

SAC-01-019

**การวิเคราะห์เปรียบเทียบประสิทธิภาพ เพื่อการแนะน่านักศึกษา
โดยใช้เทคนิคการเรียนรู้ของเครื่อง: กรณีศึกษาวิทยาลัยอาชีวศึกษาอุดรธานี
Comparative Performance Analysis for Student Guidance Using
Machine Learning Techniques: A Case Study of Udon Thani Vocational College**

สุพิชชา สุวรรณศรี^{1*} และ กริช สมกันธา²
Supitcha Suwannasri^{1*} และ Krit Somkantha²
มหาวิทยาลัยราชภัฏอุดรธานี^{1,2}
Udon Thani Rajabhat University^{1,2}
e-mail: Yingsuwanasri@gmail.com¹, dr_krit@udru.ac.th²

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์ เพื่อ (1) พัฒนาแบบจำลองการเรียนรู้ของเครื่องสำหรับทำนายแนวโน้มการเลือกสาขาวิชาเรียนของนักเรียนระดับประกาศนียบัตรวิชาชีพ (ปวช.) และ (2) เปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่อง ได้แก่ อัลกอริทึมการถดถอยโลจิสติก อัลกอริทึมป่าสุ่ม และอัลกอริทึมเกรเดียนต์บูสต์ติง ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลนักเรียนระดับประกาศนียบัตรวิชาชีพ (ปวช.) ของวิทยาลัยอาชีวศึกษาอุดรธานี จำนวน 1,789 รายการ โดยแบ่งเป็นข้อมูลชุดข้อมูลฝึกสอน และชุดข้อมูลทดสอบ 80/20 ประกอบด้วย 13 คุณลักษณะ โดยแบ่งกลุ่มข้อมูลออกเป็น 3 สาขาวิชา ได้แก่ สาขาวิชาการบัญชี จำนวน 504 รายการ สาขาวิชาเทคโนโลยีสารสนเทศ จำนวน 601 รายการ และสาขาวิชาอาหารและโภชนาการ จำนวน 684 รายการ ก่อนสร้างแบบจำลองได้ดำเนินการเตรียมข้อมูล และคัดเลือกคุณลักษณะด้วยเทคนิคการตัดตัวแปรออกทีละตัวจนเหลือชุดที่ดีที่สุด (RFECV) จากนั้น ประเมินประสิทธิภาพของแบบจำลองด้วยวิธีการตรวจสอบความถูกต้องแบบไขว้ 10 ส่วน โดยใช้ตัวชี้วัด ได้แก่ ความถูกต้อง ความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล

ผลการวิจัยพบว่า แบบจำลองอัลกอริทึมการถดถอยโลจิสติก มีประสิทธิภาพสูงที่สุด โดยมีค่าความถูกต้อง เท่ากับ 92.46% ค่าความแม่นยำ เท่ากับ 90.78% ค่าความระลึก เท่ากับ 92.24% และค่าความถ่วงดุล เท่ากับ 91.50% พร้อมทั้ง ใช้เวลาในการฝึกแบบจำลองเพียง 0.04 วินาที ขณะที่อัลกอริทึมเกรเดียนต์บูสต์ติงมีค่าความถูกต้อง เท่ากับ 88.27% ค่าความแม่นยำ เท่ากับ 85.36% ค่าความระลึก เท่ากับ 88.38% และค่าความถ่วงดุล เท่ากับ 86.85% ส่วนอัลกอริทึมป่าสุ่ม มีค่าความถูกต้อง เท่ากับ 88.83% ค่าความแม่นยำ เท่ากับ 85.09 % ค่าความระลึก เท่ากับ 89.70% และค่าความถ่วงดุล เท่ากับ 87.33% จากผลการเปรียบเทียบสรุปได้ว่า แบบจำลองอัลกอริทึมการถดถอยโลจิสติกเป็นแบบจำลองที่มีประสิทธิภาพสูงที่สุด

คำสำคัญ: การเรียนรู้ของเครื่อง อัลกอริทึมการถดถอยโลจิสติก อัลกอริทึมป่าสุ่ม อัลกอริทึมเกรเดียนต์บูสต์ติง การคัดเลือกคุณลักษณะ

Abstract

This research aimed to (1) develop machine learning models for predicting the academic program selection trends of Vocational Certificate (VC) students and (2) compare the performance of Logistic Regression, Random Forest, and Gradient Boosting to identify the most suitable prediction model. The dataset consisted of 1,789 records with 13 attributes collected from Vocational Certificate students at Udon Thani Vocational College. The data were divided into training and testing sets using an 80:20 ratio (1,431 and 358 records, respectively) and categorized into three academic programs: Accounting (504 records), Information Technology (601 records), and Food and Nutrition (684 records). Data preprocessing was performed prior to model development, and feature selection was conducted using Recursive Feature Elimination with Cross-Validation (RFECV). The prediction models were developed using Logistic Regression, Random Forest, and Gradient Boosting, and their performance was evaluated using 10-fold cross-validation and a testing dataset based on Accuracy, Precision, Recall, and F1-score.

The results indicated that the Logistic Regression model achieved the highest predictive performance, with an Accuracy of 92.46%, Precision of 90.78%, Recall of 92.24%, and an F1-score of 91.50%, while requiring only 0.04 seconds for model training. Overall, Logistic Regression outperformed the Random Forest and Gradient Boosting models and was identified as the most suitable model for predicting the academic program selection trends of Vocational Certificate students using the dataset employed in this study.

Keywords: Machine Learning, Logistic Regression, Random Forest, Gradient Boosting, Feature Selection

บทนำ

ประเทศไทยกำลังขับเคลื่อนการพัฒนาประเทศด้วยกำลังคนที่มีทักษะในสาขาอาชีพเฉพาะ ซึ่งวิทยาลัยอาชีวศึกษามีบทบาทสำคัญอย่างยิ่งในการผลิตและพัฒนากำลังคนระดับกลาง เพื่อสนับสนุนการพัฒนาประเทศตามยุทธศาสตร์ของสำนักงานคณะกรรมการการอาชีวศึกษา (พ.ศ. 2555–2569) ให้ผู้เรียนมีโอกาสรบความสำเร็จในอาชีพ และตอบสนองต่อความต้องการของตลาดแรงงานอย่างแท้จริง การแนะแนวการศึกษาต่อและอาชีพจึงถือเป็น “บริการหลัก” ที่มีความจำเป็นต่อการช่วยให้นักเรียนตัดสินใจเลือกเส้นทางการศึกษาที่เหมาะสมกับความถนัด ความสนใจ และศักยภาพของตนเอง โดยเฉพาะอย่างยิ่งในยุคปัจจุบันที่มีทางเลือกและสาขาวิชาที่หลากหลายและซับซ้อนมากขึ้น

