บทความวิจัย

การพยากรณ์การออกกลางคันของนักศึกษามหาวิทยาลัยจากการปรับปรุงด้วยการคัดเลือกคุณลักษณะร่วมกับวิธีโครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้น

อนุวัฒน์ เปพาทย์[1] วงกต ศรีอุไร[2] และณัฏฐ์ ดิษเจริญ[2,*]

[1]นักศึกษาหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี จังหวัดอุบลราชธานี

[2]ภาควิชาคณิตศาสตร์ สถิติและคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี จังหวัดอุบลราชธานี

**Email: nadh.d@ubu.ac.th*

รับบทความ: 1 กุมภาพันธ์ 2565 แก้ไขบทความ: 19 กุมภาพันธ์ 2565 ยอมรับตีพิมพ์: 21 กุมภาพันธ์ 2565

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์หาคุณลักษณะที่เป็นปัจจัยในการพยากรณ์การออกกลางคันของนักศึกษามหาวิทยาลัย เปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์ และปรับปรุงเพื่อเพิ่มประสิทธิภาพของแบบจำลองด้วยการคัดเลือกคุณลักษณะร่วมกับวิธีโครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้น ข้อมูลที่ใช้ในการวิจัยได้มาจากระบบงานทะเบียนนักศึกษาและประมวลผล กองบริการการศึกษา มหาวิทยาลัยอุบลราชธานี ประกอบด้วยข้อมูล 3 ส่วน ได้แก่ (1) ข้อมูลพื้นฐาน จำนวน 1,029 รายการ (2) ข้อมูลรายภาคการศึกษา จำนวน 6,826 รายการ และ (3) ข้อมูลผลการเรียน จำนวน 29,790 รายการ หลังจากผ่านกระบวนการเตรียมข้อมูลตามวิธีการคริปสดีเอ็ม คงเหลือข้อมูลเพื่อสร้างแบบจำลอง จำนวน 882 รายการ และคุณลักษณะจำนวน 14 แอททริบิวต์ แบบจำลองในการพยากรณ์การออกกลางคันของนักศึกษาสร้างด้วยวิธีต้นไม้ตัดสินใจ นาอีฟเบย์ โครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้น และซัพพอร์ตเวกเตอร์แมชชีน ร่วมกับการคัดเลือกคุณลักษณะด้วยวิธีอัตราส่วนเกน ไคสแควร์ และการคัดเลือกคุณลักษณะบนฐานสหสัมพันธ์ (ซีเอฟเอส) หาประสิทธิภาพของแบบจำลองด้วยวิธีการตรวจสอบไขว้สิบส่วน เพื่อเปรียบเทียบค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก ค่าความถ่วงดุล และค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ ผลการวิจัย พบว่า คุณลักษณะที่เป็นปัจจัยสำคัญในการพยากรณ์การออกกลางคันของนักศึกษามี 5 แอททริบิวต์ ได้แก่ (1) ผลการเรียนเฉลี่ยสะสม (2) ผลการเรียนเฉลี่ยสะสมกลุ่มวิชาในคณะ (3) ผลการเรียนเฉลี่ยสะสมกลุ่มวิชานอกคณะ (4) การกู้ยืมหรือทุนการศึกษา และ (5) สถานภาพนักศึกษา (คลาสเป้าหมาย) แบบจำลองในการพยากรณ์การออกกลางคันของนักศึกษาที่มีค่าความถูกต้องสูงที่สุด สร้างด้วยวิธีโครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้นร่วมกับการคัดเลือกคุณลักษณะด้วยวิธีซีเอฟเอส ที่ได้ค่าความถูกต้องหลังจากปรับค่าพารามิเตอร์ในการสร้างแบบจำลองแล้ว เท่ากับ 90.26% จากผลการวิจัยแสดงให้เห็นว่าสามารถใช้เทคนิคการคัดเลือกคุณลักษณะปรับปรุงประสิทธิภาพของแบบจำลองโครงข่ายประสาทเทียมเพื่อใช้ในการพยากรณ์การออกกลางคันของนักศึกษามหาวิทยาลัยได้แม่นยำมากขึ้นและสามารถนำแบบจำลองไปใช้พัฒนาระบบพยากรณ์การออกกลางคันของนักศึกษาต่อไปได้

คำสำคัญ: การออกกลางคัน การพยากรณ์ข้อมูล การคัดเลือกคุณลักษณะ การทำเหมืองข้อมูล

อ้างอิงบทความนี้

อนุวัฒน์ เปพาทย์, วงกต ศรีอุไร และณัฏฐ์ ดิษเจริญ. (2565). การพยากรณ์การออกกลางคันของนักศึกษามหาวิทยาลัยจากการปรับปรุงด้วยการคัดเลือกคุณลักษณะร่วมกับวิธีโครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้น. **วารสารวิทยาศาสตร์และวิทยาศาสตร์ศึกษา**, 5(1), 39-48. http://doi.org/10.14456/jsse.2022.4

Research Article

University student dropout prediction improved by feature selection with Multilayer Perceptron Neural Network

Anuwat Paphat,[1] Wongkot Sriurai,[2] and Nadh Ditcharoen[2,*]

[1]M.Sc. Student, Major of Information Technology, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani

[2]Department of Mathematics, Statistics and Computer, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani

**Email: nadh.d@ubu.ac.th*

Received <1 February 2022>; Revised <19 February 2022>; Accepted <21 February 2022>

Abstract

The objectives of this research were to analyze the factors affected university student dropout prediction, to compare the efficiency of the prediction model, and to improve the model efficiency by using feature selection with multilayer perceptron neural network. Data used in this research were collected from the Office of Student Registrar and Evaluation, Division of Educational Service, Ubon Ratchathani University. They were consisted of 3 parts: (1) 1,029 records of basic data, (2) 6,826 records of semester data, and (3) 29,790 records of grade data. After data preparation process by CRISP-DM method, 882 records remained with 14 attributes. The models of university student dropout prediction were developed using decision tree, Naive Bayes, multilayer perceptron neural network, and support vector machine integrated with feature selection methods including gain ratio, Chi-square, and correlation-based feature selection (CFS). The efficiency of the generated models was measured by 10-folds cross validation to compare accuracy, precision, recall, f-measure, and mean absolute error. The experiment results revealed that the key factors to predict the university student dropout were 5 attributes including (1) grade point average (GPA), (2) GPA of courses in the student's faculty, (3) GPA of courses outside the student's faculty, (4) student loan or scholarship, and (5) student's graduation status (target class). The best model for predicting university student dropout was developed by multilayer perceptron neural network improved with CFS. The accuracy of the prediction model, after model development with parameter tuning, was 90.39%. The results indicated that the feature selection could improve the efficiency of the prediction model developed by neural network for predicting university student dropout accurately. The developed model can be further used to develop the system for predicting university student dropout.

