การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา

นายเสกสรรค์ วิลัยลักษณ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
แขนงวิชาเทคโนโลยีสารสนเทศและการสื่อสาร สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช
พ.ศ. 2558

Utilizing Data Mining Techniques to Forecast Student Academic Achievement of Kasetsart University Laboratory School Kamphaeng Saen Campus Educational Research and Development Center

Mr. Seksan Vilailuck

A Thesis Submitted in Partial Fulfillment of the Requirements for

the Degree of Master of Science in Information and Communication Technology

School of Science and Technology

Sukhothai Thammathirat Open University

2015

หัวข้อวิทยานิพนธ์ การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา

ชื่อและนามสกุล นายเสกสรรค์ วิลัยลักษณ์

แขนงวิชา เทคโนโลยีสารสนเทศและการสื่อสาร

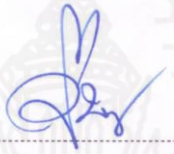
สาขาวิชา วิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช

อาจารย์ที่ปรึกษา 1. รองศาสตราจารย์ ดร. วิภา เจริญภัณฑารักษ์

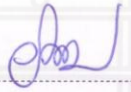
2. อาจารย์ ดร. ดวงดาว วิชาดากุล

วิทยานิพนธ์นี้ ได้รับความเห็นชอบให้เป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรระดับปริญญาโท เมื่อวันที่ 5 พฤศจิกายน 2558

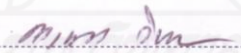
คณะกรรมการสอบวิทยานิพนธ์



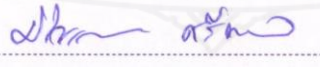
..... ประธานกรรมการ
(อาจารย์วิทยา พรพัชรพงศ์)



..... กรรมการ
(รองศาสตราจารย์ ดร. วิภา เจริญภัณฑารักษ์)



..... กรรมการ
(อาจารย์ ดร. ดวงดาว วิชาดากุล)



..... ประธานกรรมการบัณฑิตศึกษา
(ศาสตราจารย์ ดร. สิริวรรณ ศรีพหล)

ชื่อวิทยานิพนธ์ การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน
โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน
ศูนย์วิจัยและพัฒนาการศึกษา
ผู้วิจัย นายเสกสรรค์ วิลัยลักษณ์ **รหัสนักศึกษา** 2569600162
ปริญญา วิทยาศาสตรมหาบัณฑิต (เทคโนโลยีสารสนเทศและการสื่อสาร)
อาจารย์ที่ปรึกษา (1) รองศาสตราจารย์ ดร. วิภา เจริญภัณฑารักษ์ (2) อาจารย์ ดร. ดวงดาว วิชาดากุล
ปีการศึกษา 2558

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาคลังข้อมูล และสร้างตัวแบบพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา โดยใช้ข้อมูลนักเรียนระดับมัธยมศึกษาระหว่างปีการศึกษา 2548 – 2556 เพื่อพัฒนาคลังข้อมูลโดยใช้โครงสร้างแบบสโนว์เฟลก สกีมาและนำเสนอรายงาน จากนั้นใช้ข้อมูลนักเรียนระดับมัธยมศึกษาปีที่ 4 ระหว่าง ปีการศึกษา 2553 – 2556 จำนวน 525 ระเบียน ประกอบด้วย 16 คุณลักษณะ มาสร้างตัวแบบพยากรณ์ผลการเรียนโดย ใช้ชุดข้อมูล 2 แบบ คือ ข้อมูลแบบไม่จัดกลุ่ม (Original Data) และข้อมูลแบบจัดกลุ่ม (Data Cluster) จากนั้นนำไปผ่านกระบวนการคัดเลือกคุณลักษณะ (Feature Selection) ซึ่งใช้วิธี Correlation-based Feature Selection (CFS) และวิธี Information Gain (IG) แล้วใช้เทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ซัพพอร์ตเวกเตอร์แมชชีน (SVM) และต้นไม้ตัดสินใจ (Decision Tree) มาสร้างตัวแบบพยากรณ์และเปรียบเทียบตัวแบบ ด้วยการทดสอบประสิทธิภาพแบบ 10-fold Cross Validation

ผลการวิจัยพบว่า คลังข้อมูลนักเรียนระดับมัธยมศึกษาที่ใช้งานโดย ผู้บริหาร หัวหน้ากลุ่มสาระการเรียนรู้ และอาจารย์ประจำชั้นมีความพึงพอใจการใช้งานคลังข้อมูลอยู่ในระดับดี และในการทำเหมืองข้อมูลสร้างตัวแบบพยากรณ์ พบว่า ชุดข้อมูลแบบไม่จัดกลุ่มนำมาคัดเลือกคุณลักษณะด้วยวิธี Correlation-based Feature Selection (CFS) ร่วมกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน ให้ค่าความถูกต้องสูงที่สุดที่ร้อยละ 94.48 และมีค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดที่ 0.1880 เหมาะสมสำหรับการสร้างระบบพยากรณ์ผลการเรียนเฉลี่ยของนักเรียน ซึ่งนักเรียนสามารถนำผลการพยากรณ์ไปประกอบการตัดสินใจและวางแผนการเรียน

คำสำคัญ: การพยากรณ์ผลการเรียน การทำเหมืองข้อมูล การคัดเลือกคุณลักษณะ

Thesis title: Utilizing Data Mining Techniques to Forecast Student Academic Achievement of Kasetsart University Laboratory School Kamphaeng Saen Campus Educational Research and Development Center.
Researcher: Mr. Seksan Vilailuck; **ID:** 2569600162;
Degree: Master of Science (Information and Communication Technology);
Thesis advisors: (1) Dr. Vipa Jaroenpuntaruk, Associate Professor;
(2) Dr. Duangdao Wichadakul; **Academic year:** 2015

Abstract

The purposes of this research were to develop the data warehouse and to construct the "Prediction Model" of high school students achievement of Kasetsart University Laboratory school, Kamphaeng Saen Campus Educational Research and Development Center, from the data during academic year 2005-2013 to develop the data warehouse using the structure of snowflake schema and presentations, and then use the Mathayomsuksa 4 students' data during academic year 2010-2013 consisted of 525 records including 16 features of attributes to construct the "Prediction Model" with 2 sets of data types : original data and data cluster. Then the features were selected by Correlation – based Feature Selection (CFS) and Information Gain (IG). After that the "Prediction Model" was constructed by data mining techniques e.g. Neural Networks, Multi-layer Perceptron (MLP), Support Vector Machine (SVM) and Decision Tree Model. Finally, the model was compared by the Performance Test with a 10-fold Cross Validation.

The results showed that the administrators, the department leaders, and the advisors were satisfied with the high school students' data warehouse at a good level and, by using data mining techniques, the set of original data, for feature selection with correlation based feature selection (CFS) combining with a Multi-layer Perceptron Concept, provided the highest accuracy at 94.48 percent and the second root of the discrepancy was the least at 0.1880 which was suitable to construct the students prediction system.

Keyword: Forecast Academic Achievement, Data mining, Feature Selection

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความกรุณาและความอนุเคราะห์จากบุคคลผู้ทรงคุณวุฒิหลายท่าน โดยเฉพาะอย่างยิ่งท่านอาจารย์ รองศาสตราจารย์ ดร.วิภา เจริญภัณฑารักษ์ อาจารย์ที่ปรึกษาหลัก และท่านอาจารย์ ดร.ดวงดาว วิชาดากุล อาจารย์ที่ปรึกษาร่วมวิทยานิพนธ์ฉบับนี้ ซึ่งได้กรุณาเสียสละเวลาในการให้คำปรึกษา ให้ข้อเสนอแนะต่างๆ ในการจัดทำวิทยานิพนธ์ตลอดจนแนะนำแนวทางในการศึกษาค้นคว้างานวิจัยและข้อมูลที่นำมาปรับปรุงแก้ไขข้อบกพร่องต่างๆ จนกระทั่งวิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงด้วยดี ผู้วิจัยขอขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอขอบคุณโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา ที่ได้ให้โอกาสในการศึกษาต่อ และขอขอบคุณท่านผู้ช่วยศาสตราจารย์ ณชพงศ์ อุดมศรี ผู้อำนวยการโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา ที่ให้ความอนุเคราะห์ข้อมูลของโรงเรียนในการวิเคราะห์และให้คำแนะนำในการทำวิทยานิพนธ์ฉบับนี้ รวมทั้งขอขอบคุณเพื่อนร่วมงานทุกท่านที่ช่วยให้คำแนะนำ และความช่วยเหลือต่างๆ รวมทั้งผู้ที่มีส่วนร่วมทุกท่านที่มิได้เอ่ยนามไว้ ณ ที่นี้

นอกจากนี้ผู้วิจัยขอขอบพระคุณคณาจารย์ ประจำสาขาวิชาวิทยาสาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช เพื่อนนักศึกษา ตลอดจนผู้ที่มีส่วนเกี่ยวข้องในการทำวิทยานิพนธ์ครั้งนี้ที่ได้กรุณาให้การสนับสนุน ช่วยเหลือ คำแนะนำ และให้กำลังใจตลอดมาจนกระทั่งการทำวิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ด้วยดี สุดท้ายนี้ผู้วิจัยหวังเป็นอย่างยิ่งว่าวิทยานิพนธ์ฉบับนี้จะเป็นประโยชน์ต่อหน่วยงานและสำหรับผู้ที่สนใจ หากมีข้อบกพร่องหรือผิดพลาดประการใด ผู้วิจัยขออภัยเป็นอย่างสูงในข้อบกพร่องและความผิดพลาดนั้นไว้ ณ ที่นี้

เสกสรรค์ วิลัยลักษณ์

พฤศจิกายน 2558

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญตาราง ..... ฌ
สารบัญภาพ ..... ฎ
บทที่ 1 บทนำ 1
ความเป็นมาและความสำคัญของปัญหา 1
วัตถุประสงค์ของการวิจัย 5
สมมุติฐานการวิจัย 6
กรอบแนวคิดการวิจัย 6
ขอบเขตของการวิจัย 7
ประโยชน์ที่คาดว่าจะได้รับ 7
นิยามคำศัพท์ 7
บทที่ 2 วรรณกรรมที่เกี่ยวข้อง 9
โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน
ศูนย์วิจัยและพัฒนาการศึกษา 9
คลังข้อมูล (Data Warehouse 14
การทำเหมืองข้อมูล (Data Mining) 24
การคัดเลือกคุณลักษณะ 38
โปรแกรม WEKA (Waikato Environment for Knowledge Analysis) 41
งานวิจัยที่เกี่ยวข้อง 42
บทที่ 3 วิธีดำเนินการวิจัย 48
ประชากรและกลุ่มตัวอย่าง 48
เครื่องมือที่ใช้ในการวิจัย 48
ขั้นตอนการดำเนินงาน 50

สารบัญ (ต่อ)

หน้า

บทที่ 4 ผลการวิเคราะห์ข้อมูล 55
การพัฒนาคลังข้อมูลผลการเรียน 55
การพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ 66
การทำเหมืองข้อมูล 72
การประเมินผล 103
บทที่ 5 สรุปผล อภิปราย และข้อเสนอแนะ 107
สรุปผลการวิจัย 107
การอภิปรายผล 109
ปัญหาและอุปสรรค 111
ข้อเสนอแนะ 112
บรรณานุกรม 113
ภาคผนวก 118
ก แบบประเมินความพึงพอใจระบบคลังข้อมูลผลการเรียน 119
ข แบบประเมินความเชื่อมั่นระบบพยากรณ์ผลการเรียน 122
ค คู่มือการติดตั้งโปรแกรมระบบ 125
ง คู่มือการใช้งานระบบพยากรณ์ผลการเรียนเฉลี่ย 140
ประวัติผู้วิจัย 144

สารบัญตาราง

หน้า

ตารางที่ 2.1 แสดงระดับผลการเรียน 12
ตารางที่ 4.1 ค่าที่ต้องการวัดและมิติของผลการเรียน 57
ตารางที่ 4.2 ตาราง Fact_Grade สำหรับเก็บข้อมูลคะแนนและผลการเรียน 58
ตารางที่ 4.3 ตาราง Fact_GPA สำหรับเก็บข้อมูลผลการเรียนเฉลี่ยและ ผลการเรียนเฉลี่ยสะสม 59
ตารางที่ 4.4 ตาราง Dim_Subject สำหรับเก็บข้อมูลวิชาเรียน 59
ตารางที่ 4.5 ตาราง Dim_Student สำหรับเก็บข้อมูลนักเรียน 60
ตารางที่ 4.6 ตาราง Dim_Semester สำหรับเก็บข้อมูลปีการศึกษาและภาคเรียน 60
ตารางที่ 4.7 ตาราง Dim_Teacher สำหรับเก็บข้อมูลอาจารย์ผู้สอน 61
ตารางที่ 4.8 ตาราง Dim_Group สำหรับเก็บข้อมูลประเภทของวิชาเรียน 61
ตารางที่ 4.9 ตาราง Dim_Gender สำหรับเก็บข้อมูลเพศ 61
ตารางที่ 4.10 ตาราง Dim_Area สำหรับเก็บข้อมูลรายชื่อจังหวัด 61
ตารางที่ 4.11 ตาราง Dim_Class สำหรับเก็บข้อมูลระดับชั้นการศึกษา 62
ตารางที่ 4.12 ตาราง Dim_Department สำหรับเก็บข้อมูลรายชื่อกลุ่มสาระการเรียนรู้ 62
ตารางที่ 4.13 แสดงคุณลักษณะหรือตัวแปร 74
ตารางที่ 4.14 การแปลงข้อมูลที่เป็นผลลัพธ์ 75
ตารางที่ 4.15 การแปลงข้อมูลคุณลักษณะ 76
ตารางที่ 4.16 ผลการคัดเลือกคุณลักษณะด้วยวิธี Correlation - based Feature Selection (CFS) 80
ตารางที่ 4.17 ผลการคัดเลือกคุณลักษณะด้วยวิธี Information Gain 82
ตารางที่ 4.18 ผลการจัดกลุ่มข้อมูลด้วยอัลกอริทึม K-Means 86
ตารางที่ 4.19 ผลการคัดเลือกคุณลักษณะข้อมูลที่จัดกลุ่มด้วยวิธี Correlation - based Feature Selection 88
ตารางที่ 4.20 ผลการคัดเลือกคุณลักษณะข้อมูลที่จัดกลุ่มด้วยวิธี Information Gain 89
ตารางที่ 4.21 ผลการทดสอบประสิทธิภาพตัวแบบพยากรณ์โดยใช้เทคนิคโครงข่ายประสาทเทียม ด้วยข้อมูลที่ไม่จัดกลุ่ม 93

สารบัญตาราง (ต่อ)

หน้า

ตารางที่ 4.22 ผลการทดสอบประสิทธิภาพตัวแบบพยากรณ์โดยใช้เทคนิค
ซัพพอร์ตเวกเตอร์แมชชีนด้วยข้อมูลที่ไม่จัดกลุ่ม 93
ตารางที่ 4.23 ผลการทดสอบประสิทธิภาพตัวแบบพยากรณ์โดยใช้เทคนิคต้นไม้ตัดสินใจ
ด้วยข้อมูลที่ไม่จัดกลุ่ม 94
ตารางที่ 4.24 เปรียบเทียบค่าความถูกต้องของการสร้างตัวแบบพยากรณ์
ด้วยชุดข้อมูลที่ไม่จัดกลุ่ม 94
ตารางที่ 4.25 เปรียบเทียบค่ารากที่สองของความคลาดเคลื่อนของการสร้างตัวแบบพยากรณ์
ด้วยชุดข้อมูลที่ไม่จัดกลุ่ม 95
ตารางที่ 4.26 ผลการทดสอบประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลที่จัดกลุ่ม 97
ตารางที่ 4.27 การเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ 99
ตารางที่ 4.28 ค่าเฉลี่ยของการประเมินความพึงพอใจระบบคลังข้อมูล 103
ตารางที่ 4.29 ผลการประเมินความพึงพอใจระบบคลังข้อมูลผลการเรียน 104
ตารางที่ 4.30 ค่าเฉลี่ยของการประเมินความเชื่อมั่นระบบพยากรณ์ผลการเรียนเฉลี่ย 105
ตารางที่ 4.31 ผลการการประเมินความเชื่อมั่นระบบพยากรณ์ผลการเรียนเฉลี่ย 106

สารบัญภาพ

หน้า

ภาพที่ 1.1 กราฟแสดงจำนวนนักเรียนที่มีผลการเรียนเป็นศูนย์ตั้งแต่ปี พ.ศ.2554 – 2556 2
ภาพที่ 1.2 กราฟแสดงจำนวนนักเรียนที่ลาออกกลางคัน ตั้งแต่ ปี พ.ศ.2554 – 2556 3
ภาพที่ 1.3 กราฟแสดงจำนวนนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่ไม่ได้ศึกษาต่อในระดับมัธยมศึกษาตอนปลาย 3
ภาพที่ 1.4 กรอบแนวคิดการวิจัย 6
ภาพที่ 2.1 โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา 11
ภาพที่ 2.2 กระบวนการคลังข้อมูล 15
ภาพที่ 2.3 ตัวอย่างของแฟกต์ (Fact), ไดเมนชัน (Dimension) และเมเชอร์ (Measure) 18
ภาพที่ 2.4 โครงสร้างแบบสตาร์สกีมา (Star Schema) 19
ภาพที่ 2.5 โครงสร้างแบบสโนว์เฟลกสกีมา (Snowflake Schema) 20
ภาพที่ 2.6 กระบวนการ ETL 21
ภาพที่ 2.7 ลักษณะของคิวบ์ 22
ภาพที่ 2.8 การ Drill Down และ Roll Up มิติข้อมูล 23
ภาพที่ 2.9 การ Slide และ Dice มิติข้อมูล 24
ภาพที่ 2.10 การค้นหาความรู้ด้วยการทำเหมืองข้อมูล 25
ภาพที่ 2.11 สถาปัตยกรรมเหมืองข้อมูล 26
ภาพที่ 2.12 ตัวอย่างข้อมูลที่ถูกจัดกลุ่มได้ 3 กลุ่ม 28
ภาพที่ 2.13 กระบวนการ CRISP-DM Model 29
ภาพที่ 2.14 โครงข่ายประสาทเทียมแบบหลายชั้น 30
ภาพที่ 2.15 โครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน 31
ภาพที่ 2.16 หลักการของซัพพอร์ตเวกเตอร์แมชชีน 32
ภาพที่ 2.17 ไฮเปอร์เพลนที่เป็นเส้นตรงแบ่งกลุ่มข้อมูล 32
ภาพที่ 2.18 ลักษณะโครงสร้างต้นไม้ตัดสินใจ 33
ภาพที่ 2.19 ข้อมูลของลูกค้าที่อาศัยในพื้นที่ต่างๆ ถูกจัดได้ 3 กลุ่ม 35
ภาพที่ 2.20 ขั้นตอนการจัดกลุ่มข้อมูลด้วย k-means 36
ภาพที่ 2.21 ตัวอย่างการแบ่งข้อมูลด้วยวิธี K - fold Cross validation 37

สารบัญภาพ (ต่อ)

หน้า

ภาพที่ 2.22 การคัดเลือกคุณลักษณะด้วยวิธี CFS 40
ภาพที่ 2.23 แสดงหน้าจอการทำงานโปรแกรม WEKA 42
ภาพที่ 2.24 โครงสร้างการทำคลังข้อมูลและเหมืองข้อมูลของโรงเรียน 46
ภาพที่ 3.1 แสดงขั้นตอนการพัฒนาคลังข้อมูล 51
ภาพที่ 3.2 กระบวนการสร้างตัวแบบพยากรณ์ 53
ภาพที่ 4.1 แสดงโครงสร้างคลังข้อมูลแบบสโนว์เฟลค 63
ภาพที่ 4.2 กระบวนการอีทีแอลด้วยเครื่องมือ SQL Server Integration Services 64
ภาพที่ 4.3 ขั้นตอนกระบวนการอีทีแอล 64
ภาพที่ 4.4 แสดงผลของกระบวนการอีทีแอล 65
ภาพที่ 4.5 แสดงจำนวนข้อมูลในคลังข้อมูล 66
ภาพที่ 4.6 แสดงผลที่ได้จากการสร้าง Data Source View 67
ภาพที่ 4.7 แสดงตัวอย่างการประมวลผลเชิงวิเคราะห์แบบออนไลน์ 68
ภาพที่ 4.8 ตัวอย่างการแสดงผลข้อมูลหลายมิติ 69
ภาพที่ 4.9 ข้อมูลแบบสรุป (Roll UP) และแบบเจาะลึก (Drill Down) 70
ภาพที่ 4.10 ตัวอย่างแดชบอร์ดของผู้บริหาร 71
ภาพที่ 4.11 ตัวอย่างแดชบอร์ดของหัวหน้ากลุ่มสาระการเรียนรู้ 71
ภาพที่ 4.12 ตัวอย่างแดชบอร์ดของอาจารย์ประจำชั้น 72
ภาพที่ 4.13 แสดงไฟล์ข้อมูลที่ได้จากคลังข้อมูล 73
ภาพที่ 4.14 แสดงผลการแปลงข้อมูล 78
ภาพที่ 4.15 แสดงชุดข้อมูลที่ไม่มีการคัดเลือกคุณลักษณะ 79
ภาพที่ 4.16 การคัดเลือกคุณลักษณะด้วยวิธี Correlation - base Feature Selection (CFS) 80
ภาพที่ 4.17 การคัดเลือกคุณลักษณะด้วยวิธี Information Gain (IG) 81
ภาพที่ 4.18 ผลการใช้เทคนิคโครงข่ายประสาทเทียมกับข้อมูลที่ไม่จัดกลุ่ม 83
ภาพที่ 4.19 ผลการใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนกับข้อมูลที่ไม่จัดกลุ่ม 84
ภาพที่ 4.20 ผลการใช้เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5 กับข้อมูลที่ไม่จัดกลุ่ม 85
ภาพที่ 4.21 ผลการจัดกลุ่มข้อมูลด้วยอัลกอริทึม K-Means 86

สารบัญภาพ (ต่อ)

หน้า

ภาพที่ 4.22 การจำแนกกลุ่มข้อมูล 87
ภาพที่ 4.23 ผลการใช้เทคนิคโครงข่ายประสาทเทียมกับข้อมูลที่จัดกลุ่ม 90
ภาพที่ 4.24 ผลการใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนกับข้อมูลที่จัดกลุ่ม 91
ภาพที่ 4.25 ผลการใช้เทคนิคต้นไม้ตัดสินใจกับข้อมูลที่จัดกลุ่ม 92
ภาพที่ 4.26 กราฟแสดงค่าความถูกต้องของการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่ไม่จัดกลุ่ม 95
ภาพที่ 4.27 กราฟแสดงค่ารากที่สองของความคลาดเคลื่อนของการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่ไม่จัดกลุ่ม 96
ภาพที่ 4.28 กราฟแสดงค่าความถูกต้องของตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่ม 98
ภาพที่ 4.29 กราฟแสดงค่ารากที่สองของความคลาดเคลื่อนตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่ม 99
ภาพที่ 4.30 กราฟเปรียบเทียบประสิทธิภาพค่าความถูกต้อง ตัวแบบพยากรณ์ 100
ภาพที่ 4.31 ตัวอย่างการใช้งานระบบพยากรณ์ผลการเรียนเฉลี่ย 101
ภาพที่ 4.32 แสดงตัวอย่างผลการพยากรณ์แผนการเรียนโครงสร้างที่ 1 101
ภาพที่ 4.33 แสดงตัวอย่างผลการพยากรณ์แผนการเรียนโครงสร้างที่ 3 102

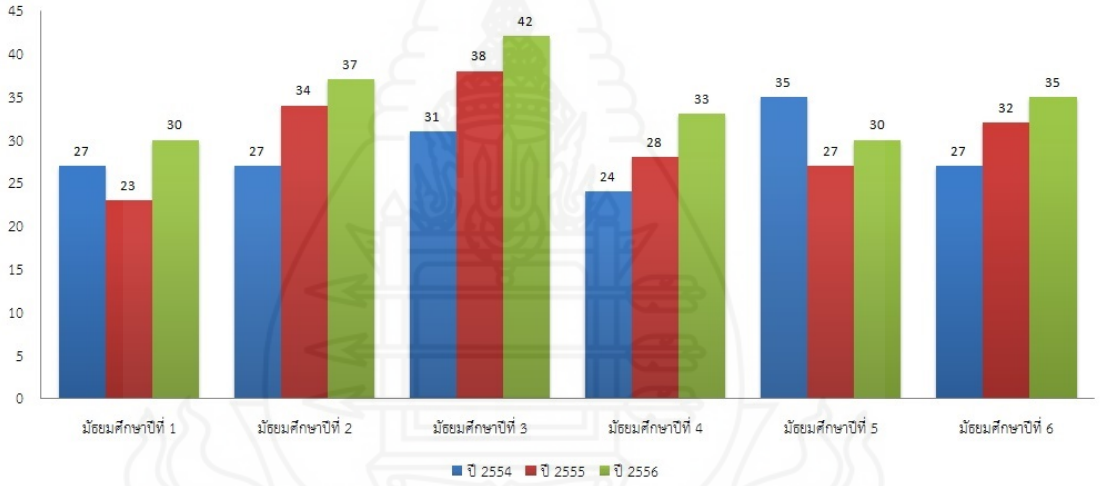
บทที่ 1
บทนำ

1. ความเป็นมาและความสำคัญของปัญหา

การศึกษาเป็นพื้นฐานของบุคคล และเป็นเครื่องมือสำคัญในการพัฒนาความสามารถ ความคิด คุณธรรมและจริยธรรม เพื่อให้บุคคลเป็นคนที่ดีมีคุณภาพสามารถใช้ชีวิตและดำรงอยู่ได้ในสังคม การศึกษาจึงมีส่วนสำคัญและมีบทบาทที่สำคัญต่อการพัฒนาประเทศ ในระบบการศึกษาขั้นพื้นฐานของไทยในปัจจุบัน ตามที่กำหนดไว้ในพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ.2542 แก้ไขเพิ่มเติม (ฉบับที่ 2) 2545 กำหนดให้มีการศึกษาขั้นพื้นฐานไม่น้อยกว่าสิบสองปีก่อนระดับอุดมศึกษา ที่เป็นการศึกษาในระบบการศึกษาขั้นพื้นฐานจะแบ่งได้ สามระดับ คือ การศึกษาก่อนระดับประถมศึกษา เป็นการจัดการศึกษาให้แก่เด็กที่มีอายุ 3 – 6 ปี การศึกษาระดับประถมศึกษา เป็นการศึกษาที่มุ่งเน้นให้ผู้เรียนมีความรู้ความสามารถพื้นฐาน โดยปกติใช้เวลาเรียน 6 ปี และการศึกษาระดับมัธยมศึกษาแบ่งออกเป็นสองระดับ คือ การศึกษาระดับมัธยมศึกษาตอนต้น จะใช้เวลาเรียน 3 ปี และการศึกษาระดับมัธยมศึกษาตอนปลาย จะใช้เวลาเรียน 3 ปี สถานศึกษาจะเป็นผู้รับผิดชอบดำเนินการให้การศึกษาขั้นพื้นฐาน ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 โดยที่ผู้เรียนต้องผ่านการวัดและประเมินผลการเรียนรู้ บนหลักการพื้นฐานสองประการคือ การประเมินเพื่อพัฒนาผู้เรียนและเพื่อตัดสินผลการเรียน เพื่อให้พัฒนาคุณภาพการเรียนของผู้เรียนให้ประสบความสำเร็จ ผลการประเมินเป็นข้อมูลและสารสนเทศที่แสดงพัฒนาการความก้าวหน้า และความสำเร็จทางการเรียนของผู้เรียน

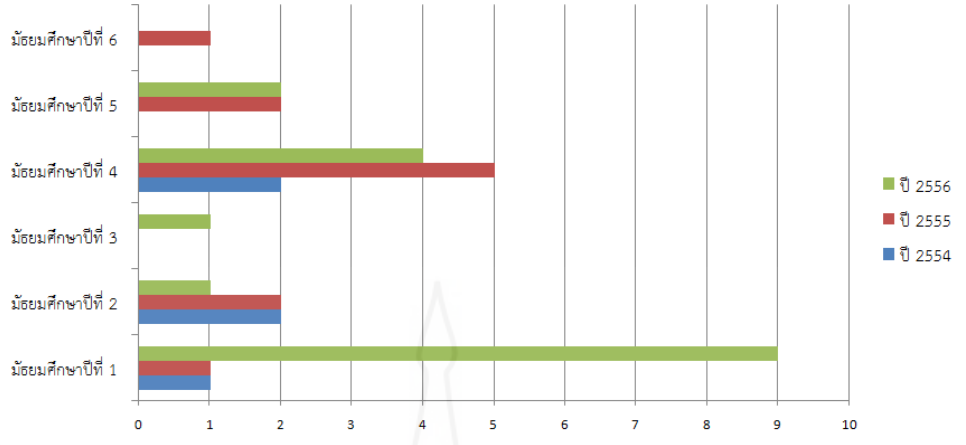
ในปี พ.ศ. 2523 ทางมหาวิทยาลัยเกษตรศาสตร์มีนโยบายที่จะดำเนินการเปิดโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน เพื่อจะได้สอดคล้องกับแผนพัฒนามหาวิทยาลัยเกษตรศาสตร์และเป็นการส่งเสริมขยายงานทางด้านการศึกษาระดับประถมศึกษาและมัธยมศึกษาในชุมชน ดังนั้นมหาวิทยาลัยเกษตรศาสตร์ จึงได้อนุมัติให้เปิดโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ขึ้น ณ วิทยาเขตกำแพงแสน ตั้งแต่วันที่ 15 กรกฎาคม 2523 โดยให้คณะศึกษาศาสตร์เป็นฝ่ายรับผิดชอบและวางเป้าหมายที่จะให้โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน เป็นศูนย์วิจัยและพัฒนาการศึกษา ที่จะทำการวิจัยในด้านการศึกษาในระดับประถมศึกษาและมัธยมศึกษา ในเขตชนบทไปพร้อมกับการจัดการเรียนการสอนตามหลักสูตรของ

กระทรวงศึกษาธิการ พร้อมทั้งให้บริการทางวิชาการแก่ชุมชน ปัจจุบันโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา มีการจัดการเรียนการสอนตั้งแต่ชั้นอนุบาลศึกษา ประถมศึกษา และมัธยมศึกษา มีนักเรียนรวมทั้งสิ้น 1,696 คน แบ่งเป็นระดับอนุบาลศึกษา จำนวน 222 คน ระดับประถมศึกษา จำนวน 693 คน ระดับมัธยมศึกษาตอนต้น จำนวน 382 คน และระดับมัธยมศึกษาตอนปลาย จำนวน 399 คน ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ทางโรงเรียนได้จัดให้มีการเรียนการสอนตามหลักสูตร นักเรียนจะต้องลงทะเบียนเรียนในรายวิชาพื้นฐาน รายวิชาบังคับเลือก และรายวิชาเลือกเสรี ที่ทางโรงเรียนกำหนดให้นักเรียนลงทะเบียนเรียนในแต่ละภาคเรียน จากการสังเกตผลการเรียนในระดับมัธยมศึกษาตอนต้น และมัธยมศึกษาตอนปลาย ตั้งแต่ปี พ.ศ. 2554 – 2556 พบว่า มีผลการเรียนไม่ผ่านเกณฑ์ในการวัดและประเมินผลการศึกษามีจำนวนมากขึ้น และแนวโน้มเพิ่มขึ้นในแต่ละปี ดังจะเห็นได้จากกราฟที่แสดงจำนวนนักเรียนที่มีผลการเรียนเป็นศูนย์ จากภาพที่ 1.1



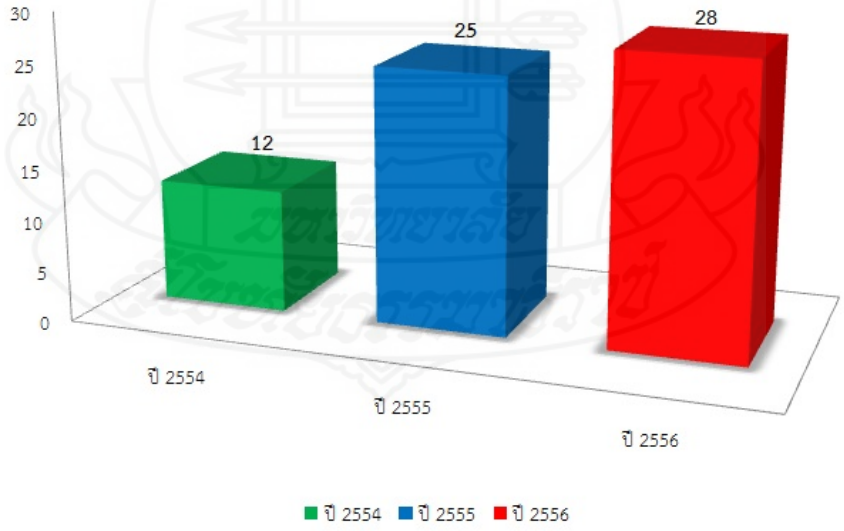
ภาพที่ 1.1 กราฟแสดงจำนวนนักเรียนที่มีผลการเรียนเป็นศูนย์ตั้งแต่ ปี พ.ศ. 2554 – 2556

จากภาพที่ 1.1 จะเห็นได้ว่านักเรียนในระดับมัธยมศึกษาตอนต้นมีนักเรียนที่มีผลการเรียนไม่ผ่านเกณฑ์ในแต่ละรายวิชาจำนวนมาก ส่วนระดับมัธยมศึกษาตอนปลายมีนักเรียนที่มีผลการเรียนไม่ผ่านเกณฑ์เพิ่มขึ้นในแต่ละปี เมื่อดูข้อมูลในส่วนของการลาออกของนักเรียนในฐานข้อมูลนักเรียน ตั้งแต่ปีการศึกษา 2554 – 2556 พบว่า นักเรียนที่ลาออกกลางคันมีจำนวนที่จะเพิ่มขึ้นในแต่ละปีการศึกษา ดังจะเห็นได้จากกราฟ จากภาพที่ 1.2



ภาพที่ 1.2 กราฟแสดงจำนวนนักเรียนที่ลาออกกลางคัน ตั้งแต่ ปี พ.ศ. 2554 – 2556

จากภาพที่ 1.2 พบว่า นักเรียนระดับมัธยมศึกษาตอนต้นและระดับมัธยมศึกษาตอนปลายมีการลาออกกลางคันเพิ่มขึ้นในแต่ละปีการศึกษา นอกจากนี้ยังพบข้อมูลของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ที่ไม่ได้ศึกษาต่อในระดับมัธยมศึกษาตอนปลายในโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา ดังจะเห็นได้จากจากภาพที่ 1.3 แสดงกราฟของจำนวนนักเรียนที่ไม่ได้ศึกษาต่อในระดับมัธยมศึกษาตอนปลาย



ภาพที่ 1.3 กราฟแสดงจำนวนนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่ไม่ได้ศึกษาต่อในระดับมัธยมศึกษาตอนปลาย

จากภาพที่ 1.3 พบว่า นักเรียนมัธยมศึกษาปีที่ 3 ไม่ได้ศึกษาต่อในระดับมัธยมศึกษาตอนปลายในโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา มีแนวโน้มเพิ่มขึ้นในแต่ละปีการศึกษา และบางคนสำเร็จการศึกษาช้ากว่าที่หลักสูตรกำหนดไว้ ในบางกรณีอาจจะต้องให้นักเรียนเรียนซ้ำชั้นในระดับการศึกษานั้นๆ เมื่อนักเรียนมีผลการเรียนที่ไม่ผ่านเกณฑ์ที่กำหนดไว้นักเรียนต้องมาลงทะเบียนซ่อมเสริมในรายวิชานั้นๆ แต่ผลการเรียนที่ได้ก็จะได้ผลการเรียนอยู่ที่ 1 (เกรด 1) จะทำผลการเรียนเฉลี่ยที่นักเรียนในแต่ละภาคนั้นต่ำลงไปด้วย ในบางกรณีนักเรียนสอบไม่ผ่านในรายวิชาที่ลงทะเบียนจำนวนมากและสะสมในหลายเทอม จนทำให้นักเรียนลาออกกลางคัน ซึ่งมีเพิ่มมากขึ้นในช่วง 2-3 ปีหลัง ปัญหาของนักเรียนคือ เมื่อมีผลการเรียนที่ต่ำในระดับมัธยมศึกษาตอนต้นทำให้นักเรียนไม่สามารถเลือกแผนการเรียนในระดับมัธยมศึกษาตอนปลายตามความสนใจและถนัดของนักเรียน ต้องเลือกแผนการเรียนที่ไม่ถนัดและไม่เหมาะสมกับความสามารถของนักเรียน ทำให้เรียนแล้วไม่ประสบความสำเร็จ มีผลการเรียนที่ต่ำ และทำให้ไม่จบตามกำหนดเวลาของหลักสูตรในระดับมัธยมศึกษาตอนปลาย ปัญหาเหล่านี้จึงเป็นอุปสรรคต่อการพัฒนาคุณภาพการศึกษาของโรงเรียน ซึ่งการให้คำแนะนำของอาจารย์ประจำชั้นและอาจารย์แนะแนว จะใช้การสอบถามและดูข้อมูลจากงานทะเบียนและประมวลผล โดยอาจารย์ประจำชั้นและอาจารย์แนะแนวไม่มีข้อมูลที่มากพอให้คำปรึกษาหรือเครื่องมือที่ช่วยในการให้คำแนะนำ ดังนั้นหากมีเครื่องมือช่วยในการวิเคราะห์คาดการณ์ ทำนายหรือพยากรณ์ผลการเรียนของนักเรียน และการเลือกแผนการเรียนในระดับมัธยมศึกษาตอนปลาย จะทำให้อาจารย์ประจำชั้นและอาจารย์แนะแนวมีข้อมูลในการให้คำแนะนำแก่นักเรียนได้อย่างมีประสิทธิภาพ นักเรียนสามารถวางแผนการเรียน ช่วยให้นักเรียนมีการปรับปรุงตัวเองมีความตั้งใจในการเรียนเพิ่มมากขึ้น ส่งผลทำให้ผลการเรียนของนักเรียนมีแนวโน้มที่ดีขึ้นหรือผลสัมฤทธิ์ทางการเรียนสูงขึ้น และช่วยพัฒนาคุณภาพการศึกษาของโรงเรียน

การพยากรณ์เป็นการคาดคะเน หรือการทำนายเหตุการณ์ที่จะเกิดขึ้นในอนาคต (ทรงศิริ แต้สมบัติ, 2549) ปกติแล้วการทำนายหรือพยากรณ์จะใช้หลักการทางสถิติเป็นส่วนใหญ่ นอกจากวิธีการทางสถิติแล้วนั้น ยังมีอีกวิธีที่นักวิจัยนิยมนำมาใช้ในการทำนายสิ่งที่ต้องการรู้ สิ่งที่จะเกิดขึ้นในอนาคต เป็นการทำนายหาค่าที่ต้องการรู้จากข้อมูลที่มีอยู่ การค้นหาความรู้ที่ซ่อนเร้นอยู่ในข้อมูลจำนวนมาก โดยใช้เทคนิคและวิธีการวิเคราะห์ข้อมูลที่ได้รับความสนใจอย่างสูงในปัจจุบัน เรียกว่า การทำเหมืองข้อมูล (Data Mining) เมื่อใช้เทคนิคการทำเหมืองข้อมูล ข้อมูลขนาดใหญ่จะถูกวิเคราะห์และสืบค้นความรู้หรือสิ่งสำคัญออกมา จากนั้นจะรวบรวมความรู้ที่ได้ให้อยู่ในรูปของฐานความรู้ (Knowledge Base) เพื่อนำไปใช้ประโยชน์ในการตัดสินใจ การทำเหมืองข้อมูลมีเทคนิคเป็นที่แพร่หลายและมีประสิทธิภาพให้ค่าที่แม่นยำในการทำนายสูง ในการทำเหมืองข้อมูลมี

ด้วยกันหลายเทคนิค เทคนิคของแบ่งกลุ่มข้อมูล (Clustering) เป็นการแบ่งข้อมูลที่มีลักษณะคล้ายคลึงกันไว้ในกลุ่มเดียวกัน เทคนิคการหาความสัมพันธ์ (Association) เป็นการหาความสัมพันธ์หรือความเกี่ยวเนื่องของข้อมูล และเทคนิคโครงข่ายประสาทเทียม (Artificial neural network) ที่ใช้ในการพยากรณ์หรือทำนายผลที่ให้ความถูกต้องในการทำนายสูงกว่าวิธีการพยากรณ์แบบอื่นๆ เมื่อข้อมูลมีรูปแบบไม่เป็นเส้น (Makridakis, Wheelwright and Hyndman. 1998) โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา มีงานทะเบียนและประมวลผลเก็บข้อมูลนักเรียนเป็นจำนวนมาก เช่น ฐานข้อมูลนักเรียน การลงทะเบียนเรียน และผลการเรียน ซึ่งสามารถนำมาสร้างคลังข้อมูลที่ช่วยในการวิเคราะห์ และใช้เทคนิคเหมืองข้อมูลในการพยากรณ์ผลการเรียนของนักเรียน

ดังนั้นผู้วิจัยจะนำข้อมูลที่มีอยู่มาพัฒนาเป็นคลังข้อมูลผลการเรียนของนักเรียน และนำเอาเทคนิคการทำเหมืองข้อมูลมาประยุกต์ใช้ใน การพยากรณ์ผลการเรียนเฉลี่ยของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา โดยผลการพยากรณ์นำไปใช้ประกอบการตัดสินใจ เพื่อหาแนวทางการศึกษาที่เหมาะสมสำหรับกับนักเรียนระดับชั้นมัธยมศึกษาตอนต้นในการเลือกแผนการเรียนที่เหมาะสมกับนักเรียน เพิ่มโอกาสสำเร็จการศึกษาตามที่หลักสูตรกำหนด และเป็นข้อมูลให้กับอาจารย์ประจำชั้นหรืออาจารย์แนะแนวช่วยในการวิเคราะห์นักเรียนได้อย่างรวดเร็วและสะดวกมากยิ่งขึ้น กรณีนักเรียนในกลุ่มเสี่ยงไม่สำเร็จการศึกษาหรือต้องลาออกกลางคันก็จะเป็นข้อมูลที่มีประโยชน์ในการให้คำปรึกษา หรือแนะนำหรือช่วยแก้ปัญหาในเบื้องต้นได้อย่างถูกต้องและช่วยให้นักเรียนมีความตั้งใจ กระตือรือร้นในการเรียนเพิ่มมากขึ้น อีกทั้งยังพัฒนาด้านคุณภาพการศึกษาช่วยยกระดับผลสัมฤทธิ์ทางการเรียนให้เพิ่มมากขึ้น

2. วัตถุประสงค์ของการวิจัย

2.1 เพื่อศึกษาและพัฒนาคลังข้อมูล ที่ใช้ในการวิเคราะห์ ผลการเรียนของนักเรียนระดับชั้นมัธยมศึกษา โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา

2.2 เพื่อศึกษาและพัฒนาตัวแบบพยากรณ์ผลการเรียนเฉลี่ยของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา ด้วยเทคนิคเหมืองข้อมูล

2.3 เพื่อประเมินผลระบบพยากรณ์ผลการเรียนของเฉลี่ยของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา

3. สมมติฐานของการวิจัย

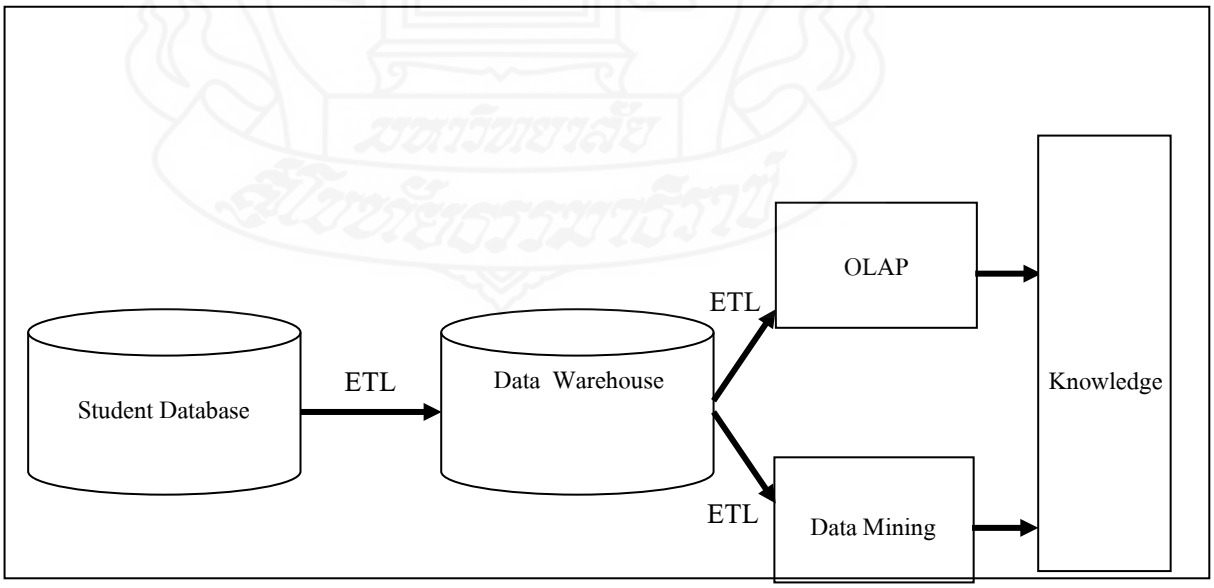
สมมติฐานสำหรับการวิจัยครั้งนี้มีดังนี้

3.1 การพัฒนาคลังข้อมูลเพื่อใช้เป็นข้อมูล ผู้ใช้งานมีความพึงพอใจอยู่ในระดับดี

3.2 ตัวแบบการพยากรณ์ผลการเรียนเฉลี่ย มีค่าความถูกต้องไม่น้อยกว่าร้อยละ 60

4. กรอบแนวคิดของการวิจัย

กรอบแนวคิดงานวิจัยประกอบด้วยข้อมูลผลการเรียน และข้อมูลทั่วไปของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา ระหว่างปีการศึกษา 2548 จนถึงปีการศึกษา 2556 นำมาพัฒนาเป็นคลังข้อมูล (Data warehouse) แล้วทำการประมวลผลเชิงวิเคราะห์ออนไลน์หรือโอแลป (OnLine Analytical Procesing - OLAP) เพื่อนำเสนอข้อมูล และการใช้เทคนิคการทำเหมืองข้อมูล เพื่อการพยากรณ์ผลการเรียนเฉลี่ยของนักเรียนดังภาพที่ 1.4



ภาพที่ 1.4 กรอบแนวคิดการวิจัย

5. ขอบเขตของการวิจัย

5.1 ข้อมูลที่ใช้ในการวิจัย คือ ข้อมูลเกี่ยวกับผลการเรียนของนักเรียนระดับมัธยมศึกษา โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา ได้แก่ ข้อมูลทั่วไปของนักเรียน ข้อมูลผลการเรียน ข้อมูลรายวิชา และอาจารย์ผู้สอน

5.2 ข้อมูลของการพัฒนาคลังข้อมูล ใช้ข้อมูลปีการศึกษา 2548 ถึงปีการศึกษา 2556

5.3 ข้อมูลของการสร้างตัวแบบพยากรณ์ผลการเรียนเฉลี่ยใช้ข้อมูลของนักเรียนศึกษาปีที่ 4 ปีการศึกษา 2553 -2556 หลักสูตรการศึกษา 2551

5.4 กลุ่มผู้ใช้งานคลังข้อมูลได้แก่ ผู้บริหาร หัวหน้ากลุ่มสาระการเรียนรู้ และอาจารย์ประจำชั้น

5.5 กลุ่มผู้ใช้ระบบพยากรณ์ผลการเรียนได้แก่ นักเรียนระดับมัธยมศึกษาตอนต้น และอาจารย์แนะแนว

6. ประโยชน์ที่คาดว่าจะได้รับ

6.1 ได้คลังข้อมูลที่ช่วยในการวิเคราะห์ ข้อมูลพื้นฐาน และผลการเรียนในรายวิชาต่างๆ ของนักเรียนระดับชั้นมัธยมศึกษา

6.2 ได้ระบบการพยากรณ์ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 4 ที่ใช้ในการประกอบการตัดสินใจเลือกแผนการเรียนของนักเรียนระดับมัธยมศึกษาตอนต้น

6.3 โรงเรียนสามารถนำความรู้ที่ได้จากการพยากรณ์ ไปพัฒนาคุณภาพการศึกษา และใช้วางแผนการให้คำปรึกษาแก่นักเรียน หรือช่วยแก้ไขปัญหาในการเรียนเบื้องต้นได้

6.4 ช่วยให้นักเรียนสามารถวางแผนการเรียน และเห็นความสำคัญในการเรียนในรายวิชาต่างๆ มากขึ้น

7. นิยามศัพท์เฉพาะ

ในการวิจัยการใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา มีนิยามศัพท์ที่สำคัญดังนี้

7.1 คลังข้อมูล (Data Warehouse) หมายถึง ระบบฐานข้อมูลที่ออกแบบมา เพื่อใช้งานด้านการวิเคราะห์ข้อมูล ข้อมูลในคลังข้อมูลจะเป็นข้อมูลที่มาจากข้อมูลที่ใช้กับงานประจำหรือระบบปฏิบัติการ จากนั้นนำมาทำให้อยู่ในรูปแบบที่เป็นข้อมูลสรุป ข้อมูลสรุปนี้จะเป็นข้อมูลในอดีตเพื่อให้สามารถนำมาทำการวิเคราะห์ข้อมูลได้อย่างมีประสิทธิภาพ

7.2 เหมืองข้อมูล (Data Mining) หมายถึง กระบวนการค้นหาความรู้ที่น่าสนใจจากข้อมูลที่มีอยู่จำนวนมาก เพื่อหารูปแบบ ความสัมพันธ์ที่มีอยู่จริงแต่ถูกซ่อนอยู่ในข้อมูลจำนวนมาก

7.3 การพยากรณ์ (Prediction) หมายถึง การคาดคะเน หรือการทำนายเหตุการณ์ล่วงหน้าที่จะเกิดขึ้นในอนาคต โดยศึกษาจากข้อมูลที่เก็บรวบรวมไว้อย่างเป็นระบบ

7.4 โครงข่ายประสาทเทียม หรือ นิวรอลเน็ตเวิร์ค (Neural Network) หมายถึง เทคนิคที่ใช้ในการจำแนกและพยากรณ์ โดยเลียนแบบการทำงานของระบบประสาทในสมองของมนุษย์ มีขั้นตอนการรับเข้าข้อมูลอินพุต และทำการคำนวณเพื่อสร้างโมเดลของการเรียนรู้

7.5 ผลสัมฤทธิ์ทางการเรียน หมายถึง ความสามารถในการเรียนของนักเรียน ในการเรียนแต่ละวิชาของแต่ละภาคการศึกษา ซึ่งอาจพิจารณาได้จากคะแนน เกรด และผลการศึกษาเฉลี่ยสะสม (Grade point average : GPA)

บทที่ 2
วรรณกรรมที่เกี่ยวข้อง

การวิจัยครั้งนี้ผู้วิจัยได้ศึกษาค้นคว้าข้อมูล ทฤษฎี และงานวิจัยที่เกี่ยวข้องจากเอกสาร บทความวิชาการและแหล่งสารสนเทศอินเทอร์เน็ต ในหัวข้อดังต่อไปนี้

1. โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา
2. คลังข้อมูล
3. การทำเหมืองข้อมูล
4. การคัดเลือกคุณลักษณะ
5. โปรแกรม WEKA (Waikato Environment for Knowledge Analysis)
6. งานวิจัยที่เกี่ยวข้อง

1. โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา

1.1 ประวัติความเป็นมาของโรงเรียน

เนื่องจากมหาวิทยาลัยเกษตรศาสตร์ มีโครงการขยายมหาวิทยาลัยเกษตรศาสตร์ขึ้น ณ วิทยาเขตกำแพงแสน อำเภอกำแพงแสน จังหวัดนครปฐม เพื่อเป็นแหล่งผลิตบุคลากรระดับปริญญาตรี ปริญญาโทและปริญญาเอก ในสาขาวิชาการเกษตรและอื่นๆ มหาวิทยาลัยเกษตรศาสตร์ จึงกำหนดเปิดใช้วิทยาเขตกำแพงแสนตั้งแต่เดือนพฤศจิกายน พ.ศ.2522 เป็นต้นมา โดยมีคณะต่างๆ เริ่มเปิดดำเนินการ คือ คณะเกษตร, คณะสัตวแพทยศาสตร์, คณะวิศวกรรมศาสตร์, คณะศึกษาศาสตร์ และบัณฑิตวิทยาลัย

ในปี พ.ศ. 2523 ทางมหาวิทยาลัยเกษตรศาสตร์มีนโยบายที่จะเปิดดำเนินการโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน เพื่อจะได้สอดคล้องกับแผนพัฒนามหาวิทยาลัยเกษตรศาสตร์ และเป็นการส่งเสริมขยายงานทางด้านการศึกษาระดับประถมศึกษาและมัธยมศึกษาในชุมชนด้วย ฉะนั้นมหาวิทยาลัยเกษตรศาสตร์ จึงได้อนุมัติให้เปิดโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ขึ้น ณ วิทยาเขตกำแพงแสน ตั้งแต่วันที่ 15 กรกฎาคม 2523 โดยให้คณะศึกษาศาสตร์เป็นฝ่ายรับผิดชอบในการจัดดำเนินการเรียนการสอน ดังนั้นคณะ

ศึกษาศาสตร์จึงวางเป้าหมายที่จะให้โรงเรียนสาธิตเกษตรฯ วิทยาเขตกำแพงแสน เป็นศูนย์วิจัยและพัฒนาการศึกษาที่จะทำการวิจัยการศึกษาในระดับประถมศึกษาและมัธยมศึกษาในเขตชนบทไป พร้อมกับการทดลองจัดการเรียนการสอนแบบสมรรถฐานและการบริการทางวิชาการแก่ชุมชนด้วย

