

การศึกษาปัจจัยที่ส่งผลต่อการสำเร็จการศึกษาตามแผนของนักศึกษา  
ระดับปริญญาตรี โดยใช้เทคนิคการคัดเลือกคุณลักษณะบนชุดข้อมูลที่ไม่สมดุล

**The Study of Factors Affecting for On-time Graduation of  
Ungraduated Student Using Feature Selection Technique on  
Imbalanced Datasets**

วรการ ใจดี\* และ นพณัฐ วรณภีร์

*Worakarn Jaidee and Noppanat Wannapee*

สาขาวิชาระบบสารสนเทศทางธุรกิจ คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคล  
ล้านนา เชียงใหม่ ตำบลช้างเผือก อำเภอเมือง จังหวัดเชียงใหม่ 50300

*Department of Business Information System, Faculty of Business Administration and Liberal Art,*

*Rajamangala University of Technology Lanna, Changpuak, Muang, Chiang Mai 50300*

Received: May 13, 2020; Revised: June 01, 2020; Accepted: June 01, 2020; Published: June 23, 2020

**ABSTRACT** – The objectives of this research were (1) to study the factors affecting the students' graduation according to the plan, and (2) to study the guidelines and improvements in teaching and learning problems of each subject affecting the graduation according to the plan of the student. Main factors or data in this research, the researcher used the grade point in each subject, GPA, the gender of the first-year and second-year students by collecting data from Computer Information System, Faculty of Business Administration and Liberal Art, Rajamangala University of Technology Lanna, Chiang Mai with a total of 358 records. As a result, it revealed that datasets were imbalanced with a higher number of students who graduated as their plan than students who did not graduate as their plan completely. Therefore, the researcher solved the problem using the method of balancing the data set by Synthetic Minority Over-Sampling Technique: SMOTE before entering the process of analyzing factors by using feature selection techniques which can divide as 3 techniques as follows: (1) Chi-Square Feature Selection (2) Information Gain Feature Selection and (3) Correlation Based Feature Selection. From the result, it was found that the most important and topnotch factor in every feature selection was Computer Programming 1 Subject This subject is an elementary and compulsory subject in the Computer Information System and most students received quite low points in this subject. Which may cause this subject to be topnotch in every feature selection techniques and maybe the main factor affecting graduation's plan of students significantly.

**KEYWORDS:** Imbalanced Datasets, Feature Selection, SMOTE

---

\*Corresponding Author: worakarn.jaidee@gmail.com

**บทคัดย่อ** -- งานวิจัยนี้ได้จัดทำขึ้นโดยมีวัตถุประสงค์ 2 ประการคือ (1) เพื่อที่ศึกษาปัจจัยที่ส่งผลต่อการสำเร็จการศึกษาตามแผนของนักศึกษา และ (2) เพื่อศึกษาแนวทางและข้อปรับปรุงในการแก้ไขปัญหาการเรียนการสอนในแต่ละรายวิชาที่ส่งผลต่อการสำเร็จการศึกษาตามแผนของนักศึกษา โดยปัจจัยหรือข้อมูลหลักในการวิจัยนี้ ผู้วิจัยจะใช้ข้อมูลระดับคะแนนในแต่ละวิชา ข้อมูลเกรดเฉลี่ย ข้อมูลเพศของนักศึกษาชั้นปีที่ 1-2 โดยผู้วิจัยได้เก็บรวบรวมข้อมูลนักศึกษาจากสาขาวิชา ระบบสารสนเทศทางคอมพิวเตอร์ คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี เชียงใหม่ จำนวน 358 คน จากการรวบรวมข้อมูลเบื้องต้นพบว่า ชุดข้อมูลมีความไม่สมดุลกัน (Imbalanced Datasets) โดยมีจำนวนกลุ่มนักศึกษาที่สำเร็จการศึกษาตามแผนมากกว่ากลุ่มนักศึกษาที่ไม่สำเร็จการศึกษาตามแผน ดังนั้นผู้วิจัยจึงได้ทำการแก้ปัญหาโดยใช้วิธีการปรับสมดุลให้กับชุดข้อมูลโดยใช้เทคนิคการสุ่มเพิ่มตัวอย่างข้อมูลกลุ่มน้อย (Synthetic Minority Over-sampling Technique : SMOTE) ก่อนที่จะนำเข้าสู่กระบวนการวิเคราะห์ปัจจัยโดยจะใช้เทคนิคการคัดเลือกคุณลักษณะ (Feature Selection) 3 เทคนิคคือ (1) Chi-Square Feature Selection (2) Information Gain Feature Selection และ (3) Correlation Based Feature Selection จากการทดลองพบว่าปัจจัยที่มีความสำคัญมากที่สุดและเป็นอันดับหนึ่งในทุกเทคนิคการคัดเลือกคุณลักษณะคือ วิชาการเขียนโปรแกรมคอมพิวเตอร์ 1 ซึ่งวิชานี้เป็นวิชาพื้นฐานและเป็นวิชาบังคับของสาขาระบบสารสนเทศทางคอมพิวเตอร์และเป็นวิชาที่นักศึกษาด้านใหญ่ได้ระดับคะแนนในวิชานี้ค่อนข้างน้อย ซึ่งอาจส่งผลให้รายวิชานี้ได้ลำดับความสำคัญเป็นอันดับหนึ่งในทุกเทคนิคการคัดเลือกคุณลักษณะและอาจเป็นปัจจัยหลักที่ส่งผลต่อการสำเร็จการศึกษาตามแผนของนักศึกษาอย่างมีนัยสำคัญ

**คำสำคัญ:** ชุดข้อมูลที่ไม่สมดุล, การคัดเลือกคุณลักษณะ, การสุ่มเพิ่มตัวอย่างข้อมูลกลุ่มน้อย

## 1. บทนำ

โดยทั่วไปแล้วปัญหาการไม่สำเร็จการศึกษาตามแผนของนักศึกษา เป็นปัญหาที่พบบ่อยในกลุ่มนักศึกษาระดับปริญญาตรี โดยเฉพาะสาขาวิชาทางคอมพิวเตอร์ที่มีอัตราการลาออกค่อนข้างสูง ซึ่งคิดเป็นร้อยละ 55.21% [1] โดยสาเหตุมาจากหลากหลายปัจจัย เช่น ปัญหาด้านครอบครัว ปัญหาด้านการเรียน ปัญหาด้านสุขภาพ เป็นต้น โดยปัญหาที่พบบ่อยมากที่สุดในปัจจุบันก็คือปัญหาด้านการเรียนซึ่งนักศึกษามีระดับคะแนนในแต่ละวิชาน้อยกว่าเกณฑ์ที่กำหนดโดยเฉพาะระดับคะแนนในวิชาพื้นฐานและวิชาบังคับ ซึ่งปัญหานี้อาจทำให้นักศึกษารู้สึกท้อและเบื่อหน่ายกับการเรียนและทำให้นักศึกษาหาไม่ทราบว่าจะดำเนินการอย่างไรหรือดำเนินการวางแผนการเรียนไปในทิศทางที่ดีขึ้นได้อย่างไร จนส่งผลให้นักศึกษาไม่สามารถที่จะสำเร็จการศึกษาตามแผน ทุกรายหรือลาออกกลางคันได้ ในงานวิจัยนี้ผู้วิจัยมีจุดมุ่งหมายและวัตถุประสงค์ 2 ประการคือ (1) เพื่อศึกษาปัจจัยที่ส่งผลต่อการสำเร็จการศึกษาตามแผนของนักศึกษา และ (2) เพื่อปรับปรุงรูปแบบการเรียนการสอนในแต่ละรายวิชาที่ส่งผลต่อการไม่สำเร็จการศึกษาของนักศึกษา โดย

