

การสร้างตัวแบบการทำนายในการเลือกศึกษาต่อในระดับอุดมศึกษา โดยใช้เทคนิค

แบบบูรณาการในการแก้ปัญหาการจำแนกข้อมูลไม่สมดุลของกลุ่มผู้เรียน

## The predictive model of higher education guidance using integrated techniques for imbalanced data of learner groups

อศวิน สุรวัชโยธิน\* และ วรภัทร ไพรีเกรง

*Atsawin Surawatchayotin\* and Worapat Paireekreng*

สาขาวิชาเทคโนโลยีสารสนเทศ วิทยาลัยนวัตกรรมและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์

*Program in Information Technology College of Innovative Technology and Engineering*

*Dhurakij Pundit University*

Received: January 31, 2021; Revised: March 25, 2021; Accepted: May 17, 2021; Published: June 25, 2021

**ABSTRACT** – An inappropriate selection of higher education entrance came from insufficient experience of student and in-depth knowledge of the subject area. Most of students rely on social norms and values to make a decision without consideration of individual skill. Therefore, in order to select the appropriate subject area, the integrated prediction model of higher education entrance based on student's multiple-intelligence and skills is necessary. According to the research, there is the imbalance data of subject areas. Hence, the integrated techniques which combined Single Imputation, SMOTE, and feature Selection were used to create prediction model. It also implemented Bagging, Boosting and Stacking in the experiments to address the problems. The experimental results found that the most importance skill using the stacking technique appeared to subject area of mathematical logic. There is an accuracrate of 77% performed better which is compared to other techniques. This subject is the fundamental of other knowledge areas. It affects the admission to higher education.

**KEYWORDS:** Imbalanced data, Recursive Feature Elimination, SMOTE, Ensemble learning-based model, Multiple-intelligence skills, Single Imputation

**บทคัดย่อ** -- การเลือกกลุ่มสายวิชาที่ไม่เหมาะสมเพื่อเรียนต่อในระดับอุดมศึกษา เกิดจากการที่นักเรียนยังขาดประสบการณ์และยังไม่วัฒนการกลุ่มสาขาวิชาที่พอ ส่วนใหญ่ใช้ความรู้สึกหรือค่านิยมเป็นหลักใหญ่ในการตัดสินใจ อาจไม่ทราบถึงความรู้เฉพาะด้านของกลุ่มสายวิชาที่เหมาะสมกับทักษะของตัวเอง ดังนั้นเพื่อให้นักเรียนสามารถเลือกเรียนในกลุ่มสายวิชาที่เหมาะสมด้วยการสร้างตัวแบบการทำนายในการเลือกศึกษาต่อตามทักษะของกลุ่มผู้เรียนในเชิงพหุปัญญาโดยใช้เทคนิคแบบบูรณาการ จากการวิจัยพบว่า ข้อมูลกลุ่มสายวิชาที่มีความไม่สมดุลกัน (Imbalanced Data) จึงแก้ปัญหาด้วยเทคนิคบูรณาการ คือ การแก้ปัญหาข้อมูลขาดหายด้วย Single Imputation การปรับสมดุลข้อมูลด้วย SMOTE การคัดเลือกลักษณะด้วย RFE และสร้างตัวแบบทำนาย 3 เทคนิค คือ Bagging, Boosting, Stacking ผลการวิจัยสรุปได้ว่าทักษะที่มีความสำคัญมากที่สุดโดยการใช้เทคนิค Stacking คือ ทักษะด้านเหตุผลและคณิตศาสตร์ มีค่าความสำคัญ 77% เมื่อ

\*Corresponding Author: [atsawinstfx8882@gmail.com](mailto:atsawinstfx8882@gmail.com)

เปรียบเทียบกับเทคนิคอื่น ซึ่งทักษะนี้เป็นรากฐานสำคัญของวิทยาการหลายสาขาส่งผลให้ทักษะนี้มีความสำคัญในการศึกษาต่อระดับอุดมศึกษา

**คำสำคัญ:** ข้อมูลไม่สมดุล, การคัดกรองคุณลักษณะแบบทำซ้ำ, การสุ่มเพิ่มตัวอย่างเพิ่มจากกลุ่มส่วนน้อยด้วยการสังเคราะห์ข้อมูลใหม่, แบบจำลองการเรียนรู้แบบรวมกลุ่ม, ทักษะการเรียนรู้ในเชิงพหุปัญญา, การแทนค่าสูญหายด้วยค่าเดียว

## 1. บทนำ

### 1.1 ที่มาของปัญหา

ถ้าจะกล่าวถึงประเด็นปัญหาความไม่สอดคล้องทางการศึกษาของการศึกษาต่อในระดับอุดมศึกษา เมื่อพิจารณาแล้วอาจเริ่มต้นมาจากการตัดสินใจในการเลือกศึกษาต่อที่ไม่เหมาะสมกับทักษะของผู้เรียน ทั้งนี้พบว่า การตัดสินใจในการเลือกศึกษาต่อในระดับอุดมศึกษาของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 ที่ไม่เหมาะสมกับทักษะของตนเองนั้น ส่งผลกระทบเชิงลบกับตัวผู้เรียนอย่างมากมาย ทั้งในแง่ของการเรียนและการประกอบอาชีพในอนาคต โดยที่นักเรียนไม่ทราบทักษะความรู้ ความถนัด หรือความต้องการที่แท้จริงของตนเอง อาจทำให้นักเรียนเลือกการเรียนและการประกอบอาชีพที่ผิด ย่อมเกิดผลเสียต่าง ๆ มากมาย เช่น ไม่สามารถเรียนและทำงานให้มีประสิทธิภาพได้ อาจเกิดความเบื่อหน่ายในสิ่งที่กำลังทำ และเกิดความล้มเหลวในการดำเนินชีวิต ส่งผลกระทบต่อด้านสังคมและเศรษฐกิจของประเทศ นอกจากนี้การทำงานในตำแหน่งที่ไม่เหมาะสมกับคุณวุฒิ การศึกษามีผลต่อการได้รับค่าจ้างต่ำกว่าคุณวุฒิ สถานการณ์ความไม่สอดคล้องกันในตลาดแรงงานส่วนหนึ่งมีสาเหตุมาจากจำนวนผู้ที่สำเร็จการศึกษาในแต่ละปีมีจำนวนมากกว่าตำแหน่งงาน ส่งผลให้มีจำนวนผู้ว่างงานมากกว่าผู้มีงานทำ โดยเฉพาะผู้สำเร็จการศึกษาในระดับอุดมศึกษา เป็นกลุ่มที่ว่างงานมากที่สุด สาขาวิชาที่ว่างงานมากที่สุดคือพาณิชยศาสตร์ คิดเป็น 20.2% รองลงมาคือ ศึกษาศาสตร์ 17.1% และ สังคมศาสตร์ 11.7% รองลงมาคือไม่สามารถหางานที่เหมาะสมได้ 14.6% และอีกส่วนหนึ่งคือการเลือกงาน สถานการณ์แรงงานที่มีการแข่งขันกันสูงดังกล่าวนำไปสู่ให้ผู้สำเร็จการศึกษาส่วนหนึ่งไม่ต้องการรองานจึงสมัครใจทำงานในตำแหน่งที่คุณวุฒิไม่เหมาะสมกับความรู้หรือความสามารถหรือทักษะของตนเอง ผลกระทบเชิงลบที่เกิดขึ้นคือการได้รับค่าจ้างต่ำกว่าคุณวุฒิ เกิดความไม่พึงพอใจในงาน และอาจเกิดความเครียดถึงขั้นทำร้ายตัวเอง [1] อีกทั้งสถานการณ์

ตลาดแรงงานไทยยังมีปัญหาความไม่สอดคล้องด้าน “Field of Study Mismatch” ด้วยเช่นกัน กล่าวคือ คนไทยนิยมเรียนด้าน สังคมศาสตร์ โดยในระดับมัธยมศึกษาเรามีนักเรียนเรียนสายศิลป์อยู่ถึง 70% และอีก 30% อยู่ในสายวิทยาศาสตร์ [2] มีงานวิจัยทางการศึกษาที่พบว่า นักเรียนส่วนใหญ่ที่ต้องการเลือกศึกษาต่อในระดับอุดมศึกษา นักเรียนต้องการเลือกเรียนคณะทางสายครุศาสตร์/ศึกษาศาสตร์ มากที่สุด รองลงมาได้แก่คณะที่เปิดสอนในสาขาวิทยาการจัดการ/บริหารธุรกิจ/พาณิชยศาสตร์ [3] อย่างไรก็ตามเพื่อเป็นการช่วยลดปัญหา Personality Mismatch รวมทั้งฝึกฝนทักษะแบบบูรณาการ ไม่ว่าจะเป็นทักษะการคิดแบบเชิงวิพากษ์ ทักษะด้านเทคโนโลยี และทักษะด้านภาษาอังกฤษ โดยเริ่มจากการคิดเป็น เป็นการนำเอาความรู้เกี่ยวกับตนเอง ความรู้ทางด้านสังคม และความรู้ทางด้านวิชาการ นำมาใช้ในการแก้ปัญหาหรือประกอบการตัดสินใจ หรือ การมีเขาวนปัญญา ในการรู้จักตนเองตามความเป็นจริงเกี่ยวกับจุดอ่อนจุดแข็งของตนเอง (Self-awareness) ตามทฤษฎีพหุปัญญา (Multiple- Intelligence Theory) และสามารถนำความรู้จักตนเองนี้ไปใช้ในการคิดพิจารณาในการกำหนดเป้าหมายและตัดสินใจนำตนเองไปสู่ความสำเร็จ [4]

### 1.2 วัตถุประสงค์ของการวิจัย

โดยในงานวิจัยนี้มีวัตถุประสงค์หลักอยู่ 2 ประการ คือ งานวิจัยนี้ได้จัดทำขึ้นโดยมีจุดมุ่งหมายและวัตถุประสงค์อยู่ 2 ประการคือ (1) เพื่อสร้างตัวแบบการทำนายในการเลือกศึกษาต่อในระดับอุดมศึกษาตามกลุ่มผู้เรียนโดยใช้เทคนิคแบบบูรณาการในการแก้ปัญหา Imbalanced Data ของกลุ่มผู้เรียน (2) เพื่อศึกษาตัวแบบความสำคัญที่เกี่ยวข้องกับการทำนายกลุ่มผู้เรียนกับกลุ่มสาขาวิชาที่เหมาะสมตามทักษะของผู้เรียนในเชิงพหุปัญญา โดยผู้วิจัยได้ทำการเก็บรวบรวมข้อมูลนักเรียน จาก 3 แผนการเรียน คือ ศิลป์คำนวณ วิทยาศาสตร์ และศิลป์ภาษา ตั้งแต่ปี พ.ศ. 2560 -