เมื่อคณะศึกษาศาสตร์ได้รับคำสั่งให้ดำเนินการเปิดการเรียนการสอนระดับประถมศึกษาขึ้นที่วิทยาเขตกำแพงแสน ในปี พ.ศ. 2523 นั้น ผู้ช่วยศาสตราจารย์ ดร.รุ่ง เจนจิต คณบดีคณะศึกษาศาสตร์ จึงได้ขออนุมัติจากมหาวิทยาลัยเพื่อให้อาจารย์ของคณะศึกษาศาสตร์มาปฏิบัติราชการประจำที่โรงเรียนสาธิตเกษตรฯ วิทยาเขตกำแพงแสน ในระยะแรกมีทั้งหมด 11 ท่าน เพื่อเตรียมการก่อนเปิดการเรียนการสอนโดยมอบหมายให้ อาจารย์อุดร รัตนภักดิ์ ปฏิบัติราชการในหน้าที่แทนอาจารย์ใหญ่

ปี พ.ศ. 2523 โรงเรียนได้เปิดดำเนินการจัดการเรียนการสอนในระดับประถมศึกษาปีที่ 1 - 3 มีนักเรียน 12 คน โดยวางวัตถุประสงค์ของโรงเรียนดังนี้

1) เพื่อพัฒนาการศึกษา วิจัย และทดลองทฤษฎีทางการศึกษา และจิตวิทยาการศึกษาที่เหมาะสมสำหรับเขตพื้นที่ชนบท เพื่อการพัฒนาการศึกษาชนบท

2) เพื่อการทดลอง วิจัยเกี่ยวกับหลักสูตรในระดับประถมศึกษาและมัธยมศึกษาในการพัฒนาอาชีพแขนงต่างๆ ให้สอดคล้องกับพื้นภูมิประเทศ

3) เพื่อเป็นแหล่งสาธิตและเผยแพร่ผลงานการศึกษา เพื่อการพัฒนาการศึกษาชนบทให้มีประสิทธิภาพ และจะเป็นโรงเรียนที่เป็นพี่เลี้ยงสำหรับโรงเรียนในชนบทในพื้นที่ใกล้เคียง

4) เพื่อเป็นสถานที่สำหรับฝึกปฏิบัติงานของนิสิตคณะศึกษาศาสตร์ในการฝึกปฏิบัติงานครู ฝึกหัดการสอน ทดลองวิจัยเกี่ยวกับแนวทางการปฏิบัติการเรียนการสอนตามหลักวิชา

5) เพื่อเป็นสถานที่สำหรับกุลบุตร กุลธิดา ในท้องที่ชนบทที่จะศึกษาเล่าเรียนตามระดับความสามารถทางสติปัญญา ความถนัด และความสนใจ ซึ่งจะเป็นการพัฒนาบุคคลในชนบทให้เป็นพลเมืองที่มีคุณภาพ พร้อมที่จะนำความรู้ ความคิด ความเข้าใจในการแก้ปัญหา เพื่อเสริมสร้างชีวิตและสังคมให้ดีขึ้นอันเป็นรากฐานของพลเมืองดี

ปี พ.ศ. 2532 มีนักเรียนตั้งแต่ชั้นประถมศึกษาปีที่ 1 - 6 และชั้นมัธยมศึกษาปีที่ 1 - 6

ปี พ.ศ. 2534 ทางโรงเรียนได้รับอนุมัติจากสภามหาวิทยาลัยเกษตรศาสตร์ ให้จัดการเรียนการสอนระดับอนุบาลศึกษาเพิ่มขึ้น

ปี พ.ศ. 2546 มหาวิทยาลัยเกษตรศาสตร์ได้อนุมัติให้เปลี่ยนชื่อโรงเรียนจากเดิมมาเป็นโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนา

การศึกษา (Kasetsart University Laboratory School Kamphaeng Saen Campus Educational Research and Development Center) โดยมีฐานะเทียบเท่าภาควิชาในคณะศึกษาศาสตร์

ปี พ.ศ. 2554 โรงเรียนได้รับอนุมัติจากกระทรวงวิทยาศาสตร์และเทคโนโลยี ให้ดำเนินการจัดการเรียนการสอนหลักสูตรห้องเรียนวิทยาศาสตร์ในโรงเรียน ในระดับชั้นมัธยมศึกษาตอนปลาย โดยการสนับสนุนของกระทรวงวิทยาศาสตร์และเทคโนโลยี ในการกำกับดูแลของมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน

ปี พ.ศ. 2555 มหาวิทยาลัยเกษตรศาสตร์ ได้อนุมัติให้โรงเรียนดำเนินงานโครงการจัดการเรียนการสอนตามหลักสูตรกระทรวงศึกษาธิการเป็นภาษาอังกฤษ (English Program)

ปัจจุบันโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน จึงมีการจัดการเรียน การสอนตั้งแต่ชั้นอนุบาลศึกษาปีที่ 1 ถึงชั้นมัธยมศึกษาปีที่ 6 จำนวน 50 ห้องเรียน คณาจารย์ 92 ท่าน และนักเรียน 1,728 คน

ภาพที่ 2.1 โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา

1.2 การจัดการศึกษา

โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา สังกัดสำนักงานคณะกรรมการการอุดมศึกษา (สกอ.) จัดการเรียนการสอนตาม

หลักสูตรหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 จัดการเรียนการสอนตั้งแต่ระดับอนุบาลศึกษา ระดับประถมศึกษา และระดับมัธยมศึกษา เป็นการศึกษาในระบบโรงเรียนที่กำหนดจุดมุ่งหมาย วิธีการศึกษา หลักสูตร ระยะเวลาของการศึกษา การวัดและการประเมินผล เป็นเงื่อนไขของการสำเร็จการศึกษาที่แน่นอน มีสาระการเรียนรู้ 8 กลุ่มประกอบด้วย กลุ่มสาระการเรียนรู้ภาษาไทย, กลุ่มสาระการเรียนรู้คณิตศาสตร์, กลุ่มสาระการเรียนรู้วิทยาศาสตร์, กลุ่มสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม, กลุ่มสาระการเรียนรู้สุขศึกษาและพลศึกษา, กลุ่มสาระการเรียนรู้ศิลปะ, กลุ่มสาระการเรียนรู้การงานอาชีพและเทคโนโลยี และกลุ่มสาระการเรียนรู้ภาษาต่างประเทศ ในการวัดและประเมินผลการเรียนของรายวิชาแต่ละกลุ่มสาระการเรียนรู้ใช้ตัวเลขแสดงระดับผลการเรียน ดังตารางที่ 2.1

ตารางที่ 2.1 แสดงระดับผลการเรียน

ระดับผลการเรียน	ร้อยละ	ความหมาย
4	80 - 100	ดีเยี่ยม
3.5	75 -79	ดีมาก
3	70 - 74	ดี
2.5	65 - 69	ค่อนข้างดี
2	60 - 64	ปานกลาง
1.5	55 - 59	พอใช้
1	50 - 54	ผ่านเกณฑ์ขั้นต่ำ
0	0 - 4 9	ต่ำกว่าเกณฑ์

การตัดสินผลการเรียนในระดับการศึกษาขั้นพื้นฐานจะใช้ระบบผ่าน และไม่ผ่าน โดยกำหนดเกณฑ์การตัดสินผ่านแต่ละวิชาที่ร้อยละ 50 จากนั้นจึงให้ระดับผลการเรียนที่ผ่านสำหรับระดับมัธยมศึกษาตอนต้น และระดับมัธยมศึกษาตอนปลาย ใช้ตัวเลขแสดงระดับผลการเรียนเป็น 8 ระดับ ส่วนการเลื่อนระดับชั้นในระดับมัธยมศึกษามีการตัดสินในหลายลักษณะ คือ การผ่านรายวิชากำหนดเป็นภาคเรียน การเลื่อนชั้นปีกำหนดเป็นปีการศึกษาและการจบระดับชั้นกำหนดเป็นระดับมัธยมศึกษาตอนต้นและระดับมัธยมศึกษาตอนปลาย หลักเกณฑ์การวัดและประเมินผลการเรียนรู้เพื่อตัดสินผลการเรียนของผู้เรียนตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ดังนี้

1) ตัดสินผลการเรียนเป็นรายวิชา ผู้เรียนต้องมีเวลาเรียนตลอดภาคเรียนไม่น้อยกว่าร้อยละ 80 ของเวลาเรียนทั้งหมดในรายวิชานั้นๆ

2) ผู้เรียนต้องได้รับการประเมินทุกตัวชี้วัดและมีผลการประเมินผ่านไม่น้อยกว่าร้อยละ 60 ของตัวชี้วัดทั้งหมด

3) รายวิชาพื้นฐาน ได้รับการตัดสินผลการเรียนผ่านทุกรายวิชา ตั้งแต่ผ่านเกณฑ์ขั้นต่ำขึ้นไป (คือตั้งแต่ 1.00 ขึ้นไป)

4) รายวิชาเพิ่มเติม ได้รับการตัดสินผลการเรียน

5) ระดับผลการเรียนเฉลี่ยในปีการศึกษานั้นได้ไม่ต่ำกว่า 1.00

6) ผู้เรียนต้องผ่านเกณฑ์การประเมินการอ่าน คิดวิเคราะห์และเขียน

7) ผู้เรียนต้องผ่านเกณฑ์การประเมินคุณลักษณะอันพึงประสงค์

8) ผู้เรียนต้องผ่านเกณฑ์การประเมินกิจกรรมพัฒนาผู้เรียนทุกกิจกรรม

ทั้งนี้รายวิชาใดที่ไม่ผ่านเกณฑ์การประเมิน โรงเรียนสามารถซ่อมเสริมผู้เรียนให้ได้รับการแก้ไขในภาคเรียนถัดไป ในส่วนของเกณฑ์การจบการศึกษาในระดับมัธยมศึกษาตอนต้น มีดังนี้

1) ผู้เรียนเรียนรายวิชาพื้นฐานและเพิ่มเติมไม่เกิน 81 หน่วยกิต โดยเป็นรายวิชาพื้นฐาน 66 หน่วยกิต และรายวิชาเพิ่มเติมตามที่สถานศึกษากำหนด

2) ผู้เรียนต้องได้หน่วยกิตตลอดหลักสูตรไม่น้อยกว่า 77 หน่วยกิต โดยเป็นรายวิชาพื้นฐาน 66 หน่วยกิต และรายวิชาเพิ่มเติมไม่น้อยกว่า 11 หน่วยกิต

3) ผู้เรียนมีผลการประเมินตั้งแต่ “ผ่าน” ขึ้นไปในการประเมินการอ่าน คิดวิเคราะห์และเขียน

4) ผู้เรียนมีผลการประเมินตั้งแต่ “ผ่าน” ขึ้นไปในการประเมินคุณลักษณะอันพึงประสงค์

5) ผู้เรียนเข้าร่วมกิจกรรมพัฒนาผู้เรียนและมีผลการประเมินเป็น “ผ่าน” ในทุกกิจกรรมของกิจกรรมพัฒนาผู้เรียน

6) ผู้เรียนต้องลงทะเบียนเรียนภาษาต่างประเทศที่สองอย่างน้อย 1 วิชา และได้รับการตัดสินผลการเรียนตั้งแต่เกณฑ์ขั้นต่ำขึ้นไป คือตั้งแต่ 1.00 ขึ้นไป

เกณฑ์การจบการศึกษาระดับมัธยมศึกษาตอนปลาย มีดังนี้

1) ผู้เรียนเรียนรายวิชาพื้นฐานและเพิ่มเติมไม่น้อยกว่า 81 หน่วยกิต โดยเป็นรายวิชาพื้นฐาน 41 หน่วยกิต และรายวิชาเพิ่มเติมตามที่สถานศึกษากำหนด

2) ผู้เรียนต้องได้หน่วยกิตตลอดหลักสูตรไม่น้อยกว่า 77 หน่วยกิต โดยเป็นรายวิชาพื้นฐาน 41 หน่วยกิต และรายวิชาเพิ่มเติมไม่น้อยกว่า 36 หน่วยกิต

3) ผู้เรียนมีผลการประเมินตั้งแต่ “ผ่าน” ขึ้นไปในการประเมินการอ่าน คิดวิเคราะห์และเขียน

4) ผู้เรียนมีผลการประเมินตั้งแต่ “ผ่าน” ขึ้นไปในการประเมินคุณลักษณะอันพึงประสงค์

5) ผู้เรียนเข้าร่วมกิจกรรมพัฒนาผู้เรียนและมีผลการประเมินเป็น “ผ่าน” ในทุกกิจกรรมของกิจกรรมพัฒนาผู้เรียน

6) ผู้เรียนต้องลงทะเบียนเรียนภาษาต่างประเทศที่สองอย่างน้อย 1 รายวิชา และได้รับการตัดสินผลการเรียนตั้งแต่เกณฑ์ขั้นต่ำขึ้นไป คือตั้งแต่ 1.00 ขึ้นไป

2. คลังข้อมูล (Data warehouse)

นับจากอดีตจนถึงปัจจุบัน ปัจจัยสำคัญที่ทำให้องค์กรต่างๆ ประสบความสำเร็จได้ก็คือข้อมูลที่มีอยู่และที่ใช้ในประจำวัน ซึ่งปริมาณเพิ่มมากขึ้นจนทำให้เกิดปัญหาสำหรับองค์กรที่จะต้องจัดเก็บข้อมูลเหล่านี้ นอกจากนี้ข้อมูลที่มากมายยังไม่เหมาะสมที่จะนำมาใช้ในการวิเคราะห์หรือช่วยในการตัดสินใจของผู้บริหาร (Decision Support System) ในด้านต่างๆ เพราะต้องใช้เวลาในการประมวลผลข้อมูลที่มากนานพอสมควร และส่งผลกระทบต่อระบบการทำงานของข้อมูลที่ใช้งานประจำวันอีกด้วย ด้วยเหตุผลดังกล่าวจึงต้องมีการพัฒนาระบบคลังข้อมูลขึ้นมาเพื่อสนับสนุนการตัดสินใจของผู้บริหารให้ตรงตามเป้าหมายหรือนโยบายขององค์กร

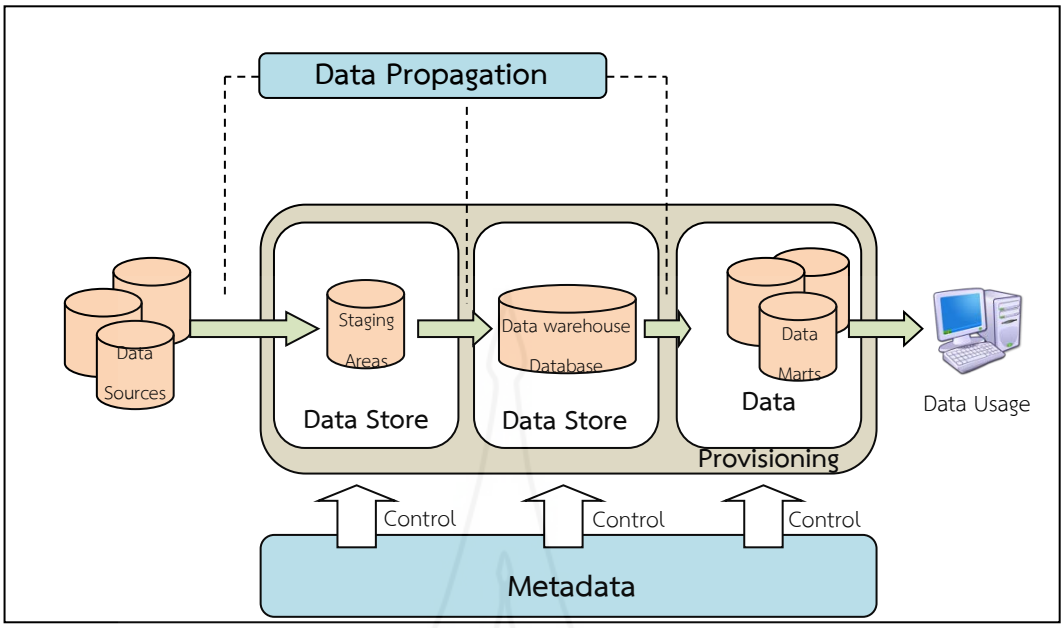
2.1 ความหมายของคลังข้อมูล

คลังข้อมูล คือ ฐานข้อมูลขนาดใหญ่ที่มีการเก็บรวบรวมข้อมูลจากแหล่งข้อมูลต่างๆ หลายช่วงเวลา ทั้งภายในและภายนอกขององค์กร “โดยจะเก็บข้อมูลเป็นระยะเวลานานและไม่มีการเปลี่ยนแปลงของชุดข้อมูลได้โดยง่าย ซึ่งเป็นข้อมูลสำหรับที่ใช้ช่วยในการสนับสนุนในการดำเนินการตัดสินใจเพื่อการบริหาร” (Roiger and Geatz, 2003, PP. 148)

นอกจากนี้ คลังข้อมูล เป็นระบบข้อมูลขนาดใหญ่ที่เอื้ออำนวยให้ผู้ใช้งานสามารถใช้ข้อมูลได้อย่างมีประสิทธิภาพ ข้อมูลที่จัดเก็บอยู่ในคลังข้อมูลต้องมีปริมาณเพียงพอ และเป็นข้อมูลที่มีคุณภาพเพียงพอต่อการนำไปวิเคราะห์ เพื่อหาคำตอบที่เหมาะสมทางธุรกิจการจัดเก็บข้อมูลต้องเอื้ออำนวยให้การใช้ข้อมูลเชิงวิเคราะห์เป็นไปได้ง่าย สามารถวิเคราะห์ข้อมูลได้อย่างรวดเร็วไม่ยุ่งยากซับซ้อน (กิตติพงษ์ กลมกล่อม 2552,น. 4)

2.2 กระบวนการคลังข้อมูล

กระบวนการต่างๆ ในคลังข้อมูลสามารถแบ่งออกเป็นกลุ่มใหญ่ๆ ได้ดังภาพที่ 2.2



ภาพที่ 2.2 กระบวนการคลังข้อมูล

2.2.1 กระบวนการจัดเก็บข้อมูล (Data Storing) คือ กระบวนการที่ดำเนินการเพื่อให้การจัดเก็บข้อมูลเป็นไปอย่างถูกต้องและมีประสิทธิภาพ ซึ่งกระบวนการเหล่านี้ ได้แก่ การสร้างแบบจำลองข้อมูล การสร้างฐานข้อมูล การจัดแบ่งพื้นที่บนฐานข้อมูล (Database Partitioning) การปรับแต่งฐานข้อมูล (Database Tuning) เป็นต้น

2.2.2 กระบวนการเคลื่อนย้ายข้อมูล (Data Propagation) คือกระบวนการเพื่อเคลื่อนย้ายข้อมูลจากที่หนึ่งไปยังอีกที่หนึ่ง เทคนิคที่ใช้ในคลังข้อมูลเรียกว่า กระบวนการอีทีแอล (Extract Transform Load : ETL)

2.2.3 กระบวนการเตรียมการเพื่อการใช้ข้อมูล (Data Provisioning) คือ กระบวนการนำเอาข้อมูลที่มีอยู่ในคลังข้อมูลมาประมวลผลให้ได้สารสนเทศที่ต้องการเพื่อรอการนำไปใช้งานต่อไป โดยอาจมีบางส่วนต้องถูกดัดแปลงหรือปรับเปลี่ยนเป็น ดาตามาร์ต (Data Mart) เพื่อประสิทธิภาพในการเรียกใช้ข้อมูล ออกรายงาน หรือวิเคราะห์ข้อมูลแบบต่างๆ เช่น รูปรายงาน หรือลูกบาศก์วิเคราะห์เชิงมิติ เป็นต้น

2.2.4 กระบวนการเพื่อการควบคุมการทำงานของระบบ (Controlling) คือ กระบวนการเพื่อควบคุมการทำงานของคลังข้อมูล เรียกว่า “เมทาดาตา (Meta Data)” โดยมีหน้าที่เก็บ ข้อมูลเกี่ยวกับ

คลังข้อมูล และสภาพแวดล้อมภายในคลังข้อมูล ซึ่งเมทาดาตาถูกออกแบบขึ้นมาเพื่อช่วยในการจัดการและดูแลรักษาคลังข้อมูลได้ง่ายขึ้น

2.3 วิธีการออกแบบคลังข้อมูล

วิธีการออกแบบนี้ถูกนำเสนอโดย Kimball ในปี 1996 เรียกว่าระเบียบวิธี 9 ขั้น หรือ Nine-Step Methodology โดยวิธีการนี้ เริ่มจากการออกแบบจากส่วนย่อยที่แสดงถึงแต่ละระบบงานขององค์กร หรือเรียกอีกอย่างหนึ่งว่า ดาต้ามาร์ท (Data Mart) โดยเมื่อออกแบบแต่ละส่วนสำเร็จแล้ว จึงนำมารวมกันเป็นคลังข้อมูลขององค์กรในขั้นสุดท้ายซึ่งขั้นตอนทั้ง 9 ขั้นตอน มีรายละเอียดในแต่ละขั้นตอนที่สำคัญดังนี้

2.3.1 กำหนดคลังข้อมูลย่อย (Data Mart) คือ การเลือกว่าจะสร้างคลังข้อมูลย่อยของระบบงานใดบ้าง และระบบงานใดเป็นระบบงานแรก โดยองค์กรจะต้องสร้าง E-R model ที่จะรวมระบบงานทุกระบบงานขององค์กรไว้แสดงการเชื่อมโยงแต่ละระบบงานอย่างชัดเจน และสิ่งที่ต้องคำนึงถึงในการเลือกระบบงานที่จะเป็นคลังข้อมูลย่อยแรกนั้น มี 3 ปัจจัยที่เกี่ยวข้อง ได้แก่ จะต้องสมารถพัฒนาออกมาได้ทันตามเวลาที่ต้องการ โดยอยู่ในกรอบงบประมาณที่กำหนดไว้ และต้องตอบปัญหาทางธุรกิจให้องค์กรได้ ดังนั้นคลังข้อมูลย่อยแรกควรจะเป็นของระบบงานที่นำรายได้เข้ามาสู่องค์กรได้ เช่น ระบบงานขาย เป็นต้น

2.3.2 กำหนดตารางข้อเท็จจริง (Fact Table) ของคลังข้อมูลย่อย คือ กำหนดเนื้อหาหลักที่ควรจะเป็นคลังข้อมูลย่อย โดยการเลือกตารางหลักๆ และกระบวนการที่เกี่ยวกับตารางนั้นๆ ออกมาจาก E-R model ขององค์กร นั้นหมายถึงว่าจะทำให้เราทราบถึงมิติขิงตารางที่ควรจะมีด้วย

2.3.3 กำหนดเขตข้อมูล (attribute) ที่จำเป็นในแต่ละตารางมิติ (Dimension Table) คือการกำหนด แอตทริบิวต์ที่บอกหรืออธิบายรายละเอียดของมิติได้ ทั้งนี้แอตทริบิวต์ที่เป็นคีย์หลัก (Primary key) ควรเป็นค่าที่คำนวณได้ กรณีที่มีดาต้ามาร์ทมากว่าหนึ่งดาต้ามาร์ทที่มีมิติเหมือนกัน นั้นหมายถึงว่า แอตทริบิวต์ในมิตินั้นจะต้องเหมือนกันทุกประการ แต่นั้นก็ไม่อาจจะแก้ไขปัญหาการจัดเก็บข้อมูลซ้ำซ้อน อันจะนำมาสู่ความแตกต่างกันของข้อมูลชุดเดียวกัน ปัญหานี้จึงเป็นการดีที่จะมีการใช้มิติตารางร่วมกันในแต่ละตารางที่จำเป็นจะต้องมีมิติดังกล่าว โดยเรียกมิติตารางลักษณะแบบนี้ว่า Conformed และเรียกตารางความจริงว่า Fact Constellation ทั้งนี้เราสามารถที่จะกำหนดข้อดีของการใช้มิติตารางร่วมกันได้

2.3.4 กำหนดเขตข้อมูลที่จำเป็นในตารางข้อเท็จจริง โดยเขตข้อมูลหลักในตารางข้อเท็จจริง ซึ่งจะมาจากกุญแจสำคัญในแต่ละตารางมิติ นอกจากนี้แล้วก็ยังสามารถมีเขตข้อมูลที่จำเป็นอื่นๆ ประกอบอยู่ด้วยเช่น เขตข้อมูลที่ได้จากการคำนวณค่าเบื้องต้น ที่มีความจำเป็นสำหรับการคงอยู่ของ

เขตข้อมูลอื่นๆ ในตารางข้อเท็จจริงหรือที่เรียกอีกอย่างหนึ่งว่าการวัด (Measure) การกำหนดเขตข้อมูลนี้ ไม่ควรเลือกเขตข้อมูลที่ไม่สามารถนำมาคำนวณได้ เช่น เป็นตัวหนังสือ หรือไม่ใช่ตัวเลข เป็นต้น และไม่ควรเลือกเขตข้อมูลที่ไม่เกี่ยวข้องกับเนื้อหาของตารางข้อเท็จจริงที่เราสนใจด้วย

2.3.5 จัดเก็บค่าจากการคำนวณเบื้องต้นในตารางข้อเท็จจริง คือ การจัดเก็บค่าที่ได้จากการคำนวณให้เป็นเขตข้อมูลในตารางข้อเท็จจริง ถึงแม้ว่าจะสามารถหาค่าได้จากเขตข้อมูลอื่นๆ ก็ตามทั้งนี้เพื่อให้การสอบถามมีประสิทธิภาพมากขึ้น และให้สามารถทำงานด้วยความรวดเร็วที่เพิ่มขึ้น เนื่องจากไม่ต้องคำนวณค่าใหม่ทั้งหมด ถึงแม้ว่าจะเกิดซ้ำซ้อนของข้อมูลในการจัดเก็บบ้างก็ตาม

2.3.6 เขียนคำอธิบายตารางมิติ ทั้งนี้ก็เพื่อให้ผู้ใช้สามารถใช้งานดาต้ามาร์ทได้อย่างมีประสิทธิภาพเพราะเกิดความเข้าใจอย่างดีในส่วนต่างๆ

2.3.7 กำหนดระยะเวลาที่ใช้ในการจัดเก็บข้อมูลในฐานข้อมูล โดยอาจเป็นการจัดเก็บเพียงช่วงระยะเวลา 1-2 ปีหรือนานกว่านั้นขึ้นอยู่กับความต้องการขององค์กร เนื่องจากองค์กรแต่ละประเภทมีความต้องการในการจัดเก็บข้อมูลต่างช่วงเวลากัน ทั้งนี้อาจขึ้นอยู่กับความจำเป็น หรือข้อกำหนดที่ใช้ในการดำเนินธุรกิจ มีข้อสังเกตอยู่ 2 ประการ ที่น่าสนใจและสำคัญสำหรับการออกแบบเขตข้อมูล ในเรื่องของการจัดเก็บ ดังนี้ ข้อมูลที่ถูกจัดเก็บไว้นานเกินมักเกิดปัญหาการอ่านหรือแปลข้อมูลนั้นๆ จากแฟ้มหรือเทปเก่า และเมื่อมีการนำรูปแบบเก่าของตารางมิติมาใช้อาจเกิดปัญหาการเปลี่ยนแปลงของมิติไปอย่างช้าๆ ได้

2.3.8 การติดตามปัญหาการเปลี่ยนแปลงของตารางมิติอย่างช้าๆ คือ การเปลี่ยนเอาเขตข้อมูลของตารางมิติเก่ามาใช้ แล้วส่งผลกระทบต่อข้อมูลปัจจุบันของตารางมิติ โดยสามารถแบ่งประเภทของปัญหาที่เกิดได้ ดังนี้คือ เกิดการเขียนทับข้อมูลใหม่ โดยข้อมูลเก่า ทำให้เกิดระเบียนใหม่ๆ ขึ้นในตารางมิติจะทำให้ระเบียนที่มีทั้งค่าเก่าและค่าใหม่ปะปนกันไป

2.3.9 การกำหนดคิวรี่ (Query) เป็นการออกแบบด้านกายภาพเพื่อให้ผู้ใช้เกิดความสะดวก ในการใช้งานและสามารถทำงานได้อย่างมีประสิทธิภาพ เมื่อได้ดำเนินการทั้ง 9 ขั้นตอนสำหรับแต่ละคลังข้อมูลย่อยเสร็จแล้ว จึงนำทั้งหมดมารวมกันเป็นภาพรวมของคลังข้อมูลขององค์กรต่อไป

2.4 การออกแบบจำลองข้อมูลเชิงมิติ

การออกแบบจำลองข้อมูลเชิงมิติ มีจุดมุ่งหมายที่จะแสดงความสัมพันธ์ที่มีอยู่จริงของข้อมูลต่างๆ ที่มีอยู่ในองค์กร แต่อย่างไรก็ตามแบบจำลองข้อมูลเชิงสัมพันธ์ นั้นในบางครั้งมีความซับซ้อนของความสัมพันธ์ค่อนข้างมาก ทำให้การค้นหาข้อมูลและคำนวณผลลัพธ์ที่ต้องการใช้ในการวิเคราะห์ระบบเป็นไปด้วยความยากลำบาก ในบางกรณีอาจต้องพัฒนาโปรแกรมที่ซับซ้อน

เพื่อดึงข้อมูล ดังนั้นเพื่อความสะดวกในการใช้ข้อมูล จึงมีการคิดค้นแบบจำลองข้อมูลแบบใหม่ที่เหมาะสมสำหรับการแสดงผลลัพธ์ที่ต้องการใช้ เรียกว่า การจำลองข้อมูลเชิงมิติ (Dimensional Modelling)

2.4.1 แบบจำลองเชิงมิติ

การจำลองข้อมูลเชิงมิติเป็นเทคนิควิธีการสร้างกรอบแนวคิด และจำลองรูปแบบกลุ่มข้อมูลที่นำมาวัดเพื่อใช้อธิบายในมุมมองธุรกิจ คลังข้อมูลถือได้ว่ามีประโยชน์อย่างยิ่งในการหาข้อมูลสรุปและมีการจัดเตรียมข้อมูลให้ใหม่ในการนำเสนอเพื่องานวิเคราะห์ที่มีความซับซ้อน โมเดลมิติได้มุ่งเน้นมิติ ตัวเลข จำนวน น้ำหนัก สมดุล และบันทึกของสิ่งที่เกิดขึ้น ซึ่งประกอบไปด้วยหลักการต่างๆ ต่อไปนี้

1) มิติหรือไดเมนชัน (Dimension) คือ เป็นที่รวบรวมสมาชิกของมุมมองประเภทเดียวกัน โดยไดเมนชันเป็นข้อมูลที่เป็นมุมมองของเมสเซอร์ (Measure) เพื่อประโยชน์ในการวิเคราะห์ข้อมูล การนำเสนอไดเมนชันในแผนภาพเป็นลักษณะแกนของมิติต่างๆ ที่รายล้อมแฟกต์ในแบบจำลองมิติ ทุกๆ จุดข้อมูลในตารางข้อเท็จจริง (Fact Table) มีความสัมพันธ์กับสมาชิกของแต่ละไดเมนชันเพียงหนึ่งเดียว

2) ข้อเท็จจริงหรือแฟกต์ (Fact) คือ กลุ่มของรายการข้อมูลที่สัมพันธ์ และสามารถใช้เพื่อการวัดประกอบด้วยเนื้อหาเชิงตัวเลขที่มาจากข้อมูลปฏิบัติงานประจำวัน ทุกตัวเลขในมิติข้อเท็จจริงสามารถใช้เพื่อการวิเคราะห์ลักษณะธุรกิจที่ดำเนินอยู่ได้ ในระบบคลังข้อมูลข้อเท็จจริงดำเนินการในตารางหลัก และทุกตัวเลขจัดเก็บลงในตารางข้อเท็จจริงนี้

3) ค่าที่ต้องการวัดหรือเมเชอร์ (Measure) คือ ข้อมูลที่ต้องการใช้เพื่อการวัดทั้งในเชิงปริมาณ (Quantitative) และเชิงคุณภาพ (Qualitative) ของสิ่งใดสิ่งหนึ่ง เช่น ยอดขายรวม กำไร จำนวนสินค้า โดยมากค่าที่ต้องการวัดจะเป็นข้อมูลที่เป็นตัวเลข

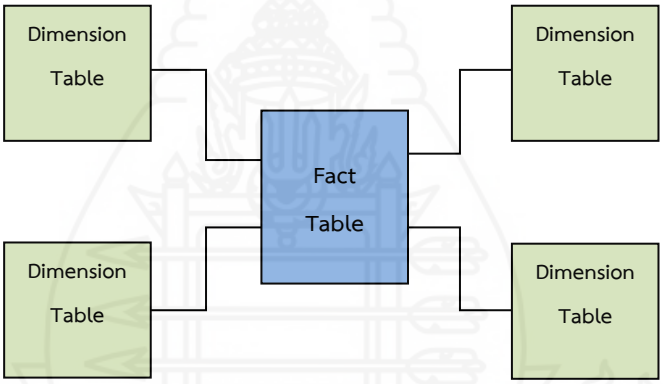
Dimension Measure

	วันที่	Ward	จำนวนครั้งที่ตาม	เวลารวมที่ตาม (นาที)
Fact ชุดที่ 1 →	1 ธ.ค. 2557	Trauma Surgery	50	311
Fact ชุดที่ 2 →	1 ธ.ค. 2557	Neurosurgery	45	538
Fact ชุดที่ 3 →	1 ธ.ค. 2557	Plastic surgery	30	432
Fact ชุดที่ 4 →	1 ธ.ค. 2557	Orthopedics	70	905

ภาพที่ 2.3 ตัวอย่างของแฟกต์ (Fact), ไดเมนชัน (Dimension) และเมเชอร์ (Measure)

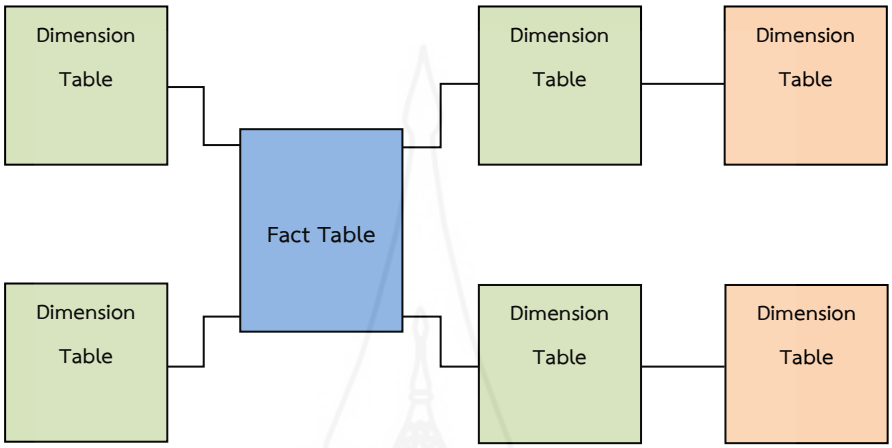
2.4.2 โครงสร้างแบบจำลองเชิงมิติ

1) โครงสร้างแบบสตาร์สกีมา (Star Schema) โครงสร้างชนิดนี้นิยมใช้ในการออกแบบคลังข้อมูล จะมีตารางข้อเท็จจริงเป็นศูนย์รวมของข้อมูลเพียงตารางเดียว และมีตารางมิติที่มีรายละเอียดของรหัสที่ใช้ในตารางข้อเท็จจริง ซึ่งตารางมิติจะมีจำนวนเท่าใดก็ได้ และมีคีย์ที่สัมพันธ์ไปยังตารางข้อเท็จจริงเท่านั้น โครงสร้างชนิดนี้จะช่วยเพิ่มความเร็วในการสืบค้นข้อมูล เนื่องจากความสัมพันธ์ระหว่าง ตารางไม่ซับซ้อน ซึ่งโครงสร้างแบบสตาร์สกีมานี้จะช่วยเพิ่มความสามารถในการคิวรี่ โดยลดปริมาณข้อมูลที่อ่านจากแหล่งข้อมูล ดังภาพที่ 2.4



ภาพที่ 2.4 โครงสร้างแบบสตาร์สกีมา (Star Schema)

2) โครงสร้างแบบสโนว์เฟลกสกีมา (Snowflake Schema) โครงสร้างแบบสโนว์เฟลกสกีมาเป็นแบบจำลองที่มีตารางข้อเท็จจริงขนาดใหญ่เพียงหนึ่งเดียวอยู่ตรงกลาง และมีตารางมิติจำนวนหนึ่งอยู่รายรอบเพื่อกำหนดมุมมองที่จะมีต่อค่าที่ต้องการวัด ในตารางข้อเท็จจริงนั้น โดยจำนวนมุมมองที่มองได้จะเท่ากับจำนวนตารางมิติที่รายล้อมอยู่ แต่จะมากกว่าจำนวนมิติที่เชื่อมต่อโดยตรงกับตารางข้อเท็จจริง และมิติที่ไม่ได้เชื่อมต่อโดยตรงกับตารางข้อเท็จจริงจะมีความสัมพันธ์กับมิติตัวอื่น ลักษณะโครงสร้างคล้ายกับโครงสร้างแบบสตาร์สกีมา แต่มีการแตกตารางมิติออกเป็นตารางย่อย เนื่องจากไม่สามารถสร้างโครงสร้างแบบสตาร์สกีมาได้ส่งผลให้ประสิทธิภาพในการคิวรี่ลดลงเมื่อเปรียบเทียบกับโครงสร้างแบบสตาร์สกีมา ดังภาพที่ 2.5



ภาพที่ 2.5 โครงสร้างแบบสโนว์เฟลกสกีมา (Snowflake Schema)

2.5 กระบวนการนำข้อมูลเข้าคลังข้อมูล

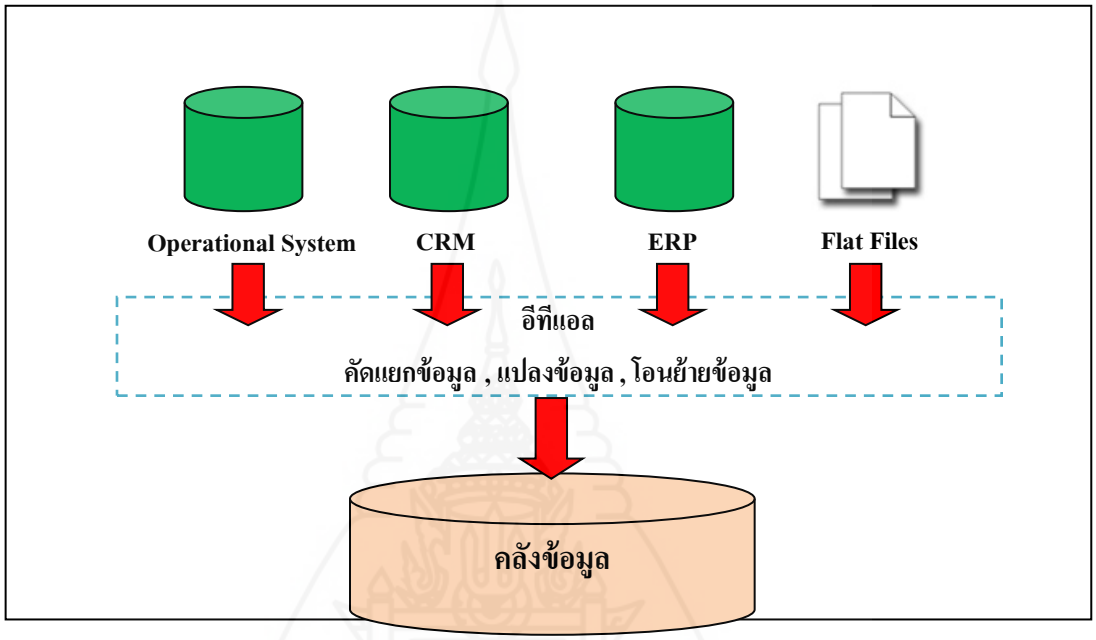
กระบวนการนำข้อมูลเข้าคลังข้อมูลเป็นกระบวนการที่ทำหน้าที่แปลงและเคลื่อนย้ายข้อมูลจากที่หนึ่งไปยังอีกที่หนึ่งในคลังข้อมูลหรือที่เรียกว่า กระบวนการอีทีแอล (Extract Transform Load: ETL) โปรแกรมประยุกต์ในปัจจุบันสามารถดำเนินการได้กับแหล่งที่มาหลายประเภท ชุดของการดำเนินงานของสถาปัตยกรรมคลังข้อมูลที่เกิดขึ้นในระยะหลังเป็นที่รู้กันในชื่อว่า การคัดแยกข้อมูล (Extraction) การแปลงข้อมูล (Transformation) และการโอนย้ายข้อมูล (Loading) กระบวนการอีทีแอลมีหน้าที่คัดแยกข้อมูลที่มาจากต่างแหล่งที่มาทำความสะอาดข้อมูลและปรับแต่งข้อมูลให้ถูกต้องสมบูรณ์ เพื่อให้เหมาะสมกับความต้องการ จากนั้นจึงโอนย้ายข้อมูลเข้าสู่คลังข้อมูล ในแต่ละขั้นตอนของกระบวนการอีทีแอล มีรายละเอียดดังนี้

2.5.1 การคัดแยกข้อมูล (Extract) คือ การดึงข้อมูลจากแหล่งข้อมูลที่แตกต่างกัน เช่น ฐานข้อมูลปฏิบัติการ ฐานข้อมูลนอกองค์กร เป็นต้น ขั้นตอนนี้มีความสำคัญมากเนื่องจากเป็นขั้นตอนที่ดึงข้อมูลที่ถูกต้องมีผลให้การดำเนินการในขั้นตอนต่อไปมีความถูกต้องด้วย

2.5.2 การแปลงข้อมูล (Transform) คือ การนำข้อมูลที่ได้มาจากการคัดแยกข้อมูลมาจัดรูปแบบให้ถูกต้องสอดคล้องกัน เช่น การจับคู่ข้อมูล (Data Mapping) การทำให้ข้อมูลที่มี

ความหมายเดียวกันแต่อยู่ในรูปแบบที่แตกต่างกันให้อยู่ในรูปแบบเดียวกัน การทำความสะอาดข้อมูล (Data Cleansing) การตรวจสอบและแก้ไขข้อมูลให้ถูกต้อง

2.5.3 การโอนย้ายข้อมูล (Load) คือ การนำข้อมูลที่ผ่านการแปลงข้อมูลแล้วเข้าสู่คลังข้อมูล ซึ่งมีการกำหนดช่วงเวลา ความถี่ในการโหลดข้อมูล

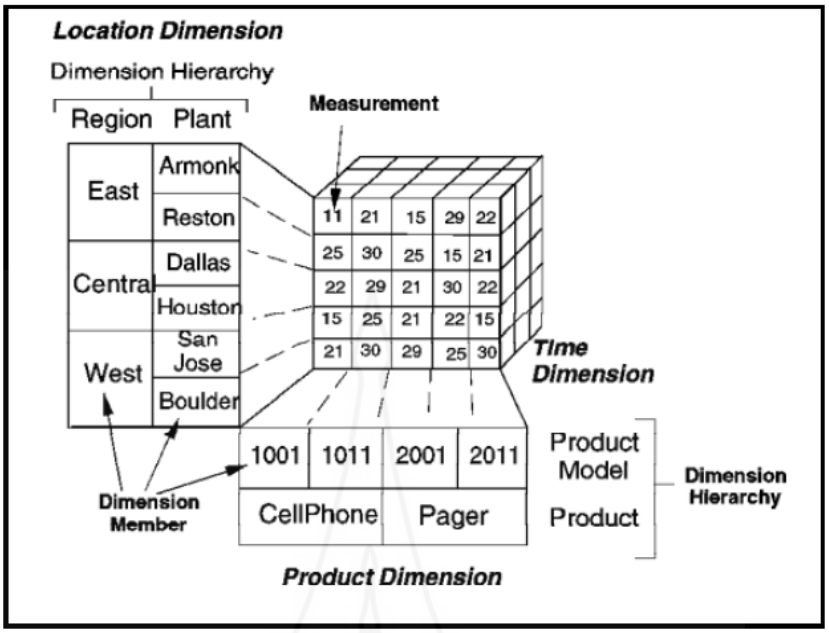


ภาพที่ 2.6 กระบวนการ ETL

2.6 การวิเคราะห์ข้อมูลในคลังข้อมูล

ในการวิเคราะห์ข้อมูลของคลังข้อมูลจะมีการดำเนินการที่นิยมทำอยู่ 3 รูปแบบ คือ รายงานและสอบถาม (Report and Query), การประมวลผลเชิงวิเคราะห์แบบออนไลน์ (Online Analytic Process: OLAP) และการทำเหมืองข้อมูล (Data Mining) โดยงานวิจัยนี้จะกล่าวถึงการประมวลผลเชิงวิเคราะห์แบบออนไลน์

การประมวลผลเชิงวิเคราะห์แบบออนไลน์ (Online Analytic Process: OLAP) เป็นเทคโนโลยีที่ประกอบด้วยเครื่องมือที่ช่วยดึงและนำเสนอข้อมูลในหลายมิติ (Multidimensional) ที่ประกอบไปด้วยมิติ (Dimension) และตารางข้อเท็จจริง (Fact Table) ในตารางข้อเท็จจริงก็จะประกอบไปด้วย Measure เป็นข้อมูลที่ต้องการวัดในเชิงปริมาณ กับข้อมูลที่ใช้เชื่อมความสัมพันธ์กับมิติ โดยที่ข้อมูลจะถูกเก็บอยู่ในรูปแบบของลูกบาศก์ที่มีหลายมิติเรียกว่า คิวบ์ (Cube) ดังภาพที่ 2.7



ภาพที่ 2.7 ลักษณะของคิวบ์

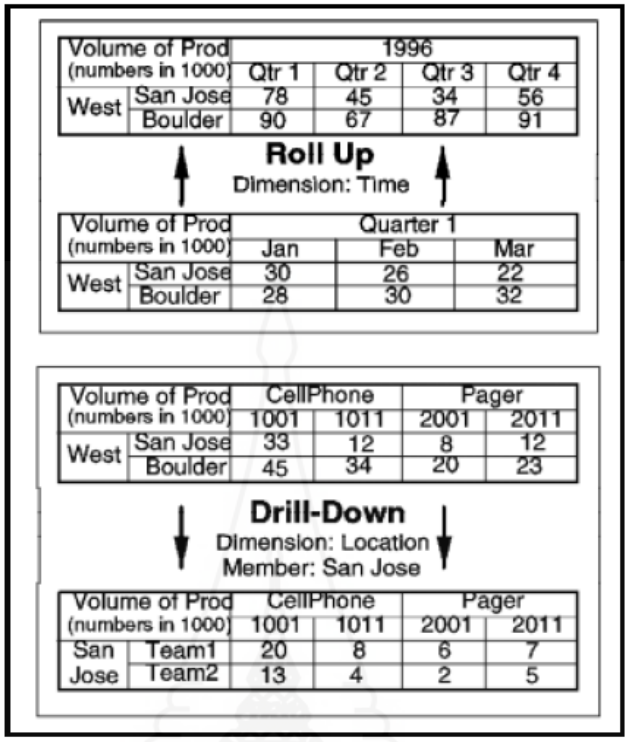
ที่มา: Ballard and others (1998: 44)

โดยที่การประมวลผลเชิงวิเคราะห์ข้อมูลเพื่อใช้ประกอบการตัดสินใจในระดับสูง ซึ่งเครื่องมือที่ช่วยในการดำเนินการกับการประมวลผลเชิงวิเคราะห์แบบออนไลน์นี้สามารถอำนวยความสะดวก โดยผู้ใช้งานไม่จำเป็นต้องมีความรู้ด้านคลังข้อมูลก็สามารถใช้งานได้ ซึ่งวิธีในการดำเนินการประมวลผลออนไลน์เชิงวิเคราะห์ ประกอบด้วย 2 คุณสมบัติ ดังนี้

2.6.1 การพิจารณาข้อมูลลงไป (Drill Down) และข้อมูลขึ้นไป (Roll Up) กระบวนการเข้าถึงข้อมูลโดยการเปลี่ยนแปลงความละเอียดของการพิจารณาข้อมูล ทั้งคู่มีจุดประสงค์คล้ายคลึงกันซึ่งสามารถอธิบายได้ ดังนี้

1) การพิจารณาข้อมูลลงไป คือ การวิเคราะห์แบบเจาะลึก โดยดูข้อมูลเริ่มต้นจากข้อมูลแบบสรุปรวม ซึ่งกระบวนการพิจารณารายละเอียดข้อมูลลงไป ด้วยระดับลึกตามระดับชั้นของความสัมพันธ์ข้อมูลที่เกี่ยวข้อง เช่น ข้อมูลสรุปลูกค้ารายปี แล้วค่อยดูข้อมูลในระดับรายละเอียด

2) การพิจารณาข้อมูลขึ้นไป คือ การวิเคราะห์แบบเป็นภาพรวมจะปรับระดับความละเอียดของข้อมูลจากระดับที่ละเอียดขึ้นมาสู่ข้อมูลที่หยาบขึ้น โดยเริ่มดูข้อมูลจากส่วนรายละเอียด แล้วค่อยดูข้อมูลแบบสรุปรวมทั้งหมด เป็นกระบวนการพิจารณารายละเอียดข้อมูลขึ้นไป จากจุดที่เล็กที่สุดไปหาจุดที่ใหญ่ที่สุดของความสัมพันธ์ข้อมูลนั้น



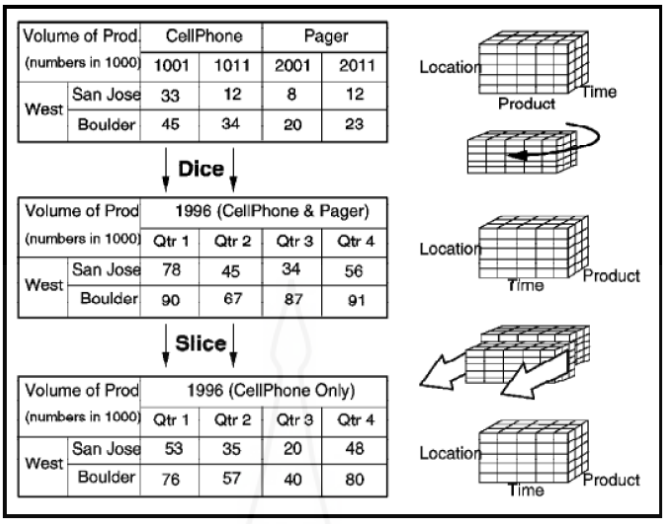
ภาพที่ 2.8 การ Drill Down และ Roll Up มิติข้อมูล

ที่มา: Ballard and others (1998: 45)

2.6.2 การนำเสนอข้อมูลบางส่วน (Slide) และการหมุนลูกบาศก์มิติ (Dice) วิธีการค้นหาข้อมูลโดยมีการปรับปรุงกลวิธีการค้นหาข้อมูลในลูกบาศก์ ซึ่งมีข้อแตกต่างระหว่างกันมีดังนี้

1) การนำเสนอข้อมูลบางส่วน คือ การวิเคราะห์แบบการพิจารณาผลลัพธ์บางส่วน เป็นการพิจารณาเฉพาะบางส่วนของคิวบ์ที่สนใจเพราะในบางครั้ง ถ้าคิวบ์มีขนาดใหญ่มากการพิจารณาข้อมูลทั้งหมดอาจจะทำให้ยากต่อการพิจารณา การวิเคราะห์โดยการแบ่งคิวบ์ออกมาเฉพาะบางส่วนจะช่วยให้ง่ายขึ้นต่อการพิจารณา

2) การหมุนลูกบาศก์มิติ คือ การวิเคราะห์แบบการพลิกแกน เป็นการพลิกแกนหรือมุมมองให้ตรงตามความต้องการของผู้ใช้ได้อย่างอิสระ เช่น การสลับข้อมูลระหว่างคอลัมน์และแถว เป็นต้น



ภาพที่ 2.9 การ Slide และ Dice มิติข้อมูล

ที่มา: Ballard and others (1998: 46)

3. การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล (Data Mining) ถือว่าเป็นศาสตร์หนึ่งที่มีความสำคัญในการค้นหาความรู้ (Knowledge Discovery) จากฐานข้อมูลหรือคลังข้อมูลที่มีอยู่ในหน่วยงาน เพื่อมาช่วยในการตัดสินใจหรือในการใช้งานด้านอื่นๆ การทำเหมืองข้อมูลสามารถค้นพบความรู้ที่ซ่อนเร้นอยู่ในข้อมูลจำนวนมาก

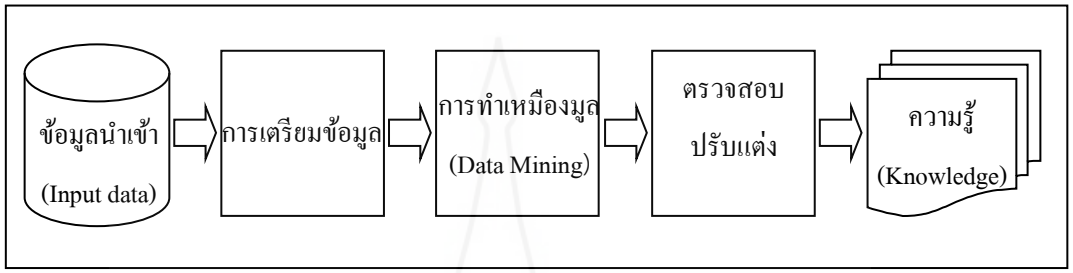
3.1 ความหมายของการทำเหมืองข้อมูล

การทำเหมืองข้อมูล คือ กระบวนการค้นหาความรู้จากข้อมูลที่มีอยู่จำนวนมากซึ่งอาจพบความสัมพันธ์ระหว่างข้อมูล ที่ถูกซ่อนไว้ภายในข้อมูลที่มีจำนวนมาก ด้วยกระบวนการและวิธีการที่เหมาะสม เพื่อนำความรู้ที่ได้ไปใช้ประโยชน์ตามวัตถุประสงค์ของการทำเหมืองข้อมูล

เอกสิทธิ์ พัชรวงศ์ศักดา (2557: 9) กล่าวว่า เหมืองข้อมูล (Data Mining) เป็นการค้นหาสิ่งที่มีประโยชน์จากฐานข้อมูลที่มีขนาดใหญ่ เช่น ข้อมูลซื้อขายสินค้าในซุปเปอร์มาร์เก็ตต่างๆ ซึ่งข้อมูลนี้จะเก็บรายการสินค้าที่ลูกค้าซื้อในแต่ละครั้งโดยเมื่อทำการวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมน์นิง แล้วจะได้สิ่งที่มีประโยชน์

