (2-2-5)
ศท142	ภาษาอังกฤษพื้นฐาน 2 Fundamental English 2	3 (2-2-5)

ตารางที่ 8 กลุ่มวิชาวิทยาศาสตร์และคณิตศาสตร์ หน่วยกิตต่ำสุด : 6

รหัสวิชา	รายวิชา	หน่วยกิต
ผษ101	เกษตรเพื่อชีวิต Agriculture for Life	3 (3-0-6)
วท101	วิทยาศาสตร์เพื่อชีวิต Science for Life	3 (2-2-5)
วท102	การพัฒนาวิทยาศาสตร์และเทคโนโลยี Development of Science and Technology	3 (2-2-5)
อก101	วิศวกรรมเบื้องต้นในชีวิตประจำวัน Basic Engineering in Daily Life	3 (3-0-6)
อก102	นานาสาระเกี่ยวกับอาหารและยา General Aspects of Food and Drug	3 (3-0-6)

ตารางที่ 9 กลุ่มวิชาแกน หน่วยกิตต่ำสุด : 30

รหัสวิชา	รายวิชา	หน่วยกิต
คพ251	คณิตศาสตร์ดีสครีต Discrete Mathematics	3 (3-0-6)
คม100	เคมีทั่วไป General Chemistry	3 (2-3-5)
คม100	เคมีทั่วไป General Chemistry	3 (2-3-5)
คศ101	แคลคูลัส 1 Calculus 1	3 (3-0-6)
คศ101	แคลคูลัส 1 Calculus 1	3 (3-0-6)
คศ102	แคลคูลัส 2 Calculus 2	3 (3-0-6)
คศ205	ตรรกศาสตร์เชิงคณิตศาสตร์เบื้องต้น Introduction to Mathematical Logic	3 (3-0-6)
ชว100	ชีววิทยาทั่วไป General Biology	3 (2-3-5)
ชว100	ชีววิทยาทั่วไป General Biology	3 (2-3-5)
ฟส107	หลักฟิสิกส์ Principles of Physics	3 (2-3-5)
ฟส107	หลักฟิสิกส์ Principles of Physics	3 (2-3-5)
ศท241	ภาษาอังกฤษเชิงวิทยาศาสตร์และเทคโนโลยี 1 English for Science and Technology 1	3 (2-2-5)

ศท241	ภาษาอังกฤษเชิงวิทยาศาสตร์และเทคโนโลยี 1 English for Science and Technology 1	3 (2-2-5)
ศท242	ภาษาอังกฤษเชิงวิทยาศาสตร์และเทคโนโลยี 2 English for Science and Technology 2	3 (2-2-5)
สต301	หลักสถิติ Principles of Statistics	3 (3-0-6)

ตารางที่ 10 กลุ่มวิชาเอกบังคับ หน่วยกิตต่ำสุด : 58

รหัสวิชา	**รายวิชา**	**หน่วยกิต**
คพ211	วิทยาการคอมพิวเตอร์ 1 Computer Science 1	3 (2-3-5)
คพ212	วิทยาการคอมพิวเตอร์ 2 Computer Science 2	3 (2-3-5)
คพ216	เทคนิคการเขียนโปรแกรมคอมพิวเตอร์ Computer Programming Techniques	3 (2-3-5)
คพ217	หลักการเขียนโปรแกรมเชิงวัตถุเบื้องต้น Introduction to the Principles of Object-Oriented Programming	3 (2-3-5)
คพ218	การเขียนโปรแกรมขั้นสูง Advanced Programming	3 (2-3-5)
คพ313	หลักการเขียนโปรแกรมบนอินเตอร์เน็ต Principles of Internet Programming	3 (2-3-5)
คพ313	หลักการเขียนโปรแกรมบนอินเทอร์เน็ต Principles of Internet Programming	3 (2-3-5)
คพ321	โครงสร้างข้อมูล Data Structures	3 (2-3-5)

คพ321	โครงสร้างข้อมูล Data Structures	3 (2-3-5)
คพ331	ระบบคอมพิวเตอร์และภาษาแอสเซมบลี Computer System and Assembly Programming	3 (2-3-5)
คพ332	สถาปัตยกรรมคอมพิวเตอร์ Computer Architecture	3 (2-3-5)
คพ342	ระบบฐานข้อมูล Database Systems	3 (2-3-5)
คพ343	การวิเคราะห์และออกแบบระบบ System Analysis and Design	3 (2-3-5)
คพ343	การวิเคราะห์และออกแบบระบบ System Analysis and design	3 (2-3-5)
คพ351	วิธีการนิวเมอริคัล Numerical Methods	3 (2-3-5)
คพ391	การบริหารโครงการคอมพิวเตอร์ Computer Project Management	1 (1-0-2)
คพ421	ระบบปฏิบัติการ Operating System	3 (2-3-5)
คพ422	วิศวกรรมซอฟแวร์ Software Engineering	3 (2-3-5)
คพ424	คอมพิวเตอร์กราฟิก Computer Graphics	3 (2-3-5)
คพ432	การสื่อสารข้อมูลและเครือข่ายคอมพิวเตอร์ Data Communication and Computer Network	3 (2-3-5)
มจ497	สหกิจศึกษา Co-operative Education	9 (0-270-0)

มจ498	การเรียนรู้อิสระ Independent Study	9 (0-270-0)
มจ499	การศึกษา หรือการฝึกงาน หรือฝึกอบรมต่างประเทศ Studying or Training Abroad	9 (0-270-0)
วท497	สหกิจศึกษา Co-operative Educatuon	9 (0-270-0)
วท498	การเรียนรู้อิสระ Independent Learning	9 (0-270-0)
วท499	การศึกษา หรือการฝึกงาน หรือฝึกอบรมต่างประเทศ Overseas Study, Training or Internship	9 (0-270-0)
คพ422	วิศวกรรมซอฟแวร์ Software Engineering	3 (2-3-5)
คพ424	คอมพิวเตอร์กราฟิก Computer Graphics	3 (2-3-5)
คพ432	การสื่อสารข้อมูลและเครือข่ายคอมพิวเตอร์ Data Communication and Computer Network	3 (2-3-5)
มจ497	สหกิจศึกษา Co-operative Education	9 (0-270-0)
มจ498	การเรียนรู้อิสระ Independent Study	9 (0-270-0)
มจ499	การศึกษา หรือการฝึกงาน หรือฝึกอบรมต่างประเทศ Studying or Training Abroad	9 (0-270-0)
วท497	สหกิจศึกษา Co-operative Educatuon	9 (0-270-0)
วท498	การเรียนรู้อิสระ Independent Learning	9 (0-270-0)

วท499	การศึกษา หรือการฝึกงาน หรือฝึกอบรมต่างประเทศ Overseas Study, Training or Internship	9 (0-270-0)

ตารางที่ 11 กลุ่มวิชาเอกเลือก หน่วยกิตต่ำสุด : 12

รหัสวิชา	**รายวิชา**	**หน่วยกิต**
คพ214	การเขียนโปรแกรมคอมพิวเตอร์และอัลกอริทึม Computer Programming and Algorithm	3 (2-3-5)
คพ312	การวิเคราะห์และออกแบบอัลกอริทึม Algorithm Analysis and Design	3 (2-3-5)
คพ324	คอมพิวเตอร์กราฟิกในการสื่อสารทางวิทยาศาสตร์ Computer Graphics in Science Communication	3 (2-3-5)
คพ341	การประมวลผลแฟ้มข้อมูล File Processing	3 (2-3-5)
คพ361	ระบบสารสนเทศภูมิศาสตร์เบื้องต้น Introduction to Geographic Information	3 (2-2-5)
คพ423	การออกแบบคอมไพเลอร์ Compiler Design	3 (2-3-5)
คพ425	การเขียนโปรแกรมเชิงวัตถุ Object - Oriented Programming	3 (2-3-5)
คพ426	เทคนิคการเขียนโปรแกรมทางคณิตศาสตร์ Mathematics Programming Techniques	3 (2-3-5)
คพ427	ซอฟต์แวร์ระบบ System Software	3 (2-3-5)
คพ427	ซอฟต์แวร์ระบบ System Software	3 (2-3-5)

คพ433	เครือข่ายคอมพิวเตอร์ Computer Networks	3 (2-3-5)
คพ441	การจำลองโดยคอมพิวเตอร์และการสร้างรูปแบบ Computer Simulation and Modeling	3 (2-3-5)
คพ442	ระบบบริหารข้อมูลเพื่อการจัดการ Management Information System	3 (2-3-5)
คพ443	ระบบชำนาญการ Expert Systems	3 (2-3-5)
คพ444	การประมวลผลเชิงภาพ Image Processing	3 (2-3-5)
คพ445	การเข้ารหัสและความปลอดภัยในเครือข่าย Cryptography and Network Security	3 (2-3-1)
คพ446	การเข้ารหัสและความปลอดภัยในเครือข่าย Cryptography and Network Security	3 (2-3-5)
คพ451	ทฤษฎีการคำนวณของคอมพิวเตอร์ Theory of Computation	3 (2-3-5)
คพ452	ปัญญาประดิษฐ์ Artificial Intelligence	3 (2-3-5)
คพ494	หัวข้อพิเศษทางวิทยาการคอมพิวเตอร์ 1 Special Topics in Computer Science 1	3 (2-3-5)
คพ495	หัวข้อพิเศษทางวิทยาการคอมพิวเตอร์ 2 Special Topics in Computer Science 2	3 (2-3-5)
วท301	ทักษะความเป็นผู้ประกอบการทางวิทยาศาสตร์ Science-based Entrepreneurship	3 (2-2-1)

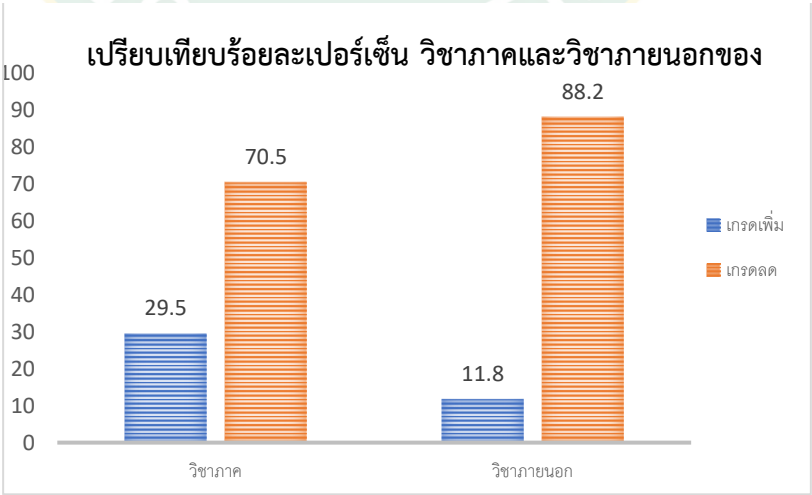
กลุ่มวิชาเลือกเสรีหน่วยกิตต่ำสุด : 6

บทที่ 3
วิธีการดำเนินงานวิจัย

ขั้นตอนวิธีการวิจัยจะดำเนินการตามในกระบวนการของ CRISP-DM (Cross-Industry Standard Process for Data Mining) ประกอบไปด้วย 6 ขั้นตอน แต่ละขั้นตอนจะเป็นขั้นตอนแบบต่อเนื่องกัน คือต้องทำขั้นตอนแรกก่อนไปขั้นตอนที่สองและสามบางขั้นตอนอาจมีการย้อนกลับมากทำในขั้นตอนก่อนหน้า จากภาพที่ 5 ตัวอย่างเช่นเมื่อได้ผลลัพธ์จากขั้นตอนการเตรียมข้อมูล (Data Preparation) แล้วจะนำไปสร้างโมเดลจำแนกประเภทข้อมูลในขั้น Modeling และหลังจากนั้นอาจย้อนกลับมาเปลี่ยนแปลงข้อมูลให้ถูกต้องมากขึ้นเพื่อที่โมเดลจะให้ค่าความถูกต้องมากขึ้น

3.1 การทำความเข้าใจปัญหา (Business Understanding)

เป็นขั้นตอนแรกในกระบวนการ CRISP-DM การเข้าใจปัญหา ในงานวิจัยนี้หมายถึงการแก้ปัญหาผลการเรียนลดลง การกระยุกต์ใช้เหมืองข้อมูลกับข้อมูลการศึกษา เพื่อช่วยแนะนำวิชาที่ควรเลือกลงทะเบียนเพราะจะทำให้ผลการเรียนออกมาในเกณฑ์ดี การแก้ปัญหาผลการเรียนลดลงโดยใช้เหมืองข้อมูลสามารถทำได้หลายวิธี ในบทความครั้งนี้เสนอเทคนิคการสร้างโมเดลทางสถิติ การสร้างโมเดลสำหรับการเลือกสาขาวิชานั้นสามารถทำได้หลายแนวทาง ในบทนี้ได้สร้างโมเดลโดยเทคนิคต่างๆทางเหมืองข้อมูล หาว่านักศึกษาสามารถเรียนวิชาที่แนะนำนั้นได้ผลการเรียนในเกณฑ์ดีหรือไม่ การวัดผลประสิทธิภาพผู้วิจัยได้แบ่งข้อมูลเป็นสองส่วนเพื่อวัดผลประสิทธิภาพของระบบ โดยข้อมูลปี 2556 เป็นข้อมูลที่นำมาวิเคราะห์ และข้อมูลปี 2555 เป็นข้อมูลที่นำมาเปรียบเทียบเพื่อวัดประสิทธิภาพของโมเดล



ภาพที่ 8 กราฟแสดงผลร้อยละของเกรดเฉลี่ยของนักศึกษาชั้นปีที่ 3

ก่อนอื่นต้องแน่ใจว่านักศึกษานั้นมีปัญหาเกี่ยวกับการเรียนจริงหรือไม่ ภาพที่ 8 แสดงถึงร้อยละของเกรดนักศึกษาชั้นปีที่ 3 เนื่องจากเป็นชั้นปีที่มีการเลือกลงทะเบียนมากที่สุด และแตกต่างกันมากที่สุดในทั้ง 4 ชั้นปี ผลปรากฏว่า นักศึกษาชั้นปีที่ 3 มีปัญหาเกี่ยวกับผลการเรียนเฉลี่ยจริงทั้งวิชาภาคที่ร้อยละ 70.5 และวิชาภายนอกที่ร้อยละ 88.2 วิชาภาคคือวิชาที่ส่วนใหญ่อยู่ในกลุ่มวิชาเอกบังคับ วิชาภายนอกคือส่วนใหญ่ไม่ได้อยู่ในกลุ่มวิชาเอกบังคับ ดังนั้นจึงมีการตัดสินใจเลือกในวิชาภายนอกมากกว่าวิชาภาค ผู้วิจัยจึงมุ่งเน้นที่จะแก้ปัญหาที่วิชาภายนอก

3.2 การเตรียมข้อมูล (Data Understanding)

เป็นขั้นตอนที่สองในกระบวนการ CRISP-DM ข้อมูลผลการลงทะเบียนของนักศึกษาจำนวน 12,000 แถว ในสาขาวิชาทางด้านคอมพิวเตอร์ศึกษาข้อมูลจากนักศึกษาที่สำเร็จการศึกษาแล้ว ประกอบได้ด้วย รหัสนักศึกษา ชื่อวิชา รหัสวิชา เกรด ดังตารางที่ 12 ขั้นตอนนี้เริ่มจากการเก็บรวบรวมข้อมูล หลังจากนั้นจะเป็นการตรวจสอบข้อมูลที่ได้ทำการรวบรวมมาได้เพื่อดูความถูกต้องของข้อมูล และพิจารณาว่าจะใช้ข้อมูลทั้งหมดหรือจำเป็นต้องเลือกข้อมูลบางส่วนมาใช้ในการวิเคราะห์ โดยผู้วิจัยได้แบ่งข้อมูลออกเป็น 2 ส่วน ส่วนแรกคือข้อมูลของนักศึกษาปีที่เข้า 2555 และส่วนที่สองคือข้อมูลของนักศึกษาปีที่เข้า 2556 โดยข้อมูลทั้งหมดที่ใช้ในการทดลองครั้งนี้ผู้วิจัยขออนุญาตกับสำนักงานทะเบียนของทางมหาวิทยาลัยเรียบร้อยแล้ว ดังแสดงในตัวอย่างในตารางที่ 12 รายละเอียดดังนี้

- Stu_code คือ รหัสนักศึกษา
- Year of study คือ ปีที่เข้าศึกษา
- branch คือ สาขาวิชา
- Subjects คือ รายวิชา
- Sub_Code คือ รหัสวิชา
- grade คือ ผลการเรียน

ตารางที่ 12 ข้อมูลดิบของนักศึกษาที่นำมาใช้ในการวิจัยครั้งนี้

Stu_code	Year of study	branch	Subjects	Sub_Code	grade
00001	2012	Computer science	Computer Science 1	CS 211	B
			Introduction to Object-Oriented Programming	CS 217	C+
00002	2013	Computer science	Principles of Thai language	TL 211	D+
...	...	...	...	...	...

3.3 การแก้ไข และดัดแปลงข้อมูล (Data Preparation)

เป็นขั้นตอนที่สามในกระบวนการ CRISP-DM เนื่องจากข้อมูลในตารางที่ 12 เป็นข้อมูลดิบที่เรียงจากรหัสนักศึกษา ต้องเปลี่ยนข้อมูลให้ให้อยู่ในลักษณะ Transection แบบเป็น Record ของนักศึกษาแต่ละคน เพื่อให้ได้โครงสร้างข้อมูลที่นำไปใช้ในการประมวลผลในการค้นหาความสัมพันธ์ของรายวิชาได้ โดยกำหนดให้คอลัมน์แรกคือรหัสนักศึกษา และคอลัมน์ถัดไปคือรายวิชาต่างๆที่นักศึกษาได้รับผลการเรียน ข้อมูลภายในคอลัมน์นี้คือผลการเรียนที่นักศึกษาได้รับ ดังแสดงในตัวอย่างในตารางที่ 13 ประกอบไปด้วย Stu_code คือ รหัสนักศึกษา ต่อด้วยรายวิชาต่างๆไปเรื่อยๆ จนครบ สำหรับนักศึกษาคนใดไม่ได้ลงทะเบียนในรายวิชานั้น ให้เว้นว่างข้อมูลในส่วนนี้ไว้

ตารางที่ 13 ตัวอย่างข้อมูลที่มีการดัดแปลงแล้ว

Stu_code	KNG304	KM100	KS(M)101	KS(M)102	KS(M)205
00001		A	C	B+	B+
00002		B+	D+	D	C+
00003	C	A	D+	D	
00004		C		C+	B
00005	B	B	C		

จากนั้นจัดกลุ่มผลการเรียนของนักศึกษา เพื่อลดการกระจายของข้อมูล โดยนำเกรดมาแบ่งตามเกณฑ์การวัดผลมาตรฐานการศึกษาของประเทศไทยตามตารางที่ 14 ได้แก่

- กลุ่ม high เกรด B B+ และ A
- กลุ่ม low เกรด D D+ C และ C+

ตารางที่ 14 เกณฑ์การวัดผลมาตรฐานการศึกษาของประเทศไทย

เกรด	ความหมาย
A	ดีเยี่ยม (Excellent)
B+	ดีมาก (Very Good)
B	ดี (Good)
C+	เกือบดี (Almost Good)
C	พอใช้ (Fair)
D+	อ่อน (Poor)
D	อ่อน (Very Poor)

จากข้อมูลในตารางที่ 13 มีการแก้ไขข้อมูลเพื่อให้ง่ายต่อการนำไปประมวลผล เช่น คอลัมน์ KNG304 แถว 00003 มีข้อมูลผลการเรียนเป็น C ในตารางที่ 15 จึงแทนที่ข้อมูลด้วย low เพราะผลการเรียนดังกล่าวอยู่ในกลุ่ม low เนื่องจากนักศึกษามีวิชาที่ลงทะเบียนไม่เหมือนกันในบางวิชา เพื่อ

ทำให้โปรแกรม Weka สามารถทำงานได้ต้องไม่มีตำแหน่งไหนของข้อมูลว่าง ผู้วิจัยจึงต้องเติมเครื่องหมาย “ ? ” ในตำแหน่งของข้อมูลที่ว่าง ซึ่งหมายถึงนักศึกษาไม่ได้ลงทะเบียนเรียนในวิชานั้น

ตารางที่ 15 ตัวอย่างข้อมูลที่มีการแก้ไขแล้ว

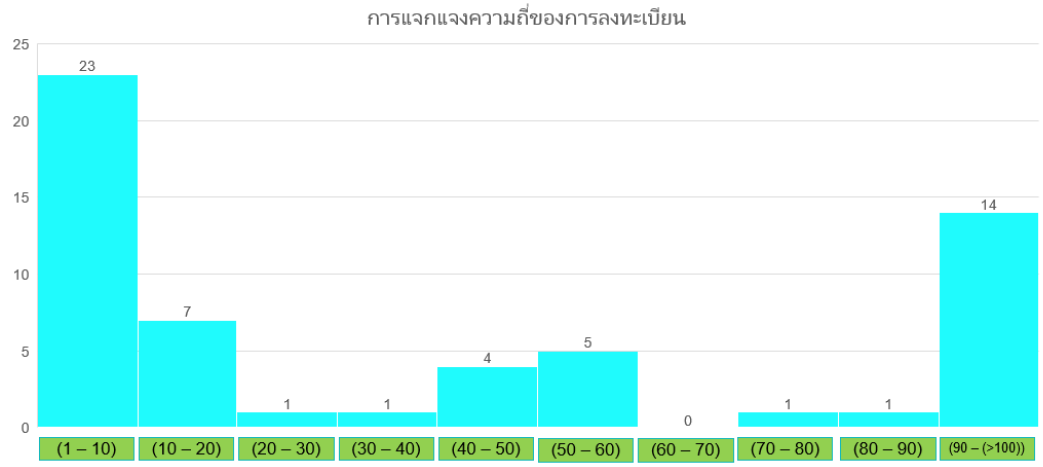
Stu_code	KNG304	KM100	KS(M)101	KS(M)102	KS(M)205
00001	?	high	low	high	high
00002	?	high	low	low	low
00003	low	A	low	low	?
00004	?	low	?	low	high
00005	high	high	low	?	?

