

การศึกษาและพัฒนาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียนโดยใช้เทคนิคเหมืองข้อมูล

สำราญ วานนท์¹, รจนา เมืองแสน²

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์ ดังนี้ 1) เพื่อศึกษาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียนโดยใช้เทคนิคเหมืองข้อมูล 2) เพื่อเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน และ 3) เพื่อพัฒนาระบบสารสนเทศพยากรณ์การเข้าศึกษาของนักศึกษาใหม่ ซึ่งศึกษาจากข้อมูลการรับสมัครนักศึกษาของมหาวิทยาลัยราชภัฏชัยภูมิ โดยใช้โปรแกรม WEKA จัดกลุ่มข้อมูลด้วยเทคนิคซิมเพิลเคมีน หากภูความสัมพันธ์ของข้อมูลด้วยอะพริออริอัลกอริทึม นำภูความสัมพันธ์ดังกล่าวไปประเมินความเหมาะสมจากผู้เชี่ยวชาญทั้งหมด 5 คน จากนั้นสร้างตัวแบบและการประเมินตัวแบบพยากรณ์ด้วยเทคนิคต้นไม้ตัดสินใจและเทคนิคแรนดอมฟอเรสเปรียบเทียบตัวแบบที่ให้ผลการพยากรณ์ที่ดีที่สุดไปพัฒนาระบบสารสนเทศ

ผลการวิจัยพบว่า ภูความสัมพันธ์ของข้อมูลจากทั้งหมด 6 กลุ่ม เลือกภูความสัมพันธ์ที่น่าสนใจกลุ่มละ 2 ภู ประเมินความเหมาะสมจากผู้เชี่ยวชาญ พบว่า ความเหมาะสมในภาพรวมอยู่ในระดับมากที่สุด การเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์ 2 เทคนิค คือ ต้นไม้ตัดสินใจและแรนดอมฟอเรส ปรากฏว่าเทคนิคแรนดอมฟอเรส ค่าความถูกต้องสูงสุดได้ 74.67% ถือว่าเป็นตัวแบบที่ยอมรับได้และนำไปพัฒนาระบบสารสนเทศพยากรณ์จำนวนนักศึกษาใหม่ส่งเสริมนโยบายการประชาสัมพันธ์เชิงรุก

คำสำคัญ: ภูความสัมพันธ์, การแบ่งกลุ่มข้อมูล, ตัวแบบพยากรณ์

ชื่อผู้ติดต่อบทความ: สำราญ วานนท์

E-mail: samran@cpru.ac.th

(Received: November 16, 2018; Revised: June 25, 2019; Accepted: August 26, 2019)

¹ อาจารย์ประจำ คณะบริหารธุรกิจ มหาวิทยาลัยราชภัฏชัยภูมิ E-Mail: samran@cpru.ac.th

² อาจารย์ประจำ คณะบริหารธุรกิจ มหาวิทยาลัยราชภัฏชัยภูมิ E-Mail: rojana@cpru.ac.th

A Study and Development of Forecasting Model for The Suitability Characteristics on The Applying Major Selection by Using Data Mining Techniques

Samran Wanon ¹, Rojjana Muangsan ²

Abstract

The purposes of this research were to: 1) study the model of qualification prediction on applying major selected by using data mining techniques; 2) compare the performance of the appropriate feature model on applying major selected subjects and; 3) develop the forecasting information system of new students to enroll in courses. The data of this study was from the student recruitment data of Chaiyaphum Rajabhat University. The researcher had used WEKA application for clustering as Simple K-Mean technique. In addition, finding association rule as Apriori algorithm and evaluate suitability by five experts, creating a model and evaluating forecasting model with the Decision Tree technique and Random Forest technique. The highest performance forecasting result of the model was developed for the information systems.

The results from the research found for all 6 clusters of the association rule it selected two interesting association rules from each cluster. It was found that the overall suitability was at the highest level. Performance model comparison was divided into 2 techniques; Decision Tree and Random Forest. As a result, the highest accuracy is Random Forest at 74.67%. It is an acceptable prediction model and applied to the development of the information systems for forecasting the number of new students and promoting proactive of public relations policies.

Keywords: Association Rules, Clustering, Prediction Model

Corresponding Author: Samran Wanon

E-mail: samran@cpru.ac.th

¹ Lecturer in Faculty of Business Administration, Chaiyaphum Rajabhat University. E-mail: samran@cpru.ac.th

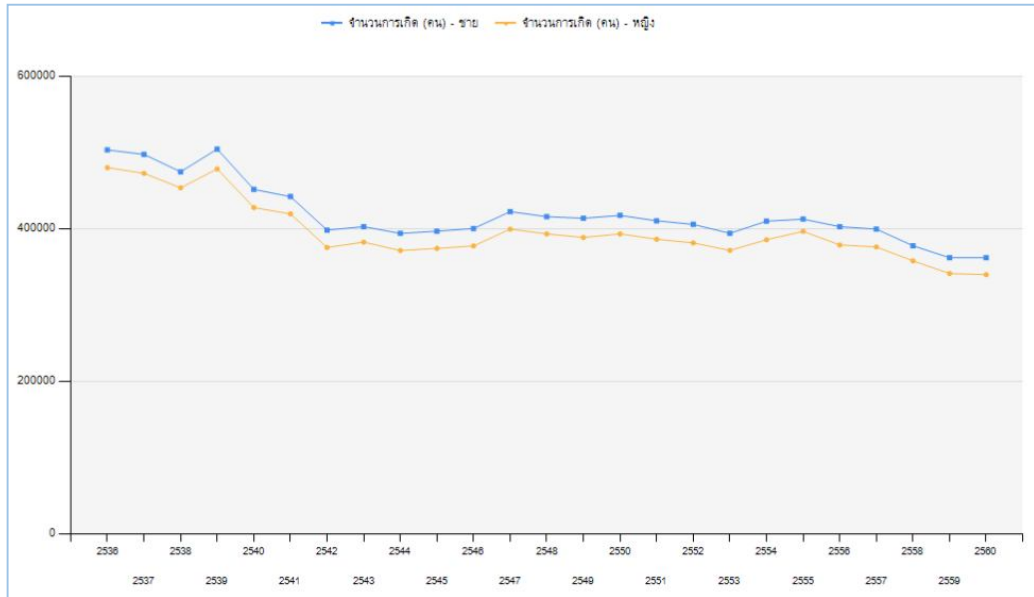
² Lecturer in Faculty of Business Administration, Chaiyaphum Rajabhat University. E-mail: rojjana@cpru.ac.th

1. บทนำ

ปัญหาจำนวนผู้เรียนที่ลดลงอย่างต่อเนื่องมีข้อมูลระบุชัดจากการสำรวจประชากรไทย เมื่อปี 2536 ถึง 2560 ย้อนหลังจะพบว่า เด็กเกิดใหม่มีแนวโน้มลดลงอย่างเห็นได้ชัด ดังแสดงตามภาพที่ 1 กรณีนี้ส่งผลต่อเนื่องไปถึงประชากรวัยเรียนในระดับมัธยมที่จะเข้าสู่มหาวิทยาลัยลดลงตามไปด้วย ขณะเดียวกันวัยเรียนอีกกลุ่มหนึ่งซึ่งครอบครัวฐานะยากจนถึงปานกลางหันไปเรียนสายอาชีพเพราะหวาดวิตกเรื่องตักเตือนและรัฐบาลเองก็เริ่มหันมาสนับสนุนให้เด็กเรียนสายอาชีพมากขึ้น สถาบันอุดมศึกษาทั้งมหาวิทยาลัย วิทยาลัย สถาบันการศึกษาทั้งของรัฐบาลในกำกับรัฐ และเอกชนที่มีอยู่ในประเทศไทยมีมากถึง 170 แห่ง ต่างแข่งขันกันจูงใจผู้เรียนมากขึ้นเพราะทราบดีว่าจำนวนผู้เรียนที่ลดลงจะทำให้เกิดปัญหาอื่นตามมา โดยเฉพาะปัญหาบางคณะวิชาของมหาวิทยาลัยที่ไม่มีชื่อเสียงดึงดูดผู้เรียน มีผู้เรียนน้อย เปิดรับหลายรอบจำนวนก็ไม่ได้เพิ่มขึ้น เมื่อห้องเรียนเริ่มไม่มีนักศึกษา ปัญหาเรื่องการจำกัดรายจ่ายโดยเฉพาะเรื่องจ้างผู้สอนก็ตามมาจากที่จ้างประจำก็เหลือเพียงจ้างไม่เต็มเวลาจนกระทบกับคุณภาพการเรียนการสอน (ชเนศน์ นุ่มมัน, 2561)