ในการทำเหมืองข้อมูลจึงต้องอาศัยวิธีการทางคอมพิวเตอร์ร่วมกับวิธีการทางสถิติเพื่อสร้างตัวแบบหรือแบบจำลองโมเดล (Modelling) ซึ่งการสร้างแบบจำลองเหล่านี้มีกฎเกณฑ์รูปแบบในการดำเนินการเพื่อให้ได้ความรู้แบบจำลองเหล่านี้อาจเป็นตัวแบบที่เรียบง่ายไปจนถึงที่

ซับซ้อน เพื่อที่จะสามารถดึงเอาความรู้ที่อยู่ในฐานข้อมูลขนาดใหญ่ โดยผู้วิเคราะห์ไม่ต้องระบุรูปแบบจำลองด้วยตนเอง มีการพยากรณ์หรือคาดการณ์ผลลัพธ์จากข้อมูลให้ได้ ดังภาพที่ 2.10 เป็นการค้นหาความรู้ด้วยการทำเหมืองข้อมูล (วิภา เจริญภัณฑารักษ์, 2555: 8)

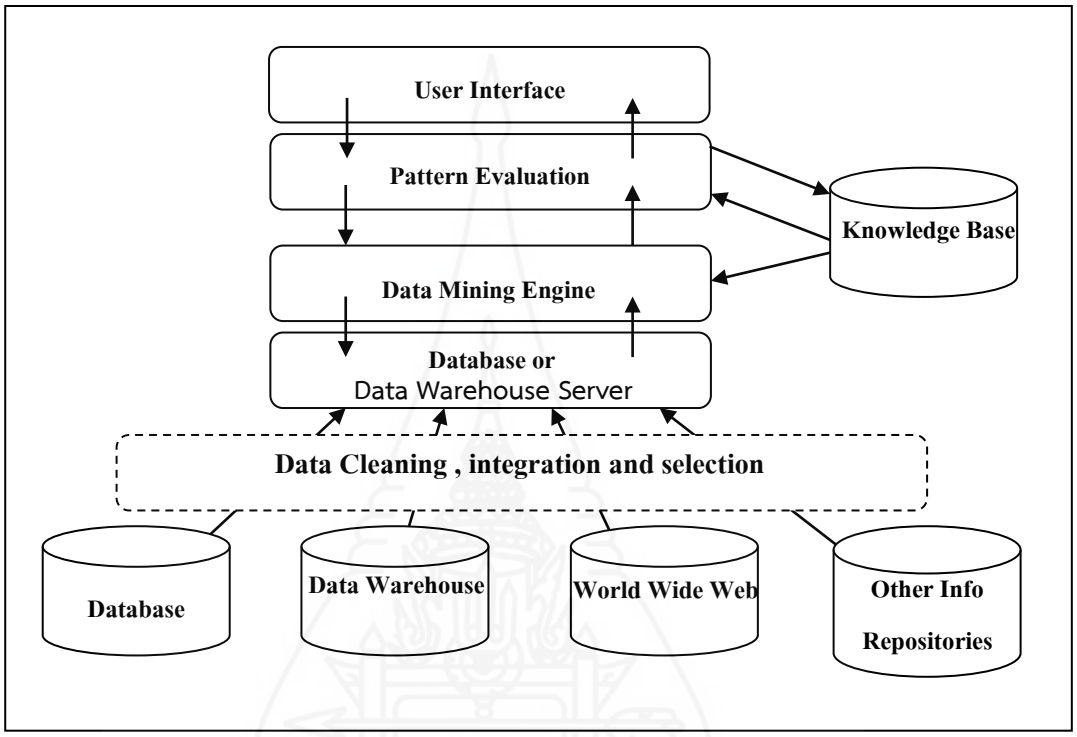


ภาพที่ 2.10 การค้นหาความรู้ด้วยการทำเหมืองข้อมูล

ที่มา: วิภา เจริญภัณฑารักษ์ (2555: 8)

3.2 สถาปัตยกรรมระบบเหมืองข้อมูล

ในการทำเหมืองข้อมูลนั้นมีสถาปัตยกรรมของการทำเหมืองข้อมูล ซึ่งประกอบด้วยส่วนต่าง ดังภาพที่ 2.11



ภาพที่ 2.11 สถาปัตยกรรมเหมืองข้อมูล

ที่มา: http://www.ijser.org/paper/Over-viewing-issues-of-data-mining-with-highlights-of- data-warehousing.html

3.2.1 แหล่งข้อมูลในการทำเหมืองข้อมูล (Data sources) คือ แหล่งข้อมูลที่จัดเก็บไว้ในฐานข้อมูล คลังข้อมูล เวิลด์ไวด์เว็บ (WWW) และคลังจัดเก็บข้อมูลประเภทอื่นๆ ในการดำเนินการเบื้องต้นเพื่อให้ข้อมูลพร้อมใช้งาน เริ่มตั้งแต่

1) การทำความสะอาดข้อมูล (Data cleansing) คือ การจัดการข้อมูลที่ไม่สมบูรณ์ ข้อมูลที่ขาดหาย ข้อมูลที่ขัดแย้งกันเอง เพื่อให้ข้อมูลสามารถใช้งานได้

2) การบูรณาการข้อมูล (Data integration) คือ การนำข้อมูลจากแหล่งข้อมูลหลายแหล่ง ที่มีความสัมพันธ์กันมารวมไว้ในที่เดียวกัน จำเป็นต้องมีการปรับรูปแบบให้เหมือนกันเพื่อสามารถทำงานร่วมกันได้

3) การคัดเลือกข้อมูล (Data selection) คือ การนำข้อมูลที่ต้องการมาใช้งาน หรือเลือกเฉพาะข้อมูลที่ต้องการเพื่อเก็บเป็นแหล่งข้อมูล

3.2.2 คลังข้อมูลหรือเซิร์ฟเวอร์ฐานข้อมูล (Data warehouse server) คือ ระบบแม่ข่ายที่ใช้ในการจัดเก็บข้อมูล ในการทำเหมืองข้อมูล

3.2.3 ฐานความรู้ (Knowledge base) คือ ความรู้ความเข้าใจในเรื่องที่ต้องการวิเคราะห์หรือค้นหาความรู้ความเข้าใจเกิดจากการเรียนรู้และประสบการณ์ในเรื่องดังกล่าว มีความเข้าใจในธรรมชาติของข้อมูลที่มีอยู่ และความรู้ในการเลือกวิธีการทำเหมืองข้อมูลได้อย่างเหมาะสมกับลักษณะงานหรือข้อมูลที่ดำเนินการ

3.2.4 กลไกการทำเหมืองข้อมูล (Data mining engine) คือ วิธีการทำเหมืองข้อมูล ได้แก่ การกำหนดคุณสมบัติข้อมูล การวิเคราะห์ความสัมพันธ์ของข้อมูล การจำแนกหมวดหมู่ การจัดกลุ่ม การค้นหาความผิดปกติของข้อมูล และเครื่องมือที่ใช้ในการทำเหมืองข้อมูลแต่ละประเภท เช่น วิธีการทางสถิติ

3.2.5 ส่วนประเมินรูปแบบ (Pattern evaluation module) คือ ในการสร้างแบบจำลองหรือตัวแบบโมเดล (Model) เพื่อแสดงถึงสมมติฐานและสามารถค้นพบความรู้ จากแบบจำลองที่ทำขึ้นมาได้ ยังต้องใช้การตรวจสอบและประเมินแบบจำลองที่จัดทำขึ้นมาด้วย เพื่อให้ผลลัพธ์ที่ได้มีความน่าเชื่อถือ เหมาะสม และถูกต้องมากที่สุด

3.2.6 ส่วนติดต่อผู้ใช้งาน (User interface) คือ ส่วนที่ผู้ใช้สามารถใช้งานซึ่งแบ่งออกเป็น ส่วนรองรับการป้อนคำสั่งข้าวของผู้ใช้งาน เช่น การค้นหา การใส่เงื่อนไขการทำเหมืองข้อมูล และส่วนนำเสนอผลลัพธ์ที่ได้จากการทำเหมืองข้อมูลในรูปแบบต่างๆ เช่น ตาราง แผนภูมิ และรูปแบบความสัมพันธ์ของข้อมูลที่ค้นหา

3.3 ประเภทการทำเหมืองข้อมูล

การทำเหมืองข้อมูลเป็นการค้นหารูปแบบที่เกิดขึ้นกับข้อมูลที่เราสนใจ ซึ่งแบ่งตามลักษณะงานที่ดำเนินการ (Data mining operation/functionality) เป็นประเภทต่างๆ ได้ดังนี้

3.3.1 การค้นหาคุณลักษณะหรือรายละเอียดของข้อมูล เป็นการค้นหาข้อมูลลักษณะเฉพาะที่เราสนใจหรือต้องการ เพื่อเข้าใจในรายละเอียดที่เกี่ยวข้องกับข้อมูลนั้น ในการทำเหมืองข้อมูลประเภทนี้ คือ การกำหนดคุณลักษณะเฉพาะข้อมูล (Data characterization) และการกำหนดคุณลักษณะแตกต่าง (Data discrimination)

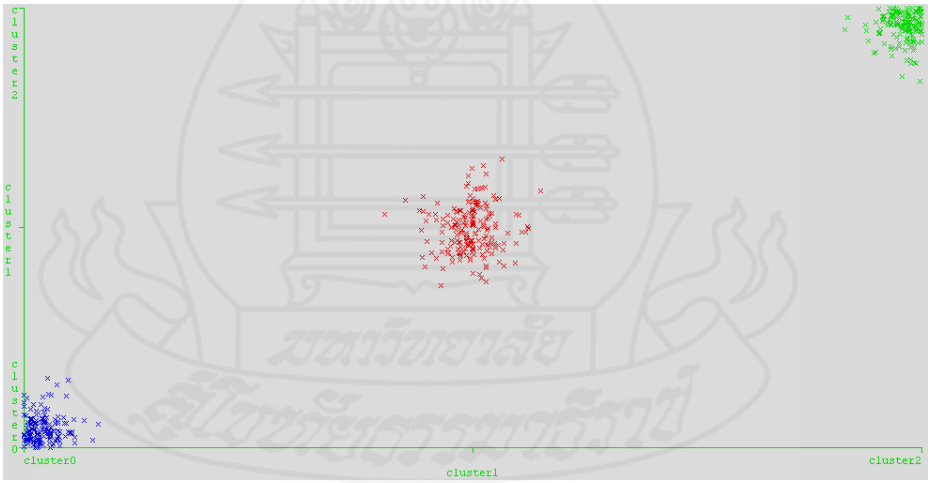
3.3.2 การค้นหาความสัมพันธ์ เป็นการทำเหมืองข้อมูลเพื่อหาความสัมพันธ์ซึ่งประกอบไปด้วย การค้นหารูปแบบที่เกิดขึ้นคู่กัน และการสร้างกฎความสัมพันธ์ที่เกิดขึ้น

3.3.3 การจำแนกประเภทและการพยากรณ์ ในการทำเหมืองข้อมูลประเภทนี้สามารถแบ่งได้ดังนี้

1) การจำแนกประเภท (Classification) เป็นการจำแนกข้อมูลที่ต้องการว่าอยู่ประเภทไหน โดยอาศัยการแบบจำลองหรือตัวแบบที่สร้างขึ้นมา ซึ่งใช้ข้อมูลที่มีอยู่จำนวนหนึ่งในการสร้างและใช้ข้อมูลอีกจำนวนหนึ่งในการทดสอบแบบจำลองนั้นเพื่อตรวจสอบความเที่ยงตรงหรือความน่าเชื่อถือของแบบจำลองที่สร้างขึ้น ข้อมูลในการจำแนกประเภทจะเป็นแบบที่มีค่าไม่ต่อเนื่อง เช่น ข้อมูลที่เป็นการตัดสินใจ ใช่ หรือไม่ใช่ เป็นต้น

2) การพยากรณ์ (Prediction) จะคล้ายกับการจำแนกประเภทแต่ข้อมูลที่ใช้มีลักษณะเป็นค่าต่อเนื่องหรือข้อมูลที่มีการลำดับการเกิดของข้อมูล เช่น ข้อมูลอายุ เงินเดือน เวลา รายได้จากการขายสินค้า เป็นต้น

3.3.4 การจัดกลุ่ม คือ การทำเหมืองข้อมูลประเภทจัดกลุ่ม (Clustering) เป็นเทคนิคจัดกลุ่มสมาชิกที่มีความคล้ายคลึงกันสูงสุด โดยไม่ได้กำหนดกลุ่มไว้ล่วงหน้า เพื่อจำแนกข้อมูลในแต่ละหน่วยของชุดข้อมูล ใช้การวัดค่าความห่างจากจุดศูนย์กลางและการวัดค่าเบี่ยงเบนเป็นการรวมหน่วยที่คล้ายคลึงกันมากที่สุดไว้เป็นกลุ่มเดียวกัน ดังภาพที่ 2.12



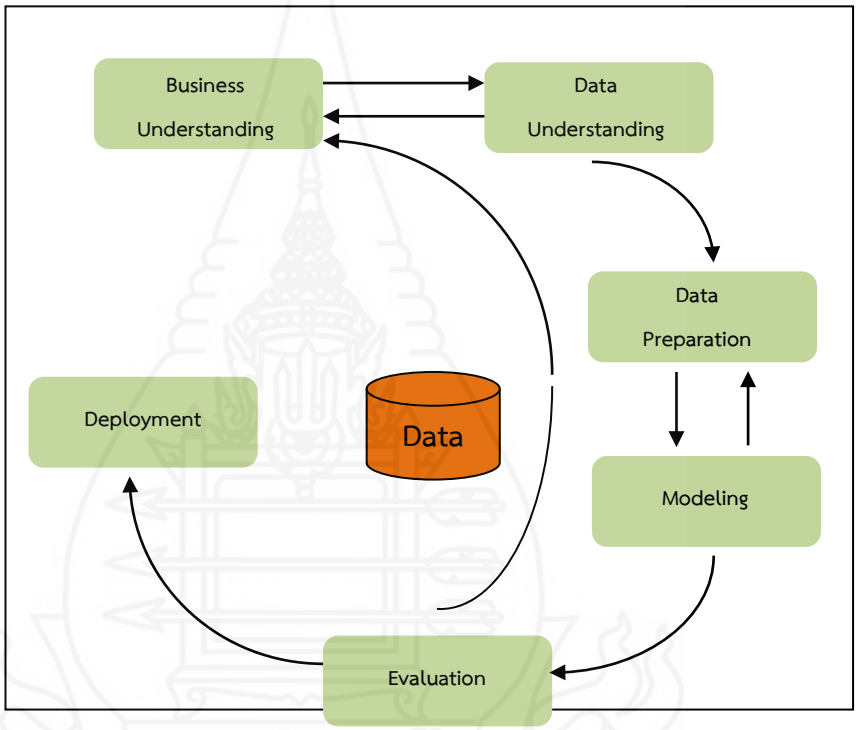
ภาพที่ 2.12 ตัวอย่างข้อมูลที่ถูกจัดกลุ่มได้ 3 กลุ่ม

3.3.5 การค้นหาหน่วยข้อมูลที่ผิดปกติ เป็นการทำเหมืองข้อมูลเพื่อค้นหาหน่วยข้อมูลที่แตกต่างหรือผิดแปลกจากคุณสมบัติข้อมูลทั่วไปที่ถูกเก็บไว้ในฐานข้อมูล สาเหตุของความผิดปกติก็เนื่องมาจากการบันทึกข้อมูล หรือความผิดพลาดอื่นๆ เมื่อพบข้อมูลที่ผิดปกติก็จะจัดการกับข้อมูลดังกล่าวอย่างเหมาะสมเพื่อใช้ในการทำเหมืองข้อมูล

3.3.6 การวิเคราะห์การเปลี่ยนแปลงที่อิงตามกาลเวลา เป็นการทำเหมืองข้อมูลเพื่อค้นหารูปแบบข้อมูลที่มีการเปลี่ยนแปลงตามเวลา (Evolution analysis) ตัวอย่างเช่น การวิเคราะห์การเปลี่ยนแปลงของราคาหุ้นแต่ละประเภทเพื่อการลงทุนที่เหมาะสม เป็นต้น

3.4 กระบวนการทำเหมืองข้อมูล

การบวนการในการทำเหมืองข้อมูลมีกระบวนการวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล ด้วยกระบวนการมาตรฐานที่เรียกว่า “Cross-Industry Standard Process for Data Mining” หรือเรียกย่อๆว่า “CRISP-DM Model” ซึ่งมีกระบวนการทำงานดังภาพที่ 2.13



ภาพที่ 2.13 กระบวนการ CRISP-DM Model

ที่มา: Umair Shafique and Haseeb Qaiser (2014: 219)

3.4.1 การวิเคราะห์ปัญหาและโอกาสทางธุรกิจ (Business Understanding) เป็นขั้นตอนการทำความเข้าใจ ระบุปัญหาหรือโอกาสในเชิงธุรกิจ เพื่อหารูปแบบที่เหมาะสมต่อการนำมาวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูล

3.4.2 การทำความเข้าใจข้อมูล (Data Understanding) เป็นขั้นตอนการรวบรวมข้อมูลที่เกี่ยวข้อง จากแหล่งข้อมูลที่ถูกต้องน่าเชื่อถือ ซึ่งข้อมูลที่ได้มีปริมาณมากพอ และเป็นข้อมูลที่เหมาะสมเพียงพอต่อการนำไปใช้ในวิเคราะห์ข้อมูล

3.4.3 การเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนการเตรียมข้อมูลเป็นขั้นตอนที่ใช้เวลานานที่สุด เนื่องจากโมเดลที่ได้จากการทำเหมืองข้อมูลจะให้ผลลัพธ์ที่ถูกต้องหรือไม่ขึ้นอยู่กับคุณภาพของข้อมูล

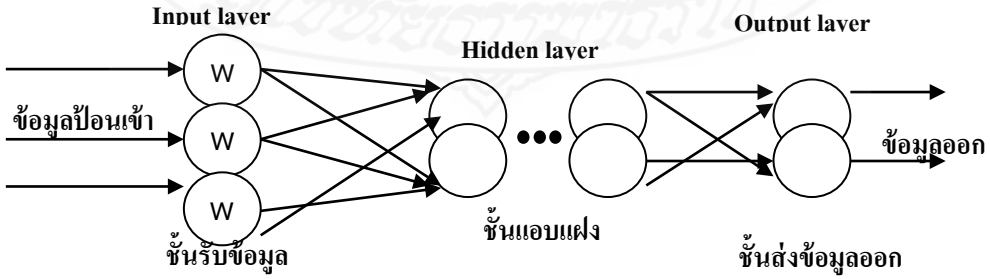
3.4.4 การสร้างโมเดล (Modeling) เป็นขั้นตอนการนำเทคนิคเหมืองข้อมูลมาใช้ในการวิเคราะห์ข้อมูล เพื่อสร้างโมเดลให้ได้ผลลัพธ์ที่ดีที่สุด

3.4.5 การวัดประสิทธิภาพ (Evaluation) เป็นขั้นตอนการประเมินหรือวัดประสิทธิภาพ ของโมเดลวิเคราะห์ข้อมูลในขั้นตอนก่อนหน้านั้น เพื่อวัดความน่าเชื่อถือของโมเดล

3.4.6 การนำโมเดลที่พัฒนาไปใช้งาน (Deployment) เป็นขั้นตอนในการนำโมเดลที่ได้จากการวิเคราะห์ข้อมูลไปใช้งานจริง เพื่อช่วยในการตัดสินใจ

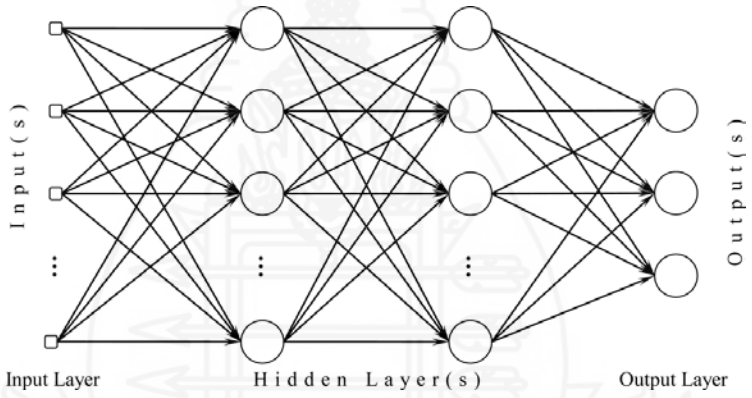
3.5 เทคนิคการทำเหมืองข้อมูล

3.5.1 เทคนิคโครงข่ายประสาทเทียม (Neural Networks) มีลักษณะคล้ายกับการส่งสัญญาณประสาทของสมองมนุษย์ โครงข่ายประสาทเทียมจะประกอบด้วยเซลล์ประสาทเทียมหรือโหนด (Node) จำนวนมากเชื่อมต่อกัน ในการเชื่อมต่อนั้นแบ่งออกเป็นกลุ่มย่อย เรียกว่า ชั้น (Layer) ชั้นแรกเป็นชั้นข้อมูลนำเข้าบางที่เรียกว่า ชั้นรับข้อมูล (Input layer) ส่วนชั้นสุดท้าย เรียกว่า ชั้นส่งข้อมูลออก (Output layer) และชั้นที่อยู่ระหว่างชั้นรับข้อมูลและขั้นส่งข้อมูลออก เรียกว่า ชั้นซ่อนหรือแอบแฝง (Hidden layer) โครงข่ายประสาทเทียมที่มีโครงข่ายแบบหลายชั้น (Multilayer) จะเป็นโครงข่ายที่มีชั้นแอบแฝงตั้งแต่หนึ่งชั้นขึ้นไป โครงข่ายแบบหลายชั้นจะใช้เมื่อมีปัญหาความซับซ้อน เพื่อเพิ่มจำนวนโหนดที่มีการคำนวณ ลักษณะของโครงข่ายแบบหลายชั้นเป็นดังภาพที่ 2.14



ภาพที่ 2.14 โครงข่ายประสาทเทียมแบบหลายชั้น

โครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (Multilayer Perceptron : MLP) เป็นรูปแบบหนึ่งของโครงข่ายประสาทเทียมที่มีโครงสร้างเป็นแบบชั้นใช้สำหรับงานที่มีความซับซ้อนซึ่งได้ผลเป็นอย่างดี โดยมีกระบวนการฝึกฝนเป็นแบบ Supervise และใช้ขั้นตอนการส่งค่าย้อนกลับ (Backpropagation) สำหรับในการฝึกฝนจะประกอบด้วย 2 ส่วนย่อย คือ การส่งผ่านไปข้างหน้า (Forward Pass) และการส่งผ่านย้อนกลับ (Backward Pass) สำหรับการส่งผ่านไปข้างหน้าข้อมูลจะผ่านเข้าชั้นของข้อมูลเข้า และจะส่งผ่านจากอีกชั้นหนึ่งไปสู่อีกชั้นหนึ่งจนกระทั่งถึงชั้นข้อมูลออก ส่วนการส่งย้อนกลับค่าน้ำหนักการเชื่อมต่อจะถูกปรับเปลี่ยนให้สอดคล้องกับกฎการแก้ข้อผิดพลาด คือ ผลต่างของผลตอบที่แท้จริงกับผลตอบเป้าหมายเกิดเป็นสัญญาณผิดพลาด ซึ่งจะถูกส่งย้อนกลับเข้าสู่โครงข่ายประสาทเทียมในทิศทางตรงกันข้ามกับการเชื่อมต่อ ค่าน้ำหนักการเชื่อมต่อจะถูกปรับจนกระทั่งผลตอบที่แท้จริงเข้าใกล้ผลตอบเป้าหมาย ดังภาพที่ 2.15

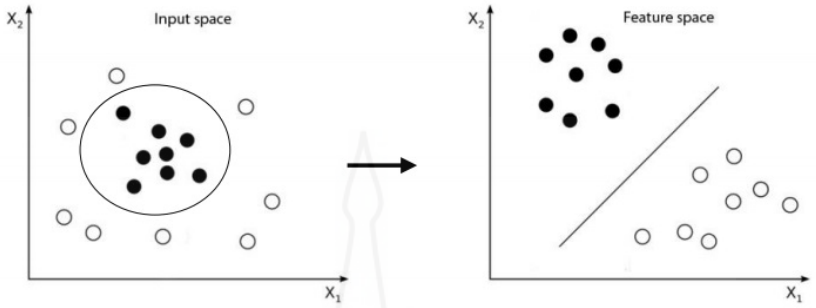


ภาพที่ 2.15 โครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน

ที่มา: http://www.mdpi.com/1424-8220/10/9/8363/htm

3.5.2 เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (*Support Vector *Machines) เป็นเทคนิคหนึ่งของโครงข่ายประสาทเทียม ที่มาช่วยแก้ปัญหาการวิเคราะห์และจำแนกข้อมูล โดยใช้หลักการสร้างระนาบหรือเส้นในการตัดสินใจแบ่งข้อมูลออกมาเป็น 2 กลุ่ม โดยพยายามสร้างเส้นแบ่งให้อยู่กึ่งกลางระหว่างสองกลุ่มให้มากที่สุด อาศัยการหาระยะทางจากเส้นไปยังข้อมูลที่อยู่ใกล้เส้นแบ่งมากที่สุดของข้อมูลในแต่ละประเภท (กชกร ณ นครพนม, 2555,น. 9-13) ในการสร้างสมการเส้นตรงของการแบ่งกลุ่มจะสร้างไฮเปอร์เพลน (Hyperplane) ที่เป็นเส้นตรงขึ้นมา และใช้

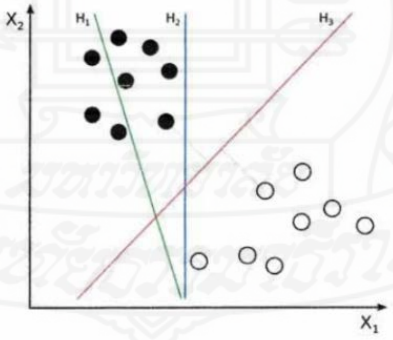
ฟังก์ชันแม็ปสำหรับย้ายข้อมูลจากอินพุทสเปซ (Input Space) ไปยังฟีเจอร์สเปซ (Feature Space) และสร้างฟังก์ชันวัดความคล้าย (Kernel Function) บนฟีเจอร์สเปซ ดังภาพที่ 2.16



ภาพที่ 2.16 หลักการของซัพพอร์ตเวกเตอร์แมชชีน

ที่มา: ฐาปนี สีหาโภชน์ (2556: 23)

จากภาพที่ 2.16 แสดงถึงแนวคิดหลักการของซัพพอร์ตเวกเตอร์แมชชีนซึ่งเป็นวิธีการที่สามารถนำมาใช้ในการจำแนกรูปแบบหรือกลุ่มข้องข้อมูลได้ โดยอาศัยระนาบในการแบ่งเขตข้อมูลออกเป็นสองกลุ่ม โดยใช้ไฮเปอร์เพลนที่เป็นเส้นตรง จะเห็นได้ว่ามีไฮเปอร์เพลนจำนวนมากที่สามารถคัดแยกได้



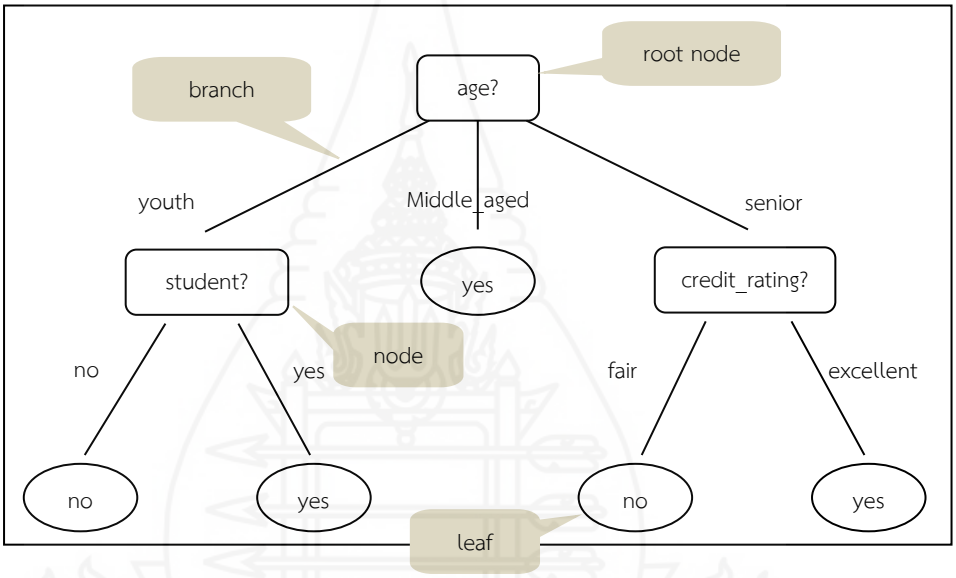
ภาพที่ 2.17 ไฮเปอร์เพลนที่เป็นเส้นตรงแบ่งกลุ่มข้อมูล

ที่มา: ฐาปนี สีหาโภชน์ (2556: 24)

จากภาพที่ 2.17 เส้น H_1, H_2, และ H_3 คือ ไฮเปอร์เพลนเป็นผลรวมของระยะห่างเส้นตรงที่เป็นไฮเปอร์เพลนถึงเส้นตรงที่ผ่านข้อมูลที่ใกล้ที่สุดและขนานกับไฮเปอร์เพลนของทั้งสองกลุ่ม ซึ่งเห็นได้ว่า H_2 จะสามารถแบ่งข้อมูลทั้งสองกลุ่มออกได้แต่ระยะในการแบ่งจากไฮเปอร์

เพลนไปถึงข้อมูลที่ใกล้สุดนั้นมีขนาดน้อย แต่จากเส้น H_3 จะเป็นเส้นที่แบ่งกลุ่มที่กว้างมากที่สุด คือให้ค่า Maximum Margin ซึ่งเรียกข้อมูลนี้ว่า Support Vector

3.5.3 เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคต้นไม้ตัดสินใจมีโครงสร้างคล้ายต้นไม้ จะประกอบด้วยโหนดในการตัดสินใจที่แสดงถึงข้อกำหนดของตัวแปรและเชื่อมต่อกันด้วยกิ่งก้าน (Branch) โดยมีโครงสร้างความสัมพันธ์เริ่มจากโหนดราก (Root node) อยู่ด้านบนสุดของแผนภูมิต้นไม้ และเชื่อมต่อโหนดในการตัดสินใจโดยเลือกแอตทริบิวต์ที่มีผลต่อการตัดสินใจแบ่งประเภทของข้อมูลไปยังกิ่งก้านต่างๆ เพื่อเชื่อมไปยังโหนดใบ (Leaf node) หรือเรียกว่า โหนดลูก ที่จะแสดงคลาสของข้อมูล ดังในภาพที่ 2.18



ภาพที่ 2.18 ลักษณะโครงสร้างต้นไม้ตัดสินใจ

ที่มา: Changala et al. (2012: 428)

จากภาพ 2.18 ด้านล่างแอตทริบิวต์ที่ถูกเลือกมาเป็นโหนดราก คือ อายุ โดยแบ่งอายุเป็น 3 รูปแบบ คือ youth, middle_aged, และ senior นอกจากนี้ยังมีการแตกกิ่งก้านไปจนถึงโหนดใบหรือโหนดลูกที่ใช้แสดงประเภทของกลุ่มลูกค้าที่ตัดสินใจซื้อคอมพิวเตอร์ (yes) และไม่ซื้อคอมพิวเตอร์ (no) เทคนิคต้นไม้ตัดสินใจมีอัลกอริทึมที่ใช้ในการจำแนกข้อมูลอยู่หลายแบบ เช่น ID3, Chi-Squared Automatic Classification and Regression Trees (CART), C4.5, Interaction Detection (CHAID), เป็นต้น ในงานวิจัยนี้ใช้อัลกอริทึม C4.5 ซึ่งเป็นอัลกอริทึมหนึ่งที่แพร่หลาย

สำหรับใช้ในการสร้างต้นไม้ตัดสินใจ ในการคัดเลือกคุณลักษณะข้อมูลจากคุณลักษณะที่มีความสัมพันธ์กับคลาสมากที่สุด ซึ่งได้จากการใช้ค่าอัตราส่วนเกน (Gain Ratio) เป็นตัวจำแนกในการตัดสินใจเลือกแอตทริบิวต์ที่ใช้รากหรือโหนด ซึ่งมีการทำงานคล้ายกับอัลกอริทึม ID3 แต่จะมีส่วนที่เพิ่มเข้ามาจาก ID3 ซึ่งสามารถแก้ไขในส่วนที่บกพร่องของ ID3 ได้ ในการเลือกคุณลักษณะที่ใช้เป็นโหนดรากบนต้นไม้ตัดสินใจนั้น ขั้นตอนวิธี ID3 ใช้ค่าเกนเป็นหลักในการเลือกแต่ขั้นตอนวิธี C4.5 ได้เพิ่มการใช้ค่ามาตรฐานอัตราส่วนเกน (Gain Ratio) ในการตัดสินใจเลือกคุณลักษณะเนื่องมาจากค่าเกนจะมีการเอนเอียงอย่างมากกับข้อมูลที่ประกอบด้วยคุณลักษณะที่มีค่าเป็นไปได้จำนวนมากๆ โดยจะต้องหาค่าสารสนเทศของการแบ่งแยก (Split Information) ก่อน ดังสมการต่อไปนี้

1) สมการ Entropy

$$\textbf{\textit{Entropy(S)}} = -\sum_{i=1}^{n} P(s_i)\log_2 P(S_i)$$

โดย $\boldsymbol{S}$ คือ แอตทริบิวต์ข้อมูล

2) สมการ Information Gain

$$\textbf{\textit{Gain(S)}} = Entropy(X) - \sum_{i=1}^{k} P(S_i)Entropy(X_{S_i})$$

โดย S คือ แอตทริบิวต์ข้อมูล

$P(S_i)$ คือ ความน่าจะเป็นของเซตย่อย $\boldsymbol{i}$ ของ $\boldsymbol{S}$

X_{S_i} คือ แอตทริบิวต์เป้าหมาย $\boldsymbol{X}$ ที่ขึ้นอยู่กับ $\boldsymbol{S}$

3) สมการ Split Information

$$\textbf{\textit{Split Info}}_{\boldsymbol{A}}\textbf{\textit{(D)}} = \sum_{j=1}^{v} \frac{|D_j|}{|D|} x Log_2\left(\frac{|D_j|}{|D|}\right)$$

โดย D คือ ชุดของข้อมูลตัวอย่าง

D_j คือ ชุดของข้อมูลตัวอย่างย่อยที่ j

D คือ จำนวนชุดตัวอย่างย่อย

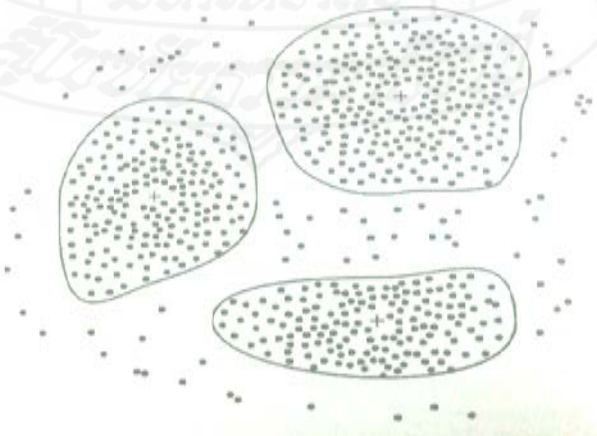
4) สมการ Gain Ratio

$$\textbf{\textit{Gain Ratio}} = \frac{InformationGain}{SplitInformation}$$

ที่มา: สัมพันธ์ ชัยภูมิ (2557: 8-9)

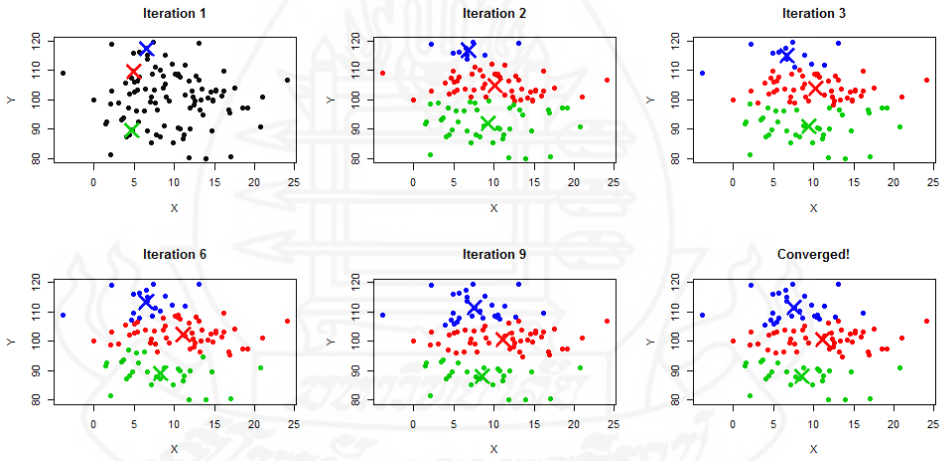
3.5.4 เทคนิคการจัดกลุ่ม (Clustering) เป็นการทำเหมืองข้อมูลในรูปแบบของการจัดกลุ่ม (Clustering) ไม่ได้กำหนดหรือแบ่งกลุ่มไว้ล่วงหน้าเป็นเทคนิคจัดกลุ่มที่มีลักษณะคล้ายคลึงกันมากที่สุด (Nearest neighbor) เพื่อจำแนกข้อมูลในแต่ละชุดข้อมูล โดยวัดค่าความห่างจากจุดศูนย์กลาง (Distance measurement) การวัดค่าเบี่ยงเบน (Deviation measurement) เป็นการรวมลักษณะที่คล้ายกันมากเข้าเป็นกลุ่มเดียวกัน

ตัวอย่างเช่น การทำเหมืองข้อมูลเพื่อวิเคราะห์ตำแหน่งที่อยู่ของลูกค้าบริษัท STOU Electronics เพื่อจัดกลุ่มลูกค้าว่าอาศัยในเขตพื้นที่ใดบ้าง ดังภาพที่ 2.19 ข้อมูลของลูกค้าที่อาศัยในพื้นที่ต่างๆ ถูกจัดได้ 3 กลุ่ม จะเห็นได้ว่าการจัดกลุ่มเหล่านี้ไม่ได้มีการกำหนดว่าต้องเป็นกลุ่มใดล่วงหน้า ในการเกิดกลุ่มเหล่านี้เป็นเพราะข้อมูลในแต่ละกลุ่มมีคุณลักษณะเป็นไปในทางเดียวกันจึงรวมตัวเป็นกลุ่มๆ ซึ่งแตกต่างจากการจำแนกประเภทที่มีการกำหนดกลุ่มหรือประเภทไว้ก่อนล่วงหน้า (วิภา เจริญภัณฑารักษ์, 2555: 27)



ภาพที่ 2.19 ข้อมูลของลูกค้าที่อาศัยในพื้นที่ต่างๆ ถูกจัดได้ 3 กลุ่ม

ในการวิจัยนี้จะใช้เทคนิคการจัดกลุ่มข้อมูลด้วยแบบเคมีน (k-means) ซึ่งเป็นที่นิยมไว้ใช้ในการแบ่งกลุ่มข้อมูลเป็นการเรียนรู้แบบไม่มีผู้สอนมีประสิทธิภาพในด้านของการเข้าใจง่าย และสามารถใช้ได้กับข้อมูลหลายประเภท โดยจะทำการแบ่งกลุ่มข้อมูลออกเป็น k กลุ่ม ตามที่ต้องการ (Han, et al., 2006) โดยในแต่ละกลุ่มจะมีข้อมูลจุดศูนย์กลางประจำกลุ่ม (Centroid) ซึ่งได้จากค่าเฉลี่ยของกลุ่ม และในการพิจารณาว่าข้อมูลจะอยู่กลุ่มใดจะดูจากระยะห่างระหว่างข้อมูลกับจุดศูนย์กลางประจำกลุ่มในแต่ละกลุ่ม ข้อมูลที่มีระยะห่างจากจุดศูนย์กลางประจำกลุ่มใดน้อยที่สุดจุถูกจัดอยู่ในกลุ่มนั้นๆ เมื่อข้อมูลทุกตัวถูกพิจารณาจัดกลุ่มแล้ว ก็จะทำการคำนวณหาค่าเฉลี่ยจุดศูนย์กลางประจำกลุ่มใหม่ หากจุดศูนย์กลางของแต่ละกลุ่มมีการเปลี่ยนตำแหน่งก็จะทำการวนซ้ำเพื่อจัดกลุ่มข้อมูลตามจุดศูนย์กลางกลุ่มที่เปลี่ยนไป และจะหยุดเมื่อจุดศูนย์กลางประจำกลุ่มของทุกกลุ่มไม่มีการเปลี่ยนแปลง สุดท้ายก็จะได้ข้อมูลที่มีลักษณะเหมือนกันอยู่ในกลุ่มเดียวกัน และข้อมูลที่มีลักษณะแตกต่างกันอยู่คนละกลุ่ม ดังภาพที่ 2.20



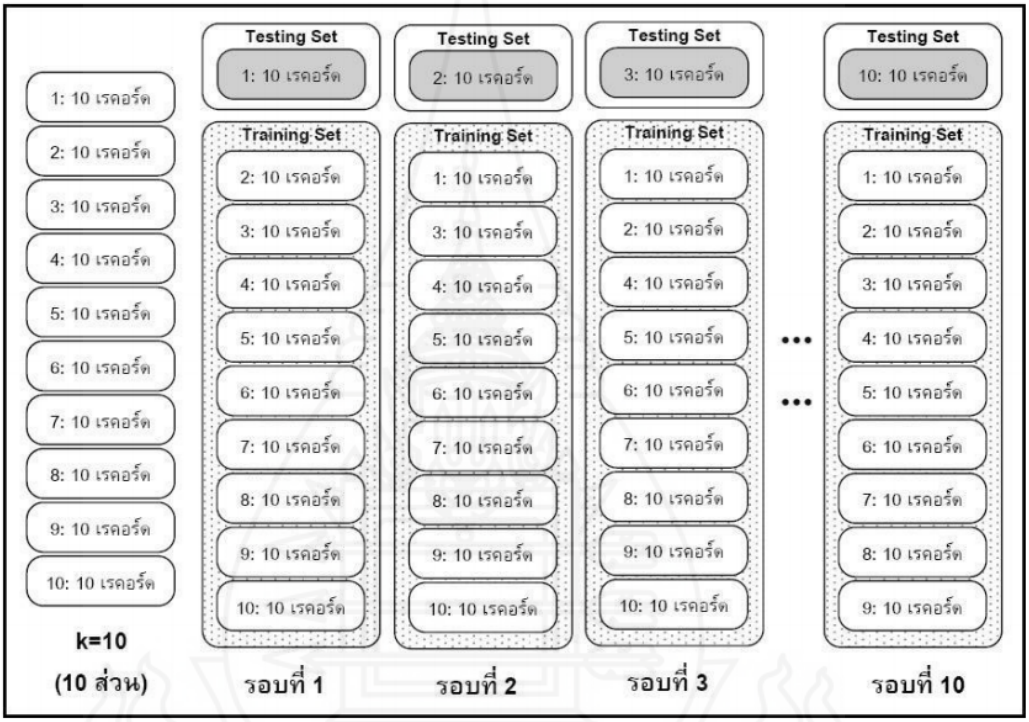
ภาพที่ 2.20 ขั้นตอนการจัดกลุ่มข้อมูลด้วย k-means

ที่มา: http://www.learnbymarketing.com/methods/k-means-clustering

3.6 การทดสอบประสิทธิภาพแบบ k-fold Cross Validation

การวัดประสิทธิภาพของแบบจำลองในวิจัยนี้ จะใช้การวัดค่าความถูกต้องของข้อมูล (Accuracy) และค่ารากที่สองของความคลาดเคลื่อน (Root Mean Square Error: RMSE) ส่วนการทดสอบทำการเลือกข้อมูลชุดสอน (Training data) และข้อมูลชุดทดสอบ (Testing data) ซึ่งใช้วิธีการทดสอบประสิทธิภาพแบบ k – fold Cross validation

การจำแนกข้อมูลโดยใช้เทคนิคของการทำเหมืองข้อมูล เช่น ต้นไม้ตัดสินใจ โครงข่ายประสาทเทียม และเบย์อย่างง่าย (Naïve Bayes Classifier) นั้นจะต้องมีการแบ่งข้อมูลออกเป็นข้อมูลชุดสอน และข้อมูลชุดทดสอบ แต่ในบางครั้งอาจเกิดปัญหาจากการเลือกข้อมูลที่ดีและง่ายมาเป็นข้อมูลชุดทดสอบ ทำให้ผลการจำแนกนั้นดีเกิดจริง ดังนั้นจึงมีการคิดวิธี k – fold Cross validation ขึ้นมาแก้ไขปัญหา ซึ่งเหมาะสำหรับชุดข้อมูลจำนวนไม่มาก โดยจะแบ่งข้อมูลออกเป็น k ชุด จำนวนเท่าๆ กัน (วรารัตน์ รุ่งวรวุฒิ, 2554) ซึ่งตัวอย่างของการแบ่งข้อมูล ดังภาพที่ 2.21



ภาพที่ 2.21 ตัวอย่างการแบ่งข้อมูลด้วยวิธี K – fold Cross validation

ที่มา: ศุภกร สระบัว (2557: 15)

ในงานวิจัยนี้ได้เลือกใช้ k – fold Cross validation โดยกำหนดค่า k เท่ากับ 10 คือ การแบ่งข้อมูลออกเป็น 10 ส่วนเท่าๆ กัน โดยการทดลองครั้งแรกข้อมูลชุดที่ 1 เป็นข้อมูลชุดทดสอบ และข้อมูลชุดที่เหลือเป็นข้อมูลชุดสอน ในการทดลองครั้งที่สองข้อมูลชุดที่ 2 เป็นข้อมูลชุดทดสอบ และข้อมูลชุดที่เหลือเป็นข้อมูลชุดสอน ทำจนกระทั่งข้อมูลทุกชุดได้ถูกนำมาเป็นข้อมูลชุดทดสอบ ตามค่า k ที่กำหนดคือ 10 ครั้ง เช่นเดียวกับการแบ่งชุดข้อมูล ส่วนการวัดประสิทธิภาพประกอบไปด้วย

3.6.1 วิธีการวัดค่าความถูกต้อง (Accuracy) ในการสร้างตัวแบบพยากรณ์ สามารถคำนวณได้จาก

$$\text{Accuracy} = \frac{\text{Correctly Classified Instance}}{\text{Total Number of Instances}} \times 100$$

โดย Correctly Classified Instance คือ จำนวนตัวอย่างที่ทำนายถูกต้อง

Total Number of Instances คือ จำนวนตัวอย่างทั้งหมด

ที่มา: C. Lakshmi Devasena (2013: 133)

3.6.2 ค่ารากที่สองของความคลาดเคลื่อน (Root Mean Square Error : RMSE) คือ การวัดค่าความแตกต่างระหว่างค่าจริงกับค่าที่ได้จากแบบจำลอง หาก RSME มีค่าน้อย แสดงว่าแบบจำลองใกล้เคียงกับค่าจริง สามารถคำนวณได้จาก

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i^{'} - Y_i\right)^2}$$

โดย Y_i แทน ค่าของข้อมูลที่เกิดขึ้นจริง

$Y_i^{'}$ แทน ค่าที่ได้จากการพยากรณ์

n แทน จำนวนนำเข้าทั้งหมด

ที่มา: Hye Rin Kim (2009: 4)

4. การคัดเลือกคุณลักษณะ

การคัดเลือกคุณลักษณะ (Feature Selection) เป็นกระบวนการหรือวิธีการหนึ่งในการทำเหมืองข้อมูลเพื่อช่วยค้นหาลักษณะที่เหมาะสมหรือดีที่สุดในการทำเหมืองข้อมูล ซึ่งในการทำ

เหมืองข้อมูลนั้นจะมีข้อมูลจำนวนมากและมีความสัมพันธ์ที่ซับซ้อนกันมาก ทำให้ในการทำเหมืองข้อมูลนั้นอาจมีคุณลักษณะที่ไม่ตรงประเด็นหรือมีคุณลักษณะที่ซ้ำซ้อนกัน ดังนั้นในการใช้วิธีการคัดเลือกคุณลักษณะจึงช่วยในการปรับปรุงคุณภาพการทำเหมืองข้อมูลและยังเป็นการลดขนาดข้อมูล อีกทั้งช่วยประหยัดเวลาในการประมวลผลข้อมูล วิธีการคัดเลือกคุณลักษณะในงานวิจัยนี้ คือ

4.1 Correlation-based Feature Subset Selection

วิธี Correlation-based Feature Subset Selection: CFS คือ การคัดเลือกคุณลักษณะด้วยอัลกอริทึมการกรองอย่างง่าย โดยจะจัดอันดับกลุ่มย่อยของมิติข้อมูล ตามความสัมพันธ์ที่อยู่บนพื้นฐานของฟังก์ชันการประมาณแบบ Heuristic ซึ่งกลุ่มย่อยของมิติข้อมูลจะมีความสัมพันธ์กันสูงกับคลาส และไม่มีความสัมพันธ์กับคลาสอื่นๆ สำหรับมิติข้อมูลที่ไม่เกี่ยวข้องก็จะถูกละทิ้งเพราะมิติของข้อมูลเหล่านี้อาจจะมีความสัมพันธ์กับคลาสมิติของข้อมูลที่ซ้ำซ้อนอาจถูกขจัดออกไปจากกลุ่มมิติของข้อมูลที่มีความสัมพันธ์สูง

สมการประเมินกลุ่มย่อยของมิติข้อมูลแบบ CFS (อาริยพร สุดใจ, 2557: 22) แสดงในสมการดังนี้

$$m_s = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}}$$

โดยที่ m_s คือ กลุ่มของ Attributes s ที่ประกอบด้วย k Attributes ที่คัดเลือก

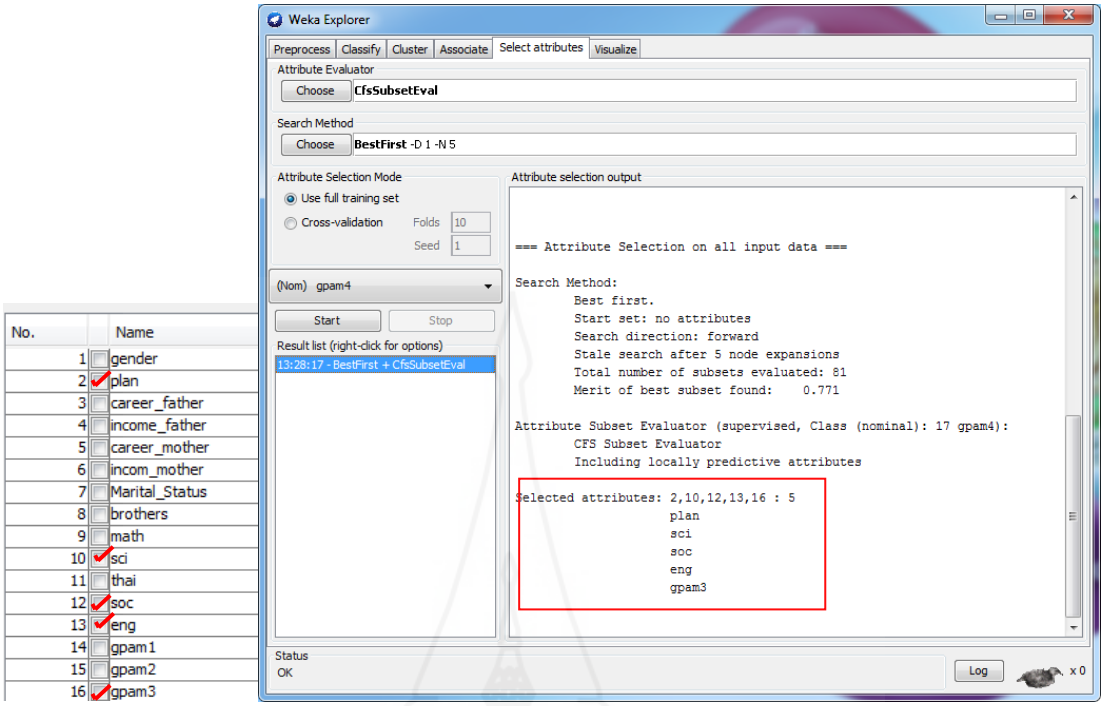
k คือ จำนวนคุณลักษณะในกลุ่ม s (เซตคุณลักษณะที่ถูกคัดเลือก)

s คือ กลุ่มคุณลักษณะที่มี k คุณลักษณะ

$\overline{r_{cf}}$ คือ ค่าเฉลี่ยความสัมพันธ์ระหว่างคุณลักษณะกับกลุ่มคุณลักษณะที่ถูกเลือก (f € s)

$\overline{r_{ff}}$ คือ ค่าเฉลี่ยความสัมพันธ์ระหว่างคุณลักษณะที่ถูกเลือก ภายในกลุ่มของคุณลักษณะที่ถูกเลือก

ตัวอย่างการคัดเลือกคุณลักษณะด้วยวิธี CFS



ภาพที่ 2.22 การคัดเลือกคุณลักษณะด้วยวิธี CFS

จากภาพ 2.22 มี 16 Attribute ได้ถูกคัดเลือกเหลือเพียง 5 Attribute คือ plan , sci , soc , eng และ gpam3

4.2 Information Gain

วิธี Information Gain : IG คือ วิธีการคัดเลือกคุณลักษณะโดยการวัดค่า Gain ของแต่ละโหนด โดยโหนดที่มีค่า Gain สูงสุดจะเลือกให้เป็นโหนดรากและคำนวณค่า Gain อีกครั้งกับข้อมูลที่เหลือ เพื่อให้ได้โหนดต่อไป โดยใช้สูตรการคำนวณ (อาริยพร สุดใจ, 2557: 22) ดังนี้

$$\textit{Entropy(S)} = - \sum_{c=1}^{N} p(S_c) x \log_2 p(S_c) \qquad (2\text{-}1)$$