3.4 การสร้างและการทดสอบตัวแบบ (Modeling)

เป็นขั้นตอนที่สี่ในกระบวนการ CRISP-DM ผู้วิจัยได้ใช้แผนภาพฮีสโตแกรม (Histogram) เป็นเครื่องมือคัดกรองข้อมูลรายวิชาที่น่าสนใจ และตัดบางรายวิชาที่มีนักศึกษาลงทะเบียนเรียนน้อยจนไม่สามารถนำมาวิเคราะห์ได้ เพราะจะมีผลให้ค่าคำนวณไม่ถูกต้องจาการสังเกต

3.4.1 การคัดกรองข้อมูลตามสถิติ

แผนภาพแสดงความถี่ของการเลือกเรียนวิชายภายนอกแกน X คือจำนวนนักศึกษาที่เลือกลงทะเบียน แกน Y คือ จำนวนวิชาที่ลงทะเบียน จากการสังเกตข้อมูลมีนักศึกษาลงทะเบียนเรียนต่ำว่า 10 คนจำนวน 23 วิชา เนื่องจากข้อมูลจำนวนนี้มีปริมาณน้อยเกินไปที่จำนำมาคำนวณเราจึงตัดข้อมูลแกนแรกวิชาที่คนลงทะเบียนไม่ถึง 10 คน จำนวน 23 วิชาทิ้งไป ดังที่แสดงในภาพที่ 9



ภาพที่ 9 ฮีสโตแกรมแสดงความถี่ของรายวิชาและจำนวนนักศึกษาที่ลงทะเบียน

กราฟแสดงเปอร์เซ็นต์ของวิชาหลังจากลบวิชาที่ไม่ได้นำมาวิเคราะห์ โดยจากข้อมูลที่ตัดออกประกอบด้วยข้อมูลของวิชาที่มีนักศึกษาลงทะเบียนต่ำกว่า 10 คน และบางวิชาที่มีการลงทะเบียนในข้อมูลปี 2555 แต่ไม่มีคนลงทะเบียนในปี 2556 ผลประกฏว่าจากวิชาทั้งหมด 56 วิชา ลดเหลือ 33 วิชา

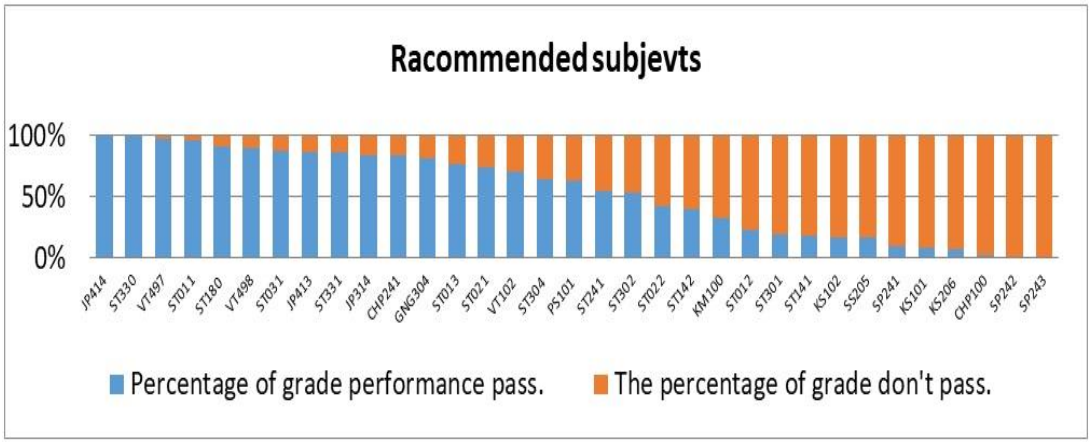
3.5 การสร้างต้นแบบจากสถิติเป็นโมเดลลำดับแรก สร้างจากการใช้สถิติ

เป็นการสร้างต้นแบบว่าหากสร้างโมเดลที่ใช้เฉพาะหลักสถิติจะให้ผลการสร้างเป็นอย่างไร จากข้อมูลการลงทะเบียนของนักศึกษาปี 2555 จำนวน 92 คน หาสัดส่วน (Proportion) ในแต่ละวิชาเพื่อหาความสัมพันธ์ว่าวิชาใดในทางสถิติแล้วมีความน่าสนใจให้นักศึกษาเลือกลงทะเบียนหรือไม่ โดยนับเฉพาะรายวิชาที่มีผลการเรียนเท่านั้น เช่น จากตารางที่ 14 วิชา KM100 มีผลการเรียนทั้งหมด 5 ผลการเรียน แบ่งออกเป็น high จำนวน 4 ผลการเรียน และ low จำนวน 1 ผลการเรียน ดังนั้น สัดส่วนของวิชา KM100 จึงมีความหมายว่า หากเลือกลงทะเบียนวิชานี้จะส่งผลให้มีผลการเรียนอยู่ในกลุ่ม high มีโอกาส 80 %

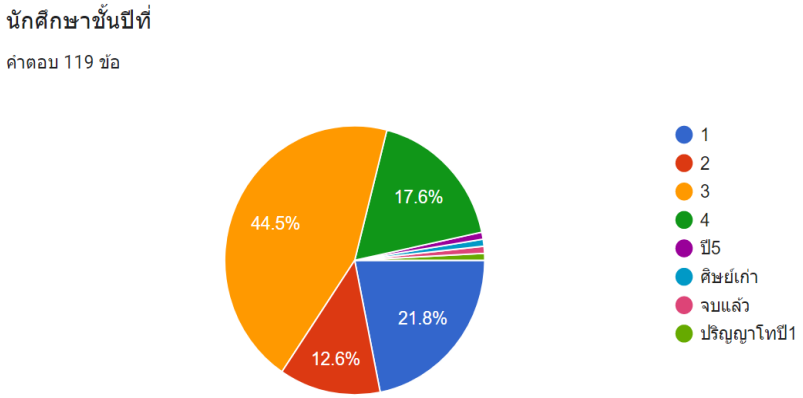
ตารางที่ 16 ตัวอย่างการหาสัดส่วน (Proportion) จากข้อมูลการลงทะเบียนของนักศึกษาปี 2555

Stu_code	KNG304	KM100	KS(M)101	KS(M)102	KS(M)205
00001	?	high	low	high	high
00002	?	high	low	low	low
00003	low	high	low	low	?
00004	?	low	?	low	high
00005	high	high	low	?	?
สัดส่วนร้อยละโอกาศเรียนได้เกณฑ์ดี	1/2 = 50	4/5 = 80	0/4 = 0	1/4 = 25	2/3 = 33

หลังจากขั้นตอนการหาสัดส่วน (Proportion) ของแต่ละวิชา ปรากฏว่าจากจำนวน 33 วิชา มี 19 วิชาที่มีสัดส่วนมากกว่าร้อยละ 50 แต่อย่างไรก็ตามต้องหาหลักการในการสร้างความเชื่อมั่นค่าสัดส่วนเหล่านี้มาเป็นเกณฑ์

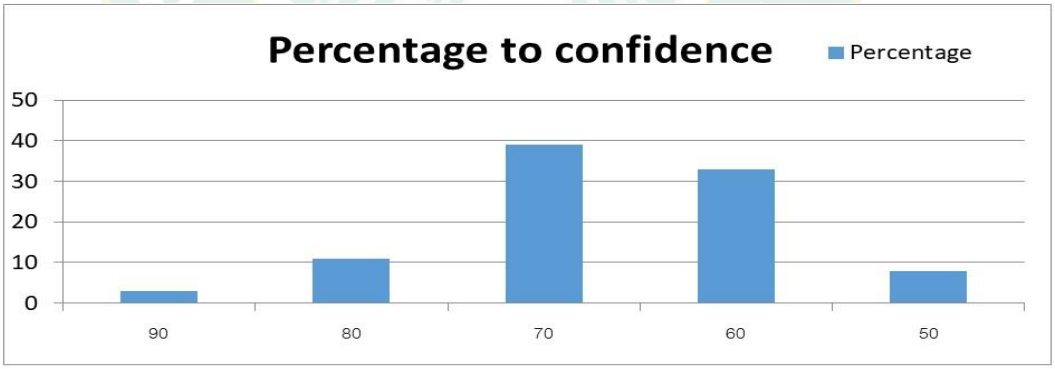


ภาพที่ 10 แสดงเปอร์เซ็นต์รายวิชาตามการคำนวณสถิติ



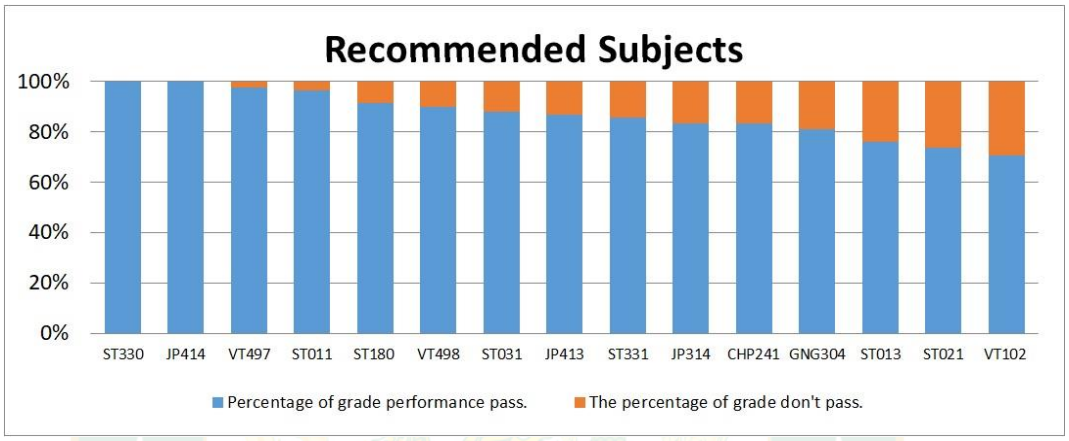
ภาพที่ 11 กราฟแสดงผู้ให้ข้อมูลแบบสอบถามของนักศึกษาตามชั้นปี

จากภาพที่ 12 เป็นการรายงานผลแบบสอบถามความต้องการระบบสนับสนุนการศึกษาของนักศึกษาสาขาวิทยาการคอมพิวเตอร์ จากคำตอบจำนวน 119 คำตอบพบว่า นักศึกษาเชื่อมั่นที่ระบบแนะนำที่ 70 % ขึ้นไป จากภาพที่ 10 สามารถบอกได้ว่าผู้ให้ข้อมูลส่วนใหญ่เป็นนักศึกษาสาขาวิชาวิทยาการคอมพิวเตอร์ชั้นปีที่ 3 ซึ่งเป็นชั้นปีที่ผู้วิจัยให้ความสำคัญเป็นอย่างยิ่ง

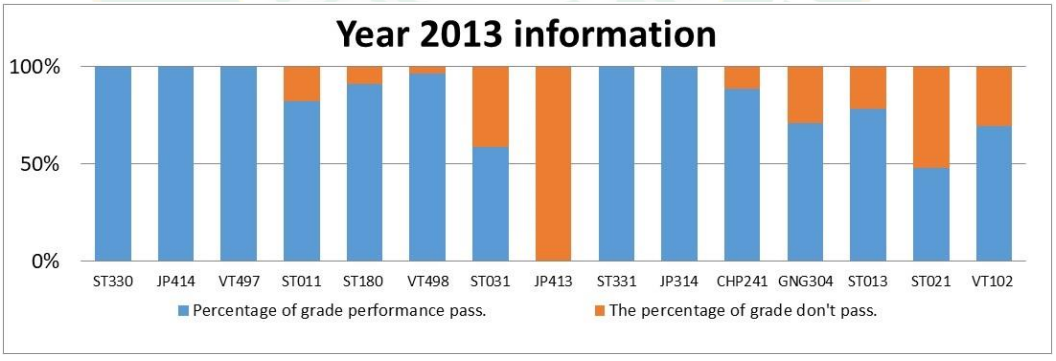


ภาพที่ 12 แบบสอบถามความพึงพอใจของผู้ทดลองใช้งาน

การทดสอบต้นแบบสถิติที่สร้างจากข้อมูลปี 2555 มาทำการแนะนำวิชาให้กับข้อมูลปี 2556 ภาพที่ 14 เลือกข้อมูลเฉพาะวิชาที่มีการลงทะเบียนเหมือนกันเพื่อเปรียบเทียบ พบว่านักศึกษาเลือกวิชาตามที่ระบบแนะนำจำนวน 15 วิชา มี 13 วิชาที่มีนักศึกษาส่วนใหญ่เรียนในเกณฑ์ดี กล่าวคือ ระบบสามารถแนะนำวิชาได้ถูกต้อง 86.67 % ในภาพที่ 13 แต่อย่างไรก็ตามการสร้างโมเดลทางสถิติอาจยังไม่มีความน่าเชื่อถือมากนักเนื่องจาก เป็นวิธีการพื้นฐานที่ยังไม่ได้เกิดจากการทำเหมืองข้อมูล ดังนั้นจะต้องมีวิธีการอื่นๆเปรียบเทียบเพื่อเพิ่มความน่าเชื่อถือของผู้ใช้งาน



ภาพที่ 13 โมเดลสถิติจากแบบสอบถามความพึงพอใจของผู้ทดลองใช้งาน



ภาพที่ 14 ข้อมูลที่นำมาทดสอบกับโมเดลที่หนึ่ง

3.6 การสร้างโมเดลที่สอง โดยใช้การทำเหมืองกฎความสัมพันธ์

ในขั้นตอนการค้นหากฎความสัมพันธ์ (Association Rules) สามารถอภิบายการค้นหาวิธีการได้ดังนี้

3.6.1 กฎความสัมพันธ์ (Association Rules)

เป็นเทคนิคหนึ่งของ Data Mining คือการค้นหาความสัมพันธ์ของข้อมูล จากข้อมูลขนาดใหญ่ที่มีอยู่เพื่อนำไปหารูปแบบที่เกิดขึ้นบ่อยๆ (frequent pattern) และใช้ในการวิเคราะห์ความสัมพันธ์หรือทำนาย ปรากฏการณ์ต่างๆ ฐานข้อมูลที่ใช้ในการทำเหมืองความสัมพันธ์ (Association Mining) มักเป็นฐานข้อมูลประเภท Transaction Database ผลลัพธ์ที่ได้เป็นกฎความสัมพันธ์ (Association Rule) สามารถเขียนได้ ในรูปเซตของรายการที่เป็นเหตุ ไปสู่เซตของรายการที่เป็นผล ซึ่งมี รากฐานมาจากการวิเคราะห์ตะกร้าตลาด (Market Basket Analysis) เช่น ลูกค้าที่ซื้อผ้าอ้อมส่วนใหญ่จะซื้อเบียร์ด้วย ข้อมูลที่นำมาใช้จะอยู่ในรูปแบบ Nominal หรือ Ordinal เท่านั้น

3.6.2 การวัดประสิทธิภาพของกฎความสัมพันธ์ (Support & Confidence)

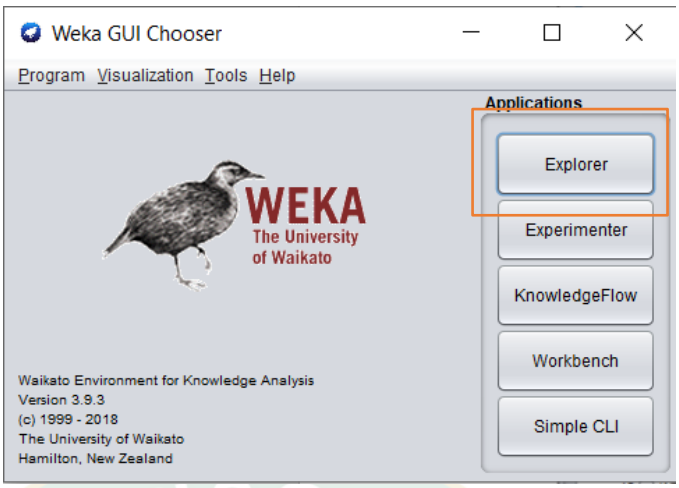
การพิจารณากฎเลือกกฎจะพิจารณาจากค่าสนับสนุน (Support) และค่าเชื่อมั่น (Confident)

3.6.3 การสร้างกฎความสัมพันธ์ด้วย Apriori Algorithm

มีการกำหนดค่าสนับสนุนขั้นต่ำ (Minimum Support) การกำหนดค่าความเชื่อมั่นขั้นต่ำ (Minimum Confidence) ในการกำหนดขั้นต่ำทั้งสองค่านี้ จะขึ้นอยู่กับผู้ใช้ระบบเป็นผู้กำหนดเอง หรือจะใช้เชี่ยวชาญ (Expert user) เป็นผู้กำหนดให้ก็ได้ โดยกฎความสัมพันธ์ที่ได้นั้นจะต้องมีค่าสนับสนุน (Support) และค่าความเชื่อมั่น (Confidence) ไม่น้อยกว่าค่าขั้นต่ำที่ได้กำหนดเอาไว้ข้างต้น

