

แบบจำลองการขุดข้อมูลทางการศึกษาโดยใช้อัลกอริทึมต้นไม้ตัดสินใจ
เพื่อพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนักเรียน
Educational Data Mining Model Using Decision Tree Algorithm
for Student Academic Performance Prediction

เยาวภา ศรีบุรินทร์¹, ชัยมงคล ปินะสา², อนุภูมิ คำยัง³, แพรวนภา เรียงริลา⁴
Yaowapha Sriburin¹, Chaimongkhon Pinasa², Anuphum Kumyoung³, Praewnapa Riengrila⁴

(วันรับบทความ 24 พฤษภาคม 2568, วันแก้ไขบทความ 20 กรกฎาคม 2568, วันตอบรับบทความ 20 กุมภาพันธ์ 2569)

บทคัดย่อ

บทความวิชาการนี้มีวัตถุประสงค์เพื่อ (1) นำเสนอความหมายและแนวคิดที่เกี่ยวข้องกับการขุดข้อมูลทางการศึกษา ต้นไม้ตัดสินใจ และโปรแกรม Altair AI Studio v.2025.0.0 และ (2) แสดงแนวทางการพัฒนาแบบจำลองด้วยอัลกอริทึมต้นไม้ตัดสินใจ โดยใช้โปรแกรม Altair AI Studio v.2025.0.0 สำหรับการพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนักเรียน รวบรวมข้อมูลนักเรียนชั้นประถมศึกษาปีที่ 6 จาก 3 โรงเรียน ได้มาโดยการเลือกแบบเจาะจง โดยใช้ระบบสารสนเทศโรงเรียน โดยยึดตามรูปแบบและโครงสร้างของข้อมูลจริง แต่ข้อมูลบางส่วนไม่สมบูรณ์ จึงได้ใช้เครื่องมือ ChatGPT จำลองข้อมูลจำนวน 1,000 ชุด กระบวนการสร้างแบบจำลอง ประกอบด้วย 5 ขั้นตอน คือ (1) ขั้นตอนการรวบรวมข้อมูล (2) ขั้นตอนเตรียมข้อมูล (3) ขั้นตอนการสร้างแบบจำลอง ด้วยอัลกอริทึมต้นไม้ตัดสินใจ แบ่งข้อมูลเป็น 2 ส่วน คือ ชุดฝึกร้อยละ 80 และชุดทดสอบร้อยละ 20 (4) ขั้นตอนการทดสอบและประสิทธิภาพของแบบจำลอง ด้วยวิธี 10-fold cross-validation และ (5) ขั้นตอนการนำแบบจำลองไปใช้งานจริง โดยพัฒนาในรูปแบบแอปพลิเคชัน One Compiler ซึ่งผู้ใช้สามารถป้อนข้อมูลนักเรียน เพื่อประมวลผลผลสัมฤทธิ์ทางการเรียนล่วงหน้าได้ ผลการทดสอบพบว่าแบบจำลองมีความแม่นยำปานกลาง โดยมีค่าความถูกต้อง (Accuracy) ร้อยละ 64.70 ค่าความคลาดเคลื่อนเฉลี่ย (RMSE) 0.478 และส่วนเบี่ยงเบนมาตรฐาน (S.D.) ± 0.010 ซึ่งสามารถนำไปประยุกต์ใช้ในการวางแผนการจัดการเรียนรู้ การให้คำปรึกษา และการยกระดับคุณภาพการศึกษาให้สอดคล้องกับบริบทของยุคดิจิทัล

คำสำคัญ: แบบจำลองการขุดข้อมูลทางการศึกษา อัลกอริทึมต้นไม้ตัดสินใจ พยากรณ์ผลสัมฤทธิ์ทางการเรียน

¹ นักศึกษาปริญญาโท สาขาวิทยนวัตกรรมการศึกษา มหาวิทยาลัยราชภัฏเลย, sm677911102@lru.ac.th

Master of Education Student, Educational Innovation Research, Loei Rajabhat University

² อาจารย์ คณะครุศาสตร์ มหาวิทยาลัยราชภัฏเลย, Chaimongkhon.Pin@lru.ac.th

Lecturer, Faculty of Education, Loei Rajabhat University

³ อาจารย์ คณะครุศาสตร์ มหาวิทยาลัยราชภัฏเลย, Anuphum.Kum@lru.ac.th

Lecturer, Faculty of Education, Loei Rajabhat University

⁴ อาจารย์ คณะครุศาสตร์ มหาวิทยาลัยราชภัฏเลย, Praewnapa.Ria@lru.ac.th

Lecturer, Faculty of Education, Loei Rajabhat University

Abstract

This article aims to (1) present conceptual frameworks of educational data mining, decision tree algorithms, and their application within Altair AI Studio; and (2) demonstrate the development of a predictive model for student academic performance utilizing decision tree algorithms in Altair AI Studio. Data were purposively collected from existing information systems of Grade 6 students across three schools. To address the issue of incomplete records, 1,000 synthetic data points were generated using ChatGPT to augment the dataset. Model development consisted of five steps: (1) data collection, (2) data preparation, (3) model building with an 80:20 training-testing split, (4) model evaluation via 10-fold cross-validation, and (5) deployment as an application on the One Compiler platform for user to input student information and predict academic performance in advance. Evaluation results indicated the model demonstrated moderate predictive accuracy, achieving an accuracy of 64.70%, a root mean square error (RMSE) of 0.478, and a standard deviation (S.D.) of ± 0.010 . This model provides valuable support for educational planning, counseling, and enhanced educational quality in the digital era.

Keywords: Educational Data Mining Model, Decision Tree Algorithm, Academic Performance Prediction

บทนำ

จากสถานการณ์การระบาดของโรคโควิด-19 ความขัดแย้งในระดับโลก และการพัฒนาทางสังคมและเทคโนโลยีที่เกิดขึ้นอย่างรวดเร็ว ได้ส่งผลให้เกิดการเปลี่ยนแปลงเชิงระบบในการจัดการศึกษาอย่างหลีกเลี่ยงไม่ได้ เพื่อตอบสนองต่อความเปลี่ยนแปลงดังกล่าว นานาประเทศจึงได้กำหนด “วาระการพัฒนาที่ยั่งยืน ค.ศ. 2030” (Sustainable Development Goals: SDGs) โดยเน้นย้ำถึงความสำคัญของการจัดการศึกษาที่มีคุณภาพสูง เพื่อเสริมสร้างศักยภาพของผู้เรียนและยกระดับคุณภาพชีวิตในอนาคต (Simionescu et al., 2024) ทั้งนี้ การเปลี่ยนแปลงทางเทคโนโลยีและการศึกษาที่มีความเป็นดิจิทัลมากขึ้นได้นำไปสู่การเก็บรวบรวมข้อมูลขนาดใหญ่จากบริบททางการศึกษา การขุดข้อมูลทางการศึกษา (Educational Data Mining: EDM) ได้เข้ามามีบทบาทสำคัญในการตรวจสอบและประยุกต์ใช้อัลกอริทึมเพื่อทำความเข้าใจพฤติกรรมของนักเรียนและบริบททางการศึกษาอย่างลึกซึ้ง อย่างไรก็ตาม ความกังวลเกี่ยวกับความโปร่งใสในการใช้ข้อมูลยังคงเป็นปัญหาสำคัญ ระหว่างบริษัทผู้วิเคราะห์ข้อมูลกับนักเรียน ผู้ปกครอง และสถานศึกษา (Askinadze & Conrad, 2018) ในช่วงสองทศวรรษที่ผ่านมา การวิจัยด้านการขุดข้อมูลทางการศึกษาได้มุ่งเน้นไปที่การพัฒนาเทคนิคเพื่อวิเคราะห์ข้อมูลจำนวนมากจากหลากหลายบริบททางการศึกษา โดยมีเป้าหมายหลักในการพยากรณ์ความก้าวหน้าและผลลัพธ์ทางการเรียนรู้ของนักเรียน เช่น การคาดการณ์อัตราการลาออก ผลการเรียนรู้ และเกรดรายวิชา (Tsiakmaki et al., 2019) ปัจจุบัน สถาบันการศึกษามีการสร้าง รวบรวม และจัดเก็บข้อมูลจากแหล่งและกระบวนการที่หลากหลาย ส่งผลให้ปริมาณข้อมูลเพิ่มขึ้นอย่างต่อเนื่อง ปัจจัยสำคัญที่ส่งเสริมการเพิ่มขึ้นของข้อมูล ได้แก่ การใช้งานคอมพิวเตอร์ การเชื่อมต่ออินเทอร์เน็ต และการนำระบบการจัดการเรียนรู้ (Learning Management System: LMS) มาใช้ในการจัดการเรียนการสอน ซึ่งทำให้เกิดการสร้างข้อมูลจำนวนมากผ่านเทคโนโลยีออนไลน์ เช่น แพลตฟอร์มการเรียนรู้ออนไลน์ เครื่องมือค้นหา (Search Engine) เครือข่ายสังคมออนไลน์ การสื่อสารทางอิเล็กทรอนิกส์ และการรับชมสื่อวิดีโอ นอกจากนี้ ยังมีข้อมูลที่ได้จากการเก็บรวบรวมโดยตรงผ่านกระบวนการประเมินผล รวมถึงการติดตามพฤติกรรมของผู้เรียนในบริบทต่าง ๆ (Corina Simionescu, 2024) อย่างไรก็ตาม ในบริบทของโรงเรียนระดับประถมศึกษา โดยเฉพาะชั้นประถมศึกษาปีที่ 6 พบปัญหาเชิงรูปธรรมหลายประการ อาทิ การทำนายผลการเรียนของนักเรียนที่ยังขาดความแม่นยำ ระบบสนับสนุนการตัดสินใจของครูที่ยังไม่สามารถตอบสนองต่อความแตกต่างระหว่างบุคคลได้อย่างมีประสิทธิภาพ และข้อจำกัดในการให้คำปรึกษาเชิงลึกแก่นักเรียนรายบุคคล ซึ่งอาจส่งผลกระทบต่อกระบวนการวางแผนการเรียนรู้และการพัฒนาศักยภาพของผู้เรียนในระยะยาว

การศึกษาเกี่ยวกับการประยุกต์ใช้เทคนิคการขุดข้อมูลในบริบททางการศึกษา เรียกว่า การขุดข้อมูลทางการศึกษา โดยมีเป้าหมายหลักในการค้นหาวิธีการแก้ปัญหาในการวิจัยทางการศึกษาเพื่อระบุคำตอบที่เหมาะสม (Conijn et al., 2016) ผลการค้นพบนี้จะช่วยนักเรียนในการพัฒนาผลลัพธ์ทางการเรียนรู้ ผ่านการให้คำแนะนำที่เหมาะสม การปรับปรุงประสิทธิภาพการสอน การประเมินผลสัมฤทธิ์ทางการเรียน และการจัดระเบียบข้อเสนอทางการศึกษา รวมถึงทรัพยากรของโรงเรียน (Romero & Ventura, 2017) ในช่วงไม่กี่ปีที่ผ่านมา นักวิชาการได้ให้ความสนใจกับการใช้เทคนิคการขุดข้อมูลทางการศึกษา การค้นหาความเชื่อมโยงที่ซ่อนอยู่ในข้อมูลทางการศึกษา และการพยากรณ์ผลลัพธ์ทางการเรียนรู้ของนักเรียน (Kostopoulos et al., 2019) และยุคที่เทคโนโลยีสารสนเทศและการวิเคราะห์ข้อมูลมีบทบาทสำคัญในการศึกษา การพยากรณ์ผลการเรียนของนักเรียนจึงเป็นเครื่องมือที่มีประโยชน์ในการวางแผนการเรียนรู้และการให้คำปรึกษาแก่นักเรียน การนำเทคนิคการขุดข้อมูลมาใช้ในการวิเคราะห์ข้อมูลนักเรียนช่วยให้พยากรณ์ผลการเรียนได้อย่างแม่นยำและมีประสิทธิภาพ โดยเทคนิคที่ได้รับความนิยม คือ ต้นไม้ตัดสินใจ (Decision

