



**การวิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษ
ของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคเหมืองข้อมูล**

**Using data mining techniques to analyze most characteristics that affect the performance of
undergraduate students learning English.**

เสนอ

ผู้ช่วยศาสตราจารย์ ดร.สุเทพ ทองงาม

จัดทำโดย

นางสาวชิตชนก แก้วพานิชย์

การวิจัยนี้เป็นส่วนหนึ่งของวิชา บทส.9000 การค้นคว้าอิสระ

หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาบริหารเทคโนโลยีสารสนเทศ

คณะ สถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์ปีการศึกษา 2563

การวิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษ
ของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคเหมืองข้อมูล

ชิดชนก แก้วพานิชย์

คณะสถิติประยุกต์

คณะกรรมการสอบการค้นคว้าอิสระ ได้พิจารณาแล้วเห็นสมควรอนุมัติให้เป็นส่วนหนึ่งของการศึกษาตาม
หลักสูตรการจัดการมหาบัณฑิต(บริหารเทคโนโลยีสารสนเทศ)

.....อาจารย์ที่ปรึกษาหลัก

(ผู้ช่วยศาสตราจารย์ ดร.สุเทพ ทองงาม)

.....อาจารย์ที่ปรึกษาร่วม

(ผู้ช่วยศาสตราจารย์ ดร.ชนาสัย สุคนธ์พันธุ์)

.....ประธานกรรมการ

()

.....กรรมการ

()

.....กรรมการ

()

ชื่อวิจัย	การวิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษา ระดับปริญญาตรี โดยใช้เทคนิคเหมืองข้อมูล
ผู้วิจัย	นางสาวชัชชนก แก้วพานิชย์
หลักสูตร	วิทยาศาสตรมหาบัณฑิต (บริหารเทคโนโลยีสารสนเทศ)
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.สุเทพ ทองงาม
ปีการศึกษา	2563

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและวิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี โดยเปรียบเทียบการทำเหมืองข้อมูลระหว่างเทคนิค Decision Tree และเทคนิค Random Forest การวิจัยนี้ใช้กลุ่มตัวอย่างข้อมูลจากนักศึกษาระดับปริญญาตรี คณะภูมิสารสนเทศศาสตร์ มหาวิทยาลัยบูรพา ที่เข้าศึกษาระหว่างปี พ.ศ.2561 - 2563 จำนวน 299 คน และใช้ข้อมูลจำนวน 32 คุณลักษณะสำหรับการทำเหมืองข้อมูล ซึ่งรวบรวมโดยแบบสอบถามที่ผ่านการตรวจสอบความเที่ยงตรงเชิงเนื้อหา IOC มีค่ามากกว่า 0.7 พัฒนาและทดสอบประสิทธิภาพของแบบจำลองด้วยวิธี 10 - fold cross validation ด้วยอัตราส่วนข้อมูล 70:30 รวมถึงตีความและวิเคราะห์ผลลัพธ์ของตัวแบบข้อมูลด้วยการคำนวณค่าความแม่นยำ (Accuracy) ค่าความถูกต้อง (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F1 score) ด้วยรูปแบบ Confusion Matrix

ผลการศึกษาวิจัย พบว่า แบบจำลองที่พัฒนาด้วยเทคนิค Random Forest ให้ค่าความแม่นยำในการทำนายประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาที่ค่าเฉลี่ยร้อยละ 85.14 ซึ่งมากกว่าเทคนิค Decision Tree ที่มีค่าความแม่นยำที่ค่าเฉลี่ยร้อยละ 78.0 โดยใช้ชุดข้อมูลเดียวกัน และอัลกอริทึม Random Forest สร้างการจัดอันดับคุณลักษณะความสัมพันธ์ระหว่างประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษและคุณลักษณะอื่นๆ พบว่า ความสนุกกับการเรียนภาษาอังกฤษมีความสัมพันธ์กับประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาที่ 0.1194 การถามอาจารย์ทันทีเมื่อไม่เข้าใจเนื้อหาที่เรียนมีความสัมพันธ์กับประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาที่ 0.1149 ความสนใจและตั้งใจเรียนวิชาภาษาอังกฤษในขณะที่อาจารย์สอนมีความสัมพันธ์กับประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาที่ 0.0884 การจับกลุ่มทบทวนบทเรียนภาษาอังกฤษกับเพื่อนเมื่อใกล้สอบมีความสัมพันธ์กับประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาที่ 0.0864 และการศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษนอกเหนือจากที่อาจารย์สอนมีความสัมพันธ์กับประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาที่ 0.0747 การวัดแต่ละครั้งเป็นผลจากการคำนวณคุณลักษณะที่ส่งผลต่อการเรียนภาษาอังกฤษในสภาพแวดล้อมของมหาวิทยาลัย งานวิจัยนี้สามารถช่วยปรับปรุงประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี และช่วยให้อาจารย์เข้าใจแนวทางการสอนที่จะส่งผลต่อประสิทธิภาพของนักเรียนต่อไป

คำสำคัญ : การทำเหมืองข้อมูล, Decision Tree, Random Forest, ประสิทธิภาพการเรียนรู้, การเรียนภาษาอังกฤษ

Title	Using data mining techniques to analyzing characteristics that affect the performance of undergraduate students learning English.
Researcher	Miss Chidchanok Kaewpanitch
Degree	MASTER OF SCIENCE (INFORMATION TECHNOLOGY MANAGEMENT)
Advisor	Assistant Professor Dr.Suthep Tongngam
Year	2020

ABSTRACT

The objective of this research was to study and analyze key features affecting the effectiveness of English language learning among undergraduate students. This research utilized data mining techniques and machine learning algorithms to compare the effectiveness and accuracy of the Decision Tree and Random Forest Classifier Supervised Machine Learning algorithms. The Faculty of Geoinformatics at Burapha University Academic year 2018 - 2020 consisted of 299 students. From this department, a survey was used to collect 32 attributes for data mining and analysis. The Content Validation Index used to calculate content validity in the questionnaire, and was greater than 0.7. The Decision Tree and Random Forest Classifier performance were evaluated using a 10-fold cross-validation method with a data ratio of 70:30. Accuracy, Precision, and balance (F1 score) with Confusion Matrix format was used to interpret and analyze the results of the data models.

The results indicated the data model developed with Random Forest Classifier produced a higher classification accuracy for students' English performance than that of the Decision Tree algorithm. The Random Forest Classifier produced an accuracy of 85.14% in classifying students' performance based upon a set of attributes in a given dataset, whereas the Decision Tree produced an accuracy of 78.0% using the same dataset. Random Forest Classifier produced an attribute ranking and correlation coefficient metrics to measure relationship strengths between student grades and other attributes. It found that students that enjoyed learning English had a 0.1194 correlation to the students' English performance. Asking teachers immediately when they didn't understand the content of the lesson had a correlation of 0.1149 to the students' English performance. Interest and attention to English showed a 0.0884 correlation to the students' English performance. Students that studied in groups to review English lessons with friends when the exam was approaching had a correlation of 0.0864 to students' English performance. Students that invested extra time into research and studying English beyond the classroom had a correlation value of 0.0747 to the students' English performance. Each of these measurements was the result of calculating the characteristics that affect English language learning in a University environment. This academic research can help to improve undergraduate students' performance when learning English and can help professors to understand how teaching methods can impact student performance.

Keyword: Data Mining, Decision Tree, Random Forest, Academic Efficiency, Learning English

กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จเรียบร้อยได้อันเนื่องมาจากความกรุณาอย่างยิ่งจาก ผู้ช่วยศาสตราจารย์ ดร.สุเทพ ทองงาม ซึ่งเป็นอาจารย์ที่ปรึกษาที่ได้สละเวลาอันมีค่าของท่านช่วยแนะนำ ให้คำปรึกษา ตลอดจนตรวจแก้ไขข้อบกพร่องต่างๆ และปรับปรุงกระบวนการทำงานจนสำเร็จลุล่วงไปได้ด้วยดี รวมถึงอาจารย์ที่ปรึกษาร่วม ผู้ช่วยศาสตราจารย์ ดร.ธนาชัย สุคนธ์พันธุ์ที่ช่วยแนะนำข้อมูลที่เป็นประโยชน์เสมอมา ข้าพเจ้าใคร่ขอขอบพระคุณท่านอาจารย์ทั้งสองไว้เป็นอย่างสูง

ขอขอบพระคุณ คณาจารย์และนักศึกษาคณะภูมิสารสนเทศศาสตร์ มหาวิทยาลัยบูรพาที่กรุณาเสียสละเวลาให้ข้อมูลที่เป็นประโยชน์ต่องานวิจัย และสนับสนุนข้อมูลต่างๆ เป็นอย่างดี

ขอขอบคุณเพื่อนนักศึกษา ที่คอยช่วยเหลือ และให้คำปรึกษาต่างๆ เป็นอย่างดี ตลอดจนคอยผลักดัน และให้กำลังใจ อยู่เบื้องหลังเสมอมา

ขอขอบพระคุณเจ้าของผลงานวิชาการต่างๆ ที่ผู้ศึกษาค้นคว้าข้อมูลและนำมาเป็นแนวทางและใช้เป็นต้นแบบให้ งานวิจัยชิ้นนี้

ขอขอบคุณครอบครัวของข้าพเจ้าที่ให้การสนับสนุนในด้านการศึกษา เป็นกำลังใจที่ยิ่งใหญ่ และช่วยผลักดัน ให้ดำเนินงานวิจัยครั้งนี้ประสบความสำเร็จ

ข้าพเจ้าหวังเป็นอย่างยิ่งว่า งานวิจัยฉบับนี้จะเป็นประโยชน์ไม่มากนักน้อย รวมถึงยินดีรับฟังคำแนะนำจากทุกท่าน หากมีสิ่งขาดตกบกพร่องหรือผิดพลาดประการใด ข้าพเจ้าขออภัยเป็นอย่างสูงในข้อบกพร่องและความผิดพลาด ณ ที่นี้ ด้วย

ชิดชนก แก้วพานิชย์

มิถุนายน 2564

สารบัญ

บทคัดย่อ.....	ก
ABSTRACT.....	ข
กิตติกรรมประกาศ.....	ค
บทที่ 1 บทนำ	
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 คำถามวิจัย.....	2
1.3 วัตถุประสงค์ของการวิจัย.....	2
1.4 ขอบเขตของการวิจัย.....	2
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	3
1.6 นิยามคำศัพท์.....	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	
2.1 ทัศนคติและแรงจูงใจในการเรียนภาษาอังกฤษ.....	4
2.2 พฤติกรรมส่งผลต่อการเรียนภาษาอังกฤษ.....	5
2.3 แนวคิดการทำเหมืองข้อมูล (Data Mining).....	6
2.4 งานวิจัยที่เกี่ยวข้อง.....	14
บทที่ 3 วิธีการดำเนินงานวิจัย	
3.1 ประชากรและกลุ่มตัวอย่าง.....	18
3.2 เครื่องมือที่ใช้ในการวิจัย.....	19
3.3 ขั้นตอนการดำเนินการ.....	19
3.4 กรอบแนวคิดการวิจัย.....	43
บทที่ 4 ผลการวิจัย.....	44
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....	51
บรรณานุกรม.....	54
ภาคผนวก	
ภาคผนวก ก แบบสอบถามงานวิจัย.....	57
ภาคผนวก ข แบบประเมินแบบสอบถามเชิงปริมาณผ่านผู้เชี่ยวชาญ.....	64
ภาคผนวก ค สรุปผลการคำนวณและแปลผลค่าแบบประเมินความเที่ยงตรง (IOC).....	72

สารบัญตาราง

ตารางที่	หน้า
ตาราง 3.1 ชุดข้อมูล (Dataset) ที่เกี่ยวข้องกับความสำเร็จทางการเรียนวิชาภาษาอังกฤษสำหรับใช้ประมวลผล	
ข้อมูล	21
ตารางที่ 4.1 แสดงผลการทำสอบแบบจำลองด้วยวิธี k-fold cross validation	44
ตารางที่ 4.2 แสดงผลการทดสอบประสิทธิภาพแบบจำลองด้วยการเปรียบเทียบ	45
ตารางที่ 4.3 แสดงคุณลักษณะ 5 อันดับแรกที่สัมพันธ์กับประสิทธิภาพการเรียนวิชาภาษาอังกฤษของนักศึกษา	46

สารบัญภาพ

ภาพที่	หน้า
ภาพที่ 2.1 มาตรฐานการทำเหมืองข้อมูลตามขั้นตอน CRISP-DM	8
ภาพที่ 2.2 ตัวอย่าง Decision Tree สำหรับการจำแนกประเภทข้อมูล	9
ภาพที่ 2.3 โครงสร้าง Random Forest	12
ภาพที่ 2.4 กระบวนการของวิธี Cross Validation	13
ภาพที่ 3.1 ตัวอย่างข้อมูลที่มีความผิดพลาดไม่สอดคล้องกับรูปแบบข้อมูลที่ต้องการใช้งาน	20
ภาพที่ 3.2 ตัวอย่างข้อมูลที่ปรับเป็นภาษาอังกฤษ	21
ภาพที่ 3.3 ตัวอย่างข้อมูลที่แปลงเป็นตัวเลข ด้วยโปรแกรม Jupyter	23
ภาพที่ 3.4 กราฟภาพรวมของข้อมูลคุณลักษณะทั้งหมด	24
ภาพที่ 3.5 กราฟภาพเปรียบเทียบระหว่างเพศ (GENDER) กับคุณลักษณะเป้าหมาย	25
ภาพที่ 3.6 กราฟภาพเปรียบเทียบระหว่างอายุ (AGE) กับคุณลักษณะเป้าหมาย	25
ภาพที่ 3.7 กราฟภาพเปรียบเทียบระหว่างมีคอมพิวเตอร์และอินเทอร์เน็ตใช้ส่วนตัว (H-COM) กับคุณลักษณะเป้าหมาย	26
ภาพที่ 3.8 กราฟภาพเปรียบเทียบระหว่างลักษณะที่ตั้งของบ้านที่อยู่อาศัย (LOC) กับคุณลักษณะเป้าหมาย	26
ภาพที่ 3.9 กราฟภาพเปรียบเทียบระหว่างจำนวนการขาดเรียนวิชาภาษาอังกฤษ (ABSENCE) กับคุณลักษณะเป้าหมาย	27
ภาพที่ 3.10 กราฟภาพเปรียบเทียบระหว่างเรียนพิเศษวิชาภาษาอังกฤษ (S-ENG) กับคุณลักษณะเป้าหมาย	27
ภาพที่ 3.11 กราฟภาพเปรียบเทียบระหว่างเวลาว่างจากการเรียน (F-TIME) กับคุณลักษณะเป้าหมาย	28
ภาพที่ 3.12 กราฟภาพเปรียบเทียบระหว่างเวลาสำหรับออกไปเที่ยวกับเพื่อน (H-EREQ) กับคุณลักษณะเป้าหมาย	28
ภาพที่ 3.13 กราฟภาพเปรียบเทียบระหว่างเวลาว่างสำหรับฝึกฝนภาษาอังกฤษด้วยตนเอง (S-STUDY) กับคุณลักษณะเป้าหมาย	29
ภาพที่ 3.14 กราฟภาพเปรียบเทียบระหว่างถามอาจารย์ทันทีเมื่อไม่เข้าใจเนื้อหาที่เรียน (A-PROF) กับคุณลักษณะเป้าหมาย	29
ภาพที่ 3.15 กราฟภาพเปรียบเทียบระหว่างวางแผนการเรียนภาษาอังกฤษ (E-PLAN) กับคุณลักษณะเป้าหมาย	30
ภาพที่ 3.16 กราฟภาพเปรียบเทียบระหว่างเตรียมตัวก่อนเข้าเรียนวิชาภาษาอังกฤษ (P-SELF) กับคุณลักษณะเป้าหมาย	30

ภาพที่ 3.17 กราฟภาพเปรียบเทียบระหว่างชอบนั่งใกล้ๆ เพื่อนที่ตั้งใจเรียน (S-FRIEND) กับคุณลักษณะเป้าหมาย	31
ภาพที่ 3.18 กราฟภาพเปรียบเทียบระหว่างมีเพื่อนเป็นชาวต่างชาติ เพื่อฝึกใช้ทักษะภาษาอังกฤษ (F-FRIEND) กับคุณลักษณะเป้าหมาย	31
ภาพที่ 3.19 กราฟภาพเปรียบเทียบระหว่างสนใจและตั้งใจเรียนวิชาภาษาอังกฤษในขณะที่อาจารย์สอน (I-ENG) กับคุณลักษณะเป้าหมาย	32
ภาพที่ 3.20 กราฟภาพเปรียบเทียบระหว่างจดบันทึกตามที่อาจารย์สอน (T-NOTES) กับคุณลักษณะเป้าหมาย	32
ภาพที่ 3.21 กราฟภาพเปรียบเทียบระหว่างตอบคำถามในขณะที่เรียนภาษาอังกฤษ (A-QUE) กับคุณลักษณะเป้าหมาย	33
ภาพที่ 3.22 กราฟภาพเปรียบเทียบระหว่างทบทวนบทเรียนภาษาอังกฤษหลังเลิกเรียน (A-SCHOOL) กับคุณลักษณะเป้าหมาย	33
ภาพที่ 3.23 กราฟภาพเปรียบเทียบระหว่างศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษนอกเหนือจากที่อาจารย์สอน (E-RES) กับคุณลักษณะเป้าหมาย	34
ภาพที่ 3.24 กราฟภาพเปรียบเทียบระหว่างจับกลุ่มทบทวนบทเรียนภาษาอังกฤษกับเพื่อนเมื่อใกล้สอบ (G-STUDY) กับคุณลักษณะเป้าหมาย	34
ภาพที่ 3.25 กราฟภาพเปรียบเทียบระหว่างความต้องการศึกษาในระดับที่สูงขึ้น (H-EDU) กับคุณลักษณะเป้าหมาย	35
ภาพที่ 3.26 กราฟภาพเปรียบเทียบระหว่างการสนับสนุนด้านการเรียนภาษาอังกฤษจากครอบครัว (F-SUPP) กับคุณลักษณะเป้าหมาย	35
ภาพที่ 3.27 กราฟภาพเปรียบเทียบระหว่างสนุกกับการเรียนภาษาอังกฤษ (E-ENG) กับคุณลักษณะเป้าหมาย	36
ภาพที่ 3.28 กราฟภาพเปรียบเทียบระหว่างสนใจเรียนภาษาอังกฤษเพื่อพัฒนาตนเอง (L-IMPROV) กับคุณลักษณะเป้าหมาย	36
ภาพที่ 3.29 กราฟภาพเปรียบเทียบระหว่างสนใจเรียนภาษาอังกฤษ เพื่อหางานที่ดีในอนาคต (L-JOB) กับคุณลักษณะเป้าหมาย	37
ภาพที่ 3.30 กราฟภาพเปรียบเทียบระหว่างสนใจเรียนภาษาอังกฤษ เพราะเป็นความต้องการของพ่อแม่ (L-PAREN) กับคุณลักษณะเป้าหมาย	37
ภาพที่ 3.31 กราฟภาพเปรียบเทียบระหว่างสนใจเรียนภาษาอังกฤษ เพื่อต้องการสอบผ่าน (L-PASS) กับคุณลักษณะเป้าหมาย	38
ภาพที่ 3.32 กราฟภาพเปรียบเทียบระหว่างสนใจเรียนภาษาอังกฤษ เพื่อใช้สื่อสารในชีวิตประจำวัน(L-DAILY) กับคุณลักษณะเป้าหมาย	38
ภาพที่ 3.33 กราฟภาพเปรียบเทียบระหว่างความสุขในการอ่านหนังสือ ดูภาพยนตร์เป็นภาษาอังกฤษ (E-MOV) กับคุณลักษณะเป้าหมาย	39

ภาพที่ 3.34 กราฟภาพเปรียบเทียบระหว่างการเรียนภาษาอังกฤษมีความสำคัญ เพราะภาษาอังกฤษช่วยให้มีความรู้มากขึ้น (E-KNOW) กับคุณลักษณะเป้าหมาย.....	39
ภาพที่ 3.35 กราฟภาพเปรียบเทียบระหว่างการเรียนภาษาอังกฤษมีความสำคัญ เพราะต้องการได้เกรดที่ดี (E-GRADE) กับคุณลักษณะเป้าหมาย.....	40
ภาพที่ 3.36 ภาพคำนวณหาค่าความสัมพันธ์ ด้วยวิธีสหสัมพันธ์แบบเพียร์สันในโปรแกรม Jupyter.....	40
ภาพที่ 3.37 ข้อมูลคุณลักษณะที่มีถูกคัดเลือกนำมาใช้สร้างตัวแบบจำลองใน โปรแกรม Jupyter.....	41
ภาพที่ 3.38 การแบ่งสัดส่วนชุดข้อมูลในโปรแกรม Jupyter.....	42
ภาพที่ 3.39 กรอบแนวคิดงานวิจัย.....	43
ภาพที่ 4.1 การทำงานการจำแนกประเภทข้อมูลด้วยเทคนิค Decision Tree และเทคนิค Random Forest.....	45
ภาพที่ 4.2 กราฟแสดงอันดับของตัวแปรคุณลักษณะที่บ่งบอกถึงประสิทธิภาพการเรียนวิชาภาษาอังกฤษของนักศึกษา.....	46
ภาพที่ 4.3 แสดงผลประสิทธิภาพของแบบจำลองโดยใช้ Confusion Metrix.....	47
ภาพที่ 4.4 แสดงอัตราส่วนนักศึกษาได้ผลการเรียนเกรดเอวิชาภาษาอังกฤษระดับพื้นฐานในการสำรวจความรู้สึกสนุกสนานเมื่อได้เรียนภาษาอังกฤษ.....	48
ภาพที่ 4.5 แสดงอัตราส่วนนักศึกษาได้ผลการเรียนเกรดเอวิชาภาษาอังกฤษระดับพื้นฐาน.....	48
ภาพที่ 4.6 แสดงอัตราส่วนนักศึกษาได้ผลการเรียนเกรดเอวิชาภาษาอังกฤษระดับพื้นฐานในการสำรวจความสนใจและตั้งใจเรียนวิชาภาษาอังกฤษในขณะที่อาจารย์สอนของนักศึกษา.....	49
ภาพที่ 4.7 แสดงอัตราส่วนนักศึกษาได้ผลการเรียนเกรดเอวิชาภาษาอังกฤษระดับพื้นฐานในการสำรวจการศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษนอกเหนือจากที่อาจารย์สอน.....	50

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

“ภาษาอังกฤษ” เป็นภาษาต่างประเทศที่นิยมใช้กันอย่างแพร่หลาย และเป็นภาษากลางที่ใช้สื่อความหมายไปเกือบทั่วโลก และมากกว่า 2,000 ล้านคน หรือ 1 ใน 3 ของประชากรของโลก ใช้ภาษาอังกฤษเป็นหลักในการสื่อสาร (ชนกร และ เสี่ยงม, 2559) ดังนั้น จำเป็นอย่างยิ่งที่จะต้องส่งเสริมให้ประชากรไทย โดยเฉพาะกลุ่มระดับอุดมศึกษาซึ่งจะเป็นกลุ่มกำลังสำคัญในอนาคตของประเทศ ควรได้รับการพัฒนาทักษะด้านภาษาอังกฤษ และยกระดับการเรียนรู้ภาษาอังกฤษให้มีประสิทธิภาพเพิ่มมากขึ้น สามารถใช้เป็นเครื่องมือในการแสวงหาความรู้ และประกอบอาชีพ นำไปสู่การแข่งขันด้านเศรษฐกิจและสังคมในเวทีสากล อย่างไรก็ตาม แม้ว่าประเทศไทยได้พยายามเพิ่มขีดความสามารถในการใช้ภาษาอังกฤษ แต่ระดับภาษาอังกฤษของคนไทยยังคงไม่ได้มีการพัฒนาเท่าที่ควร จากรายงานการจัดอันดับภาษาอังกฤษ 100 ประเทศทั่วโลกที่ไม่ได้ใช้ทักษะภาษาอังกฤษเป็นภาษาหลักของสถาบัน EF ซึ่งเกณฑ์การวัดระดับภาษาอังกฤษ แบ่งออกเป็น 5 ระดับ เริ่มจากการใช้ภาษาอังกฤษดีมาก (Very high), ดี (High), ปานกลาง (Moderate), แย่ (Low), แย่มาก (Very low) พบว่าภาษาอังกฤษของคนไทยจัดอยู่ในกลุ่มสุดท้าย คือ แย่มาก (Very Low) ได้คะแนนเพียง 47.61 ซึ่งน้อยลงมา 3 ปีซ้อน คือ ลงจากปี 2561 ได้ 48.54 คะแนน ลงจากปี 2560 ได้ 49.7 คะแนน และผลการจัดอันดับภาษาอังกฤษของคนไทยในปี 2563 พบว่า การใช้ภาษาอังกฤษของคนไทย ยังคงตกต่ำอย่างต่อเนื่อง อันดับลดลงอยู่อันดับที่ 89 (Education First, 2020)

ดังนั้น การวางแผนส่งเสริมทักษะภาษาอังกฤษในระดับอุดมศึกษาตั้งแต่วิชาบังคับพื้นฐาน โดยเริ่มจากการทำความเข้าใจพฤติกรรมและแรงจูงใจที่ส่งผลต่อการเรียนวิชาภาษาอังกฤษ อาจเป็นตัวช่วยหนึ่งที่จะช่วยส่งเสริมให้ทักษะภาษาอังกฤษของคนไทยในอนาคตมีแนวโน้มที่ดีขึ้น เพราะหากสามารถเข้าใจพฤติกรรมเชิงลึกถึงสาเหตุที่ทำให้นักศึกษาบางคนมีผลการเรียนที่ดี จะทำให้คาดการณ์พฤติกรรมล่วงหน้าเพื่อเตรียมความพร้อมการเรียนการสอนที่เหมาะสม รวมถึงพัฒนากลุ่มนักศึกษาอื่นๆ ให้มีผลการเรียนภาษาอังกฤษในระดับที่ดีขึ้นได้ มีการศึกษาวิจัยเกี่ยวกับความสัมพันธ์ระหว่างปัจจัยทัศนคติ แรงจูงใจ และพฤติกรรมการเรียนภาษาอังกฤษของนักศึกษามหาวิทยาลัยราชภัฏภูเก็ต ผลการศึกษพบว่า ปัจจัยด้านทัศนคติ แรงจูงใจภายใน และพฤติกรรมการเรียนมีความเชื่อมโยงกับความสำเร็จในการเรียนวิชาภาษาอังกฤษ ซึ่งทัศนคติที่ดีมีอิทธิพลมากที่สุดต่อผลสัมฤทธิ์ทางการศึกษาวิชาภาษาอังกฤษ (สุภิกา, 2561) และแรงจูงใจยังมีบทบาทสำคัญในการเรียนภาษาต่างประเทศหรือภาษาที่สอง ให้ประสบความสำเร็จอีกด้วย แม้ว่าแรงจูงใจของผู้เรียนจะสามารถเปลี่ยนแปลงได้เสมอขึ้นอยู่กับบริบทของการเรียนรู้ภาษาก็ตาม (Mitra, 2016) ในขณะเดียวกันเทคนิคเหมืองข้อมูล (Data mining) เริ่มได้รับความสนใจนำมาใช้ค้นหาความสัมพันธ์ที่อยู่ภายในชุดข้อมูลมากขึ้น และสามารถนำไปประยุกต์ใช้ได้หลากหลายด้าน รวมถึงด้านการศึกษา ยกตัวอย่างเช่น การใช้เทคนิค Decision tree ทำนายผลการเรียนของนักศึกษาวิศวกรรมศาสตร์ในประเทศอินเดีย เพื่อทราบข้อมูลของนักศึกษาที่มีแนวโน้มจะไม่ประสบความสำเร็จในการเรียนล่วงหน้า และวางแผนดำเนินการที่จำเป็นก่อนที่จะเกิดขึ้น (Kabra & Bichkar, 2011) อีกทั้งนักวิจัยคนไทยได้เปรียบเทียบเทคนิคเหมืองข้อมูลหลายวิธี เพื่อหารูปแบบการทำนายที่มีประสิทธิภาพที่สุด สำหรับการทำนายผลของการสอบภาษาอังกฤษวิชาสุดท้ายก่อนจบการศึกษา ซึ่งพบว่า เทคนิค Decision tree มีประสิทธิภาพสูงสุด และตัวกำหนดที่สำคัญของการทำนายผล การสอบภาษาอังกฤษ คือ การทดสอบวัดระดับภาษาอังกฤษ นักศึกษาที่สอบผ่านการทดสอบวัดระดับภาษาอังกฤษในปีแรก มักจะสอบผ่านในปีสุดท้าย (Puarungroj, Boonsirisumpun, Pongpatrakant, และ Phromkhot, 2017) จะเห็นได้ว่าการทำเหมืองข้อมูลเป็นเครื่องมือที่จะช่วยเพิ่มโอกาสในการศึกษาวิเคราะห์ข้อมูลเชิงลึก ทำให้ทราบปัจจัยสำคัญเพื่อการวางแผนรองรับและสร้างกระบวนการเรียนรู้ระหว่างความสัมพันธ์ต่าง ๆ อย่างมีประสิทธิภาพ

ดังนั้น ผู้วิจัยจึงเกิดแนวคิดในการศึกษาความเป็นไปได้ที่ใช้ประโยชน์จากข้อมูลด้านพฤติกรรม ทักษะคิดและแรงจูงใจ ประยุกต์ร่วมกับกระบวนการทำเหมืองข้อมูล เพื่อจำแนกประเภทคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อประสิทธิภาพการเรียนรู้ภาษาอังกฤษ และทำนายแนวโน้มคุณลักษณะที่จะส่งผลให้นักศึกษามีผลการเรียนวิชาภาษาอังกฤษที่ดี โดยการศึกษาครั้งนี้ใช้กรณีศึกษาจากนักศึกษา คณะภูมิสารสนเทศศาสตร์ มหาวิทยาลัยบูรพา มีวัตถุประสงค์เพื่อมุ่งหวังที่จะเข้าถึงผลสัมฤทธิ์ทางการเรียนวิชาภาษาอังกฤษพื้นฐานของนักศึกษาที่ไม่ได้เรียนภาษาอังกฤษเป็นวิชาหลัก รวมถึงเพื่อเตรียมความพร้อมและยกระดับระดับความรู้ทางด้านภาษาอังกฤษให้ดียิ่งขึ้นทั้งในชีวิตประจำวัน และการใช้ในการพัฒนาตนเองในด้านต่างๆ ต่อไปได้อย่างมีประสิทธิภาพ

1.2 คำถามในการวิจัย

- 1.2.1 การทำเหมืองข้อมูลด้วยเทคนิค Decision Tree หรือเทคนิค Random Forest มีความเหมาะสมกับการนำมาใช้วิเคราะห์คุณลักษณะที่ส่งผลต่อประสิทธิภาพการเรียนรู้ภาษาอังกฤษของนักศึกษาระดับปริญญาตรี
- 1.2.2 คุณลักษณะแบบใดเป็นสาเหตุทำให้นักศึกษาระดับปริญญาตรี มีประสิทธิภาพการเรียนรู้ภาษาอังกฤษพื้นฐานที่ดี

1.3 วัตถุประสงค์การวิจัย

เพื่อศึกษาและวิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนรู้ภาษาอังกฤษของนักศึกษาระดับปริญญาตรี โดยเปรียบเทียบการทำเหมืองข้อมูลระหว่างเทคนิค Decision Tree และเทคนิค Random Forests

1.4 ขอบเขตการวิจัย

1.4.1 ขอบเขตด้านเนื้อหา

งานวิจัยนี้ศึกษาคุณลักษณะที่ส่งผลต่อความสำเร็จทางการเรียนวิชาภาษาอังกฤษระดับพื้นฐานของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยบูรพา ซึ่งครอบคลุม 4 มิติ ได้แก่ 1) ข้อมูลด้านการเรียนภาษาอังกฤษของนักศึกษา ประกอบด้วย เกรดวิชาภาษาอังกฤษระดับพื้นฐานของมหาวิทยาลัย และจำนวนการขาดเรียนวิชาภาษาอังกฤษ 2) ข้อมูลพื้นฐานทั่วไปของนักศึกษา 3) ข้อมูลด้านพฤติกรรมที่ส่งผลต่อการเรียนวิชาภาษาอังกฤษ และ 4) ทักษะคิดและแรงจูงใจในการเรียนภาษาอังกฤษ

1.4.2 ขอบเขตด้านประชากร

งานวิจัยนี้เก็บข้อมูลกลุ่มตัวอย่างจากนักศึกษาระดับปริญญาตรี คณะภูมิสารสนเทศศาสตร์ มหาวิทยาลัยบูรพา ที่เข้าศึกษาระหว่างปี พ.ศ.2561 - 2563 และมีผลเกรดวิชาภาษาอังกฤษระดับพื้นฐานของมหาวิทยาลัย จำนวน 299 คน

1.4.3 ขอบเขตด้านวิธีการ

งานวิจัยนี้ใช้การประมวลผลและเปรียบเทียบเทคนิคเหมืองข้อมูล (Data Mining) 2 เทคนิค คือเทคนิค Decision Tree และเทคนิค Random Forest ในการวิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อการเรียนวิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี และวัดประสิทธิภาพแบบจำลองด้วย Cross Validation และ Confusion Matrix รวบรวมข้อมูลโดยแบบสอบถามปลายปิด

1.4.4 ขอบเขตด้านระยะเวลา

งานวิจัยนี้ใช้ระยะเวลาเวลาในการศึกษา เก็บข้อมูล และประมวลผล เป็นระยะเวลา 6 เดือน ตั้งแต่เดือนมกราคม ถึงเดือนมิถุนายน พ.ศ.2564

1.5 ประโยชน์ที่คาดว่าจะได้รับ

- 1.5.1 ผลการวิจัยทำให้ทราบถึงคุณลักษณะหรือพฤติกรรมของนักศึกษาระดับปริญญาตรีที่มีประสิทธิภาพในการเรียนวิชาภาษาอังกฤษ ซึ่งสามารถนำมาใช้เป็นเป็นแนวทางพัฒนาประสิทธิภาพการเรียนภาษาของนักศึกษาให้ดีขึ้นได้
- 1.5.2 ผลการศึกษาค้นคว้าจะเป็นประโยชน์ในการวิเคราะห์ต่อความต้องการความรู้ถึงวิธีการเพิ่มประสิทธิภาพการเรียนภาษาอังกฤษ รวมถึงสามารถนำไปแนวทางในการเพิ่มประสิทธิภาพการเรียนรายวิชาอื่นๆ ของนักศึกษาได้
- 1.5.3 ผลการศึกษาค้นคว้าสามารถนำไปใช้เป็นข้อมูลประกอบการปรับปรุงการเรียนการสอน หรือออกแบบการเรียนวิชาภาษาอังกฤษที่เหมาะสมสำหรับนักศึกษา