ปัจจุบันการรับสมัครนักศึกษาใหม่ในระดับปริญญาตรีของแต่ละสถาบันอุดมศึกษาต่างก็มียุทธวิธีหรือกลยุทธ์ในการประชาสัมพันธ์และการรับสมัครนักศึกษาที่มีความแตกต่างกันไปมหาวิทยาลัยราชภัฏชัยภูมิเองก็มีการประชาสัมพันธ์และการรับสมัครที่เป็นกระบวนการให้ได้มาซึ่งจำนวนนักศึกษาใหม่ โดยมีการประชาสัมพันธ์โดยทางสังคมออนไลน์ เว็บไซต์ ผ่านสื่อวิทยุ สื่อสิ่งพิมพ์ รวมถึงการออกไปประชาสัมพันธ์หรือการแนะนำนอกสถาบันการศึกษาของอาจารย์และบุคลากรแต่ละคณะซึ่งมีการเดินทางไปตามโรงเรียนหรือจังหวัดต่างๆ เพื่อให้ได้ผู้สมัครเพื่อเข้าศึกษาต่อในหลักสูตรของตนให้มีปริมาณที่มากขึ้น และเป็นการเพิ่มโอกาสในการคัดเลือกผู้ที่สมัครเรียนที่มีความสามารถเป็นนักศึกษาใหม่ได้ดียิ่งขึ้น อาจารย์และบุคลากรผู้มีหน้าที่ออกประชาสัมพันธ์การศึกษาต่อของนักศึกษาใหม่ต่างประสบปัญหาในการออกปฏิบัติหน้าที่ คือ การออกแนะนำหรือประชาสัมพันธ์พร้อมการรับสมัครนักศึกษานอกสถานที่แต่ละครั้งมีการใช้เวลา ใช้งบประมาณในจำนวนที่สูงและอาจได้รับผู้สมัครที่ไม่ตรงสาขาวิชาที่เปิดสอน ไม่ตรงกับกลุ่มเป้าหมาย หรือในหลายๆ ครั้งรับสมัครนักศึกษาใหม่แล้วเมื่อถึงเวลาเข้าศึกษาจริงก็ไม่ได้มีการรายงานตัวทำให้สิ่งที่ได้ลงทุนไปเสียเปล่า ดังนั้น ผู้วิจัยจึงได้มีแนวคิดที่ศึกษาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียนของมหาวิทยาลัยราชภัฏชัยภูมิโดยใช้เทคนิคเหมืองข้อมูล เพื่อส่งเสริมนโยบายการประชาสัมพันธ์เชิงรุกและได้มาซึ่งสารสนเทศที่พร้อมสำหรับการวางแผนและตัดสินใจสำหรับผู้บริหารต่อไป

การศึกษาและพัฒนาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน



ภาพที่ 1 แนวโน้มอัตราการเกิดของประชากรไทยตั้งแต่ปี 2536-2560

ที่มา: สำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ (2561)

2. วัตถุประสงค์

- 2.1 เพื่อศึกษาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียนโดยใช้เทคนิคเหมืองข้อมูล
- 2.2 เพื่อเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน
- 2.3 เพื่อพัฒนาระบบสารสนเทศพยากรณ์การเข้าศึกษาของนักศึกษาใหม่

3. การทบทวนวรรณกรรม และกรอบแนวคิด

3.1 การทบทวนวรรณกรรม

3.1.1 การทำเหมืองข้อมูล

เหมืองข้อมูลเป็นการใช้ประโยชน์จากข้อมูลที่ถูกเก็บไว้ในรูปแบบอิเล็กทรอนิกส์ โดยผ่านกระบวนการทางคอมพิวเตอร์ ในการค้นหารูปแบบ กฎเกณฑ์ของความสัมพันธ์ที่ซ่อนอยู่ให้ได้มาซึ่งความรู้ที่ต้องการเพื่อนำมาใช้ในการตัดสินใจ ถือเป็นเครื่องมือชิ้นใหม่ที่สามารถค้นหาข้อมูลในฐานข้อมูลขนาดใหญ่หรือข้อมูลที่เป็นประโยชน์ในการบริหาร (จามรกุล เหล่าเกียรติกุล, 2558, หน้า 1) การทำเหมืองข้อมูลเป็นขบวนการที่มีขั้นตอนในการปฏิบัติ เพื่อให้ได้มาซึ่งผลลัพธ์ที่มีความถูกต้องสมบูรณ์

การศึกษาและพัฒนาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน

3.1.1.1 กระบวนการทำเหมืองข้อมูล (Han, J. & Kamber, M, 2006, p. 7) การทำเหมืองข้อมูล เป็นวิธีการที่ให้ได้มาซึ่งองค์ความรู้ที่ซ่อนอยู่ในฐานข้อมูลขนาดใหญ่ โดยมีกระบวนการทำเหมืองข้อมูลดังนี้

1) การทำความสะอาดข้อมูล (Data Cleaning and Pre-Processing) เป็นขั้นตอนในการกรองข้อมูลเพื่อเลือกข้อมูลที่สามารถนำไปสู่ความรู้ที่ต้องการได้ตรงประเด็นหรือตรงความต้องการ โดยไม่นำข้อมูลที่ไมครบสมบูรณ์มาใช้ในการประมวลผลเหมืองข้อมูล ทำให้ได้ข้อมูลที่มีคุณภาพ

2) การรวบรวมข้อมูล (Data Integration) การจัดทำเหมืองข้อมูลต้องใช้ข้อมูลปริมาณมาก อาจจำเป็นต้องมีการรวบรวมข้อมูลทั้งหมดที่ต้องการจากหลายๆ แหล่งข้อมูล ซึ่งข้อมูลอาจจัดเก็บอยู่ในรูปแบบที่แตกต่างกันออกไปหรืออยู่ในหลายฐานข้อมูล ซึ่งต้องมีการรวบรวมข้อมูลแล้วทำให้ข้อมูลอยู่ในรูปแบบเดียวกันก่อน

3) การคัดเลือกข้อมูล (Data Selection) เป็นวิธีการเลือกและดึงเอาเฉพาะข้อมูลที่เกี่ยวข้องหรือที่ต้องการจะนำไปวิเคราะห์มาใช้ โดยระบุถึงแหล่งของข้อมูลที่จะนำมาทำเหมืองข้อมูลและแนวทางการนำข้อมูลที่ต้องการออกจากฐานข้อมูล เพื่อนำไปสร้างกลุ่มของข้อมูลสำหรับใช้ในการพิจารณาหรือค้นหาคำตอบความรู้ต่อไป

4) การแปลงข้อมูล (Data Transformation) เป็นวิธีการปรับเปลี่ยนข้อมูลให้มีค่าที่เหมาะสมสำหรับการทำเหมืองข้อมูลในด้านการคำนวณและการสืบค้นในกระบวนการทำเหมืองข้อมูล ตลอดจนให้ข้อมูลนั้นสื่อความหมายในการประกอบการตัดสินใจ

5) การจัดรูปแบบข้อมูล (Data Transaction Identification) เป็นวิธีการจัดข้อมูลให้อยู่ในรูปแบบที่เหมาะสมและถูกต้องเพื่อให้สามารถนำไปเข้าสู่กระบวนการจัดทำเหมืองข้อมูล วิธีนี้นิยมใช้ในการทำข้อมูลให้อยู่ในรูปแบบตาราง (Table) มีลักษณะเป็นแถวและสดมภ์หรือคอลัมน์ที่มีความสัมพันธ์กัน

6) การค้นหารูปแบบ (Pattern Discovery) เป็นวิธีการกำหนดรูปแบบในการวิเคราะห์และสืบค้นเพื่อให้ได้ผลลัพธ์ที่ต้องการ สามารถแบ่งเป็น รูปแบบการวิเคราะห์ กฎจำแนกเชิงความสัมพันธ์ การจำแนกประเภทข้อมูลและการพยากรณ์ การแบ่งกลุ่มข้อมูลแบบอัตโนมัติ และรูปแบบการทำงานตามลำดับ

7) การวิเคราะห์รูปแบบ (Pattern Analysis) เป็นการนำผลลัพธ์จากการสืบค้นข้อมูลที่สัมพันธ์กันที่ผ่านกระบวนการจัดทำเหมืองข้อมูลมาทำการวิเคราะห์เพื่อช่วยประกอบการตัดสินใจหรือการวางแผนการดำเนินงานต่างๆ ในอนาคต

3.1.1.2 เทคนิคการทำเหมืองข้อมูล (เอกสิทธิ์ พัทธวงค์ศักดิ์ดา, 2557, หน้า 15)

1) เทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เป็นรูปแบบการนำเอาข้อมูลมาวิเคราะห์โดยพิจารณาความสัมพันธ์ของชุดข้อมูลนั้น เพื่อนำเอารูปแบบผลลัพธ์จากการวิเคราะห์ข้อมูลที่ได้ไปใช้ประโยชน์โดยตรงกับชุดข้อมูลเดียวกัน รูปแบบการเรียนรู้แบบไม่มีผู้สอน คือ การวิเคราะห์ความสัมพันธ์ และการวิเคราะห์การจัดกลุ่ม

2) เทคนิคการเรียนรู้แบบมีผู้สอน (Supervised Learning) เป็นรูปแบบการรวบรวมเอาชุดข้อมูลในอดีต นำมาวิเคราะห์โดยพิจารณาความสัมพันธ์ หรือรูปแบบเฉพาะของชุดข้อมูลนั้นๆ เพื่อนำเอารูปแบบผลลัพธ์จากการวิเคราะห์ข้อมูลที่ได้ไปใช้ในการคาดการณ์ หรือทำนายสิ่งที่จะเกิดขึ้นในอนาคต เทคนิควิธีการวิเคราะห์ข้อมูลนี้คือการวิเคราะห์จำแนกประเภทข้อมูล

การศึกษาและพัฒนาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน

3.1.1.3 เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ต้นไม้ตัดสินใจถูกพัฒนาโดย Quinlan (1986, p. 1) เป็นวิธีการที่ได้รับความนิยมอย่างแพร่หลายเนื่องจากโมเดลที่ได้สามารถแปลความหมายและเข้าใจง่าย ลักษณะของรูปแบบข้อมูล (Pattern) ซึ่งแต่ละโหนด (Node) จะแสดงคุณลักษณะหรือแอตทริบิวต์ (Attribute) ที่ใช้ทดสอบข้อมูล แต่ละกิ่งจะแสดงผลในการทดสอบ และลีฟโหนด (Leaf Node) จะแสดงกลุ่มหรือคลาส (Class) ที่กำหนดไว้ อัลกอริทึมของเทคนิคต้นไม้ตัดสินใจส่วนใหญ่ไม่รองรับข้อมูลแบบต่อเนื่อง จึงต้องมีการแบ่งให้เป็นข้อมูลแบบไม่ต่อเนื่องเสียก่อน อัลกอริทึมสำหรับการแบ่งให้ข้อมูลเป็นแบบไม่ต่อเนื่อง คือ ID3, C4.5 และ C5.0 อัลกอริทึมของเทคนิคต้นไม้ตัดสินใจ C4.5 (J48) เป็นเทคนิคการสร้างตัวแบบต้นไม้ตัดสินใจ โดยการพิจารณาค่าংশี่ความเหมาะสมของแอตทริบิวต์ เรียกว่า Gini (Gini Index)

3.1.1.4 ตัวแบบจำแนกประเภทข้อมูลหลายๆ ตัว (Ensemble) ใช้ตัวแบบจำแนกประเภทข้อมูลหลายๆ ตัวมารวมกันตัดสินใจเพื่อบอกประเภทของข้อมูล โดยผลลัพธ์ของการตัดสินใจร่วมกันของตัวแบบจำแนกประเภทข้อมูลหลายๆ ตัวจะให้ความถูกต้องแม่นยำกว่าการใช้ตัวแบบเพียงตัวเดียว (สุรพงษ์ เอื้อวัฒนามงคล, 2559, หน้า. 67) วิธีแบบแรนดอมฟอเรส (Random Forest) เป็นวิธีการสุ่มข้อมูลตัวอย่างและการเพิ่มการสร้าง ความหลากหลายของโมเดลด้วยการสุ่มแอตทริบิวต์ ซึ่งเทคนิคที่ใช้ในการสร้างโมเดลก็เป็นเทคนิคต้นไม้ตัดสินใจ อย่างเดียว แม้ว่าจะเป็นเทคนิคต้นไม้ตัดสินใจเหมือนกันแต่ข้อมูลและแอตทริบิวต์ที่ใช้ในการสร้างโมเดลต่างกันก็ ทำให้โมเดลที่สร้างขึ้นมามีลักษณะที่ต่างกัน จากกระบวนการของต้นไม้ตัดสินใจ (Tree Classification Algorithm) ถูกสร้างขึ้นจนกลายเป็นป่า (Forest) จนกระทั่งวิเคราะห์การตัดสินใจจากต้นไม้แต่ละต้นที่อยู่ในป่า ขั้นตอนการทำงานของอัลกอริทึมต้นไม้หลายๆ ต้นตัดสินใจดังนี้ 1) กำหนดจำนวนข้อมูลเรียนรู้ n จำนวน 2) กำหนดจำนวนตัวแปร m ตัวเพื่อใช้ในการตัดสินใจโหนดของต้นไม้ 3) เลือกจำนวนข้อมูลมาเรียนรู้แล้วหาค่าความผิดพลาดจากการคาดเดาคลาส (3.1) สร้าง bootstrap จากกลุ่มตัวอย่างของ n และทำการสร้างต้นไม้จากกลุ่มตัวอย่าง bootstrap (3.2) เมื่อต้นไม้แต่ละโหนดมีการเจริญเติบโตแล้ว จะทำการสุ่มเลือกตัวแปร m เพื่อใช้ในการตัดสินใจในการจำแนกหาค่าความผิดพลาดและค่าที่ดีที่สุดออกมา 4) รวบรวมคะแนนโหวตจากต้นไม้ทุกต้นในป่า จากนั้นใช้การลงคะแนนเสียงส่วนใหญ่ในการตัดสินใจจากแต่ละคลาส

3.1.1.5 วิธีการจัดกลุ่มข้อมูลแบบเคย์มีน (K-Means) อัลกอริทึมเคย์มีนจะตัดแบ่ง (Partition) วัตถุออกเป็น K กลุ่ม โดยแทนแต่ละกลุ่มด้วยค่าเฉลี่ยของกลุ่ม ซึ่งใช้เป็นจุดศูนย์กลาง (Centroid) ของกลุ่มในการวัดระยะห่างของข้อมูลในกลุ่มเดียวกัน ในขั้นแรกของการจัดกลุ่มโดยการหาค่าเฉลี่ยแบบเคย์มีนต้องกำหนดจำนวนกลุ่ม (K) ที่ต้องการ และกำหนดจุดศูนย์กลางเริ่มต้นจำนวน K จุด สิ่งสำคัญในการกำหนดจุดศูนย์กลางเริ่มต้นของแต่ละกลุ่มนี้ ควรจะถูกกำหนดด้วยวิธีที่เหมาะสม เพราะตำแหน่งจุดศูนย์กลางเริ่มต้นที่แตกต่างกันทำให้ได้ผลลัพธ์สุดท้ายแตกต่างกัน ดังนั้นในทางที่ดีควรจะกำหนดจุดศูนย์กลางนี้ให้ห่างจากจุดศูนย์กลางอื่นๆ ขั้นตอนต่อไป คือ สร้างกลุ่มข้อมูลและความสัมพันธ์กับจุดศูนย์กลางที่ใกล้มากที่สุด โดยแต่ละจุดจะถูกกำหนดไปยังจุดศูนย์กลางที่ใกล้เคียงที่สุดจนครบหมดทุกจุด และคำนวณจุดศูนย์กลางใหม่ โดยการหาค่าเฉลี่ยทุกวัตถุที่อยู่ในกลุ่ม หากจุดศูนย์กลางในแต่ละกลุ่มถูกเปลี่ยนตำแหน่ง จะได้จุดมีความสัมพันธ์กับกลุ่มใหม่และใกล้กับจุดศูนย์กลางใหม่ ทำซ้ำแบบนี้ไปเรื่อยๆ จะสังเกตเห็นว่าผลลัพธ์จากการทำซ้ำแบบนี้ทำให้จุดศูนย์กลางเปลี่ยนตำแหน่งทุกรอบ จนกระทั่งจุดศูนย์กลางจำนวน K จุด ไม่มีการเปลี่ยนแปลงจึงจะสิ้นสุดกระบวนการ วิธีนี้จะต้องมีการกำหนดจำนวนกลุ่มข้อมูล (K) ล่วงหน้าซึ่งอาจเป็นข้อจำกัดอย่างหนึ่งของวิธีนี้ (สุรพงษ์ เอื้อวัฒนามงคล, 2559, หน้า. 80)

การศึกษาและพัฒนาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน

3.1.1.6 การหากฎความสัมพันธ์ (Association Rules) (Han, J. & Kamber, M, 2006) การค้นหากฎความสัมพันธ์ของข้อมูลเป็นการค้นหารูปแบบความสัมพันธ์ที่เกิดขึ้นร่วมกันในรูปแบบของความสัมพันธ์ในการเกิดความสัมพันธ์ที่ถูกเรียกว่าสิ่งที่ปรากฏเด่นชัด (Frequent Itemsets) และมีรูปแบบทั่วไปของกฎความสัมพันธ์อยู่ในลักษณะ $A \Rightarrow B$ โดยที่ A และ B คือ ชุดของข้อมูลที่เกิดขึ้นในรายการข้อมูล (Item Sets) สำหรับการวัดค่าความสำคัญของกฎที่ถูกค้นหาได้ ในชุดข้อมูลจะใช้ค่าความเชื่อมั่น (Confidence) และค่าสนับสนุน (Support) อัลกอริทึมอะพริออริ (Apriori) เป็นประเภทของการค้นหาความสัมพันธ์แบบ Boolean Association Rule ซึ่งเป็นอัลกอริทึมที่ใช้กันอย่างแพร่หลาย คือ เทคนิควิธีที่ใช้สำหรับค้นหาสิ่งที่ปรากฏเด่นชัดจากข้อมูลที่กำหนดโดยมีหลักการทํางาน คือ อัลกอริทึมอะพริออริ ทำหน้าที่สร้างเซตไอเท็มหรือกลุ่มข้อมูลที่ต้องการวิเคราะห์ที่เป็นไปได้ทั้งหมดที่มีค่าสนับสนุนมากกว่าค่าสนับสนุนขั้นต่ำโดยเป็นการทํางานในแบบล่างขึ้นบน (Bottom Up) โดยมีขั้นตอนดังนี้ 1) อ่านข้อมูลทั้งหมดและสร้างไอเท็มเซตที่ผ่านค่าสนับสนุนขั้นต่ำ ความยาว 1 ไอเท็ม (Frequent 1-Itemset) 2) สร้างเซตไอเท็มทดสอบ (Candidate Itemset) ที่มีความยาว 2 ไอเท็มจากเซตไอเท็มที่ปรากฏเด่นชัดความยาว 1 ไอเท็มในขั้นตอนแรกและนำไปหาค่าสนับสนุนเพื่อค้นหาไอเท็มเซตที่ปรากฏเด่นชัดความยาว 2 ไอเท็ม ขั้นตอนวิธีอะพริออริ จะวนรอบทํางานจนกระทั่งไม่พบไอเท็มเซตที่ผ่านค่าสนับสนุนขั้นต่ำจึงจบการทํางานไอเท็มเซตที่ผ่านค่าสนับสนุน ขั้นต่ำในแต่ละรอบคือสิ่งที่ปรากฏเด่นชัดจากข้อมูล