ปัจจุบันวิทยาลัยอาชีวศึกษาหลายแห่ง รวมถึง วิทยาลัยอาชีวศึกษาอุดรธานี ยังคงประสบปัญหาในการบริหารจัดการงานแนะแนวในรูปแบบดั้งเดิม ซึ่งก่อให้เกิดข้อจำกัดหลายประการ เช่น การจัดการข้อมูล ผลการเรียนรู้ ความถนัด และข้อมูลศิษย์เก่า เป็นต้น มักถูกจัดเก็บอย่างกระจัดกระจาย ทำให้ยากต่อการนำมาวิเคราะห์เชิงลึกและใช้ประโยชน์ในการให้คำปรึกษาอย่างมีประสิทธิภาพ ข้อจำกัดด้านบุคลากร ครูแนะแนวต้องรับผิดชอบภาระงานสอนและงานอื่น ๆ ร่วมด้วย ส่งผลให้มีเวลาไม่เพียงพอในการให้คำปรึกษาเชิงลึกแบบรายบุคคล (ภัสรา โกมลเจียร, 2559) และความไม่แม่นยำของการแนะแนว ส่วนใหญ่อาศัยประสบการณ์ส่วนบุคคลของครู ซึ่งขาดการวิเคราะห์ข้อมูลเชิงประจักษ์ ส่งผลให้นักเรียนบางส่วนตัดสินใจเลือกศึกษาต่อในสาขาที่ไม่สอดคล้องกับความถนัดของตนเอง และอาจนำไปสู่ปัญหาการเปลี่ยนสาขาหรือลาออกในภายหลัง เพื่อแก้ไขข้อจำกัดดังกล่าว การนำเทคโนโลยีสารสนเทศ เข้ามาประยุกต์ใช้ในการแนะแนวการศึกษาต่อจึงมีความสำคัญอย่างยิ่ง ระบบสารสนเทศเพื่อแนะแนวการศึกษาต่อสามารถเป็นเครื่องมือสำคัญในการจัดเก็บและบริหารข้อมูลอย่างเป็นระบบ รวดเร็ว และถูกต้อง รวมถึงเป็นแหล่งข้อมูลที่เชื่อถือได้สำหรับนักเรียนและครูแนะแนว (ธนารักษ์ สารเถื่อนแก้ว, 2566) โดยระบบดังกล่าวสามารถรวบรวมข้อมูลหลักสูตร สถิติการศึกษาต่อในระดับที่สูงขึ้น และแนวโน้มของตลาดแรงงาน เพื่อให้ผู้ใช้เข้าถึงข้อมูลที่จำเป็นได้อย่างสะดวกและเท่าเทียม ซึ่งระบบสารสนเทศส่วนใหญ่ในปัจจุบันยังคงจำกัดอยู่เพียงการจัดเก็บและนำเสนอข้อมูลเท่านั้น โดยขาดความสามารถในการให้คำแนะนำเชิงรุก (Proactive Recommendation) หรือการคาดการณ์แนวโน้ม (Prediction) ที่สามารถสนับสนุนการตัดสินใจได้อย่างมีประสิทธิภาพ

การพัฒนาแนะแนวในยุคดิจิทัลจึงจำเป็นต้องนำเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) เข้ามาผนวกเพื่อเพิ่มศักยภาพในการวิเคราะห์และทำนายแนวโน้มของผู้เรียน (สำนักงานเลขาธิการสภาการศึกษา, 2563) ความก้าวหน้าของเทคนิคการเรียนรู้ของเครื่องสามารถช่วยวิเคราะห์รูปแบบที่ซับซ้อนจากข้อมูลขนาดใหญ่ (Big Data) ของนักเรียน เช่น ผลการเรียนรู้ บุคลิกภาพ หรือประวัติศิษย์เก่าที่ประสบความสำเร็จ เป็นต้น เพื่อนำมาสร้างแบบจำลองเชิงทำนาย (Predictive Model) ที่ช่วยให้ระบบสามารถคาดการณ์ความเหมาะสมในการศึกษาต่อในแต่ละสาขาได้อย่างแม่นยำ นอกจากนี้ ยังสามารถให้คำแนะนำเฉพาะบุคคล (Personalized Guidance) ที่สอดคล้องกับคุณลักษณะเฉพาะของนักเรียนแต่ละคนได้อย่างเป็นกลาง

จากเหตุผลและความสำคัญดังกล่าว ผู้วิจัยจึงมีแนวคิดในการสร้างตัวแบบสำหรับการแนะแนวการศึกษาต่อโดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) เพื่อเพิ่มประสิทธิภาพในการวิเคราะห์และทำนายแนวโน้มการเลือกสาขาวิชาที่เหมาะสมกับศักยภาพของผู้เรียนแต่ละคน การดำเนินการวิจัยในครั้งนี้มุ่งเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่อง จำนวน 3 แบบ ได้แก่ อัลกอริทึมการถดถอยโลจิสติก อัลกอริทึมป่าสุ่ม และอัลกอริทึมเกรเดียนต์บูสต์ติง เพื่อหาวิธีการทำนายที่มีความแม่นยำและสอดคล้องกับบริบทของสถานศึกษาได้ดีที่สุด ผลลัพธ์ที่ได้จะถูกนำไปใช้ในการพัฒนา เครื่องมือสนับสนุนการตัดสินใจ สำหรับนักเรียนและครูแนะแนวให้สามารถให้คำแนะนำเชิงวิเคราะห์และเฉพาะบุคคลได้อย่างมีประสิทธิภาพ โดยเลือกใช้วิทยาลัยอาชีวศึกษาอุดรธานี เป็นพื้นที่กรณีศึกษาในการพัฒนาและทดลองใช้ระบบต้นแบบ ซึ่งคาดว่า ผลการวิจัยจะช่วยยกระดับคุณภาพของงานแนะแนวให้มีความเป็นระบบและมีพื้นฐานทางวิทยาศาสตร์มากยิ่งขึ้น ตอบสนองต่อความเปลี่ยนแปลงของยุคดิจิทัล และสามารถเป็นแนวทางให้กับวิทยาลัยอาชีวศึกษาอื่น ๆ ในการนำเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence: AI) มาประยุกต์ใช้ เพื่อพัฒนากำลังคนอาชีวศึกษาที่มีคุณภาพ มีศักยภาพในการแข่งขัน และสามารถเลือกเส้นทางการศึกษาหรืออาชีพได้อย่างมั่นคงในอนาคต

วัตถุประสงค์

1. เพื่อสร้างตัวแบบสำหรับการแนะแนวและสนับสนุนการตัดสินใจเลือกแผนกวิชาสำหรับนักศึกษาใหม่ โดยประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่อง
2. เพื่อเปรียบเทียบประสิทธิภาพของโมเดลอัลกอริทึมการถดถอยโลจิสติก อัลกอริทึมป่าสุ่ม และอัลกอริทึมเกรเดียนต์บูสต์ติงในการแนะแนวการศึกษาต่อของผู้เรียน