1.6 คำนิยามศัพท์

- 1.6.1 ประสิทธิภาพการเรียนวิชาภาษาอังกฤษ หมายถึง นักศึกษาที่มีผลดีมาก ซึ่งได้รับเกรดเอในวิชาเรียนภาษาอังกฤษระดับพื้นฐานของมหาวิทยาลัยบูรพา
- 1.6.2 สถาบัน EF หมายถึง สถาบันสอนภาษาในต่างประเทศชั้นนำ (EF Education First) เป็นองค์กรด้านการศึกษานานาชาติใหญ่ที่เป็นผู้จัดอันดับประเทศจากความสามารถด้านการใช้ภาษาอังกฤษที่ได้รับการยอมรับจากทั่วโลก
- 1.6.3 การประยุกต์ใช้เทคนิคเหมืองข้อมูล หมายถึง การนำกระบวนการเทคนิคเหมืองข้อมูลมาใช้ประโยชน์ โดยปรับให้มีความเหมาะสมกับนักศึกษาระดับปริญญาตรีที่มีผลการเรียนวิชาภาษาอังกฤษระดับพื้นฐานของมหาวิทยาลัยบูรพา
- 1.6.4 ข้อมูลคุณลักษณะ หมายถึง ข้อมูลคุณสมบัติหรือพฤติกรรมเฉพาะที่เกี่ยวข้องกับชุดข้อมูลที่จะนำมาใช้ในการสร้างอัลกอริทึมด้วยเทคนิคเหมืองข้อมูล

บทที่ 2

แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้เป็นการศึกษา วิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนรู้ภาษาต่างประเทศระดับพื้นฐาน กลุ่มตัวอย่างจากนักศึกษาระดับปริญญาตรี คณะมนุษยศาสตร์ มหาวิทยาลัยบูรพา โดยใช้เทคนิค Decision Tree และเทคนิค Random Forest ซึ่งมีการรวบรวมแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง แบ่งเป็น 4 เรื่องที่เกี่ยวข้อง ดังนี้

- 2.1 ทักษะและแรงจูงใจในการเรียนภาษาอังกฤษ
- 2.2 พฤติกรรมส่งผลต่อการเรียนภาษาอังกฤษ
- 2.3 แนวคิดการทำเหมืองข้อมูล (Data Mining)
- 2.4 งานวิจัยที่เกี่ยวข้อง

2.1 ทักษะและแรงจูงใจที่ส่งผลต่อการเรียนภาษาอังกฤษ

งานวิจัยทางการเรียนรู้ภาษาหลายชิ้นได้ให้นิยามคำว่า ทักษะ ว่าเป็นความคิดของบุคคลที่กำหนดโดยความเชื่อ และนำไปสู่การแสดงออกทางพฤติกรรมตามความเชื่อนั้น โดยหากมีความเชื่อเชิงบวกต่อสิ่งหนึ่ง พฤติกรรมที่แสดงออกต่อสิ่งนั้นจะเป็นเชิงบวกเช่นกัน (Montano and Kasparzyk, 2008) นอกจากนี้ Triandis (1971) ได้ให้นิยามองค์ประกอบของ ทักษะ ว่าเป็น 3 ส่วน ได้แก่ 1) ด้านการรับรู้ ความเชื่อ และค่านิยม 2) ด้านอารมณ์หรือความรู้สึกจากการรับรู้ และ 3) พฤติกรรมที่แสดงออกมาต่อสิ่งใดสิ่งหนึ่ง สอดคล้องกับความรู้สึกชอบหรือไม่ชอบต่อสิ่งนั้น ดังนั้น สามารถสรุปได้ว่า ทักษะ ความรู้สึก และพฤติกรรม มีความสัมพันธ์ระหว่างกันไม่ทางใดก็ทางหนึ่ง

ทักษะ เป็นสิ่งที่ยังคงอยู่ต่อการเรียนรู้ภาษาอังกฤษ โดย Visser (2008) กล่าวว่า ความสำเร็จในการเรียนภาษา ไม่ได้เพียงเกิดจากความฉลาดหรือความเชี่ยวชาญเท่านั้น แต่เกิดจากทัศนคติที่ดีต่อการเรียนรู้ เช่นเดียวกับงานวิจัยของ Gardner (1975) ซึ่งกล่าวว่าทัศนคติเป็นปัจจัยสำคัญในการเรียนรู้ภาษาต่างประเทศเพื่อให้ประสบความสำเร็จ เช่นเดียวกับการเรียนรู้สาขาอื่นๆ ซึ่งต้องเริ่มจากการมีทัศนคติที่ดีต่อเป้าหมาย เพื่อนำไปสู่กระบวนการเรียนรู้ที่มีประสิทธิภาพ ส่งผลให้ พฤติกรรมที่ดีของผู้เรียนด้วย Gardner & Lambert (1972) ยังได้เสนอว่า ทักษะในการเรียนภาษาต่างประเทศสามารถแบ่งออกเป็น 2 ประเภท ได้แก่ 1) ทักษะเชิงบูรณาการ เป็นความรู้สึกชื่นชอบต่อภาษา และความต้องการสร้างความคุ้นเคยเพื่อเป็นส่วนหนึ่งของสังคมที่ใช้ภาษานั้น และ 2) ทักษะเชิงใช้ประโยชน์ เป็นความต้องการใช้ภาษาเพื่อวัตถุประสงค์ต่อตนเองในการพัฒนาด้านต่างๆ เช่น ใช้ในการทำงาน เป็นเครื่องมือในการค้นคว้าหาข้อมูล สร้างความร่วมมือ เป็นต้น นอกจากนี้ สุดฤทัย มุขวงศา (2533) พบว่า ปัจจัยที่ทำให้เกิดทัศนคติเชิงบวกหรือเชิงลบต่อการเรียน ได้แก่ ครูผู้สอน การศึกษา การเรียนการสอน และเพื่อน สอดคล้องกับการศึกษาของ Mantle-Bromley (1995) ที่กล่าวว่าทัศนคติเชิงบวกต่อ องค์ประกอบต่างๆ ของภาษา สัมพันธ์กับความสำเร็จในการศึกษาภาษานั้น ดังนั้น หากผู้เรียนมีทัศนคติที่ดีต่อภาษาอังกฤษ ทำให้สนใจและอยากเรียนรู้ ในขณะที่ผู้เรียนมักไม่สนใจเรียน หากมีทัศนคติที่ไม่ดีต่อภาษาอังกฤษ

นอกจากทัศนคติแล้ว แรงจูงใจก็เป็นปัจจัยสำคัญที่นำไปสู่ความสำเร็จในการเรียนภาษาอังกฤษ โดย Spraat et al. (2005) อธิบายว่า แรงจูงใจมีอิทธิพลต่อความสำเร็จหรือความล้มเหลวในการเรียนภาษาอย่างมีนัยสำคัญ สอดคล้องกับ Broussard & Garrison (2004) ที่กล่าวว่า ผู้เรียนที่มีระดับแรงจูงใจในการเรียนภาษาอังกฤษสูง จะมีผลสัมฤทธิ์ทางการเรียน สูงเช่นกัน โดยแรงจูงใจสามารถแบ่งออกเป็น 2 ประเภท ได้แก่ แรงจูงใจภายใน (Intrinsic motivation) ซึ่งเกิดจากการผลักดันภายในของบุคคล โดยทัศนคติ ความเชื่อ ความสนใจที่มีต่อสิ่งหนึ่ง โดยไม่ได้เกิดจากสิ่งเร้าภายนอก และแรงจูงใจ ภายนอก (Extrinsic Motivation) ซึ่งเกิดจากการกระตุ้นจากสภาพแวดล้อมภายนอก เช่น คะแนน รางวัล หรือปริญญาบัตร)

2.2 พฤติกรรมที่ส่งผลต่อการเรียนภาษาอังกฤษ

Intaraprasert (2000) ได้กล่าวถึง พฤติกรรมในการเรียนรู้ภาษาอังกฤษที่พึงประสงค์ของผู้เรียนเพื่อให้บรรลุเป้าหมายในการเรียน ทั้งพฤติกรรมในห้องเรียน เช่น การเตรียมพร้อมก่อนเข้าเรียน ตั้งใจฟังผู้สอน มีส่วนร่วมในกิจกรรม มีสมาธิในการเรียน พยายามทบทวนบทเรียน และพฤติกรรมนอกห้องเรียน เช่น การเรียนรู้คำศัพท์ใหม่ด้วยตนเอง โดยจดศัพท์จากการอ่าน ฟัง ดูสื่อภาษาอังกฤษในชีวิตประจำวัน โดยจากการศึกษาของ Wu (2008) ยังพบว่า ผู้เรียนที่มีความสามารถทางภาษาสูง มักมีพฤติกรรมการเรียนที่ดีกว่าผู้เรียนที่มีความสามารถทางภาษาดำ เนื่องจากมีการวางแผนการเรียน และการพูดคุยกับเจ้าของภาษาเสมอ มีงานวิจัยในต่างประเทศเกี่ยวกับทัศนคติ แรงจูงใจ และพฤติกรรมการเรียนภาษาอังกฤษอีกมาก ยกตัวอย่างเช่น Chen & Huang (2003) พบว่านักเรียนที่มีทักษะด้านภาษาสูงมีพฤติกรรมและกลวิธีในการเรียนรู้มากและบ่อยกว่านักเรียนที่มีทักษะด้านภาษาดำกว่า สอดคล้องกับผลวิจัยของ Wu (2008) ที่พบว่า นักเรียนในที่มีทักษะด้านภาษาสูง จากกลุ่มตัวอย่างประเทศไต้หวัน จะมีกลวิธีในการเรียน เช่น มีการวางแผน การติดต่อสื่อสาร และการเรียนรู้วัฒนธรรมภาษามากกว่า นอกจากนี้ Broussard & Garrison (2004) ศึกษาความสัมพันธ์ระหว่างแรงจูงใจและผลสัมฤทธิ์การเรียน ภาษาอังกฤษของกลุ่มตัวอย่างนักเรียนชั้นประถมศึกษาปีที่ 1 และ 3 และพบว่าแรงจูงใจในการเรียนมีผลกระทบบวกต่อผลสัมฤทธิ์ทางการเรียนภาษา และ Pedtrides (2006) พบว่าเด็กที่ชอบเรียนภาษาอังกฤษมักมีทัศนคติเชิงบวก คือเห็นว่าเป็นเรื่องสนุก น่าสนใจ และมีผลการเรียนที่ดีกว่าเด็กกลุ่มอื่นๆ จากการทบทวนวรรณกรรม กล่าวได้ว่าทัศนคติ แรงจูงใจ และพฤติกรรมที่ดี มีผลกระทบเชิงบวกต่อความสำเร็จในการเรียนรู้ภาษาอังกฤษ

2.3 แนวคิดการทำเหมืองข้อมูล (Data Mining)

2.3.1 ความหมายของเทคนิคเหมืองข้อมูล (Data Mining)

ปัจจุบันนักสถิติมีความคิดเห็นว่าเหมืองข้อมูล คือ การสร้างแบบจำลองทางสถิติ โดยการแจกแจงพื้นฐานที่มาจาก การดึงข้อมูลที่มองเห็นได้ และบางคนถือว่าการทำเหมืองข้อมูล (Data Mining) มีความคล้ายคลึงกับการเรียนรู้ของเครื่อง (Machine Learning) เนื่องจากเทคนิคเหมืองข้อมูลบางส่วนมีการใช้อัลกอริทึมจากการเรียนรู้ของเครื่อง เช่น วิธี Bayes Nets, วิธี Support-Vector Machines, วิธี Decision Trees และวิธีอื่นๆ อีกมากมาย (Leskovec, Rajaraman, & Ullman, 2010) อย่างไรก็ตาม มีผู้ให้ความหมายการทำเหมืองข้อมูล (Data Mining) ไว้อย่างจำนวนมาก ยกตัวอย่างเช่น

Jeffrey (2013) ได้อธิบายการทำเหมืองข้อมูล คือ การใช้อัลกอริทึมและคอมพิวเตอร์ในการค้นหารูปแบบใหม่ที่น่าสนใจภายในข้อมูล ประกอบด้วย 4 กระบวนการ ได้แก่ 1) การเตรียมข้อมูล ขั้นตอนที่ใช้เวลามากที่สุด เนื่องจากต้องตรวจสอบให้แน่ใจว่าข้อมูลถูกจัดระเบียบอย่างถูกต้อง 2) การวิเคราะห์สำรวจข้อมูล ขั้นตอนทำความเข้าใจกับข้อมูล โดยใช้ฮิสโตแกรม (Histogram) และเครื่องมือสร้างแผนภาพอื่นๆ (Visualization) 3) การพัฒนาแบบจำลอง ขั้นตอนการเลือกเทคนิคเหมืองข้อมูลที่เหมาะสมที่สุด ซึ่งขึ้นอยู่กับโครงสร้างของชุดข้อมูลนั้นๆ และ 4) การตีความผลลัพธ์ ขั้นตอนสำคัญที่ต้องทำความเข้าใจถึงสิ่งที่อัลกอริทึมเหมืองข้อมูลสร้างขึ้น นอกจากนี้ Springer Link (2013) ก็ได้ให้ความหมายการทำเหมืองข้อมูล เป็นกระบวนการที่ใช้วิธีการคำนวณกับข้อมูลขนาดใหญ่ ในการดึงข้อมูลที่ซับซ้อนและสำคัญ โดยจุดประสงค์เพื่อค้นหา รูปแบบภายในข้อมูล สำรวจชุดข้อมูลขนาดใหญ่ สร้างแบบจำลองที่อธิบายคุณสมบัติที่สำคัญของข้อมูล และคาดการณ์ แนวโน้มตามข้อมูลนั้น ซึ่งมีความคล้ายคลึงกับ Daniel และ Larose (2014) ที่อธิบายการทำเหมืองข้อมูล คือ กระบวนการ ค้นหารูปแบบที่ซ่อนอยู่ และแนวโน้มที่เป็นประโยชน์ในชุดข้อมูลขนาดใหญ่ โดยใช้การจดจำรูปแบบทางสถิติและคณิตศาสตร์ และยังมีนักวิชาการ นักวิจัยคนไทยให้ความหมายการทำเหมืองข้อมูลอีกมากมาย เช่น ชนวัฒน์ ศรีสอาน (2551) ได้ให้ความหมายของการทำเหมืองข้อมูล หมายถึง กระบวนการสกัดความรู้ที่น่าสนใจจากข้อมูลปริมาณมาก โดยเป็นความรู้ที่ไม่ได้ปรากฏเห็นเด่นชัดหรือบ่งบอกเป็นนัย ไม่ทราบล่วงหน้าว่าข้อมูลเหล่านั้นมีศักยภาพที่จะนำไปใช้ ประโยชน์ และการทำเหมืองข้อมูลยังเป็นการค้นหารูปแบบและความสัมพันธ์ทั้งหมดที่มีอยู่ภายในฐานข้อมูลด้วย และ วันวิสาข์ ชนะประเสริฐ (2559) ให้ความหมายของการทำเหมืองข้อมูลเป็นการค้นหาข้อมูลความรู้ (Knowledge Discovery) จากแหล่งฐานข้อมูลขนาดใหญ่ที่มีอยู่ โดยนำมาวิเคราะห์ด้วยวิธีการทางสถิติทางคณิตศาสตร์ ร่วมกับการประมวลผลทางคอมพิวเตอร์ (AI) ที่มุ่งเน้นกระบวนการหารูปแบบแนวโน้มหาความสัมพันธ์ สร้างสมมติฐานของกฎ และการพยากรณ์ โดยนำผลสรุปที่ได้ค้นพบความรู้ที่ซ่อนอยู่ออกมานำเสนอให้ปรากฏมุมมองที่ชัดเจน ซึ่งนับเป็นความรู้ที่สามารถใช้ประโยชน์ ในด้านต่างๆ และสร้างคุณค่าให้กับข้อมูลให้มีประสิทธิภาพมากที่สุด

จากความหมายข้างต้นสามารถสรุปได้ว่า การทำเหมืองข้อมูลเป็นกระบวนการทางสถิติร่วมกับการวิเคราะห์ข้อมูล ด้วยคอมพิวเตอร์ ที่มีวัตถุประสงค์หลักเพื่อสกัดข้อมูล ค้นหารูปแบบหรือความสัมพันธ์เชิงลึกที่ซ่อนอยู่ภายในชุดข้อมูล รวมถึงสร้างแบบจำลองสำหรับใช้คาดการณ์แนวโน้มสิ่งที่จะเกิดขึ้นจากชุดข้อมูลนั้น นอกจากนี้ เทคนิคเหมืองข้อมูล สามารถใช้คู่กับอัลกอริทึมการเรียนรู้ของเครื่องเพื่อวิเคราะห์ชุดข้อมูลและสร้างการคาดการณ์ชุดข้อมูลใหม่ได้อีกด้วย

2.3.2 ประเภทของการทำเหมืองข้อมูล

เทคนิคเหมืองข้อมูล เป็นการวิเคราะห์ข้อมูลเพื่อนำไปใช้ประโยชน์ ตามวัตถุประสงค์ต่างๆ โดย (สุรพงศ์, 2018) ได้แบ่งประเภทของการทำเหมืองข้อมูล ออกเป็น 2 ลักษณะ ดังนี้

2.3.2.1 การทำเหมืองข้อมูลเชิงทำนาย (Predictive Data Mining) เป็นการเรียนรู้ที่ได้รับการแนะนำ (Supervised Learning) ที่ใช้ชุดข้อมูลการสอน (Training Data) สร้างตัวแบบที่มีการจำแนกประเภทแล้ว นำมาใช้ในการหาค่าหรือทำนาย ค่าของตัวแปรเป้าหมาย (Target Variable) ยกตัวอย่างเช่น

- 1) การจำแนกประเภทข้อมูล (Classification) เป็นการสร้างตัวแบบจำแนกข้อมูลจากชุดข้อมูลเรียนรู้ ที่มีการจำแนกประเภทแล้ว มาใช้จำแนกข้อมูลใหม่ที่ยังไม่ทราบประเภท เช่น การจำแนกคนไข้โรคต่างๆ โดยพิจารณาจากอาการที่เกิดขึ้น การจำแนกโอกาสที่ลูกค้าจะเลิกใช้บริการต่างๆ เป็นต้น
- 2) การคาดคะเน (Estimation) เป็นการสร้างตัวแบบ เพื่อใช้คาดคะเนค่าของตัวแปรเป้าหมาย โดยดูจากค่าตัวแปรอื่นๆ ซึ่งข้อแตกต่างจากการจำแนกข้อมูล คือ ตัวแปรที่ถูกคาดคะเนมีลักษณะเป็นตัวเลข (Numerical Data)

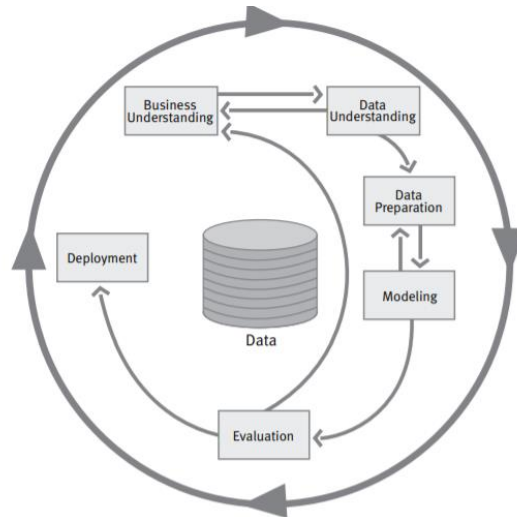
2.3.2.2 การทำเหมืองข้อมูลเชิงพรรณนา (Descriptive Data Mining) จัดเป็นการเรียนรู้ที่ไม่ต้องการการแนะนำ (Unsupervised Learning) โดยใช้วิธีการค้นหารูปแบบ (Pattern) หรือลักษณะของข้อมูลภายในชุดข้อมูล ที่ไม่ต้องอาศัยชุดข้อมูลเรียนรู้สำหรับค้นหารูปแบบที่ต้องการ ยกตัวอย่างเช่น

- 1) การจับกลุ่มข้อมูล (Clustering) การแบ่งข้อมูลออกเป็นกลุ่ม ข้อมูลที่มีลักษณะคล้ายกันจะถูกจัดอยู่ในกลุ่มเดียวกัน และข้อมูลที่แตกต่างจะถูกแยกออกไป ซึ่งการจับกลุ่มข้อมูลนี้ต้องอาศัยมาตรวัดความเหมือนหรือความแตกต่างระหว่างข้อมูลในการจัดหมวดหมู่
- 2) การสร้างกฎความสัมพันธ์ (Association Rule) เป็นการหากฎที่บอกถึงความสัมพันธ์ระหว่างเหตุการณ์ ที่ส่วนใหญ่เกิดขึ้นพร้อมกันเสมอ โดยกฎความสัมพันธ์ที่สร้างได้จะระบุความสัมพันธ์เมื่อพบว่า เหตุการณ์หนึ่งเกิดขึ้น จะมีโอกาสสูงที่เหตุการณ์อีกอย่างหนึ่งจะเกิดขึ้นด้วย
- 3) การค้นหาข้อมูลที่มีความผิดปกติ (Anomaly Detection) เป็นวิธีสำหรับใช้ค้นหารูปแบบที่มีลักษณะผิดปกติหรือมีความแตกต่างจากข้อมูลส่วนใหญ่ (Outlier Data) ซึ่งข้อมูลผิดปกติเหล่านี้จะมีจำนวนน้อยที่รวมอยู่ในข้อมูลจำนวนมาก ซึ่งเราสามารถใช้ประโยชน์ในงานบางอย่างที่ต้องการตรวจสอบข้อมูลที่มีความผิดปกติได้

จะเห็นได้ว่าการทำงานเหมืองข้อมูลแต่ละประเภทมีความเหมาะสมกับข้อมูลที่แตกต่างกันออกไป การเลือกเทคนิคต่างๆ ใช้งานจำเป็นต้องเลือกให้เหมาะสมตามวัตถุประสงค์และประเภทของข้อมูล เพื่อให้ได้ผลลัพธ์ตามต้องการและมีประสิทธิภาพ

2.3.3 กระบวนการทำเหมืองข้อมูล

ปัจจุบันหลายภาคส่วนไม่ว่าจะเป็นภาคอุตสาหกรรม การบริการ การตลาด หรือการศึกษา ฯลฯ ได้หันมาสนใจ และใช้ประโยชน์จากการทำเหมืองข้อมูลเพิ่มมากขึ้น ดังนั้น กระบวนการที่เป็นมาตรฐานเป็นสิ่งจำเป็นที่จะช่วยให้การทำงานเป็นไปตามขั้นตอน ข้อมูลมีประสิทธิภาพ และได้รับผลลัพธ์ที่ถูกต้อง โดยกระบวนการทำเหมืองข้อมูลมีหลายรูปแบบ สามารถเลือกนำไปใช้ปฏิบัติ ซึ่งกระบวนการหนึ่งที่ได้รับการยอมรับ คือ CRISP-DM (Cross-Industry Standard Process for Data Mining) เป็นหนึ่งในกระบวนการที่เป็นมาตรฐานของการทำเหมืองข้อมูล พัฒนาขึ้นในปี ค.ศ. 1996 โดยนักวิเคราะห์ที่เป็นตัวแทนของ Daimler Chrysler, SPSS และ NCR ประกอบด้วย 6 ขั้นตอนแสดงในภาพที่ 2.1 และอธิบายแต่ละขั้นตอน (สุรพงศ์, 2561) ดังนี้



ภาพที่ 2.1 มาตรฐานการทำเหมืองข้อมูลตามขั้นตอน CRISP-DM

ที่มา: (TheCRISP-DMConsortium, 2000)

2.3.3.1 การทำความเข้าใจธุรกิจ (Business/Research Understanding Phase) ขั้นตอนแรกที่เน้นทำความเข้าใจเชิงธุรกิจ โดยเริ่มจากการกำหนดเป้าหมายและวัตถุประสงค์ของการทำเหมืองข้อมูล มีการวางแผนการดำเนินงาน เช่น แหล่งข้อมูลที่จะนำมาใช้ในการวิเคราะห์ ประเภทข้อมูล และความต้องการที่จะนำผลลัพธ์ไปใช้ประโยชน์ เป็นต้น เพื่อให้การดำเนินงานตรงตามวัตถุประสงค์

3.2.3.2 การทำความเข้าใจข้อมูล (Data Understanding Phase) ขั้นตอนของการรวบรวมข้อมูล การคัดเลือกข้อมูลต่างๆ ที่จะนำมาวิเคราะห์ รวมทั้งประเมินคุณภาพและลักษณะของข้อมูลเพื่อตรวจสอบความสมบูรณ์และความถูกต้องก่อนทำการวิเคราะห์ข้อมูล ในกรณีพบเป้าหมายที่จะวิเคราะห์เพิ่มเติมในระหว่างขั้นตอนนี้ก็สามารถย้อนกลับไปขั้นตอนที่ 1 เพื่อแก้ไขเป้าหมายและวัตถุประสงค์ของการทำเหมืองข้อมูลใหม่ได้

3.2.3.3 การเตรียมข้อมูล (Data Preparation Phase) ขั้นตอนการคัดเลือกข้อมูลและกำหนดตัวแปรที่จะนำไปวิเคราะห์ในเครื่องมือสร้างแบบจำลอง ในขั้นตอนนี้ต้องใช้เวลาและอาจต้องมีการดำเนินการหลายครั้ง เนื่องจากอาจมีการเปลี่ยนแปลงข้อมูลให้อยู่ในรูปแบบที่นำมาใช้วิเคราะห์ได้ และ/หรือ จัดการกับข้อมูลที่มีลักษณะผิดปกติ ที่จะส่งผลให้การวิเคราะห์เกิดข้อผิดพลาด

3.2.3.4 การสร้างโมเดล (Modeling Phase) ขั้นตอนของการวิเคราะห์ข้อมูลเพื่อให้ได้ผลลัพธ์ที่ดีที่สุด โดยเลือกวิธีการวิเคราะห์ข้อมูลและอัลกอริทึมที่เหมาะสมนำมาสร้างแบบจำลอง ซึ่งอาจใช้เทคนิควิธีการต่างๆ ร่วมกันหลายๆ เทคนิคได้ เนื่องจากบางครั้งอาจเกิดปัญหาในระหว่างการวิเคราะห์ เช่น บางเทคนิคอาจมีข้อจำกัดเฉพาะเกี่ยวกับรูปแบบของข้อมูล หรือบางข้อมูลอาจเหมาะกับเทคนิคอีกแบบหนึ่งก็เป็นได้ ดังนั้น บางกรณีสามารถย้อนกลับไปขั้นตอนที่ 3 เพื่อเตรียมและคัดเลือกตัวอย่างข้อมูล และเพิ่มเติมตัวแปรในการวิเคราะห์ได้

3.2.3.5 การประเมินประสิทธิภาพของโมเดล (Evaluation Phase) เป็นขั้นตอนประเมินผลลัพธ์จากการทำเหมืองข้อมูล รวมถึงเปรียบเทียบประสิทธิภาพแบบจำลองที่ได้จากขั้นตอนที่ 4 ว่าสามารถให้ผลลัพธ์ที่จะนำไปใช้งานตามวัตถุประสงค์ได้มากน้อยเพียงใด และในขั้นตอนนี้สามารถย้อนกลับไปขั้นตอนก่อนหน้า เพื่อสร้างรูปแบบใหม่ที่มีความสมบูรณ์มากยิ่งขึ้น

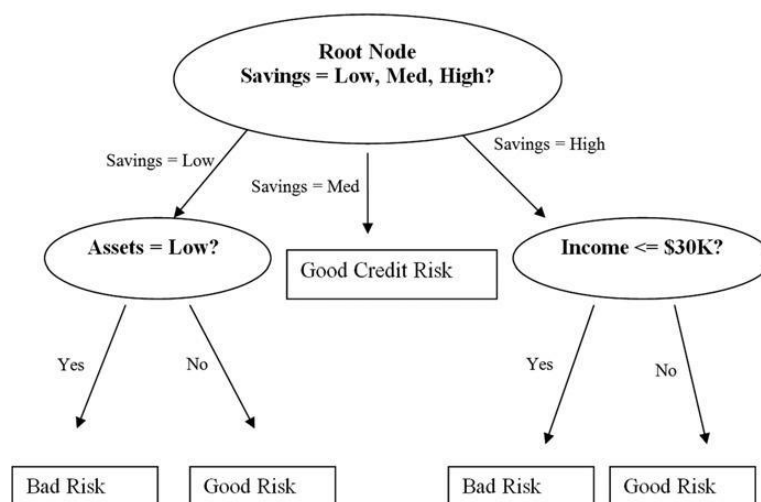
3.2.3.6 การนำโมเดลไปใช้งานจริง (Deployment) ขั้นตอนสุดท้ายหลังจากได้ผลลัพธ์ที่สมบูรณ์และนำไปใช้งานจริง โดยผลลัพธ์จากการใช้งานสามารถนำมาใช้ในการวางแผนเพื่อเริ่มการโครงการทำเหมืองข้อมูลต่อไปในอนาคตตามกระบวนการ 6 ขั้นตอนนี้

จะเห็นว่า ขั้นตอนการทำเหมืองข้อมูลตาม CRISP-DM มีลักษณะเป็นวงรอบ สามารถปรับเปลี่ยน(Adaptive) การดำเนินการได้เสมอ โดยย้อนกลับไปปรับแก้ไขในขั้นตอนที่ก่อนหน้านี้โดยไม่ต้องจบรอบเพื่อให้ได้ผลลัพธ์ที่ดีขึ้น และจากนั้นก็ดำเนินตามขั้นตอนลำดับต่อไป

2.3.4 เทคนิคเหมืองข้อมูลใช้ในการศึกษา

2.3.4.1 เทคนิค Decision tree

Decision Tree เป็นวิธีการจำแนกประเภทข้อมูล (Classification) ประเภทหนึ่งที่แบบจำลองมีลักษณะคล้ายต้นไม้ แต่ละโหนดของ Decision Tree จะประกอบด้วยเงื่อนไขที่ใช้ตัวแปรใดตัวแปรหนึ่งของข้อมูลทำการตัดสินใจเลือกโหนดลูก (Child Node) ที่จะเป็นโหนดการตัดสินใจต่อไป การตัดสินใจจะเริ่มจากโหนดราก (Root Node) ของต้นไม้ไล่ลำดับลงไปยังโหนดลูกจนถึงจุดได้สิ้นสุดในโหนดใบ (Leaf Node) ซึ่งเป็นโหนดที่จะบอกประเภทของข้อมูลนั้น (สุรพงศ์, 2561) โดยแต่ละโหนด (Node) จะเชื่อมต่อกันเรียกว่า กิ่ง (Link) ซึ่งข้อมูลที่ต้องการจำแนกจะวิ่งไปตามเงื่อนไขของแต่ละโหนดเรื่อยๆ จนได้คำตอบ ทั้งนี้ โหนดราก (Root node) จะถูกเลือกจากโหนดที่มีคุณสมบัติที่ดีที่สุด โดยพิจารณาจากค่าความหลากหลายของประเภทข้อมูล เป้าหมายของการสร้าง Decision Tree คือ เพื่อให้ได้ข้อมูลที่ถูกต้องมากที่สุดในการจำแนกประเภทของทั้งส่วนข้อมูลเรียนรู้ (Training Data) และข้อมูลทดสอบ (Test Data) (Larose และ Larose, 2014) ตัวอย่างการใช้เทคนิค Decision Tree สำหรับการจำแนกประเภทข้อมูล แสดงให้เห็นในภาพที่ 2.9



ภาพที่ 2.2 ตัวอย่าง Decision Tree สำหรับการจำแนกประเภทข้อมูล

ที่มา (Larose และ Larose, 2014)

จากตัวอย่างข้างต้นจะเห็นว่า การใช้เทคนิค Decision Tree จำแนกข้อมูลมีจุดมุ่งหมายที่ต้องการจัดประเภทผู้ที่มีโอกาสจะเป็นลูกค้าในอนาคต โดยกำหนดตัวแปรเป้าหมาย (Target Variable) เป็นประเภทความเสี่ยงด้านเครดิต คือ ลูกค้าที่มีเครดิตดี และลูกค้าที่มีความเสี่ยงจะเป็นผู้ไม่เครดิตไม่ดี กำหนดตัวแปรทำนาย (Predictor Variables) ได้แก่ การออมเงิน (ต่ำ กลาง และสูง) สินทรัพย์ (ต่ำ หรือไม่ต่ำ) และรายได้ ($\leq \$30,000$ หรือ $> \$30,000$) ในที่นี้ เลือกแอตทริบิวต์การออมเงินเป็นโหนดราก อย่างไรก็ตาม การเลือกตัวแปรที่จะใช้ในแต่ละโหนดในต้นไม้ ต้องมีการวัดความหลากหลายของประเภทข้อมูล หรือเรียกว่า Impurity ภายในชุดข้อมูล ซึ่งโหนดที่มี Impurity น้อยที่สุดจะถูกเลือกเป็นโหนดราก โดย (สุรพงศ์, 2561) อธิบายตัวชี้วัดสำหรับวัดค่าความหลากหลายของประเภทข้อมูล ไว้ดังนี้

1) **Entropy** เป็นการวัดความไม่แน่นอนตามทฤษฎีสารสนเทศ (Information Theory) กล่าวคือ หากข้อมูลมีหลายประเภทและแต่ละประเภทมีสัดส่วนพอๆ กัน ความน่าจะเป็นที่จะทำนายสารสนเทศของข้อมูลใดข้อมูล

หนึ่งทำได้น้อยเพราะมีความไม่แน่นอนสูง ในทางตรงกันข้าม หากข้อมูลเป็นประเภทเดียว ความน่าจะเป็นที่จะทำนายสารสนเทศของข้อมูลใดข้อมูลหนึ่งทำได้มาก เนื่องจากข้อมูลเป็นประเภทเดียวกันทั้งหมดและค่อนข้างมีความแน่นอน สามารถวัดด้วยค่า Entropy ดังสมการที่ 1

$$\text{Entropy} = - \sum_j^n (j) \log(P(j))$$

โดยที่ $P(j)$ คือ สัดส่วนของข้อมูลที่มีประเภท
 n เป็นจำนวนประเภทข้อมูล

จากสมการของ Entropy ข้างต้น เป็นการคำนวณค่า Entropy ซึ่งจะอยู่ระหว่าง 0 กับ 1 โดยค่า 0 เกิดขึ้นเมื่อชุดข้อมูลมีข้อมูลประเภทเดียวกันทั้งหมด หรือ มี Impurity น้อยที่สุด และจะมีค่าสูงขึ้นถ้าชุดข้อมูลมีความหลากหลายของประเภทข้อมูล หรือ มี Impurity เพิ่มขึ้น ดังนั้น ค่า Entropy จะมีค่าสูงขึ้นเท่ากับ $\log n$ เมื่อชุดข้อมูลประกอบด้วยข้อมูลแต่ละประเภทมีสัดส่วนเท่าๆ กัน คือ $\frac{1}{n}$