3.1.1.7 การประเมินตัวแบบการจำแนกข้อมูล การประเมินตัวแบบการจำแนกข้อมูลเป็นขั้นตอนสำคัญขั้นตอนหนึ่งในขบวนการทำเหมืองข้อมูล เพื่อให้ทราบประสิทธิภาพของตัวแบบที่ถูกสร้างขึ้นรวมทั้งสามารถเปรียบเทียบตัวแบบหลายตัว เพื่อเลือกตัวแบบที่ดีที่สุดในการนำไปใช้งานด้วยตาราง Confusion Matrix ซึ่งใช้ตัวชี้วัดต่อไปนี้ (สุรพงศ์ เอื้อวัฒนามงคล, 2559, หน้า 70)

1) Accuracy คือ ความแม่นยำในการจำแนกข้อมูล ได้แก่ สัดส่วนระหว่างจำนวนข้อมูลทั้งหมดที่จำแนกประเภทถูกต้องทั้งประเภท Positive และ Negative กับจำนวนข้อมูลทั้งหมดที่ถูกจำแนกประเภท

2) Precision คือ ความถูกต้องของการจำแนกประเภทข้อมูลว่าเป็น Positive ได้แก่ สัดส่วนที่ตัวแบบจำแนกประเภทข้อมูลแบบ Positive ได้ถูกต้องเทียบกับจำนวนข้อมูลทั้งหมดที่ถูกจำแนกประเภทว่าเป็น Positive

3) Recall คือ ความสามารถของตัวแบบในการจำแนกประเภทข้อมูลที่เป็น Positive เมื่อเทียบกับจำนวนข้อมูลที่เป็น Positive ทั้งหมด

4) F-measure เป็นตัวชี้วัดที่คำนวณจากค่า Precision และ Recall เพื่อให้สามารถวัดค่าทั้งสองได้พร้อมกันด้วยตัวชี้วัดเพียงตัวเดียว ค่า F-measure จะมีค่าเข้าใกล้หนึ่งถ้าทั้ง Recall และ Precision มีค่าเข้าใกล้หนึ่ง

3.1.2 งานวิจัยที่เกี่ยวข้อง

งานวิจัยของ ธีรพงษ์ สังข์ศรี (2557) ได้ศึกษาเรื่องการวิเคราะห์พฤติกรรมสำหรับการเลือกสมัครสาขาวิชาเรียนและการเปรียบเทียบตัวแบบพยากรณ์จำนวนนักศึกษาใหม่โดยใช้เทคนิคการทำเหมืองข้อมูล งานวิจัยชิ้นนี้ได้นำแนวความคิดการทำเหมืองข้อมูลมาช่วยในการวิเคราะห์ข้อมูลผู้สมัครเข้าศึกษาต่อโดยสามารถแบ่งส่วนการทำงานออกได้เป็น 2 ส่วนหลักประกอบด้วย 1) การวิเคราะห์พฤติกรรมสำหรับการเลือกสมัครสาขาวิชาเรียนเป็นการประยุกต์ใช้เทคนิคการจัดกลุ่มข้อมูลแบบ Simple K-means ด้วยเทคนิคอะพริออริ และ 2) การเปรียบเทียบ

การศึกษาและพัฒนาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน

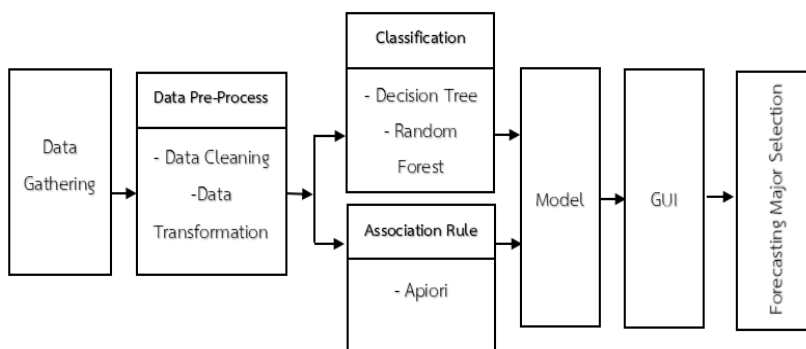
เทียบตัวแบบพยากรณ์จำนวนนักศึกษาใหม่ โดยตัวแบบพยากรณ์ที่ถูกสร้างขึ้นด้วยเทคนิคต้นไม้ช่วยตัดสินใจ และตัวแบบที่ถูกสร้างขึ้นด้วยเทคนิคโครงข่ายประสาทเทียมแบบเพอร์เซ็ปตรอนหลายชั้น

งานวิจัยของ อัจฉราภรณ์ จุฑาผาด (2559) ได้ศึกษาเรื่องการพัฒนาาระบบสารสนเทศเพื่อการพยากรณ์จำนวนนักศึกษาใหม่ โดยใช้กฎการจำแนกต้นไม้ตัดสินใจซึ่งงานวิจัยนี้มีวัตถุประสงค์ คือ 1) เพื่อพัฒนาระบบสารสนเทศเพื่อใช้ในการพยากรณ์จำนวนนักศึกษาใหม่ที่จะเข้าศึกษาต่อในระดับปริญญาตรี คณะบริหารธุรกิจและการบัญชี มหาวิทยาลัยราชภัฏร้อยเอ็ด โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจ และ 2) เพื่อวัดประสิทธิภาพและความถูกต้องของตัวต้นแบบที่พัฒนาขึ้นที่นำมาใช้ในการพยากรณ์จำนวนนักศึกษาใหม่ของคณะบริหารธุรกิจและการบัญชี มหาวิทยาลัยราชภัฏร้อยเอ็ด

งานวิจัยของ ชันทอง ประชุมชาติ และพิมรินทร์ คีรินทร์ (2558) ได้ศึกษาวิจัยเรื่องการวิเคราะห์พฤติกรรมกรรมการเลือกสมัครสาขาวิชาเรียนของนักศึกษาใหม่ โดยใช้เทคนิคการเหมืองข้อมูลวิเคราะห์พฤติกรรมกรรมการเลือกสมัครสาขาวิชาเรียนของนักศึกษาใหม่ ด้วยการใช้อัลกอริทึมการสมัครเข้าศึกษาต่อของมหาวิทยาลัยราชภัฏพิบูลสงครามเป็นกรณีศึกษา โดยใช้ซอฟต์แวร์ WEKA เทคนิคการจำลองการแบ่งกลุ่มด้วยขั้นตอนเคมีน เพื่อใช้ในการจัดกลุ่มของผู้สมัคร และสร้างตัวแบบจำลองการหากฎความสัมพันธ์ด้วยขั้นตอนออพริออริ เพื่อหากฎความสัมพันธ์ของพฤติกรรมกรรมการเลือกสาขาวิชาเรียนของผู้สมัคร

จากที่กล่าวมาแล้วงานวิจัยที่ศึกษาเกี่ยวกับพฤติกรรมกรรมการเลือกสมัครสาขาวิชาเรียนของนักศึกษาใหม่ ด้วยการแบ่งกลุ่มของข้อมูลและหากฎความสัมพันธ์ที่สอดคล้องกันมากที่สุด โดยใช้กรอบแนวคิดในการวิจัยที่ธีรพงษ์ สังข์ศรี (2557), ชันทอง ประชุมชาติ และพิมรินทร์ คีรินทร์ (2558) ได้นำเสนอด้วยเทคนิคการแบ่งกลุ่มของข้อมูล Simple K-Means ก่อนจากนั้นจะหากฎความสัมพันธ์ของข้อมูลด้วยขั้นตอนออพริออริ ในส่วนของการสร้างตัวแบบสำหรับนำไปพัฒนาระบบสารสนเทศเพื่อการพยากรณ์จำนวนนักศึกษาใหม่ที่จะเข้าศึกษา ให้ได้ค่าความแม่นยำที่สุด โดยใช้กรอบแนวคิดในการวิจัยที่ อัจฉราภรณ์ จุฑาผาด (2559) ได้นำเสนอด้วยเทคนิคต้นไม้ตัดสินใจ และจากการศึกษาพบว่า เทคนิคแรนดอมฟอเรสจะให้ประสิทธิภาพที่สูงกว่าการใช้เทคนิคต้นไม้ตัดสินใจ (สุรพงศ์ เอื้อวัฒน์มงคล, 2559)

3.2 กรอบแนวคิดในการวิจัย



ภาพที่ 2 กรอบแนวคิดในการวิจัย

การศึกษาและพัฒนาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน

นิยามศัพท์เฉพาะ

ตัวแบบพยากรณ์ หมายถึง ความสัมพันธ์ของคุณลักษณะของข้อมูลที่ได้จากวิธีการจำแนกด้วยเทคนิคเหมืองข้อมูลที่เปรียบเทียบผลการพยากรณ์ที่ดีที่สุดจากเทคนิคต้นไม้ตัดสินใจกับเทคนิคแรนดอมฟอเรส

