



Student Academic Success Prediction System Using Random Forest Technique

Wichit Sombat¹, Chayaporn Kaensar^{1,*}, Supawadee Hiranpongsin¹ and Sirinya Boakam¹

¹*Department of Mathematics, Statistics and Computer, Faculty of Science, Ubon Ratchathani University*

**Email: chayaporn.k@ubu.ac.th*

Received <4 May 2024>; Revised <4 July 2024>; Accepted <10 July 2024>

Abstract

The challenges of academic underperformance and rising dropout rates have contributed to declining student retention rates, resulting in implications such as wasted time, increased expenses, and missed job opportunities. Consequently, the need to analyze and prepare student data is paramount. This study aims to address these challenges by: 1) investigating and constructing a predictive model of student academic performance within the Data Science and Software Innovation (DSSI), Information Technology and Communications (ICT), Chemistry, and Biology programs at the Faculty of Science, Ubon Ratchathani University, utilizing the Random Forest technique, and 2) creating a prototype web application to apply the model and summarize the findings, developing with Django and PostgreSQL. The dataset used for model construction encompasses academic records from both secondary school and university for 1,336 students enrolled in the programs spanning academic years 2017-2021. This dataset includes 11 factors: educational programs, first-term university GPA, and academic performance in 8 core subject groups: Thai Language, Mathematics, Science, Social Studies, Religion and Culture, Arts, Career and Technology, Foreign Languages, alongside high school GPA. The results of the prototype model research for DSSI, ICT, Chemistry, and Biology programs demonstrates the ability to predict student learning outcomes, with F1-measure values of 88.02%, 86.18%, 84.04% and 85.18% respectively. Consequently, the implementation of the developed system holds potential to benefit students, educators, and educational institutions in the design and implementation of more effective and tailored teaching methodologies.

Keywords: Academic performance, Dropout rates, Predictive model, Random Forest technique, Web application

ระบบพยากรณ์การสำเร็จการศึกษาของผู้เรียนด้วยเทคนิคป่าสุ่ม

วิจิต สมบัติ¹ ชยาพร แก่นสาร^{1,*} สุภาวดี หิรัญพงศ์สิน¹ และ สิริณญา เบ้าคำ¹

¹ภาควิชาคณิตศาสตร์ สถิติ และคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี

*Email: chayaporn.k@ubu.ac.th

รับบทความ: 4 พฤษภาคม 2567 แก้ไขบทความ: 4 กรกฎาคม 2567 ยอมรับตีพิมพ์: 10 กรกฎาคม 2567

บทคัดย่อ

ปัญหาผลการเรียนที่อยู่ในระดับค่อนข้างน้อยและการตกออกเพิ่มขึ้น ทำให้อัตราการคงอยู่ของนักศึกษาลดลง ทำให้ส่งผลกระทบต่อหลายด้าน เช่น สูญเสียเวลา ค่าใช้จ่าย และโอกาสในการได้งาน การเตรียมความพร้อมและการวิเคราะห์ข้อมูลผู้เรียนจึงเป็นสิ่งสำคัญ งานวิจัยนี้มีวัตถุประสงค์เพื่อ 1) ศึกษาและสร้างแบบจำลองการพยากรณ์ประสิทธิภาพการเรียนรู้ของนักศึกษา 4 สาขาวิชา ได้แก่ วิทยาการข้อมูลและนวัตกรรมซอฟต์แวร์ เทคโนโลยีสารสนเทศและการสื่อสาร เคมี และจุลชีววิทยา คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี ด้วยเทคนิคป่าสุ่ม และ 2) พัฒนาระบบต้นแบบในรูปแบบเว็บแอปพลิเคชันเพื่อใช้งานแบบจำลองและสรุปผลด้วยจังก์เฟรมเวิร์กและจัดเก็บข้อมูลในฐานข้อมูลโพสต์เกรสคิวเอล ชุดข้อมูลที่ใช้ในการสร้างแบบจำลองเป็นผลการเรียนจากทั้งโรงเรียนและมหาวิทยาลัยในอดีตของนักศึกษา 4 สาขาวิชา คณะวิทยาศาสตร์ ในปีการศึกษา 2560-2564 จำนวน 1,336 คน โดยพิจารณาข้อมูล 11 ปัจจัย ได้แก่ สาขาวิชาการศึกษา เกรดเฉลี่ยสะสมปีแรกในมหาวิทยาลัย ผลการเรียนใน 8 กลุ่มสาระหลัก ได้แก่ กลุ่มวิชาภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ สังคมศึกษา ศาสนา และวัฒนธรรม ศิลปะ การงานอาชีพและเทคโนโลยี ภาษาต่างประเทศ และเกรดเฉลี่ยในโรงเรียน ผลการวิจัยแบบจำลองต้นแบบของสาขาวิชา วิทยาการข้อมูลและนวัตกรรมซอฟต์แวร์ เทคโนโลยีสารสนเทศและการสื่อสาร เคมี และจุลชีววิทยา สามารถพยากรณ์ผลการเรียนของผู้เรียนโดยมีค่า F1-measure เท่ากับ 88.02% 86.18% 84.04% และ 85.18% ตามลำดับ ดังนั้นการนำระบบที่พัฒนาขึ้นนี้ไปใช้จะมีประโยชน์ต่อผู้เรียน ผู้สอนในสาขาวิชา และสถาบันการศึกษา เพื่อใช้ในการวางแผนและการจัดการสอนให้เหมาะสมและมีประสิทธิภาพมากยิ่งขึ้น

คำสำคัญ: ผลการเรียน อัตราการตกออก แบบจำลองการพยากรณ์ เทคนิคป่าสุ่ม เว็บแอปพลิเคชัน

บทนำ

ในช่วงปี พ.ศ. 2560-2566 คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี พบปัญหานักศึกษาใหม่มีผลการเรียนค่อนข้างน้อย ตกออก หรือแม้แต่ไม่สำเร็จการศึกษาตามแผนที่กำหนดเพิ่มมากขึ้น ทำให้อัตราการคงอยู่ของนักศึกษาตลอดหลักสูตรลดน้อยลงอย่างต่อเนื่อง ส่งผลเสียและผลกระทบในหลายด้านทั้งในส่วนบุคคลและผู้ปกครองที่สูญเสียเวลา ค่าใช้จ่าย และเสียโอกาสในการได้งานทำ อีกทั้งหลักสูตรหรือคณะมีค่าใช้จ่ายและงบประมาณเพิ่มขึ้น หรือแม้แต่ประเทศที่อาจจะขาดแคลนบุคลากรด้านวิทยาศาสตร์และเทคโนโลยีที่ไปช่วยขับเคลื่อนเศรษฐกิจได้ทัน ดังนั้นจึงนับเป็นปัญหาด้านการศึกษาที่ต้องเร่งแก้ไขอย่างเร่งด่วน จากการศึกษาของ Valverde-Berrococo, Acevedo-Borrega and Cerezo-Pizarro (2022) ที่ได้ศึกษาแนวทางและปัจจัยสำคัญที่มีผลต่อการจัดการสอนเพื่อส่งเสริมคุณภาพผู้เรียนในระดับอุดมศึกษา พบว่าการเสริมทักษะทางวิชาการและสังคม เพื่อเตรียมความพร้อมผู้เรียนก่อนเข้าสู่บทเรียนหรือหลักสูตรใหม่ และการนำข้อมูลของผู้เรียนมาประมวลผลด้วยนวัตกรรมทางการศึกษาเพื่อเข้าใจถึงบริบทและภูมิหลังผู้เรียนจะช่วยให้ผู้สอนรวมทั้งสถาบันการศึกษาสามารถจัดการผู้เรียนและพัฒนาคุณภาพการศึกษาได้อย่างมีประสิทธิภาพมากยิ่งขึ้น โดยในส่วนการเตรียมความพร้อมผู้เรียนนั้น ในคณะหรือมหาวิทยาลัยต่างๆ ได้มีการดำเนินการโดยจัดกิจกรรมเตรียมความพร้อมและส่งเสริมทักษะพื้นฐานด้านวิชาการและอื่น ๆ ให้กับนักศึกษาใหม่อย่างต่อเนื่อง แต่ก็พบว่าประสิทธิภาพการเรียนรู้ของผู้เรียนกลับลดลง ทำให้ในหลายคณะหรือมหาวิทยาลัยมุ่งเน้นในส่วนการนำข้อมูลผู้เรียนมาช่วยวิเคราะห์ สรุปผล วิเคราะห์เชิงลึก คาดการณ์ผลการเรียนล่วงหน้าเพิ่ม อีกทั้งยังนำมาใช้สร้างเครื่องมือและรายงานสารสนเทศที่มีประโยชน์เพื่อสร้างความได้เปรียบและช่วยวางแผนการตัดสินใจด้านการศึกษาในมหาวิทยาลัยมากขึ้น