Keywords: Student dropout, prediction, feature selection, data mining

Cite this article:

Paphat, A., Sriurai, W. and Ditcharoen, N. (2022). University student dropout prediction improved by feature selection with Multilayer Perceptron Neural Network (in Thai). **Journal of Science and Science Education**, 5(1), 39-48. http://doi.org/10.14456/jsse.2022.4

บทนำ

การออกกลางคันของนักศึกษาเกิดจากหลายสาเหตุ เช่น ผลการเรียนต่ำกว่าเกณฑ์ ไม่ชำระค่าลงทะเบียน (Ubon Ratchathani University, 2018) หรือการลาออกเองด้วยเหตุผลอื่นๆ เช่น ต้องการเปลี่ยนหลักสูตร เปลี่ยนสถาบันการศึกษา เป็นต้น ซึ่งปัญหาการออกกลางคันของนักศึกษาดังกล่าว อาจส่งผลกระทบต่อการประกันคุณภาพการศึกษา งบประมาณ และการจัดการศึกษา ปัจจุบันมีการเก็บข้อมูลทางการศึกษาอย่างเป็นระบบ ช่วยให้สามารถติดตามผลการศึกษาของนักศึกษา รวมถึงมีระบบตรวจสอบทำนาย พยากรณ์โอกาสที่นักศึกษาอาจจะออกกลางคันเพื่อช่วยให้นักศึกษาได้เตรียมพร้อม หรือหาทางช่วยเหลือก่อนที่จะพ้นสภาพนักศึกษา ซึ่งจะเป็นการลดปัญหาและความเสี่ยงเหล่านั้นได้ จากงานวิจัยที่ผ่านมาพบว่ามีการนำเทคนิคต่างๆ มาใช้ในการวิเคราะห์ทำนาย หรือสร้างแบบจำลอง เช่น การใช้กฎความสัมพันธ์เพื่อวิเคราะห์ความเสี่ยงการออกกลางคันของนักศึกษา หรือหาความสัมพันธ์ของรายวิชาที่มีผลต่อประสิทธิภาพทางการศึกษาของนักศึกษา (Paruechanon and Sriurai, 2018; Pheunpha, 2020) นอกจากนี้ยังพบว่ามีการใช้เทคนิคการจำแนกข้อมูลมาสร้างแบบจำลองเพื่อพยากรณ์การออกกลางคันของนักศึกษา เช่น การใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) วิธีการเรียนรู้เบย์อย่างง่ายหรือนาอีฟเบย์ (Naïve Bayes) วิธีเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors) โครงข่ายประสาทเทียม (Neural Network) และการวิเคราะห์การถดถอยโลจิสติค (Logistic Regression) เป็นต้น (Iamprik and Sudadet, 2017) รวมถึงมีการประยุกต์ใช้การคัดเลือกคุณลักษณะ (Feature Selection) ร่วมกับวิธีการจำแนกข้อมูลเพื่อเพิ่มประสิทธิภาพให้แม่นยำมากขึ้น ดังปรากฏในงานวิจัยของ Rawengwan and Seresangtakul (2017) Sittichat (2017) และ Boonprasom and Sanrach (2018) ที่ประยุกต์ใช้วิธี Filter Ranker Method หรือ Correlation-based Feature Selection (CFS) ในการคัดเลือกคุณลักษณะที่สำคัญร่วมกับการจำแนกข้อมูล งานวิจัยดังกล่าวมาข้างต้น ได้มีการศึกษา รวบรวมปัจจัยหรือคุณลักษณะที่นำมาใช้วิเคราะห์ที่แตกต่างกันไปตามแต่ละบริบทและชุดข้อมูลที่มีอยู่ ซึ่งข้อมูลและกลุ่มตัวอย่างส่งผลต่อการออกกลางคันค่อนข้างหลากหลาย บางปัจจัยอาจส่งผลต่อความแม่นยำในการพยากรณ์แตกต่างกันไป รวมทั้งเทคนิคในการจำแนกหรือทำนายในแต่ละบริบทก็ให้ประสิทธิภาพที่แตกต่างกัน ดังนั้น งานวิจัยนี้จึงนำวิธีการคัดเลือกคุณลักษณะมาประยุกต์ใช้วิเคราะห์ปัจจัยที่มีความสำคัญต่อการออกกลางคันของนักศึกษามหาวิทยาลัยร่วมกับเทคนิคจำแนกข้อมูลในการพยากรณ์การออกกลางคันของนักศึกษา เพื่อเพิ่มประสิทธิภาพของแบบจำลองให้มีความถูกต้องแม่นยำขึ้น

วัตถุประสงค์การวิจัย

1. เพื่อวิเคราะห์หาคุณลักษณะที่เป็นปัจจัยในการพยากรณ์การออกกลางคันของนักศึกษามหาวิทยาลัย
2. เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์การออกกลางคันของนักศึกษามหาวิทยาลัยจากการปรับปรุงด้วยวิธีการคัดเลือกคุณลักษณะร่วมกับวิธีสร้างแบบจำลอง
3. เพื่อเพิ่มประสิทธิภาพของแบบจำลองในการพยากรณ์การออกกลางคันของนักศึกษาด้วยการคัดเลือกคุณลักษณะร่วมกับการสร้างแบบจำลองด้วยวิธีโครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้น

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การคัดเลือกคุณลักษณะ (Feature Selection)

การเลือกคุณลักษณะ เป็นวิธีการที่ช่วยลดจำนวนคุณลักษณะหรือแอททริบิวต์ (Attribute) ซึ่งจะช่วยเพิ่มประสิทธิภาพและความแม่นยำของแบบจำลองในการจำแนกข้อมูล โดยการคัดเลือกคุณลักษณะแบ่งได้ 3 ประเภท ได้แก่ Filter Approach, Wrapper Approach และ Embedded Approach (Pavya and Srinivasan, 2017) วิธี Filter Approach เป็นการเลือกคุณลักษณะโดยไม่ขึ้นกับประเภทของแบบจำลอง ซึ่งจะใช้เป็นเกณฑ์จัดอันดับที่เหมาะสม และตัวแปรที่มีค่าน้ำหนักต่ำกว่าค่าเกณฑ์บางค่าจะถูกลบออก การคำนวณหาค่าน้ำหนักซึ่งอาจจะเป็นค่าความสัมพันธ์ระหว่างแต่ละคุณลักษณะและคลาสเป้าหมาย ข้อดีของวิธีนี้คือมีความรวดเร็ว ไม่ซับซ้อนและไม่ขึ้นกับประเภทของแบบจำลองที่ใช้ แต่มีข้อควรระวังคือการไม่ขึ้นต่อกันของคุณลักษณะ แต่ละคุณลักษณะจะพิจารณาแยกกัน ส่วนในวิธี Wrapper Approach คุณลักษณะจะขึ้นกับแบบจำลองที่ใช้ โดยจะใช้ผลลัพธ์จากแบบจำลองในการพิจารณาความเหมาะสมของคุณลักษณะที่กำหนด และวิธี Embedded Approach จะค้นหาซับเซตของคุณลักษณะที่เหมาะสมที่สุดที่สร้างแบบจำลอง ซึ่งจะใช้การคำนวณที่น้อยกว่าวิธี Wrapper ซึ่งในงานวิจัยนี้ผู้วิจัยเลือกใช้อัลกอริทึมประเภท Filter Approach ในการคัดเลือกคุณลักษณะ คือ วิธีอัตราส่วนเกน (Gain Ratio) วิธีไคสแควร์ (Chi-Square) และวิธีคัดเลือกคุณลักษณะบนฐานสหสัมพันธ์ (Correlation-based Feature Selection: CFS)