Tree) ซึ่งมีความสามารถในการจำแนกข้อมูลและแสดงผลในรูปแบบที่เข้าใจง่าย สามารถระบุปัจจัยที่มีผลต่อผลการเรียนของนักเรียนได้อย่างชัดเจน (Varade & Thankanchan, 2021) และเป็นแบบจำลองที่สร้างจากชุดของกฎการจัดประเภท ซึ่งต้นไม้ตัดสินใจถูกนำมาใช้ในการศึกษาเพื่อวิเคราะห์และประเมินข้อมูลที่มีขนาดใหญ่โดยใช้วิธีการขุดข้อมูลทางการศึกษา (Martínez-Abad et al., 2020)

การพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนักเรียนมีความสำคัญอย่างยิ่งต่อการทำความเข้าใจอัตราความก้าวหน้าของนักเรียน ดังคำกล่าวที่ว่า “กันไว้ดีกว่าแก้” ซึ่งครูสามารถช่วยเหลือและให้คำแนะนำที่ถูกต้องแก่นักเรียน (Hasan et al., 2019) ในการวิเคราะห์ชุดข้อมูล ได้มีการนำเครื่องมือ Altair AI Studio ซึ่งเป็นซอฟต์แวร์โอเพนซอร์ส (Open Source) มาใช้ โดยเครื่องมือนี้รองรับการวิเคราะห์ข้อมูลเชิงพยากรณ์และการขุดข้อมูล (Data Mining) สามารถประมวลผลและแสดงผลลัพธ์ได้อย่างรวดเร็ว ช่วยให้ผู้ใช้สามารถพัฒนาแบบจำลองตามวัตถุประสงค์ได้อย่างมีประสิทธิภาพ (Abdullahi et al., 2023) และเป็นเครื่องมือในการวิเคราะห์ข้อมูลที่มีอินเทอร์เฟซ (Interface) แบบลากและวาง (Drag and Drop) ซึ่งเหมาะสำหรับผู้ที่ไม่มีความรู้พื้นฐานในการเขียนโปรแกรม โดยสามารถนำเทคนิคต้นไม้ตัดสินใจมาใช้ในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนักเรียนได้อย่างมีประสิทธิภาพ เครื่องมือ Altair AI Studio ได้รับการยกย่องว่าเป็นหนึ่งในบริการแพลตฟอร์มและระบบวิเคราะห์ขั้นสูงที่มีประสิทธิภาพมากที่สุดในสาขาปัญญาประดิษฐ์ ช่วยในการสร้างแบบจำลองการพยากรณ์ที่มีประสิทธิภาพสูงสุด สามารถขยายและเปิดให้ผสานรวมกับข้อมูลแอปพลิเคชัน และภาษาโปรแกรม เช่น R และ Python (Bjaoui et al., 2020)

เนื่องจากอัลกอริทึมต้นไม้ตัดสินใจสามารถจำแนกข้อมูลและแสดงผลในรูปแบบที่เข้าใจง่าย มีความยืดหยุ่นในการประมวลผลข้อมูลทั้งเชิงปริมาณและเชิงคุณภาพ อีกทั้งยังสามารถระบุปัจจัยสำคัญที่ส่งผลต่อผลการเรียนได้อย่างเป็นระบบ การวิจัยครั้งนี้จึงมุ่งศึกษาในบริบทของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยประยุกต์ใช้เทคนิคการขุดข้อมูลทางการศึกษา ผ่านอัลกอริทึมต้นไม้ตัดสินใจ เพื่อพยากรณ์ผลสัมฤทธิ์ทางการเรียน และสนับสนุนการตัดสินใจของครูและผู้บริหารสถานศึกษาได้อย่างมีประสิทธิภาพ

วัตถุประสงค์

1. เพื่อนำเสนอความหมายและแนวคิดที่เกี่ยวข้องกับการขุดข้อมูลทางการศึกษา ต้นไม้ตัดสินใจ และโปรแกรม Altair AI Studio v.2025.0.0
2. เพื่อแสดงแนวทางการพัฒนาแบบจำลองด้วยอัลกอริทึมต้นไม้ตัดสินใจ โดยใช้โปรแกรม Altair AI Studio v.2025.0.0 สำหรับการพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนักเรียน

ขอบเขตของบทความ

รวบรวมข้อมูลของนักเรียนระดับชั้นประถมศึกษาปีที่ 6 จากระบบสารสนเทศของโรงเรียน แต่ข้อมูลบางส่วนไม่สมบูรณ์หรือขาดคุณลักษณะที่จำเป็นต่อการวิเคราะห์ การศึกษาครั้งนี้ผู้วิจัยได้ใช้เครื่องมือปัญญาประดิษฐ์ ChatGPT เพื่อสร้างข้อมูลจำลอง 1,000 ชุด โดยอ้างอิงโครงสร้างและรูปแบบจากข้อมูลจริงที่ได้จากระบบสารสนเทศของสถานศึกษา ทั้งนี้ ผู้วิจัยได้พิจารณาประเด็นด้านจริยธรรมอย่างรอบคอบ โดยมีได้นำข้อมูลส่วนบุคคลหรือข้อมูลระบุตัวบุคคลใดมาใช้โดยตรง อีกทั้งได้ตรวจสอบความสมเหตุสมผล ความสอดคล้อง และความเป็นไปได้ของข้อมูลจำลอง เพื่อให้มั่นใจว่าข้อมูลที่สร้างขึ้นมีลักษณะใกล้เคียงกับข้อมูลจริงในเชิงสถิติและสามารถนำไปใช้ในการสร้างแบบจำลองได้โดยไม่ละเมิดสิทธิส่วนบุคคลหรือก่อให้เกิดความเสียหายต่อผู้ใด อันเป็นไปตามหลักจริยธรรมการวิจัยและข้อกำหนดด้านการคุ้มครองข้อมูลส่วนบุคคล

แนวคิดที่เกี่ยวข้อง

1. แบบจำลองการขุดข้อมูลทางการศึกษา (Educational Data Mining Model)

เทคนิคการขุดข้อมูล (Data Mining) คือ เทคนิคการขุดข้อมูลจากข้อมูลจำนวนมากมหาศาล แล้วนำมาวิเคราะห์เพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ ซึ่งอาจเป็นประโยชน์ในการตัดสินใจ (Varade & Thankanchan, 2021) การขุดข้อมูลทางการศึกษา คือ กระบวนการรวบรวมและวิเคราะห์ข้อมูลจากระบบการศึกษา (Chen & Lin, 2023) โดยประยุกต์ใช้เทคนิคการขุดข้อมูลกับการขุดกฎความสัมพันธ์ (Association Rule Mining) การจัดกลุ่ม (Clustering) และการจำแนกประเภท (Classification) ร่วมกับเครื่องมือวิเคราะห์ข้อมูล เช่น Weka, Orange, SPSS และ Altair AI Studio เพื่อพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนักเรียนที่เรียนรู้ซ้ำ และพยากรณ์อัตราการลาออกของนักเรียน โดยมีเทคนิคที่ใช้ในการจำแนกข้อมูล ดังนี้ (Dol & Jawandhiya, 2024)

1.1 การขุดกฎความสัมพันธ์ (Association Rule Mining) ใช้เพื่อกำหนดการพึ่งพาของรายการข้อมูลหนึ่งต่อรายการข้อมูลหนึ่ง เช่น ความน่าจะเป็นที่บุคคลจะซื้อกางเกงยีนส์หากเขาซื้อเสื้อยืด ตัวอย่างของอัลกอริทึมกฎความสัมพันธ์ เช่น เทอร์ติอุส (Tertius), อาปริโอรี (Apriori) และการเจริญเติบโตที่ผ่านการกรอง (Filtered Growth)

1.2 การจำแนกประเภท (Classification) แบ่งชุดข้อมูลออกเป็น 1) ชุดข้อมูลฝึกสอน ใช้เพื่อจำแนกข้อมูลลงในคลาส และ 2) ชุดข้อมูลทดสอบ ใช้เพื่อพยากรณ์คลาส โดยเทคโนโลยีนี้ถูกนำไปใช้ในงานต่าง ๆ เช่น การจัดหมวดหมู่เอกสาร การจำแนกอีเมลสแปมกับอีเมลที่ไม่ใช่สแปม โดยเทคนิคการจัดหมวดหมู่ เช่น นาอิวเบย์ (Naïve Bayes), ป่าสุ่ม (Random Forest), เครื่องเวกเตอร์สนับสนุน (Support Vector Machine) และต้นไม้ตัดสินใจ (Decision Tree)

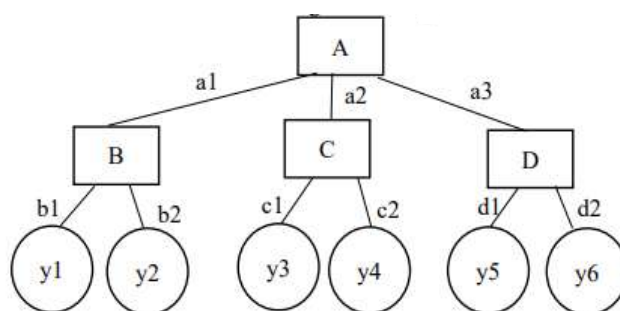
1.3 การจัดกลุ่ม (Clustering) จะแยกชุดข้อมูลออกเป็นกลุ่ม โดยที่ข้อมูลในกลุ่มหนึ่ง มีลักษณะร่วมกับข้อมูลในกลุ่มเดียวกันและมีลักษณะเฉพาะที่แตกต่างจากข้อมูลในกลุ่มอื่น เช่น เทคนิคการแบ่งกลุ่ม (Partitioning Techniques) การจัดกลุ่มแบบลำดับขั้น (Hierarchical Clustering) การจัดกลุ่มแบบมีข้อจำกัด (Constraint Based Clustering) การจัดกลุ่มแบบอิงการกระจาย (Distribution Based Clustering) การจัดกลุ่มแบบอิงความหนาแน่น (Density Based Clustering) และการจัดกลุ่มแบบฟัซซี (Fuzzy Clustering)