ดังนั้นจึงเห็นได้จากมีงานวิจัยจำนวนมากได้นำเสนอแนวคิดที่ประยุกต์ใช้เทคนิคทางคอมพิวเตอร์ต่าง ๆ มานำเสนอโครงสร้างระบบ สร้างแบบจำลองพยากรณ์ผล รวมทั้งพัฒนาเครื่องมือต่าง ๆ ที่จะช่วยวิเคราะห์และพยากรณ์ประสิทธิภาพของผู้เรียนเพื่อนำมาใช้ประโยชน์ในการศึกษามากขึ้น เช่น Shou *et al.* (2024) ที่นำข้อมูลผู้เรียน คือ เพศ ประวัติครอบครัว รายได้ การใช้งานอินเทอร์เน็ต ระยะเวลาการเดินทาง และเวลาที่ใช้ในคลาสของผู้เรียนมาสร้างแบบจำลองเพื่อพยากรณ์ประสิทธิภาพการเรียนรู้ของผู้เรียน เช่นเดียวกับ Kaensar and Wongnin (2023) และ Nachouki *et al.* (2023) ที่นำข้อมูลคะแนนผู้เรียนและบริบทของโรงเรียนมาสร้างแบบจำลองและวิเคราะห์ปัจจัยที่มีผลต่อความสำเร็จของผู้เรียนร่วมด้วย ซึ่งผลการทดสอบระบบพบว่าผู้สอนสามารถใช้ข้อมูลในการคาดการณ์และวางแผนการสอน หรือติดตามผู้เรียนที่มีความเสี่ยงได้ดียิ่งขึ้น นอกจากนี้ งานวิจัย Alboaneen *et al.* (2022) ยังต่อยอดโดยการนำเสนอเว็บแอปพลิเคชันที่นำแบบจำลองพยากรณ์ผลผู้เรียนที่สร้างจากข้อมูลผู้เรียนและคะแนนสอบเข้ามาแทนเป็นองค์ความรู้ในระบบ ซึ่งข้อดี คือ เป็นเครื่องมือที่ผู้สอน หลักสูตร (สาขาวิชา) และมหาวิทยาลัยสามารถใช้พยากรณ์และสร้างรายงานผลการเรียนรู้ของผู้เรียนได้ นอกจากนี้ Kaensar and Wongnin (2024) ได้พัฒนาโมเดลด้วยเทคนิคการเรียนรู้เชิงลึกและนำไปใช้งานผ่านเว็บแอปพลิเคชันเพื่อเป็นเครื่องมือในการทำนายความสำเร็จของนักศึกษามหาวิทยาลัยอุบลราชธานี เช่นเดียวกับ Riansut (2016) ที่ได้ค้นหาค่าปัจจัยที่มีอิทธิพลต่อเกรดเฉลี่ยสะสมของนิสิตระดับปริญญาตรี มหาวิทยาลัยทักษิณ ซึ่งผลการทดสอบพบว่าเกรดเฉลี่ยสะสมและการเลือกแผนการเรียนในระดับมัธยมศึกษาตอนปลาย มีผลต่อเกรดเฉลี่ยในมหาวิทยาลัยอย่างมีนัยสำคัญ

ดังนั้น งานวิจัยนี้จึงเสนอการสร้างแบบจำลองการพยากรณ์ผู้เรียนและพัฒนาเครื่องมือเพื่อช่วยวิเคราะห์ประสิทธิภาพการเรียนรู้ของนักศึกษาใหม่ โดยใช้ข้อมูลนักศึกษาที่ศึกษาในปีการศึกษา 2560-2564 จำนวน 1,336 คน ของ 4 สาขาวิชา ได้แก่ วิทยาการข้อมูลและนวัตกรรมซอฟต์แวร์ (Data Science and Software Innovation: DSSI) เทคโนโลยีสารสนเทศและการสื่อสาร (Information Technology and Communications: ICT) เคมี (Chemistry) และจุลชีววิทยา (Biology) คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี ซึ่งประกอบด้วยการทำงาน 2 ส่วน ได้แก่ 1) ส่วนการสร้างแบบจำลองเพื่อพยากรณ์ผลการเรียนนักศึกษาใหม่ด้วยเทคนิคป่าสุ่ม (Random Forest) โดยใช้ข้อมูลผลการเรียนในอดีตของนักศึกษา จำนวน 11 ปัจจัย ซึ่งประกอบด้วย ข้อมูลสาขาวิชา ข้อมูลคะแนนจากโรงเรียน ได้แก่ เกรดเฉลี่ย ข้อมูลคะแนนใน 8 กลุ่มสาระ (1. กลุ่มภาษาไทย 2. คณิตศาสตร์ 3. วิทยาศาสตร์ 4. สังคมศึกษา ศาสนา และวัฒนธรรม 5. สุขศึกษาและพลศึกษา 6. ศิลปะ 7. การงานอาชีพและเทคโนโลยี และ 8. ภาษาต่างประเทศ) และข้อมูลคะแนนเกรดเฉลี่ยปี 1 ในมหาวิทยาลัย 2) ส่วนการพัฒนาเว็บแอปพลิเคชันซึ่งถูกพัฒนาในรูปแบบเว็บแอปพลิเคชันด้วยจังก์โพรแกรมเวิร์ก (Django Framework) และจัดเก็บข้อมูลในฐานข้อมูลโพสทเกรสคิวเอล (PostgreSQL) เพื่อนำแบบจำลองต้นแบบมาใช้งานและสรุปเป็นรายงานต่าง ๆ โดยประโยชน์ของงานวิจัยนี้จะช่วยอำนวยความสะดวกให้กับผู้ใช้งานระดับต่าง ๆ กล่าวคือ ผู้เรียนสามารถใช้พยากรณ์ประสิทธิภาพการเรียนรู้ของนักศึกษาใหม่ได้ ผู้สอนทราบข้อมูลผู้เรียนที่มีความเสี่ยงในการตกออกหรือมีผลการเรียนน้อยในปีแรกได้ ซึ่งผู้สอนสามารถวางแผนและจัดการสอนที่เหมาะสมในรายบุคคลได้ ในขณะที่สาขาวิชาหรือคณะ สามารถใช้เป็นข้อมูลในการจัดกิจกรรมเตรียมความพร้อมและเสริมทักษะให้กับผู้เรียนได้สอดคล้องมากขึ้น

ในงานวิจัยนี้แบ่งเนื้อหาออกเป็นส่วนคำถามงานวิจัย แนวคิดและงานวิจัยที่เกี่ยวข้อง โครงสร้างและวิธีดำเนินงานในระบบ ผลการทดสอบระบบและการอภิปรายผล รวมทั้งสรุปและแนวทางการพัฒนางานวิจัยในอนาคต

วัตถุประสงค์การวิจัย

1. เพื่อศึกษาและสร้างแบบจำลองการพยากรณ์ประสิทธิภาพการเรียนรู้ของนักศึกษาจำนวน 4 สาขาวิชา ได้แก่ วิทยาการข้อมูลและนวัตกรรมซอฟต์แวร์ เทคโนโลยีสารสนเทศและการสื่อสาร เคมี และจุลชีววิทยา คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี ด้วยเทคนิคป่าสุ่ม
2. เพื่อนำเสนอโครงสร้างระบบต้นแบบและพัฒนาแอปพลิเคชันในการพยากรณ์ประสิทธิภาพการเรียนรู้ด้วย จังโก้ เฟรมเวิร์กและโพสต์เกรสคิวเอล ของนักศึกษาจำนวน 4 สาขาวิชา ได้แก่ วิทยาการข้อมูลและนวัตกรรมซอฟต์แวร์ เทคโนโลยีสารสนเทศและการสื่อสาร เคมี และจุลชีววิทยา คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี

ทฤษฎีที่เกี่ยวข้อง

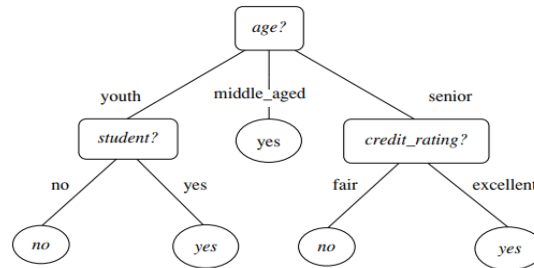
แบบจำลองการจำแนกประเภท (Classification Model)

เป็นการวิเคราะห์ข้อมูลเพื่อพยากรณ์ประเภทของข้อมูล ที่จัดอยู่ในแบบจำลองประเภทการเรียนรู้แบบมีผู้สอน (Supervised Learning Model) โดยนำชุดข้อมูลที่มีอยู่มาพัฒนาแบบจำลองเพื่อการพยากรณ์และประยุกต์ใช้หาคำตอบให้กับชุดข้อมูลใหม่ที่เข้ามา (Naresuan University Publishing House, 2023) ประเภทของแบบจำลองนี้ แบ่งเป็น 3 ประเภทหลัก (Data Innovation and Governance Institute, 2022) ได้แก่

