

การพยากรณ์การต้อออกของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏพิบูลสงคราม ด้วยเทคนิคเหมืองข้อมูล

Forecasting Dropout of Undergraduates Pibulsongkram Rajabhat University with Data Mining Technique

ธนพร คล้ายทอง¹, ชุตินันท์ ศรีสวัสดิ์^{1*}

Tanaphorn Klaythong¹, Chutipphon Srisawat^{1*}

¹ สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏพิบูลสงคราม พิษณุโลก 65000 ประเทศไทย

¹ Department of Computer Science, Faculty of Science and Technology, Pibulsongkram Rajabhat University, Phitsanulok 65000, Thailand

* Corresponding Author: Chutipphon Srisawat, chutipphon@psru.ac.th

Received:

26 February 2022

Revised:

31 March 2022

Accepted:

27 April 2022

คำสำคัญ:

การต้อออก, เทคนิคต้นไม้ตัดสินใจ, เทคนิคนาอิวเบย์, กฎการอุปนัย

Keywords:

Fall off, Decision Tree, Naïve Bayes, Rule Induction

บทคัดย่อ: งานวิจัยนี้มีวัตถุประสงค์เพื่อ 1) วิเคราะห์คุณลักษณะและศึกษาปัจจัยที่เกี่ยวข้องในการต้อออกของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏพิบูลสงคราม 2) สร้างโมเดลและเปรียบเทียบประสิทธิภาพของโมเดล และ 3) เพื่อพัฒนาระบบการพยากรณ์การต้อออกของนักศึกษา โดยรวบรวมข้อมูลของนักศึกษาระดับปริญญาตรีจากกองบริการการศึกษา มหาวิทยาลัยราชภัฏพิบูลสงคราม ตั้งแต่ปีการศึกษา 2558-2563 จำนวนทั้งสิ้น 20,093 ชุดข้อมูล มีแอททริบิวต์ทั้งหมด 16 แอททริบิวต์ วิธีการคัดเลือกคุณลักษณะ (Feature Selection) ด้วยเทคนิค Filter Approach โดยใช้ค่า Information Gain ถูกนำมาประยุกต์ใช้เพื่อการวิเคราะห์หาปัจจัยที่มีผลต่อการต้อออกของนักศึกษา พบว่ามีปัจจัยที่เกี่ยวข้องในการต้อออกของนักศึกษาจำนวน 10 ปัจจัย ได้แก่ เกรดเฉลี่ยรวม ปีที่เข้าศึกษา สาขาวิชา หลักสูตร อาชีพมารดา อาชีพบิดา คณะ ระดับการศึกษา ภาค และความถนัด/ความสามารถพิเศษ จากนั้นนำปัจจัยที่เกี่ยวข้องไปสร้างแบบจำลอง โดยใช้เทคนิคเหมืองข้อมูล 3 เทคนิค คือ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคนาอิวเบย์ (Naïve Bayes) และกฎการอุปนัย (Rule Induction) ในการทดสอบประสิทธิภาพของโมเดล ด้วยวิธีการ 5-Fold Cross-Validation และ 10-Fold Cross-Validation ซึ่งวัดประสิทธิภาพแบบจำลองด้วยค่าความถูกต้อง (Accuracy) และค่าความผิดพลาด (Mean Absolute Error: MAE) ผลการทดสอบประสิทธิภาพพบว่าแบบจำลองที่ใช้เทคนิคต้นไม้ตัดสินใจได้ค่าสูงที่สุดในการแบ่งข้อมูลทดสอบ 10-Fold Cross-Validation ให้ค่าความถูกต้อง (Accuracy) 97.81% และค่าความผิดพลาด (MAE) เท่ากับ 0.026 ผู้วิจัยจึงได้ศึกษาออกแบบ และพัฒนาระบบ

ในรูปแบบเว็บแอปพลิเคชัน ผลการประเมินระบบจากการสุ่มตัวอย่างข้อมูลมาทดสอบในระบบ ค่าเฉลี่ยรวมของการทำนายถูกต้องของระบบคิดเป็นร้อยละ 86.29%

Abstract: The objectives of this research were 1) to analyze characteristics and study factors related to the dropout of undergraduate students. Pibulsongkram Rajabhat University 2) to create a model and compare the efficiency of the model and 3) to develop a forecasting system for student dropout. The data of undergraduate students were obtained from the Education Services Division, Pibulsongkram Rajabhat University, between the 2015 and 2020 academic years, a total of 20,093 data sets containing 16 attributes. The filter approach feature selection method based on information gain was presented to analyze the factors affecting student dropout. It was found that ten factors influence student dropout, namely total GPA, year attended in university, programs of study, course of study, mother's occupation, father's occupation, faculty, educational level, region, and aptitude/talent. Then, the relevant factors were used to create the model using 3 data mining techniques: Decision Tree, Naïve Bayes, and Rule Induction technique. The model's performance was investigated using the 5-Fold Cross-Validation and 10-Fold Cross-Validation methods. The model's accuracy and mean absolute error (MAE) were also quantified. The results showed that the decision tree technique model had the highest value in the 10-Fold Cross-Validation. The accuracy and MAE of the model were 97.81% and 0.026, respectively. The researcher also designed and developed the model system using the form of a web application. The results from system evaluation data found the total mean of the correct prediction of the system. Accounted for 86.29%

1. บทนำ

ในปัจจุบันการศึกษาถือว่าเป็นกระบวนการพัฒนาให้คนในสังคมเกิดการเรียนรู้ ความสามารถในการนำไปพัฒนาอาชีพ สังคมและประเทศชาติ ซึ่งการศึกษาในระดับอุดมศึกษา โดยเฉพาะอย่างยิ่งการศึกษาในระดับปริญญาตรี หากนักศึกษาไม่มีความพร้อม ไม่มีความเข้าใจในสาขาที่ตนเองเลือก อาจส่งผลให้เกิดการต้อออกหรือการพ้นสภาพนักศึกษา ทำให้เสียเวลา เสียค่าใช้จ่ายเป็นจำนวนมากโดยเปล่าประโยชน์ อีกทั้งส่งผลไปถึงสถานศึกษาในการประเมินคุณภาพการศึกษาของมหาวิทยาลัย ในการกำหนดหลักสูตร การจัดทำแผนพัฒนาการศึกษา ทำให้เสียเวลาการบริหารจัดการและเสียทรัพยากรในการลงทุน ซึ่งระบบสนับสนุนการตัดสินใจในการพ้นสภาพการ

เป็นนักศึกษา เป็นสื่อกลางเชื่อมโยงไปยังนักศึกษา ให้นักศึกษาตรวจสอบตนเอง เพื่อพัฒนาตนเองให้ดีขึ้น และสามารถแก้ปัญหาเรื่องพ้นสภาพนักศึกษาได้ (Che-doloh, Kaheng, & Kepan, 2018) อีกทั้งสถานศึกษาสามารถนำข้อมูลไปพิจารณาในการกำหนดหลักสูตร และการจัดการบริหารกับนักศึกษาในอนาคต โดยการนำเทคโนโลยีทางด้านการทำเหมืองข้อมูลมาประยุกต์ใช้

การนำเทคโนโลยีทางด้านปัญญาประดิษฐ์ ในด้านการทำเหมืองข้อมูลมาใช้เกี่ยวกับการศึกษา การวิเคราะห์การลาออกกลางคันของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคเหมืองข้อมูล ด้วยเทคนิค Rule Induction, K-Nearest Neighbor, Decision Tree และ Naïve Bayes เพื่อวิเคราะห์หาปัจจัยที่เกี่ยวข้องในการลาออกกลางคันของนักศึกษาระดับ

ปริญญาตรี และเปรียบเทียบประสิทธิภาพของโมเดล (Laopilai & Sanrach, 2019) หรือ การใช้เทคนิคการทำเหมืองข้อมูล เพื่อพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา (Vilailuck, Jaroenpuntaruk, & Wichadukul, 2015) และวิธีการทำเหมืองข้อมูลเพื่อศึกษาประสิทธิภาพในการเรียนของนักศึกษาระดับปริญญาตรี (Asif *et al.*, 2017) ดังนั้นผู้วิจัยจึงมีแนวคิดในการพัฒนาระบบการพยากรณ์การคัดออกของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏพิบูลสงครามด้วยเทคนิคเหมืองข้อมูล เพื่อนำไปช่วยเหลือ ติดตาม และหาแนวทางการแก้ปัญหาการคัดออกของนักศึกษาในแต่ละปีการศึกษา

2. ทฤษฎีที่เกี่ยวข้อง

2.1 เหมืองข้อมูล (Data Mining)

เหมืองข้อมูล คือการสกัดความรู้จากฐานข้อมูลที่มีปริมาณมาก เพื่อค้นหารูปแบบแนวทางในการหาความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ ซึ่งมีหลักการทางคณิตศาสตร์ สถิติ การรู้จำ และการเรียนรู้ทางเครื่องจักร เพื่อนำความรู้ที่ได้มาวางแผน แก้ไขปัญหาและประกอบการตัดสินใจในอนาคต เทคนิคเหมืองข้อมูลแบ่งได้ 2 ประเภทคือ 1) เทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เช่น เทคนิคการค้นหากฎความสัมพันธ์ (Association

