

การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน
โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา^{*}

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

เสกสรรค์ วิลัยลักษณ์ (Seksan Vilailuck)^{**}

วิภา เจริญภักดิ์ (Vipa Jaroenpuntaruk)^{***}

ดวงดาว วิชาดากุล (Duangdao Wichadaku)^{****}

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาลงข้อมูล และสร้างตัวแบบพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา โดยใช้ข้อมูล นักเรียนระดับมัธยมศึกษาระหว่างปีการศึกษา 2548 – 2556 เพื่อพัฒนาลงข้อมูลโดยใช้โครงสร้างแบบสโรว์ เฟลท สกีมานำเสนอรายงาน จากนั้นใช้ข้อมูลนักเรียนระดับมัธยมศึกษาปีที่ 4 ระหว่าง ปีการศึกษา 2553 – 2556 จำนวน 525 ระเบียบ ประกอบด้วย 16 คุณลักษณะ มาสร้างตัวแบบพยากรณ์ผลการเรียนโดยใช้ชุดข้อมูล 2 แบบ คือ ข้อมูลแบบไม่จัดกลุ่ม (Original Data) และข้อมูลแบบจัดกลุ่ม (Cluster Data) จากนั้นนำไปผ่าน กระบวนการคัดเลือกคุณลักษณะ (Feature Selection) ซึ่งใช้วิธี Correlation-based Feature Selection (CFS) และวิธี Information Gain (IG) แล้วใช้เทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียมแบบมัลติเลเยอร์

^{*} บทความนี้เป็นส่วนหนึ่งของวิทยานิพนธ์ เรื่อง การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา

^{**} นักศึกษาระดับปริญญาโทหลักสูตรวิทยาศาสตรมหาบัณฑิต แขนงวิชาเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยสุโขทัยธรรมาธิราช

M.Sc. Information and Communication Technology, Sukhothai Thamathirat Open University. E-mail : seksan.v@ku.th

^{***} อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก รองศาสตราจารย์ ดร. สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช

ADVISORS: Asst.Prof. Dr. at School of Science and Technology, Sukhothai Thamathirat Open University. E-mail : vipa.jar@stou.ac.th

^{****} อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ดร. ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย CO-ADVISORS: Dr. at Computer Engineering, Faculty of Engineering, Chulalongkorn University.

E-mail : duangdao.w@chula.ac.th

เพอร์เซ็ปตรอน (MLP) ซัพพอร์ตเวกเตอร์แมชชีน (SVM) และต้นไม้ตัดสินใจ (Decision Tree) มาสร้างตัวแบบพยากรณ์และเปรียบเทียบตัวแบบ ด้วยการทดสอบประสิทธิภาพแบบ 10-fold Cross Validation

ผลการวิจัยพบว่า คลังข้อมูลนักเรียนระดับมัธยมศึกษาที่ใช้งานโดย ผู้บริหาร หัวหน้ากลุ่มสาระการเรียนรู้ และอาจารย์ประจำชั้นมีความพึงพอใจการใช้งานคลังข้อมูลอยู่ในระดับดี และในการทำเหมืองข้อมูลพยากรณ์ พบว่า ชุดข้อมูลแบบไม่จัดกลุ่มนำมาคัดเลือกคุณลักษณะด้วยวิธี Correlation-based Feature Selection (CFS) ร่วมกับเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน ให้ ค่าความถูกต้องสูงที่สุดที่ร้อยละ 94.48 และมีค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดที่ 0.1880 เหมาะสมสำหรับการสร้างระบบพยากรณ์ผลการเรียนของนักเรียน

คำสำคัญ : การพยากรณ์ผลการเรียน การทำเหมืองข้อมูล การคัดเลือกคุณลักษณะ

Abstract

The purpose of this research were to develop the data warehouse from the structure of Snowflake Schema and to construct the “Prediction Model” of high school students achievement from the data during academic year 2005-2013 of Kasetsart University Laboratory school, Kamphaeng Saen Campus Educational Research and Development Center. The sample consisted of 525 records of Matayom Suksa 4 students’ achievement including 16 features of attributes to construct the tradition 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 like 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

บทนำ

การศึกษาเป็นกระบวนการพัฒนาคนในสังคมให้มีความรู้ความสามารถ ระบบการศึกษาถือว่าเป็นปัจจัยหนึ่งที่สำคัญต่อการเจริญเติบโตและการพัฒนาของประเทศ ประเทศไทยนี้มีระบบการศึกษาขั้นพื้นฐาน ซึ่งจัดไม่น้อยกว่าสิบสองปี ก่อนที่จะเข้าระดับอุดมศึกษาซึ่งแบ่งออกเป็น 3 ระดับคือ การศึกษาก่อนระดับประถมศึกษา การศึกษาระดับประถมศึกษา และการศึกษาระดับมัธยมศึกษาที่แบ่งเป็นสองระดับ คือการศึกษาระดับมัธยมศึกษาตอนต้นและระดับมัธยมศึกษาตอนปลาย (สำนักงานเลขาธิการสภาการศึกษา, 2553: ออนไลน์)

โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา สังกัดสำนักงานคณะกรรมการการอุดมศึกษา จัดการเรียนการสอนขั้นพื้นฐานตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ในระยะหลังพบว่านักเรียนระดับมัธยมศึกษาที่มีผลการเรียน ไม่ผ่านเกณฑ์ในการวัดและประเมินผลการศึกษา มีจำนวนมากขึ้น และแนวโน้มเพิ่มขึ้นในแต่ละปี เมื่อดูผลการศึกษาต่อของนักเรียนชั้นมัธยมศึกษาปีที่ 3 พบว่าไม่ได้ศึกษาต่อในระดับมัธยมศึกษาตอนปลายในโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา มีแนวโน้มเพิ่มขึ้น และนักเรียนบางคนสำเร็จการศึกษาช้ากว่าที่หลักสูตรกำหนดไว้ ในบางกรณีอาจจะต้องเรียนซ้ำชั้นในระดับการศึกษานั้นๆ ปัญหาของนักเรียนคือ เมื่อมีผลการเรียนที่ต่ำในระดับมัธยมศึกษาตอนต้นทำให้นักเรียนไม่สามารถเลือกแผนการเรียนในระดับมัธยมศึกษาตอนปลายตามความสนใจและถนัดของนักเรียน ต้องเลือกแผนการเรียนที่ไม่ถนัดและไม่เหมาะสมกับความสามารถของนักเรียน ทำให้เรียนแล้วไม่ประสบความสำเร็จ มีผลการเรียนที่ต่ำ สาเหตุทั้งหลายของปัญหาที่เกิดขึ้นอาจเกิดจากการที่นักเรียนไม่มีข้อมูลและประสบการณ์ เกี่ยวกับการวางแผนการเรียน ไม่ตั้งใจเรียน และไม่ทราบถึงแผนการเรียนที่เหมาะสมกับตัวนักเรียน ซึ่งทางโรงเรียนมีข้อมูลผลการเรียนของนักเรียนเหล่านี้ที่ถูกจัดเก็บไว้เป็นจำนวนมาก ข้อมูลเหล่านี้ไม่ได้ถูกนำมาใช้ประโยชน์ การสร้างคลังข้อมูลช่วยให้การสอบถามข้อมูลและสร้างรายงาน (Rifaie, Blas, and Muhsen, 2008) และการทำเหมืองข้อมูล (Data Mining) เป็นวิธีการหนึ่งในการวิเคราะห์ข้อมูลที่มีปริมาณมาก (Han, and Kamber, 2006) จากการทบทวนวรรณกรรมและงานวิจัยพบว่ามีการวิจัยได้นำเทคนิคเหมืองข้อมูลมาใช้ในการพยากรณ์ ดังเช่นงานวิจัยของ สุพัฒน์กุล ภัคโชค (2555: 105) ใช้เทคนิคเหมืองข้อมูลสร้างตัวแบบเลือกแผนการเรียนระดับมัธยมศึกษาตอนปลาย โรงเรียนสตรีศรีสุริโยทัย ปีการศึกษา 2555 จำนวน 850 คน ให้ค่าความถูกต้องในการแนะนำ แผนการเรียนร้อยละ 79.76 ต่อมา พรรณีญา บุตรเอก และสุรเดช บุญลือ (2557: 40) ได้ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (SVM) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ (BP-ANN) มาเปรียบเทียบเพื่อสร้างตัวแบบพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาพบว่า ตัวแบบพยากรณ์ที่ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน ให้ความแม่นยำสูงที่สุดที่ร้อยละ 89.13

ปัญหาที่สำคัญในการทำเหมืองข้อมูลอย่างหนึ่ง คือ การใช้ข้อมูลซึ่งประกอบด้วยคุณลักษณะ (Attribute) ที่มีจำนวนมากและคุณลักษณะเหล่านี้อาจมีบางส่วนที่ไม่เกี่ยวข้องกับการสร้างตัวแบบพยากรณ์ ทำให้ซ้ำซ้อน (วิภา เจริญภักดิ์, 2555: 19) ดังนั้นขั้นตอนในการคัดเลือกคุณลักษณะ (Feature Selection) จึงเป็นขั้นตอนสำคัญของการทำเหมืองข้อมูล อย่างที่ พรเทพ คงไชย และรัชฎา คงคะจันทร์ (2554: 2012) ได้ใช้

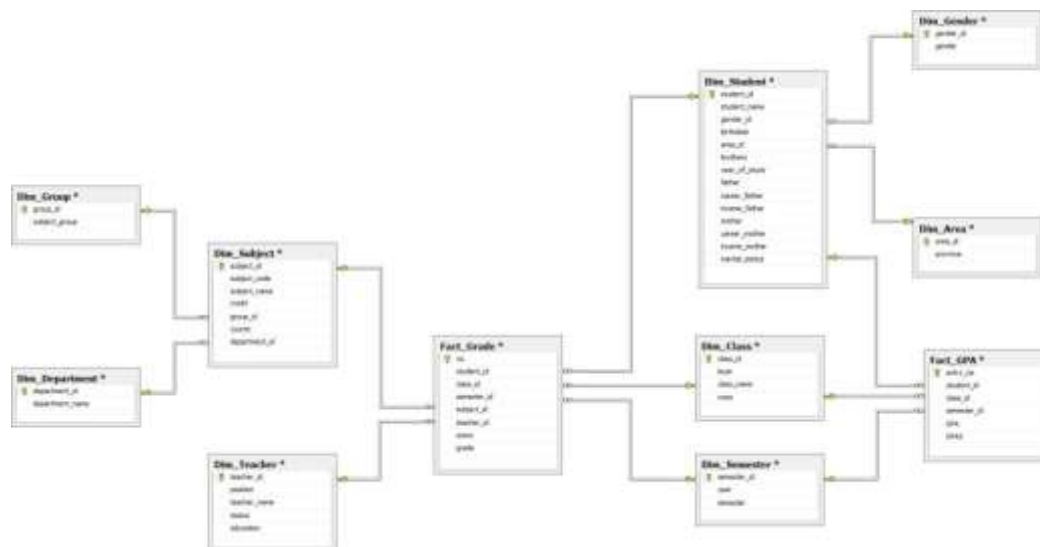
การเปรียบเทียบการคัดเลือกคุณลักษณะ แบบ Filter และแบบ Wrapper ร่วมกับเทคนิคเหมืองข้อมูลแบบ SVM , MLP และ Bayesnet พบว่า ค่าความถูกต้องในการจำแนกเท่ากับ 95.10% ร่วมกับเทคนิคเหมืองข้อมูลแบบ MLP ดังนั้นผู้วิจัยจึงมีแนวคิดที่จะใช้ข้อมูลผลการเรียนเหล่านี้ มาพัฒนาเป็นคลังข้อมูล (Data Warehouse) และนำมาวิเคราะห์ด้วยเทคนิคเหมืองข้อมูลในการทำนายผลการเรียนของนักเรียน โดยใช้ข้อมูลในอดีต ซึ่งใช้วิธีการคัดเลือกคุณลักษณะร่วมกับเทคนิคเหมืองข้อมูล แล้วนำมาเปรียบเทียบเพื่อให้ได้ตัวแบบพยากรณ์ที่ดีที่สุด ในการพัฒนาระบบพยากรณ์ที่มีประสิทธิภาพในการพยากรณ์ผลการเรียนของนักเรียน เพื่อใช้ประกอบการตัดสินใจเลือกแผนการเรียนในระดับมัธยมศึกษาตอนปลายและวางแผนการเรียนของนักเรียน

วิธีการวิจัย

วิธีในการวิจัยครั้งนี้แบ่งออกเป็น 4 ขั้นตอน ประกอบด้วย 1) การพัฒนาคลังข้อมูล 2) การพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ 3) การทำเหมืองข้อมูล 4) การประเมินผล

1. การพัฒนาคลังข้อมูล (Data Warehouse)

ผู้วิจัยได้วิเคราะห์ความต้องการของผู้ใช้งานซึ่งได้แก่ ผู้บริหาร หัวหน้ากลุ่มสาระการเรียนรู้ และอาจารย์ประจำชั้น ด้วยการสัมภาษณ์ถึงปัญหาและความต้องการ เพื่อใช้ในการออกแบบคลังข้อมูลและสร้างรายงาน ในการออกแบบคลังข้อมูลผู้วิจัยได้กำหนดมุมมอง (Dimension) ค่าที่ต้องการวัด (Measure) ประเภทตารางข้อเท็จจริง (Fact Table) และตารางมิติ (Dimension Table) โดยใช้โครงสร้างคลังข้อมูลแบบสโนว์เฟลก สกีมา (Snowflake Schema) ดังภาพที่ 1 จากนั้นนำข้อมูลจากฐานข้อมูลผลการเรียน และประวัตินักเรียนระดับชั้นมัธยมศึกษา ตั้งแต่ปีการศึกษา 2548 -2556 เข้าคลังข้อมูลซึ่งใช้โปรแกรม Microsoft SQL Server 2008 R2 ด้วยกระบวนการอีทีแอล (Extract Transform and Load : ETL) ตามที่ออกแบบไว้