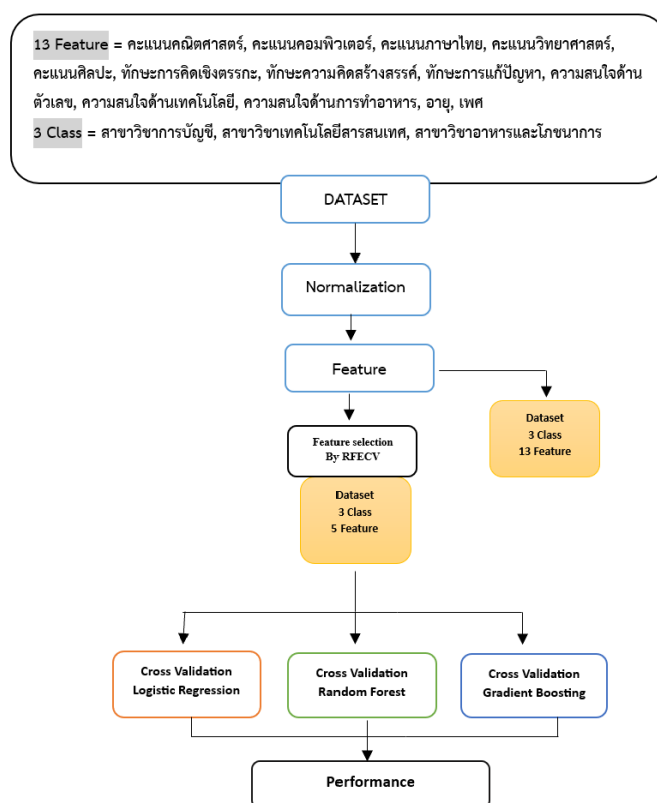
วิธีดำเนินการวิจัย

1. การเตรียมข้อมูลสำหรับการทดสอบประสิทธิภาพแบบจำลองโมเดล

การเก็บข้อมูลในการวิจัยนี้ ได้รวบรวมข้อมูลจากนักเรียนระดับประกาศนียบัตรวิชาชีพ (ปวช.) ชั้นปีที่ 1 ถึงชั้นปีที่ 3 วิทยาลัยอาชีวศึกษาอุดรธานี เป็นนักเรียนสาขาวิชาการบัญชี จำนวน 504 คน สาขาวิชาเทคโนโลยีสารสนเทศ จำนวน 601 คน และสาขาวิชาอาหารและโภชนาการ จำนวน 684 คน รวมจำนวนทั้งสิ้น 1,789 คน จากนักเรียนทั้งหมด 3,061 คน เมื่อทำความสะอาดข้อมูล ด้วยการลบข้อมูลรายการที่ไม่สมบูรณ์ออกทำให้เหลือข้อมูล 1,789 รายการ และประกอบไปด้วย 13 คุณลักษณะ ดังนี้ (1) คะแนนคณิตศาสตร์ (2) คะแนนคอมพิวเตอร์ (3) คะแนนภาษาไทย (4) คะแนนวิทยาศาสตร์ (5) คะแนนศิลปะ (6) ทักษะการคิดเชิงตรรกะ (7) ทักษะความคิดสร้างสรรค์ (8) ทักษะการแก้ปัญหา (9) ความสนใจด้านตัวเลข (10) ความสนใจด้านเทคโนโลยี (11) ความสนใจด้านการทำอาหาร (12) อายุ และ (13) เพศ และเป้าหมาย จำนวน 3 เป้าหมาย คือ (1) สาขาวิชาการบัญชี (2) สาขาวิชาเทคโนโลยีสารสนเทศ และ (3) สาขาวิชาอาหารและโภชนาการ

2. ขั้นตอนการสร้างแบบจำลองโมเดล

การวิจัยในครั้งนี้ได้สร้างโมเดล ดังภาพที่ 1 แสดงให้เห็นถึงการนำเข้าสู่ข้อมูลทั้งหมด 1,789 รายการ โดยแบ่งเป็นข้อมูลสำหรับฝึกสอนและข้อมูลสำหรับทดสอบ 80/20 โดยมี 13 คุณลักษณะ เพื่อนำไปประมวลผลข้อมูลด้วยเทคนิคการเรียนรู้ของเครื่อง และประเมินประสิทธิภาพของแบบจำลอง ด้วยวิธีการตรวจสอบความถูกต้องแบบไขว้ 10 ส่วน และคัดเลือกคุณลักษณะด้วยเทคนิคการตัดตัวแปรออกทีละตัวจนเหลือชุดที่ดีที่สุด (RFECV) ซึ่งมีโมเดลในการทำนาย 3 โมเดล คือ อัลกอริทึมการถดถอยโลจิสติก อัลกอริทึมป่าสุ่ม และอัลกอริทึมเกรเดียนต์บูสต์ติง โดยมีกระบวนการวิธีดังดังนี้



ภาพที่ 1 แสดงแผนผังขั้นตอนการทดลองโมเดล

จากภาพที่ 1 แสดงขั้นตอนการดำเนินการงานวิจัย ในขั้นตอนการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง หลังจากที่ได้ทำการศึกษาแล้ว จึงได้มีการวางแผนการขั้นตอนการดำเนินการขั้นตอนการสร้างแบบจำลอง ตามแผนผังแสดงดังภาพที่ 1 โดยมีรายละเอียดดังนี้

2.1 การเตรียมข้อมูล จำนวน 1,789 เรคคอร์ด ในรูปแบบของไฟล์ Excel ซึ่งยังไม่สามารถนำมาวิเคราะห์ได้

2.2 การทำ Normalization ผู้วิจัยได้ปรับข้อมูลให้อยู่ในช่วงเดียวกัน เพื่อปรับค่าข้อมูลให้อยู่ในช่วงมาตรฐานเดียวกัน เพื่อลดอคติของคุณลักษณะที่อาจส่งผลกระทบต่อกระบวนการเรียนรู้ของเครื่อง โดยใช้เทคนิคการปรับข้อมูลให้อยู่ในรูปมาตรฐานด้วยวิธีซีสกอร์ ซึ่งจะทำให้ข้อมูลมีค่าเฉลี่ยใกล้เคียงศูนย์ และส่วนเบี่ยงเบนมาตรฐานใกล้เคียงหนึ่ง

2.3 การแบ่งชุดข้อมูล หลังจากได้ข้อมูลที่ได้ผ่านการปรับขนาดและทำการปรับข้อมูลให้อยู่ในช่วงเดียวกันเรียบร้อยแล้ว ผู้วิจัยได้ดำเนินการแบ่งชุดข้อมูลออกเป็น 2 ส่วน ได้แก่

2.3.1 ชุดข้อมูลสำหรับฝึกสอนแบบจำลอง จำนวน 1,431 รายการ คิดเป็นร้อยละ 80 ของข้อมูลทั้งหมด

2.3.2 ชุดข้อมูลสำหรับทดสอบแบบจำลอง จำนวน 358 รายการ คิดเป็นร้อยละ 20 ของข้อมูลทั้งหมด

การแบ่งชุดข้อมูลในลักษณะนี้มีวัตถุประสงค์เพื่อให้แบบจำลองสามารถเรียนรู้รูปแบบของข้อมูลจากชุดฝึก และตรวจสอบความสามารถในการทำนายข้อมูลใหม่จากชุดทดสอบได้อย่างเป็นอิสระ เพื่อป้องกันการเกิดการจำข้อมูลฝึกมากเกินไป จนไม่สามารถทำนายข้อมูลใหม่ได้อย่างมีประสิทธิภาพ