ผู้วิจัยได้เก็บรวบรวมข้อมูลระดับคะแนนของนักศึกษาชั้นปีที่ 1-2 จากสาขาวิชา ระบบสารสนเทศทางคอมพิวเตอร์ คณะบริหารธุรกิจและศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี เชียงใหม่ จำนวน 358 คน [2] จากการรวบรวมข้อมูลเบื้องต้นพบว่า ชุดข้อมูลมีความไม่สมดุลกัน (Imbalanced Datasets) โดยมีจำนวนกลุ่มนักศึกษาที่สำเร็จการศึกษาตามแผนมากกว่ากลุ่มนักศึกษาที่ไม่สำเร็จการศึกษาตามแผน

ปัญหาความไม่สมดุลของคลาส เป็นปัญหาที่พบบ่อยในชุดข้อมูลทั่วไปและสามารถเกิดขึ้นได้ในหลากหลายโดเมนข้อมูล [3] เช่น ข้อมูลทางการแพทย์ ข้อมูลด้านการศึกษา ข้อมูลด้านการจัดการความเสี่ยง ข้อมูลด้านการรู้จำใบหน้า เป็นต้น ปัญหาความไม่สมดุลของคลาสโดยทั่วไปแล้วอาจเกิดจากสาเหตุที่ข้อมูลมีการทับซ้อนกันหรือมีความซับซ้อนของข้อมูลสูง (Class Overlapping) ซึ่งทำให้ยากต่อการจำแนก ปัญหาข้อมูลในการเรียนรู้ไม่เพียงพอ (Small Sample Size) และปัญหาที่พบอย่างแพร่หลายในงานวิจัยคือปัญหาการกระจายไม่สม่ำเสมอเมื่อเทียบกับคลาสหลัก (Class Distribution) ตัวอย่างเช่น [4] ชุดข้อมูลการสำเร็จการศึกษานี้มีจำนวนข้อมูลนักศึกษา

ทั้งหมด 100 คน โดยมีข้อมูลของคลาสที่สำเร็จการศึกษาตามแผนจำนวน 85 คน และมีข้อมูลคลาสที่ไม่สำเร็จการศึกษาตามแผนจำนวน 15 คน โดยจะเรียกข้อมูลคลาสที่สำเร็จการศึกษาตามแผนที่ซึ่งมีขนาดใหญ่กว่าว่าคลาสส่วนมาก (Majority class) และเรียกคลาสที่ไม่สำเร็จการศึกษาตามแผนที่ซึ่งมีขนาดเล็กกว่าว่าคลาสส่วนน้อย (Minority class) ซึ่งปัญหาความไม่สมดุลของคลาสเหล่านี้ อาจทำให้การนำข้อมูลมาใช้ในการทำเหมืองข้อมูลด้วยอัลกอริทึมพื้นฐานต่างๆ เช่น การคัดเลือกคุณลักษณะ (Feature Selection) การจำแนกประเภทข้อมูล (Data Classification) ไม่มีประสิทธิภาพหรือมีประสิทธิผลต่ำ

การคัดเลือกคุณลักษณะ (Feature Selection) เป็นวิธีการในการคัดเลือกคุณลักษณะที่มีความสำคัญและใช้เพื่อลดขนาดหรือมิติของข้อมูลลง การคัดเลือกคุณลักษณะจึงเป็นหัวใจสำคัญในการทำเหมืองข้อมูล (Data Mining) หรือการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งในปัจจุบันมีเทคนิคหรือวิธีการที่หลากหลายที่นักวิจัยได้นำมาใช้ในงานวิจัย ตัวอย่างเช่น [5] เทคนิค Correlation Based Feature Selection เป็นเทคนิคที่ใช้การหาคลุ่มของคุณลักษณะที่ได้รับการประเมินค่าจากความสามารถในการคาดการณ์ โดยคุณลักษณะที่ถูกคัดเลือกจะถูกนำไปใช้ในการในขั้นตอนต่อไป เช่น การจำแนกประเภทข้อมูล (Data Classification) เป็นต้น เทคนิคนี้ยังสามารถจัดการกับคุณลักษณะที่ไม่เกี่ยวข้อง โดยจะจัดอันดับกลุ่มย่อยของมิติข้อมูลและทำการคัดเลือกกลุ่มย่อยของมิติข้อมูลที่มีความสัมพันธ์กันสูงกับคลาส และไม่มีความสัมพันธ์กับคลาสอื่นๆ เทคนิค chi-square เป็นเทคนิคการวัดโดยการใช้สถิติประมาณความสัมพันธ์ร่วมระหว่างคุณลักษณะเฉพาะกับคลาสของคุณลักษณะเฉพาะ เทคนิค Information Gain เป็นเทคนิคการพิจารณาจากค่าความน่าจะเป็นของแต่ละคุณลักษณะที่เป็นไปได้แล้ววัดค่าความไร้ระเบียบ (Entropy) เพื่อคัดเลือกคุณลักษณะที่มีความสำคัญในการจำแนกกลุ่มได้ดีที่สุด โดยทั่วไปแล้วเทคนิคการคัดเลือกคุณลักษณะด้วย Information Gain มักจะนำมาใช้ร่วมกับอัลกอริทึมหรือตัวแบบต้นไม้ตัดสินใจ (Decision Tree)

ดังนั้นในงานวิจัยนี้ผู้วิจัยจึงสังเกตเห็นว่าการใช้เทคนิคในการจัดการชุดข้อมูลที่ไม่สมดุล โดยใช้เทคนิคการสุ่มเพิ่มตัวอย่างข้อมูลกลุ่มน้อย (Synthetic Minority Over-sampling Technique : SMOTE) [6] ร่วมกับวิธีการคัดเลือกคุณลักษณะด้วยเทคนิค Chi-Square, Information Gain และเทคนิค Correlation Based