ภาพที่ 1 โครงสร้างคลังข้อมูลแบบสโนว์เฟลก สกีมา (Snowflake Schema)

2. การพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ (Online Analytical Processing)

ในการพัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ (OLAP) เป็นกระบวนการวิเคราะห์ข้อมูลด้วยตารางมิติมุมมอง (Dimension) และตารางข้อเท็จจริง (Fact Table) ทำให้เรียกดูรายละเอียดของข้อมูลแต่ละมิติได้ โดยใช้โปรแกรม Microsoft SQL Server 2008 Analysis Services (SSAS) เป็นเครื่องมือสำหรับการวิเคราะห์ข้อมูลหลายมิติ (กุลเทพ จิรลักข์ และเอกสิทธิ์ วิวัฒนาประสิทธิ์, 2555: 56) และสร้างมุมมองของผู้ใช้งาน 3 กลุ่ม คือ ผู้บริหารใช้ดูจำนวนนักเรียนในแต่ละปีการศึกษา หัวหน้ากลุ่มสาระการเรียนรู้ใช้ดูผลการเรียนเฉลี่ยของแต่ละกลุ่มสาระการเรียนรู้ในแต่ละภาคการศึกษา และอาจารย์ประจำชั้นใช้ดูผลการเรียนและคะแนนของนักเรียนในแต่ละระดับชั้นของภาคการศึกษา และใช้โปรแกรม QlikView เป็นเครื่องมือในการสร้างรายงาน

3. การทำเหมืองข้อมูล (Data Mining)

ในส่วนการทำเหมืองข้อมูลผู้วิจัยเลือกใช้โปรแกรม WEKA Version 3.6.12 เป็นเครื่องมือที่ใช้ในการสร้างตัวแบบพยากรณ์และใช้วิธีการวัดประสิทธิภาพแบบ 10-fold Cross Validation ซึ่งขั้นตอนการทำเหมืองข้อมูลมีดังนี้

3.1 การจัดเตรียมข้อมูล

ข้อมูลที่ใช้ในการทำเหมืองข้อมูลเป็นข้อมูลจากคลังข้อมูลที่พัฒนาขึ้นมา โดยใช้ข้อมูลนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 ระหว่างปีการศึกษา 2553 – 2556 จำนวน 525 ระเบียบ ประกอบด้วยข้อมูลประวัติส่วนตัวและข้อมูลผลการเรียนของนักเรียนเป็นข้อมูลนำเข้า จำนวน 16 คุณลักษณะ ดังตารางที่ 1 และข้อมูลออก 1 คุณลักษณะ คือ ผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 4 โดยแบ่งระดับผลการเรียนเฉลี่ยออกเป็น 3 คลาส ได้แก่ High, Medium, Low ดังตารางที่ 2 เพื่อใช้ในการสร้างตัวแบบพยากรณ์ (Model) ซึ่งช่วยลดการกระจายของข้อมูล นำข้อมูลที่ได้มาทำการตรวจสอบความสมบูรณ์ และทำการแทนค่าข้อมูลเพื่อใช้ในการสร้างตัวแบบพยากรณ์

ตารางที่ 1 คุณลักษณะที่ใช้ในการสร้างตัวแบบพยากรณ์

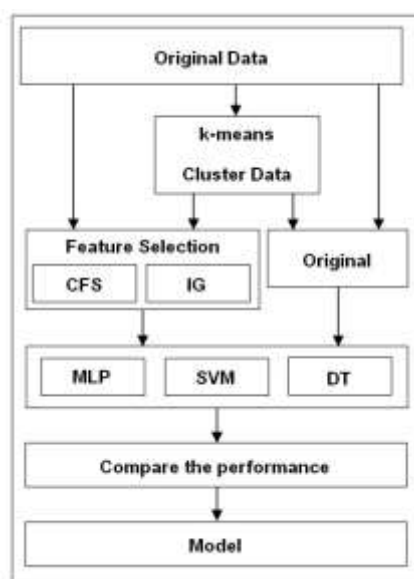
ข้อมูลประวัติส่วนตัว	ข้อมูลผลการเรียนของนักเรียน
1. เพศ	9. ผลการเรียนเฉลี่ยคณิตศาสตร์
2. แผนการเรียน	10. ผลการเรียนเฉลี่ยวิทยาศาสตร์
3. อาชีพของบิดา	11. ผลการเรียนเฉลี่ยภาษาไทย
4. รายได้เฉลี่ยต่อปีของบิดา	12. ผลการเรียนเฉลี่ยสังคมศึกษา
5. อาชีพของมารดา	13. ผลการเรียนเฉลี่ยภาษาอังกฤษ
6. รายได้เฉลี่ยต่อปีของมารดา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

ตารางที่ 2 เกรดเฉลี่ยในแต่ละคลาส

คลาส	ระดับเกรดเฉลี่ย
High	3.01 – 4.00
Medium	2.01 – 3.00
Low	1.00 – 2.00

3.2 การสร้างตัวแบบพยากรณ์

ในขั้นตอนการสร้างตัวแบบพยากรณ์ โดยข้อมูลแบ่งเป็น 2 แบบ ได้แก่ ชุดข้อมูลที่ไม่จัดกลุ่ม (Original Data) เป็นข้อมูลนักเรียนจำนวน 525 ระเบียบ และชุดข้อมูลที่จัดกลุ่ม (Cluster Data) เป็นข้อมูลนักเรียนที่นำมาผ่านการจัดกลุ่มข้อมูลด้วยอัลกอริทึม k-means ผู้วิจัยเลือกใช้ค่าผิดพลาดในการจำแนกกลุ่ม (Incorrectly Clustered Instances : ICI) ที่น้อยที่สุดในการแบ่งกลุ่มข้อมูล จากนั้นนำไปผ่านกระบวนการคัดเลือกคุณลักษณะ (Feature Selection) ซึ่งใช้วิธี Correlation-based Feature Selection (CFS) และวิธี Information Gain (IG) ก่อนเพื่อให้ได้คุณลักษณะที่สำคัญ (Bi, Zhang, and Cheng, 2009) นำไปสร้างตัวแบบพยากรณ์ หลังจากผ่านกระบวนการคัดเลือกคุณลักษณะ จึงใช้เทคนิคเหมือนข้อมูลโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (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 2