2562 จากโรงเรียนเซนต์ฟรังซิสเซเวียร์ จำนวน 312 คน ข้อมูลที่เก็บรวบรวมมาจากแบบสอบถามออนไลน์ ผ่านการประเมินความเที่ยงตรง (IOC) จากผู้เชี่ยวชาญ เพื่อนำเข้าสู่กระบวนการสร้างตัวแบบในการทำนายการเลือกศึกษาต่อในระดับอุดมศึกษาเบื้องต้นจากการเก็บรวบรวมข้อมูลผู้วิจัยพบว่า จำนวนข้อมูลของกลุ่มผู้เรียน กับกลุ่มสายวิชาไม่มีความไม่สมดุลกัน (Imbalanced Data) โดยบางกลุ่มสายวิชานักเรียนมีความสนใจน้อยกว่าบางกลุ่มสายวิชา ดังนั้นในการศึกษางานวิจัยนี้ทางผู้วิจัยได้เล็งเห็นว่าพื้นที่สำคัญที่จะช่วยในการแก้ปัญหาความไม่สอดคล้องของการเลือกศึกษาต่อในระดับอุดมศึกษาของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 กับทักษะของผู้เรียน ด้วยการใช้เทคนิคแบบบูรณาการในการแก้ปัญหา Imbalanced Data ของกลุ่มผู้เรียน อีกทั้งเพื่อเป็นการเพิ่มสมรรถนะสูงสุดให้กับตัวแบบการทำนายทางผู้วิจัยจึงได้นำวิธีการคัดเลือกคุณลักษณะ (Feature Selection: FS) ของตัวแปรที่สำคัญและเกี่ยวข้องกับการทำนายกลุ่มผู้เรียนกับกลุ่มสายวิชา ด้วยเทคนิคแบบทำซ้ำ (Recursive Feature Elimination : RFE) และเพิ่มประสิทธิภาพความแม่นยำของตัวแบบการทำนายด้วยเทคนิคของแบบจำลองการเรียนรู้แบบรวมกลุ่มด้วย 3 เทคนิค คือ Bagging, Boosting และ Stacking ซึ่งจะส่งเสริมให้ประสิทธิภาพในการทำนายกลุ่มผู้เรียนกับกลุ่มสายวิชาที่ความแม่นยำและถูกต้องมากยิ่งขึ้น

## 2. วิธีดำเนินการวิจัย

### 2.1 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

กลุ่มสาขาวิชา (Subject Area) ในปัจจุบันมีการจำแนกออกเป็นหลายสาขา ซึ่งในแต่ละสาขาวิชา มีการศึกษารายละเอียดที่ลึกซึ้งและแยกออกเป็นสาขาย่อยอีกเป็นจำนวนมาก การจำแนกสาขาวิชาสามารถแบ่งออกได้เป็น 3 ประเภทใหญ่ ได้แก่ สาขาวิทยาศาสตร์ (Sciences) สาขาวิชาสังคมศาสตร์ (Social Sciences) และสาขาวิชามนุษยศาสตร์ (Humanities) โดยแบ่งตามลักษณะและขอบเขตของข้อเท็จจริงที่ศึกษา สาเหตุของข้อเท็จจริง และการประเมินคุณค่าของข้อเท็จจริงนั้น รวมทั้งวิธีการศึกษา และการแสวงหาความรู้ในแต่ละสาขาวิชา [5]

จากการเก็บรวบรวมของสำนักงานคณะกรรมการการอุดมศึกษา ได้มีการสรุปข้อมูลนักศึกษาใหม่ปี 2563 ในสถาบันอุดมศึกษาทั้งหมด โดยแบ่งตามกลุ่มสาขาวิชาของ UNESCO ไว้ดังนี้ Education จำนวน 29,315 จำนวน Arts and

humanities 45,694 จำนวน Social sciences จำนวน 28,204 Business administration จำนวน 104,525 Natural science จำนวน 15,874 ICT จำนวน 18,190 Engineering จำนวน 49,221, Agriculture จำนวน 9,213 Health and welfare จำนวน 28,301 Services จำนวน 16,301 รวมทั้งสิ้น 345,038 คน [6] ทั้งนี้พบว่าในแต่ละมหาวิทยาลัยมีการเปิดสอนตามหลักสูตรที่ไม่เหมือนกัน จึงมีชื่อคณะ และสาขาที่แตกต่างกัน ดังนั้นทางผู้วิจัยจึงมีการวิเคราะห์จำแนกในแต่ละประเภทกลุ่มสาขาวิชาจากข้อมูลของสำนักงานคณะกรรมการการอุดมศึกษา โดยแบ่งออกเป็น 8 กลุ่มสายวิชาดังนี้ กลุ่มสายวิชาวิทยาศาสตร์, กลุ่มสายวิชาวิศวกรรมศาสตร์และเทคโนโลยี, กลุ่มสายวิชาวิทยาศาสตร์สุขภาพ, กลุ่มสายวิชาบริหารธุรกิจ, กลุ่มสายวิชาภาษา, กลุ่มสายวิชาสังคม, กลุ่มสายวิชาการศึกษา และ กลุ่มสายวิชาศิลปะ เพื่อนำเข้าสู่กระบวนการวิเคราะห์และสร้างตัวแบบการทำนายต่อไป

ความไม่สมดุลของข้อมูล (Imbalanced data) เป็นปัญหาของข้อมูลที่มีจำนวนข้อมูลในกลุ่มหนึ่งมากกว่าจำนวนข้อมูลของอีกกลุ่มหนึ่งเป็นจำนวนมาก อาจเกิดขึ้นเนื่องจากลักษณะทางธรรมชาติของข้อมูลที่มีความแตกต่างของจำนวนในแต่ละกลุ่ม หรืออาจเกิดจากข้อจำกัดในการเก็บข้อมูล ซึ่งเกิดขึ้นได้ทั่วไปในหลายๆ สถานการณ์ [7] อาทิเช่น ได้มีงานวิจัยทางด้านการแพทย์ที่นำวิธีการแก้ปัญหาค่าความไม่สมดุลของข้อมูลมาใช้กับข้อมูลผู้ประกอบการร้านอาหาร แบ่งเป็นผู้ประกอบการที่มีรายได้ น้อยและกลุ่มผู้ประกอบการที่มีรายได้สูง โดยมีจำนวน 72 และ 39 ตามลำดับ ซึ่งมีจำนวนในแต่ละกลุ่มที่แตกต่างกันเกือบเท่าตัว ดังนั้นวิธีการจำแนกประเภทจะให้ผลที่เอนเอียงเนื่องจากจำนวนของทั้งสองกลุ่มแตกต่างกันมาก จึงควรมีการปรับข้อมูลให้มีความสมดุล และเนื่องจากจำนวนข้อมูลทั้งหมดนั้นมีน้อย ดังนั้นการสุ่มตัวอย่างเพิ่มเป็นทางเลือกที่เหมาะสมด้วยการสุ่มตัวอย่างเพิ่มกลุ่มส่วนน้อยด้วยการสังเคราะห์โดยวิธี SMOTE การที่ข้อมูลไม่สมดุลจะส่งผลกระทบต่อการทำนายประเภทกลุ่มส่วนน้อย (Minority Class) ที่เป็นกลุ่มหายาก เนื่องจากวิธีการจำแนกข้อมูลทั่วไปจะมีประสิทธิภาพในการทำนายข้อมูลก็ต่อเมื่อข้อมูลในแต่ละกลุ่มมีจำนวนใกล้เคียงกัน ซึ่งหากข้อมูลไม่สมดุลวิธีการจำแนกประเภททั่วไปมีความเอนเอียงที่จะทำนายเป็นกลุ่มส่วนมาก (Majority Class) การทำให้ข้อมูลทั้งสองกลุ่มที่มีจำนวนใกล้เคียงกันทำได้โดยการสุ่มตัวอย่างซ้ำ (Re-sampling) ซึ่งทำได้ใน 2 ลักษณะ ได้แก่ การสุ่มตัวอย่างลด (Under-sampling) กลุ่มส่วนมาก และการสุ่มตัวอย่างเพิ่ม (Over-

sampling) กลุ่มส่วนน้อย โดยตัวอย่างที่เพิ่มหรือลดสามารถทำได้ด้วยการสุ่มตัวอย่างเพิ่มจากข้อมูลเริ่มต้นอย่างสุ่ม (Random Over-sampling: ROS) หรือการสุ่มตัวอย่างลดอย่างสุ่ม (Random Under-sampling: RUS) การสุ่มตัวอย่างลดจะเหมาะกับข้อมูลขนาดใหญ่ ซึ่งมีข้อเสียคือทำให้สูญเสียสารสนเทศของข้อมูล ส่วนการสุ่มตัวอย่างเพิ่มเหมาะกับข้อมูลที่ขนาดไม่ใหญ่ แต่มีข้อเสียคือทำให้ใช้เวลาในการประมวลผลเพิ่มขึ้น [8] ได้มีงานวิจัยทางด้านเหมืองข้อมูลที่น่าสนใจที่นำข้อมูลจำนวนกลุ่มนักศึกษาที่ไม่สำเร็จการศึกษาตามแผนมากกว่ากลุ่มนักศึกษาที่ไม่สำเร็จการศึกษาตามแผน ผู้วิจัยแก้ปัญหาโดยการนำวิธีการสุ่มเพิ่มตัวอย่างข้อมูลกลุ่มน้อย (SMOTE) เป็นเทคนิคการสุ่มตัวอย่างโดยใช้ข้อมูลเดิมแต่ละทำการสังเคราะห์ข้อมูลขึ้นมาใหม่จากข้อมูลเดิมที่มีอยู่ด้วยหลักการเพื่อช่วยในการปรับสมดุลของข้อมูลเพื่อให้จำนวนของคลาสรอง (Minority Class) มีจำนวนเทียบเท่าหรือใกล้เคียงกับจำนวนข้อมูลที่เป็นคลาสหลัก (Majority Class) [9] จากงานวิจัยที่กล่าวมาข้างต้นพบว่าถ้าข้อมูลมีปัญหาความไม่สมดุล ควรมีการปรับข้อมูลให้มีความสมดุลก่อน ซึ่งจะช่วยให้การจำแนกประเภทข้อมูลสามารถทำงานได้อย่างมีประสิทธิภาพมากยิ่งขึ้น และจากการเปรียบเทียบผลการทดลองของงานวิจัยที่ได้นำมาทบทวน ทางผู้วิจัยพบว่าเทคนิคการแก้ปัญหาความไม่สมดุลของข้อมูลที่มีประสิทธิภาพที่สุดนั่นก็คือ เทคนิคการสังเคราะห์ข้อมูลใหม่ (SMOTE)