เป้าหมายของการขุดข้อมูลทางการศึกษา คือ การสร้างข้อมูลเชิงลึกใหม่จากสภาพแวดล้อมการเรียนรู้ เพื่อประโยชน์ของครู นักเรียน และผู้มีส่วนได้ส่วนเสีย (Toivonen & Jormanainen, 2019) ในการพัฒนาคุณภาพการเรียนรู้ และค้นหาความสัมพันธ์ระหว่างผลสัมฤทธิ์ทางการเรียนกับพฤติกรรม การเรียนรู้ วัตถุประสงค์การสอน และวิธีการจัดการเรียนการสอน (Dutt et al., 2017) โดยมีการศึกษาแนวทางการเรียนรู้ของเครื่องจักรสำหรับการติดตามและพยากรณ์ผลการเรียนของนักศึกษา เพื่อทำนายผลการเรียนเฉลี่ยสะสม (GPA) สุดท้ายของนักศึกษา โดยพิจารณาจากสถานะผลการเรียนและการพยากรณ์วิชาที่ยังไม่ได้เรียน เพื่อให้สามารถพยากรณ์ผลการเรียนแบบก้าวหน้าได้ (Xu et al., 2017) และได้ศึกษาอัลกอริทึมการถดถอยแบบกึ่งกำกับดูแลสำหรับการพยากรณ์ผลการเรียนของนักศึกษาในหลักสูตรการเรียนทางไกล เพื่อทำนายผลการเรียนสุดท้าย และระบุความเสี่ยงของนักศึกษาที่จะสอบตกหรือพยากรณ์ผลสัมฤทธิ์ทางการเรียน (Kostopoulos et al., 2019) การขุดข้อมูลทางการศึกษา เป็นเครื่องมือที่ช่วยวิเคราะห์พฤติกรรมและแนวโน้มการเรียนรู้ของนักเรียน เพื่อยกระดับประสิทธิภาพการจัดการเรียนการสอน อย่างไรก็ตาม ยังมีข้อจำกัดด้านความเป็นส่วนตัวของข้อมูลและความแม่นยำของแบบจำลอง อีกทั้งการนำไปใช้ควรคำนึงถึงบริบททางการศึกษาในแต่ละพื้นที่ (Joy et al., 2019; Romero & Ventura, 2020)

แม้ว่าการชุดข้อมูลทางการศึกษา จะมีศักยภาพสูงในการยกระดับการเรียนรู้ แต่ผู้ประพันธ์มองว่าการประยุกต์ใช้จริงยังเผชิญความท้าทายสำคัญ ประการแรกคือ คุณภาพและความสมบูรณ์ของข้อมูล ซึ่งเป็นปัจจัยพื้นฐานต่อความน่าเชื่อถือของแบบจำลอง ประการที่สอง การตีความผลลัพธ์เชิงลึก จำเป็นต้องอาศัยผู้เชี่ยวชาญที่มีความเข้าใจทั้งด้านเทคนิคการชุดข้อมูลทางการศึกษา และหลักการทางการศึกษา เพื่อแปลงการวิเคราะห์ทางสถิติไปสู่ข้อเสนอแนะเชิงปฏิบัติที่เหมาะสมกับบริบทการเรียนรู้ที่หลากหลายได้อย่างแท้จริง

2. อัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree Algorithm)

อัลกอริทึมต้นไม้ตัดสินใจเป็นวิธีการเรียนรู้ของเครื่อง (Machine Learning) สำหรับการสร้างแบบจำลองในการพยากรณ์ โดยสร้างแบบจำลองคล้ายต้นไม้ เพื่อจับคู่ข้อมูลนำเข้ากับตัวแปรเป้าหมาย วิธีการต้นไม้ตัดสินใจจะแบ่งข้อมูลซ้ำ ๆ ตามคุณลักษณะที่ทำให้การแบ่งที่ดีที่สุดหรือการได้รับข้อมูลมากที่สุด โดยใช้วิธีการแบ่งแยกและเอาชนะ (Divide and Conquer Approach) วิธีการนี้ถูกสร้างขึ้นจากบนลงล่าง ดังแสดงในภาพที่ 1 โดยแต่ละโหนดใบ ($y_1, y_2, \dots, y_6$) แสดงถึงการพยากรณ์ และแต่ละโหนดภายใน (โหนด B, C และ D) แสดงถึงตัวเลือกตามคุณลักษณะ ชุดฝึกอบรมทั้งหมดของตัวอย่างจะรวมอยู่ในโหนดบนสุด A ซึ่งบางครั้งเรียกว่า โหนดราก



ภาพที่ 1 แบบจำลองคล้ายต้นไม้ในอัลกอริทึมต้นไม้ตัดสินใจ
ที่มา: Chen and Lin (2023)

อัลกอริทึมจะเพิ่มค่าการได้รับข้อมูล (Information Gain) ดัชนีจินี (Gini Index) หรือเอนโทรปี (Entropy) ให้สูงสุดที่แต่ละโหนดภายใน เพื่อกำหนดคุณลักษณะการแบ่งข้อมูลออกเป็นเซตย่อยที่บริสุทธิ์ (Purest Subsets) ที่สุด เซตย่อยที่มีป้ายกำกับคลาสเดียวกันถือว่าบริสุทธิ์ที่สุด ซึ่งช่วยอำนวยความสะดวกในการพยากรณ์ที่แม่นยำ อัลกอริทึมต้นไม้ตัดสินใจ จะทำการแบ่งข้อมูลต่อไปจนกว่าโหนดใบทั้งหมดจะเป็นข้อมูลบริสุทธิ์หรือตรงตามเงื่อนไขการหยุดที่กำหนดไว้ เมื่อสร้างต้นไม้เสร็จสมบูรณ์แล้ว สามารถใช้เพื่อพยากรณ์ข้อมูลใหม่ได้ โดยการไล่ตามเส้นทางจากโหนดรากไปยังโหนดใบตามค่าคุณลักษณะของข้อมูลนำเข้า และคืนค่าการพยากรณ์ซึ่งก็คือป้ายกำกับคลาสที่เก็บไว้ในโหนดใบนั้น อัลกอริทึมการจำแนกประเภทต้นไม้แบบดั้งเดิม ได้แก่ ID3, C4.5 และ CART (Chen & Lin, 2023) โดยมีการศึกษาการพยากรณ์ผลการเรียนของนักศึกษาระดับปริญญาตรีโดยใช้อัลกอริทึมต้นไม้ตัดสินใจที่แสดงให้เห็นว่าอัลกอริทึมต้นไม้ตัดสินใจสามารถใช้พยากรณ์ผลการเรียนของนักศึกษาชั้นปีที่ 1 และ 2 ได้อย่างแม่นยำ ครูผู้สอนสามารถสร้างแบบฝึกหัดเพื่อเสริมสร้างความสามารถในการรับรู้ของนักศึกษาในกลุ่มอ่อนและให้ความสนใจเป็นพิเศษเพื่อพัฒนาผลสัมฤทธิ์ทางการเรียนของนักศึกษา (Varade & Thankanchan, 2021) และการพัฒนาตัวแบบทำนายทักษะการคิดเชิงคำนวณของนักเรียนชั้นมัธยมศึกษาตอนต้น โรงเรียนแห่งหนึ่งในจังหวัดสุราษฎร์ธานี โดยใช้เทคนิคแมชชีนเลิร์นนิง เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบทำนายทักษะการคิดเชิงคำนวณ

ผลการพัฒนาตัวแบบทำนายทักษะการคิดเชิงคำนวณพบว่า ตัวแบบทำนายที่พัฒนาจากเทคนิคนาอ์ฟเบย์ เทคนิคต้นไม้ตัดสินใจ เทคนิคเพื่อนบ้านใกล้สุด มีค่าความถูกต้อง เท่ากับ ร้อยละ 60.05 ร้อยละ 74.95 และ ร้อยละ 68.48 ตามลำดับ ผลการเปรียบเทียบประสิทธิภาพของตัวแบบทำนายทักษะการคิดเชิงคำนวณพบว่า ตัวแบบทำนายที่พัฒนาจากเทคนิคต้นไม้ตัดสินใจสามารถทำนายผลระดับทักษะการคิดเชิงคำนวณได้อย่างมีประสิทธิภาพสูงที่สุดจาก 3 ตัวแบบ ต้นไม้ตัดสินใจ เป็นเทคนิคที่ได้รับความนิยมเนื่องจากเข้าใจง่าย และสามารถอธิบายเหตุผลในการตัดสินใจได้อย่างชัดเจน อย่างไรก็ตาม เทคนิคนี้มีความเสี่ยงต่อการเรียนรู้มากเกินไป (Overfitting) โดยเฉพาะในกรณีที่ข้อมูลมีความซับซ้อน ทั้งนี้สามารถนำไปประยุกต์ใช้ในการพยากรณ์ผลสัมฤทธิ์ทางการเรียน การจำแนกกลุ่มนักเรียน และการพัฒนาระบบแนะนำการเรียนรู้เฉพาะบุคคล (Hu, 2022) ข้อได้เปรียบสำคัญของเทคนิคต้นไม้ตัดสินใจ คือ ความสามารถในการแสดงผลลัพธ์ในรูปแบบที่เข้าใจง่าย โดยผลลัพธ์ของการวิเคราะห์จะปรากฏในลักษณะของผังต้นไม้ที่ประกอบด้วยโหนด (Nodes) และทางเลือก (Branches) ทำให้ครู ผู้บริหาร หรือผู้ไม่มีพื้นฐานทางเทคนิคสามารถตีความผลได้โดยไม่ต้องอาศัยความรู้ด้านคณิตศาสตร์ขั้นสูง นอกจากนี้ เทคนิคต้นไม้ตัดสินใจยังสามารถระบุปัจจัยที่มีอิทธิพลต่อผลลัพธ์ได้อย่างชัดเจน เช่น พฤติกรรมการเรียนที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียน ซึ่งเป็นข้อมูลที่สามารถนำไปใช้ในการวางแผนการสอนหรือให้คำปรึกษากับนักเรียนได้ทันที อีกข้อได้เปรียบหนึ่งที่สำคัญ คือ ความยืดหยุ่นในการประมวลผลข้อมูลทั้งเชิงปริมาณและเชิงคุณภาพ โดยไม่จำเป็นต้องมีการแปลงข้อมูลให้เป็นมาตรฐานเดียวกัน ซึ่งแตกต่างจากเทคนิคอื่น เช่น เครื่องจักรเวกเตอร์สนับสนุน (Support Vector Machine) หรือ โครงข่ายประสาทเทียม (Neural Networks) ที่มักต้องการการปรับขนาดข้อมูล การจัดรูปแบบ หรือการตั้งค่าพารามิเตอร์ที่ซับซ้อน และยากต่อการตรวจสอบย้อนกลับ นอกจากนี้ เทคนิคต้นไม้ตัดสินใจยังไม่ทำหน้าที่เป็นกล่องดำ (Black Box) เหมือนกับโมเดลเชิงลึก (Deep Learning) ทำให้สามารถตรวจสอบเหตุผลของการตัดสินใจของแบบจำลองได้ ซึ่งเหมาะกับบริบทการศึกษาที่ต้องการความโปร่งใสและสามารถอธิบายได้

แม้ว่าอัลกอริทึมต้นไม้ตัดสินใจจะเข้าใจง่ายและเป็นที่นิยมในด้านการศึกษา แต่ผู้ประพันธ์เห็นว่าการนำไปใช้ยังมีประสิทธิภาพยังมีข้อจำกัดสำคัญ โดยเฉพาะความเสี่ยงต่อการเกิด Overfitting เมื่อข้อมูลมีขนาดเล็ก ซับซ้อน หรือหลากหลายเกินไป ทำให้แบบจำลองจับสัญญาณรบกวนแทนที่จะจับรูปแบบแท้จริง ส่งผลให้ประสิทธิภาพการพยากรณ์ข้อมูลใหม่ลดลง การสร้างต้นไม้ที่ซับซ้อนเกินไปยังทำให้ตีความยากและขาดความสามารถในการสรุปผล แม้จะมีเทคนิคอย่างการตัดกิ่ง (Pruning) แต่ก็ยังต้องระมัดระวังในการเลือกพารามิเตอร์ เพื่อรักษาสมดุลระหว่างความแม่นยำและการอธิบายผลอย่างมีเหตุผล