ต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจ เป็นเทคนิคหนึ่งที่ให้ผลลัพธ์และอธิบายความสัมพันธ์ได้ง่าย พัฒนาขึ้นมาโดย J. Ross Quinlan เทคนิคที่นิยมใช้ในการสร้างต้นไม้ตัดสินใจ ได้แก่ อัลกอริทึม C4.5 (J48) ใช้หลักการของ Information Gain (IG) หรือ Entropy Reduction เพื่อจำแนกโหนด (Node) โดยคัดเลือกคุณลักษณะที่มีความสัมพันธ์กับคลาสมากที่สุดมาเป็นราก (Root Node) จากนั้นหาคุณลักษณะไป

เรื่อยๆ โหนดถัดไปจะมีค่า Gain ลดหลั่นกันไป ซึ่งแต่ละโหนดจะแสดงถึงการตัดสินใจบนข้อมูลของคุณสมบัติต่างๆ ของกิ่งไม้ โดยข้อมูลชั้นล่างสุดของต้นไม้ตัดสินใจจะแสดงถึงกลุ่มของข้อมูล (Class) (Pacharawongsakda, 2014; Habusaya and Ditcharoen, 2020) ในการหาความสัมพันธ์ของคุณลักษณะจะใช้ค่า IG ซึ่งคำนวณได้จาก

$$IG(parent, child) = Entropy(parent) - [p(c_1) \times Entropy(c_1) + p(c_2) \times Entropy(c_2) + ...] \quad (1)$$

โดยที่

$Entropy(c_i)$ คือ $-p(c_i) \log p(c_i)$

$p(c_i)$ คือ ค่าความน่าจะเป็นของค่า c_i

c คือ คลาสเป้าหมาย (Class)

ค่า Entropy จะใช้ในการวัดความแตกต่างกันของข้อมูล ถ้าข้อมูลมีความแตกต่างกันน้อย ค่า Entropy จะมีค่าต่ำ แต่ถ้าข้อมูลมีความแตกต่างกันมาก ค่า Entropy จะมีค่าสูง ดังนั้นถ้าข้อมูล Entropy ของโหนดลูก (child) สามารถแบ่งแยกข้อมูลได้ดี จะมีค่า Entropy ต่ำและจะทำให้ค่า IG มีค่าสูงเมื่อเทียบกับโหนดบน (parent)

นาอีฟเบย์ (Naive Bayes)

นาอีฟเบย์ เป็นเทคนิคหนึ่งที่ได้รับความนิยมเนื่องจากสร้างแบบจำลองได้ง่ายและไม่ซับซ้อน โดยอาศัยทฤษฎีความน่าจะเป็นเป็นหลัก (Pacharawongsakda, 2014; Techapanurak, 2021) ซึ่งหาคำนวณได้จาก

$$P(C|A) = \frac{P(A|C)P(C)}{P(A)} \quad (2)$$

โดยที่

A คือ คุณลักษณะ (Attributes)

C คือ คลาสเป้าหมาย (Class)

P(C|A) คือ ค่าความน่าจะเป็นที่ข้อมูลที่มีแอททริบิวต์เป็น A จะมีคลาส C

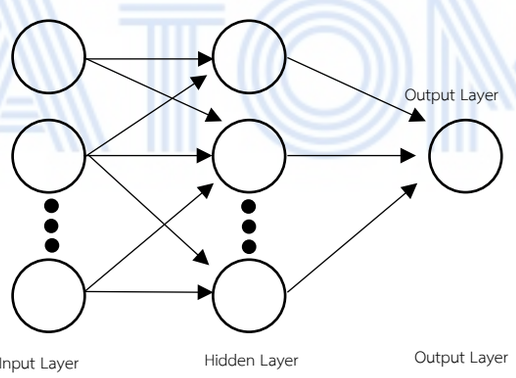
P(A|C) คือ ค่าความน่าจะเป็นที่ข้อมูลฝึกฝน (Training Data) ที่มีคลาส C และมีแอททริบิวต์ A

โดยที่ $A = a_1 \cap a_2 \ldots \cap a_M$ และ M คือ จำนวนแอททริบิวต์ในชุดข้อมูลฝึกฝน

P(C) คือ ค่าความน่าจะเป็นของคลาส C

โครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้น (Multilayer Perceptron Neural Network)

โครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้น เป็นโครงข่ายประสาทเทียมแบบ Feed Forward ซึ่งเป็นแบบจำลองทางคณิตศาสตร์ที่ใช้อย่างแพร่หลาย และได้รับความสนใจนำมาศึกษาวิจัยอย่างมากในหลากหลายสาขาวิชา พัฒนาขึ้นเพื่อจำลองการทำงานของโครงข่ายประสาทในสมองมนุษย์ประกอบด้วยเซลล์ประสาทเทียมหรือโหนดจำนวนมากเชื่อมต่อกัน ซึ่งการเชื่อมต่อแบ่งออกเป็นกลุ่มย่อย เรียกว่า ชั้น (Layer) ชั้นแรก เป็นชั้นรับข้อมูลป้อนเข้า (Input Layer) ส่วนชั้นสุดท้ายเรียกว่า ชั้นส่งข้อมูลออก (Output Layer) และชั้นที่อยู่ระหว่างชั้นรับข้อมูลป้อนเข้าและชั้นส่งข้อมูลออกเรียกว่า ชั้นซ่อน (Hidden Layer) ซึ่งโดยทั่วไปชั้นซ่อนอาจมีมากกว่า 1 ชั้นก็ได้ โครงสร้างแสดงดังภาพที่ 1 ด้วยเหตุนี้ จึงสามารถแบ่งประเภทของโครงข่ายประสาทเทียมตามจำนวนชั้นของโครงข่ายแบบกว้าง ๆ ได้ 2 แบบ ได้แก่ โครงข่ายแบบชั้นเดียว (Single Layer) และ โครงข่ายแบบหลายชั้น (Multilayer) (Pacharawongsakda, 2014)