การคัดเลือกคุณลักษณะ (Feature Selection) เป็นวิธีการในการคัดเลือกคุณลักษณะที่มีความสำคัญและใช้เพื่อลดขนาดหรือมิติของข้อมูลลง การคัดเลือกคุณลักษณะจึงเป็นหัวใจสำคัญในการทำเหมืองข้อมูล (Data Mining) ซึ่งในปัจจุบันมีวิธีที่น่าสนใจอยู่ด้วยกันหลายวิธี โดยในงานวิจัยนี้จะกล่าวถึงวิธีการหลักๆ อยู่ด้วยกัน 3 วิธี คือ (1) วิธีฝังตัว (Embedded methods) เป็นวิธีการคัดเลือกคุณลักษณะเป็นส่วนหนึ่งที่อยู่ภายในกระบวนการจำแนกประเภทด้วย เช่น กระบวนการสร้างต้นไม้การตัดสินใจ (Decision Tree) โดยจะทำการคัดเลือกเซตย่อยของคุณลักษณะที่เหมาะสมในแต่ละขั้นตอนวิธี [10] (2) วิธีแบบกรอง (Filter-based) เป็นวิธีที่ใช้การประเมินความมีประโยชน์ของแต่ละลักษณะเฉพาะต่อการจำแนกข้อมูล โดยเรียกว่า ค่าคะแนนลักษณะเฉพาะ (Feature Score) โดยขั้นตอนนี้จะเลือกคุณลักษณะที่มีค่าคะแนนสูง [Top Score Feature] มากกว่าคุณลักษณะที่มีค่าคะแนนต่ำ [11] (3) วิธีการคัดเลือกแบบ

ห่อหุ้ม (Wrapper-based) จะใช้อัลกอริทึมการค้นหา (Searching Algorithm) และอัลกอริทึมการเรียนรู้ (Learning Algorithm) ที่เตรียมไว้แล้วสำหรับกลุ่มของลักษณะเฉพาะที่เหมาะสมกับอัลกอริทึมการเรียนรู้ นั้น ทำให้วิธีการนี้มีความแม่นยำในการจำแนกข้อมูลได้ดีกว่าตัวกรอง (Part, 2017, 101-102) จากงานวิจัยที่กล่าวมาข้างต้นพบว่าวิธีคัดเลือกแบบห่อหุ้ม (Wrapper-based) เป็นวิธีการคัดเลือกคุณลักษณะที่มีความแม่นยำในการจำแนกข้อมูลสำหรับปัญหาที่เฉพาะเจาะจง ดังนั้นในกระบวนการคัดเลือกคุณลักษณะเนื่องจากวิธีการคัดเลือกคุณลักษณะแบบห่อหุ้มให้ความแม่นยำสูง ทางผู้วิจัยจึงเลือกใช้วิธีการดังกล่าว ด้วยเทคนิค Recursive Feature Elimination เพื่อคัดเลือกคุณลักษณะที่สำคัญและเกี่ยวข้องกับการทำนายกลุ่มผู้เรียนกับกลุ่มสายวิชา ที่เหมาะสมตามทักษะของผู้เรียนในเชิงพหุปัญญา เพิ่มประสิทธิภาพความแม่นยำของตัวแบบการทำนาย

แบบจำลองการเรียนรู้แบบเชิงเดี่ยว (Individual Standard Model) เป็นแบบจำลองการเรียนรู้จากเพียงตัวจำแนกข้อมูล (Classifier) เพียงตัวเดียวโดยไม่ใช้ควบคู่กับตัวอื่น ได้มีงานวิจัยทางด้านเหมืองข้อมูลที่น่าสนใจในการพัฒนาคุณภาพการศึกษาคณะวิศวกรรมศาสตร์ โดยใช้ข้อมูลประวัติส่วนตัวของนิสิตและข้อมูลการลงทะเบียนของนิสิตมาช่วยพิจารณาหารูปแบบความสัมพันธ์ของข้อมูล และนำไปประยุกต์ใช้ในการช่วยค้นหาภาควิชาที่เหมาะสมกับนิสิต เพื่อเป็นแนวทางในการแก้ไขปัญหาคือการเลือกสาขาวิชาที่ไม่ตรงกับความสามารถ และปัญหาผลการเรียนที่ตกต่ำจนต้องออกจากสถาบันการศึกษา [12]

แบบจำลองการเรียนรู้แบบรวมกลุ่ม (Ensemble Learning-based Model) เป็นการสร้างแบบจำลองการเรียนรู้แบบรวมกลุ่มที่มีการใช้ตัวจำแนก (Classifier) มากกว่าหนึ่งตัวในการเรียนรู้ แต่ละตัวจำแนกจะมีกระบวนการทำงานของตัวเองและทุกตัวจำแนกจะกระทำกับข้อมูลเดียวกัน เมื่อได้ผลการจำแนกของแต่ละตัวจำแนกแล้ว ก็จะนำผลลัพธ์เหล่านั้นมาผ่านวิธีการรวบรวม (Combination หรือ Vote) และสุดท้ายนำไปตัดสินใจโดยเทคนิคดังกล่าวมีจุดมุ่งหมายในการเพิ่มประสิทธิภาพความแม่นยำให้กับแบบจำลองการเรียนรู้ [13] โดยที่ Ensemble Classification มีเทคนิคอยู่ด้วยกัน 3 แบบ คือ Bagging, Boosting, และ Stacking จากการทบทวนวรรณกรรมมีงานวิจัยที่มีการประยุกต์แบบจำลองแบบรวมกลุ่มในการจำแนกข้อมูล เช่น ได้มีงานวิจัยทางการศึกษามีการวิเคราะห์ความคิดเห็น

เกี่ยวกับการจัดการเรียนการสอนด้วยการวิเคราะห์ความรู้สึกโดยใช้เทคนิคการเรียนรู้แบบรวมกลุ่มในการจำแนกข้อมูล เพื่อเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลจากการทดลองพบว่า เทคนิควิธีการรวมกลุ่มเพื่อจำแนกข้อมูลแบบ Majority Vote มีประสิทธิภาพมากที่สุด [14] ได้มีงานวิจัยทางด้านเหมืองข้อมูลที่มีการประยุกต์ใช้ข้อมูลเกี่ยวกับธุรกิจด้านอสังหาริมทรัพย์ร่วมกับเทคนิคการเรียนรู้แบบกลุ่ม (Ensemble Learning-based Model) ในการกำหนดระดับของโครงการอสังหาริมทรัพย์เพื่อช่วยในการตัดสินใจให้กับนักลงทุนในการกู้เงินเพื่อลงทุน จากผลการทดลองพบว่า การจำแนกข้อมูลด้วยเทคนิคการเรียนรู้แบบกลุ่มมีประสิทธิภาพที่ดีกว่าการจำแนกข้อมูลด้วยเทคนิคอื่น ๆ [15] ได้มีงานวิจัยทางด้านเหมืองข้อมูลที่มีการประยุกต์ใช้แบบจำลองการเรียนรู้แบบรวมกลุ่มเพื่อพยากรณ์แนวโน้มผลตอบแทนการลงทุนในตลาดหลักทรัพย์แห่งประเทศไทย โดยใช้ข้อมูลหลักทรัพย์ที่อยู่ในดัชนี SET 50 ของตลาดหลักทรัพย์แห่งประเทศไทย และข้อมูลดัชนีชี้วัดทางเทคนิคเพื่อเปรียบเทียบประสิทธิภาพ และผลตอบแทนในการลงทุน โดยผลการทดลองให้เห็นว่า แบบจำลองการเรียนรู้แบบกลุ่ม ด้วยวิธีการ Weight Average ช่วยเพิ่มประสิทธิภาพในความแม่นยำได้ดีกว่าการจำแนกข้อมูลด้วยเทคนิคอื่น ๆ [16]

การเรียนรู้แบบกลุ่มด้วยเทคนิคลงจำแนก Bagging Classification เป็นเทคนิคสำหรับช่วยเพิ่มประสิทธิภาพในการสร้างแบบจำลองโดยทำงานร่วมกับเทคนิคการเรียนรู้ต่าง ๆ เช่น ต้นไม้ตัดสินใจ (Decision Tree) โดยเริ่มจากการสุ่มเลือกข้อมูลด้วยหลักการสุ่มแบบ Bootstrap ขึ้นมาหลายๆ ชุด ในจำนวนเท่าข้อมูลเดิม แล้วนำข้อมูลที่สุ่มขึ้นมาไปใช้สร้างแบบจำลองทุกชุด จากนั้นใช้หลักการโหวต (Vote) เพื่อหาแบบจำลองที่สามารถตอบข้อมูลทดสอบได้ดีที่สุดมาเป็นแบบจำลองสำหรับการจำแนกข้อมูลชุดนั้น [17] ได้มีงานวิจัยทางการแพทย์ที่นำมาใช้เทคนิคการสุ่มตัวอย่างให้กับข้อมูลที่ไม่สมดุล ซึ่งเป็นข้อมูลทางการแพทย์ที่มีความไม่สมดุลระหว่างผู้ป่วยและผู้ไม่ป่วย และมีการใช้เทคนิค Bagging เพื่อเปรียบเทียบเทคนิควิธีการจำแนกข้อมูลที่มีประสิทธิภาพที่เหมาะสมกับข้อมูลของผู้ป่วย [18]

การเรียนรู้แบบกลุ่มด้วยเทคนิค Boosting Classification เป็นแบบจำลองการเรียนรู้แบบกลุ่มที่ใช้วิธีการเรียนรู้แบบเป็นลำดับ โดยนำเอาข้อผิดพลาดที่ได้จากการเรียนรู้ของตัวจำแนก

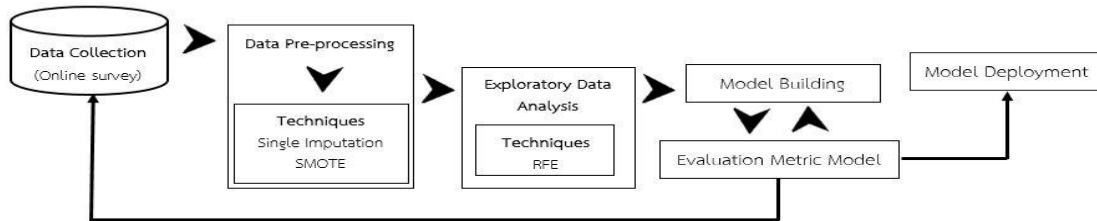
ตัวแรกมาปรับปรุงกับตัวจำแนกข้อมูลตัวถัดไป เพื่อช่วยลดข้อผิดพลาด ส่งผลให้มีค่าความถูกต้องเพิ่มมากขึ้น แนววิธีแบบชุดของการทำงานระบบการเรียนรู้หลายระบบ ได้มีงานวิจัยทางด้านเหมืองข้อมูลที่น่าเทคนิคต้นไม้มันช่วยตัดสินใจ ร่วมกับเทคนิค Boosting โดยได้ทดสอบประสิทธิภาพของแบบจำลองกับข้อมูลจาก UCI ผลลัพธ์ที่ได้แสดงให้เห็นว่าเทคนิคดังกล่าวให้ประสิทธิภาพที่ดีกว่าแบบจำลองการเรียนรู้แบบเชิงเดี่ยว [19]