3. การพยากรณ์ผลสัมฤทธิ์ทางการเรียน (Academic Performance Prediction)

การพยากรณ์ผลสัมฤทธิ์ทางการเรียน เป็นแนวทางหนึ่งในด้านการวิจัยทางการศึกษา ซึ่งมุ่งวิเคราะห์และคาดการณ์ผลการเรียนของนักเรียนโดยอาศัยข้อมูลจากพฤติกรรมการเรียนรู้ เช่น คะแนนสอบ การเข้าเรียน หรือข้อมูลด้านประชากรศาสตร์ ด้วยเทคนิคการขุดข้อมูลทางการศึกษา (Educational Data Mining) และอัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning) เช่น Decision Tree, Random Forest และ Support Vector Machine (Al-Barrak & Al-Razgan, 2015; Sankar et al., 2020) จุดมุ่งหมายของการพยากรณ์คือ เพื่อระบุผู้เรียนที่มีความเสี่ยงต่อการมีผลสัมฤทธิ์ต่ำ และเพื่อช่วยในการออกแบบแผนการเรียนรู้เฉพาะบุคคล (Salam et al., 2022) การพยากรณ์ยังช่วยส่งเสริมการตัดสินใจเชิงนโยบายทางการศึกษาให้แม่นยำและมีประสิทธิภาพยิ่งขึ้น

แม้การพยากรณ์ผลสัมฤทธิ์ทางการเรียนจะมีศักยภาพสูงในการช่วยนักเรียนและกำหนดนโยบาย แต่ผู้ประพันธ์เห็นว่าการนำไปใช้จริงยังมีข้อจำกัดสำคัญด้านคุณภาพและความสมบูรณ์ของข้อมูล ซึ่ง

ส่งผลต่อความแม่นยำของการพยากรณ์ นอกจากนี้ ควรพิจารณาว่าแบบจำลองเป็นเพียงเครื่องมือสนับสนุน ไม่ใช่การตัดสินใจเด็ดขาด ควรใช้ควบคู่กับการพิจารณาปัจจัยเชิงคุณภาพอื่น ๆ และบริบทส่วนตัวของนักเรียน เพื่อให้การแทรกแซงทางการศึกษาเกิดประโยชน์สูงสุดอย่างแท้จริง

4. โปรแกรม Altair AI Studio v.2025.0.0 (เดิมชื่อ RapidMiner)

Altair AI Studio เป็นแพลตฟอร์มซอฟต์แวร์ที่ช่วยในการสร้างแบบจำลองเชิงพยากรณ์ อัลกอริทึมการเรียนรู้ของเครื่อง การวิเคราะห์ข้อมูล และการเตรียมข้อมูลแบบครบวงจร โดย RapidMiner มี 3 รุ่น ได้แก่ 1) Altair AI Studio เป็นโปรแกรมแบบสแตนด์อโลน (Stand Alone) ใช้งานได้แบบเดี่ยวไม่ต้องเชื่อมต่อกับเซิร์ฟเวอร์หรือระบบเครือข่ายอื่น 2) RapidMiner Server เป็นสถาปัตยกรรมแบบไคลเอนต์เซิร์ฟเวอร์ (Client Server Architecture) แบ่งบทบาทระหว่างเครื่องผู้ใช้ (Client) และเครื่องแม่ข่าย (Server) ที่ให้บริการทรัพยากร การประมวลผล และการจัดการข้อมูล และ 3) RapidMiner Hadoop เป็นแพลตฟอร์มสำหรับการจัดเก็บและประมวลผลข้อมูลขนาดใหญ่ในลักษณะกระจาย (Distributed Computing) ประกอบด้วยสภาพแวดล้อมการทำงานแบบเห็นภาพที่สร้างขึ้นบนเวิร์กโฟลว์ (Workflow Based) ช่วยให้ผู้ใช้เข้าถึงข้อมูลทั้งแบบมีโครงสร้างและไม่มีโครงสร้าง เช่น ไฟล์มัลติมีเดีย ข้อความ และรูปภาพ รองรับรูปแบบไฟล์ CSV (Comma Separated Values), สเปรดชีตเอกเซล (Excel Spreadsheet), SAS, Stata10, รูปแบบ WEKA และการเชื่อมต่อกับฐานข้อมูล เช่น Access, SQL Server, Oracle, MySQL และ Postgres นอกจากนี้ยังสามารถเข้าถึงฐานข้อมูล NoSQL เช่น Cassandra และ MongoDB ได้ Altair AI Studio สามารถใช้เพื่อแก้ปัญหาการจัดประเภท การถดถอย การจัดกลุ่ม ความสัมพันธ์ และนำเสนอเทคนิคการเตรียมข้อมูลที่หลากหลาย เช่น การจัดการค่าที่หายไป การเลือกคุณลักษณะ และการแปลงข้อมูล เทคนิคการเรียนรู้ของเครื่องมีให้เลือกใช้อย่างหลากหลายและเข้าถึงได้ง่าย เช่น โครงข่ายประสาทเทียม ต้นไม้ตัดสินใจ การจัดกลุ่ม และการถดถอย นอกจากนี้ยังสามารถรันสคริปต์ที่เขียนด้วยภาษา R และ Python (Campos & Campos, 2024) โดยเครื่องมือนี้มีคลังไลบรารีที่ครอบคลุมอัลกอริทึมและฟังก์ชันการเรียนรู้ของเครื่องมากกว่า 1,500 รายการ (Bjaoui et al., 2020)

Altair AI Studio ใช้สำหรับการสร้างแบบจำลองการพยากรณ์และการออกแบบแบบกราฟิกที่ไม่ต้องใช้โค้ดของ Altair AI Studio ทำให้ใช้งานง่าย การสร้างแบบจำลองการพยากรณ์ใช้วิธีการลากและวางแบบจำลองการถดถอยและการจำแนกประเภททุกรูปแบบสามารถเข้าถึงได้หลังจากตั้งค่าและนำเข้าข้อมูลไปยัง Altair AI Studio แล้วลากและวางแบบจำลองที่ต้องการลงในหน้าต่างการออกแบบ ส่วนผลลัพธ์จะแสดงในแบบจำลองสุดท้าย นอกจากนี้ยังช่วยอำนวยความสะดวกในการวิเคราะห์ทางสถิติของข้อมูล เพื่อประเมินความแม่นยำของแบบจำลอง นำเสนอการแสดงผลแบบกราฟิก และยังเป็นประโยชน์สำหรับผู้เริ่มต้น (Batool et al., 2023) อย่างไรก็ตาม การประมวลผลข้อมูลขนาดใหญ่อาจมีข้อจำกัดด้านประสิทธิภาพเมื่อเทียบกับแพลตฟอร์มที่พัฒนาโดยใช้โค้ดโดยตรง (Klein et al., 2021) ปัจจุบันการวิเคราะห์และประมวลผลแบบสำรวจเพื่อค้นหารูปแบบเป็นสิ่งสำคัญอย่างยิ่งในการแก้ปัญหา (Moreira et al., 2025) นอกจากความสามารถในการจัดเก็บข้อมูลขนาดใหญ่แล้ว เครื่องมือ Altair AI Studio ยังได้รับการยกย่องว่าเป็นหนึ่งในบริการแพลตฟอร์มและระบบวิเคราะห์ขั้นสูงที่มีประสิทธิภาพมากที่สุดในสาขาปัญญาประดิษฐ์ ธุรกิจ บริการทางการเงิน พลังงาน โลจิสติกส์ วิทยาศาสตร์ชีวภาพ และการดูแลสุขภาพ (Bjaoui et al., 2020)

แม้ว่า Altair AI Studio จะช่วยให้การสร้างแบบจำลองและการวิเคราะห์ข้อมูลทำได้ง่ายโดยไม่ต้องเขียนโค้ด แต่ผู้ประพันธ์มองว่ายังมีข้อจำกัดด้านประสิทธิภาพในการประมวลผล ข้อมูลขนาดใหญ่ และความยืดหยุ่นในการปรับแต่งอัลกอริทึมที่ซับซ้อน ซึ่งอาจไม่ตอบโจทย์งานวิจัยเฉพาะทางอย่างเต็มที่

การพึ่งพาแพลตฟอร์มสำเร็จรูปอาจทำให้ผู้ใช้ขาดความเข้าใจเชิงลึกในกลไกเบื้องหลัง ดังนั้น การเลือกใช้เครื่องมือนี้ควรพิจารณาจากขนาดข้อมูลและความซับซ้อนของปัญหา เพื่อให้ได้ผลลัพธ์ที่เหมาะสมที่สุด

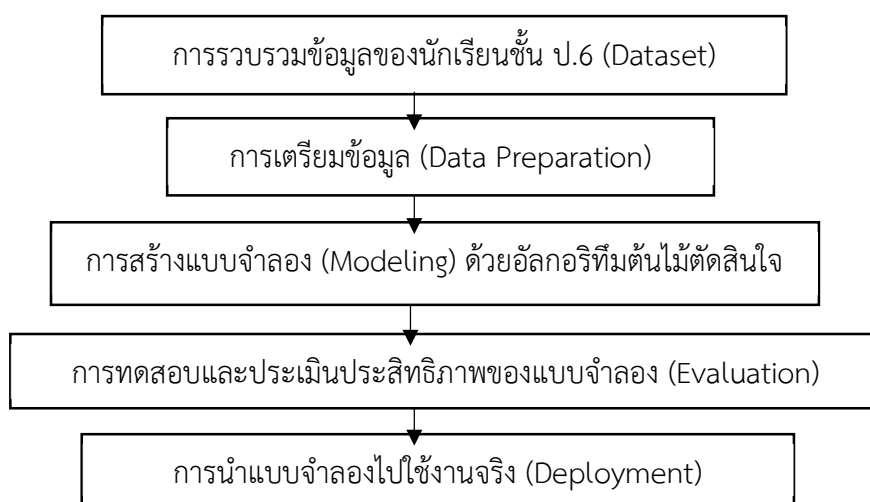
5. แอปพลิเคชัน One Compiler

แอปพลิเคชัน One Compiler เป็นแพลตฟอร์มรันโค้ดออนไลน์ที่มีความสามารถในการรองรับภาษาโปรแกรมหลากหลายชนิด เช่น Python, Java, C++, JavaScript และภาษาอื่น ๆ อีกมากกว่า 70 ภาษา โดยไม่จำเป็นต้องติดตั้งซอฟต์แวร์เพิ่มเติมในเครื่องผู้ใช้ ซึ่งเป็นการอำนวยความสะดวกสำหรับทั้งผู้เรียนและผู้สอนในบริบทของการเรียนรู้ด้านการเขียนโปรแกรมและโครงสร้างคอมไพเลอร์ ด้วยความสามารถในการให้ผลลัพธ์แบบเรียลไทม์ ผู้ใช้สามารถทดสอบ แก้ไข และปรับปรุงโค้ดของตนได้อย่างต่อเนื่อง นอกจากนี้ One Compiler ยังมีการออกแบบระบบให้สามารถฝัง (Embed) เข้าในเว็บไซต์หรือระบบการเรียนรู้ขององค์กรผ่าน API ได้ ซึ่งเอื้อให้เกิดการพัฒนาแอปพลิเคชันด้านการศึกษาหรือการทดสอบเชิงโต้ตอบ (Interactive Learning) ได้อย่างมีประสิทธิภาพ อีกทั้งแอปพลิเคชัน One Compiler ยังให้ความสำคัญกับความปลอดภัยในการรันโค้ดผ่านกลไก Sandbox ที่จำกัดสิทธิ์ในการเข้าถึงทรัพยากรของระบบ ซึ่งเป็นแนวทางปฏิบัติที่พบบ่อยในงานวิจัยด้านความปลอดภัยของคอมไพเลอร์ออนไลน์