2.4 การคัดเลือกคุณลักษณะ เพื่อเพิ่มประสิทธิภาพในการสร้างแบบจำลอง ผู้วิจัยได้ทำการคัดเลือกคุณลักษณะ โดยใช้วิธีเทคนิค RFECV (Recursive Feature Elimination with Cross Validation) ซึ่งเป็นกระบวนการในการคัดเลือกคุณลักษณะที่สำคัญ โดยเริ่มต้นจากคุณลักษณะทั้งหมด จากนั้นจะค่อย ๆ ตัดคุณลักษณะที่มีความสำคัญน้อยที่สุดออกทีละตัว พร้อมประเมินประสิทธิภาพของแบบจำลองด้วยวิธี Cross Validation ในแต่ละรอบ เพื่อเลือกชุดคุณลักษณะที่ให้ผลการทำนายดีที่สุด

2.5 การสร้างแบบจำลองการเรียนรู้ของเครื่อง

2.5.1 แบบจำลองที่ 1 นำข้อมูลที่ได้จากกระบวนการคัดกรองคุณลักษณะด้วยเทคนิค RFECV 1,431 รายการ นำเข้ากระบวนการวัดประสิทธิภาพโมเดล โดยใช้ 10-Fold Cross Validation มีการกำหนดพารามิเตอร์ให้ทำ 10 รอบ ซึ่งใช้อัลกอริทึมการถดถอยโลจิสติก เพื่อหาค่าประสิทธิภาพของตัวแบบทำนาย

2.5.2 แบบจำลองที่ 2 นำข้อมูล ที่ได้จากกระบวนการคัดกรองคุณลักษณะด้วยเทคนิค RFECV 1,431 รายการ นำเข้ากระบวนการวัดประสิทธิภาพโมเดลโดยใช้ 10-Fold Cross Validation มีการกำหนดพารามิเตอร์ให้ทำ 10 รอบ ซึ่งใช้ อัลกอริทึมป่าสุ่ม เพื่อหาค่าประสิทธิภาพของตัวแบบทำนาย

2.5.3 แบบจำลองที่ 3 นำข้อมูลที่ได้จากกระบวนการคัดกรองคุณลักษณะด้วยเทคนิค RFECV 1,431 รายการ นำเข้ากระบวนการวัดประสิทธิภาพโมเดลโดยใช้ 10-Fold Cross Validation มีการกำหนดพารามิเตอร์ให้ทำ 10 รอบ ซึ่งใช้อัลกอริทึมเกรเดียนต์บูสต์ติง เพื่อหาค่าประสิทธิภาพของตัวแบบทำนาย

3. การประเมินและหาประสิทธิภาพ

ในการวัดประสิทธิภาพของโมเดลที่สร้างขึ้น ผู้วิจัยได้กำหนดเกณฑ์ในการประเมินประสิทธิภาพโดยใช้ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าความถ่วงดุล จากสมการที่ 1-4 ตามลำดับดังนี้

3.1 Confusion Matrix คือ เครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนาย หรือ Prediction ที่ทำนายจาก Model ที่สร้างขึ้น ในการเรียนรู้ของเครื่อง โดยมีความคิดมาจากการวัดว่า สิ่งที่เกิดขึ้นจริง โดยจะสามารถบอกค่า Accuracy, Precision, Recall, และ F1-Score ดังสมการที่ (1), (2), (3) และ (4) ซึ่งวัดว่า โมเดลนั้นมีประสิทธิภาพเพียงใด

3.2 Accuracy คือ ค่าความถูกต้อง ค่าที่บอกว่า โปรแกรมสามารถทำนายได้แม่นยำขนาดไหน เป็นการวัดความถูกต้องของ Model โดยพิจารณาจากทุกคลาส

$$\text{Accuracy} = \frac{TP_A + TP_B + TP_C}{\text{Total}} \quad (1)$$

3.3 Precision คือ ค่าความแม่นยำ ค่าที่บอกว่า โปรแกรมทำนายว่าจริง ถูกต้องเท่าไร เป็นการวัดความแม่นยำของข้อมูล โดยพิจารณาแยกทีละคลาส

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \quad (2)$$

3.4 Recall คือ ค่าความระลึก ค่าที่บอกว่าโปรแกรมทำนายได้จริง เป็นอัตราส่วนเท่าไรของค่าจริงทั้งหมดเป็นการวัดความถูกต้องของ Model โดยพิจารณาแยกทีละคลาส

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \quad (3)$$

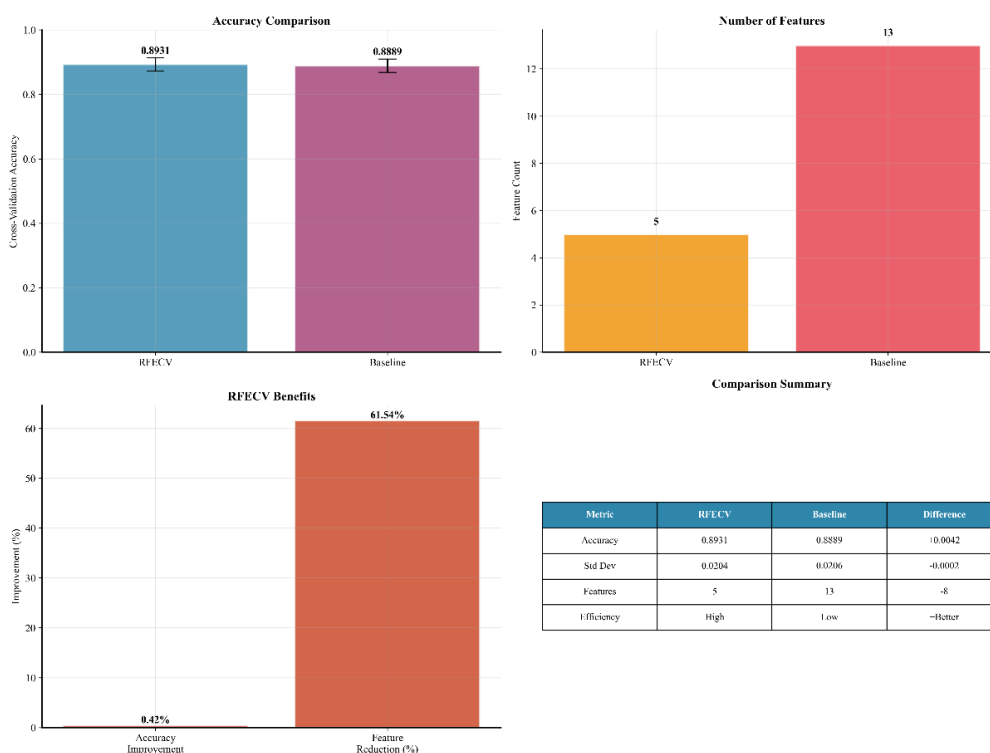
3.5 f1_measure คือ ค่าความถ่วงดุล เป็นการนำค่า Precision กับ Recall มาคำนวณค่าเฉลี่ยกันสำหรับพิจารณาร่วมกัน เพื่อความแม่นยำและความครบถ้วน

$$\text{F1 - Score}_i = 2 \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \quad (4)$$