การเรียนรู้แบบกลุ่มด้วยเทคนิค Stacking Classification เป็นแบบจำลองการเรียนรู้แบบกลุ่ม โดยรวมตัวจำแนกข้อมูลการทำนายที่แตกต่างกันมากกว่า 1 ประเภทมารวมกันไว้ในที่เดียวและมีขั้นตอนการทำนายมากกว่า 1 ชั้น และมีการใช้ K-fold Cross validation เพื่อสร้างการรวมการถ่วงน้ำหนักที่ดีที่สุดของการทำนาย ได้มีงานวิจัยทางการแพทย์นำเสนอแบบจำลองการเรียนรู้แบบหลายชั้น (Multi Kernels on Multiple Layers:MKML) ซึ่งเป็นเป็นการเรียนรู้แบบกลุ่มประเภทหนึ่งแต่มีหลายชั้นหรือเรียกว่า Stacking Ensemble Model ผลการทดลองพบว่าแบบจำลองการเรียนรู้แบบ MKML ช่วยเพิ่มประสิทธิภาพความเที่ยงตรงมากกว่าแบบจำลองการเรียนรู้การทำนายแบบเดี่ยว อีกทั้งยังสามารถนำไปใช้เพื่อช่วยเหลือแพทย์ในการวินิจฉัยผู้ป่วยที่มีความเสี่ยงต่อการเป็นโรคเบาหวานได้แม่นยำยิ่งขึ้น [20] กล่าวโดยสรุป เมื่อพิจารณาอย่างรอบด้านแล้วผู้วิจัยตระหนักว่าในงานวิจัยนี้ควรมีการใช้เทคนิคในการจัดการข้อมูลที่ไม่สมดุล ด้วยการใช้เทคนิคการสุ่มเพิ่มตัวอย่างข้อมูลจากกลุ่มส่วนน้อยด้วยเทคนิคการสังเคราะห์ข้อมูลใหม่ (SMOTE) ร่วมกับกระบวนการคัดเลือกคุณลักษณะ ด้วยเทคนิค Recursive Feature Elimination เพื่อช่วยลดขนาดหรือมิติของข้อมูลลงเพื่อให้ได้คุณลักษณะที่สำคัญและเกี่ยวข้องกับกลุ่มผู้เรียนกับกลุ่มสายวิชา ที่เหมาะสมตามทักษะของผู้เรียนในเชิงพหุปัญญา และก็นำข้อมูลที่ได้เข้าสู่กระบวนการจำแนกข้อมูลด้วยวิธีการเรียนรู้แบบกลุ่ม (Ensemble Learning-based Model) ด้วยเทคนิค Bagging, Boosting และ Stacking เพื่อเป็นการเปรียบเทียบประสิทธิภาพของแต่ละเทคนิคในการทำนายผลลัพธ์จากคุณลักษณะที่สำคัญต่อการเลือกกลุ่มสายวิชาในการศึกษาต่อในระดับอุดมศึกษา ได้อย่างถูกต้องและมีประสิทธิภาพ

## 2.2 กรอบแนวคิดในการวิจัย

ผู้วิจัยได้ใช้เทคนิคแบบบูรณาการ เริ่มต้นจากการแก้ปัญหาข้อมูลขาดหายและข้อมูลไม่สมดุล ด้วยเทคนิค Single Imputation การสังเคราะห์ข้อมูลใหม่ ด้วยเทคนิค SMOTE การคัดเลือกคุณลักษณะด้วยเทคนิค RFE

และการสร้างตัวแบบการทำนายในการจำแนกข้อมูลด้วยการสร้างแบบจำลองการเรียนรู้แบบรวมกลุ่ม (Ensemble Learning-based Model) ด้วย 3 เทคนิค Bagging, Boosting และ Stacking โดยมีขั้นตอนการดำเนินการดังรูปที่ 1



รูปที่ 1. กรอบแนวคิดการสร้างตัวแบบการทำนายการเลือกศึกษาต่อระดับอุดมศึกษาด้วยเทคนิคแบบบูรณาการ

## 2.3 ขั้นตอนการเก็บรวบรวมข้อมูล (Data Collection)

งานวิจัยนี้เริ่มต้นเก็บรวบรวมข้อมูลจากแบบสอบถามออนไลน์ที่ผ่านการประเมินความเที่ยงตรงจากผู้เชี่ยวชาญ (IOC) ของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 6 จาก 3 แผนการเรียน ศิลป์คำนวณ วิทยาศาสตร์ และศิลป์ภาษา ตั้งแต่ปี พ.ศ. 2560 - 2562 จำนวน 312 คน ประกอบด้วยข้อมูลด้านประชากรศาสตร์

(Demographic data) ข้อมูลทักษะการเรียนรู้ในเชิงพหุปัญญา และข้อมูลกลุ่มสายวิชา ในแง่ของสถิติผู้วิจัยกำหนดขนาดกลุ่มตัวอย่างตามแนวคิดของ Taro Yamane[21] จำนวนกลุ่มตัวอย่างที่ได้ 175 คน โดยข้อมูลจะประกอบไปด้วย เพศ ทักษะการเรียนรู้ในเชิงพหุปัญญาทั้ง 9 ด้าน และกลุ่มสายวิชาแบ่งเป็นคลาสจำนวน 8 กลุ่ม โดยสรุปมีรายละเอียด ดังตารางที่ 1 และ 2

ตารางที่ 1. แสดงตัวอย่างชุดข้อมูลของกลุ่มผู้เรียน

| Gender | Lin   | Log   | Spa   | Mus   | Bod   | Inter | Intra | Nat   | Exi   | Class |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Male   | 3     | 5     | 3     | 3     | 5     | 4     | 4     | 5     | 5     | 2     |
| Female | 5     | 4     | 4     | 3     | 3     | 5     | 4     | 5     | 4     | 5     |
| Male   | 4     | 4     | 1     | 3     | 3     | 4     | 5     | 5     | 4     | 3     |
| .....  | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... |
| Female | 1     | 4     | 4     | 1     | 1     | 4     | 3     | 3     | 4     | 3     |
| Female | 5     | 4     | 3     | 4     | 4     | 3     | 4     | 4     | 4     | 5     |
| Male   | 5     | 4     | 4     | 1     | 3     | 4     | 5     | 3     | 4     | 4     |
| .....  | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... | ..... |

## 2.4 ขั้นตอนการเตรียมข้อมูล (Data Pre-processing)

ในเบื้องต้นจากการเก็บรวบรวมข้อมูล พบว่าข้อมูลมีความไม่สมบูรณ์และมีค่าที่ขาดหายไป (Missing Data) จึงทำให้ข้อมูลของผู้เรียนบางรายไม่ครบหรือขาดหายไป อีกทั้งชนิดของข้อมูลยังไม่เหมาะสมกับการนำไปวิเคราะห์ เนื่องจากข้อมูลที่เก็บรวบรวมมาจากการวัดทัศนคติเป็นข้อมูลเชิงคุณภาพ (Qualitative data) เช่น ข้อมูลเพศ ข้อมูลระดับคะแนน ที่มีช่วงเท่าแต่ไม่สามารถนำมาคำนวณได้ (Categorical Data) จึงทำให้

ไม่สามารถนำข้อมูลเข้าสู่กระบวนการคัดเลือกคุณลักษณะ ได้ ซึ่งปัญหาดังกล่าวอาจส่งผลกระทบต่อประสิทธิภาพในขั้นตอนการวิเคราะห์ได้ ทางผู้วิจัยจึงได้ทำการแก้ปัญหาด้วยการแปลงข้อมูลเชิงคุณภาพให้เป็นข้อมูลเชิงปริมาณแบบต่อเนื่อง (Continuous data) หลังจากแปลงชนิดข้อมูลแล้วผู้วิจัยพบปัญหาข้อมูลที่ขาดหายไปจึงแก้ปัญหาด้วยวิธี Single Imputation ซึ่งมีข้อดี คือ ทำให้จำนวนตัวอย่างไม่ลดลงและวิธีที่นิยมใช้มากที่สุด คือ การแทนค่าด้วยวิธีการคำนวณค่าเฉลี่ย ซึ่งวิธีการนี้ไม่ส่งผลกระทบต่อ

ต่อการวิเคราะห์ข้อมูล เนื่องจากงานวิจัยนี้ไม่ได้มีการวิเคราะห์ ซึ่งข้อมูลจะประกอบไปด้วย เพศ ระดับคะแนนของทักษะการ  
ความสัมพันธ์ระหว่างตัวแปร อีกทั้งวิธีการ Single Imputation เรียนรู้ของกลุ่มผู้เรียนในเชิงพหุปัญญา ดังตารางที่ 3  
เป็นโมเดลที่นิยมใช้สำหรับการทำงานด้านเหมืองข้อมูล[22]

ตารางที่ 2. แสดงข้อมูลคุณลักษณะเฉพาะของกลุ่มผู้เรียน

| Feature              | Description                                                                                                                               |
|----------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| Gender               | เพศ                                                                                                                                       |
| Linguistic           | เชาว์ปัญญาทางด้านภาษาศาสตร์                                                                                                               |
| Logical-mathematical | เชาว์ปัญญาทางด้านตรรกะและคณิตศาสตร์                                                                                                       |
| Spatial              | เชาว์ปัญญาทางด้านมิติสัมพันธ์                                                                                                             |
| Musical              | เชาว์ปัญญาทางด้านดนตรี                                                                                                                    |
| Bodily-kinesthetic   | เชาว์ปัญญาทางด้านร่างกายและการเคลื่อนไหว                                                                                                  |
| Interpersonal        | เชาว์ปัญญาทางด้านความสัมพันธ์ระหว่างบุคคล                                                                                                 |
| Intrapersonal        | เชาว์ปัญญาทางด้านการรู้จักและเข้าใจตนเอง                                                                                                  |
| Naturalistic         | เชาว์ปัญญาทางด้านธรรมชาติวิทยา                                                                                                            |
| Existential          | เชาว์ปัญญาทางด้านการคงอยู่ของชีวิต                                                                                                        |
| Class1-8             | กลุ่มสายวิชาประกอบด้วย 1.วิทยาศาสตร์ 2.วิศวกรรมศาสตร์และเทคโนโลยี<br>3.วิทยาศาสตร์สุขภาพ 4.บริหารธุรกิจ 5.ภาษา 6.สังคม 7.การศึกษา 8.ศิลปะ |