Feature Selection จะทำให้ได้ปัจจัยที่ส่งผลต่อการสำเร็จการศึกษาของนักศึกษาได้อย่างถูกต้องและมีประสิทธิภาพ

## 2. วิธีการวิจัย

ในการศึกษาวิจัยครั้งนี้ ผู้วิจัยได้ทำการศึกษาปัจจัยที่ส่งผลต่อการสำเร็จการศึกษาของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคการคัดเลือกคุณลักษณะ บนชุดข้อมูลที่ไม่สมดุล โดยผู้วิจัยได้มีการดำเนินการ 5 ขั้นตอนดังต่อไปนี้

### 2.1 ขั้นตอนการรวบรวมข้อมูล (Data Gathering)

งานวิจัยนี้ผู้วิจัยได้รวบรวมข้อมูลประวัติผลการเรียนของนักศึกษาสาขาวิชาระบบสารสนเทศทางคอมพิวเตอร์ มหาวิทยาลัยราชภัฏวชิรเวศน์ เชียงใหม่ ตั้งแต่ปี พ.ศ. 2557 – 2559 จำนวน 358 คน โดยใช้ข้อมูลผลการเรียนและระดับคะแนนในแต่ละวิชาใน 4 เทอมแรก (ชั้นปีที่ 1 – 2) สาขาที่ใช้ข้อมูลผลการเรียนและระดับคะแนนในแต่ละวิชาใน 4 เทอมแรก เนื่องจากในช่วงก่อนขึ้นชั้นปีที่ 3 นักศึกษาส่วนใหญ่จะมีอัตราการย้ายสาขา การลาออกหรือพ้นสภาพการเป็นนักศึกษาก่อนข้างมาก จึงทำให้ข้อมูลของนักศึกษบางรายไม่ครบหรือขาดหายไป ซึ่งอาจส่งผลต่อประสิทธิภาพในขั้นตอนของการวิจัยได้ โดยข้อมูลจะประกอบไปด้วย ระดับคะแนนในแต่ละวิชาเกรดเฉลี่ย เพศ โดยมีจำนวนคุณลักษณะทั้งหมด 27 คุณลักษณะ ดังตารางที่ 1

ตารางที่ 1. แสดงรายละเอียดคุณลักษณะข้อมูล

ลำดับ	ชื่อคุณลักษณะ	คำอธิบาย
1	Gender	เพศนักศึกษา
2	เศรษฐศาสตร์จุลภาค	ระดับคะแนน
3	การวิเคราะห์ธุรกิจเชิงปริมาณ	ระดับคะแนน
4	การเงินธุรกิจ	ระดับคะแนน
5	การบัญชีขั้นต้น	ระดับคะแนน
6	องค์การและการจัดการ	ระดับคะแนน
7	การจัดการการผลิตและการปฏิบัติการ	ระดับคะแนน
8	หลักการตลาด	ระดับคะแนน
9	การใช้งานระบบสารสนเทศในธุรกิจ	ระดับคะแนน
10	การเขียนโปรแกรมคอมพิวเตอร์ 1	ระดับคะแนน

ตารางที่ 1. แสดงรายละเอียดคุณลักษณะข้อมูล (ต่อ)

ลำดับ	ชื่อคุณลักษณะ	คำอธิบาย
11	โครงสร้างข้อมูลและอัลกอริทึม	ระดับคะแนน
12	ระบบจัดการฐานข้อมูล	ระดับคะแนน
13	การเขียนโปรแกรมบนเว็บ	ระดับคะแนน
14	การวิเคราะห์และออกแบบระบบ	ระดับคะแนน
15	ระบบสารสนเทศในองค์กร	ระดับคะแนน
16	พลศึกษา	ระดับคะแนน
17	ภาษาอังกฤษเพื่ออาชีพ	ระดับคะแนน
18	ภาษาอังกฤษเพื่อการสื่อสาร	ระดับคะแนน
19	ภาษาอังกฤษผ่านสื่อและ เทคโนโลยี	ระดับคะแนน
20	ภาษาอังกฤษในชีวิตประจำวัน	ระดับคะแนน
21	ภาษาไทยเพื่อการสื่อสาร	ระดับคะแนน
22	การพัฒนาคุณภาพชีวิตและสังคม	ระดับคะแนน
23	ปรัชญาเศรษฐกิจพอเพียงเพื่อการ พัฒนาที่ยั่งยืน	ระดับคะแนน
24	สถิติพื้นฐาน	ระดับคะแนน
25	การคิดและการตัดสินใจเชิง วิทยาศาสตร์	ระดับคะแนน
26	GPA	เกรดเฉลี่ย
27	Status	0 = สำเร็จ การศึกษา 1 = ไม่สำเร็จ การศึกษา

## 2.2 ขั้นตอนการเตรียมข้อมูล (Data Pre-processing)

จากการรวบรวมข้อมูลเบื้องต้น ผู้วิจัยพบว่าข้อมูลมีความไม่สมบูรณ์และมีค่าที่ขาดหายไป (Missing Value) จึงทำให้ไม่สามารถนำข้อมูลเข้าสู่กระบวนการคัดเลือกคุณลักษณะได้ ผู้วิจัยจึงได้ทำการแก้ปัญหาข้อมูลที่ขาดหายไปโดยการเติมค่าที่ขาดหายไปด้วยวิธีการหาเพื่อนบ้านใกล้เคียงที่สุด (k-Nearest Neighbor Imputation) [7] เทคนิคการเติมข้อมูลที่ขาดหายไปด้วยวิธีเพื่อนบ้านใกล้เคียงเป็นเทคนิคที่มีประสิทธิภาพสูงและเป็นการประยุกต์อัลกอริทึมเพื่อนบ้านใกล้เคียงที่สุดมาใช้ในการ

เติมข้อมูลที่ขาดหายไป โดยข้อมูลจะถูกเติมจากข้อมูลใกล้เคียงของแต่ละเรคคอร์ดนั้นๆ และยังสามารถแทนที่ข้อมูลที่หายไปด้วยค่าที่เป็นไปได้ที่ใกล้เคียงที่สุดกับค่าจริง โดยวิธีการเติมข้อมูลที่ขาดหายไปด้วยวิธีนี้จะมีหลักการทำงานเหมือนกับอัลกอริทึมเพื่อนบ้านใกล้เคียงที่สุดคือ จะใช้วิธีการวัดระยะห่างแบบ Euclidean Distance ซึ่งเกิดจากรากที่สองของผลต่างระหว่างแอดทริบิวต์ต่างๆ ยกกำลังสอง ดังสมการที่ 1

$$\begin{aligned}
 D_{Euclidean}(p, q) &= D_{Euclidean}(q, p) \\
 &= \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \\
 &= \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)
 \end{aligned}$$