การพยากรณ์ หมายถึง การทำนายเหตุการณ์ในอนาคต โดยนำข้อมูลการรับสมัครนักศึกษาใหม่ของมหาวิทยาลัยราชภัฏชัยภูมิตั้งแต่ปี 2559 - 2560 มาพยากรณ์อนาคตโดยใช้เทคนิคทางเหมืองข้อมูล

คุณลักษณะ หมายถึง องค์ประกอบของข้อมูลนักศึกษาที่สมัครเข้าสาขาวิชาเรียนซึ่งอยู่ในรูปของฟิลต์หรือคอลัมน์ ได้แก่ เพศ ผลการเรียน โรงเรียน หลักสูตรสมัคร 1 หลักสูตรสมัคร 2 วุฒิการศึกษา วิชาเอก และผลการเข้าศึกษา

ความเหมาะสม หมายถึง ค่าความถูกต้องของการตัวแบบการพยากรณ์อยู่ในระดับที่สูงที่สุด

เหมืองข้อมูล หมายถึง กระบวนการที่กระทำกับข้อมูล เพื่อค้นหารูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์

4. วิธีดำเนินการวิจัย

การศึกษาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียนโดยใช้เทคนิคเหมืองข้อมูล ผู้วิจัยได้ดำเนินการวิจัยตามขั้นตอนดังนี้

4.1 การจัดเตรียมข้อมูล

จัดเตรียมข้อมูลโดยใช้ข้อมูลการสมัครเข้าศึกษาต่อจากงานบริการการศึกษา มหาวิทยาลัยราชภัฏชัยภูมิ ตั้งแต่ปีการศึกษา 2559-2560 เป็นจำนวน 13,629 รายการ จากนั้นได้ทำการเตรียมข้อมูลดังนี้

1) ทำความสะอาดข้อมูล ด้วยการตัดรายการข้อมูลที่มีคุณลักษณะไม่ครบหรือไม่สมบูรณ์ทั้งไป เช่น วุฒิการศึกษาที่ไม่มีค่าหรือเป็นค่าที่ไม่สามารถอธิบายได้ จากข้อมูลทั้งหมด 13,690 รายการปรับลดลงเหลือจำนวนทั้งหมด 13,629 รายการ แสดงดังภาพที่ 3

31	59200030	08/01/2559	1	นาย	ธนพงษ์	ศาสร์	13/02/2538	21	1.2095E+12	ม.6	รับ-ดผล	2.78	โรงเรียนเทพารักษ์ราชวิทยา
32	59200031	08/01/2559	1	นาย	ธนพงษ์	ศาสร์	13/02/2538	21	1.2095E+12	ม.6	รับ-ดผล	2.78	โรงเรียนเทพารักษ์ราชวิทยา
33	59200032	01/10/2559	1	นาย	ธนากร	อภัย	16/06/2529	30	1.3601E+12	ม.6	ดีเยี่ยม-ทั่วไป	2.96	ศูนย์การศึกษาอิสระและก
34	59200033	01/11/2559	1	นาย	ศุภกฤต	วิเศษพิศล	06/05/2537	21	1.3099E+12	ม.6	ดีเยี่ยม-ทั่วไป	2.73	ศูนย์การศึกษาอิสระและก
35	59200034	01/11/2559	1	นางสาว	วิลาวรรณ	พินโสภา	06/10/2540	18	1.3022E+12	ม.6	รับ-ดผล	-	โรงเรียนหนองบุงมากประสงค
36	59200035	01/11/2559	1	นาย	เกียรติศักดิ์	สวัสดิ์	06/03/2538	20	1.3111E+12	ม.6	รับ-ดผล	3.09	ชุมพลวิทยาสรร
37	59200036	01/11/2559	1	นาย	พีเชษฐ	หาญนอก	19/12/2521	37	3.3005E+12	ปวส.	แผนกวิชาคอมพิวเตอร์	2.62	โรงเรียนเทคโนโลยีพระพลชน
38	59200037	01/11/2559	1	นาย	วิระ	ใจเที่ยง	06/07/2534	24	2.4101E+12	ม.6	ดีเยี่ยม-ทั่วไป	2.91	ศูนย์การศึกษาอิสระและก
39	59200038	13/01/2559	1	นางสาว	ดวงฤทัย	ชัยประสิทธิ์	24/08/2540	18	1.3699E+12	-	รับ-ดผล	2.88	สตรีชัยภูมิ
40	59200039	13/01/2559	1	นางสาว	ดวงใจ	ชัยประสิทธิ์	24/08/2540	18	1.3699E+12	-	รับ-ดผล	2.69	สตรีชัยภูมิ
41	59200040	13/01/2559	1	นาย	แสนภูมิ	สิงทองทิพย์	11/11/2538	20	1.3699E+12	ปวช.	แผนกวิชาช่างกลโรงงาน	3.07	วิทยาลัยเทคนิคชัยภูมิ
42	59200041	13/01/2559	1	นาย	ชาญวิทย์	บุญโชค	07/10/2540	18	1.3612E+12	ม.6	ดีเยี่ยม-ดีเยี่ยม	3.07	โรงเรียนพรตพิทยพยัต
43	59200042	14/01/2559	1	นางสาว	ปิยธิดา	กองโสกลาง	03/06/2540	18	1.3097E+12	ม.6	รับ-ดผล	2.95	โรงเรียนเมืองจ
44	59200043	15/01/2559	1	นาง	จริพร	ยอดแก่น	07/09/2530	28	1.4016E+12	ม.6	รับ-ดผล	2.5	โรงเรียนสตรีชัยภูมิ

ภาพที่ 3 ตัวอย่างข้อมูลที่ไม่สมบูรณ์

2) รวบรวมข้อมูลที่แยกการเก็บระหว่างนักศึกษาภาคปกติ ภาคการศึกษาเพื่อปวงชน ปีการศึกษา และรอบการเปิดรับสมัครนักศึกษา ซึ่งจะอยู่ในรูปแบบของ Excel File

3) คัดเลือกคุณลักษณะที่ต้องการเพื่อนำไปสู่วิธีการแปลงข้อมูลและจัดให้อยู่ในรูปแบบที่เหมาะสมต่อการวิเคราะห์ ผู้วิจัยได้ทำการคัดเลือกคุณลักษณะ ดังตารางที่ 1

การศึกษาและพัฒนาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน

ตารางที่ 1 คุณลักษณะที่คัดเลือกเพื่อการวิเคราะห์สำหรับการจำแนกข้อมูล

ดับที่	คุณลักษณะ	ตัวอย่างข้อมูล	ข้อมูลเดิม	ข้อมูลแทน
1	เพศ	นาย	นาย นางสาว นาง	M F F
2	ผลการเรียน	2.41	0.00 – 1.99 2.00 – 2.99 3.00 – 4.00	LOW MID GOOD
3	โรงเรียน	อำเภอเมือง จังหวัดชัยภูมิ	อำเภอเมือง จังหวัดชัยภูมิ อำเภออื่น จังหวัดชัยภูมิ อำเภอเมือง จังหวัดอื่น อำเภออื่น จังหวัดอื่น	S1 S2 S3 S4
4	หลักสูตรสมัคร 1	รัฐประศาสนศาสตร์	เทคโนโลยีการจัดการอุตสาหกรรม สาธารณสุขศาสตร์ วิทยาการคอมพิวเตอร์ ศิลปะและการออกแบบ การศึกษาปฐมวัย คณิตศาสตร์ ภาษาอังกฤษ คอมพิวเตอร์ศึกษา วิทยาศาสตร์ทั่วไป ภาษาไทย พลศึกษา รัฐประศาสนศาสตร์ การบริหารทรัพยากรมนุษย์ รัฐศาสตร์ การจัดการ คอมพิวเตอร์ธุรกิจ การเงินและการธนาคาร การจัดการธุรกิจค้าปลีก วิศวกรรมการผลิต	01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19

การศึกษาและพัฒนาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน

ตารางที่ 1 คุณลักษณะที่คัดเลือกเพื่อการวิเคราะห์สำหรับการจำแนกข้อมูล (ต่อ)

ลำดับที่	คุณลักษณะ	ตัวอย่างข้อมูล	ข้อมูลเดิม	ข้อมูลแทน
			วิศวกรรมเครื่องกล	20
			นิติศาสตร์	21
			สหวิทยาการเพื่อการพัฒนาท้องถิ่น	22
			พยาบาลศาสตร์	23
5	หลักสูตรสมัคร 2	การจัดการ	*** แทนข้อมูลลำดับที่ 4	
6	วุฒิการศึกษา	ม.6	สายสามัญ (ม.6) สายอาชีวะ (ปวช. ปวส.)	M R
7	วิชาเอก	วิทย์-คณิต	วิทยาศาสตร์ ศิลปศาสตร์ อื่น ๆ	SCI ART OTH
8	การเข้าศึกษา	ไม่ผ่าน	เข้าศึกษา ไม่เข้าศึกษา	Yes No