ตารางที่ 3. ตารางเปรียบเทียบการแก้ปัญหาข้อมูลที่ขาดหายไปด้วยวิธี Single Imputation ด้วยการแทนค่าด้วยการคำนวณค่าเฉลี่ย

| Original Data               |            |         |         |         |        |
|-----------------------------|------------|---------|---------|---------|--------|
| Gender                      | Linguistic | Logical | Spatial | Musical | Bodily |
| Male                        | 3          | 5       | NaN     | 3       | 5      |
| Female                      | 5          | NaN     | 4       | 3       | 3      |
| Male                        | 3          | 4       | 1       | 3       | 3      |
| Female                      | NaN        | 2       | NaN     | 4       | 5      |
| Female                      | 3          | 3       | 3       | NaN     | 2      |
| Single Imputation with Mean |            |         |         |         |        |
| Male                        | 3          | 5       | 2       | 3       | 5      |
| Female                      | 5          | 3       | 4       | 3       | 3      |
| Male                        | 3          | 4       | 1       | 3       | 3      |
| Female                      | 3          | 2       | 2       | 4       | 5      |
| Female                      | 3          | 3       | 3       | 3       | 2      |

## 2.5 ขั้นตอนการปรับความสมดุลข้อมูลด้วยเทคนิคการสุ่ม เพิ่มตัวอย่างข้อมูลจากกลุ่มส่วนน้อยด้วยการสังเคราะห์ ข้อมูลใหม่ (Synthetic Minority Over-sampling

### Technique: SMOTE)

หลังจากผ่านขั้นตอนการจัดเตรียมข้อมูลเพื่อเป็นการเตรียมความพร้อมสำหรับการนำข้อมูลไปใช้ในการจำแนก ผู้วิจัยพบว่าชุดข้อมูลมีความไม่สมดุลกัน โดยชุดข้อมูลในกลุ่มสาขาวิชาในแต่ละสาขาวิชาจากการวัดทัศนคติของนักเรียนนั้นมีจำนวนที่ไม่เท่ากันในทุกกลุ่มสาขาวิชา ดังนั้นผู้วิจัยจึงแก้ปัญหาด้วยวิธีการปรับความสมดุลของข้อมูล ผ่านเทคนิคการสุ่มเพิ่มตัวอย่างข้อมูลจากกลุ่มส่วนน้อยด้วยการสังเคราะห์ข้อมูลใหม่ ช่วยในการปรับสมดุลของชุดข้อมูลเพื่อให้จำนวนของคลาสย่อย

(Minority Class) มีจำนวนเทียบเท่าหรือใกล้เคียงกับจำนวนข้อมูลที่เป็นคลาสหลัก (Majority Class) ก่อนผ่านชุดข้อมูลเพื่อเข้าสู่กระบวนการคัดเลือกคุณลักษณะต่อไป

เทคนิคการปรับสมดุลข้อมูลด้วยวิธีการ SMOTE มีหลักการทำงานเหมือนกับอัลกอริทึมเพื่อนบ้านใกล้เคียงที่สุดทำการสุ่มเลือกแล้วเก็บไว้ในตัวแปร  $x_i$  จากนั้นทำการสุ่มค่าตัวแปร  $\sigma$  ให้มีค่า 0 หรือ 1 เพื่อใช้ในการคำนวณ โดยมีสมการในการสุ่มเพิ่มตัวอย่างข้อมูลกลุ่มส่วนน้อยด้วยการสังเคราะห์ข้อมูลใหม่ดังแสดงในสมการที่ 1

$$x_{\text{new}} = x_i + (x - x_i) * \sigma \quad (1)$$

และผลของการปรับสมดุลข้อมูลด้วยเทคนิค SMOTE ดังตารางที่ 4

ตารางที่ 4. ตารางเปรียบเทียบจำนวนข้อมูลของกลุ่มผู้เรียนที่มีความสนใจกับกลุ่มผู้เรียนที่ไม่มีความสนใจในการเรียนกลุ่มสาขาวิชานั้น ๆ ด้วยเทคนิค SMOTE โดยพบว่าจำนวนของคลาสรอง (Minority Class) มีจำนวนเทียบเท่าหรือใกล้เคียงกับจำนวนข้อมูลที่เป็นคลาสหลัก (Majority Class) ได้อย่างถูกต้องมากยิ่งขึ้น

| Subject Area            | Original   |              | SMOTE      |              |          |
|-------------------------|------------|--------------|------------|--------------|----------|
|                         | Interested | Uninterested | Interested | Uninterested | Accuracy |
| Science                 | 3          | 136          | 136        | 136          | 98%      |
| Engineering Technology  | 26         | 113          | 113        | 113          | 93%      |
| Health Sciences         | 35         | 104          | 104        | 104          | 79%      |
| Business Administration | 22         | 117          | 117        | 117          | 81%      |
| Linguistic              | 31         | 108          | 108        | 108          | 71%      |
| Social                  | 7          | 132          | 132        | 132          | 93%      |
| Education               | 3          | 136          | 136        | 136          | 98%      |
| Art                     | 12         | 127          | 127        | 127          | 90%      |

## 2.6 ขั้นตอนการคัดเลือกคุณลักษณะด้วยเทคนิค RFE

การคัดเลือกคุณลักษณะ ถือได้ว่าเป็นขั้นตอนที่สำคัญเป็นอย่างยิ่ง เนื่องจากคุณลักษณะที่เก็บรวบรวมมานั้นมีจำนวนที่มาก และมีหลายคุณลักษณะที่ไม่มีความสำคัญต่อการจำแนกข้อมูลเพื่อใช้ในการสร้างตัวแบบการทำนายการศึกษาต่อในระดับอุดมศึกษา โดยจะใช้เทคนิค RFE โดยใช้หลักการค้นหาคุณลักษณะที่มีความเกี่ยวข้อง (Relevant features) กัน โดยคุณลักษณะ  $X_i$  ที่ได้

จะถือว่ามีความเกี่ยวข้องกับตัวแปรกลุ่มอย่างเข้ม (Strongly relevant) ก็ต่อเมื่อการแจกแจงความน่าจะเป็นอย่างมีเงื่อนไขของกลุ่ม เมื่อกำหนดเซตย่อยของคุณลักษณะ  $S$  (ซึ่งรวม  $X_i$ ) นั้นเปลี่ยนไป ถ้า  $X_i$  ถูกกำจัด คุณลักษณะใดที่ไม่ได้มีความเกี่ยวข้องอย่างเข้มหรืออย่างอ่อน จะเรียกว่าคุณลักษณะที่ไม่มีความเกี่ยวข้องกัน (Irrelevant features) แสดงผลลัพธ์การคัดเลือกคุณลักษณะที่สำคัญดังตารางที่ 5 และ 6

ตารางที่ 5. ตารางเปรียบเทียบการปรับลดคุณลักษณะทักษะการเรียนรู้จาก 9 คุณลักษณะด้วยเทคนิค RFE พบว่าตัวแบบของคลาสที่ 1-4 ที่ใช้คุณลักษณะเพียง 5 คุณลักษณะยังมีสมรรถนะการทำงานที่ใกล้เคียงกับการใช้คุณลักษณะแบบดั้งเดิม

| Result   | Science  |     |     | Engineering and Technology |     |     | Health Sciences |     |     | Business and Administration |     |     |
|----------|----------|-----|-----|----------------------------|-----|-----|-----------------|-----|-----|-----------------------------|-----|-----|
|          | Original | RFE | +/- | Original                   | RFE | +/- | Original        | RFE | +/- | Original                    | RFE | +/- |
| Accuracy | 82%      | 81% | 1%  | 78%                        | 75% | 3%  | 81%             | 77% | 4%  | 84%                         | 79% | 5%  |
| F1       | 81%      | 77% | 4%  | 72%                        | 69% | 3%  | 62%             | 60% | 2%  | 76%                         | 78% | 2%  |

ตารางที่ 6. ตารางเปรียบเทียบการปรับลดคุณลักษณะทักษะการเรียนรู้จาก 9 คุณลักษณะด้วย เทคนิค RFE พบว่าตัวแบบของคลาสที่ 5 – 8 ที่ใช้คุณลักษณะเพียง 5 คุณลักษณะยังมีสมรรถนะการทำงานที่ใกล้เคียงกับการใช้คุณลักษณะแบบดั้งเดิม

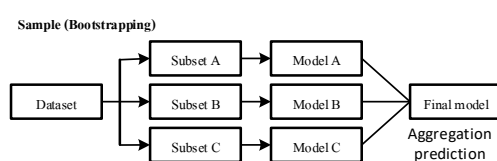
| Result   | Linguistic |     |     | Social   |     |     | Education |     |     | Art      |     |     |
|----------|------------|-----|-----|----------|-----|-----|-----------|-----|-----|----------|-----|-----|
|          | Original   | RFE | +/- | Original | RFE | +/- | Original  | RFE | +/- | Original | RFE | +/- |
| Accuracy | 73%        | 72% | 1%  | 89%      | 86% | 3%  | 92%       | 88% | 4%  | 88%      | 85% | 3%  |
| F1       | 77%        | 72% | 5%  | 84%      | 80% | 4%  | 94%       | 89% | 5%  | 82%      | 78% | 4%  |

## 2.7 ขั้นตอนการจำแนกข้อมูลด้วยแบบจำลองแบบรวมกลุ่ม (Ensemble Learning-based Model)

เป็นเทคนิคหลักที่นำเสนอในงานวิจัยนี้เป็นขั้นตอนที่สำคัญในการวิเคราะห์ผลลัพธ์ที่ได้มาจากการคัดเลือกคุณลักษณะด้วยเทคนิค RFE ที่เกี่ยวข้องกับการทำนายกลุ่มผู้เรียนกับกลุ่มสาขาวิชาที่เหมาะสมตามทักษะของผู้เรียนในเชิงพุทธิปัญญา โดยในขั้นตอนนี้ผู้วิจัยจะใช้เทคนิคการสร้างแบบจำลองการเรียนรู้โดยนำผลลัพธ์ที่ได้จากการสร้างแบบจำลองการเรียนรู้แบบตัวจำแนกเดี่ยวมาไว้รวมกันเพื่อเพิ่มประสิทธิภาพในการทำนายให้กับแบบจำลองการเรียนรู้แบบรวมกลุ่ม โดยมี 3 เทคนิค ดังนี้

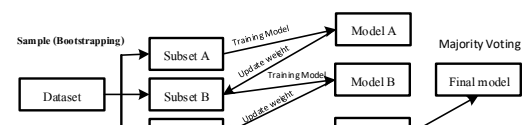
(1) Bagging Classification โดยในงานวิจัยนี้ใช้ตัวแบบจำลองการเรียนรู้ในการจำแนกข้อมูลคือ Decision Tree, Random Forest, ExtraTrees, AdaBoost, Gradient Boosting, XGB เทคนิค Bagging เริ่มจากการสุ่มเลือกข้อมูลด้วยหลักการสุ่มแบบ แล้วนำข้อมูลที่สุ่มขึ้นมาไปเพื่อสร้างแบบจำลองทุกชุดข้อมูล จากนั้นใช้หลักการ