ผลการวิจัย

ผลการศึกษาในครั้งนี้เป็นการเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่องสำหรับการทำนายแนวโน้มการเลือกสาขาวิชาของนักเรียน โดยใช้ข้อมูลนักเรียนจาก 3 สาขาวิชา ได้แก่ สาขาวิชาการบัญชี สาขาวิชาเทคโนโลยีสารสนเทศ และสาขาวิชาอาหารและโภชนาการ จำนวนทั้งสิ้น 1,789 รายการ ซึ่งประกอบด้วยคุณลักษณะจำนวน 13 คุณลักษณะ ก่อนการสร้างแบบจำลองได้ดำเนินการคัดเลือกคุณลักษณะด้วยเทคนิค Recursive Feature Elimination with Cross-Validation (RFECV) ซึ่งเป็นกระบวนการคัดเลือกคุณลักษณะโดยเริ่มจากใช้คุณลักษณะทั้งหมด จากนั้นจึงค่อย ๆ ตัดคุณลักษณะที่มีความสำคัญน้อยที่สุดออกทีละตัว พร้อมประเมินประสิทธิภาพของแบบจำลองด้วยวิธี 10-Fold Cross Validation ในแต่ละรอบ เพื่อค้นหาชุดคุณลักษณะที่ให้ประสิทธิภาพการทำนายดีที่สุด จากผลการคัดเลือกคุณลักษณะ พบว่า เหลือคุณลักษณะที่สำคัญจำนวน 5 คุณลักษณะ ได้แก่ คะแนนคณิตศาสตร์ คะแนนคอมพิวเตอร์ คะแนนภาษาไทย คะแนนวิทยาศาสตร์ และคะแนนศิลปะ ซึ่งถูกนำไปใช้ในการสร้างและเปรียบเทียบประสิทธิภาพของแบบจำลองการถดถอยโลจิสติก ป่าสุ่ม และเกรเดียนต์บูสต์ติง โดยประเมินประสิทธิภาพด้วยวิธี 10-Fold Cross Validation และชุดข้อมูลทดสอบ เพื่อเปรียบเทียบค่าความถูกต้อง ความแม่นยำ ความระลึก และค่าความถ่วงดุล ของแต่ละแบบจำลอง ดังแสดงผลการทดลองต่อไปนี้

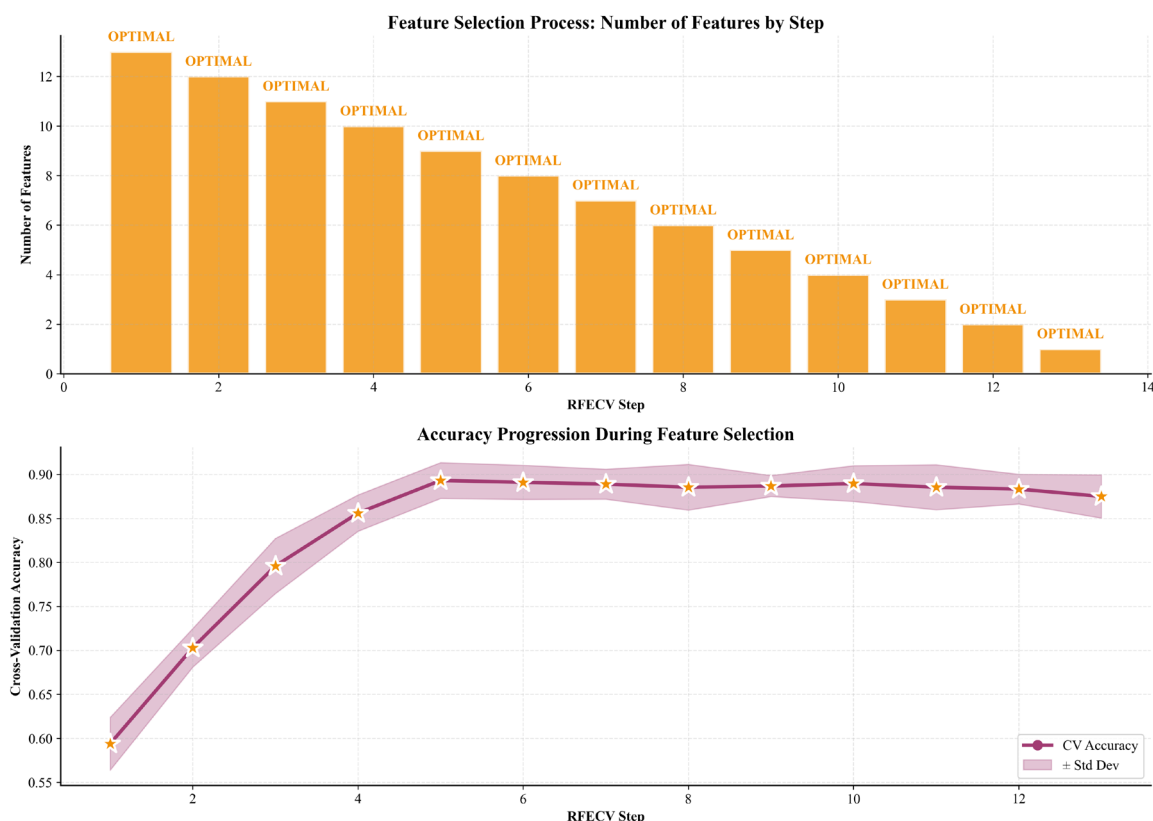
1. ผลการเปรียบเทียบการคัดเลือกคุณลักษณะ



ภาพที่ 1 ผลการเปรียบเทียบประสิทธิภาพการคัดเลือกคุณลักษณะด้วยเทคนิค RFECV

จากภาพที่ 1 ผลการเปรียบเทียบการคัดเลือกคุณลักษณะด้วยเทคนิค Recursive Feature Elimination with Cross-Validation (RFECV) กับการใช้คุณลักษณะทั้งหมด (Baseline) พบว่า RFECV สามารถลดจำนวนคุณลักษณะที่ใช้ในการสร้างแบบจำลองจาก 13 คุณลักษณะ เหลือเพียง 5 คุณลักษณะ คิดเป็นการลดลง ร้อยละ 61.54% โดยมีค่าการใช้เทคนิค RFECV มีค่าของความถูกต้อง เท่ากับ 89.31% ขณะที่การใช้คุณลักษณะทั้งหมด (Baseline) ให้ค่าความถูกต้อง เท่ากับ 88.89% ซึ่งสูงกว่าประมาณ 0.42% แสดงให้เห็นว่า การคัดเลือกคุณลักษณะสามารถช่วยคัดกรองเฉพาะคุณลักษณะที่มีความสำคัญต่อการทำนาย ทำให้แบบจำลองมีประสิทธิภาพในการเรียนรู้เพิ่มขึ้น