แม้ว่าแอปพลิเคชัน One Compiler จะเป็นเครื่องมือที่มีประโยชน์ในการเรียนรู้การเขียนโปรแกรม เนื่องจากใช้งานง่ายและรองรับหลากหลายภาษา แต่ผู้ประพันธ์มองว่ายังมีข้อจำกัดที่ควรพิจารณาการประมวลผลโค้ดที่ซับซ้อนหรือต้องการทรัพยากรสูงอาจมีข้อจำกัดด้านประสิทธิภาพ การเลือกใช้ควรคำนึงถึงความซับซ้อนของงานและความต้องการเฉพาะด้าน

ขั้นตอนการพัฒนาแบบจำลองอัลกอริทึมต้นไม้ตัดสินใจ

การศึกษานี้ได้เสนอการพยากรณ์ผลสัมฤทธิ์ทางการเรียนโดยใช้อัลกอริทึมต้นไม้ตัดสินใจของนักเรียนชั้นประถมศึกษาปีที่ 6 ประกอบด้วย 5 ขั้นตอน (ชัยมงคล ปิ่นสา และคณะ, 2567) ดังนี้ 1) การรวบรวมข้อมูล (Dataset) 2) การเตรียมข้อมูล (Data Preparation) 3) การสร้างแบบจำลอง (Modeling) ด้วยอัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree Algorithm) 4) การทดสอบและประเมินประสิทธิภาพของแบบจำลอง (Evaluation) และ 5) การนำแบบจำลองไปใช้งานจริง (Deployment) แสดงในภาพที่ 2



ภาพที่ 2 ขั้นตอนการพัฒนาแบบจำลองอัลกอริทึมต้นไม้ตัดสินใจ

1. ขั้นการรวบรวมข้อมูล (Dataset)

นำข้อมูลที่ถูกรวบรวมจากระบบสารสนเทศของโรงเรียนแล้วใช้ ChatGPT ช่วยจำลองข้อมูล แม้ผู้เขียนบทความจะมีข้อมูลเบื้องต้นจากระบบสารสนเทศของโรงเรียน แต่ข้อมูลบางส่วนไม่สมบูรณ์หรือขาดคุณลักษณะที่จำเป็นต่อการวิเคราะห์ อาทิ รายวิชา หน่วยกิต และสถานะการเรียน ผู้เขียนบทความจึงเลือกใช้เครื่องมือ ChatGPT เพื่อจำลองข้อมูลจำนวน 1,000 ชุด เพื่อใช้ในการฝึกสอนและทดสอบแบบจำลองเชิงเทคนิค โดยยึดรูปแบบและโครงสร้างของข้อมูลจริง ทั้งนี้ได้คำนึงถึงจริยธรรมการวิจัยและความเหมาะสมด้านความน่าเชื่อถือของข้อมูลจำลองอย่างรอบคอบ โดยเขียน Prompt “ช่วยจำลองข้อมูลที่ประกอบด้วย ชื่อ-นามสกุล เลขประจำตัว เพศเกรดเฉลี่ยสะสม จำนวนชั่วโมงในการเรียน และสถานะทางการเรียน จากไฟล์ที่แนบ จำนวน 1,000 ชุดข้อมูล” ซึ่งข้อมูลจะได้รับการตรวจสอบความถูกต้องและแปลงให้อยู่ในรูปแบบที่เหมาะสมต่อการนำไปวิเคราะห์ด้วยเทคนิคอัลกอริทึมต้นไม้ตัดสินใจ โดย GPA คือคะแนนเฉลี่ยสะสมที่นักเรียนได้รับในแต่ละภาคเรียน ถ้ามีค่า 0 เป็นคะแนนต่ำสุด และค่า 4 เป็นคะแนนสูงสุด (Wiyono et al., 2019) และ SKS คือ หน่วยกิตภาคเรียนที่นักเรียนได้ลงทะเบียนเรียน (Novianto & Suhirman, 2024) โดยชื่อและนามสกุลเก็บรวมกันใน Attribute เดียวข้อมูลที่มีอยู่ได้รับมาในรูปแบบรวมผู้วิจัยได้ดำเนินการแยก (Parsing) Attribute ในขั้นตอนการเตรียมข้อมูล (Data Preprocessing) ก่อนนำไปวิเคราะห์หรือสร้างโมเดล ข้อมูลนักเรียนและคุณลักษณะข้อมูลต่อไปนี้อาจดูได้ในภาพที่ 3

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1	First Name	Last Name	Gender	GPA	Seme	GPA	Seme	GPA	Seme	GPA	Seme	Credits	Ser	Credits	Ser	Credits	Ser	Final GPA	Status
2	Nebsr	Wngschali	Male	3.77	3.52	3.56	3.34	3.34	3.66	24	24	24	24	22	23	3.531667	1		
3	Echsd	Aekwdng	Male	3.93	4	3.23	3.61	4	3.82	24	24	24	24	22	23	3.765	1		
4	Chnakant	Oankrothk	Female	4	4	4	4	4	4	24	24	24	24	22	23	4	1		
5	Pphinthphy	Rachphath	Female	4	4	4	4	4	4	24	24	24	24	22	23	4	1		
6	Nichda	Swinbuy	Female	4	4	4	4	4	4	24	24	24	24	22	23	4	1		
7	Phathchrpt	Sukhng	Male	3.34	3.8	3.52	2.73	3.28	3.75	23	23	23	23	23	24	3.4	0		
8	Phirwathn	Aessim	Male	2.97	3.32	3	2.39	3.32	3.54	23	23	23	23	23	24	3.09	0		
9	Tharwa	Wangkhiri	Male	3.78	3.89	3.73	3.5	3.78	4	23	23	23	23	23	24	3.78	1		
10	Phkhnanth	Esrmsthar	Female	3	3.32	3.17	2.06	3.04	3.6	23	23	23	23	23	24	3.03	0		
11	Siriny	Enthibutr	Female	3.65	3.97	3.93	3.26	3.86	4	23	23	23	23	23	24	3.78	1		
12	Charupich	Sismbat	Female	3.97	3.79	3.58	3.76	2.88	3.61	24	24	24	23	22	22	3.598333	1		
13	Suphnu	Bawbanbul	Female	3.93	4	3.75	3.82	3.63	3.93	24	24	24	23	22	22	3.843333	1		
14	Srayya	Yodathng	Female	3.85	3.89	4	3.8	3.22	3.79	24	24	24	23	22	22	3.758333	1		
15	Wrrya	Chanthrekl	Female	3.7	3.77	3.2	3.54	2.63	3.65	24	24	24	23	22	22	3.415	0		
16	Wrinth	Osthrmmi	Male	3.31	3.79	3.12	2.47	2.61	3.06	24	24	24	23	22	22	3.06	0		
17	Wsuphl	Aekwaesnt	Male	3.31	3.79	2.72	3.34	2.54	3.11	24	24	24	23	22	22	3.135	0		
18	Suphedch	Erueonga	Male	4	4	4	3.91	3.25	3.81	24	24	24	23	22	22	3.828333	1		
19	Nebsr	Wngschali	Male	3.77	3.52	3.56	3.34	3.34	3.66	24	24	24	24	22	23	3.531667	1		
20	Echsd	Aekwdng	Male	3.93	4	3.23	3.61	4	3.82	24	24	24	24	22	23	3.765	1		
21	Chnakant	Oankrothk	Female	4	4	4	4	4	4	24	24	24	24	22	23	4	1		
22	Pphinthphy	Rachphath	Female	4	4	4	4	4	4	24	24	24	24	22	23	4	1		
23	Nichda	Swinbuy	Female	4	4	4	4	4	4	24	24	24	24	22	23	4	1		
24	Phathchrpt	Sukhng	Male	3.34	3.8	3.52	2.73	3.28	3.75	23	23	23	23	23	24	3.4	0		
GPA DATA 1000 English																			

ภาพที่ 3 ชุดข้อมูลนักเรียน

2. ขั้นการเตรียมข้อมูล (Data Preparation)

ข้อมูลได้รับการเตรียมพร้อมก่อนการวิเคราะห์ ซึ่งเป็นขั้นตอนที่ข้อมูลถูกจัดเตรียมเพื่อใช้ในการทดสอบแบบจำลองอัลกอริทึม โดยข้อมูลที่รวบรวมได้จะผ่านกระบวนการทำความสะอาดข้อมูล (Data Cleaning) (Putri & Waspada, 2018) โดยการลบข้อมูลที่ซ้ำซ้อน กำจัดค่าที่หายไป (Missing Values) และตรวจสอบค่าผิดปกติ (Outliers) จากนั้นดำเนินการแปลงข้อมูล (Data Transformation) เพื่อเตรียมความพร้อมสำหรับการประมวลผลด้วยเทคนิคการขุดข้อมูล (Switrayana et al., 2023) ชุดข้อมูลที่ใช้ในกระบวนการขุดข้อมูลนี้มีคุณลักษณะ 16 รายการ รายละเอียดของคุณลักษณะที่ใช้สามารถดูได้ในตารางที่ 1

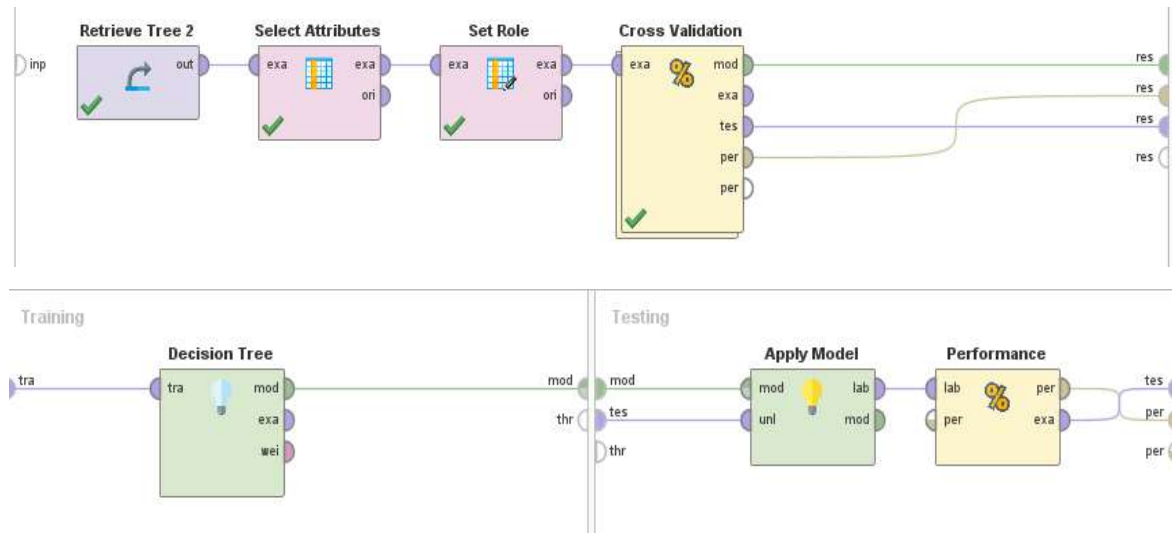
ตารางที่ 1 คุณลักษณะข้อมูล (Attributes) ที่ใช้ในการสร้างแบบจำลอง