4.2 การวิเคราะห์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน

นำข้อมูลที่เตรียมไว้ตามคุณลักษณะตามตารางที่ 1 มาทำการจัดกลุ่มข้อมูลด้วยเทคนิค Simple K-Means จากนั้นทำการหาความสัมพันธ์ของข้อมูลด้วย อะพริออริอัลกอริทึม โดยใช้โปรแกรม WEKA จากข้อมูลทั้งหมดจำนวน 13,629 รายการ แยกเป็นประเภทการรับสมัครรอบโควตาจำนวน 4,072 รายการและประเภทการรับสมัครรอบทั่วไปจำนวน 9,557 รายการ นำแต่ละประเภทไปจัดกลุ่มข้อมูล และหาความสัมพันธ์ของข้อมูลจากกลุ่มที่จัดไว้แล้ว

4.3 การสร้างตัวแบบและการประเมินตัวแบบพยากรณ์จำนวนนักศึกษาใหม่

นำข้อมูลที่เตรียมไว้ตามคุณลักษณะตามตารางที่ 1 มาทำการจำแนกข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจ ภายใต้ัลกอริทึม C4.5 สร้างและทดสอบตัวแบบด้วยการแบ่งข้อมูล Cross Validation 10-Folds กำหนดค่าความเชื่อมั่นที่ 0.25 และเทคนิคแรนดอมฟอเรส สร้างและทดสอบตัวแบบด้วยการแบ่งข้อมูล Cross Validation 10-Folds โดยใช้โปรแกรม WEKA จากนั้นทำการเปรียบเทียบผลจากทั้งสองวิธีเพื่อเลือกตัวแบบที่ให้ผลการพยากรณ์ที่ดีที่สุด

4.4 การพัฒนาระบบพยากรณ์การเข้าศึกษาของนักศึกษาใหม่

นำตัวแบบการพยากรณ์จากข้อ 4.3 โดยทำการบันทึกตัวแบบให้อยู่ในรูปแบบของไฟล์นามสกุล .model มาพัฒนาระบบการพยากรณ์การเข้าศึกษาของนักศึกษาใหม่ ในรูปแบบเว็บแอปพลิเคชัน ด้วยเครื่องมือพัฒนาระบบภาษา JAVA

การศึกษาและพัฒนาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน

4.5 การประเมินความคิดเห็นของผู้เชี่ยวชาญ

แบบสอบถามประเมินความคิดเห็นของผู้เชี่ยวชาญจำนวน 5 คน เป็นอาจารย์ที่รับผิดชอบหลักสูตรที่มีการเลือกสมัครเรียนตามผลการหาความสัมพันธ์ของข้อมูลการเลือกสมัครเรียนที่น่าสนใจ ได้แก่ หลักสูตรพยาบาลศาสตร์ หลักสูตรสาธารณสุขศาสตร์ หลักสูตรวิทยาศาสตร์ทั่วไป หลักสูตรรัฐประศาสนศาสตร์ หลักสูตรภาษาอังกฤษ ที่มีต่อความเหมาะสมของกฎความสัมพันธ์

5. ผลการวิจัยและอภิปรายผล

5.1 ผลการวิจัย

5.1.1 ผลการวิเคราะห์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน ผลการวิเคราะห์ที่แสดงได้ดังต่อไปนี้ โดยผลการจัดกลุ่มข้อมูลแบ่งออกเป็น 2 ประเภท

1) ประเภทการรับสมัครรอบโควตาทำการจัดกลุ่มข้อมูลด้วย Simple K-Means ผลออกมาเป็นกลุ่มย่อยๆ ดังตารางที่ 2

ตารางที่ 2 การจัดกลุ่มข้อมูลประเภทการรับรอบโควตา

กลุ่มที่	คุณสมบัติ
1	เพศหญิง วิชาเอกวิทยาศาสตร์ ผลการเรียนระดับกลาง สถานศึกษาอยู่นอกอำเภอเมืองและจังหวัดชัยภูมิ
2	เพศหญิง วิชาเอกวิทยาศาสตร์ ผลการเรียนระดับดี สถานศึกษาอยู่นอกอำเภอเมืองและจังหวัดอื่น
3	เพศชาย วิชาเอกวิทยาศาสตร์ ผลการเรียนระดับดี สถานศึกษาอยู่นอกอำเภอเมืองและจังหวัดอื่น

2) ประเภทการรับสมัครรอบทั่วไปทำการจัดกลุ่มข้อมูลด้วย Simple K-Means ผลออกมาเป็นกลุ่มย่อยๆ ดังตารางที่ 3

ตารางที่ 3 การจัดกลุ่มข้อมูลประเภทการรับรอบทั่วไป

กลุ่มที่	คุณสมบัติ
4	เพศหญิง วิชาเอกวิทยาศาสตร์ ผลการเรียนระดับกลาง สถานศึกษาอยู่นอกอำเภอเมืองและจังหวัดอื่น
5	เพศหญิง วิชาเอกวิทยาศาสตร์ ผลการเรียนระดับดี สถานศึกษาอยู่นอกอำเภอเมืองและจังหวัดอื่น
6	เพศหญิง วิชาเอกศิลปศาสตร์ ผลการเรียนระดับดี สถานศึกษาอยู่นอกอำเภอเมืองและจังหวัดอื่น

5.1.2 ผลการหาความสัมพันธ์ของข้อมูลการเลือกสมัครเรียนที่แตกต่างกันแยกตามกลุ่มดังต่อไปนี้

1) การเลือกสมัครเรียนสาขาวิชา กลุ่มที่ 1 ได้จำนวนกฎทั้งสิ้น 27 กฎ ซึ่งมีกฎที่น่าสนใจต่อไปนี้ คือ ถ้าประเภทการสอบเข้าเป็นโควตา ผู้สมัครมีคุณลักษณะวิชาเอกวิทยาศาสตร์ จะมีการเลือกสมัครสาขาพยาบาลศาสตร์ ด้วยความเชื่อมั่นร้อยละ 100 และถ้าประเภทการสอบเข้าเป็นโควตา ผู้สมัครมีคุณลักษณะวิชาเอกวิทยาศาสตร์ และผลการเรียนอยู่ในระดับปานกลาง จะมีการเลือกสมัครสาขาสาธารณสุขศาสตร์ ด้วยความเชื่อมั่นร้อยละ 95

การศึกษาและพัฒนาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน

2) การเลือกสมัครเรียนสาขาวิชา กลุ่มที่ 2 ได้จำนวนกฎทั้งสิ้น 73 กฎ ซึ่งมีกฎที่น่าสนใจต่อไปนี้เป็น 3 คือ ถ้าประเภทการสอบเข้าเป็นโควตา ผู้สมัครมีคุณลักษณะวิชาเอกวิทยาศาสตร์ ผลการศึกษาอยู่ในระดับดี และเพศหญิง จะมีการเลือกสมัครสาขาพยาบาลศาสตร์ ด้วยความเชื่อมั่นร้อยละ 100 และถ้าประเภทการสอบเข้าเป็นโควตา ผู้สมัครมีคุณลักษณะวิชาเอกวิทยาศาสตร์ และเพศหญิง จะมีการเลือกสมัครสาขาพยาบาลศาสตร์ ด้วยความเชื่อมั่นร้อยละ 95

3) การเลือกสมัครเรียนสาขาวิชา กลุ่มที่ 3 ได้จำนวนกฎทั้งสิ้น 174 กฎ ซึ่งมีกฎที่น่าสนใจต่อไปนี้เป็น 3 คือ ถ้าประเภทการสอบเข้าเป็นโควตา ผู้สมัครมีคุณลักษณะวิชาเอกวิทยาศาสตร์ และผลการศึกษาอยู่ในระดับดี จะมีการเลือกสมัครสาขาวิทยาศาสตร์ทั่วไป ด้วยความเชื่อมั่นร้อยละ 100 และถ้าประเภทการสอบเข้าเป็นโควตา ผู้สมัครมีคุณลักษณะวิชาเอกวิทยาศาสตร์ ผลการศึกษาอยู่ในระดับดีและเพศหญิง จะมีการเลือกสมัครสาขาวิทยาศาสตร์ทั่วไป ด้วยความเชื่อมั่นร้อยละ 100

4) การเลือกสมัครเรียนสาขาวิชา กลุ่มที่ 4 ได้จำนวนกฎทั้งสิ้น 34 กฎ ซึ่งมีกฎที่น่าสนใจต่อไปนี้เป็น 3 คือ ถ้าประเภทการสอบเข้าเป็นรอบทั่วไป ผู้สมัครมีคุณลักษณะวิชาเอกวิทยาศาสตร์ ผลการศึกษาอยู่ในระดับปานกลางและเพศหญิง จะมีการเลือกสมัครสาขาพยาบาลศาสตร์ ด้วยความเชื่อมั่นร้อยละ 100 และถ้าประเภทการสอบเข้าเป็นรอบทั่วไป ผู้สมัครวิชาเอกวิทยาศาสตร์และผลการเรียนอยู่ในระดับปานกลาง จะมีการเลือกสมัครสาขาพยาบาลศาสตร์ ด้วยความเชื่อมั่นร้อยละ 100

5) การเลือกสมัครเรียนสาขาวิชา กลุ่มที่ 5 ได้จำนวนกฎทั้งสิ้น 83 กฎ ซึ่งมีกฎที่น่าสนใจต่อไปนี้เป็น 3 คือ ถ้าประเภทการสอบเข้าเป็นรอบทั่วไป ผู้สมัครมีคุณลักษณะวิชาเอกวิทยาศาสตร์ และผลการศึกษาอยู่ในระดับดี จะมีการเลือกสมัครสาขาพยาบาลศาสตร์ ด้วยความเชื่อมั่นร้อยละ 100 และถ้าประเภทการสอบเข้าเป็นรอบทั่วไป ผู้สมัครมีผลการศึกษาอยู่ในระดับดี วิชาเอกวิทยาศาสตร์ และเพศหญิง จะมีการเลือกสมัครสาขาพยาบาลศาสตร์ ด้วยความเชื่อมั่นร้อยละ 100