1. การจำแนกประเภทแบบไบนารี (Binary Classification) เป็นการวิเคราะห์ข้อมูลนำเข้าเพื่อพยากรณ์คลาสคำตอบเพียง 2 คลาส เช่น การพยากรณ์การยืนยันสิทธิ์เข้าศึกษาของนักศึกษาใหม่ในระดับมหาวิทยาลัยช่วงสถานการณ์แพร่ระบาดของโควิด-19 ที่มีข้อมูลนำเข้าเป็นข้อมูลการรับเข้าของนักศึกษา โดยให้แบบจำลองพยากรณ์ว่านักศึกษาใหม่จะยืนยันสิทธิ์ หรือไม่ยืนยันสิทธิ์ (Kongchai, Hiranpongsin and Ditcharoen, 2024)
2. การจำแนกประเภทแบบหลายคลาส (Multi-Class Classification) เป็นการวิเคราะห์ข้อมูลที่คล้ายกับการจำแนกประเภทแบบไบนารี เพียงแต่จะมีคลาสคำตอบมากกว่า 2 คลาส ซึ่งจะจำแนกคุณสมบัติของข้อมูลออกเป็นหลายประเภทที่มีความแตกต่างกัน เช่น การจำแนกโรคใบขาวโพด ด้วยข้อมูลนำเข้าที่เป็นภาพของใบขาวโพดที่เป็นโรค 3 ประเภท ได้แก่ โรคใบไหม้แผลใหญ่ โรคราสนิม และโรคใบจุดสีเทา และภาพของใบขาวโพดปกติ รวมเป็น 4 คลาสคำตอบ โดยแบบจำลองจะพยากรณ์ว่าภาพที่นำเข้านั้นจัดอยู่ใน 1 ใน 4 คลาสใด (Mekha et al., 2023)
3. การจำแนกประเภทแบบหลายเลเบล (Multi-Label Classification) เป็นการจำแนกที่มีคลาสคำตอบมากกว่า 2 คลาส และต่างจากการจำแนกประเภทแบบหลายคลาส เนื่องจากข้อมูลชุดหนึ่งอาจมีข้อมูลที่มีคุณสมบัติเหมือนกันแต่สามารถให้ผลลัพธ์ที่มีความแตกต่างกัน เช่น รูปภาพเกี่ยวกับทะเล ภาพหนึ่งจะถูกจำแนกเป็นคลาสคำตอบใดคลาสคำตอบหนึ่ง เช่น ภาพถ่าย ภาพวาด หรือภาพพระบายสี ด้วยการจำแนกประเภทแบบหลายคลาส ในขณะที่การจำแนกประเภทแบบหลายเลเบลภาพ ๆ นั้น จะถูกเลเบลว่าในภาพมี นก น้ำทะเล เรือ หรือ ท้องฟ้า หรือไม่ (Brownlee, 2020)

เทคนิคต้นไม้ตัดสินใจ (Decision Tree Technique)

การนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบโครงสร้างต้นไม้ สามารถสร้างแบบจำลองการจำแนกประเภทได้จากกลุ่มตัวอย่างข้อมูลที่กำหนดไว้ล่วงหน้า และพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจำแนกประเภทได้ด้วยรูปแบบของต้นไม้ โครงสร้างของต้นไม้ ประกอบด้วยโหนดภายในต่าง ๆ ซึ่งแสดงถึงเงื่อนไขหรือคุณลักษณะหนึ่ง ๆ ของข้อมูล โดยที่แต่ละกิ่งของโหนดหนึ่ง ๆ หมายถึงค่าที่เป็นไปได้จากการทดสอบกับคุณลักษณะนั้น ๆ และโหนดใบซึ่งจะมีประเภทของข้อมูลถูกจัดเก็บไว้ ตัวอย่างต้นไม้ตัดสินใจแสดงดังภาพที่ 1 ซึ่งแสดงการพยากรณ์การซื้อคอมพิวเตอร์ด้วยข้อมูลนำเข้าคือคุณลักษณะของลูกค้า โดยโหนดภายในต่าง ๆ จะถูกแทนด้วยสี่เหลี่ยม และโหนดใบจะถูกแทนด้วยวงรี จากรูปจะพบว่าโหนดใบเป็นโหนดที่บ่งบอกถึงประเภทของคำตอบที่ต้องการซึ่งมี 2 คลาส ได้แก่ “yes” หมายถึง ลูกค้าจะซื้อคอมพิวเตอร์ และ “no” หมายถึง ลูกค้าจะไม่ซื้อคอมพิวเตอร์ (Han, Kamber and Pei, 2012)



ภาพที่ 1 ต้นไม้ตัดสินใจแสดงการพยากรณ์การซื้อคอมพิวเตอร์
ที่มา: Han, Kamber and Pei (2012)

การจำแนกประเภทของต้นไม้ตัดสินใจจะใช้ Gini Impurity (Gini) (Jamjumrat, 2022) ซึ่งเป็นหนึ่งในค่าที่ใช้ในการตัดสินใจเลือกโหนดที่ดีที่สุดในการแบ่งข้อมูล (Split point) เพื่อสร้างกิ่งหรือทางแยก โดยค่า Gini เป็นการนำผลรวมของค่าความน่าจะเป็นของเหตุการณ์ที่สนใจมาคูณด้วยผลต่างระหว่าง 1 และค่าความน่าจะเป็นของเหตุการณ์ที่สนใจ ซึ่งคำนวณได้จากสมการที่ (1)

$$GiNi = \sum_{i=1}^C p(i) \times (1 - p(i)) \quad (1)$$

โดยที่

- C คือ จำนวนคลาสคำตอบทั้งหมด
p(i) คือ ความน่าจะเป็นของเหตุการณ์ที่สนใจของคลาสที่ i

หลังจากได้ค่า Gini ของแต่ละกลุ่มในทุก ๆ ตัวแปรแล้ว จะทำการหาค่าถ่วงน้ำหนักของค่า Gini (Weighted Gini) เพื่อเลือกตัวแปรที่มีค่าถ่วงน้ำหนักของค่า Gini ต่ำที่สุดมาใช้ในการตัดสินใจ (โหนดที่ดีที่สุดเพื่อสร้างทางแยก) ก่อน เนื่องจากสามารถแบ่งข้อมูลได้ดีที่สุด โดยค่า Weighted Gini เป็นผลรวมของจำนวนข้อมูลของคลาสที่ i ในตัวแปรที่สนใจหารด้วยจำนวนข้อมูลทั้งหมดในตัวแปรที่สนใจและคูณกับค่า Gini ในเหตุการณ์ที่สนใจ ซึ่งคำนวณได้จากสมการที่ (2)

$$Weighted\ GiNi = \sum_{i=1}^C \frac{n_i}{n} Gini(i) \quad (2)$$

โดยที่

- C คือ จำนวนคลาสคำตอบทั้งหมด
n(i) คือ จำนวนข้อมูลของคลาสที่ i ในตัวแปรที่สนใจ
n คือ จำนวนข้อมูลทั้งหมดในตัวแปรที่สนใจ
Gini(i) คือ ค่า Gini ในเหตุการณ์ที่สนใจของคลาสที่ i

เทคนิคป่าสุ่ม (Random Forest Technique)

ป่าสุ่มเป็นหนึ่งในอัลกอริทึมที่เป็นตัวอย่างของวิธีการเอนเซมเบิล (Ensemble) ซึ่งใช้หลักการรวมหลายอัลกอริทึมเพื่อเพิ่มประสิทธิภาพการจำแนกประเภทข้อมูล และจัดอยู่ในกลุ่มต้นไม้ตัดสินใจ โดยป่าสุ่มจะถูกสร้างจากต้นไม้ตัดสินใจหลาย ๆ ต้น ซึ่งจะเรียนรู้ข้อมูลไม่เหมือนกันจากกลุ่มตัวอย่างย่อยแบบสุ่มจากชุดข้อมูลฝึกฝนและคุณลักษณะแบบสุ่ม ในการพยากรณ์ผลเป็นการพยากรณ์จากต้นไม้ตัดสินใจแต่ละต้นร่วมกัน โดยต้นไม้ตัดสินใจเหล่านั้นจะพยากรณ์ผลของตัวเอง และคำนวณผลของการพยากรณ์จากการโหวต (Vote) ซึ่งช่วยลดปัญหาการเกิดโอเวอร์ฟิตติ้ง (Overfitting) ของแบบจำลองและมีแนวโน้มที่การพยากรณ์จะมีความแม่นยำมากกว่าการพยากรณ์ด้วยต้นไม้ตัดสินใจเพียงต้นเดียว (Han, Kamber and Pei, 2012; Shalev-Shwartz and Ben-David, 2014)

ตัวชี้วัดสำหรับการประเมินประสิทธิภาพ (Confusion Matrix)

การประเมินประสิทธิภาพของแบบจำลองเป็นส่วนสำคัญที่จะประเมินการพยากรณ์ของแบบจำลองว่าสามารถพยากรณ์ได้ดีเพียงใด ตัวชี้วัดสำหรับประเมินการจำแนกประเภท ได้แก่ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ

(Precision) ค่าความระลึก (Recall) และค่า F1-Score (Han, Kamber and Pei, 2012; Soontranon, 2023) ซึ่งคำนวณได้จากสมการที่ (3) - (6) ตามลำดับ

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{3}$$

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{6}$$

โดยที่

True Positive (TP) คือ พยากรณ์ตรงกับสิ่งที่เกิดขึ้นจริง กรณี พยากรณ์ว่าจริงและสิ่งที่เกิดขึ้น คือ จริง

True Negative (TN) คือ พยากรณ์ตรงกับสิ่งที่เกิดขึ้นจริง กรณี พยากรณ์ว่าไม่จริงและสิ่งที่เกิดขึ้น คือ ไม่จริง

False Positive (FP) คือ พยากรณ์ไม่ตรงกับสิ่งที่เกิดขึ้นจริง กรณี พยากรณ์ว่าจริงแต่สิ่งที่เกิดขึ้น คือ ไม่จริง