โหวต (Vote) เพื่อหาแบบจำลองที่สามารถตอบข้อมูลทดสอบได้ดีที่สุดมาเป็นแบบจำลอง ดังรูปที่ 2



รูปที่ 2. กระบวนการทำงานของ Bagging

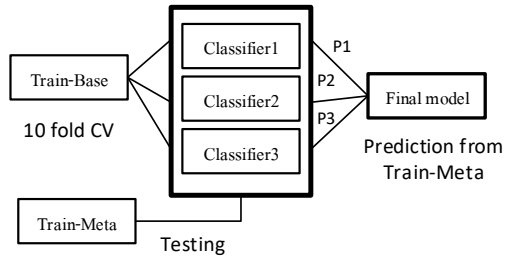
(2) Boosting Classification โดยในงานวิจัยนี้ใช้ตัวแบบจำลองการเรียนรู้ในการจำแนกข้อมูลคือ Decision Tree, Random Forest, ExtraTrees, AdaBoost, Gradient Boosting, XGB Boosting เทคนิค Boosting เป็น การ สร้าง แบบจำลองที่มีความต่อเนื่องกัน โดยแต่ละแบบจะได้รับการฝึกจากชุดข้อมูลเดียวกัน แล้วนำผลลัพธ์ของแบบจำลองที่มีค่าความแม่นยำต่ำ (Weak Classifier) มาทำการแก้ไขข้อผิดพลาดที่มีอยู่ในแบบจำลองก่อนหน้านี้ นั่นโดยผลรวมของแบบจำลองทุกตัวจะถูกถ่วงน้ำหนักตามความสำเร็จ จากนั้นจะรวมผลลัพธ์ทั้งหมดด้วยหลักการโหวตจนได้แบบจำลองการเรียนรู้ตัวใหม่ขึ้น ดังรูปที่ 3



รูปที่ 3. กระบวนการทำงานของเทคนิค Boosting

(3) Stacking Classification ในงานวิจัยนี้ใช้ตัวแบบจำลองการเรียนรู้ด้วยเทคนิค Stacking โดยใช้ตัวจำแนกข้อมูล คือ Decision Tree, Random Forest, ExtraTrees, AdaBoost, Gradient Boosting และ XGB Boosting เทคนิคนี้มีความคล้ายคลึงกับเทคนิค Boosting โดยมีการแบ่งตัวจำแนกไว้ 6 ตัวเป็นการนำผลลัพธ์ที่ได้ในแต่ละแบบจำลองมาสร้างเป็นข้อมูล

input แล้วส่งต่อผลลัพธ์ให้ตัวจำแนกตัวต่อไปเรียนรู้ และมีการใช้ K-fold cross validation ในการรวมการถ่วงน้ำหนักที่ดีที่สุดของการทำนายไว้ ดังรูปที่ 4



รูปที่ 4. กระบวนการทำงานของ Stacking

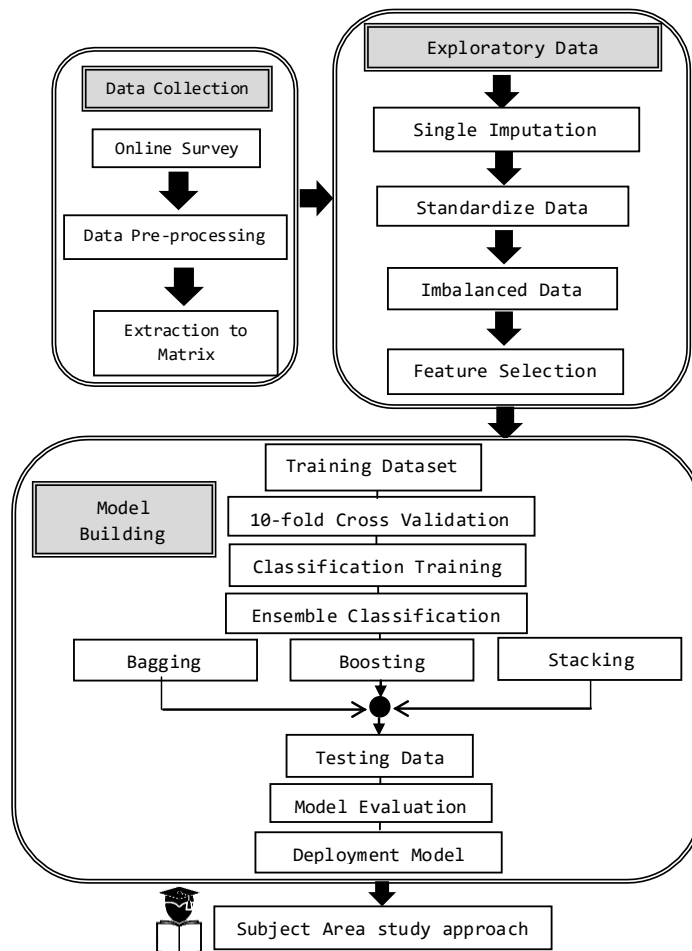
## 2.8 ขั้นตอนการวิเคราะห์และสรุปผล (Result Analysis)

ขั้นตอนนี้ ผู้วิจัยทำการสรุปผลการสร้างแบบจำลองการเรียนรู้แบบกลุ่มด้วยการใช้เทคนิคทั้ง 3 เทคนิค (1) Bagging (2) Boosting และ (3) Stacking กับการทำนายกลุ่มผู้เรียนกับกลุ่มสาขาวิชาที่เหมาะสมกับทักษะของผู้เรียนในเชิงพหุปัญญาดังรูปที่ 5

โดยทำการหาข้อสรุปจากการเปรียบเทียบประสิทธิภาพของแบบจำลอง 3 เทคนิค

## 3. ผลการทดลอง

ผลการวิจัยการวิเคราะห์และสรุปผล การสร้างตัวแบบการทำนายในการจำแนกข้อมูลด้วยวิธีการสร้างแบบจำลองการเรียนรู้แบบกลุ่มด้วยการใช้ 3 เทคนิค (1) Bagging (2) Boosting และ (3) Stacking โดยทำการหาข้อสรุปจากการเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้แบบรวมกลุ่มทั้ง 3 เทคนิค โดยผลของการวิจัยด้วยเทคนิคแบบบูรณาการเริ่มจากการจัดการกับปัญหาข้อมูลที่ขาดหายไป การปรับความสมดุลให้กับข้อมูลด้วยเทคนิคการสังเคราะห์ข้อมูลใหม่ การคัดเลือกคุณลักษณะด้วยเทคนิค RFE และขั้นตอนของการเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้แบบรวมกลุ่มด้วย 3 เทคนิค เพื่อสร้างตัวแบบการทำนายกลุ่มผู้เรียนกับกลุ่มสาขาวิชาที่เหมาะสมกับทักษะของผู้เรียนในเชิงพหุปัญญาดังรูปที่ 5



รูปที่ 5. แสดงขั้นตอนการดำเนินการสร้างตัวแบบการทำนายกลุ่มผู้เรียนกับกลุ่มสาขาวิชาตามทักษะการเรียนรู้ในเชิงพหุปัญญาดังรูปที่ 5

(1) การใช้เทคนิค Bagging Classification ในการประเมินสมรรถนะของตัวแบบการทำนายจากผลการทดลองโดยใช้เทคนิค Bagging Classification ผ่านการวิเคราะห์ด้วย 6 ตัวจำแนก เพื่อประเมิน 5 คุณลักษณะสำคัญที่เกี่ยวข้องกับการทำนายกลุ่มผู้เรียนกับกลุ่มสาขาวิชาตามทักษะการเรียนรู้ของผู้เรียนในเชิงพหุปัญญา โดยทักษะการเรียนรู้ที่มีค่าความสำคัญสูงสุดอยู่ที่ 75% และค่าเฉลี่ยความแม่นยำและความถูกต้องอยู่ที่ 76% คือ ทักษะการเรียนรู้ด้านเหตุผลและคณิตศาสตร์ (Logical-mathematical) ดังรูปที่ 6 และ 7

(2) การใช้เทคนิค Boosting Classification ในการประเมินสมรรถนะของตัวแบบการทำนายจากผลการทดลองโดยใช้เทคนิค Boosting Classification ผ่านการวิเคราะห์ด้วย 6 ตัวจำแนก เพื่อประเมิน 5 คุณลักษณะสำคัญที่เกี่ยวข้องกับการทำนายกลุ่มผู้เรียนกับกลุ่มสาขาวิชาตามทักษะการเรียนรู้ในเชิงพหุปัญญา ด้วยวิธีการสร้างแบบจำลองที่มีความต่อเนื่องกันแล้วฝึกฝนจากชุดข้อมูลเดียวกันแล้วนำผลลัพธ์ของแบบจำลองที่มีค่าความแม่นยำต่ำมาทำการแก้ไขข้อผิดพลาด จากนั้นจะรวมผลลัพธ์ทั้งหมดด้วยหลักการโหวต (Vote) โดยทักษะการเรียนรู้ที่มีค่าความสำคัญสูงสุดอยู่ที่ 72% และค่าเฉลี่ยความแม่นยำและความถูกต้องอยู่ที่ 71% คือ ทักษะการเรียนรู้ด้านเหตุผลและคณิตศาสตร์ (Logical-mathematical) ดังรูปที่ 8 และ 9

(3) การใช้เทคนิค Stacking Classification ในการประเมินสมรรถนะของแบบจำลองจากผลการทดลองโดยใช้เทคนิค Boosting Classification ผ่านการวิเคราะห์ด้วย 6 ตัวจำแนก เพื่อประเมิน 5 คุณลักษณะสำคัญที่เกี่ยวข้องกับการทำนายกลุ่มผู้เรียนกับกลุ่มสาขาวิชาตามทักษะการเรียนรู้ของผู้เรียนในเชิงพหุปัญญา ด้วยวิธีการแบ่งตัวจำแนกสำหรับการวิเคราะห์ไว้ 6 ตัว แล้วนำผลลัพธ์ที่ได้ในแต่ละแบบจำลองมาสร้างเป็นข้อมูลนำเข้าแล้วส่งต่อผลลัพธ์ให้ตัวจำแนกตัวต่อไปเรียนรู้ โดยทักษะการเรียนรู้ที่มีค่าความสำคัญสูงสุดอยู่ที่ 77% และค่าเฉลี่ยความแม่นยำและความถูกต้องอยู่ที่ 75% คือ ทักษะการเรียนรู้ด้านเหตุผลและคณิตศาสตร์ (Logical-mathematical) ดังรูปที่ 10 และ 11