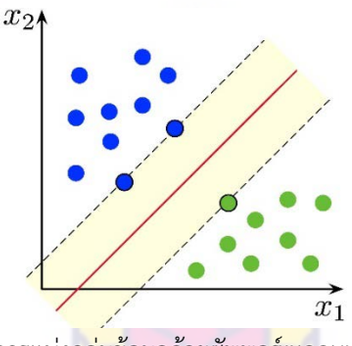
ภาพที่ 1 โครงสร้างและการทำงานของโครงข่ายประสาทเทียม
ที่มา : https://th.wikipedia.org/wiki/โครงข่ายประสาทเทียม

โครงข่ายประสาทเทียมแบบหลายชั้น ใช้สำหรับงานที่มีความซับซ้อนได้ผลเป็นอย่างดี โดยมีกระบวนการฝึกฝนเป็นแบบมีผู้สอน (Supervised Learning) และใช้ขั้นตอนการส่งค่าย้อนกลับ (Backpropagation) การฝึกฝนกระบวนการส่งค่าย้อนกลับ

ประกอบด้วย 2 ส่วนย่อยคือ การส่งผ่านไปข้างหน้า (Forward Pass) การส่งผ่านย้อนกลับ (Backward Pass) สำหรับการส่งผ่านไปข้างหน้า ข้อมูลจะผ่านเข้าโครงข่ายประสาทเทียมที่ชั้นข้อมูลเข้า และจะส่งผ่าน จากอีกชั้นหนึ่งไปสู่อีกชั้นหนึ่งจนกระทั่งถึงชั้นข้อมูลออก ส่วนการส่งผ่านย้อนกลับค่าน้ำหนักการเชื่อมต่อจะถูกปรับเปลี่ยนให้สอดคล้องกับกฎการแก้ข้อผิดพลาด (Error-Correction) คือผลต่างของผลตอบที่แท้จริง (Actual Response) กับผลตอบเป้าหมาย (Target Response) เกิดเป็นสัญญาณผิดพลาด (Error Signal) ซึ่งสัญญาณผิดพลาดนี้จะถูกส่งย้อนกลับเข้าสู่โครงข่ายประสาทเทียมในทิศทางตรงกันข้ามกับการเชื่อมต่อ และค่าน้ำหนักของการเชื่อมต่อจะถูกปรับจนกระทั่งผลตอบที่แท้จริงเข้าใกล้ผลตอบเป้าหมาย (Prakobpol, 2009)

ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM)

ซัพพอร์ตเวกเตอร์แมชชีน เป็นอัลกอริทึมในกลุ่มวิธีการเรียนแบบมีผู้สอน (Supervised Learning) โดยการนำค่าของกลุ่มข้อมูลวางลงในฟีเจอร์สเปซ (Feature Space) จากนั้นจึงหาเส้นที่ใช้แบ่งข้อมูลทั้งสองออกจากกันโดยสร้างเส้นแบ่ง (Hyperplane) ที่เป็นเส้นตรงขึ้นมาเพื่อแบ่งกลุ่มข้อมูลสองกลุ่มออกจากกัน ดังภาพที่ 2 เดิมซัพพอร์ตเวกเตอร์แมชชีนนำมาใช้กับข้อมูลเชิงเส้น มีประสิทธิภาพในการจำแนกข้อมูลที่มีมิติจำนวนมากได้ นอกจากนี้การใช้ฟังก์ชันเคอร์เนล (Kernel Function) เพื่อแปลงข้อมูลไปยังมิติที่สูงขึ้น สามารถจำแนกข้อมูลที่มีความคลุมเครือได้อย่างมีประสิทธิภาพ โดยใช้หลักการหาเส้นตรงที่มีมาร์จินที่โตที่สุด (Maximum Margin) ที่สามารถแบ่งข้อมูลออกเป็น 2 คลาส โดยมีความผิดพลาดน้อยที่สุด (Boonprasom and Sanrach, 2018)



ภาพที่ 2 การแบ่งกลุ่มข้อมูลด้วยซัพพอร์ตเวกเตอร์แมชชีน
ดัดแปลงจาก: https://en.wikipedia.org/wiki/Support-vector_machine

วิธีดำเนินการวิจัย

งานวิจัยนี้ใช้กระบวนการคริปสดีเอ็ม (CRISP-DM: Cross-Industry Standard Process for Data Mining) (Pacharawongsakda, 2014) ซึ่งประกอบด้วยขั้นตอนการดำเนินงาน ดังนี้

การทำความเข้าใจปัญหา (Problem Understanding)

ผู้วิจัยวิเคราะห์ปัญหาการออกกลางคันของนักศึกษาจากข้อมูลพื้นฐานของนักศึกษา การลงทะเบียนในแต่ละภาคเรียน และผลการเรียน โดยอ้างอิงจากระเบียบข้อบังคับของมหาวิทยาลัยอุบลราชธานี (Ubon Ratchathani University, 2018) โครงสร้างหลักสูตรของนักศึกษา เพื่อหาปัจจัยที่ส่งผลต่อการออกกลางคันของนักศึกษา จากชุดข้อมูลการลงทะเบียนเรียนของนักศึกษา

การทำความเข้าใจข้อมูล (Data Understanding)

ข้อมูลที่ใช้ในการวิจัยได้มาจากระบบงานทะเบียนนักศึกษาและประมวลผล กองบริการการศึกษา มหาวิทยาลัยอุบลราชธานี ระหว่างปีการศึกษา 2553-2563 ในรูปแบบไฟล์ Excel ประกอบด้วยข้อมูล 3 ส่วน ได้แก่

1. ข้อมูลพื้นฐาน จำนวน 1,029 รายการ ประกอบด้วย รหัสนักศึกษา ผลการเรียนเฉลี่ยสะสม รหัสสถานภาพ สถานภาพนักศึกษา โรงเรียนเดิม ผลการเรียนเฉลี่ยแรกเข้า อาชีพบิดา รายได้บิดา อาชีพมารดา รายได้มารดา และจำนวนพี่น้อง

2. ข้อมูลรายภาคการศึกษา จำนวน 6,826 รายการ ประกอบด้วย รหัสนักศึกษา ปีการศึกษา ภาคเรียน รหัสสถานภาพสถานภาพนักศึกษา ผลการเรียนเฉลี่ย และผลการเรียนเฉลี่ยสะสม

3. ข้อมูลผลการเรียน จำนวน 29,790 รายการ ประกอบด้วย รหัสนักศึกษา ปีการศึกษา ภาคเรียน รหัสวิชา ชื่อวิชา หน่วยกิต และระดับผลการเรียน

การเตรียมข้อมูล (Data Preparation)

เป็นขั้นตอนการจัดการข้อมูลเพื่อเตรียมพร้อมสำหรับการวิเคราะห์และสร้างแบบจำลองในการพยากรณ์ข้อมูล ได้แก่