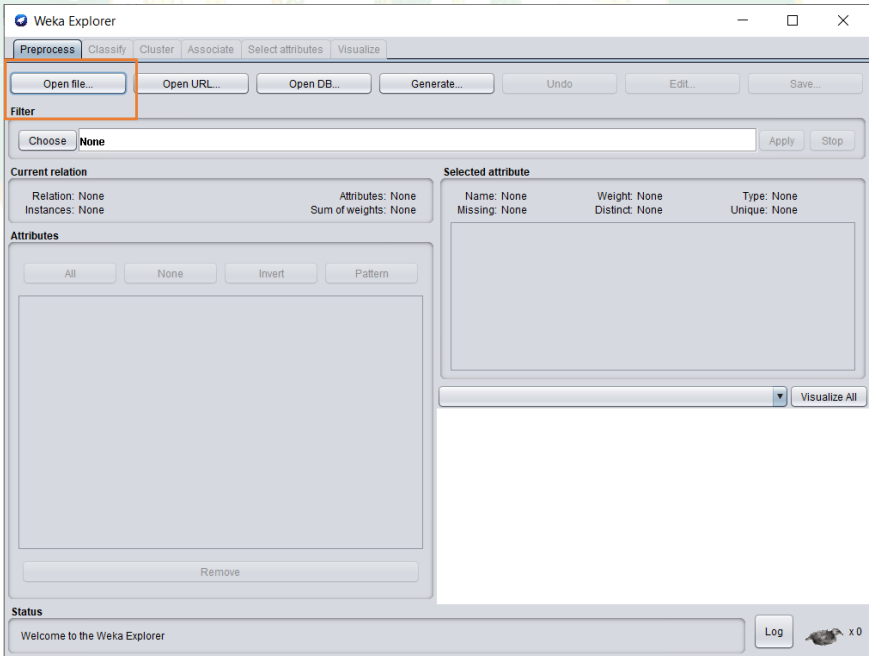
3.6.4 การหากฎความสัมพันธ์ Apriori โดยโปรแกรม Weka

ขั้นตอนที่ 1 เริ่มการใช้งานโปรแกรม และแสดงหน้าโปรแกรมตามภาพที่ 15 และเลือกที่เมนู Application คลิกที่ Explorer

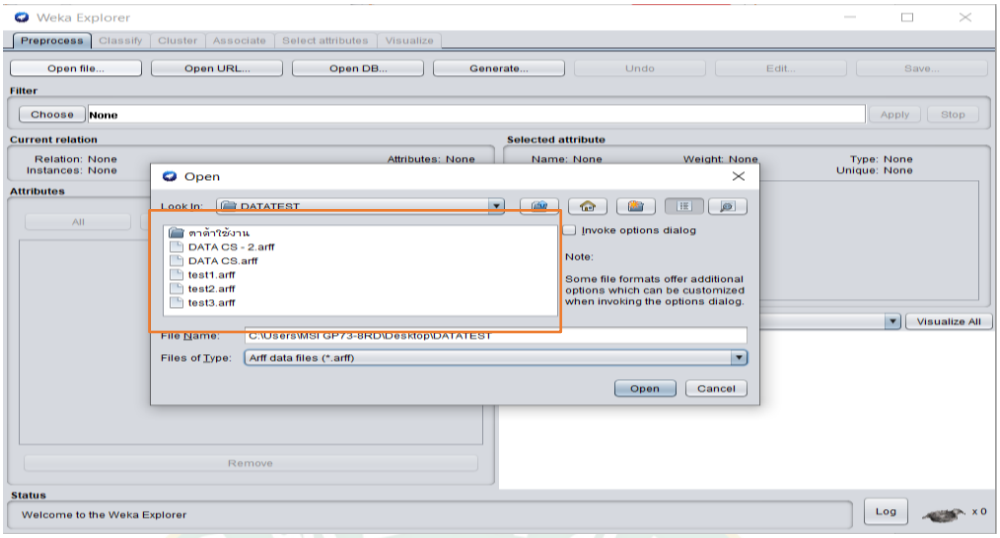


ภาพที่ 15 หน้าจอเริ่มใช้งานของโปรแกรม Weka

ขั้นตอนที่ 2 นำเข้าข้อมูลที่เตรียมไว้จากตารางที่ 15 โดยการเลือก Open file ในภาพที่ 16 เลือกข้อมูลที่จะนำเข้าในภาพที่ 17 ตัวอย่างข้อมูลภายในไฟล์ excel อยู่ในภาพที่ 18



ภาพที่ 16 หน้าจอเลือกใช้งาน

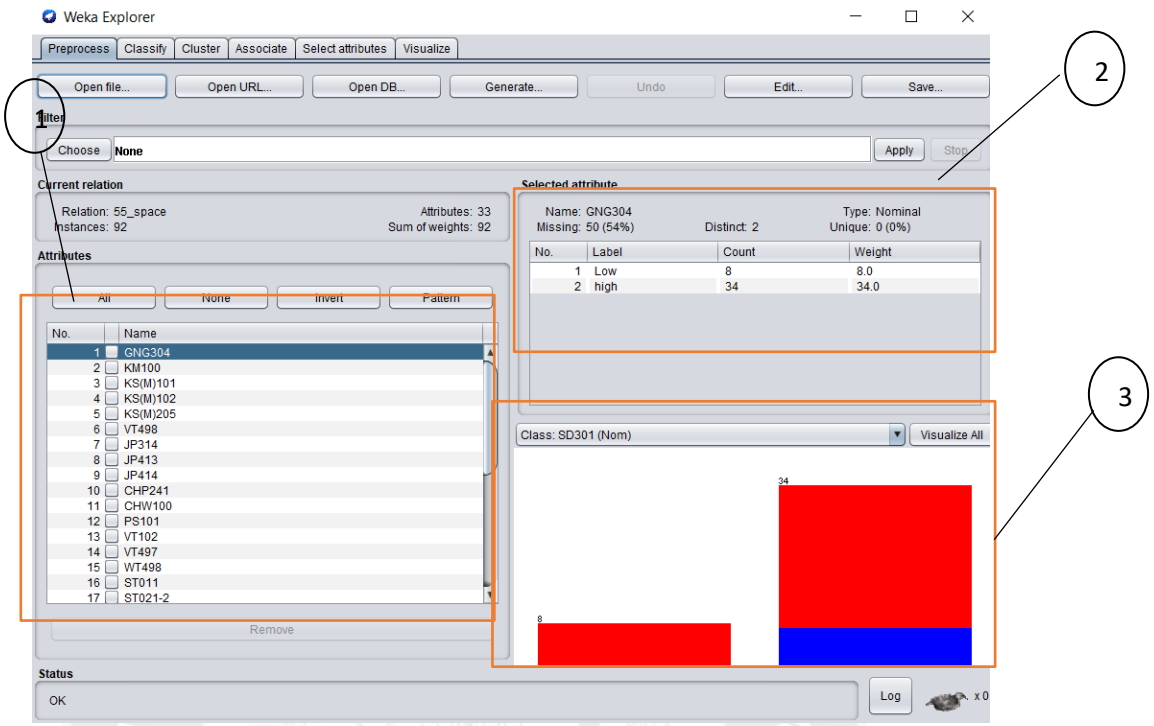


ภาพที่ 17 เมนูเลือกไฟล์ข้อมูลที่นำเข้าไปประมวลผล

GNG304	KM100	KS(M)101	KS(M)102	KS(M)205	VT498	JP314	JP413	JP414	CHP241	CHW100	PS101	VT102	VT497	WT498	ST011	ST012	ST013	ST021	ST022	ST031
?	high	low	high	high	low	?	?	?	?	?	high	high	?	high	high	?	high	high	?	high
?	high	low	low	low	low	?	?	?	?	high	high	high	?	high	high	?	high	high	?	high
low	low	low	low	low	low	?	?	?	?	?	low	low	high	?	?	low	?	high	low	high
?	?	low	low	low	low	?	?	high	?	low	low	low	?	high	?	low	low	low	low	high
?	low	low	low	high	low	high	?	?	high	?	low	low	?	high	?	?	?	low	low	high
?	?	low	low	low	low	?	?	?	?	low	high	high	?	high	high	?	high	high	?	high
?	?	high	high	high	low	?	?	?	?	low	low	high	?	high	high	?	high	high	?	high
?	low	low	high	high	high	?	?	?	?	?	high	high	?	high	?	?	high	high	?	high
high	low	high	low	low	low	?	?	?	?	?	high	high	high	?	?	?	high	high	low	high
?	?	low	low	low	low	?	?	high	?	low	low	low	?	low	?	?	low	high	high	high
?	high	low	high	low	low	?	?	?	?	?	high	high	?	high	?	?	high	high	?	high
high	?	low	high	low	high	?	?	high	?	?	low	high	high	?	high	?	high	high	low	high
?	low	low	low	low	low	?	high	?	?	?	high	high	?	low	?	low	?	high	low	high
?	low	low	low	low	low	?	?	?	?	?	low	low	?	high	high	?	high	high	low	high
high	low	low	low	low	low	?	?	?	?	?	high	high	high	?	?	?	high	high	?	high
?	low	low	low	low	low	?	low	?	?	?	low	high	?	high	?	?	low	high	?	high
?	low	low	low	low	low	?	high	?	?	?	low	low	?	high	high	?	low	high	low	high
?	low	low	low	low	low	?	high	?	?	?	high	high	?	high	?	low	?	high	high	high
high	low	low	low	low	high	?	?	?	?	?	high	high	high	?	high	?	high	high	high	high
high	?	low	low	low	low	?	?	?	?	low	low	low	high	?	?	?	high	high	low	high
low	?	low	low	low	low	?	high	?	?	low	high	high	high	?	high	?	high	high	high	high
high	?	low	low	low	low	?	?	?	?	low	high	high	low	?	?	low	?	high	low	high
low	low	high	low	low	low	?	?	high	?	?	low	low	high	?	?	low	low	low	high	high
?	?	low	low	low	low	low	?	?	low	?	low	low	?	high	?	?	high	low	?	high
high	?	low	low	low	low	?	?	?	?	low	high	low	high	?	high	?	low	high	low	high
low	?	low	low	low	low	?	high	?	?	low	high	high	high	?	high	?	high	high	low	high
high	low	low	low	low	low	high	?	?	high	?	high	high	high	?	?	?	?	high	?	high
high	?	low	low	low	low	?	?	high	?	low	high	high	high	?	?	low	?	low	low	high
?	?	low	low	low	low	?	?	high	?	low	low	?	?	high	?	low	high	low	low	high

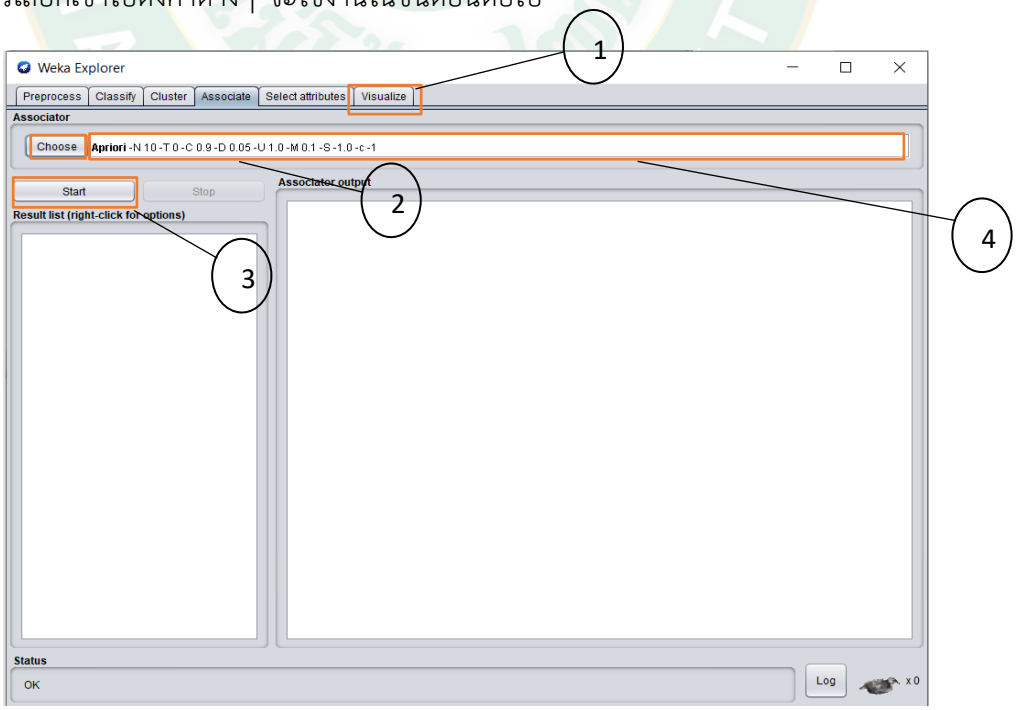
ภาพที่ 18 ไฟล์ข้อมูลที่นำเข้าไปประมวลผล

ภาพที่ 19 เป็นภาพหน้าจอการทำงานหลังเลือกข้อมูลที่จะนำมาประมวลผลแล้ว จุดที่ 1 คือ รายละเอียดจากไฟล์ที่อัพโหลดในที่นี้คือข้อมูลผลการเรียนของนักศึกษา ประกอบไปด้วย No. คือ ลำดับ Name คือชื่อตารางในที่นี้คือรายวิชา และจุดที่ 2 คือ ปรากฏรายละเอียดแต่ละตาราง หรือ รายวิชา ประกอบไปด้วย No. คือลำดับข้อมูล Label คือ ข้อมูลภายในตาราง หรือรายวิชา Count และ Weight คือจำนวน และน้ำหนัก จุดที่ 3 คือ แผนภาพโดยรวมของจุดที่ 2 โดยข้อมูลจะสัมพันธ์กัน



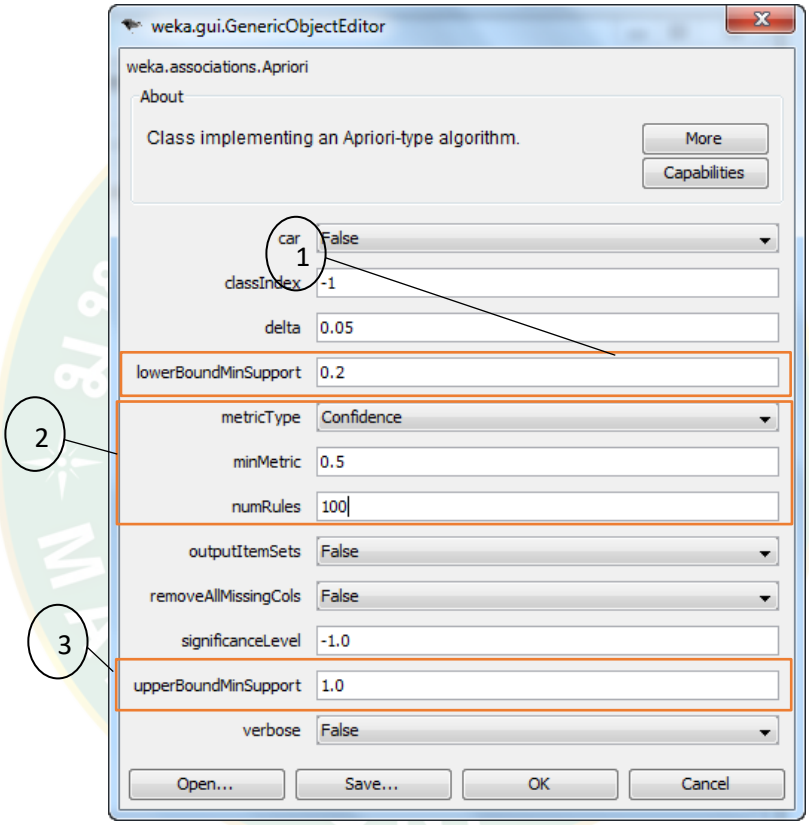
ภาพที่ 19 โปรแกรมโดยข้อมูลจะปรากฏขึ้นหลังจากเลือกไฟล์

ขั้นตอนที่ 3 เลือกที่เมนู Associate ตรงจุดที่ 1 จะปรากฏหน้าจอในภาพที่ 20 โดยที่ จุดที่ 2 คือ การเลือกวิธีการค้นหากฎความสัมพันธ์ในการวิจัยนี้คือ Apriori จุดที่ 3 คือ ปุ่มเริ่มการค้นหา จุดที่ 4 คือการเลือกเข้าไปตั้งค่าต่างๆ จะใช้งานในขั้นตอนต่อไป



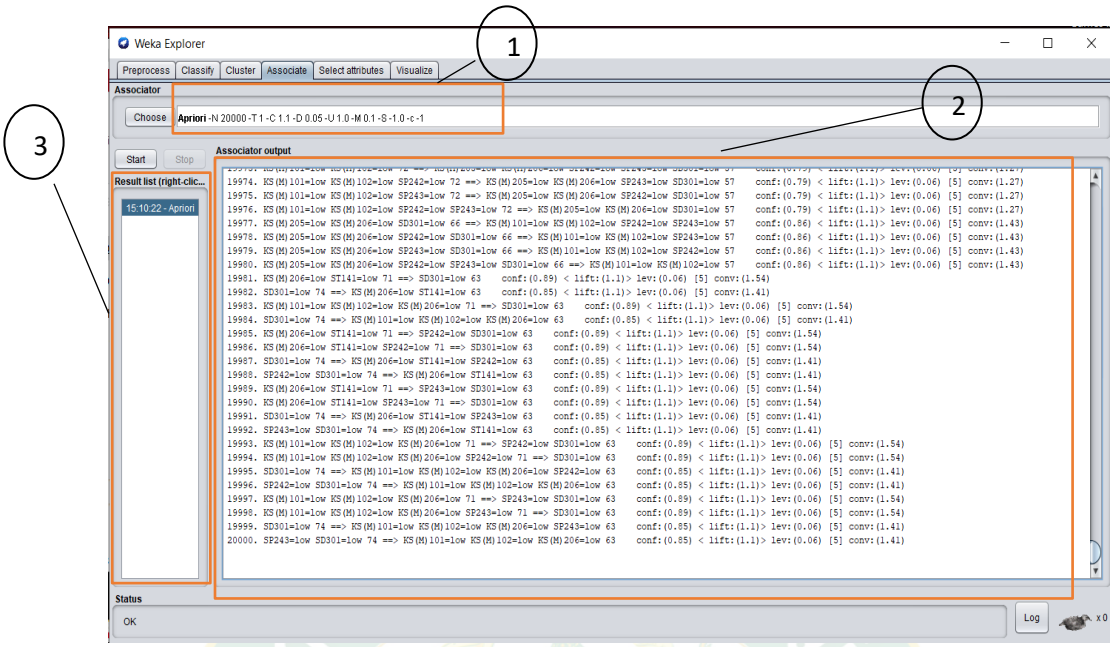
ภาพที่ 20 หน้าจอการใช้งาน Associate

ขั้นตอนที่ 4 เป็นการกำหนดค่าต่างๆให้กับอัลกอลิทึม Apriori โดยมีการตั้งค่าที่สำคัญ เช่น จุดที่ 1 Lower Bound MinSupport คือ การตั้งค่ากำหนดค่าสนับสนุนขั้นต่ำ (Minimum Support) ให้กับอัลกอลิทึม จุดที่ 2 metricType คือ การกำหนดให้กฎที่ค้นหา แสดงรายการโดยคำนึงถึงค่าใด minMetric ค่าขั้นต่ำของ metricType ที่กำหนด ในภาพ20 คือการกำหนดค่าความเชื่อมั่นขั้นต่ำ (Minimum Confidence) เพราะ metricType คือค่า Confidence และ numRules คือ จำนวนกฎที่ต้องการให้ปรากฏในภาพที่ 21



ภาพที่ 21 หน้าจอการตั้งค่า Apriori algorithm

ขั้นตอนที่ 5 หลักจากการตั้งค่าแล้วเริ่มใช้งานการค้นหากฎความสัมพันธ์ จะปรากฏข้อมูลในภาพที่ 22 โดยจุดที่ 1 คือ ข้อมูลสรุปจากการตั้งค่าในภาพที่ 21 จุดที่ 2 คือ กฎความสัมพันธ์ที่ได้จากการใช้โปรแกรม Weka โดยจะปรากฏข้อมูล จุดที่ 3 คือ ลำดับการประมวลผล โดยจะเรียงลำดับเวลาจากข้างล่างขึ้นข้างบน