ที่	ตัวแปร	ชนิดของข้อมูล	ที่	ตัวแปร	ชนิดของข้อมูล
1	ชื่อ-นามสกุล	Text	9	หน่วยกิต ป.1	integer
2	เพศ*	Polynomial	10	หน่วยกิต ป.2	integer
3	เกรดเฉลี่ยชั้น ป.1	Real	11	หน่วยกิต ป.3	integer
4	เกรดเฉลี่ยชั้น ป.2	Real	12	หน่วยกิต ป.4	integer
5	เกรดเฉลี่ยชั้น ป.3	Real	13	หน่วยกิต ป.5	integer
6	เกรดเฉลี่ยชั้น ป.4	Real	14	หน่วยกิต ป.6	integer
7	เกรดเฉลี่ยชั้น ป.5	Real	15	เกรดเฉลี่ยรวม	Real
8	เกรดเฉลี่ยชั้น ป.6	Real	16	สถานะทางการเรียน**	Polynomial

*ชาย-หญิง, **กำลังศึกษา-จบการศึกษา-ลาออก, Polynomial คือ ตัวแปรที่เป็นหมวดหมู่, Real คือ ตัวแปรที่มีค่าต่อเนื่อง หรือมีค่าจำนวนจริง, Integer คือ ตัวแปรที่เป็นตัวเลขจำนวนเต็ม

3. ขั้นตอนการสร้างแบบจำลอง (Modeling) ด้วยอัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree Algorithm)

ข้อมูลที่ผ่านการเตรียมแล้วจะถูกนำเข้าสู่อัลกอริทึม Altair AI Studio เพื่อเข้าสู่กระบวนการวิเคราะห์ โดยแบ่งข้อมูลออกเป็นชุดฝึก (Training Set) ร้อยละ 80 และชุดทดสอบ (Testing Set) ร้อยละ 20 จากนั้นใช้กระบวนการสร้างแบบจำลองด้วยอัลกอริทึมต้นไม้ตัดสินใจ โดยประเมินประสิทธิภาพด้วยเทคนิค 10 cross-validation ตั้งค่า Maximal Depth เท่ากับ 10 และจำนวนข้อมูลขั้นต่ำในแต่ละโหนด เพื่อควบคุมความซับซ้อนและป้องกันภาวะ Overfitting หลังจากนั้นทำการรันแบบจำลองออกมา (Adekita & Salau, 2019) โดยมีกระบวนการสร้างแบบจำลองประกอบด้วยชุด Operator ที่ทำงานร่วมกันอย่างเป็นระบบ เริ่มต้นจากการใช้ Operator สำหรับ Read Data เพื่อนำเข้าชุดข้อมูลนักเรียนที่ผ่านการเตรียมความพร้อมแล้ว ตามด้วย Set Role Operator ซึ่งมีหน้าที่กำหนดบทบาทของคอลัมน์ในชุดข้อมูล โดยเฉพาะอย่างยิ่งการระบุตัวแปรเป้าหมายสำหรับการพยากรณ์ จากนั้น Split Data Operator จะถูกใช้เพื่อแบ่งข้อมูลออกเป็นชุดฝึกและ ชุดทดสอบในอัตราส่วนที่เหมาะสม เพื่อรองรับการฝึกและการประเมินแบบจำลองอย่างเป็นอิสระ ขั้นตอนที่สำคัญคือการประยุกต์ใช้ Cross Validation Operator ซึ่งภายในประกอบด้วย Decision Tree Operator สำหรับการสร้างแบบจำลอง โดยมีการปรับแต่งพารามิเตอร์เชิงลึก Maximal Depth และเกณฑ์ขั้นต่ำของข้อมูลในแต่ละโหนด เพื่อควบคุมความซับซ้อนและป้องกันภาวะ Overfitting นอกจากนี้ยังมี Apply Model Operator เพื่อนำแบบจำลองไปใช้ทำนายผล และ Performance Operator สำหรับการประเมินประสิทธิภาพของแบบจำลองด้วยเมตริกที่เกี่ยวข้อง ดังภาพที่ 4

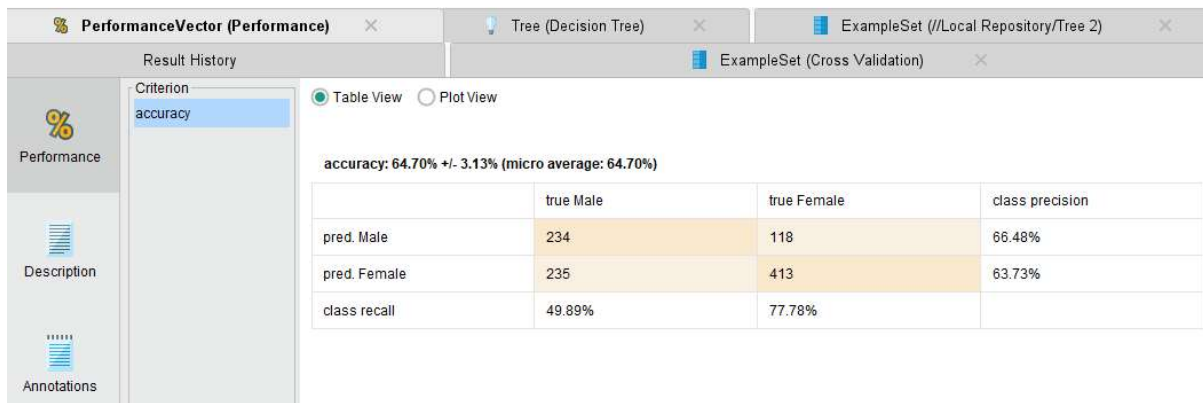


ภาพที่ 4 แบบจำลองอัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree Algorithm)

ผลลัพธ์การจำแนกประเภทโดยแบ่งข้อมูล 80% เป็นข้อมูลชุดฝึก และ 20% เป็นข้อมูลชุดทดสอบ โดยใช้วิธีอัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree Algorithm) และใช้คุณลักษณะเพศของนักเรียน, GPA 1, GPA 2, GPA 3, GPA 4, GPA 5, GPA 6, SKS 1, SKS 2, SKS 3, SKS 4, SKS 5, SKS 6 และ GPA รวมถึงสถานะ

4. ขั้นตอนการทดสอบและประเมินประสิทธิภาพของแบบจำลอง (Evaluation)

ทดสอบและประเมินประสิทธิภาพของแบบจำลองโดยใช้เทคนิค 10-fold cross-validation เนื่องจากเป็นจำนวน Fold ที่ได้รับการยอมรับและใช้กันอย่างแพร่หลายในงานวิจัยด้าน Machine Learning โดยให้ความสมดุลที่ระหว่างความน่าเชื่อถือของผลการประเมินและการคำนวณ (Kohavi, 1995) เป็นการประเมินแบบจำลองโดยแบ่งชุดข้อมูลออกเป็น 10 ส่วน สลับกันฝึกและทดสอบ แล้วหาค่าเฉลี่ย เพื่อเพิ่มความแม่นยำในการประเมินและลดความเอนเอียงของผลการพยากรณ์ โดยใช้ชุดข้อมูลทดสอบที่ถูกแยกไว้ล่วงหน้าแบบสุ่ม (Random Split) นอกจากนี้จะประเมินผลลัพธ์ด้วยค่าตัวชี้วัดต่าง ๆ ได้แก่ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-measure) และการวิเคราะห์กราฟ Confusion Matrix เพื่อเปรียบเทียบค่าที่พยากรณ์กับค่าจริง แบบจำลองต้นไม้ตัดสินใจ ที่ใช้ในการวิเคราะห์ให้ค่าความแม่นยำ (Accuracy) เท่ากับ 64.70% และค่า RMSE เท่ากับ 0.478 ซึ่งสะท้อนถึงประสิทธิภาพในระดับปานกลาง ขณะที่ค่าเบี่ยงเบนมาตรฐานของผลลัพธ์อยู่ในช่วง -0.01 ถึง 0.01 แสดงถึงความเสถียรของโมเดล ทั้งนี้ยังสามารถปรับปรุงประสิทธิภาพเพิ่มเติมได้โดยการปรับพารามิเตอร์หรือทดลองใช้เทคนิคอื่น ๆ เพื่อเปรียบเทียบผลลัพธ์ ดังภาพที่ 5

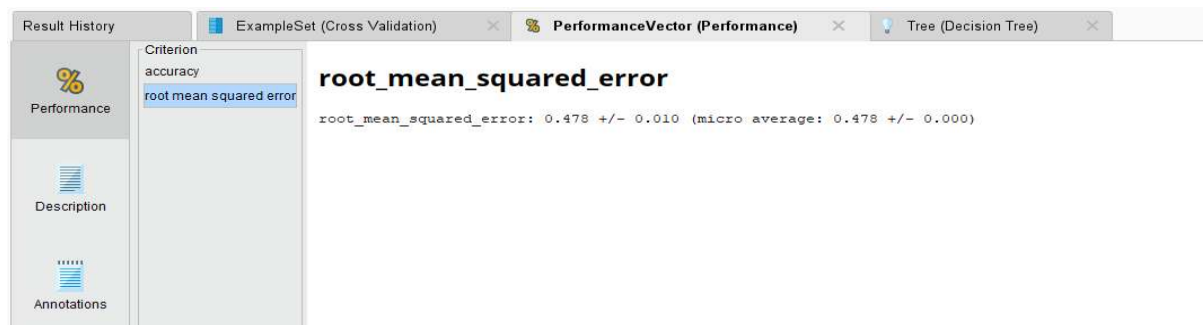


The screenshot shows the PerformanceVector (Performance) window with the 'Table View' selected. The 'Criterion' dropdown is set to 'accuracy'. The main display shows 'accuracy: 64.70% +/- 3.13% (micro average: 64.70%)'. Below this is a table with the following data:

	true Male	true Female	class precision
pred. Male	234	118	66.48%
pred. Female	235	413	63.73%
class recall	49.89%	77.78%	

ภาพที่ 5 ค่าความถูกต้อง

ผลลัพธ์การพยากรณ์การเรียนรู้ของนักเรียนมีความแม่นยำ 64.70% โดยมีค่าความคลาดเคลื่อนมาตรฐาน ($\pm$) อยู่ที่ 3.13% โดยการวิเคราะห์จากตาราง Confusion Matrix พบว่า กลุ่มตัวอย่างที่มีเพศจริงเป็นผู้ชาย จำแนกถูกต้องเป็นผู้ชาย จำนวน 234 ชุดข้อมูล และจำแนกผิดเป็นผู้หญิง จำนวน 235 ชุดข้อมูล กลุ่มตัวอย่างที่มีเพศจริงเป็นผู้หญิง จำแนกถูกต้องเป็นผู้หญิง จำนวน 413 ชุดข้อมูล และถูกจำแนกผิดเป็นผู้ชาย จำนวน 118 ชุดข้อมูล ค่าวัดประสิทธิภาพเชิงชั้นเรียน (Class Precision) พบว่า เพศชาย (Male) มีค่า Precision เท่ากับ 66.48% และเพศหญิง (Female) มีค่า Precision เท่ากับ 63.73% ค่าความสามารถในการดึงข้อมูล (Class Recall) พบว่า เพศชายมีค่า Recall เท่ากับ 49.89% และเพศหญิงมีค่า Recall เท่ากับ 77.78% และแสดงค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Squared Error) ดังภาพที่ 6