1. การคัดเลือกข้อมูล (Data Selection) โดยเลือกคุณลักษณะ (ปัจจัยหรือแอททริบิวต์) จากข้อมูลต้นฉบับ ซึ่งผู้วิจัยคัดเลือกข้อมูลจากทั้งสามส่วนโดยตัดข้อมูลส่วนซ้ำซ้อนออก คงเหลือข้อมูลดังนี้ รหัสนักศึกษา ผลการเรียนเฉลี่ยสะสม สถานภาพนักศึกษา โรงเรียนเดิม ผลการเรียนเฉลี่ยแรกเข้า อาชีพบิดา รายได้บิดา อาชีพมารดา รายได้มารดา จำนวนพี่น้อง ปีการศึกษา ภาคเรียน รหัสวิชา ชื่อวิชา หน่วยกิต ระดับผลการเรียน

2. การกลั่นกรองข้อมูล (Data Cleaning) เป็นการตัดข้อมูลที่ผิดพลาดและไม่สมบูรณ์ออก คงเหลือข้อมูลพื้นฐานที่ใช้สำหรับนำไปประมวลผลต่อได้จำนวน 1,025 รายการ

3. การแปลงข้อมูล (Data Transformation) ให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ต่อได้ ดังนี้

3.1 รวมข้อมูลจากสามส่วนให้เป็นตารางเดียวกัน โดยใช้ข้อมูลพื้นฐานเป็นตารางหลัก

3.2 กำหนดแอททริบิวต์ระยะเวลาการศึกษาซึ่งคำนวนจากแอททริบิวต์ปีการศึกษา ภาคเรียน ตรวจสอบระยะเวลาว่าเป็นไปตามแผนการศึกษาหรือไม่ ทั้งนี้ใช้ข้อมูลจากแอททริบิวต์ประเภทของสถานศึกษาเดิมสำหรับพิจารณาระยะเวลาศึกษาตามแผน

3.3 รวมแอททริบิวต์ระดับผลการเรียนรายวิชา คำนวณผลการเรียนเฉลี่ยรายวิชา แล้วจัดเป็น 3 กลุ่ม คือ (1) กลุ่มรายวิชาในคณะ (2) กลุ่มรายวิชานอกคณะ (3) รายวิชากลุ่มภาษา และแยกเป็นแอททริบิวต์รายวิชาที่ได้รับผลการเรียน W

3.4 แทนค่าและแปลงข้อมูลทั้งหมดให้อยู่ในรูปแบบ Nominal ทั้งหมด

3.5 จัดกลุ่มสถานภาพนักศึกษาเป็น 2 กลุ่มคือ สำเร็จการศึกษา พ้นสภาพนักศึกษา และกำหนดเป็นคลาสเป้าหมาย

เมื่อผ่านการแปลงข้อมูลแล้ว คงเหลือรายการทั้งสิ้นจำนวน 882 รายการ (จาก 1,025 รายการ) และผลของการเตรียมข้อมูลทำให้คงเหลือคุณลักษณะ (แอททริบิวต์) ทั้งหมด 14 แอททริบิวต์ ดังตารางที่ 1

ตารางที่ 1 รายการคุณลักษณะที่ใช้ในการวิเคราะห์ข้อมูล

คุณลักษณะ	คำอธิบาย	ค่าข้อมูลที่เป็นไปได้
GPA	ผลการเรียนเฉลี่ยสะสม	low = 0.00-1.50, mid = 1.51-2.50, hi = 2.51-4.00
GPAin	ผลการเรียนเฉลี่ยแรกเข้า	low = 0.00-1.50, mid = 1.51-2.50, hi = 2.51-4.00
oldSchool	สถานศึกษาเดิม	school=โรงเรียน, collage = วิทยาลัย, nfe = กศน.
fathercareer	อาชีพบิดา	unknown = ไม่ระบุ noneincome = ไม่มีเงินได้ own_business = ค้าขาย/ธุรกิจส่วนตัว agriculture = เกษตรกร/ประมง freelance = อาชีพอิสระ/รับจ้าง government = รับราชการ employee = พนักงาน/ลูกจ้างหน่วยงานเอกชน emp_government = พนักงานราชการ/ลูกจ้างหน่วยงานราชการ state_enterprises = รัฐวิสาหกิจ etc = อื่นๆ
mothercareer	อาชีพมารดา	
fatherincome	รายได้บิดา	unknown = ไม่ระบุ none = ไม่มีรายได้ low < 150,000 บาทต่อปี medium = 150,000-300,000 บาทต่อปี high > 300,000 บาทต่อปี
motherincome	รายได้มารดา	
sibling	จำนวนพี่น้อง	none = ไม่มีพี่น้อง, one = 1 คน, two = 2 คน, three = 3 คน, morethan.three = 3 คนขึ้นไป
loan	สถานะทุน/กู้ยืม	yes = ได้รับทุน/กู้ยืม, no = ไม่ได้รับทุน/กู้ยืม
havew	มีการถอนรายวิชา (ได้เกรด W)	yes = มีผลการเรียน W, no = ไม่มีผลการเรียน W
GPAInside	ผลการเรียนเฉลี่ยกลุ่มวิชาในคณะ	low = 0.00-1.50, mid = 1.51-2.50, hi = 2.51-4.00
GPAoutside	ผลการเรียนเฉลี่ยกลุ่มวิชานอกคณะ	low = 0.00-1.50, mid = 1.51-2.50, hi = 2.51-4.00
GPAlanguage	ผลการเรียนเฉลี่ยกลุ่มวิชาภาษาต่างประเทศ	low = 0.00-1.50, mid = 1.51-2.50, hi = 2.51-4.00
graduate	สถานภาพนักศึกษา	yes = สำเร็จการศึกษา, no = พ้นสภาพนักศึกษา

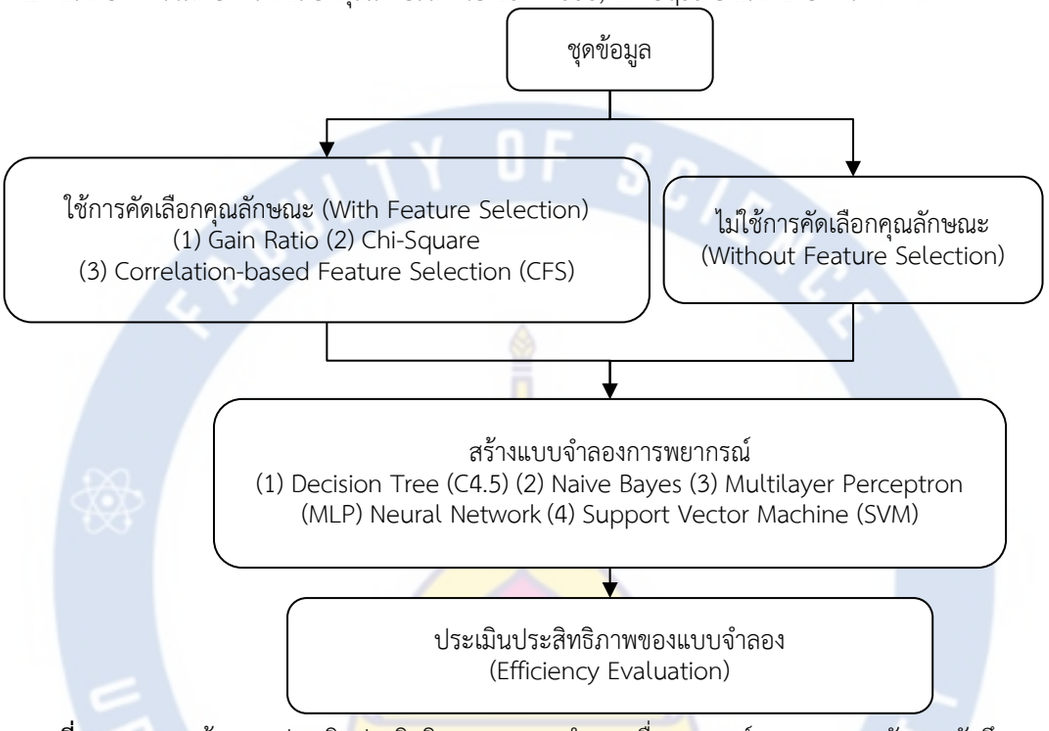
การสร้างแบบจำลอง (Modeling)

