การวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาโดยใช้เทคนิคเหมืองข้อมูล กรณีศึกษา นักศึกษาระดับปริญญาตรี คณะบริหารธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

Analysis on Factors Affecting Student Termination Using Data Mining Techniques A Case of Undergraduate Students, Faculty of Business Administration, Rajamangala University of Technology Thanyaburi

ชาลี จิตรีผ่อง[1] และ สลิตตา สาริบุตร[2*]
(Chalee Jittreephong[1] and Salitta Saributr[2*])

[1-2]คณะบริหารธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี กรุงเทพมหานคร
[1-2]Faculty of Business Administration,
Rajamangala University of Technology Thanyaburi, Bangkok, Thailand
*Corresponding author E-mail: salitta_s@rmutt.ac.th

Article history
Received 26 February 2025 Revised 28 April 2025
Accepted 2 May 2025 SIMILARITY INDEX = 5.44 %

บทคัดย่อ

การพ้นสภาพของนักศึกษาในระดับอุดมศึกษาเป็นประเด็นสำคัญที่ส่งผลต่อประสิทธิภาพการบริหารจัดการของสถาบันการศึกษาและคุณภาพของบัณฑิต งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะบริหารธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี และเพื่อพัฒนาแบบจำลองพยากรณ์ความเสี่ยงโดยประยุกต์ใช้เทคนิคเหมืองข้อมูล

ข้อมูลที่ใช้เป็นข้อมูลทุติยภูมิของนักศึกษา 5,034 คน ที่เข้าศึกษาระหว่างปีการศึกษา 2562-2565 การวิเคราะห์ดำเนินการตามกระบวนการ KDD (Knowledge Discovery in Databases) ประกอบด้วยการทำความสะอาดข้อมูล การแปลงข้อมูล และการวิเคราะห์ด้วยอัลกอริทึม FP-Growth เพื่อหากฎความสัมพันธ์ร่วมกับการจำแนกประเภทโดยใช้อัลกอริทึม Decision Tree, Naïve Bayes และ k-Nearest Neighbors (KNN) พร้อมประเมินประสิทธิภาพด้วยค่า Accuracy, Precision, Recall, F1-Score และ AUC

ผลการวิจัยพบว่าแบบจำลอง Naïve Bayes มีค่า AUC สูงสุดที่ 1.000 แสดงถึงความสามารถในการจำแนกนักศึกษาที่มีความเสี่ยงได้อย่างแม่นยำ ส่วนแบบจำลอง Decision Tree มีค่า Accuracy สูงสุดที่ร้อยละ 97.45 และมีข้อได้เปรียบด้านการตีความง่ายและนำไปใช้จริง สำหรับปัจจัยสำคัญที่สัมพันธ์กับการพ้นสภาพ ได้แก่ เกรดเฉลี่ยต่ำกว่า 2.00 เกรดต่ำในรายวิชาหลัก การลงทะเบียนเรียนซ้ำ และผลการเรียนในชั้นปีที่ 1 ซึ่งสามารถนำไปใช้พัฒนาระบบแจ้งเตือนล่วงหน้าได้อย่างมีประสิทธิภาพ

คำสำคัญ: การพ้นสภาพนักศึกษา เหมืองข้อมูล ระบบแจ้งเตือนล่วงหน้า ผลสัมฤทธิ์ทางการศึกษา

ABSTRACT

Student attrition in higher education is an important issue that affects the efficiency of university management and the quality of graduates. This research aimed to analyze the factors that influence student attrition among undergraduate students at the Faculty of Business Administration, Rajamangala University of Technology Thanyaburi, and to develop a risk prediction model using data mining techniques.

The study used secondary data from 5,034 students who enrolled between the academic years 2019 and 2022. The analysis followed the Knowledge Discovery in Databases (KDD) process, which included data cleaning, data transformation, and the FP-Growth algorithm to find association rules. Classification models were created using Decision Tree, Naïve Bayes, and k-Nearest Neighbors (KNN), and their performance was evaluated using Accuracy, Precision, Recall, F1-Score, and AUC.

The results showed that the Naïve Bayes model achieved the highest AUC value at 1.000, showing a strong ability to identify students at risk. The Decision Tree model gave the highest accuracy at 97.45% and had the advantage of being easy to interpret and apply. The important factors related to student attrition were GPA lower than 2.00, low grades in core courses, repeated course enrollment, and poor academic performance in the first year. These findings can be applied to improve early warning systems and proactive student advising effectively.

Keywords: Student Attrition, Data Mining, Early Warning System, Academic Performance

1. บทนำ

สถาบันอุดมศึกษาเป็นเสมือนขุมพลังในการผลิตทรัพยากรมนุษย์ที่มีคุณภาพ เพื่อตอบสนองต่อความต้องการของสังคมและภาคอุตสาหกรรม โดยทำหน้าที่เป็นศูนย์กลางในการถ่ายทอดองค์ความรู้ พัฒนาทักษะและปลูกฝังคุณลักษณะสำคัญที่จำเป็นต่อการประกอบอาชีพและการดำรงชีวิตในยุคที่เปลี่ยนแปลงอย่างรวดเร็ว นอกจากนี้สถาบันอุดมศึกษายังมีบทบาทสำคัญในการสร้างนวัตกรรมและงานวิจัยที่ส่งเสริมความก้าวหน้าทางเศรษฐกิจและสังคม อันเป็นรากฐานสำคัญของการพัฒนาประเทศอย่างยั่งยืน (สุกัญญา ทารส และทรงศักดิ์ ภูสีอ่อน, 2563) ซึ่งมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี เป็นหนึ่งในสถาบันอุดมศึกษาที่มุ่งเน้นการพัฒนากำลังคนทางด้านบริหารธุรกิจที่มีทักษะความชำนาญในด้านวิชาชีพในระดับสากล เพื่อให้ได้ทุนมนุษย์ที่มีมูลค่าเพิ่มให้กับประเทศชาติ โดยมีพันธกิจหลัก 4 ด้าน ได้แก่ การจัดการเรียนการสอน การวิจัย การให้บริการทางวิชาการแก่สังคม และการทำนุบำรุงศิลปะและวัฒนธรรม (มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี, 2564) นอกจากนี้ ยังมีการกำหนดตัวบ่งชี้ในการประเมินคุณภาพของการศึกษา ซึ่งหนึ่งในเกณฑ์ที่สำคัญคือจำนวนนักศึกษาที่สำเร็จการศึกษาตามระยะเวลาที่หลักสูตรกำหนดไว้ไม่เกินร้อยละ 80 จากจำนวนนักศึกษาทั้งหมด โดยข้อมูลย้อนหลังในช่วงปีการศึกษา 2562-2565 พบว่า นักศึกษาจำนวนทั้งหมด 5,034 คน สถานะการเป็นนักศึกษา “กำลังศึกษา” จำนวน 4,559 คน หรือคิดเป็นร้อยละ 90.56 ในขณะที่สถานะการเป็นนักศึกษา “พ้นสภาพการเป็นนักศึกษา” จำนวน 475 คน หรือคิดเป็นร้อยละ 9.44 (ระบบบริการการศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี, 2565) ถึงแม้ว่าตัวเลขดังกล่าวจะแสดงถึงคุณภาพของการศึกษาที่ยังคงอยู่ในเกณฑ์ที่ดี แต่ปัญหาการพ้นสภาพของนักศึกษาในระดับอุดมศึกษายังเป็นประเด็นสำคัญที่อาจบั่นทอนศักยภาพบัณฑิตและความยั่งยืนของระบบการศึกษา ดังนั้นมหาวิทยาลัยยังคงมุ่งมั่นที่จะแก้ไขปัญหา เพื่อลดการเสียโอกาสทางการศึกษา และความสูญเปล่าของการลงทุนในการเข้ารับการศึกษา ยิ่งไปกว่านั้นคือการได้ผลิตทรัพยากรมนุษย์ที่สำคัญให้แก่ประเทศชาติ