ภาพที่ 2 กระบวนการสร้างตัวแบบพยากรณ์

3.3 การวัดประสิทธิภาพ

ในงานวิจัยนี้ใช้การทดสอบแบบไขว้พับ (k - fold cross Validation) แบบ 10 ส่วนแล้วทำการทดสอบเพื่อประเมินประสิทธิภาพ (Ron, 1995) โดยวัดค่าความถูกต้อง (Accuracy) ดังสมการ (1) และค่ารากที่สองของความคลาดเคลื่อน (Root Mean Square Error : RMSE) ดังสมการ (2)

$$\text{Accuracy} = \frac{\text{Total Number of Instances}}{\text{Correctly Classified Instance}} \times 100 \quad (1)$$

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

4. การประเมินผล

ในการประเมินผลการใช้งานผู้วิจัยใช้แบบสอบถามในการประเมิน ซึ่งมีมาตราวัดอัตราส่วน 5 ระดับ (Rating Scale) ได้แก่ ระดับดีมากเท่ากับ 5, ระดับดีเท่ากับ 4, ระดับปานกลางเท่ากับ 3, ระดับพอใช้เท่ากับ 2, และระดับปรับปรุงเท่ากับ 1 เพื่อประเมินความพึงพอใจของผู้ใช้งานที่มีต่อคลังข้อมูลและระบบพยากรณ์ผลการเรียน โดยแบ่งเป็น 2 แบบ ได้แก่

4.1 แบบประเมินความพึงพอใจผู้ใช้งานคลังข้อมูล

โดยสอบถามจากผู้ใช้งานจำนวน 10 คน โดยแบ่งเป็น ผู้บริหาร 2 คน หัวหน้ากลุ่มสาระการเรียนรู้ 3 คน และอาจารย์ประจำชั้น 5 คน เพื่อหาค่าเฉลี่ย (\bar{x}) ดังสมการ (3)

4.2 แบบประเมินความเชื่อมั่นผู้ใช้งานระบบพยากรณ์ผลการเรียน

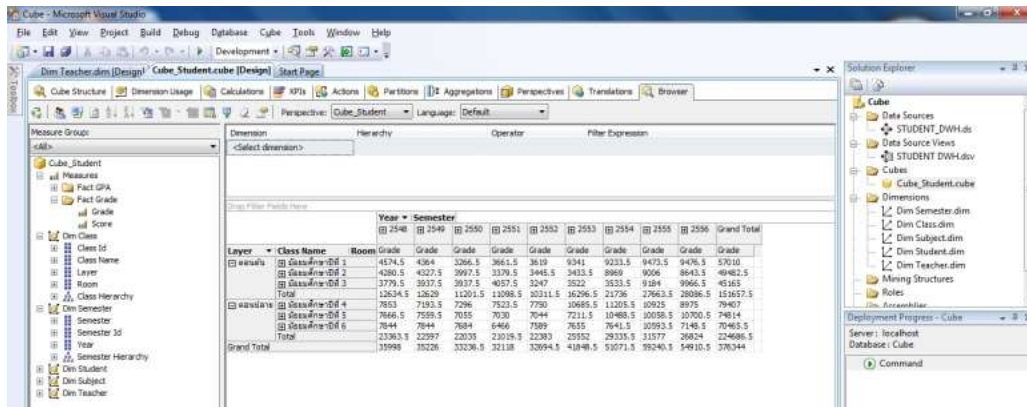
โดยสอบถามจากผู้ใช้งานจำนวน 40 คน โดยแบ่งเป็น อาจารย์แนะแนว 2 คน นักเรียนระดับมัธยมศึกษาปีที่ 3 จำนวน 19 คน และนักเรียนระดับมัธยมศึกษาปีที่ 4 จำนวน 19 คน เพื่อหาค่าเฉลี่ย (\bar{x}) ดังสมการ (3)

$$\bar{x} = \frac{\sum x}{N} \quad (3)$$

ผลการวิจัย

1. การพัฒนาคลังข้อมูล

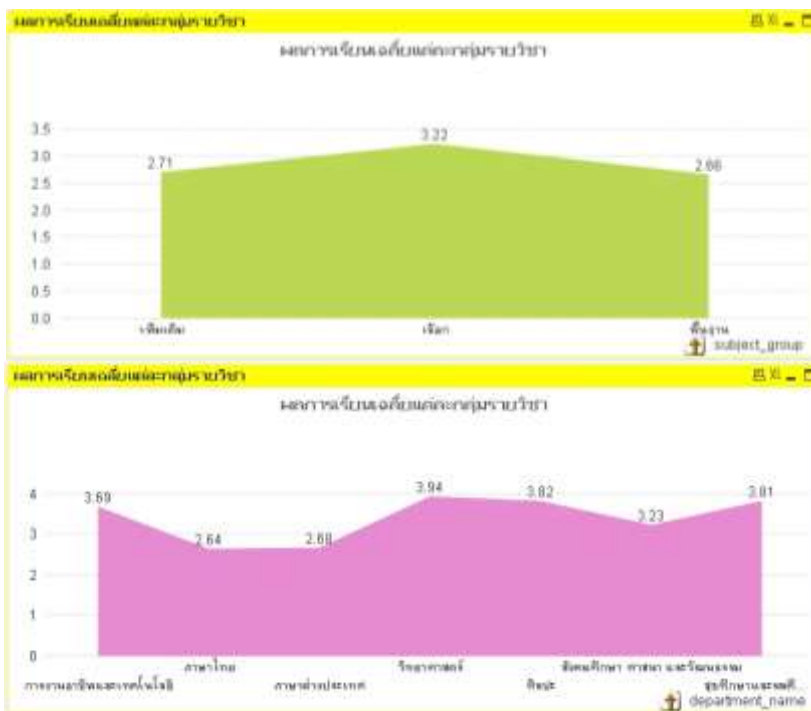
การพัฒนาคลังข้อมูลนักเรียนระดับมัธยมศึกษา ได้พัฒนาระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ (OLAP) ที่สามารถดูมิติมุมมอง (Dimension) ต่างๆ ได้ดังนี้ มิติปีการศึกษา มิติวิชาเรียน มิตินักเรียน มิติระดับชั้น และมีติอาจารย์ โดยมีค่าการวัด (Measure) คือ คะแนน ผลการเรียน และผลการเรียนเฉลี่ย ดังภาพที่ 3