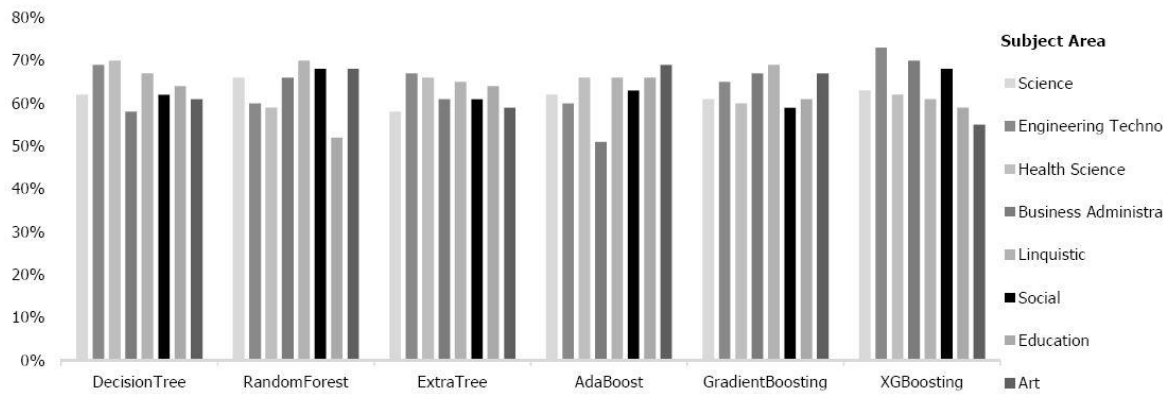
#### 4. บทสรุปและการอภิปราย

งานวิจัยนี้นำเสนอการสร้างตัวแบบการทำนายในการเลือกศึกษาต่อในระดับอุดมศึกษาตามกลุ่มผู้เรียนโดยใช้เทคนิคแบบบูรณาการในการแก้ปัญหา Imbalanced Data ของกลุ่มผู้เรียน และเพื่อ

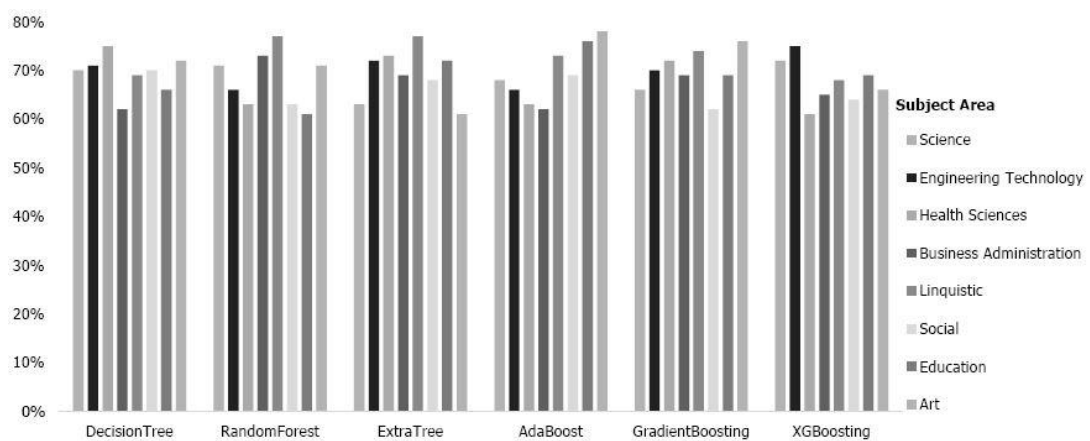
ศึกษาตัวแบบความสำคัญที่เกี่ยวข้องกับการทำนายกลุ่มผู้เรียนกับกลุ่มสาขาวิชาที่เหมาะสมตามทักษะของผู้เรียนในเชิงพหุปัญญา ข้อมูลเก็บรวบรวมจากแบบสอบถามออนไลน์ การสำรวจข้อมูลมีการตรวจสอบข้อมูลมีการแทนค่าที่หายไปด้วย Single Imputation การปรับสมดุลข้อมูลด้วยเทคนิค SMOTE ทำให้สามารถจำแนกข้อมูลของคลาสย่อย (Minority Class) ให้มีจำนวนที่เทียบเท่าหรือใกล้เคียงกับจำนวนข้อมูลที่เป็นคลาสหลัก (Majority Class) ถัดมาการคัดเลือกคุณลักษณะ เนื่องจากชุดข้อมูลที่เก็บรวบรวมมานั้น ทางผู้วิจัยพบว่ามีความหลากหลายมาก ทางผู้วิจัยจึงต้องวิเคราะห์ว่าคุณลักษณะแบบไหนที่มีความสำคัญจริงๆ ในการสร้างตัวแบบการทำนายด้วยการปรับลดคุณลักษณะผ่านเทคนิค RFE พบว่าประสิทธิภาพการทำงานของตัวแบบมีความใกล้เคียงกับตัวแบบที่ใช้คุณลักษณะแบบดั้งเดิมผลการทดสอบคุณลักษณะการเรียนรู้ในเชิงพหุปัญญาที่มีความสำคัญมากที่สุด คือ ทักษะการเรียนรู้ด้านเหตุผลและคณิตศาสตร์ ซึ่งมีความสำคัญมาเป็นอันดับหนึ่งในทุกเทคนิคของการจำแนกข้อมูลในส่วนของคุณลักษณะการเรียนรู้ในด้านอื่น ๆ ที่ติดอันดับหนึ่งถึงห้าในทุกเทคนิคของการจำแนกข้อมูลด้วยแบบจำลองการเรียนรู้แบบรวมกลุ่ม ได้แก่ ด้านมนุษยสัมพันธ์ ด้านความเข้าใจในตน ด้านมิติสัมพันธ์และ ด้านดนตรี ซึ่งทักษะการเรียนรู้ที่ติดอันดับหนึ่งถึงห้าล้วนแต่เป็นทักษะพื้นฐานที่เป็นรากฐานที่สำคัญของวิทยาการความรู้ในด้านต่าง ๆ และที่สำคัญนอกเหนือจากการที่ผู้เรียนจะเข้าใจหลักสูตรและเข้าใจความสำคัญในการเรียนแล้ว นั่นก็คือการเข้าใจทักษะที่จำเป็นสำหรับการศึกษาต่อในแต่ละกลุ่มสาขาวิชา ซึ่งเมื่อผู้เรียนมีความเข้าใจและนำไปใช้ในทางที่ถูกต้องก็จะเป็นสิ่งที่ช่วยให้ผู้เรียนก้าวไปสู่ความสำเร็จได้อีกขั้นหนึ่งทั้งนี้พบว่าจากแนวคิดการศึกษาตัวแบบความสำคัญที่เกี่ยวข้องกับการทำนายกลุ่มผู้เรียนกับกลุ่มสาขาวิชาที่เหมาะสมตามทักษะของผู้เรียนในเชิงพหุปัญญาด้วยเทคนิคแบบบูรณาการด้วยการเพิ่มสมรรถนะให้ตัวแบบด้วยการปรับลดคุณลักษณะที่จำเป็นและปรับสมดุลข้อมูล โดยใช้ตัวจำแนกทำงานร่วมกัน และต่างชนิดกัน สามารถเพิ่มประสิทธิภาพการทำงานของการทำงานกลุ่มสาขาวิชาที่เหมาะสมให้กับกลุ่มผู้เรียนที่ต้องอาศัยอำนาจการจำแนกที่มีความถูกต้องและแม่นยำสูง ซึ่งเป็นตัวบ่งบอกถึงความสามารถที่แตกต่างจากการใช้ตัวแบบการทำนายที่ใช้ตัวจำแนกเพียงตัวเดียวอย่างเห็นได้ชัดอีกทั้งยังสามารถวิเคราะห์ข้อมูลได้อย่างรวดเร็วซึ่งผลการวิจัยพบว่าเทคนิคที่มีประสิทธิภาพสูงที่สุดต่อ

การสร้างตัวแบบการทำนาย คือ เทคนิคการจำแนกข้อมูลแบบ Stacking มีค่าความสำคัญอยู่ที่ 77% ส่วนอันดับที่สอง คือ เทคนิคการจำแนกข้อมูลแบบ Bagging มีค่าความสำคัญอยู่ที่

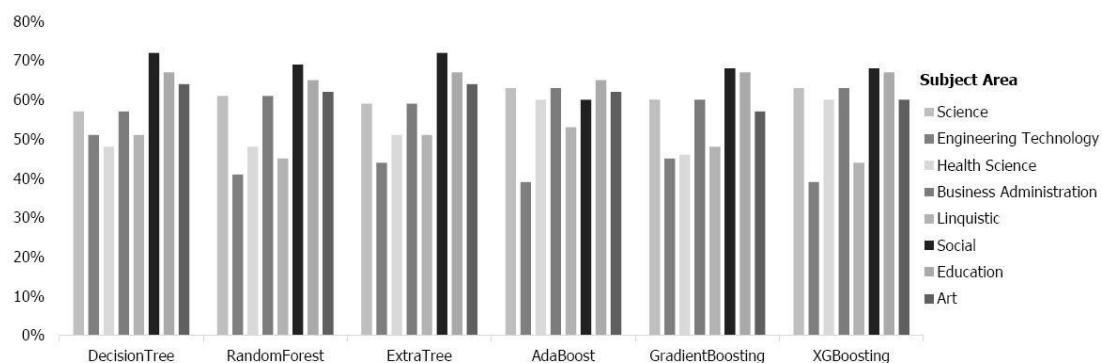
75% และเทคนิคการจำแนกข้อมูลแบบ Boosting มีค่าความสำคัญอยู่ที่ 72% ตามลำดับ ซึ่งจากการค้นพบครั้งนี้สามารถนำไปประยุกต์ใช้กับการวิเคราะห์ข้อมูลขนาดใหญ่ทางด้านการศึกษาต่อไปได้ในอนาคต



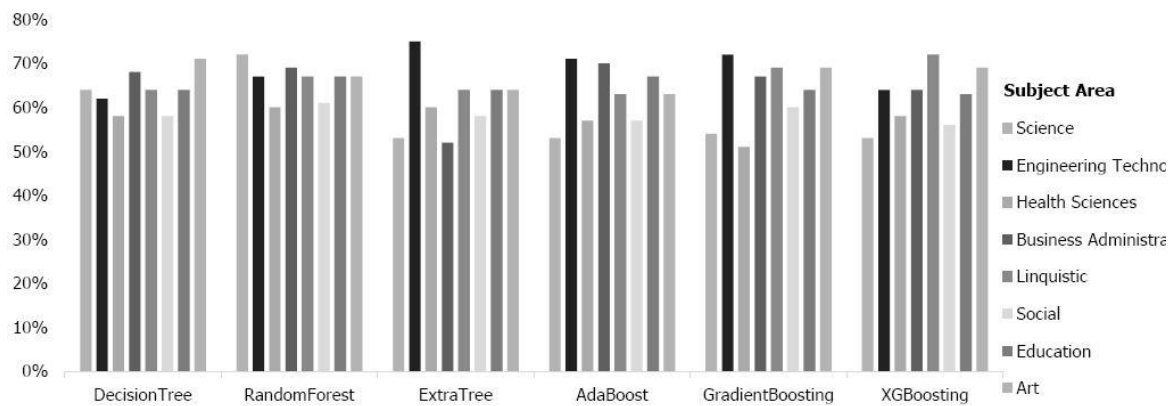
รูปที่ 6. แสดงค่าความสำคัญ Accuracy score ของตัวแบบการทำนายด้วยเทคนิค Bagging Classification