โดยที่ $\sum_c$ คือ ผลรวมของความน่าจะเป็นของค่า S ที่เกิดในคลาส c

$p(S_c)$ คือ ค่าความถี่ที่มีความสัมพันธ์ของกลุ่ม S กับโหนด c

$$\textit{Gain(S,V)} = \textit{Entropy(S)} - \sum_{v \in value(v)} \frac{|S_v|}{|S|} \text{ x } \textit{Entropy}(S_v) \qquad (2\text{-}2)$$

โดยที่ค่า ***Entropy(S)*** คือ ค่าของ Entropy ของตัว Root

$\sum_{v \in value(v)} \frac{|S_v|}{|S|}$ ค่า Entropy ในแต่ละโหนดย่อยของการคำนวณ

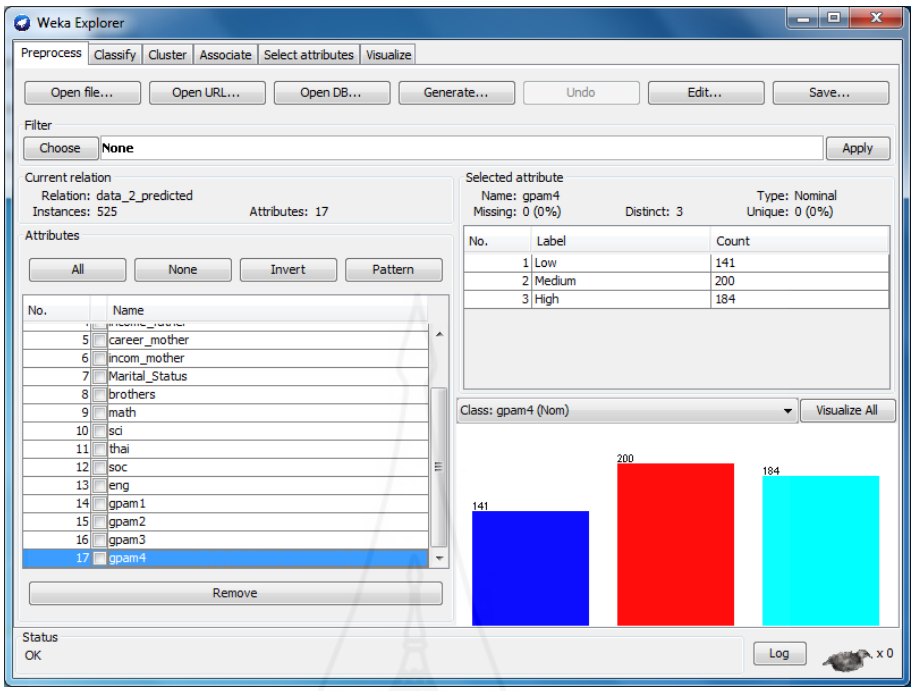
$$\textbf{\textit{SplitInfo(S,V)}} = -\sum_{i=1}^{m} \frac{|S_i|}{|S|} \text{ x } \log_2 \frac{|S_i|}{|S|} \qquad (2\text{-}3)$$

$$\textbf{\textit{Gain(S,V)}} = \frac{Gain(S,V)}{SplitInfo(S,V)} \qquad (2\text{-}4)$$

โดยนำค่าที่ได้จาก (2-2) และ (2-3) ใสคำนวณหา Gain Ratio ตามสมการที่ (2-4)

5. โปรแกรม WEKA (Waikato Environment for Knowledge Analysis)

โปรแกรม WEKA (Waikato Environment for Knowledge Analysis) เป็นซอฟต์แวร์ที่สามารถใช้งานได้ฟรีภายใต้เงื่อนไข GPL license โดยซอฟต์แวร์นี้พัฒนาขึ้นด้วยภาษาจาวา ซึ่งปรกอบไปด้วยการวิเคราะห์ข้อมูลและการใช้ทำเหมืองข้อมูล โดยเน้นการเรียนรู้ด้วยเครื่อง (Machine Learning) กับการทำเหมืองข้อมูล การแสดงผลเป็นแบบ GUI และสามารถทำงานได้ในหลายระบบปฏิบัติการ เป็นซอฟต์แวร์ที่ใช้งานในด้านการทำเหมืองข้อมูล รวบรวมอัลกอริทึมหลากหลาย ข้อดีของการใช้โปรแกรม WEKA คือ มีอัลกอริทึมที่รู้จักและนิยมใช้ในการทำเหมืองข้อมูลให้เลือกใช้อย่างครบถ้วน แฟ้มข้อมูลในการนำเข้า ได้แก่ แฟ้มข้อมูลที่อยู่ในรูปแบบ arff, cvs เป็นต้น นอกจากนี้ยังสามารถใช้ข้อมูลที่อยู่ในฐานข้อมูลในการนำเข้าได้ ในการวิจัยนี้ใช้โปรแกรม WEKA ในการทำเหมืองข้อมูลซึ่งมีหน้าจอการทำงาน ดังภาพที่ 2.23



ภาพที่ 2.23 แสดงหน้าจอการทำงานโปรแกรม WEKA

ในส่วนของเมนูหลักโปรแกรม WEKA Explorer มีดังนี้

1) Preprocess เป็นการเตรียมข้อมูล
2) Classify เป็นโมดูลของการทำเหมืองข้อมูลแบบจำแนกกลุ่มหรือพยากรณ์
3) Cluster เป็นโมดูลของการทำเหมืองข้อมูลแบบการจัดกลุ่ม
4) Associate เป็นโมดูลของการทำเหมืองข้อมูลแบบกฎเชื่อมโยง
5) Select attributes เป็นโมดูลสำหรับการวิเคราะห์ความเกี่ยวพันของคุณลักษณะ
6) Visualize เป็นการนำเสนอข้อมูลด้วยภาพนามธรรมสองมิติ

6. งานวิจัยที่เกี่ยวข้อง

ทิพย์หทัย ทองธรรมชาติ (2554) ได้ทำการศึกษาเรื่อง การพยากรณ์ผลสัมฤทธิ์ทางการเรียนด้วยเทคนิคเหมืองข้อมูล มีวัตถุประสงค์เพื่อนำเทคนิคเหมืองข้อมูลมาประยุกต์ใช้ในการพยากรณ์หาปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนของนิสิตสาขาวิทยาการคอมพิวเตอร์ และนิสิตสาขาวิชาเทคโนโลยีสารสนเทศ โดยใช้เทคนิคเหมืองข้อมูล 4 เทคนิค ได้แก่ เทคนิคเคเนียเรสเนเบอร์ (K-Nearest Neighbor) เทคนิคเนอีฟเบย์ (Naïve Bayes) เทคนิคต้นไม้ตัดสินใจ (Decision

Tree) และเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ (Back propagation Neural Network) ข้อมูลที่ใช้ในการวิจัยครั้งนี้เป็นข้อมูลภูมิหลัง และข้อมูลผลการเรียนรายวิชาที่อยู่ในแผนการเรียนชั้นปีที่ 1 และปีที่ 2 ของนิสิตสาขาวิทยาการคอมพิวเตอร์ และสาขาวิชาเทคโนโลยีสารสนเทศ มหาวิทยาลัยนเรศวร ปีการศึกษา 2551-2552 จำนวน 236 คน จากผลการวิจัยพบว่า ปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนของนิสิตสาขาวิทยาการคอมพิวเตอร์ และนิสิตสาขาเทคโนโลยีสารสนเทศ มีความแตกต่างกันชุดตัวแปรสำคัญที่คัดเลือกโดยพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน และสัมประสิทธิ์อีต้าเมื่อนำมาใช้จะทำให้ผลพยากรณ์สูงกว่าชุดตัวแปรที่ได้จากอัตราส่วนแกนร่วมกับวิธีการค้นหาแบบจัดลำดับ และพบว่าข้อมูลภูมิหลังไม่อยู่ในตัวแปรสำคัญที่คัดเลือกทั้ง 2 วิธี และในขั้นตอนการพยากรณ์นั้น ผลการเปรียบเทียบประสิทธิภาพพบว่าเทคนิคเนอีฟเบย์และเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับให้ค่าความถูกต้องในการพยากรณ์สูงใกล้เคียงกันและสูงกว่าเทคนิคเคเนียเรสเนเบอร์ และเทคนิคต้นไม้ตัดสินใจ

พรเทพ คงไชย และรัชฎา คงคะจันทร์ (2554) ได้ทำการศึกษาเกี่ยวกับเรื่อง การศึกษาเปรียบเทียบในการคัดเลือกคุณลักษณะที่เหมาะสมสำหรับการทำเหมืองข้อมูลเพื่อพยากรณ์โอกาสการสำเร็จการศึกษาของนักศึกษา มีจุดประสงค์เพื่อนำเสนอรูปแบบและวิธีการทดลองเชิงเปรียบเทียบการคัดเลือกคุณลักษณะแบบ Filter โดยใช้ Ranker Method และ Wrapper โดยใช้ Genetic Algorithm ซึ่งทั้ง 2 รูปแบบใช้การจำแนกข้อมูล (Classifiers) คุณลักษณะแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (Multi-layer Perceptron: MLP) แบบซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines : SVM) และแบบเบย์เซี่ยน (Bayesian Network) เพื่อการจำแนกความถูกต้องโดยอาศัยค่า Precision ค่า Recall และค่า F-Measure ผลการวิจัยพบว่า เมื่อทำการเปรียบเทียบค่าความถูกต้องสูงสุดจากการจำแนก คือค่า Precision เท่ากับ 95.10% ค่า Recall เท่ากับ 94.90% และค่า F-Measure เท่ากับ 94.90% ลดการใช้คุณลักษณะที่ไม่เหมาะสมลง 75% และลดเวลาในการทำเหมืองข้อมูลเพื่อพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษา

สุพัฒน์กุล ภัคโชค (2555) ได้ทำการศึกษาเกี่ยวกับเรื่อง ตัวแบบแผนการเรียนระดับมัธยมศึกษาตอนปลาย โดยการพิจารณาผลการเรียนรายวิชาหลัก ด้วยเทคนิคเหมืองข้อมูล เพื่อค้นหาปัจจัยที่มีผลต่อการเลือกแผนการเรียน และความสามารถในการศึกษาแผนการเรียนนั้นได้อย่างประสบความสำเร็จ โดยใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) อัลกอริทึม C4.5 ในการหาความสัมพันธ์ ใช้เครื่องมือแบบสอบถามในการเก็บข้อมูลการตัดสินใจเลือกแผนการเรียนของนักเรียนระดับมัธยมศึกษาตอนปลาย โดยมีจำนวนกลุ่มตัวอย่างจากนักเรียนระดับมัธยมศึกษาตอนปลาย โรงเรียนสตรีศรีสุริโยทัย ปีการศึกษา 2555 ทั้งสิน 850 คน ผลการวิจัยพบว่าตัวแบบเลือกแผนการเรียนระดับมัธยมศึกษาตอนปลายสามารถบ่งบอกได้ว่าปัจจัยใดมีผลต่อการตัดสินใจเลือก

แผนการเรียนของนักเรียน และให้ค่าความถูกต้องในการแนะนำแผนการเรียนร้อยละ 79.76 จากตัวแบบดังกล่าวทำให้ทราบว่าวิชาพื้นฐานในระดับมัธยมศึกษาตอนต้น ได้แก่ วิชาภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ สังคมศึกษา ศาสนาและวัฒนธรรม และภาษาอังกฤษ รวมทั้งผลการเรียนเฉลี่ยสะสมระดับมัธยมศึกษาตอนต้นเป็นปัจจัยหลักที่มีผลโดยตรงต่อการตัดสินใจเลือกแผนการเรียนของนักเรียนและสามารถศึกษาในแผนการเรียนนั้นได้อย่างประสบความสำเร็จ

อาทิตยาพร โรจรัตน์ (2556) ได้ทำการศึกษาเกี่ยวกับ การทำนายผลการศึกษาเพื่อการวางแผนการลงทะเบียนของนักศึกษา โดยใช้การทำเหมืองข้อมูล ในงานวิจัยนี้มีวัตถุประสงค์เพื่อเสนอรูปแบบการทำนายผลการศึกษาเพื่อใช้ในการวางแผนการลงทะเบียนเรียน โดยใช้เทคนิคกฎความสัมพันธ์ (Association Rules) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคของการเรียนรู้เบย์อย่างง่าย (Naïve Bayes) นำมาวิเคราะห์และเปรียบเทียบในการหาความสัมพันธ์ที่ทำให้เกิดการทำนายผลการศึกษาของนักศึกษาสาขาเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยขอนแก่น เป็นกรณีศึกษา ผลการวิจัยพบว่า การวิเคราะห์ความสัมพันธ์ของทั้ง 3 เทคนิค มีรูปแบบที่น่าสนใจ เทคนิคกฎความสัมพันธ์มีทั้งหมด 26 กฎความสัมพันธ์ที่สามารถนำไปใช้งานได้จริง เทคนิคต้นไม้ตัดสินใจได้กฎทั้งหมด 26 กฎความสัมพันธ์ แบ่งเป็นการวิเคราะห์โดยใช้รายวิชาทั้งหมด 13 กฎความสัมพันธ์ มีค่าความถูกต้องอยู่ที่ 77.92% ประเมินโดยการใช้ 10-fold cross validation และการวิเคราะห์โดยแยกเอาแต่รายวิชาที่เป็นวิชาเลือกทั้งหมด 13 กฎความสัมพันธ์ มีค่าความถูกต้องอยู่ที่ 68.56% โดยใช้การประเมิน 40- fold cross validation และเทคนิคสุดท้ายการเรียนรู้เบย์อย่างง่ายที่ทำการประมวลผลข้อมูลเพื่อหาค่าความถูกต้องที่มากที่สุด โดยวัดค่าความถูกต้องได้ 85.77% โดยใช้การประเมิน 10- fold cross validation อย่างไรก็ตามเทคนิคกฎความสัมพันธ์และเทคนิคต้นไม้ตัดสินใจจะเห็นความสัมพันธ์ของรายวิชาเพื่อการลงทะเบียนได้ดีกว่าเทคนิคของการเรียนรู้เบย์อย่างง่าย

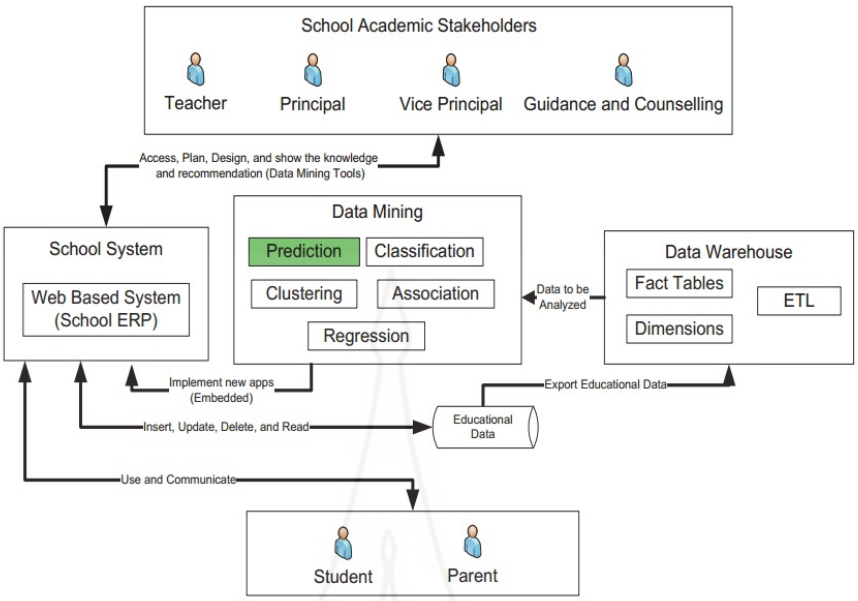
พรรณิภา บุตรเอก (2557) ได้ทำการศึกษาเกี่ยวกับ การพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน มีวัตถุประสงค์เพื่อพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษา โดยใช้เทคนิคเหมืองข้อมูลแบบซัพพอร์ตเวกเตอร์แมชชีน ชุดข้อมูลกลุ่มตัวอย่างที่นำมาใช้ทดสอบเป็นข้อมูลของนักศึกษาระดับปริญญาตรีหลักสูตร 4 ปี สาขาวิชาวิทยาการคอมพิวเตอร์ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยกรุงเทพ ระหว่างปีการศึกษา 2547-2551 จำนวนทั้งสิ้น 138 ระเบียบ 18 คุณลักษณะ ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคต้นไม้ตัดสินใจ และเทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ ใช้ในการสร้างตัวแบบพยากรณ์แล้วเปรียบเทียบเพื่อเลือกตัวแบบที่มีประสิทธิภาพสูงที่สุด โดยการทดสอบประสิทธิภาพแบบใช้วิธีไขว้ทบ 15 ส่วน ผลการวิจัยพบว่า ตัวแบบพยากรณ์ที่ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน ให้ความ

แม่นยำในการพยากรณ์สูงที่สุดที่ร้อยละ 89.13 โดยมีผลการประเมินค่าความแม่นยำเท่ากับ 0.878 ค่าระลึกเท่ากับ 0.891 และค่าประสิทธิภาพโดยรวมเท่ากับ 0.860 นอกจากนี้ยังมีค่า RMSE น้อยที่สุดเท่ากับ 0.3297 โดยที่ตัวแบบพยากรณ์ซัพพอร์ตเวกเตอร์แมชชีนใช้เคอร์เนลฟังก์ชั่นแบบ Polynomial Kernel เป็นตัวแบบที่เหมาะสมที่สุดสำหรับการพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษา

William W. Guo (2010) ได้ศึกษา เรื่อง การรวมวิธีการทางสถิติและโครงข่ายประสาทสำหรับการวิเคราะห์ความพึงพอใจ และการทำนายหลักสูตรการศึกษาของนักศึกษา ซึ่งเป็นการวิเคราะห์และพยากรณ์ถึงปัจจัยที่สำคัญในการพยากรณ์ความพึงพอใจทางการเรียนเพื่อส่งเสริมการเลือกสถาบันศึกษาของนักศึกษาใหม่ที่อยู่ในท้องถิ่น และเป็นการดึงดูดนักศึกษาในปัจจุบันไม่ให้ย้ายไปที่อื่น ซึ่งผู้วิจัยได้ทำการเก็บรวบรวมข้อมูลจากหลักสูตรใน 11 ภาคการศึกษา ระหว่างปีการศึกษา ค.ศ. 2002 – 2007 โดยใช้หลักการทางสถิติสมการถดถอยเชิงเส้นร่วมกับโครงข่ายประสาทเทียม สำหรับสร้างตัวแบบวิเคราะห์ความพึงพอใจ และทำนายหลักสูตรของนักศึกษา ผลการวิจัยพบว่า โครงข่ายประสาทเทียมให้ผลลัพธ์การพยากรณ์ดีกว่าสมการถดถอยเชิงเส้น

Buathong and Meesad (2013) ได้ศึกษาและเสนอวิธีการลดมิติข้อมูลโดยใช้เทคนิคการคัดเลือกคุณลักษณะของข้อมูลด้วยวิธี Linear SVM Weight จำนวน 2 รอบ (Double Linear SVM Weight) เปรียบเทียบกับวิธีแบบ ReliefF, Information Gain, Gain Ratio และ Gini Gain โดยใช้อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) ร่วมกับการคัดเลือกคุณลักษณะ ซึ่งผลการศึกษาวิจัยพบว่า เทคนิคการคัดเลือกคุณลักษณะแบบ Double Linear SVM Weight ให้ค่าความถูกต้อง 100% ของข้อมูลที่แอตทริบิวต์ลดลงเหลือที่จำนวน 60, 50 และ 40 ซึ่งการคัดเลือกคุณลักษณะช่วยให้ค่าความถูกต้องสูงขึ้น

Kurniawan and Halim (2013) ได้ศึกษาและวิจัยเกี่ยวกับการพัฒนาคลังข้อมูลและการใช้เทคนิคเหมืองข้อมูลในการวิเคราะห์และพยากรณ์ผลการเรียน เพื่อช่วยให้นักเรียนมีผลการเรียนที่ดีขึ้น ผลการวิจัยพบว่า คลังข้อมูลที่ออกแบบและพัฒนาขึ้นใช้โครงสร้างแบบ Star schema นำเสนอข้อมูลในรูปแบบของเว็บ และใช้การดึงข้อมูลจากคลังข้อมูลเพื่อมาทำเหมืองข้อมูลด้วยการพยากรณ์ การจัดกลุ่ม การจำแนกข้อมูล และการหาความสัมพันธ์ ผลการพยากรณ์ที่ได้จะแสดงบนเว็บของโรงเรียน ดังภาพที่ 2.24 ซึ่งผลที่ได้นั้นแสดงให้เห็นว่าการทำเหมืองข้อมูลสามารถพยากรณ์ผลการเรียนของนักเรียนได้ และสามารถในโรงเรียนเพื่อเป็นการเพิ่มประสิทธิภาพของผลสัมฤทธิ์ทางการเรียนให้กับโรงเรียนและนักเรียน



ภาพที่ 2.24 โครงสร้างการทำคลังข้อมูลและเหมืองข้อมูลของโรงเรียน

ที่มา: Kurniawan and Halim (2013: 102)

Rai, Saini and Jain (2014) ได้ศึกษาการออกกลางคันของนักศึกษาชั้นปีที่ 1 ในระดับปริญญาตรี โดยใช้เทคนิคต้นไม้ตัดสินใจอัลกอริทึม ID3 ร่วมกับการคัดเลือกคุณลักษณะด้วยวิธี Correlation-based Feature Subset Selection ซึ่งใช้ข้อมูลผลการเรียนและประวัติส่วนตัวของนักศึกษาชั้นปีที่ 1 จากการสุ่มในมหาวิทยาลัยต่าง ได้กลุ่มตัวอย่างจำนวน 220 คน ผลการศึกษาพบว่า การใช้วิธีการคัดเลือกคุณลักษณะแบบ Correlation-based Feature Subset Selection คัดเลือกคุณลักษณะไว้ 12 คุณลักษณะ จากทั้งหมด 34 คุณลักษณะ ทำให้การทำนายโดยใช้เทคนิคต้นไม้ตัดสินใจอัลกอริทึม ID3 ให้ความแม่นยำ 98% ซึ่งจะเห็นได้ว่าการคัดเลือกคุณลักษณะมีส่วนช่วยในการเลือกคุณลักษณะที่สำคัญจึงทำให้ผลการทำนายมีความแม่นยำมากขึ้น

Venkatramaphanikumar S., (2015) ได้ทำการศึกษาเกี่ยวกับการพยากรณ์อาชีพที่นักศึกษามีโอกาสได้ทำหลังจบการศึกษา โดยใช้เทคนิคเหมืองข้อมูลโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน, เทคนิคต้นไม้ตัดสินใจอัลกอริทึม J48 และเทคนิคของการเรียนรู้เบย์อย่างง่าย มาเปรียบเทียบประสิทธิภาพของการพยากรณ์ โดยใช้ข้อมูลผลการเรียน ประวัติส่วนตัว วิชาเอก ของนักศึกษา ผลการวิจัยพบว่า เทคนิคเหมืองข้อมูลโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอนให้ค่าความถูกต้องดีที่สุด เทคนิคต้นไม้ตัดสินใจอัลกอริทึม J48 และและเทคนิคของการเรียนรู้เบย์อย่างง่ายตามลำดับ

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องดังกล่าวนั้น มีการใช้เทคนิคเหมืองข้อมูลในการพยากรณ์หรือทำนายผลการเรียนจำนวนมาก ซึ่งในการทำเหมืองข้อมูลส่วนใหญ่จะใช้โปรแกรม WEKA เป็นเครื่องมือในการวิเคราะห์ข้อมูลและทดสอบ ซึ่งมีการใช้เทคนิคโครงข่ายประสาทเทียม เทคนิคต้นไม้ตัดสินใจ และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน มาใช้ในการวิเคราะห์สร้างตัวแบบพยากรณ์ ในบ้างครั้งที่ข้อมูลเหล่านั้นมีคุณลักษณะที่ใช้ในการวิเคราะห์จำนวนมากทำให้ต้องมีการคัดเลือกคุณลักษณะที่สำคัญเพื่อให้ได้ตัวแบบพยากรณ์ที่ถูกต้องและดีที่สุด โดยใช้เทคนิควิธีการในการคัดเลือกคุณลักษณะเข้ามาช่วยก่อนการใช้เทคนิคเหมืองข้อมูล แต่งานวิจัยส่วนใหญ่จะใช้ในการพยากรณ์ผลการเรียนในระดับอุดมศึกษาหรือแนะแนวในการเลือกวิชา คณะในระดับมหาวิทยาลัยผู้วิจัยจึงสนใจและศึกษาเกี่ยวกับการใช้เทคนิคเหมืองข้อมูลมาช่วยในการพยากรณ์ผลการเรียนของนักเรียนในระดับการศึกษาขั้นพื้นฐาน โดยนำข้อมูลผลการเรียนและประวัติของนักเรียนมาสร้างคลังข้อมูลด้วยโปรแกรม Microsoft SQL Server 2008 R2 จากนั้นใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนตามขั้นตอนของ CRISP-DM เพื่อหาตัวแบบพยากรณ์ที่ดีที่สุดไปใช้ในการพยากรณ์ผลการเรียนของนักเรียน ซึ่งจะสามารถช่วยให้นักเรียนสามารถวางแผนการเรียนและใช้ประกอบการตัดสินใจในการเรียนต่อไป

บทที่ 3
วิธีดำเนินการวิจัย

งานวิจัยนี้เป็นงานวิจัยเชิงพัฒนา (Development research) โดยเป็นการพัฒนาคลังข้อมูลผลการเรียนของนักเรียนเพื่อนำมาวิเคราะห์หาความรู้ใหม่ โดยการทำเหมืองข้อมูลตามขั้นตอนแบบจำลองของคริสพ์ – ดีเอ็ม (CRISP – DM Model) ซึ่งรายละเอียดของวิธีดำเนินการวิจัยประกอบด้วย 3 ส่วน ได้แก่

1. ประชากรและกลุ่มตัวอย่าง
2. เครื่องมือที่ใช้ในการดำเนินงานวิจัย
3. ขั้นตอนการดำเนินงาน

1. ประชากรและกลุ่มตัวอย่าง

1.1 ประชากร

ข้อมูลที่ใช้ในงานวิจัยเป็นข้อมูลทุติยภูมิ (Secondary data) ของนักเรียนระดับชั้นมัธยมศึกษาโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา ซึ่งถูกจัดเก็บไว้ในฐานข้อมูลของโรงเรียน

1.2 กลุ่มตัวอย่าง

ฐานข้อมูลผลการเรียน และประวัตินักเรียนระดับชั้นมัธยมศึกษา โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา ตั้งแต่ปีการศึกษา 2548 -2556

2. เครื่องมือที่ใช้ในการดำเนินงาน

2.1 ฮาร์ดแวร์

2.1.1 เครื่องคอมพิวเตอร์แม่ข่าย ที่มีคุณสมบัติดังนี้

1) หน่วยประมวลกลาง (CPU) ความเร็วไม่น้อยกว่า 2.00 GHz
2) หน่วยความจำหลัก (RAM) ความจุไม่น้อยกว่า 2 GB

3) อุปกรณ์บันทึกข้อมูล (Hard disk) ความจุไม่น้อยกว่า 350 GB

2.1.2 เครื่องคอมพิวเตอร์ลูกข่าย ที่มีคุณสมบัติดังนี้

1) หน่วยประมวลกลาง (CPU) ความเร็วไม่น้อยกว่า 1.00 GHz

2) หน่วยความจำหลัก (RAM) ความจุไม่น้อยกว่า 1 GB

3) อุปกรณ์บันทึกข้อมูล (Hard disk) ความจุไม่น้อยกว่า 80 GB

2.2 ซอฟต์แวร์

2.2.1 เครื่องคอมพิวเตอร์แม่ข่าย ที่มีคุณสมบัติดังนี้

1) ระบบปฏิบัติการไมโครซอฟต์วินโดวส์เซิร์ฟเวอร์ 2008 (Microsoft Windows Server 2008) เป็นระบบปฏิบัติการ

2) โปรแกรมไมโครซอฟต์เอสคิวแอลเซิร์ฟเวอร์ 2008 อาร์ทู (Microsoft SQL Server R2) เป็นระบบฐานข้อมูลในการพัฒนาคลังข้อมูล

3) โปรแกรมไมโครซอฟต์เอสคิวแอลเซิร์ฟเวอร์อนาไลซิสเซอร์วิสส์ (Microsoft SQL Server Analysis Services) เป็นเครื่องมือสร้างเดต้า (Data)

2.2.2 เครื่องคอมพิวเตอร์ลูกข่าย ที่มีคุณสมบัติดังนี้

1) ระบบปฏิบัติการไมโครซอฟต์วินโดวส์ตั้งแต่ Windows 7 ขึ้นไป

2) โปรแกรม Internet Explorer 10 ขึ้นไป

3) โปรแกรมไมโครซอฟต์เอสคิวแอลเซิร์ฟเวอร์ 2008 อาร์ทู (Microsoft SQL Server R2) เป็นระบบฐานข้อมูลในการพัฒนาคลังข้อมูล

4) โปรแกรมไมโครซอฟต์เอสคิวแอลเซิร์ฟเวอร์อนาไลซิสเซอร์วิสส์ (Microsoft SQL Server Analysis Services) เป็นเครื่องมือสร้างเดต้า (Data)

5) โปรแกรม ไมโครซอฟต์เอกซ์เซล 2007 (Microsoft Excel 2007)

6) โปรแกรมเวก้า (WEKA)

2.3 แบบสอบถามเพื่อประเมินผล

2.3.1 แบบสอบถามความพึงพอใจการใช้งานคลังข้อมูลผลการเรียนของนักเรียนระดับมัธยมศึกษา

2.3.2 แบบสอบถามความเชื่อมั่นของระบบพยากรณ์ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 4 เพื่อใช้ในการประกอบการตัดสินใจเลือกแผนการเรียนในระดับมัธยมศึกษาตอนปลาย

3. ขั้นตอนการดำเนินงาน

งานวิจัยนี้ประกอบด้วยขั้นตอนสำคัญ 5 ขั้นตอน ได้แก่

3.1 การพัฒนาคลังข้อมูลนักเรียน

3.2 การพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ

3.3 การทำเหมืองข้อมูล

3.4 การประเมินผล

3.5 สรุปผลและข้อเสนอแนะ

3.1 การพัฒนาคลังข้อมูลนักเรียน

3.1.1 วิเคราะห์ความต้องการของผู้ใช้งาน

วิเคราะห์ความต้องการของผู้ใช้งานซึ่งได้แก่ ผู้บริหาร หัวหน้ากลุ่มสาระการเรียนรู้ และอาจารย์ประจำชั้น โดยการสัมภาษณ์ถึงปัญหาและความต้องการ เพื่อใช้ในการออกแบบคลังข้อมูล

3.1.2 การออกแบบคลังข้อมูล

ผู้วิจัยออกแบบคลังข้อมูล โดยออกแบบโครงสร้างคลังข้อมูลด้วยข้อมูลที่ได้จากฐานข้อมูลงานทะเบียนและประมวลผล ทำการกำหนดมุมมองหรือมิติ ค่าที่ต้องการวัด ขอบเขต ประเภท และขนาดของตาราง เพื่อใช้ในการออกแบบตารางข้อเท็จจริง (Fact Table) และตารางมิติ (Dimension Table) จากนั้นกำหนดความสัมพันธ์ระหว่างตารางข้อเท็จจริงและตารางมิติ จากข้อมูลค่าที่ต้องการวัดและมิติของข้อมูล ซึ่งใช้โครงสร้างคลังข้อมูลแบบสโนว์เฟลค (Snowflake Schema)

3.1.3 การพัฒนาคลังข้อมูล

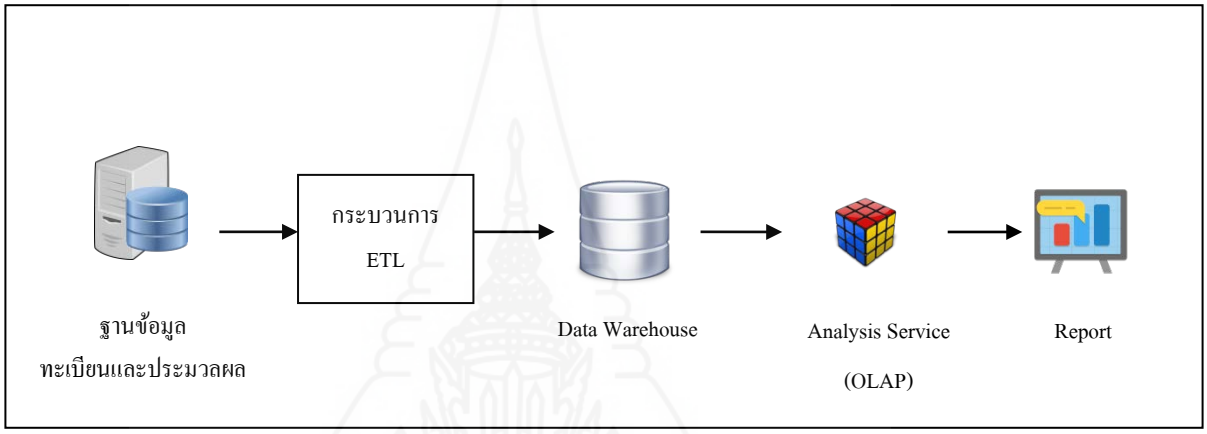
กระบวนการที่สำคัญในการนำข้อมูลเข้าคลังข้อมูลคือ กระบวนการอีทีแอล (Extract Transform and Load : ETL) ซึ่งประกอบด้วย 3 ขั้นตอน คือ

1) การคัดแยกข้อมูล (Extract) เป็นการคัดแยกและเลือกข้อมูลนักเรียนที่ต้องการจากฐานข้อมูลงานทะเบียนและประมวลผลที่มี

2) การแปลงข้อมูล (Transform) เป็นการปรับรูปแบบและโครงสร้างข้อมูลนักเรียนผลการเรียนที่ได้ และขจัดข้อมูลที่ไม่ถูกต้องหรือไม่สอดคล้องกันของข้อมูล

3) การนำเข้าข้อมูล (Load) หลังจากที่มีคัดแยกและจัดรูปแบบข้อมูลเรียบร้อยแล้ว เข้าสู่คลังข้อมูลเพื่อวิเคราะห์และแสดงผลตามที่ได้ออกแบบไว้

งานวิจัยนี้ผู้วิจัยนำข้อมูลมาจากฐานข้อมูลของโปรแกรมงานทะเบียนและประมวลผล โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา นำข้อมูลมาผ่านกระบวนการอีทีแอลเพื่อเลือกหรือคัดกรองข้อมูลที่ต้องการ จัดรูปแบบข้อมูล และทำความสะอาดข้อมูลให้มีความถูกต้องสอดคล้องกัน จากนั้นจึงนำข้อมูลที่ได้เข้าสู่คลังข้อมูลเพื่อจัดเก็บ แล้วจึงดำเนินการนำข้อมูลในคลังข้อมูลไปวิเคราะห์ข้อมูลหลายมิติ สำหรับเครื่องมือที่ใช้ในการวิจัยนี้คือ Microsoft SQL Server Analysis Services ในการทำรายงานหรือนำเสนอข้อมูล ดังภาพที่ 3.1



ภาพที่ 3.1 แสดงขั้นตอนการพัฒนาคลังข้อมูล

3.2 การพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ

3.2.1 การสร้างคิวบ์ข้อมูล (Data Cube)

ผู้วิจัยได้ใช้โปรแกรม Microsoft SQL Server 2008 Analysis Services (SSAS) ซึ่งเป็นเครื่องมือสำหรับการประมวลผลเชิงวิเคราะห์ออนไลน์ (OnLine Analytical Procesing - OLAP) ในการสร้างคิวบ์ข้อมูล และสร้างมุมมองตามความต้องการของกลุ่มผู้ใช้งาน

3.2.2 การสร้างรายงาน (Report)

การสร้างรายงานผู้วิจัยใช้โปรแกรม QlikView ในการนำเสนอข้อมูลตามที่ได้ทำการสร้างคิวบ์ข้อมูล และสร้างมุมมอง ซึ่งรายงานจะออกแบบตามความต้องการของผู้ใช้งาน ซึ่งแบ่งออกเป็น 3 กลุ่ม ได้แก่ ผู้บริหาร หัวหน้ากลุ่มสาระการเรียนรู้ และอาจารย์ประจำชั้น

3.3 การทำเหมืองข้อมูล

งานวิจัยนี้ใช้การทำเหมืองข้อมูลตามแบบจำลองของคริสพ์ – ดีเอ็ม (CRISP – DM Model) ซึ่งประกอบด้วย 6 ขั้นตอน ดังนี้

3.3.1 ทำความเข้าใจปัญหา (Business Understanding)

วัตถุประสงค์ของงานวิจัยนี้เพื่อใช้เทคนิคเหมืองข้อมูลในการสร้างระบบพยากรณ์ผลการเรียนเฉลี่ย เพื่อใช้ในการประกอบการตัดสินใจของนักเรียนมัธยมศึกษาตอนต้นในการเลือกแผนการเรียนระดับมัธยมศึกษาตอนปลาย

3.3.2 ทำความเข้าใจข้อมูล (Data Understanding)

ข้อมูลในการทำเหมืองข้อมูลเป็นข้อมูลที่มาจากคลังข้อมูลนักเรียนด้วยกระบวนการ อีทีแอล (Extract Transform and Load: ETL) โดยใช้เครื่องมือ SQL Server Integration Services (SSIS) ของโปรแกรม Microsoft SQL Server 2008 R2 เพื่อนำข้อมูลออกมาให้อยู่ในรูปแบบของ Excel File ซึ่งประกอบด้วยข้อมูลของนักเรียนในระดับชั้นมัธยมศึกษาปีที่ 4 ระหว่างปีการศึกษา 2553 – 2556 ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน ปี 2551 จำนวน 525 เรคคอร์ด

3.3.3 เตรียมข้อมูล (Data Preparation)

ข้อมูลที่ได้จากคลังข้อมูลถูกนำออกมาอยู่ในรูปแบบของไฟล์ Excel ยังไม่อยู่ในรูปแบบที่สามารถนำมาวิเคราะห์ด้วยวิธีการทำเหมืองข้อมูลได้ เพื่อให้ได้ผลการวิจัยที่มีประสิทธิภาพ ผู้วิจัยได้ทำขั้นตอนการเตรียมข้อมูล (Data Preparation) เพื่อใช้ในการทำเหมืองข้อมูลดังต่อไปนี้

1) การคัดกรองข้อมูล (Data Cleaning) คัดกรองข้อมูลที่ไม่เกี่ยวข้อง และข้อมูลที่ผิดปกติออกไป โดยข้อมูลนักเรียนที่นำมาใช้เป็นข้อมูลอยู่ระหว่างปีการศึกษา 2553 – 2556 ระดับชั้นมัธยมศึกษาปีที่ 4 มีจำนวน 525 คน

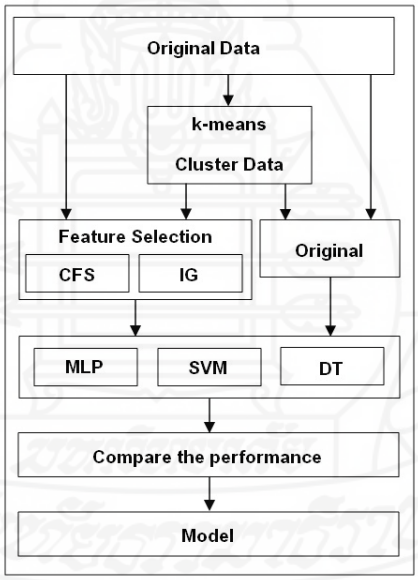
2) การเลือกข้อมูล (Data Selection) ข้อมูลที่ใช้ในการวิเคราะห์ ผู้วิจัยได้คัดเลือกเลือกคุณลักษณะหรือตัวแปรที่เกี่ยวข้องกับผลการเรียนเฉลี่ยในการวิจัยครั้งนี้

3) การแปลงข้อมูล (Data Transformation) เป็นการแปลงข้อมูลให้เหมาะสมสำหรับการทำเหมืองข้อมูลด้วยโปรแกรม WEKA Version 3.6.12 ใช้ฟังก์ชันในโปรแกรม Microsoft Excel 2007 ในการแปลงข้อมูลผลการเรียนเฉลี่ย ให้อยู่ในรูปแบบของไฟล์ CSV จากนั้นจึงนำไฟล์เข้าโปรแกรม WEKA เพื่อบันทึกให้อยู่ในรูปแบบไฟล์ arff

3.3.4 สร้างแบบจำลองหรือโมเดล (Modeling)

การสร้างตัวแบบพยากรณ์ โดยข้อมูลออกเป็น 2 แบบ ได้แก่ ชุดข้อมูลที่ไม่จัดกลุ่ม (Original Data) เป็นข้อมูลนักเรียนจำนวน 525 ระเบียน และชุดข้อมูลที่จัดกลุ่ม (Cluster Data) เป็นข้อมูลนักเรียนที่นำมาผ่านการจัดกลุ่มข้อมูลด้วยอัลกอริทึม k-means ผู้วิจัยเลือกใช้ค่า

ผิดพลาดในการจำแนกกลุ่ม (Incorrectly Clustered Instances : ICI) ที่น้อยที่สุดในการแบ่งกลุ่มข้อมูล จากนั้นนำไปผ่านกระบวนการคัดเลือกคุณลักษณะ (Feature Selection) ซึ่งใช้วิธี Correlation-based Feature Selection (CFS) และวิธี Information Gain (IG) ก่อนเพื่อที่ให้ได้คุณลักษณะที่สำคัญ (Bi, Zhang, and Cheng, 2009) นำไปสร้างตัวแบบพยากรณ์ หลังจากผ่านกระบวนการคัดเลือกคุณลักษณะ จึงใช้เทคนิคเหมืองข้อมูลโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (Multi-layer Perceptron : MLP) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) โดยใช้เคอร์เนลฟังก์ชันแบบ Normalise Polynomial Kernel (SVM-NP) Polynomial Kernel (SVM-PK) Radial Basis Function Kernel (SVM-RBF) Pearson VII function-based universal kernel (SVM-PUK) และเทคนิคต้นไม้ตัดสินใจ (Decision Tree) ใช้อัลกอริทึมแบบ C4.5 เพื่อเปรียบเทียบหาค่าความถูกต้อง (Accuracy) และ ค่ารากที่สองของความคลาดเคลื่อน (Root Mean Square Error : RMSE) ที่ดีที่สุดในการพยากรณ์แล้วนำไปพัฒนาเป็นระบบพยากรณ์ผลการเรียน ดังภาพที่ 3.2



ภาพที่ 3.2 กระบวนการสร้างตัวแบบพยากรณ์

3.3.5 ประเมินประสิทธิภาพ (Evaluation)

การประเมินประสิทธิภาพตัวแบบพยากรณ์ ใช้การประเมินประสิทธิภาพของแบบ K-fold Cross Validation โดยกำหนดค่า k เท่ากับ 10 เพื่อแบ่งชุดข้อมูลสำหรับการสอนและการทดสอบ ใช้วิธีการวัดค่าความถูกต้อง (Accuracy) ในการสร้างตัวแบบพยากรณ์ และค่าราก

ที่สองของความคลาดเคลื่อน (Root Mean Square Error : RMSE) ส่วนการจัดกลุ่มข้อมูล (Clustering) ผู้วิจัยใช้การประเมินประสิทธิภาพของตัวแบบในการจัดกลุ่ม โดยวัดค่าผิดพลาดในการจำแนกกลุ่ม (Incorrectly Clustered Instances : ICI) เพื่อใช้ในการเลือกตัวแบบที่มีความผิดพลาดน้อยที่สุด ด้วยอัลกอริทึม K-Means Clustering โดยใช้โปรแกรม WEKA

3.3.6 การนำไปใช้งาน (Deployment)

เป็นการนำตัวแบบพยากรณ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูลมาพัฒนาสร้างระบบพยากรณ์ผลการเรียนเฉลี่ย เพื่อนำไปใช้กับนักเรียนระดับมัธยมศึกษาตอนต้น เพื่อใช้ประกอบการตัดสินใจเลือกแผนการเรียน และวางแผนการเรียนเพื่อพัฒนาผลการเรียนให้ดีขึ้น

3.4 การประเมินผล

ในงานวิจัยจะใช้แบบสอบถามเพื่อประเมินความพึงพอใจระบบคลังข้อมูลผลการเรียนและแบบประเมินความเชื่อมั่นของระบบพยากรณ์ผลการเรียน โดยมีขั้นตอนดังนี้

1) การออกแบบเครื่องมือการประเมิน ในออกแบบเครื่องมือประเมินความพึงพอใจของผู้ใช้ระบบคลังข้อมูลผลการเรียน และประเมินความเชื่อมั่นของระบบพยากรณ์ผลการเรียน ซึ่งใช้แบบสอบถามเป็นเครื่องมือในประเมิน แบบสอบถามที่ใช้เป็นแบบมาตราส่วนแบบ Rating Scale 5 ระดับ คือ มากที่สุด, มาก, ปานกลาง, น้อย และน้อยที่สุด

2) การประเมินผล เป็นการดำเนินการหลังจากนำแบบสอบถามที่ได้จากการประเมินของกลุ่มผู้ใช้งาน มาวิเคราะห์และแปลผล โดยใช้ค่าสถิติการวัดหาค่ากลางของข้อมูล ได้แก่ ค่าเฉลี่ยเลขคณิต (Arithmetic Mean)

3) เกณฑ์การพิจารณา ในการพิจารณาแบบประเมิน ได้กำหนดเกณฑ์การพิจารณาคุณภาพ โดยใช้ค่าเฉลี่ยเลขคณิต ($\bar{x}$) แล้วนำค่าเฉลี่ยที่ได้มาเปรียบเทียบกับเกณฑ์ของ เบสท์ (Best.,1963: 204-208)

3.5 สรุปผลการวิจัยและข้อเสนอแนะ

สรุปผลการวิจัยที่ได้ จากการพัฒนาคลังข้อมูล และการใช้เทคนิคเหมืองข้อมูลมาสร้างระบบพยากรณ์ผลการเรียน และอภิปรายผลงานวิจัย รวมทั้งข้อเสนอแนะ

บทที่ 4
ผลการวิเคราะห์ข้อมูล

การศึกษาวิจัยเรื่อง การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา ผู้วิจัยได้พัฒนาคลังข้อมูลผลการเรียนของนักเรียนระดับมัธยมศึกษาโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา ในการนำเสนอรายงาน และนำมาวิเคราะห์หาความรู้ใหม่โดยการทำเหมืองข้อมูลตามรูปแบบและขั้นตอนของแบบจำลอง คริสพ์ – ดีเอ็ม (CRISP – DM Model) ซึ่งผลการวิจัยและการวิเคราะห์ข้อมูลประกอบด้วย 4 ส่วน ได้แก่

1. การพัฒนาคลังข้อมูลผลการเรียน
2. การพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ
3. การทำเหมืองข้อมูล
4. การประเมินผล

1. การพัฒนาคลังข้อมูลผลการเรียน

1.1 การวิเคราะห์ความต้องการของผู้ใช้งาน

จากการสัมภาษณ์และสอบถามข้อมูลจาก ผู้บริหาร หัวหน้ากลุ่มสาระการเรียนรู้และอาจารย์ประจำชั้น ที่เป็นผู้ใช้งานข้อมูลผลการเรียนของนักเรียนในระดับมัธยมศึกษา พบว่าปัญหาและความต้องการของผู้ใช้งานสามารถสรุปได้ดังนี้

1.1.1 โปรแกรมที่ใช้อยู่ในปัจจุบันเป็นการใช้งานบนระบบปฏิบัติการดอส (Dos) ซึ่งในปัจจุบันไม่ค่อยมีผู้ใช้งานระบบปฏิบัติการนี้ ไม่รองรับกับฮาร์ดแวร์ที่ใช้งานอยู่ และมีขั้นตอนที่ยุ่งยากในการใช้งาน ร่วมถึงการเข้าถึงข้อมูลของผู้ใช้งานอย่างผู้บริหาร หัวหน้ากลุ่มสาระการเรียนรู้ และอาจารย์ประจำชั้น ต้องให้เจ้าหน้าที่นำข้อมูลที่ต้องการออกมาให้ ผู้ใช้งานอยากได้ระบบที่ทำงานบนระบบปฏิบัติการวินโดว์ ที่ใช้งานได้ง่าย

1.1.2 ผู้ใช้งานสามารถดูข้อมูลผลการเรียนที่เป็นรายงานในรูปแบบของกระดาษ และต้องรอให้เจ้าหน้าที่สั่งพิมพ์รายงานข้อมูลออกมาให้เท่านั้น ซึ่งความต้องการของผู้ใช้งานต้องการรายงานที่สามารถดูได้จากคอมพิวเตอร์

1.1.3 ข้อมูลผลการเรียนไม่ได้นำมาใช้ประโยชน์อย่างอื่น นอกจากรายงานผลให้กับนักเรียนได้ทราบ ซึ่งผู้ใช้งานต้องการให้ข้อมูลผลการเรียนสามารถนำมาใช้ประโยชน์ เช่น การเป็นข้อมูลเปรียบเทียบผลการเรียนเฉลี่ยของแต่ละปี รวมถึงจำนวนนักเรียนในแต่ละปีการศึกษา และคะแนนของนักเรียนที่ทำได้ในแต่ละวิชาของแต่ละภาคการศึกษา เป็นต้น

จากผลที่ได้ผู้วิจัยได้ทำการศึกษา วิเคราะห์ความต้องการ และศึกษาจากงานวิจัยต่างๆ ที่เกี่ยวข้อง จากงานวิจัยของ สุประพล เลาวพงศ์ (2553: 1-3) ได้พัฒนาคลังข้อมูลด้านนักศึกษาของมหาวิทยาลัยพายัพเพื่อแก้ปัญหาระบบงานเดิมที่ไม่ได้ออกแบบการจัดเก็บข้อมูลของนักศึกษา และรูปแบบของข้อมูลที่แตกต่างกัน ไม่สามารถนำมาใช้ประโยชน์ในด้านข้อมูลได้ จึงพัฒนาคลังข้อมูลมาเพื่อแก้ปัญหา ซึ่งผลจากพัฒนาคลังข้อมูลสามารถเป็นใช้ประโยชน์ของข้อมูลในการตัดสินใจของผู้บริหาร และผู้ใช้งานได้ ดังนั้นผู้วิจัยจึงพัฒนาคลังข้อมูลผลการเรียนของนักเรียนเพื่อแก้ปัญหาที่เกิดขึ้น

1.2 การออกแบบคลังข้อมูล

ในการออกแบบคลังข้อมูล ผู้วิจัยได้ออกแบบโครงสร้างคลังข้อมูล โดยข้อมูลที่ได้จากฐานข้อมูลงานทะเบียนและประมวลผล ทำจัดการข้อมูลก่อนนำเข้าสู่คลังข้อมูลที่มีความสัมพันธ์กัน ทั้งการจัดเก็บข้อมูลในฐานข้อมูลเดิมและความต้องการข้อมูลของการวิเคราะห์ข้อมูล ดังนั้นจำเป็นต้องทราบถึงลักษณะของข้อมูลที่ถูกจัดเก็บในฐานข้อมูลระบบผลการเรียน แล้วนำมาแบ่งกลุ่มข้อมูลที่จัดเก็บในคลังข้อมูล สามารถแบ่งกลุ่มข้อมูลได้ดังนี้

1) ข้อมูลรายละเอียดนักเรียน เช่น รหัสนักเรียน ชื่อ นามสกุล วันเดือนปีเกิด รายได้ของพ่อแม่

2) ข้อมูลอาจารย์ผู้สอน เช่น ตำแหน่ง ชื่อ นามสกุล วุฒิการศึกษา

3) ข้อมูลวิชาเรียน เช่น รหัสวิชา ชื่อวิชา ตัวย่อวิชา หน่วยกิต กลุ่มสาระการเรียนรู้ หลักสูตร

4) ข้อมูลปีการศึกษา เช่น ปีการศึกษา ภาคเรียน

5) ข้อมูลผลการเรียน

6) ข้อมูลผลการเรียนเฉลี่ย

จากข้อมูลที่ได้จากฐานข้อมูลผลการเรียนทำการกำหนดมุมมองหรือมิติ และค่าที่ต้องการวัด เพื่อใช้ในการออกแบบตารางข้อเท็จจริงและตารางมิติ ดังนี้

ตารางที่ 4.1 ค่าที่ต้องการวัดและมิติของผลการเรียน

ลำดับ	ประเภท	ชื่อ	ตัวอย่างข้อมูล
1	มุมมองหรือมิติ	นักเรียน	รหัสนักเรียน 2017 กานต์ธิดา ใจดี เพศ หญิง
2	มุมมองหรือมิติ	อาจารย์ผู้สอน	ผศ. ขจรรัตน์ อุดมศรี
3	มุมมองหรือมิติ	วิชา	ท 21101 ภาษาไทย 1 1.5 หน่วยกิต
4	มุมมองหรือมิติ	ประเภทของวิชาเรียน	วิชาพื้นฐาน, วิชาเลือก
5	มุมมองหรือมิติ	ระดับชั้นการศึกษา	มัธยมศึกษาปีที่ 1 ห้อง 1
6	มุมมองหรือมิติ	ปีการศึกษาและภาคเรียน	ปีการศึกษา 2556 ภาคเรียนที่ 1
7	มุมมองหรือมิติ	รายชื่อจังหวัด	นครปฐม, สุพรรณบุรี เป็นต้น
8	มุมมองหรือมิติ	รายชื่อกลุ่มสาระการเรียนรู้	ภาษาไทย, วิทยาสาสตร์, คณิตศาสตร์ เป็นต้น
9	มุมมองหรือมิติ	เพศ	ชาย, หญิง
10	ค่าที่ต้องการวัด	คะแนน	ตัวเลขจำนวนเต็ม เช่น 54 , 70 , 86 เป็นต้น
11	ค่าที่ต้องการวัด	ผลการเรียน (เกรด)	จำนวนทศนิยม 0 – 4 เช่น 2.54 , 3.78 เป็นต้น
12	ค่าที่ต้องการวัด	ผลการเรียนเฉลี่ย	จำนวนทศนิยม 0 – 4 เช่น 2.54 , 3.78 เป็นต้น
13	ค่าที่ต้องการวัด	ผลการเรียนเฉลี่ยสะสม	จำนวนทศนิยม 0 – 4 เช่น 2.54 , 3.78 เป็นต้น