ผู้วิจัยใช้โปรแกรม Weka ในการสร้างแบบจำลอง และใช้เทคนิคการวิเคราะห์ค่าน้ำหนักด้วยวิธีการ Gain Ratio, Chi-Square และ Correlation-based Feature Selection (CFS) ในการคัดเลือกคุณลักษณะที่สำคัญร่วมกับการสร้างแบบจำลองเพื่อเปรียบเทียบ

ประสิทธิภาพแบบจำลอง ได้แก่ (1) Decision Tree (C4.5) (2) Naive Bayes (3) Multilayer Perceptron (MLP) Neural Network (4) Support Vector Machine (SVM) ดังแสดงในภาพที่ 3

การหาประสิทธิภาพของแบบจำลอง (Evaluation)

ผู้วิจัยหาประสิทธิภาพของแบบจำลองด้วยวิธี K-Fold Cross Validation (10-Folds) โดยเปรียบเทียบค่า Accuracy, Precision, Recall, F-Measure และ Mean Absolute Error (MAE) จากการทดสอบแบบจำลองที่สร้างด้วยวิธี Decision Tree (C4.5), Naive Bayes, MLP และ SVM ร่วมกับการคัดเลือกคุณลักษณะด้วย Gain Ratio, Chi-Square และ CFS ดังภาพที่ 3



ภาพที่ 3 กระบวนสร้างและประเมินประสิทธิภาพของแบบจำลองเพื่อพยากรณ์การออกกลางคันของนักศึกษา

ผู้วิจัยออกแบบการทดลองที่สอดคล้องกับวัตถุประสงค์การวิจัย ดังนี้

1. สร้างแบบจำลองด้วยวิธี Decision Tree (C4.5), Naive Bayes, MLP และ SVM จากคุณลักษณะทั้งหมด 14 แอททริบิวต์ (ดังตารางที่ 1) แล้วเปรียบเทียบประสิทธิภาพด้วยค่า Accuracy, Precision, Recall, F-Measure และ MAE

2. คัดเลือกคุณลักษณะด้วยวิธี Gain Ratio, Chi-Square โดยใช้เกณฑ์การค้นหาและจัดอันดับ (Ranker) แบบ GreedyStepwise และวิธี CFS ซึ่งใช้ Filters AttributeSelection ของโปรแกรม Weka

3. สร้างแบบจำลองด้วยวิธี Decision Tree (C4.5), Naive Bayes, MLP และ SVM จากคุณลักษณะที่ได้จากการคัดเลือกคุณลักษณะซึ่งพิจารณาจำนวนแอททริบิวต์ที่ได้จากการจัดอันดับด้วยวิธี Gain Ratio และ Chi-Square ในสัดส่วน 50% ของแอททริบิวต์ทั้งหมด และจำนวนที่เท่ากับผลที่ได้จาก CFS ที่มีค่ามากกว่า 0% แล้วเปรียบเทียบประสิทธิภาพด้วยค่า Accuracy, Precision, Recall, F-Measure และ MAE

4. ปรับค่าพารามิเตอร์ของวิธีสร้างแบบจำลองในข้อ 3 ที่มีค่า Accuracy สูงที่สุด เพื่อเพิ่มประสิทธิภาพของแบบจำลองในการพยากรณ์การออกกลางคันของนักศึกษา แล้วเปรียบเทียบประสิทธิภาพด้วยค่า Accuracy, Precision, Recall, F-Measure และ MAE

การนำแบบจำลองไปใช้งาน (Deployment)

เมื่อได้แบบจำลองที่ถูกต้องแม่นยำแล้ว สามารถนำไปพัฒนาระบบพยากรณ์การออกกลางคัน ในรูปแบบแอปพลิเคชัน ให้นักศึกษา ผู้บริหาร ผู้รับผิดชอบหลักสูตร ใช้ประกอบการตัดสินใจในการบริหารหลักสูตร จัดการเรียนการสอน รวมถึงบริหารงบประมาณต่อไป

ผลการวิจัย

ผลการวิจัย แบ่งออกเป็น 4 ส่วน ดังนี้

1. ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองที่สร้างด้วยวิธี Decision Tree (C4.5), Naive Bayes, MLP และ SVM จากคุณลักษณะทั้งหมด 14 แอททริบิวต์ (ไม่ได้ใช้การคัดเลือกคุณลักษณะ) ได้ผลแสดงดังตารางที่ 2 ซึ่งพบว่าวิธีต้นไม้ตัดสินใจ (C4.5) ให้ผลความถูกต้องสูงที่สุด

ตารางที่ 2 ผลการหาประสิทธิภาพของแบบจำลองที่ไม่ได้ใช้การคัดเลือกคุณลักษณะ

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F-Measure	MAE
Naïve Bayes	87.59	86.90	93.80	87.40	0.1384
Decision Tree (C4.5)	**88.81**	**89.30**	**88.80**	**88.50**	**0.1597**
Multilayer Perceptron	86.86	86.80	86.90	86.80	0.1418
Support Vector Machine	88.44	89.90	88.40	88.00	0.1156

2. ผลการคัดเลือกคุณลักษณะด้วยวิธี Gain Ratio (GR) และ Chi-Square (CS) โดยการจัดอันดับ (Ranker) และวิธี Correlation-based Feature Selection (CFS) ซึ่งพบว่า GR และ CS ได้ผลลัพธ์ใกล้เคียงกันโดยแอททริบิวต์ 6 อันดับแรก คือ GPA, GPAInside, GPAOutside, GPALanguage, loan และ sibling ส่วนวิธี CFS ที่มีค่าน้ำหนักมากกว่า 0% มีเพียง 4 แอท-ทริบิวต์ ได้แก่ GPA, GPAInside, GPAOutside และ loan ดังตารางที่ 3 ทั้งนี้แอททริบิวต์ที่ 14 คือคลาสเป้าหมาย