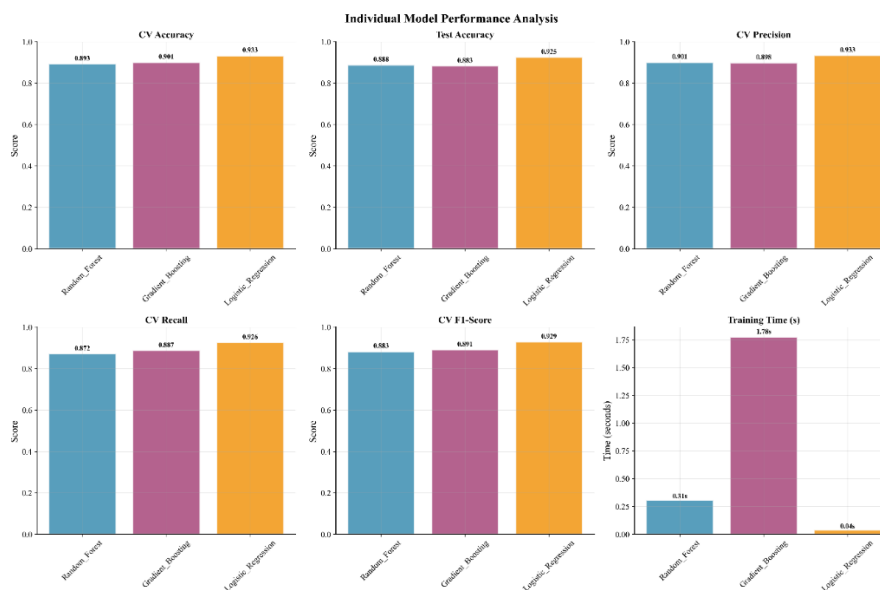
2. ผลการคัดเลือกคุณลักษณะ



ภาพที่ 2 ผลการคัดเลือกคุณลักษณะด้วยเทคนิค
Recursive Feature Elimination with Cross-Validation (RFECV)

จากภาพที่ 2 ผลการวิเคราะห์ RFECV แสดงขั้นตอนการลดคุณลักษณะ กราฟด้านล่างแสดงให้เห็นว่า การตัดคุณลักษณะที่ไม่เกี่ยวข้องในช่วงแรก ทำให้ความแม่นยำเฉลี่ยเพิ่มขึ้นอย่างรวดเร็ว จากประมาณ 60% ไปสู่ระดับสูงสุดที่ประมาณ 88% ซึ่งเป็นแสดงให้เห็นว่า การคัดเลือกคุณลักษณะช่วยเพิ่มประสิทธิภาพของแบบจำลองได้อย่างชัดเจน ซึ่งให้ความแม่นยำสูงสุดในขณะที่มีความซับซ้อนต่ำ ทำให้ได้แบบจำลองที่เรียบง่ายและเสถียรเพิ่มขึ้น

3. ผลการวิเคราะห์ประสิทธิภาพแบบจำลอง



ภาพที่ 3 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่อง

จากภาพที่ 3 ผลการเปรียบเทียบประสิทธิภาพของชุดข้อมูลสำหรับฝึกสอนแบบจำลอง จำนวน 1,431 รายการ ของแบบจำลองการเรียนรู้ของเครื่อง ได้แก่ การถดถอยโลจิสติก ป่าสุ่ม และเกรเดียนต์บูสต์ติง พบว่า การถดถอยโลจิสติก มีประสิทธิภาพดีที่สุด โดยให้ค่าความถูกต้องจากการตรวจสอบความถูกต้องแบบไขว้ เท่ากับ 93.3% และค่าความถูกต้องของชุดข้อมูลทดสอบเท่ากับ 92.5% ค่าความแม่นยำ เท่ากับ 93.3% ความระลึก เท่ากับ 92.6% และค่าความถ่วงดุล เท่ากับ 92.9% เวลาในการฝึกแบบจำลอง 0.04 วินาที ซึ่งน้อยกว่าแบบจำลองเกรเดียนต์บูสต์ติงและป่าสุ่ม ให้ผลการทำนายในระดับที่ดี แต่ใช้เวลาในการฝึกแบบจำลองมากกว่า โดยเฉพาะเกรเดียนต์บูสต์ติง ซึ่งใช้เวลาในการฝึกแบบจำลองสูงที่สุด คือ 1.78 วินาที

4. ผลการวัดประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่อง

ตารางที่ 1

ผลการวัดประสิทธิภาพข้อมูลแบบจำลองการถดถอยโลจิสติก

	สาขาวิชา การบัญชี	สาขาวิชาเทคโนโลยี สารสนเทศ	สาขาวิชาอาหารและ โภชนาการ	Class precision
สาขาวิชาการบัญชี	67	3	5	89.33%
สาขาวิชาเทคโนโลยีสารสนเทศ	4	81	9	86.17%
สาขาวิชาอาหารและโภชนาการ	2	4	183	81.92%
Class recall	91.78%	92.05%	92.89%	

จากตารางที่ 1 ผลการวัดประสิทธิภาพแบบจำลองการถดถอยโลจิสติก ข้อมูลทดสอบแบบจำลอง จำนวน 358 รายการ ค่าความถูกต้องมีค่า TP = 67, TN = 277, FP = 8, FN = 6 เมื่อแทนค่าในสมการที่ (1), (2), (3) และ (4) จะได้ค่าความถูกต้อง เท่ากับ 92.46% ค่าความแม่นยำ เท่ากับ 90.78% ค่าความระลึก เท่ากับ 92.24% ค่าความถ่วงดุล เท่ากับ 91.50% แสดงให้เห็นถึงผลการวัดประสิทธิภาพข้อมูลด้วยวิธีการถดถอยโลจิสติก

ตารางที่ 2

ผลการวัดประสิทธิภาพข้อมูลแบบจำลองเกรเดียนต์บูสต์ติง

	สาขาวิชา การบัญชี	สาขาวิชาเทคโนโลยี สารสนเทศ	สาขาวิชาอาหาร และโภชนาการ	Class precision
สาขาวิชาการบัญชี	60	6	9	80.00%
สาขาวิชาเทคโนโลยีสารสนเทศ	3	76	15	80.85%
สาขาวิชาอาหารและโภชนาการ	4	5	180	95.24%
Class recall	89.55%	87.36%	88.24%	

จากตารางที่ 2 ผลการวัดประสิทธิภาพแบบจำลองเกรเดียนต์บูสต์ติง ข้อมูลทดสอบแบบจำลอง จำนวน 358 รายการ ค่าความถูกต้อง มีค่า TP = 60, TN = 276, FP = 15, FN = 7 เมื่อแทนค่าในสมการที่ (1), (2), (3) และ (4) จะได้ค่าความถูกต้อง เท่ากับ 88.27% ค่าความแม่นยำ เท่ากับ 85.36% ค่าความระลึก เท่ากับ 88.38% ค่าความถ่วงดุล เท่ากับ 86.85% แสดงให้เห็นถึงผลการวัดประสิทธิภาพข้อมูลด้วยวิธีเกรเดียนต์บูสต์ติง

ตารางที่ 3

ผลการวัดประสิทธิภาพข้อมูลแบบจำลองแบบป่าสุ่ม

	สาขาวิชา การบัญชี	สาขาวิชาเทคโนโลยี สารสนเทศ	สาขาวิชาอาหาร และโภชนาการ	Class precision
สาขาวิชาการบัญชี	56	6	12	76.00%
สาขาวิชาเทคโนโลยีสารสนเทศ	3	77	14	81.91%
สาขาวิชาอาหารและโภชนาการ	2	3	184	97.35%
Class recall	91.94%	89.53%	87.62%	