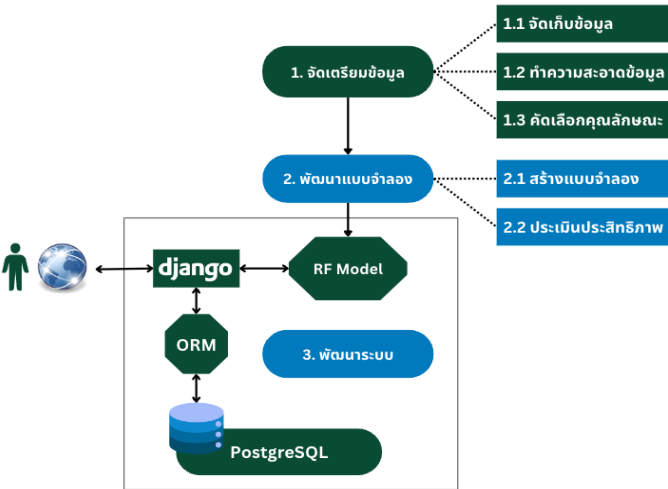
False Negative (FN) คือ พยากรณ์ไม่ตรงกับที่ที่เกิดขึ้นจริง กรณี พยากรณ์ว่าไม่จริงแต่สิ่งที่เกิดขึ้น คือ จริง

จังกโเฟรมเวิร์ก (Django Framework)

เฟรมเวิร์กสำหรับสร้างเว็บแอปพลิเคชันฝั่งแบ็กเอนด์ (Back End) ที่พัฒนาด้วยภาษาไพธอน (Python) มีเป้าหมายหลักในการทำให้การสร้างเว็บไซต์ที่ทำงานร่วมกับฐานข้อมูลที่มีความซับซ้อนในง่ายขึ้น จังกโใช้สถาปัตยกรรมแบบ Model-View-Template (MVT) โดย Model ทำหน้าที่เป็นอินเทอร์เฟซ (Interface) ระหว่างฐานข้อมูลและเซิร์ฟเวอร์ (Server) View เป็นส่วนที่ใช้สำหรับเขียนฟังก์ชันต่าง ๆ จัดการกับการร้องขอ/การตอบกลับ (Request/Response) คำขอของผู้ใช้ การแสดงผล (Render) หน้า HTML ไปแสดงผลที่ส่วนเทมเพลต (Template) รวมถึงการทำการสอบถาม (Query) ข้อมูลจากฐานข้อมูล จังกโเฟรมเวิร์กมีการจัดระเบียบเป็นอย่างดี และมีโครงสร้างพร้อมใช้งานสำหรับงานพัฒนาเว็บทั่วไป เช่น การยืนยันตัวตนผู้ใช้ การจัดการเนื้อหา แผนผังเว็บไซต์ และเว็บฟีดอาร์เอสเอส (RSS feed) (Amazon Web Services, 2024)

วิธีดำเนินการวิจัย

งานวิจัยนี้แบ่งการดำเนินการวิจัยออกเป็น 3 ขั้นตอนหลัก ได้แก่ การจัดเตรียมข้อมูล การพัฒนาแบบจำลอง และการพัฒนาระบบ โดยมีลำดับการดำเนินการตามภาพที่ 2 รายละเอียดในแต่ละขั้นตอนดังต่อไปนี้



ภาพที่ 2 วิธีดำเนินการวิจัย

การจัดเตรียมข้อมูล (Data preparation)

1. **จัดเก็บข้อมูลดิบ (Raw Data)** ในขั้นตอนนี้ เป็นการเก็บรวบรวมข้อมูลที่เกี่ยวข้องกับการวิจัย เช่น ข้อมูลการศึกษาของนักเรียน ข้อมูลเกี่ยวกับกลุ่มสาระ ข้อมูลเสริมเพิ่มเติมที่อาจจะมีผลต่อการพยากรณ์ โดยมีการรวบรวมข้อมูลดังนี้

1.1 ข้อมูลนำเข้า เนื่องจากงานวิจัยของ Wongsakoonkan *et al.* (2021) ระบุว่าผลสัมฤทธิ์ทางการเรียนในระดับมัธยมศึกษาตอนปลายมีความสัมพันธ์ในทิศทางบวกกับผลสัมฤทธิ์ในมหาวิทยาลัยอย่างมีนัยสำคัญ อีกทั้งหลักสูตรต่าง ๆ ในคณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี มีการใช้ความรู้จากรายวิชาต่าง ๆ ในกลุ่มวิทยาศาสตร์ สังคมศาสตร์ คณิตศาสตร์ วิทยาศาสตร์สุขภาพ และกลุ่มเทคโนโลยี เป็นพื้นฐานในการศึกษาต่อในแต่ละสาขาที่แตกต่างกัน นอกจากนี้การประเมินผลการเรียนในแต่ละโรงเรียนก็มีความแตกต่างกันด้วย ดังนั้นในงานวิจัยนี้จึงใช้ข้อมูลนำเข้าประเภทคะแนนสอบ ซึ่งประกอบด้วย เกรดเฉลี่ยในโรงเรียน คะแนนใน 8 กลุ่มสาระ และผลการเรียนในระดับชั้นปีที่ 1 ในมหาวิทยาลัย มาใช้เพื่อช่วยทำนายผลสัมฤทธิ์ของผู้เรียนในระดับมหาวิทยาลัย

โดยข้อมูลนำเข้ามีจำนวนปัจจัยทั้งหมด 11 ปัจจัย ประกอบด้วยข้อมูลสาขาวิชา และข้อมูลคะแนนจากโรงเรียน ได้แก่ ข้อมูลสาขาวิชา ข้อมูลคะแนนจากโรงเรียน เช่น เกรดเฉลี่ย ข้อมูลคะแนนใน 8 กลุ่มสาระ (1. กลุ่มภาษาไทย 2. คณิตศาสตร์ 3. วิทยาศาสตร์ 4. สังคมศึกษา ศาสนา และวัฒนธรรม 5. สุขศึกษาและพลศึกษา 6. ศิลปะ 7. การงานอาชีพและเทคโนโลยี และ 8. ภาษาต่างประเทศ) และข้อมูลคะแนนเกรดเฉลี่ยปี 1 ในมหาวิทยาลัย

1.2 ข้อมูลเป้าหมาย เป็นสถานะการศึกษาเมื่อสิ้นสุดภาคเรียนที่ 2 ในระดับมหาวิทยาลัย ซึ่งประกอบด้วย 2 สถานะ ได้แก่ ผ่าน (ได้ผลการเรียนมากกว่า 2.00) และตก (ได้ผลการเรียนน้อยกว่าหรือเท่ากับ 2.00 และสถานะอื่น ๆ เช่น ลาออก ย้ายสาขา)

2. **ทำความสะอาดข้อมูล (Data Cleansing)** หลังจากการเก็บข้อมูลเรียบร้อยแล้ว จะมีการทำความสะอาดข้อมูล โดยการตรวจสอบความสมบูรณ์และความถูกต้องของข้อมูล เช่น ตรวจสอบข้อมูลที่ขาดหายไป ตรวจสอบข้อมูลซ้ำซ้อน และการจัดระเบียบข้อมูลให้อยู่ในรูปแบบที่เหมาะสมสำหรับการนำเข้าแบบจำลอง

3. **คัดเลือกคุณลักษณะและแปลงข้อมูล (Feature Selection and Feature Engineering)** ในขั้นตอนสุดท้าย เป็นการคัดเลือกคุณลักษณะที่มีผลต่อการพยากรณ์และนำเข้าสู่ระบบ ซึ่งจากการพิจารณาข้อมูลที่รวบรวมมา งานวิจัยนี้เลือกใช้ทั้ง 11 ปัจจัย และการทำการแปลงข้อมูลให้อยู่ในรูปแบบที่เหมาะสมสำหรับการพยากรณ์ต่อไป

ตารางที่ 1 รายละเอียดข้อมูลหลังจากทำการเตรียมข้อมูล

ลำดับ	แอตทริบิวต์	ชนิดข้อมูล	รายละเอียด	ค่าที่เป็นไปได้
1	branch	ตัวอักษร	สาขาวิชา	{DSSI, ICT, CHEM, BIO}
2	admission_grade	ตัวเลข	เกรดเฉลี่ยรับเข้า	{very poor, poor, medium, good, very good, excellent} very poor = เกรด < 1.50, poor = เกรด 1.50 - 1.99, medium = เกรด 2.00 - 2.49, good = เกรด 2.50 -2.99, very good = เกรด 3.00 - 3.49, excellent = เกรด 3.50 – 4.00
3	gpa_year_1	ตัวเลข	เกรดเฉลี่ยชั้นปี 1	
4	thai	ตัวเลข	เกรดวิชาภาษาไทย	
5	math	ตัวเลข	เกรดวิชาคณิตศาสตร์	
6	sci	ตัวเลข	เกรดวิชาวิทยาศาสตร์	
7	society	ตัวเลข	เกรดวิชาสังคมศึกษา ศาสนา และวัฒนธรรม	
8	hygiene	ตัวเลข	เกรดวิชาสุขศึกษา และพลศึกษา	
9	art	ตัวเลข	เกรดวิชาศิลปะ	
10	career	ตัวเลข	เกรดวิชาการงานอาชีพ และเทคโนโลยี	
11	languages	ตัวเลข	เกรดวิชาภาษาต่างประเทศ	
12	status	ตัวอักษร	สถานะผลการเรียนเมื่อสิ้นสุดภาคเรียนที่ 2 ในระดับมหาวิทยาลัย	{Pass, Fail} โดย Pass คือมีผลการเรียนมากกว่า 2.00 และ Fail คือสถานะอื่นๆ เช่น มีผลการเรียนน้อยกว่าหรือเท่ากับ 2.00 หรือ ลาออก ย้ายสาขา