ในขณะเดียวกัน ยุคปัจจุบันข้อมูลกลายเป็นทรัพยากรอันล้ำค่าของทุกภาคส่วน รวมไปถึงสถาบันทางการศึกษาที่ไม่อาจละเลยศักยภาพของข้อมูลขนาดใหญ่ (Big Data) เพื่อการขับเคลื่อนคุณภาพการศึกษา (Williamson, 2017) นำไปสู่หนึ่งในความท้าทายที่สำคัญของการรักษาอัตราการคงอยู่ของนักศึกษา (Student Retention) และลดอัตราการออกกลางคัน (Dropout Rate) ซึ่งเป็นปัญหาที่ส่งผลกระทบทั้งต่อคุณภาพการศึกษาของมหาวิทยาลัยและอนาคตของนักศึกษา ซึ่งการทำเหมืองข้อมูล (Data Mining) เป็นหนึ่งในเครื่องมือที่ทรงพลังในการวิเคราะห์พฤติกรรมของนักศึกษา และยังสามารถคาดการณ์ได้ว่า นักศึกษาคนใดมีแนวโน้มที่จะประสบความสำเร็จหรือเผชิญกับความท้าทายในเส้นทางการศึกษา (Dutt et al., 2017) เมื่อพิจารณาในแง่มุมของช่องว่างการวิจัยจะพบว่า มีงานวิจัยจำนวนหนึ่งที่ศึกษาปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาในประเทศไทย แต่การใช้เทคนิคเหมืองข้อมูลในการวิเคราะห์จากฐานข้อมูล Thai Journals Online (ThaiJO) มีเพียง 5 รายการเท่านั้น เช่นงานวิจัยของ จีระนันต์ เจริญรัตน์ (2559) ที่พบว่า ปัจจัยที่สำคัญต่อการพ้นสภาพของนักศึกษา ได้แก่ เกรดเฉลี่ยสะสม อาชีพของผู้ปกครอง และสาขาวิชาที่เรียน ในขณะที่งานวิจัยของ ซอและ เกปัน และคณะ (2561) พบว่า ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษา ได้แก่ ผลการเรียนในรายวิชา และผลการเรียนเฉลี่ย แสดงให้เห็นว่า การใช้เทคนิคเหมืองข้อมูลเพื่อวิเคราะห์ข้อมูลภายในของแต่ละสถาบันหรือเฉพาะคณะยังคงมีอยู่อย่างจำกัด โดยเฉพาะอย่างยิ่งในบริบทของคณะบริหารธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี ซึ่งยังไม่มีการศึกษาที่ครอบคลุม

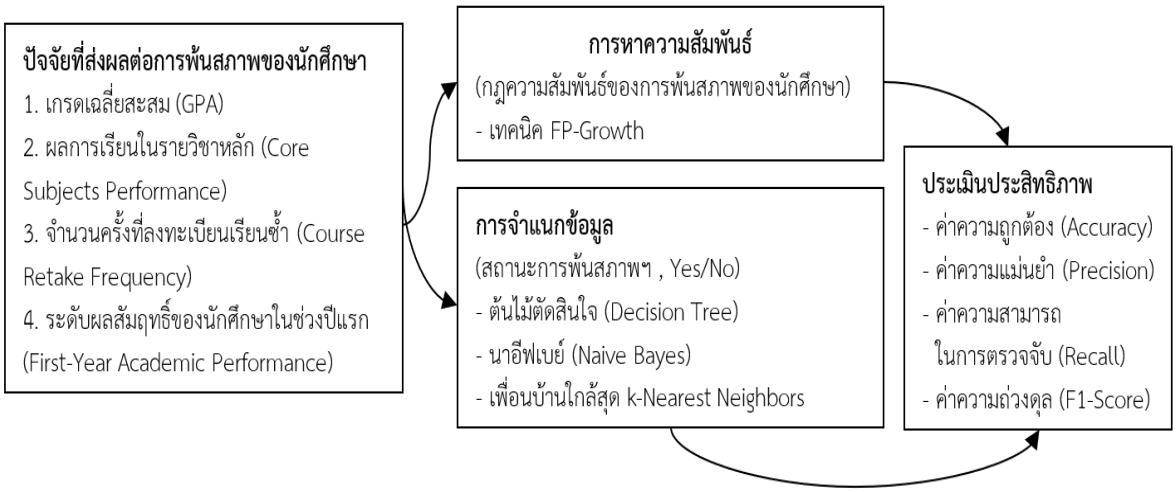
ดังนั้นงานวิจัยนี้จึงมุ่งเน้นการนำเทคนิคเหมืองข้อมูลมาใช้เพื่อพยากรณ์อนาคตของนักศึกษา ผ่านแบบจำลองเชิงคาดการณ์ที่สามารถให้คำตอบแก่ผู้บริหารและอาจารย์ในการพัฒนาแนวทางสนับสนุนนักศึกษา