จากตารางที่ 4.1 ทำการกำหนดมุมมองหรือมิติ ค่าที่ต้องการวัด ขอบเขต ประเภท และขนาดของตาราง เพื่อใช้ในการออกแบบตารางข้อเท็จจริง (Fact Table) และตารางมิติ (Dimension Table) ซึ่งมีตารางข้อเท็จจริง 2 ตาราง คือ Fact_Grade และ Fact_GPA และมีตารางมิติ 9 ตาราง คือ Dim_Area, Dim_Class, Dim_Department, Dim_Group, Dim_Gender, Dim_Semester, Dim_Student, Dim_Teacher และ Dim_Subject

ตารางที่ 4.2 ตาราง Fact_Grade สำหรับเก็บข้อมูลคะแนนและผลการเรียน

ลำดับที่	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	no	interger	ลำดับรายการ	Primary Key
2	student_id	interger	รหัสนักเรียน	Foreign Key (Dim_Student)
3	class_id	interger	รหัสระดับชั้น	Foreign Key (Dim_Class)
4	subject_id	interger	รหัสวิชา	Foreign Key (Dim_Subject)
5	teacher_id	interger	รหัสอาจารย์	Foreign Key (Dim_Teacher)
6	semester_id	interger	รหัสปีการศึกษา	Foreign Key (Dim_Semester)
7	score	interger	คะแนน	
8	grade	float	ผลการเรียน(เกรด)	

ตารางที่ 4.3 ตาราง Fact_GPA สำหรับเก็บข้อมูลผลการเรียนเฉลี่ยและผลการเรียนเฉลี่ยสะสม

ลำดับที่	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	no	interger	ลำดับรายการ	Primary Key
2	student_id	interger	รหัสนักเรียน	Foreign Key (Dim_Student)
3	class_id	interger	รหัสระดับชั้น	Foreign Key (Dim_Class)
4	semester_id	interger	รหัสปีการศึกษา	Foreign Key (Dim_Semester)
5	GPA	float	ผลการเรียนเฉลี่ย	
6	GPAX	float	ผลการเรียนเฉลี่ยสะสม	

ตารางที่ 4.4 ตาราง Dim_Subject สำหรับเก็บข้อมูลวิชาเรียน

ลำดับที่	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	subject_id	interger	รหัสวิชา	Primary Key
2	subject_code	varchar	ตัวย่อวิชา	
3	subject_name	varchar	ชื่อวิชา	
4	credit	float	หน่วยกิต	
5	department_id	interger	รหัสกลุ่มสาระการเรียนรู้	Foreign Key (Dim_Department)
6	group_id	interger	รหัสประเภทวิชา	Foreign Key (Dim_Group)
7	course	varchar	หลักสูตรการศึกษา	

ตารางที่ 4.5 ตาราง Dim_Student สำหรับเก็บข้อมูลนักเรียน

ลำดับที่	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	student_id	interger	รหัสนักเรียน	Primary Key
2	student_name	varchar	ชื่อนักเรียน	
3	birthdate	varchar	วัน เดือน ปี เกิด	
4	gender_id	interger	เพศ	Foreign Key (Dim_Gender)
5	area_id	interger	จังหวัด	Foreign Key (Dim_Area)
6	brothers	interger	จำนวนพี่น้อง	
7	year_of_study	interger	ปีที่เข้าเรียน	
8	father	varchar	ชื่อบิดา	
9	career_father	varchar	อาชีพบิดา	
10	income_father	interger	รายได้ต่อปีของบิดา	
11	mother	varchar	ชื่อมารดา	
12	career_mother	varchar	อาชีพมารดา	
13	income_mother	interger	รายได้ต่อปีของมารดา	
14	marital_status	nvarchar	สถานภาพ สมรส	

ตารางที่ 4.6 ตาราง Dim_Semester สำหรับเก็บข้อมูลปีการศึกษาและภาคเรียน

ลำดับที่	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	semester_id	interger	รหัสปีการศึกษา	Primary Key
2	semester	varchar	ภาคเรียน	
3	year	varchar	ปีการศึกษา	

ตารางที่ 4.7 ตาราง Dim_Teacher สำหรับเก็บข้อมูลอาจารย์ผู้สอน

ลำดับที่	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	teacher_id	interger	รหัสอาจารย์	Primary Key
2	position	varchar	ตำแหน่ง	
3	teacher_name	varchar	ชื่ออาจารย์	
4	status	varchar	สถานะ	
5	education	varchar	วุฒิการศึกษา	

ตารางที่ 4.8 ตาราง Dim_Group สำหรับเก็บข้อมูลประเภทของวิชาเรียน

ลำดับที่	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	group_id	interger	รหัสประเภทวิชา	Primary Key
2	subject_group	varchar	ชื่อประเภทวิชา	

ตารางที่ 4.9 ตาราง Dim_Gender สำหรับเก็บข้อมูลเพศ

ลำดับที่	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	gender_id	interger	รหัสเพศ	Primary Key
2	gender	varchar	เพศ	

ตารางที่ 4.10 ตาราง Dim_Area สำหรับเก็บข้อมูลรายชื่อจังหวัด

ลำดับที่	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	area_id	interger	รหัสจังหวัด	Primary Key
2	province	varchar	จังหวัด	

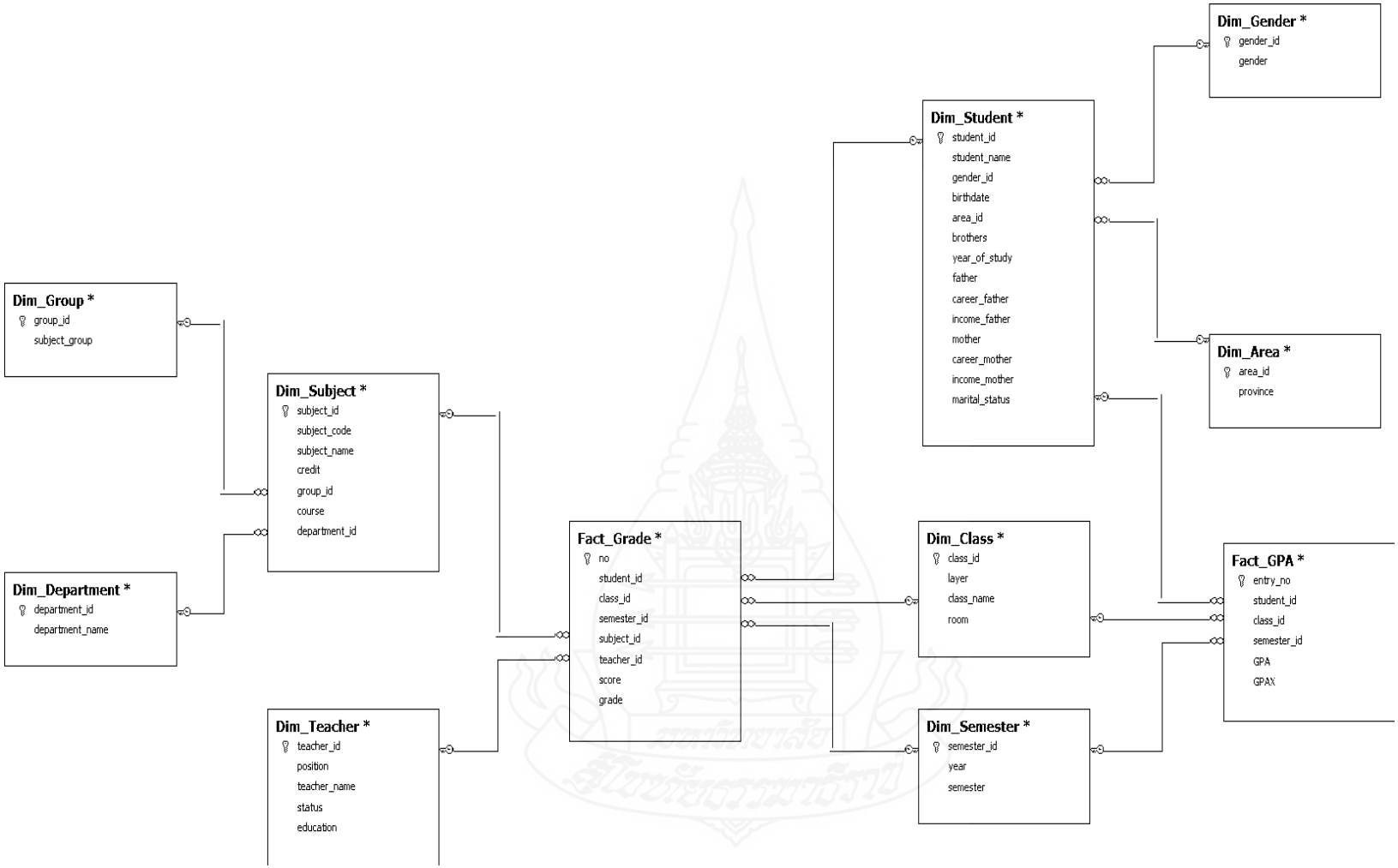
ตารางที่ 4.11 ตาราง Dim_Class สำหรับเก็บข้อมูลระดับชั้นการศึกษา

ลำดับที่	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	class_id	interger	รหัสระดับชั้น	Primary Key
2	class_name	varchar	ชื่อระดับชั้นการศึกษา	
3	room	varchar	ห้อง	
4	layer	varchar	ช่วงชั้นการศึกษา	

ตารางที่ 4.12 ตาราง Dim_Department สำหรับเก็บข้อมูลรายชื่อกลุ่มสาระการเรียนรู้

ลำดับที่	ชื่อฟิลด์	ประเภทข้อมูล	คำอธิบาย	Key & Index
1	department_id	interger	รหัสกลุ่มสาระการเรียนรู้	Primary Key
2	department_name	varchar	ชื่อกลุ่มสาระการเรียนรู้	

ผู้วิจัยได้ออกแบบโครงสร้างคลังข้อมูลโดยใช้โครงสร้างคลังข้อมูลแบบสโนว์เฟลค (Snowflake Schema) ที่มีตารางข้อเท็จจริง 2 ตาราง และตารางมิติ 9 ตาราง อยู่โดยรอบ ซึ่งตารางมิติบางตารางจะมีตารางมิติเชื่อมโยงกัน คือ ตารางมิตินักเรียน (Dim_Student) มีตารางมิติเพศ (Dim_Gender) และตารางจังหวัด (Dim_Area) เชื่อมโยงกัน และตารางวิชาเรียน (Dim_Subject) มีตารางประเภทวิชาเรียน (Dim_Group) และตารางกลุ่มสาระการเรียนรู้ (Dim_Department) เชื่อมโยงกัน ซึ่งความสัมพันธ์ระหว่างตารางข้อเท็จจริงและตารางมิติเป็นดังภาพที่ 4.1



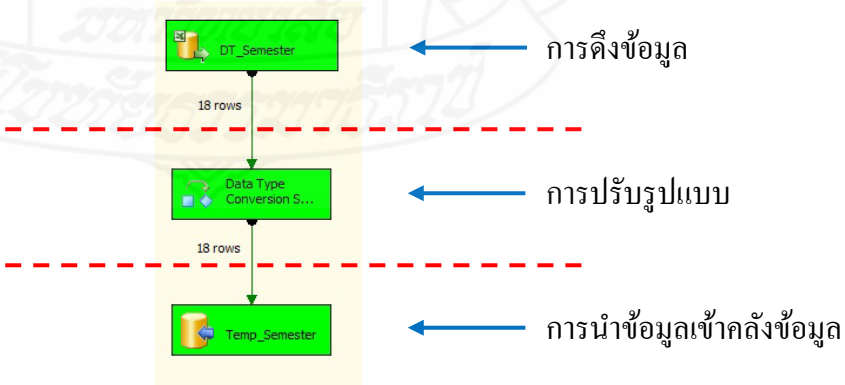
ภาพที่ 4.1 แสดงโครงสร้างคลังข้อมูลแบบสโนว์เฟลค

1.3 การพัฒนาคลังข้อมูล

ผู้วิจัยนำข้อมูลเข้าสู่คลังข้อมูลด้วยกระบวนการอีทีแอล (Extract Transform and Load : ETL) โดยใช้เครื่องมือ SQL Server Integration Services (SSIS) ของโปรแกรม Microsoft SQL Server 2008 R2 ดังภาพที่ 4.2

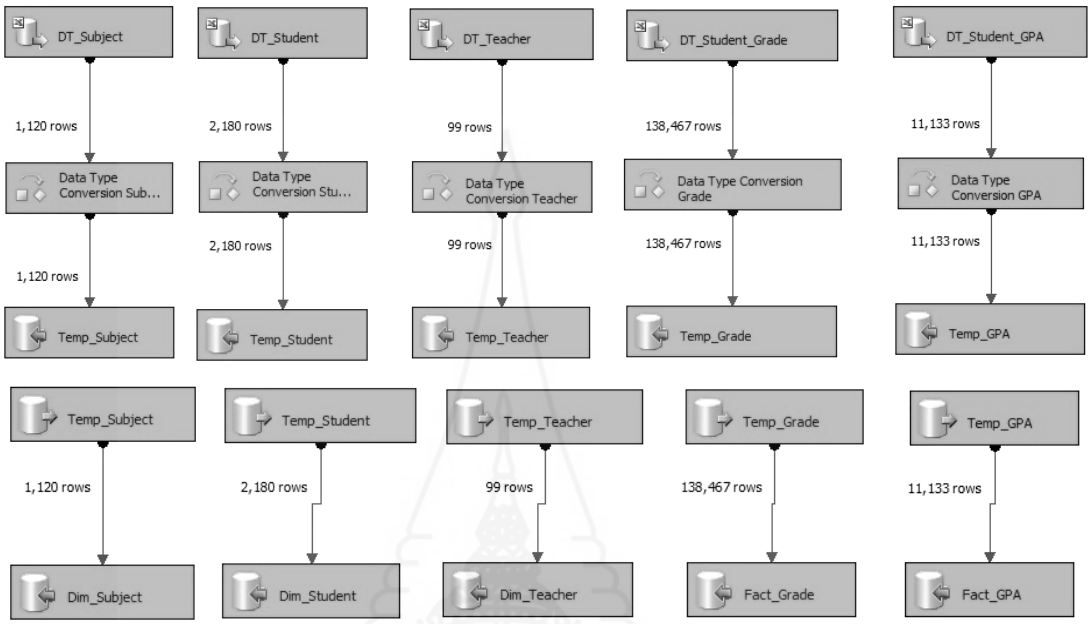
ภาพที่ 4.2 กระบวนการอีทีแอลด้วยเครื่องมือ SQL Server Integration Services

โดยกระบวนการนำเข้าข้อมูลแบ่งออกเป็น การถ่ายโอนนำข้อมูลเข้าสู่ที่พักข้อมูล จากนั้นทำการคัดลอกข้อมูลจากที่พักเข้าสู่ตารางข้อเท็จจริง และข้อมูลตารางมิติหรือมุมมองที่ออกแบบไว้ และขั้นสุดท้ายเป็นการล้างข้อมูลออกจากที่พักข้อมูล ดังภาพที่ 4.3



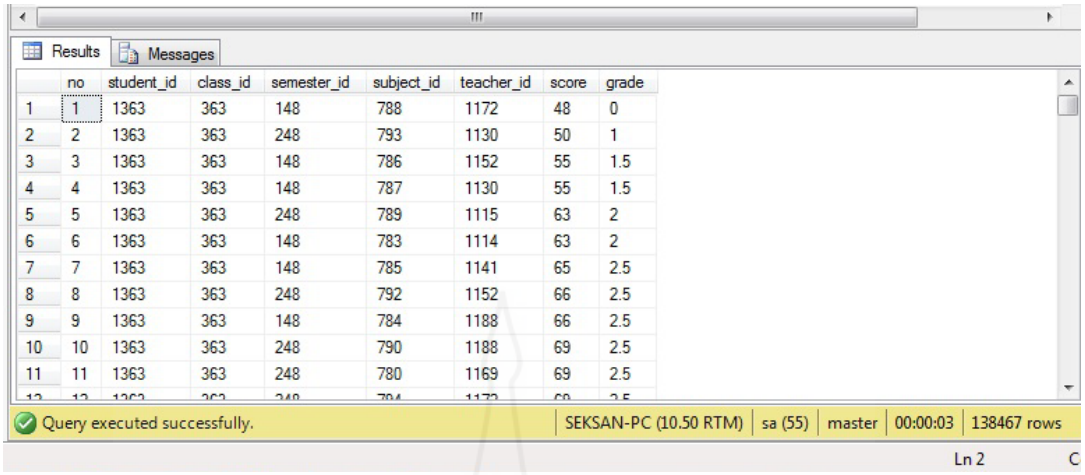
ภาพที่ 4.3 ขั้นตอนกระบวนการอีทีแอล

ผลของกระบวนการอีทีแอลด้วยเครื่องมือ SQL Server Integration Services (SSIS) เป็น ดังภาพที่ 4.4



ภาพที่ 4.4 แสดงผลของกระบวนการอีทีแอล

หลังจากนำข้อมูลเข้าคลังข้อมูล พบว่า มีจำนวนข้อมูลในตารางผลการเรียน (Fact_Grade) มีจำนวน 138,467 เรคคอร์ด ข้อมูลในตารางผลการเรียนเฉลี่ย (Fact_GPA) มีจำนวน 11,133 เรคคอร์ด ข้อมูลในตารางรายวิชา (Dim_Subject) มีจำนวน 1,120 เรคคอร์ด ข้อมูลในตารางนักเรียน (Dim_Student) มีจำนวน 2,180 เรคคอร์ด ข้อมูลในตารางอาจารย์ (Dim_Teacher) มีจำนวน 99 เรคคอร์ด ดังภาพที่ 4.5



Results | Messages

	no	student_id	class_id	semester_id	subject_id	teacher_id	score	grade
1	1	1363	363	148	788	1172	48	0
2	2	1363	363	248	793	1130	50	1
3	3	1363	363	148	786	1152	55	1.5
4	4	1363	363	148	787	1130	55	1.5
5	5	1363	363	248	789	1115	63	2
6	6	1363	363	148	783	1114	63	2
7	7	1363	363	148	785	1141	65	2.5
8	8	1363	363	248	792	1152	66	2.5
9	9	1363	363	148	784	1188	66	2.5
10	10	1363	363	248	790	1188	69	2.5
11	11	1363	363	248	780	1169	69	2.5

Query executed successfully. | SEKSAN-PC (10.50 RTM) | sa (55) | master | 00:00:03 | 138467 rows

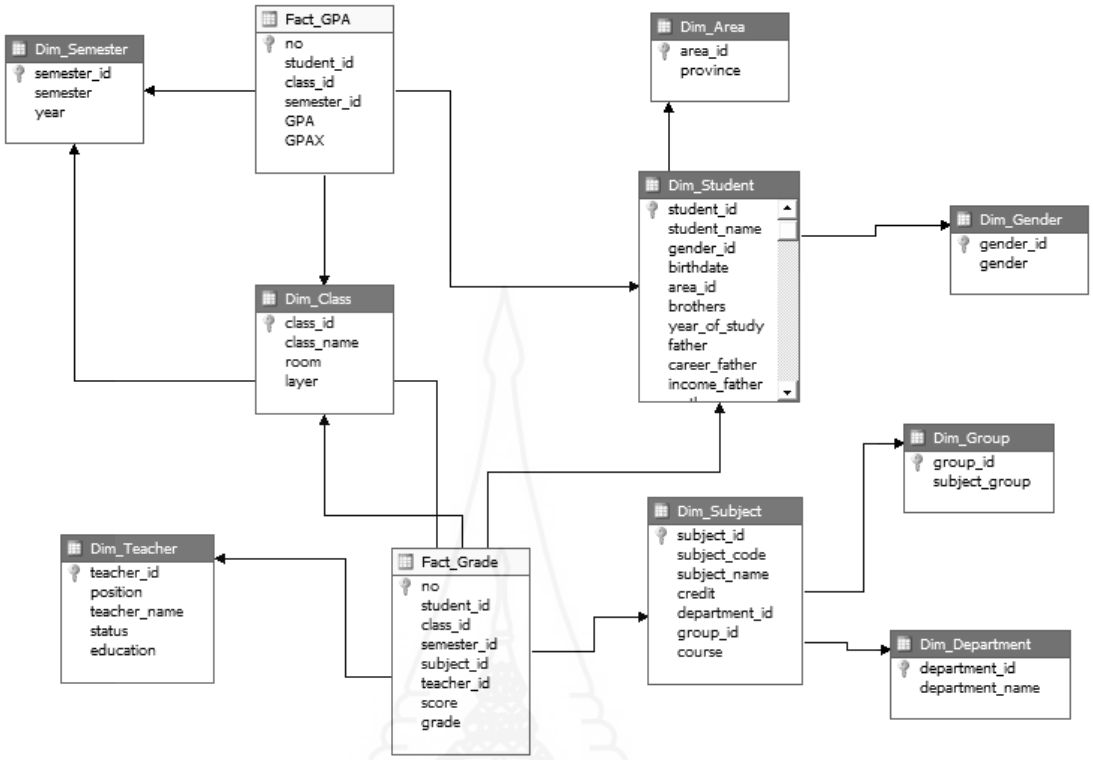
Ln 2 C

ภาพที่ 4.5 แสดงจำนวนข้อมูลในคลังข้อมูล

2. การพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ

2.1 การสร้างคิวบ์ข้อมูล

ผลจากการวิจัยได้สร้าง Data Source ในส่วนของ SQL Server Analysis Service (SSAS) เชื่อมต่อกับคลังข้อมูล จากนั้นสร้าง Data Source View ซึ่งประกอบด้วยมิติมุมมอง (Dimension) คือ มิติภาคเรียน (Dim Semester) , มิติระดับชั้น (Dim Class), มิติวิชา (Dim Subject), มิตินักเรียน (Dim Student) และมิติอาจารย์ (Dim Teacher) โดยมีเมเชอร์หรือค่าที่ต้องการวัด (Measure) คือ คะแนน (Score) , ผลการเรียน (Grade) และผลการเรียนเฉลี่ย (GPA) ดังภาพที่ 4.6



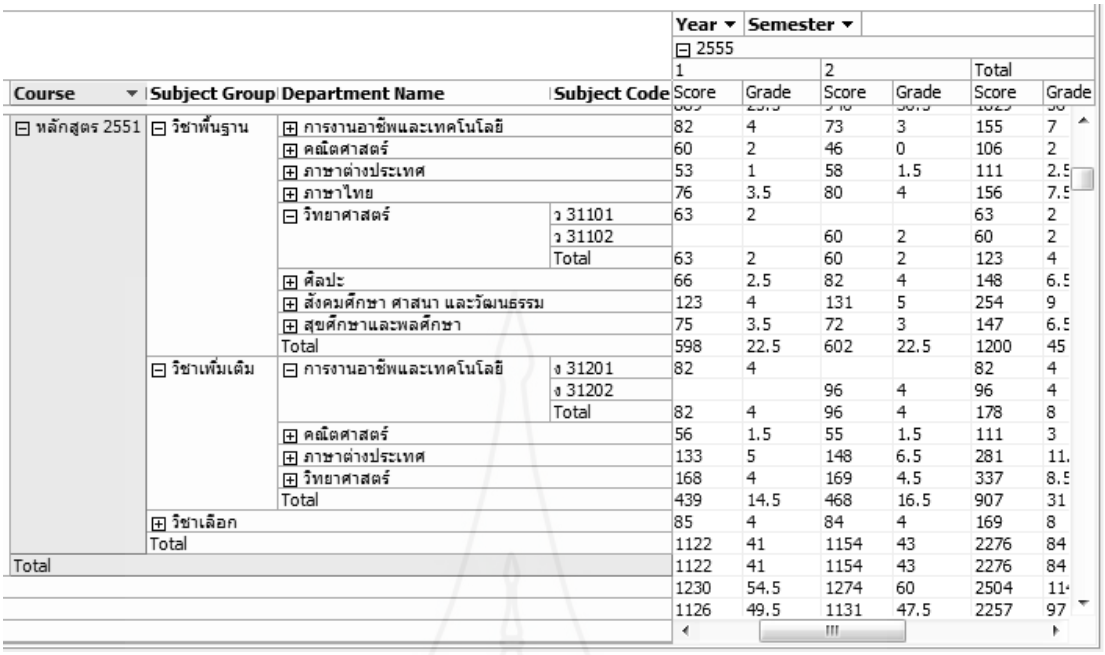
ภาพที่ 4.6 แสดงผลที่ได้จากการสร้าง Data Source View

จากนั้นสร้างคิวบ์ข้อมูล จะสามารถดูข้อมูลที่เป็นการประมวลผลเชิงวิเคราะห์ออนไลน์ในลักษณะหลายมิติได้ ซึ่งผู้ใช้งานสามารถเปลี่ยนมุมมอง (Dimension) คือ คือ มิติภาคเรียน (Dim_Semester), มิติระดับชั้น (Dim_Class), มิติวิชา (Dim_Subject), มิตินักเรียน (Dim_Student) และมิติอาจารย์ (Dim_Teacher) และเมเชอร์หรือค่าที่ต้องการวัด (Measure) คือ คะแนน (Score), ผลการเรียน (Grade) และผลการเรียนเฉลี่ย (GPA) ดังภาพที่ 4.7

		Year			Semester						
		2554			2555			2556			Grand Total
		1	2	Total	1	2	Total	1	2	Total	
Class Name	Subject Code	Grade	Grade	Grade	Grade	Grade	Grade	Grade	Grade	Grade	Grade
⊞ มัธยมศึกษาปีที่ 3			43.5	43.5							43.5
⊟ มัธยมศึกษาปีที่ 4	ค 31101				1.5		1.5				1.5
	ค 31102					0	0				0
	ค 31201				0		0				0
	ค 31204					4	4				4
	ง 31101				4		4				4
	ง 31102					4	4				4
	ง 31201				4		4				4
	ง 31202					4	4				4
	ท 31101				4		4				4
	ท 31102					4	4				4
	ฝ 30001				4		4				4
	ฝ 31202					2	2				2
	พ 31101				4		4				4
	พ 31102					3.5	3.5				3.5
	ว 31101				1.5		1.5				1.5
	ว 31102					3	3				3
	ว 31201				1		1				1
	ว 31221				2		2				2
	ว 31241				3		3				3
	ว 31282					4	4				4
	ศ 30043					4	4				4
	ศ 31101				4		4				4
	ศ 31102					4	4				4

ภาพที่ 4.7 แสดงตัวอย่างการประมวลผลเชิงวิเคราะห์แบบออนไลน์

จากภาพที่ 4.7 แสดงให้เห็นข้อมูลในมิติที่ด้านบนหรือคอลัมน์เป็นข้อมูลของปีการศึกษาและภาคเรียน ส่วนด้านซ้ายหรือแถวเป็นข้อมูลของนักเรียนในแต่ละระดับชั้น แสดงให้เห็นถึงผลการเรียนในแต่ละระดับชั้น เป็นรายวิชาในแต่ละปีการศึกษาแต่ละภาคเรียน เช่น ผลการเรียนในระดับชั้นมัธยมศึกษาปีที่ 4 ในรายวิชา ค 31101 ปีการศึกษา 2555 ภาคเรียนที่ 1 มีผลการเรียนคือ 1.5 เป็นต้น ซึ่งในการแสดงผลข้อมูลหลายมิติสามารถปรับเปลี่ยนข้อมูลหรือมิติในการดูข้อมูลได้ ดังแสดงในภาพที่ 4.8 และหากต้องการดูข้อมูลลึกลงไปอีกก็สามารถปรับเปลี่ยนรูปแบบการแสดงผลได้เช่นกัน โดยการเลือกข้อมูลที่ต้องการแสดงผล

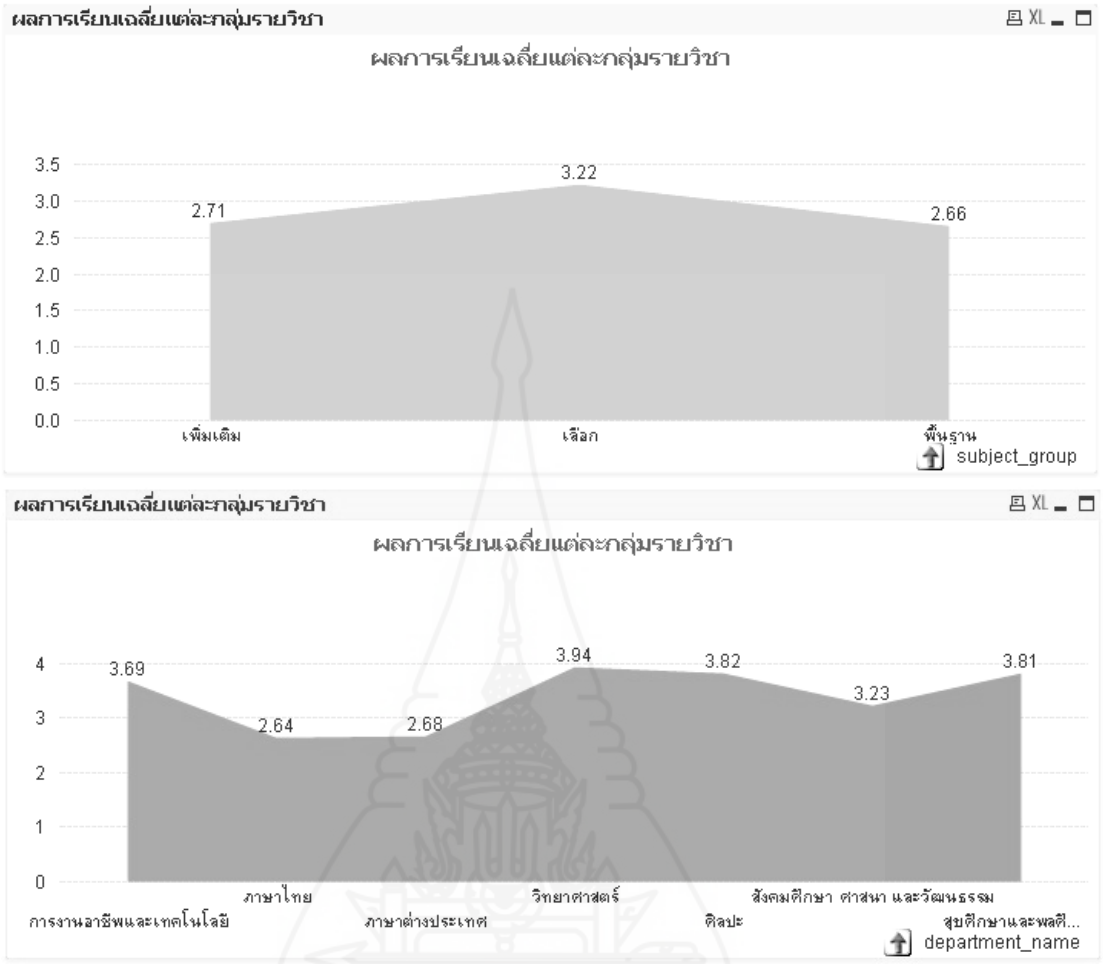


Course	Subject Group	Department Name	Subject Code	Year 2555 Semester 1 Score	Grade	Semester 2 Score	Grade	Total Score	Grade
หลักสูตร 2551	วิชาพื้นฐาน	การงานอาชีพและเทคโนโลยี		82	4	73	3	155	7
		คณิตศาสตร์		60	2	46	0	106	2
		ภาษาต่างประเทศ		53	1	58	1.5	111	2.5
		ภาษาไทย		76	3.5	80	4	156	7.5
		วิทยาศาสตร์	ว 31101	63	2			63	2
			ว 31102			60	2	60	2
			Total	63	2	60	2	123	4
		ศิลปะ		66	2.5	82	4	148	6.5
		สังคมศึกษา ศาสนา และวัฒนธรรม		123	4	131	5	254	9
		สุขศึกษาและพลศึกษา		75	3.5	72	3	147	6.5
		Total		598	22.5	602	22.5	1200	45
	วิชาเพิ่มเติม	การงานอาชีพและเทคโนโลยี	ง 31201	82	4			82	4
			ง 31202			96	4	96	4
			Total	82	4	96	4	178	8
		คณิตศาสตร์		56	1.5	55	1.5	111	3
		ภาษาต่างประเทศ		133	5	148	6.5	281	11.
		วิทยาศาสตร์		168	4	169	4.5	337	8.5
		Total		439	14.5	468	16.5	907	31
	วิชาเลือก			85	4	84	4	169	8
	Total			1122	41	1154	43	2276	84
Total				1122	41	1154	43	2276	84
				1230	54.5	1274	60	2504	11
				1126	49.5	1131	47.5	2257	97

ภาพที่ 4.8 ตัวอย่างการแสดงผลข้อมูลหลายมิติ

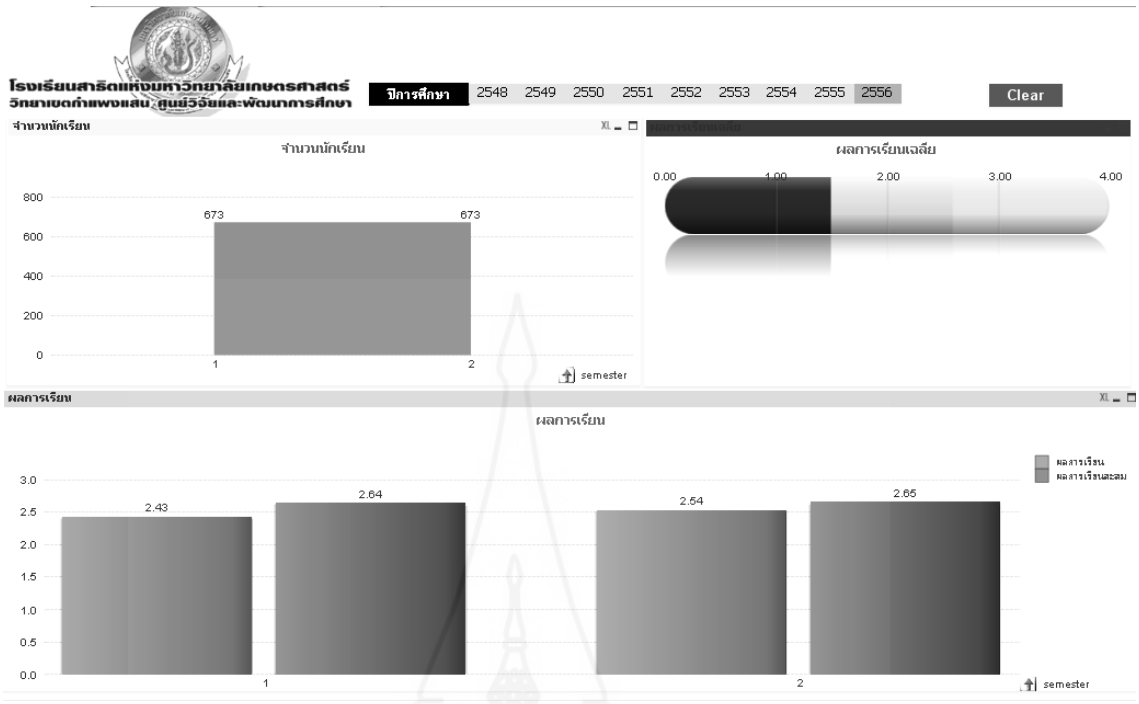
2.2 การสร้างรายงาน

ในการนำเสนอรายงานข้อมูลใช้โปรแกรม QlikView สร้างส่วนของการติดต่อกับคลังข้อมูล เพื่อทำการคิวรีข้อมูลโอแลปมาสร้างเป็นมุมมองของผู้ใช้งาน จำนวน 3 กลุ่ม คือ ผู้บริหาร หัวหน้ากลุ่มสาระการเรียนรู้ และอาจารย์ประจำชั้น ซึ่งผู้ใช้งานสามารถดูข้อมูลแบบแผนภูมิและแดชบอร์ดตามมุมมองที่ออกแบบไว้ สามารถเรียกดูข้อมูลแบบสรุป (Roll UP) และแบบเจาะลึก (Drill Down) ดังภาพที่ 4.9



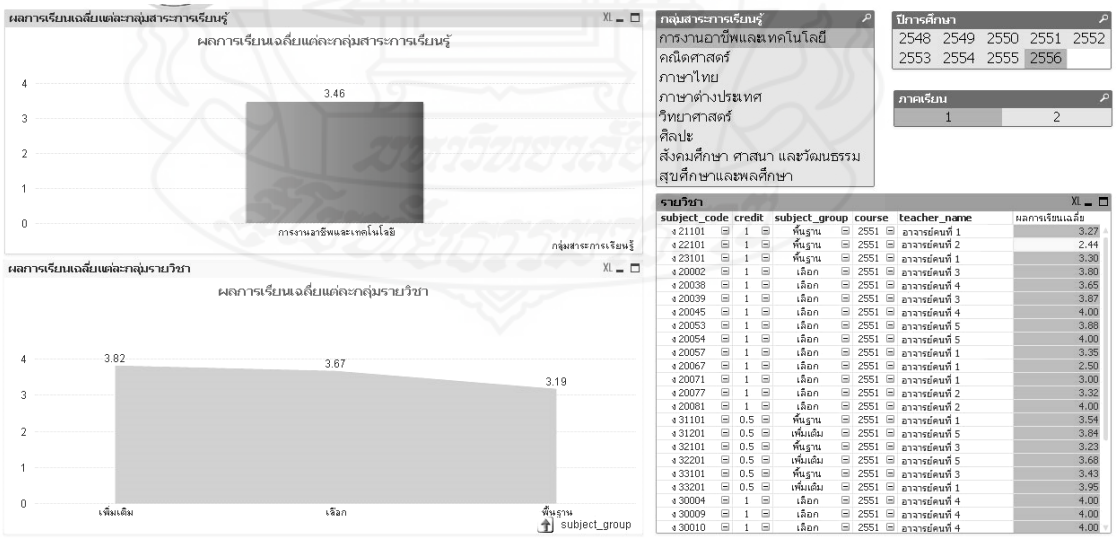
ภาพที่ 4.9 ข้อมูลแบบสรุป (Roll UP) และแบบเจาะลึก (Drill Down)

ผู้ใช้งานกลุ่มผู้บริหารเป็นการนำเสนอข้อมูลจำนวนนักเรียนในแต่ละปีการศึกษา สามารถ แบบสรุป และปรับเปลี่ยนระดับจำนวนนักเรียนจากปีการศึกษาเป็นภาคเรียนที่เป็นแบบเจาะลึก นอกจากนี้ยังนำเสนอผลการเรียนเฉลี่ยในแต่ละปีการศึกษา และ KPI ของผลการเรียนเฉลี่ยในรูปดูข้อมูลแบบแผนภูมิและแดชบอร์ด ดังภาพที่ 4.10



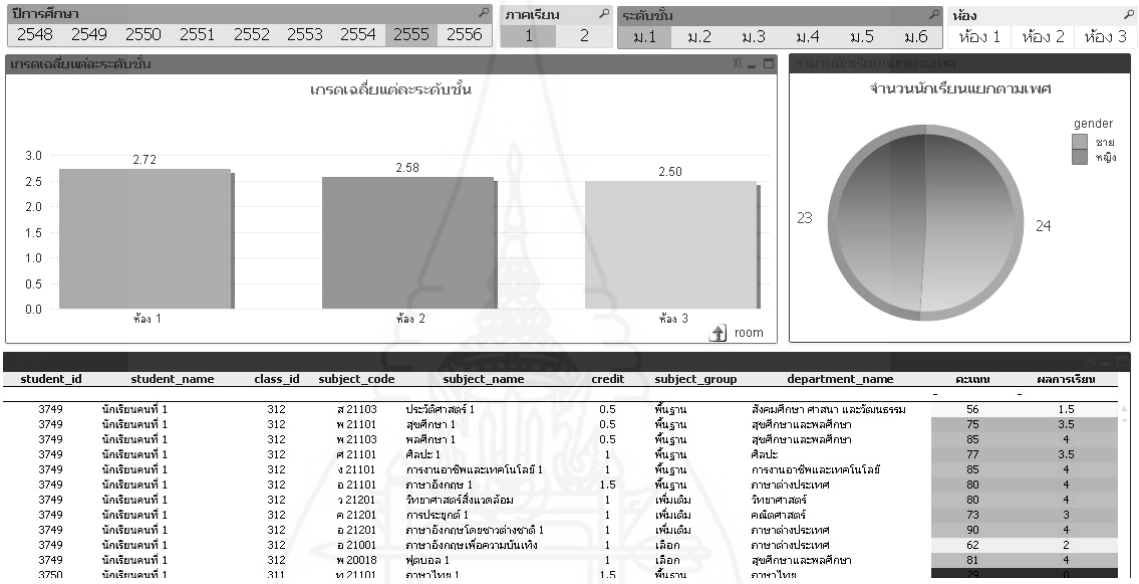
ภาพที่ 4.10 ตัวอย่างแดชบอร์ดของผู้บริหาร

ผู้ใช้งานหัวหน้ากลุ่มสาระการเรียนรู้ เป็นการนำเสนอข้อมูลผลการเรียนเฉลี่ยของแต่ละกลุ่มสาระการเรียนรู้ในแต่ละปีการศึกษา ผลการเรียนเฉลี่ยของแต่ละวิชาของกลุ่มสาระการเรียนรู้ ทำ KPI ผลการเรียนเฉลี่ยเป็นรายวิชา ดังภาพที่ 4.11



ภาพที่ 4.11 ตัวอย่างแดชบอร์ดของหัวหน้ากลุ่มสาระการเรียนรู้

ผู้ใช้งานกลุ่มอาจารย์ประจำชั้น เป็นการนำเสนอข้อมูลผลการเรียนและคะแนนของนักเรียนแต่ละระดับชั้นของภาคเรียนในแต่ละปีการศึกษา สามารถปรับเปลี่ยนระดับจากปีการศึกษาเป็นภาคเรียนเพื่อดูผลการเรียนเฉลี่ย หรือการดูสรุปของแต่ละปีการศึกษา และจำนวนนักเรียนในแต่ระดับชั้น สามารถ ปรับเปลี่ยนระดับจากข้อมูลสรุปเป็นระดับชั้นและห้องเรียน ดังภาพที่ 4.12



ภาพที่ 4.12 ตัวอย่างแดชบอร์ดของอาจารย์ประจำชั้น

3. การทำเหมืองข้อมูล

งานวิจัยนี้ดำเนินการทำเหมืองข้อมูลตามแบบจำลองของคริสพ์ – ดีเอ็ม (CRISP – DM Model) ซึ่งประกอบด้วย 6 ขั้นตอน ดังนี้

3.1 ทำความเข้าใจปัญหา (Business Understanding)

ผู้วิจัยได้ศึกษางานวิจัยที่เกี่ยวข้อง และศึกษาข้อมูลเกี่ยวกับการพยากรณ์ผลการเรียน และวิเคราะห์ปัญหาที่เกิดขึ้นกับนักเรียนระดับมัธยมศึกษาพบว่า นักเรียนมีผลการเรียนที่ต่ำกว่าเกณฑ์ในการวัดและประเมินผลเพิ่มขึ้นเรื่อยๆ และเมื่อจบการศึกษาในระดับมัธยมศึกษาตอนต้น มีผลการเรียนเฉลี่ยที่ไม่สามารถเลือกแผนการเรียนที่ตนเองต้องการได้ บางครั้งต้องย้ายไปเรียน โรงเรียนอื่น เนื่องจากผลการเรียนไม่ถึงเกณฑ์ที่กำหนดและสอบไม่ผ่านการคัดเลือก เพราะนักเรียนไม่สามารถ

พยากรณ์ผลการเรียนเฉลี่ยได้ทำให้ขาดข้อมูลในการประกอบการตัดสินใจของนักเรียนเพื่อเตรียมการวางแผนล่วงหน้า หากมีเครื่องมือที่ใช้ในการพยากรณ์หรือให้ข้อมูลนักเรียนจะได้วางแผนและพัฒนาผลการเรียนเพื่อเลือกแผนการเรียนในระดับมัธยมศึกษาตอนปลายได้อย่างเหมาะสม และเป็นเครื่องมือที่ให้อาจารย์ช่วยเหลือนักเรียนในเรื่องของการเรียน

3.2 ทำความเข้าใจข้อมูล (Data Understanding)

ผู้วิจัยรวบรวมข้อมูลที่เกี่ยวข้องเพื่อนำไปวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล ผู้วิจัยจึงได้สร้างคลังข้อมูลขึ้นมาดังกล่าวแล้วนั้น ซึ่งเป็นข้อมูลนักเรียน ข้อมูลผลการเรียน ข้อมูลรายวิชา ข้อมูลอาจารย์ ในการทำเหมืองข้อมูลใช้ข้อมูลที่มาจากคลังข้อมูลที่สร้างขึ้นนักเรียนด้วยกระบวนการอีทีแอล โดยใช้เครื่องมือ SQL Server Integration Services (SSIS) ของโปรแกรม Microsoft SQL Server 2008 R2 เพื่อนำข้อมูลออกมาให้อยู่ในรูปแบบของไฟล์เอ็กเซล (Excel) ซึ่งประกอบด้วยข้อมูลของนักเรียนในระดับชั้นมัธยมศึกษาปีที่ 4 ระหว่าง ปีการศึกษา 2553 – 2556 ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐานปี 2551 จำนวน 525 เรคคอร์ด ดังภาพที่ 4.13

C	D	E	F	G	H	I	J	K	L	M	N
plan	career_father	income_father	career_mother	incom_mother	Marital_Status	brothers	math	sci	thai	soc	eng
โครงสร้างการเรียนที่ 2(ภาษาอังกฤษ-ภาษาที่สาม)	เกษตรกรรม	100000	อื่น ๆ	200000	อยู่ด้วยกัน	2	0.11	0.00	0.00	0.00	0.26
โครงสร้างการเรียนที่ 1(วิทยาศาสตร์-คณิตศาสตร์)	นักธุรกิจ-ค้าขาย	300000	นักธุรกิจ-ค้าขาย	300000	อยู่ด้วยกัน	2	4.00	3.22	3.16	3.16	3.60
โครงสร้างการเรียนที่ 1(วิทยาศาสตร์-คณิตศาสตร์)	นักธุรกิจ-ค้าขาย	600000	นักธุรกิจ-ค้าขาย	600000	อยู่ด้วยกัน	2	1.61	1.33	1.83	2.16	1.80
โครงสร้างการเรียนที่ 1(วิทยาศาสตร์-คณิตศาสตร์)	นักธุรกิจ-ค้าขาย	500000	นักธุรกิจ-ค้าขาย	300000	อยู่ด้วยกัน	3	3.66	2.16	3.33	2.33	3.00
โครงสร้างการเรียนที่ 3(ภาษาไทย-สังคมศึกษา)	นักธุรกิจ-ค้าขาย	600000	ไม่ได้ประกอบอาชีพ	0	อยู่ด้วยกัน	2	0.11	0.11	0.83	0.50	0.26
โครงสร้างการเรียนที่ 3(ภาษาไทย-สังคมศึกษา)	ข้าราชการ	250000	ข้าราชการ	480000	อยู่ด้วยกัน	3	0.27	0.11	0.00	0.33	0.00
โครงสร้างการเรียนที่ 3(ภาษาไทย-สังคมศึกษา)	นักธุรกิจ-ค้าขาย	700000	นักธุรกิจ-ค้าขาย	700000	อยู่ด้วยกัน	3	0.88	0.88	1.33	1.50	1.03
โครงสร้างการเรียนที่ 1(วิทยาศาสตร์-คณิตศาสตร์)	รับจ้าง	500000	ข้าราชการ	300000	อยู่ด้วยกัน	2	1.33	1.16	2.50	2.33	1.93
โครงสร้างการเรียนที่ 1(วิทยาศาสตร์-คณิตศาสตร์)	พนักงานรัฐวิสาหกิจ	360000	ข้าราชการ	300000	อยู่ด้วยกัน	2	2.77	2.33	3.50	3.16	2.96
โครงสร้างการเรียนที่ 1(วิทยาศาสตร์-คณิตศาสตร์)	นักธุรกิจ-ค้าขาย	800000	นักธุรกิจ-ค้าขาย	800000	อยู่ด้วยกัน	3	2.00	1.72	2.83	2.16	2.58
โครงสร้างการเรียนที่ 1(วิทยาศาสตร์-คณิตศาสตร์)	ข้าราชการ	240000	อื่น ๆ	100000	อยู่ด้วยกัน	2	1.88	1.83	3.33	2.66	2.90
โครงสร้างการเรียนที่ 1(วิทยาศาสตร์-คณิตศาสตร์)	ข้าราชการ	270000	อื่น ๆ	120000	อยู่ด้วยกัน	2	2.50	2.22	3.00	2.77	2.93
โครงสร้างการเรียนที่ 3(ภาษาไทย-สังคมศึกษา)	นักธุรกิจ-ค้าขาย	480000	ข้าราชการ	360000	อยู่ด้วยกัน	2	0.66	0.61	1.50	0.83	1.38
โครงสร้างการเรียนที่ 3(ภาษาไทย-สังคมศึกษา)	เสียชีวิต	0	นักธุรกิจ-ค้าขาย	300000	บิดาถึงแก่กรรม	3	0.55	0.61	2.50	2.16	1.26

ภาพที่ 4.13 แสดงไฟล์ข้อมูลที่ได้จากคลังข้อมูล

3.3 การเตรียมข้อมูล (Data Preparation)

3.3.1 การคัดกรองข้อมูล (Data Cleaning) เป็นการคัดกรองข้อมูลที่ไม่เกี่ยวข้อง และข้อมูลที่ผิดปกติออกไป เพื่อลดความผิดพลาดที่อาจจะเกิดขึ้น ซึ่งผลที่ตามมาคือ ทำให้เกิดค่าสูญหายของการไม่มีข้อมูล (Missing Values) โดยข้อมูลนักเรียนที่นำมาใช้เป็นข้อมูลอยู่ระหว่างปีการศึกษา 2553 – 2556 ระดับชั้นมัธยมศึกษาปีที่ 4 มีจำนวน 525 เรคคอร์ด

3.3.2 การเลือกข้อมูล (Data Selection) ในการเลือกข้อมูลที่ใช้ในการวิเคราะห์ ผู้วิจัยได้ศึกษางานวิจัยของพรรณิภา บุตรเอก และสุรเดช บุญลือ (2557: 43) ในการพยากรณ์ผลการศึกษา

ของนักศึกษา โดยใช้คุณลักษณะของนักศึกษาที่เป็นข้อมูลประวัติส่วนตัว เช่น เพศ รายได้ของบิดา รายได้มารดา สถานบิดาสถานมารดา เป็นต้น และข้อมูลผลการเรียน เช่น เกรดเฉลี่ยภาคเรียนที่ 1 เกรดเฉลี่ยภาคเรียนที่ 2 กลุ่มการเรียน เป็นต้น และงานวิจัยของสุพัฒน์กุล ภัคโชค (2556: 27) ที่พัฒนาตัวแบบเลือกแผนการเรียนระดับมัธยมศึกษาตอนปลายด้วยเทคนิคเหมืองข้อมูล โดยใช้คุณลักษณะผลการเรียนเฉลี่ยระดับมัธยมศึกษาตอนต้นวิชาภาษาไทย ผลการเรียนเฉลี่ยระดับมัธยมศึกษาตอนต้นวิชาวิทยาศาสตร์ ผลการเรียนเฉลี่ยระดับมัธยมศึกษาตอนต้นวิชาสังคมศึกษา ผลการเรียนเฉลี่ยระดับมัธยมศึกษาตอนต้นวิชาภาษาอังกฤษ ผลการเรียนเฉลี่ยสะสมระดับมัธยมศึกษาตอนต้น เป็นต้น ซึ่งงานวิจัยนี้ใช้คุณลักษณะหรือตัวแปร ดังตารางที่ 4.13

ตารางที่ 4.13 แสดงคุณลักษณะหรือตัวแปร

ลำดับที่	คุณลักษณะหรือตัวแปร
1	เพศ
2	แผนการเรียน
3	อาชีพของบิดา
4	รายได้เฉลี่ยต่อปีของบิดา
5	อาชีพของมารดา
6	รายได้เฉลี่ยต่อปีของมารดา
7	สถานภาพของบิดามารดา
8	จำนวนพี่น้อง
9	ผลการเรียนเฉลี่ยคณิตศาสตร์
10	ผลการเรียนเฉลี่ยวิทยาศาสตร์
11	ผลการเรียนเฉลี่ยภาษาไทย
12	ผลการเรียนเฉลี่ยสังคมศึกษา
13	ผลการเรียนเฉลี่ยภาษาอังกฤษ
14	ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 1
15	ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 2
16	ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3
17	ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 4

3.3.3 การแปลงข้อมูล (Data Transformation) ในขั้นตอนการแปลงข้อมูลใช้ฟังก์ชันของโปรแกรม Microsoft Excel 2007 ในการแปลงข้อมูล ซึ่งข้อมูลผลการเรียนเฉลี่ยมีความละเอียดมากเกินไปผู้วิจัยจึงจัดกลุ่มระดับผลการเรียนเฉลี่ยใหม่ โดยแบ่งระดับผลการเรียนเฉลี่ยออกเป็น 3 กลุ่ม เพื่อใช้ในการสร้างตัวแบบพยากรณ์ (Model) ซึ่งช่วยลดการกระจายของข้อมูลดังนี้

High คือ ผลการเรียนเฉลี่ยของนักเรียนตั้งแต่ 3.01 – 4.00

Medium คือ ผลการเรียนเฉลี่ยของนักเรียนตั้งแต่ 2.01 – 3.00

Low คือ ผลการเรียนเฉลี่ยของนักเรียนตั้งแต่ 1.00 – 2.00

รูปแบบของการแปลงข้อมูลที่เป็นผลลัพธ์ที่ต้องการ (Output) ดังตารางที่ 4.14 และข้อมูลที่เป็นคุณลักษณะ (Input) มีรายละเอียดของการแปลงข้อมูลดังตารางที่ 4.15

ตารางที่ 4.14 การแปลงข้อมูลที่เป็นผลลัพธ์

คุณลักษณะ	**คำอธิบาย**	**ข้อมูลที่ถูกแปลง**
gpam4	ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 4	High = ผลการเรียนเฉลี่ย 3.01 – 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 – 3.00 Low = ผลการเรียนเฉลี่ย 1.00 – 2.00