การพัฒนาแบบจำลอง (Model development)

1. สร้างแบบจำลอง ทีมวิจัยได้พัฒนาคำสั่ง create-model เพื่อสร้างแบบจำลองโดยใช้รูปแบบการสร้างคำสั่งเสริมบน จังโก้ ทำให้สามารถใช้คำสั่งเพื่อเรียกใช้ชุดข้อมูลที่จัดเก็บไว้ในตารางของระบบจัดการฐานข้อมูลโพสต์เกรสคิวเอล โดยใช้เงื่อนไขการกรองข้อมูลตามสาขาที่ต้องการแล้วแยกตัวแปรต้นและตัวแปรตามเพื่อผ่านเข้าขั้นตอนการแปลงข้อมูล (Pre-processing) จึงสามารถป้อนข้อมูลให้แบบจำลองป่าสุ่ม ดังแสดงในภาพที่ 3

```

1 data = TrainingData.objects.filter(
2     Q(branch_id__contains=settings.BRANCHES[0])
3     | Q(branch_id__contains=settings.BRANCHES[1])
4 ).values()
5 df = pd.DataFrame(data)
6 features = settings.FEATURES
7 target = settings.TARGET
8 X = df[features]
9 y = df[target]
10 pipeline = Pipeline(steps=[ ('preprocessor', ColumnTransformer(
11     remainder='passthrough',
12     transformers=[('categories',
13         OneHotEncoder(handle_unknown='ignore'), features)])),
14     ('trainer', RandomForestClassifier(
15         criterion='gini',
16         n_estimators=100,
17         random_state=123,
18         class_weight={'status': 'balanced'},
19         max_depth=5))
20     ] )
21

```

ภาพที่ 3 คำสั่งกรองข้อมูล แยกตัวแปร แปลงข้อมูลและสร้างแบบจำลอง

2. การฝึกและการทดสอบแบบจำลอง หลังจากการสร้างแบบจำลองเสร็จสิ้น จะส่งข้อมูลที่แปลงแล้วให้กับแบบจำลองเพื่อทำการฝึกโดยใช้คำสั่ง fit ก่อนจะทำการทดสอบแบบจำลอง โดยแบ่งข้อมูลเป็นชุดฝึกและชุดทดสอบเพื่อวัดประสิทธิภาพของแบบจำลองด้วยการใช้ 10-Fold cross-validation แล้วแสดงค่าเฉลี่ยของคะแนนความแม่นยำทั้ง 10 รอบการทดลอง เพื่อป้องกันการโอเวอร์ฟิตติ้ง หรืออันเดอร์ฟิตติ้ง (Underfitting) ของแบบจำลอง ตามคำสั่งที่แสดงในภาพที่ 4

```

1 model = pipeline.fit(X, y)
2
3 cv_result = cross_validate(pipeline, X, y, cv=10)
4 avg_accuracy = cv_result['test_score'].mean()
5

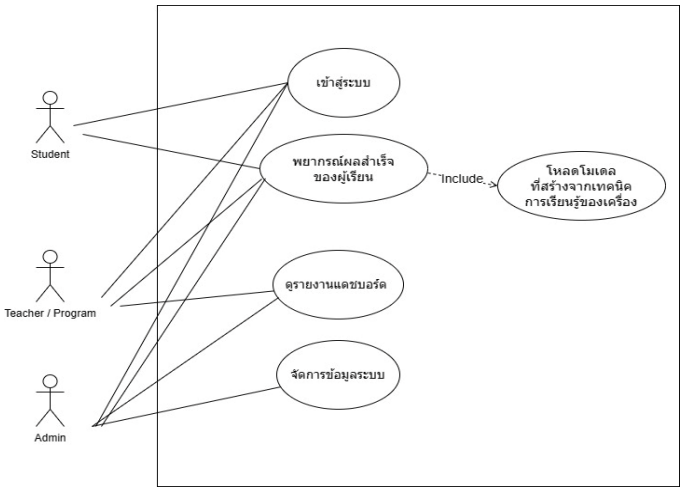
```

ภาพที่ 4 คำสั่งฝึกและการทดสอบ

การพัฒนาเว็บ (Web application development)

ในขั้นตอนนี้จะเป็นการนำแบบจำลองการพยากรณ์ผลที่ได้ในขั้นตอนก่อนหน้ามาประยุกต์และใช้งานภายในสภาพแวดล้อมของเว็บแอปพลิเคชันเพื่อต่อยอดและแทนองค์ความรู้ในระบบ โดยจะใช้จังโก้เฟรมเวิร์กมาช่วยจัดการและประมวลผลผลลัพธ์ในระบบ และใช้โพสต์เกรสคิวเอลเพื่อช่วยจัดการฐานข้อมูล แบบจำลองและข้อมูลผู้ใช้ต่าง ๆ ซึ่งสามารถสรุปการทำงานได้ดังนี้

1. การวางแผนและออกแบบระบบ (Planning and Design) เป็นการวิเคราะห์และออกแบบฟังก์ชันการทำงานให้กับระบบ ซึ่งสามารถสรุปการทำงานของระบบได้ดังแผนภาพยูสเคส (Use Case Diagram) ดังภาพที่ 5



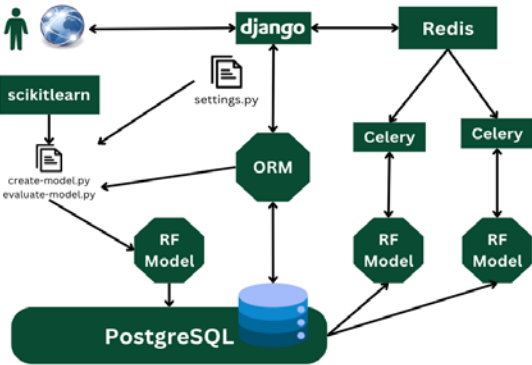
ภาพที่ 5 สรุปลักษณะการทำงานของระบบผ่านแผนภาพยูสเคส

จากภาพที่ 5 แสดงฟังก์ชันการทำงานของระบบ โดยสามารถสรุปและอธิบายฟังก์ชันการทำงานได้ดังตารางที่ 2

ตารางที่ 2 รายละเอียดฟังก์ชันการทำงาน

ฟังก์ชัน	รายละเอียดการทำงาน
เข้าสู่ระบบ	ตรวจสอบสิทธิ์ผู้ใช้ โดยจะมีการเรียกใช้การยืนยันตน (LDAP Authentication) ผ่านบริการผู้ใช้ (User) และรหัสผ่าน (Password) ของมหาวิทยาลัย
พยากรณ์ผลสำเร็จของผู้เรียน	พยากรณ์ผลสำเร็จของผู้เรียน โดยจะตรวจสอบจากข้อมูลนำเข้า 11 ปัจจัย และนำมาวิเคราะห์ประมวลผล กับแบบจำลองการพยากรณ์ผล เพื่อแสดงสถานะการสำเร็จการศึกษา
ดูรายงานแดชบอร์ด	นำผลการพยากรณ์มาแสดงรายงานแดชบอร์ดเพื่อแจ้งเตือนผู้เรียนและผู้สอนในสาขาวิชา ในการติดตามและประเมินความเสี่ยงของผู้เรียน
จัดการข้อมูลในระบบ	จัดการค่าข้อมูลแบบจำลอง ข้อมูลผู้ใช้ และค่าต่าง ๆ ในระบบ

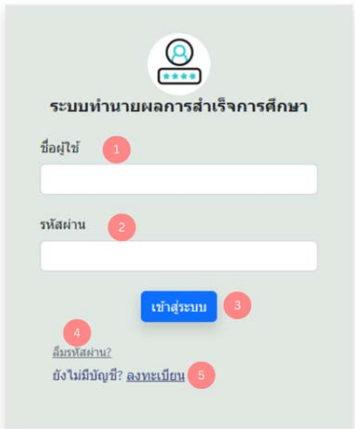
2. การออกแบบสถาปัตยกรรมในการทำงานร่วมกัน โดยจะใช้จังก์เฟรมเวิร์กซึ่งทำงานบนพื้นฐานของ MVT (Model-View-Template) เข้ามาพัฒนาเว็บแอปพลิเคชันตามฟังก์ชันที่กำหนดในแผนภาพยูสเคส ซึ่งจะมีการนำแบบจำลองการพยากรณ์ที่ได้จากขั้นตอนการพัฒนาแบบจำลองข้างต้นมาใช้งานร่วมด้วย โดยเมื่อผู้ใช้ติดต่อและร้องขอข้อมูลไปยังส่วนจังก์ เช่น การพยากรณ์ผลการเรียนรู้ หรือการขอดูรายงานแดชบอร์ดในฝั่งเซิร์ฟเวอร์จะมีการส่งรายการไปยังเรดิส (Redis) จากนั้นรายการจะถูกแบ่งเป็นงานย่อยแบบขนานโดยเซเลอรี (Celery) ซึ่งในขั้นตอนนี้จะมีการโหลดแบบจำลองการพยากรณ์ในฐานข้อมูลเพื่อหาผลลัพธ์สถานะการเรียนรู้ของผู้เรียนจากข้อมูลนำเข้าที่สอดคล้องกัน ก่อนส่งผลลัพธ์ที่ได้กลับไปยังส่วนแสดงผลในเบราว์เซอร์ในฝั่งไคลเอนต์ (Client) เพื่อแสดงผลการพยากรณ์การเรียนรู้ หรือรายงานแดชบอร์ดที่เกี่ยวข้องต่อไป ซึ่งแสดงขั้นตอนได้ดังภาพที่ 6