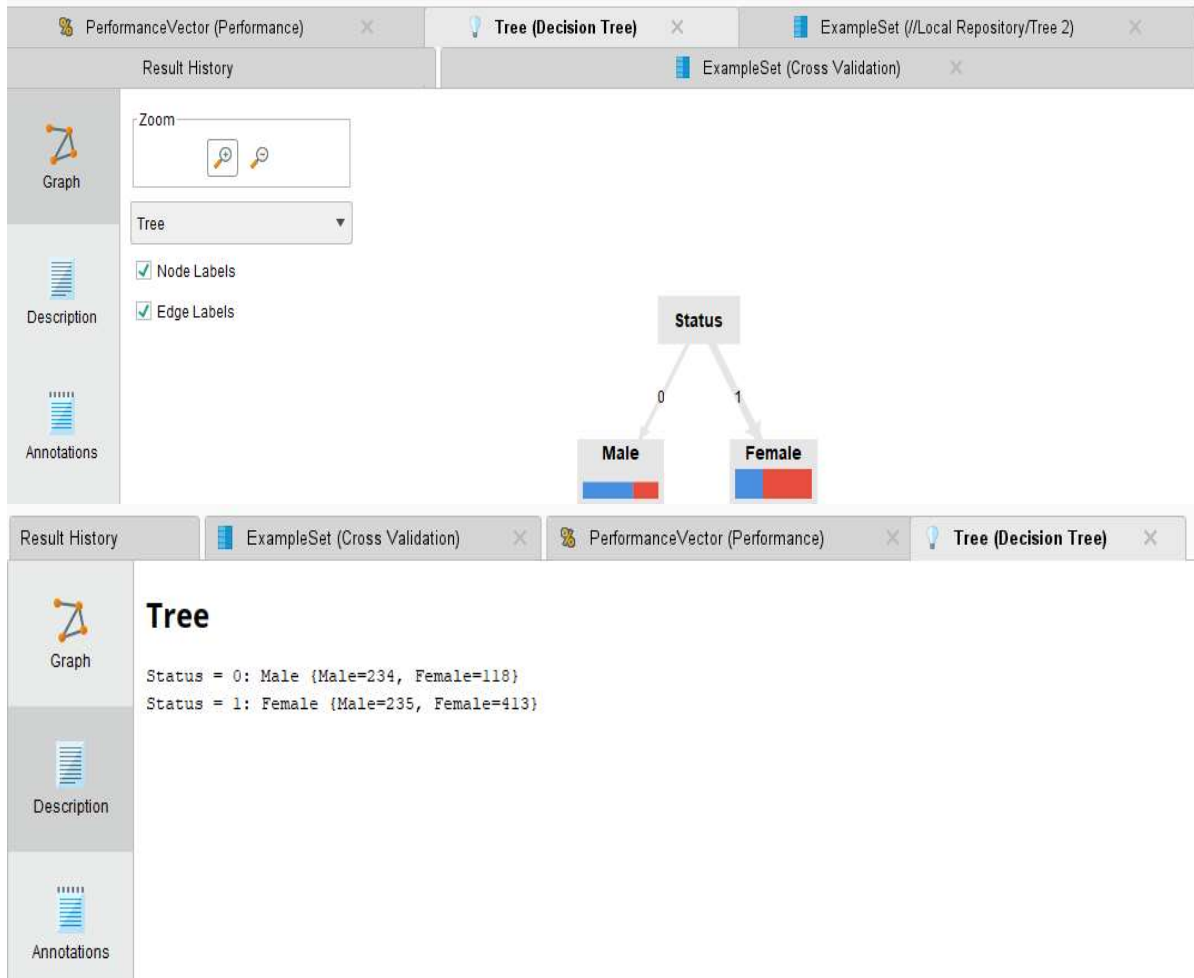
The screenshot shows the PerformanceVector (Performance) window with the 'Table View' selected. The 'Criterion' dropdown is set to 'root mean squared error'. The main display shows 'root_mean_squared_error: 0.478 +/- 0.010 (micro average: 0.478 +/- 0.000)'.

ภาพที่ 6 ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง

ผลลัพธ์การพยากรณ์การเรียนรู้ของนักเรียนด้วยแบบจำลองต้นไม้ตัดสินใจพบว่า ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง มีค่าเฉลี่ยเท่ากับ 0.478 โดยมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ ± 0.010 ทั้งนี้ค่าเฉลี่ยไมโคร (Micro Average) ของค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง มีค่าเท่ากับ 0.478 ± 0.000 สรุปได้ว่า ความคลาดเคลื่อนโดยเฉลี่ยระหว่างค่าที่พยากรณ์ได้กับค่าจริงอยู่ในระดับที่ยอมรับได้

5. ขั้นการนำแบบจำลองไปใช้งานจริง (Deployment)

ลักษณะของแบบจำลองอัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree Algorithm) ดังภาพที่ 7



ภาพที่ 7 แบบจำลองต้นไม้ตัดสินใจ

จากภาพแสดงโครงสร้างของต้นไม้ตัดสินใจ ซึ่งสร้างขึ้นเพื่อจำแนกเพศของกลุ่มตัวอย่างโดยใช้ตัวแปร "Status" เป็นตัวแปรในการแยกข้อมูลโหนดราก (Root Node) เมื่อค่า Status = 0 กลุ่มข้อมูลจะถูกจำแนกกว่าเป็นเพศ "Male" ซึ่งภายในกลุ่มนี้มีตัวอย่างที่เป็นเพศชายจำนวน 234 ชุดข้อมูล และเพศหญิงจำนวน 118 ชุดข้อมูล และเมื่อค่า Status = 1 กลุ่มข้อมูลจะถูกจำแนกกว่าเป็นเพศ "Female" ซึ่งภายในกลุ่มนี้มีตัวอย่างที่เป็นเพศชายจำนวน 235 ชุดข้อมูล และเพศหญิงจำนวน 413 ชุดข้อมูล

สำหรับการประยุกต์ใช้งานผ่านแพลตฟอร์ม One Compiler ที่ปรากฏในบทความเป็นเพียงตัวอย่างเบื้องต้นเพื่อสาธิตรูปแบบอินเทอร์เฟซผู้ใช้ โดยยังไม่เชื่อมต่อกับแบบจำลองที่สร้างจากอัลกอริทึมต้นไม้ตัดสินใจโดยตรง ซึ่งแบบจำลองหลักในบทความนี้ได้พัฒนาบน Altair AI Studio และสามารถส่งออกในรูปแบบ PMML (Predictive Model Markup Language) เป็นภาษามาตรฐานแบบ XML ที่ใช้ในการแลกเปลี่ยนแบบจำลองการเรียนรู้ของเครื่อง (Machine Learning Models) ระหว่างซอฟต์แวร์และแพลตฟอร์มต่าง ๆ โดยไม่ต้องสร้างโมเดลใหม่ตั้งแต่ต้น หรือ JavaScript เพื่อใช้งานร่วมกับระบบจริงในภายหลัง โดยจะประมวลผลการพยากรณ์จากโครงสร้างของต้นไม้ตัดสินใจอย่างครบถ้วน สามารถเข้าถึงได้จาก Link: <https://shorturl.asia/rnlfI> ตามที่แสดงในภาพที่ 8

พยากรณ์ผลสัมฤทธิ์ทางการเรียน ป.6

ชื่อ:

นามสกุล:

วิชา:

เลือกอัลกอริทึม:

Decision Tree

ผลการเรียนก่อนหน้า:

อัตราการเข้าเรียน (%):

การมีส่วนร่วมของอุปกรณ์ (%):

เวลาเรียนเฉลี่ยต่อสัปดาห์ (ชั่วโมง):

3

ผลการพยากรณ์: ไม่ผ่าน (Accuracy: 50.90%)

ตารางสรุปข้อมูลผู้กรอกข้อมูล

ชื่อ	วิชา	เกรดเฉลี่ย	เข้าเรียน (%)	ผู้ปกครองมีส่วนร่วม (%)	เรียนเฉลี่ย (ชั่วโมง)	อัลกอริทึม	ผลพยากรณ์	Accuracy
มานิจ โจติ	คณิตศาสตร์	3.50	95.0	80.0	4.0	decision_tree	ผ่าน	63.20%
มานะ โจงาม	คณิตศาสตร์	2.50	80.0	70.0	4.0	decision_tree	ผ่าน	54.20%

ภาพที่ 8 การประยุกต์ใช้ร่วมกับแอปพลิเคชัน One Compiler

บทสรุป

บทสรุปนี้มุ่งนำเสนอผลการศึกษาและองค์ความรู้ที่เกิดขึ้นจากงานวิจัยที่เกี่ยวกับการประยุกต์ใช้เทคนิคการขุดข้อมูลทางการศึกษาโดยอัลกอริทึมต้นไม้ตัดสินใจ เพื่อพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนักเรียน โดยมีวัตถุประสงค์เพื่ออธิบายแนวคิดพื้นฐานของการขุดข้อมูลทางการศึกษา อัลกอริทึมต้นไม้ตัดสินใจ และการใช้โปรแกรม Altair AI Studio ในการพัฒนาแบบจำลองเชิงพยากรณ์

ผลการวิจัยชี้ให้เห็นว่าแบบจำลองที่พัฒนาขึ้นมีความแม่นยำในระดับปานกลาง ซึ่งแสดงให้เห็นถึงศักยภาพในการนำไปใช้เป็นเครื่องมือสนับสนุนการตัดสินใจสำหรับบุคลากรทางการศึกษาในการวางแผนการเรียนการสอน การให้คำปรึกษาแก่นักเรียน และการพัฒนาคุณภาพการศึกษาในยุคปัจจุบัน อย่างไรก็ตาม การพยากรณ์ผลสัมฤทธิ์ทางการเรียนนั้นมีความซับซ้อน เนื่องจากปัจจัยที่ส่งผลต่อการเรียนรู้ของนักเรียนมีหลากหลายและไม่อาจจำกัดอยู่เพียงข้อมูลเชิงปริมาณที่นำมาใช้ในการสร้างแบบจำลองได้ ปัจจัยภายนอก อาทิ สภาพแวดล้อมทางสังคม เศรษฐกิจ อารมณ์ และพฤติกรรมการเรียนรู้ส่วนบุคคล ล้วนมีผลกระทบที่อาจนำไปสู่ความคลาดเคลื่อนในการพยากรณ์ เพื่อเพิ่มความน่าเชื่อถือของแบบจำลอง จำเป็นต้องพิจารณาใช้ข้อมูลที่มีคุณภาพและความสมบูรณ์สูง รวมถึงการบูรณาการข้อมูลเชิงคุณภาพที่หลากหลาย และเน้นย้ำว่าแบบจำลองเป็นเพียงเครื่องมือสนับสนุนการตัดสินใจที่ต้องควบคู่ไปกับการพิจารณาบริบทเฉพาะบุคคลและการตีความโดยผู้เชี่ยวชาญ