6) การเลือกสมัครเรียนสาขาวิชา กลุ่มที่ 6 ได้จำนวนกฎทั้งสิ้น 42 ซึ่งมีกฎที่น่าสนใจต่อไปนี้เป็น 3 คือ ถ้าประเภทการสอบเข้าเป็นรอบทั่วไป ผู้สมัครมีคุณลักษณะวิชาเอกศิลปศาสตร์ จะมีการเลือกสมัครสาขารัฐประศาสนศาสตร์ ด้วยความเชื่อมั่นร้อยละ 100 และถ้าประเภทการสอบเข้าเป็นรอบทั่วไป ผู้สมัครมีผลการศึกษาอยู่ในระดับดี และเพศหญิง จะมีการเลือกสมัครสาขาภาษาอังกฤษ ด้วยความเชื่อมั่นร้อยละ 97

จากกฎความสัมพันธ์การเลือกสมัครเรียนสาขาวิชาที่ได้สอดคล้องกับความคิดเห็นของผู้เชี่ยวชาญ ซึ่งได้ให้ความเหมาะสมอยู่ในระดับมากที่สุด ดังตารางที่ 4

ตารางที่ 4 ผลการประเมินความเหมาะสมของความสัมพันธ์การเลือกสมัครเรียนสาขาวิชาโดยผู้เชี่ยวชาญ

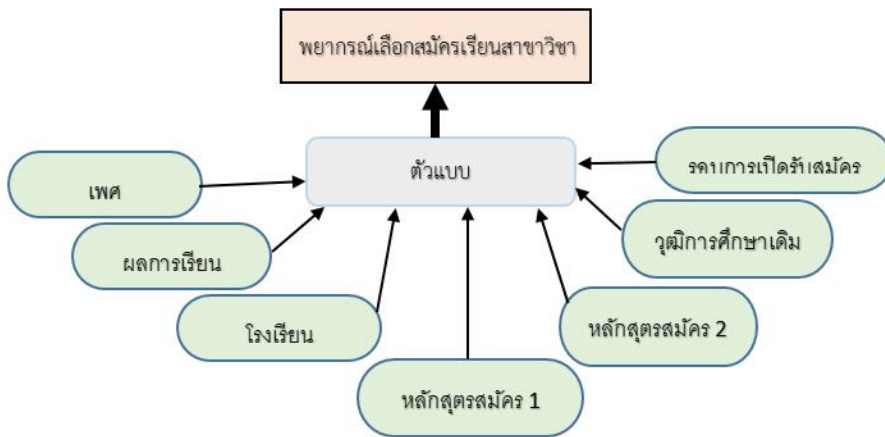
รายการประเมิน	\bar{X}	S.D.	ความหมาย
คุณลักษณะที่คัดเลือกเพื่อการวิเคราะห์สำหรับการจัดกลุ่มข้อมูล	4.80	0.48	มากที่สุด
การจัดกลุ่มข้อมูลประเภทการรับรองโควตาและรอบทั่วไป	4.20	0.48	มาก
จำนวนกลุ่มที่จัดแบ่งออกเป็นจำนวน 6 กลุ่ม	4.60	0.55	มากที่สุด
ผลของความสัมพันธ์ของข้อมูลการสมัครเรียนสาขาวิชา	4.60	0.55	มากที่สุด
เฉลี่ยรวม	4.55	0.32	มากที่สุด

การศึกษาและพัฒนาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน

จากตารางที่ 4 พบว่า คุณลักษณะที่คัดเลือกเพื่อการวิเคราะห์สำหรับการจัดกลุ่มข้อมูล จำนวนกลุ่มที่จัดแบ่งออกเป็นจำนวน 6 กลุ่มและผลของความสัมพันธ์ของข้อมูลการสมัครเรียนสาขาวิชา มีระดับความเหมาะสมมากที่สุด ส่วนการจัดกลุ่มข้อมูลประเภทการรับรองโควตาและรอบทั่วไป มีระดับความเหมาะสมมาก ตามลำดับ

5.1.3 ผลการสร้างตัวแบบและการประเมินตัวแบบพยากรณ์การเข้าศึกษาของนักศึกษาใหม่

นำข้อมูลที่เตรียมไว้ มาทำการจำแนกข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจและเทคนิคแรนดอมฟอเรส แบ่งข้อมูลออกเป็น 2 ส่วน 1) ส่วนเรียนรู้ (Training Set) และ 2) ส่วนทดสอบ (Test Set) แบบ Cross Validation 10-Folds จากนั้นทำการเปรียบเทียบผลการจำแนกข้อมูลจากทั้งสองตัวแบบ ได้ผลดังภาพที่ 4 และตารางที่ 5



ภาพที่ 4 ตัวแบบพยากรณ์

ตารางที่ 5 ตารางการเปรียบเทียบผลการวิเคราะห์ค่าความถูกต้องของแต่ละตัวแบบ

ที่	Algorithms	Accuracy (%)	RMSE	MAE
1	ต้นไม้ตัดสินใจ	73.61	0.43	0.34
2	แรนดอมฟอเรส	74.67	0.41	0.31

จากตารางที่ 5 พบว่าตัวแบบการพยากรณ์การเข้าศึกษาต่อด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอเรสให้ค่าความถูกต้อง (Accuracy) สูงที่สุดร้อยละ 74.67 ตามด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจให้ค่าความถูกต้องร้อยละ 73.61 มีความคลาดเคลื่อนเฉลี่ยกำลังสอง (RMSE) เท่ากับ 0.41 และ 0.43 มีค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ (MAE) เท่ากับ 0.34 และ 0.31 ตามลำดับ เปรียบเทียบค่าความถูกต้องและนำไปใช้ในการดำเนินการในขั้นตอนต่อไป

5.1.4 ผลการพัฒนากระบวนการพยากรณ์การเข้าศึกษาของนักศึกษาใหม่

จากการสร้างตัวแบบและทดสอบตัวแบบการพยากรณ์ผู้เข้าศึกษาต่อด้วยเทคนิคแรนดอมฟอเรส เป็นวิธีที่ให้ค่าประสิทธิภาพสูงกว่าตัวแบบการพยากรณ์อื่น นำตัวแบบแรนดอมฟอเรสดังกล่าวไปใช้ในการพยากรณ์ในระบบพยากรณ์การเข้าศึกษาของนักศึกษาใหม่ โดยผลการพยากรณ์จะจำแนกผู้สมัครแต่ละรายว่าสามารถสอบผ่านหรือสอบไม่ผ่านการคัดเลือกเป็นผู้เข้าศึกษาต่อ แสดงตัวอย่างดังภาพที่ 5

การศึกษาและพัฒนาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน

ภาพที่ 5 หน้าจอแสดงผลการพยากรณ์การเข้าศึกษาของนักศึกษาใหม่

5.1.5 องค์ความรู้ใหม่

จากผลการจัดกลุ่มของการเลือกสมัครสาขาวิชาเรียนจะพบว่านักเรียนที่เรียนในเขตที่ตั้งของโรงเรียนที่อยู่นอกอำเภอเมืองของจังหวัดชัยภูมิและจังหวัดอื่น เรียนวิชาเอกวิทยาศาสตร์ จะสมัครเข้าเรียนสาขาวิชาพยาบาลศาสตร์เป็นลำดับต้นๆ ส่วนสาขาวิชารัฐประศาสนศาสตร์จะเป็นนักเรียนที่เรียนวิชาเอกศิลปศาสตร์เป็นส่วนใหญ่ ซึ่งโดยรวมแล้วผลการเรียนจะอยู่ในระดับดี ดังนั้นมหาวิทยาลัยควรจะเน้นการวางแผนแนะแนวประชาสัมพันธ์เป้าหมายหลักไปยังโรงเรียนที่อยู่นอกเขตอำเภอเมืองของจังหวัดชัยภูมิและจังหวัดอื่น