Rule) และการแบ่งกลุ่มข้อมูล (Clustering) 2) เทคนิคการเรียนรู้แบบมีผู้สอน (Supervised Learning) เช่น การจำแนกประเภทข้อมูล (Classification) และการประมาณค่าข้อมูล (Regression) ซึ่งในงานวิจัยของ Daoruang, Sanrach, & Mingkwan (2021) เป็นการนำเทคนิคเหมืองข้อมูลมาเปรียบเทียบประสิทธิภาพของโมเดลในการจำแนกประเภทข้อมูล ซึ่งใช้ในการทำนายแขนงวิชาเรียนของนักศึกษาภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปทุมธานี เพื่อค้นคว้าหาตัวแบบที่ใช้ในการจำแนกประเภทข้อมูลที่ดีที่สุด

2.2 การคัดเลือกคุณลักษณะ (Feature Selection)

การคัดเลือกคุณลักษณะเป็นวิธีการค้นหา กลุ่มคุณลักษณะย่อยที่มีขนาดเล็กที่สุด และไม่ทำให้ประสิทธิภาพของแบบจำลองลดลงอย่างมีนัยสำคัญ ซึ่งเป็นขั้นตอนที่สำคัญสำหรับการสร้างแบบจำลองต่างๆ เพราะนอกจากจะช่วยลดมิติข้อมูล และลดเวลาในการประมวลผลแล้วยังสามารถกำจัดข้อมูลที่ไม่เกี่ยวข้องหรือไม่สัมพันธ์กับผลลัพธ์ ทำให้สามารถเพิ่มประสิทธิภาพของแบบจำลองได้อีกด้วย (Jetpipattanapong, 2018) วิธีการคัดเลือกคุณลักษณะสามารถแบ่งได้ 2 วิธี ได้แก่ Filter Approach และ Wrapper Approach ดังตาราง 1

ตาราง 1 ตารางเปรียบเทียบเทคนิคคุณลักษณะที่นำมาใช้

| เทคนิคคุณลักษณะ | คำอธิบาย |
|------------------|---|
| Filter Approach | เป็นการเลือกฟีเจอร์ตามค่าน้ำหนักที่คำนวณได้และเรียงลำดับค่าน้ำหนักที่มากนำมาใช้งานที่คำนวณจากค่าความสัมพันธ์ระหว่างฟีเจอร์ เช่น เทคนิค Information Gain (IG) และเทคนิค Chi-Square |
| Wrapper Approach | เป็นการคัดเลือกฟีเจอร์ด้วยการสร้างโมเดลขึ้นมาจากเซตของฟีเจอร์ที่กำหนดไว้และวัดประสิทธิภาพการทำงานของโมเดล และเลือกเซตของฟีเจอร์ที่ทำให้โมเดลมีประสิทธิภาพมากที่สุดมาใช้งาน เช่น เทคนิค Forward Selection และเทคนิค Backward Elimination |

2.3 เทคนิคต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจเป็นการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ ในรูปแบบโครงสร้างต้นไม้สามารถสร้างแบบจำลองการจัดหมวดหมู่ได้จากกลุ่มตัวอย่างข้อมูลที่กำหนดไว้ล่วงหน้า และพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้ด้วยรูปแบบของ Tree โครงสร้างประกอบด้วย Root Node, Child และ Leaf Node (Daoruang, Sanrach, & Mingkhwan, 2021) จากงานวิจัยของ Chareonrat (2016) เป็นการวิเคราะห์ปัจจัยที่ส่งผลต่อการพัฒนาของนักศึกษาที่มีผลการเรียนปกติโดยใช้ต้นไม้ตัดสินใจ เป็นการนำกฎการจำแนกข้อมูลที่ได้มาใช้ในการพัฒนาระบบพยากรณ์การพัฒนาของนักศึกษาต่อไป

2.4 เทคนิคนาอิวเบย์ (Naïve Bayes)

นาอิวเบย์เป็นเครื่องจักรเรียนรู้ที่อาศัยหลักการความน่าจะเป็น (Probability) ตามทฤษฎีของเบย์ (Bayes' Theorem) ซึ่งมีอัลกอริทึมที่ไม่ซับซ้อนเป็นขั้นตอนวิธีในการจำแนกข้อมูล โดยการเรียนรู้ปัญหาที่เกิดขึ้น เพื่อนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่ (Laopilai & Sanrach, 2019) จากงานวิจัยของ Boonprasom & Sanrach (2018) ได้นำเทคนิคนาอิวเบย์ไปใช้ในการวิเคราะห์การทำนายการลาออกกลางคันของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏอุบลราชธานี เพื่อเปรียบเทียบ

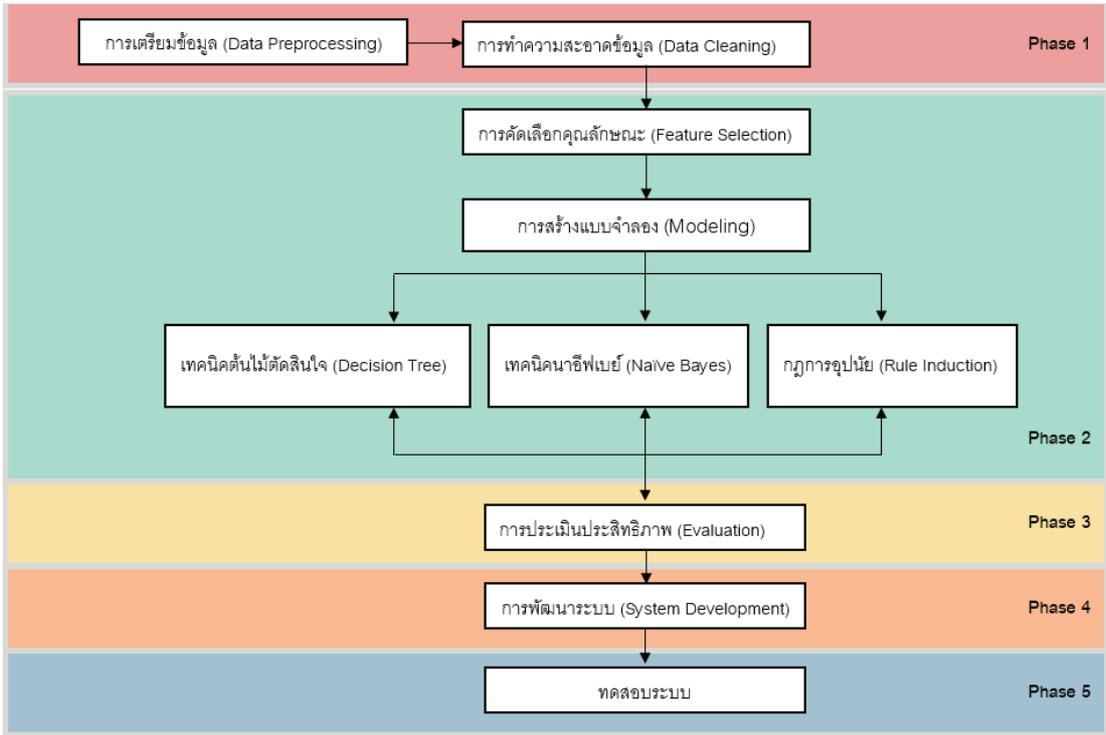
ประสิทธิภาพการจำแนกข้อมูล พบว่าแบบจำลองที่สร้างด้วยเทคนิคนาอิวเบย์มีประสิทธิภาพมากที่สุด มีค่าเฉลี่ยความถูกต้อง 93.58%

2.5 กฎการอุปนัย (Rule Induction)

กฎการอุปนัยเป็นวิธีการสำหรับการดึงเอาชุดกฎเกณฑ์ต่างๆ เพื่อจัดแบ่งเงื่อนไขหรือกรณีโครงสร้างต้นไม้สามารถสร้างชุดของกฎต่างๆ และขณะที่บางครั้งเรียกวิธีการนี้ว่า การสร้างกฎใหม่จากตัวอย่าง แต่วิธีการนี้ยังมีความหมายที่แตกต่างกันเนื่องจากวิธีการใช้การอุปนัยจะสร้างชุดของกฎที่เป็นอิสระ ซึ่งไม่จำเป็นต้องอยู่ในรูปโครงสร้างของต้นไม้ เพราะตัวสร้างการอนุมานกฎ (Rule Induction) ไม่ได้บังคับการแตกข้อมูลเป็นแต่ละระดับแต่อาจจะสามารถค้นหารูปแบบ (Pattern) ที่แตกต่างกันได้ (Kladchuen & Sanrach, 2018) จากงานวิจัยของ Laopilai & Sanrach (2019) ในการสร้างโมเดลด้วยเทคนิคเหมืองข้อมูล 4 เทคนิคจากการทำนายการลาออกกลางคันของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน พบว่าโมเดลที่สร้างด้วยเทคนิควิธี Rule Induction มีประสิทธิภาพสูงสุดมีค่าเฉลี่ยความถูกต้อง 94.70%