จากตารางที่ 3 ผลการวัดประสิทธิภาพแบบจำลองแบบป่าสุ่ม ข้อมูลทดสอบแบบจำลอง จำนวน 358 รายการ ค่าความถูกต้อง มีค่า TP = 57, TN = 278, FP = 18, FN = 5 เมื่อแทนค่าในสมการที่ (1), (2), (3) และ (4) จะได้ค่าความถูกต้อง เท่ากับ 88.27% ค่าความแม่นยำ เท่ากับ 85.36% ค่าความระลึก เท่ากับ 88.38% ค่าความถ่วงดุล เท่ากับ 86.85% แสดงให้เห็นถึงผลการวัดประสิทธิภาพข้อมูลด้วยวิธีแบบป่าสุ่ม

ตารางที่ 4

สรุปผลการเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่อง

ตัวแบบ	ค่าความถูกต้อง	ค่าความแม่นยำ	ค่าความระลึก	ค่าความถ่วงดุล	เวลา (วินาที)
การถดถอยโลจิสติก	92.46	90.78	92.24	91.50	0.04
เกรเดียนต์บูสต์ติง	88.27	85.36	88.38	86.85	1.78
ป่าสุ่ม	88.83	85.09	89.70	87.33	0.31

จากตารางที่ 4 ผลการทดลอง พบว่า แบบจำลองการถดถอยโลจิสติกมีประสิทธิภาพโดยรวมดีที่สุด โดยมีค่าความถูกต้อง เท่ากับ 92.46% ซึ่งเป็นค่าสูงที่สุดเมื่อเปรียบเทียบกับแบบจำลองอื่น นอกจากนี้ ยังมีค่าความแม่นยำ เท่ากับ 90.78%, ค่าความระลึก เท่ากับ 92.24% และ ค่าความถ่วงดุล เท่ากับ 91.50% พร้อมทั้ง ใช้เวลาในการฝึกแบบจำลอง 0.04 วินาที ซึ่งน้อยที่สุด แสดงให้เห็นถึงความสามารถในการจำแนกข้อมูลได้อย่างมีประสิทธิภาพและรวดเร็ว

สรุปผลการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาและเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่องในการทำนายแนวโน้มการเลือกสาขาวิชาเรียนของนักเรียนระดับประกาศนียบัตรวิชาชีพ (ปวช.) โดยใช้ข้อมูลนักเรียนจำนวน 1,789 ตัวอย่าง ประกอบด้วย 13 คุณลักษณะ โดยแบ่งเป็นชุดข้อมูลสำหรับฝึกสอน และชุดข้อมูลสำหรับทดสอบ 80/20 และมุ่งเน้นการจำแนกข้อมูลออกเป็น 3 สาขาวิชา ได้แก่ สาขาวิชาการบัญชี สาขาวิชาเทคโนโลยีสารสนเทศ และสาขาวิชาอาหาร ผู้วิจัยได้ทำการเตรียมข้อมูลและคัดเลือกคุณลักษณะด้วยเทคนิคการตัดตัวแปรออกทีละตัว จนเหลือชุดที่ดีที่สุด (RFECV) ก่อนนำข้อมูลไปสร้างแบบจำลองการเรียนรู้ของเครื่อง จำนวน 3 แบบ ได้แก่ อัลกอริทึมการถดถอยโลจิสติก อัลกอริทึมป่าสุ่ม และอัลกอริทึมเกรเดียนต์บูสต์ติง จากนั้นประเมินประสิทธิภาพของแบบจำลองด้วยวิธี 10-Fold Cross Validation และชุดข้อมูลทดสอบ โดยพิจารณาจาก ค่าความถูกต้อง ความแม่นยำ ความระลึก และค่าความถ่วงดุล

ผลการวิจัย พบว่า แบบจำลองอัลกอริทึมการถดถอยโลจิสติกให้ประสิทธิภาพสูงที่สุด โดยมีค่าความถูกต้อง เท่ากับ 92.46% ค่าความแม่นยำ เท่ากับ 90.78% ค่าความระลึก เท่ากับ 92.24% และค่าความถ่วงดุล เท่ากับ 91.50% อีกทั้ง ยังใช้เวลาในการฝึกแบบจำลองเพียง 0.04 วินาที ซึ่งน้อยที่สุดเมื่อเปรียบเทียบกับแบบจำลองอื่น ขณะที่อัลกอริทึมเกรเดียนต์บูสต์ติงและอัลกอริทึมป่าสุ่มให้ผลการทำนายในระดับที่ดี แต่มีค่าประสิทธิภาพโดยรวมต่ำกว่าอัลกอริทึมการถดถอยโลจิสติก และใช้เวลาในการฝึกแบบจำลองมากกว่า จากผลการเปรียบเทียบสามารถสรุปได้ว่า แบบจำลองอัลกอริทึมการถดถอยโลจิสติกมีความเหมาะสมที่สุดสำหรับการทำนายแนวโน้มการเลือกสาขาวิชาเรียนของนักเรียนระดับประกาศนียบัตรวิชาชีพ ในชุดข้อมูลที่ใช้ในการศึกษาครั้งนี้

การอภิปรายผล

ผลการวิจัยแสดงให้เห็นว่า แบบจำลองอัลกอริทึมการถดถอยโลจิสติกมีประสิทธิภาพสูงกว่าอัลกอริทึมป่าสุ่มและอัลกอริทึมเกรเดียนต์บูสต์ติง ทั้งในด้านความแม่นยำ ความถูกต้องในการจำแนกข้อมูล และระยะเวลาในการประมวลผล ทั้งนี้ อาจเนื่องมาจากข้อมูลที่ใช้ในการวิจัยมีคุณลักษณะที่ผ่านการคัดเลือกด้วยเทคนิคการตัดตัวแปรออกทีละตัวจนเหลือชุดที่ดีที่สุด (RFECV) ทำให้เหลือเฉพาะคุณลักษณะที่มีความสัมพันธ์กับตัวแปรเป้าหมาย ส่งผลให้แบบจำลองอัลกอริทึมการถดถอยโลจิสติกสามารถเรียนรู้รูปแบบของข้อมูลได้อย่างมีประสิทธิภาพ โดยไม่เกิดความซับซ้อนของแบบจำลองมากเกินไป

เมื่อพิจารณาผลการตรวจสอบด้วยวิธีการตรวจสอบแบบไขว้ พบว่า อัลกอริทึมการถดถอยโลจิสติกมีค่าความแม่นยำเฉลี่ยสูงที่สุด และมีการกระจายตัวของค่าความแม่นยำในแต่ละส่วนย่อยค่อนข้างต่ำ แสดงถึงความเสถียรของแบบจำลองและความสามารถในการนำไปประยุกต์ใช้กับข้อมูลใหม่ได้ดี นอกจากนี้ ผลการวิเคราะห์ภาวะการเรียนรู้มากเกินไปยังพบว่า อัลกอริทึมการถดถอยโลจิสติกมีความแตกต่างระหว่างค่าความถูกต้องของชุดข้อมูลฝึก และค่าความถูกต้องของชุดข้อมูลทดสอบต่ำกว่าแบบจำลองอื่น สะท้อนให้เห็นว่าแบบจำลองสามารถสร้างสมดุลระหว่างการเรียนรู้ข้อมูลฝึกและการทำนายข้อมูลใหม่ได้อย่างเหมาะสม สำหรับอัลกอริทึมเกรเดียนต์บูสต์ติงและอัลกอริทึมป่าสุ่ม แม้ว่าจะสามารถจำแนกข้อมูลได้ในระดับที่ดี แต่ยังมีข้อจำกัดในด้านเวลาในการฝึกแบบจำลองและมีแนวโน้มเกิดภาวะการเรียนรู้มากเกินไปมากกว่าอัลกอริทึมการถดถอยโลจิสติก โดยเฉพาะอัลกอริทึมเกรเดียนต์บูสต์ติงที่ใช้เวลาในการฝึกสูงที่สุด ขณะที่อัลกอริทึมป่าสุ่มมีเวลาในการฝึกน้อยกว่าอัลกอริทึมเกรเดียนต์บูสต์ติง แต่มีค่าความระลึกและค่าความถ่วงดุลที่ต่ำกว่า