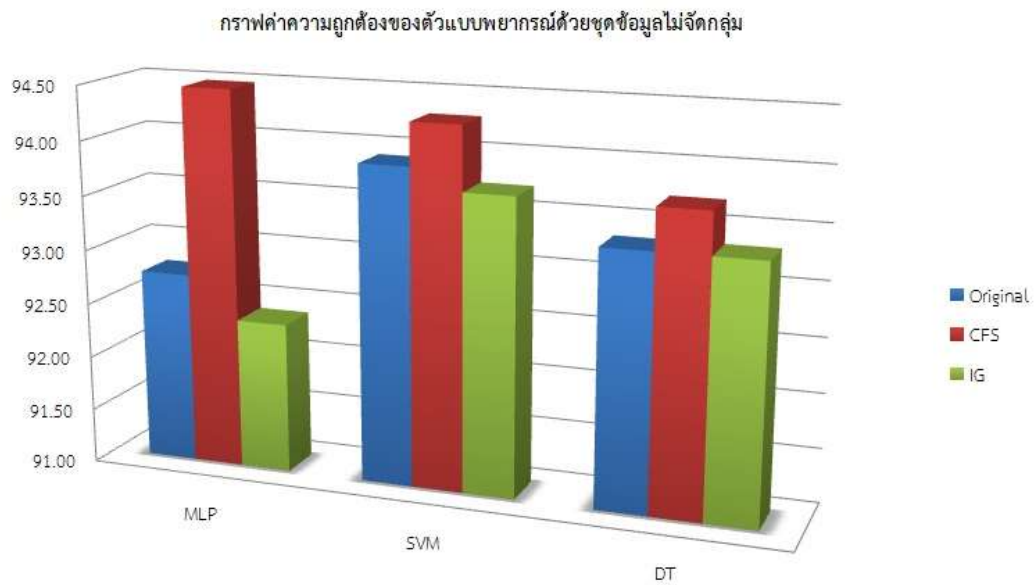
ภาพที่ 3 คลังข้อมูลและมุมมองข้อมูลในมิติต่างๆ

2. ระบบสนับสนุนการวิเคราะห์ข้อมูลหลายมิติ

การแสดงรายงานด้วยโปรแกรม QlikView เป็นการแสดงรายงานให้ผู้ใช้งานทั้ง 3 กลุ่มได้แก่ ผู้บริหาร หัวหน้ากลุ่มสาระการเรียนรู้ และอาจารย์ประจำชั้นตามมุมมองที่ออกแบบไว้สามารถเรียกดูข้อมูลแบบสรุป (Roll UP) และแบบเจาะลึก (Drill Down) ดังภาพที่ 4

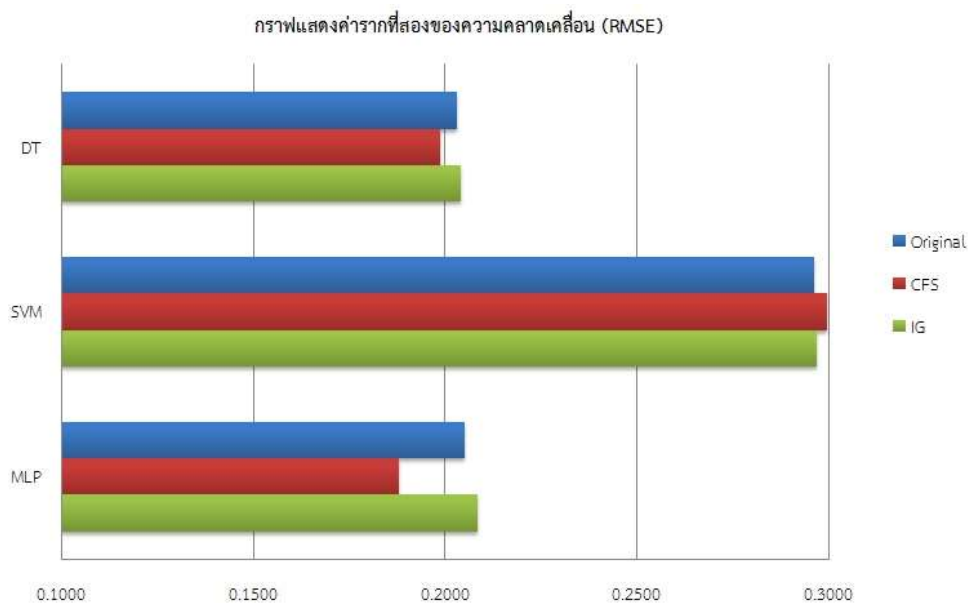


ภาพที่ 4 ข้อมูลแบบสรุป (Roll UP) และแบบเจาะลึก (Drill Down)



ภาพที่ 8 กราฟแสดงค่าความถูกต้องของตัวแบบพยากรณ์ด้วยชุดข้อมูลไม่จัดกลุ่ม

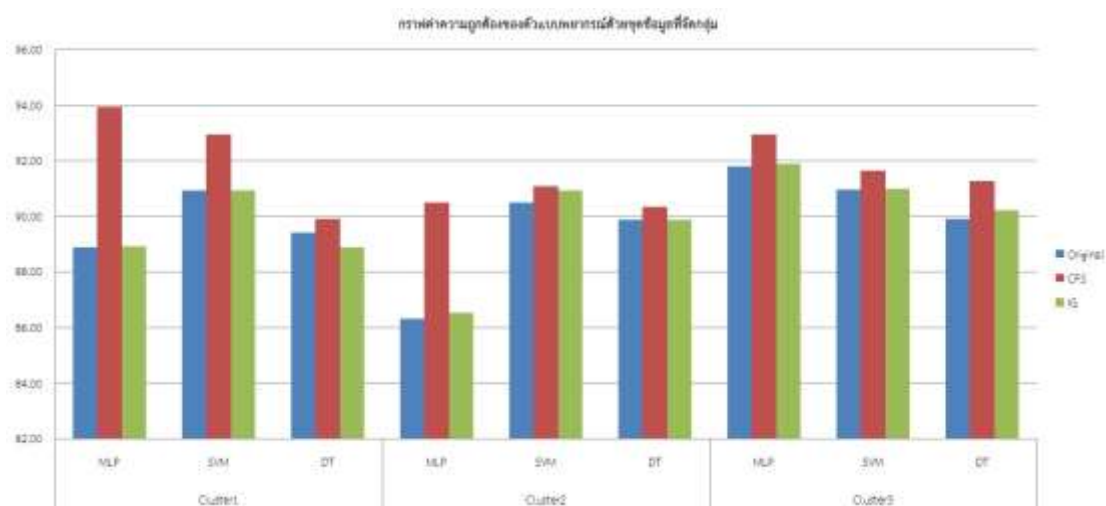
นอกจากนี้ เมื่อทำการเปรียบเทียบค่ารากที่สองของความคลาดเคลื่อน (RMSE) พบว่า ตัวแบบพยากรณ์โครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ร่วมกับวิธีการคัดเลือกคุณลักษณะแบบ Correlation-based Feature Selection (CFS) ให้ค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดเท่ากับ 0.1880 ส่วนตัวแบบพยากรณ์ซัพพอร์ตเวกเตอร์แมชชีน (SVM) มีค่าเท่ากับ 0.2960 และตัวแบบพยากรณ์ต้นไม้ตัดสินใจ (DT) มีค่าเท่ากับ 0.1988 ดังภาพที่ 9