ตารางที่ 4.15 การแปลงข้อมูลคุณลักษณะ

ลำดับ	คุณลักษณะ	คำอธิบาย	ข้อมูลที่ถูกแปลง
1	plan	แผนการเรียน	A = โครงสร้างการเรียนที่ 1 (วิทยาศาสตร์-คณิตศาสตร์) B = โครงสร้างการเรียนที่ 2 (ภาษาอังกฤษ-ภาษาที่สาม) C = โครงสร้างการเรียนที่ 3 (ภาษาไทย-สังคมศึกษา)
2	gender	เพศ	Male = ชาย Female = หญิง
3	career_father	อาชีพบิดา	Government = ข้าราชการ เจ้าหน้าที่รัฐ หรือรัฐวิสาหกิจ Employ = พนักงานเอกชน ลูกจ้าง หรือรับจ้าง Trade = ค้าขาย หรือธุรกิจส่วนตัว Farmer = เกษตรกร Other = ไม่ได้ประกอบอาชีพ หรือเสียชีวิต
4	income_father	รายได้ของบิดา	L = น้อยกว่า 300,000 บาท ต่อปี M = 300,000 – 500,000 บาท ต่อปี H = มากกว่า 500,000 บาท ต่อปี
5	career_mother	อาชีพมารดา	Government = ข้าราชการ เจ้าหน้าที่รัฐ หรือรัฐวิสาหกิจ Employ = พนักงานเอกชน ลูกจ้าง หรือรับจ้าง Trade = ค้าขาย หรือธุรกิจส่วนตัว Farmer = เกษตรกร Other = ไม่ได้ประกอบอาชีพ หรือเสียชีวิต
6	income_mother	รายได้มารดา	L = น้อยกว่า 300,000 บาท ต่อปี M = 300,001 – 500,000 บาท ต่อปี H = มากกว่า 500,001 บาท ต่อปี
7	Marital_Status	สถานภาพของบิดามารดา	Cohabit = บิดามารดาอยู่ด้วยกัน Divorce = บิดามารดาอย่าร้างหรือแต่งงานใหม่ Dead = บิดา หรือ มารดาเสียชีวิต หรือทั้งหมด

ตารางที่ 4.15 (ต่อ)

ลำดับ	คุณลักษณะ	คำอธิบาย	ข้อมูลที่ถูกแปลง
8	brothers	จำนวนพี่น้อง	None = ไม่มีพี่น้อง 1 = มีพี่น้อง 1 คน >1 = มีพี่น้องมากว่า 1 คน
9	math	ผลการเรียนเฉลี่ย คณิตศาสตร์	High = ผลการเรียนเฉลี่ย 3.01 – 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 – 3.00 Low = ผลการเรียนเฉลี่ย 1.00 – 2.00
10	sci	ผลการเรียนเฉลี่ย วิทยาศาสตร์	High = ผลการเรียนเฉลี่ย 3.01 – 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 – 3.00 Low = ผลการเรียนเฉลี่ย 1.00 – 2.00
11	thai	ผลการเรียนเฉลี่ย ภาษาไทย	High = ผลการเรียนเฉลี่ย 3.01 – 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 – 3.00 Low = ผลการเรียนเฉลี่ย 1.00 – 2.00
12	soc	ผลการเรียนเฉลี่ยสังคม ศึกษา	High = ผลการเรียนเฉลี่ย 3.01 – 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 – 3.00 Low = ผลการเรียนเฉลี่ย 1.00 – 2.00
13	eng	ผลการเรียนเฉลี่ย ภาษาอังกฤษ	High = ผลการเรียนเฉลี่ย 3.01 – 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 – 3.00 Low = ผลการเรียนเฉลี่ย 1.00 – 2.00
14	gpam1	ผลการเรียนเฉลี่ย มัธยมศึกษาปีที่ 1	High = ผลการเรียนเฉลี่ย 3.01 – 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 – 3.00 Low = ผลการเรียนเฉลี่ย 1.00 – 2.00
15	gpam2	ผลการเรียนเฉลี่ย มัธยมศึกษาปีที่ 2	High = ผลการเรียนเฉลี่ย 3.01 – 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 – 3.00 Low = ผลการเรียนเฉลี่ย 1.00 – 2.00
16	gpam3	ผลการเรียนเฉลี่ย มัธยมศึกษาปีที่ 3	High = ผลการเรียนเฉลี่ย 3.01 – 4.00 Medium = ผลการเรียนเฉลี่ย 2.01 – 3.00

Low = ผลการเรียนเฉลี่ย 1.00 – 2.00

ทำการแปลงข้อมูลด้วยโปรแกรม Microsoft Excel 2007 บันทึกข้อมูลให้อยู่ในรูปแบบของไฟล์ CSV เพื่อให้สามารถใช้งานกับโปรแกรม WEKA ได้ จากนั้นนำไฟล์เข้าโปรแกรม WEKA เพื่อบันทึกให้อยู่ในรูปแบบไฟล์ arff เพื่อใช้ในการวิเคราะห์ข้อมูล ซึ่งแสดงดังภาพที่ 4.14

```
@attribute gender {Male,Female}
@attribute plan {A,B,C}
@attribute career_father {Government,Employ,Trade,Farmer,Other}
@attribute income_father {L,M,H}
@attribute career_mother {Government,Employ,Trade,Farmer,Other}
@attribute incom_mother {L,M,H}
@attribute Marital_Status {Cohabit,Divorce,Dead}
@attribute brothers {>1,1,None}
@attribute math {Low,Medium,High}
@attribute sci {Low,Medium,High}
@attribute thai {Low,Medium,High}
@attribute soc {Low,Medium,High}
@attribute eng {Low,Medium,High}
@attribute gpam1 {Low,Medium,High}
@attribute gpam2 {Low,Medium,High}
@attribute gpam3 {Low,Medium,High}
@attribute gpam4 {Low,Medium,High}

@data
Male,A,Trade,L,Trade,L,Cohabit,>1,High,High,High,High,High,High,High,High,High
Male,A,Trade,H,Trade,H,Cohabit,>1,Low,Low,Low,Medium,Low,Medium,Low,Medium,Medium
Male,A,Trade,M,Trade,L,Cohabit,>1,High,Medium,High,Medium,Medium,Medium,Medium,High,High
Male,A,Employ,M,Government,L,Cohabit,>1,Low,Low,Medium,Medium,Low,Low,Medium,Medium,Medium
Male,A,Government,M,Government,L,Cohabit,>1,Medium,Medium,High,High,Medium,Medium,Medium,High,High
Female,A,Trade,H,Trade,H,Cohabit,>1,Low,Low,Medium,Medium,Medium,Low,Medium,Medium,Medium
```

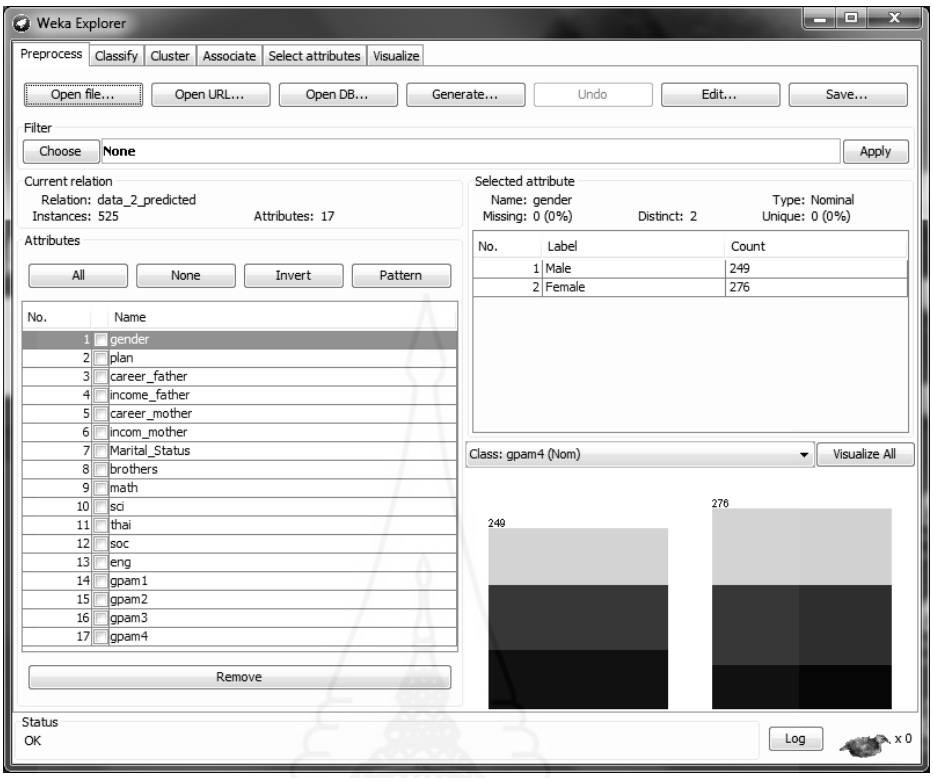
ภาพที่ 4.14 แสดงผลการแปลงข้อมูล

3.4 การสร้างตัวแบบพยากรณ์ (Modeling)

ข้อมูลที่ใช้ในการสร้างตัวแบบพยากรณ์มีจำนวน 525 เรคคอร์ด แบ่งเป็น 2 แบบ คือ ชุดข้อมูลที่ไม่จัดกลุ่ม และข้อมูลที่จัดกลุ่มด้วยอัลกอริทึม k-means ผลที่ได้จากการวิจัยเป็นดังนี้

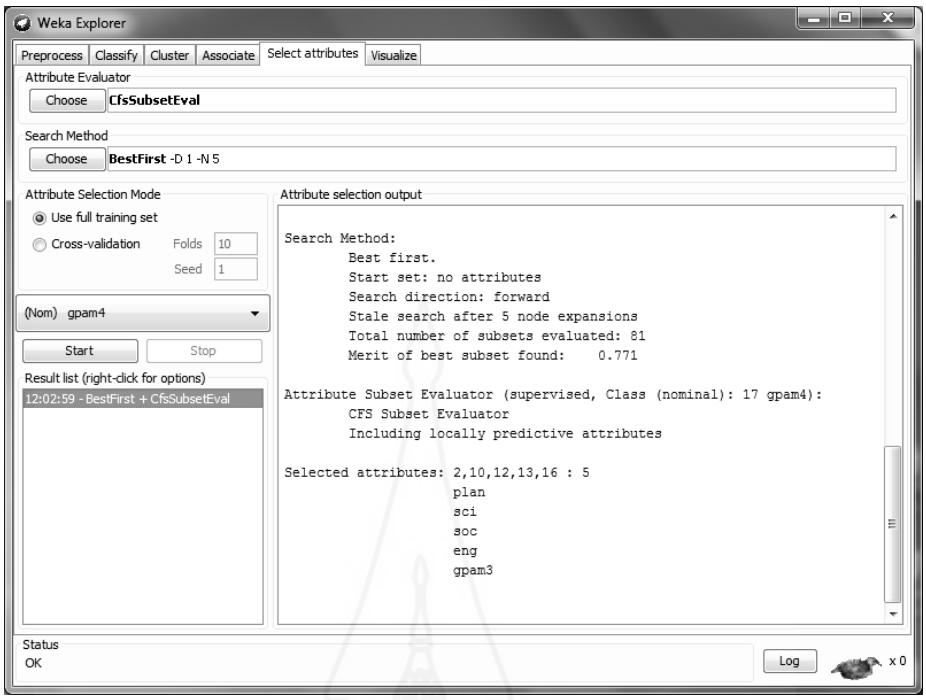
3.4.1 การสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่ไม่จัดกลุ่ม

1) กระบวนการคัดเลือกคุณลักษณะ การสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่ไม่จัดกลุ่ม ผู้วิจัยใช้ชุดข้อมูล (Data set) ที่ไม่มีการคัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะหรือตัวแปร (Feature Selection) ด้วยวิธี Correlation – based Feature Selection (CFS) และวิธี Information Gain (IG) ซึ่งชุดข้อมูลที่ไม่มีการคัดเลือกคุณลักษณะ ดังภาพที่ 4.15



ภาพที่ 4.15 แสดงชุดข้อมูลที่ไม่มีการคัดเลือกคุณลักษณะ

จากการคัดเลือกคุณลักษณะหรือตัวแปร (Feature Selection) ด้วยวิธี Correlation – based Feature Selection (CFS) โดยใช้โปรแกรม WEKA พบว่า คุณลักษณะที่ถูกคัดเลือกจาก 16 คุณลักษณะ เหลือเพียง 5 คุณลักษณะ ดังภาพที่ 4.16



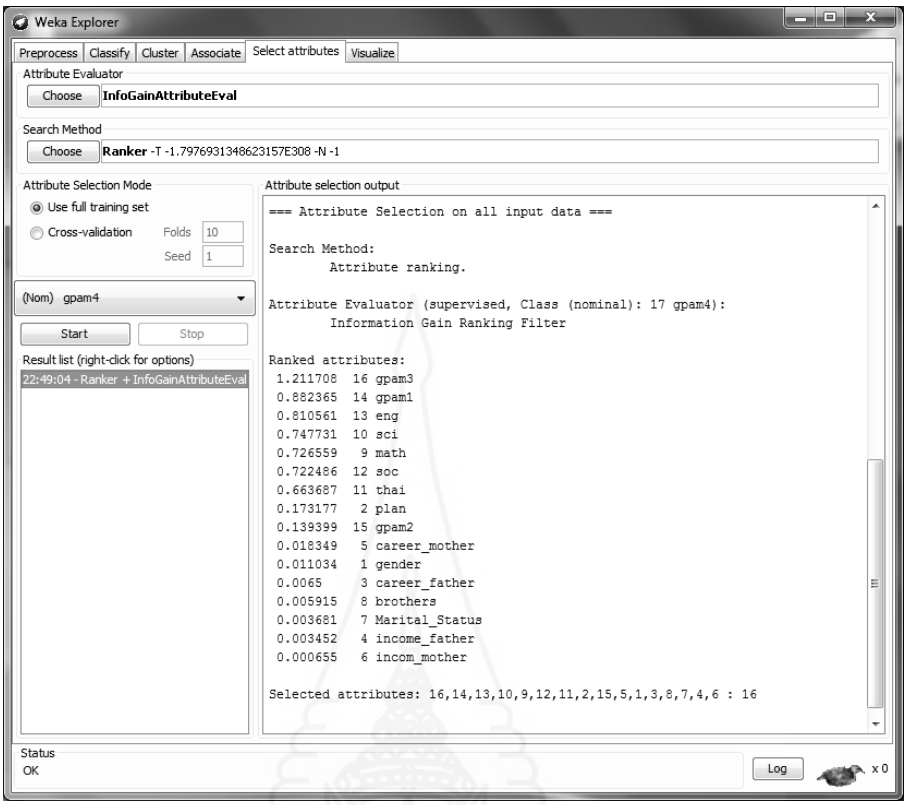
ภาพที่ 4.16 การคัดเลือกคุณลักษณะด้วยวิธี Correlation – based Feature Selection (CFS)

ผลการคัดเลือกคุณลักษณะหรือตัวแปร ด้วยวิธี Correlation – based Feature Selection (CFS) ดังตารางที่ 4.16

ตารางที่ 4.16 ผลการคัดเลือกคุณลักษณะด้วยวิธี Correlation – based Feature Selection (CFS)

ลำดับที่	คุณลักษณะ	คำอธิบาย
1	plan	แผนการเรียน
2	sci	ผลการเรียนเฉลี่ยวิทยาศาสตร์
3	soc	ผลการเรียนเฉลี่ยสังคมศึกษา
4	eng	ผลการเรียนเฉลี่ยภาษาอังกฤษ
5	gpam3	ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3

จากการคัดเลือกคุณลักษณะหรือตัวแปร ด้วยวิธี Information Gain (IG) พบว่า มีการจัดเรียงลำดับคุณลักษณะที่สำคัญไว้ตาม Ranked ซึ่งผู้วิจัยได้ทำคัดเลือกคุณลักษณะโดยพิจารณาจากค่า Ranked ที่มีค่าระหว่าง 1 – 0.1 ตั้งแต่ลำดับที่ 1 ถึง 9 ดังภาพที่ 4.17



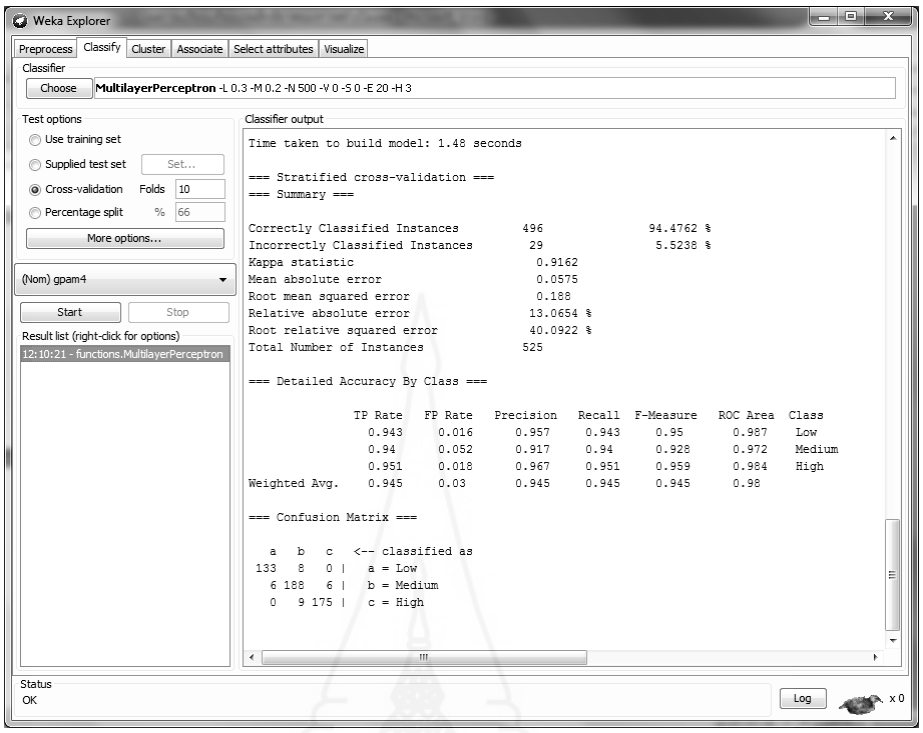
ภาพที่ 4.17 การคัดเลือกคุณลักษณะด้วยวิธี Information Gain (IG)

ผลการคัดเลือกคุณลักษณะหรือตัวแปร (Feature Selection) ด้วยวิธี Information Gain (IG) ดังตารางที่ 4.17

ตารางที่ 4.17 ผลการคัดเลือกคุณลักษณะด้วยวิธี Information Gain

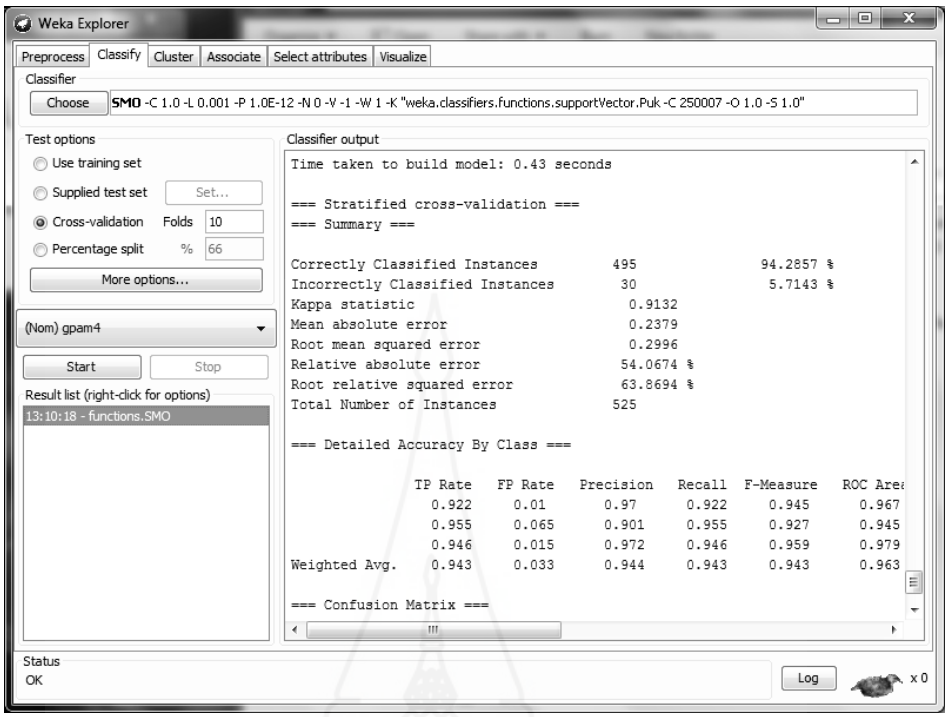
ลำดับที่	คุณลักษณะ	คำอธิบาย
1	gpam3	ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3
2	gpam1	ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 1
3	eng	ผลการเรียนเฉลี่ยภาษาอังกฤษ
4	sci	ผลการเรียนเฉลี่ยวิทยาศาสตร์
5	math	ผลการเรียนเฉลี่ยคณิตศาสตร์
6	soc	ผลการเรียนเฉลี่ยสังคมศึกษา
7	thai	ผลการเรียนเฉลี่ยภาษาไทย
8	plan	แผนการเรียน
9	gpam2	ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 2

2) การใช้เทคนิคเหมืองข้อมูลโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (Multi-layer Perceptron: MLP) ร่วมกับวิธีที่ไม่คัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะด้วย Correlation – based Feature Selection (CFS) และวิธี Information Gain ในการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่ไม่จัดกลุ่ม ในการสร้างตัวแบบพยากรณ์ผู้วิจัยได้กำหนดค่า Moment 0.2 Training Time 500 แล้วปรับค่า Hidden layer ระหว่าง 3 – 6 และค่า Learning ระหว่าง 0.1 – 0.3 ดังภาพที่ 4.18



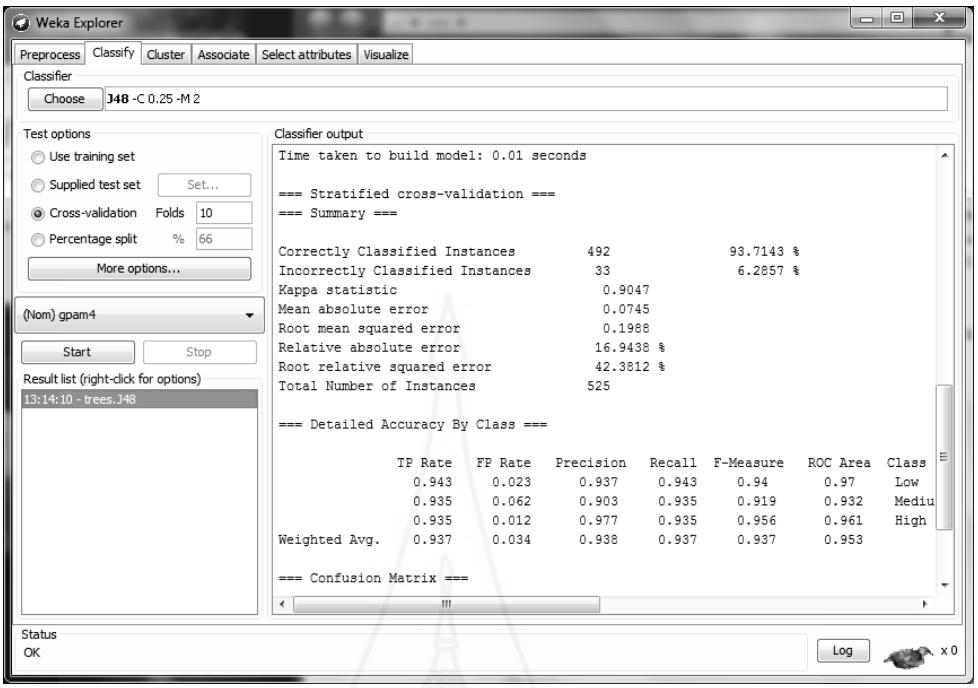
ภาพที่ 4.18 ผลการใช้เทคนิคโครงข่ายประสาทเทียมกับข้อมูลที่ไม่จัดกลุ่ม

3) การใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) ในการสร้างตัวแบบพยากรณ์ โดยใช้เคอร์เนล (Kernel) ฟังก์ชันแบบ Normalise Polynomial Kernel (SVM-NP) Polynomial Kernel (SVM-PK) Radial Basis Function Kernel (SVM-RBF) Pearson VII function-based universal kernel (SVM-PUK) ร่วมกับวิธีที่ไม่คัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะด้วยวิธี Correlation – based Feature Selection (CFS) และวิธี Information Gain ดังภาพที่ 4.19



ภาพที่ 4.19 ผลการใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนกับข้อมูลที่ไม่จัดกลุ่ม

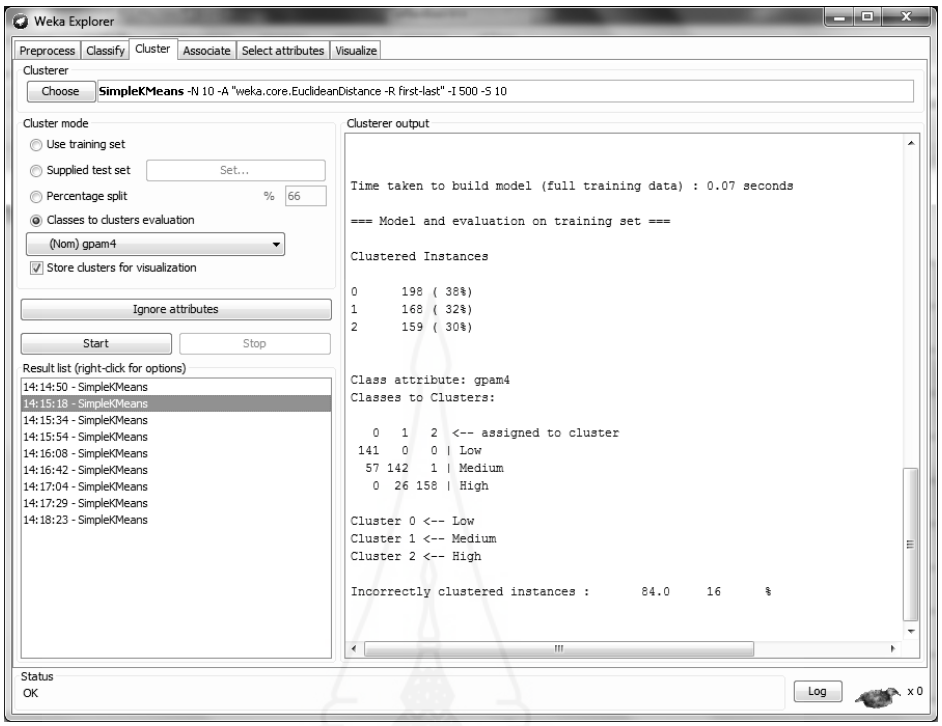
4) การใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึมแบบ C4.5 ในการสร้างตัวแบบพยากรณ์ ร่วมกับวิธีที่ไม่คัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะด้วยวิธี Correlation – based Feature Selection (CFS) และวิธี Information Gain ดังภาพที่ 4.20



ภาพที่ 4.20 ผลการใช้เทคนิคต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5 กับข้อมูลที่ไม่จัดกลุ่ม

3.4.2 การสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่ม

1) การใช้เทคนิคเหมืองข้อมูลในการจัดกลุ่มข้อมูล (Clustering) ด้วยอัลกอริทึม K-Means ด้วยโปรแกรม WEKA ดังภาพที่ 4.21 เพื่อจัดกลุ่มข้อมูลที่ใช้ในการสร้างตัวแบบพยากรณ์ โดยผู้วิจัยได้กำหนดค่า Num Cluster ระหว่าง 2 - 10 ซึ่งเป็นค่าที่ใช้ในการแบ่งกลุ่มข้อมูลผลที่ได้ ดังตารางที่ 4.18



ภาพที่ 4.21 ผลการจัดกลุ่มข้อมูลด้วยอัลกอริทึม K-Means

ตารางที่ 4.18 ผลการจัดกลุ่มข้อมูลด้วยอัลกอริทึม K-Means

ลำดับที่	NumCluster	Class	ICI
1	2	2	32.57%
2	3*	3*	16.00%
3	4	4	27.24%
4	5	5	36.57%
5	6	6	46.29%
6	7	7	51.05%
7	8	8	53.52%
8	9	9	58.29%
9	10	10	60.57%

จากตารางที่ 4.18 เมื่อพิจารณาค่าความผิดพลาดในการจำแนกกลุ่ม (Incorrectly Clustered Instances: ICI) พบว่า การจำแนกกลุ่มโดยใช้ค่า Num Cluster เท่ากับ 3 มีค่าความผิดพลาดในการจำแนกข้อมูลน้อยที่สุด เท่ากับ 16.00% โดยนำค่าที่ได้มาทำการการจำแนกกลุ่มข้อมูล (Class) ได้ 3 กลุ่ม ดังภาพที่ 4.22

Viewer

Relation: data_2_predicted-weka.filters.unsupervised.attribute.AddCluster-Wweka.clusterers.SimpleKMeans -N 3 -A "weka.core.EuclideanDistance -R first-last" -I 500 -S 10

No.	gender Nominal	plan Nominal	career_father Nominal	income_father Nominal	career_mother Nominal	incom_mother Nominal	Marital_Status Nominal	brothers Nominal	math Nominal	sci Nominal	thai Nominal	soc Nominal	eng Nominal	gpam1 Nominal	gpam2 Nominal	gpam3 Nominal	gpam4 Nominal	**cluster** Nominal
1	Male	A	Trade	L	Trade	L	Cohabit	)1	High	High	High	High	High	High	High	High	High	cluster3
2	Male	A	Trade	H	Trade	H	Cohabit	)1	Low	Low	Low	Medium	Low	Medium	Low	Medium	Medium	cluster1
3	Male	A	Trade	M	Trade	L	Cohabit	)1	High	Medium	High	Medium	Medium	Medium	Medium	High	High	cluster2
4	Male	A	Employ	M	Government	L	Cohabit	)1	Low	Low	Medium	Medium	Low	Low	Medium	Medium	Medium	cluster1
5	Male	A	Government	M	Government	L	Cohabit	)1	Medium	Medium	High	High	Medium	Medium	Medium	High	High	cluster2
6	Female	A	Trade	H	Trade	H	Cohabit	)1	Low	Low	Medium	Medium	Medium	Low	Medium	Medium	Medium	cluster2
7	Female	A	Government	L	Other	L	Cohabit	)1	Low	Low	High	Medium	Medium	Medium	Medium	Medium	Medium	cluster2
8	Female	A	Government	L	Other	L	Cohabit	)1	Medium	Medium	Medium	Medium	Medium	Medium	Medium	High	Medium	cluster2
9	Female	A	Government	L	Government	L	Divorce	)1	High	High	High	High	High	High	High	High	High	cluster3
10	Female	A	Farmer	L	Trade	L	Cohabit	)1	Medium	Low	High	High	High	High	Medium	High	High	cluster3
11	Female	A	Other	L	Trade	L	Cohabit	)1	High	High	High	Medium	High	High	High	High	High	cluster3
12	Female	A	Trade	L	Other	L	Cohabit	)1	Medium	Low	Medium	Medium	Medium	High	High	Medium	Medium	cluster2
13	Male	A	Trade	H	Government	M	Cohabit	)1	High	Medium	High	High	Medium	High	Medium	High	High	cluster3
14	Male	A	Government	L	Government	L	Cohabit	)1	Medium	Low	Medium	Medium	Medium	Medium	Medium	Medium	Medium	cluster2
15	Male	A	Government	M	Government	L	Cohabit	1	Low	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	cluster2
16	Male	A	Government	M	Government	M	Cohabit	None	High	Medium	Medium	Medium	High	Medium	High	High	High	cluster3
17	Male	A	Government	M	Other	L	Dead	)1	High	Low	Medium	Low	Low	Medium	Medium	Medium	Medium	cluster1

ภาพที่ 4.22 การจำแนกกลุ่มข้อมูล

ผลการจำแนกกลุ่มข้อมูลที่ได้คือ Cluster1 มีจำนวนข้อมูล 198 เรคคอร์ด Cluster2 มีจำนวนข้อมูล 168 เรคคอร์ด และ Cluster3 มีจำนวนข้อมูล 159 เรคคอร์ด

2) กระบวนการคัดเลือกคุณลักษณะ การสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่ม ผู้วิจัยใช้ชุดข้อมูล (Data set) ที่ไม่มีผ่านการคัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะด้วยวิธี Correlation – based Feature Selection (CFS) และวิธี Information Gain (IG) ผลการคัดเลือกคุณลักษณะด้วยวิธี Correlation – based Feature Selection (CFS) ดังตารางที่ 4.19 และผลการคัดเลือกคุณลักษณะด้วยวิธี Information Gain (IG) ดังตารางที่ 4.20

ตารางที่ 4.19 ผลการคัดเลือกคุณลักษณะข้อมูลที่จัดกลุ่มด้วยวิธี Correlation – based Feature Selection

กลุ่มข้อมูล	คุณลักษณะ	คำอธิบาย
Cluster1	plan	แผนการเรียน
	math	ผลการเรียนเฉลี่ยคณิตศาสตร์
	soc	ผลการเรียนเฉลี่ยสังคมศึกษา
	eng	ผลการเรียนเฉลี่ยภาษาอังกฤษ
	gpam3	ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3
Cluster2	math	ผลการเรียนเฉลี่ยคณิตศาสตร์
	soc	ผลการเรียนเฉลี่ยสังคมศึกษา
	eng	ผลการเรียนเฉลี่ยภาษาอังกฤษ
	gpam3	ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3
Cluster3	marital_Status	สถานภาพของบิดามารดา
	brothers	จำนวนพี่น้อง
	sci	ผลการเรียนเฉลี่ยวิทยาศาสตร์
	gpam3	ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3

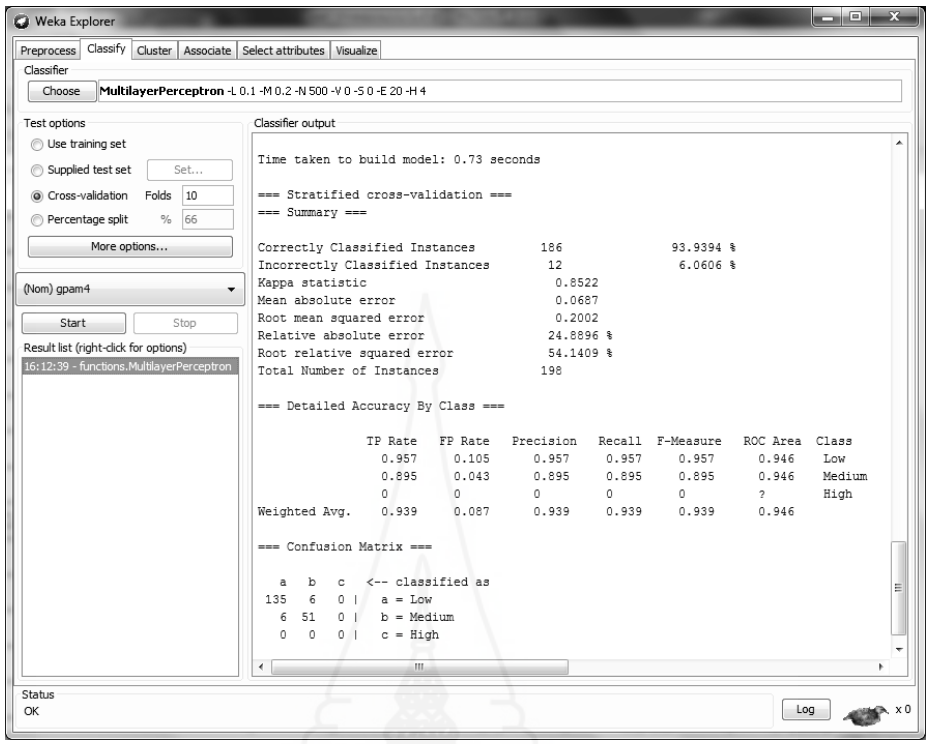
จากตารางที่ 4.19 การคัดเลือกคุณลักษณะด้วยวิธี Correlation – based Feature Selection (CFS) พบว่า กลุ่มที่ 1 (Cluster1) คุณลักษณะที่ถูกคัดเลือก 5 คุณลักษณะ ได้แก่ แผนการเรียน ผลการเรียนเฉลี่ยคณิตศาสตร์ ผลการเรียนเฉลี่ยสังคมศึกษา ผลการเรียนเฉลี่ยภาษาอังกฤษ และผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3 กลุ่มที่ 2 (Cluster2) คุณลักษณะที่ถูกคัดเลือก 4 คุณลักษณะ ได้แก่ ผลการเรียนเฉลี่ยคณิตศาสตร์ ผลการเรียนเฉลี่ยสังคมศึกษา ผลการเรียนเฉลี่ยภาษาอังกฤษ และผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3 และกลุ่มที่ 3 (Cluster3) คุณลักษณะที่ถูกคัดเลือก 4 คุณลักษณะ ได้แก่ สถานภาพของบิดามารดา จำนวนพี่น้อง ผลการเรียนเฉลี่ยวิทยาศาสตร์ และผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3

ตารางที่ 4.20 ผลการคัดเลือกคุณลักษณะข้อมูลที่จัดกลุ่มด้วยวิธี Information Gain

กลุ่มข้อมูล	คุณลักษณะ	คำอธิบาย
Cluster1	gpam3	ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3
	gpam1	ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 1
	plan	แผนการเรียน
	eng	ผลการเรียนเฉลี่ยภาษาอังกฤษ
	soc	ผลการเรียนเฉลี่ยสังคมศึกษา
Cluster2	gpam3	ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3
	math	ผลการเรียนเฉลี่ยคณิตศาสตร์
	sci	ผลการเรียนเฉลี่ยวิทยาศาสตร์
Cluster3	gpam3	ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3
	sci	ผลการเรียนเฉลี่ยวิทยาศาสตร์
	math	ผลการเรียนเฉลี่ยคณิตศาสตร์
	marital_Status	สถานภาพของบิดามารดา

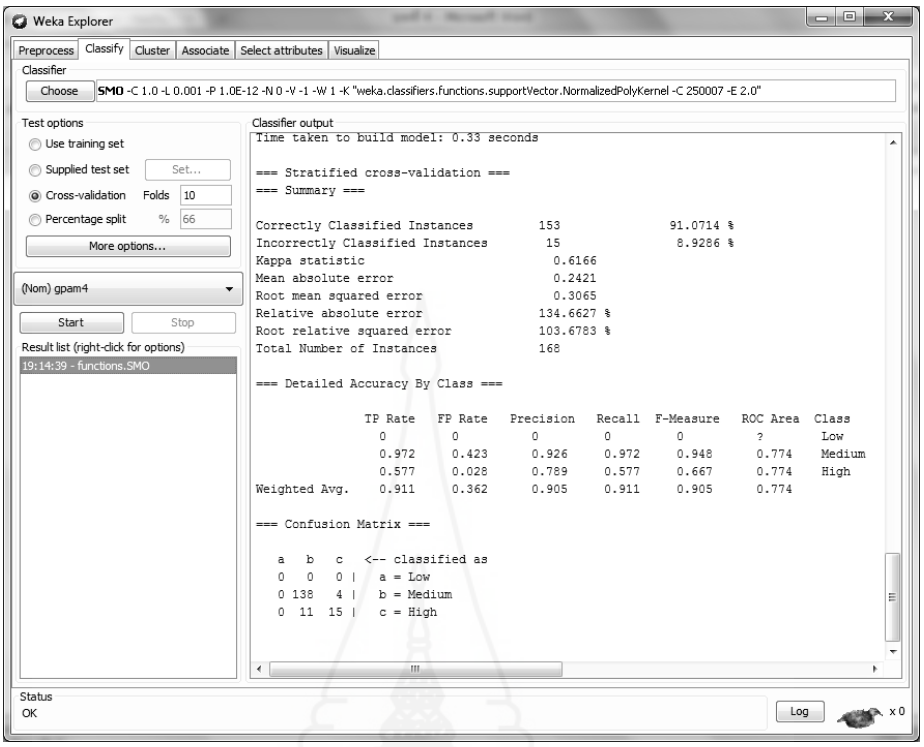
จากตารางที่ 4.20 การคัดเลือกคุณลักษณะด้วยวิธี Information Gain (IG) พบว่า กลุ่มที่ 1 (Cluster1) คุณลักษณะที่ถูกคัดเลือก 5 คุณลักษณะ ได้แก่ ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3 ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 1 แผนการเรียน ผลการเรียนเฉลี่ยภาษาอังกฤษ และผลการเรียนเฉลี่ยสังคมศึกษา กลุ่มที่ 2 (Cluster2) คุณลักษณะที่ถูกคัดเลือก 3 คุณลักษณะ ได้แก่ ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3 ผลการเรียนเฉลี่ยคณิตศาสตร์ และผลการเรียนเฉลี่ยวิทยาศาสตร์ และกลุ่มที่ 3 (Cluster3) คุณลักษณะที่ถูกคัดเลือก 4 คุณลักษณะ ได้แก่ ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3 ผลการเรียนเฉลี่ยวิทยาศาสตร์ ผลการเรียนเฉลี่ยคณิตศาสตร์ และสถานภาพของบิดามารดา

3) การใช้เทคนิคเหมืองข้อมูลโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (Multi-layer Perceptron: MLP) ร่วมกับวิธีที่ไม่คัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะด้วย Correlation – based Feature Selection (CFS) และวิธี Information Gain ในการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่ม ผู้วิจัยได้กำหนดค่า Moment 0.2 Training Time 500 แล้วปรับค่า Hidden layer ระหว่าง 3 – 6 และค่า Learning ระหว่าง 0.1 – 0.3 ดังภาพที่ 4.23



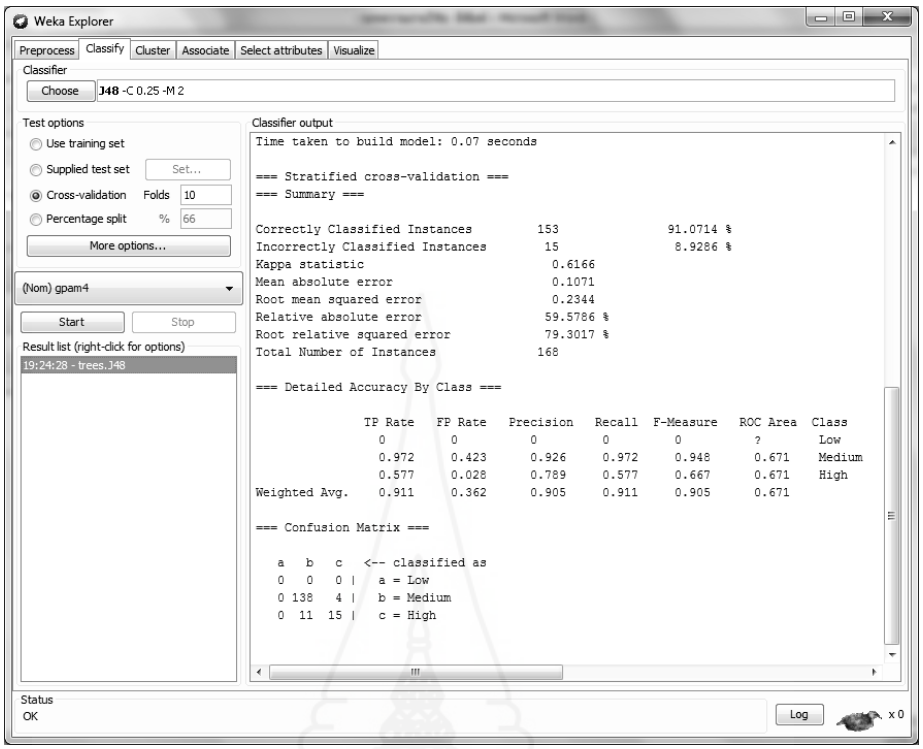
ภาพที่ 4.23 ผลการใช้เทคนิคโครงข่ายประสาทเทียมกับข้อมูลที่จัดกลุ่ม

4) การใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) ร่วมกับวิธีที่ไม่คัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะด้วย Correlation – based Feature Selection (CFS) และวิธี Information Gain ในการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่ม โดยใช้เคอร์เนล (Kernel) ฟังก์ชันแบบ Normalise Polynomial Kernel (SVM-NP) Polynomial Kernel (SVM-PK) Radial Basis Function Kernel (SVM-RBF) Pearson VII function-based universal kernel (SVM-PUK) ดังภาพที่ 4.24



ภาพที่ 4.24 ผลการใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนกับข้อมูลที่จัดกลุ่ม

5) การใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึมแบบ C4.5 ร่วมกับวิธีที่ไม่คัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะด้วย Correlation – base Feature Selection (CFS) และวิธี Information Gain ในการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่มดังภาพที่ 4.25



ภาพที่ 4.25 ผลการใช้เทคนิคต้นไม้ตัดสินใจกับข้อมูลที่จัดกลุ่ม

3.5 การประเมินประสิทธิภาพ (Evaluation)

3.5.1 การสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่ไม่จัดกลุ่ม

จากการทดสอบประสิทธิภาพด้วยวิธี K-fold Cross Validation โดยกำหนดค่า k เท่ากับ 10 ซึ่งเป็นวิธีการที่ให้ค่าความแม่นยำ ด้วยการแบ่งข้อมูลออกเป็น 10 ส่วน โดยแบ่งข้อมูลการสร้างตัวแบบเป็น 9 ส่วน และตัวทดสอบ 1 ส่วน แล้วสลับสับเปลี่ยนข้อมูลในการสร้างตัวแบบและทดสอบ ใช้วิธีการวัดค่าความถูกต้อง (Accuracy) จากการสร้างตัวแบบพยากรณ์ สามารถคำนวณได้จาก

$$\text{Accuracy} = \frac{\text{Correctly Classified Instance}}{\text{Total Number of Instances}} \times 100$$

และค่ารากที่สองของความคลาดเคลื่อน (Root Mean Square Error: RMSE) สามารถคำนวณได้จาก

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i^{'} - Y_i\right)^2}$$

ผลทดสอบประสิทธิภาพโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (Multi-layer Perceptron: MLP) ร่วมกับวิธีที่ไม่คัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะด้วย Correlation – based Feature Selection (CFS) และวิธี Information Gain ด้วยข้อมูลที่ไม่จัดกลุ่มในการสร้างตัวแบบพยากรณ์ ดังตารางที่ 4.21

ตารางที่ 4.21 ผลการทดสอบประสิทธิภาพตัวแบบพยากรณ์โดยใช้เทคนิคโครงข่ายประสาทเทียมด้วยข้อมูลที่ไม่จัดกลุ่ม

Feature Selection	Hidden layer	Learning	Accuracy	RMSE
Original	6	0.1	92.76	0.2051
CFS	3*	0.3*	94.48*	0.1880*
IG	6	0.1	92.38	0.2085

ผลทดสอบประสิทธิภาพเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines: SVM) ร่วมกับวิธีที่ไม่คัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะด้วย Correlation – based Feature Selection (CFS) และวิธี Information Gain ด้วยข้อมูลที่ไม่จัดกลุ่มในการสร้างตัวแบบพยากรณ์ ดังตารางที่ 4.22

ตารางที่ 4.22 ผลการทดสอบประสิทธิภาพตัวแบบพยากรณ์โดยใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนด้วยข้อมูลที่ไม่จัดกลุ่ม

Feature Selection	Kernel	Accuracy	RMSE
Original	RBF	93.90	0.2960
CFS	Puk*	94.29*	0.2996*
IG	RBF	93.71	0.2967

ผลทดสอบประสิทธิภาพเทคนิคต้นไม้ตัดสินใจ (Decision Tree) ด้วยอัลกอริทึม C4.5 ร่วมกับวิธีที่ไม่คัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะด้วย Correlation – based Feature

Selection (CFS) และวิธี Information Gain ด้วยข้อมูลที่ไม่จัดกลุ่มในการสร้างตัวแบบพยากรณ์ ดังตารางที่ 4.23

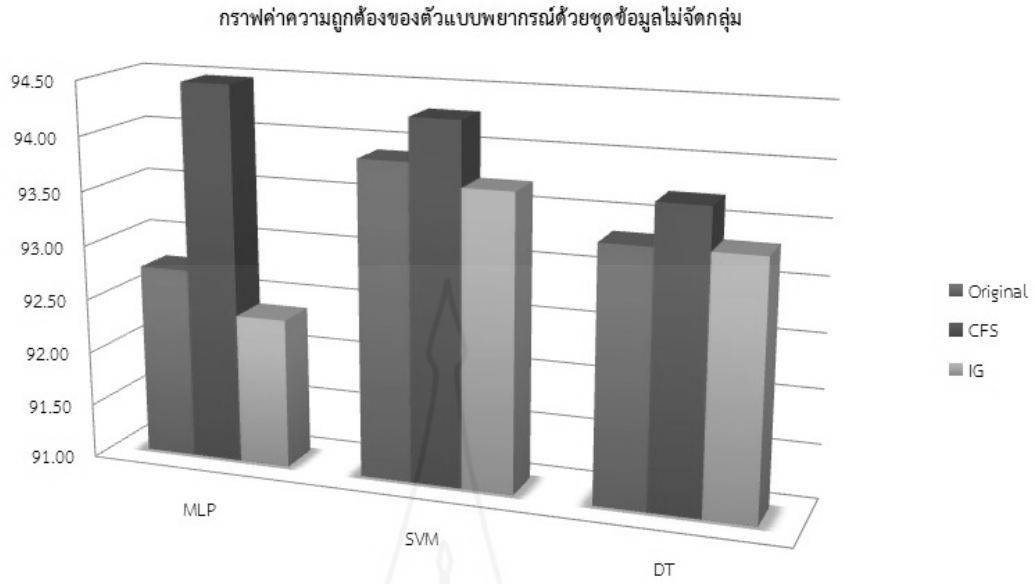
ตารางที่ 4.23 ผลการทดสอบประสิทธิภาพตัวแบบพยากรณ์โดยใช้เทคนิคต้นไม้ตัดสินใจ ด้วยข้อมูลที่ไม่จัดกลุ่ม

Feature Selection	**Accuracy**	**RMSE**
Original	93.33	0.2031
CFS	93.71*	0.1988*
IG	93.33	0.2041

จากผลการทดสอบประสิทธิภาพด้วยวิธี 10 fold Cross-Validation พบว่าการจำแนกข้อมูลด้วยวิธีการคัดเลือกตัวแปรหรือคุณลักษณะด้วยวิธีการ Correlation – based Feature Selection (CFS) ร่วมกับเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอนให้ค่าความถูกต้อง เท่ากับร้อยละ 94.48 ซึ่งให้ค่าความถูกต้องสูงกว่าการใช้งานร่วมกับเทคนิค ซัพพอร□ตเวกเตอร□แมชชีน และเทคนิคต้นไม้ตัดสินใจ อีกทั้งยังสูงกว่าวิธีการคัดเลือกตัวแปรหรือคุณลักษณะด้วย Information Gain (IG) ร่วมกับเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน เทคนิคซัพพอร□ตเวกเตอร□แมชชีน และเทคนิคต้นไม้ตัดสินใจ อีกทั้งสูงกว่าวิธีที่ใช้เทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน เทคนิคซัพพอร□ตเวกเตอร□แมชชีน และเทคนิคต้นไม้ตัดสินใจเพียงอย่างเดียว ดังตารางที่ 4.24

ตารางที่ 4.24 เปรียบเทียบค่าความถูกต้องของการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่ไม่จัดกลุ่ม

algorithm	Original	CFS	IG
MLP	92.76	94.48*	92.38
SVM	93.90	94.29	93.71
DT	93.33	93.71	93.33

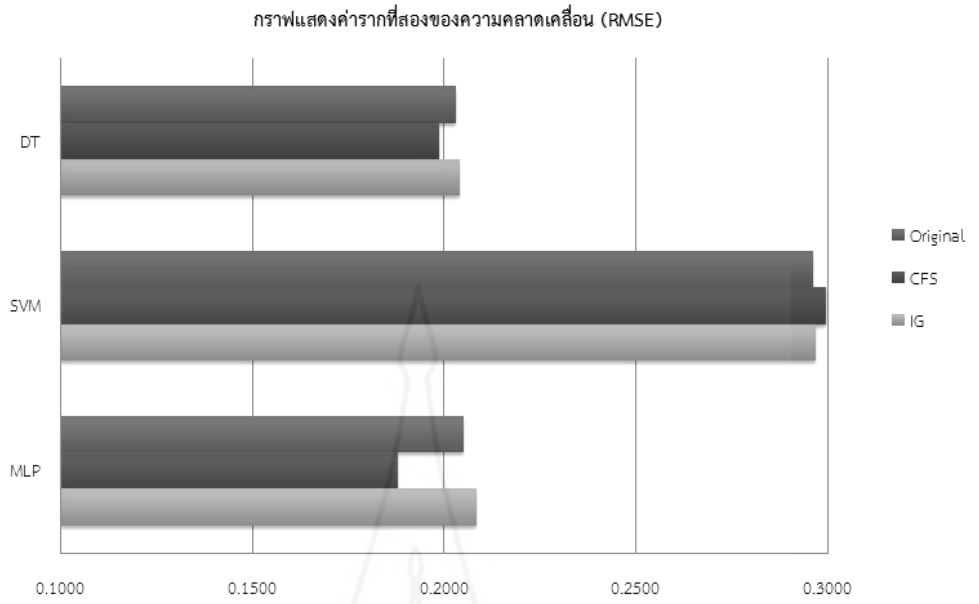


ภาพที่ 4.26 กราฟแสดงค่าความถูกต้องของการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่ไม่จัดกลุ่ม