โดยที่ p คือข้อมูลที่ต้องการจำแนก  
q คือข้อมูลในแต่ละกลุ่ม

## 2.3 ขั้นตอนการเพิ่มความสมดุลของข้อมูลด้วยวิธีการสุ่มเพิ่มตัวอย่างข้อมูลกลุ่มน้อย (SMOTE)

หลักจากขั้นตอนการเตรียมข้อมูลเรียบร้อยแล้ว ผู้วิจัยพบว่าข้อมูลมีความไม่สมดุลกัน โดยมีจำนวนกลุ่มนักศึกษาที่สำเร็จการศึกษาตามแผนมากกว่ากลุ่มนักศึกษาที่ไม่สำเร็จการศึกษาตามแผน ดังนั้นผู้วิจัยจึงแก้ปัญหาโดยการนำวิธีการสุ่มเพิ่มตัวอย่างข้อมูลกลุ่มน้อย เพื่อช่วยในการปรับสมดุลของข้อมูลเพื่อให้จำนวนของคลาสรอง (Minority Class) มีจำนวนเทียบเท่าหรือใกล้เคียงกับจำนวนข้อมูลที่เป็นคลาสหลัก (Majority Class) ก่อนนำข้อมูลเข้าสู่กระบวนการคัดเลือกคุณลักษณะต่อไป

เทคนิคการสุ่มเพิ่มตัวอย่างข้อมูลกลุ่มน้อย (SMOTE: Synthetic Minority Over-Sampling Technique) มีหลักการการทำงานเหมือนกับอัลกอริทึมเพื่อนบ้านใกล้เคียงที่สุด โดยจะทำการสุ่มเลือกแล้วเก็บในตัวแปร  $x_i$  โดยทั่วไปแล้วตัวแปร  $x_i$  และตัวแปร  $x'$  จะถูกเรียกว่า (Seed Sample) จากนั้นจะทำการสุ่มค่าตัวแปร  $\sigma$  ให้มีค่า 0 หรือ 1 เพื่อใช้ในการคำนวณ โดยมีสมการในการสุ่มเพิ่มตัวอย่างข้อมูลกลุ่มน้อยดังสมการที่ 2

$$x_{new} = x_i + (x' - x_i) * \sigma \quad (2)$$

ตารางที่ 2. ตารางเปรียบเทียบข้อมูลที่ผ่านมากระบวนการสุ่มเพิ่ม  
ตัวอย่างข้อมูลกลุ่มน้อย

กลุ่ม	original	smote
กลุ่มที่สำเร็จการศึกษา	203	203
กลุ่มที่ไม่สำเร็จการศึกษา	155	203
จำนวนทั้งหมด	358	406

## 2.4 ขั้นตอนการคัดเลือกคุณลักษณะข้อมูล (Feature Selection)

ในขั้นตอนการคัดเลือกคุณลักษณะเป็นขั้นตอนในการสำคัญในงานวิจัยนี้เพื่อที่จะศึกษาว่าปัจจัยใดส่งผลกระทบต่อความสำเร็จการศึกษาตามแผนของนักศึกษามากที่สุด โดยในขั้นตอนนี้ผู้วิจัยจะใช้เทคนิค (1) Chi-Square (2) Information Gain และ (3) Correlation Based Feature Selection

(1) Chi-Square ( $\chi^2$ ) เป็นเทคนิคการประเมินค่าของคุณลักษณะโดยใช้การคำนวณค่า chi-square ทางสถิติ เพื่อศึกษาว่าการแจกแจงความถี่ของตัวแปรคุณลักษณะเป็นไปตามรูปแบบที่กำหนดไว้หรือไม่ ดังแสดงในสมการที่ 3

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (3)$$

โดยที่  $O_1, O_2 \dots O_n$  เป็นความถี่ของตัวแปรที่ได้จากการศึกษา

$E_1, E_2 \dots E_n$  เป็นความถี่ที่คาดหวัง (หรือความถี่ที่ควรจะเป็น)

(2) Information Gain เป็นเทคนิคที่ใช้ในการชี้วัดเพื่อเลือกคุณลักษณะข้อมูลที่ต้องการที่น้อยที่สุดที่เหมาะสมในการนำไปใช้ระบุ โดยการคำนวณค่า Gain สำหรับแต่ละมิติข้อมูลซึ่งมิติข้อมูลใดมีค่า Gain สูงสุด จะถูกคัดเลือกเพื่อนำไปใช้ระบุ สมการที่ 4 แสดงการคำนวณค่า Information Gain หรือค่า Entropy ของชุดข้อมูลทั้งหมด สมการที่ 5 แสดงการคำนวณค่า Entropy ของชุดมิติข้อมูลในแต่ละคุณลักษณะ สมการที่ 6 เป็นสมการในการคำนวณค่า Information Gain สำหรับพิจารณามิติข้อมูลคุณลักษณะ A

$$E(D) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (4)$$

$$E_A(D) = \sum_{j=1}^m \frac{|D_j|}{D} x E(D_j) \quad (5)$$

$$Gain(A) = E(D) - E_A(D) \quad (6)$$

โดยที่  $p_i$  คือค่าความน่าจะเป็นที่เรคคอร์ดหนึ่งๆ จะมีหมวดหมู่ของข้อมูล หรือกล่าวว่าการคำนวณค่า Information Gain หรือการวัดค่า Entropy ก่อนที่จะมีการแบ่งข้อมูลออกตามมิติข้อมูลและหลักการแบ่งว่ามีประสิทธิภาพดีขึ้นหรือไม่ ถ้ามีประสิทธิภาพดีขึ้นค่า information gain ก็จะมีค่าสูง

(3) (Correlation Based Feature Selection : CFS) เป็นเทคนิคในการเลือกคุณสมบัติของคุณลักษณะโดยใช้การพิจารณาบนพื้นฐานความสัมพันธ์ของกลุ่มคุณลักษณะที่ได้จากการประเมินค่าความสามารถในการคาดการณ์ และยังสามารถจัดการกับคุณลักษณะที่ไม่เกี่ยวข้อง โดย CFS จะจัดอันดับกลุ่มย่อยของมิติข้อมูลและทำการคัดเลือกกลุ่มย่อยของมิติข้อมูลที่มีความสัมพันธ์กันสูงกับคลาสและไม่มีความสัมพันธ์กับคลาสอื่นๆ สมการประเมินกลุ่มย่อยของมิติข้อมูลแบบ CFS แสดงในสมการที่ 7

$$M_s = \frac{k r_{ef}}{\sqrt{(k+k(k-1) r_{ef})}} \quad (7)$$

โดยที่  $M_s$  คือค่าที่ค้นหาได้ของมิติข้อมูลกลุ่มย่อย S ซึ่งประกอบด้วยมิติข้อมูล K