ภาพที่ 22 ตัวอย่างกฎความสัมพันธ์ที่ได้จากการใช้โปรแกรม

ผู้วิจัยได้ทดลองปรับเปลี่ยนค่า Support และค่า Confidence โดยผลการทดลองเป็นไปตามตารางที่ 18 โดยกำหนดค่า Support เริ่มต้น ขั้นต่ำที่ 0.5 และค่า Confidence เริ่มต้นที่ 0.95 แล้วเพิ่มค่า Lower Bound ไปเรื่อยๆตามลำดับ เพื่อเปรียบเทียบกฎที่ค้นพบและนำกฎนั้นไปบันทึกลงในฐานข้อมูล

ตารางที่ 17 จำนวนกฎที่ได้จากการค้นหากฎความสัมพันธ์ ด้วยค่า Support และ Confidence ต่างกัน ในการหมวดวิชาภายนอก

Lower Bound MinSupport	Upper Bound MinSupport	Lift Min	จำนวนกฎ	จำนวนวิชา
0.5	1	0.95	30000	15
0.7	1	0.95	2548	11
0.8	1	0.95	278	10
0.9	1	0.95	32	5
0.95	1	0.95	2	2

เนื่องจากต้องการทราบจำนวนจำนวนวิชาที่จะปรากฏในกฎความสัมพันธ์จากตารางที่ 18 ปรากฏว่าการตั้งค่าในแถวแรกมีวิชาที่แนะนำมากที่สุดจำนวน 15 วิชา การวิจัยในครั้งนี้จึงเลือกใช้การตั้งค่านี้ในการวิจัยในขั้นต่อไป สรุปการตั้งค่าได้ดังนี้

- Lower Bound MinSupport = 0.5
- Upper Bound MinSupport = 1
- Lift Min = 0.95

ตารางที่ 18 ตัวอย่างกฎความสัมพันธ์ 10 อันดับเรียงจากค่า Confidence จากมากไปน้อย

ลำดับ	กฎ	conf	lift
1	SP243=Low ==> SP242=Low	1	1
2	SP242=Low ==> SP243=Low	1	1
3	VT498=Low ==> SP242=Low	1	1
4	VT498=Low ==> SP243=Low	1	1
5	VT498=Low SP243=Low ==> SP242=Low	1	1
6	VT498=Low SP242=Low ==> SP243=Low	1	1
7	VT498=Low ==> SP242=Low SP243=Low	1	1
8	KS(M)101=low ==> SP242=low	1	1
9	KS(M)101=low ==> SP243=low	1	1
10	SP241=low ==> SP242=low	1	1

หลังขั้นตอนการสร้างกฎความสัมพันธ์ และการตั้งค่าการค้นหากฎความสัมพันธ์ จากตัวอย่างตารางที่ 19 สามารถอธิบายตัวอย่างกฎความสัมพันธ์จำนวน 10 กฎ โดยเรียงจากค่า Confidence จากมากไปน้อยได้ดังนี้

กฎลำดับที่ 1 SP243=Low ==> SP242=Low ถ้านักศึกษาเรียนรายวิชา SP243 ได้เกรดน้อยกว่า B แล้วจะเรียนวิชา SP243 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 2 SP242=Low ==> SP243=Low ถ้านักศึกษาเรียนรายวิชา SP242 ได้เกรดน้อยกว่า B แล้วจะเรียนวิชา SP242 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 3 VT498=Low ==> SP242=Low ถ้านักศึกษาเรียนรายวิชา VT498 ได้เกรดน้อยกว่า B แล้วจะเรียนวิชา SP242 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 4 VT498=Low ==> SP243=Low ถ้านักศึกษาเรียนรายวิชา VT498 ได้เกรดน้อยกว่า B แล้วจะเรียนวิชา SP243 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 5 VT498=Low SP243=Low ==> SP242=Low ถ้านักศึกษาเรียนรายวิชา VT498ได้เกรดน้อยกว่า B และเรียนวิชา SP243 ได้เกรดน้อยกว่า B แล้วเรียนวิชา SP242 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 5 VT498=Low SP242=Low ==> SP243=Low ถ้านักศึกษาเรียนรายวิชา VT498ได้เกรดน้อยกว่า B และเรียนวิชา SP242 ได้เกรดน้อยกว่า B แล้วจะเรียนวิชา SP243 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 7 VT498=Low ==> SP242=Low SP243=Low ถ้านักศึกษาเรียนรายวิชา VT498ได้เกรดน้อยกว่า B แล้วจะเรียนวิชา SP243 และวิชา SP243 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 8 KS(M)101=low ==> SP242=low ถ้านักศึกษาเรียนรายวิชา KS(M)101 ได้เกรดน้อยกว่า B แล้วเรียนวิชา SP242 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 9 KS(M)101=low ==> SP243=low ถ้านักศึกษาเรียนรายวิชา SP242ได้เกรดน้อยกว่า B แล้วจะเรียนวิชา SP243ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 10 SP241=low ==> SP242=low ถ้านักศึกษาเรียนรายวิชา SP241 ได้เกรดน้อยกว่า B แล้วจะเรียนวิชา SP242 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

ตารางที่ 19 ตัวอย่างกฎความสัมพันธ์ 10 อันดับเรียงจากค่า lift จากมากไปน้อย

ลำดับ	กฎความสัมพันธ์	conf	lift
1	KS(M)102=low ==> KS(M)206=low	0.99	1.06
2	KS(M)206=low ==> KS(M)102=low	0.97	1.06
3	KS(M)102=low ==> KS(M)206=low SP242=low	0.99	1.06
4	KS(M)206=low ==> KS(M)102=low SP242=low	0.97	1.06
5	KS(M)102=low SP242=low ==> KS(M)206=low	0.99	1.06
6	KS(M)206=low SP242=low ==> KS(M)102=low	0.87	1.06
7	KS(M)102=low ==> KS(M)206=low SP243=low	0.99	1.06
8	KS(M)206=low ==> KS(M)102=low SP243=low	0.98	1.06
9	KS(M)102=low SP243=low ==> KS(M)206=low	0.99	1.06
10	KS(M)206=low SP243=low ==> KS(M)102=low	0.7	1.06

หลังขั้นตอนการสร้างกฎความสัมพันธ์ และการตั้งค่าการค้นหากฎความสัมพันธ์ จากตัวอย่างตารางที่ 20 สามารถอธิบายตัวอย่างกฎความสัมพันธ์จำนวน 10 กฎ โดยเรียงจากค่า **lift** จากมากไปน้อยได้ดังนี้

กฎลำดับที่ 1 KS(M)102=low ==> KS(M)206=low ถ้านักศึกษาเรียนรายวิชา SP243 ได้เกรดน้อยกว่า B แล้วเรียนวิชา SP242 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 2 KS(M)206=low ==> KS(M)102=low ถ้านักศึกษาเรียนรายวิชา SP242 ได้เกรดน้อยกว่า B แล้วเรียนวิชา SP243ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1กฎลำดับที่ 3 SP243=Low ==> SP242=Low ถ้านักศึกษาเรียนรายวิชา SP243 ได้เกรดน้อยกว่า B แล้วเรียนวิชา SP242 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 3 KS(M)102=low ==> KS(M)206=low SP242=low ถ้านักศึกษาเรียนรายวิชา KS(M)102 ได้เกรดน้อยกว่า B และ KS(M)206 ได้เกรดน้อยกว่า B แล้วจะเรียนวิชา SP242 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 4 KS(M)206=low ==> KS(M)102=low SP242=low ถ้านักศึกษาเรียนรายวิชา KS(M)106 ได้เกรดน้อยกว่า B แล้วจะเรียนวิชา SP242 ได้เกรดน้อยกว่า B และ KS(M)206 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 5 KS(M)102=low SP242=low ==> KS(M)206=low ถ้านักศึกษาเรียนรายวิชา KS(M)102 ได้เกรดน้อยกว่า B และเรียน SP242 น้อยกว่า B แล้วจะเรียนวิชา KS(M)206 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 6 KS(M)206=low SP242=low ==> KS(M)102=low ถ้านักศึกษาเรียนรายวิชา KS(M)206 ได้เกรดน้อยกว่า B และเรียน SP242 ได้เกรดน้อยกว่า B แล้วจะเรียนวิชา SP242 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 7 KS(M)102=low ==> KS(M)206=low SP243=low ถ้านักศึกษาเรียนรายวิชา KS(M)102 ได้เกรดน้อยกว่า B แล้วจะเรียนวิชา KS(M)206 และวิชา SP243 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

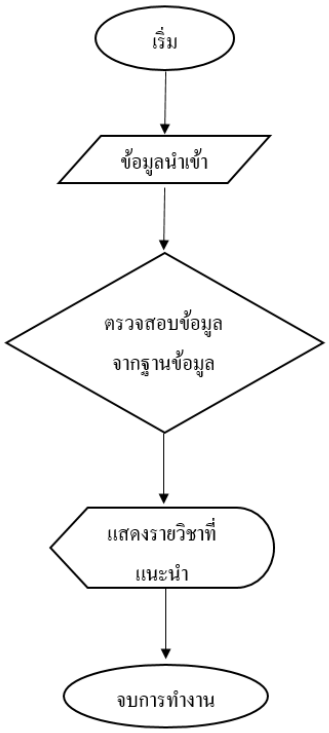
กฎลำดับที่ 8 KS(M)206=low ==> KS(M)102=low SP243=low ถ้านักศึกษาเรียนรายวิชา KS(M)206 ได้เกรดน้อยกว่า B และเรียน KS(M)102 ได้เกรดน้อยกว่า B แล้วเรียนวิชา SP243 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 9 KS(M)102=low SP243=low ==> KS(M)206=low ถ้านักศึกษาเรียนรายวิชา KS(M)102 ได้เกรดน้อยกว่า B และเรียน SP243ได้เกรดน้อยกว่า B แล้วเรียน KS(M)206 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

กฎลำดับที่ 10 KS(M)206=low SP243=low ==> KS(M)102=low ถ้านักศึกษาเรียนรายวิชา KS(M)206 ได้เกรดน้อยกว่า B และเรียน SP243 น้อยกว่า B แล้วเรียนวิชา KS(M)102 ได้เกรดน้อยกว่า B ที่ค่าความเชื่อมั่นที่ 1 และ มีค่าความสัมพันธ์ที่ 1

3.7 การออกแบบการทำงานของโปรแกรม

นำโมเดลที่สองมาสร้างโปรแกรมแนะนำการเลือกลงทะเบียนของนักศึกษามีวัตถุประสงค์เพื่อให้คำแนะนำวิชาต่างๆที่นักศึกษาเลือกแล้วจะได้ผลการเรียนตั้งแต่ B ขึ้นไป



ภาพที่ 23 ผังงาน (Flowchart) แสดงการทำงานของโปรแกรมแนะนำการศึกษา *

3.7.1 การนำกฎความสัมพันธ์จำนวนมากเข้าสู่ฐานข้อมูล

ผู้วิจัยเลือกตั้งค่าโปรแกรมที่ Lower Bound MinSupport เท่ากับ 0.5 Upper Bound MinSupport เท่ากับ 1 Lift Min เท่ากับ 0.95 และได้จำนวนกฎ 50000 กฎ ผู้วิจัยใช้วิธีนำกฎความสำพันธ์ทั้งหมดใส่ไว้ในไฟล์ excel และใช้ ฟังก์ชัน fopen ในภาษา php เป็นตัวกลางในการนำกฎความสัมพันธ์เหล่านี้จัดเก็บไว้ในฐานข้อมูล ตัวอย่างโค้ดที่ใช้งานเกี่ยวกับการอ่านไฟล์ excel ด้วยภาษา php

บรรทัดที่ 1: if (($handle = fopen("csv/".$_FILES["filupload"]["name"], "r")) !== FALSE) {

บรรทัดที่ 2 : while (($data = fgetcsv($handle, 1000, ",")) !== FALSE) {

บรรทัดที่ 3 : }

บรรทัดที่ 4 : }

บรรทัดที่ 1 คือเงื่อนไขพร้อมกับการเรียกใช้ฟังก์ชัน fopen นำมากเก็บไว้ในตัวแปร handle โดยหากเป็นจริงจะทำบรรทัดที่ 2 ใช้ลูป while ในการช่วยอ่านค่าจากฟังก์ชัน fgetcsv โดยข้อมูลแต่ละ record จะนำมาเก็บไว้ในตัวแปร data จากนั้นเขียนคำสั่ง sql เพื่อนำข้อมูลแต่ละครั้งของลูป while บันทึกข้อมูลลงในฐานข้อมูล

3.7.2 การออกแบบข้อมูลนำเข้า

เนื่องจากเราไม่สามารถกำหนดข้อมูลนำเข้า หรือวิชาที่นักศึกษาเรียนแล้วมีผลการเรียนได้ เพราะแต่ละคนเลือกลงทะเบียนไม่เหมือนกัน อาจเกินจากคนละชั้นปีหรือชั้นปีเดียวกันแต่เลือกลงเรียนไม่เหมือนกัน ดังนั้นรูปแบบข้อมูลนำเข้าคือให้นักศึกษาแต่ละคนเลือกวิชาที่ตนเองมีผลการเรียนแล้ว

ตารางที่ 21 ตัวอย่างข้อมูลนำเข้า

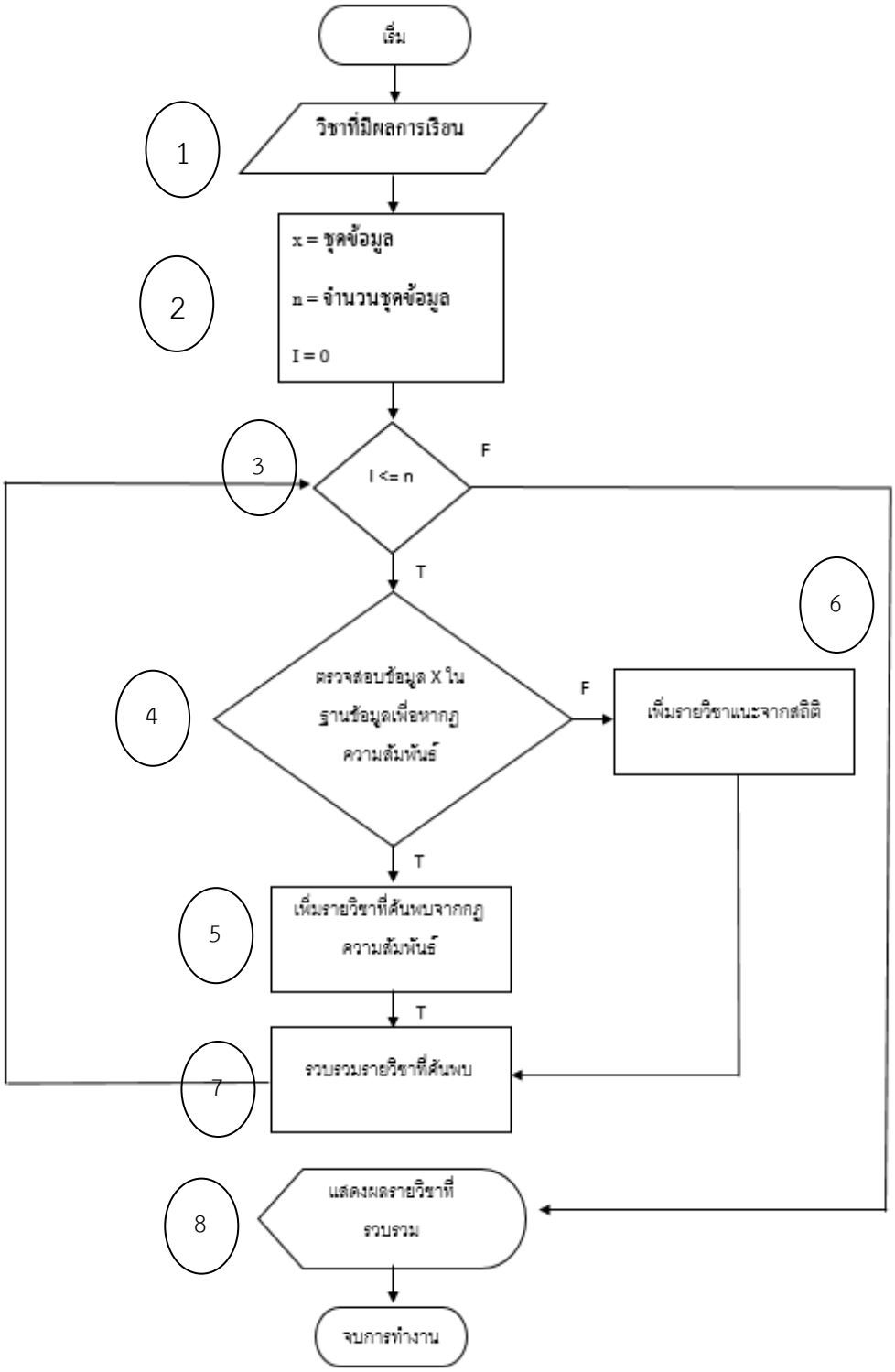
รหัสนักศึกษา	ผลการเรียน
56xxxxxx01	KM100high , KS(M)101low , SP241low , SP242low
56xxxxxx02	KM100low , KS(M)101low , SP241low

3.6.3 การออกแบบเงื่อนไข

กฎความสัมพันธ์ต่าง ๆ ถูกเก็บไว้ในฐานข้อมูลดังนั้นต้องมีการเตรียมข้อมูลเพื่อนำไปค้นหา ผู้วิจัยจึงเลือกใช้ตัวแปรอาเรย์ (Array) คือ ประเภทของข้อมูลที่สามารถเก็บข้อมูลประเภทเดียวกันแบบเป็นลำดับได้ Element ของอาร์เรย์ ลำดับแรกจะเป็น 0 เสมอ โดยข้อมูลนั้นจะอยู่ในตัวแปรตัวเดียวกัน เช่น จากตารางที่ 21 รหัสนักศึกษา 56xxxxxx01 มีข้อมูลในอาเรย์คือ {KM100high , KS(M)101low , SP241low , SP242low} การนำข้อมูลนำเข้าไปข้นหากฎความสัมพันธ์ในฐานข้อมูลนั้นทำได้ด้วยจากการค้นหาทีละรายวิชา ค้นหาที่ละสองวิชา และค้นหาที่ละสามวิชา จากข้อมูลตัวแปรอาเรย์โดยการสลับตำแหน่งไปเรื่อย ๆ

ตารางที่ 20 การแยกข้อมูลเพื่อนำไปประมวลผล

ข้อมูลชุดที่	ข้อมูล	จำนวนข้อมูล
1	{KM100high} , { KS(M)101low} , { SP241low} , {SP242low}	4
2	{KM100high , KS(M)101low} , { SP241low , SP242low}	2
3	{KM100high , KS(M)101low , SP241low } , { KM100high , KS(M)101low, SP242low}	2



ภาพที่ 24 ผังงาน (Flowchart) แสดงเงื่อนไขการทำงานภายในโปรแกรม

จากภาพที่ 24 สามารถอภิบายการทำงานของโปแกรมได้ดังนี้

ขั้นตอนที่ 1 ผู้ใช้งานในที่นี้คือนักศึกษาสามารถเลือกใส่รายวิชาทั้งหมดที่มีผลการเรียน เช่นในตัวอย่างจากตางที่ 19 ข้อมูลที่ 1 รหัสนักศึกษา 56xxxxxx01 มีการใส่ข้อมูลรายวิชาเป็นรหัสวิชาจำนวน 4 วิชา คือ KM100 KS(M)101 SP241 SP242 มีผลการเรียนคือ high = มากกว่า B low = น้อยกว่า B low = น้อยกว่า B low = น้อยกว่า B ตามลำดับ