ภาพที่ 6 การออกแบบสถาปัตยกรรมในการทำงานร่วมกันในระบบ

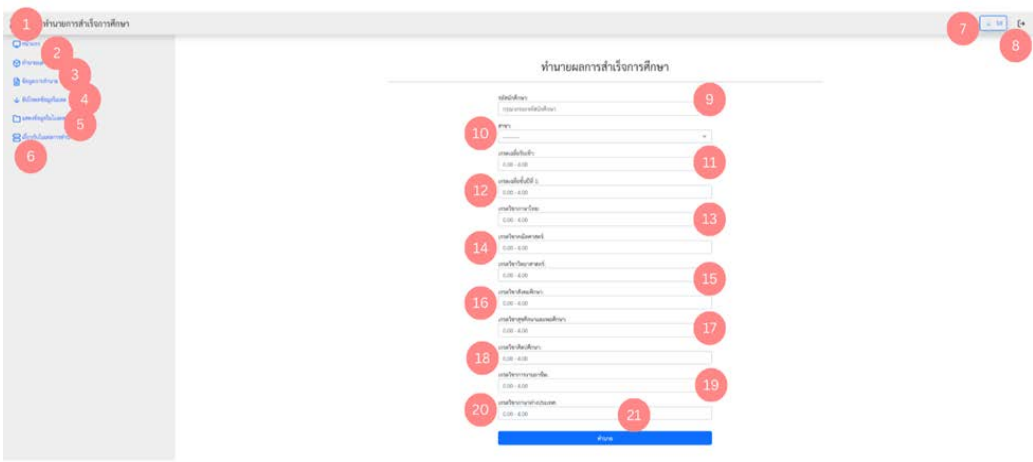
3. การสร้างเว็บแอปพลิเคชัน โดยนำกลไกการทำงานข้างต้นมาจัดการและพัฒนาระบบ ซึ่งสรุปการทำงานได้ 4 ส่วนหลัก ดังนี้

ส่วนที่ 1: ส่วนเข้าสู่ระบบ เป็นส่วนตรวจสอบสิทธิ์การใช้งานระบบของผู้ใช้ ซึ่งส่วนนี้จะมีการเรียกบริการของ Lightweight Directory Access Protocol (LDAP) Authentication ของมหาวิทยาลัยเพื่อตรวจสอบสิทธิ์ของผู้ใช้งานในระบบ ดังแสดงในภาพที่ 7

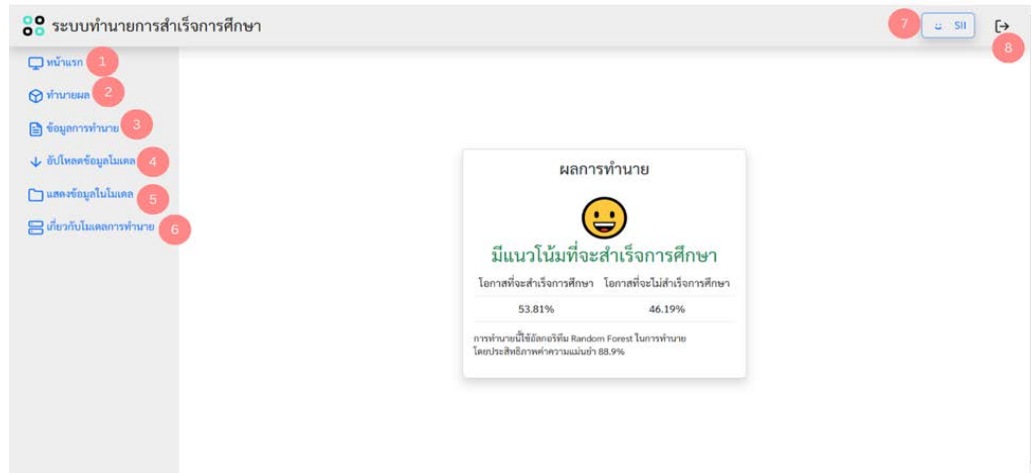


ภาพที่ 7 ส่วนเข้าสู่ระบบ

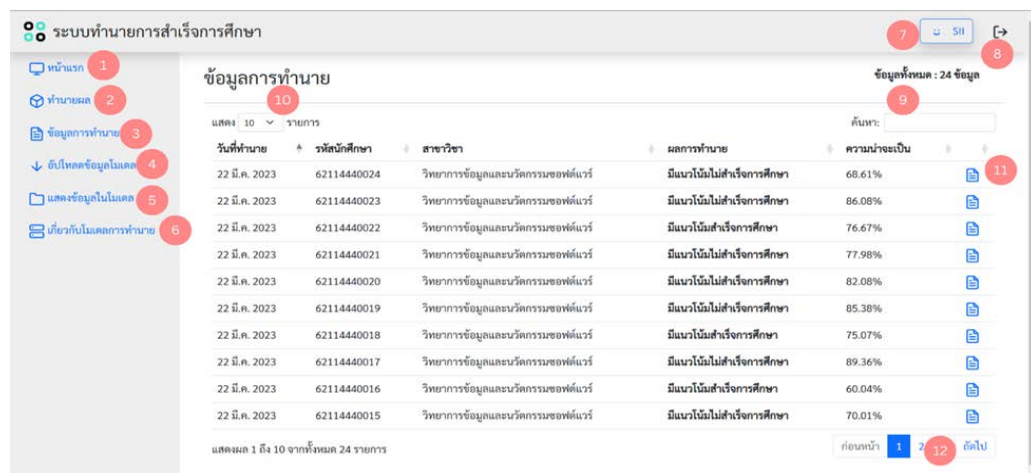
ส่วนที่ 2: ส่วนพยากรณ์ผลการเรียนรู้ของผู้เรียน เป็นส่วนที่ทำหน้าที่พยากรณ์ผลสำเร็จของผู้เรียน โดยนำข้อมูลเข้าแต่ละรายการประกอบด้วย 11 ปัจจัย เพื่อไปประมวลผลผ่านแบบจำลองการพยากรณ์ ดังแสดงในภาพที่ 8 จากนั้นแสดงผลการพยากรณ์ซึ่งสามารถดำเนินการได้ 2 รูปแบบ แบบแรก คือ การพยากรณ์แบบรายบุคคล ที่ระบุข้อมูลและพยากรณ์ประสิทธิภาพของผู้เรียนแบบรายบุคคลได้ ดังแสดงในภาพที่ 9 ในขณะที่แบบที่ 2 คือ การพยากรณ์แบบรายกลุ่ม ที่สามารถอัปโหลดไฟล์ข้อมูลผู้เรียนแบบหลายคนเข้าไปในระบบ ก่อนจะสรุปผลการพยากรณ์ความสำเร็จในรูปแบบตาราง ดังแสดงในภาพที่ 10