งานวิชาการนี้ได้สร้างองค์ความรู้ใหม่หลายประการ ได้แก่ การนำเสนอแนวทางการประยุกต์ใช้ปัญญาประดิษฐ์ โดยเฉพาะอย่างยิ่ง ChatGPT ในการสร้างข้อมูลจำลองสำหรับงานขุดข้อมูลทางการศึกษา ซึ่งเป็นแนวทางที่ยังไม่แพร่หลายในบริบทงานวิจัยด้านการศึกษา นอกจากนี้ การใช้ Altair AI Studio เป็นการสาธิตให้เห็นถึงศักยภาพของเครื่องมือที่พัฒนาขึ้นสำหรับภาคธุรกิจ ซึ่งสามารถนำมาประยุกต์ใช้ในบริบทของการศึกษาได้อย่างมีประสิทธิภาพในการสร้างแบบจำลองเชิงพยากรณ์เพื่อสนับสนุนการวางแผนการจัดการเรียนรู้และการให้คำปรึกษา และที่สำคัญ การประยุกต์ใช้อัลกอริทึมต้นไม้ตัดสินใจได้แสดงให้เห็นถึงประโยชน์ในแง่ของความสามารถในการตีความผลลัพธ์ได้อย่างโปร่งใส ทำให้ครูและผู้บริหารสามารถเข้าใจเหตุผลเบื้องหลังการตัดสินใจของแบบจำลองได้ง่าย ซึ่งเอื้อต่อการนำข้อมูลไปประยุกต์ใช้ในการวางแผนและดำเนินการแทรกแซงทางการศึกษาได้อย่างเหมาะสม โมเดล Decision Tree ที่พัฒนาในงานนี้ จึงสามารถนำไปใช้เป็นเครื่องมือช่วยประเมินความเสี่ยงของนักเรียนที่อาจมีผลการเรียนต่ำ โดยครู ผู้บริหารสถานศึกษา

หรือแม้กระทั่งผู้ปกครอง สามารถใช้ข้อมูลเชิงลึกที่ได้จากโมเดลนี้เพื่อวางแผนการสอน การให้คำปรึกษา การจัดกิจกรรมเสริม หรือการช่วยเหลือทางวิชาการได้อย่างตรงจุดและทันท่วงที ซึ่งจะช่วยให้การจัดการเรียนรู้มีประสิทธิภาพและตอบสนองความต้องการของนักเรียนแต่ละบุคคลได้ดียิ่งขึ้น องค์ความรู้เหล่านี้ถือเป็นรากฐานสำคัญในการขับเคลื่อนการพัฒนาการศึกษาให้ก้าวหน้ายิ่งขึ้นในยุคดิจิทัล

โดยสรุปแล้ว แม้ว่าข้อมูลด้านการเรียนการสอนที่มีนักเรียนเป็นศูนย์กลางจะมีความซับซ้อนและมีปัจจัยแทรกซ้อนที่ท้าทาย แต่ด้วยการประยุกต์ใช้หลักการของการขุดข้อมูลทางการศึกษาอย่างรอบด้าน ทั้งในด้านการรวบรวมข้อมูล การสร้างตัวแปร การเลือกใช้แบบจำลอง และที่สำคัญที่สุดคือการตีความผลลัพธ์ควบคู่กับบริบทการศึกษา จะสามารถเพิ่มความน่าเชื่อถือและความแม่นยำของการพยากรณ์ผลลัพธ์ทางการเรียนได้อย่างมีนัยสำคัญ และนำไปสู่การสนับสนุนการเรียนรู้ของนักเรียนได้อย่างมีประสิทธิภาพมากขึ้น

ข้อจำกัดและข้อเสนอแนะสำหรับการศึกษาในอนาคต

ข้อจำกัดที่สำคัญของงานวิจัยนี้คือ การใช้ข้อมูลจำลองที่สร้างขึ้น ซึ่งอาจยังไม่สามารถสะท้อนความซับซ้อนและความหลากหลายของข้อมูลจริงที่เกิดขึ้นในภาคสนามได้อย่างครบถ้วนสมบูรณ์ ดังนั้น เพื่อเป็นการต่อยอดงานวิจัยในอนาคต จึงมีข้อเสนอแนะดังนี้ (1) ควรมีการทดลองใช้โมเดลที่พัฒนาขึ้นกับข้อมูลจริงในสถานการณ์จริง เพื่อประเมินประสิทธิภาพและความแม่นยำของโมเดลภายใต้บริบทที่แท้จริง และ (2) ควรพิจารณาพัฒนาอินเทอร์เฟซ (Interface) หรือระบบที่สามารถเชื่อมโยงกับโมเดลที่สร้างขึ้น ให้สามารถใช้งานได้จริงในสถานศึกษา เพื่ออำนวยความสะดวกในการเข้าถึงและนำข้อมูลเชิงลึกไปใช้ประโยชน์ในระดับปฏิบัติการได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

เอกสารอ้างอิง

- ชัยมงคล ปินะสา, สำราญ มีแจ้ง, และน้ำทิพย์ งามอาภาณิชย์. (2567). การประยุกต์ใช้โครงข่ายประสาทเทียมในการคัดเลือกข้อสอบข้อถัดไป สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์แบบพหุมิติ. *วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม*. 30(1); 1-15.
- Abdullahi, A. M., Htet, S. A., Naing, N. N. N., Hossain, M. S., Ismail, S., & Zaaba, M. A. M. (2024). A prediction of customer segmentation using random forest in RapidMiner. *2023 IEEE 21st Student Conference on Research and Development (SCoReD)*, 33-38. <https://doi.org/10.1109/SCoReD60679.2023.10563451>
- Adekitan, A. I., & Salau, O. (2019). The impact of engineering students' performance in the first three years on their graduation result using educational data mining. *Heliyon*, 5(2), 1-21. <https://doi.org/10.1016/j.heliyon.2019.e01250>
- Askinadze, A., & Conrad, S. (2018). Respecting data privacy in educational data mining: an approach to the transparent handling of student data and dealing with the resulting missing value problem. *2018 IEEE 27th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, 160-164. <https://doi.org/10.1109/WETICE.2018.00037>
- Batool, S., Rashid, J., Nisar, M. W., Kim, J., Kwon, H.-Y., & Hussain, A. (2023). Educational data mining to predict students' academic performance: A survey study. *Education and Information Technologies*, 28(1), 905-971. <https://doi.org/10.1007/s10639-022-11152-y>

- Bjaoui, M., Sakly, H., Said, M., Kraiem, N., & Bouhlef, M. S. (2020). Depth insight for data scientist with RapidMiner «an innovative tool for AI and big data towards medical applications». *Proceedings of the 2nd international conference on digital tools & uses congress*, 28, 905-971. <https://doi.org/10.1145/3423603.3424059>
- Chen, S., & Lin, X. (2023). Application of decision tree algorithm in educational data mining. *Curriculum and Teaching Methodology*, 6(8), 120-127. <https://doi.org/10.23977/curtm.2023.060818>
- Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2016). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies*, 10(1), 17-29. <https://doi.org/10.1109/TLT.2016.2616312>
- Corina Simionescu, M. D., Bogdănel Constantin Grădinaru, Marius Silviu Măciucă. (2024). Educational data mining in European union – achievements and challenges: A systematic literature review. *International Journal of Advanced Computer Science and Applications*, 15(3). <https://doi.org/10.14569/IJACSA.2024.0150386>
- Dol, S. M., & Jawandhiya, P. M. (2024). Systematic review and analysis of EDM for predicting the academic performance of students. *Journal of The Institution of Engineers (India): Series B*, 105(4), 1021-1071. <https://doi.org/10.1007/s40031-024-00998-0>
- Dutt, A., Ismail, M. A., & Herawan, T. (2017). A systematic review on educational data mining. *IEEE Access*, 5, 15991-16005. <https://doi.org/10.1109/ACCESS.2017.2654247>
- Hasan, H. R., Rabby, A. S. A., Islam, M. T., & Hossain, S. A. (2019). Machine learning algorithm for student's performance prediction. *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 27, 1-7. <https://doi.org/10.1109/ICCCNT45670.2019.8944629>
- Hu, Y.-H. (2022). Effects and acceptance of precision education in an AI-supported smart learning environment. *Education and Information Technologies*, 27(2), 2013-2037. <https://doi.org/10.1007/s10639-021-10664-3>
- Joy, T. T., Rana, S., Gupta, S., & Venkatesh, S. (2019). A flexible transfer learning framework for Bayesian optimization with convergence guarantee. *Expert Systems with Applications*, 115, 656-672. <https://doi.org/10.1016/j.eswa.2018.08.023>
- Klein, M., Gerlitz, L., & Sychalska-Wojtkiewicz, M. (2021). Cultural and creative industries as boost for innovation and sustainable development of companies in cross innovation process. *Procedia Computer Science*, 192, 4218-4226.

- Kostopoulos, G., Kotsiantis, S., Fazakis, N., Koutsonikos, G., & Pierrakeas, C. (2019). A semi-supervised regression algorithm for grade prediction of students in distance learning courses. *International Journal on Artificial Intelligence Tools*, 28(4), 1940001. <https://doi.org/10.1142/S0218213019400013>
- Martínez-Abad, F., Gamazo, A., & Rodríguez-Conde, M.-J. (2020). Educational data mining: identification of factors associated with school effectiveness in PISA assessment. *Studies in Educational Evaluation*, 66, 100875. <https://doi.org/10.1016/j.stueduc.2020.100875>
- Moreira, G. P., Talledo, O. L. M., Rodríguez, H. M., Moreira, M. J. P., & Morales, J. J. P. (2025). Using classification and prediction algorithms to process surveys in the Altair® RapidMiner data mining software. *Prohominum*, 7(1), 136-153. <https://doi.org/10.47606/ACVEN/PH0319>
- Novianto, E., & Suhirman, S. (2024). Comparison of K-Nearest neighbor classification methods and support vector machine in predicting students' study period. *Journal of Education and Science*, 33(1), 32-45. <https://doi.org/10.33899/edusj.2023.144865.1408>
- Putri, R. P. S., & Waspada, I. (2018). Penerapan Algoritma C4. 5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika. *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, 4(1), 1-7. <https://doi.org/10.23917/khif.v4i1.5975>
- Romero, C., & Ventura, S. (2017). Educational data science in massive open online courses. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(1), e1187. <https://doi.org/10.1002/widm.1187>
- Switrayana, I. N., Ashadi, D., Hairani, H., & Aminuddin, A. (2023). Sentiment analysis and topic modeling of Kitabisa applications using Support Vector Machine (SVM) and Smote-Tomek Links Methods. *International Journal of Engineering and Computer Science Applications (IJECSA)*, 2(2), 87-98.
- Toivonen, T., & Jormanainen, I. (2019). Evolution of decision tree classifiers in open ended educational data mining. *Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality*, León, 290-296. <https://doi.org/10.1145/3362789.3362880>
- Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., & Ragos, O. (2019). Implementing AutoML in educational data mining for prediction tasks. *Applied Sciences*, 10(1), 90. <https://doi.org/10.3390/app10010090>

- Varade, R. V., & Thankanchan, B. (2021). Academic performance prediction of undergraduate students using decision tree algorithm. *SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology*, 13(1), 97-100.
<https://doi.org/10.18090/samriddhi.v13iS1.22>
- Wiyono, S., Abidin, T., Wibowo, D., Hidayatullah, M., & Dairoh, D. (2019). Comparative study of machine learning knn, svm, and decision tree algorithm to predict student's performance. *International Journal of Research-Granthaalayah*, 7(1), 190-196.
<https://doi.org/10.29121/granthaalayah.v7.i1.2019.1048>
- Xu, J., Moon, K. H., & Van Der Schaar, M. (2017). A machine learning approach for tracking and predicting student performance in degree programs. *IEEE Journal of Selected Topics in Signal Processing*, 11(5), 742-753. <https://doi.org/10.1109/JSTSP.2017.2692560>