ขั้นตอนที่ 2 การเตรียมข้อมูลเพื่อนำไปประมวลผลเนื่องจากข้อมูลใส่เข้ามานั้นนักศึกษาแต่ละคนอาจมีรายวิชาที่มีผลการเรียนไม่เท่ากัน อีกทั้งไม่สามารถกำหนดลำดับรายวิชาที่นักศึกษาใส่เข้ามาได้ จึงต้องมีการปรับเปลี่ยนข้อมูลให้เหมาะสมกับการนำไปประมวลผล ตารางที่ 20 มีการนำข้อมูลของนักศึกษารหัส 56xxxxxx01 จำนวน 4 วิชา คือ KM100 KS(M)101 SP241 SP242 นำข้อมูลทั้งหมดมาใส่ในตัวแปรอาเรย์คือ {KM100high , KS(M)101low , SP241low , SP242low} ตามลำดับ แบบข้อมูลเป็นจำนวน 3 ชุด ชุดแรกจำนวน 4 ข้อมูล ชุดที่สองและชุดที่สาม มีจำนวน 2 ข้อมูล รวมเป็น 6 ข้อมูล โดย

x คือ ข้อมูลในรอบนั้นๆ

n คือ จำนวนชุดข้อมูล

l คือ รอบการทำงาน

ขั้นตอนที่ 3 และ ขั้นตอนที่ 4 จำนวนรอบการและตรวจสอบข้อมูลในฐานข้อมูล การกำหนดจำนวนรอบการทำงานโดยจำนวนรอบเป็นจำนวนข้อมูลจากขั้นตอนที่ 2 ในรอบการที่ 1-4 เป็นการนำรหัสวิชาเพียง 1 วิชาไปค้นหากฎความสัมพันธ์หรือสถิติที่มีในฐานข้อมูล ในรอบการทำงานที่ 5-8 เป็นการนำรหัสวิชาจำนวน 2 วิชาไปค้นหากฎความสัมพันธ์หรือสถิติที่ที่มีในฐานข้อมูลโดยรอบที่ 5 และ 6 เป็นการสลับตำแหน่งกัน และรอบที่ 7 และ 8 เป็นการสลับตำแหน่งข้อมูลเช่นกัน ในรอบการทำงานที่ 9-21 เป็นการนำรหัสวิชาจำนวน 2 ชุดชุดละ 3 วิชาไปค้นหากฎความสัมพันธ์หรือสถิติโดยสลับตำแหน่งของแต่ละวิชาไปเลื่อยๆ ดังตารางที่ 24

ตารางที่ 21 การแจกแจงรอบการทำงานจากตัวอย่างขั้นตอนที่ 2

รอบการทำงานที่	ข้อมูล
1	KM100high
2	KS(M)101low
3	SP241low
4	SP242low
5	KM100high , KS(M)101low
6	KS(M)101low , KM100high
7	SP241low , SP242low
8	SP242low , SP241low
9	KM100high , KS(M)101low , SP241low
10	KM100high , SP241low , KS(M)101low
11	SP241low , KS(M)101low , KM100high
12	SP241low , KM100high , KS(M)101low
14	KS(M)101low , KM100high , SP241low
15	KS(M)101low , SP241low , KM100high
16	KM100high , KS(M)101low , SP242low
17	KM100high , SP242low , KS(M)101low
18	KS(M)101low , SP242low , KS(M)101low
19	KS(M)101low , KM100high , SP242low
20	SP242low , KS(M)101low , KM100high
21	SP242low , KM100high , KS(M)101low

ขั้นตอนที่ 5 ขั้นตอนที่ 6 เพิ่มรายวิชาที่ค้นพบในฐานข้อมูลจากกฎความสัมพันธ์ และสถิติ ตัวอย่าง เช่น หาค้นพบกฎความสัมพันธ์ KS(M)206=low SP243=low ==> KS(M)102=low หมายความว่าหลังเครื่องหมาย “==>” คือวิชาที่จะนำไปแนะนำนักศึกษาว่าควรหรือไม่ควรเลือก

ขั้นตอนที่ 7 รวบรวมรายวิชาที่ค้นพบทั้งจากกฎความสัมพันธ์ และสถิติ

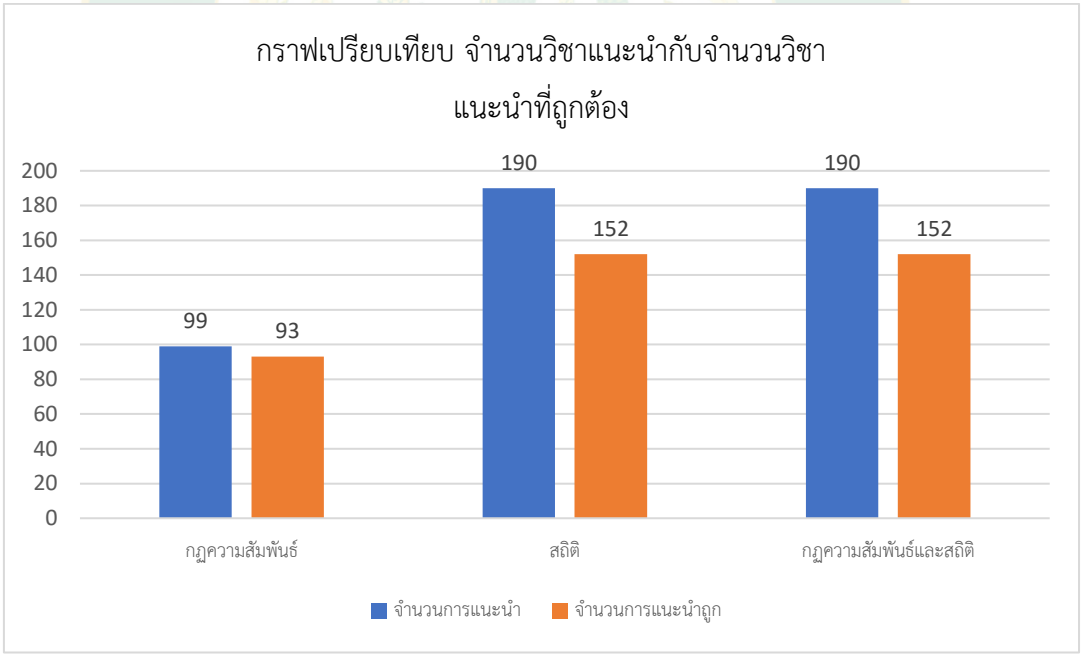
ขั้นตอนที่ 8 แสดงรายวิชาที่ค้นพบเป็นวิชาที่ระบบแนะนำ โดยวิชาที่จะแนะนำต้องไม่ซ้ำกับรายวิชาที่ใส่เข้ามา หรือไม่ซ้ำกับขั้นตอนที่ 1 นั้นเอง

3.8 วิธีการวัดประสิทธิภาพข้อมูล

สุ่มข้อมูลนำเข้าจากข้อมูลที่แบ่งไว้สำหรับทดสอบตัวต้นแบบ โดยคละรายวิชาให้มีความแตกต่างให้เหมือนกับมีนักศึกษามาใช้งานจริง เพราะในการใช้งานจริงเราไม่สามารถกำหนดรายวิชาที่นักศึกษามีผลการเรียนได้ ดังนั้นข้อมูลที่ทดสอบจึงเป็นข้อมูลรูปแบบสุ่มเฉลี่ยจำนวน 4 วิชา และเปรียบเทียบความถูกต้องจากโปรแกรม

3.9 ผลการวัดประสิทธิภาพ

เนื่องจากข้อมูลที่นำมาทดสอบเป็นข้อมูลที่เตรียมไว้จึงไม่สามารถกำหนดรายวิชานำเข้า และวิชาที่แนะนำได้ ดังนั้นจึงมีบางรายวิชาที่ระบบแนะนำแล้วแต่ไม่มีผลเฉลย ผู้วิจัยจะไม่นำข้อมูลส่วนนี้มาวิเคราะห์ร่วมด้วย นำเฉพาะข้อมูลที่มีผลเฉลยเท่านั้น และได้ผลวิจัยว่ามีวิชาที่มีผลทำนาย 33 วิชา แบ่งเป็น กฎความสัมพันธ์จำนวน 9 วิชา และสถิติ 33 วิชาจากภาพที่ 10



ภาพที่ 25 กราฟเปรียบเทียบจำนวนวิชาที่แนะนำ และจำนวนที่แนะนำถูกต้องของแต่ละวิธี

จำนวนการแนะนำวิชาทั้งหมดจำนวน 479 วิชา แนะนำถูกต้อง 397 ครั้ง คิดเป็น 82.88 เปอร์เซ็นต์ แบ่งออกเป็นจากกฎความสัมพันธ์ 93 ครั้ง จากสถิติ 152 ครั้ง จากกฎความสัมพันธ์และสถิติ 152 ครั้ง ดังภาพที่ 25 โดยกฎความสัมพันธ์สามารถแนะนำได้ทั้งหมด 8 วิชา สถิติ 8 วิชา กฎความสัมพันธ์และสถิติ 16 วิชา ดังตารางที่ 21 จึงสามารถบอกได้ว่าการแนะนำวิชาด้วยกฎความสัมพันธ์มีเปอร์เซ็นต์ถูกมากกว่าวิธีอื่น แต่ในเรื่องของจำนวนวิชาที่แนะนำไม่สามารถสู้วิธีอื่นได้ ดังนั้นจึงต้องมีการประยุกต์การใช้งานให้เหมาะสม เช่น อาจต้องมีการเปรียบเทียบวิชาที่เหมือนกันของสถิติ และกฎความสำพันธ์ในแต่ละครั้งที่ค้นหา อาจต้องพิจารณาเลือกเชื่อกฎความสัมพันธ์ในเฉพาะวิชาที่มีผล และเลือกเชื่อสถิติต่อเมื่อวิชานั้นไม่มีอยู่ในกฎความสัมพันธ์

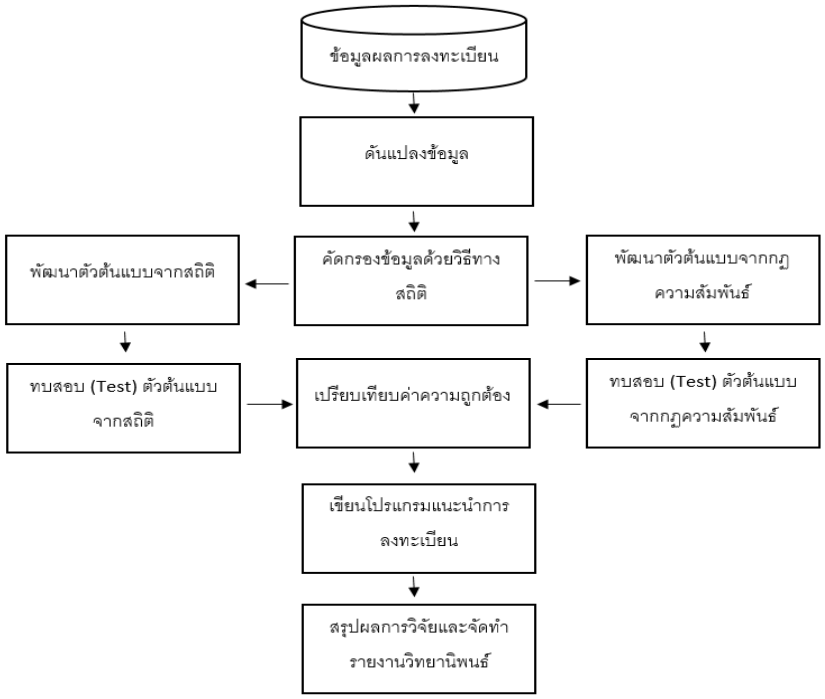
ตารางที่ 22 การเปรียบเทียบจำนวนเปอร์เซ็นต์ถูกของแต่ละวิธี

วิธี	ร้อยละวิชาที่มีผล	เปอร์เซ็นต์แนะนำถูก
กฎความสัมพันธ์	30	94.09
สถิติ	100	79.89
กฎความสัมพันธ์และสถิติ	100	79.74

หากเปรียบเทียบที่ความถูกต้องนั้นการแนะนำวิชาแบบใช้กฎความสัมพันธ์มีเปอร์เซ็นต์ความถูกต้องมากกว่าวิธีอื่นที่ร้อยละ 94.09 แต่สามารถแนะนำได้เพียงร้อยละ 30 ของวิชาทั้งหมด 33 วิชา หรือประมาณ 9 วิชาเท่านั้น

3.10 สรุปผลการวิจัยและจัดทำรายงานวิทยานิพนธ์

หลังการวิจัยตามวัตถุประสงค์แล้ว ผู้วิจัยจะดำเนินการสรุปผลการวิจัยและจัดทำรายงานวิทยานิพนธ์ แผนภาพที่ 26



ภาพที่ 26 แผนผังสรุปการวิจัย

บทที่ 4
ผลการดำเนินการวิจัย

การศึกษาได้ดำเนินงานตามวัตถุประสงค์ การสร้างโมเดลโดยวิธีการทางการทำเหมืองข้อมูล (Data Mining) ที่มาจากผลการเรียนของนักศึกษา เปรียบเทียบและประยุกต์ใช้ในการพัฒนาระบบแนะนำการลงทะเบียนของนักศึกษา เพื่อช่วยลดโอกาสการมีผลการเรียนไม่ถึงเกณฑ์ดีของนักศึกษา

ผู้วิจัยได้ดำเนินการทดลองตามกระบวนการ CRISP-DM ในบทที่ 3 กล่าวถึงขั้นตอนที่ 4 เป็นขั้นตอนที่สร้างตัวต้นแบบและการสร้างโปรแกรม โดยบทที่ 4 จะกล่าวถึงขั้นตอนที่เหลือ และการวิเคราะห์ เช่น ขั้นตอนการทดสอบความแม่นยำของระบบ และขั้นตอนการนำไปใช้งาน เป็นต้น จึงจะอภิบายเป็นหัวข้อดังนี้

4.1 การวิเคราะห์ปัญหา และการเลือกปัญหาจากข้อมูลที่มี

ในงานวิจัยนี้กล่าวถึงการเลือกลงทะเบียนเรียน และการแก้ปัญหาผลการเรียนลดลง จากการหาข้อมูลพบว่านักศึกษาชั้นปีที่ 3 มีปัญหาเกี่ยวกับผลการเรียนเฉลี่ยลดลงในวิชาภาคที่ร้อยละ 70.5 และวิชาภายนอกที่ร้อยละ 88.2 ดังนั้นจึงมีการตัดสินใจเลือกแก้ปัญหาที่วิชาภายนอกมากกว่าวิชาภาค และบางครั้งการเลือกเรียนวิชาของนักศึกษาอาจต้องใช้การแนะนำเพื่อประกอบการตัดสินใจ การแนะนำที่ดีจึงต้องบอกความเป็นไปได้ของผลลัพธ์ในวิชานั้น หากวิชาที่เลือกลงทะเบียนมีคำแนะนำเป็นไม่แนะนำ นักศึกษาสามารถนำคำแนะนำของระบบเป็นการเตือนว่าวิชาที่เลือกลงทะเบียนนั้นยากให้เตรียมความพร้อมให้ดียิ่งขึ้น

4.2 การเตรียมข้อมูล

เมื่อทราบปัญหา และกำหนดวิธีการแก้ไขปัญหาแล้ว การเตรียมข้อมูลเป็นสิ่งสำคัญอย่างยิ่งในการใช้แก้ไขปัญหา ข้อมูลที่ได้มามักเป็นข้อมูลดิบ (Raw Data) การเตรียมข้อมูลจึงมักเกี่ยวข้องกับเครื่องมือที่ใช้ในการแก้ไขปัญหา ดังนั้นการเลือกเครื่องมือที่จะใช้จึงเป็นเรื่องที่ควบคู่กับการเตรียมข้อมูล บางครั้งการเลือกเครื่องมือที่ใช้งานจึงต้องคำนึงถึงข้อมูลดิบ (Raw Data) ที่มี หรือการเตรียมข้อมูลจากข้อมูลดิบ (Raw Data) จะต้องคำนึงถึงเครื่องมือที่ใช้งาน

4.3 การเลือกเทคนิคที่เหมาะสม

ในงานวิจัยนี้เลือกใช้หลักสถิติ และกฎความสัมพันธ์ (Association rule) ในการเปรียบเทียบโอกาสที่จะเกิดขึ้นของข้อมูล นั้นคือโอกาสที่นักศึกษาเลือกเรียนวิชาที่ระบบแนะนำแล้วมีความเป็นไปได้ที่จะได้ผลการเรียนอยู่ในเกณฑ์ดี

4.4 ผลการทดสอบความแม่นยำของตัวต้นแบบ (Evaluation)

เป็นขั้นตอนที่ 5 ในกระบวนการ CRISP-DM ผลการทดสอบระบบกับข้อมูลชุดเดิมที่ใช้ในการสร้างตัวต้นแบบ พบว่าวิธีกฎความสัมพันธ์สามารถแนะนำวิชาถูกที่ร้อยละ 86.22 ร้อยละวิชาที่มีผลเท่ากับ 46.87 วิธีสถิติสามารถแนะนำวิชาถูกที่ร้อยละ 78.94 ร้อยละวิชาที่มีผลเท่ากับ 100 หากนำวิธีกฎความสัมพันธ์ และวิธีสถิติรวมกันสามารถแนะนำวิชาถูกที่ร้อยละ 86.22 ร้อยละวิชาที่มีผลเท่ากับ 100 ดังตารางที่ 23 จึงสรุปได้ว่าผลการทดสอบกับข้อมูลชุดเดิมที่ใช้ในการสร้างตัวต้นแบบนั้น วิธีกฎความสัมพันธ์สามารถแนะนำวิชาถูกได้มากกว่าวิธีสถิติที่ร้อยละ 7.28 แต่วิธีกฎความสัมพันธ์นั้นสามารถแสดงวิชาที่มีผลได้ร้อยละ 46.87 ต่างจากวิธีทางสถิติที่ร้อยละ 53.13 แต่หากนำวิธีทั้งสองมาแนะนำ ระบบมีร้อยละที่แนะนำถูกที่ 79.17 และ ร้อยละวิชาที่มีผลที่ 100 จึงกล่าวได้ว่าการนำทั้งสองวิธีใช้ร่วมกันกับข้อมูลชุดเดิมจะมีร้อยละที่แนะนำถูกมากที่สุด และทีร้อยละวิชาที่มีผลมากที่สุด

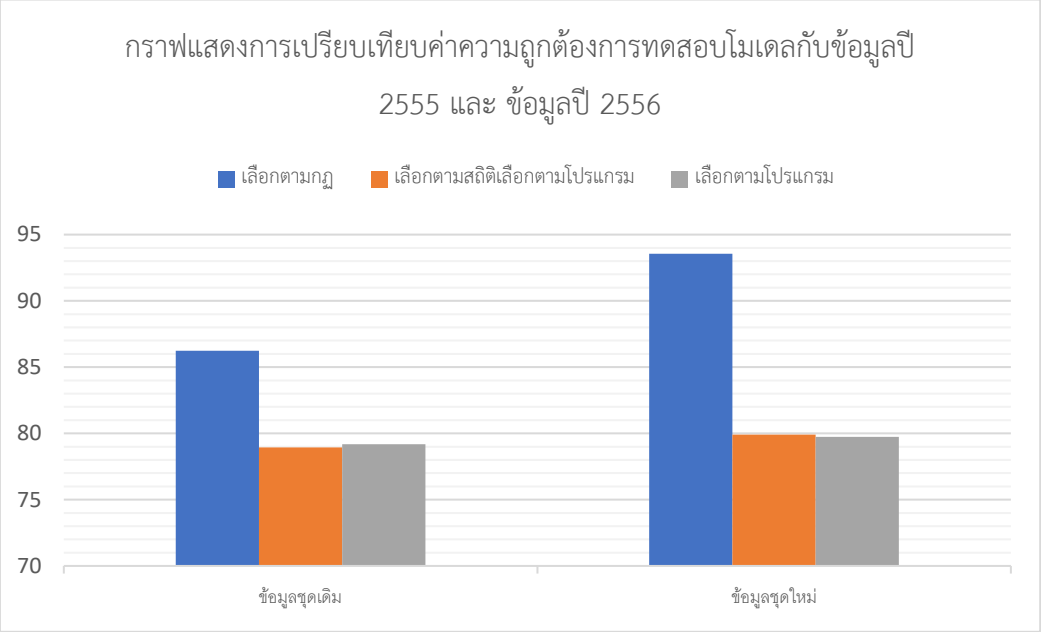
ตารางที่ 23 ผลการทดสอบกับข้อมูลชุดเดิม

วิธี	ร้อยละที่แนะนำถูก	ร้อยละวิชาที่มีผล
กฎความสัมพันธ์	86.22	46.87
สถิติ	78.94	100
กฎความสัมพันธ์ และสถิติ	79.17	100