2) **Gini Index** มีลักษณะคล้าย Entropy ที่ใช้การคำนวณสัดส่วนของชุดข้อมูลแต่ละประเภท โดยสามารถวัดค่า Gini Index ดังสมการที่ 2

$$\text{Gini} = 1 - \sum_{j=1}^n p^2(j)$$

โดยที่ $P(j)$ คือ สัดส่วนของข้อมูลประเภท j
 n เป็นจำนวนประเภทข้อมูล

ในกรณีที่ชุดข้อมูลมีข้อมูลประเภทเดียวกันทั้งหมดจะมีค่า Impurity ต่ำสุดที่ 0 และค่า Impurity จะสูงเพิ่มขึ้นเรื่อยๆ เมื่อประเภทข้อมูลมีความหลากหลายมากขึ้น และเมื่อข้อมูลแต่ละประเภทมีสัดส่วนเท่ากัน คือ $P(j) = \frac{1}{n}$ มีค่า Impurity สูงสุดเท่ากับ $1 - \frac{1}{n^2} * n = 1 - \frac{1}{n}$

3) **Misclassification Error** เป็นการทำนายประเภทข้อมูลโดยกำหนดให้ประเภทข้อมูลที่มีสัดส่วนมากที่สุดเป็นประเภทข้อมูลหลัก (Majority Class) ซึ่งหากกำหนดให้ทุกตัวในชุดข้อมูลเป็นประเภทข้อมูลหลักทั้งหมด จะทำให้การทำนายข้อมูลประเภทอื่นจะไม่ถูกต้อง โดยการคำนวณอัตราความผิดพลาดของการทำนายประเภทข้อมูลใดข้อมูลหนึ่งในชุดข้อมูล มีดังสมการที่ 3

$$\text{Misclassification Error} = 1 - \text{Max}_{j=1}^n (P(j))$$

โดยที่ $P(j)$ คือ สัดส่วนของประเภทข้อมูล j
 n คือ จำนวนประเภทข้อมูล

เมื่อข้อมูลในชุดข้อมูลเป็นประเภทเดียวกันทั้งหมด จะมีค่า Misclassification Error เท่ากับ 0 และค่า Misclassification Error จะเพิ่มขึ้นถ้าข้อมูลมีความหลากหลายมากขึ้น และเมื่อข้อมูลแต่ละประเภทมีสัดส่วนเท่ากัน คือ $\frac{1}{n}$ จะมีค่าสูงสุดเท่ากับ $1 - \frac{1}{n}$

ในการจำแนกประเภทข้อมูลนิยมใช้ Entropy กับ Gini อย่างไรก็ดี ตัววัด Entropy มีเงื่อนไขการตัดสินใจที่แบ่งข้อมูลเรียนรู้ออกเป็นหลายส่วน แต่ละส่วนมีข้อมูลจำนวนน้อย จึงใช้กับ C4.5 หรือ ID3 Algorithm ทำให้ได้ Decision Tree ในรูป N – ary Tree และส่วน Gini มีเงื่อนไขการตัดสินใจที่แบ่งข้อมูลการเรียนรู้ออกเป็น 2 ส่วนเท่านั้น แต่ละส่วนมีข้อมูลจำนวนมาก Gini จึงใช้กับ Classification and Regression Tree (CART) ทำให้ได้ Decision Tree ในรูป Binary Tree การใช้เทคนิค Decision Tree จะมีผลออกมาใช้งานในลักษณะ “ถ้า เงื่อนไขแล้วผลลัพธ์” เช่น (พยุร, 2548)

“If Income = High and Married = No THEN Risk = Poor”

“If Income = High and Married = Yes THEN Risk = Good”

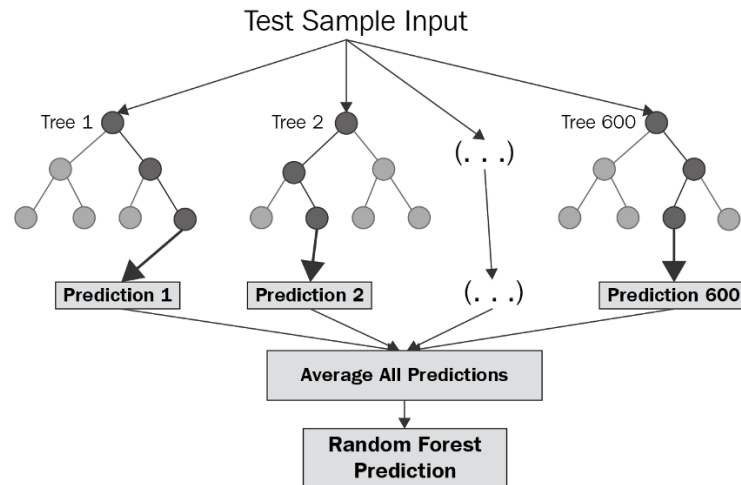
2.3.4.2 อัลกอริทึม C4.5 (J48)

อัลกอริทึม C4.5 เป็นอัลกอริทึมหนึ่งในเทคนิค Decision Tree โดยการสร้างกฎที่มีการพัฒนาต่อจากอัลกอริทึม ID3 ออกแบบโดย Quinlan (1992) เป็นกระบวนการจำแนกกลุ่มของข้อมูลที่มีลักษณะสร้างเป็นเงื่อนไข (If Then...) และนำไปใช้ประโยชน์ในการทำนายข้อมูลเป้าหมายที่ต้องการ การสร้าง Decision Tree ด้วยอัลกอริทึม C4.5 มีลักษณะการตัดสินใจในรูป Binary Tree ซึ่งมีการคำนวณ ค่า Entropy และ ค่า Gain Ratio โดย Split Info(x) ที่เป็ค่าที่ได้จากการแบ่งข้อมูลจากการทำ Normalization เป็น n subset และ Gain Ratio(x) ที่เป็นการวัดว่าการแบ่งข้อมูลโดยใช้ข้อมูลใดข้อมูลหนึ่งก่อให้เกิดประโยชน์ต่อการทำนายหรือไม่ (กัลยณภัส พนิตา และเฉลิมชัย, 2559)

2.3.4.3 เทคนิค Random Forest

Random Forest เป็นหนึ่งในเทคนิคที่ใช้การเรียนรู้ของเครื่อง (Machine learning) ที่มีคุณสมบัติจำแนกข้อมูลโครงสร้างแบบต้นไม้ เพื่อสร้างแบบจำลองและนำไปใช้ทำนายผลข้อมูลเป้าหมาย เทคนิค Random Forest ได้รับการยอมรับอย่างมากจากงานวิจัยด้านชีวการแพทย์ และเกษตรกรรม เพราะมีประสิทธิภาพในการทำนายและมีความถูกต้องสูง เนื่องจากกฎไม่ได้ทำให้เกิดความเฉพาะเจาะจง หรือ over fit เมื่อใช้ข้อมูลจำนวนมากหรือการเพิ่มจำนวนข้อมูลมากขึ้น และเมื่อทำการสุ่มหลายๆ ครั้งบนข้อมูลชุดเดียวกันจะทำให้การจำแนกประเภทข้อมูล (classifier) และตัวกดถอย (Repressor) มีความแม่นยำ อย่างไรก็ตาม อัลกอริทึม Random Forests เป็นตัวจำลองทำนายรูปแบบต้นไม้ที่แต่ละต้นจะขึ้นอยู่กับค่าของเวกเตอร์ที่สุ่มขึ้นมาอย่างอิสระต่อกันด้วยการกระจายแบบเดียวกันของต้นไม้ทั้งหมด ทำให้เมื่อต้นไม้มีจำนวนเยอะขึ้นค่าจะเข้าสู่ข้อจำกัด (Limit) และเกิดความผิดพลาดในการจำแนกได้ รวมถึงการจำแนกด้วยเทคนิค Random Forest ขึ้นอยู่กับค่าความแข็งแรง (strength) ของต้นไม้แต่ละต้นและความสัมพันธ์ (correlation) ระหว่างกัน ซึ่งจะส่งผลให้เกิดความผิดพลาดในการจำแนกได้ด้วยเช่นเดียวกัน นภาพร ศิริกุลวิริยะ (2554) อ้างอิงใน (วันวิสาข์, 2017)

การสุ่มเลือกใช้ข้อมูลและคุณลักษณะตามแบบ Decision Tree ของ Random Forest นั้นสร้างจากการนำข้อมูลไปสุ่มเลือกตัวอย่างแบบเลือกแล้วใส่กลับ (Sampling with Replacement) และนำไปสร้างเป็น Decision Tree แต่ละครั้งจะเลือกส่วนของข้อมูลไปเรียนรู้ไม่เหมือนกัน ทำให้มีตัวอย่างบางส่วนที่ไม่ถูกเลือก เรียกว่า Out-of-Bag (OOB) จะถูกนำมาใช้ในการทดสอบ กลไกนี้เรียกว่า Bagging โดยผลลัพธ์ที่ได้จากการเรียนรู้ที่เป็นอิสระต่อกันจากต้นไม้แต่ละต้นจะถูกนำมาคิดเป็นผลการโหวต ซึ่งผลโหวตที่มากที่สุดจะกลายเป็นการทำนายโดยรวมของแบบจำลอง (พัฒน์พงษ์ และ ทอจารี, 2017) แสดงภาพที่ 2.3



ภาพที่ 2.3 โครงสร้าง Random Forest
ที่มา (Chakure, 2019)

อาจกล่าวได้ว่า Random Forest เป็นเทคนิคที่พัฒนาต่อยอดจากเทคนิคต้นไม้การตัดสินใจ (Decision Tree) ที่ใช้เพียงต้นเดียวในการจำแนกประเภทข้อมูลที่ต้องการ เพื่อให้แบบจำลองมีประสิทธิภาพความแม่นยำในการทำนายมากที่สุด

2.3.5 การวัดประสิทธิภาพแบบจำลอง

ในกระบวนการพัฒนาแบบจำลอง จำเป็นต้องมีการทดสอบประสิทธิภาพแบบจำลองที่จะนำไปใช้งานด้วยว่ามีความแม่นยำและความถูกต้องมากน้อยแค่ไหน ดังนั้น กระบวนการวัดประสิทธิภาพแบบจำลองจึงถือเป็นขั้นตอนที่สำคัญและสามารถใช้วิธีการวัดประสิทธิภาพหลายๆ วิธีมาเปรียบเทียบเพื่อเลือกแบบจำลองที่ดีที่สุดไปใช้งานได้ โดย (สุรพงศ์, 2561) กล่าวว่า การประเมินประสิทธิภาพแบบจำลองการจำแนกประเภทข้อมูล (Classification) สามารถใช้เครื่องมือได้หลายแบบ โดยในการศึกษานี้ขออธิบายตัววัด 2 แบบ ดังนี้

2.3.4.1 Confusion Matrix เป็นตารางแสดงข้อมูลของจำนวนในแต่ละประเภทที่ถูกจำแนกด้วยตัวแบบโดย ตัวชี้วัดที่คำนวณได้จาก Confusion Matrix มี 4 แบบ คือ

แบบที่ 1 ค่าความแม่นยำ (Accuracy) ในการจำแนกประเภทที่ตรงกับข้อมูลที่เกิดขึ้นจริงที่ดังสมการที่ 4

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$$

โดยที่ TP คือ ทำนายเกิดขึ้นเป็นจริงและสิ่งที่เกิดขึ้นก็เป็นจริง

TN คือ ทำนายเกิดขึ้นเป็นไม่จริงและสิ่งที่เกิดขึ้นไม่จริง

FP คือ ทำนายเกิดขึ้นเป็นจริงแต่สิ่งที่เกิดขึ้นเป็น ไม่จริง

FN คือ ทำนายเกิดขึ้นเป็นไม่จริงแต่สิ่งที่เกิดขึ้นเป็นจริง

แบบที่ 2 ค่าความถูกต้อง (Precision) ในการทำนายถูกของการจำแนกประเภทข้อมูล โดยคำนวณเฉพาะสัดส่วนของข้อมูลที่ถูกต้องการทำนายเกิดขึ้นเป็นจริงและสิ่งที่เกิดขึ้นก็เป็นจริง (TP) กับ การทำนายเกิดขึ้นเป็นจริงแต่สิ่งที่เกิดขึ้นเป็น ไม่จริง (FP) ดังสมการที่ 5

$$\text{Precision} = \frac{TP}{TP+FP}$$

แบบที่ 3 ค่าความระลึก (Recall) คือ ค่าความสามารถของตัวแปรในการจำแนกประเภทข้อมูลที่อัลกอริธึมทำนายถูกต้องว่าเกิดขึ้นจริงเปรียบเทียบกับผลรวมของอัลกอริธึมทำนายถูกต้องว่าเกิดขึ้นจริงกับที่อัลกอริธึมทำนายว่าเกิดขึ้นไม่จริงแต่ในความเป็นจริงคือเกิดขึ้นจริง สมการที่ 6

$$\text{Recall} = \frac{TP}{TP+FN}$$

แบบที่ 4 ค่าความถ่วงดุล (F-Measure) เป็นตัววัดประสิทธิภาพโดยรวมคำนวณจาก Precision และ Recall และเฉลี่ยค่าเข้าด้วยกัน ดังสมการที่ 7

$$F_1 = 2 * \frac{\text{precision} * \text{recall}}{\text{Precision} + \text{recall}}$$

2.3.4.2 Cross Validation วิธีการแบ่งชุดข้อมูลสำหรับสร้างและทดสอบแบบจำลอง โดยแบ่งออกเป็น K Disjoint Subsets โดยที่แต่ละ Subsets จะมีข้อมูลไม่ซ้ำกัน จากนั้นจะเริ่มจากข้อมูลของ k-1 Subsets ในการพัฒนาแบบจำลอง และให้ส่วน 1 Subset ที่เหลือ เป็นข้อมูลทดสอบ ทำวนซ้ำเป็นจำนวน k ครั้ง ดังนั้นในการพัฒนาแบบจำลองแต่ละครั้งจะใช้ k - 1 Subset และ 1 Subset ที่เหลือที่ไม่ใช่ข้อมูลเดียวกันประเมินประสิทธิภาพของแบบจำลอง โดยจะใช้ค่าเฉลี่ยของการประเมินทั้ง k ครั้ง ซึ่งวิธีการนี้เรียกว่า k-fold cross validation จะเห็นได้ว่าแต่ละข้อมูลจะถูกใช้เป็นข้อมูลชุดทดสอบเพียง 1 ครั้ง และร่วมในการสร้างตัวแบบ k - 1 ครั้ง ในกรณีที่ข้อมูลมีจำนวนไม่มาก อาจกำหนดให้ k = n โดยที่ n คือ จำนวนข้อมูลทั้งหมด ซึ่งหมายถึง ในแต่ละครั้งจะใช้ข้อมูล n - 1 ตัวในการสร้างตัวแบบ และใช้ข้อมูลที่เหลือ 1 ตัว เป็นตัวพัฒนาแบบจำลอง (สุระสิทธิ์, 2558) โดยกระบวนการของวิธี Cross Validation ดังภาพที่ 2.4



ภาพที่ 2.4 กระบวนการของวิธี Cross Validation
ที่มา (Gunjal, 2020)

2.4 งานวิจัยที่เกี่ยวข้อง

2.4.1 R. R. Kabra และ R. S. Bichka (2011) ได้วิจัย “Performance Prediction of Engineering Students using Decision Trees” การใช้เทคนิคต้นไม้วิเคราะห์ข้อมูลผลการเรียนในอดีตของนักศึกษาคณะวิศวกรรมศาสตร์ และสร้างแบบจำลองเพื่อทำนายนักเรียนที่จะเข้าศึกษาในคณะวิศวกรรมศาสตร์ที่มีแนวโน้มจะเรียนไม่ผ่าน โดยใช้ข้อมูลนักศึกษาจากวิทยาลัยวิศวกรรมศาสตร์และการจัดการข้อมูลในประเทศอินเดีย จำนวน 346 คนที่เข้าศึกษาในชั้นปีที่ 1 คณะวิศวกรรมศาสตร์ ในปี 2009-2010 และ ปี 2010-2011 เก็บรวบรวมผ่านแบบฟอร์มการลงทะเบียนที่กรอกช่วงการรับเข้าศึกษา ประกอบด้วย ข้อมูลส่วนตัวทั่วไป เช่น เพศ เขตที่อยู่อาศัย อาชีพของบิดา อาชีพของมารดา เป็นต้น และข้อมูลผลการเรียนของนักศึกษาในอดีต เช่น หลักสูตรสาขาที่เรียน ผลการเรียนวิชาฟิสิกส์ เคมี และคณิตศาสตร์ เปอร์เซ็นต์ของคะแนนในวิชาคณิตศาสตร์ วิทยาศาสตร์ ประเภทการรับเข้าเรียน และข้อมูลผลการเรียนปีหนึ่งในระบบการศึกษา เป็นต้น งานวิจัยนี้ใช้โปรแกรม Weka ในการประมวลผลข้อมูลและการสร้างแบบจำลองการทำนาย เลือกใช้อัลกอริทึม J48 ในการประมวลผล ใช้ตัวแปรเป้าหมายคือ ข้อมูลผลการเรียนของชั้นปีที่ 1 โดยแบ่งการจำแนกประเภทเป็นผ่าน ไม่ผ่าน และ ATKT แบบจำลองสร้างและทดสอบโดยวิธี Cross Validation แบบ 10- fold cross validation และวัดประสิทธิภาพแบบจำลองด้วย Confusion Matrix ผลการวิจัยพบว่า แบบจำลองนี้ให้ค่าความแม่นยำที่ 60.46% โดยที่กลุ่มตัวอย่าง 209 ระเบียบถูกจัดประเภทอย่างถูกต้องจากกลุ่มตัวอย่างทั้งหมด 346 ระเบียบ และคุณลักษณะที่สำคัญในการทำนายผลการเรียนของนักเรียนที่จะเข้าศึกษาต่อในคณะวิศวกรรมศาสตร์ คือ คุณลักษณะทางสังคม เช่น อาชีพของผู้ปกครอง ที่อยู่อาศัย ในขณะเดียวกัน ปัจจัยเพศ และข้อมูลในระดับมัธยมศึกษามีความเกี่ยวข้องน้อย และงานวิจัยได้อธิบายเพิ่มเติมว่า ผลการศึกษาก่อนเข้ามีความสมเหตุสมผล กล่าวคือนักศึกษาที่มีแนวโน้มมีผลการศึกษาดี ถ้าถ้าพ่อแม่ของเขามีการศึกษาที่ดี มีการทำงานดี และ/หรือมีรายได้ที่ดี แต่อย่างไรก็ตามอาจมีข้อยกเว้นบางประการตามบริบทที่แตกต่างกัน

2.4.2 Paulo Cortez และ Alice Silva (2008) ได้วิจัย “Using data mining to predict secondary school student performance” เพื่อศึกษาและทำนายปัจจัยที่ส่งผลต่อความสำเร็จและล้มเหลวของนักเรียนระดับมัธยมศึกษา ประเทศโปรตุเกส รวมถึงความเป็นไปได้ที่จะทำนายผลการเรียนของนักเรียนด้วยเทคนิคการทำเหมืองข้อมูล 4 แบบ (Decision Trees, Random Forest, Neural Networks, Support Vector Machines) กลุ่มตัวอย่างในการวิจัยนี้ใช้ข้อมูลจากนักเรียนสองคลาสเรียน คือ คณิตศาสตร์และภาษาโปรตุเกส จำนวน 1044 ระเบียบ โดยรวบรวมข้อมูลจากรายงานของโรงเรียนและแบบสอบถามปลายปิด ประกอบด้วย เกรดของนักเรียน ข้อมูลส่วนตัวทั่วไป ข้อมูลด้านสังคมและอารมณ์ที่เกี่ยวข้อง จำนวน 33 ลักษณะ ใช้การนำเข้าข้อมูลทั้งแบบมีเกรดและไม่มีเกรดในการประมวลผลด้วยเป้าหมายการทำเหมืองข้อมูล 3 วิธี 1) การจำแนกประเภทแบบเลขฐานสอง (binary classification) ผ่าน/ไม่ผ่าน 2) การจำแนกประเภทแบบ 5 ระดับ (ระดับ 1 ดีมาก ถึง ระดับ 5 ไม่เพียงพอ) และ 3) การถดถอย (Regression) ผลลัพธ์เป็นตัวเลขที่อยู่ในช่วงระหว่างศูนย์ (0%) ถึง 20 (100%) ในการประเมินประสิทธิภาพแบบจำลองสำหรับนำไปใช้เชิงทำนาย ใช้วิธี Cross Validation แบบ 10- fold cross-validation จำนวน 20 รอบจากการจำลองทั้งหมด 200 ระเบียบ ข้อมูลถูกสุ่มแบ่งเป็น 10 ชุดย่อย (Subset) ที่มีขนาดเท่ากัน โดยชุดย่อยหนึ่งชุดได้รับการทดสอบด้วย 10% ของข้อมูล และข้อมูลที่เหลือที่ใช้เป็นข้อมูลเรียนรู้ โดยการวิจัยนี้ใช้โปรแกรม Rminer ผลการศึกษพบว่า เทคนิค Random Forest และ Decision Tree เป็นตัวเลือกที่ดีที่สุด และนำมาใช้ในการอภิปรายผล โดย Random Forest แสดงผลลัพธ์ที่เกี่ยวข้องมากที่สุด 5 ลำดับ คือ เกรดของนักเรียน จำนวนความล้มเหลวในอดีต จำนวนการขาดเรียน การสนับสนุนจากครอบครัว และความต้องการศึกษาในระดับที่สูงขึ้น อย่างไรก็ตาม ผลกระทบโดยรวมของ 5 ปัจจัยนี้มีค่าระหว่าง 43% ถึง 77% ซึ่งบ่งชี้ว่ามีข้อมูลนำเข้าไม่เกี่ยวข้องจำนวนมาก รวมถึงผลการวิเคราะห์ทั้ง 3 วิธีมีความแตกต่างกันเล็กน้อย เช่น เกรดในช่วงชั้นที่ 2 มีความสำคัญเมื่อใช้จำแนกประเภทแบบเลขฐานสอง ในขณะที่เกรดช่วงแรกมีความสำคัญอย่างมากกับการจำแนกประเภทแบบ 5 ระดับ นอกจากนี้ จำนวนความล้มเหลวในอดีต ยังเป็นปัจจัยที่สำคัญที่สุดเมื่อไม่มีผลคะแนนของนักเรียน ในส่วน Decision Tree นำผลดีที่สุดที่ได้จากการวิเคราะห์ที่มาสสร้างเป็นกฎ โดย

กำหนดให้คะแนนของนักเรียนเป็นโหนดราก เพราะมีคุณลักษณะที่เกี่ยวข้องมากที่สุดใช้ และใช้ตัวแปรงานของมารดา การออกไปเที่ยวกับเพื่อน และจำนวนการขาดเรียนสร้างเป็นกฎเงื่อนไข ดังนั้น สรุปได้ว่าเทคนิคการทำเหมืองข้อมูลด้วยการประมวลผลทั้ง 3 วิธี สามารถทำได้ได้อย่างแม่นยำต้องมีเกรดด้วย ถึงแม้ว่าผลสัมฤทธิ์ทางการเรียนของนักเรียนมีอิทธิพลอย่างสูงก็ตาม และยังมีคุณลักษณะที่เกี่ยวข้องอื่นๆ ด้วย เช่น จำนวนครั้งที่ขาดเรียน งานและการศึกษาของผู้ปกครอง การบริโภคเครื่องดื่มแอลกอฮอล์

2.4.3 Shameem Ahmed (2015) ได้วิจัย “Attitudes towards English Language Learning among EFL Learners at UMSKAL” การวิจัยเชิงคุณภาพเพื่อศึกษาทัศนคติและมุมมองของผู้เรียนต่อการใช้ภาษาอังกฤษในด้านต่างๆ รวมถึงสาเหตุที่อาจมีอิทธิพลต่อผลการเรียนภาษาอังกฤษ กลุ่มตัวอย่าง คือ นักศึกษาที่ไม่ได้เรียนวิชาภาษาอังกฤษเป็นวิชาเอก (English as a Foreign Language) จำนวน 238 คนในประเทศมาเลเซีย โดยเครื่องมือที่นำมาใช้ในงานวิจัยนี้ใช้แบบสอบถามในการเก็บรวบรวมข้อมูลประกอบด้วยคำถามปรนัย 19 คำถามและอัตนัย 2 คำถาม แบ่งการตอบคำถาม 5 ระดับ คือ 1= เห็นด้วย 2= ค่อนข้างเห็นด้วย 3= เป็นกลาง 4= ไม่เห็นด้วยบ้าง 5= ไม่เห็นด้วย มีการจัดแบ่งข้อคำถามเพื่อวิเคราะห์ทัศนคติในแบบต่างๆ เช่น ทัศนคติเชิงบวก ทัศนคติเชิงลบ ประสิทธิภาพที่ผ่านมาของผู้เรียนเกี่ยวกับภาษาอังกฤษ ทักษะทางภาษาในหลักสูตรการสอน ทำให้เห็นภาพรวมของทัศนคติของผู้นักศึกษาที่มีต่อการเรียนรู้ภาษาอังกฤษในทุกด้าน ผลการศึกษาพบว่า ทัศนคติเชิงบวกของผู้เรียนมีความสำคัญอย่างมากต่อการเรียนรู้ภาษาอังกฤษ ถึงแม้จะมีหลักสูตรภาษาอังกฤษในสถาบันการศึกษาแล้วก็ตาม และแรงจูงใจส่งผลต่อความสำเร็จหรือความล้มเหลว

2.4.4 พัฒนพงษ์ ดลรัตน์ และจารี ทองคำ (2017) ได้วิจัยเรื่อง “การเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์ความสำเร็จการศึกษาของนักเรียนระดับประกาศนียบัตรวิชาชีพ” โดยเปรียบเทียบเทคนิคเหมืองข้อมูลจำนวน 6 วิธี ได้แก่ วิธี Decision Tree (C4.5) วิธี Random Forest วิธี Random Tree วิธี Reduced Error Pruning (REP Tree), วิธี k-Nearest Neighbors (k-NN) และ Support Vector Machine โดยใช้ตัวแปรนำเข้า 17 ตัวแปร คือ เพศ อำเภอ ตำบลหมู่บ้าน อาชีพบิดามารดา สถานะครอบครัว เกรดเฉลี่ยกลุ่มงานอาชีพและเทคโนโลยี เกรดเฉลี่ยกลุ่มคณิตศาสตร์ เกรดเฉลี่ยกลุ่มสังคม เกรดเฉลี่ยกลุ่มศิลปะ เกรดเฉลี่ยกลุ่มวิทยาศาสตร์ เกรดเฉลี่ยกลุ่มภาษาต่างประเทศ เกรดเฉลี่ยกลุ่มภาษาไทย กลุ่มสุขศึกษาและพลศึกษา ประเภทวิชาที่จบ กำหนดตัวแปรเป้าหมาย คือ สถานะจบการศึกษา หรือไม่จบการศึกษา เก็บรวบรวมข้อมูลจากสถานศึกษาอาชีวศึกษาเอกชนในจังหวัดกาฬสินธุ์ จำนวน 12 แห่ง จำนวนนักเรียนรวมทั้งสิ้น 1266 คน และวัดประสิทธิภาพของแบบจำลองด้วย วิธี Cross validation แบบ 10- fold cross-validation โดยทำการแบ่งข้อมูลเป็น 10 ชุดเท่าๆกัน และทำการทดสอบทั้งหมด 10 รอบ เป็นอัตราส่วน 10:90 และวัดประสิทธิภาพแบบจำลองด้วย Confusion Matrix ซึ่งผลการศึกษาพบว่า วิธี Decision Tree (C4.5) มีความเหมาะสมที่สุดสำหรับใช้พยากรณ์ความสำเร็จการศึกษาในสถานศึกษาอาชีวศึกษาเอกชน จังหวัดกาฬสินธุ์ โดยมีค่าความแม่นยำ (Accuracy) ที่ 95.36% ค่าความถูกต้อง (Precision) ที่ 95.4%

2.4.5 นนทวัฒน์ ทวีชาติ อรยา เฟื่องประจัญ วิไลรัตน์ ยาทองไชย และ ชุศักดิ์ ยาทองไชย (2562) ได้วิจัย “ระบบทำนายการฟื้นฟูสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนิค การทำเหมืองข้อมูล” โดยทำการวิจัยแบ่งออกเป็น 2 ส่วน คือ การพัฒนาแบบจำลองการทำนายการฟื้นฟูสภาพของนักศึกษาระดับปริญญาตรี และนำผลที่ได้จากการสร้างแบบจำลองมาพัฒนาระบบการทำนายการฟื้นฟูสภาพของนักศึกษาระดับปริญญาตรี งานวิจัยนี้ใช้กลุ่มตัวอย่างนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ที่ศึกษาระหว่างปี พ.ศ. 2556 - 2561 จำนวน 3,650 ระเบียน โดยในส่วนของโครงสร้างแบบจำลองใช้โปรแกรม WEKA ในการวิเคราะห์และประมวลผลด้วยวิธี Decision Tree นำเข้าข้อมูลตัวแปรทั้งหมด 11 ตัวแปร คือ สาขาวิชาที่ศึกษาในคณะวิทยาศาสตร์ เกรดเฉลี่ยในภาคเรียนที่ 1 – 6 เกรดเฉลี่ยจากโรงเรียนมัธยม แผนการเรียนที่ศึกษาในโรงเรียนมัธยม ขนาดโรงเรียน สถานะการกู้ยืมเพื่อการศึกษา และสถานะการฟื้นฟูสภาพ ทดสอบประสิทธิภาพแบบจำลองด้วยวิธี Cross Validation แบบ 10- fold cross-validation และวัดประสิทธิภาพแบบจำลองด้วย Confusion Matrix ผลการศึกษาพบว่า แบบจำลองให้ค่าความถูกต้องที่ 95.57% และค่าความ

แม่นยำที่ 95.4% สำหรับในส่วนการพัฒนาระบบการทำนายการฟื้นสภาพของนักศึกษาระดับปริญญาตรีนั้น งานวิจัยนี้ใช้พัฒนาระบบในรูปแบบการใช้งานผ่านระบบเครือข่ายผ่านเว็บเบราว์เซอร์ และประเมินผลการทำงานของระบบด้วยแบบสอบถามความพึงพอใจจากกลุ่มตัวอย่างจำนวน 30 คน จากอาจารย์ที่ปรึกษา และนักศึกษาคณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ซึ่งโดยรวมผู้ที่มีความพึงพอใจระดับมาก

2.4.6 ศุภิกา นิธิติชัย (2561) ได้วิจัย “ทัศนคติ แรงจูงใจ และพฤติกรรมการเรียนภาษาอังกฤษของนักศึกษามหาวิทยาลัยราชภัฏบุรีรัมย์” ศึกษาความสัมพันธ์ระหว่างทัศนคติ แรงจูงใจ และพฤติกรรมการเรียนภาษาอังกฤษ และปัจจัยที่มีอิทธิพลต่อผลสัมฤทธิ์ทางการเรียนวิชาภาษาอังกฤษเพื่อการสื่อสารทั่วไป โดยเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างจากนักศึกษามหาวิทยาลัยราชภัฏบุรีรัมย์ ชั้นปีที่ 1 จำนวน 473 คน และใช้แบบสอบถาม ข้อสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาภาษาอังกฤษ และแบบสัมภาษณ์ถึงโครงสร้าง โดยแบบสอบถามได้แบ่งคำถามออกเป็น 5 ส่วน ประกอบด้วย ข้อมูลส่วนตัวของผู้ตอบแบบสอบถาม ทัศนคติของนักศึกษาต่อภาษาอังกฤษ แรงจูงใจของนักศึกษาต่อภาษาอังกฤษ และพฤติกรรมการเรียนภาษาอังกฤษ โดยข้อคำถามได้ผ่านการตรวจสอบความถูกต้องในด้านภาษา และเชิงเนื้อหา ด้วยวิธี Item-Objective Congruence Index – IOC และการทดสอบ (pre-test) กับนักศึกษา จำนวน 60 คน ได้ค่าสัมประสิทธิ์แอลฟาสูงกว่า 0.7 และใช้วิธีวิเคราะห์ถดถอยเชิงพหุ (Multiple Regression Analysis) ในการหาปัจจัยที่มีอิทธิพลต่อผลสัมฤทธิ์ทางการเรียนวิชาภาษาอังกฤษเพื่อการสื่อสารทั่วไป ผลการศึกษา พบว่า ความสัมพันธ์ระหว่างทัศนคติและแรงจูงใจภายในมีระดับสูงสุด ($r = .925, p < .01$) และผลการศึกษาปัจจัยมีเป็นอิทธิพล ทั้ง 3 ตัวแปรแสดงให้เห็นถึงความสัมพันธ์ที่แต่ละปัจจัยมีต่อกันอย่างชัดเจนทั้งมีนัยยะสำคัญ และไม่มีนัยยะสำคัญ โดยทัศนคติมีผลมากที่สุด