ภาพที่ 9 กราฟแสดงค่ารากที่สองของความคลาดเคลื่อน (RMSE)

3.2 ผลการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่ม

ผลการสร้างตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่มแบ่งออกเป็น 3 กลุ่ม กลุ่ม Cluster1 พบว่า ตัวแบบพยากรณ์โครงข่ายประสาทเทียมแบบมีลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ร่วมกับวิธี การคัดเลือกคุณลักษณะแบบ Correlation-based Feature Selection (CFS) ซึ่งคัดเลือกคุณลักษณะไว้ 5 คุณลักษณะ คือ แผนการเรียน, ผลการเรียนเฉลี่ยคณิตศาสตร์, ผลการเรียนเฉลี่ยสังคมศึกษา, ผลการเรียนเฉลี่ยภาษาอังกฤษ และผล การเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3 ให้ค่าความถูกต้องร้อยละ 93.94 สูงที่สุด และค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดเท่ากับ 0.2002 ส่วนกลุ่ม Cluster2 พบว่า ตัวแบบพยากรณ์ ซัพพอร์ตเวกเตอร์แมชชีน (SVM) ร่วมกับวิธีการคัดเลือกคุณลักษณะแบบ Correlation-based Feature Selection (CFS) ซึ่งคัดเลือกคุณลักษณะไว้ 4 คุณลักษณะ คือ ผลการเรียนเฉลี่ยคณิตศาสตร์, ผลการเรียนเฉลี่ยสังคมศึกษา, ผลการเรียนเฉลี่ย ภาษาอังกฤษ และผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3 ให้ค่าความถูกต้องร้อยละ 91.07 สูงที่สุด และค่า รากที่สองของความคลาดเคลื่อนน้อยที่สุดเท่ากับ 0.2301 และกลุ่ม Cluster3 พบว่า ตัวแบบพยากรณ์โครงข่าย ประสาทเทียมแบบมีลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ร่วมกับวิธีการคัดเลือกคุณลักษณะแบบ Correlation-based Feature Selection (CFS) ซึ่งคัดเลือกคุณลักษณะไว้ 5 คุณลักษณะ คือ สถานภาพของบิดามารดา, จำนวนพี่น้อง, ผลการเรียนเฉลี่ยวิทยาศาสตร์ และผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3 ให้ค่าความ ถูกต้องร้อยละ 92.93 สูงที่สุด และค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดเท่ากับ 0.2761 ดังภาพที่ 10 และภาพที่ 11



ภาพที่ 10 กราฟแสดงค่าความถูกต้องของตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่ม



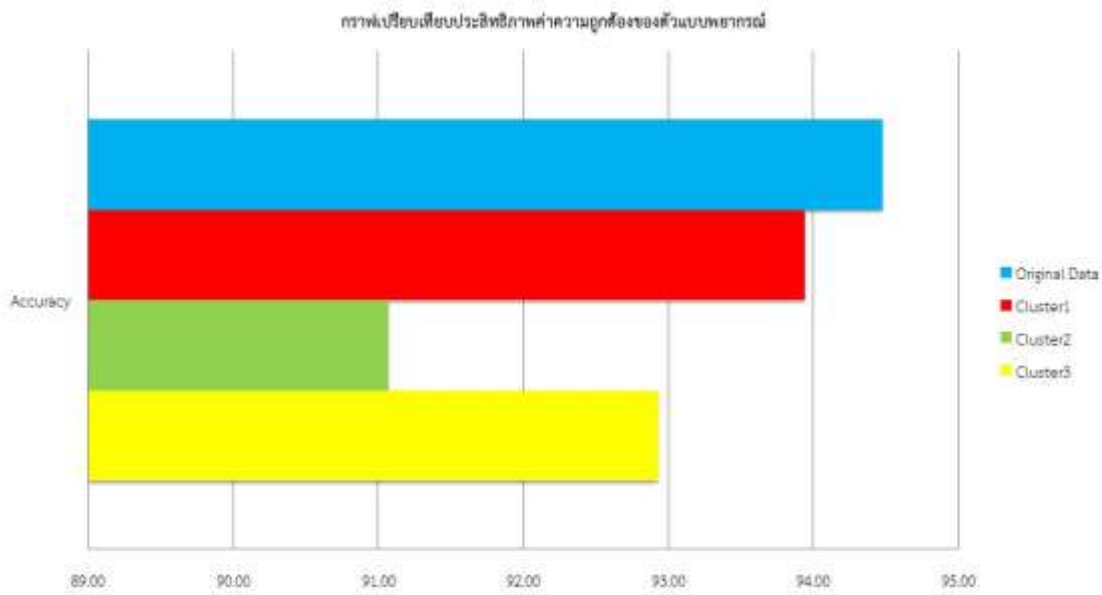
ภาพที่ 11 กราฟแสดงค่ารากที่สองของความคลาดเคลื่อน (RMSE)

3.3 การเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์

นำผลที่ได้จากตัวแบบพยากรณ์ด้วยชุดข้อมูลที่ไม่จัดกลุ่มที่มีค่าความถูกต้องสูงที่สุดมาเปรียบเทียบกับตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่มที่มีค่าความถูกต้องสูงที่สุดในแต่ละกลุ่ม ผลการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ พบว่า ตัวแบบพยากรณ์ด้วยชุดข้อมูลที่ไม่จัดกลุ่มด้วยเทคนิคโครงข่ายประสาทเทียมแบบมีลติเลเยอร์เพอร์เซ็ปตรอน (MLP) ร่วมกับวิธีการคัดเลือกคุณลักษณะแบบ Correlation-based Feature Selection (CFS) ซึ่งคัดเลือกคุณลักษณะไว้ 5 คุณลักษณะ มีค่าความถูกต้อง (Accuracy) ร้อยละ 94.48 สูงกว่าตัวแบบพยากรณ์ด้วยชุดข้อมูลที่จัดกลุ่ม ดังภาพที่ 12 ส่วนค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดเท่ากับ 0.1880 ดังตารางที่ 3

ตารางที่ 3 ผลการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์

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



ภาพที่ 12 กราฟเปรียบเทียบประสิทธิภาพค่าความถูกต้องของตัวแบบพยากรณ์

ดังนั้นในการสร้างระบบพยากรณ์ผู้วิจัยจึงใช้ตัวแบบพยากรณ์ที่ใช้ชุดข้อมูลที่ไม่จัดกลุ่ม โดยใช้วิธีการคัดเลือกคุณลักษณะแบบ Correlation-based Feature Selection (CFS) ร่วมกับเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (MLP) เหมาะสมที่สุดในการพยากรณ์ผลการเรียนของนักเรียน เนื่องจากให้ค่าความถูกต้องของการพยากรณ์ที่ดีที่สุด และค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุด จากนั้นนำตัวแบบพยากรณ์ที่ดีที่สุดไปพัฒนาด้วยภาษา PHP ในการเขียนแอปพลิเคชัน และใช้โปรแกรมฐานข้อมูล MySQL เช่นเดียวกับงานวิจัยของชิตชนก ศรีชัยวงศ์, ไพศาล ตระกูลสุข และสุรเดช บุญลือ(2557: 1) ดังภาพที่ 12