ที่มีความเสี่ยงได้อย่างทันท่วงที โดยเทคนิคที่นำมาศึกษาในครั้งนี้ประกอบด้วย ต้นไม้ตัดสินใจ (Decision Tree) นาอีฟเบย์ (Naïve Bayes) และ k-Nearest Neighbors (KNN) ซึ่งเป็นเทคนิคที่ได้รับการยอมรับอย่างแพร่หลายในวงการวิทยาศาสตร์ข้อมูล การเปรียบเทียบประสิทธิภาพของเทคนิคเหล่านี้จะช่วยให้สามารถระบุวิธีที่เหมาะสมที่สุดในการพยากรณ์สถานภาพของนักศึกษา อันจะนำไปสู่การปรับปรุงระบบการศึกษา และการพัฒนามาตรการป้องกันการออกกลางคันที่มีประสิทธิภาพมากขึ้น ว่าอะไรคือปัจจัยสาเหตุที่ทำให้นักศึกษาออกกลางคัน เช่น เกรดเฉลี่ยต่ำ การลงทะเบียนเรียนล่าช้า หรือพฤติกรรมบางอย่างที่ยังซ่อนเร้น ซึ่งงานวิจัยนี้จะสามารถเปิดเผยความลับของข้อมูลเพื่อค้นหากุญแจสำคัญที่ช่วยให้มหาวิทยาลัยสามารถรักษานักศึกษาไว้ในระบบได้อย่างมีประสิทธิภาพ

วัตถุประสงค์ของการวิจัย

1. เพื่อวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาโดยใช้เทคนิคเหมืองข้อมูล และพัฒนาโมเดลการทำนายความเสี่ยงในการพ้นสภาพ
2. เพื่อเสนอแนวทางในการลดอัตราการพ้นสภาพของนักศึกษา และพัฒนาระบบสนับสนุนการเรียนการสอนให้มีประสิทธิภาพยิ่งขึ้น

กรอบกระบวนการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล



แผนภาพที่ 1 กรอบกระบวนการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล

แผนภาพนี้แสดงกระบวนการวิเคราะห์ข้อมูลที่ใช้ในงานวิจัย โดยเริ่มจากการระบุปัจจัยที่เกี่ยวข้องกับการพ้นสภาพของนักศึกษา ได้แก่ เกรดเฉลี่ยสะสม ผลการเรียนในรายวิชาหลัก จำนวนครั้งที่ลงทะเบียนเรียนซ้ำ และผลสัมฤทธิ์ในช่วงปีแรก จากนั้นนำข้อมูลที่ผ่านการทำความสะอาดและแปลงให้อยู่ในรูปแบบที่เหมาะสมไปวิเคราะห์ด้วยเทคนิคเหมืองข้อมูล ประกอบด้วย การหากฎความสัมพันธ์ด้วยอัลกอริทึม FP-Growth และการจำแนกประเภทด้วยอัลกอริทึม Decision Tree, Naïve Bayes และ k-Nearest Neighbors โดยผลที่ได้จากกระบวนการนี้จะนำไปสู่การพัฒนาแบบจำลองเชิงพยากรณ์และแนวทางปฏิบัติสำหรับการป้องกันการพ้นสภาพของนักศึกษา

2. การทบทวนวรรณกรรม

2.1. แนวคิดเกี่ยวกับปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษา

งานวิจัยในอดีตจำนวนมากได้วิเคราะห์ถึงปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษา โดยพบว่าปัจจัยหลักมักเกี่ยวข้องกับเกรดเฉลี่ยสะสม (GPA) ผลการเรียนในรายวิชาหลัก (Core Subjects Performance) จำนวนครั้งที่ลงทะเบียนเรียนซ้ำ (Course Retake Frequency) และผลสัมฤทธิ์ของนักศึกษาในช่วงปีแรก (First-Year Academic Performance) ซึ่งเป็นตัวบ่งชี้ที่สำคัญต่อความสำเร็จทางการศึกษา ซึ่งงานวิจัยของ Tinto (2012) และ Pascarella and Terenzini (2005) พบว่า เกรดเฉลี่ยสะสมต่ำมีความสัมพันธ์กับโอกาสพ้นสภาพที่สูงขึ้น เนื่องจากสะท้อนถึงปัญหาทางวิชาการและแรงจูงใจที่ลดลง ในขณะที่ Kuh et al. (2008) พบว่า นักศึกษาที่มีผลการเรียนต่ำในรายวิชาหลัก เช่น คณิตศาสตร์ และวิทยาศาสตร์ มีแนวโน้มลาออกสูงขึ้น เนื่องจากทักษะพื้นฐานที่จำเป็นไม่เพียงพอ เช่นเดียวกันกับ Astin (1999) ชี้ให้เห็นว่าผลการเรียนในรายวิชาหลักเป็นตัวบ่งชี้ถึงความพร้อมในการศึกษาในระดับอุดมศึกษา นอกจากนี้ การลงทะเบียนเรียนซ้ำหลายครั้งยังยิ่งเพิ่มความเสี่ยงต่อการพ้นสภาพของนักศึกษา เนื่องจากภาระค่าใช้จ่ายที่เพิ่มขึ้นและความล่าช้าในการสำเร็จการศึกษา (Robbins et al., 2004) โดย Bean and Eaton (2000) ยังกล่าวว่า การเรียนซ้ำมีผลกระทบทางจิตใจ ทำให้นักศึกษารู้สึกหมดกำลังใจและตัดสินใจลาออกก่อนจบการศึกษา

ยิ่งไปกว่านั้น ผลสัมฤทธิ์ในช่วงปีแรกของนักศึกษายังเป็นอีกหนึ่งตัวแปรที่สำคัญ เนื่องจากนักศึกษาที่มีผลสัมฤทธิ์ต่ำในปีแรกมีแนวโน้มสูงที่จะลาออกก่อนจบหลักสูตร (Yorke & Longden, 2004; Tinto, 2017) ซึ่งมีความสอดคล้องกับการศึกษาในประเทศไทยอย่าง ขวัญจิตร สงวนโรจน์ และภัคสุภาส์ จิตโกศลวณิชย์ (2564) ที่พบว่า ปัจจัยด้านผลการเรียน ความรู้พื้นฐาน และการสนับสนุนจากครอบครัวเป็นตัวแปรสำคัญที่ส่งผลต่อนักศึกษาที่ออกกลางคัน นอกจากนี้ งานวิจัยของ จีระนันต์ เจริญรัตน์ (2559) ที่ใช้การจำแนกประเภทด้วย ต้นไม้ตัดสินใจ (Decision Tree) พบว่านักศึกษาที่มีเกรดเฉลี่ยสูงกว่า 3.50 มักได้รับอิทธิพลจากวุฒิการศึกษาเดิม ขณะที่นักศึกษาที่มีเกรดเฉลี่ยต่ำกว่า 2.50 มีแนวโน้มพ้นสภาพเนื่องจากปัญหาทางเศรษฐกิจและสถานภาพครอบครัว จากการทบทวนวรรณกรรมจึงสรุปได้ว่า เกรดเฉลี่ยสะสมต่ำ ผลการเรียนในรายวิชาหลักที่ไม่ดี การลงทะเบียนเรียนซ้ำหลายครั้ง และผลสัมฤทธิ์ในช่วงปีแรกที่ต่ำ ล้วนเป็นปัจจัยสำคัญที่เพิ่มโอกาสในการพ้นสภาพของนักศึกษา เนื่องจากสะท้อนถึงปัญหาทางวิชาการ แรงจูงใจที่ลดลง และภาระทางเศรษฐกิจที่ส่งผลต่อความสามารถในการสำเร็จการศึกษา