2.4.7 วันวิสาข์ ชนะประเสริฐ (2559) ได้วิจัย “การประยุกต์ใช้เทคนิคเหมืองข้อมูลเพื่อแนะนำอาชีพสำหรับนักศึกษาปริญญาตรี คณะโบราณคดี มหาวิทยาลัยศิลปากร” โดยศึกษา วิเคราะห์ถึงคุณลักษณะที่มีความเหมาะสมสำหรับนำไปใช้พัฒนาระบบการแนะนำอาชีพของนักศึกษามหาวิทยาลัยศิลปากร คณะโบราณคดี มหาวิทยาลัยศิลปากร โดยงานวิจัยมีขั้นตอนการดำเนินงานแบ่งเป็น 3 ส่วน ประกอบด้วย 1) เปรียบเทียบเกณฑ์สำหรับใช้คัดเลือกคุณสมบัติ 3 แบบ ได้แก่ InfoGainAttributeEval, CfsSubsetEval และการไม่คัดเลือก 2) นำคุณลักษณะที่ได้จากการเปรียบเทียบเกณฑ์ตาม 1) มาพัฒนาแบบจำลองด้วยเทคนิค Random Forest, Neural Network และ Naive Bayes เพื่อเปรียบเทียบหาแบบจำลองที่เหมาะสมต่อการคัดเลือกคุณลักษณะดังกล่าว และ 3) นำแบบจำลองมาพัฒนาระบบแนะนำอาชีพ โดยประเมินระบบจากแบบสอบถามความพึงพอใจของผู้ใช้งานที่มีต่อระบบกับของกลุ่มตัวอย่างจำนวน 199 คน งานวิจัยมีการเก็บรวบรวมจากฐานข้อมูลประวัติและการศึกษา และแบบสอบถามของนักศึกษาที่สำเร็จการศึกษาปริญญาตรี คณะโบราณคดี มหาวิทยาลัยศิลปากร และได้งานทำหรือประกอบอาชีพอิสระ จำนวน 400 คน สัดส่วนชุดข้อมูลการเรียนรู้ (Training) ร้อยละ 70% และชุดข้อมูลการทดสอบ (Testing) ร้อยละ 30% ประกอบด้วย 22 คุณลักษณะ ผลการศึกษาพบว่า ประสิทธิภาพแบบจำลองที่ใช้การคัดเลือกด้วยวิธี InfoGainAttributeEval ร่วมกับเทคนิค Neural Network มีความถูกต้องในการจำแนกข้อมูลในเกณฑ์ที่ดีที่สุด และเมื่อนำผลมาพัฒนาระบบแนะนำอาชีพสำหรับนักศึกษาฯ ในกลุ่มตัวอย่างจำนวน 199 คน มีความพึงพอใจภาพรวมระบบอยู่ในระดับมาก

จากการศึกษางานวิจัยต่างๆ ที่เกี่ยวข้อง การนำข้อมูลมาวิเคราะห์และพัฒนาแบบจำลองด้วยเทคนิคการทำเหมืองข้อมูล นักวิจัยส่วนใหญ่ทำการเปรียบเทียบประสิทธิภาพแบบจำลองประเภทต่างๆ เพื่อคัดเลือกแบบจำลองที่เหมาะสมกับชุดข้อมูล และการทำเหมืองข้อมูลด้วยวิธีการจำแนกประเภทโดยเทคนิคต้นไม้ Decision Tree และเทคนิค Random Forest ได้รับความนิยมนำมาใช้ในการจำแนกประสิทธิภาพการเรียน รวมถึงใช้ทำนายแนวโน้มผลการศึกษาและค้นหาคุณลักษณะที่สำคัญที่อยู่ภายในชุดข้อมูล นอกจากนี้ การแบ่งสัดส่วนข้อมูลในการประมวลผลส่วนใหญ่ใช้วิธี Cross Validation เพื่อเพิ่มประสิทธิภาพการประมวลผลข้อมูล และวัดประสิทธิภาพแบบจำลอง เช่น ค่าความแม่นยำ (Accuracy) ค่าความถูกต้องของการทำนาย (Precision) ค่าความระลึกได้ (Recall) เป็นต้น สำหรับในการศึกษานี้ผู้วิจัยประยุกต์ใช้เทคนิคเหมืองข้อมูล

เปรียบเทียบระหว่างเทคนิค Decision Tree และเทคนิค Random Forest เพื่อวิเคราะห์คุณลักษณะที่มีอิทธิพลที่ส่งผลให้นักศึกษามีผลการเรียนวิชาภาษาอังกฤษพื้นฐานดีมาก โดยจะใช้การแบ่งสัดส่วนข้อมูลในการประมวลผลด้วยวิธี Cross Validation และประสิทธิภาพแบบจำลอง เพื่อให้ได้ตัวแบบที่เหมาะสมกับการค้นหาคุณลักษณะที่สำคัญที่ส่งผลต่อประสิทธิภาพการเรียนวิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี คณะภูมิสารสนเทศศาสตร์ มหาวิทยาลัยบูรพา

บทที่ 3

วิธีการดำเนินการวิจัย

การวิจัยนี้เป็นการวิจัยเชิงปริมาณ (Quantitative Research) โดยประยุกต์ใช้การวิเคราะห์ข้อมูลด้วยวิธีการเรียนรู้แบบมีผู้สอน (Supervised Learning) ซึ่งมีรายละเอียดวิธีการดำเนินการวิจัย ดังนี้

- 3.1 ประชากรและกลุ่มตัวอย่าง
- 3.2 เครื่องมือที่ใช้ในการวิจัย
- 3.3 การเก็บรวบรวมข้อมูล
- 3.4 ขั้นตอนการดำเนินการ
- 3.5 กรอบแนวคิดการวิจัย

3.1 ประชากรและกลุ่มตัวอย่าง

3.1.1 ประชากร

การวิจัยนี้กำหนดประชากร คือ นักศึกษาที่เข้าศึกษาในคณะภูมิสารสนเทศศาสตร์ มหาวิทยาลัยบูรพา ระหว่างปีการศึกษา พ.ศ.2561 - 2563 และมีผลการเรียนวิชาภาษาอังกฤษระดับพื้นฐานของมหาวิทยาลัย จำนวน 462 คน (ข้อมูลจาก กองทะเบียนและประมวลผลการศึกษา มหาวิทยาลัยบูรพา ณ วันที่ 4 พฤษภาคม 2564)

3.1.2 กลุ่มตัวอย่าง

จำนวนกลุ่มตัวอย่างสำหรับการศึกษานี้ใช้การคำนวณสัดส่วนกลุ่มตัวอย่างตาม Krejcie และ Morgan ดังนี้ (ธีรวุฒิ, 2543)

$$n = \frac{x^2 N p (1 - p)}{e^2 (N - 1) + x^2 p (1 - p)}$$

โดย n = ขนาดของกลุ่มตัวอย่าง

N = ขนาดของประชากร

e = ระดับความคลาดเคลื่อนของการสุ่มตัวอย่างที่ยอมรับได้

x^2 = ค่าไคสแควร์ที่ df เท่ากับ 1 และระดับความเชื่อมั่น 95% ($x^2=3.841$)

p = สัดส่วนของลักษณะที่สนใจในประชากร (ถ้าไม่ทราบให้กำหนด $p = 0.5$)

โดยกำหนดระดับการยอมรับให้เกิดความคลาดเคลื่อนของการสุ่มตัวอย่างคือ 5% ที่ระดับความเชื่อมั่น 95% และสัดส่วนของลักษณะที่สนใจในประชากรเท่ากับ 0.5 แทนค่าในสมการที่

$$n = \frac{3.841 \times 462 \times 0.5 \times 0.5}{(.05^2) \times (462 - 1) + 3.841 \times 0.5 \times 0.5}$$
$$n = 209.98014166 = 210 \text{ หน่วย}$$

ดังนั้น ขนาดกลุ่มตัวอย่างน้อยที่สุด คือ 210 คน อย่างไรก็ตาม เพื่อลดคลาดเคลื่อนจากการเก็บข้อมูลและเพื่อให้ได้ผลลัพธ์ที่ถูกต้องมากขึ้น ผู้วิจัยจึงเพิ่มขนาดกลุ่มตัวอย่างเป็น 299 คน

3.2 เครื่องมือที่ใช้ในการวิจัย

3.2.1 ฮาร์ดแวร์

- 1) Intel(R) Core (TM) i5-7300HQ
- 2) RAM 12 GB
- 3) CPU@2.50 GHz 2.50 GHz
- 4) HDD 500 GB

3.2.2 ซอฟต์แวร์

- 1) โปรแกรมจัดการตารางงาน Microsoft Excel
- 2) โปรแกรมสำหรับช่วยเขียนโค้ด Jupyter โดยใช้ภาษา Python

3.2.3 แบบสอบถามออนไลน์ (Google Form) ที่ผ่านการตรวจสอบความเที่ยงตรงเชิงเนื้อหา (IOC) จากการประเมินของผู้เชี่ยวชาญจำนวน 3 ท่าน

3.3 ขั้นตอนการดำเนินงาน

ในการศึกษาและวิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคการทำเหมืองข้อมูล ผู้วิจัยประยุกต์ใช้กระบวนการพัฒนาเหมืองข้อมูล (CRISP-DM) เพื่อเป็นแนวทางดำเนินงานวิจัย แบ่งวิธีการดำเนินงานออกเป็น 5 ขั้นตอน ดังนี้

3.4.1 ศึกษาและรวบรวมข้อมูล

ในการศึกษาครั้งนี้ใช้แบบสอบถามปลายปิดในการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างนักศึกษาคณะมนุษยศาสตร์ มหาวิทยาลัยบูรพาที่เข้าศึกษาระหว่างปี พ.ศ. 2561 – 2563 จำนวน 299 คน โดยแบบสอบถามผ่านการตรวจสอบความสมบูรณ์ของข้อมูลทั้งหมดด้วยการประเมินความเที่ยงตรงเชิงเนื้อหา (IOC) จากผู้เชี่ยวชาญจำนวน 3 ท่าน ก่อนนำเข้าสู่การวิเคราะห์ด้วยเทคนิคเหมืองข้อมูล แบบสอบถามแบ่งออกเป็น 4 ส่วน ดังนี้

1) ข้อมูลด้านการเรียนภาษาอังกฤษของนักศึกษา ประกอบด้วย เกรดวิชาภาษาอังกฤษระดับพื้นฐานของมหาวิทยาลัย และจำนวนการขาดเรียนวิชาภาษาอังกฤษ

2) ข้อมูลพื้นฐานทั่วไปของนักศึกษา ประกอบด้วย เพศ อายุ มีคอมพิวเตอร์และอินเทอร์เน็ตใช้ส่วนตัว ลักษณะที่ตั้งของบ้านที่อยู่อาศัย

3) ข้อมูลด้านพฤติกรรมที่ส่งผลต่อการเรียนวิชาภาษาอังกฤษ ประกอบด้วย เรียนพิเศษวิชาภาษาอังกฤษ เวลาว่างจากการเรียน เวลาสำหรับออกไปเที่ยวกับเพื่อน เวลาว่างสำหรับฝึกฝนภาษาอังกฤษด้วยตนเอง ถามอาจารย์ทันทีเมื่อไม่เข้าใจเนื้อหาที่เรียน วางแผนการเรียนภาษาอังกฤษ เตรียมตัวก่อนเข้าเรียนวิชาภาษาอังกฤษ ชอบนั่งใกล้ๆ เพื่อนที่ตั้งใจเรียน มีเพื่อนเป็นชาวต่างชาติเพื่อฝึกใช้ทักษะภาษาอังกฤษ สนใจและตั้งใจเรียนวิชาภาษาอังกฤษในขณะที่อาจารย์สอน จดบันทึกตามที่อาจารย์สอน ตอบคำถามในขณะที่เรียนภาษาอังกฤษ ทบทวนบทเรียนภาษาอังกฤษหลังเลิกเรียน ศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษนอกเหนือจากที่อาจารย์สอน และจับกลุ่มทบทวนบทเรียนภาษาอังกฤษกับเพื่อนเมื่อใกล้สอบ

4) ข้อมูลทัศนคติและแรงจูงใจในการเรียนภาษาอังกฤษ ประกอบด้วย ความต้องการศึกษาในระดับที่สูงขึ้น การสนับสนุนด้านการเรียนภาษาอังกฤษจากครอบครัว สนุกกับการเรียนภาษาอังกฤษ สนใจเรียนภาษาอังกฤษเพื่อพัฒนาตนเอง สนใจเรียนภาษาอังกฤษเพื่อหางานที่ดีในอนาคต สนใจเรียนภาษาอังกฤษเพราะเป็นความต้องการของพ่อแม่ สนใจเรียนภาษาอังกฤษเพื่อต้องการสอบผ่าน สนใจเรียนภาษาอังกฤษเพื่อใช้สื่อสารในชีวิตประจำวัน ความสุขในการอ่านหนังสือหรือ

คุณภาพนตร์เป็นภาษาอังกฤษ การเรียนภาษาอังกฤษมีความสำคัญเพราะต้องการได้เกรดที่ดี และการเรียนภาษาอังกฤษมีความสำคัญ เพราะภาษาอังกฤษช่วยให้มีความรู้มากขึ้น

ทั้งนี้ จุดมุ่งหมายของการเก็บรวบรวมข้อมูลจากคำถามดังกล่าว คือ เพื่อให้ได้ชุดข้อมูลที่จะสามารถนำมาใช้ค้นหาคุณลักษณะสำคัญที่จะบ่งบอกได้ว่าคุณลักษณะแบบใดที่เป็นสาเหตุทำให้นักศึกษาระดับปริญญาตรีมีประสิทธิภาพการเรียนวิชาภาษาอังกฤษที่ดี โดยใช้กระบวนการวิเคราะห์ด้วยเทคนิคเหมืองข้อมูล ประยุกต์ใช้วิธีการจำแนกคุณลักษณะเป้าหมายแบบเลขฐานสอง (Binary Classification) (Cortez and Silva, 2014) ชุดข้อมูลที่จะนำมาวิเคราะห์ประกอบไปด้วย 31 คุณลักษณะ (Input Attribute) และ 1 คุณลักษณะเป้าหมาย (Target Attribute) โดยใช้เกรควิชาภาษาอังกฤษพื้นฐานของนักศึกษาจำแนกแบบเลขฐานสอง แบ่งเป็นนักศึกษาที่ได้รับเกรดเอเป็นกลุ่มนักศึกษาที่มีประสิทธิภาพการเรียนวิชาภาษาอังกฤษ และนักศึกษาที่ได้ผลการเรียนอื่นๆ เป็นกลุ่มนักศึกษาที่มีผลการเรียนวิชาภาษาอังกฤษระดับพื้นฐานทั่วไป

3.4.2 เตรียมข้อมูล

1) การตรวจสอบข้อมูล เป็นขั้นตอนการตรวจสอบความผิดพลาด หรือความไม่สมบูรณ์ของข้อมูลที่ได้จากการเก็บรวบรวมแบบสอบถามทั้งหมดจำนวน 299 ตัวอย่าง โดยนำข้อมูลจัดเก็บในโปรแกรมจัดการตารางงาน Microsoft Excel และพิจารณาตัดส่วนที่ไม่เกี่ยวข้อง ได้แก่ เวลาการตอบแบบสอบถาม และรหัสนักศึกษา จากนั้นจัดการข้อมูลที่มีความไม่สอดคล้องกับรูปแบบข้อมูลที่ต้องการใช้งาน ตัวอย่างดังภาพที่ 3.1 และปรับแก้ไขภาษาข้อมูลคำถามจากภาษาไทยเป็นภาษาอังกฤษทั้งหมด ตัวอย่างดังภาพที่ 3.2 นอกจากนี้ ชุดข้อมูลสำหรับการศึกษาคครั้งนี้ไม่มีค่าข้อมูลที่ขาดหายไป (Missing Value) เนื่องจากในแบบสอบถามผู้วิจัยได้กำหนดความจำเป็นในการตอบคำถามทุกข้อ

2. เพศ	3. อายุขล	4. ดัชนีคลอ	5. ลักษณะที่ตั้งของบ้านที่ อยู่อาศัยของเงิน	6. ผลการเรียนวิชาภาษาอังกฤษระดับ มหาวิทยาลัย (วิชาบังคับระดับต้นพื้นฐาน)	7. ดัชนีไม่ได้เข้าเรียนวิชาภาษาอังกฤษ (จำนวนครั้ง)	8. ดัชนีเรียนพิเศษภาษาอังกฤษ	9. เวลาว่างจากการเรียนของเงิน (ไม่นับช่วงเวลา ทำงานพิเศษหรือทำงานบ้านหรือการนอนพัก)	10. ดัชนี ออกใจ	
หญิง	23 มี		ตั้งอยู่ในเมือง		B	0	ไม่ใช้	มาก	ปานกลาง
ชาย	20 มี		ตั้งอยู่ในเมือง		C	0	ไม่ใช้	น้อยมาก	ปานกลาง
หญิง	19 มี		อยู่นอกเมืองหรืออยู่ในเขตชน		B	0	ไม่ใช้	น้อย	น้อย
หญิง	20 มี		อยู่นอกเมืองหรืออยู่ในเขตชน		C	0	ไม่ใช้	ปานกลาง	ปานกลาง
ชาย	20 มี		อยู่นอกเมืองหรืออยู่ในเขตชน		D	0	ไม่ใช้	ปานกลาง	ปานกลาง
ชาย	20 มี		อยู่นอกเมืองหรืออยู่ในเขตชน		C	0	ไม่ใช้	ปานกลาง	ปานกลาง
หญิง	20 มี		อยู่นอกเมืองหรืออยู่ในเขตชน		C	2	ไม่ใช้	ปานกลาง	ปานกลาง
หญิง	21 มี		ตั้งอยู่ในเมือง		C	3	ไม่ใช้	น้อย	น้อย
ชาย	20 มี		อยู่นอกเมืองหรืออยู่ในเขตชน		C	2	ไม่ใช้	ปานกลาง	ปานกลาง
หญิง	20 มี		อยู่นอกเมืองหรืออยู่ในเขตชน		C	6	ไม่ใช้	ปานกลาง	มาก
หญิง	21 มี		ตั้งอยู่ในเมือง		A	0	ใช่	น้อยมาก	น้อย
หญิง	20 มี		ตั้งอยู่ในเมือง		C	0	ไม่ใช้	ปานกลาง	ปานกลาง
หญิง	21 มี		อยู่นอกเมืองหรืออยู่ในเขตชน		A	0	ไม่ใช้	น้อย	น้อย
หญิง	21 มี		อยู่นอกเมืองหรืออยู่ในเขตชน		A	0	ใช่	น้อยมาก	ปานกลาง
หญิง	21 มี		อยู่นอกเมืองหรืออยู่ในเขตชน		C	0	ไม่ใช้	ปานกลาง	มาก
หญิง	21 มี		อยู่นอกเมืองหรืออยู่ในเขตชน		C	3	ไม่ใช้	ปานกลาง	น้อย
หญิง	19 ปี		ตั้งอยู่ในเมือง		A	0	ไม่ใช้	ปานกลาง	ปานกลาง
หญิง	20 มี		อยู่นอกเมืองหรืออยู่ในเขตชน		B	3	ไม่ใช้	มาก	ปานกลาง
หญิง	27 มี		ตั้งอยู่ในเมือง		C	3	ไม่ใช้	น้อยมาก	น้อย

ภาพที่ 3.1 ตัวอย่างข้อมูลที่มีความผิดพลาดไม่สอดคล้องกับรูปแบบข้อมูลที่ต้องการใช้งาน

จากภาพที่ 3.1 จะเห็นได้ว่าเป็นข้อมูลที่ไม่สอดคล้อง เนื่องจากมีกลุ่มนักศึกษาตัวอย่างบางคนใส่ตัวอักษร ปี เข้ามาพร้อมกับอายุที่ต้องการเก็บข้อมูลในลักษณะตัวเลขเท่านั้น ทำให้ต้องทำการตรวจหาและแก้ไขข้อมูลให้เป็นตัวเลขทั้งหมด

GENDER	AGE	H-COM	LOC	GRADE	ABSENCE	S-ENG	F-TIME	H-FREQ	S-STUDY	A-PROF	E-PLAN	P-SELF	S-FRIEND	F-FRIEND	I-EN
F	23	Yes	urban	B		0	No	high	moderate	moderate	moderate	Yes	No	No	mod
M	20	Yes	urban	C		0	No	very little	moderate	moderate	No	No	Yes	Yes	mod
F	19	Yes	rural	B		0	No	little	very little	moderate	little	No	Yes	No	high
F	20	Yes	rural	C		0	No	moderate	moderate	moderate	Yes	Yes	Yes	No	very
M	20	Yes	rural	D		0	No	moderate	moderate	very little	little	No	Yes	No	mod
M	20	Yes	rural	C		0	No	moderate	moderate	very little	very little	No	No	Yes	No
F	20	Yes	rural	C		2	No	moderate	moderate	very little	very little	No	No	Yes	No
F	21	Yes	urban	C		3	No	little	little	little	No	No	Yes	No	high
M	20	Yes	rural	C		2	No	moderate	moderate	little	little	No	No	No	mod
F	20	Yes	rural	C		6	No	moderate	high	very little	very little	No	No	Yes	No
F	21	Yes	urban	A		0	Yes	very little	little	moderate	very little	No	No	No	high
F	20	Yes	urban	C		0	No	moderate	moderate	very little	very little	No	No	No	little
F	21	Yes	rural	A		0	No	little	little	moderate	little	No	No	Yes	No
F	21	Yes	rural	A		0	Yes	very little	moderate	little	little	No	Yes	No	mod
F	21	Yes	rural	C		0	No	moderate	high	little	little	Yes	No	Yes	No
F	21	Yes	rural	C		3	No	moderate	very little	little	very little	Yes	Yes	Yes	No
F	19	Yes	urban	A		0	No	moderate	moderate	moderate	Yes	Yes	No	No	mod
F	20	Yes	urban	B		3	No	high	moderate	very little	No	No	Yes	No	mod
F	23	Yes	urban	C		3	No	very little	little	very little	very little	No	No	No	mod
F	23	Yes	rural	C		2	Yes	moderate	little	moderate	Yes	Yes	Yes	No	high
F	21	Yes	urban	D		0	No	very little	very little	moderate	little	Yes	No	No	high

ภาพที่ 3.2 ตัวอย่างข้อมูลที่ปรับเป็นภาษาอังกฤษ

จากภาพที่ 3.2 แสดงการปรับแก้ไขชุดข้อมูลเป็นภาษาอังกฤษทั้งหมดเพื่อให้อยู่ในรูปแบบที่จะนำมาวิเคราะห์ด้วยเทคนิคเหมืองข้อมูลได้ ยกตัวอย่างเช่น ข้อมูลเพศชายและเพศหญิง เปลี่ยนเป็น Male และ Female เปลี่ยนจากคำตอบ 5 ระดับ น้อยมาก น้อย ปานกลาง มาก และมากที่สุด เป็น Very little, Little, Moderate, High และ Very high เป็นต้น โดยละเอียดชุดข้อมูลที่ปรับแก้ไขก่อนนำเข้าสู่โปรแกรม แสดงดังตาราง 3.1

ตาราง 3.1 ชุดข้อมูล (Dataset) ที่เกี่ยวข้องกับความสำเร็จทางการเรียนวิชาภาษาอังกฤษสำหรับใช้ประมวลผลข้อมูล

คุณลักษณะ	คำอธิบาย	ค่าที่ปรับแก้
GENDER	เพศของนักศึกษา	{F, M}
AGE	อายุของนักศึกษา	{numeric 18 to 25}
H-COM	มีคอมพิวเตอร์และอินเทอร์เน็ตใช้ส่วนตัว	{yes, no}
LOC	ลักษณะที่ตั้งของบ้านที่อยู่อาศัย	{urban, rural}
ABSENCE	จำนวนการขาดเรียนวิชาภาษาอังกฤษ	{numeric 1 to 10}
S-ENG	เรียนพิเศษวิชาภาษาอังกฤษ	{yes, no}
F-TIME	เวลาว่างจากการเรียน (ไม่นับช่วงเวลาทำงานพิเศษหรือทำงานบ้านหรือการนอนพัก)	{very little, little, moderate, high, very high}
H-FREQ	เวลาสำหรับออกไปเที่ยวกับเพื่อน	{very little, little, moderate, high, very high}
S-STUDY	เวลาว่างสำหรับฝึกฝนภาษาอังกฤษด้วยตนเอง	{very little, little, moderate, high, very high}
A-PROF	ถามอาจารย์ทันทีเมื่อไม่เข้าใจเนื้อหาที่เรียน	{very little, little, moderate, high, very high}
E-PLAN	วางแผนการเรียนภาษาอังกฤษ	{yes, no}
P-SELF	เตรียมตัวก่อนเข้าเรียนวิชาภาษาอังกฤษ	{yes, no}
S-FRIEND	ชอบนั่งใกล้ๆ เพื่อนที่ตั้งใจเรียน	{yes, no}
F-FRIEND	มีเพื่อนเป็นชาวต่างชาติ เพื่อฝึกใช้ทักษะภาษาอังกฤษ	{yes, no}

คุณลักษณะ	คำอธิบาย	ค่าที่ปรับแก้
I-ENG	สนใจและตั้งใจเรียนวิชาภาษาอังกฤษในขณะที่อาจารย์สอน	{very little, little, moderate, high, very high}
T-NOTES	จดบันทึกตามที่อาจารย์สอน	{very little, little, moderate, high, very high}
A-QUE	ตอบคำถามในขณะที่เรียนภาษาอังกฤษ	{very little, little, moderate, high, very high}
A-SCHOOL	ทบทวนบทเรียนภาษาอังกฤษหลังเลิกเรียน	{very little, little, moderate, high, very high}
E-RES	ศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษนอกเหนือจากที่อาจารย์สอน	{very little, little, moderate, high, very high}
G-STUDY	จับกลุ่มทบทวนบทเรียนภาษาอังกฤษกับเพื่อนเมื่อใกล้สอบ	{very little, little, moderate, high, very high}
H-EDU	ความต้องการศึกษาในระดับที่สูงขึ้น	{very little, little, moderate, high, very high}
F-SUPP	การสนับสนุนด้านการเรียนภาษาอังกฤษจากครอบครัว	{yes, no}
E-ENG	สนุกกับการเรียนภาษาอังกฤษ	{very little, little, moderate, high, very high}
L-IMPROV	สนใจเรียนภาษาอังกฤษเพื่อพัฒนาตนเอง	{very little, little, moderate, high, very high}
L-JOB	สนใจเรียนภาษาอังกฤษ เพื่อหางานที่ดีในอนาคต	{very little, little, moderate, high, very high}
L-PAREN	สนใจเรียนภาษาอังกฤษ เพราะเป็นความต้องการของพ่อแม่	{very little, little, moderate, high, very high}
L-PASS	สนใจเรียนภาษาอังกฤษ เพื่อต้องการสอบผ่าน	{very little, little, moderate, high, very high}
L-DAILY	สนใจเรียนภาษาอังกฤษ เพื่อใช้สื่อสารในชีวิตประจำวัน	{very little, little, moderate, high, very high}
E-MOV	ความสุขในการอ่านหนังสือ ดูภาพยนตร์เป็นภาษาอังกฤษ	{very little, little, moderate, high, very high}
L-GRADE	การเรียนภาษาอังกฤษมีความสำคัญ เพราะต้องการได้เกรดที่ดี	{very little, little, moderate, high, very high}
L-KNOW	การเรียนภาษาอังกฤษมีความสำคัญ เพราะภาษาอังกฤษช่วยให้มีความรู้มากขึ้น	{very little, little, moderate, high, very high}
Grade	เกรดวิชาภาษาอังกฤษพื้นฐานของมหาวิทยาลัย	{A , B, C, D, F }

2) การแปลงข้อมูล ขั้นตอนการแปลงข้อมูลเพื่อให้มีความเหมาะสมก่อนเข้าสู่กระบวนการทำเหมืองข้อมูล ซึ่งปกติแล้วอัลกอริทึมการเรียนรู้ของเครื่องจะวิเคราะห์ข้อมูลเฉพาะค่าตัวเลข ดังนั้น ในขั้นตอนนี้จึงต้องทำการปรับสภาพข้อมูลให้อยู่ในรูปแบบที่เหมาะสมก่อนนำไปวิเคราะห์ เช่น 1) วิธีการเข้ารหัสข้อมูลแบบ One-hot encoding เพื่อแปลงข้อมูลจากข้อความ (Label) ให้เป็นตัวเลขฐานสอง ใช้ในกรณีที่ค่าจริงของข้อมูลไม่ได้มีความสัมพันธ์ต่อเนื่องเป็นลำดับขั้น เพื่อป้องกันอัลกอริทึมตีความค่าของตัวเลขใกล้เคียงกันมีคุณค่าข้อมูลเหมือนกัน เช่น จังหวัด เป็นต้น 2) วิธีการเข้ารหัสแบบ Factorization ใช้การแทนค่าข้อมูลเป็นตัวเลขเรียงลำดับ เช่น ต่ำ แทนค่าด้วย 0 ปานกลาง แทนค่าด้วย 1 และ สูง แทนค่าด้วย 2 จะใช้ในกรณีที่ค่าจริงของข้อมูลมีความสัมพันธ์เชิงคุณค่าที่ต่อเนื่องกัน และ 3) การลดการกระจายตัวของข้อมูลด้วยการปรับขนาดข้อมูล (ต้องทำให้ข้อมูลเป็นตัวเลขทั้งหมดก่อน) ให้มีช่วงข้อมูลที่เท่ากันในกรณีที่ชุดข้อมูลมีค่ากระจายตัวแตกต่างกันมาก

สำหรับการศึกษานี้ผู้วิจัยใช้โปรแกรมสำหรับช่วยเขียนโค้ด Jupyter ผ่านไลบรารี Pandas ของ Python นำเข้าไฟล์ชุดข้อมูลเพื่อทำการปรับข้อมูลด้วยวิธีการเข้ารหัสแบบ Factorization แทนค่าในแต่ละชุดข้อมูลคุณลักษณะที่เป็นชนิดข้อความ (Label) ให้อยู่ในรูปตัวเลข (Numerical) เนื่องจากข้อมูลมีความสัมพันธ์แบบต่อเนื่องกัน และไม่มีมีการกระจายข้อมูลที่แตกต่างกันมาก แสดงในภาพ ที่ 3.3

In [80]: df.head()

Out[80]:

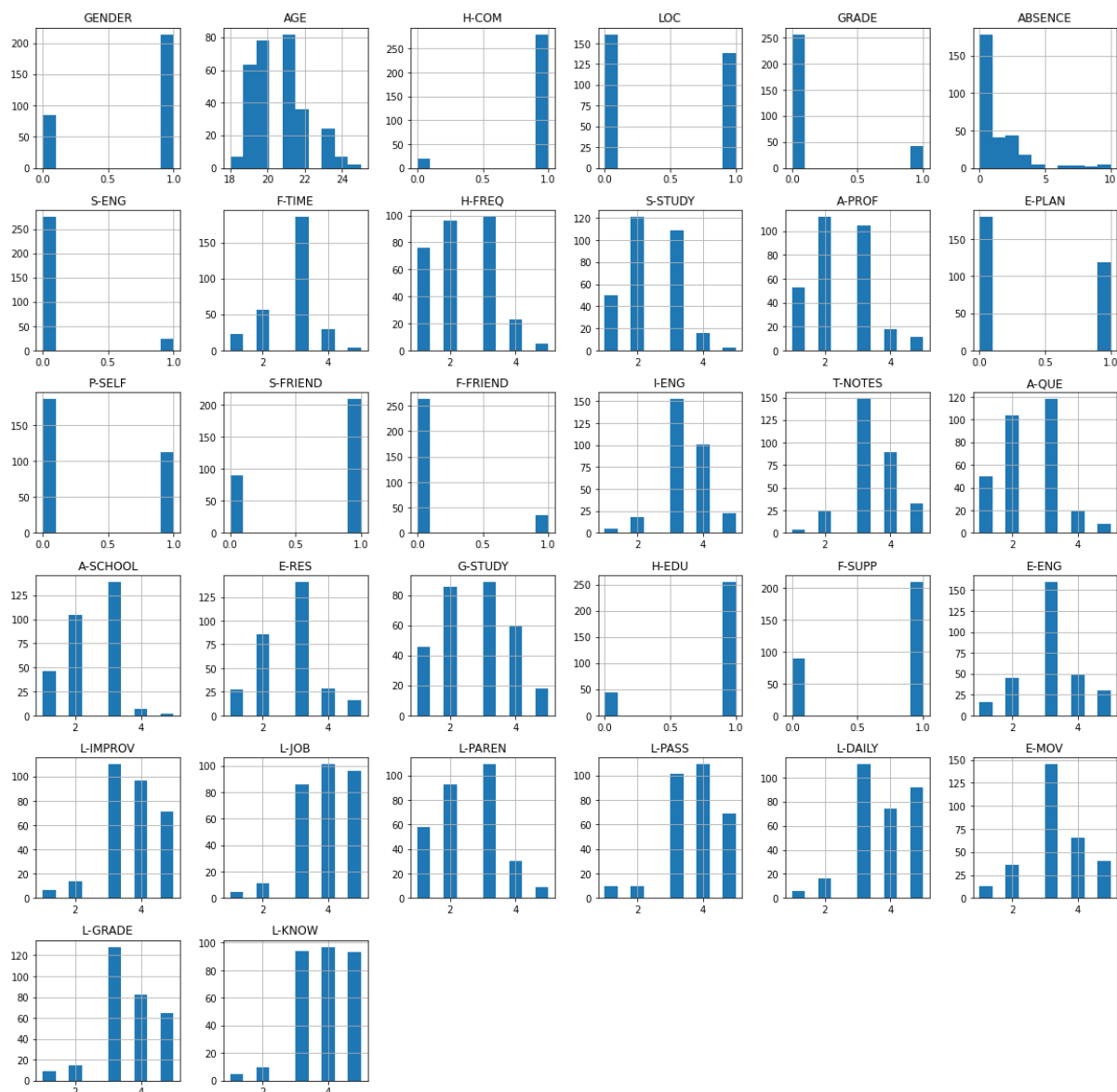
	GENDER	AGE	H-COM	LOC	GRADE	ABSENCE	S-ENG	F-TIME	H-FREQ	S-STUDY	...	F-SUPP	E-ENG	L-IMPROV	L-JOB	L-PAREN	L-PASS	L-DAILY	E-MOV	L-GRADE	L-KNOW
0	1	23	1	1	0	0	0	4	3	3	...	0	3	4	3	2	4	3	3	3	3
1	0	20	1	1	0	0	0	1	3	3	...	1	3	4	3	3	4	4	3	3	3
2	1	19	1	0	0	0	0	2	1	3	...	0	4	5	5	3	4	5	3	4	5
3	1	20	1	0	0	0	0	3	3	3	...	1	5	5	5	2	2	2	2	3	4
4	0	20	1	0	0	0	0	3	3	1	...	1	2	3	3	2	4	3	3	4	3