ภาพที่ 8 การนำข้อมูลเข้าระบบเพื่อประมวลผลผ่านแบบจำลองการพยากรณ์

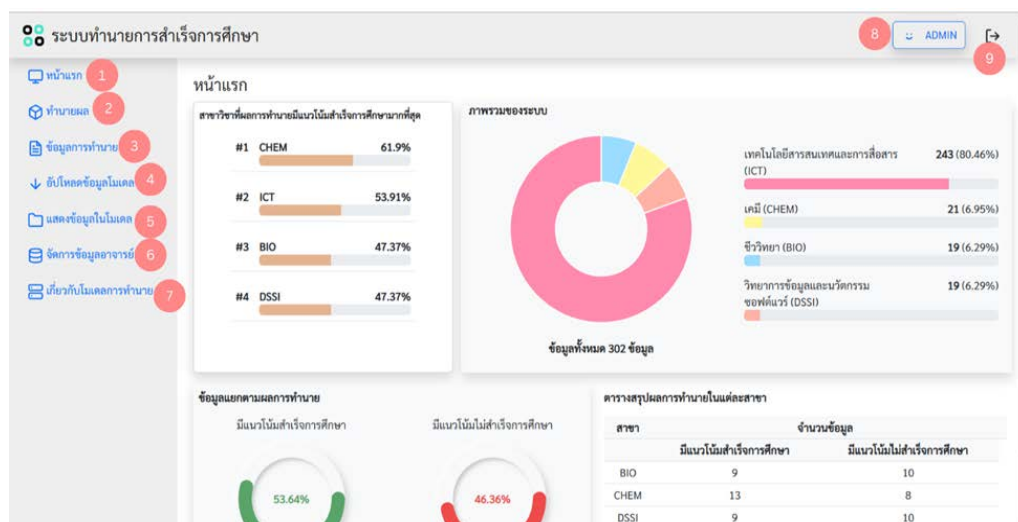


ภาพที่ 9 การพยากรณ์แบบรายบุคคล



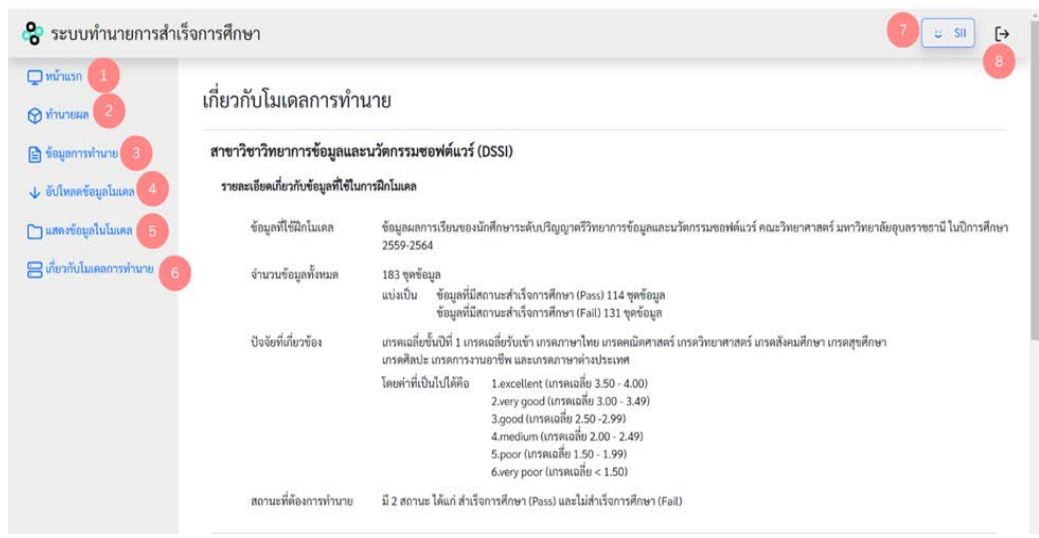
ภาพที่ 10 การพยากรณ์แบบรายกลุ่ม

ส่วนที่ 3: ส่วนรายงานผลแดชบอร์ด เป็นส่วนงานที่ทำหน้าที่แสดงผลรายงานผลการพยากรณ์ ปัจจัยที่ส่งผลต่อความสำเร็จของผู้เรียนในระบบจำแนกตามสาขาวิชา โดยจะแสดงอยู่ในรูปแบบของกราฟต่าง ๆ ซึ่งการเข้าถึงข้อมูลจะเป็นไปตามสิทธิ์ของผู้ใช้ที่แตกต่างกัน ดังแสดงในภาพที่ 11



ภาพที่ 11 ส่วนรายงานผลแดชบอร์ด

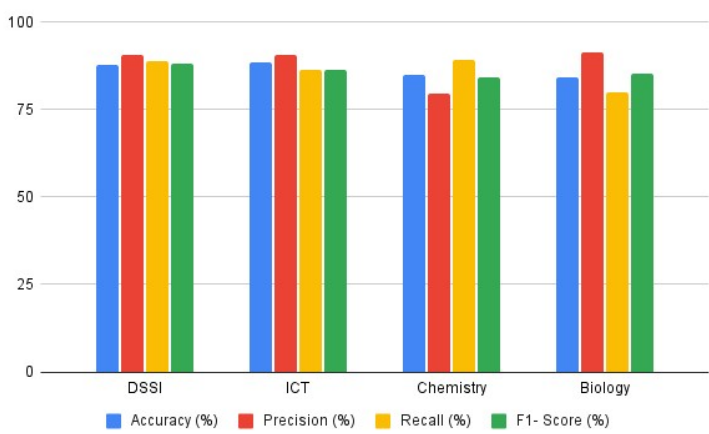
ส่วนที่ 4: ส่วนจัดการค่าต่าง ๆ ในระบบ เป็นการจัดการแก้ไขข้อมูลผู้ใช้ ข้อมูลแบบจำลองและค่าต่าง ๆ ในระบบ ดังแสดงในภาพที่ 12



ภาพที่ 12 ส่วนจัดการค่าต่าง ๆ ในระบบ

ผลการวิจัยและอภิปรายผล

ในการวัดประสิทธิภาพการทำงาน จะดำเนินการ 2 ส่วน ได้แก่ การวัดประสิทธิภาพของแบบจำลองการพยากรณ์ผล โดยใช้ตัวชี้วัดสำหรับการประเมินประสิทธิภาพเพื่อประเมินความแม่นยำในการพยากรณ์ผลของแบบจำลองด้วย อัลกอริทึมป่าสุ่ม ได้แก่ ค่า Precision ค่า Accuracy ค่า Recall และ ค่า F1-measure เพื่อปรับและประเมินแบบจำลองให้มีประสิทธิภาพที่สูงสุดเท่าที่เป็นไปได้ โดยประสิทธิภาพของแบบจำลองเมื่อทดสอบแยกตามสาขาวิชา ได้ผลการทดสอบดังตารางที่ 3 และแสดงการเปรียบเทียบผลการทดสอบได้ดังภาพที่ 13



ภาพที่ 13 เปรียบเทียบผลการทดสอบประสิทธิภาพแบบจำลองแยกตามสาขาวิชา

จากตารางที่ 3 พบว่า อัลกอริทึมป่าสุ่มสามารถสร้างแบบจำลองการพยากรณ์ผลที่ให้ความแม่นยำและความถูกต้อง โดยมีค่า F1-measure สูงเท่ากับ 84% - 88% ซึ่งเป็นผลจากการจัดเตรียมข้อมูล การสร้างและทดสอบแบบจำลอง รวมทั้งการใช้อัลกอริทึมป่าสุ่มในการสร้างแบบจำลองที่มีประสิทธิภาพและเหมาะสมกับบริบทข้อมูลของนักศึกษาคณะวิทยาศาสตร์ทั้ง 4 สาขาวิชา ซึ่งสอดคล้องกับผลของงานวิจัยของ Hailu and Murty (2019) ที่ระบุว่าเทคนิคป่าสุ่มสามารถสร้างแบบจำลองที่ใช้พยากรณ์ประสิทธิภาพการเรียนรู้ของผู้เรียน ซึ่งให้ความแม่นยำสูงถึง 95% อย่างไรก็ตาม การจัดการเวลาในการสร้างแบบจำลองด้วยเทคนิคป่าสุ่มยังมีการดำเนินการที่สูงอยู่ (Time-consuming process)

ตารางที่ 3 ผลการทดสอบประสิทธิภาพแบบจำลอง

สาขาวิชา	Accuracy (%)	Precision (%)	Recall (%)	F1- Score (%)
DSSI	87.70 %	90.64 %	88.77 %	88.02 %
ICT	88.48 %	90.59 %	86.31 %	86.18 %
Chemistry	84.88 %	79.63 %	88.97 %	84.04 %
Biology	84.12 %	91.27 %	79.86 %	85.18 %

การประเมินและทดสอบการทำงานของระบบ

ในการทดสอบการทำงานของเว็บแอปพลิเคชัน จะทดสอบการทำงานของฟังก์ชันในรูปแบบการทดสอบกล่องดำ (Black box testing) เพื่อประเมินว่าฟังก์ชันสามารถทำงาน ประมวลผล และสร้างผลลัพธ์ได้ตามข้อมูลนำเข้าได้ตามวัตถุประสงค์ของฟังก์ชันนั้น ๆ โดยสรุปได้ดังตารางที่ 4

ตารางที่ 4 ผลการทดสอบการทำงานของระบบ

แผนการทดสอบ (Test Case)	ข้อมูลนำเข้า	ผลลัพธ์ที่คาดหวัง	ผลสรุป
เข้าสู่ระบบ	ข้อมูลผู้ใช้และรหัสผ่าน	เรียกใช้บริการ LDAP Authentication และได้รับผล ข้อมูลเป็นประเภทหรือสถานะผู้ใช้	✓
พยากรณ์ผลการสำเร็จ การศึกษา	ข้อมูลนำเข้าของผู้ใช้ จำแนกตามสาขา	ผลการพยากรณ์ความสำเร็จของผู้เรียนที่สอดคล้องกับ สาขาวิชา	✓
รายงานสรุปผลผู้เรียน	ข้อมูลสาขา หรือชั้นปีที่ ต้องการตรวจสอบ	รายงานที่แสดงผลการพยากรณ์จำแนกตามชั้นปี และ สาขาวิชาที่ถูกต้อง	✓
การจัดการข้อมูลระบบ	ข้อมูลผู้ใช้ ข้อมูล แบบจำลอง	ข้อมูลผู้ใช้และข้อมูลแบบจำลองที่ถูกอัปเดตแล้ว	✓

จากตารางที่ 4 ผลการทดสอบการทำงานของระบบ สรุปได้ว่า ระเบียบวิธีที่เสนอซึ่งประกอบด้วยการจัดเตรียมข้อมูล พัฒนาแบบจำลอง ซึ่งนำมาใช้งานและทำงานภายใต้สภาพแวดล้อมของจังก์ท์เฟรมเวิร์ก และนำเสนอผลลัพธ์ผ่านเว็บแอปพลิเคชันสามารถทำงานได้ ซึ่งสอดคล้องกับงานวิจัยของ Li and Zhao (2020) ที่พัฒนาแบบจำลองการพยากรณ์ผลการตกออกและ นำผลลัพธ์มาใช้งานในเว็บแอปพลิเคชัน ซึ่งผลการทำงานคือระบบสามารถทำงานร่วมกันและประมวลผลผลลัพธ์ได้ถูกต้อง