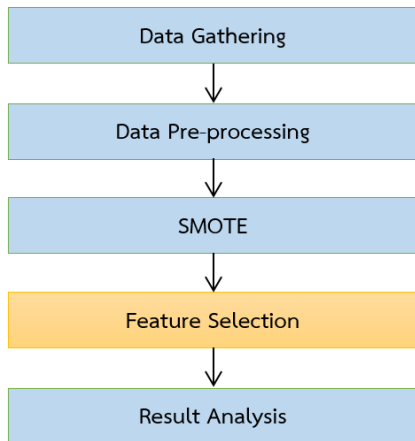
— คือค่าเฉลี่ยความสัมพันธ์ของตัวแปรกับคลาส

$(f \in S)$

— คือค่าเฉลี่ยความสัมพันธ์ระหว่างมิติข้อมูล

## 2.5 ขั้นตอนการวิเคราะห์และสรุปผล (Result Analysis)

ในขั้นตอนของการวิเคราะห์และสรุปผล ผู้วิจัยจะทำการสรุปผลการใช้เทคนิคการคัดเลือกคุณลักษณะทั้ง 3 เทคนิค (1) Chi-Square (2) Information Gain และ (3) Correlation Based Feature Selection โดยทำการหาข้อสรุปว่าปัจจัยใดที่ส่งผลกระทบต่อความสำเร็จการศึกษาตามแผนของนักศึกษามากที่สุด



รูปที่ 1. แสดงขั้นตอนการทำงานของงานวิจัย

## 2.6 เครื่องมือที่ใช้ในการวิจัย

ในงานวิจัยนี้ผู้วิจัยได้ใช้ซอฟต์แวร์ Jupyter Notebook โดยเขียนด้วยภาษาไพธอน (Python) ในขั้นตอนของกระบวนการทดลอง ตั้งแต่ขั้นตอนของการเตรียมข้อมูล (Data Pre-processing) จนถึงขั้นตอนของการคัดเลือกคุณลักษณะ (Feature Selection)

## 3. ผลการทดลอง

ผลการวิจัยการศึกษาปัจจัยที่ส่งผลต่อการสำเร็จการศึกษาตามแผนของนักศึกษาโดยใช้เทคนิคการคัดเลือกคุณลักษณะทั้งหมด 3 เทคนิค คือ (1) Chi-Square (2) Information Gain และ (3) Correlation Based Feature Selection โดยผลการวิจัยของแต่ละเทคนิคหลังจากผ่านกระบวนการในการจัดการกับข้อมูลที่ขาดหายไป (Missing Value) และขั้นตอนการเพิ่มความสมดุลของข้อมูลด้วยวิธีการสุ่มเพิ่มตัวอย่างข้อมูลกลุ่มน้อย (SMOTE) เรียบร้อยแล้ว สามารถสรุปผลการทดลองได้ดังนี้

### 3.1 การใช้เทคนิค Chi-Square ในการคัดเลือกคุณลักษณะ

จากผลการทดลองโดยใช้เทคนิค Chi-Square ในการคัดเลือกคุณลักษณะ โดยผู้วิจัยได้จัดอันดับคุณลักษณะที่มีค่าความสำคัญที่คำนวณได้ 10 อันดับแรก ดังตารางที่ 3

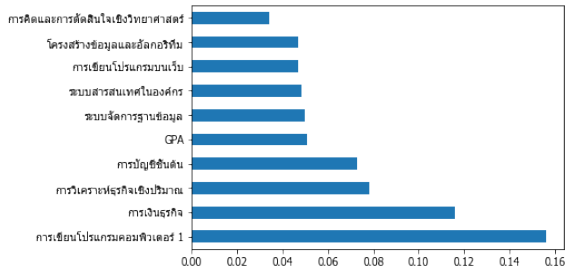
ตารางที่ 3. คุณลักษณะ 10 อันดับแรกที่มีความสำคัญมากที่สุด โดยการใช้เทคนิค Chi-Square

อันดับ	วิชา	ค่าความสำคัญที่คำนวณได้
1	การเขียนโปรแกรมคอมพิวเตอร์ 1	3506.270
2	การเงินธุรกิจ	1892.362
3	การวิเคราะห์ธุรกิจเชิงปริมาณ	1805.420
4	การบัญชีขั้นต้น	1802.137
5	ภาษาอังกฤษผ่านสื่อและเทคโนโลยี	1045.797
6	สถิติพื้นฐาน	963.232
7	ระบบสารสนเทศในองค์กร	799.03
8	ระบบจัดการฐานข้อมูล	708.430
9	ภาษาอังกฤษเพื่อการสื่อสาร	657.455
10	การเขียนโปรแกรมบนเว็บ	550.181

จากตารางที่ 3 คุณลักษณะ 10 อันดับแรกที่มีความสำคัญมากที่สุด โดยการใช้เทคนิค Chi-Square พบว่า วิชาที่เป็นปัจจัยหลักที่ส่งผลต่อการสำเร็จการศึกษาตามแผนของนักศึกษาคือ วิชา การเขียน โปรแกรมคอมพิวเตอร์ 1 ซึ่งเป็นวิชาหลักของสาขาวิชาระบบสารสนเทศทางคอมพิวเตอร์มีค่าความสำคัญที่ 3506.270 อันดับที่สองถึงอันดับที่สี่คือ วิชาการเงินธุรกิจมีค่าความสำคัญที่ 1892.362 วิชา การวิเคราะห์ธุรกิจเชิงปริมาณมีค่าความสำคัญที่ 1805.420 และ วิชา การบัญชีขั้นต้น มีค่าความสำคัญที่ 1802.137 ซึ่งทั้งสามวิชานี้เป็นวิชาหลักของทางคณะบริหารธุรกิจและศิลปศาสตร์ อันดับห้าวิชาภาษาอังกฤษผ่านสื่อและเทคโนโลยีมีค่าความสำคัญที่ 1045.797 อันดับหก วิชาสถิติพื้นฐานมีค่าความสำคัญที่ 963.232 อันดับเจ็ดวิชา ระบบสารสนเทศในองค์กรมีค่าความสำคัญที่ 799.03 อันดับแปดวิชา ระบบจัดการฐานข้อมูลมีค่าความสำคัญที่ 708.430 อันดับเก้าภาษาอังกฤษเพื่อการสื่อสาร มีค่าความสำคัญที่ 657.455 และ อันดับสิบ วิชา การเขียน โปรแกรมบนเว็บ มีค่าความสำคัญที่ 550.181

### 3.2 การใช้เทคนิค Information Gain ในการคัดเลือกคุณลักษณะ

จากผลการทดลองโดยใช้เทคนิค Information Gain ในการคัดเลือกคุณลักษณะ โดยผู้วิจัยได้จัดอันดับคุณลักษณะที่มีความสำคัญ 10 อันดับแรก ดังรูปที่ 2