2.2. เทคนิคเหมืองข้อมูลในการวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษา

เหมืองข้อมูล (Data Mining) เป็นเครื่องมือสำคัญในการวิเคราะห์ข้อมูลขนาดใหญ่และค้นหารูปแบบที่ซ่อนอยู่ในพฤติกรรมของนักศึกษา งานวิจัยหลายฉบับนำเทคนิคเหมืองข้อมูลมาใช้เพื่อระบุปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษา สุรวัชร ศรีเปารยะ และสายชล สินสมบูรณ์ทอง (2560) แบ่งเทคนิคเหมืองข้อมูลออกเป็นสองประเภทหลัก ได้แก่ (1) เทคนิคพยากรณ์ (Predictive Modeling) เช่น การจำแนกประเภท (Classification) และการถดถอย (Regression), และ (2) เทคนิคพรรณนา (Descriptive Modeling) เช่น การหากฎความสัมพันธ์ (Association Rule Mining) และการจัดกลุ่ม (Clustering)

ในแง่ของการจำแนกประเภท ชณิดาภา บุญประสม และจรัญ แสนราช (2561) ใช้เทคนิค Decision Tree, Naïve Bayes และ K-Nearest Neighbors (KNN) เพื่อทำนายการพ้นสภาพของนักศึกษา พบว่าการกู้ยืมเพื่อการศึกษา สาขาวิชา และอาชีพของบิดามารดาเป็นตัวแปรสำคัญ ขณะที่ ซอและ เกปัน และคณะ (2561) ศึกษาการพ้นสภาพของนักศึกษาในสาขาวิทยาการคอมพิวเตอร์โดยใช้ Neural Network และ

Support Vector Machine (SVM) และพบว่าผลการเรียนในวิชาพื้นฐาน เช่น คณิตศาสตร์และโปรแกรมมิ่ง เป็นตัวแปรหลักที่มีอิทธิพล

นอกจากนี้ ยังมีการหากฎความสัมพันธ์โดยนำอัลกอริทึม Apriori และ FP-Growth มาใช้ค้นหาความสัมพันธ์ของรายวิชาที่มีผลต่อการพ้นสภาพของนักศึกษา พบว่ารายวิชาพื้นฐานวิชาชีพมีผลต่อการพ้นสภาพของนักศึกษาในระดับที่มีนัยสำคัญ (บุษราภรณ์ มหัทธนชัย และคณะ, 2559) อีกทั้ง ปฏิพัทธ์ ปฤชานนท์ และวงกต ศรีอุไร (2561) พบว่าเกรดในรายวิชาศึกษาทั่วไปมีความสัมพันธ์กับโอกาสการออกกลางคัน ซึ่งบ่งชี้ว่านักศึกษาที่มีผลการเรียนต่ำในวิชาพื้นฐานมีแนวโน้มพ้นสภาพสูงกว่า

2.3. งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการพ้นสภาพของนักศึกษาในระดับมหาวิทยาลัยสะท้อนให้เห็นถึงความสำคัญของปัจจัยหลายด้าน เช่น นนทวัฒน์ ทวีชาติ และคณะ (2562) ได้ทำการวิเคราะห์ปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษาโดยใช้ Decision Tree และพบว่า การกู้ยืมเพื่อการศึกษาและเกรดเฉลี่ยจากโรงเรียนมัธยมมีผลต่อการพ้นสภาพของนักศึกษาอย่างมีนัยสำคัญ สิรินยา โมสิกะ (2562) ได้ศึกษาความสัมพันธ์ของรายวิชาในสาขาวิชาบัญชีโดยใช้ Apriori Algorithm และพบว่ารายวิชาพื้นฐานทางการเงินเป็นตัวแปรสำคัญที่มีผลต่อความสำเร็จของนักศึกษา อัจจิมา ปุ่นสุวรรณ และฐิมาพร เพชรแก้ว (2564) ได้ใช้เทคนิค Association Rule Mining ในการวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาคณะนิติศาสตร์ พบว่า ภูมิหลังการศึกษาเป็นปัจจัยหลัก ที่ส่งผลต่อการพ้นสภาพของนักศึกษา ขณะที่ เอกวิจัย เมยไธสง และคณะ (2565) ศึกษาการพยากรณ์ผลการเรียนของนักศึกษาโดยใช้ Neural Network, Decision Tree และ KNN และพบว่า Neural Network ให้ค่าความถูกต้องสูงสุดในการพยากรณ์แนวโน้มการพ้นสภาพ

โดยจากการทบทวนวรรณกรรมที่เกี่ยวข้องทั้งหมดเหล่านี้ชี้ให้เห็นว่า เกรดเฉลี่ยสะสม ผลการเรียนในรายวิชาหลัก จำนวนครั้งที่ลงทะเบียนเรียนซ้ำ และผลสัมฤทธิ์ของนักศึกษาในช่วงปีแรก ล้วนเป็นปัจจัยสำคัญที่ส่งผลต่อการพ้นสภาพของนักศึกษา โดยการนำเทคนิคเหมืองข้อมูล เช่น Decision Tree, Naïve Bayes, KNN และ Association Rule Mining เหมาะสมสำหรับการใช้ในการพยากรณ์และวิเคราะห์ความสัมพันธ์ของข้อมูล ซึ่งงานวิจัยนี้จะนำ อัลกอริทึม FP-Growth มาใช้ในการค้นหากฎความสัมพันธ์ของรายวิชา และสร้างตัวแบบจำแนกประเภทด้วย Decision Tree, Naïve Bayes และ KNN เพื่อนำไปใช้ในการพยากรณ์และพัฒนากลยุทธ์ในการลดอัตราการพ้นสภาพของนักศึกษาในอนาคต