5 rows x 32 columns

ภาพที่ 3.3 ตัวอย่างข้อมูลที่แปลงเป็นตัวเลข ด้วยโปรแกรม Jupyter

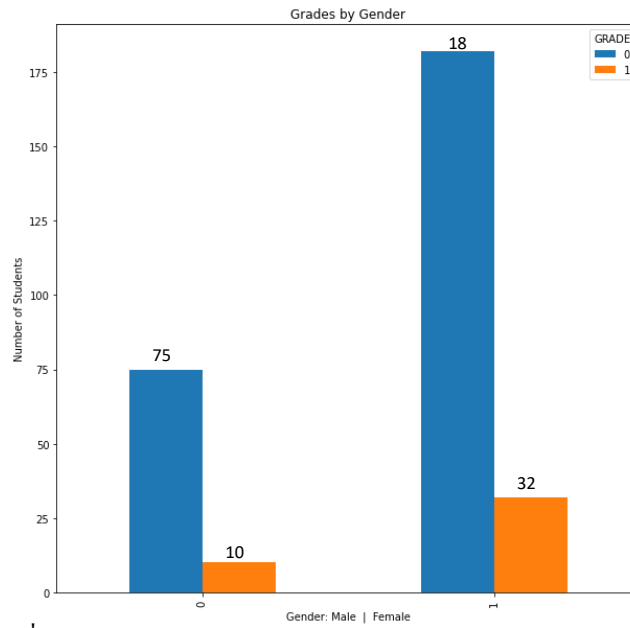
จากภาพที่ 3.3 ผู้วิจัยทำการปรับค่าของข้อมูลคุณลักษณะเป้าหมายเกรดวิชาภาษาอังกฤษ (A แทนค่าด้วย 1 คือกลุ่มนักศึกษาที่มีประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษระดับพื้นฐาน และ B, C, D, F แทนค่าด้วย 0 คือกลุ่มนักศึกษาที่มีผลการเรียนวิชาภาษาอังกฤษระดับพื้นฐานทั่วไป) เพื่อวิเคราะห์คุณลักษณะที่บ่งบอกถึงประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาที่มีผลการเรียนดีมาก (นักศึกษาที่ได้รับเกรดเอ) ตามการจำแนกตัวแปรเป้าหมายแบบเลขฐานสอง (Binary Classification) และปรับค่าข้อมูลคุณลักษณะที่มีค่า 5 ระดับ (Very little แทนค่าด้วย 1, Little แทนค่าด้วย 2, Moderate แทนค่าด้วย 3, High แทนค่าด้วย 4 และ Very high แทนค่าด้วย 5) คุณลักษณะเพศ (M แทนค่าด้วย 0 และ F แทนค่าด้วย 1) และข้อมูลคุณลักษณะอื่นๆ ที่มีค่า 2 ระดับ (Yes แทนค่าด้วย 1 และ No แทนค่าด้วย 0)

3) การสำรวจข้อมูล เป็นขั้นตอนการพิจารณาภาพรวมของข้อมูล โดยจัดทำข้อมูลให้แสดงในลักษณะแผนภาพหรือกราฟต่างๆ เพื่อให้ง่ายต่อการทำความเข้าใจ และนำมาใช้พิจารณาแนวโน้มทางสถิติภายในชุดข้อมูล ซึ่งการศึกษานี้ผู้วิจัยทำการสำรวจข้อมูลทั้งหมด จำนวน 32 คุณลักษณะ โดยใช้ไลบรารี Matplotlib, ของ Python ในการจัดทำข้อมูลให้แสดงในลักษณะกราฟต่างๆ ผ่านโปรแกรม Jupyter จากการสำรวจ พบว่า สัดส่วนกลุ่มนักศึกษาที่มีประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษมีจำนวนทั้งหมด 42 คน และกลุ่มนักศึกษาที่มีผลการเรียนวิชาภาษาอังกฤษระดับพื้นฐานทั่วไป จำนวน 257 คน ซึ่งกราฟภาพรวมของข้อมูลคุณลักษณะทั้งหมดแสดงในภาพที่ 3.4 และกราฟเปรียบเทียบระหว่างข้อมูล 31 คุณลักษณะ กับข้อมูลคุณลักษณะเป้าหมาย (Target Attribute) ดังภาพที่ 3.5 - ภาพที่ 3.35

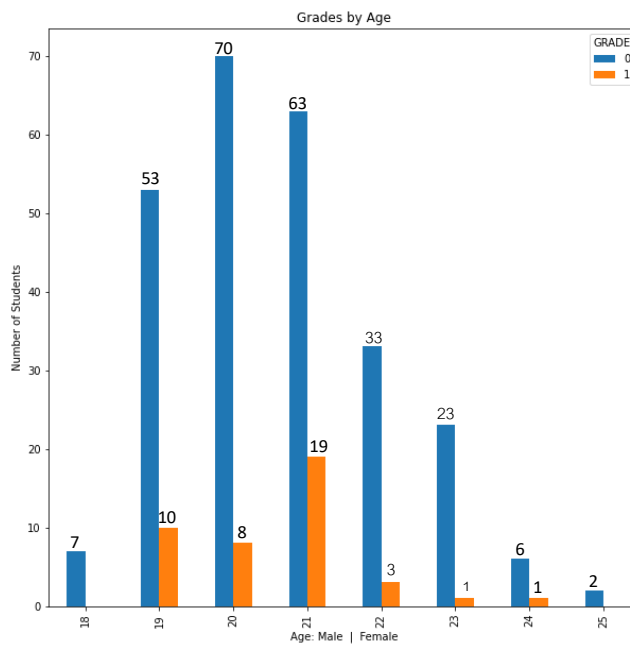


ภาพที่ 3.4 กราฟภาพรวมของข้อมูลคุณลักษณะทั้งหมด

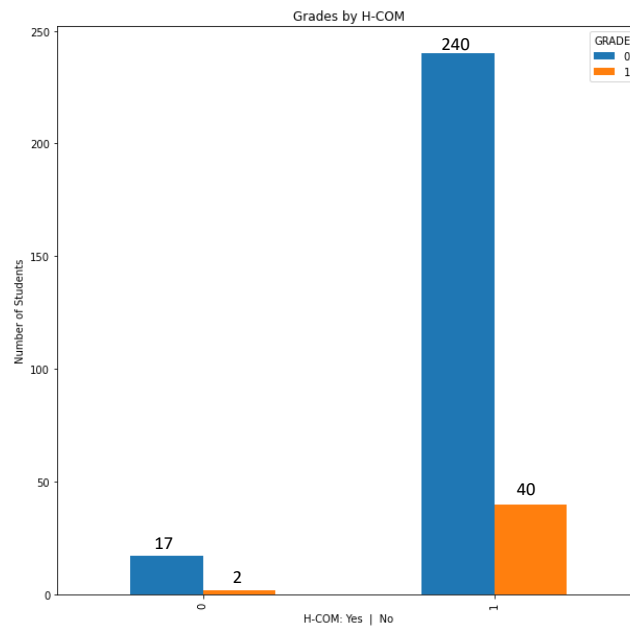
จากภาพที่ 3.4 เมื่อนำข้อมูลคุณลักษณะทั้ง 32 ตัวมาจัดแสดงให้อยู่ในรูปกราฟ สามารถวิเคราะห์เบื้องต้นได้ว่า ผู้กลุ่มตัวอย่างที่ตอบแบบสอบถามส่วนใหญ่เป็นผู้หญิงมากกว่าผู้ชาย และกลุ่มตัวอย่างเกือบทุกคนมีคอมพิวเตอร์และอินเทอร์เน็ตใช้ส่วนตัว รวมถึงไม่ได้เรียนพิเศษวิชาภาษาอังกฤษ เป็นต้น



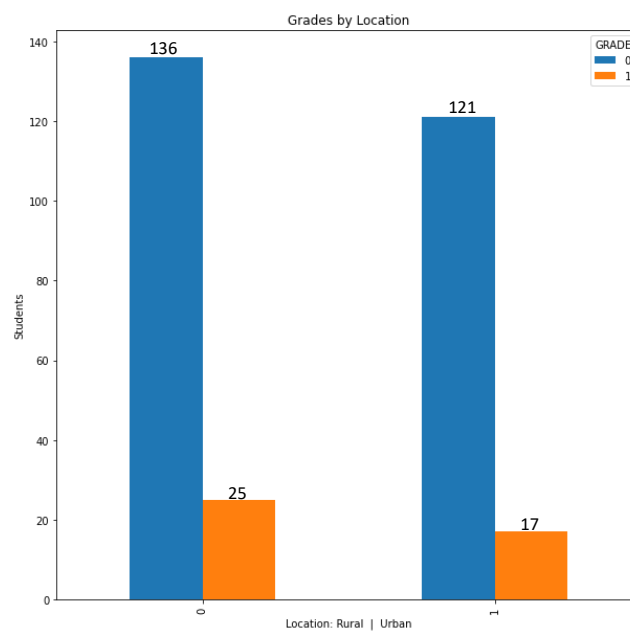
ภาพที่ 3.5 กราฟภาพเปรียบเทียบระหว่างเพศ (GENDER) กับคุณลักษณะเป้าหมาย



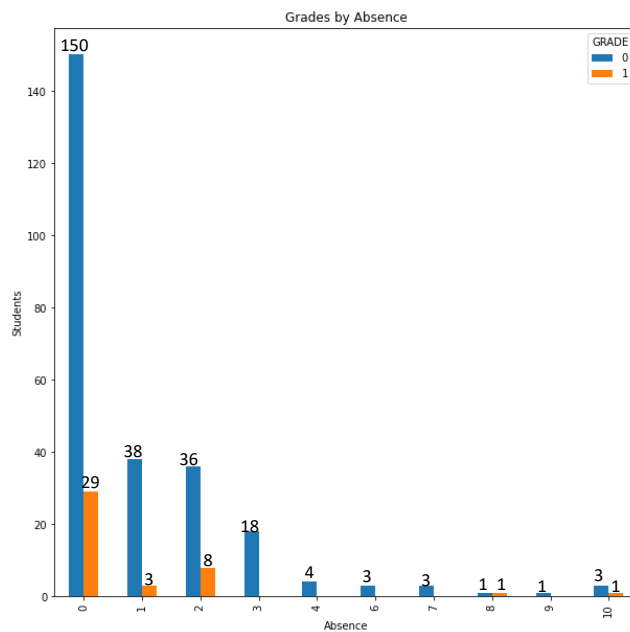
ภาพที่ 3.6 กราฟภาพเปรียบเทียบระหว่างอายุ (AGE) กับคุณลักษณะเป้าหมาย



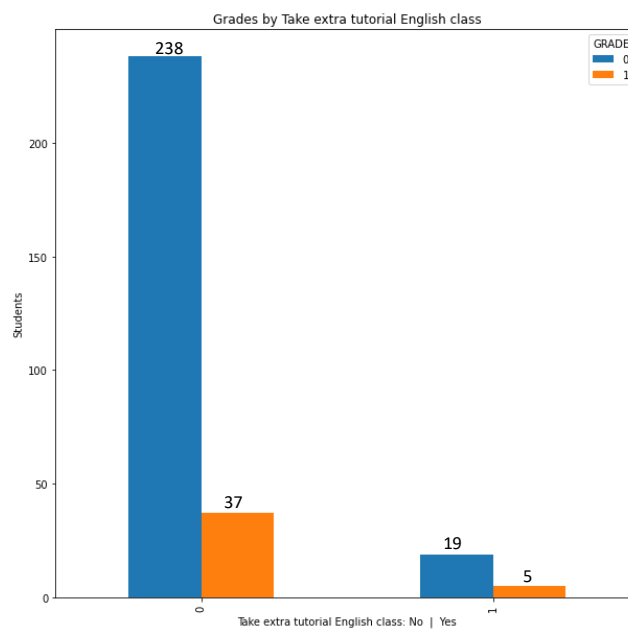
ภาพที่ 3.7 กราฟภาพเปรียบเทียบระหว่างมีคอมพิวเตอร์และอินเทอร์เน็ตใช้ส่วนตัว (H-COM) กับคุณลักษณะเป้าหมาย



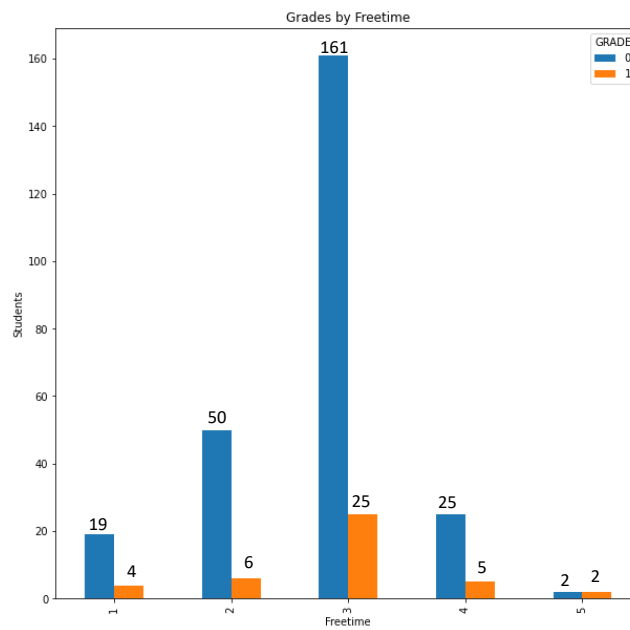
ภาพที่ 3.8 กราฟภาพเปรียบเทียบระหว่างลักษณะที่ตั้งของบ้านที่อยู่อาศัย (LOC) กับคุณลักษณะเป้าหมาย



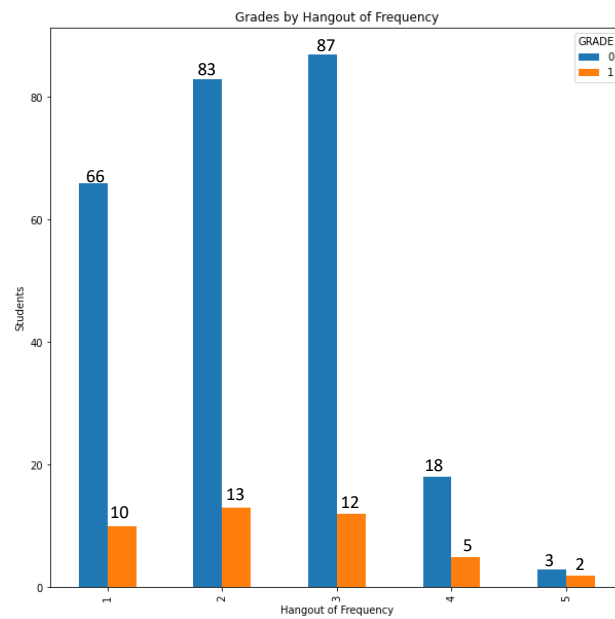
ภาพที่ 3.9 กราฟภาพเปรียบเทียบระหว่างจำนวนการขาดเรียนวิชาภาษาอังกฤษ (ABSENCE) กับคุณลักษณะเป้าหมาย



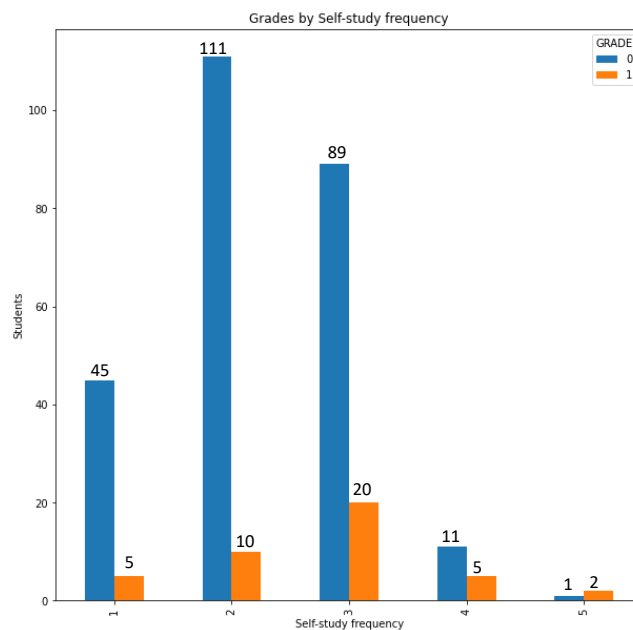
ภาพที่ 3.10 กราฟภาพเปรียบเทียบระหว่างเรียนพิเศษวิชาภาษาอังกฤษ (S-ENG) กับคุณลักษณะเป้าหมาย



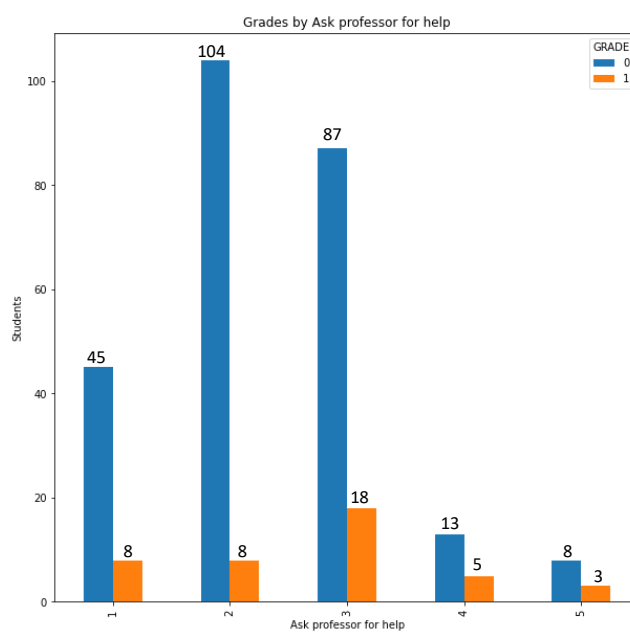
ภาพที่ 3.11 กราฟภาพเปรียบเทียบระหว่างเวลาว่างจากการเรียน (F-TIME) กับคุณลักษณะเป้าหมาย



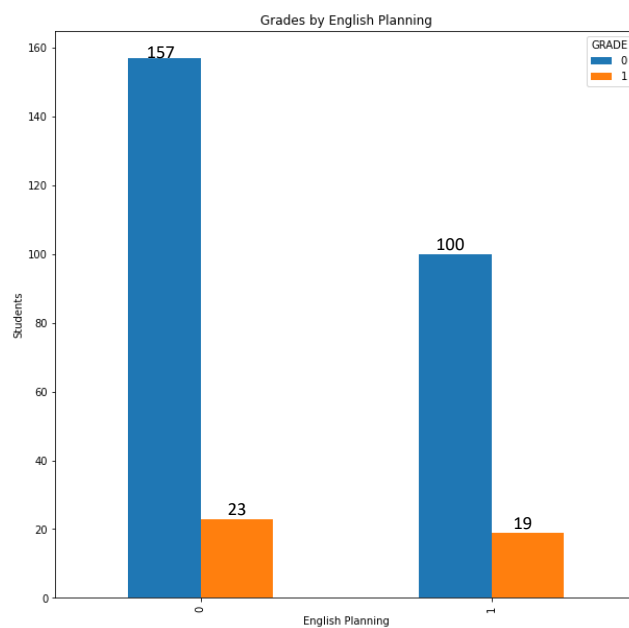
ภาพที่ 3.12 กราฟภาพเปรียบเทียบระหว่างเวลาสำหรับออกไปเที่ยวกับเพื่อน (H-EREQ) กับคุณลักษณะเป้าหมาย



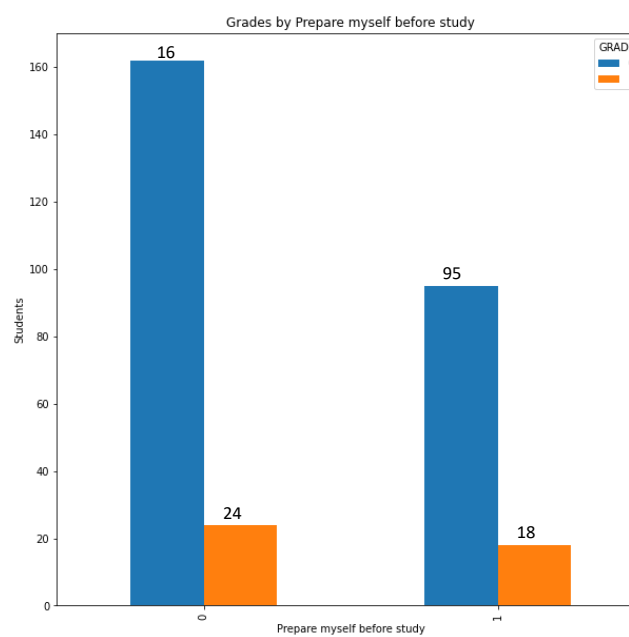
ภาพที่ 3.13 กราฟภาพเปรียบเทียบระหว่างเวลาว่างสำหรับฝึกฝนภาษาอังกฤษด้วยตนเอง (S-STUDY) กับคุณลักษณะเป้าหมาย



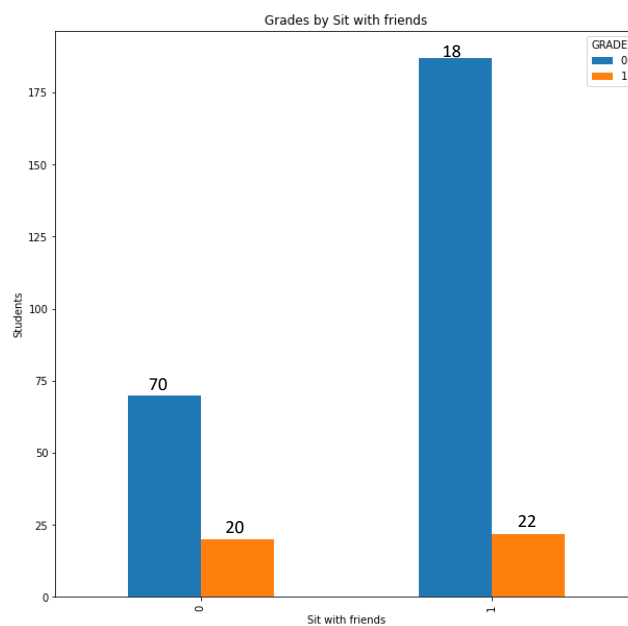
ภาพที่ 3.14 กราฟภาพเปรียบเทียบระหว่างถามอาจารย์ทันทีเมื่อไม่เข้าใจเนื้อหาที่เรียน (A-PROF) กับคุณลักษณะเป้าหมาย



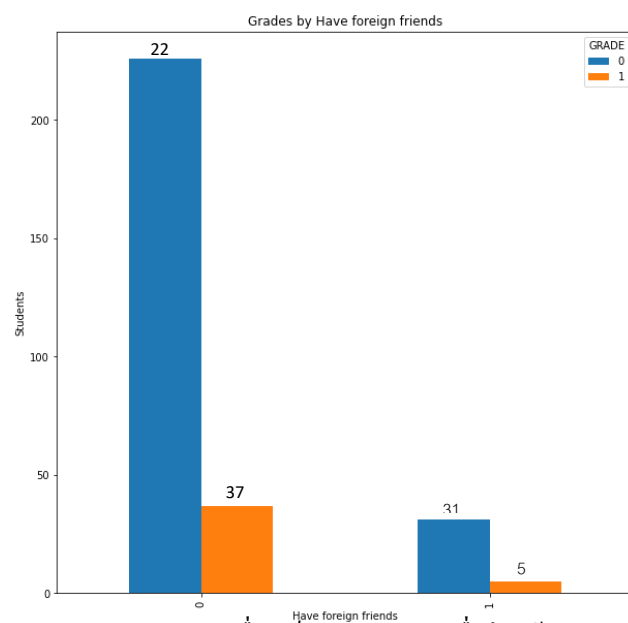
ภาพที่ 3.15 กราฟภาพเปรียบเทียบระหว่างวางแผนการเรียนภาษาอังกฤษ (E-PLAN) กับคุณลักษณะเป้าหมาย



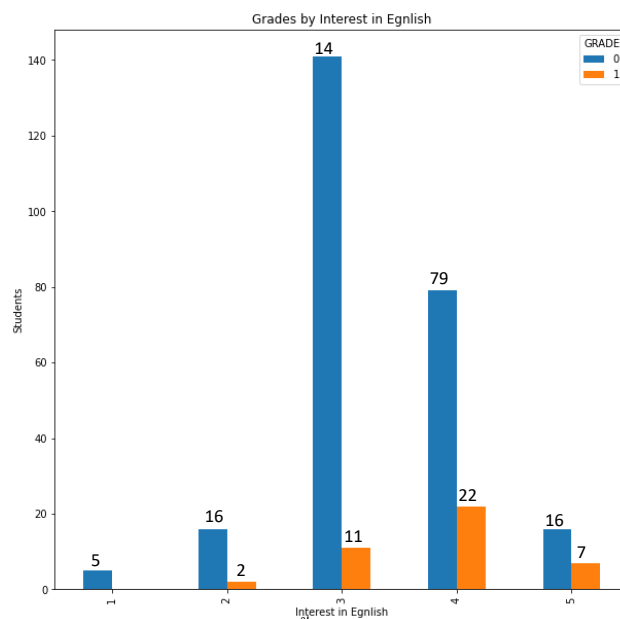
ภาพที่ 3.16 กราฟภาพเปรียบเทียบระหว่างเตรียมตัวก่อนเข้าเรียนวิชาภาษาอังกฤษ (P-SELF) กับคุณลักษณะเป้าหมาย



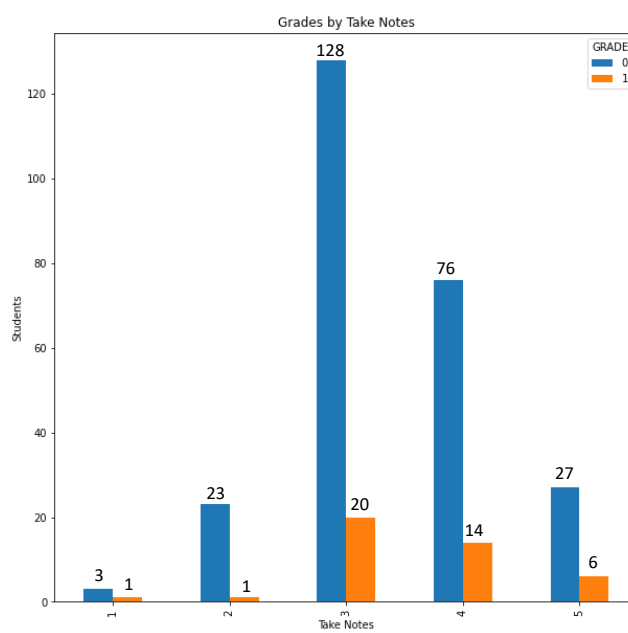
ภาพที่ 3.17 กราฟภาพเปรียบเทียบระหว่างชอบนั่งใกล้ๆ เพื่อนที่ตั้งใจเรียน (S-FRIEND) กับคุณลักษณะเป้าหมาย



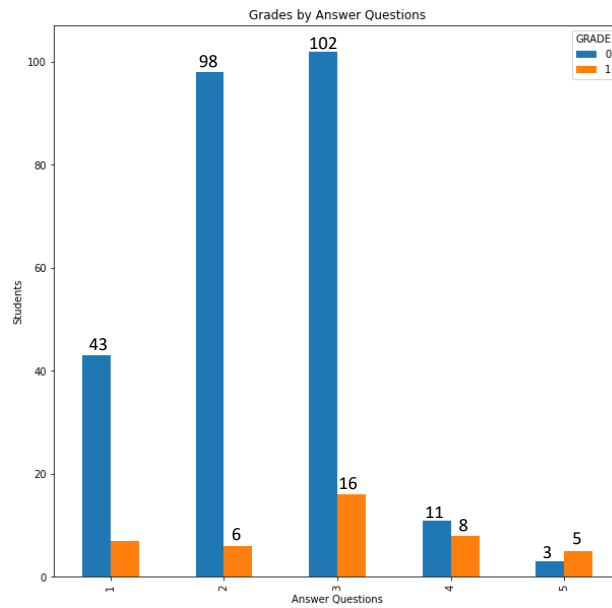
ภาพที่ 3.18 กราฟภาพเปรียบเทียบระหว่างมีเพื่อนเป็นชาวต่างชาติ เพื่อฝึกใช้ทักษะภาษาอังกฤษ (F-FRIEND) กับคุณลักษณะเป้าหมาย



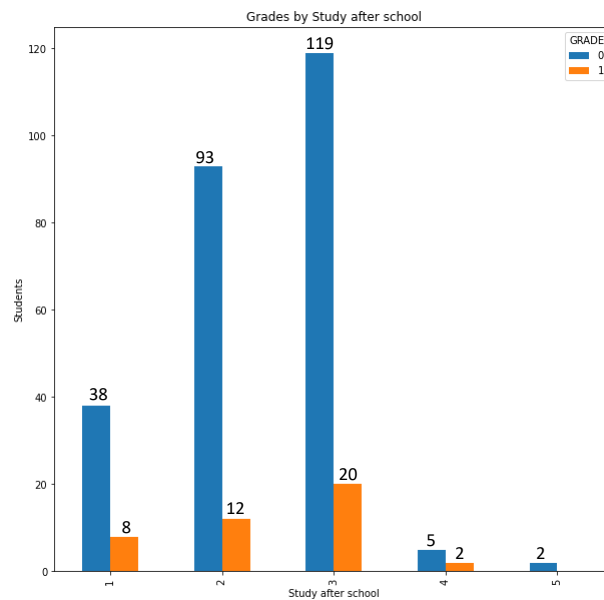
ภาพที่ 3.19 กราฟภาพเปรียบเทียบระหว่างสนใจและตั้งใจเรียนวิชาภาษาอังกฤษในขณะที่อาจารย์สอน (I-ENG) กับคุณลักษณะเป้าหมาย



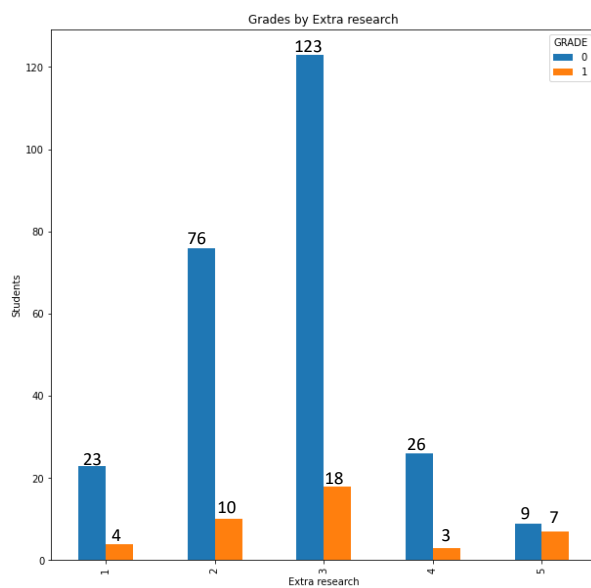
ภาพที่ 3.20 กราฟภาพเปรียบเทียบระหว่างจดบันทึกตามที่อาจารย์สอน (T-NOTES) กับคุณลักษณะเป้าหมาย



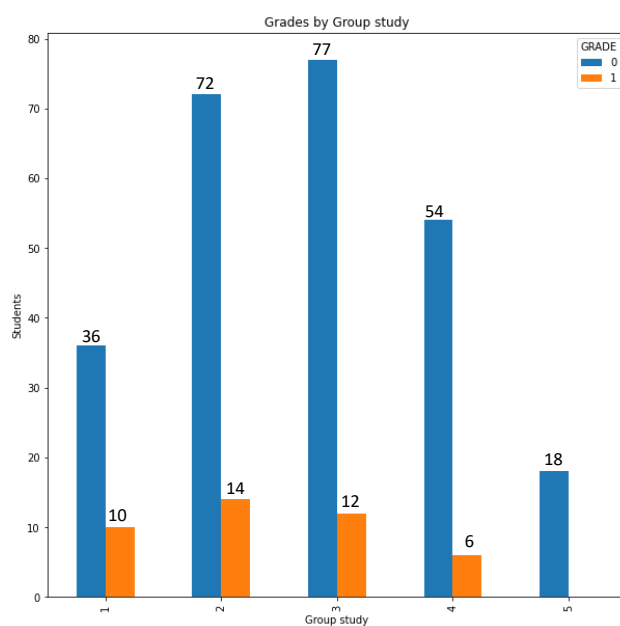
ภาพที่ 3.21 กราฟภาพเปรียบเทียบระหว่างตอบคำถามในขณะเรียนภาษาอังกฤษ (A-QUE) กับคุณลักษณะเป้าหมาย



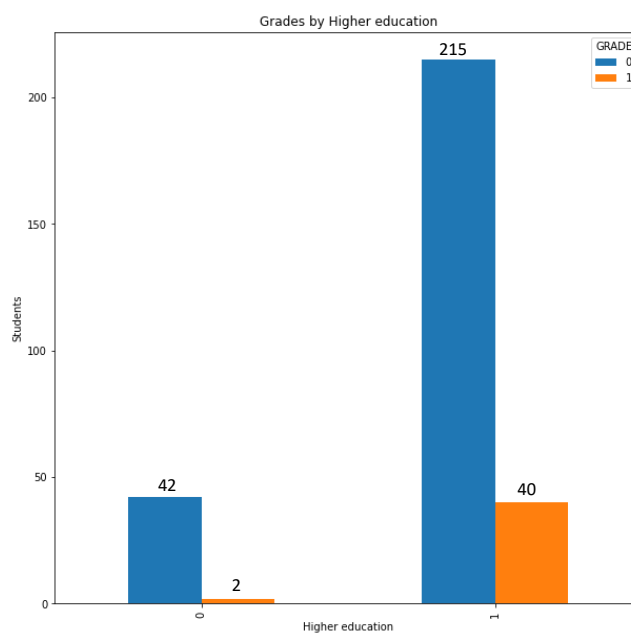
ภาพที่ 3.22 กราฟภาพเปรียบเทียบระหว่างทบทวนบทเรียนภาษาอังกฤษหลังเลิกเรียน (A-SCHOOL) กับคุณลักษณะเป้าหมาย



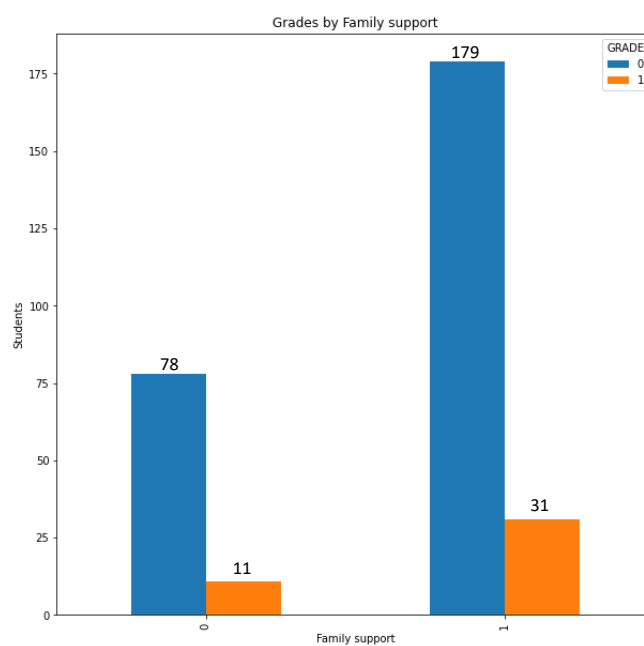
ภาพที่ 3.23 กราฟภาพเปรียบเทียบระหว่างศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษนอกเหนือจากที่อาจารย์สอน (E-RES) กับคุณลักษณะเป้าหมาย