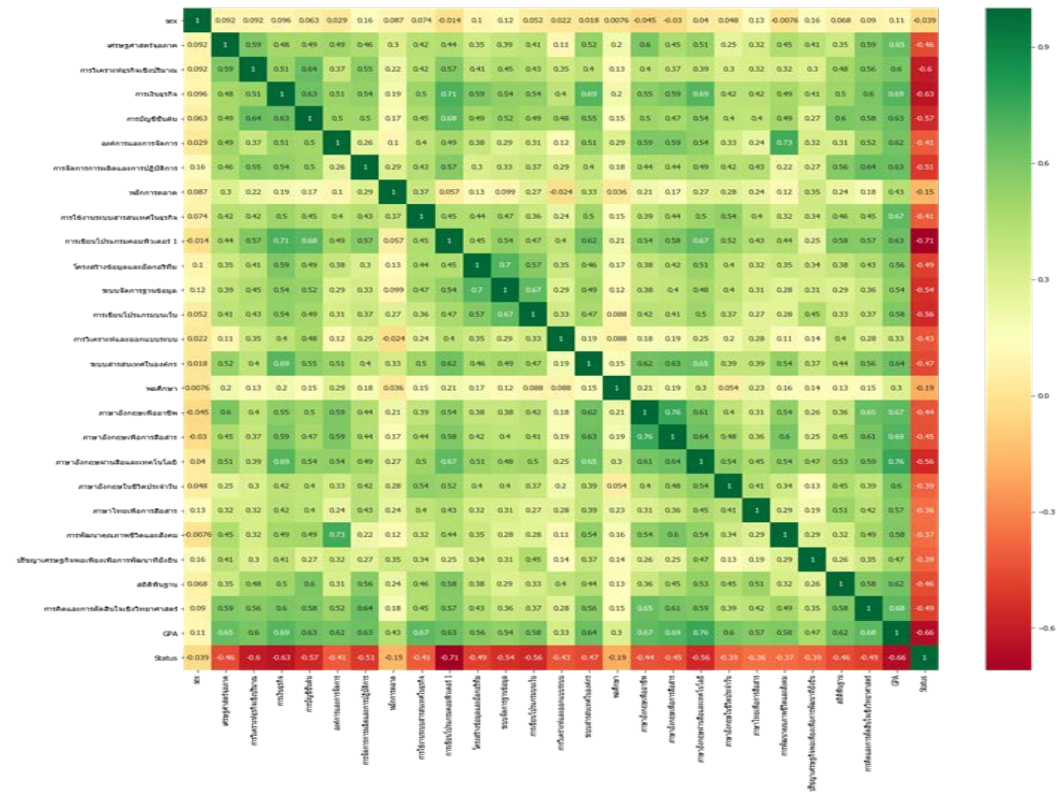
รูปที่ 2. แสดงคุณลักษณะ 10 อันดับแรกที่มีความสำคัญมากที่สุด โดยการใช้เทคนิค Information Gain

จากภาพที่ 2 คุณลักษณะ 10 อันดับแรกที่มีความสำคัญมากที่สุด โดยการใช้เทคนิค Information Gain พบว่าวิชาที่เป็นปัจจัยหลักที่ส่งผลต่อการสำเร็จการศึกษาตามแผนของนักศึกษา คือวิชา การเขียนโปรแกรมคอมพิวเตอร์ 1 ซึ่งเป็นวิชาหลักของ

สาขาวิชาระบบสารสนเทศทางคอมพิวเตอร์ และอันดับสองถึงอันดับสี่ คือวิชาวิชาการเงินธุรกิจ วิชาการวิเคราะห์ธุรกิจเชิงปริมาณ และการบัญชีขั้นต้น ซึ่งทั้งสามวิชานี้เป็นวิชาหลักของทางคณะบริหารธุรกิจและศิลปศาสตร์ และในอันดับที่ห้า คือระดับคะแนนเฉลี่ย (GPA) และในอันดับที่หกถึงอันดับที่เก้า ได้แก่วิชาระบบจัดการฐานข้อมูล วิชาระบบสารสนเทศในองค์กร วิชาการเขียนโปรแกรมบนเว็บ วิชาโครงสร้างข้อมูลและอัลกอริทึม ซึ่งทั้งสี่วิชานี้เป็นวิชาหลักและวิชาเลือกของหลักสูตรสาขาวิชาระบบสารสนเทศทางคอมพิวเตอร์ ส่วนในอันดับที่สิบคือ วิชาการคิดและการตัดสินใจเชิงวิทยาศาสตร์ที่มีผลต่อการสำเร็จการศึกษาของนักศึกษาน้อยที่สุดใน 10 อันดับแรก

### 3.3 การใช้เทคนิค Correlation Based Feature Selection ในการคัดเลือกคุณลักษณะ

ในการคัดเลือกคุณลักษณะด้วยเทคนิค Correlation Based Feature Selection ผู้วิจัยได้ใช้วิธีการแสดงผลข้อมูลของผลการทดลองในรูปแบบของ Correlation Matrix ร่วมกับ Heatmap ซึ่งได้ผลการทดลองดังรูปที่ 3



รูปที่ 3. ผลการทดลองโดยการใช้เทคนิค Correlation Matrix ร่วมกับ Heatmap

จากรูปที่ 3 การคัดเลือกคุณลักษณะโดยการวิเคราะห์สหสัมพันธ์ (Correlation Based Feature Selection) จะเห็นว่าวิชาที่มีความสัมพันธ์กับสถานการณ์สำเร็จการศึกษาตามแผนของนักศึกษา (Status) มากที่สุดคือวิชา การเขียน โปรแกรมคอมพิวเตอร์ 1 มีค่าสหสัมพันธ์ -0.71 ส่วนอันดับที่สองคือระดับคะแนนเฉลี่ย (GPA) ของนักศึกษาในสองปีแรกมีค่าสหสัมพันธ์ -0.66 และในอันดับที่สามถึงอันดับที่ห้าคือวิชาวิชาการเงินธุรกิจที่มีค่าสหสัมพันธ์ -0.63 วิชาการวิเคราะห์ธุรกิจเชิงปริมาณมีค่าสหสัมพันธ์ -0.6 และวิชาการบัญชีขั้นต้นมีค่าสหสัมพันธ์ -0.57 ตามลำดับ ซึ่งทั้งสามวิชานี้เป็น วิชาหลักของทางคณะบริหารธุรกิจและศิลปศาสตร์