และยังพบว่าการจำแนกข้อมูลด้วยวิธีการคัดเลือกตัวแปรหรือคุณลักษณะด้วยวิธีการ Correlation – based Feature Selection (CFS) ร่วมกับเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอนให้ค่ารากที่สองของความคลาดเคลื่อน (RMSE) น้อยที่สุด มีค่าเท่ากับ 0.1880 ซึ่งให้ค่ารากที่สองของความคลาดเคลื่อนน้อยกว่าการใช้งานร่วมกับเทคนิคซัพพอร□ตเวกเตอร□แมชชีน และเทคนิคต้นไม้ตัดสินใจ อีกทั้งยังน้อยกว่าวิธีการคัดเลือกตัวแปรหรือคุณลักษณะด้วย Information Gain (IG) ร่วมกับเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน เทคนิคซัพพอร□ตเวกเตอร□แมชชีน และเทคนิคต้นไม้ตัดสินใจ และน้อยกว่าวิธีที่ใช้เทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน เทคนิคซัพพอร□ตเวกเตอร□แมชชีน และเทคนิคต้นไม้ตัดสินใจเพียงอย่างเดียว ดังตารางที่ 4.25

ตารางที่ 4.25 เปรียบเทียบค่ารากที่สองของความคลาดเคลื่อนของการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่ไม่จัดกลุ่ม

algorithm	Original	Correlation-based Feature	Information Gain
MLP	0.2051	0.1880*	0.2085
SVM	0.2960	0.2996	0.2967
DT	0.2031	0.1988	0.2041



ภาพที่ 4.27 กราฟแสดงค่ารากที่สองของความคลาดเคลื่อนของการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่ไม่จัดกลุ่ม

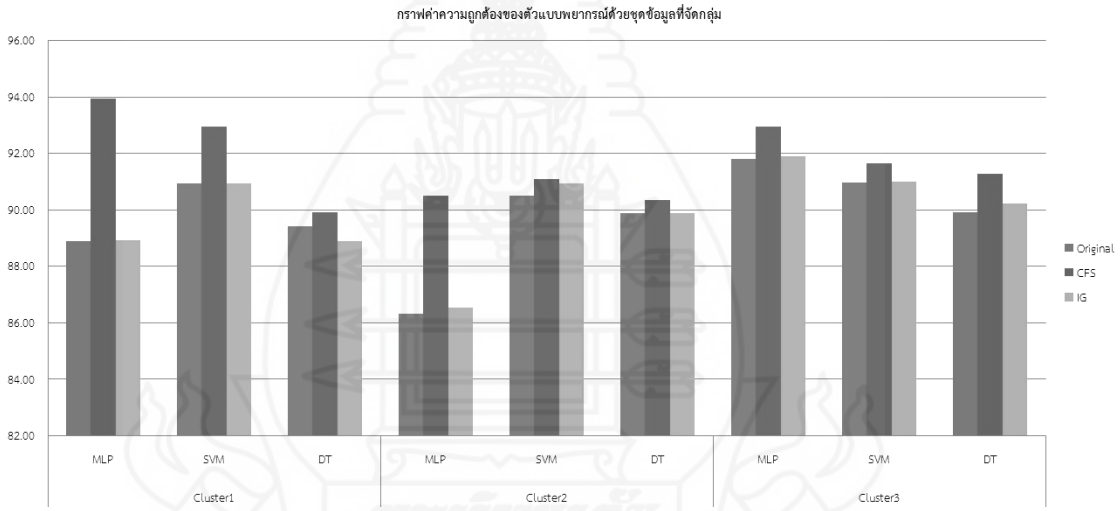
3.5.2 การสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่ม

จากการทดสอบประสิทธิภาพตัวแบบพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียม เทคนิคซัพพอร□ตเวกเตอร□แมชชีน และเทคนิคต้นไม้ตัดสินใจร่วมกับวิธีที่ไม่คัดเลือกคุณลักษณะ และวิธีการคัดเลือกคุณลักษณะด้วยข้อมูลที่จัดกลุ่มผลที่ได้ ดังตารางที่ 4.26

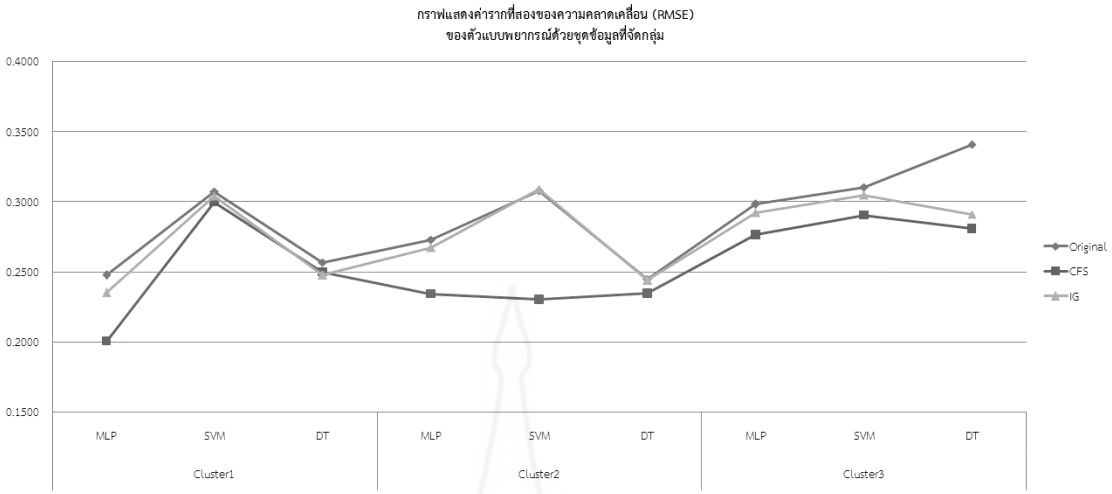
ตารางที่ 4.26 ผลการทดสอบประสิทธิภาพตัวแบบพยากรณ์ด้วยข้อมูลที่จัดกลุ่ม

Data Set	*Algorithm*	Feature Selection	Attribute	Accuracy	RMSE
Cluster1	MLP	Original	16	88.89	0.2474
		CFS	5	93.94*	0.2002*
		IG	10	88.90	0.2350
	SVM	Original	16	90.91	0.3070
		CFS	5	92.93	0.2996
		IG	10	90.93	0.3041
	C4.5	Original	16	89.39	02565
		CFS	5	89.90	0.2497
		IG	10	88.89	0.2474
Cluster2	MLP	Original	16	86.30	0.2729
		CFS	4	90.48	0.2339
		IG	10	86.52	0.2670
	SVM	Original	16	90.48	0.3076
		CFS	4	91.07*	0.2301*
		IG	10	90.92	0.3091
	C4.5	Original	16	89.88	0.2448
		CFS	4	90.34	0.2344
		IG	10	89.87	0.2437
Cluster3	MLP	Original	16	91.78	0.2980
		CFS	4	92.93*	0.2761*
		IG	9	91.89	0.2921
	SVM	Original	16	90.94	0.3102
		CFS	4	91.63	0.2901
		IG	9	90.98	0.3045
	C4.5	Original	16	89.89	0.3408
		CFS	4	91.27	0.2806
		IG	9	90.20	0.2908

จากตารางที่ 4.26 ผลการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่มแบ่งออกเป็น 3 กลุ่ม กลุ่ม Cluster1 พบว่า ตัวแบบพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ร่วมกับวิธีการคัดเลือกคุณลักษณะแบบ Correlation-based Feature Selection (CFS) ให้ค่าความถูกต้องร้อยละ 93.94 สูงที่สุด และค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดเท่ากับ 0.2002 ส่วนกลุ่ม Cluster2 พบว่า ตัวแบบพยากรณ์ด้วยเทคนิค ซัพพอร์ตเวกเตอร์แมชชีน (SVM) ร่วมกับวิธีการคัดเลือกคุณลักษณะแบบ Correlation-based Feature Selection (CFS) ให้ค่าความถูกต้องร้อยละ 91.07 สูงที่สุด และค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดเท่ากับ 0.2301 และกลุ่ม Cluster3 พบว่า ตัวแบบพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ร่วมกับวิธีการคัดเลือกคุณลักษณะแบบ Correlation-based Feature Selection (CFS) ให้ค่าความถูกต้องร้อยละ 92.93 สูงที่สุด และค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดเท่ากับ 0.2761 ดังภาพที่ 4.28 และภาพที่ 4.29



ภาพที่ 4.28 กราฟแสดงค่าความถูกต้องของตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่ม



ภาพที่ 4.29 กราฟแสดงค่ารากที่สองของความคลาดเคลื่อนตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่ม

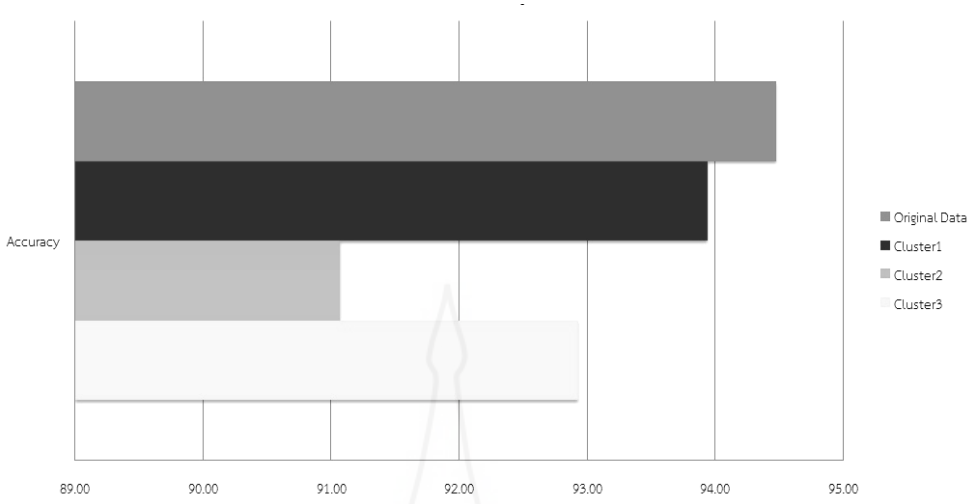
3.5.2 การเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ด้วยชุดข้อมูลที่ไม่จัดกลุ่มกับชุดข้อมูลที่จัดกลุ่ม

นำผลที่ได้จากตัวแบบพยากรณ์ด้วยชุดข้อมูลที่ไม่จัดกลุ่มที่มีค่าความถูกต้องสูงที่สุดมาเปรียบเทียบกับตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่มที่มีค่าความถูกต้องสูงที่สุดในแต่ละกลุ่ม ผลการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ ดังตารางที่ 4.27

ตารางที่ 4.27 การเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์

Data Set	*Algorithm*	Feature Selection	Attribute	Accuracy	RMSE
Original Data	MLP	CFS	5	94.48*	0.1880*
Cluster1	MLP	CFS	5	93.94	0.2002
Cluster2	SVM	CFS	4	91.07	0.2301
Cluster3	MLP	CFS	4	92.93	0.2761

จากตารางที่ 4.27 การเปรียบเทียบตัวแบบพยากรณ์พบว่า ตัวแบบพยากรณ์ด้วยชุดข้อมูลที่ไม่จัดกลุ่มด้วยเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ร่วมกับวิธีการคัดเลือกคุณลักษณะแบบ Correlation-based Feature Selection (CFS) ซึ่งคัดเลือกคุณลักษณะไว้ 5 คุณลักษณะ มีค่าความถูกต้อง (Accuracy) ร้อยละ 94.48 สูงกว่าตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่ม ดังภาพที่ 4.30 ส่วนค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดเท่ากับ 0.1880

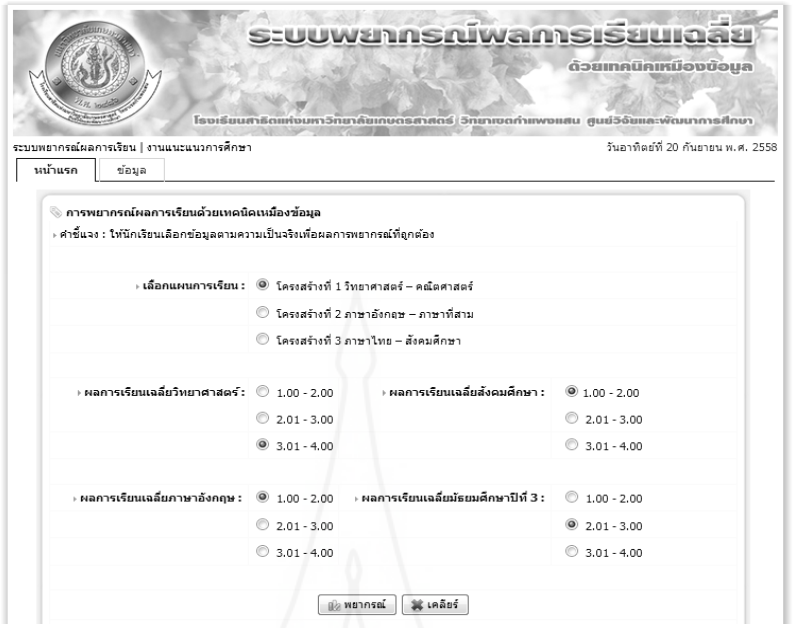


ภาพที่ 4.30 กราฟเปรียบเทียบประสิทธิภาพค่าความถูกต้อง ตัวแบบพยากรณ์

ดังนั้นในการสร้างระบบพยากรณ์ผู้วิจัยจึงใช้ตัวแบบพยากรณ์ที่ใช้ชุดข้อมูลที่ไม่จัดกลุ่ม โดยใช้วิธีการคัดเลือกคุณลักษณะแบบ Correlation-based Feature Selection (CFS) ร่วมกับเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) เหมาะสมที่สุดในการพยากรณ์ผลการเรียนของนักเรียน เนื่องจากให้ค่าความถูกต้องของการพยากรณ์ดีที่สุด และค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุด

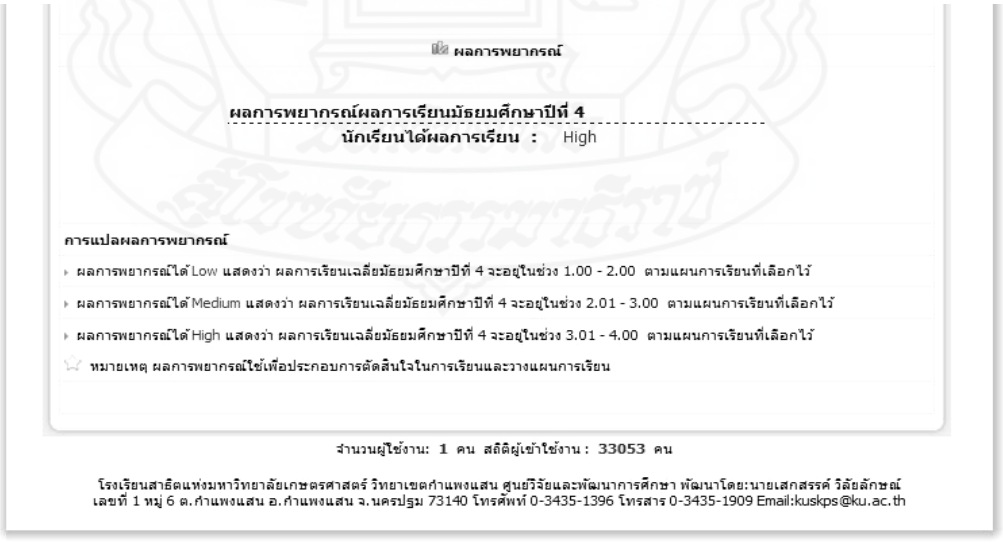
3.6 การนำไปใช้งาน (Deployment)

ระบบพยากรณ์ผลการเรียนเฉลี่ย ผู้วิจัยได้ใช้รูปแบบการแสดงผลบนเว็บ โดยใช้โปรแกรมภาษา PHP ร่วมกับโปรแกรม Adobe Dreamweaver CS 6.0 ในการพัฒนาเพื่อความสามารถสะดวกในการใช้งาน ซึ่งเป็นการนำตัวแบบพยากรณ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล นำมาใช้งานทำให้นักเรียนสามารถพยากรณ์ผลการเรียนเฉลี่ยในระดับมัธยมศึกษาปีที่ 4 เพื่อประกอบการตัดสินใจเลือกแผนการเรียนระดับมัธยมศึกษาตอนปลาย และวางแผนการเรียนเพื่อให้ได้ผลการเรียนที่สูงขึ้น นักเรียนสามารถนำผลการพยากรณ์ไปใช้ในการประกอบการตัดสินใจหรือเป็นแนวทางในการเรียน รวมทั้งอาจารย์แนะแนวใช้เป็นเครื่องมือช่วยนักเรียนในการแนะนำการเลือกแผนการเรียน และยังช่วยแนะแนวทางการเรียนให้นักเรียนได้ ผู้วิจัยจึงนำระบบการพยากรณ์ผลการเรียนเฉลี่ยมาใช้งาน เพื่อเป็นข้อมูลและประกอบการตัดสินใจของนักเรียน ดังภาพที่ 4.31



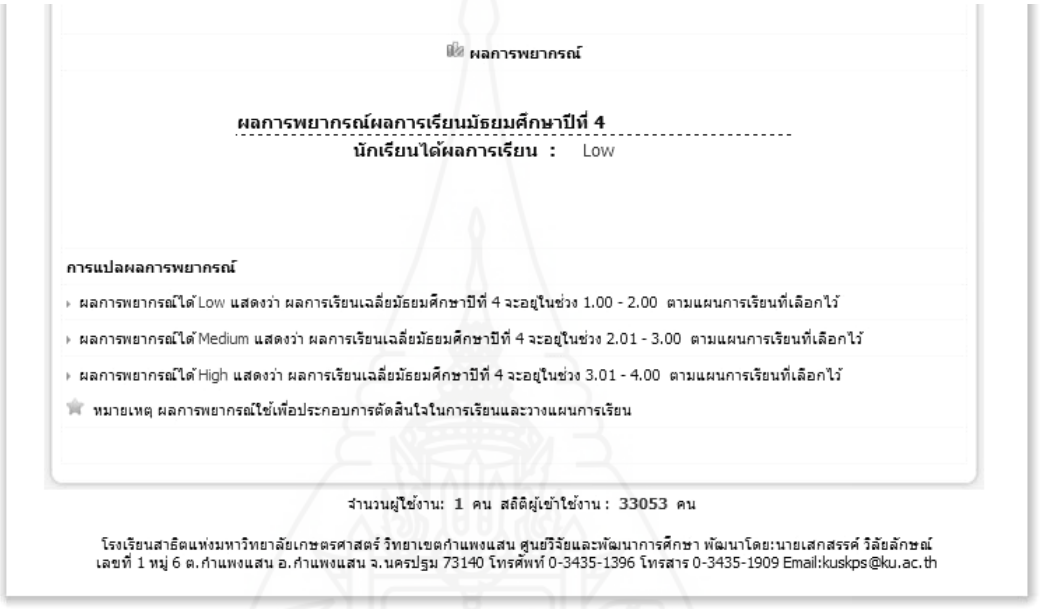
ภาพที่ 4.31 ตัวอย่างการใช้งานระบบพยากรณ์ผลการเรียนเฉลี่ย

จากภาพที่ 4.31 ถ้านักเรียนเลือกแผนการเรียน โครงสร้างที่ 1 วิทยาศาสตร์ –คณิตศาสตร์ มีผลการเรียนเฉลี่ยวิทยาศาสตร์อยู่ระหว่าง 3.01 – 4.00 (High) ผลการเรียนเฉลี่ยสังคมศึกษาอยู่ระหว่าง 1.00 – 2.00 (Low) ผลการเรียนเฉลี่ยภาษาอังกฤษอยู่ระหว่าง 1.00 – 2.00 (Low) และผลการเรียนเฉลี่ยมัธยมศึกษาปีที่ 3 อยู่ระหว่าง 2.01 – 3.00 (Medium) ผลการพยากรณ์เป็นดังภาพที่ 4.32



ภาพที่ 4.32 แสดงตัวอย่างผลการพยากรณ์แผนการเรียน โครงสร้างที่ 1

จากภาพที่ 4.32 ผลการพยากรณ์ที่ได้ของนักเรียนอยู่ในระดับสูง (High) ซึ่งหมายความว่า ถ้านักเรียนเลือกเรียนแผนการเรียนโครงสร้างที่ 1 วิทยาศาสตร์ – คณิตศาสตร์ จะมีผลการเรียนระดับชั้นมัธยมศึกษาปีที่ 4 อยู่ระหว่าง 3.01 – 4.00 (High) แต่ถ้านักเรียนลองเปลี่ยนแผนการเรียนในระบบพยากรณ์จากโครงสร้างที่ 1 วิทยาศาสตร์ – คณิตศาสตร์ เป็นโครงสร้างที่ 3 ภาษาไทย – สังคม ผลที่ได้ ดังภาพที่ 4.33



ภาพที่ 4.33 แสดงตัวอย่างผลการพยากรณ์แผนการเรียนโครงสร้างที่ 3

จากภาพที่ 4.33 ผลการพยากรณ์ที่ได้ของนักเรียนอยู่ในระดับต่ำ (Low) ซึ่งหมายความว่า ถ้านักเรียนเลือกเรียนแผนการเรียนโครงสร้างที่ 3 ภาษาไทย – สังคม จะมีผลการเรียนระดับชั้นมัธยมศึกษาปีที่ 4 อยู่ระหว่าง 1.00 – 2.00 (Low) ดังนั้นนักเรียนก็จะมีข้อมูลที่ใช้ในการประกอบการตัดสินใจในการเลือกแผนการเรียน และพัฒนาผลการเรียนให้ดีขึ้นเพื่อที่จะได้แผนการเรียนที่นักเรียนต้องการ

4. การประเมินผล

4.1 ประเมินความพึงพอใจระบบคลังข้อมูลผลการเรียน

การประเมินความพึงพอใจระบบคลังข้อมูลผลการเรียน โดยใช้แบบสอบถามเป็นเครื่องมือในการประเมิน ซึ่งมีมาตรวัดอัตราส่วน 5 ระดับ (Rating Scale) ได้แก่

ดีมาก (ระดับ 5) หมายถึง ระดับความพึงพอใจมากที่สุด
ดี (ระดับ 4) หมายถึง ระดับความพึงพอใจมาก
ปานกลาง (ระดับ 3) หมายถึง ระดับความพึงพอใจปานกลาง
พอใช้ (ระดับ 2) หมายถึง ระดับความพึงพอใจน้อย
ปรับปรุง (ระดับ 1) หมายถึง ระดับความพึงพอใจน้อยที่สุด

จากข้อมูลที่ได้จากการตอบแบบสอบถามของกลุ่มผู้ใช้งาน จำนวน 10 คน ซึ่งแบ่งเป็นผู้บริหาร 2 คน หัวหน้ากลุ่มสาระการเรียนรู้ 3 คน และอาจารย์ประจำชั้น 5 คน นำมาหาจำนวนและค่าเฉลี่ยของผู้ตอบในแต่ละข้อคำถาม ดังตารางที่ 4.28

ตารางที่ 4.28 ค่าเฉลี่ยของการประเมินความพึงพอใจระบบคลังข้อมูล

รายการประเมิน	ดีมาก	ดี	ปานกลาง	พอใช้	ปรับปรุง	N	$\bar{x}$
1. มีความสะดวกและง่ายต่อการใช้งาน	3	3	3	1	0	10	3.80
2. การออกแบบรายงานเข้าใจได้ง่าย	3	4	2	1	0	10	3.90
3. ความถูกต้องของข้อมูล	3	3	4	0	0	10	3.90
4. ช่วยลดขั้นตอนในการทำงาน	5	3	2	0	0	10	4.30
5. ข้อมูลตรงกับความต้องการผู้ใช้งาน	5	4	1	0	0	10	4.40
รวม	19	17	12	2	0	50	4.06

จากตารางที่ 4.28 นำค่าเฉลี่ยที่ได้มาเปรียบเทียบกับเกณฑ์ของ เบสท์ (Best.,1963: 204-208) ดังนี้

ค่าเฉลี่ย 4.51 – 5.00 หมายถึง ความคิดเห็นอยู่ในระดับดีมาก
ค่าเฉลี่ย 3.51 – 4.50 หมายถึง ความคิดเห็นอยู่ในระดับดี
ค่าเฉลี่ย 2.51 – 3.50 หมายถึง ความคิดเห็นอยู่ในระดับปานกลาง
ค่าเฉลี่ย 1.51 – 2.50 หมายถึง ความคิดเห็นอยู่ในระดับน้อย
ค่าเฉลี่ย 1.01 – 1.50 หมายถึง ความคิดเห็นอยู่ในระดับน้อยมาก

ผลที่ได้จากการวิเคราะห์ข้อมูลจากแบบสอบถาม จำนวน 10 ชุด นำมาทำการแปลผลเทียบกับเกณฑ์ ดังตารางที่ 4.29

ตารางที่ 4.29 ผลการประเมินความพึงพอใจระบบคลังข้อมูลผลการเรียน

รายการประเมิน	ค่าเฉลี่ย ($\bar{x}$)	การแปลผล
1. มีความสะดวกและง่ายต่อการใช้งาน	3.80	ดี
2. การออกแบบรายงานเข้าใจได้ง่าย	3.90	ดี
3. ความถูกต้องของข้อมูล	3.90	ดี
4. ช่วยลดขั้นตอนในการทำงาน	4.30	ดี
5. ข้อมูลตรงกับความต้องการผู้ใช้งาน	4.40	ดี
ความพึงพอใจโดยภาพรวม	4.06	ดี

จากตารางที่ 4.29 ผลของการประเมินระดับความพึงพอใจระบบคลังข้อมูลผลการเรียน พบว่า มีข้อมูลตรงกับความต้องการผู้ใช้งานมีค่าเฉลี่ยเท่ากับ 4.40 ช่วยลดขั้นตอนในการทำงานมีค่าเฉลี่ยเท่ากับ 4.30 ในการออกแบบรายงานเข้าใจได้ง่ายมีค่าเฉลี่ยเท่ากับ 3.90 ส่วนความถูกต้องของข้อมูลในการรายงานมีค่าเฉลี่ยเท่ากับ 3.90 และคลังข้อมูลมีความสะดวกและง่ายต่อการใช้งานค่าเฉลี่ยเท่ากับ 3.80 ผลการประเมินความพึงพอใจของผู้ใช้งานมีค่าเฉลี่ยรวมเท่ากับ 4.06 ซึ่งเมื่อแปลผลมีระดับความพึงพอใจอยู่ในระดับดี คลังข้อมูลผลการเรียนของนักเรียนมีประสิทธิภาพสามารถนำไปใช้งานได้

4.2 แบบประเมินความเชื่อมั่นของระบบพยากรณ์ผลการเรียนเฉลี่ย

การประเมินความเชื่อมั่นระบบพยากรณ์ผลการเรียนเฉลี่ย โดยใช้แบบสอบถามเป็นเครื่องมือในการประเมิน ซึ่งมีมาตรวัดอัตราส่วน 5 ระดับ (Rating Scale) ได้แก่

มากที่สุด (ระดับ 5) หมายถึง ระดับความเชื่อมั่นมากที่สุด
มาก (ระดับ 4) หมายถึง ระดับความเชื่อมั่นมาก
ปานกลาง (ระดับ 3) หมายถึง ระดับความเชื่อมั่นปานกลาง
น้อย (ระดับ 2) หมายถึง ระดับความเชื่อมั่นน้อย
น้อยที่สุด (ระดับ 1) หมายถึง ระดับความเชื่อมั่นน้อยที่สุด

จากข้อมูลที่ได้จากการตอบแบบสอบถามของกลุ่มผู้ใช้งาน จำนวน 40 คน ซึ่งได้แก่ อาจารย์แนะแนว 2 คน นักเรียนระดับมัธยมศึกษาปีที่ 3 จำนวน 19 คน และนักเรียนระดับมัธยมศึกษาปีที่ 4 จำนวน 19 คน นำมาหาจำนวน และค่าเฉลี่ย ของผู้ตอบในแต่ละข้อคำถาม ดังตารางที่ 4.30

ตารางที่ 4.30 ค่าเฉลี่ยของการประเมินความเชื่อมั่นระบบพยากรณ์ผลการเรียนเฉลี่ย

รายการประเมิน	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยสุด	N	$\bar{x}$
1. ความสะดวกและง่ายต่อการใช้งาน	17	13	8	2	0	40	4.33
2. ความน่าเชื่อถือของผลการพยากรณ์	14	14	9	3	0	40	4.20
3. ผลการพยากรณ์ตรงกับความต้องการ	15	12	11	2	0	40	4.25
4. ช่วยประกอบการตัดสินใจ	16	13	8	3	0	40	4.28
รวม	62	52	36	10	0	160	4.27

จากตารางที่ 4.30 นำค่าเฉลี่ยที่ได้มาเปรียบเทียบกับเกณฑ์ของ เบสท์ (Best.,1963: 204-208) ดังนี้

ค่าเฉลี่ย 4.51 – 5.00 หมายถึง ความคิดเห็นอยู่ในระดับดีมาก
ค่าเฉลี่ย 3.51 – 4.50 หมายถึง ความคิดเห็นอยู่ในระดับดี
ค่าเฉลี่ย 2.51 – 3.50 หมายถึง ความคิดเห็นอยู่ในระดับปานกลาง
ค่าเฉลี่ย 1.51 – 2.50 หมายถึง ความคิดเห็นอยู่ในระดับน้อย
ค่าเฉลี่ย 1.01 – 1.50 หมายถึง ความคิดเห็นอยู่ในระดับน้อยมาก

ผลที่ได้จากการวิเคราะห์ข้อมูลจากแบบสอบถาม จำนวน 40 ชุด นำมาทำการแปลผลเทียบกับเกณฑ์ ดังตารางที่ 4.31

ตารางที่ 4.31 ผลการการประเมินความเชื่อมั่นระบบพยากรณ์ผลการเรียนเฉลี่ย

รายการประเมิน	**ค่าเฉลี่ย ($\bar{x}$)**	**การแปลผล**
1. ความสะดวกและง่ายต่อการใช้งาน	4.33	ดี
2. ความน่าเชื่อถือของผลการพยากรณ์	4.20	ดี
3. ผลการพยากรณ์ตรงกับความต้องการ	4.25	ดี
4. ช่วยประกอบการตัดสินใจ	4.28	ดี
ความเชื่อมั่นโดยภาพรวม	4.27	ดี

จากตารางที่ 4.30 ผลของการประเมินความเชื่อมั่นระบบพยากรณ์ผลการเรียนเฉลี่ย พบว่า ในเรื่องของความสะดวกและง่ายต่อการใช้งานมีค่าเฉลี่ยเท่ากับ 4.33 สามารถนำไปประกอบการตัดสินใจมีค่าเฉลี่ยเท่ากับ 4.28 ส่วนการผลการพยากรณ์ตรงกับความต้องการมีค่าเฉลี่ยเท่ากับ 4.25 และมีความน่าเชื่อถือของผลการพยากรณ์มีค่าเฉลี่ยเท่ากับ 4.20 ผลการประเมินความความเชื่อมั่นมีค่าเฉลี่ยรวมเท่ากับ 4.27 เมื่อเทียบกับเกณฑ์อยู่ในระดับดี ระบบพยากรณ์ผลการเรียนของนักเรียนมีความน่าเชื่อถือสามารถนำไปใช้งานได้

บทที่ 5
สรุปผลการวิจัย

การศึกษาวิจัยเรื่อง การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา ผู้วิจัยได้พัฒนาคลังข้อมูลผลการเรียนของนักเรียนระดับมัธยมศึกษา ในการนำเสนอรายงาน และนำมาวิเคราะห์ข้อมูลหาความรู้ใหม่ด้วยการทำเหมืองข้อมูล ได้ข้อสรุปผลการวิจัย การอภิปรายผล และข้อเสนอแนะดังนี้

1. สรุปผลการวิจัย

1.1 การพัฒนาคลังข้อมูลผลการเรียนของนักเรียน

งานวิจัยนี้ผู้วิจัยนำข้อมูลผลการเรียนจากฐานข้อมูลงานทะเบียนและประมวลผลที่จัดเก็บข้อมูลนักเรียน อาจารย์ รายวิชา และผลการเรียนของนักเรียนระดับมัธยมศึกษา โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา นำข้อมูลที่ได้มาสร้างคลังข้อมูล โดยใช้โปรแกรม Microsoft SQL Server 2008 R2 และออกแบบโครงสร้างแบบสโนว์เฟลก (Snowflake Schema) ผ่านกระบวนการอีทีแอล (Extract Transform and Load : ETL) ซึ่งผู้วิจัยได้ออกแบบส่วนของการนำเข้าข้อมูล ซึ่งข้อมูลที่ได้จากฐานข้อมูลอยู่ในรูปแบบของไฟล์เอ็กเซล (Excel) โดยทำการคัดแยกข้อมูลที่ไม่เกี่ยวข้อง จัดโครงสร้างและรูปแบบข้อมูลให้อยู่ในรูปแบบเดียวกัน ถือเป็นการทำความสะอาดข้อมูลให้มีความสมบูรณ์และถูกต้องก่อนนำข้อมูลเข้าคลังข้อมูล ซึ่งในการทำกระบวนการอีทีแอล (ETL) ผู้วิจัยใช้ Microsoft SQL Server Integration Service เป็นเครื่องมือนำข้อมูลเข้าคลังข้อมูล

1.2 การพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ

ผู้วิจัยได้ใช้เครื่องมือ Microsoft SQL Server Analysis Service (SSAS) สร้าง Analysis Service Project สำหรับการวิเคราะห์และสร้างคิวบ์ข้อมูล (Data Cube) ที่เป็นฐานข้อมูลแบบหลายมิติ และกำหนดข้อมูลจากตารางข้อมูลที่มีอยู่ในคลังข้อมูลที่ได้ออกแบบโครงสร้างคลังข้อมูลแบบสโนว์เฟลก สกีม่า (Snowflake Schema) ซึ่งในการประมวลผลเชิงวิเคราะห์ออนไลน์ (Online Analytical Processing : OLAP) ประกอบด้วยมิติมุมมอง (Dimension) คือ มิติภาคเรียน (Dim Semester), มิติ

ระดับชั้น (Dim Class), มิติวิชา (Dim Subject), มิตินักเรียน (Dim Student) และมิติอาจารย์ (Dim Teacher) โดยมีเมเชอร์หรือค่าที่ต้องการวัด (Measure) คือ คะแนน (Score), ผลการเรียน (Grade) และผลการเรียนเฉลี่ย (GPA) ซึ่งสามารถนำข้อมูลมาแสดงแบบหลายมิติได้ โดยในงานวิจัยนี้ใช้โปรแกรม QlikView ในการติดต่อกับคลังข้อมูลเพื่อทำการคิวรีข้อมูลของ OLAP มาสร้างเป็นมุมมองของกลุ่มผู้ใช้งาน จำนวน 3 กลุ่ม คือ ผู้บริหารแสดงรายงานในภาพรวมของผลการเรียนของโรงเรียนและจำนวนนักเรียน หัวหน้ากลุ่มสารการเรียนรู้แสดงรายงานของรายวิชา ผลการเรียนเฉลี่ยในแต่ละภาคเรียน และอาจารย์ประจำชั้นแสดงรายงานของคะแนน ผลการเรียนแต่ละวิชาของนักเรียนในแต่ละระดับชั้น จากผลการประเมินความพึงพอใจคลังข้อมูลผลการเรียน โดยผู้ใช้งานจำนวน 10 คน พบว่า ผลการประเมินความความพึงพอใจโดยรวมมีค่าเฉลี่ยเท่ากับ 4.04 ถือว่าคลังข้อมูลผลการเรียนที่พัฒนาขึ้นมีความพึงพอใจต่อการใช้งาน โดยรวมอยู่ในระดับดี สามารถนำไปใช้เป็นเครื่องมือในการนำเสนอข้อมูลเพื่อประกอบการตัดสินใจหรือแก้ปัญหาของผู้ใช้งานในระดับผู้บริหาร หัวหน้ากลุ่มสาระการเรียนรู้ และอาจารย์ประจำชั้นได้

1.3 การทำเหมืองข้อมูล

ข้อมูลที่ใช้ในการทำเหมืองข้อมูลได้มาจากคลังข้อมูลผลการเรียนของนักเรียนที่สร้างขึ้นมา ซึ่งใช้เทคนิคการทำเหมืองข้อมูลมาสร้างระบบพยากรณ์ผลการเรียนเฉลี่ย โดยผู้วิจัยดำเนินงานวิจัยในการทำเหมืองข้อมูลตามแบบจำลองของคริสพ์ - ดีเอ็ม (CRISP - DM Model) ใช้ข้อมูลนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 ระหว่างปีการศึกษา 2553 – 2556 ประกอบด้วยข้อมูลประวัติส่วนตัวและข้อมูลผลการเรียนของนักเรียนเป็นข้อมูลนำเข้า จำนวน 16 คุณลักษณะ คือ เพศ, แผนการเรียน, อาชีพของบิดา, รายได้เฉลี่ยต่อปีของบิดา, อาชีพของมารดา, รายได้เฉลี่ยต่อปีของมารดา, สถานภาพของบิดามารดา, จำนวนพี่น้อง, ผลการเรียนเฉลี่ยคณิตศาสตร์, ผลการเรียนเฉลี่ยวิทยาศาสตร์, ผลการเรียนเฉลี่ยภาษาไทย, ผลการเรียนเฉลี่ยสังคมศึกษา, ผลการเรียนเฉลี่ยภาษาอังกฤษ, ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 1, ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 2 และผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3

การสร้างตัวแบบพยากรณ์ได้ใช้ข้อมูลเป็น 2 แบบ คือ ชุดข้อมูลที่ไม่จัดกลุ่ม (Original Data) เป็นข้อมูลนักเรียนจำนวน 525 ระเบียน และชุดข้อมูลที่จัดกลุ่ม (Cluster Data) เป็นข้อมูลที่ผ่านการจัดกลุ่มข้อมูลด้วยอัลกอริทึม k-means ได้ 3 กลุ่ม คือ Cluster1 มีจำนวนข้อมูล 198 ระเบียน Cluster2 มีจำนวนข้อมูล 168 ระเบียน และ Cluster3 มีจำนวนข้อมูล 159 ระเบียน มาผ่านวิธีการคัดเลือกคุณลักษณะ 2 วิธี คือ วิธี Correlation-based Feature Selection (CFS) และวิธี Information Gain (IG) ร่วมกับเทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน, เทคนิคซัพพอร์ตเวกเตอร์แมชชีน และเทคนิคต้นไม้ตัดสินใจ ผลการวิจัยพบว่า ตัว

แบบพยากรณ์ด้วยชุดข้อมูลที่ไม่จัดกลุ่มด้วยวิธีการคัดเลือกคุณลักษณะแบบ Correlation-based Feature Selection (CFS) ร่วมกับเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) มีค่าความถูกต้อง (Accuracy) ร้อยละ 94.48 และมีค่ารากที่สองของความคลาดเคลื่อนเท่ากับ 0.1880 ซึ่งมีมากกว่าตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่ม และค่าความถูกต้องถือว่าเป็นค่าที่มากกว่าสมมติฐานที่ตั้งไว้ คือ ตัวแบบการพยากรณ์ผลการเรียนเฉลี่ย มีค่าความถูกต้องไม่น้อยกว่าร้อยละ 60 ผู้วิจัยจึงนำตัวแบบไปพัฒนาระบบพยากรณ์ผลการเรียนเฉลี่ย

การพัฒนาระบบพยากรณ์ผลการเรียนเฉลี่ย ผู้วิจัยได้ใช้รูปแบบการแสดงผลบนเว็บ โดยใช้โปรแกรมภาษา PHP ในการพัฒนาร่วมกับโปรแกรม Adobe Dreamweaver CS 6.0 นอกจากนี้ยังใช้โปรแกรม AppServ 2.5.10 ในการพัฒนาเพื่อความสามารถในการใช้งาน จากผลการประเมินความเชื่อมั่นของระบบพยากรณ์ผลการเรียนเฉลี่ย โดยผู้ใช้งานจำนวน 40 คน พบว่า ผลการประเมินความเชื่อมั่นโดยรวมมีค่าเฉลี่ยเท่ากับ 4.27 ถือว่าระบบพยากรณ์ผลการเรียนที่พัฒนาขึ้นมีความเชื่อมั่นของการใช้งาน โดยรวมอยู่ในระดับดี สามารถนำไปใช้เป็นเครื่องมือช่วยนักเรียนประกอบการตัดสินใจเลือกแผนการเรียน และวางแผนการเรียนของนักเรียนได้

2. การอภิปรายผล

2.1 การพัฒนาคลังข้อมูลผลการเรียนของนักเรียน

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาคลังข้อมูลผลการเรียนของนักเรียน ซึ่งในการพัฒนาคลังข้อมูลใช้ซอฟต์แวร์ไมโครซอฟท์เอสคิวแอลเซิร์ฟเวอร์ 2008 อาร์ทู นั้นสามารถสร้างคลังข้อมูลผลการเรียน จากแหล่งข้อมูลภายในหน่วยงานของโรงเรียน โดยรวบรวมข้อมูล ทำการวิเคราะห์และออกแบบคลังข้อมูล ด้วยโปรแกรมไมโครซอฟท์เอสคิวแอลเซิร์ฟเวอร์อินทิเกรชันเซอร์วิสในการสร้างกระบวนการอีทีแอล ในการดึงข้อมูล เปลี่ยนรูปแบบข้อมูล และนำข้อมูลเข้าคลังข้อมูล และสามารถพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ ซึ่งเป็นการประมวลผลเชิงวิเคราะห์ออนไลน์หรือโอแลป (OLAP) และการสร้างรายงานด้วยโปรแกรม QlikView เป็นเครื่องมือที่ช่วยดึงข้อมูลและนำเสนอข้อมูลในหลายมิติ(Multidimensional) จากหลายๆ มุมมอง โดยที่ออกแบบมาสำหรับกลุ่มผู้ใช้งานในระดับผู้บริหารที่ต้องใช้ข้อมูลนักเรียนเพื่อประกอบการตัดสินใจ ระดับหัวหน้ากลุ่มสาระการเรียนรู้ที่ใช้ข้อมูลในการประเมินผลระดับกลุ่มสาระการเรียนรู้ และระดับอาจารย์ประจำชั้นที่ต้องใช้ข้อมูลผลการเรียน คะแนนของนักเรียนในการช่วยเหลือในด้านการเรียน ซึ่งสอดคล้องกับงานวิจัยของ สุประพล เลาวพงศ์ (2553 :67) ที่พัฒนาคลังข้อมูลด้านนักศึกษาของมหาวิทยาลัยพายัพ เพื่อสนับสนุนการตัดสินใจของผู้บริหาร ซึ่งผลของการวิจัยพบว่า

การพัฒนาคลังข้อมูลด้านนักศึกษาสามารถทำงานได้ตามวัตถุประสงค์ และความต้องการของผู้บริหาร และยังสอดคล้องกับงานวิจัยของ

2.2 การทำเหมืองข้อมูล

ในงานวิจัยนี้ นำเสนอการใช้เทคนิคเหมืองข้อมูลในการสร้างตัวแบบพยากรณ์ผลการเรียนเฉลี่ยของนักเรียนด้วยวิธีการคัดเลือกคุณลักษณะร่วมกับเทคนิคเหมืองข้อมูล เพื่อใช้มาประกอบการตัดสินใจของนักเรียน และเป็นข้อมูลการให้คำปรึกษาทางการเรียนของอาจารย์แนะแนว ซึ่งผลการวิจัยสอดคล้องกับงานวิจัยของ ทิพย์หทัย ทองธรรมชาติ (2555: 112-113) ที่ใช้เทคนิคเหมืองข้อมูลมาพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนิสิตสาขาวิทยาการคอมพิวเตอร์ และสาขาเทคโนโลยีสารสนเทศ ซึ่งพบว่าเทคนิคเนอีฟเบย์ และเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับให้ค่าความถูกต้องในการพยากรณ์สูงใกล้เคียงกัน และ สุพัฒน์กุล ภัคโชค (2556: 73-74) เป็นการค้นหาความสัมพันธ์ที่มีผลต่อการเลือกแผนการเรียน ใช้แบบสอบถามเป็นเครื่องมือเก็บข้อมูล โดยใช้เทคนิคเหมืองข้อมูลในการหาความสัมพันธ์ในการเลือกแผนการเรียนในระดับมัธยมศึกษาตอนปลาย พบว่า อัลกอริทึมต้นไม้ตัดสินใจให้ผลลัพธ์มีค่าความถูกต้องเท่ากับ 79.76% นำตัวแบบที่ได้ไปพัฒนาระบบเลือกแผนการเรียนระดับมัธยมศึกษาตอนปลาย ใช้เป็นเครื่องมือช่วยตัดสินใจเลือกแผนการเรียนได้ และของ วงกต ศรีอุไร (2557: 101) ที่ได้สร้างโมเดลในการจำแนกผู้ป่วยโรคอ้วนลงพุง โดยใช้วิธีการคัดเลือกคุณลักษณะร่วมกับเทคนิคเหมืองข้อมูล พบว่าวิธีการคัดเลือกคุณลักษณะด้วย Correlation-based Feature Selection (CFS) ร่วมกับเทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน ในการสร้างโมเดลจำแนกข้อมูลให้ค่าความถูกต้องในการจำแนกสูงที่สุด

นอกจากนี้ กระบวนการทำเหมืองข้อมูลเป็นการวิเคราะห์ข้อมูลเพื่อค้นหาความรู้ที่ซ่อนเร้นอยู่ในข้อมูลที่มีจำนวนมาก Berry&Linoff กล่าวว่า "*การทำเหมืองข้อมูลเป็นกระบวนการของการคัดแยกและจัดรูปแบบชุดข้อมูลจากแหล่งข้อมูลหลายๆ แหล่งให้มาเป็นรูปแบบเดียวกัน จากนั้นมีการนำชุดข้อมูลดังกล่าวเข้าสู่คลังข้อมูลที่ได้ออกแบบไว้ และสร้างแบบจำลองขึ้นมาวิเคราะห์ชุดข้อมูลนั้น ผลลัพธ์ที่ได้จากการวิเคราะห์คือความรู้ใหม่ที่ไม่เคยทราบมาก่อน และสามารถนำไปใช้ประโยชน์ได้*" ซึ่งงานวิจัยนี้ได้นำเทคนิคการทำเหมืองข้อมูลมาวิเคราะห์ข้อมูลผลการเรียนเฉลี่ยของนักเรียน สำหรับการพยากรณ์ผลการเรียนเฉลี่ยของนักเรียนในระดับมัธยมศึกษาชั้นปีที่ 4 ผลที่ได้จากการพยากรณ์ของตัวแบบที่สร้างขึ้นนั้น มีร้อยละของความถูกต้องค่อนข้างสูง โดยผลการพยากรณ์ถูกนำเสนอเป็นแนวโน้มของผลการเรียนเฉลี่ยว่าจะสูง ปานกลาง หรือต่ำ เพื่อเป็นข้อมูลช่วยในการตัดสินใจของนักเรียนในการเลือกแผนการเรียนในระดับมัธยมศึกษาตอนปลาย และเป็นข้อมูลให้คำปรึกษาของอาจารย์แนะแนว ที่จะช่วยแนะนำแผนการเรียนที่

เหมาะสมกับความถนัดและความสามารถของนักเรียน ที่ได้ผลการเรียนเฉลี่ยต่ำ ส่งผลให้นักเรียนสามารถเลือกแผนการเรียนที่เหมาะกับตนเอง วางแผนการเรียนในขณะศึกษาอยู่ในระดับมัธยมศึกษาตอนต้น และลดการย้ายโรงเรียนหรือการลาออกกลางคันเนื่องจากการมีผลการเรียนต่ำกว่าเกณฑ์การประเมินผล

ซึ่งผลการพยากรณ์ที่ได้ มิได้หมายความว่านักเรียนที่ใช้งานระบบการพยากรณ์ จะมีผลการเรียนตามผลการพยากรณ์ร้อยเปอร์เซ็นต์ แต่หมายความว่าข้อมูลของนักเรียนคนนั้น มีผลการเรียนเฉลี่ยในรูปแบบเดียวกับกลุ่มของแผนการเรียนนั้นๆ ซึ่งเป็นเพียงข้อมูลที่ใช้ประกอบการตัดสินใจ นอกจากนี้ยังต้องอาศัยความเอาใจใส่ การตั้งใจเรียน ความรับผิดชอบต่อการเรียนของนักเรียนถึงจะทำให้ผลการเรียนสูงขึ้น และยังมีปัจจัยต่างๆ ที่ส่งผลต่อการเรียนอีก เช่น ความเอาใจใส่ของผู้ปกครอง การสอนของอาจารย์แต่ละวิชา สิ่งแวดล้อมในการเรียน เป็นต้น

3. ปัญหาและอุปสรรค

3.1 ในการนำข้อมูลออกมาจากโปรแกรมงานทะเบียนและประมวลผล มีความซับซ้อนและใช้เวลานาน เนื่องมาจากโปรแกรมที่ใช้ในการจัดเก็บข้อมูลนั้นเป็นโปรแกรมที่ใช้งานในระบบดอส การนำข้อมูลออกมาจึงมีขั้นตอนที่ซับซ้อน และข้อมูลบางส่วนเกิดการเสียหาย

3.2 ข้อมูลที่ใช้เป็นข้อมูลที่มีรายชื่ออาจารย์และนักเรียนซึ่งไม่สามารถเปิดเผยได้ เนื่องจากผิดข้อกฎหมายของการละเมิดสิทธิและเสรีภาพของบุคคล จึงต้องใช้นามสมมติแทนชื่อและนามสกุลจริง

3.3 ข้อมูลที่ใช้ในการทำตัวแบบพยากรณ์ผลการเรียนยังมีน้อย ทั้งนี้เนื่องมาจากมีการปรับเปลี่ยนมาใช้หลักสูตรการศึกษา 2551 ได้เพียง 3 ปี

3.4 ข้อมูลที่ใช้ในการสร้างตัวแบบพยากรณ์ผลการเรียนได้จากคลังข้อมูลที่พัฒนาขึ้น มีการคัดเลือกข้อมูล (Data Cleansing) ที่ไม่สมบูรณ์ออกไป ทำให้เหลือข้อมูลที่ใช้ในการสร้างตัวแบบเพียง 525 เรคคอร์ด

3.5 การพัฒนาระบบพยากรณ์ที่ได้ใช้ได้เฉพาะระบบอินทราเน็ตภายในโรงเรียนเท่านั้น

4. ข้อเสนอแนะ

4.1 ข้อเสนอแนะในการนำผลการวิจัยไปใช้

4.1.1 การศึกษาวิจัยครั้งนี้เป็นการประยุกต์ใช้เทคนิคเหมืองข้อมูลมาสร้างระบบพยากรณ์ ผลการเรียนที่นำไปใช้กับนักเรียนในระดับมัธยมศึกษาตอนต้น ดังนั้นนักเรียนสามารถวางแผนการเรียนล่วงหน้าเพื่อให้ได้ผลการเรียนให้อยู่ในระดับที่ดีขึ้น และอาจารย์แนะแนวสามารถให้คำปรึกษาเกี่ยวกับวิชาเรียนของนักเรียน

4.1.2 การพัฒนาคลังข้อมูลผลการเรียนของนักเรียน ซึ่งสามารถนำข้อมูลผลการเรียนของนักเรียนนำเข้ามาจัดเก็บในคลังข้อมูล เพื่อใช้เป็นฐานข้อมูลในการวิเคราะห์หรือออกรายงานต่อไป ดังนั้นทางโรงเรียนจึงควรจัดหาซอฟต์แวร์ และเครื่องแม่ข่ายที่รองรับการทำงานของคลังข้อมูลเพื่อใช้ในการเก็บข้อมูลต่อไปในอนาคต

4.1.3 การพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ ซึ่งเป็นรายงานที่ผู้บริหารโรงเรียนสามารถดูภาพรวมของผลการเรียน หัวหน้ากลุ่มสาระการเรียนรู้สามารถดูผลการเรียนในกลุ่มสาระการเรียนรู้ หรือเปรียบเทียบกับกุล่มสาระการเรียนรู้อื่น เพื่อจัดการรายวิชาพื้นฐานและวิชาเพิ่มเติมให้มีผลการเรียนที่ดียิ่งขึ้น และอาจารย์ประจำชั้นสามารถดูข้อมูลผลการเรียนของนักเรียนเป็นรายบุคคล ดังนั้นควรนำไปใช้ในระบบสารสนเทศของโรงเรียน เพื่อให้ผู้ใช้งานในกลุ่มต่างๆ สามารถเรียกดูข้อมูลได้แบบออนไลน์ เพื่อใช้ในการตัดสินใจหรือเป็นข้อมูลในการบริหารโรงเรียนอย่างมีประสิทธิภาพ

4.2 ข้อเสนอแนะในการวิจัยครั้งต่อไป

4.2.1 การวิจัยครั้งต่อไปควรนำการพัฒนาคลังข้อมูลไปใช้ในการเก็บข้อมูลพื้นฐานของนักเรียน พฤติกรรมและกิจกรรมของนักเรียน เพื่อใช้ในการวิเคราะห์ข้อมูล

4.2.2 การวิจัยครั้งต่อไปควรนำเทคนิคเหมืองข้อมูลไปใช้ในการพยากรณ์ ในการเลือกคณะ หรือการศึกษาต่อในมหาวิทยาลัยของนักเรียนชั้นมัธยมศึกษาปีที่ 6

4.2.3 การวิจัยครั้งต่อไปควรศึกษาคุณลักษณะที่เกี่ยวข้องกับพฤติกรรมการเรียนของนักเรียน เพื่อนำมาใช้ในการวิเคราะห์หาพฤติกรรมที่เกี่ยวข้องกับผลการเรียนของนักเรียน

ภาคผนวก

ภาคผนวก ก

แบบประเมินความพึงพอใจระบบคลังข้อมูลผลการเรียน

แบบประเมินความพึงพอใจ
ระบบคลังข้อมูลผลการเรียน

คำชี้แจง

1. แบบสอบถามชุดนี้สำหรับผู้ใช้งาน เป็นแบบประเมินความพึงพอใจการใช้งานคลังข้อมูลผลการเรียนของนักเรียน เพื่อนำไปใช้ในการปรับปรุงต่อไป

2. การแสดงความคิดเห็นของผู้ใช้งานที่มีต่อคลังข้อมูลผลการเรียนของนักเรียนที่ได้พัฒนาขึ้นจะประกอบด้วยส่วนของรายการประเมินที่อยู่ด้านซ้ายมือ และมาตราส่วนประมาณค่าอยู่ด้านขวามือ จำนวน 5 ช่อง โปรดทำเครื่องหมาย √ ลงในช่องทางด้านขวามือที่ตรงกับความคิดเห็นของท่าน โดยกำหนดค่าความหมายดังนี้

5 หมายถึง ระดับความพึงพอใจมากที่สุด
4 หมายถึง ระดับความพึงพอใจมาก
3 หมายถึง ระดับความพึงพอใจปานกลาง
2 หมายถึง ระดับความพึงพอใจน้อย
1 หมายถึง ระดับความพึงพอใจน้อยที่สุด

รายละเอียดแบบสอบถาม

แบบสอบถามแบ่งเป็น 2 ตอน ดังนี้

ตอนที่ 1 ความพึงพอใจการใช้งาน

ตอนที่ 2 ความคิดเห็นและข้อเสนอแนะ

ตอนที่ 1 ความพึงพอใจการใช้งาน

กลุ่มผู้ใช้งาน [ ] ผู้บริหาร [ ] หัวหน้ากลุ่มสาระการเรียนรู้
[ ] อาจารย์ประจำชั้น

รายการประเมิน	ระดับการประเมิน				
	มากที่สุด (5)	มาก (4)	ปานกลาง (3)	น้อย (2)	น้อยที่สุด (1)
1. มีความสะดวกและง่ายต่อการใช้งาน					
2. การออกแบบรายงานเข้าใจได้ง่าย					
3. ความถูกต้องของข้อมูล					
4. ช่วยลดขั้นตอนในการทำงาน					
5. ข้อมูลตรงกับความต้องการผู้ใช้งาน					

ตอนที่ 2 ความคิดเห็นและข้อเสนอแนะ

.....