3. วิธีการดำเนินการวิจัย

ขั้นตอนการดำเนินการวิจัยประกอบด้วย 5 ระยะ คือ ระยะที่ 1 การเตรียมข้อมูล (Data Preprocessing) ระยะที่ 2 การคัดเลือกคุณลักษณะ



ภาพประกอบ 1 แสดงขั้นตอนการดำเนินการวิจัย

และการสร้างแบบจำลอง (Modeling) ระยะเวลาที่ 3 การประเมินประสิทธิภาพ (Evaluation) ระยะเวลาที่ 4 การพัฒนาระบบ (System Development) และระยะเวลาที่ 5 ทดสอบระบบ โดยมีรายละเอียด ดังภาพประกอบ 1

ระยะที่ 1 การเตรียมข้อมูล

1.1 การเตรียมข้อมูล (Data Preprocessing)

ข้อมูลที่ใช้ในการพยากรณ์การตออกของนักศึกษา คือ ข้อมูลของนักศึกษาย้อนหลัง 6 ปี โดยรวบรวมมาจากกองบริการการศึกษา มหาวิทยาลัยราชภัฏพิบูลสงคราม ระหว่างปีการศึกษา 2558-2563 จำนวน 25,132 ชุดข้อมูล โดยมีข้อมูลนักศึกษาอยู่ 7 ระดับ คือประกาศนียบัตรบัณฑิต ปริญญาตรี 2 ปี (ต่อเนื่อง) ปริญญาตรี 4 ปี (เทียบโอน) ปริญญาตรี 4 ปี ปริญญาตรี 5 ปี ปริญญาโท และปริญญาเอก ในการนำข้อมูลไปใช้ผู้วิจัยเลือกใช้ข้อมูลเฉพาะนักศึกษา ระดับปริญญาตรี จำนวนทั้งสิ้น 20,093 ชุดข้อมูล

1.2 การทำความสะอาดข้อมูล (Data Cleaning)

การแก้ไขความไม่สมบูรณ์ของข้อมูล เพื่อจัดการข้อมูลที่สูญหาย (Missing Value) และลดข้อมูลรบกวน (Noise) หรือข้อมูลที่มีค่าผิดปกติ ข้อมูลที่ได้ทำการรวบรวมในงานวิจัยครั้งนี้ พบว่ามีข้อมูลที่เป็นค่าว่าง ซึ่งการทำความสะอาดข้อมูล โดยกำจัดค่าว่าง และแทนที่ค่าที่ผิดปกติที่ไม่สามารถนำข้อมูลเข้าโปรแกรม RapidMiner Studio 9.9 ชุดข้อมูลที่นำมาวิเคราะห์การหาปัจจัยที่เกี่ยวข้องในการตออกของนักศึกษาประกอบด้วย 15 แอททริบิวต์ และผลลัพธ์ 1 แอททริบิวต์ ได้แก่ 1) เพศ 2) ภาคการศึกษา 3) ระดับการศึกษา 4) ความถนัด/ความสามารถพิเศษ 5) จังหวัด 6) ความพิการ 7) อาชีพบิดา 8) อาชีพมารดา 9) ปีที่เข้าศึกษา 10) กลุ่มเรียน 11) คณะ 12) หลักสูตร 13) สาขาวิชา 14) เกรดแต่ละภาคการศึกษา 15) เกรดเฉลี่ยรวม และ 16) สถานะนักศึกษา ดังภาพประกอบ 2

| รหัสนักศึกษา | สถานศึกษา | ระดับ | ความถนัด/จังหวัด | ความพิการ | อาชีพบิดา | อาชีพมารดา | ปีเข้าศึกษา | กลุ่มเรียน | คณะ | หลักสูตร | สาขาวิชา | 1/58 | 1/56-1/58 | 2/58 | 1/56-2/58 | 3/58 | 1/56-3/58 | 1/59 |
|--------------|----------------|---------------|------------------|-----------|-----------|------------|-------------|------------|-----------------------------------|----------|-------------|------|-----------|------|-----------|------|-----------|------|
| 1 | ส่วจังหวัด | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | ค้าขาย | ธุรกิจ | ธุรกิจ | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.17 | 2.17 | 2.5 | 2.33 | 0 | 2.33 | 2.47 |
| 2 | ไม่มาติดต่อ | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | เกษตร | ประมง | เกษตร | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.05 | 2.05 | 1.91 | 1.98 | 0 | 1.98 | 2.17 |
| 3 | สำเร็จการศึกษา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | เกษตร | ประมง | เกษตร | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.14 | 2.14 | 2.23 | 2.19 | 0 | 2.19 | 2.5 |
| 4 | สำเร็จการศึกษา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | เกษตร | ประมง | เกษตร | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.91 | 2.91 | 2.35 | 2.6 | 0 | 2.6 | 2.82 |
| 5 | ไม่มาติดต่อ | ปริญญตรี 4 ปี | สุโขทัย | ไม่พิการ | เกษตร | ประมง | เกษตร | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 1.82 | 1.82 | 2 | 1.9 | 0 | 1.9 | 2.2 |
| 6 | ไม่มาติดต่อ | ปริญญตรี 4 ปี | สุโขทัย | ไม่พิการ | ค้าขาย | ธุรกิจ | ธุรกิจ | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 1.91 | 1.91 | 2.21 | 2.04 | 0 | 2.04 | 2.1 |
| 7 | ย้ายสาขา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | ค้าขาย | ธุรกิจ | ธุรกิจ | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.35 | 2.35 | 0 | 2.35 | 0 | 2.35 | 2.17 |
| 8 | ลาออก | ปริญญตรี 4 ปี | เพชรบูรณ์ | ไม่พิการ | เกษตร | ประมง | เกษตร | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.11 | 2.11 | 2.32 | 2.32 | 0 | 2.32 | 2.3 |
| 9 | สำเร็จการศึกษา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | เกษตร | ประมง | เกษตร | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.47 | 2.47 | 3.05 | 2.78 | 0 | 2.78 | 2.95 |
| 10 | สำเร็จการศึกษา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | ค้าขาย | ธุรกิจ | ธุรกิจ | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 3.14 | 3.14 | 2.7 | 2.92 | 0 | 2.92 | 2.95 |
| 11 | ย้ายสาขา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | ค้าขาย | ธุรกิจ | ธุรกิจ | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.41 | 2.41 | 2.61 | 2.51 | 0 | 2.51 | 2.55 |
| 12 | ย้ายสาขา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | ค้าขาย | ธุรกิจ | ธุรกิจ | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 1.94 | 1.94 | 2.42 | 2.16 | 0 | 2.16 | 2.28 |
| 13 | ย้ายสาขา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | ค้าขาย | ธุรกิจ | ธุรกิจ | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.32 | 2.32 | 2.3 | 2.31 | 0 | 2.31 | 2.4 |
| 14 | สำเร็จการศึกษา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | เกษตร | ประมง | เกษตร | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.32 | 2.32 | 2.64 | 2.48 | 0 | 2.48 | 2.5 |
| 15 | สำเร็จการศึกษา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | เกษตร | ประมง | เกษตร | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.32 | 2.32 | 2.58 | 2.45 | 0 | 2.45 | 2.44 |
| 16 | สำเร็จการศึกษา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | เกษตร | ประมง | เกษตร | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.5 | 2.5 | 2.35 | 2.42 | 0 | 2.42 | 2.64 |
| 17 | สำเร็จการศึกษา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | เกษตร | ประมง | เกษตร | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.41 | 2.41 | 2.79 | 2.6 | 0 | 2.6 | 3.14 |
| 18 | สำเร็จการศึกษา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | ค้าขาย | ธุรกิจ | ธุรกิจ | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.47 | 2.47 | 2.46 | 2.46 | 0 | 2.46 | 2.5 |
| 19 | ย้ายสาขา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | ไม่ระบุ | ค้าขาย | ธุรกิจ | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.2 | 2.2 | 2.63 | 2.37 | 0 | 0 | 0 |
| 20 | ย้ายสาขา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | พนักงาน | พนักงาน | พนักงาน | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 1.94 | 1.94 | 2.27 | 2.07 | 0 | 2.07 | 1.86 |
| 21 | สำเร็จการศึกษา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | ไม่ระบุ | พนักงาน | พนักงาน | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.64 | 2.64 | 2.37 | 2.5 | 0 | 2.5 | 2.57 |
| 22 | ไม่มาติดต่อ | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | ค้าขาย | ธุรกิจ | ธุรกิจ | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.14 | 2.14 | 0 | 0 | 0 | 0 | 0 |
| 23 | ไม่มาติดต่อ | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | ค้าขาย | ธุรกิจ | ธุรกิจ | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 1.92 | 1.92 | 2 | 1.96 | 0 | 1.96 | 1.73 |
| 24 | สำเร็จการศึกษา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | ไม่ระบุ | เกษตร | ประมง | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.14 | 2.14 | 2.21 | 2.17 | 0 | 2.17 | 1.88 |
| 25 | สำเร็จการศึกษา | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | ไม่ระบุ | ไม่ระบุ | ไม่ระบุ | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.08 | 2.08 | 2 | 2.04 | 0 | 2.04 | 1.89 |
| 26 | ไม่มาติดต่อ | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | อื่นๆ | อื่นๆ | อื่นๆ | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.64 | 2.64 | 2.77 | 2.69 | 0 | 2.69 | 2.97 |
| 27 | ลาออก | ปริญญตรี 4 ปี | พิษณุโลก | ไม่พิการ | อื่นๆ | อื่นๆ | อื่นๆ | 2558 | เทคโนโลยี วิทยาศาสตร์ เกษตรศาสตร์ | | เกษตรศาสตร์ | 2.64 | 2.64 | 2.77 | 2.69 | 0 | 2.69 | 2.97 |