ผลการทดสอบระบบกับข้อมูลชุดที่สองที่ใช้ในการสร้างตัวต้นแบบ พบว่าวิธีกฎความสัมพันธ์สามารถแนะนำวิชาถูกที่ร้อยละ 93.56 ร้อยละวิชาที่มีผลเท่ากับ 50 วิธีสถิติสามารถแนะนำวิชาถูกที่ร้อยละ 79.89 ร้อยละวิชาที่มีผลเท่ากับ 100 หากนำกฎความสัมพันธ์ และสถิติรวมกันสามารถแนะนำวิชาถูกที่ร้อยละ 79.74ร้อยละวิชาที่มีผลเท่ากับ 100 ดังตารางที่ 24 จึงสรุปได้ว่าผลการทดสอบกับข้อมูลชุดใหม่นั้น วิธีกฎความสัมพันธ์สามารถแนะนำวิชาถูกได้มากกว่าวิธีสถิติที่ร้อยละ 13.67 แต่วิธีกฎความสัมพันธ์นั้นสามารถแสดงวิชาที่มีผลได้ร้อยละ 50 ต่างจากวิธีทางสถิติที่สามารถแนะนำรายวิชาได้ทั้งหมด แต่หากนำวิธีทั้งสองมาแนะนำ ระบบมีร้อยละที่แนะนำถูกที่ 79.17 และแนะนำรายวิชาได้ทั้งหมด จึงกล่าวได้ว่าการนำทั้งสองวิธีใช้ร่วมกันกับข้อมูลชุดที่เตรียมไว้จะมีร้อยละที่แนะนำถูกมากที่สุดที่ร้อยละ 79.89 และทีร้อยละวิชาที่มีผลมากที่สุดคือทุกวิชา

ตารางที่ 24 ผลการทดสอบกับข้อมูลอีกชุด

วิธี	ถูกเฉลี่ย	ร้อยละวิชาที่มีผล
กฎความสัมพันธ์	93.56	50
สถิติ	79.89	100
กฎความสัมพันธ์ และสถิติ	79.74	100



ภาพที่ 27 การเปรียบเทียบค่าความถูกต้องการทดสอบโมเดลกับข้อมูลชุดเดิม และชุดใหม่

4.6 การนำไปใช้งาน (Deployment)

เป็นขั้นตอนสุดท้ายในกระบวนการ CRISP-DM กล่าวถึงการนำความรู้ที่ได้ มาใช้เป็นเงื่อนไขในการเขียนโปรแกรม ระบบสนับสนุนการตัดสินใจเพื่อการลงทะเบียนของนักศึกษาสาขาวิทยาการคอมพิวเตอร์ เพื่อเป็นเครื่องมือช่วยนักศึกษาสาขาวิทยาการคอมพิวเตอร์ในการตัดสินใจเลือกลงทะเบียนวิชาใดๆ

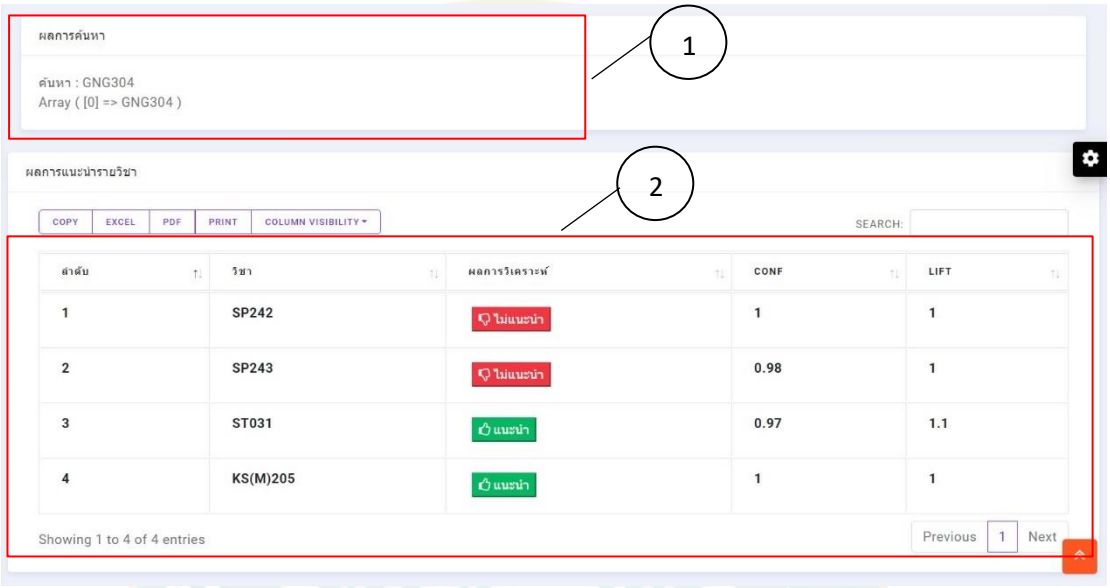
4.6.1 การออกแบบข้อมูลนำเข้า

ผู้วิจัยใช้ภาษา PHP เป็นเครื่องมือในการสร้างพร้อมออกแบบระบบ และติดต่อกับฐานข้อมูล My SQL เพื่อบันทึกข้อมูลกฎความสัมพันธ์ลงในฐานข้อมูล สร้างเงื่อนไขการค้นหา ประกอบไปด้วย ชื่อวิชา และผลการเรียน โดยนักศึกษาสามารถเพิ่มรายวิชาได้ตามที่มีผลการเรียนในภาพที่ 28 โปรแกรมจะวิเคราะห์และแสดงผลลัพธ์วิชาที่แนะนำลงทะเบียนเรียน พร้อมรายละเอียด เช่น ค่า Support ค่า Confidence และ ค่า Lift เป็นต้น

ภาพที่ 28 หน้าจอการกรอกข้อมูลที่ต้องการนำไปค้นหาวิชาที่แนะนำ

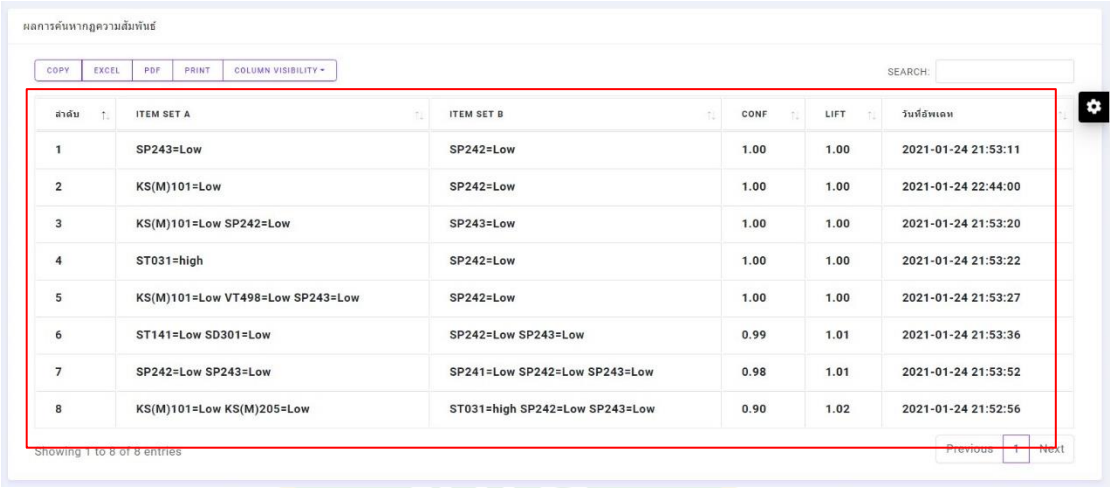
4.6.2 การแสดงผล

ในภาพที่ 29 จุดที่ 1 แสดงผลการค้นหา และจุดที่ 2 แสดงผลการค้นหารายวิชาที่ควรเลือกและไม่ควรเลือก โดยหากควรเลือกจะเป็นปุ่มสีเขียว หากไม่ควรเลือกจะเป็นปุ่มสีแดง แสดงรายละเอียดดังนี้จุดที่ 1 วิชาที่นำเข้า จุดที่ 2 คือลำดับรายวิชาที่แนะนำ วิชา คือ วิชาที่ค้นพบ confident คือค่าความเชื่อมั่นสูงสุดที่ค้นพบ กรอบหมายเลขสามแสดงกฎทั้งหมดที่ค้นพบ ประกอบไปด้วย ลำดับ กฎความสัมพันธ์ (frequent item set) conf และ Lift



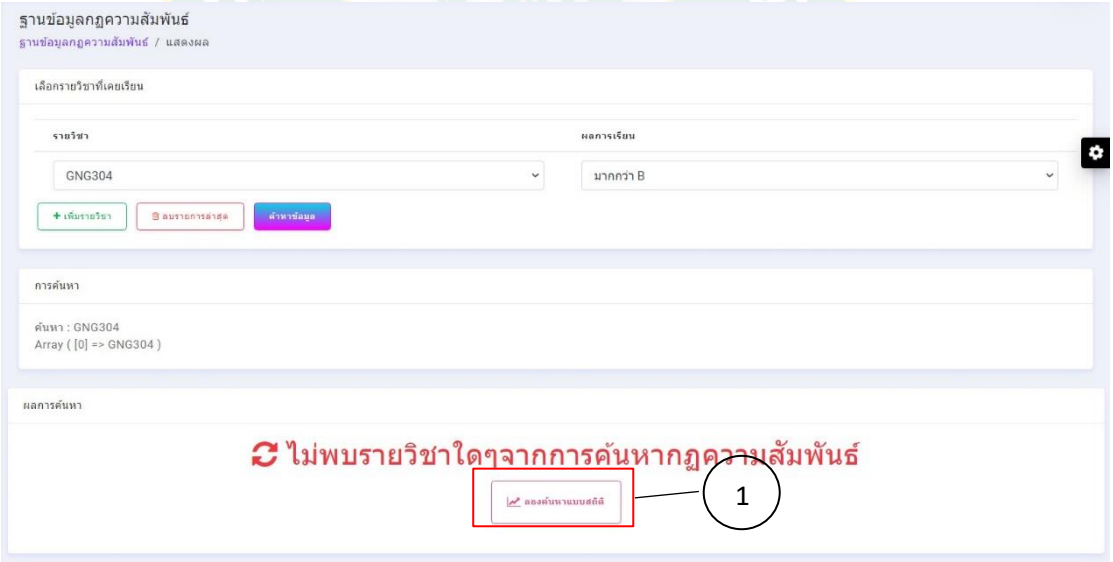
ภาพที่ 29 หน้าจอแสดงรายงานผลการค้นหารายวิชาที่ควรเลือก

ในภาพที่ 30 เป็นภาพต่อเนื่องจากภาพที่ 29 แสดงผลในหน้าจอเดียวกัน โดยแสดงผลลัพธ์จากกฎความสัมพันธ์ที่พบจากการค้นหา เพื่อให้ผู้ใช้งานสามารถตรวจสอบผลลัพธ์จากการค้นหา หรืออาจพิจารณาวิชาที่ต้องการจากกฎความสัมพันธ์ได้โดยตรง โดยผู้ใช้งานอาจต้องมีความเข้าใจเรื่องกฎความสัมพันธ์เบื้องต้นด้วย

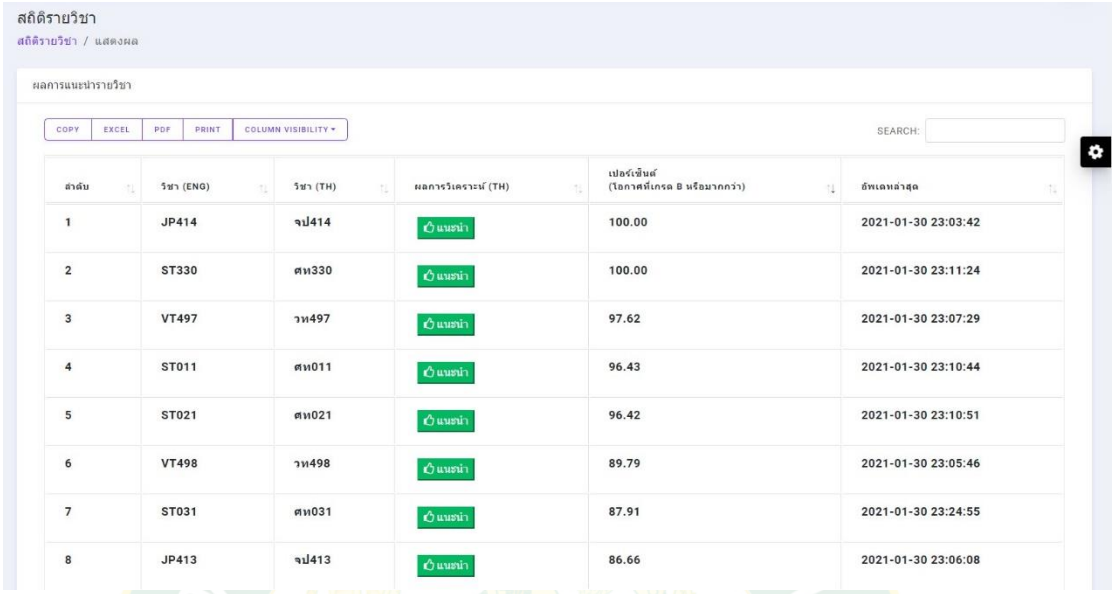


ภาพที่ 30 หน้าจอแสดงรายงานผลการค้นหารายวิชาที่ควรเลือกแบบกฎความสัมพันธ์

หากผลการค้นหาด้วยกฎไม่พบจะมีปุ่มให้เลือกค้นหาต่อไปในรูปแบบการโมเดลทางสถิติ ในภาพที่ 31 จุดที่ 1 ในกรณีที่เลือกค้นหารายวิชาจากสถิติแสดงผลการค้นหารายวิชาที่ควรเลือก และไม่ควรเลือก ในภาพที่ 32 โดยหากควรเลือกจะเป็นปุ่มสีเขียว หากไม่ควรเลือกจะเป็นปุ่มสีแดง แสดงรายละเอียดดังนี้ ลำดับ คือลำดับรายวิชาที่แนะนำ วิชา คือ วิชาที่ค้นพบ ผลการวิเคราะห์ เปอร์เซ็นต์ที่เรียนได้มากกว่าเกรด B และวิชาที่ปรากฏจะไม่ซ้ำกับวิชาที่เป็นข้อมูลนำเข้า



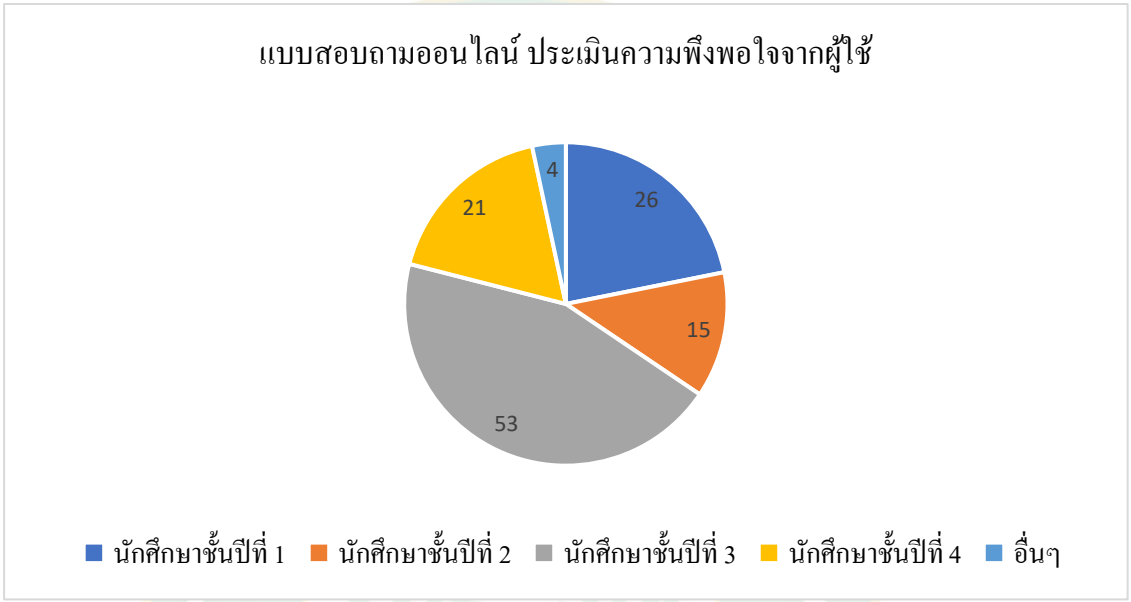
ภาพที่ 31 หากไม่ค้นพบรายวิชาจะปรากฏปุ่มเพื่อค้นหาแบบสถิติ



ภาพที่ 32 หน้าจอแสดงผลการค้นหารายวิชาจากสถิติ

4.6.2 การประเมินความพึงพอใจจากผู้ใช้

แบบสอบถามออนไลน์โดยนักศึกษาสาขาวิทยาการคอมพิวเตอร์ โดยผู้ตอบแบบสอบถามจำนวน 119 คน แบ่งเป็น เพศหญิง 56 คน และชาย 63 คน แบ่งเป็น ในภาพที่ 33 แบ่งเป็นนักศึกษาชั้นปี 1 จำนวน 26 คน นักศึกษาชั้นปี 2 จำนวน 15 คน นักศึกษาชั้นปี 3 จำนวน 53 คน นักศึกษาชั้นปี 4 จำนวน 21 คน และ อื่นๆ 4 คน ซึ่งจำนวนมากที่สุดคือนักศึกษาชั้นที่ 3



ภาพที่ 33 หน้าจอแสดงผลการค้นหารายวิชาจากสถิติ

ตารางที่ 25 แสดงผลการประเมินความพึงพอใจของนักศึกษาที่ทดลองใช้งานระบบ โดยแบ่งออกเป็น 4 ประเด็นความพึงพอใจ คือ

ประเด็นแรก “สามารถช่วยให้ผลการศึกษาดีขึ้น” เพื่อวัดความพึงพอใจโดยภาพรวมว่าระบบสามารถช่วยให้นักศึกษามีผลการศึกษาดีขึ้นจากการแนะนำรายวิชา โดยผลประเมินพบว่า มีผลมากที่สุด ที่ร้อยละ 51.26 มีผลเล็กน้อย ที่ร้อยละ 41.17 และไม่มีผล ที่ร้อยละ 7.56

ประเด็นที่สอง “สามารถช่วยให้ฉันเลือกลงทะเบียนรายวิชาอย่างถูกต้อง สอดคล้องกับความเป็นจริง และมีความน่าเชื่อถือ” เพื่อวัดรายวิชาที่ระบบแนะนำนั้น สามารถแนะนำวิชาได้ถูกต้องเหมาะสม และมีความน่าเชื่อถือหรือไม่ โดยผลประเมินพบว่า มีผลมากที่สุด ที่ร้อยละ 53.78 มีผลเล็กน้อย ที่ร้อยละ 36.97 และไม่มีผล ที่ร้อยละ 9.24