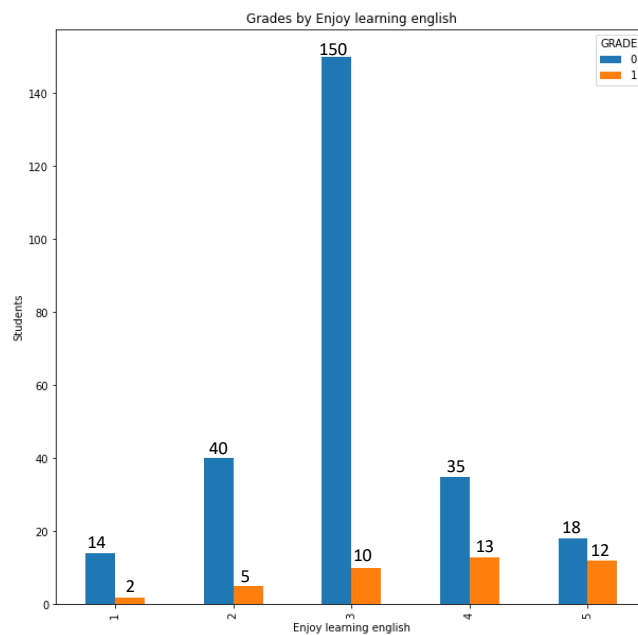
ภาพที่ 3.24 กราฟภาพเปรียบเทียบระหว่างจับกลุ่มทบทวนบทเรียนภาษาอังกฤษกับเพื่อนเมื่อใกล้สอบ (G-STUDY) กับคุณลักษณะเป้าหมาย



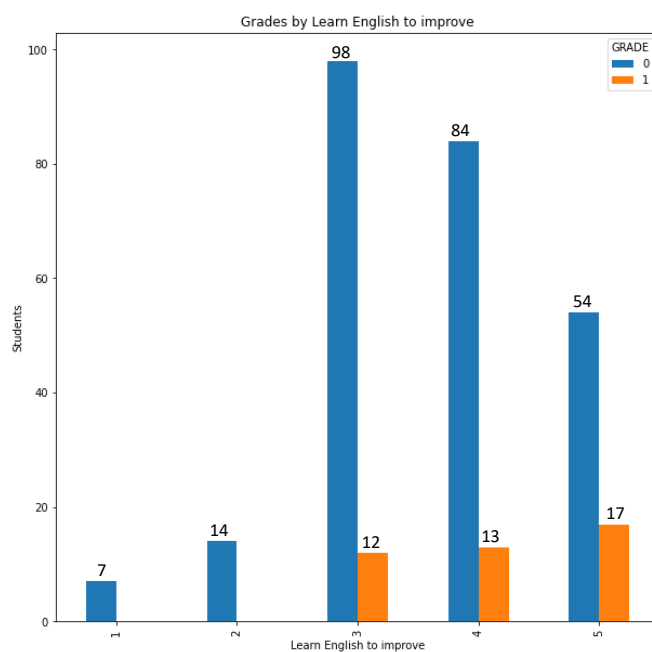
ภาพที่ 3.25 กราฟภาพเปรียบเทียบระหว่างความต้องการศึกษาในระดับที่สูงขึ้น (H-EDU) กับคุณลักษณะเป้าหมาย



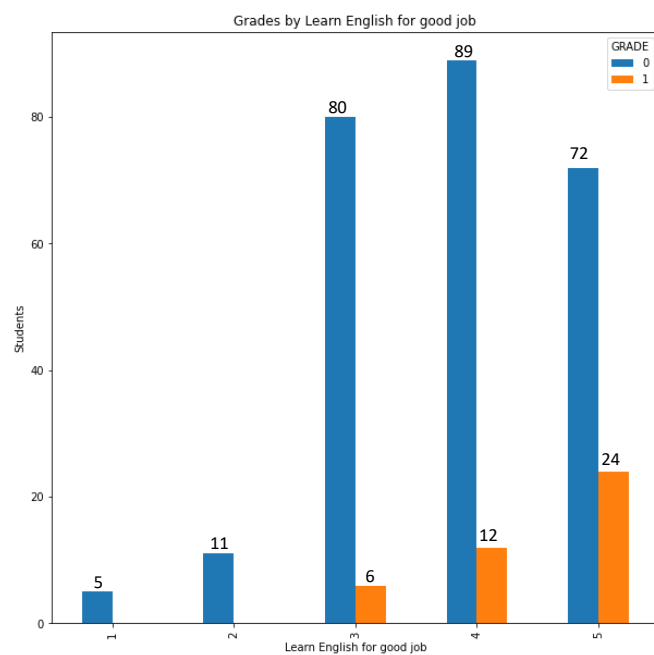
ภาพที่ 3.26 กราฟภาพเปรียบเทียบระหว่างการสนับสนุนด้านการเรียนภาษาอังกฤษจากครอบครัว (F-SUPP) กับคุณลักษณะเป้าหมาย



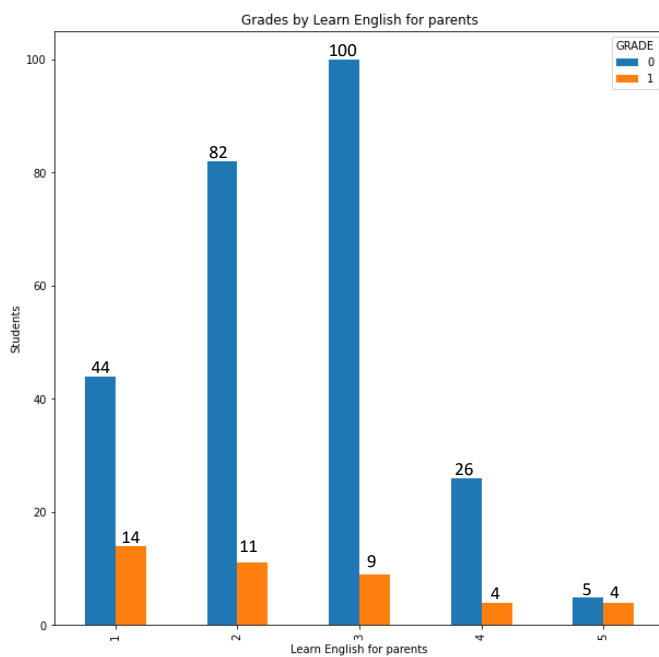
ภาพที่ 3.27 กราฟภาพเปรียบเทียบระหว่างสนุกกับการเรียนภาษาอังกฤษ (E-ENG) กับคุณลักษณะเป้าหมาย



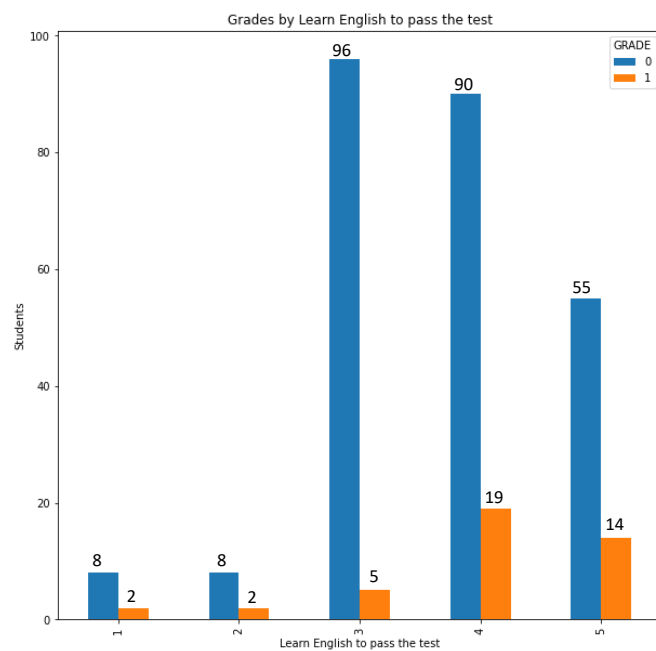
ภาพที่ 3.28 กราฟภาพเปรียบเทียบระหว่างสนใจเรียนภาษาอังกฤษเพื่อพัฒนาตนเอง (L-IMPROV) กับคุณลักษณะเป้าหมาย



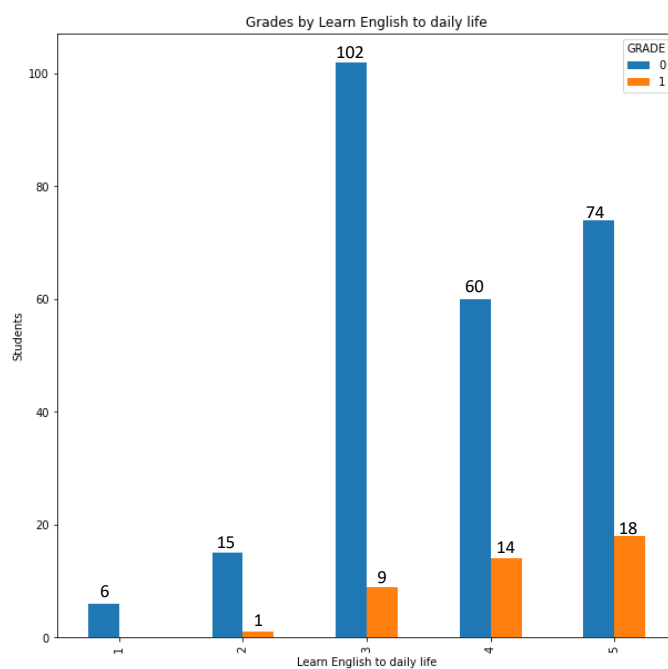
ภาพที่ 3.29 กราฟภาพเปรียบเทียบระหว่างสนใจเรียนภาษาอังกฤษ เพื่อหางานที่ดีในอนาคต (L-JOB)
กับคุณลักษณะเป้าหมาย



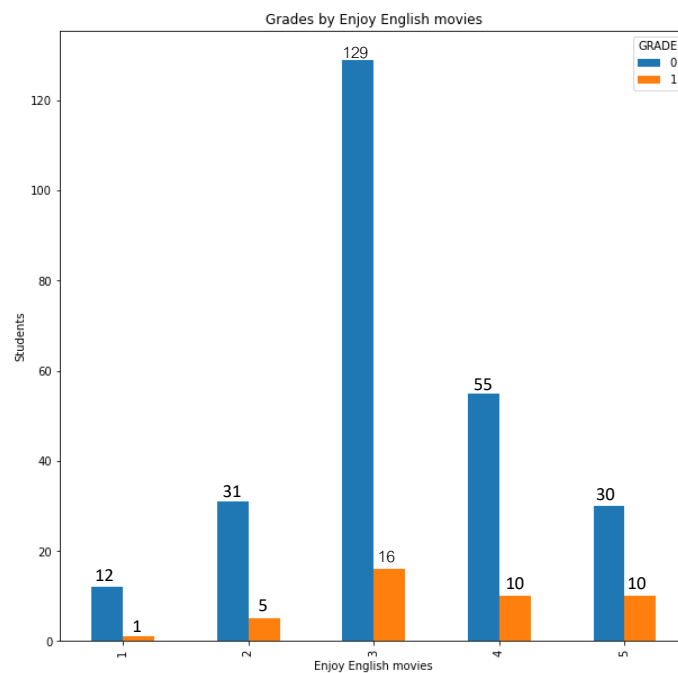
ภาพที่ 3.30 กราฟภาพเปรียบเทียบระหว่างสนใจเรียนภาษาอังกฤษ เพราะเป็นความต้องการของพ่อแม่ (L-PAREN)
กับคุณลักษณะเป้าหมาย



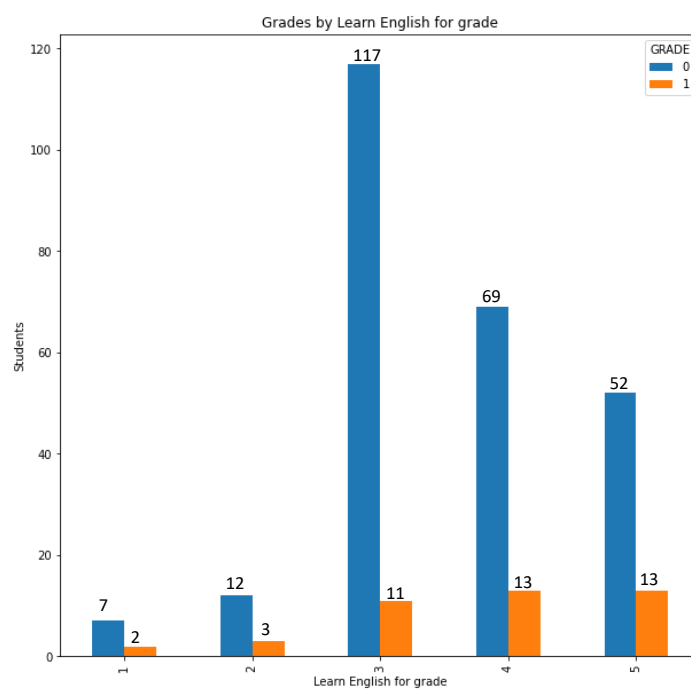
ภาพที่ 3.31 กราฟภาพเปรียบเทียบระหว่างสนใจเรียนภาษาอังกฤษ เพื่อต้องการสอบผ่าน (L-PASS) กับคุณลักษณะเป้าหมาย



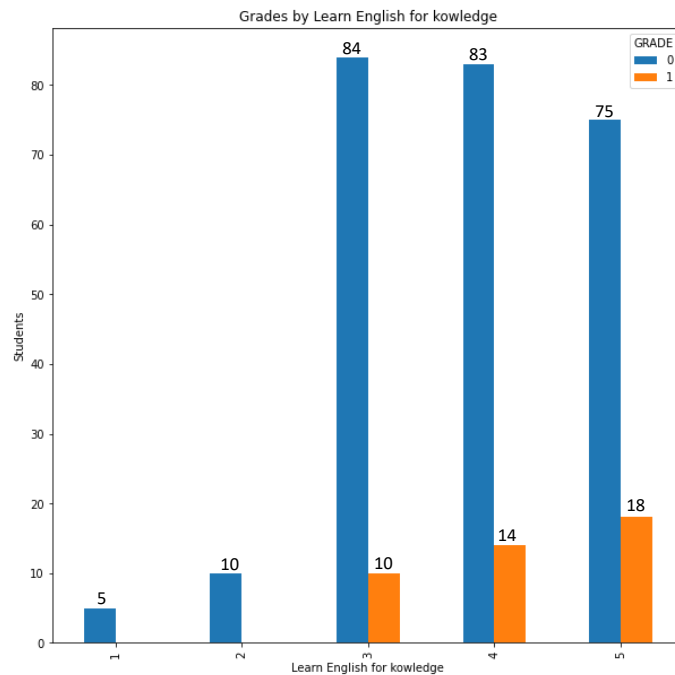
ภาพที่ 3.32 กราฟภาพเปรียบเทียบระหว่างสนใจเรียนภาษาอังกฤษ เพื่อใช้สื่อสารในชีวิตประจำวัน (L-DAILY) กับคุณลักษณะเป้าหมาย



ภาพที่ 3.33 กราฟภาพเปรียบเทียบระหว่างความสุขในการอ่านหนังสือ ภาพยนตร์เป็นภาษาอังกฤษ (E-MOV) กับคุณลักษณะเป้าหมาย

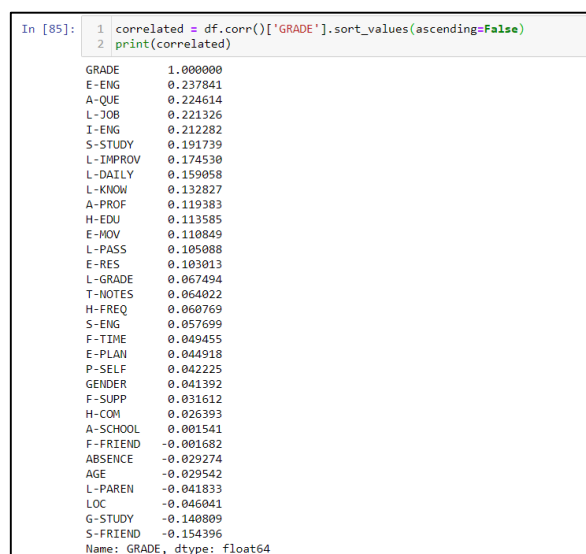


ภาพที่ 3.34 กราฟภาพเปรียบเทียบระหว่างการเรียนรู้ภาษาอังกฤษมีความสำคัญ เพราะภาษาอังกฤษช่วยให้มีความรู้มากขึ้น (E-KNOW) กับคุณลักษณะเป้าหมาย



ภาพที่ 3.35 กราฟภาพเปรียบเทียบระหว่างการเรียนภาษาอังกฤษมีความสำคัญ เพราะต้องการได้เกรดที่ดี (E-GRADE) กับคุณลักษณะเป้าหมาย

4) การคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ เป็นขั้นตอนค้นหาความสัมพันธ์ระหว่างข้อมูลคุณลักษณะต่างๆ ภายในชุดข้อมูล เพื่อคัดเลือกเฉพาะคุณลักษณะที่สำคัญมาเป็นตัวแทนของชุดข้อมูลทั้งหมดสำหรับใช้วิเคราะห์และสร้างตัวแบบจำลอง ซึ่งการลดจำนวนคุณลักษณะที่ไม่เกี่ยวข้องหรือไม่มีความสัมพันธ์กับคุณลักษณะเป้าหมาย สามารถลดความผิดพลาด และช่วยลดเวลาการทำงานของแบบจำลองได้ (สุรพงศ์, 2557) โดยการศึกษาใช้วิธีการคำนวณค่าสหสัมพันธ์แบบเพียร์สัน (Pearson product moment Correlation Coefficient) ในการหาความสัมพันธ์ระหว่างคุณลักษณะ 2 ตัว (คุณลักษณะอื่นๆ กับคุณลักษณะเป้าหมาย) ดังภาพ 3.36



ภาพที่ 3.36 ภาพคำนวณหาความสัมพันธ์ ด้วยวิธีสหสัมพันธ์แบบเพียร์สันในโปรแกรม Jupyter

จากภาพ 3.36 จะเห็นได้ว่า คุณลักษณะสันทกกับการเรียนภาษาอังกฤษ (E-ENG) คุณลักษณะตอบคำถามในขณะเรียนภาษาอังกฤษ (A-QUE) และคุณลักษณะสนใจเรียนภาษาอังกฤษเพื่อหางานที่ดีในอนาคต (L-JOB) มีความสัมพันธ์เชิงบวกอย่างมากกับคุณลักษณะเป้าหมาย ในขณะที่คุณลักษณะชอบนั่งใกล้ๆ เพื่อนที่ตั้งใจเรียน (S-FRIEND) และคุณลักษณะจับกลุ่มทบทวนบทเรียนภาษาอังกฤษกับเพื่อนเมื่อใกล้สอบ (G-STUDY) มีความสัมพันธ์เชิงลบกับคุณลักษณะเป้าหมาย อย่างไรก็ตาม ผู้วิจัยทำการคัดเลือกเฉพาะคุณลักษณะที่มีความสัมพันธ์กับคุณลักษณะเป้าหมายมากที่สุด โดยเลือกตัวแปรที่มีค่าความสัมพันธ์มากกว่า 0.1 มาใช้ในการวิเคราะห์ ทั้งนี้ เพื่อให้ได้ข้อมูลที่มีคุณภาพมากที่สุด และลดความผิดพลาดที่อาจเกิดจากการมีตัวแปรในการวิเคราะห์มากเกินไป แสดงดังภาพที่ 3.37

```
In [119]: 1 # Consider correlations only with the target variable
2 cor = df.corr()
3 cor_target = abs(cor['GRADE'])
4 correlated = cor_target[cor_target>0.1]
5 most_correlated = df.loc[:,correlated.index]
6 print(correlated)

GRADE      1.000000
S-STUDY     0.191739
A-PROF      0.119383
S-FRIEND    0.154396
I-ENG       0.212282
A-QUE       0.224614
E-RES       0.103013
G-STUDY     0.140809
H-EDU       0.113585
E-ENG       0.237841
L-IMPROV    0.174530
L-JOB       0.221326
L-PASS      0.105088
L-DAILY     0.159058
E-MOV       0.110849
L-KNOW      0.132827
Name: GRADE, dtype: float64
```

ภาพที่ 3.37 ข้อมูลคุณลักษณะที่มีถูกคัดเลือกลำมาใช้สร้างตัวแบบจำลองในโปรแกรม Jupyter

3.4.3 สร้างตัวแบบจำลอง

ขั้นตอนการเลือกใช้อัลกอริทึมเทคนิคเหมืองข้อมูลในการ สร้างตัวแบบจำลองคุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนวิชาภาษาอังกฤษของนักเรียนระดับปริญญาตรี โดยผู้วิจัยใช้ไลบรารี Sklearn ของ Python ทำการกำหนดสัดส่วนชุดข้อมูลกลุ่มตัวอย่างจำนวน 299 ตัวอย่าง ออกเป็น 2 ส่วน แบ่งเป็นชุดข้อมูลเรียนรู้ 70% สำหรับทำการฝึกสอน (จำนวน 209 ตัวอย่าง) และชุดข้อมูลทดสอบ 30% สำหรับใช้ทดสอบตัวแบบจำลอง (จำนวน 90 ตัวอย่าง) แสดงดังภาพที่ 3.38 เนื่องจากการกำหนดอัตราส่วนพื้นฐานของการทดสอบข้อมูล (Alls, Karakurt, และ Melli, 2000) จากนั้นเลือกอัลกอริทึมเทคนิค Decision Tree และเทคนิค Random Forest ทำการทดลองและปรับพารามิเตอร์ และใช้ชุดข้อมูลจำนวน 15 คุณลักษณะ (ตัวแปร X) ที่ได้จากการหาค่าความสัมพันธ์ เข้าสู่กระบวนการวิเคราะห์ พร้อมกำหนดให้เกรดวิชาภาษาอังกฤษระดับพื้นฐานของนักศึกษาเป็นคุณลักษณะเป้าหมาย (ตัวแปร Y) เพื่อทำการจำแนกประสิทธิภาพการเรียนวิชาภาษาอังกฤษ โดยกำหนดให้นักศึกษาที่มีเกรดเอเป็นกลุ่มนักศึกษาที่มีประสิทธิภาพการเรียนวิชาภาษาอังกฤษ และนักศึกษาที่ได้ผลการเรียนอื่นๆ เป็นกลุ่มนักศึกษาที่มีผลการเรียนวิชาภาษาอังกฤษระดับพื้นฐานทั่วไป

```
Number of rows,columns in X_train dataset: (209, 15)
Number of rows in y_train dataset: (209,)
Number of rows,columns in X_test dataset: (90, 15)
Number of rows in y_test dataset: (90,)
```

ภาพที่ 3.38 การแบ่งสัดส่วนชุดข้อมูลในโปรแกรม Jupyter

3.4.4 ประเมินประสิทธิภาพแบบจำลอง

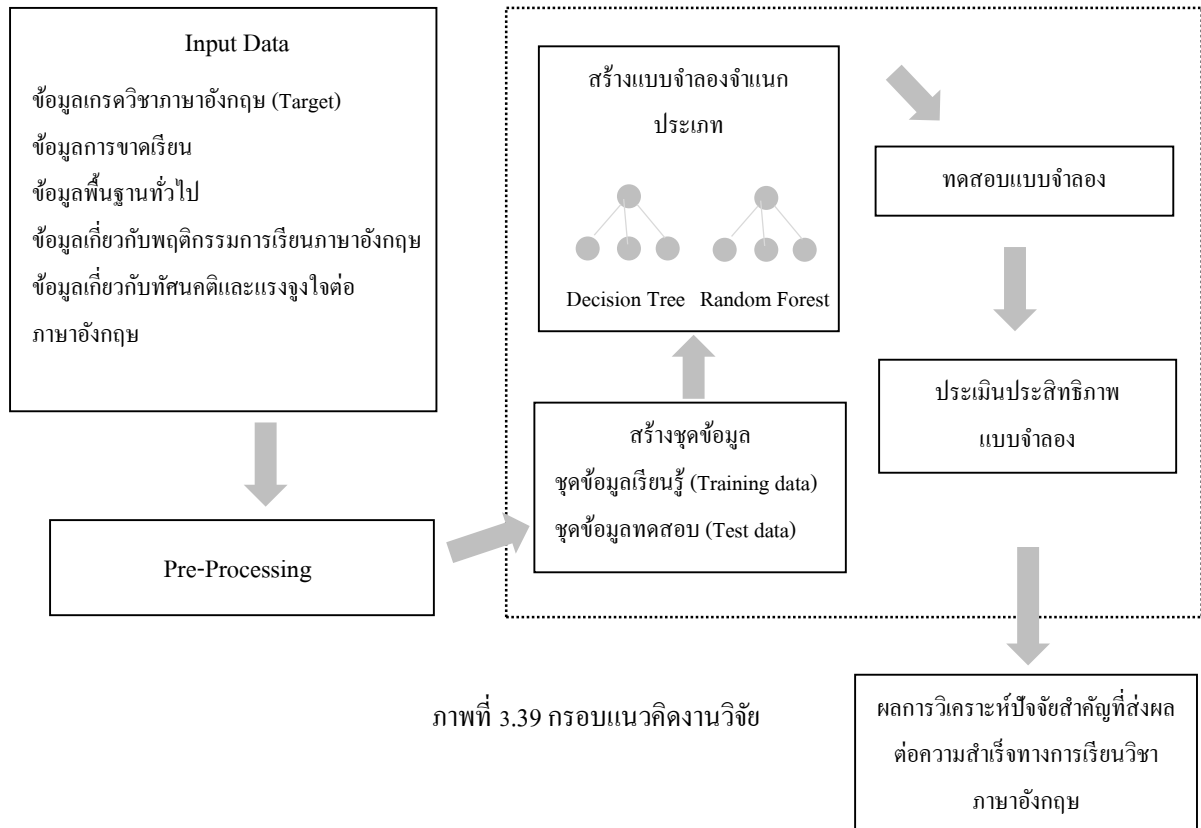
ขั้นตอนการทดสอบประสิทธิภาพของตัวแบบจำลอง เพื่อพิจารณาตัวแบบจำลองข้อมูลที่เหมาะสมที่สุดระหว่างเทคนิค Decision Tree และเทคนิค Random Forest ซึ่งผู้วิจัยใช้วิธี k-fold cross validation ในการแบ่งชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบเป็นจำนวน 10 ส่วนเท่าๆ กัน ข้อมูล 9 ส่วนจะถูกใช้เป็นข้อมูลเรียนรู้ และข้อมูล 1 ส่วนจะใช้สำหรับทดสอบ ด้วยอัตราส่วน 70:30 โดยแต่ละส่วนมาจากการสุ่ม ทำการเรียนรู้และทดสอบวนซ้ำจนครบ 10 ครั้ง วิธีนี้ใช้เพื่อเปรียบเทียบระหว่างแบบจำลองเพื่อเลือกแบบจำลองที่มีความเหมาะสมกับข้อมูล พิจารณาจากตัวชี้วัดประสิทธิภาพแบบจำลองสำหรับอัลกอริทึมจำแนกประเภท นอกจากนี้ผู้วิจัยใช้ Confusion Matrix เครื่องมือการประเมินผลลัพธ์สรุปประสิทธิภาพของแบบจำลองการจำแนกประเภท โดยในการประเมินประสิทธิภาพของตัวแบบจำลองด้วยการวัดค่าความแม่นยำของการจำแนกข้อมูล (Accuracy) วัดค่าความถูกต้องของการทำนาย (Precision) วัดค่าความระลึก (Recall) และวัดค่าความถ่วงดุล (F1 score) เพื่อหาความแม่นยำของแบบจำลองที่มีค่ามากที่สุด

3.4.5 การนำไปใช้ประโยชน์

ผลการวิจัยที่ได้จากการวิเคราะห์จะทำให้ทราบถึงคุณลักษณะหรือพฤติกรรมของนักศึกษาระดับปริญญาตรีที่มีประสิทธิภาพในการเรียนวิชาภาษาอังกฤษ ซึ่งข้อมูลคุณลักษณะที่ปรากฏจากการวิเคราะห์ด้วยเทคนิคเหมืองข้อมูลสามารถมาประกอบวางแผนการเรียนภาษาของนักศึกษา รวมถึงจะเป็นประโยชน์ในการวิเคราะห์ต่อยอดองค์ความรู้ถึงวิธีการเพิ่มประสิทธิภาพการเรียนรายวิชาอื่นๆ ของนักศึกษา

3.4 กรอบแนวคิดการวิจัย

จากแนวคิดและทฤษฎีที่เกี่ยวข้อง สามารถกำหนดกรอบวิจัยของงานวิจัยการศึกษา วิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักเรียนระดับปริญญาตรี โดยใช้เทคนิคเหมืองข้อมูล แสดงดังภาพที่ 3.39



ภาพที่ 3.39 กรอบแนวคิดงานวิจัย

บทที่ 4

ผลการวิจัย

ในการศึกษาเรื่องการวิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักเรียนระดับปริญญาตรี โดยใช้เทคนิคเหมืองข้อมูล ผู้วิจัยได้นำเอาโปรแกรม Jupyter และภาษา python ไลบรารี sklearn มาเป็นเครื่องมือในการสร้างแบบจำลองด้วยเทคนิค Decision Tree และเทคนิค Random Forests โดยใช้กลุ่มตัวอย่างจากนักเรียนระดับปริญญาตรี คณะมนุษยศาสตร์ มหาวิทยาลัยบูรพา ปีการศึกษา พ.ศ.2561 - 2563 จำนวน 299 คน และข้อมูล 32 คุณลักษณะที่ได้จากแบบสอบถามปลายปิด การศึกษานี้วัดประสิทธิภาพแบบจำลองด้วยค่าความแม่นยำ (Accuracy) ค่าความถูกต้อง (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F1 score) ซึ่งผลการวิจัยสามารถสรุปแบ่งเป็น 2 หัวข้อดังต่อไปนี้

4.1 ผลการพัฒนาและเปรียบเทียบแบบจำลอง

จากการจัดเตรียมข้อมูลก่อนการประมวลผลแบบจำลอง ผู้วิจัยได้ทำการคัดเลือกตัวแปรคุณลักษณะที่มีความสัมพันธ์กับประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษ พิจารณาจากค่าสหสัมพันธ์ที่มีค่ามากกว่า 0.1 โดยผู้วิจัยกำหนดให้เกรดวิชาภาษาอังกฤษระดับพื้นฐานของนักศึกษาเป็นคุณลักษณะเป้าหมายในการจำแนกประเภทประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษา ผลการหาค่าสหสัมพันธ์สหสัมพันธ์ พบว่าจำนวน 15 คุณลักษณะมีความสัมพันธ์กับเกรดวิชาภาษาอังกฤษระดับพื้นฐานของนักศึกษา ในการศึกษาครั้งนี้ผู้วิจัยจึงใช้คุณลักษณะทั้ง 15 ตัว ในการทดสอบการเทรนข้อมูลเพื่อหาค่าตัวแปรที่แสดงคุณลักษณะของผู้ที่มีแนวโน้มมีทักษะภาษาอังกฤษด้วยเทคนิค Decision Tree และเทคนิค Random Forests

ผลการพัฒนาแบบจำลองด้วยเทคนิค Decision Tree และเทคนิค Random Forests ประเมินได้จากค่าความแม่นยำ (Accuracy) ของแบบจำลองที่สามารถจำแนกประเภทประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษได้อย่างถูกต้อง โดยการศึกษาครั้งนี้ ผู้วิจัยใช้วิธี k-fold cross validation กำหนดสัดส่วนข้อมูล 10 ส่วน สำหรับสร้างและทดสอบประสิทธิภาพของแบบจำลอง และกำหนดอัตราส่วนข้อมูลที่ 70:30 ทำให้ชุดย่อยที่แตกต่างกันจำนวน 1 ส่วนจะได้รับการทดสอบด้วย 30% ของข้อมูล วิธีนี้จะทำให้การสร้างและทดสอบซ้ำจำนวนครบจำนวน 10 ครั้ง และคำนวณความแม่นยำจากค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของผลทดสอบทั้งหมด ทั้งนี้ ผลจากการสร้างและทดสอบแบบจำลองแสดงให้เห็นว่า เทคนิค Random Forest ให้ค่าความแม่นยำในการทำนายประเภทประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาเท่ากับ 85.14% ในขณะที่เทคนิค Decision Tree ค่าความแม่นยำในการทำนาย 78% แสดงดังตารางที่ 4.1

ตารางที่ 4.1 แสดงผลการทำสอบแบบจำลองด้วยวิธี k-fold cross validation

Algorithm	Accuracy Mean	Accuracy STD
Random Forest	85.14	6.26
Decision Tree Classifier	78.00	5.26

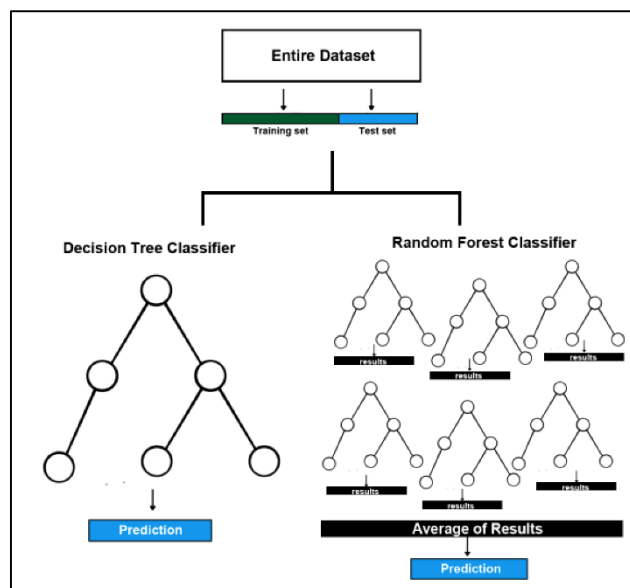
นอกจากนี้ ผู้วิจัยได้ทำการตรวจสอบความแม่นยำอีกครั้ง โดยการหาค่าความแม่นยำจากสัดส่วนการจำแนกประเภทชุดข้อมูลการเรียนรู้ (Training) 70% และชุดข้อมูลการทดสอบ (Testing) 30% เพื่อบ่งบอกระดับความแม่นยำในการจำแนกประเภทข้อมูล วิธีนี้เป็นการวัดค่าความแม่นยำรูปแบบหนึ่งของ Confusion Metrix โดยคำนวณจากสัดส่วนจำนวนที่