5.2 อภิปรายผล

การศึกษาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียนโดยใช้เทคนิคเหมืองข้อมูล ผู้วิจัยได้สอบถามความเหมาะสมของความสัมพันธ์การเลือกสมัครเรียนสาขาวิชา กับ กฎความสัมพันธ์จากผู้เชี่ยวชาญ จำนวน 5 คน พบว่า โดยภาพรวมอยู่ในระดับมากที่สุด ซึ่งผู้วิจัยและผู้เชี่ยวชาญมีความเห็นไปในทางเดียวกันสอดคล้องกับ ชันทอง ประชุมชาติ และพิมรินทร์ ศิริรินทร์ (2558) ได้ทำวิจัยการวิเคราะห์พฤติกรรมเลือกสมัครสาขาวิชาเรียนของนักศึกษาใหม่ โดยใช้เทคนิคการเหมืองข้อมูล แสดงให้เห็นถึงพฤติกรรมเลือกสมัครสาขาวิชาเรียนของผู้สมัคร และผลเปรียบเทียบตัวแบบการพยากรณ์ด้วยเทคนิค แรנדอมฟอเรสต์ค่าความถูกต้องสูงกว่าเทคนิคต้นไม้ตัดสินใจเพียง 1.06 % สอดคล้องกับ ชีรพงษ์ สังข์ศรี (2557) ได้ทำวิจัยเปรียบเทียบตัวแบบพยากรณ์จำนวนนักศึกษาใหม่พบว่าเทคนิคต้นไม้ตัดสินใจให้ค่าความถูกต้องสูงกว่าเทคนิคเพอร์เซ็ปตรอนหลายชั้นสอดคล้องกับ อัจฉราภรณ์ จุฑาผาด (2559) ได้ทำวิจัยการพัฒนาาระบบสารสนเทศเพื่อการพยากรณ์จำนวนนักศึกษาใหม่โดยใช้กฎการจำแนกต้นไม้ตัดสินใจพบว่าระดับค่าความถูกต้องอยู่ที่มากที่สุด และสอดคล้องกับ สุรพงศ์ เอื้อวัฒนามงคล (2559) ที่กล่าวว่าเทคนิคแรนดอมฟอเรสต์เป็นเทคนิคที่นำต้นไม้ตัดสินใจหลาย ๆ ต้นเข้ามาช่วยในการจำแนกข้อมูลยอมให้ประสิทธิภาพดีกว่าการใช้ต้นไม้ตัดสินใจเพียงต้นเดียว

6. สรุปผลการวิจัย

ผลการศึกษาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียนโดยใช้เทคนิคเหมืองข้อมูลใช้ข้อมูลการสมัครเข้าศึกษาต่อจากงานบริการการศึกษา มหาวิทยาลัยราชภัฏชัยภูมิ ตั้งแต่ปีการศึกษา 2559-2560 เป็นจำนวน 13,629 รายการ ทำการจัดกลุ่มข้อมูลแยกเป็นประเภทการรับสมัครรอบ โควตาจำนวน 4,072 รายการ และประเภทการรับสมัครรอบทั่วไปจำนวน 9,557 รายการ ซึ่งแต่ละประเภทการรับสมัครนำไปกำหนดจำนวนกลุ่มด้วย Simple K-Means นำแต่ละกลุ่มไปหาความสัมพัทธ์ของข้อมูลจากทั้งหมด 6 กลุ่ม พร้อมกับเลือกความสัมพัทธ์ของข้อมูลที่นำเสนอกลุ่มละ 2 กฎ นำกฎความสัมพัทธ์ของข้อมูลไปประเมินความเหมาะสมจากผู้เชี่ยวชาญทั้งหมด 5 คน พบว่า ความเหมาะสมในภาพรวมอยู่ในระดับมากที่สุดและสร้างตัวแบบพยากรณ์การเข้าศึกษาของนักศึกษาด้วย 2 เทคนิค คือ เทคนิคต้นไม้ตัดสินใจ และเทคนิค แรนดอมฟอเรสแล้วประเมินประสิทธิภาพตัวแบบจากทั้ง 2 เทคนิค พบว่า เทคนิคแรนดอมฟอเรสให้ประสิทธิภาพในการพยากรณ์สูงที่สุด คือ ให้ค่าความถูกต้องร้อยละ 74.67 ค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง 0.41 และค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ 0.31 เมื่อได้ตัวแบบการพยากรณ์ที่มีประสิทธิภาพสูงที่สุดนำไปพัฒนาระบบสารสนเทศสำหรับพยากรณ์การเข้าศึกษาต่อในสาขาวิชาที่ได้สมัครเรียน โดยสามารถนำไปช่วยในการพยากรณ์ผู้สมัครเข้าศึกษาต่อในระดับปริญญาตรีของสถาบันการศึกษาได้ และสามารถนำไปพัฒนาต่อไปเพื่อให้มีประสิทธิภาพดียิ่งขึ้น

7. ข้อเสนอแนะ

7.1 ข้อเสนอแนะจากการวิจัย

การศึกษาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียนโดยใช้เทคนิคเหมืองข้อมูล จากกฎความสัมพัทธ์สามารถนำมาใช้ในการวางแผนสำหรับการแนะแนวการศึกษา และระบบสารสนเทศสำหรับการพยากรณ์การเลือกสมัครสาขาวิชาเรียนจะช่วยในการพยากรณ์การเข้าศึกษาของนักศึกษาที่จะเข้าศึกษาในสาขาวิชาได้ สาขาวิชาจะได้เตรียมแผนการในการรับนักศึกษาให้มีจำนวนเพิ่มขึ้นต่อไป

7.2 ข้อเสนอแนะในการวิจัยครั้งต่อไป

การทำวิจัยในครั้งต่อไป ควรพิจารณาเทคนิคการสร้างตัวแบบการพยากรณ์ที่หลากหลาย การปรับแต่งค่าพารามิเตอร์ และจำนวนข้อมูล เพื่อให้การพยากรณ์มีความถูกต้องและแม่นยำมาก ยิ่งขึ้น

8. กิตติกรรมประกาศ

การวิจัยในครั้งนี้ สำเร็จลุล่วงได้ด้วยดี เพราะความกรุณาและอนุเคราะห์ช่วยเหลือเป็นอย่างดีจากงานบริการการศึกษา มหาวิทยาลัยราชภัฏชัยภูมิ ที่ให้ความกรุณาด้านอุปกรณ์และด้านข้อมูลเป็นที่ปรึกษา

การศึกษาและพัฒนาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน

แนะนำ แนวทางเป็นอย่างดี แนวทางการดำเนินการวิจัย จึงทำให้งานวิจัยครั้งนี้เสร็จสมบูรณ์ไปได้ด้วยดี ผู้วิจัยจึงกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอกราบขอบพระคุณ บิดา มารดา บุรพจารย์ของผู้วิจัยทุกท่าน ซึ่งไม่อาจกล่าวได้หมดตลอดจนผู้ที่ยกยอช่วยเหลือให้กำลังใจเอาใจใส่ดูแลทุกอย่าง จนทำให้งานวิจัยนี้สำเร็จ ซึ่งความดีงามทั้งหมดทั้งมวลขอขอบมอบให้แก่ทุกท่าน

9. เอกสารอ้างอิง

ชั้นทอง ประชুমชาติ และพิมรินทร์ คีรินทร์. (2558). การวิเคราะห์พฤติกรรมกรรมการเลือกสมัครสาขาวิชาเรียนของนักศึกษาใหม่ โดยใช้เทคนิคการเหมืองข้อมูล. *การประชุมวิชาการ Kamphaeng Phet Rajabhat University National Conference*. ครั้งที่ 2, กำแพงเพชร. 174-185.

จามรกุล เหล่าเกียรติกุล. (2558). *เหมืองข้อมูลเบื้องต้น*. กรุงเทพมหานคร: แดเน็กซ์ อินเทอร์เน็ตโปรดักชั่น.

ธีรพงษ์ สังข์ศรี. (2557). การวิเคราะห์พฤติกรรมสำหรับการเลือกสมัครสาขาวิชาเรียนและการเปรียบเทียบ ตัวแบบพยากรณ์จำนวนนักศึกษาใหม่โดยใช้เทคนิคการทำเหมืองข้อมูล. *National Conference on Computing and Information Technology*, 10(1), 963-968.

ธเนศน์ นุ่มมัน. (2561, 04 เมษายน). วิฤติมหาลัยใหญ่ที่ต้องเร่งปรับตัว. *โพสต์ทูเดย์*. สืบค้นเมื่อ 4 เมษายน 2561, จาก <https://www.posttoday.com/politic/report/546661>

สุรพงศ์ เอื้อวัฒนามงคล. (2559). *การทำเหมืองข้อมูล*. กรุงเทพฯ: สถาบันบัณฑิตพัฒนบริหารศาสตร์.

สำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ. (2561). *แนวโน้มอัตราการเกิดของประชากรไทยตั้งแต่ปี 2536-2560*. เข้าถึงได้จาก <http://social.nesdb.go.th>

อัจฉราภรณ์ จุฑาผาด. (2559). การพัฒนาระบบสารสนเทศเพื่อการพยากรณ์จำนวนนักศึกษาใหม่โดยใช้กฎ การจำแนกต้นไม้ตัดสินใจ. *นเรศวรวิจัย*, 12(1), 267-278.

เอกสิทธิ์ พัทธวงศ์ศักดิ์. (2557). *การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมนิงเบื้องต้น*. กรุงเทพฯ: ดาต้า คิวบ์.

Han, J. & Kamber, M. (2006). *Data Mining Concepts and Techniques*. San Francisco, CA : Morgan Kaufmann Publishers.

Quinlan, J.R. (1986). Induction of Decision Tree. *Machine Learning*, 1(1), 81-106.

การศึกษาและพัฒนาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียน

ประวัติแนบท้ายบทความ



Name and Surname: Samran Wanon
 Highest Education: Master of Science (Applied Statistics and Information Technology)
 Major Computer Science
 University or Agency: National Institute of Development Administration
 Field of Expertise: Information Technology
 Address: 288 Tambon Nafai Amphur Muang Chaiyaphum 36000
 E-mail: samran@cpru.ac.th



Name and Surname: Rojjana Muangsan
 Highest Education: Master of Business (Business Administration)
 University or Agency: Ramkhamhaeng University
 Expertise: Digital Marketing, Business
 Address: 39 Nong Kham Tambon Ta Noen Amphur Noen Sa-nga
 Chaiyaphum 36130
 E-mail: rojjana@cpru.ac.th