ภาพที่ 12 ระบบพยากรณ์ผลการเรียนของนักเรียน

3. การประเมินผลการใช้งาน

3.1 การประเมินความพึงพอใจผู้ใช้งานคลังข้อมูล

ผลการประเมินความพึงพอใจ โดยให้ผู้ใช้งานจำนวน 10 คน โดยแบ่งเป็น ผู้บริหาร 2 คน หัวหน้ากลุ่มสาระการเรียนรู้ 3 คน และอาจารย์ประจำชั้น 5 คน ผลการประเมินความพึงพอใจของผู้ใช้คลังข้อมูลพบว่า คลังข้อมูลมีความสะดวกและง่ายต่อการใช้งานค่าเฉลี่ยเท่ากับ 3.80 ในการออกแบบรายงานเข้าใจได้ง่ายมีค่าเฉลี่ยเท่ากับ 3.90 ส่วนความถูกต้องของข้อมูลในการรายงานมีค่าเฉลี่ยเท่ากับ 3.90 ช่วยลดขั้นตอนในการทำงานมีค่าเฉลี่ยเท่ากับ 4.30 และข้อมูลตรงกับความต้องการผู้ใช้งานมีค่าเฉลี่ยเท่ากับ 4.40 ผลการประเมินความพึงพอใจของผู้ใช้งานมีค่าเฉลี่ยรวมเท่ากับ 4.06 เมื่อเทียบกับเกณฑ์อยู่ในระดับดี คลังข้อมูลนักเรียนมีประสิทธิภาพสามารถนำไปใช้งานได้

3.2 การประเมินความเชื่อมั่นของผู้ใช้งานระบบพยากรณ์

ผลการประเมินความความเชื่อมั่นในระบบพยากรณ์ โดยผู้ใช้งาน จำนวน 40 คน แบ่งเป็น อาจารย์แนะแนว 2 คน นักเรียนระดับมัธยมศึกษาปีที่ 3 จำนวน 19 คน และนักเรียนระดับมัธยมศึกษาปีที่ 4 จำนวน 19 คน พบว่า ในเรื่องของความสะดวกและง่ายต่อการใช้งานมีค่าเฉลี่ยเท่ากับ 4.33 ส่วนความน่าเชื่อถือของผลการพยากรณ์มีค่าเฉลี่ยเท่ากับ 4.20 ผลการพยากรณ์ตรงกับความต้องการมีค่าเฉลี่ยเท่ากับ 4.25 และสามารถนำไปประกอบการตัดสินใจมีค่าเฉลี่ยเท่ากับ 4.28 ผลการประเมินความความเชื่อมั่นมีค่าเฉลี่ยรวมเท่ากับ 4.27 เมื่อเทียบกับเกณฑ์อยู่ในระดับดี ระบบพยากรณ์ผลการเรียนของนักเรียนมีประสิทธิภาพสามารถนำไปใช้งานได้

สรุปและอภิปรายผล

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาคลังข้อมูลและการใช้เทคนิคเหมืองข้อมูลมาประยุกต์ใช้ในการพยากรณ์ผลการเรียนของนักเรียน ในการพัฒนาคลังข้อมูลใช้โปรแกรม Microsoft SQL Server 2008 R2 และออกแบบโครงสร้างแบบสโนว์เฟลก (Snowflake Schema) สร้างคิวบ์ข้อมูลนักเรียน ที่สามารถประมวลผลเชิงวิเคราะห์ออนไลน์ (OLAP) ผู้ใช้งานสามารถเรียกดูรายงานผ่านโปรแกรม QlikView ซึ่งผู้ใช้งานมีความพึงพอใจต่อคลังข้อมูลที่พัฒนาขึ้นมีค่าเฉลี่ยรวมเท่ากับ 4.06 อยู่ในระดับดี ส่วนการสร้างตัวแบบพยากรณ์ผลการเรียนของนักเรียนจะใช้ชุดข้อมูลแบบไม่แบ่งกลุ่มและชุดข้อมูลแบบแบ่งกลุ่มมาผ่านวิธีการคัดเลือกคุณลักษณะ 2 วิธี ได้แก่ Correlation-based Feature Selection (CFS) และวิธี Information Gain (IG) ร่วมกับเทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียมแบบมีลิตเลเยอร์เพอร์เซ็ปตรอน เทคนิคซัพพอร์ตเวกเตอร์แมชชีน และเทคนิคต้นไม้ตัดสินใจ จากผลการวิจัยพบว่า การสร้างตัวแบบพยากรณ์ข้อมูลแบบไม่แบ่งกลุ่มข้อมูลโดยใช้การคัดเลือกคุณลักษณะด้วยวิธี Correlation-based Feature Selection (CFS) ร่วมกับโครงข่ายประสาทเทียมแบบมีลิตเลเยอร์เพอร์เซ็ปตรอน ให้ค่าความถูกต้องสูงที่สุดมีค่าร้อยละ 94.48 และมีค่ารากที่สองของความคลาดเคลื่อนน้อยที่สุดมีค่าเท่ากับ 0.1880 ซึ่งผลการวิจัยสอดคล้องกับงานวิจัยของ วงกต ศรีอุไร (2557: 101) ที่ได้สร้างโมเดลในการจำแนกผู้ป่วยโรคอ้วนลงพุงโดยใช้วิธีการคัดเลือกคุณลักษณะ ร่วมกับเทคนิคเหมืองข้อมูล ผลของงานวิจัยนี้พบว่าวิธีการคัดเลือกคุณลักษณะด้วย Correlation-based Feature Selection (CFS) ร่วมกับเทคนิคเหมือง

ข้อมูลแบบโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน ในการสร้างโมเดลจำแนกข้อมูลให้ค่าความถูกต้องในการจำแนกสูงที่สุด มีค่าเท่ากับ 92.56%