.....

.....

.....

.....

.....

*** ขอขอบคุณสำหรับการตอบแบบประเมิน ***

ภาคผนวก ข

แบบประเมินความเชื่อมั่นระบบพยากรณ์ผลการเรียน

แบบประเมินความเชื่อมั่น
ระบบพยากรณ์ผลการเรียน

คำชี้แจง

1. แบบประเมินชุดนี้สำหรับผู้ใช้งานระบบพยากรณ์ เป็นแบบประเมินความเชื่อมั่นในระบบพยากรณ์ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 4

2. การแสดงความคิดเห็นของผู้ใช้งานที่มีต่อระบบพยากรณ์ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 4 ที่ได้พัฒนาขึ้นจะประกอบด้วยส่วนของรายการประเมินที่อยู่ด้านซ้ายมือ และมาตราส่วนประมาณค่าอยู่ด้านขวามือ จำนวน 5 ช่อง โปรดทำเครื่องหมาย √ ลงในช่องทางด้านขวามือที่ตรงกับความคิดเห็นของท่าน โดยกำหนดค่าความหมายดังนี้

5 หมายถึง ระดับความเชื่อมั่นมากที่สุด
4 หมายถึง ระดับความเชื่อมั่นมาก
3 หมายถึง ระดับความเชื่อมั่นปานกลาง
2 หมายถึง ระดับความเชื่อมั่นน้อย
1 หมายถึง ระดับความเชื่อมั่นน้อยที่สุด

รายละเอียดแบบสอบถาม

แบบสอบถามแบ่งเป็น 2 ตอน ดังนี้

ตอนที่ 1 ความเชื่อมั่นของระบบพยากรณ์

ตอนที่ 2 ความคิดเห็นและข้อเสนอแนะ

ตอนที่ 1 ความเชื่อมั่นของระบบพยากรณ์

กลุ่มผู้ใช้งาน [ ] อาจารย์แนะแนว [ ] นักเรียน

รายการประเมิน	ระดับการประเมิน				
	มากที่สุด (5)	มาก (4)	ปานกลาง (3)	น้อย (2)	น้อยที่สุด (1)
1. ความสะดวกและง่ายต่อการใช้งาน					
2. ความน่าเชื่อถือของผลการพยากรณ์					
3. ผลการพยากรณ์ตรงกับความต้องการ					
4. ช่วยประกอบการตัดสินใจ					

ตอนที่ 2 ความคิดเห็นและข้อเสนอแนะ

.....

.....

.....

.....

.....

.....

.....

*** ขอขอบคุณสำหรับการตอบแบบสอบถาม ***

ภาคผนวก ค

คู่มือการติดตั้งโปรแกรมระบบ

การติดตั้งโปรแกรมระบบ

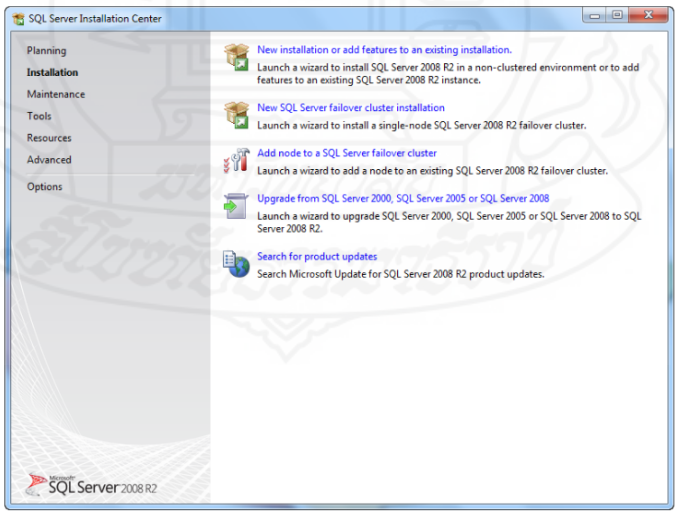
1. การติดตั้ง Microsoft SQL Server 2008 R2

เครื่องมือในการพัฒนาคลังข้อมูล คือ โปรแกรม Microsoft SQL Server 2008 R2 ซึ่งขั้นตอนในการติดตั้งโปแกรม มีดังนี้

1. ก่อนการติดตั้งโปรแกรมต้องทำการตรวจสอบเครื่องที่จะติดตั้งก่อนว่าที่เครื่องมีการติดตั้งโปรแกรมดอตเน็ตเฟรมเวิร์ก 3.5 (.Net Framework 3.5) ไว้หรือไม่ หากยังไม่ได้ติดตั้งให้ทำการติดตั้ง โดยดาวน์โหลดไฟล์จากเว็บไซต์ของไมโครซอฟท์แล้วทำการติดตั้ง

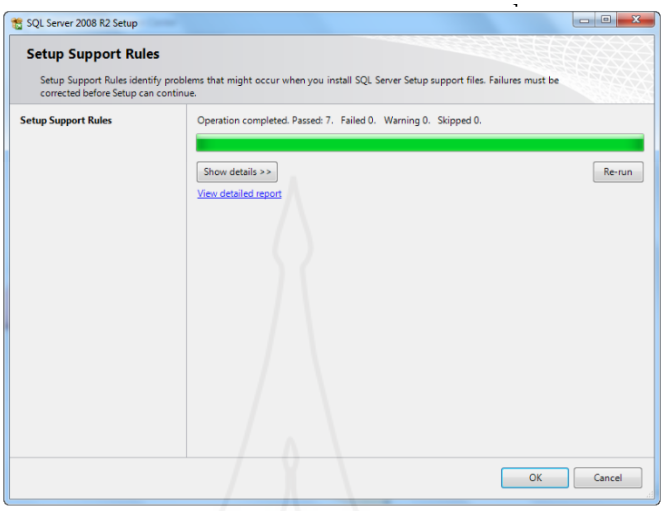
ภาพที่ 1.1 ไฟล์ติดตั้งโปรแกรมดอตเน็ตเฟรมเวิร์ก 3.5

2. เริ่มต้นติดตั้งโปรแกรม โดยใส่แผ่นติดตั้งโปรแกรม Microsoft SQL Server 2008 R2 โปรแกรมจะแสดงหน้าเมนูหลักในการติดตั้งขึ้นมาอัตโนมัติ เลือกหัวข้อ “New installation or and feature to an existing installation”



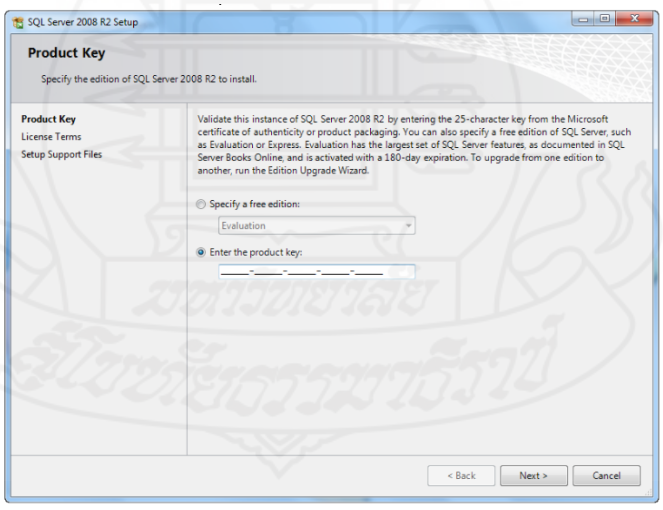
ภาพที่ 1.2 แสดงหน้าเมนูหลักในการติดตั้งโปรแกรม Microsoft SQL Server 2008 R2

3. โปรแกรมจะทำการตรวจสอบระบบที่ต้องการ เมื่อเสร็จเรียบร้อยแล้วจะให้กดปุ่ม “OK”



ภาพที่ 1.3 การตรวจสอบระบบก่อนการติดตั้ง

4. ทำการป้อน Product Key จากนั้นคลิกปุ่ม “Next”

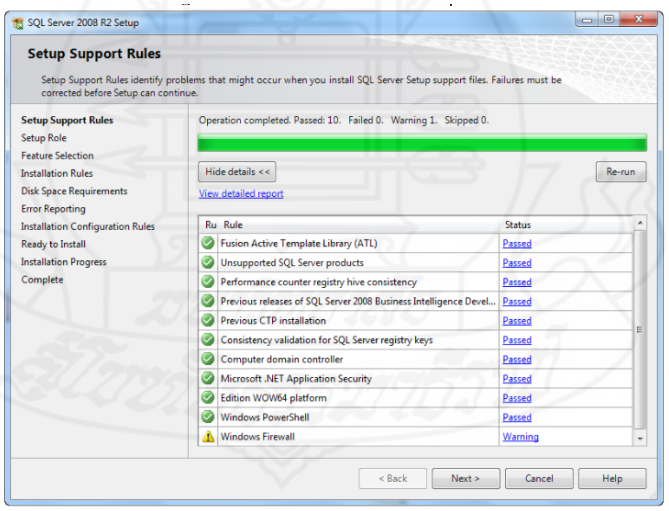


ภาพที่ 1.4 หน้าจอการป้อน Product Key

5. จากนั้นเป็นข้อตกลงของผลิตภัณฑ์ เลือก “I accept the license terms” แล้วคลิกปุ่ม “Next”

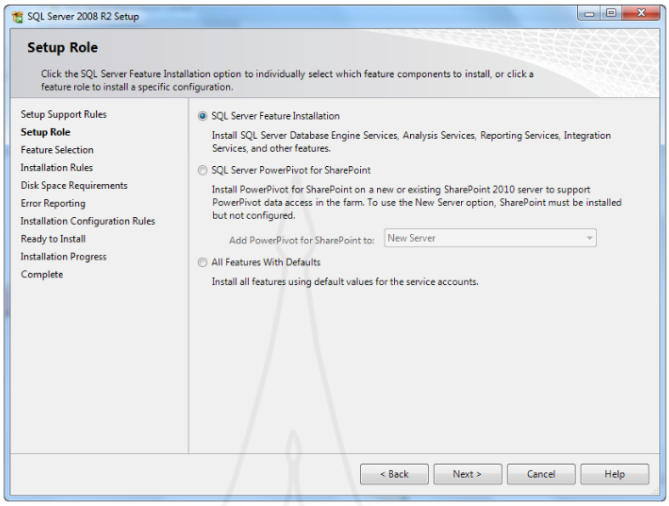
ภาพที่ 1.5 ข้อตกลงของการใช้ผลิตภัณฑ์

6. โปรแกรมจะทำการตรวจสอบกฎการติดตั้ง เมื่อตรวจสอบเสร็จแล้วให้คลิกที่ปุ่ม “Next”



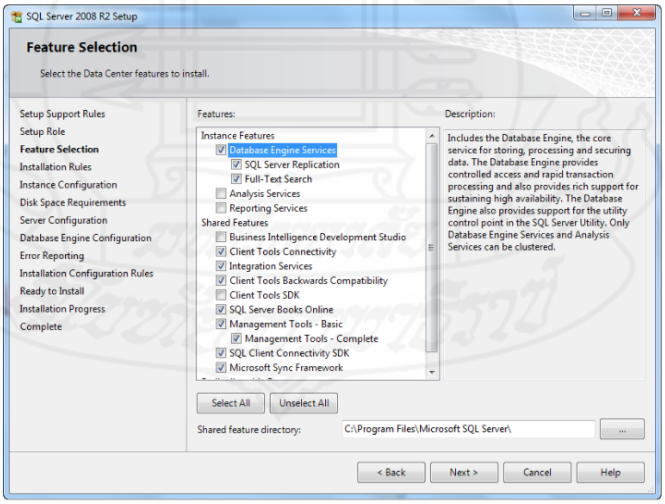
ภาพที่ 1.6 หน้าจอการตรวจสอบกฎการติดตั้ง

7. เลือก “SQL Server Feature Installation” แล้วคลิกปุ่ม “Next”



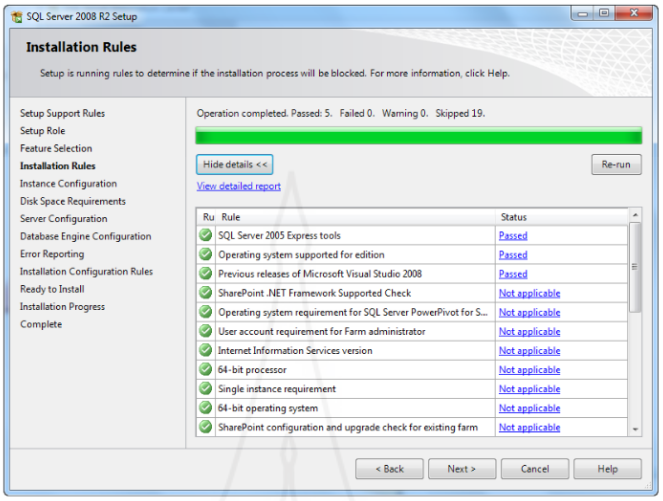
ภาพที่ 1.7 หน้าจอการเลือกรูปแบบการติดตั้งโปรแกรม

8. เลือก “Feature” ที่ต้องการ และเลือก “Directory” ที่ต้องการจะติดตั้งโปรแกรม แล้วคลิกปุ่ม “Next”



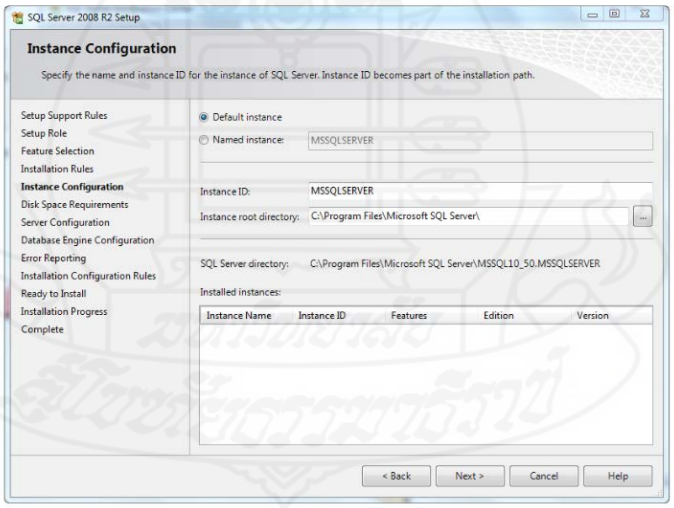
ภาพที่ 1.8 หน้าจอการเลือก Feature และDirectory ที่ต้องการติดตั้ง

9. โปรแกรมจะทำการตรวจสอบข้อมูลที่จะติดตั้งอีกครั้ง เมื่อเสร็จแล้วให้คลิกปุ่ม “Next”



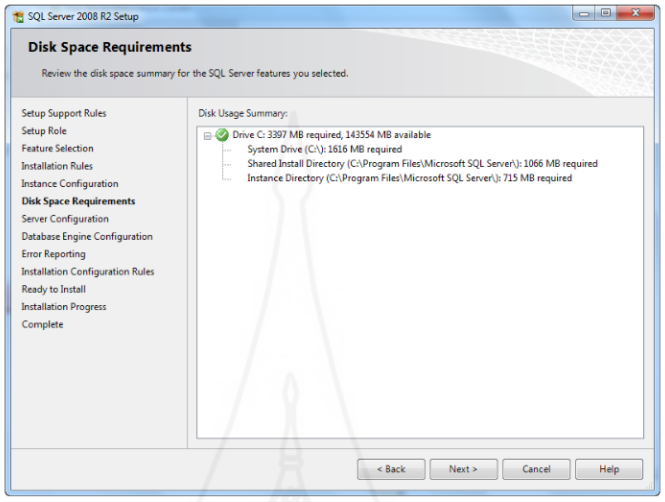
ภาพที่ 1.9 หน้าจอการตรวจสอบข้อมูลที่จะติดตั้งอีกครั้ง

10. เลือก Default Instance แล้วคลิกปุ่ม “Next”



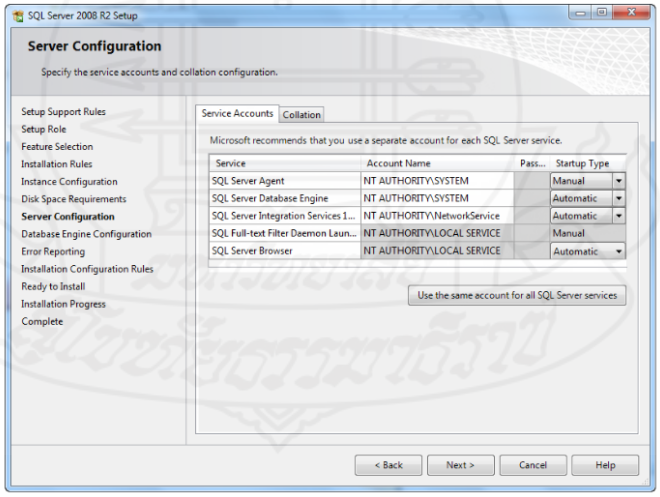
ภาพที่ 1.10 หน้าจอกำหนดค่า Instance

11. โปรแกรมติดตั้งแจ้งจำนวนพื้นที่ใน Hard disk ที่ต้องใช้และจำนวนคงเหลือ จากนั้นคลิกปุ่ม “Next”



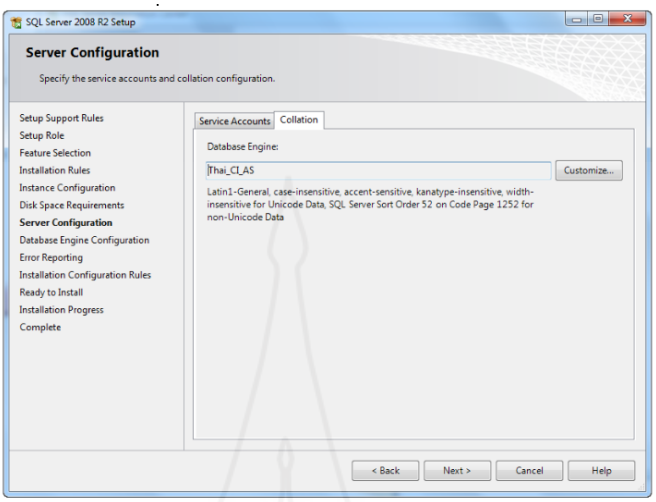
ภาพที่ 1.11 หน้าจอแสดงพื้นที่ใน Hard disk

12. การปรับตั้งค่า Service Accounts จากนั้นคลิกปุ่ม “Next”



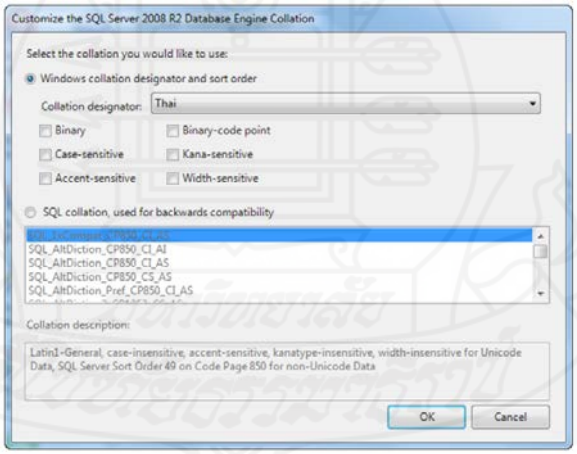
ภาพที่ 1.12 หน้าจอปรับตั้งค่า Service Accounts

13. เลือก Tab Collation แล้วคลิกที่ปุ่ม “Customize”



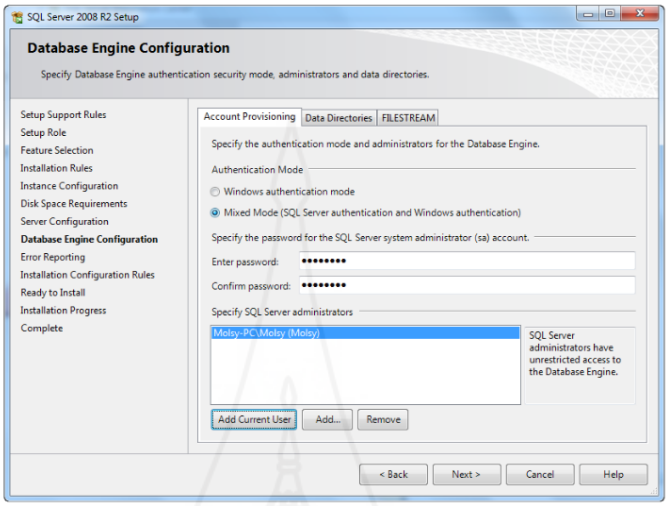
ภาพที่ 1.13 หน้าจอ Tab Collation

14. การตั้งค่า Collation เลือก “Thai” คลิกปุ่ม “OK”



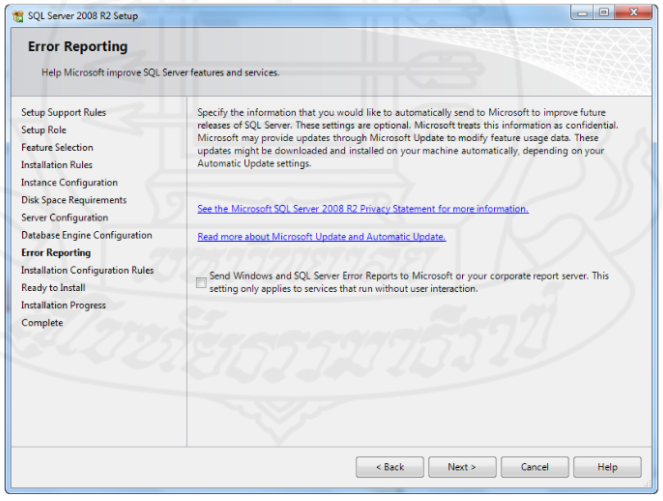
ภาพที่ 1.14 การตั้งค่า Collation

15. เลือก “Authentication Mode เป็น Mixed Mode แล้วกรอก “Password” ของ SA ในช่อง Enter password และ Confirm password คลิกปุ่ม “Add Current User” แล้วคลิกปุ่ม “Next”



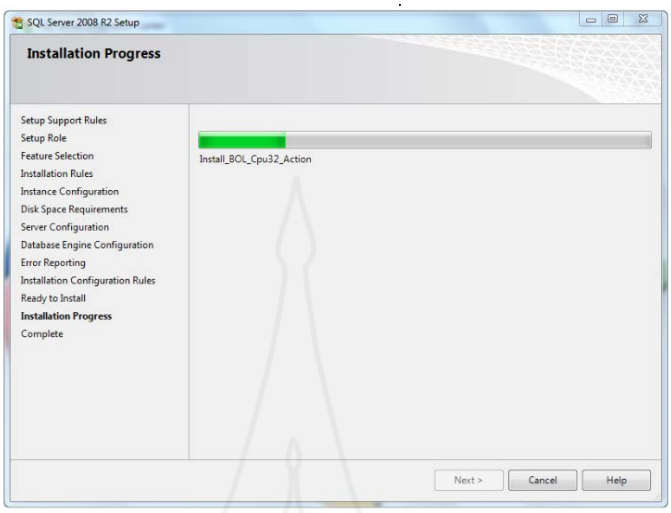
ภาพที่ 1.15 หน้าจอการตั้งค่ารหัสผ่านบัญชีผู้ใช้งาน

16. คลิกปุ่ม “Next” เพื่อติดตั้งโปรแกรม



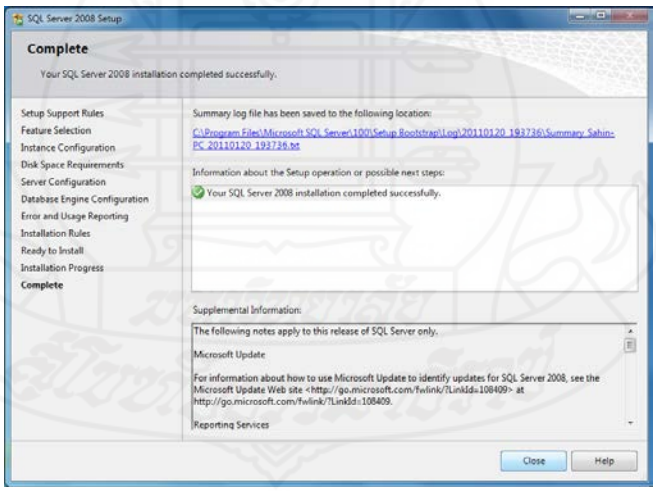
ภาพที่ 1.16 หน้าจอก่อนติดตั้งโปรแกรม

17. โปรแกรมจะทำการติดตั้ง



ภาพที่ 1.17 หน้าจอกำลังติดตั้งโปรแกรม

18. เมื่อเสร็จสิ้น ให้คลิกปุ่ม “Close”



ภาพที่ 1.18 หน้าจอการติดตั้งโปรแกรมเสร็จสิ้น

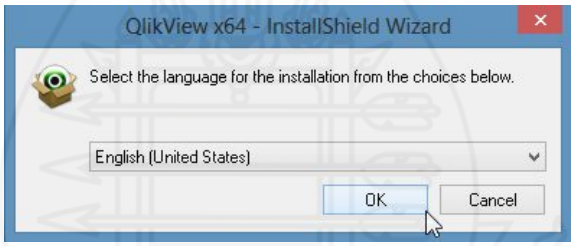
2. การติดตั้งโปรแกรม QlikView

การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ใช้โปรแกรม QlikView ในการแสดงรายงานของคลังข้อมูล ซึ่งขั้นตอนในการติดตั้งโปรแกรม QlikView มีดังนี้

1. การติดตั้งโปรแกรม โดยดาวน์โหลดไฟล์จากเว็บไซต์ของ QlikView แล้วทำการติดตั้ง

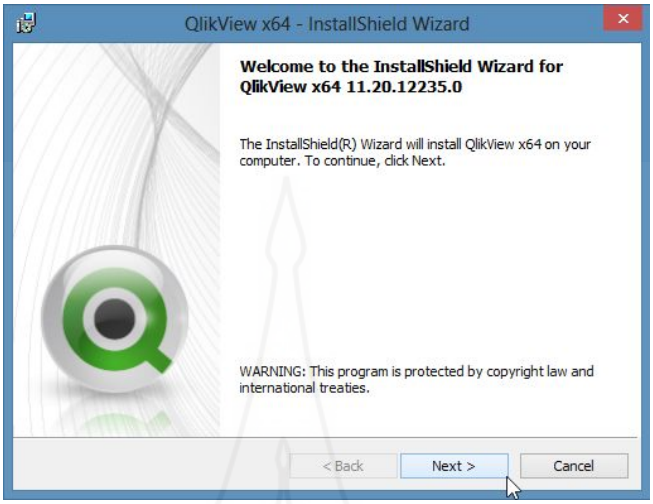
ภาพที่ 2.1 ไฟล์ติดตั้ง QlikView

2. เลือกภาษาในการติดตั้ง จากนั้นคลิกที่ปุ่ม “OK”



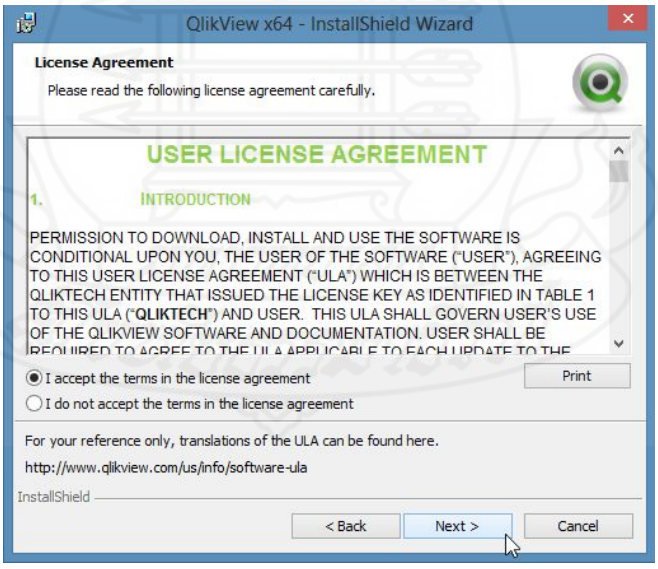
ภาพที่ 2.2 เลือกภาษาในการติดตั้ง

3. เข้าสู่หน้าการติดตั้งโปรแกรม QlikView คลิกปุ่ม “Next”



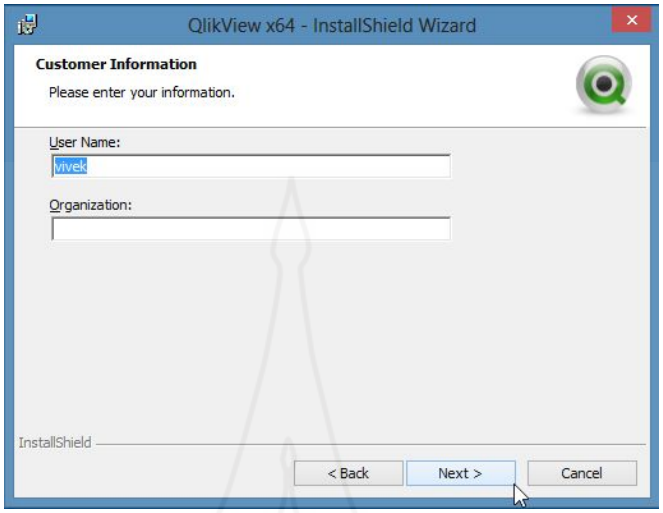
ภาพที่ 2.3 หน้าจอการติดตั้งโปรแกรม QlikView

4. รายละเอียดข้อตกลงการใช้งานโปรแกรม เลือก “I accept the terms in the license agreement” จากนั้นคลิก “Next”



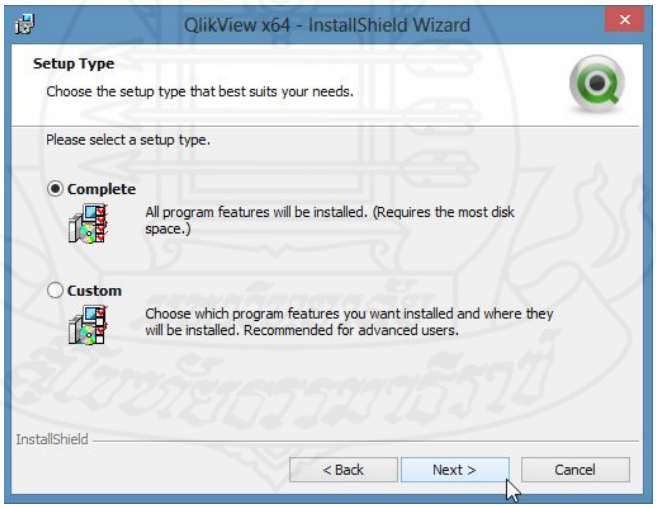
ภาพที่ 2.4 หน้าจอข้อตกลงเกี่ยวกับโปรแกรม

5. กรอกข้อมูลชื่อผู้ใช้และองค์กร เสร็จแล้วคลิก “Next”



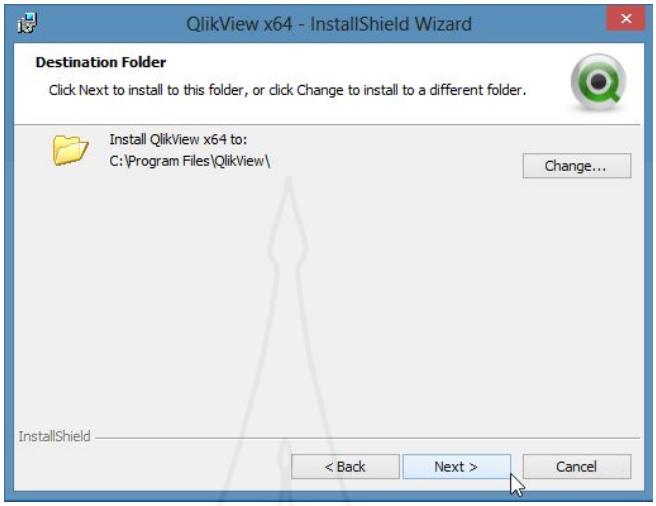
ภาพที่ 2.5 หน้าจอใส่ชื่อผู้ใช้งานและองค์กร

6. เลือก “Complete” จากนั้นคลิก “Next”



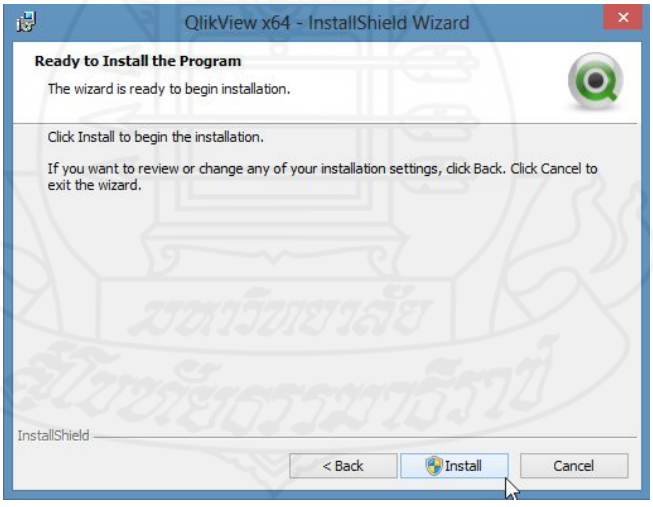
ภาพที่ 2.6 เลือกรูปแบบของการติดตั้ง

7. กำหนดที่อยู่ของ “Folder” ในการติดตั้งโปรแกรม จากนั้นคลิก “Next”



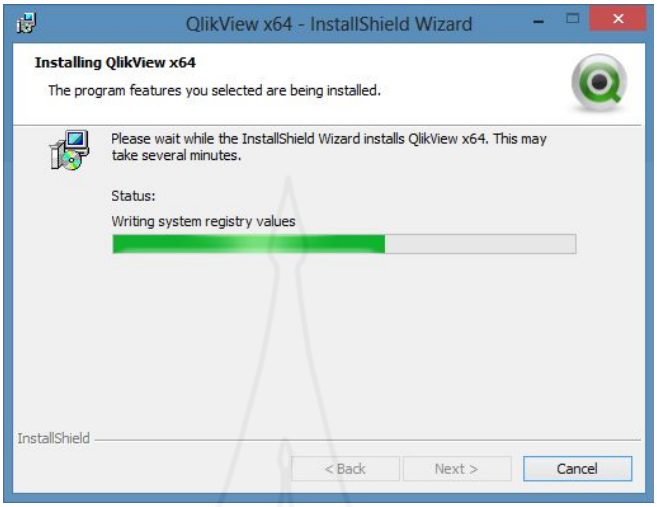
ภาพที่ 2.7 กำหนดที่อยู่ของโปรแกรม

8. โปรแกรมติดตั้งพร้อมทำการติดตั้ง คลิก “Install”



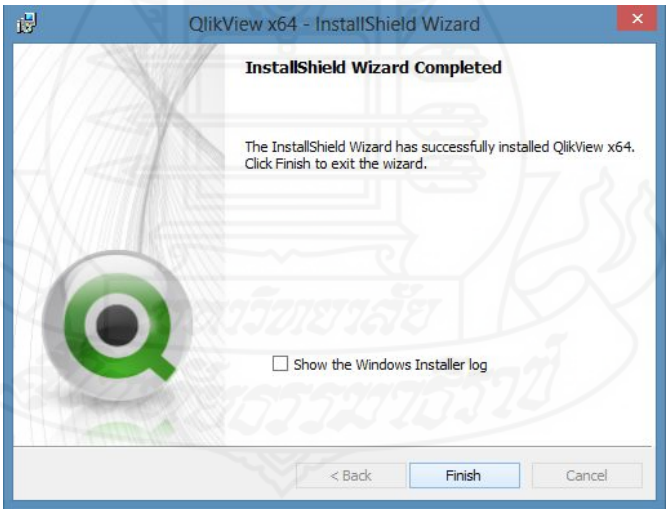
ภาพที่ 2.8 โปรแกรมพร้อมติดตั้ง

9. กำลังติดตั้งโปรแกรม



ภาพที่ 2.9 กำลังติดตั้งโปรแกรม

10. เมื่อติดตั้งเสร็จสมบูรณ์ คลิก “Finish”



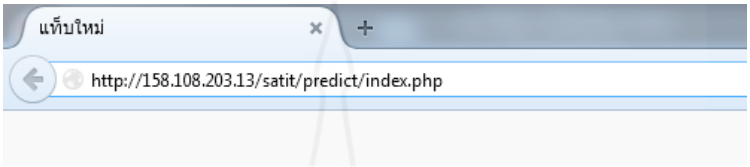
ภาพที่ 2.10 การติดตั้งโปรแกรมเสร็จสิ้น

ภาคผนวก ง

คู่มือการใช้งานระบบพยากรณ์ผลการเรียนเฉลี่ย

คู่มือการใช้งานระบบพยากรณ์ผลการเรียนเฉลี่ย

1. เปิดโปรแกรมเบราว์เซอร์พิมพ์ที่อยู่ของระบบพยากรณ์ผลการเรียนเฉลี่ย คือ http://158.108.203.13/satit/predict

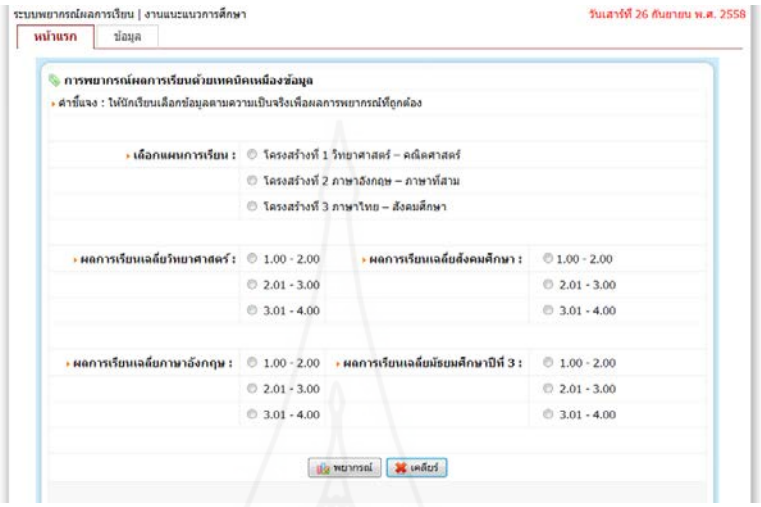


ภาพที่ 3.1 พิมพ์ที่อยู่ของระบบพยากรณ์ผลการเรียนเฉลี่ย

2. เข้าสู่หน้าเว็บของระบบพยากรณ์ผลการเรียนเฉลี่ย จากนั้นคลิกที่ปุ่ม “Enter” เพื่อใช้งานระบบพยากรณ์ผลการเรียนเฉลี่ย

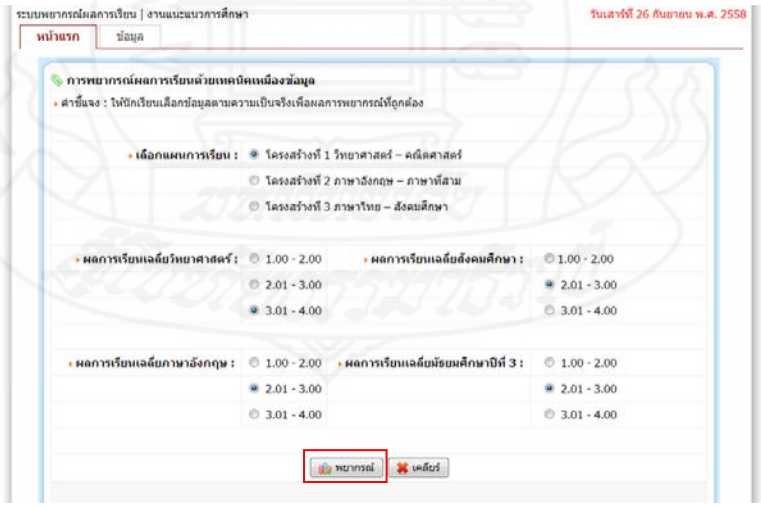
ภาพที่ 3.2 หน้าแรกของระบบพยากรณ์ผลการเรียนเฉลี่ย

3. เข้าสู่หน้าของการเลือกข้อมูลเพื่อใช้ในการพยากรณ์ผลการเรียนเฉลี่ยตามแผนการเรียนในระดับชั้นมัธยมศึกษาปีที่ 4



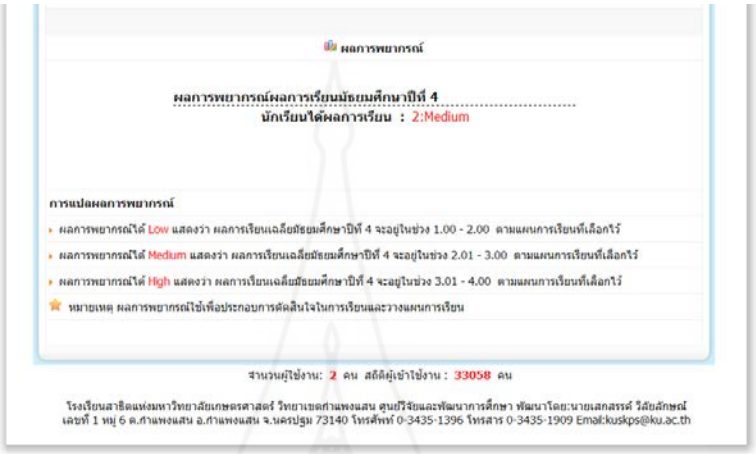
ภาพที่ 3.3 แสดงหน้าการเลือกข้อมูลเพื่อพยากรณ์

4. ทำการเลือกข้อมูลในการพยากรณ์ผลการเรียนเฉลี่ยและเลือกแผนการเรียนที่ต้องการในระดับชั้นมัธยมศึกษาปีที่ 4 จากนั้นคลิกที่ปุ่ม “พยากรณ์” เพื่อดูผลการพยากรณ์



ภาพที่ 3.4 เลือกข้อมูลผลการเรียนและแผนการเรียนเพื่อพยากรณ์

5. ผลพยากรณ์จะปรากฏขึ้นมา จากนั้นไปอ่านการแปลผลการพยากรณ์ที่ได้จากด้านล่าง ตัวอย่าง เช่น ผลพยากรณ์ได้ “Medium” การแปลผลการพยากรณ์ที่ได้คือ “นักเรียนจะมีผลการเรียนเฉลี่ยมัธยมศึกษาปีที่ 4 จะอยู่ในช่วง 2.01 – 3.00 ตามแผนการเรียนที่เลือกไว้”



ภาพที่ 3.5 ผลการพยากรณ์

บรรณานุกรม

บรรณานุกรม

กชกร ณ นครพนม. (2555). “การทำเหมืองข้อมูลด้วยการจำแนกประเภทและการพยากรณ์” ใน *ประมวลสาระชุดวิชาคลังข้อมูล เหมืองข้อมูล และธุรกิจอัจฉริยะ (Data warehouse data mining and business intelligence)* หน่วยที่ 9 หน้า 9-13 นนทบุรี มหาวิทยาลัยสุโขทัยธรรมาธิราช บัณฑิตศึกษา สาขาวิชาวิทยาศาสตร์และเทคโนโลยี.

กิตติพงษ์ กลมกล่อม. (2552).การออกแบบและพัฒนาคลังข้อมูล (Data warehouse) กรุงเทพมหานคร : เคทีพีคอมพ์แอนด์ คอนซัลท์.

กุลเทพ จิรลักข์ และเอกสิทธิ์ วิวัฒนาประสิทธิ์. (2555). “กรณีศึกษาการทำคลังข้อมูล.” ใน *ประมวลสาระชุดวิชาคลังข้อมูล เหมืองข้อมูล และธุรกิจอัจฉริยะ (Data warehouse data mining and business intelligence)* หน่วยที่ 7 หน้า 299-383 นนทบุรี มหาวิทยาลัยสุโขทัยธรรมาธิราช บัณฑิตศึกษา สาขาวิชาวิทยาศาสตร์และเทคโนโลยี.

ชินพัฒน์ แก้วชินพร. (2556). K-fold cross validation. สืบค้นเมื่อ 18 กรกฎาคม 2558, จาก http://scriptslines.com/blog/k-fold-cross-validation.

ฐาปนี สีหาโภชน์. (2556). “การรู้จำอารมณ์จากเสียงพูดโดยใช้ซัพพอร์ตเวคเตอร์แมชชีน” (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยขอนแก่น, ขอนแก่น.

ทรงศิริ แต้สมบัติ. (2549). เทคนิคการพยากรณ์เชิงปริมาณ. มหาวิทยาลัยเกษตรศาสตร์ กรุงเทพฯ.

ทิพย์หทัย ทองธรรมชาติ. (2554). “การพยากรณ์ผลสัมฤทธิ์ทางการเรียนด้วยเทคนิคเหมืองข้อมูล” (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยนเรศวร, พิษณุโลก.

พรเทพ คงไชย และรัชฎา คงคะจันทร์. (2554). “การศึกษาเปรียบเทียบในการคัดเลือกคุณลักษณะที่เหมาะสมสำหรับการทำเหมืองข้อมูลเพื่อพยากรณ์โอกาสการสำเร็จการศึกษาของนักศึกษา” การประชุมวิชาการข่ายงานวิศวกรรมอุตสาหกรรม ประจำปี 2554 (20-21 ตุลาคม 2554).

พรรณิภา บุตรเอก และสุรเดช บุญลือ. (2557). “การพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน” (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยนอร์ทกรุงเทพ, กรุงเทพ.

วงกต ศรีอุไร. (2557). “การจำแนกผู้ป่วยโรคอ้วนลงพุงโดยใช้วิธีการคัดเลือกคุณลักษณะและโครงข่ายประสาทเทียม” *วารสารวิทยาศาสตร์มหาวิทยาลัยศรีนครินทรวิโรฒ*. ฉบับที่ 1 (มิถุนายน): 91-102.

วรารัตน์ รุ่งวรวุฒิ. (2554). เอกสารประกอบการบรรยาย วิชาการทำเหมืองข้อมูล (Data Mining) ขอนแก่น:มหาวิทยาลัยขอนแก่น.

วิภา เจริญภัณฑารักษ์. (2555) “หลักการพื้นฐานของการทำเหมืองข้อมูล” *ใน ประมวลสาระชุดวิชาคลังข้อมูล เหมืองข้อมูล และธุรกิจอัจฉริยะ (Data warehouse data mining and business intelligence)* หน่วยที่ 8 หน้า 1-62 นนทบุรี มหาวิทยาลัยสุโขทัยธรรมาธิราช บัณฑิตศึกษา สาขาวิชาวิทยาศาสตร์และเทคโนโลยี.

ศุภกร สระบัว. (2557). “การตรวจจับมัลแวร์ประเภทม้าโทรจันอย่างรวดเร็วโดยเทคนิคเหมืองข้อมูล” (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ.

สุประพล เลาวพงศ์. (2553). “การพัฒนาคลังข้อมูลด้านนักศึกษาของมหาวิทยาลัยพายัพ” (การค้นคว้าอิสระปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยเชียงใหม่, เชียงใหม่.

สุพัฒน์กุล ภัคโชค. (2555). “ตัวแบบแผนการเรียนระดับมัธยมศึกษาตอนปลาย โดยการพิจารณาผลการเรียนรายวิชาหลัก ด้วยเทคนิคเหมืองข้อมูล” (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยธุรกิจบัณฑิต, กรุงเทพฯ.

สุวรรณี อัศวกุลชัย. (2555) “หลักการพื้นฐานของคลังข้อมูล” ใน *ประมวลสาระชุดวิชาคลังข้อมูล เหมืองข้อมูล และธุรกิจอัจฉริยะ (Data warehouse data mining and business intelligence)* หน่วยที่ 1 หน้า 1-45 นนทบุรี มหาวิทยาลัยสุโขทัยธรรมาธิราช บัณฑิตศึกษา สาขาวิชาวิทยาศาสตร์.

สัมพันธ์ ชัยภูมิ. (2557). “การวิเคราะห์ประสิทธิภาพและเปรียบเทียบการจำแนกข้อมูลด้วยเทคนิค C4.5 , CART และ Neural Network เพื่อพยากรณ์สาขาวิชาที่มีความเหมาะสมกับผู้สมัครคัดเลือกเข้าศึกษาต่อระดับปริญญาตรี” (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ.

อาทิตยาพร โรจรัตน์. (2556). “การทำนายผลการศึกษาเพื่อการวางแผนการลงทะเบียนของนักศึกษา โดยใช้การทำเหมืองข้อมูล” (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยขอนแก่น, ขอนแก่น.

อาริยพร สุดใจ. (2557). “การเปรียเทียบประสิทธิภาพการคัดเลือกคุณลักษณะโดยใช้วิธีปกติ Information gain และวิธี CfsSubseteval” (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต ไม่ได้ตีพิมพ์). มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ.

เอกสิทธิ์ พัชรวงศ์ศักดา. (2557) *การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไมน์นิง เบื้องต้น (An Introduction to Data Mining Techniques)* กรุงเทพมหานคร เอเชีย ดิจิตอลการพิมพ์.

Ballard, C. and others. (1998). Data Modelling Techniques for data warehouse, California: IBM Corporation, International Technical Support Organization.

Berry, M. and Linoff, G. (2004). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management.* Indianapolis: Wiley.

Best, John W. (1963). Research in Education. New Jersey: Prentice-Hall.

C.Lakshmi Devasena. (2013). Classification of multivariate data sets without missing values using memory based Classifiers-An Effectiveness evaluation. International Journal of Artificial Intelligence & Applications. VOL. 4, No. 1, (2013): 129-142.

Changala et al. (2012). Classification by Decision Tree Induction Algorithm to Learn Decision Trees from the class-Labeled Training Tuples. International Journal of Advanced Research in Computer Sciences and Software Engineering. VOL. 2, Issue 4, (2012): 428-434.

Han, Jiawei and Kamber, Micheline. (2006). *Data Mining: Concepts and Techniques* (2nd ed). New York: Morgan Kaufmann.

http://www.ijser.org/paper/Over-viewing-issues-of-data-mining-with-highlights-of-data-warehousing.html Retrieved June 21, 2015.

http://www.learnbymarketing.com/methods/k-means-clustering Retrieved June 21, 2015.

http://www.mdpi.com/1424-8220/10/9/8363/htm Retrieved June 21, 2015.

Hye Rin Kim. (2009). Prediction of Forest Fires using Data Mining Methods (Master's Project). Retrieved from http://www.stats.uwo.ca/faculty/aim/2011/FireProject/R/project.pdf.

Jing Bi, Kun Zhang, and Xiaojing Cheng, (2009). “Intrusion Detection Base on RBF Neural Network” Internation Symposium on Information Engineering and Electronic Commerce, 362-365.

Kurniawan and Halim. (2013). Use Data Warehouse and Data Mining to Predict Student Academic Performance in School: A Case Study. IEEE International Conference on Teaching 2013, Kuta, Indonesia.

Makridakis, S., and Wheelwright, and R.J. Hyndman. (1998). Forecasting: Methods and Applications. New York. Jonh Wiley & Sons.

Quadri, MN, and NV Kalyankar. (2010). Drop out Feature of Student Data for Academic. Performance Using Decision Tree Techniques. Global Journal of Computer Science and Technology 10. no. 2 (2010): 2-5.

Rai, Saini and Jain (2014). Decision Tree Algorithm Implementation Using Education Data. Internation Journal of Computer-Aided technologied. Vol.1, No.1 (2014): 31-41.

Roiger and Geatz. (2003). Data mining A tutorial-based primer. Pearson Internation Edition, USA: Addison-Wesley.

Umair Shafique and Haseeb Qaiser. (2014). A Comparative Study of Data Mining Process Model(KDD, CRISP-DM and SEMMA). International Journal of Innovation and Scientific Research. VOL. 2, No. 1, (2014): 217-222.

Venkatramaphanikumar S., (2015). A Novel Prediction Model for Academic Emotional Progression of Graduates. ARPN Journal of Engineering and Applied Sciences. VOL. 10, NO.6, (2015): 2561-2569.

WEKA Machine Learning Group at University of Waikato. Weka 3. form http://www.cs.waikato.ac.nz/ml/weka Retrieved June 21, 2015.

William W. Guo. (2010). Incorporating statiscal and neural network approaches for student course satisfaction analysis and prediction. Expert Systems with Applications. VOL. 37, (2010): 3358-3365.

ประวัติผู้วิจัย

ชื่อ	นายเสกสรรค์ วิลัยลักษณ์
วัน เดือน ปีเกิด	2 มิถุนายน 2520
สถานที่เกิด	อำเภอบางคนที จังหวัดสมุทรสงคราม
ประวัติการศึกษา	วิทยาศาสตรบัณฑิต สถาบันราชภัฏนครปฐม พ.ศ. 2544
สถานที่ทำงาน	โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรสาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา
ตำแหน่ง	อาจารย์