ภาพประกอบ 2 แสดงคุณลักษณะตัวอย่างของข้อมูลนักศึกษา

จากชุดข้อมูลข้างต้น สถานะนักศึกษา เป็นผลลัพธ์ในการหาปัจจัยที่เกี่ยวข้องในการต้อออกของนักศึกษา ประกอบไปด้วย ไม่มาติดต่อ ครบหลักสูตร ไม่ลงทะเบียนเรียน ย้ายสาขา ย้ายข้ามภาค รอเรียนเพิ่ม ลงทะเบียนเรียน ลาออก ออกตามระเบียบ สำเร็จการศึกษา และเสียชีวิต ซึ่งสถานะนักศึกษาบางสถานะไม่สามารถทำการพยากรณ์การต้อออกของนักศึกษาได้ เนื่องจากปัจจัยที่มีอยู่ไม่เพียงพอต่อการพยากรณ์ บางสถานะต้องใช้จำนวนหน่วยกิตในการคำนวณรายวิชาที่ต้องเรียน ข้อมูลผลของการประเมินในแต่ละรายวิชา เช่น การเรียนครบหลักสูตร ย้ายสาขา ย้ายข้ามภาค และรอเรียนเพิ่ม รวมไปถึงประวัติส่วนตัวของนักศึกษา ในการระบุสาเหตุของการลาออก ไม่มาติดต่อ และการเสียชีวิต ผู้วิจัยจึงเลือกทำนายสถานะนักศึกษา 3 สถานะ ได้แก่ 1) สำเร็จการศึกษา 2) ออกตามระเบียบและ 3) ลงทะเบียนเรียน เพื่อให้ได้ผลลัพธ์ในการพยากรณ์ สำหรับการเลือกสถานะของนักศึกษาทั้ง 3 สถานะ เป็นสถานะที่สามารถนำมาทำนายสถานะปัจจุบันของนักศึกษาได้ ซึ่งเป็นแนวทางในการพัฒนาตนเอง และในชุดข้อมูลมีปัจจัยที่สามารถนำไปทำนายสถานะของนักศึกษาได้ทั้ง 3 สถานะ

ระยะที่ 2 การคัดเลือกคุณลักษณะและการสร้างแบบจำลอง (Modeling)

2.1 การคัดเลือกคุณลักษณะ (Feature Selection)

เป็นเทคนิคในการช่วยลดจำนวนตัวแปรคัดเลือกปัจจัยที่ส่งผลต่อการพยากรณ์ กระบวนการนี้เป็นขั้นตอนสำคัญสำหรับการเตรียมข้อมูลในการทำเหมืองข้อมูล ซึ่งทำให้การนำข้อมูลไปสร้างตัวแบบพยากรณ์ได้อย่างรวดเร็ว และมีประสิทธิภาพมากยิ่งขึ้น การคัดเลือกคุณลักษณะแบ่งออกเป็น 2 วิธี ได้แก่ Filter approach และ Wrapper approach (Wichian *et al.*, 2020; Pacharawongsakda, 2015) ในงานวิจัยครั้งนี้เลือกใช้วิธีเทคนิค Filter Approach ด้วยค่า Information Gain หรือการวัดค่า Entropy ในการชี้วัดการเลือกคุณลักษณะข้อมูล โดยการคำนวณค่า Entropy สำหรับแต่ละมิติข้อมูล ซึ่งข้อมูลใดมีค่า Entropy สูงที่สุดจะถูกคัดเลือกคุณลักษณะ เพื่อนำไปใช้ ดังสมการที่ 1 แสดงการคำนวณค่า Information Gain หรือค่า Entropy ของชุดข้อมูลทั้งหมด สมการที่ 2 แสดงการคำนวณค่า Entropy ของชุดมิติข้อมูลในแต่ละลักษณะ สมการที่ 3 เป็นการคำนวณค่า Information Gain สำหรับการพิจารณามิติข้อมูลคุณลักษณะ

$$E(D) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (1)$$

$$E_A(D) = - \sum_{j=1}^m \frac{|D_j|}{D} \times E(D_j) \quad (2)$$

$$Gain(A) = E(D) - E_A(D) \quad (3)$$

โดย คือ ค่าความน่าจะเป็นที่เรีคคอร์ดหนึ่งๆ จะมีหมวดหมู่ของข้อมูลหรือกล่าวว่าการคำนวณค่า Information Gain หรือการวัดค่า Entropy ก่อนที่จะมีการแบ่งข้อมูลออกตามมิติข้อมูล และหลักการแบ่งว่ามีประสิทธิภาพดีขึ้นหรือไม่ ถ้ามีประสิทธิภาพดีขึ้นค่า Information Gain ก็จะมีค่าสูง (Jaidee & Wannapee, 2020)

การวิเคราะห์หาปัจจัยที่เกี่ยวข้องในการตออกของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏพิบูลสงคราม โดยการคัดเลือกคุณลักษณะ (Feature Selection) ด้วยค่า Information Gain การคัดเลือกคุณลักษณะ Operator ที่เกี่ยวข้อง ได้แก่ 1) Read Excel ใช้สำหรับอ่านไฟล์ข้อมูลประเภท Excel 2) Weight by Information Gain ใช้สำหรับคำนวณค่าน้ำหนักของแอททริบิวต์ ด้วยเทคนิค Information Gain และ 3) Selection by Weight ใช้สำหรับการเลือกแอททริบิวต์ตามค่าน้ำหนัก

2.2 การสร้างแบบจำลอง (Modeling)

เป็นการจำแนกกลุ่มข้อมูลด้วยกระบวนการสร้างแบบจำลอง ซึ่งในงานวิจัยครั้งนี้ ต้องนำข้อมูลที่ได้มาทำการสร้างแบบจำลองและเลือกเทคนิคที่เหมาะสม เพื่อนำใช้ในการวิเคราะห์ทำการทดสอบความถูกต้อง และหาค่าความผิดพลาดของแบบจำลอง แล้วจึงนำไปพัฒนาเป็นระบบการพยากรณ์การตออกของนักศึกษา มหาวิทยาลัยราชภัฏพิบูลสงคราม โดยทำการเปรียบเทียบเทคนิคที่นำมาใช้ในงานวิจัยครั้งนี้ซึ่งสรุปข้อดีและข้อจำกัด (Nettleton, 2014; Sintawichai, 2015) ดังตาราง 2

ขั้นตอนการสร้างแบบจำลองจะใช้ข้อมูลจากการหาปัจจัยที่เกี่ยวข้องในการตออกของนักศึกษา ทำการวิเคราะห์ข้อมูลด้วยวิธีการทำเหมืองข้อมูล (Data Mining) เป็นวิธีการหนึ่งในการวิเคราะห์ข้อมูลที่มีปริมาณมาก ด้านข้อมูลการศึกษาถือว่าเป็นข้อมูลขนาดใหญ่ที่ถูกจัดเก็บรวบรวมอยู่เป็นจำนวนมาก ซึ่งงานวิจัยครั้งนี้ใช้เครื่องมือในการทำเหมืองข้อมูลด้วยโปรแกรม RapidMiner Studio 9.9 ในการสร้างแบบจำลองการพยากรณ์ด้วยเทคนิคเหมืองข้อมูล 3 เทคนิค ดังนี้

ตาราง 2 ตารางเปรียบเทียบข้อดีและข้อจำกัดของแต่ละเทคนิคที่นำมาใช้

| เทคนิคที่นำมาใช้ | ข้อดี | ข้อจำกัด |
|------------------|--|--|
| Decision Tree | - มีความชัดเจน ง่ายต่อการ Implement และการแปลผล | - เรื่องของ Over Fitting ที่โมเดลจะ Fit เข้ากับ Data จนทำให้เกิดประสิทธิภาพของการทำงานที่ต่ำ |
| Naïve Bayes | - ง่ายต่อการนำไปใช้เพราะใช้การคำนวณที่ง่าย - ได้ผลลัพธ์ที่สามารถนำไปประยุกต์ใช้ได้ดี | - ใช้ได้กับ Attribute ที่เป็นอิสระกันเท่านั้น |
| Rule Induction | - สร้างกฎที่เป็นอิสระ - ไม่ได้บังคับการแตกข้อมูลแต่ละระดับ - การสร้างกฎใหม่จากตัวอย่างไม่จำเป็นต้องอยู่ในรูปโครงสร้างของต้นไม้ | - โมเดลต้องมีชุดตัวแปรอินพุตและตัวแปรเอาต์พุต ตั้งแต่ 1 ตัวขึ้นไป - การสร้างแบบจำลองที่สร้างขึ้นจากกฎของประเภท If-then-else |

เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เป็นการจำแนกกฎ โดยมีลักษณะการทำงานเหมือนโครงสร้างต้นไม้ ซึ่งการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ มีการเรียนรู้ข้อมูลแบบมีผู้สอน และทำการคัดเลือกแอททริบิวต์ที่มีความสัมพันธ์กับคลาสมากที่สุดขึ้นมาเป็นโหนดบนสุดของต้นไม้ จากนั้นจะเลือกคุณสมบัติที่มีความสัมพันธ์ถัดไปเรื่อยๆ จากการคำนวณ Information Gain (IG) โดยเลือกคุณสมบัติที่มีค่า IG สูงที่สุด การสร้างแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจ Operator ที่เกี่ยวข้อง (Kumjit *et al.*, 2022) ได้แก่ 1) Decision Tree ใช้สำหรับสร้างโมเดล Decision Tree 2) Apply Model ใช้สำหรับ predict ข้อมูลใหม่และ 3) Performance ใช้สำหรับแสดงตัวชี้วัดของโมเดล Classification และการ set ค่า parameter ด้วยเทคนิคต้นไม้ตัดสินใจ ดังนี้ Criterion เลือก Information_gain, ความลึกสูงสุด maximal depth = 10, ค่า confidence = 0.1 minimal gain = 0.01, ขนาดใบน้อยที่สุด minimal leaf size = 2, ขนาดขั้นต่ำสำหรับการแบ่ง minimal size for split = 4 และจำนวนทางเลือกในการตัดแต่งกิ่ง number of prepruning alternatives = 5 ในการปรับค่าพารามิเตอร์เป็นสิ่งที่จำเป็นเพื่อให้ได้ผลลัพธ์ที่ดีที่สุด

เทคนิคนาอิวเบย์ (Naïve Bayes) เป็นเทคนิควิเคราะห์ความสัมพันธ์ระหว่างตัวแปร เพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นของแต่ละความสัมพันธ์ในการแก้ปัญหาแบบจำแนกประเภทที่สามารถคาดการณ์ผลลัพธ์ได้ และมีการทำงานที่ไม่ซับซ้อนมีขั้นตอน การสร้างแบบจำลองด้วยเทคนิคนาอิวเบย์ Operator ที่เกี่ยวข้อง ได้แก่ 1) Naïve Bayes ใช้สำหรับสร้างโมเดล Naïve Bayes 2) Apply Model ใช้สำหรับ predict ข้อมูลใหม่และ 3) Performance ใช้สำหรับแสดงตัวชี้วัดของโมเดล Classification

กฎการอุปนัย (Rule Induction) ดึงเอาชุดกฎเกณฑ์ต่างๆ มาสร้างเป็นเงื่อนไขที่อยู่ในรูปแบบเป็นอิสระ ซึ่งไม่จำเป็นต้องอยู่ในรูปแบบของโครงสร้างต้นไม้ เพื่อหาความสัมพันธ์ของข้อมูลมีขั้นตอนการสร้างแบบจำลองด้วยกฎการอุปนัย Operator ที่เกี่ยวข้อง ได้แก่ 1) Rule Induction ใช้สำหรับสร้างโมเดล Rule Induction 2) Apply Model ใช้สำหรับ predict ข้อมูลใหม่และ 3) Performance ใช้สำหรับแสดงตัวชี้วัดของโมเดล Classification

ระยะที่ 3 การประเมินประสิทธิภาพ (Evaluation)

3.1 การประเมินประสิทธิภาพ

การทดสอบประสิทธิภาพของแบบจำลองด้วยวิธีการ 5-Fold Cross-Validation และ 10-Fold Cross-Validation จะแบ่งข้อมูลออกเป็น 5 และ 10 ส่วน (แสดงด้วยค่า) โดยแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ 1) ข้อมูลชุดฝึกหัด (Training Data) นำไปสร้างตัวต้นแบบ ซึ่งมีข้อมูลร้อยละ 70 ของข้อมูลทั้งหมด 2) ชุดทดสอบ (Test Data) นำไปทำนายตัวแบบ ซึ่งมีข้อมูลร้อยละ 30 ของข้อมูลทั้งหมด ซึ่ง Operator ที่ใช้ในการประเมินประสิทธิภาพ คือ Cross-Validation เป็นการแบ่งข้อมูลสำหรับสร้างโมเดลและทดสอบโมเดล (Nasritha *et al.*, 2018)

การวัดประสิทธิภาพของแบบจำลองที่ได้จากข้อมูลชุดเรียนรู้มาทดสอบด้วยข้อมูลชุดทดสอบ วัดจากค่าความถูกต้อง (Accuracy) และค่าความผิดพลาด (Mean Absolute Error: MAE) ในการวัดประสิทธิภาพของแบบจำลอง ดังนี้

Accuracy คือ การวัดค่าความถูกต้อง โดยรวมของระบบระหว่างค่าจริงและค่าการทำนาย ถ้าหากค่า Accuracy มีค่ามากนั้นหมายถึง ค่าการทำนายนั้นสามารถทำนายได้ถูกต้องใกล้เคียงกับค่าจริง

(4) ระยะเวลาที่ 4 การพัฒนาระบบ (System Development)

จากสมการ

Accuracy คือ ค่าความถูกต้อง

TP คือ ค่าการทำนายที่ทำนายว่าจริงซึ่งตรงกับค่าจริง

TN คือ ค่าการทำนายที่ทำนายว่าไม่จริงซึ่งตรงกับค่าจริง

FP คือ ค่าการทำนายที่ทำนายว่าไม่จริงซึ่งไม่ตรงกับค่าจริง

FN คือ ค่าการทำนายที่ทำนายว่าจริงซึ่งไม่ตรงกับค่าจริง

ค่าความผิดพลาด (Mean Absolute Error: MAE) คือ เป็นค่าเฉลี่ยของความแตกต่างสัมบูรณ์ระหว่างค่าพยากรณ์และค่าจริง หากค่า MAE มีค่าน้อยแสดงว่าแบบจำลองสามารถประมาณค่าได้ใกล้เคียงกับค่าจริง (Lake, 2018)

(5)

จากสมการ

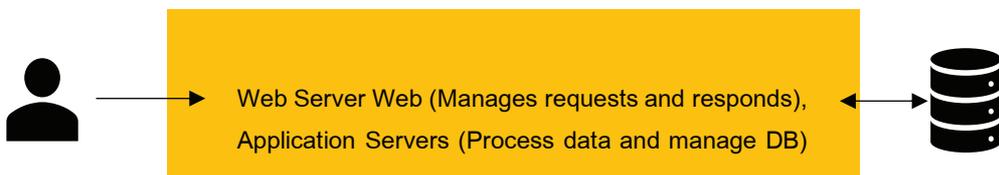
T_i คือ ค่าจริง

F_i คือ ค่าพยากรณ์

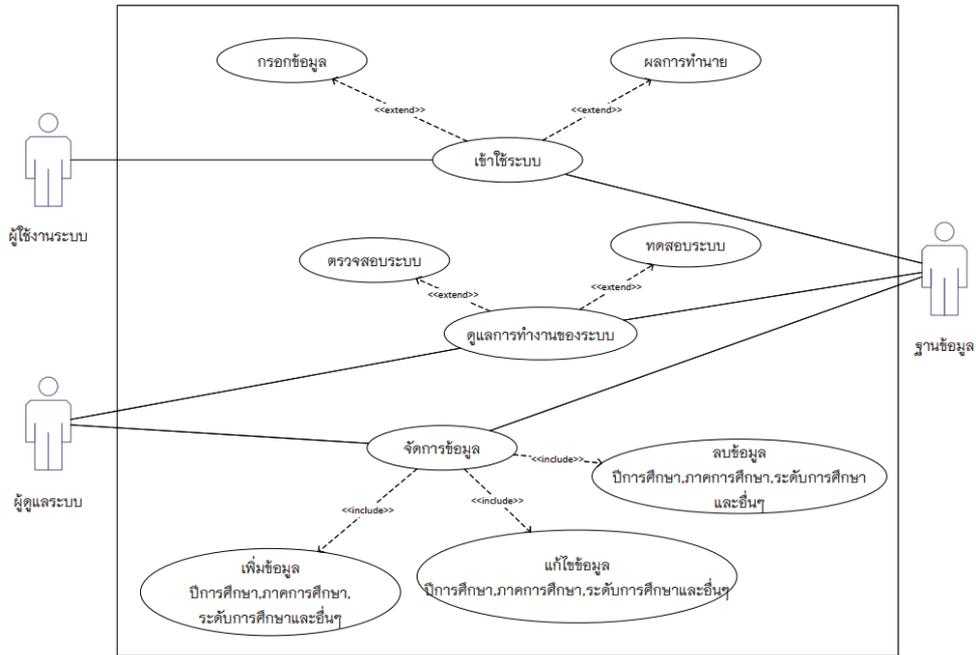
N คือ จำนวนข้อมูลในชุดข้อมูล

ในการพัฒนาระบบการพยากรณ์การตอกออกของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยพิษณุโลกด้วยเทคนิคเหมืองข้อมูล เป็นการนำข้อมูลที่วิเคราะห์ปัจจัยที่ส่งผลต่อการตอกออกนำมาพัฒนาระบบ โดยใช้เครื่องมือในการพัฒนา AppServ สำหรับการจำลองเครื่องแม่ข่ายภาษา PHP สำหรับการเขียน Script โปรแกรม RapidMiner Studio 9.9 สำหรับสร้างต้นแบบพยากรณ์ และMySQL สำหรับจัดการฐานข้อมูล (Saemmasu, Sitimanung, & Kepan, 2018) ผู้วิจัยจึงได้พัฒนาระบบในรูปแบบเว็บแอปพลิเคชันซึ่งมีภาพรวมของระบบ ดังภาพประกอบ 3