อันดับที่หกและเจ็ดมีค่าสหสัมพันธ์เท่ากับ -0.56 ได้แก่ วิชาการเขียน โปรแกรมบนเว็บและวิชาภาษาอังกฤษผ่านสื่อเทคโนโลยี อันดับแปดวิชาระบบจัดการฐานข้อมูลมีค่าสหสัมพันธ์ -0.54 อันดับเก้าและอันดับที่สิบมีค่าสหสัมพันธ์เท่ากับ -0.49 ได้แก่วิชาโครงสร้างข้อมูลและอัลกอริทึมและ วิชาการคิดและการตัดสินใจเชิงวิทยาศาสตร์

ผลการวิจัยพบว่าปัจจัยที่มีความสำคัญต่อการส่งผลการต่อ การสำเร็จการศึกษาตามแผนของนักศึกษามากที่สุดคือ ระดับคะแนนในรายวิชาการเขียน โปรแกรมคอมพิวเตอร์ 1 ซึ่งจากผลการทดลองด้วยการใช้เทคนิคการคัดเลือกคุณลักษณะทั้ง 3 เทคนิคพบว่ารายวิชาการเขียน โปรแกรมคอมพิวเตอร์ 1 มีความสำคัญมาเป็นอันดับหนึ่งในทุกเทคนิคของการคัดเลือกคุณลักษณะ ซึ่งอาจส่งผลให้รายวิชานี้ได้ลำดับความสำคัญมาเป็นอันดับหนึ่งในทุกเทคนิคการคัดเลือกคุณลักษณะและอาจเป็นปัจจัยที่ส่งผลการสำเร็จการศึกษาตามแผนอย่างมีนัยสำคัญ ด้วยเหตุนี้ผู้วิจัยได้สรุปสถิติ รวมถึงความถี่ของแต่ละปัจจัยใน 10 อันดับแรกจากการคัดเลือกคุณลักษณะทั้งสามเทคนิค ดังตารางที่ 4

ตารางที่ 4. แสดงสรุปสถิติ ความถี่ของแต่ละปัจจัยใน 10 อันดับแรกจากการคัดเลือกคุณลักษณะทั้งสามเทคนิค

ลำดับ	วิชา	ความถี่
1	การเขียน โปรแกรมคอมพิวเตอร์ 1	3
2	การเงินธุรกิจ	3
3	การวิเคราะห์ธุรกิจเชิงปริมาณ	3
4	การบัญชีขั้นต้น	3
5	ภาษาอังกฤษผ่านสื่อและเทคโนโลยี	2

ตารางที่ 4. แสดงสรุปสถิติ ความถี่ของแต่ละปัจจัยใน 10 อันดับแรกจากการคัดเลือกคุณลักษณะทั้งสามเทคนิค (ต่อ)

ลำดับ	วิชา	ความถี่
6	สถิติพื้นฐาน	1
7	ระบบสารสนเทศในองค์กร	2
8	ระบบจัดการฐานข้อมูล	3
9	ภาษาอังกฤษเพื่อการสื่อสาร	1
10	การเขียน โปรแกรมบนเว็บ	3
11	การคิดและการตัดสินใจเชิงวิทยาศาสตร์	2
12	GPA	2
13	โครงสร้างข้อมูลและอัลกอริทึม	2

จากตารางที่ 4 สามารถสรุปได้ว่าปัจจัยที่ติดอันดับหนึ่งถึงสิบครบทุกเทคนิคการคัดเลือกคุณลักษณะได้แก่ วิชาการเขียนโปรแกรมคอมพิวเตอร์ 1 วิชาการเงินธุรกิจ วิชาการวิเคราะห์ธุรกิจเชิงปริมาณ วิชาการบัญชีขั้นต้น วิชาระบบจัดการฐานข้อมูล และวิชาการเขียน โปรแกรมบนเว็บ ซึ่งวิชาที่ติดอันดับหนึ่งถึงสิบครบทุกเทคนิคการคัดเลือกคุณลักษณะล้วนแต่เป็นวิชาพื้นฐานและเป็นวิชาบังคับของคณะบริหารธุรกิจและศิลปศาสตร์และสาขาระบบสารสนเทศทางคอมพิวเตอร์ และวิชาที่ติดอันดับทั้งหมดนี้ นักศึกษาส่วนใหญ่จะได้ระดับคะแนนในวิชานี้ค่อนข้างน้อย ซึ่งอาจส่งผลให้แต่ละรายวิชาได้ลำดับความสำคัญในอันดับหนึ่งถึงสิบครบทุกเทคนิคอย่างมีนัยสำคัญ ส่วนปัจจัยที่ติดอันดับหนึ่งถึงสิบเพียงสองเทคนิคการคัดเลือกคุณลักษณะได้แก่ วิชาภาษาอังกฤษผ่านสื่อและเทคโนโลยี วิชาระบบสารสนเทศในองค์กร วิชาการคิดและการตัดสินใจเชิงวิทยาศาสตร์ วิชาโครงสร้างข้อมูลและอัลกอริทึม และระดับคะแนนเฉลี่ย (GPA) และปัจจัยที่ติดอันดับหนึ่งถึงสิบเพียงหนึ่งเทคนิคการคัดเลือกคุณลักษณะได้แก่ วิชาภาษาอังกฤษเพื่อการสื่อสารและวิชาสถิติพื้นฐาน

จะเห็นได้ว่าทั้งสามเทคนิคที่ใช้ให้ผลลัพธ์ที่ใกล้เคียงกันและมีจำนวนวิชาที่ติดมาเป็นอันดับหนึ่งถึงสิบมากถึงหกวิชาที่เหมือนกัน ซึ่งรายวิชาเหล่านี้ล้วนแล้วแต่เป็นวิชาหลักของทางคณะฯ และทางสาขาวิชาฯ ซึ่งอาจเป็นปัจจัยหลักที่ส่งผลการสำเร็จการศึกษาตามแผนของนักศึกษาอย่างมีนัยสำคัญ