แม้ว่าจะมีงานวิจัยจำนวนมากที่ศึกษาปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษา โดยเฉพาะในระดับมหาวิทยาลัยหรือระดับสาขาวิชา แต่ยังพบว่า งานวิจัยที่มุ่งวิเคราะห์ข้อมูลเชิงลึกในระดับคณะ โดยเฉพาะในบริบทของคณะบริหารธุรกิจ ซึ่งมีลักษณะเฉพาะทั้งในด้านหลักสูตร การบริหารจัดการ และกลุ่มผู้เรียน ยังมีอยู่อย่างจำกัด นอกจากนี้ งานวิจัยที่มีการใช้เทคนิคเหมืองข้อมูลในการสร้างแบบจำลองพยากรณ์ความเสี่ยงส่วนใหญ่มักเน้นการใช้เทคนิคขั้นสูง เช่น Neural Network หรือ Support Vector Machine ซึ่งแม้จะให้ค่าความแม่นยำสูง แต่มีข้อจำกัดด้านการตีความและการนำไปใช้จริงในระดับผู้บริหารหรืออาจารย์ที่ปรึกษา เนื่องจากโครงสร้างของโมเดลที่ซับซ้อนและไม่สามารถอธิบายได้อย่างตรงไปตรงมา อีกประเด็นที่ยังขาดการศึกษาอย่างชัดเจน คือ การประยุกต์ใช้เทคนิคเหมืองข้อมูลในลักษณะที่ตีความง่าย (interpretable models) เช่น การใช้ต้นไม้ตัดสินใจ (Decision Tree) หรือการหากฎความสัมพันธ์ (Association Rules) ร่วมกับฐานข้อมูลนักศึกษาในระดับหลักสูตร เพื่อพัฒนาเครื่องมือที่สามารถนำไปใช้ในการเฝ้าระวัง ติดตาม และให้คำปรึกษานักศึกษาที่มีความเสี่ยงได้อย่างเป็นระบบ ดังนั้น งานวิจัยฉบับนี้จึงมุ่งเน้นการเติมเต็มช่องว่างดังกล่าว โดย

วิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาในระดับปริญญาตรี คณะบริหารธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี ด้วยเทคนิคเหมืองข้อมูลที่เน้นความสามารถในการตีความ และออกแบบโมเดลที่เหมาะสมกับการนำไปประยุกต์ใช้จริงในเชิงบริหารและการให้คำปรึกษาเชิงรุกภายในคณะบริหารธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรีอย่างเป็นรูปธรรม

3. วิธีดำเนินการวิจัย

งานวิจัยนี้มีเป้าหมายในการสร้างแบบจำลองเชิงพยากรณ์เพื่อคาดการณ์โอกาสที่นักศึกษาจะสามารถศึกษาต่อจนสำเร็จการศึกษาหรือมีแนวโน้มออกกลางคัน โดยใช้เทคนิคเหมืองข้อมูล (Data Mining) ซึ่งเป็นเครื่องมืออันทรงพลังในการวิเคราะห์และสกัดความรู้จากข้อมูลเชิงซ้อน การศึกษานี้ครอบคลุมนักศึกษาระดับปริญญาตรี คณะบริหารธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรีที่ลงทะเบียนเรียนใน ปีการศึกษา 2562-2565 และดำเนินการผ่านกระบวนการวิเคราะห์ข้อมูลที่เป็นระบบ เพื่อให้ได้แบบจำลองที่มีประสิทธิภาพสูงสุด

3.1. ประชากรและกลุ่มตัวอย่าง

ประชากรและกลุ่มตัวอย่างที่ใช้ในการวิจัย คือ ข้อมูลของนักศึกษาระดับปริญญาตรี คณะบริหารธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี ที่เข้าศึกษาระหว่างปี 2562 - 2565 จำนวน 5,034 คน

3.2. การเก็บรวบรวมข้อมูล

ข้อมูลที่ใช้ในงานวิจัยนี้เป็น ข้อมูลทุติยภูมิ (Secondary Data) ที่ได้จากระบบบริการการศึกษา มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี ซึ่งผ่านกระบวนการทำความสะอาดข้อมูล (Data Cleaning) เพื่อลบค่าผิดพลาด (Erroneous Data) และกำจัดค่าที่ขาดหายไป (Missing Values) ก่อนนำเข้าสู่กระบวนการแปลงข้อมูล (Data Transformation) ให้เหมาะสมกับการวิเคราะห์ ซึ่งประกอบไปด้วยข้อมูลที่เกี่ยวข้องกับพฤติกรรมการเรียนของนักศึกษา โดยเน้นไปที่ตัวแปรสำคัญที่สามารถนำมาใช้เป็น ตัวชี้วัดความสำเร็จหรือความเสี่ยงในการออกกลางคัน ตัวแปรที่ได้รับการคัดเลือก ได้แก่ เกรดเฉลี่ยสะสม เป็นปัจจัยหลักที่บ่งชี้ถึงความสามารถทางการเรียนของนักศึกษา ผลการเรียนรายวิชา โดยเน้นรายวิชาหลักที่มีความสำคัญต่อหลักสูตร จำนวนหน่วยกิตที่ลงทะเบียนในแต่ละภาคการศึกษา แสดงให้เห็นถึงภาระการเรียนของนักศึกษา สถิติการถอนรายวิชา และระดับผลสัมฤทธิ์ของนักศึกษาในช่วงปีแรก โดยข้อมูลทั้งหมดผ่านกระบวนการ ทำความสะอาดข้อมูล (Data Cleaning) และ แปลงข้อมูล (Data Transformation) เพื่อให้พร้อมสำหรับการนำไปวิเคราะห์ด้วยเทคนิคเหมืองข้อมูล

3.3. เครื่องมือวิจัยและกระบวนการวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลใช้ซอฟต์แวร์ RapidMiner Studio ซึ่งเป็นแพลตฟอร์มชั้นนำด้านการวิเคราะห์ข้อมูล โดยใช้เทคนิคการจำแนกประเภท (Classification) เพื่อพัฒนารูปแบบเชิงคาดการณ์ โดยเปรียบเทียบประสิทธิภาพของ 3 เทคนิคหลัก ได้แก่