แบบจำลองจำแนกประเภทถูกต้องกับจำนวนข้อมูลทั้งหมด พบว่าเทคนิค Random Forest ยังมีค่าความแม่นยำสูงอยู่เช่นเดิม โดยมีค่าความแม่นยำเท่ากับ 0.84 โดยที่ค่าความแม่นยำของเทคนิค Decision Tress เท่ากับ 0.69 แสดงดังตารางที่ 4.2

ตารางที่ 4.2 แสดงผลการทดสอบประสิทธิภาพแบบจำลองด้วยการเปรียบเทียบ

Algorithm	Accuracy Score
Random Forest	0.84
Decision Tree Classifier	0.69

ความแตกต่างของความแม่นยำของแบบจำลองเกิดจากหลายปัจจัย รวมถึงค่าความแม่นยำถือเป็นตัวชี้วัดประสิทธิภาพที่ใช้บ่อยสำหรับอัลกอริทึมการเรียนรู้แบบจำแนกประเภท (Classification) ทั้งนี้ ข้อจำกัดหนึ่งของแบบจำลองที่พัฒนาด้วยเทคนิค Decision Tree คือ การทำงานโดยใช้ Decision Tree เพียงต้นเดียวในการสร้างแบบจำลอง ทำให้ประสิทธิภาพในการจำแนกและทำนายผลไม่สูงมากนัก เมื่อเปรียบเทียบกับเทคนิค Random Forest ซึ่งผู้วิจัยใช้จำนวน 75 Decision tree ในการสร้างและทดสอบแบบจำลอง รวมถึงสามารถปรับปรุงความถูกต้องของแบบจำลองโดยการปรับค่าพารามิเตอร์เพื่อให้เหมาะสมกับข้อมูลได้อีกด้วย จึงทำให้ค่าความแม่นยำมีประสิทธิภาพมากกว่า แสดงลักษณะการทำงานของการทำงานการจำแนกประเภทข้อมูลด้วยเทคนิค Decision Tree และเทคนิค Random Forest ดังภาพที่ 4.1

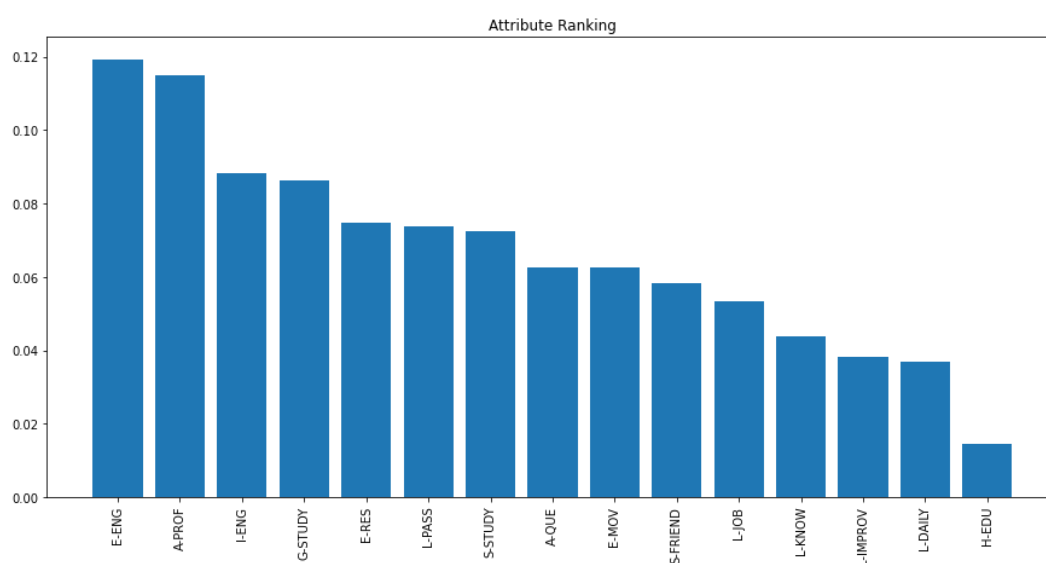


ภาพที่ 4.1 การทำงานการจำแนกประเภทข้อมูลด้วยเทคนิค Decision Tree และเทคนิค Random Forest

ดังนั้น จากผลการเปรียบเทียบประสิทธิภาพแบบจำลอง ทำให้ทราบว่าแบบจำลองที่พัฒนาด้วยเทคนิค Random Forest ให้ความแม่นยำในการนำมาวิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนรู้ภาษาอังกฤษของนักศึกษาระดับปริญญาตรีมากกว่าแบบจำลองที่พัฒนาด้วยเทคนิค Decision Tree ผู้วิจัยจึงเลือกแบบจำลองจากเทคนิค Random Forest มาใช้ในวิเคราะห์ 15 คุณลักษณะสำคัญ

4.2 ผลวิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี ด้วยเทคนิค Random Forest

จากการประมวลผลด้วยเทคนิค Random Forest เพื่อวิเคราะห์คุณลักษณะที่บ่งบอกถึงประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษา โดยใช้กลุ่มตัวอย่างนักศึกษาคณะภูมิสารสนเทศ มหาวิทยาลัยบูรพา จำนวน 299 คน ประกอบด้วย 15 คุณลักษณะ (ตัวแปร X) แสดงให้เห็นว่า ผลลัพธ์ของตัวแปรคุณลักษณะ 5 อันดับแรกที่มีความเกี่ยวข้องกับประสิทธิภาพของนักศึกษามากที่สุด คือ คุณลักษณะสนุกกับการเรียนภาษาอังกฤษ (E-ENG) คุณลักษณะถามอาจารย์ทันทีเมื่อไม่เข้าใจเนื้อหาที่เรียน (A-PROF) คุณลักษณะสนใจและตั้งใจเรียนวิชาภาษาอังกฤษในขณะที่อาจารย์สอน (I-ENG) คุณลักษณะจับกลุ่มทบทวนบทเรียนภาษาอังกฤษกับเพื่อนเมื่อใกล้สอบ (G-STUDY) และคุณลักษณะศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษนอกเหนือจากที่อาจารย์สอน (E-RES) แสดงให้เห็นดังภาพที่ 4.2



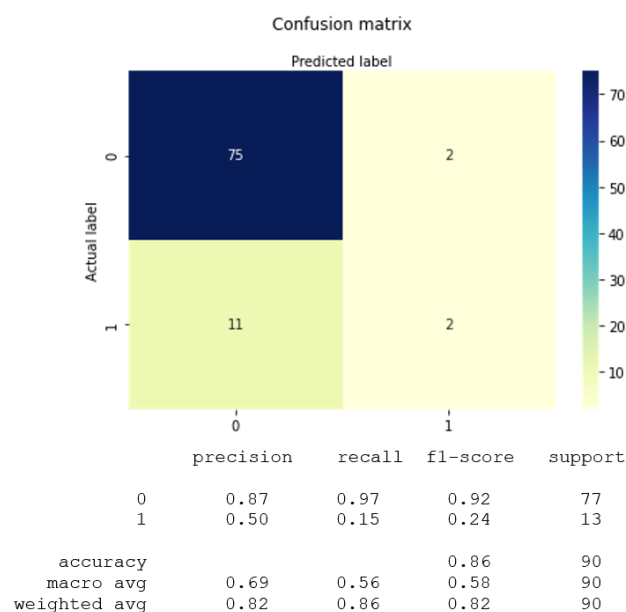
ภาพที่ 4.2 กราฟแสดงอันดับของตัวแปรคุณลักษณะที่บ่งบอกถึงประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษา

ตารางที่ 4.1 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของคุณลักษณะแต่ละตัวที่วัดโดยเทคนิค Random forest เพื่อชี้แจงการวิเคราะห์ข้อมูลเพิ่มเติม จะเห็นได้ว่า ผลลัพธ์ของตัวแปรคุณลักษณะ 5 อันดับแรกมีความเกี่ยวข้องกับประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาโดยค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง 0.07 - 0.1 ซึ่งบ่งชี้ว่าคุณลักษณะดังกล่าวมีแนวโน้มส่งผลต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษ

ตารางที่ 4.3 แสดงคุณลักษณะ 5 อันดับแรกที่สัมพันธ์กับประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษา

No	Variable	Coefficient
1	E-ENG	0.119383
2	A-PROF	0.114909
3	I-ENG	0.088360
4	G-STUDY	0.086430
5	E-RES	0.074748

ผลจากการตรวจสอบประสิทธิภาพของแบบจำลองที่พัฒนาด้วยเทคนิค Random Forest โดยใช้ Confusion Matrix ในการวัดประสิทธิภาพ พบว่า ค่าความแม่นยำ (Accuracy Score) ของแบบจำลองในการจำแนกประเภท เท่ากับ 86% และ วัดผลประสิทธิภาพด้วยค่าความถูกต้อง (Precision) ค่าความระลึก (Recall) และค่าถ่วงดุล (F-Score) ให้ค่าเท่ากับ 87%, 94% และ 92% ตามลำดับ แสดงให้เห็นมุมมองความสามารถของแบบจำลองในการจำแนกกลุ่มนักศึกษาที่มีผลการเรียนวิชา ภาษาอังกฤษระดับพื้นฐานทั่วไป (กำหนดเกรดวิชาภาษาอังกฤษระดับพื้นฐาน B, C, D, F) ในขณะที่มุมมองความสามารถ ของแบบจำลองในการจำแนกกลุ่มนักศึกษาที่มีประสิทธิภาพการเรียนวิชาภาษาอังกฤษ (กำหนดเกรดวิชาภาษาอังกฤษ ระดับพื้นฐาน A) ให้ค่าเท่ากับ 50% 15% และ 24% ตามลำดับ ซึ่งในส่วนนี้ชี้ให้เห็นว่า คุณลักษณะที่บ่งบอกถึง ประสิทธิภาพการเรียนภาษาอังกฤษ ข้อมูลอาจมีจำนวนน้อยเกินไป ภาพการวัดประสิทธิภาพด้วย Confusion Matrix แสดง ดังภาพที่ 4.3



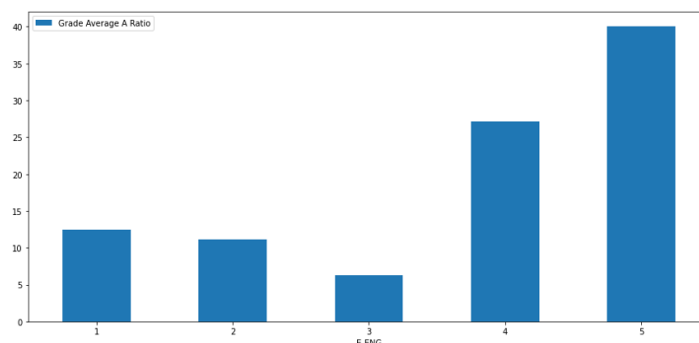
ภาพที่ 4.3 แสดงผลประสิทธิภาพของแบบจำลองโดยใช้ Confusion Matrix

อย่างไรก็ตาม คุณลักษณะ 5 อันดับแรกสามารถสะท้อนให้เห็นถึงคุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพของ การเรียนวิชาภาษาอังกฤษระดับพื้นฐานของนักศึกษา โดยพิจารณาจากผลการวิเคราะห์คุณลักษณะที่บ่งบอกถึง ประสิทธิภาพการเรียนวิชาภาษาอังกฤษของนักศึกษา เปรียบเทียบอัตราส่วนระหว่างกลุ่มนักศึกษาที่มีประสิทธิภาพการ เรียนวิชาภาษาอังกฤษ (ผลการเรียนเกรดเอวิชาภาษาอังกฤษพื้นฐาน) ดังนี้

- 4.2.1 ความรู้สึกสนุกสนานเมื่อได้เรียนภาษาอังกฤษ (E-ENG)
- 4.2.2 การถามอาจารย์ทันทีเมื่อไม่เข้าใจเนื้อหาที่เรียน (A-PROF)
- 4.2.3 ความสนใจและตั้งใจเรียนวิชาภาษาอังกฤษในขณะที่อาจารย์สอน (I-ENG)
- 4.2.4 การจับกลุ่มทบทวนบทเรียนภาษาอังกฤษกับเพื่อนเมื่อใกล้สอบ (G-STUDY)
- 4.2.5 การศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษนอกเหนือจากที่อาจารย์สอน (E-RES)

4.2.1 ความรู้สึกสนุกสนานเมื่อได้เรียนภาษาอังกฤษ (E-ENG)

จากข้อมูลการสำรวจความรู้สึกสนุกสนานเมื่อได้เรียนภาษาอังกฤษของนักศึกษาได้ผลการเรียนเกรดเอวิชาภาษาอังกฤษระดับพื้นฐานเมื่อเทียบกับจำนวนนักศึกษาทั้งหมดได้ผลลัพธ์แสดงดังภาพที่ 4.4

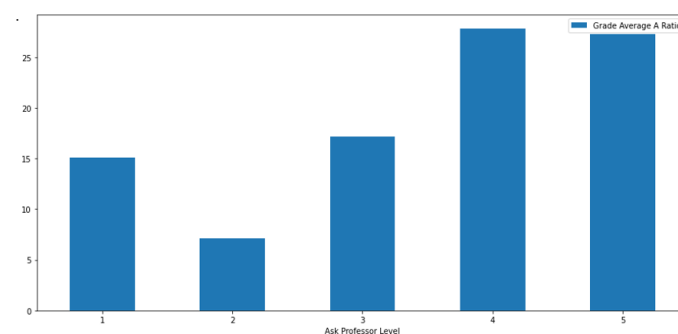


คำตอบ	ร้อยละ
น้อยที่สุด	12.5
น้อย	11.11
ปานกลาง	6.25
มาก	27.08
มากที่สุด	40.0

ภาพที่ 4.4 แสดงอัตราส่วนนักศึกษาได้ผลการเรียนเกรดเอวิชาภาษาอังกฤษระดับพื้นฐานในการสำรวจความรู้สึกสนุกสนานเมื่อได้เรียนภาษาอังกฤษ

4.2.2 การถามอาจารย์ทันทีเมื่อไม่เข้าใจเนื้อหาที่เรียน (A-PROF)

จากข้อมูลการสำรวจการสอบถามอาจารย์ทันทีเมื่อไม่เข้าใจเนื้อหาที่เรียนของนักศึกษาได้ผลการเรียนเกรดเอวิชาภาษาอังกฤษระดับพื้นฐานเมื่อเทียบกับจำนวนนักศึกษาทั้งหมดได้ผลลัพธ์แสดงดังภาพที่ 4.5

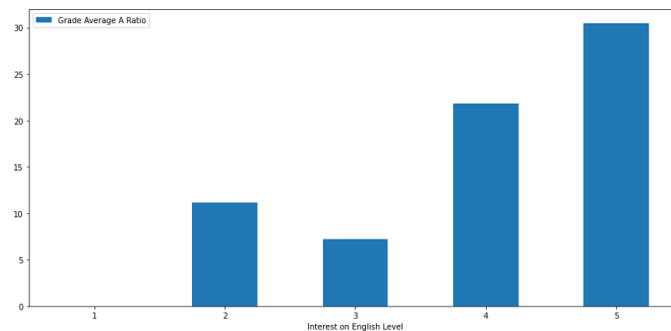


คำตอบ	ร้อยละ
น้อยที่สุด	15.09%
น้อย	7.14%
ปานกลาง	17.14%
มาก	27.77%
มากที่สุด	27.27%

ภาพที่ 4.5 แสดงอัตราส่วนนักศึกษาได้ผลการเรียนเกรดเอวิชาภาษาอังกฤษระดับพื้นฐานในการสำรวจการสอบถามอาจารย์ทันทีเมื่อไม่เข้าใจเนื้อหาที่เรียน

4.2.3 ความสนใจและตั้งใจเรียนวิชาภาษาอังกฤษในขณะที่ยังสอน (I-ENG)

จากข้อมูลการสำรวจความสนใจและตั้งใจเรียนวิชาภาษาอังกฤษในขณะที่ยังสอนของนักศึกษาได้ผลการเรียนเกรดเอวิชาภาษาอังกฤษระดับพื้นฐานเมื่อเทียบกับจำนวนนักศึกษาทั้งหมดได้ผลลัพธ์แสดงดังภาพที่ 4.6

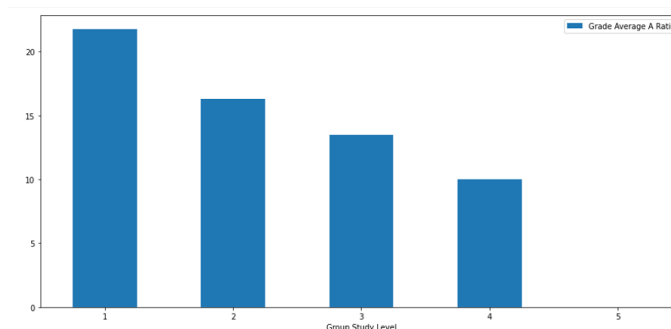


คำตอบ	ร้อยละ
น้อยที่สุด	0%
น้อย	11.11%
ปานกลาง	7.23%
มาก	21.78%
มากที่สุด	30.43%

ภาพที่ 4.6 แสดงอัตราส่วนนักศึกษาได้ผลการเรียนเกรดเอวิชาภาษาอังกฤษระดับพื้นฐานในการสำรวจความสนใจและตั้งใจเรียนวิชาภาษาอังกฤษในขณะที่ยังสอนของนักศึกษา

4.2.4 การจับกลุ่มทบทวนบทเรียนภาษาอังกฤษกับเพื่อนเมื่อใกล้สอบ (G-STUDY)

จากข้อมูลการสำรวจการจับกลุ่มทบทวนบทเรียนภาษาอังกฤษกับเพื่อนเมื่อใกล้สอบของนักศึกษาได้ผลการเรียนเกรดเอวิชาภาษาอังกฤษระดับพื้นฐานเมื่อเทียบกับจำนวนนักศึกษาทั้งหมดได้ผลลัพธ์แสดงดังภาพที่ 4.6

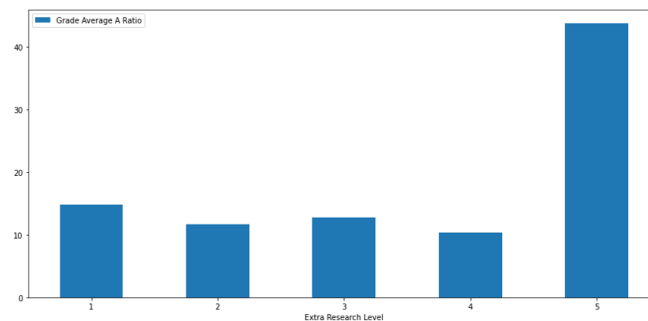


คำตอบ	ร้อยละ
น้อยที่สุด	21.74%
น้อย	16.28%
ปานกลาง	13.48%
มาก	10.00%
มากที่สุด	0.00%

ภาพที่ 4.6 แสดงอัตราส่วนนักศึกษาได้ผลการเรียนเกรดเอวิชาภาษาอังกฤษระดับพื้นฐานในการสำรวจการจับกลุ่มทบทวนบทเรียนภาษาอังกฤษกับเพื่อนเมื่อใกล้สอบ

4.2.5 การศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษนอกเหนือจากที่อาจารย์สอน (E-RES)

จากข้อมูลการสำรวจการศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษนอกเหนือจากที่อาจารย์สอนของนักศึกษาได้ผลการเรียนเกรดเอวิชาภาษาอังกฤษระดับพื้นฐานเมื่อเทียบกับจำนวนนักศึกษาทั้งหมดได้ผลลัพธ์แสดงดังภาพที่ 4.7



ลำดับ	ร้อยละ
น้อยที่สุด	14.81%
น้อย	11.26%
ปานกลาง	12.76%
มาก	10.34%
มากที่สุด	43.75%

ภาพที่ 4.7 แสดงอัตราส่วนนักศึกษาได้ผลการเรียนเกรดเอวิชาภาษาอังกฤษระดับพื้นฐานในการสำรวจการศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษนอกเหนือจากที่อาจารย์สอน

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยเรื่อง การวิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคเหมืองข้อมูล มีวัตถุประสงค์เพื่อศึกษาและวิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี โดยเปรียบเทียบการทำเหมืองข้อมูลระหว่างเทคนิค Decision Tree และเทคนิค Random Forests งานวิจัยนี้ใช้กลุ่มตัวอย่างข้อมูลจากนักศึกษาระดับปริญญาตรี คณะภูมิสารสนเทศศาสตร์ มหาวิทยาลัยบูรพา ปีการศึกษา พ.ศ.2561 - 2563 จำนวน 299 คน และเก็บรวบรวมข้อมูลคุณลักษณะทั้งหมด 32 คุณลักษณะจากแบบสอบถามปลายปิด ซึ่งชุดข้อมูลประกอบด้วย ข้อมูลด้านการเรียนภาษาอังกฤษของนักศึกษา ข้อมูลพื้นฐานทั่วไปของนักศึกษา ข้อมูลด้านพฤติกรรมที่ส่งผลต่อการเรียนรู้วิชาภาษาอังกฤษ และข้อมูลทัศนคติและแรงจูงใจในการเรียนภาษาอังกฤษ ทำการจำแนกข้อมูลเพื่อสร้างและทดสอบแบบจำลองของเทคนิค Decision Tree และเทคนิค Random Forests ด้วยการแบ่งชุดข้อมูลเรียนรู้ (Training) 209 ตัวอย่าง คิดเป็นร้อยละ 70 สำหรับเป็นข้อมูลฝึกสอน และชุดข้อมูลการทดสอบ (Testing) จำนวน 90 ตัวอย่าง คิดเป็นร้อยละ 30 ใช้ค่าสหสัมพันธ์แบบเพียร์สันในการคัดเลือกคุณลักษณะที่มีความสัมพันธ์กับประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษ ซึ่งกำหนดให้เกรดวิชาภาษาอังกฤษระดับพื้นฐานเป็นตัวจำแนกประเภทกลุ่มนักศึกษาก่อเป็น 2 กลุ่ม สำหรับนำเข้าสู่การวิเคราะห์และสร้างตัวแบบจำลองของทั้งสองเทคนิค และใช้การทดสอบประสิทธิภาพของแบบจำลองด้วยวิธี 10 - fold Cross - validation และค่าความแม่นยำ (Accuracy) ค่าความถูกต้อง (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F1 score) ด้วยรูปแบบ Confusion Matrix

5.1 สรุปผลการวิจัย

งานวิจัยเรื่อง การวิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคเหมืองข้อมูล สามารถสรุปผลได้ดังนี้

5.1.1 ผลการพัฒนาและเปรียบเทียบแบบจำลอง

การเปรียบเทียบประสิทธิภาพแบบจำลองระหว่างเทคนิค Decision Tree และเทคนิค Random Forest ด้วยการนำเข้าสู่ข้อมูลคุณลักษณะ 15 คุณลักษณะที่มีความสัมพันธ์กับประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี ซึ่งผู้วิจัยได้ใช้เกรดวิชาภาษาอังกฤษระดับพื้นฐานเป็นตัวกำหนดประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษานั้น ผลการสร้างและทดสอบตัวแบบจำลองแสดงให้เห็นว่า เทคนิค Random Forest ให้ค่าความแม่นยำในการทำนายประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาที่ค่าเฉลี่ยร้อยละ 85.14 ซึ่งมากกว่าเทคนิค Decision Tree ที่มีค่าความแม่นยำที่ค่าเฉลี่ยร้อยละ 78 จากการทดสอบประสิทธิภาพโดยใช้วิธี 10 - fold cross validation และเมื่อผลการวัดประสิทธิภาพด้วยค่าความแม่นยำ (Accuracy Score) ด้วย Confusion Matrix เทคนิค Random Forest ยังคงให้ค่าความแม่นยำมากกว่า เทคนิค Decision Tree ที่ 0.84 ในขณะที่เทคนิค Decision Tree ให้ค่าความแม่นยำที่ 0.69

5.1.1 ผลวิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษา ระดับปริญญาตรี ด้วยเทคนิค Random Forest

จากการนำเข้าสู่ข้อมูลกลุ่มตัวอย่างนักศึกษาจำนวน 299 คน โดยสัดส่วนของกลุ่มนักศึกษาที่มีประสิทธิภาพการเรียนรู้ภาษาอังกฤษระดับพื้นฐาน จำนวน 42 ตัวอย่าง และกลุ่มนักศึกษาที่มีผลการเรียนรู้วิชาภาษาอังกฤษระดับพื้นฐานทั่วไปจำนวน 257 ตัวอย่าง ชุดข้อมูลที่คัดเลือกจำนวน 15 คุณลักษณะเข้าสู่กระบวนการวิเคราะห์ข้อมูลด้วยเทคนิค Random Forest แสดงให้เห็นถึง 5 คุณลักษณะสำคัญที่มีความเกี่ยวข้องกับประสิทธิภาพทางการเรียนรู้วิชาภาษาอังกฤษระดับพื้นฐานของนักศึกษาประกอบด้วย ความรู้สึกสนุกสนานเมื่อได้เรียนภาษาอังกฤษ การถามอาจารย์ทันทีเมื่อไม่เข้าใจเนื้อหาที่เรียน

ความสนใจและตั้งใจเรียนวิชาภาษาอังกฤษในขณะที่อาจารย์สอน การจับกลุ่มทบทวนบทเรียนภาษาอังกฤษกับเพื่อนเมื่อใกล้สอบ และการศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษนอกเหนือจากที่อาจารย์สอน โดยผลการวัดประสิทธิภาพแบบจำลองที่พัฒนาโดยเทคนิค Random Forest ให้ค่าความแม่นยำ (Accuracy Score) ค่าเฉลี่ยร้อยละ 86 และประสิทธิภาพแบบจำลองในมุมมองความสามารถของแบบจำลองในการจำแนกกลุ่มนักศึกษาที่มีผลการเรียนวิชาภาษาอังกฤษระดับพื้นฐานทั่วไปให้ค่าความถูกต้อง (Precision) ค่าเฉลี่ยร้อยละ 87 ค่าความระลึก (Recall) ค่าเฉลี่ยร้อยละ 94 และค่าถ่วงดุล (F-Score) ค่าเฉลี่ยร้อยละ 92 ในขณะที่ แบบจำลองในมุมมองความสามารถของแบบจำลองในการจำแนกกลุ่มนักศึกษาที่มีประสิทธิภาพการเรียนวิชาภาษาอังกฤษ ให้ค่าความถูกต้อง (Precision) ค่าเฉลี่ยร้อยละ 50 ค่าความระลึก (Recall) ค่าเฉลี่ยร้อยละ 15 และค่าถ่วงดุล (F-Score) ค่าเฉลี่ยร้อยละ 24

5.2 อภิปรายผลการวิจัย

การทำเหมืองข้อมูลช่วยให้สามารถสกัดความรู้ที่น่าสนใจจากข้อมูล และนำมาใช้ประโยชน์ได้อย่างกว้างขวาง การศึกษาวิจัยจำนวนมากได้นำการทำเหมืองข้อมูลมาใช้ในการปรับปรุงคุณภาพการศึกษาและเพิ่มประสิทธิภาพการเรียนการสอนต่างๆ สำหรับการศึกษาครั้งนี้ ได้ทำการทดลองเปรียบเทียบประสิทธิภาพแบบจำลองระหว่างเทคนิค Decision Tree และเทคนิค Random Forest เพื่อคัดเลือกแบบจำลองที่เหมาะสมกับการนำมาใช้วิเคราะห์คุณลักษณะที่ส่งผลต่อประสิทธิภาพการเรียนวิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี แบบจำลองได้รับการพัฒนาโดยใช้ข้อมูลที่รวบรวมโดยวิธีแบบสอบถามหลังจากได้รับการตรวจสอบถูกต้องจากผู้เชี่ยวชาญ ผลการศึกษาพบว่า เทคนิค Random Forest มีความแม่นยำ และมีประสิทธิภาพมากกว่าเทคนิค Decision Tree ซึ่งสอดคล้องกับงานวิจัยของ (R, 2015) ที่อธิบายว่าเทคนิค Random Forest เป็นกลุ่มของแผนผังต้นไม้การตัดสินใจ (Decision Tree) ประกอบด้วยหลาย Decision Tree โดยอัลกอริทึมจะเปรียบเทียบและทำการวิเคราะห์แยกกันแต่ละ Decision Tree ทำให้อัตราความผิดพลาดของเทคนิค Random Forest น้อยกว่าเทคนิค Decision Tree จึงช่วยในการทำนายข้อมูลได้อย่างแม่นยำกว่า

สำหรับผลลัพธ์ของอัลกอริทึมการจัดอันดับคุณลักษณะสำคัญที่มีความสัมพันธ์กับประสิทธิภาพการเรียนวิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี ด้วยเทคนิค Random Forest สามารถสรุปผลการการศึกษาที่สำคัญจากการวิเคราะห์ 5 อันดับแรก ดังนี้

- 1) ความรู้สึกสนุกกับการเรียนภาษาอังกฤษ ผลการศึกษาพบว่า ระดับความรู้สึกสนุกสนานในการเรียนภาษาอังกฤษที่สูงมีความสัมพันธ์โดยตรงกับนักศึกษาที่มีประสิทธิภาพทางการเรียนวิชาภาษาอังกฤษ ซึ่งแสดงให้เห็นว่ากลุ่มนักศึกษาที่มีประสิทธิภาพการเรียนวิชาภาษาอังกฤษ มักจะรู้สึกสนุกกับการเรียนภาษาอังกฤษ สอดคล้องกับงานวิจัยของ (Lucardie, 2014) ที่กล่าวว่า แรงจูงใจและความเพลิดเพลินในการเรียนวิชาใดวิชาหนึ่งจะเชื่อมโยงกับผลผลิต ประสิทธิภาพ และความสำเร็จในการเรียนรู้วิชานั้น ทั้งนี้ มีหลายปัจจัยที่สามารถสร้างสภาพแวดล้อมการเรียนรู้เชิงบวกกระตุ้นความสนุกสนานและความเพลิดเพลินในการเรียนภาษาอังกฤษ ซึ่งอาจารย์สามารถใช้ผลลัพธ์ของการศึกษานี้เพื่อปรับเปลี่ยนหลักสูตรในชั้นเรียน การรวมกิจกรรมที่สนุกสนาน เช่น แบบฝึกหัดภาษาอังกฤษ วิดีโอเกมต่างๆ และกิจกรรมทางสังคมจะช่วยสร้างบรรยากาศที่เป็นมิตร และสนับสนุนการเรียนรู้ภาษาอังกฤษของนักศึกษาได้ดียิ่งขึ้น

- 2) การถามอาจารย์ทันทีเมื่อไม่เข้าใจเนื้อหาที่เรียน ผลการศึกษาพบว่า นักศึกษาที่มีแนวโน้มจะขอความช่วยเหลือหรือถามข้อสงสัยต่างๆ จากอาจารย์เมื่อเรียนภาษาอังกฤษมักจะเป็นกลุ่มนักศึกษาที่มีประสิทธิภาพทางการเรียนวิชาภาษาอังกฤษ สอดคล้องกับ (ประชาณ, 2537) กล่าวถึง เทคนิคในการสร้างความเข้าใจในการเรียนว่าควรเตรียมความล่วงหน้า และถามคำถาม เพื่อทำให้เกิดความรู้ ความเข้าใจอย่างถ่องแท้ ดังนั้น อาจารย์สามารถจัดสรรเวลาเฉพาะเพื่อสร้าง

แบบฝึกหัดหรือกิจกรรมการตั้งคำถาม กิจกรรมนี้สามารถเปิดโอกาสให้ทั้งนักศึกษาและอาจารย์ได้ถามคำถาม นอกจากนี้ การถามคำถามช่วยให้นักเรียนประสบความสำเร็จ และยังช่วยให้อาจารย์เรียนรู้นักเรียนอีกด้วย

3) ความสนใจเรียนภาษาอังกฤษ ผลการศึกษาพบว่า นักศึกษาที่มีประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษมีความสนใจเรียนในในระดับที่สูงเมื่อเทียบกับกลุ่มนักศึกษาที่มีผลการเรียนปกติ สอดคล้องกับการวิจัยเชิงวิชาการของ (Renninger et al, 1992) อธิบายว่านักเรียนที่สนใจเรียนมากกว่ามีแนวโน้มที่จะมีผลการเรียนสูงกว่านักเรียนที่มีความสนใจเรียนระดับต่ำ ดังนั้น การจัดรูปแบบการเรียนการสอนให้เหมาะสมและสร้างบรรยากาศในการเรียนที่น่าสนใจ อาจเป็นเครื่องมือหนึ่งที่ช่วยเพิ่มความสนใจการเรียนรู้วิชาภาษาอังกฤษได้ อาจารย์สามารถจัดกิจกรรมที่นักศึกษาสามารถเรียนภาษาอังกฤษนอกเหนือจากรูปแบบการเรียนตามตำราทั่วไป ทำให้การเรียนภาษาอังกฤษเป็นเรื่องที่น่าสนใจและสนุกสนานมากยิ่งขึ้น ซึ่งมีหลายวิธีในการจัดรูปแบบชั้นเรียนให้น่าสนใจสำหรับนักเรียนมากขึ้น เช่น เรียนรู้ผ่านภาพยนตร์ และ รายการทีวี เป็นต้น

4) การจับกลุ่มทบทวนบทเรียนภาษาอังกฤษกับเพื่อนเมื่อใกล้สอบ ผลการศึกษาพบว่า นักศึกษาที่มีประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษมักไม่จับกลุ่มทบทวนบทเรียน พฤติกรรมนี้อาจวิเคราะห์ได้ว่า นักศึกษาที่มีประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษอาจมีการจัดสรรเวลาของตัวเองชัดเจน รวมถึงทบทวนบทเรียนและศึกษาค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษ ดังนั้น จึงไม่จำเป็นต้องทบทวนบทเรียนในช่วงเวลาใกล้สอบมากนัก มีงานวิจัยทางวิชาการแสดงให้เห็นว่ากลยุทธ์การจัดกลุ่มสามารถส่งผลทั้งด้านลบและด้านบวกต่อการเรียนรู้และการพัฒนาของนักเรียน ขึ้นอยู่กับวิธีการนำไปใช้ในห้องเรียน (Chen, Hwang, & Lin, 2019) การทบทวนบทเรียนช่วยใกล้สอบอาจมีผลกระทบทั้งในแง่บวกและด้านลบ ซึ่งอาจารย์สามารถใช้ข้อมูลนี้ในการสังเกตพฤติกรรมของนักศึกษา เช่น จากการส่งการบ้าน การมอบหมายงาน หรือกระตุ้นให้นักศึกษาจัดสรรเวลาเพื่อเตรียมตัวสอบ เป็นต้น

5) การศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษนอกเหนือจากที่อาจารย์สอน ผลการศึกษาพบว่า นักศึกษาที่มีประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษสนใจที่จะศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษ ซึ่งเป็นพฤติกรรมการเรียนรู้ที่สำคัญ การใช้เวลาเรียนภาษาอังกฤษนอกห้องเรียนเพื่อทบทวนบทเรียนมากขึ้น สอดคล้องกับแนวคิดของ (Maxwell-Jolly, 2011) แสดงให้เห็นว่าการใช้เวลาพิเศษในการศึกษานอกห้องเรียนสามารถปรับปรุงแรงจูงใจและการมีส่วนร่วมของนักเรียน และส่งเสริมความสัมพันธ์ทางสังคมระหว่างนักเรียน ครอบครัวได้ ผลลัพธ์นี้สามารถนำไปใช้จัดการเรียนการสอน ที่อาจเน้นโดยจัดหาแหล่งข้อมูลเพิ่มเติมและเครื่องมือที่จำเป็นในการผ่านชั้นเรียน

5.4 ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

5.2.1. งานวิจัยนี้ทำการศึกษาทดลองโดยใช้กลุ่มตัวอย่างจากคณะมนุษยศาสตร์ มหาวิทยาลัยบูรพา ทำให้ผลลัพธ์จากการพัฒนาแบบจำลองเป็นเพียงกลุ่มตัวอย่างเดียว ซึ่งอาจไม่ครอบคลุมเท่าที่ควร ดังนั้น เพื่อประสิทธิภาพในการทำนายของแบบจำลองในการศึกษาครั้งต่อไปอาจต้องกระจายกลุ่มตัวอย่าง พร้อมทั้งเพิ่มจำนวนตัวอย่าง เพื่อเก็บข้อมูลให้ครอบคลุมมากยิ่งขึ้น

5.2.2. การวัดประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี อาจมีปัจจัยเพิ่มเติมที่สามารถนำมาใช้กำหนดเป็นเกณฑ์ในการวัดประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาได้คือ เกรดรายวิชา เช่น การสร้างแบบทดสอบวัดระดับภาษาอังกฤษ และเปรียบเทียบผลคะแนนกับเกรดที่ได้จากมหาวิทยาลัย เป็นต้น

5.2.3. ควรทำการศึกษาตัวแปรเชิงสาเหตุอื่นๆ ที่คาดว่าจะส่งผลกระทบต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษ เพื่อเปรียบเทียบและวางแผนประสิทธิภาพการเรียนรู้ที่ครอบคลุม อาจมีปัจจัยอื่นๆ ที่ส่งผลกระทบต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาสามารถนำมาวิเคราะห์เพิ่มเติม เช่น ผลการเรียนในรายวิชาอื่นๆ ความถนัด จำนวนชั่วโมงเรียนต่อวัน ปัจจัยด้านสังคม คุณภาพการสอน เป็นต้น

บรรณานุกรม

- Afroz Chakure. (29 June 2019). *Random Forest Regression*. เข้าถึงได้จาก medium: <https://medium.com/swlh/random-forest-and-its-implementation-71824ced454f>
- Alizadeh Mitra. (2016). The Impact of Motivation on English Language Learning. *International Journal of Research in English Education*, 11.
- Ann K Renninger, Suzanne Hidi, , Andreas Krapp, และ 1 st Edition. (1992). *The Role of interest in Learning and Development*. New York: 17 January 2014.
- Beiyin Chen, Gwo-Haur Hwang, และ Tze-Shiun Lin. (2019). Impacts of a dynamic grouping strategy on students' learning effectiveness and experience value in an item bank-based collaborative practice system. *British Journal of Educational Technology*, 36-52.
- Broussard, S. C., & Garrison, M. E. B. (2004). The relationship between classroom motivation and academic achievement in elementary school-aged children. *Family and Consumer Sciences Research Journal* 33(2): 106-120.
- Cambridge Dictionary*. (2021). เข้าถึงได้จาก Dictionary: <https://dictionary.cambridge.org/dictionary/english/attitude>
- Gardner, R. C. (1975). Motivational variables in second language learning. *Proceedings of the Canadian Association of Applied Linguistics*, 45-73.
- Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second-language learning*. Rowley, Massachusettes: Newbury House.
- Chen, I., & Huang, S. (2003). Language learning strategy use differences between high and low proficiency learners – an example from a technology college in Taiwan. *Journal of Humanities of Changhua Teachers College*, 2, 301-321.
- Daniel T. Larose, และ Chantal D. Larose. (2014). *Discovering Knowledge in Data _An Introduction to Data Mining (2nd ed.) Data Warehouse and Mining*. (June 2021). เข้าถึงได้จาก Techblogmu: <https://techblogmu.blogspot.com/2018/05/describe-various-functionalities-of.html>
- Dorothy Lucardie. (2014). The impact of fun and enjoyment on adult's learning. *Dorothy Lucardie / Procedia - Social and Behavioral Sciences* 142 (2014) 439 – 446, 439-446.
- EducationFrist. (2020). *EF English Proficiency Index*. Switzerland: 2020.
- Ellis, R. (1997). *The study of second language acquisition*. Oxford University Press.
- Intaraprasert, C. (2000). *Language learning strategies employed by engineering students learning English at the tertiary level in Thailand*. Doctoral dissertation. School of Education. University of Leeds, the United Kingdom.
- Jeffrey Stanton. (2013). Hi Ho, Hi Ho - Data Mining We Go. ใน *An Introduction to Data Science* (หน้า 173).
- Julie Maxwell-Jolly. (2011). English Learners and Out-of-School Time Programs: The Potential of OST Programs to Foster EL Success. *ENGLISH LEARNERS AND OUT-OF-SCHOOL TIME PROGRAMS*, 1-12.
- Jure Leskovec, Anand Rajaraman, and Jeffrey D Ullman. (2010). What is Data Mining? *Mining of Massive Datasets* (หน้า 1).

- Mantle-Bromley, Corinne. (1995). Positive Attitudes and Realistic Beliefs: Links to Proficiency. *The Modern Language Journal*, 79(3).
- Montano, D. E. & Kasprzyk, D. (2008). Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. *Health behavior and health education: Theory, research, and practice*. San Francisco, CA: Jossey-Bass, 67-96.
- Omer Faruk Alls, Ertan Karakurt, และ Piero Melli. (2000). *Data Mining For Database Marketing At Garanti Bank. WIT Transactions on Information and Communication Technologies*. 25: 93-107.
- Oxford, R.L. (1990). *Language Learning Strategies: What every teacher should know*. New York: Newsbury House/Harper Collins.
- Paulo Cortez, และ Alice Silva. (2014). Using data mining to predict secondary school student performance.
- Pedtrides, J. R. (2006). Attitudes and Motivation and their impact on the performance of young English Language Learners. *Journal of Language and Learning*, 5(1), 1-20.
- Prajwala T R. (2015). A Comparative Study on Decision Tree and Random Forest Using R Tool. *International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 1, January 2015*, 196-199.
- R. R. Kabra, และ R. S. Bichkar. (2011). Performance Prediction of Engineering Students using Decision Trees. *International Journal of Computer Applications (0975 – 8887)*, 8.
- Satish Gunjal. (2020). *K Fold Cross Validation*. เข้าถึงได้จาก Quality Tech Tutorials: <https://satishgunjal.com/kfold/#5SatishGunjal>
- Spraat, M., Pulvemess, A., & William, M. (2005). The TMT teaching knowledge test course. Cambridge: Cambridge University Press.
- SpringerLink. (2013). *Encyclopedia of Systems Biology*. เข้าถึงได้จาก SpringerLink: https://link.springer.com/referenceworkentry/10.1007%2F978-1-4419-9863-7_599
- Tamimi, A. A., & Shuib, M. (2009). Motivation and attitudes towards learning English: A study of petroleum engineering undergraduates at Hadhramout university of sciences and technology. *GEMA Online Journal of Language Studies*, 9(2), 29-55.
- TheCRISP-DMConsortium. (2000). *The CRISP-DM reference model*.
- Triandis, H. C. (1971). *Attitude and attitude change*. New York: John Wiley & Sons Inc.
- Visser, P.S., Holbrook, A. L., & Krosnick, J. A., (2008). Knowledge and Attitudes. *Handbook of Public Opinion Research*. Thousand Oaks, CA: Sage Publications.
- Wichai Puarungroj, Narong Boonsirisumpun, Pathapong Pongpatrakant, และ Suchada Phromkhot. (2017). A preliminary implementation of data mining approaches for predicting the results of English exit exam. *2017 2nd International Conference on Information Technology (INCIT)At: Nakhonpathom, Thailand*. Thailand.
- Wu, M.M.F. (2008). Language learning strategy use of Chinese ESL learners of Hong Kong- findings from a qualitative study. *Electronic Journal of Foreign Language Teaching*, 5(1), 68-83.

- กัลยณภัส สุขพลับพลา, พนิดา หล่อวงศ์ตระกูล, และ เฉลิมชัย หล่อวงศ์ตระกูล. (2559). ตัวแบบเลือกแผนการเรียนในระดับมัธยมศึกษาตอนปลาย โดยใช้เทคนิคเหมืองข้อมูล. *การประชุมวิชาการวิศวกรรมศาสตร์ วิทยาศาสตร์ เทคโนโลยี และสถาปัตยกรรมศาสตร์ ครั้งที่ 7 วันที่ 25-26 กรกฎาคม 2559*, 715.
- ชนวัฒน์ ศรีตออัน. (2551). *ฐานข้อมูล คลังข้อมูล และเหมืองข้อมูล (พิมพ์ครั้งที่ 2)*. Bangkok.
- ชนกร สุวรรณพุด, และ เสี่ยม โดรัตน์. (2559). การพัฒนารูปแบบการสอนกลยุทธ์การเรียนรู้ภาษาต่างประเทศ เพื่อเสริมสร้างความสามารถในการฟัง-พูดภาษาอังกฤษ ส สำหรับนักศึกษาระดับปริญญาตรี. *Veridian E-Journal, Silpakorn University*, 2.
- ธีรวุฒิ เอกะกุล. (2543). *ระเบียบวิธีวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์*. อุบลราชธานี.
- ณภัทร วุฒิวงศา. (2557). กลยุทธ์สร้างแรงจูงใจ :การพัฒนาทักษะภาษาอังกฤษ.วารสารนักบริหาร, ปีที่ 34 ฉบับที่ 1 (ม.ค.-มิ.ย. 2557) หน้า 89-97.
- ประธาน วัฒนวาณิชย์. (2537). *เรื่องไม่ยาก ถ้าอยากเรียนเก่ง*. กรุงเทพฯ : โรงพิมพ์เจริญผล.
- ปิติภูมิ โปสาวัง, สถิตย์ โชคโพธิ์สอาด, และ วสันต์ ภัทรธิดม. (2560). *ก้าวสู่โลกแห่งการระบุดาแห่งที่แม่นยำ:พัฒนาการในปัจจุบันและอนาคตของระบบดาวเทียมนำทางทั่วโลกหลายระบบ*. วารสารวิชาการ วิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครสวรรค์ ปีที่ 9 ฉบับที่ 10 กรกฎาคม – ธันวาคม 2560, 3.
- พยุณ พาณิชย์กุล. (2548). *การพัฒนาระบบค้ำไมน์นิ่งโดยใช้ Decision Tree*. โครงการพัฒนาระบบงานปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ แขนงวิทยาการสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าคุณทหารลาดกระบัง.
- พรพมิล ชัยวุฒิศักดิ์, และ ชูชาติ, ถอดอมวิเศษ. (2562). การพัฒนาการท นานายผลการเรียนของนักศึกษาชั้นปี ที่1 โดยใช้เทคนิคการท าเหมืองข้อมูล1. วารสารวิจัยรามคำแหง (วิทยาศาสตร์และเทคโนโลยี) ปีที่ 22 ฉบับที่ 2 กรกฎาคม – ธันวาคม 2562, 43.
- พรพจน์ เพ็ชรทวีพรเดช. (2547). ปัจจัยด้านพฤติกรรมที่เกี่ยวข้องกับผลสัมฤทธิ์ทางการเรียนของนักศึกษามหาวิทยาลัยกรุงเทพ. ปริญญานิพนธ์การศึกษามหาบัณฑิต. มหาวิทยาลัยศรีนครินทรวิโรฒ.
- พัฒนพงษ์ คลรัตน์, และ จาริ ทองคำ. (2560). การเปรียบเทียบประสิทธิภาพของแบบจำลองในการพยากรณ์ความสำเร็จการศึกษาของนักเรียนระดับประกาศนียบัตรวิชาชีพ. *The Comparison of Performance Models for Predicting Students Success in Vocation Education*, 380.
- วันวิสาข์ ชนะประเสริฐ,. (2559). การประยุกต์ใช้เทคนิคเหมืองข้อมูลเพื่อแนะนำอาชีพสำหรับนักศึกษาปริญญาตรี คณะโบราณคดีมหาวิทยาลัยศิลปากร. *การประชุมวิชาการบัณฑิตศึกษาระดับชาติและนานาชาติครั้งที่ 7 “ประเทศไทย 4.0 นวัตกรรมสร้างสรรค์สู่การพัฒนาที่ยั่งยืน” (NGSC&IGSC)*.
- ศุภิกา นิรัตติชัย. (2561). Attitudes, Motivation, and English Learning Behavior. *วารสารมนุษยศาสตร์ และสังคมศาสตร์ ปีที่ 9 ฉบับที่ 2 กรกฎาคม – ธันวาคม 2561*, 160.
- สุรพงศ์, เอื้อพัฒนามงคล. (2561). *การทำเหมืองข้อมูล*. Bangkok.
- สุระสิทธิ์ ทองมัว. (2558). การเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจ รุ่นที่ 3 ที่ใช้เทคนิคการแบ่งข้อมูลที่แตกต่างกัน. *SDU Research Journal*, 29-45.

ภาคผนวก ก
แบบสอบถามงานวิจัย

แบบสอบถามเพื่องานวิจัยด้านการเรียนภาษาอังกฤษของนักศึกษาระดับปริญญาตรี

แบบสอบถามนี้เป็นส่วนหนึ่งของการศึกษาระดับปริญญาโท คณะสถิติประยุกต์สาขาวิชาบริหารเทคโนโลยีสารสนเทศ สถาบันบัณฑิตพัฒนบริหารศาสตร์ ในรายวิชา บทส.9000 การค้นคว้าอิสระ ปีการศึกษา 2563 และมีวัตถุประสงค์เพื่อประกอบการศึกษาวิจัยเรื่องการเปรียบเทียบประสิทธิภาพของอัลกอริทึม เพื่อวิเคราะห์คุณลักษณะที่ส่งผลต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคการทำเหมืองข้อมูล

คำชี้แจง >> แบบสอบถามนี้ แบ่งออกเป็น 3 ส่วน

ส่วนที่ 1 แบบสอบถามเกี่ยวกับข้อมูลและคุณสมบัติทั่วไป

ส่วนที่ 2 แบบสอบถามเกี่ยวกับพฤติกรรมการเรียนภาษาอังกฤษ

ส่วนที่ 3 แบบสอบถามเกี่ยวกับคุณลักษณะด้านทัศนคติและแรงจูงใจ

ส่วนที่ 1 แบบสอบถามเกี่ยวกับข้อมูลและคุณสมบัติทั่วไป

ข้อ	คำถาม
1	รหัสนักศึกษา -----
2	เพศของนักศึกษา ☐ ชาย ☐ หญิง
3	อายุของนัก -----
4	ฉันมีคอมพิวเตอร์และอินเทอร์เน็ตใช้ส่วนตัว ☐ ใช่ ☐ ไม่ใช่
5	ลักษณะที่ตั้งของบ้านที่อยู่อาศัยของฉัน ☐ อยู่ในเมือง ☐ อยู่นอกเมืองหรือชนบท
6.	ผลการเรียนวิชาภาษาอังกฤษระดับมหาวิทยาลัย (วิชาบังคับระดับพื้นฐาน) ☐ A ☐ B ☐ C ☐ D ☐ F ☐ I
7.	ฉันไม่ได้เข้าเรียนวิชาภาษาอังกฤษ (จำนวนครั้ง) ☐ 0 ครั้ง ☐ 1 ครั้ง ☐ 2 ครั้ง ☐ 3 ครั้ง ☐ 4 ครั้ง ☐ 5 ครั้ง ☐ 6 ครั้ง

ข้อ	คำถาม
	☐ 7 ครั้ง ☐ 8 ครั้ง ☐ 9 ครั้ง ☐ 10 ครั้ง

ส่วนที่ 2 แบบสอบถามเกี่ยวกับพฤติกรรมการเรียนภาษาอังกฤษ (15 ข้อ)

ข้อ	ข้อคำถาม
1.	ฉันเรียนพิเศษวิชาภาษาอังกฤษ ☐ ใช่ ☐ ไม่ใช่
2.	เวลาว่างจากการเรียนของฉัน (ไม่นับช่วงเวลาทำงานพิเศษหรือทำงานบ้านหรือการนอนพัก) ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด
3.	ฉันใช้เวลาสำหรับออกไปเที่ยวกับเพื่อน ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด
4.	ฉันใช้เวลาว่างของฉันฝึกฝนภาษาอังกฤษด้วยตนเอง ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด

ข้อ	ข้อความ
5.	ฉันมักจะถามอาจารย์ทันทีเมื่อฉันรู้สึกไม่เข้าใจเนื้อหาที่เรียน (วิชาภาษาอังกฤษ) ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด
6.	ฉันวางแผนการเรียนภาษาอังกฤษ เช่น ทำตารางอ่านหนังสือ ทำรายงาน หรือทำการบ้านภาษาอังกฤษในแต่ละวันและทำตามนั้น ☐ ใช่ ☐ ไม่ใช่
7.	ฉันเตรียมตัวก่อนเข้าเรียนวิชาภาษาอังกฤษ เช่น คำศัพท์ อ่านบทเรียนล่วงหน้า ☐ ใช่ ☐ ไม่ใช่
8.	ฉันชอบนั่งใกล้ๆ เพื่อนที่ตั้งใจเรียน ☐ ใช่ ☐ ไม่ใช่
9.	ฉันมีเพื่อนเป็นชาวต่างชาติ เพื่อฝึกใช้ทักษะภาษาอังกฤษ ☐ ใช่ ☐ ไม่ใช่
10.	ฉันให้ความสนใจและตั้งใจเรียนวิชาภาษาอังกฤษในขณะที่อาจารย์สอน ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด
11.	ฉันจดบันทึกตามที่อาจารย์สอน ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด

ข้อ	ข้อคำถาม
12	ฉันมักตอบคำถามในขณะที่เรียนภาษาอังกฤษ ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด
13	ฉันทบทวนบทเรียนภาษาอังกฤษหลังเลิกเรียน ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด
14	ฉันศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษนอกเหนือจากที่อาจารย์สอน ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด
15	ฉันจับกลุ่มทบทวนบทเรียนภาษาอังกฤษกับเพื่อนเมื่อใกล้สอบ ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด

แบบสอบถามเกี่ยวกับคุณลักษณะด้านทัศนคติและแรงจูงใจ (11 ข้อ)

ข้อ	ข้อคำถาม
1.	ฉันมีความต้องการศึกษาในระดับที่สูงขึ้น ☐ ใช่ ☐ ไม่ใช่
2.	ฉันได้รับการสนับสนุนด้านการเรียนภาษาอังกฤษจากครอบครัว ☐ ใช่ ☐ ไม่ใช่

ข้อ	ข้อความ
3.	ฉันสนุกกับการเรียนภาษาอังกฤษ ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด
4.	ฉันสนใจเรียนภาษาอังกฤษเพื่อพัฒนาตนเอง ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด
5.	ฉันสนใจเรียนภาษาอังกฤษเพื่อหางานที่ดีในอนาคต ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด
6.	ฉันสนใจเรียนภาษาอังกฤษ เพราะเป็นความต้องการของพ่อแม่ ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด
7.	ฉันสนใจเรียนภาษาอังกฤษ เพื่อต้องการสอบผ่าน ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด

ข้อ	ข้อความ
8.	ฉันสนใจเรียนภาษาอังกฤษ เพื่อใช้สื่อสารในชีวิตประจำวัน ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด
9.	ฉันมีความสุขในการอ่านหนังสือ หรือ ดูข่าว ดูภาพยนตร์ที่เป็นภาษาอังกฤษ ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด
10.	การเรียนภาษาอังกฤษมีความสำคัญกับฉัน เพราะฉันต้องการ ได้เกรดที่ดี ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด
11.	การเรียนภาษาอังกฤษมีความสำคัญกับฉัน เพราะภาษาอังกฤษช่วยให้ฉันมีความรู้มากขึ้น ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด

ขอขอบคุณนักศึกษาทุกท่านที่สละเวลาตอบแบบสอบถามนี้ ซึ่งจะช่วยให้เราสามารถนำมาใช้ประเมินคุณลักษณะ
ที่ส่งผลต่อประสิทธิภาพการเรียนวิชาภาษาอังกฤษ เพื่อนำมาเป็นแนวทางการพัฒนาภาษาอังกฤษให้กับนักศึกษาในอนาคต
ต่อไป

แบบสอบถามออนไลน์



ภาคผนวก ข
แบบประเมินแบบสอบถามเชิงปริมาณผ่านผู้เชี่ยวชาญ (ioc)

แบบประเมินความคิดเห็นของผู้เชี่ยวชาญ
การวิเคราะห์ความสอดคล้องของวัตถุประสงค์และหัวข้อการออกแบบสอบถาม

คำชี้แจง ขอให้ท่านผู้เชี่ยวชาญได้กรุณาแสดงความคิดเห็นของท่านที่มีต่อข้อมูลสำหรับใช้วิเคราะห์ในงานวิจัย โดยใส่เครื่องหมาย (✓) ลงในช่องความคิดเห็นของท่าน พร้อมเขียนข้อเสนอแนะเพื่อเป็นประโยชน์ในการนำไปพิจารณาปรับปรุงต่อไป (ส่วนที่ 1 ประเมินความสอดคล้องของข้อมูลและแบบสอบถามสำหรับงานวิจัย) และกำหนดเกณฑ์คะแนนวิชาภาษาอังกฤษ เพื่อใช้สำหรับแบ่งกลุ่มประสิทธิภาพการเรียนรู้ภาษาอังกฤษของนักศึกษา (ส่วนที่ 2 กำหนดเกณฑ์คะแนนวิชาภาษาอังกฤษ)

การหาค่าดัชนีความสอดคล้อง (IOC) โดยการพิจารณาเพื่อหาความสอดคล้อง ดังนี้

- +1 หมายถึง แน่ใจว่าหัวข้อคำถามสอดคล้องกับวัตถุประสงค์
- 0 หมายถึง ไม่แน่ใจว่าหัวข้อคำถามสอดคล้องกับวัตถุประสงค์
- 1 หมายถึง แน่ใจว่าหัวข้อคำถามไม่สอดคล้องกับวัตถุประสงค์

หมายเหตุ แบบสอบถามเป็นลักษณะคำถามปลายปิด เพื่อผู้วิจัยจะสามารถรวบรวมเป็นชุดข้อมูล (Dataset) และนำมาใช้ในการทำนายปัจจัยสำคัญที่ส่งผลต่อความสำเร็จทางการเรียนวิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี ด้วยเทคนิคเหมืองข้อมูล (Data mining)

การวิจัยส่วน 1 แบบสอบถามเกี่ยวกับข้อมูลและคุณสมบัติทั่วไป

ข้อ	ข้อความคำถาม	ความคิดเห็นของผู้เชี่ยวชาญ			ข้อเสนอแนะ
		+1	0	-1	
1	รหัสประจำตัวนักศึกษา -----				
2	เพศของนักศึกษา ☐ ชาย ☐ หญิง				
3	อายุของนักศึกษา -----				
4	ฉันมีคอมพิวเตอร์และอินเทอร์เน็ตใช้ส่วนตัว ☐ ใช่ ☐ ไม่ใช่				
5	ลักษณะที่ตั้งของบ้านที่อยู่อาศัยของฉัน ☐ อยู่ในเมือง ☐ อยู่นอกเมืองหรือชนบท				
6	เกรดผลการเรียนวิชาภาษาอังกฤษระดับมหาวิทยาลัย ☐ A ☐ B ☐ C				

ข้อ	ข้อความ	ความคิดเห็นของ ผู้เชี่ยวชาญ			ข้อเสนอแนะ
		+1	0	-1	
	☐ D ☐ F				
7	จำนวนการขาดเรียนวิชาภาษาอังกฤษ ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10				

วัตถุประสงค์การวิจัยส่วนที่ 3: คำถามเกี่ยวกับพฤติกรรมการเรียนของผู้ตอบแบบสอบถาม (จากแบบสอบถาม)

ข้อ	ข้อความ	ความคิดเห็นของ ผู้เชี่ยวชาญ			ข้อเสนอแนะ
		+1	0	-1	
8	ถิ่นเรียนพิเศษวิชาภาษาอังกฤษ ☐ ใช่ ☐ ไม่ใช่				
9	เวลาว่างจากการเรียนของถิ่น (ไม่นับช่วงเวลาทำงานพิเศษ หรือทำงานบ้านหรือการนอนพัก) ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				
10	ถิ่นใช้เวลาสำหรับออกไปเที่ยวกับเพื่อน ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				

ข้อ	ข้อความ	ความคิดเห็นของ ผู้เชี่ยวชาญ			ข้อเสนอแนะ
		+1	0	-1	
11	ฉันใช้เวลาว่างของฉันฝึกฝนภาษาอังกฤษด้วยตนเอง ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				
12	ฉันมักจะถามอาจารย์ทันทีเมื่อฉันรู้สึกไม่เข้าใจเนื้อหาที่เรียน (วิชาภาษาอังกฤษ) ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				
13	ฉันวางแผนการเรียนภาษาอังกฤษ เช่น ทำตารางอ่านหนังสือ ทำรายงาน หรือทำการบ้านภาษาอังกฤษในแต่ละวันและทำตามนั้น ☐ ใช่ ☐ ไม่ใช่				
14	ฉันเตรียมตัวก่อนเข้าเรียนวิชาภาษาอังกฤษ เช่น คำศัพท์อ่านบทเรียนล่วงหน้า ☐ ใช่ ☐ ไม่ใช่				
15	ฉันชอบนั่งใกล้ๆ เพื่อนที่ตั้งใจเรียน ☐ ใช่ ☐ ไม่ใช่				
16	ฉันมีเพื่อนเป็นชาวต่างชาติ เพื่อฝึกใช้ทักษะภาษาอังกฤษ ☐ ใช่ ☐ ไม่ใช่				
17	ฉันให้ความสนใจและตั้งใจเรียนวิชาภาษาอังกฤษในขณะที่อาจารย์สอน ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				

ข้อ	ข้อความ	ความคิดเห็นของ ผู้เชี่ยวชาญ			ข้อเสนอแนะ
		+1	0	-1	
18	ฉันจดบันทึกตามที่อาจารย์สอน ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				
19	ฉันมักตอบคำถามในขณะที่เรียนภาษาอังกฤษ ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				
20	ฉันทบทวนบทเรียนภาษาอังกฤษหลังเลิกเรียน ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				
21	ฉันศึกษาหรือค้นคว้าข้อมูลเพิ่มเติมเกี่ยวกับภาษาอังกฤษ นอกเหนือจากที่อาจารย์สอน ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				
22	ฉันจับกลุ่มทบทวนบทเรียนภาษาอังกฤษกับเพื่อนเมื่อใกล้ สอบ ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				

การวิจัยส่วนที่ 4: คำถามเกี่ยวกับด้านทัศนคติและแรงจูงใจ

ข้อ	ข้อความ	ความคิดเห็นของ ผู้เชี่ยวชาญ			ข้อเสนอแนะ
		+ 1	0	-1	
23	ฉันมีความต้องการศึกษาในระดับที่สูงขึ้น ☐ ใช่ ☐ ไม่ใช่				
24	ฉันได้รับการสนับสนุนด้านการเรียนภาษาอังกฤษจากครอบครัว ☐ ใช่ ☐ ไม่ใช่				
25	ฉันสนุกกับการเรียนภาษาอังกฤษ ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				
26	ฉันสนใจเรียนภาษาอังกฤษเพื่อพัฒนาตนเอง ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				
27	ฉันสนใจเรียนภาษาอังกฤษเพื่อหางานที่ดีในอนาคต ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				
28	ฉันสนใจเรียนภาษาอังกฤษ เพราะเป็นความต้องการของพ่อแม่ ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				
29	ฉันสนใจเรียนภาษาอังกฤษ เพื่อต้องการสอบผ่าน ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				
30	ฉันสนใจเรียนภาษาอังกฤษ เพื่อใช้สื่อสารในชีวิตประจำวัน				

ข้อ	ข้อความ	ความคิดเห็นของ ผู้เชี่ยวชาญ			ข้อเสนอแนะ
		+ 1	0	-1	
	☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				
31	ฉันมีความสุขในการอ่านหนังสือ หรือ ดูว่า ดูภาพยนตร์ที่เป็น ภาษาอังกฤษ ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				
32	การเรียนรู้ภาษาอังกฤษมีความสำคัญกับฉัน เพราะฉันต้องการได้เกรด ที่ดี ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				
33	การเรียนรู้ภาษาอังกฤษมีความสำคัญกับฉัน เพราะภาษาอังกฤษช่วย ให้ฉันมีความรู้มากขึ้น ☐ น้อยมาก ☐ น้อย ☐ ปานกลาง ☐ มาก ☐ มากที่สุด				

ส่วนที่ 2 กำหนดเกณฑ์คะแนนวิชาภาษาอังกฤษ

กลุ่มนักศึกษา	ช่วงคะแนน		ข้อเสนอแนะเพิ่มเติม
	ต่ำสุด	สูงสุด	
นักศึกษาที่มีผลการเรียนดี (เรียนเก่ง)			
นักศึกษาที่มีผลการเรียนปานกลาง			
นักศึกษาที่มีผลการเรียนต้องปรับปรุง			

ขอแสดงความขอบคุณอย่างยิ่ง

(นางสาวชิตชนก แก้วพานิชย์)

สาขาบริหารเทคโนโลยีสารสนเทศ คณะสถิติประยุกต์

ลงชื่อ.....

(.....)

ผู้ทรงคุณวุฒิ

ภาคผนวก ก
สรุปผลการคำนวณและแปลผลค่าแบบประเมินความเที่ยงตรง (IOC)

สรุปผลการคำนวณและแปลผลค่าแบบประเมินความเที่ยงตรงเชิงเนื้อหา (IOC) ของเครื่องมือการวิจัยเรื่อง “การวิเคราะห์คุณลักษณะสำคัญที่ส่งผลต่อประสิทธิภาพการเรียนรู้วิชาภาษาอังกฤษของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิคเหมืองข้อมูล” โดยผู้เชี่ยวชาญที่ให้ความกรุณาทำแบบประเมินเพื่อใช้สำหรับงานวิจัยนี้ มีรายนาม ดังต่อไปนี้

คนที่ 1 ดร.กฤษณีย์ เจริญจิตร คณบดี คณะภูมิสารสนเทศศาสตร์

คนที่ 2 อ.วุฒิชัย แก้วแหวน รองคณบดีฝ่ายบริหารและพัฒนางานองค์กร

คนที่ 3 ดร.ปัทมา พอดี รองคณบดีฝ่ายวิชาการและบริการวิชาการ

ข้อคำถาม	คะแนนความเห็นผู้เชี่ยวชาญตรวจสอบเครื่องมือวิจัย			รวม	ค่า IOC	สรุปผล
	คนที่ 1	คนที่ 2	คนที่ 3			
1	+1	0	0	1	0.33	ไม่ได้
2	+1	+1	+1	3	1	ใช้ได้
3	+1	+1	+1	3	1	ใช้ได้
4	+1	+1	+1	3	1	ใช้ได้
5	+1	0	+1	2	0.67	ใช้ได้
6	+1	+1	+1	3	1	ใช้ได้
7	+1	+1	+1	3	1	ใช้ได้
8	+1	+1	+1	3	1	ใช้ได้
9	0	0	+1	1	0.33	ใช้ได้
10	+1	+1	+1	3	1	ใช้ได้
11	+1	+1	+1	3	1	ใช้ได้
12	+1	+1	+1	3	1	ใช้ได้
13	+1	+1	+1	3	1	ใช้ได้
14	+1	+1	+1	3	1	ใช้ได้
15	0	0	+1	1	0.33	ใช้ได้
16	+1	+1	+1	3	1	ใช้ได้
17	+1	+1	+1	3	1	ใช้ได้
18	+1	+1	+1	3	1	ใช้ได้
19	+1	+1	+1	3	1	ใช้ได้
20	+1	+1	+1	3	1	ใช้ได้
21	+1	+1	+1	3	1	ใช้ได้
22	+1	+1	+1	3	1	ใช้ได้
23	+1	+1	+1	3	1	ใช้ได้
24	+1	+1	+1	3	1	ใช้ได้
25	+1	+1	+1	3	1	ใช้ได้
26	+1	+1	+1	3	1	ใช้ได้

ข้อคำถาม	คะแนนความเห็นผู้เชี่ยวชาญตรวจสอบเครื่องมือวิจัย			รวม	ค่า IOC	สรุปผล
	คนที่ 1	คนที่ 2	คนที่ 3			
27	+1	+1	+1	3	1	ใช้ได้
28	+1	+1	+1	3	1	ใช้ได้
29	+1	+1	+1	3	1	ใช้ได้
30	+1	+1	+1	3	1	ใช้ได้
31	+1	+1	+1	3	1	ใช้ได้
32	+1	+1	+1	3	1	ใช้ได้
33	+1	+1	+1	3	1	ใช้ได้

ส่วนที่ 2 กำหนดเกณฑ์คะแนนวิชาภาษาอังกฤษ

กลุ่มนักศึกษา	ช่วงคะแนน		ข้อเสนอแนะเพิ่มเติม
	ต่ำสุด	สูงสุด	
นักศึกษาที่มีผลการเรียนดี (เรียนเก่ง)			ควรใช้คะแนนแบบอิงเกณฑ์เป็นตัววัด
นักศึกษาที่มีผลการเรียนปานกลาง			
นักศึกษาที่มีผลการเรียนต้องปรับปรุง			