สรุปผลการวิจัยและข้อเสนอแนะจากการวิจัย

งานวิจัยนี้นำเสนอระบบพยากรณ์การสำเร็จการศึกษาของผู้เรียนด้วยเทคนิคป่าสุ่ม ซึ่งพัฒนาขึ้นเพื่อศึกษาและสร้างแบบจำลองการพยากรณ์การสำเร็จการศึกษาของผู้เรียนด้วยเทคนิคป่าสุ่ม โดยได้รวบรวมและวิเคราะห์ข้อมูลจากนักศึกษาคณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี จำนวน 4 สาขาวิชา ของนักศึกษาจำนวน 1,336 คน ที่ศึกษาระหว่างปีการศึกษา 2560-2564 โดยใช้ข้อมูล 11 ปัจจัย ได้แก่ ข้อมูลสาขาวิชา ข้อมูลคะแนนจากโรงเรียน ได้แก่ เกรดเฉลี่ย ข้อมูลคะแนนใน 8 กลุ่มสาระ (1. กลุ่มภาษาไทย 2. คณิตศาสตร์ 3. วิทยาศาสตร์ 4. สังคมศึกษา ศาสนา และวัฒนธรรม 5. สุขศึกษาและพลศึกษา 6. ศิลปะ 7. การงานอาชีพและเทคโนโลยี และ 8. ภาษาต่างประเทศ) และข้อมูลคะแนนเกรดเฉลี่ยปี 1 ในมหาวิทยาลัย ซึ่งผลลัพธ์จะได้แบบจำลองการพยากรณ์ผลความสำเร็จของผู้เรียนที่สามารถนำไปใช้แทนความรู้เพื่อสร้างเครื่องมือพยากรณ์ผลและรายงานในรูปแบบเว็บแอปพลิเคชันที่พัฒนาด้วยจังก์ท์เฟรมเวิร์กและจัดเก็บข้อมูลในฐานข้อมูลโพสต์เกรสควิเอล ซึ่งจะช่วยสนับสนุนผู้ใช้งานในระดับต่าง ๆ ได้ เช่น ผู้เรียนและผู้สอนในสาขาวิชา สามารถใช้เป็นเครื่องมือในการพยากรณ์ผลความสำเร็จ ดูรายงานแดชบอร์ดเพื่อประเมินความเสี่ยง หรือติดตามสถานะผู้เรียนในระยะแรกได้ เพิ่มอัตราคงอยู่ รวมทั้งวางแผนการจัดการเรียนการสอนในสาขาวิชาได้ นอกจากนี้ระเบียบวิธีที่นำเสนอสามารถนำไปใช้เป็นต้นแบบหรือต่อยอดในสาขาวิชาหรือคณะที่มีบริบทของข้อมูลที่คล้ายกันได้ ซึ่งจะเป็นประโยชน์ต่อการพัฒนาคุณภาพการศึกษาต่อไป

แนวทางการพัฒนาในอนาคต ทีมวิจัยจะมีการคัดกรองกลุ่มนักศึกษาตัวอย่างและนำข้อมูลส่วนอื่น ๆ ของผู้เรียน และสาขาวิชาอื่น ๆ มาวิเคราะห์เพิ่มเติม เช่น ข้อมูลส่วนตัว ข้อมูลครอบครัว หรือพฤติกรรมการใช้งานรายวิชาในระบบการจัดการ

เรียนรู้ (Learning Management System: LMS) รวมถึงปัจจัยที่เกิดจากผลกระทบภายนอกในสถานการณ์ไม่ปกติได้แก่ การแพร่ระบาดของโรคติดต่อร้ายแรง อุบัติภัยร้ายแรง หรืออื่น ๆ ซึ่งส่งผลกระทบโดยตรงต่อสภาพความเป็นอยู่ รายได้ รวมทั้งสถานะทางการเงินของผู้เรียน นอกจากนี้จะประยุกต์อัลกอริทึมอื่น ๆ ทั้งในส่วนของเทคนิคการเรียนรู้ของเครื่องแบบดั้งเดิม (Traditional Machine Learning) และเทคนิคการเรียนรู้เชิงลึก (Deep Learning) เพื่อสร้างแบบจำลองการพยากรณ์ผลที่มีประสิทธิภาพและมีความแม่นยำในการพยากรณ์ผลมากยิ่งขึ้น

กิตติกรรมประกาศ

ขอขอบคุณ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี สำหรับการสนับสนุนการดำเนินงานวิจัยในครั้งนี้

References

- Alboaneen, D., Almelihi, M., Alsubaie, R., Alghamdi, R., Alshehri, L. and Alharthi, R. (2022). Development of a web-based prediction system for students' academic performance. **Data**, 7(2), 21.
- Amazon Web Services. (2024). What is Django?. Retrieved February 22, 2024, from AWS: <https://aws.amazon.com/th/what-is/django/>
- Brownlee, J. (2020). 4 Types of Classification Tasks in Machine Learning. Retrieved February 22, 2024, from **Machine Learning Mastery**: https://machinelearningmastery.com/types-of-classification-in-machine-learning/?fbclid=IwAR2gbsHUbY_YUdt4l4RBvHAQyze3YlMYVsnyjq99ZJW36lb5qjMISvYcZ2A
- Data Innovation and Governance Institute. (2022). Understand in 5 minutes! What is a Classification Model? (in Thai). Retrieved February 22, 2024, from **DIGI**: <https://digi.data.go.th/blog/what-is-classification-model/>
- Hailu, B. and Murty, M. N. (2019). Performance analysis and prediction of students using random forest algorithm. **International Journal of Engineering Research and Technology**, 12(2), 129-136.
- Han, J., Kamber, M. and Pei, J. (2012). **Data Mining: Concepts and Techniques**. Melbourne: Elsevier Inc.
- Jamjumrat, C. (2022). What Is A Decision Tree?. Retrieved 22 February 2024, from **borntoDev**: <https://www.borntodev.com/2022/09/15/รู้จักกับ-decision-tree/>
- Kaensar, C. and Wongnin, W. (2023). Predicting new student performances and identifying important attributes of admission data using machine learning techniques with hyperparameter tuning. **Eurasia Journal of Mathematics, Science and Technology Education**, 19(12), em2369.
- Kaensar, C. and Wongnin, W. (2024). Enhancing prediction of student learning success in higher education using deep learning (in Thai). **Journal of Science and Science Education**, 7(1), 21–36.
- Kongchai, P., Hiranpongsin, S. and Ditcharoen, N. (2024). Creating a prediction model for prospective university student admissions during the COVID-19 pandemic situation: A case study of the faculty of science at Ubon Ratchathani university (in Thai). **The Journal of King Mongkut's University of Technology North Bangkok**, 34(2), 1–10.
- Li, H. and Zhao, C. (2020). Research on the Application of Machine Learning in Web Application Development. In **Proceedings of Smart Grid and Electrical Automation International Conference 2020**, (pp. 478-482). Zhangjiajie: China.
- Mekha, P., Musikong, P., Palakong, N., Pramokchon, P. and Kasemsumran, P. (2023). Performance comparison of image classification models for corn leaf disease (in Thai). **Maejo Information Technology and Innovation Journal**, 9(2), 1–16.
- Nachouki, M., Elfadil, A.M., Mehdi, R. and Naaj, M (2023). Student course grade prediction using the random forest algorithm: Analysis of predictors' importance. **Trends in Neuroscience and Education**, 33, 100214.
- Naresuan University Publishing House. (2023). Data Mining (in Thai). Retrieved February 22, 2024. from **NUPRESS**: <https://www.nupress.grad.nu.ac.th/data-mining/#3--classification>

- Riansut, W. (2016). Factors influencing on grade point average of thaksin university's undergraduate students (in Thai). **Science Journal of Phetchaburi Rajabhat University**, 13(2), 23-38.
- Shalev-Shwartz, S. and Ben-David, S. (2014). **Understanding Machine Learning**. Cambridge: Cambridge University Press.
- Shou, Z., Xie, M., Mo, J. and Zhang, H (2024). Predicting student performance in online learning: a multidimensional time-series data analysis approach. **Applied Sciences**. 14(6), 1-16.
- Soontranon, N. (2023). Evaluation Metrics for Classification Model (in Thai). Retrieved February 22, 2024, from **NERD**: <https://www.nerd-data.com/classification-metrics/>
- Valverde-Berrocso, J., Acevedo-Borrega J. and Cerezo-Pizarro, M. (2022), Educational technology and student performance: a systematic review. **Frontiers in Education**, 7(1), 1-13.
- Wongsakoonkan, W., Ju-Ngam, B., Komuttiban, P., Klunbut, P. and Alongkorn, V. (2021). A comparison of learning achievement of the undergraduate students admitted through direct examination and admission case study: department of occupational health and safety, faculty of science and technology: Valaya Alongkorn Rajabhat University Under the Royal Patronage (in Thai). **Valaya Alongkorn Review Journal**, 11(1), 1-12.