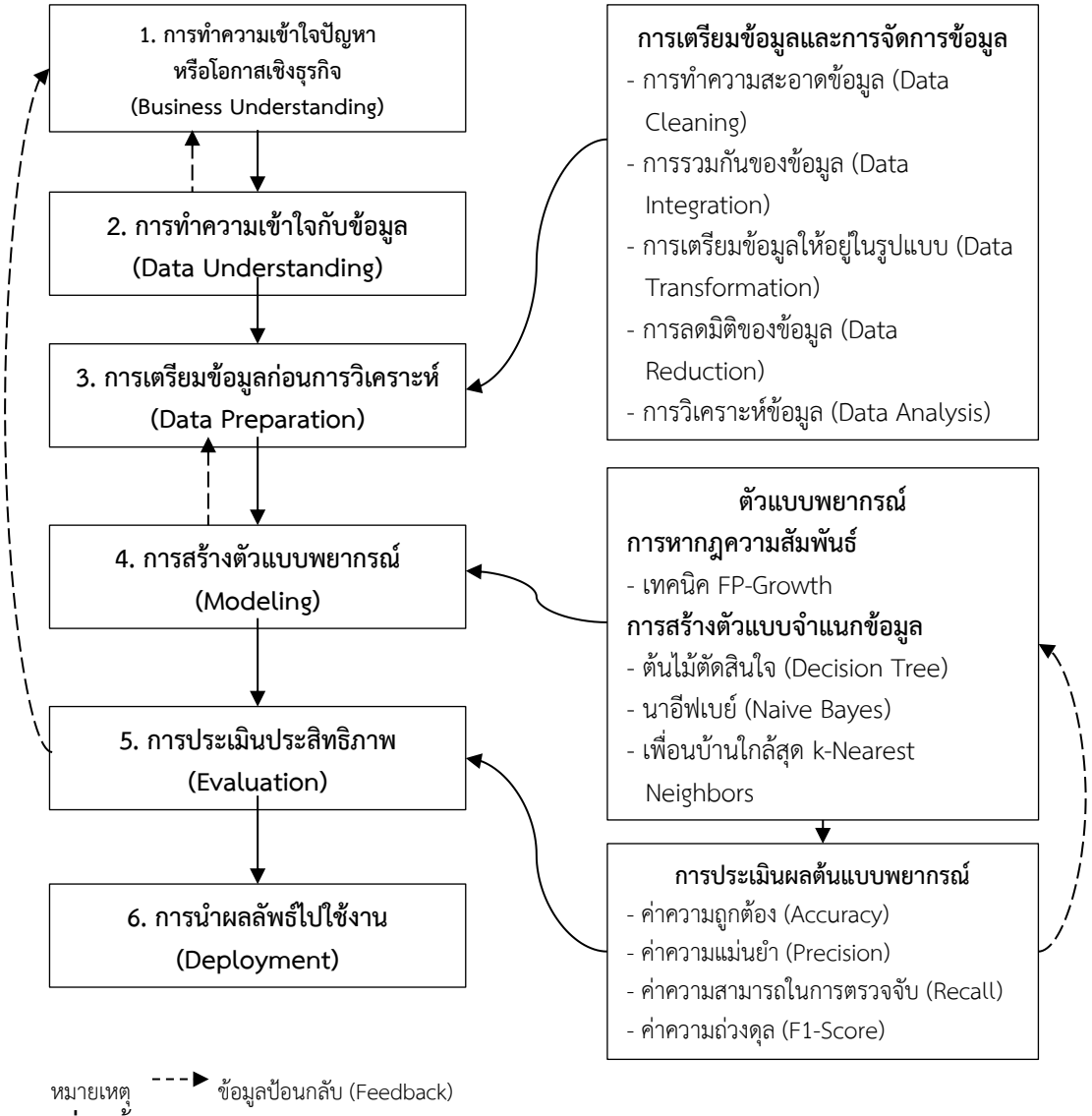
1. ต้นไม้ตัดสินใจ (Decision Tree) เป็นโมเดลที่ใช้โครงสร้างลำดับชั้น (Hierarchical Structure) ในการจำแนกข้อมูลโดยอาศัยกฎ If-Then Rules ซึ่งช่วยให้สามารถเข้าใจการตัดสินใจของโมเดลได้อย่างชัดเจน ข้อดี คือ เข้าใจง่ายและตีความได้ตรงไปตรงมา สามารถจัดการกับตัวแปรเชิงปริมาณและเชิงคุณภาพได้ดี แต่อาจมีข้อจำกัดที่อาจเกิดปัญหา Overfitting หากไม่มีการควบคุมความซับซ้อนของต้นไม้

2. นาอีฟเบย์ (Naïve Bayes) เป็นโมเดลที่อาศัยทฤษฎีความน่าจะเป็น (Bayes' Theorem) โดยมีสมมติฐานว่า แต่ละตัวแปรเป็นอิสระต่อกัน (Independence Assumption) ข้อดี คือ สามารถทำงานได้ดีแม้

ในกรณีที่มีข้อมูลจำนวนน้อย ประมวลผลรวดเร็ว และมีความแม่นยำสูงในกรณีที่สมมติฐานเป็นจริง แต่อาจมีข้อจำกัดในเรื่องของสมมติฐานของความเป็นอิสระของตัวแปรอาจไม่สอดคล้องกับข้อมูลจริง

3. เพื่อนบ้านใกล้สุด (k-Nearest Neighbors : KNN) เป็นโมเดลที่ใช้ระยะห่างทางคณิตศาสตร์ (Euclidean Distance) เพื่อหาตัวอย่างที่คล้ายคลึงกันมากที่สุด และใช้ข้อมูลเหล่านั้นเป็นฐานในการคาดการณ์ ข้อดี คือ ไม่มีการสร้างแบบจำลองล่วงหน้า ทำให้สามารถรองรับข้อมูลที่มีการเปลี่ยนแปลงได้ดี สามารถจัดการกับข้อมูลที่มีโครงสร้างไม่เป็นเชิงเส้น (Non-Linear Data) แต่อาจมีข้อจำกัดในเรื่องของความซับซ้อนในการคำนวณเมื่อขนาดของข้อมูลเพิ่มขึ้น อ่อนไหวต่อค่าผิดปกติ (Outliers) และต้องมีการเลือกค่าพารามิเตอร์ k อย่างเหมาะสม

นอกเหนือจากการใช้โมเดลเชิงคาดการณ์ งานวิจัยนี้ยังนำ กฎเชิงสัมพันธ์ (Association Rule Mining) มาใช้เพื่อค้นหา รูปแบบ (Patterns) และกฎ (Rules) ที่แอบแฝงอยู่ในข้อมูล ตัวอย่างของกฎที่น่าสนใจ ได้แก่ หากนักศึกษามี GPA ต่ำกว่า 2.00 และสอบตกในรายวิชาคุณภาพชีวิตที่ดีของพลเมืองยุคใหม่ (C009) หรือสอบตกในรายวิชาทักษะการใช้คอมพิวเตอร์และเทคโนโลยีสารสนเทศ (C110) นักศึกษามีโอกาสออกกลางคันสูงถึงร้อยละ 85 เป็นต้น