การทำงานของระบบการพยากรณ์การตอกออกของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยพิษณุโลกด้วยเทคนิคเหมืองข้อมูล ได้ใช้ Use Case Diagram เป็นเครื่องมือในการวิเคราะห์ระบบเพื่อแสดงให้เห็นถึงการทำงานของผูู้ดูแลระบบและผูู้ใช้งานระบบ ดังภาพประกอบ 4



ภาพประกอบ 3 ภาพรวมของการพัฒนาระบบ



ภาพประกอบ 4 แสดงการทำงานของระบบ

ลักษณะการทำงานของ Use Case Diagram แสดงการทำงานของระบบ โดยแบ่งออกเป็น 2 ส่วน คือ 1) ส่วนของผู้ใช้งานระบบสามารถเข้าใช้งานระบบ และกรอกข้อมูลตามแบบฟอร์มในการนำเข้าข้อมูล เมื่อกรอกข้อมูลถูกต้องครบถ้วนจะสามารถทำนายผลสถานะนักศึกษาได้ ระบบจะแสดงผลการทำนายสถานะภาพของนักศึกษา และ 2) ส่วนของผู้ดูแลระบบสามารถจัดการข้อมูลได้ โดยการเพิ่ม ลบ แก้ไข ข้อมูล ปีการศึกษา ภาคการศึกษา ระดับการศึกษาและอื่นๆ รวมถึงการดูแลระบบสามารถตรวจสอบระบบและทดสอบระบบได้ซึ่งข้อมูลทั้งหมดถูกจัดเก็บอยู่ในฐานข้อมูล

ระยะที่ 5 ทดสอบระบบ

การทดสอบระบบการพยากรณ์การตออกของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏพิบูลสงครามด้วยเทคนิคเหมืองข้อมูล ทำการทดสอบ

ระบบจากการสุ่มตัวอย่าง ซึ่งแบ่งออกเป็น 3 สถานะ ได้แก่ 1) มีโอกาสสำเร็จการศึกษา 2) ออกตามระเบียบ และ 3) ลงทะเบียนเรียน โดยมีการแบ่งปีการศึกษาออกเป็น 6 ปีการศึกษา และทำการสุ่มตัวอย่างข้อมูลในแต่ละสถานะออกเป็นปีละ 15 ชุดข้อมูล จำนวนข้อมูลในแต่ละสถานะมีจำนวน 90 ชุดข้อมูล ซึ่งรวมชุดข้อมูลที่นำมาทดสอบในระบบจำนวน 270 ชุดข้อมูล จากข้อมูลทั้งหมด 20,093 ชุดข้อมูล

4. ผลการวิจัย

ผลการวิเคราะห์การหาปัจจัยที่เกี่ยวข้องในการตออกของนักศึกษาระดับปริญญาตรี ด้วยวิธีการคัดเลือกคุณลักษณะ (Feature Selection) โดยใช้เทคนิค Filter Approach การจัดลำดับคุณสมบัติวิเคราะห์ค่าน้ำหนักของ แอททริบิวต์ทั้งหมด 15 แอททริบิวต์ พบว่ามีปัจจัยที่เกี่ยวข้องในการตออก 10 ปัจจัยที่สำคัญตามลำดับ ดังตาราง 3

ตาราง 3 ค่าน้ำหนักของปัจจัยที่เกี่ยวข้อง

| ลำดับ | ชื่อแอททริบิวต์ | ค่าน้ำหนัก |
|-------|--------------------------|------------|
| 1. | เกรดเฉลี่ยรวม | 0.642 |
| 2. | ปีที่เข้าศึกษา | 0.561 |
| 3. | สาขาวิชา | 0.145 |
| 4. | หลักสูตร | 0.086 |
| 5. | อาชีพมารดา | 0.086 |
| 6. | อาชีพบิดา | 0.068 |
| 7. | คณะ | 0.051 |
| 8. | ระดับการศึกษา | 0.027 |
| 9. | ภาค | 0.009 |
| 10. | ความถนัด/ความสามารถพิเศษ | 0.004 |

ค่าน้ำหนักของแอททริบิวต์ที่ถูกตัดออก 5 แอททริบิวต์ มีค่าน้ำหนักเท่ากับ 0.000 จึงถูกตัดออกจากปัจจัยที่เกี่ยวข้อง เนื่องจากปัจจัยนั้น ไม่มีผลต่อการทำนายสถานภาพของนักศึกษา

ผลการเปรียบเทียบประสิทธิภาพแบบจำลอง โดยการเปรียบเทียบการทำงานของ 3 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคนาอิวเบย์ (Naïve Bayes) และกฎการอุปนัย (Rule Induction) ผลการทดสอบด้วยวิธีการแบบ 5-Fold Cross-Validation และ 10-Fold Cross-Validation

พบว่าตัวแบบการพยากรณ์ที่ใช้เทคนิคต้นไม้ตัดสินใจ ซึ่งได้ค่าสูงที่สุดในการแบ่งข้อมูลทดสอบ 10-Fold Cross-Validation ให้ค่าความแม่นยำและค่าความผิดพลาดดีที่สุดในการรวมข้อมูลนักศึกษาในระดับปริญญาตรีทั้งหมด ได้ค่าความถูกต้อง (Accuracy) 97.81% และค่าความผิดพลาด (MAE) เท่ากับ 0.026 กฎการอุปนัยได้ค่าความถูกต้อง 93.37% และค่าความผิดพลาดเท่ากับ 0.122 และเทคนิคนาอิวเบย์ได้ค่าความถูกต้อง 92.40% และค่าความผิดพลาดเท่ากับ 0.076 ตามลำดับ ดังตาราง 4

ตาราง 4 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลอง 5-Fold Cross-Validation และ 10-Fold Cross-Validation

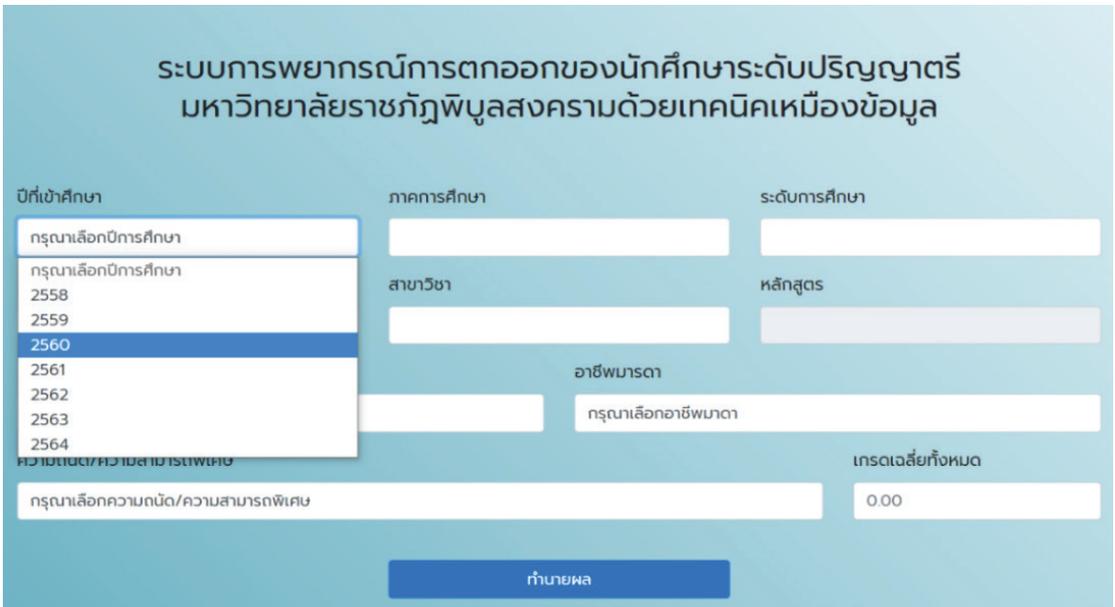
| Domain | Data Set | Classification Algorithm | | | | | | | | | | | |
|-----------------|----------|--------------------------|-------|----------|-------|-------------|-------|----------|-------|----------------|-------|----------|-------|
| | | Decision Tree | | | | Naïve Bayes | | | | Rule Induction | | | |
| | | 5-Fold | | 10-Fold | | 5-Fold | | 10-Fold | | 5-Fold | | 10-Fold | |
| | | Accuracy | MAE | Accuracy | MAE | Accuracy | MAE | Accuracy | MAE | Accuracy | MAE | Accuracy | MAE |
| ปีการศึกษา 2558 | 3,724 | 98.95 | 0.016 | 98.63 | 0.017 | 93.82 | 0.064 | 92.53 | 0.074 | 91.70 | 0.134 | 91.65 | 0.129 |
| ปีการศึกษา 2559 | 3,831 | 98.51 | 0.019 | 98.38 | 0.020 | 94.13 | 0.057 | 94.02 | 0.057 | 90.70 | 0.137 | 89.84 | 0.178 |
| ปีการศึกษา 2560 | 3,442 | 89.22 | 0.124 | 89.10 | 0.122 | 86.25 | 0.142 | 85.96 | 0.148 | 88.23 | 0.174 | 86.98 | 0.182 |
| ปีการศึกษา 2561 | 3,031 | 96.70 | 0.056 | 96.86 | 0.051 | 96.20 | 0.036 | 96.07 | 0.038 | 92.51 | 0.131 | 92.84 | 0.128 |
| ปีการศึกษา 2562 | 2,936 | 99.73 | 0.005 | 99.76 | 0.004 | 97.24 | 0.026 | 97.55 | 0.024 | 95.95 | 0.077 | 95.95 | 0.077 |
| ปีการศึกษา 2563 | 3,129 | 99.78 | 0.005 | 99.78 | 0.004 | 99.78 | 0.002 | 99.78 | 0.002 | 99.78 | 0.004 | 99.78 | 0.004 |
| Total | 20,093 | 97.79 | 0.027 | 97.81 | 0.026 | 92.39 | 0.076 | 92.40 | 0.076 | 93.09 | 0.124 | 93.37 | 0.122 |