จึงสรุปได้ว่า การคัดเลือกคุณลักษณะด้วยวิธี Correlation -based Feature Selection (CFS) ร่วมกับการใช้เทคนิคเหมืองข้อมูลแบบโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน เหมาะสมที่สุดกับข้อมูลที่ใช้ในการสร้างตัวแบบพยากรณ์ในงานวิจัยนี้มากที่สุด และเมื่อคุณลักษณะที่ส่งผลในการพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา จากการคัดเลือกคุณลักษณะพบว่ามี 5 ปัจจัย ได้แก่ แผนการเรียน, ผลการเรียนเฉลี่ยวิทยาศาสตร์, ผลการเรียนเฉลี่ยสังคมศึกษา, ผลการเรียนเฉลี่ยภาษาอังกฤษ และผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3 ที่ส่งผลต่อผลการเรียนในระดับมัธยมศึกษาปีที่ 4 ซึ่งผลในการพยากรณ์สามารถนำไปใช้ในการแนะนำหรือประกอบการตัดสินใจของนักเรียนระดับชั้นมัธยมศึกษาตอนต้นในการเลือกแผนการเรียนในระดับมัธยมศึกษาตอนปลาย และเป็นการวางแผนการเรียน แรงจูงใจในการศึกษาเพื่อทำให้นักเรียนมีผลการเรียนที่สูงขึ้น ทำให้สามารถเลือกแผนการเรียนที่นักเรียนต้องการได้ อีกทั้งเป็นเครื่องมือให้อาจารย์แนะแนวเพื่อให้คำแนะนำในการศึกษาต่อแก่นักเรียนได้

ข้อเสนอแนะในการวิจัยพบว่า การใช้เทคนิคเหมืองข้อมูลเป็นการค้นหาความรู้จากฐานข้อมูลที่มีอยู่มาพยากรณ์ผลการเรียนของนักเรียน ซึ่งในการวิจัยครั้งต่อไปควรมีการเพิ่มปัจจัยอื่นๆ ที่น่าจะมีผลต่อการเรียนของนักเรียนมาพิจารณาเป็นปัจจัยในการพยากรณ์ และควรมีข้อมูลมากพอในการทำเหมืองข้อมูล อาจใช้เทคนิคเหมืองข้อมูลเทคนิคอื่นๆ มาเปรียบเทียบเพื่อให้ได้ผลการพยากรณ์ที่มีประสิทธิภาพมากขึ้น

เอกสารอ้างอิง

ภาษาไทย

กุลเทพ จิรลักข์ และเอกสิทธิ์ วิวัฒนาประสิทธิ์. (2555). กรณีศึกษาการทำคลังข้อมูล. ใน **ประมวลการสนชชุดวิชาคลังข้อมูล เหมืองข้อมูล และธุรกิจอัจฉริยะ หน่วยที่ 1-7 สาขาวิทยาศาสตร์และเทคโนโลยี สารสนเทศ มหาวิทยาลัยสุโขทัยธรรมาธิราช**, 299-383. นนทบุรี: สำนักพิมพ์มหาวิทยาลัยสุโขทัยธรรมาธิราช.

ชิตชนก ศรีชัยวงศ์ ไพศาล ตระกูลสุข และสุรเดช บุญลือ.(2557). **ระบบสนับสนุนการตัดสินใจเพื่อวินิจฉัยโรคใบลำไยด้วยเทคนิคต้นไม้ตัดสินใจ**. Veridian E-Journal Science and Technology Silpakorn University ฉบับที่ 6 (พฤศจิกายน – ธันวาคม): 1-14.

พรเทพ คงไชย และรัชฎา คงคะจันทร์. (2554). **การศึกษาเชิงเปรียบเทียบในการคัดเลือกคุณลักษณะที่เหมาะสมสำหรับการทำเหมืองข้อมูลเพื่อพยากรณ์โอกาสการสำเร็จการศึกษาของนักศึกษา**. การประชุมวิชาการข่ายงานวิศวกรรมอุตสาหกรรม ประจำปี 2554 , (ตุลาคม): 2012- 2015.

พรณิภา บุตรเอก และสุรเดช บุญลือ.(2557). **การพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้ซอฟต์แวร์เวกเตอร์แมชชีน**. Veridian E-Journal Science and Technology Silpakorn University ฉบับที่ 6 (พฤศจิกายน – ธันวาคม): 40-49.

- วงศ ศรีอุไร. (2557). การจำแนกผู้ป่วยโรคอ้วนลงพุงโดยใช้วิธีการคัดเลือกคุณลักษณะและโครงข่ายประสาทเทียม. วารสารวิทยาศาสตร์ มศว. ฉบับที่ 1 (มิถุนายน): 91-102.
- วิภา เจริญภักดิ์ธารักษ์ (2555). หลักการพื้นฐานของการทำเหมืองข้อมูล. ใน **ประมวลการสอบชุดวิชาคลังข้อมูล เหมืองข้อมูล และธุรกิจอัจฉริยะ หน่วยที่ 8-15 สาขาวิทยาศาสตร์และเทคโนโลยีสารสนเทศ มหาวิทยาลัยสุโขทัยธรรมาธิราช**, 1-62. นนทบุรี: สำนักพิมพ์มหาวิทยาลัยสุโขทัยธรรมาธิราช.
- สำนักงานเลขาธิการศึกษา. (2553). **นิยามทางการศึกษา**. เข้าถึงเมื่อ 19 กรกฎาคม 2558.เข้าถึงได้จาก http://www.onec.go.th/onec_web/page.php?mod=Category&categoryID=CNT0000145
- สุพัฒน์กุล ภัคโชค. (2555). **ตัวแบบเลือกแผนการเรียนระดับมัธยมศึกษาตอนปลาย โดยการพิจารณาผลการเรียนรายวิชาหลัก ด้วยเทคนิคเหมืองข้อมูล** สารนิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิศวกรรมเว็บ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยธุรกิจบัณฑิตย์.
- สุวรรณี อัครกุลชัย. (2555). หลักการพื้นฐานของคลังข้อมูลข้อมูล. ใน **ประมวลการสอบชุดวิชาคลังข้อมูล เหมืองข้อมูล และธุรกิจอัจฉริยะ หน่วยที่ 1-7 สาขาวิทยาศาสตร์และเทคโนโลยีสารสนเทศ มหาวิทยาลัยสุโขทัยธรรมาธิราช**, 1-45. นนทบุรี: สำนักพิมพ์มหาวิทยาลัยสุโขทัยธรรมาธิราช.

ภาษาต่างประเทศ

- Han, J., and Kamber, M. (2006). **Data mining concepts and techniques (2th ed.)**. United States of America. Morgan Kaufman Publishers.
- Jing Bi, Kun Zhang, and Xiaojing Cheng. (2009). **Intrusion Detection Base on RBF Neural Network**. International Symposium on Information Engineering and Electronic Commerce.
- Mohammad Rifaie, Erwin J. Bls and Abdel-Rahman M. Muhsen. (2008). **Data Warehouse Architecture for GIS Applications**. Proceeding of iiWAS2008, Linz, Austria.
- Ron Kohavi. (1995). **A study of cross-validation and bootstrap for accuracy estimation And model selection**. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence. Vol. 2, no. 12, pp. 1137-1143.