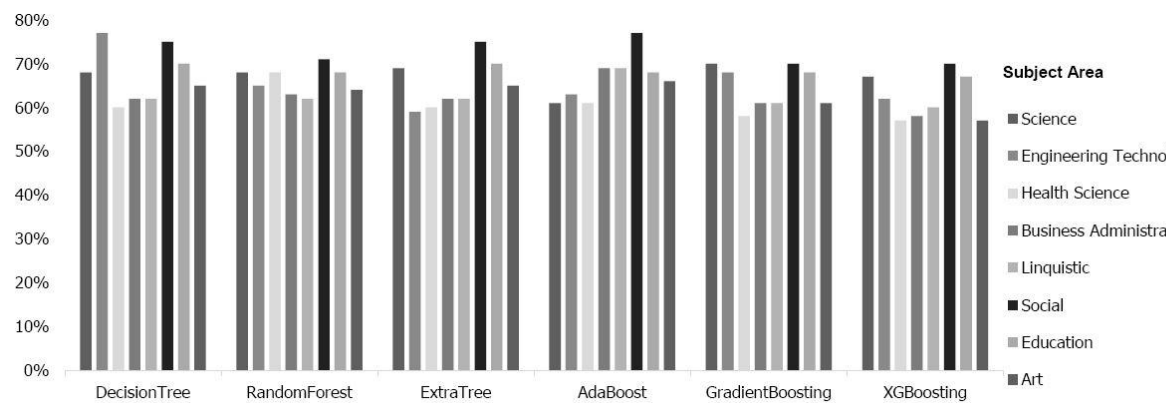
รูปที่ 7. แสดงค่าเฉลี่ยความแม่นยำและความถูกต้อง F1 score ของตัวแบบการทำนายด้วยเทคนิค Bagging Classification



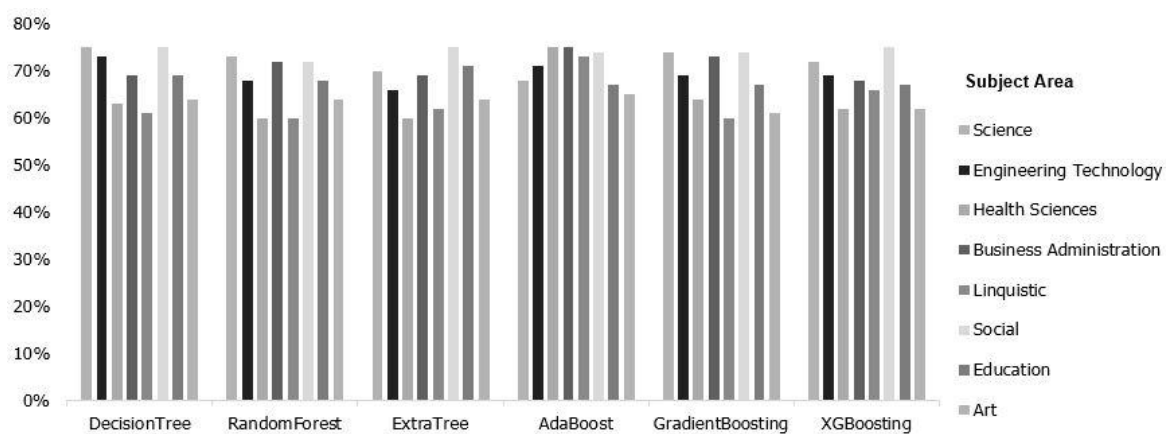
รูปที่ 8. แสดงค่าความสำคัญ Accuracy score ของตัวแบบการทำนายด้วยเทคนิค Boosting Classification



รูปที่ 9. แสดงค่าเฉลี่ยความแม่นยำและความถูกต้อง  $F1$  score ของตัวแบบการทำนายด้วยเทคนิค Boosting Classification



รูปที่ 10. แสดงค่าความสำคัญ  $Accuracy$  score ของตัวแบบการทำนายด้วยเทคนิค Stacking Classification



รูปที่ 11. แสดงค่าเฉลี่ยความแม่นยำและความถูกต้อง  $F1$  score ของตัวแบบการทำนายด้วยเทคนิค Stacking Classification

## 5. ข้อเสนอแนะ

ในการนำเทคนิคแบบบูรณาการมาแก้ปัญหา Imbalanced Data ของกลุ่มผู้เรียน เพื่อศึกษาตัวแบบความสัมพันธ์ที่เกี่ยวข้องกับการทำนายกลุ่มผู้เรียนกับกลุ่มสาขาวิชาที่เหมาะสมตามทักษะของผู้เรียนพบว่ายังมีข้อจำกัดในเรื่องของปัจจัยข้อมูลที่มีเพียงข้อมูลทักษะการเรียนรู้ เพศ และกลุ่มสาขาวิชา ซึ่งอาจจะมีปัจจัยในด้านอื่น ๆ ที่มีความเกี่ยวข้องและสำคัญต่อการเลือกกลุ่มสาขาวิชา เช่น ข้อมูลด้านความถนัด คะแนนสอบตามเกณฑ์มาตรฐาน และอาชีพในอนาคต เป็นต้น ในอนาคตหากมีการพัฒนางานวิจัยทางด้าน Data Mining for Education ในแง่มุมของการเพิ่มประสิทธิภาพให้กับตัวแบบการทำนาย อาจจะต้องมีการเก็บรวบรวมข้อมูลเพิ่มมากขึ้นเพื่อให้ได้ปัจจัยที่มีปริมาณและคุณภาพที่สูงขึ้นใช้เทคนิคในการคัดเลือกคุณลักษณะและการจำแนกข้อมูลแบบอื่น ๆ เพื่อช่วยเพิ่มสมรรถนะการจำแนกข้อมูลและก่อให้เกิดผลลัพธ์ที่มีประสิทธิภาพมากยิ่งขึ้น

## เอกสารอ้างอิง

- [1] I. Tongsamsi, "Labor Market and Educational Mismatch A Case Study of New Community Development Workers," presented at The 10<sup>th</sup> Hatyai National and International Conference, Songkla, Thailand, 2019.
- [2] S. Jantapong, "The Skills Mismatch and the Challenges to Education 4.0," MER 2018, Bangkok, Thailand, 2018.
- [3] P. Kanjanasamranwong, "THE NEEDS OF ATTENDING IN GRADUATE INSTITUTE OF GRADE 12," Journal of Education Research Faculty of Education, vol.11, no.1, pp. 75-88, 2018.
- [4] W. SAENKOSIK, "ENHANCING INTRAPERSONAL INTELLIGENCE AMONG PRIMARY SCHOOL STUDENTS USING AN INTRAPERSONAL TRAINING APPLICATION," Ph.D., RESEARCH AND STATISTICS IN COGNITIVE SCIENCE, Burapha University, Chonburi, 2018.
- [5] N. Supanatsetakul, "Interdisciplinary studies: humanities, social sciences and medicine for applying in the medical instruction," AMJAM , vol. 13, no. 3, pp. 380-392, 2013.
- [6] Office of the Higher Education Commission, "New student information," OHEC, 2020. [Online]. Available: <http://www.info.mua.go.th/info/> [Accessed: Jan.12, 2021].
- [7] A. Phacobang, "Adjusting the Imbalanced Data with 5 Classification Methods," Thai Journal of Science and Technology, vol. 9, no.4, July - August, 2020.
- [8] D. Noppamas, "Comparison of Imbalanced Data Problem Solving for Income Classification of Type I Pharmacies Entrepreneur," in The 9<sup>th</sup> STOU National Research Conference, pp. 1578-1586, 2018.
- [9] W. Jaidee, "The Study of Factors Affecting for On-time Graduating of Ungraduated Student Using Feature Selection Technique on Imbalanced Datasets," Journal of Information Science and Technology, vol.10, no. 1, pp. 75-84, 2020.
- [10] N. Akarachantachote, "Feature Selection for High-dimensional data in Classification," Research Methodology & Cognitive Science, vol. 8, no. 2, pp. 1-13, October 2010 – March 2011.
- [11] P. Pramokchon, "Filter-Based Feature Selection for Data Classification in IoT," FEU Academic Review, vol. 11, no. 3, pp. 98-113, 2017.
- [12] C. Songsiri, T. Rakthanmanon, and K. Waiyamai, "Applying a data mining technique to help students in selecting their majors," In Proc. The 39<sup>th</sup> Kasetsart University Annual Conference, pp. 43-50, 2001.
- [13] W. Pimpakun, "Ensemble Learning Model for Cardiotocography Classification," In Proc. Graduate Research Conference, pp. 333-340, 2012.
- [14] P. Promla, "The Comparison of Efficiency on The Analysis of Satisfaction Teaching Performance using Sentiment Analysis by Ensemble Technique," KRU Research Journal, vol. 20, no. 4, pp. 140-149, 2020.
- [15] W. Paireekreng, "An ensemble learning based model for real estate project classification," ScienceDirect, vol. 3, pp. 3852 – 3859, 2015.

- [16] P. Thanvanon, "Predict stock price trends in Stock Exchange of Thailand using Ensemble Model," JOURNAL OF INFORMATION SCIENCE AND TECHNOLOGY, vol. 7, no. 1, pp. 12-21, 2017.
- [17] L. Breiman, "Bagging Predictors," Machine Learning, vol. 24, pp. 123-140, 1996.
- [18] K. Nasritha, "Comparison of sampling techniques for imbalanced data classification," JOURNAL Of APPLIED INFORMATICS AND TECHNOLOGY, vol. 1, pp. 20-35, 2018.
- [19] S. Wang, "Negative correlation learning for classification ensembles," presented at the International Joint Conference on Neural Networks, IJCNN 2010, Barcelona, Spain, 2010.
- [20] C. Onsang, "Combining Multiple Models for Predicting the Risk of Diabetes Using Multi-Kernel Learning," In Proc. NCCIT2019, pp. 1-6, 2019.
- [21] N. Sritrakul and T. Hudakorn, "The economic value and satisfaction of substituting LPG in households by a biogas network," Energy Reports, pp. 565-571, 2020.
- [22] W. Sookcharoen, "Dealing with Missing Data," ROMPHRUEK JOURNAL KRIRK UNIVERSITY, vol. 33, no. 2, pp. 11-32, 2015.