ประเด็นที่สาม “สามารถช่วยให้ฉันเลือกลงทะเบียนรายวิชาทั่วไปง่ายขึ้น” เพื่อวัดว่าระบบสามารถช่วยให้นักศึกษาที่ใช้งานสามารถเลือกลงทะเบียนง่ายขึ้นหรือไม่ โดยผลประเมินพบว่า มีผลมากที่สุด ที่ร้อยละ 57.14 มีผลเล็กน้อย ที่ร้อยละ 35.29 และไม่มีผล ที่ร้อยละ 7.5

ประเด็นสุดท้าย “สามารถช่วยให้ฉันเลือกลงทะเบียนวิชาทั่วไปที่สอดคล้องกับความสามารถของฉัน”เพื่อวัดรายวิชาที่ระบบแนะนำนั้น สามารถแนะนำวิชาที่สอดคล้องกันกับความสามารถของนักศึกษาที่ใช้งานหรือไม่ โดยผลประเมินพบว่า มีผลมากที่สุด ที่ร้อยละ 52.94 มีผลเล็กน้อย ที่ร้อยละ 38.65 และไม่มีผล ที่ร้อยละ 8.4

ตารางที่ 25 ความพึงพอใจของนักศึกษาที่ทดลองใช้งานระบบ

ประเด็นความพึงพอใจ	**มีผลมากที่สุด (ร้อยละ)**	**มีผลเล็กน้อย (ร้อยละ)**	**ไม่มีผล (ร้อยละ)**
1. สามารถช่วยให้ผลการศึกษาดีขึ้น	61 (51.26)	49 (41.17)	9 (7.56)
2. สามารถช่วยให้ฉันเลือกลงทะเบียนรายวิชาอย่างถูกต้อง สอดคล้องกับความเป็นจริง และมีความน่าเชื่อถือ	64 (53.78)	44 (36.97)	11 (9.24)
3. สามารถช่วยให้ฉันเลือกลงทะเบียนรายวิชาทั่วไปง่ายขึ้น	68 (57.14)	42 (35.29)	9 (7.5)
4. สามารถช่วยให้ฉันเลือกลงทะเบียนวิชาทั่วไปที่สอดคล้องกับความสามารถของฉัน	63 (52.94)	46 (38.65)	10 (8.4)

บทที่ 5
สรุปและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

ระบบสนับสนุนการตัดสินใจเพื่อการลงทะเบียนของนักศึกษาสาขาวิทยาการคอมพิวเตอร์ มีวัตถุประสงค์เพื่อแนะนำว่าวิชาไหนควรเลือกลงทะเบียน ลงทะเบียนแล้วจะได้ผลการเรียนที่ดี หรือถ้าไม่อาจเลี่ยงการลงทะเบียนแล้ว สามารถใช้ข้อมูลในระบบเป็นการเตือนว่าวิชานั้นง่ายหรือยาก และให้นักศึกษาระมัดระวังในการเรียนเพิ่มขึ้นได้

ในขั้นตอนการพัฒนาแบบจำลองได้ใช้กระบวนการทำเหมืองข้อมูลแบบ CRISP-DM เป็นแนวทางในการวิจัย โดยใช้วิธีการหากฎความสัมพันธ์ และสถิติ เพื่อนำเป็นระบบสนับสนุนการตัดสินใจการลงทะเบียนเรียนของนักศึกษาสาขาวิทยาการคอมพิวเตอร์ โดยเริ่มจากการวิเคราะห์ปัญหา ศึกษาข้อจำกัดต่างๆ การรวบร่วมแก้ไขข้อมูล

โมเดลแรกผลจากการสร้างการทดสอบต้นแบบสถิติที่สร้างจากข้อมูลปี 2555 มาทำการแนะนำวิชาให้กับข้อมูลปี 2556 ภาพที่ 14 เลือกข้อมูลเฉพาะวิชาที่มีการลงทะเบียนเหมือนกันเพื่อเปรียบเทียบ พบว่านักศึกษาเลือกวิชาตามที่ระบบแนะนำจำนวน 15 วิชา มี 13 วิชาที่มีนักศึกษาส่วนใหญ่เรียนในเกณฑ์ดี กล่าวคือระบบสามารถแนะนำวิชาได้ถูกต้อง 86.67 % ในภาพที่ 13 แต่อย่างไรก็ตามการสร้างโมเดลทางสถิติอาจยังไม่สามารถความน่าเชื่อมากนักเนื่องจาก เป็นวิธีการพื้นฐาน ดังนั้นจะต้องมีวิธีการอื่นๆเปรียบเทียบเพื่อเพิ่มความน่าเชื่อถือของผู้ใช้งานระบบ

โมเดลที่สองผลจาการสร้างโปรแกรมระบบสนับสนุนการตัดสินใจ หลังจากนั้นต้องทดสอบและวัดประสิทธิภาพของโปรแกรม ผลปรากฏว่าจากเทคนิคกฎความสัมพันธ์สามารถแนะนำวิชาถูกที่ร้อยละ 86.22 ร้อยละวิชาที่มีผลเท่ากับ 50 สถิติสามารถแนะนำวิชาถูกที่ร้อยละ 78.94 ร้อยละวิชาที่มีผลเท่ากับ 100 หากนำกฎความสัมพันธ์ และสถิติรวมกันสามารถแนะนำวิชาถูกที่ร้อยละ 86.22 ร้อยละวิชาที่มีผลเท่ากับ 100

การเลือกใช้โมเดลจากต้นแบบสถิติอาจเหมาะกับสามารถนำไปแนะนำในวงกว้างของนักศึกษาอาจมีข้อผิดพลาดเรื่องความถูกต้องบ้างเนื่องจากไม่ได้พิจารณาความสัมพันธ์การเชื่อมโยงของแต่ละรายวิชา

โมเดลที่สองที่นำไปสร้างโปรแกรมระบบสนับสนุนการตัดสินใจสามารถสร้างความเชื่อมันได้มากกว่าโมเดลจากต้นแบบสถิติ โดยเฉพาะการแนะนำรายวิชาที่เจาะจง เพราะพิจารณาจากรายวิชาที่นักศึกษาคนนั้นๆเคยเรียนผ่านมาแล้วร่วมด้วย

5.2 ปัญหาและอุปสรรค

อัลกอริทึม Apriori คือการนำข้อมูลทั้งหมดรวมกัน แล้วจึงจะสามารถนำข้อมูลทั้งหมดนั้น ไปค้นหากฎความสัมพันธ์ใหม่ทั้งหมด โดยไม่สามารถนำกฎความสัมพันธ์ที่หาได้จากกลุ่มข้อมูลเก่าก่อนหน้ามาใช้ง่านร่วมได้ ทำให้ต้องเสียเวลาในการประมวลผลใหม่เพื่อค้นหากฎความสัมพันธ์ทั้งหมดใหม่

การวิจัยนี้มีการใช้ข้อมูลของนักศึกษาจำนวนจำไม่ได้คำนึงถึงปัจจัยภายนอก เช่น เพศ ฐานะทางการเงิน ซึ่งอาจเหมาะสมกับตัวอย่าง หรือรูปแบบข้อมูลนี้ ซึ่งอาจมีผลทำให้แบบจำลองอาจมีเพิ่มหรือลดค่าความแม่นยำในการแนะนำวิชาของระบบ

เนื่องจากการปรับเปลี่ยนหลักสูตรให้ทันสมัย จำนวนข้อมูลจึงมีไม่มากเพื่อการนำมาวิเคราะห์อาจมีผลในค่าความถูกต้องของระบบ

ในวิธีการค้นหาความสัมพันธ์ของข้อมูลนั้น อาจได้กฎความสัมพันธ์เป็นจำนวนมาก จึงทำให้เลือกใช้งานกฎความสัมพันธ์ยาก ซึ่งการเลือกนำกฎสัมพันธ์มาใช้อาจต้องมีการพิจารณาที่ละเอียดจากผู้เชี่ยวชาญในการเลือกกฎความสัมพันธ์ และตัดกฎความสัมพันธ์ที่ไม่ต้องการใช้งานออกไป ขั้นตอนเหล่านี้อาจมีผลกับประสิทธิภาพของตัวต้นแบบ แต่ยังไม่ได้ถูกกล่าวถึงในงานวิจัยนี้

ในวิธีการสร้างตัวต้นแบบจากสถิตินั้น ถึงแม้จะสามารถบอกถึงจำนวนรายวิชาได้มากกว่ากฎความสัมพันธ์ แต่ไม่สามารถบอกถึงพฤติกรรมที่แท้จริงของนักศึกษาได้ จึงต้อมีการประยุกต์ใช้กับวิธีการอื่นเพื่อเพิ่มประสิทธิภาพและความแม่นยำ

5.3 ข้อเสนอแนะ

การพัฒนาการศึกษานั้นสามารถทำได้หลายวิธี การหากฎความสัมพันธ์เป็นหนึ่งในวิธีที่นิยมใช้มาช่วยในการสร้างระบบการตัดสินใจ เนื่องจากสร้างขึ้นจากความถี่ของข้อมูล แต่หากต้องการการประสิทธิของตัวต้นแบบที่ดีที่สุดต้องมีการเปรียบเทียบวิธีการอื่นๆ ควรคู่ไปได้ เช่น เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เป็นเทคนิคหนึ่งที่ใช้ในการช่วยตัดสินใจ เทคนิคต้นไม้ตัดสินใจเป็นเทคนิคหนึ่งในการจำแนกประเภท (Classification) ซึ่งเป็นวิธีการแบ่งประเภทหรือแยกหมวดหมู่ข้อมูล โดยการจำแนกประเภทนั้น เป็นเทคนิคหนึ่งของการทำเหมืองข้อมูล แต่ในงานวิจัยเลือกที่จะใช้เทคนิค

ค้นหาความสัมพันธ์เนื่องจากเป็นวิธีการที่จะได้รูปแบบ (Pattern) ของวิชาที่จะมีผลต่อกันออกมาในรูปแบบ ไอเทมเซ็ต X และ Y ทำให้ง่ายต่อการนำไปใช้กับฐานข้อมูล เพราะมีการแยกรูปแบบที่ชัดเจนต่างจากเทคนิคต้นไม้ตัดสินใจ ที่อาจต้องหาหารูปแบบเองภายหลัง

ปริมารณข้อมูลที่ใช้ในการวิจัยนี้มีไม่มากนักอาจส่งผลกับประสิทธิภาพของระบบ และข้อมูลที่น่าสนใจในการวิเคราะห์อื่น เช่น ผู้สอนที่อาจมีอิทธิพล (influence) ที่จะมีผล (dominate) กับผลการเรียน ขาดการพิจารณารายวิชาที่มีผลการเรียนเป็น F และ การลงทะเบียนเรียนรายวิชาซ้ำเป็นต้น

ระบบนี้อาจมีรูปแบบการใช้งานแบบอื่น เช่น อาจให้อาจารย์ที่ปรึกษาใช้ในการแนะนำนักษาเป็นต้น

ระบบอาจมีการเพิ่มเติมการใช้งาน (feature) เช่น การเลือก (selection) การเชื่อมต่อข้อมูลส่วนบุคคลของนักศึกษา เป็นต้น

หากต้องการเพิ่มหลักสูตรอื่นนอกเหนือจากสาขาวิชาคอมพิวเตอร์ ควรกลับไปทำในขั้นตอน Data preparation โดยการสร้างตัวต้นแบบที่หลากหลาย (Multi Model) ที่เหมาะกับแต่ละหลักสูตรหรืออาจเพิ่มหลักสูตรอื่นทั้งมหาวิทยาลัย เพื่อนำข้อมูลไปสร้างตัวต้นแบบ (Train Data)ได้เมื่อนำไปใช้กับระบบ แล้วให้ระบบคัดกรองผลลัพธ์ (filter)

ภาคพนวก ก

การนำเสนอผลงานวิชาการระดับนานาชาติ

DECISION SUPPORT SYSTEM FOR SUBJECTS REGISTRATION OF COMPUTER SCIENCE STUDENT

Naratsaporn Heamra
Digital Technology Innovation Program
Faculty of Science, Maejo University
Chiang Mai, Thailand
mju6004308012@mju.ac.th

Part Pramokchon
Digital Technology Innovation Program
Faculty of Science, Maejo University
Chiang Mai, Thailand
part@gmaejo.mju.ac.th

***Abstract*—Nowadays, the concept and theory to support the educational system have been continuously developed. One of the major problems of the educational system that has to be solved is that the grade point average (GPA) of student are reducing when studying at higher levels. This problem arises from students choosing to study subjects that are not consistent with their learning ability. This paper presents the process of analyzing and solving the problem of gradual decline. Based on the statistical technique, the method presented in this article can recommend students that should register any subject that will likely get good grades performance. The results of performance investigation show that the proposed method can accurately recommend student choosing subjects that will achieve good or higher grades with 86.67% accuracy.**

Keywords—educational progression, student support, data mining, course registration suggestion, learning performance

I. INTRODUCTION

At present, the information is increasing more and more based on information technology developed especially in educational institutions. Analysis of educational data is very popular nowadays. The people are known as knowledge discovery from very large databases or data mining. It is very popular nowadays. It has been used extensively in various forms, particularly in relation to education. Such factors that affect student learning level, find the most appropriate department for students. This paper presents techniques using statistical models. The objective is to solve the problem of lower grades students. Because the results grades are important, it is an opportunity to graduate. Scholarship and choose a future job.

The curriculum standard for computer science student requires that each student must register at least 40 subjects. Information of computer science students between 2012 – 2013 it was found that the most registered period was the second to third-year students. In addition, it was found that 35% of students with GPA lower. The result of this search, this knowledge will allow students to choose subjects that make their academic performance is good quality.

Data mining has been applied in many fields, such as business education, medical. In this paper, we present the data mining techniques used in the development of education.

This paper presents an effort to solve the problem of having a reduced grade of students. By introducing the method for selecting the subjects that students are expected to achieve a good quality academic result. The good quality favorably means grade is B or up by measuring the level of basic education courses in Thailand.

This paper is organized as followers. Section 2 we present the literature review. In Section 3 we present the method and experiment. Conclusions and future work are presented in Section 4.

II. LITERATURE

A. Education Development

Past research has used data mining techniques to analyze information about the school. Such as the analysis of relationship student factors affecting the level of academic [1]. This paper presents Decision Analysis Techniques with Decision Tree Techniques as one technique for data mining. This objective relies on the relationship of the student to indicate the grade of the student.

The association rule mining is used to predict Student Learning Outcome [2]. The system can suggest students adjust the method of learning the remaining subjects. The results show that the system increases the overall average outcome of students. This system supports students to be able to completely graduate their courses.

Applying a data mining technique can also help students in selecting their majors [3]. The objective is students can choose based on their skills. This paper is using data mining techniques using K-fold cross-validation and find out an important factor influencing the decision of students choose main study program.

Use data mining techniques to analyze pre-test results and skills to increase student potential[4]. This paper applies K-mean clustering algorithm for grouping students into two groups, students learning bad or good. The first objective is to provide information to faculty, advisors or faculty members to take advantage of student care. The second objective is to reduce the risk of students who fail to success the educational criteria.

B. CRISP-DM Framework

CRISP-DM is like a blueprint that is widely used. Similar to the ISO process in industrial plants CMMI is a standard process in software development. Standard process for data analysis in data mining Developed in 1996, the collaboration of 3 companies is DaimlerChrysler SPSS and NCR This process is called "Cross-Industry Standard Process for Data

978-1-5386-8072-8/19/$31.00 ©2019 IEEE

Mining" or "CRISP-DM"[5]. This process consists of 6 steps as shown in Figure 1.

At present, research has become more popular with CRISP-DM in data mining, especially in education. Such as the application of data mining technology to predict student success and failure[6]. Along with offering the use of CRISP-DM with data sets. The objective is to separate the hidden form from student information. These patterns can be seen in relation to the variables of the students. After that, the model was created and tested using the sample data set of undergraduate students.

Predictive analysis of the academic performance of student [7] selected CRISP-DM process for the design of data mining. The objective is to predict student data at the end of one year of study. In this process, descriptive statistics are analyzed to obtain in-depth information. Then compare the results to predict the class with other models.

Analyzing college graduation rates in higher education institutions has become the main focus in measuring institutional performance [8]. This article a hybrid data envelopment analysis (DEA) combined with the Cross-Industry Standard Process for Data (CRISP-DM).The objectives to evaluate the effectiveness of undergraduate college students. These analyzes provide useful information and policy support for university administrators.

III. EXPERIMENT

From student data from computer science graduated during the academic year 2012 – 2013. We divide the operation method into 6 parts according to the CRISP-DM Framework as shown in Figure 1.

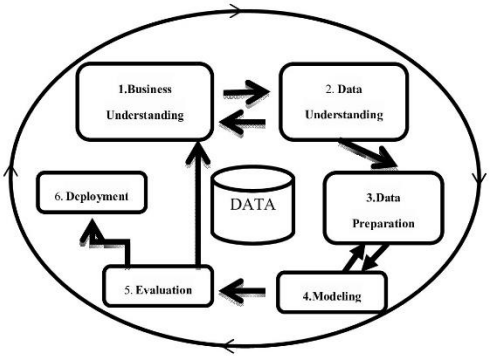


Fig. 1. CRISP-DM Framework

A. Business Understanding

The first step in the CRISP-DM process is to understand the problem. This paper means to solve the grades lower. We discuss the application of data mining and education data. The objective is to help guide the subjects that should be registered because it will result in good grades results. The solution reduced academic performance by using data mining can be done in various methods. This paper presents statistical modeling techniques.

We propose performance measurement. We have divided the data into two parts to measure the efficiency of the system. The year 2013 data is analyzed and the 2012 data is compared to measure the performance of the statistical model.

B. Data Understanding

The data used in the analysis is the registration students of 12,000 row. Information is computer education information. We choose from only the information of the graduate students. The information consists of student code, major, subject name, subject code, grade as shown in Table I. This step starts from collecting data. After that, it will check the data that has been collected to see the accuracy of the data. And consider whether to use all the data or need to select some data for analysis.

TABLE I. DISPLAY ALL STUDENT REGISTRATION INFORMATION.

Stu_code	Year of study	Major	Subjects	Sub_Code	grade
00001	2012	Computer science	Computer Science 1	CS 211	B
			Introduction to Object-Oriented Programming	CS 217	C+
00002	2013	Computer science	Principles of Thai language	TL 211	D+
...	...	…	…	...	...

C. Data Preparation

This step is a step to convert the data that has been collected. The objective is to become the next step in the analysis. This data conversion must be done correctly by deleting and edit information. Such as converting data to the same range or missing information. This step is the most a long time step of the CRISP-DM process. Table II is an example of the data that has already been converted. The first variable is stu_code, is the student code. The next variable is the subject code. The objective is to collect data on grades.

TABLE II. DISPLAY ALL STUDENT REGISTRATION INFORMATION.

number	Stu_code	CS 321	CS 200	JP 313	JP 315	...
1	00001	A	B+	B	C	...
2	00002	B	B+	C	C+	...
...	…	…	…	…	…	...

The data in Table III are classified into two sections. The first part is Subjects in the field, meaning subjects that are in the group that compels students to study. Most of them are subjects that are directly related to computer science. The second part is Subjects outside the field. Most subjects require students to register from other disciplines. The next step after dividing the course and separate the data is the 2nd and 3rd academic year for easy analysis.

TABLE III DISPLAY ALL STUDENT REGISTRATION INFORMATION.

Year of study 2012	
Number	**Stu_code**
1	00001
2	00002
…	…

Subjects in the field					
2nd academic year			3rd academic year		
CS 321	**GCH 321**	**…**	**CS 200**	**GD 200**	**…**
A	C+	…	B+	C+	…
…	…	…	..	…	..