ตารางที่ 3 ผลการคัดเลือกคุณลักษณะด้วยวิธี Gain Ratio (GR), Chi-Square (CS) และ CFS

Gain Ratio (GR)	Chi-Square (CS)	CFS
จัดลำดับ		**ไม่จัดลำดับ**
1. GPA	1. GPA	1. GPA (100%)
2. GPAInside	2. GPAInside	2. GPAInside (100%)
3. GPAOutside	3. GPAOutside	3. GPAOutside (90%)
4. GPALanguage	4. GPALanguage	4. loan (10%)
5. loan	5. sibling	5. sibling (0%)
6. sibling	6. loan	6. GPALanguage (0%)
7. oldschool	7. mothercareer	7. oldschool (0%)
8. havew	8. oldschool	8. havew (0%)
9. mothercareer	9. havew	9. mothercareer (0%)
10. GPAin	10. fathercareer	10. GPAin (0%)
11. fathercareer	11. fatherincome	11. fathercareer (0%)
12. fatherincome	12. GPAin	12. fatherincome (0%)
13. motherincome	13. motherincome	13. motherincome (0%)

3. ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองที่สร้างด้วยวิธี Decision Tree (C4.5), Naive Bayes, MLP และ SVM จากแอททริบิวต์ที่ได้จากการคัดเลือกคุณลักษณะโดยพิจารณาจำนวน 7 แอททริบิวต์ และ 5 แอททริบิวต์ ได้ผลแสดงดังตารางที่ 4 ซึ่งพบว่าวิธี MLP ที่ผ่านการคัดเลือกคุณลักษณะด้วยวิธี CFS และลดจำนวนคุณลักษณะเหลือ 5 แอททริบิวต์ให้ผลความถูกต้องสูงที่สุด

4. จากผลการทดลองในข้อ 3 ผู้วิจัยจึงปรับค่าพารามิเตอร์ในการสร้างแบบจำลองด้วยวิธี MLP เพื่อสร้างและทดสอบหาประสิทธิภาพของแบบจำลอง ดังนี้

4.1 ปรับค่า Learning Rate ให้มีค่าเท่ากับ 0.3, 0.5 และ 0.7 ตามลำดับ

4.2 ปรับจำนวนโหนดชั้นซ่อน (Hidden Layer) ของ MLP ให้มีค่าเท่ากับ 1, 3 และ 5 ตามลำดับ

4.3 กำหนดค่า Momentum เท่ากับ 0.2

4.4 กำหนดค่าชั้นรับข้อมูลป้อนเข้า (Input Layer) เท่ากับ 5

4.5 กำหนดค่าชั้นส่งข้อมูลออก (Output Layer) เท่ากับ 2

ซึ่งผลการทดลอง พบว่า แบบจำลองที่มีผลความถูกต้องสูงที่สุด คือ แบบจำลองที่สร้างจาก Input Layer = 5, Hidden Layer = 1, Output Layer = 2, Learning Rate = 0.5 และ Momentum = 0.2 ดังตารางที่ 5

ตารางที่ 4 ผลการหาประสิทธิภาพของแบบจำลองร่วมกับการคัดเลือกคุณลักษณะ

Algorithm	Feature Selection	#Attribute	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)	MAE
Naïve Bayes	GR/CS	7	88.20	88.40	88.20	88.00	0.1366
		5	87.23	87.20	87.20	87.10	0.1419
	CFS	5	88.69	89.00	88.70	88.50	0.1338
Decision Tree (C4.5)	GR/CS	7	88.56	90.10	88.60	88.10	0.1611
		5	88.81	90.40	88.80	88.40	0.1602
	CFS	5	88.81	90.40	88.80	88.40	0.1602
Multilayer Perceptron	GR/CS	7	89.17	89.20	89.20	89.10	0.1389
		5	89.05	89.50	89.10	88.80	0.1397
	CFS	**5**	**89.90**	**90.90**	**89.90**	**89.60**	**0.1431**
Support Vector Machine	GR/CS	7	88.44	89.90	88.40	88.00	0.1156
		5	88.44	89.90	88.40	88.00	0.1156
	CFS	5	88.44	89.90	88.40	88.00	0.1156

ตารางที่ 5 ผลการหาประสิทธิภาพของแบบจำลองที่สร้างด้วยวิธี MLP ร่วมกับการคัดเลือกคุณลักษณะด้วยวิธี CFS

Model	Learning Rate	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)	MAE
5:1:2	0.3	90.15	91.00	90.10	89.90	0.1689
	0.5	**90.26**	**91.10**	**90.30**	**90.00**	**0.1675**
	0.7	89.90	90.50	89.90	89.70	0.1708
5:3:2	0.3	90.02	91.00	90.00	89.70	0.1450
	0.5	90.02	91.00	90.00	89.70	0.1405
	0.7	89.90	91.10	89.90	89.60	0.1391
5:5:2	0.3	90.02	91.00	90.00	89.70	0.1432
	0.5	90.02	91.00	90.00	89.70	0.1394
	0.7	89.90	90.80	89.99	89.60	0.1372

สรุปและอภิปรายผลการวิจัย

งานวิจัยนี้ได้ประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลในการวิเคราะห์ปัจจัยที่ส่งผลต่อการออกกลางคันของนักศึกษามหาวิทยาลัย ผู้วิจัยศึกษาและเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์การออกกลางคันของนักศึกษา ปรับปรุงแบบจำลองเพื่อเพิ่มประสิทธิภาพด้วยการคัดเลือกคุณลักษณะ และปรับค่าพารามิเตอร์ที่ทำให้แบบจำลองที่พัฒนาด้วยโครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้นมีความถูกต้องและแม่นยำในการพยากรณ์สูงที่สุด โดยการวิเคราะห์ข้อมูลดำเนินการตามขั้นตอน CRISP-DM ประกอบด้วย 6 ขั้นตอน ได้แก่ (1) การทำความเข้าใจปัญหา (2) การทำความเข้าใจข้อมูล (3) การเตรียมข้อมูล (4) การสร้างแบบจำลอง (5) การหาประสิทธิภาพของแบบจำลอง และ (6) การนำแบบจำลองไปใช้งาน ข้อมูลที่นำมาใช้พัฒนาแบบจำลองรวบรวมมาจากงานทะเบียนนักศึกษาและประมวลผล กองบริการการศึกษา มหาวิทยาลัยอุบลราชธานี จำนวน 1,029 รายการ หลังจากผ่านกระบวนการเตรียมข้อมูลลดลงเหลือ 882 รายการ ประกอบด้วย 14 แอททริบิวต์ ซึ่งจำแนกข้อมูลออกเป็น 2 กลุ่ม คือ กลุ่มสำเร็จการศึกษา และกลุ่มพ้นสถานภาพนักศึกษา วิธีที่ใช้ในการสร้างแบบจำลอง ได้แก่ (1) Naive Bayes (2) Decision Tree (C4.5) (3) Multilayer Perceptron (MLP) Neural Network และ (4) Support Vector Machine เมื่อใช้ร่วมกับการคัดเลือกคุณลักษณะด้วยวิธี Gain Ratio, Chi-Square และ Correlation-based Feature Selection (CFS) พบว่า ความถูกต้อง (Accuracy) ของแบบจำลองเพิ่มขึ้นเมื่อแอททริบิวต์ลดลงเหลือ 5 แอททริบิวต์ และใช้ร่วมกับการคัดเลือกด้วยวิธี CFS มีความถูกต้องสูงที่สุด โดยเฉพาะเมื่อใช้ CFS ร่วมกับ MLP จะได้ค่าความถูกต้องในการพยากรณ์การออกกลางคันของนักศึกษาเท่ากับ 89.90% และถึงแม้ว่าการปรับค่าพารามิเตอร์ของ MLP จะไม่ส่งผลต่อความถูกต้องมากนัก แต่ค่าที่ดีที่สุดที่ทำให้ความถูกต้องของแบบจำลองเพิ่มขึ้นเป็น 90.26% คือการปรับค่า