แผนภาพที่ 2 ขั้นตอนการดำเนินการวิจัยตามกระบวนการ KDD (Knowledge Discovery in Databases)

จากแผนภาพที่ 2 แสดงลำดับขั้นตอนการดำเนินงานวิจัยภายใต้แนวคิดกระบวนการค้นหาความรู้จากฐานข้อมูล (KDD) ซึ่งแบ่งออกเป็น 6 ขั้นตอนหลัก ได้แก่ (1) การทำความเข้าใจปัญหา (Business Understanding) เพื่อกำหนดเป้าหมายและขอบเขตของการวิจัย (2) การทำความเข้าใจกับข้อมูล (Data Understanding) เพื่อสำรวจคุณลักษณะของข้อมูลนักศึกษา (3) การเตรียมข้อมูล (Data Preparation) ซึ่งรวมถึงการทำความสะอาดข้อมูล การแปลงรูปแบบข้อมูล และการเลือกตัวแปรที่เหมาะสม (4) การสร้างโมเดล (Modeling) โดยใช้เทคนิคเหมืองข้อมูลในการวิเคราะห์ (5) การประเมินผล (Evaluation) เพื่อเปรียบเทียบความแม่นยำของโมเดลด้วยตัวชี้วัดต่างๆ และ (6) การนำผลลัพธ์ไปใช้งานจริง (Deployment) เช่น การพัฒนาระบบแจ้งเตือนล่วงหน้าหรือแนวทางการให้คำปรึกษาเชิงรุก ทั้งนี้การดำเนินการในแต่ละขั้นตอนมีความสำคัญต่อการสร้างแบบจำลองที่มีประสิทธิภาพและสามารถนำไปใช้ได้จริง

3.4 ระเบียบวิธีวิจัย

งานวิจัยฉบับนี้ดำเนินการภายใต้แนวทางของกระบวนการค้นหาความรู้จากฐานข้อมูล (Knowledge Discovery in Databases: KDD) ซึ่งเป็นกระบวนการวิเคราะห์ข้อมูลที่เป็นระบบและได้รับการยอมรับอย่างกว้างขวางในสาขาการทำเหมืองข้อมูล โดยครอบคลุมตั้งแต่การทำความเข้าใจปัญหา การจัดเตรียมข้อมูล การประมวลผลข้อมูล ไปจนถึงการสร้างแบบจำลองและประเมินประสิทธิภาพ ทั้งนี้ ผู้วิจัยได้ดำเนินการในแต่ละขั้นตอน ดังแสดงในตารางที่ 1

ตารางที่ 1 ขั้นตอนกระบวนการ KDD ที่ใช้ในงานวิจัย

ขั้นตอน	รายละเอียด
1. การทำความเข้าใจปัญหา	ระบุปัญหาการพ้นสภาพของนักศึกษาระดับปริญญาตรี และวัตถุประสงค์ในการสร้างโมเดลทำนายความเสี่ยง
2. การทำความเข้าใจกับข้อมูล	ศึกษาโครงสร้างและเนื้อหาของข้อมูล เช่น เพศ ระบบรับเข้า สาขา GPA รายวิชา และสถานภาพ
3. การเตรียมข้อมูล	ทำความสะอาด แปลง และลดข้อมูล เพื่อให้เหมาะสมกับการวิเคราะห์ - Data Cleaning ลบข้อมูลซ้ำ ข้อมูลผิดพลาด เช่น ค่าคะแนน > 4.00 หรือข้อมูลที่มีค่า Null มากเกินไป - การจัดการ Missing Values แทนค่าที่หายไปด้วยค่ากลางหรือค่าที่พบบ่อยในกลุ่ม - การแปลงข้อมูล (Data Transformation) ดำเนินการดังนี้ 1. การจัดกลุ่มค่าตัวแปรเชิงปริมาณ (Discretization of Continuous Variables) แปลงค่า GPA (เกรดเฉลี่ยสะสม) จากค่าต่อเนื่องให้เป็นกลุ่ม โดยจัดแบ่งออกเป็นช่วง เช่น ต่ำกว่า 2.00 ระหว่าง 2.00-3.00 และมากกว่า 3.00 เพื่อให้สามารถนำมาใช้ในกระบวนการจำแนกประเภทได้ง่ายขึ้น และเพื่อเพิ่มความสามารถในการตีความผลลัพธ์ของโมเดล 2. สำหรับตัวแปรเชิงกลุ่ม เช่น สาขาวิชา และ ระบบการรับเข้าศึกษา ถูกแปลงให้อยู่ในรูปแบบรหัสตัวเลข (Label Encoding) ซึ่งช่วยให้สามารถนำตัวแปรเหล่านี้เข้าสู่โมเดลเชิงคณิตศาสตร์ได้อย่างถูกต้อง โดยไม่ทำให้โมเดลเกิดความลำเอียงจากค่าประเภท (Nominal Values)

ตารางที่ 1 (ต่อ)

ขั้นตอน	รายละเอียด
3. การเตรียมข้อมูล (ต่อ)	3. ตัวแปรที่เป็นรายวิชาหรือผลการเรียนรายวิชาได้ถูกจัดให้อยู่ในโครงสร้างที่สามารถนำเข้าสู่กระบวนการหากฎความสัมพันธ์ด้วยเทคนิค FP-Growth ได้อย่างมีประสิทธิภาพ เช่น การแทนค่าผลการเรียนเป็นระดับ (เช่น “สอบผ่าน” “สอบตก”) เพื่อใช้เป็น input ในกระบวนการ mining - Data Reduction คัดเลือกเฉพาะรายวิชาที่มีจำนวนนักศึกษาเรียนมากกว่า 30 คน และตัดข้อมูลที่ไม่มีผลต่อสถานภาพออก
4. การสร้างแบบจำลอง	ใช้อัลกอริทึม FP-Growth, Decision Tree, Naïve Bayes และ KNN
5. การประเมินผล	ใช้ 5-fold และ 10-fold cross-validation พร้อมตัวชี้วัด ประสิทธิภาพด้วยค่า Accuracy, Precision, Recall, F1-Score, ROC Curves และค่า AUC
6. การนำผลลัพธ์ไปใช้	พัฒนาแนวทางการแจ้งเตือนและให้คำปรึกษานักศึกษาที่มีความเสี่ยง