Subjects outside the field					
2nd academic year			3rd academic year		
CS 321	**GCH 321**	**…**	**CS216**	**JP400**	**…**
B	B+	…	C+	D+	…
…	….	…	…	…	…

D. Modeling

This step will be the process of analyzing data with techniques. Sometimes, it may be necessary to go back to the Data Preparation step to convert the data appropriately. From the information mentioned in the introduction .It was found that 35% of the students in the grades decreased during the 2nd and 3rd academic year because this period from the survey results was the most registered period. For this reason we use this information to analyze the cause of the problem. Figure 2 is an analysis of the data from Table 3, by means of a numerical grade A = 8, B + = 7, B = 6, or F = 1. Next step, then find the percentage of each group of subjects. Found that 169 computer science students percentage of students in Subjects in the field reduced to 119 up to 50. And percentage of students in Subjects outside the field reduced to 149 up to 50. We therefore choose to introduce the subjects registered in the Subjects outside the field because there is a result that results in a reduced in student results.

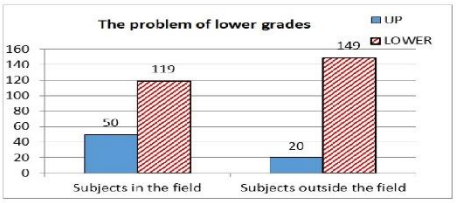


Fig. 2. Compare subjects in and outside of the field.

Figure 3 is an image histogram to study Subjects outside the field. The X axis is the number of students enrolled. The y axis is the number of subjects that students register. Because there are some subjects that are taught less and therefore can't be analyzed in this section. This graph shows data from our study is registered at least 10 of the 23 subjects. For this reason, we therefore take subjects that is in this frequency range discarded.

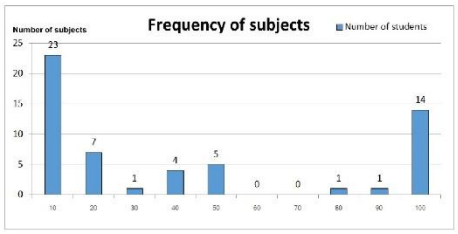


Fig. 3. Graph showing the frequency of study selection subjects outside the field

Figure 4 is a graph showing the percentage of subjects after removing subjects that haven't been analyzed. The data from the eliminated consisted of subjects with a student enrollment of less than 10, and some of which are registered in the year 2012, but not registered in 2013. The results of the study subjects 56 reduced to just 36.

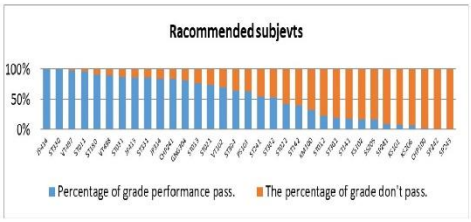


Fig. 4. Graph showing percentage of study selection subjects outside the field ascending.

Figure 5 is a survey of students studying in computer science from a total of 94 data. Topics that if the system supports the introduction of any one subject. The topic of the question is "How much will choose to believe that percentages are recommended?". The results show that most students choose to believe that the recommended percentage is 70 percent. We therefore used this result as a criterion for cutting subjects with a percentage of students getting good grades below 70 percent leaving as shown in Figure 6.

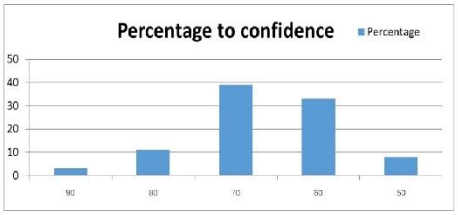


Fig. 5. Survey of students to choose percentage to confidence.

E. Evaluation

Figure 6 uses the model created from the year 2012 data to introduce the course to the year 2013 data. Figure 7 select the specific information of the subjects that are registered as well to compare. Found that if students choose 15 subjects as recommended by the system. There are 13 subjects that most

students study on a good basis. That is system can recommend the correct subjects 86.67%

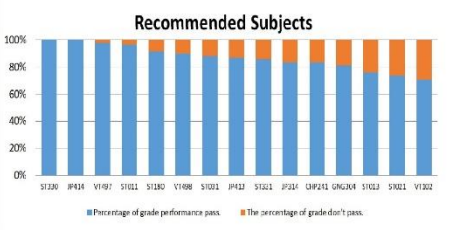


Fig. 6. Recommended subjects compared to the year 2013 information

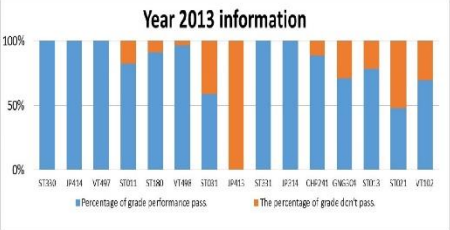


Fig. 7. Recommended subjects compared to the year 2013 information

F. Deployment

The results of analyzes of computer science students. We can recommend subjects to new students through the website. Students can choose to study subjects that will enable students to achieve good grade results (grade B up) with the subject guidance system 86.67% accurate.

IV. Conclusion

This research applies data mining techniques to solve the problem of lowering the academic performance of computer science students. The proposed method can recommend subjects that students can get a good or higher academic grade. By using prior data to analyze and create decision support systems for subject registration in the later years. The results of this recommendation are quite satisfactory. Statistic based method can select some subjects that students can study well with 86.67% accuracy. However, this paper does not mention the analysis of various factors that may affect the performance of the model, such as gender, age, etc. Thus, the proposed method still has to be improved in order to develop the suggestion system to be the most effective. The performance compared with relevant research is also investigated in the future.

Refesrences

[1] Osmanbegović, Edin & Suljic, Mirza. (2012). DATA MINING APPROACH FOR PREDICTING STUDENT PERFORMANCE. Journal of Economics & Business/Economic Review. 10. 3-12.

[2] Al-Hatem A.I., Masood M., Al-Samarraie H.Fostering student nurses' self-regulated learning with the Second Life environment: An empirical studyJournal of Information Technology Education: Research, Volume 17, 2018

[3] Sugiyarti, Eka & Jasmi, Kamarul Azmi & Basiron, Bushrah & Huda, Miftachul & Maseleno, Andino. (2018). Decision Support System of Scholarship Grantee Selection using Data Mining. International Journal of Pure and Applied Mathematics. 119.

[4] Bansal, Arpit & Sharma, Mayur & Goel, Shalini. (2017). Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining. International Journal of Computer Applications. 157. 35-40. 10.5120/ijca2017912719.

[5] Leventhal, B. (2010). An introduction to data mining and other techniques for advanced analytics. Journal of Direct, Data and Digital Marketing Practice, 12(2), 137-153, doi:10.1057/dddmp. 2010.35.

[6] Behzad Nakhkobl, M. K. (2016). "Predicted Increase Enrollment in Higher Education Using Neural Networks and Data Mining Techniques." Journal of Advances in Computer Research

[7] Asif, R., Merceron, A., Ali, S., & Haider, N. (2017). Analyzing undergraduate students' performance using educational data mining. Computers & Education, 113(Supplement C), 177–194.

[8] Chen, Ya & Chen, Yao & Oztekin, Asil. (2016). A hybrid data envelopment analysis approach to analyse college graduation rate at higher education institutions. INFOR: Information Systems and Operational Research. 55. 1-23. 10.1080/03155986.2016.1262584.

ภาคพนวก ข

ข้อมูลวิชาที่ปรากฏในงานวิจัย

รายละเอียดรายวิชาที่ปรากฏในงานวิจัย

ตารางที่ 26 แสดงรายละเอียดวิชาที่ปรากฏในงานวิจัย

วิชาที่	รหัสวิชา	รหัสวิชาที่ใช้	ชื่อวิชา	สังกัด / คณะ	หน่วยกิต
1	จป414	JP414	การประกวดและการตัดสินสัตว์น้ำ	เทคโนโลยีการประมงและทรัพยากรทางน้ำ , การประมง	3 (2-3-5)
2	ศท330	ST330	ภาษาถิ่นล้านนา	ภาษาไทยและภาษาตะวันออก , ศิลปะศาสตร์	3 (3-0-0)
3	วท497	WT497	สหกิจศึกษา	วิทยาศาสตร์, ส่วนกลาง (คณะวิทย์)	9 (0-270-0)
4	ศท011	ST011	มนุษย์กับความงามทางศิลปะ	การพัฒนาสุขภาพชุมชน, ศิลปะศาสตร์	3 (3-0-6)
5	ศท180	ST180	ศิลปะกับความคิดสร้างสรรค์	การพัฒนาสุขภาพชุมชน, ศิลปะศาสตร์	3 (1-4-4)
6	วท498	WT498	การเรียนรู้อิสระ	วิทยาศาสตร์, ส่วนกลาง (คณะวิทย์)	9 (0-270-0)
7	ศท031	ST031	การใช้ภาษาไทย	ภาษาไทยและภาษาตะวันออก , ศิลปศาสตร์	3 (1-4-4)
8	จป413	JP413	การประกวดและตัดสินสัตว์น้ำ	เทคโนโลยีการประมงและทรัพยากรทางน้ำ ,	3 (3-0-6)

				การประมง	
9	ศท331	ST331	คติชนวิทยา	ภาษาไทยและภาษาตะวันออก , ศิลปศาสตร์	3 (3-0-0)
10	จป314	JP314	การอนุรักษ์ทรัพยากรประมง	เทคโนโลยีการประมงและทรัพยากรทางน้ำ , การประมง	3 (3-0-6)
11	ชป241	CHP241	ทรัพยากรสัตว์น้ำและความหลากหลายทางชีวภาพ	เทคโนโลยีการประมงและทรัพยากรทางน้ำ , การประมง	3 (3-0-6)
12	กง304	KNG304	การวิเคราะห์รายงานทางการเงิน	การเงิน ,บริหารธุรกิจ	3 (2-2-5)
13	ศท013	ST013	สุขภาพเพื่อการดำรงชีวิต	การพัฒนาสุขภาพชุมชน , ศิลปะศาสตร์	3 (1-4-4)
14	ศท021	ST021	สังคมศาสตร์ในชีวิตประจำวัน	กลุ่มวิชาสหวิทยาการสังคมศาสตร์ ,ศิลปะศาสตร์	3 (3-0-6)
15	วท102	WT102	การพัฒนาวิทยาศาสตร์และเทคโนโลยี	วิทยาศาสตร์, ส่วนกลาง (คณะวิทย์)	3 (2-2-5)
16	ศท304	ST304	ศาสตร์และศิลป์แห่งปัญญาชน	กลุ่มวิชาสหวิทยาการสังคมศาสตร์ , ศิลปะศาสตร์	3 (2-2-5)
17	ผษ101	PS101	เกษตรเพื่อชีวิต	ผลิตกรรมการเกษตร, ส่วนกลาง(คณะผลิตฯ)	3 (3-0-6)
18	ศท241	ST241	ภาษาอังกฤษเชิงวิทยาศาสตร์และเทคโนโลยี	ภาษาตะวันตก , ศิลปะศาสตร์	3 (2-2-5)
19	ศท302	ST302	สังคมและวัฒนธรรมไทย	กลุ่มวิชาสหวิทยาการสังคมศาสตร์ , ศิลปะศาสตร์	3 (3-0-6)

20	ศท022	ST022	อารยธรรมโลก	กลุ่มวิชาสหวิทยาการ สังคมศาสตร์ , ศิลปะศาสตร์	3 (3-0-6)
21	ศท142	ST142	ภาษาอังกฤษพื้นฐาน 2	ภาษาตะวันตก , ศิลปะศาสตร์	3 (2-2-5)
22	คม100	KM100	เคมีทั่วไป	เคมี ,วิทยาศาสตร์	3 (2-3-5)
23	ศท012	ST012	จิตวิทยากับพฤติกรรมมนุษย์	กลุ่มวิชาสหวิทยาการ สังคมศาสตร์ , ศิลปศาสตร์	3 (3-0-6)
24	สต301	SD301	หลักสถิติ	สถิติ,วิทยาศาสตร์	3 (3-0-6)
25	ศท141	ST141	ภาษาอังกฤษพื้นฐาน 1	ภาษาตะวันตก , ศิลปะศาสตร์	3 (2-2-5)
26	คศ102	KS(M)102	แคลคูลัส 2	คณิตศาสตร์,วิทยาศาสตร์	3 (3-0-6)
27	คศ205	KS(M)205	ตรรกศาสตร์เชิงคณิตศาสตร์เบื้องต้น	ตรรกศาสตร์เชิงคณิตศาสตร์เบื้องต้น	3 (3-0-6)
28	ศป241	SP241	การพัฒนาทักษะภาษาอังกฤษ 1	ภาษาตะวันตก , ศิลปะศาสตร์	1 (0-2-1)
29	คศ101	KS(M)101	แคลคูลัส 1	คณิตศาสตร์,วิทยาศาสตร์	3 (3-0-6)
30	คศ206	KS(M)206	พีชคณิตเชิงเส้น	คณิตศาสตร์,วิทยาศาสตร์	3 (3-0-6)
31	ชว100	CHW100	ชีววิทยาทั่วไป	เทคโนโลยีชีวภาพ ,วิทยาศาสตร์	3 (2-3-5)
32	ศป242	SP242	การพัฒนาทักษะภาษาอังกฤษ 2	ภาษาตะวันตก , ศิลปะศาสตร์	1 (0-2-1)
33	ศป243	SP243	การพัฒนาทักษะภาษาอังกฤษ 3	ภาษาตะวันตก , ศิลปะศาสตร์	1 (0-2-1)

บรรณานุกรม

Al-Hatem, Masood และAl-Samarraie. 2561. Self-Regulated Learning with the Second Life Environment An Empirial Study. Journal of Information Technology Education: Research 2018,Volume 17

Asif, Raheela, Merceron, Agethe, Ali, Syed Abbas และHaider, Najmu Ghani. 2560. Analyzing undergraduate students performance using educational data mining 2017. Commputers & Education An International Journal.

Bansal, Arpit, Sharma, Mayur และGoel, Shalini. 2560. Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining International Journal of Computer Applications January 2017 157 – No 6(

Chen, Ya, Chen, Yao และOztekin, Asil. 2559. A hybrid data envelopment analysis approach to analyse college graduation rate at higher education institutions Information Systems and Operational Research 2016.

Leventhal, Barry. 2553. An introduction to data mining and other techniques for advanced analytics 2010. Journal of Direct, Data and Digital Marketing Practice.

Nakhkob, Behzad และKhademi, Maryam. 2558. Predicted Increase Enrollment in Higher Education Using Neural Networks and Data Mining Techniques Journal of Advances in Computer Research Quarterly pISSN 2015.

Osmanbegović, Edin และSuljić, Mirza. 2555. Data mining approach for predicting student performance. Journal of Economics and Business, Vol. X, Issue 1, May 2012.

Sugiyarti, Eka, Jasmi, Kamarul Azmi, Basiron, Bushrah, Huda, Miftachul, K, Shankar และ Maseleno, Andino. 2561. Decision Support System of Scholarship Grantee Selection using Data Mining. International Journal of Pure and Applied Mathematics 2018,119(

เปรมปรีดิ์, ปรีดีชนก. 2560. Weka (Waikato Environment for Knowledge Analysis). [ระบบออนไลน์]. แหล่งที่มา http://preedeechanok.blogspot.com/2017/12/weka-waikato-environment-for-knowledge.html

โพธิ์ผลิ, วรรณนภา. 2552. การใช้เทคนิคเหมืองข้อมูลเพื่อการวิเคราห์ความสัมพันธ์ของรายวิชา. มหาวิทยาลัยศิลปากร.

คงทิม, ณรงคศ์กักดิ์ และภูบุญอบ, จิรัฏฐา. 2552. การประยุกต์ใช้เอฟพี-กโรธกับงานแนะแนวการศึกษาต่อในระดับอุดมศึกษา CIT2011 & UniNOMS2011.

คูณทวี, จารุภรณ์, แก้วตา, รสรินทร์, ใจเที่ยง, ณรงค์ และสุทธธง, ณกมล. 2563 ปัจจัยที่มีความสัมพันธ์กับพฤติกรรมการบริโภคผลิตภัณฑ์เสริมอาหาร ของนิสิต มหาวิทยาลัยแห่งหนึ่งในภาคเหนือ Public Health Policy & Laws Journal Vol. 6 No.1 January - June 2020

ชัยวุฒิศักดิ์, พรพิมล และกล่อมวิเศษ, ยุวดี. 2562. การพัฒนากฎการทำนายผลการเรียนของนักศึกษาชั้นปีที่ 1 โดยใช้เทคนิคการทำเหมืองข้อมูล น. ใน วารสารวิจัยรามคำแหง (วิทยาศาสตร์และเทคโนโลยี) ปีที่ 22 ฉบับที่ 2

ชื่นมัจฉา, ณัฏญาพร. 2559. การสร้างแบบจำลองกฎความสัมพันธ์สำหรับฐานข้อมูลการสั่งซื้อสินค้าโดยใช้เทคนิค เอฟพี-กโรธ. มหาวิทยาลัยศรีปทุม.

ทองคำ, จารี. 2561. การเปรียบเทียบประสิทธิภาพของเทคนิค Apriori และ FP-Growth ในการสร้างกฎความสัมพันธ์ของโรคมะเร็งต่อมลูกหมาก JOURNAL Of APPLIED INFORMATICS AND TECHNOLOGY 1(2) 2018

ปฤชานนท์, ปฏิพัทธ์ และศรีอุไร, วงกต. 2561. การประยุกต์ใช้กฎความสัมพันธ์เพื่อวิเคราะห์ความเสี่ยงการออกกลางคัน ของนักศึกษาสาขาเทคโนโลยีสารสนเทศ Journal of Science and Science Education Vol. 1 No. 2

พลอาสา, นิสานันท์. 2558. การสร้างแบบจำลองการขายผลิตภัณฑ์และพยากรณ์ยอดขายประกันชีวิตโดยเทคนิคกการทำเหมืองข้อมูล กรณีศึกษาบริษัทประกันชีวิตแห่งหนึ่ง. มหาวิทยาลัยธรรมศาสตร์.

มหัทธนชัย, บุษราภรณ์, มาลัยวงศ์, ครรชิต, สมหอม, เสมอแข และตันตรานนท์, ณัฐิยา. 2559. กฎความสัมพันธ์ของรายวิชาที่มีผลต่อการพ้นสภาพนักศึกษาโดยใช้อัลกอริทึมอพริโอริ. น. ในการประชุมวิชาการระดับชาติ มหาวิทยาลัยราชภัฏกำแพงเพชร ครั้งที่ 3.

รักผกาวงศ์, พงษ์เทพ และรักผกาวงศ์, อุไรวรรณ์. 2558. การใช้กฎความสัมพันธ์ ร่วมกับฟัซซี่กฎความสัมพันธ์ เพื่อคาดการณ์ผลการเรียนของนักศึกษา น. ใน วารสารวิทยากศาสตร์ มข. ปีที่ 43 ฉบับที่ 3.

ส่งศิริ, ชิดชนก, รักธรรมานนท์, ธนาวินท์ และไวยมัย, กฤษณะ. 2547. การใช้เทคนคิ Data Mining เพื่อค้นหาภาควิชาที่เหมาะสมที่สุดให้กับนิสิต น. ใน การประชุมทางวิชาการของมหาวิทยาลัยเกษตรศาสตร์ ครั้งที่ 39.

สุวรรณโณ, ณัฐธิดา และสิงห์เอี่ยม, อันธิกา. 2554. การหาปัจจัยที่ส่งผลต่อความเสี่ยงของนักศึกษา

เรียนอ่อน ด้วยเทคนิคกฎความสัมพันธ์ กรณีศึกษา : มหาวิทยาลัยสงขลานครินทร์ น. ใน วารสารวิทยาการจัดการ ปีที่ 28.

มหาวิทยาลัยแม่โจ้
MAEJO UNIVERSITY

ประวัติผู้วิจัย

ชื่อ-สกุล	ว่าที่ ร.ต. ณรัฐศภรณ์ เหมรา
เกิดเมื่อ	07/06/2538
ประวัติการศึกษา	ปริญญาตรี สาขาวิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยแม่โจ้
ประวัติการทำงาน	ปัจจุบัน กลุ่มบริษัท : Benchachinda Group บริษัท : Brainergy Company Limited ฝ่าย : Software Optimeze Division ตำแหน่ง : Full Stack Developer