ผลการวิจัยครั้งนี้จึงแสดงให้เห็นว่า การเลือกแบบจำลองที่เหมาะสมควรพิจารณาทั้งด้านความแม่นยำ ความเสถียร และประสิทธิภาพในการประมวลผลร่วมกัน ไม่ใช่พิจารณาจากค่าความแม่นยำเพียงอย่างเดียว

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 สถานศึกษาสามารถนำแบบจำลองอัลกอริทึมการถดถอยโลจิสติก ไปประยุกต์ใช้ในการพัฒนาระบบสนับสนุนการแนะแนวการศึกษาต่อของนักเรียน เพื่อช่วยวิเคราะห์แนวโน้มการเลือกสาขาวิชาได้อย่างรวดเร็วและแม่นยำ

1.2 ผลการวิจัยสามารถใช้เป็นข้อมูลประกอบการวางแผนการรับนักเรียน การจัดสรรทรัพยากร และการพัฒนาหลักสูตรให้สอดคล้องกับแนวโน้มความต้องการของผู้เรียนในอนาคต

2. ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

2.1 ควรเพิ่มจำนวนข้อมูลจากหลายสถานศึกษา เพื่อให้แบบจำลองสามารถเรียนรู้รูปแบบข้อมูลที่หลากหลายและเพิ่มความสามารถในการนำผลไปใช้ในบริบทอื่น

2.2 ควรศึกษาการเปรียบเทียบกับแบบจำลองการเรียนรู้ของเครื่องประเภทอื่น เช่น แบบจำลองซัพพอร์ตเวกเตอร์แมชชีน เอ็กซ์จีบูสต์ หรือโครงข่ายประสาทเทียม เป็นต้น

2.3 ควรพัฒนาระบบต้นแบบสำหรับทำนายแนวโน้มการเลือกสาขาวิชาเรียน โดยเชื่อมโยงแบบจำลองที่ได้จากงานวิจัย เข้ากับระบบสารสนเทศหรือระบบแนะแนวของสถานศึกษา เพื่อให้สามารถนำไปใช้งานได้จริง

เอกสารอ้างอิง

- กาญจน์เชจร ชูชีพ. (2561). การถดถอยโลจิสติก (Logistic Regression). *Remote Sensing Technical Note*, (5), 1-10.
- ชัชวาล สิงคะลิง. (2564). *การพยากรณ์ความเสี่ยงของการพ้นสภาพจากการเป็นนักศึกษา โดยใช้อัลกอริทึม การเรียนรู้ของเครื่อง: กรณีศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา* [การค้นคว้าอิสระปริญญามหาบัณฑิต ไม่ได้ตีพิมพ์]. มหาวิทยาลัยเชียงใหม่.
- ชุดิธารัฐ อุดมมะสิริเสนี และพรสิริ ชาตปรีชา. (2562). เทคนิคเหมืองข้อมูลสำหรับพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจ สำหรับทำนายทักษะต้นโปรแกรมประยุกต์ของนักศึกษาระดับปริญญาตรีในยุคดิจิทัล. *วารสารวิทยาลัยเทคโนโลยีหัวเฉียวเฉลิมพระเกียรติ*, 5(2), 1-10.
- โชคดี วชิรบุญโญ. (2562). การแนะแนวการศึกษาและอาชีพของนิสิตพุทธศาสตรบัณฑิต คณะครุศาสตร์ มหาวิทยาลัยมหาจุฬาลงกรณราชวิทยาลัย. *วารสารมหาวิทยาลัยมหาจุฬาลงกรณราชวิทยาลัย*, 6(1), 5-10.
- ธนารักษ์ สารเลื่อนแก้ว. (2566). การใช้เทคโนโลยีเพื่อการแนะแนวในโรงเรียน. *วารสารวิชาการสถาบันเทคโนโลยีภาคตะวันออกแห่งสุวรรณภูมิ*, 2(1), 14-15.
- ภัสรา โกมณเทียร. (2559). การพัฒนาระบบสารสนเทศเพื่อการแนะแนวการศึกษาของโรงเรียนระดับมัธยมศึกษา ในจังหวัดมหาสารคามที่ร่วมเป็นโรงเรียนเครือข่ายกับทางมหาวิทยาลัยมหาสารคาม. *วารสารศึกษาศาสตร์*, 27(1), 3-4.
- ภุมริน หรั่งน้อย และคุณัญญา สัมเกลี้ยง. (2563). การประยุกต์ใช้เทคนิคเหมืองข้อมูล เพื่อแนะนำอาชีพด้านไอทีสำหรับนักศึกษาระดับปริญญาตรี กรณีศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี. ใน *การประชุมมหาดใหญ่วิชาการระดับชาติและนานาชาติ มหาวิทยาลัยมหาดใหญ่* (น.201-212). มหาวิทยาลัยมหาดใหญ่.
- วรเทพ วันกาล. (2564). *ระบบพยากรณ์เพื่อสนับสนุนการตัดสินใจศึกษาต่อหลังสำเร็จมัธยมศึกษาตอนต้น โดยใช้อัลกอริทึมป่าสุ่ม* [วิทยานิพนธ์ปริญญามหาบัณฑิต ไม่ได้ตีพิมพ์]. มหาวิทยาลัยสุโขทัยธรรมาธิราช.
- สำนักงานคณะกรรมการการอาชีวศึกษา. (2555). *แผนพัฒนาการอาชีวศึกษาพ.ศ. 2560 – 2579*.
https://www.kkwtech.ac.th/data_2371
- สำนักงานเลขาธิการสภาการศึกษากระทรวงศึกษาธิการ. (2563). *AI เพื่อพัฒนาการเรียนรู้*. <https://backoffice.onec.go.th/uploads/Book/1805-file.pdf>
- Aman. (2025, September 14). *Feature selection in machine learning*. Analytics Vidhya.
<https://www.analyticsvidhya.com/blog/2020/10/feature-selection-techniques-in-machine-learning/>
- Belcic, I., & Stryker, C. (2025, September 14). *What is feature selection?* IBM. <https://www.ibm.com/think/topics/feature-selection>
- Breiman, L. (2001). *Random forests*. University of California Berkeley.
- Kavlakoglu, E. (2025, September 12). *What is random forest?* IBM. <https://www.ibm.com/think/topics/random-forest>