#### 4. สรุปและการอภิปราย

การวิจัยนี้มีวัตถุประสงค์เพื่อที่ศึกษาปัจจัยที่ส่งผลต่อการสำเร็จ การศึกษาตามแผน และเพื่อปรับปรุงรูปแบบการเรียนการสอน ในแต่ละรายวิชาที่ส่งผลต่อการไม่สำเร็จการศึกษาของนักศึกษา จากผลการวิจัยนี้สามารถสรุปได้ว่าจากการใช้เทคนิคการ คัดเลือกคุณลักษณะ ทั้งสามเทคนิค (1) chi-square feature selection (2) information gain feature selection และ (3) correlation based feature selection ปัจจัยที่ส่งผลต่อการสำเร็จ การศึกษาของนักศึกษามากที่สุดได้แก่ วิชาการเขียนโปรแกรม คอมพิวเตอร์ 1 ซึ่งมีความสำคัญมาเป็นอันดับหนึ่งในทุกเทคนิค ของการคัดเลือกคุณลักษณะ และในส่วนของปัจจัยอื่นๆ ที่คิด อันดับหนึ่งถึงสิบในทุกเทคนิคของการคัดเลือกคุณลักษณะ ได้แก่ วิชาการเงินธุรกิจ วิชาการวิเคราะห์ธุรกิจเชิงปริมาณ วิชาการบัญชีขั้นต้น วิชาการบริหารฐานข้อมูล และวิชาการ เขียนโปรแกรมบนเว็บ ซึ่งวิชาที่คิดอันดับหนึ่งถึงสิบครบทุก เทคนิคของการคัดเลือกคุณลักษณะล้วนแต่เป็นวิชาพื้นฐานและ เป็นวิชาบังคับของคณะบริหารธุรกิจและศิลปศาสตร์และสาขา ระบบสารสนเทศทางคอมพิวเตอร์ และนักศึกษาส่วนใหญ่ได้ ระดับคะแนนในแต่ละวิชาค่อนข้างน้อย ซึ่งอาจส่งผลให้รายวิชา เหล่านี้คิดอันดับหนึ่งถึงสิบในทุกเทคนิคของการคัดเลือก คุณลักษณะและอาจส่งผลต่อการสำเร็จการศึกษาตามแผนของ นักศึกษาอย่างมีนัยสำคัญ

ปัจจัยที่คิดอันดับหนึ่งถึงสิบเพียงสองเทคนิคของการ คัดเลือกคุณลักษณะได้แก่ วิชาภาษาอังกฤษผ่านสื่อและ เทคโนโลยี วิชาการระบบสารสนเทศในองค์กร วิชาการคิดและการ ตัดสินใจเชิงวิทยาศาสตร์ วิชาโครงสร้างข้อมูลและอัลกอริทึม และระดับคะแนนเฉลี่ย (GPA) โดยวิชาการระบบสารสนเทศใน องค์กรและวิชาโครงสร้างข้อมูลและอัลกอริทึมเป็นวิชาพื้นฐาน และเป็นวิชาบังคับของทางสาขาฯ และนักศึกษาส่วนใหญ่ได้ ระดับคะแนนในแต่ละวิชาค่อนข้างน้อยเช่นกัน ซึ่งอาจส่งผลต่อ การสำเร็จการศึกษาตามแผนของนักศึกษาอย่างมีนัยสำคัญ และ ปัจจัยที่คิดอันดับหนึ่งถึงสิบเพียงหนึ่งเทคนิคการคัดเลือก คุณลักษณะได้แก่ วิชาภาษาอังกฤษเพื่อการสื่อสารและวิชาสถิติ พื้นฐาน

#### 5. ข้อเสนอแนะ

เทคนิคการคัดเลือกคุณลักษณะทั้งสามเทคนิคที่ผู้วิจัยได้นำเสนอ นั้นเป็นเพียงบางเทคนิคของการคัดเลือกคุณลักษณะจาก หลากหลายเทคนิคที่มีอยู่ในปัจจุบัน และยังมีข้อจำกัดในเรื่อง ของปัจจัยข้อมูลที่มีเพียงข้อมูลระดับคะแนนรายวิชา เกรดเฉลี่ย เพศของนักศึกษาเท่านั้น ซึ่งอาจจะมีปัจจัยในด้านอื่นๆ ที่อาจ ส่งผลต่อการสำเร็จการศึกษาของนักศึกษาได้ เช่น ข้อมูลด้าน สถานภาพทางครอบครัว สถานการณ์ผู้ยืมเงิน ข้อมูลการศึกษา ในอดีต เป็นต้น ในอนาคตหากมีการพัฒนางานวิจัยให้มี ประสิทธิภาพมากขึ้น อาจจะต้องมีการเก็บรวบรวมข้อมูลเพื่อให้ ได้ปัจจัยที่เพิ่มมากขึ้นและใช้เทคนิคในการคัดเลือกคุณลักษณะ ในหลายๆ เทคนิคมากขึ้นเพื่อสามารถนำผลของการทดลองไป ใช้ได้อย่างมีประสิทธิภาพสูงสุด และในงานวิจัยนี้ผู้วิจัยยัง ไม่ได้ มีกำหนัดในการนำคุณลักษณะที่คัดเลือกได้ไปหาค่าความ แม่นยำในความถูกต้องของแต่ละเทคนิคการคัดเลือกคุณลักษณะ ที่ใช้ ซึ่งผู้วิจัยจะมีการดำเนินการในหัวข้อดังกล่าวในงานวิจัย ถัดไป

#### เอกสารอ้างอิง

- [1] วรการ ใจดี, “ตัวแบบการทำนายผลการสำเร็จการศึกษาตาม แผน โดยใช้เทคนิคเอสเอ็มไอทีอีและตัวแบบการเรียนรู้ แบบรวมกลุ่ม บนชุดข้อมูลที่ไม่สมดุล สำหรับข้อมูลการ สำเร็จการศึกษา,” รายงานสืบเนื่องจากการประชุมสัมมนา ทางวิชาการ มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา ตะวันออก ระดับชาติครั้งที่ 12, 2562, หน้า 325-335.
- [2] พิชัย ระวังวัน, “โมเดลเพื่อการพยากรณ์สถานภาพทางการ ศึกษาของนักศึกษา,” รายงานสืบเนื่องจากการประชุม วิชาการเสนอผลงานวิจัยบัณฑิตศึกษา ระดับชาติและ นานาชาติ มหาวิทยาลัยขอนแก่น, 2560, หน้า 273-283.
- [3] เบนจรณ์ จันทรวงกุล, “วิธีการที่เหมาะสมสำหรับการ แบ่งกลุ่มข้อมูลที่ไม่สมดุลสูง,” วิทยาศาสตร์สุขภาพบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์, มหาวิทยาลัยบูรพา, 2557.
- [4] กิตติพงศ์ ชมบุญ, “เทคนิคการจำแนกประเภทข้อมูลส่วน น้อยบน ข้อมูลไม่สมดุลด้วยวิธีการแบ่งข้อมูล,” วิศวกรรมศาสตร์สุขภาพบัณฑิต สาขาวิชาวิศวกรรม คอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีสุรนารี, 2558.

- [5] อัจจิมา มณฑาพันธุ์, “การเปรียบเทียบวิธีการคัดเลือกคุณลักษณะที่สำคัญในการปรับปรุงการพยากรณ์มะเร็งเต้านม,” Royal Thai Air Force Medical Gazette, 2562, หน้า 49-56.
- [6] Zheng Z, Cai Y, Li Y, “Oversampling Method for Imbalanced Classification,” Computing and Informatics, Vol. 34, 2015, pp. 1017-1037.
- [7] Beretta L, Santaniello A, “Nearest Neighbor Imputation Algorithm: A Critical Evaluation,” The 5<sup>th</sup> Translational Bioinformatics Conference, 2015, pp. 197-208.