ผลจากการพัฒนาระบบการพยากรณ์การตออกของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏพิบูลสงครามด้วยเทคนิคเหมืองข้อมูล ซึ่งเว็บ แอปพลิเคชันประกอบไปด้วยการทำงาน 3 ส่วน คือ 1) หน้าหลักของเว็บแอปพลิเคชัน

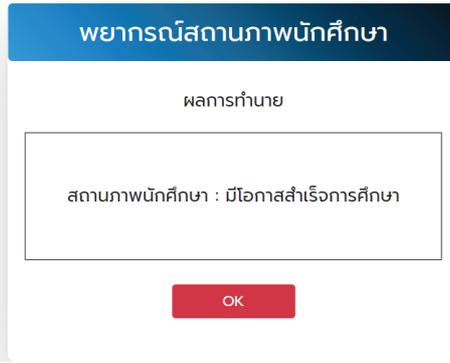
2) หน้ากรพัฒนาแบบฟอร์มในการนำเข้าข้อมูล (Input) เพื่อนำปัจจัยของนักศึกษาเข้าสู่ระบบในการทำนายสถานภาพของนักศึกษาและ 3) หน้าผลการทำนายสถานภาพของนักศึกษา (Output) ดังภาพประกอบ 5-7



ภาพประกอบ 5 หน้าหลักของเว็บแอปพลิเคชัน



ภาพประกอบ 6 หน้ากรพัฒนาแบบฟอร์มในการนำเข้าข้อมูล



ภาพประกอบ 7 หน้าของผลการทำนายสถานภาพของนักศึกษา (Output)

ตาราง 5 ผลการทดสอบระบบ

| สถานภาพ | จำนวนชุดข้อมูล | ทำนายถูกต้อง | ค่าเฉลี่ยความถูกต้อง |
|----------------------|----------------|--------------|----------------------|
| มีโอกาสำเร็จการศึกษา | 90 | 90 | 100% |
| ออกตามระเบียบ | 90 | 87 | 96.66% |
| ลงทะเบียนเรียน | 90 | 56 | 62.22% |
| ค่าเฉลี่ยรวม | | | 86.29% |

จากภาพประกอบ 6 หน้าการพัฒนาแบบฟอร์มในการนำเข้าสู่ข้อมูล เป็นการดึงข้อมูลมาจากฐานข้อมูลโดยใช้ Dropdown Menu ในการเลือกข้อมูล ซึ่งประกอบไปด้วย ปีที่เข้าศึกษา ภาคการศึกษา ระดับการศึกษา คณะ สาขาวิชา หลักสูตร อาชีพบิดา อาชีพมารดา ความถนัด/ความสามารถพิเศษ และเกรดเฉลี่ยทั้งหมด

การประเมินผลของระบบการพยากรณ์การตกรอกของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏพิบูลสงครามด้วยเทคนิคเหมือนข้อมูล ทำการประเมินผลจากการสุ่มตัวอย่างข้อมูลมาทดสอบในระบบจำนวน 270 ชุดข้อมูลจากข้อมูลทั้งหมด 20,093 ชุดข้อมูล แบ่งออกเป็น 3 สถานะ ได้แก่ 1) มีโอกาสำเร็จการศึกษา 2) ออกตามระเบียบ และ 3) ลงทะเบียนเรียน ซึ่งใช้การคำนวณค่าสถิติ ค่าร้อยละ (Percentage) หรือเปอร์เซ็นต์ ในการหาค่าเฉลี่ยความถูกต้องของการทำนายสถานภาพนักศึกษา ดังตาราง 5

5. วิเคราะห์ผลการวิจัย

ในงานวิจัยนี้เป็นการพยากรณ์การตกรอกของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏพิบูลสงครามด้วยเทคนิคเหมือนข้อมูล ซึ่งวิธีการในการวิเคราะห์หาปัจจัยที่เกี่ยวข้องในการตกรอกของนักศึกษาระดับปริญญาตรี โดยใช้วิธีการคัดเลือกคุณลักษณะ (Feature Selection) ด้วยเทคนิค Filter Approach พบว่าปัจจัยที่มีผลต่อการตกรอก คือ เกรดเฉลี่ยรวม ปีที่เข้าที่ศึกษา สาขาวิชา หลักสูตร อาชีพมารดา อาชีพบิดา คณะ ระดับการศึกษา ภาค และความถนัด/ความสามารถพิเศษ การสร้างแบบจำลองด้วยเทคนิคเหมือนข้อมูล 3 เทคนิค ได้แก่ 1) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) 2) เทคนิคนาอิวเบย์ (Naive Bayes) และ 3) กฎการอุปนัย (Rule Induction) ซึ่งผู้วิจัยได้อภิปรายผลการวิจัยไว้ข้างต้น

จากการศึกษาวิจัยครั้งนี้มีผลในการวิเคราะห์ข้อมูล โดยมีรายละเอียดดังนี้

5.1 การคัดเลือกคุณลักษณะ (Feature Selection)

จากการศึกษาการวิเคราะห์ในงานวิจัยครั้งนี้ ปัจจัยที่เกี่ยวข้องในการต้อออกของนักศึกษาระดับปริญญาตรี มีทั้งหมด 10 ปัจจัย โดยใช้วิธีการคัดเลือกคุณลักษณะ เพื่อลดมิติของข้อมูล (Attribute Selection) ในการคำนวณหาค่าน้ำหนักของแอททริบิวต์ สอดคล้องกับวิธีการดำเนินงานวิจัยและผลงานวิจัยของ (Laopilai & Sanrach, 2019; Pukkhem, Sitthisam, & Kanjanasamranwong, 2016; Thaweecat *et al.*, 2020) พบว่ามีปัจจัยที่ใกล้เคียงกัน คือ เกรดเฉลี่ย ปีที่เข้าศึกษา และสาขาวิชา เป็นปัจจัยที่สำคัญที่มีส่วนทำให้เกิดการพ้นสภาพของนักศึกษา

5.2 การสร้างแบบจำลอง (Modeling)

การสร้างแบบจำลองในงานวิจัยครั้งนี้ ผู้วิจัยได้ทำการนำปัจจัยที่เกี่ยวข้องในการต้อออกของนักศึกษาระดับปริญญาตรี นำมาสร้างแบบจำลองด้วยเทคนิคเหมือนข้อมูล 3 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคนาอิวเบย์ และกฎอุปนัย โดยการทดสอบประสิทธิภาพของแบบจำลอง ด้วยวิธีการ K-Fold Cross-Validation พบว่าเทคนิคต้นไม้ตัดสินใจ ให้ค่าความถูกต้องสูงที่สุด ซึ่งสอดคล้องกับงานวิจัยของ (Jantakoon, 2006; Panthayak, 2019; Rawengwan & Seresangtakul, 2017) ในการทดสอบประสิทธิภาพแบบจำลองพบว่าเทคนิคต้นไม้ตัดสินใจให้ค่าความถูกต้องสูงที่สุดเช่นเดียวกัน

ทั้งนี้ผู้วิจัยได้นำเสนอ เทคนิคการคัดเลือกคุณลักษณะและเทคนิคการทำเหมืองข้อมูล เพื่อวิเคราะห์ในการหาปัจจัยที่เกี่ยวข้องในการต้อออกและการพยากรณ์ ซึ่งเป็นเพียงบางเทคนิคที่มีหลากหลายอยู่ในปัจจุบัน รวมถึงยังมีขีดจำกัดในเรื่องของปัจจัยที่มีอยู่ไม่เพียงพอ จึงทำให้สามารถทำนาย