Learning Rate เท่ากับ 0.5 ซึ่งสอดคล้องกับงานวิจัยของ Sriurai (2014) ที่มีการเปรียบเทียบแบบจำลองโครงข่ายประสาทเทียมร่วมกับการคัดเลือกคุณลักษณะแบบต่างๆ ในการจำแนกผู้ป่วยโรคอ้วนลงพุง ซึ่งการคัดเลือกคุณลักษณะด้วยวิธี CFS ให้ค่าความถูกต้องสูงที่สุด จากผลการวิจัย พบว่า ปัจจัยสำคัญที่มีผลต่อการออกกลางคันของนักศึกษามหาวิทยาลัย ได้แก่ ผลการเรียนเฉลี่ยสะสม (GPA) ผลการเรียนเฉลี่ยสะสมกลุ่มวิชาในคณะ (GPAInside) ผลการเรียนเฉลี่ยสะสมกลุ่มวิชานอกคณะ (GPAOutside) นอกจากนี้ยังพบว่าผลการเรียนเฉลี่ยสะสมกลุ่มวิชาภาษา (GPALanguage) การกู้ยืมหรือทุนการศึกษา (loan) ก็เป็นปัจจัยรองที่มีผลต่อการออกกลางคันของนักศึกษาเช่นกัน และยังมีปัจจัยของจำนวนพี่น้อง (sibling) ที่ยังคงต้องศึกษาให้แน่ชัดต่อไป ทั้งนี้สามารถกล่าวโดยสรุป จากผลการวิจัยแสดงให้เห็นว่าสามารถใช้เทคนิคการคัดเลือกคุณลักษณะเพื่อปรับปรุงประสิทธิภาพของแบบจำลองในการพยากรณ์การออกกลางคันของนักศึกษา และยังสามารถเพิ่มประสิทธิภาพของแบบจำลองโครงข่ายประสาทเทียมให้มีความถูกต้องแม่นยำมากขึ้น ซึ่งสามารถนำแบบจำลองไปใช้พัฒนาระบบการพยากรณ์การออกกลางคันของนักศึกษามหาวิทยาลัยได้

งานวิจัยนี้สามารถพัฒนาต่อยอดได้โดยการวิเคราะห์ความสัมพันธ์ของปัจจัยที่มีผลต่อการออกกลางคันของนักศึกษาซึ่งอาจประยุกต์ใช้เทคนิคกฎความสัมพันธ์ของวิธีการทำเหมืองข้อมูล และยังสามารถพัฒนาต่อยอดเพื่อสร้างแบบจำลองในการพยากรณ์การสำเร็จการศึกษาภายในระยะเวลาที่หลักสูตรกำหนดเพื่อให้ผู้เกี่ยวข้องสามารถติดตามสถานภาพการศึกษาของนักศึกษา ให้ความช่วยเหลือนักศึกษาให้สำเร็จการศึกษาตามแผนการศึกษาได้

เอกสารอ้างอิง

Boonprasom, C. and Sanrach, C. (2018). Predictive analytic for student dropout in undergraduate using data mining technique (in Thai). **Technical Education Journal King Mongkut's University of Technology North Bangkok**, 9(1), 142-151.

Habusaya, S. and Ditcharoen, N. (2020). Analysis of ICD-10 coding errors in 43 files database systems for medical record department using data mining techniques. **KKU Science Journal,** 48(1), 142 – 155.

Iamprik, C. and Sudadet, K. (2017). Predicting student's data for performance by data mining. **Proceedings of The Thirteenth National Conference on Computing and Information Technology NCCIT2017** (pp. 32-37).

Pacharawongsakda, E. (2014). **An introduction to data mining techniques** (in Thai). Bangkok: Asia Digital Press Company Limited.

Paruechanon, P. and Sriurai, W. (2018). Applying association rule to risk analysis for dropout students of Information Technology Department (in Thai). **Journal of Science and Science Education**, 1(2), 123-133.

Pavya, K., and Srinivasan, D. B. (2017). Feature Selection Techniques in Data Mining: A Study. **International Journal of Scientific Development and Research (IJSDR),** 2(6), 594-598.

Pheunpha, P. (2020). Drop-out Factors of Students of Business Administration Faculty (in Thai). **Journal of Education, Mahasarakham University**. 14(2), 144-158.

Prakobpol, T. (2009). Artificial Neural Networks. **HCU Journal**, 12(24). 73-87.

Rawengwan, P. and Seresangtakul, P. (2017). A model for forecasting educational status of students (in Thai). **Proceedings of The National and International Graduate Research Conference 2017**. (pp 273 – 283). Khon Kaen University.

Sittichat, S. (2017). Study of Educational Attributes Using Data Mining Technique (in Thai). **Information Technology Journal**, 13(2), 20–28.

Sriurai, W. (2014). Patients Classification of Metabolic Syndrome Using Feature Selection and Artificial Neural Network (in Thai). **Srinakharinwirot Science Journal,** 30(1), 91-102.

Techapanurak, E. (2021). Bayesian Neural Network (in Thai). Retrieved September 25, 2021, from medium.com : https://medium.com/@dopplerz/bayesian-neural-network-ตอนที่-1-ทฤฎีความน่าจะเป็นแบบเบย์ๆ-99deeab8c206

Ubon Ratchathani University. (2018). **Rules and Regulations of Ubon Ratchathani University on Bachelor's Degree Education B.E. 2561 (2018)** (in Thai). Retrieved December 30, 2020, from https://www.ubu.ac.th /web/files_up/46f20 19042610283611.pdf.