ได้เพียงบางสถานะของนักศึกษาเท่านั้น หากมีการเก็บรวบรวมข้อมูลของนักศึกษาเพิ่มขึ้น และเพิ่มข้อมูลปัจจัยที่เกี่ยวข้องของนักศึกษา เช่น สถานภาพของครอบครัว สถานะกู้ยืมระหว่างเรียน และการหารายได้ระหว่างกำลังศึกษา เป็นต้น จะส่งผลในการวิเคราะห์และคัดเลือกข้อมูล เพื่อนำไปสร้างตัวแบบจำลองในการพยากรณ์ และพัฒนาระบบให้มีประสิทธิภาพมากยิ่งขึ้น

6. สรุปผลการวิจัย

การคัดเลือกคุณลักษณะ (Feature Selection) โดยใช้เทคนิค Filter approach การจัดลำดับคุณสมบัติวิเคราะห์ค่าน้ำหนักของแอททริบิวต์ พบว่าปัจจัยที่เกี่ยวข้องในการต้อออกของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏพิบูลสงคราม มีจำนวน 10 ปัจจัย คือ เกรดเฉลี่ยรวม ปีที่เข้าศึกษา สาขาวิชา หลักสูตร อาชีพมารดา อาชีพบิดา คณะ ระดับการศึกษา ภาค และความถนัด/ความสามารถพิเศษ การวัดประสิทธิภาพแบบจำลองด้วยวิธีการ K-Fold Cross-Validation ทั้ง 3 เทคนิค จะเห็นได้ว่าแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจมีค่าประสิทธิภาพดีกว่าแบบจำลองเทคนิคนาอิวเบย์ และกฎการอุปนัย ซึ่งได้ค่าสูงที่สุดจากการแบ่งข้อมูลทดสอบออกเป็น 10 ชุด โดยมีค่าความถูกต้อง (Accuracy) 97.81% และค่าความผิดพลาด (MAE) เท่ากับ 0.026 ผลการทดสอบระบบค่าเฉลี่ยรวมของการทำนาย ความถูกต้องคิดเป็นร้อยละ 86.29% จากการสุ่มตัวอย่างของข้อมูลมาทดสอบในระบบจำนวน 270 ชุดข้อมูล จากผลการทดลองพบว่าปัจจัยที่มีผลต่อการต้อออกของนักศึกษาที่กล่าวมาข้างต้น โดยเฉพาะเกรดเฉลี่ยของนักศึกษา ทำให้ทราบถึงความยากของสาขาวิชาที่นักศึกษา กำลังศึกษาอยู่ ในการให้ความช่วยเหลือควรมีการจัดทำ การสอนเพิ่มเติม หรือการสอบถามนักศึกษา ให้มีส่วนร่วมเกี่ยวกับแนวทางการจัดการเรียนการสอนที่สามารถทำให้นักศึกษาพัฒนาตนเองได้ รวมไปถึงอาจารย์ประจำสาขาวิชาต้องคอยติดตาม กำกับดูแลนักศึกษาอย่าง

ใกล้ขีด และดูแลการลงทะเบียนเรียนของนักศึกษา เพื่อลดความเสี่ยงในการพ้นสภาพการเป็นนักศึกษา และด้านอาชีพบิดาอาชีพมารดาที่ไม่มีรายได้หรือรายได้ไม่เพียงพอต่อค่าใช้จ่ายของนักศึกษา ทางมหาวิทยาลัย อาจมีการช่วยเหลือทางด้านทุนทรัพย์ โดยพิจารณา ผ่อนผันการชำระค่าเทอม หรือการพิจารณานุมัติ การกู้ยืมกองทุนเพื่อการศึกษา การมอบทุนการศึกษา ให้กับนักศึกษาที่มีผลการเรียนดีแต่ขาดแคลน ทุนทรัพย์

7. กิตติกรรมประกาศ

ขอขอบคุณกองบริการการศึกษา มหาวิทยาลัย ราชภัฏพิบูลสงคราม ที่อนุเคราะห์ข้อมูลเพื่อนำมาใช้ ในงานวิจัยในครั้งนี้

8. เอกสารอ้างอิง

- Asif, R., Merceron, A., Ali, S. A., & Haider, N.G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education* 133, 177-194. <https://doi.org/10.1016/j.compedu.2017.05.007>
- Boonprasom, C. & Sanrach, C. (2018). Predictive analytic for student dropout in undergraduate using data mining technique. *Technical Education Journal King Mongkut's University of Technology North Bangkok*, 9(1), 142-151.
- Chareonrat, J. (2016). Analysis on factors affecting normal-grade student dismissal using decision tree. *SNRU Journal of Science and Technology*, 8(2), 256-267.
- Che-doloh, N., Kaheng, F., & Kepan, S. (2018). Decision support system of student retirement: A case study of Faculty of Science and Technology and Agriculture of Yala Rajabhat University. *The 3th Nation Science and Technology Conference, Yala Rajabhat University, Yala, Thailand, February 11-12, 2018*, 310-323.
- Daoruang, B., Sanrach, C. & Mingkhwan, A. (2021). The comparison of data classification efficiency for decision in the selection of information technology students' academic subjects, Faculty of Technology and Industrial Management, King Mongkut's University of Technology North Bangkok, Prachinburi Campus. *Technical Education Journal: King Mongkut's University of Technology North Bangkok*, 12(2), 136-144.
- Jaidee, W. & Wannapee, N. (2020). The study of factors affecting for on-time graduation of ungraduated student using feature selection technique on imbalanced datasets. *Journal of Information Science and Technology*. 10(1), 75-84.

- Jantakoon, T. (2016). Classification model for selection of program studies in Faculty of Information Technology in Rajabhat MahaSarakham University Using Data Mining Techniques. *Proceedings The 9th National Conference on Technical Education, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand, November 24, 2016*, 336-343.
- Jetpipattanapong, D. (2018). Feature selection for human activity classification from skeleton data using two-level selection technique. *KMUTT Research and Development Journal*, 41(4), 401-420.
- Kladchuen, R. & Sanrach, C. (2018). An efficiency comparison of algorithms and feature selection methods to predict the learning achievement of vocational students. *Research Journal Rajamangala University of Technology Thanyaburi*, 17(1), 1-10.
- Kumjit, K., Jaikoomkao, D., Phumirang, W., Sattanako, A., & Sukprasert, A. (2022). The efficiency of data mining technique for the prognosis of Cerebrovascular Disease. *Journal of Applied Informatics and Technology*, 4(2), 87-98.
- Lake, P. (2018). Forecasting foreign exchange rate using time series analysis with data mining techniques. *Apheit Journal (Science and Technology)*, 7(1), 28-45.
- Laopilai, P. & Sanrach, C. (2019). Analysis for student dropout in undergraduate using data mining technique. *The Science Journal of Phetchaburi Rajabhat University*, 16(2), 61-71.
- wichian, P. N., Manair, P., Chuchuen, Y. & Mak-on, S. (2020). Optimization feature selection for classification of manuscript grouping. *Journal of Science and Technology Songkhla Rajabhat University*, 1(1), 55-66.
- Nasritha, K., Kerdprasop, K., & Kerdprasop, N. (2018). Comparison of sampling techniques for imbalanced data classification. *Journal of Applied Informatics and Technology*, 1(1), 20-37.
- Nettleton, D. (2014). *Commercial data mining: Processing, analysis and modeling for predictive analytics projects*. Amsterdam: Elsevier.
- Pacharawongsakda. E. (2015). *Feature selection by information gain method*. Retrieved January 22, 2022, from <https://th.linkedin.com/pulse/การคัดเลือก-feature-selection-ด้วยวิธี-information-gain-pacharawongsakda>.
- Panthayak. C. (2019). *Data analysis of O-NET exam scores case study Bangpakok Wittayakhom School with data mining techniques*. Retrieved January 22, 2022, from <http://coms.kru.ac.th/tee/Projects/ShowPdf?name=637211803876993643.pdf&chk=False>.

- Pukkhem, N., Sitthisarn, S. & Kanjanasamranwong, P. (2016). The Comparison of classification techniques for predicting an on-time graduation. *TSU CONFERENCE 26, Hat Yai, Songkhla, Thailand, May 26-29, 2016*, 253-261.
- Rawengwan, P. & Seresangtakul, P. (2017). A model for forecasting educational status of students. *The National and International Graduate Research Conference 2017*. Pote Sarasin Building: Khon Kaen University, Thailand, March 10, 2017, 1-11.
- Saemmasu, N., Sitimanung, A. & Kepan, S. (2018). Decision support system for teacher and student development: A case study of Yala Province. *The 3th Nation Science and Technology Conference, Yala Rajabhat University, Yala, Thailand, February 11-12*, 324-337.
- Sintawichai. N. (2015). Data Mining. Retrieved January 22, 2022, from <https://slideplayer.in.th/slide/2717498/>.
- Thaweechat, N., Pengprachan, O., Yathongchai, W. & Yathongchai, C. (2020). A prediction system for undergraduate student dropout at Faculty of Science, Buriram Rajabhat University using data mining techniques. *Journal of Science and Technology Buriram Rajabhat University*, 4(1), 47-60.
- Vilailuck, S., Jaroenpuntaruk, V. & Wichadaku, D. (2015). Utilizing data mining techniques to forecast student academic achievement of Kasetsart University Laboratory School Kamphaeng Saen Campus Educational Research and Development Center. *Veridian E-Journal, Science and Technology Silpakorn University*, 2(2), 1-17.