จากตารางที่ 1 แสดงการดำเนินการตามกรอบกระบวนการค้นหาความรู้จากฐานข้อมูล (Knowledge Discovery in Databases: KDD) โดยกระบวนการเริ่มจากการทำความเข้าใจปัญหาการพ้นสภาพของนักศึกษาระดับปริญญาตรี และกำหนดวัตถุประสงค์เพื่อสร้างโมเดลทำนายความเสี่ยง จากนั้นได้ศึกษาโครงสร้างข้อมูลซึ่งประกอบด้วยตัวแปรต่างๆ เช่น เพศ ระบบรับเข้า สาขาวิชา เกรดเฉลี่ย ผลการเรียนรายวิชา และสถานภาพการศึกษา ในขั้นตอนการเตรียมข้อมูล ผู้วิจัยได้ทำความสะอาดข้อมูลโดยลบข้อมูลซ้ำและข้อมูลที่มีความผิดพลาด เช่น ค่าคะแนนที่เกิน 4.00 หรือข้อมูลที่มีค่า Null มากเกินไป ส่วนข้อมูลที่หายไปได้รับการแทนค่าด้วยค่ากลางหรือค่าที่พบบ่อยในกลุ่ม นอกจากนี้ยังมีการแปลงข้อมูลโดยจัดกลุ่มค่า GPA เป็นช่วง (ต่ำกว่า 2.00, 2.00–3.00, มากกว่า 3.00) และแปลงตัวแปรสาขาและระบบรับเข้าเป็นรหัสตัวเลข รวมถึงการลดขนาดข้อมูลโดยคัดเลือกเฉพาะรายวิชาที่มีนักศึกษาเรียนมากกว่า 30 คน และตัดข้อมูลที่ไม่ส่งผลต่อสถานภาพ

การวิเคราะห์ข้อมูลในงานวิจัยนี้ประกอบด้วยสองเทคนิคหลัก คือ การหากฎความสัมพันธ์และการจำแนกประเภทข้อมูล สำหรับการหากฎความสัมพันธ์ ผู้วิจัยเลือกใช้อัลกอริทึม FP-Growth เนื่องจากมีประสิทธิภาพในการจัดการข้อมูลขนาดใหญ่ได้อย่างรวดเร็ว โดยกำหนดค่าความเชื่อมั่นขั้นต่ำที่ ร้อยละ 95 และคัดเลือกเฉพาะกฎที่มีค่า Lift มากกว่า 1 เพื่อความน่าเชื่อถือ ส่วนการจำแนกประเภทข้อมูลนั้น ผู้วิจัยได้ใช้อัลกอริทึม 3 ประเภท ได้แก่ Decision Tree (C4.5) ซึ่งเหมาะกับข้อมูลที่ตีความเป็นลำดับชั้นและให้ผลลัพธ์ที่เข้าใจง่าย Naïve Bayes ที่มีความแม่นยำสูงสำหรับข้อมูลหลายมิติและประมวลผลได้รวดเร็ว และ k-Nearest Neighbors (KNN) ที่ใช้งานง่ายและเหมาะกับการแบ่งกลุ่มที่ไม่ซับซ้อน ทั้งนี้ อัลกอริทึมเหล่านี้เป็นแบบจำลองน้ำหนักเบา (Lightweight Models) ที่ใช้ทรัพยากรประมวลผลต่ำ ตีความง่าย สามารถนำไปประยุกต์ใช้ได้จริง และเปิดโอกาสให้มีการพัฒนาต่อยอดได้ในอนาคต

ในการประเมินประสิทธิภาพของแบบจำลอง ผู้วิจัยใช้วิธีการ Cross-Validation ทั้งแบบ 5-Fold และ 10-Fold เพื่อแบ่งข้อมูลสำหรับการฝึกและทดสอบอย่างเป็นระบบ โดยใช้ตัวชี้วัดหลายตัว ได้แก่ ค่าความถูกต้อง (Accuracy) ซึ่งแสดงความถูกต้องโดยรวมของแบบจำลอง ค่าความแม่นยำ (Precision) ที่วัดความถูกต้องในการทำนายกลุ่มที่พ้นสภาพ ค่าความสามารถในการตรวจจับ (Recall) ที่แสดงความสามารถในการตรวจพบนักศึกษาที่พ้นสภาพ และค่าความถ่วงดุล (F1-Score) ซึ่งเป็นค่าที่สะท้อนความสมดุลระหว่างค่าความแม่นยำ (Precision)และค่าความสามารถในการตรวจจับ (Recall) และเพิ่มการวิเคราะห์ด้วยแผนภาพ ROC Curve

และพิจารณาค่า AUC โดยเฉพาะในกรณีที่ข้อมูลมีความไม่สมดุลเพื่อให้การประเมินประสิทธิภาพของแบบจำลองมีความสมบูรณ์ยิ่งขึ้น

4. ผลการวิจัย

ผลการวิเคราะห์ข้อมูล

งานวิจัยฉบับนี้มุ่งศึกษาปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะบริหารธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี โดยประยุกต์ใช้เทคนิคเหมืองข้อมูล เพื่อออกแบบระบบเฝ้าระวังและป้องกันการพ้นสภาพอย่างมีประสิทธิภาพ ผลการวิเคราะห์สามารถนำเสนอได้ดังนี้

1. การเปรียบเทียบประสิทธิภาพของแบบจำลองการทำนาย

การวิจัยได้ทดสอบประสิทธิภาพของแบบจำลองเหมืองข้อมูล 3 เทคนิค ได้แก่ Decision Tree, Naïve Bayes และ k-Nearest Neighbors (KNN) โดยเปรียบเทียบประสิทธิภาพด้วยเกณฑ์วัดต่างๆ พบว่า Naïve Bayes มีประสิทธิภาพโดยรวมดีที่สุด ดังแสดงในตารางที่ 2

ตารางที่ 2 การเปรียบเทียบประสิทธิภาพแบบจำลองที่ใช้ในการพยากรณ์

เทคนิคเหมืองข้อมูล	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	96.62	91.47	70.87	79.86
Naïve Bayes	98.07	90.77	88.59	89.67
k-Nearest Neighbors (KNN)	67.57	20.52	84.68	33.04

จากการเปรียบเทียบประสิทธิภาพของแบบจำลองการจำแนกประเภททั้งสามเทคนิค ได้แก่ Decision Tree, Naïve Bayes และ k-Nearest Neighbors (KNN) พบว่าแบบจำลอง Naïve Bayes มีค่าความถูกต้อง (Accuracy) สูงที่สุด อยู่ที่ร้อยละ 98.07 สะท้อนถึงความสามารถของแบบจำลองในการจำแนกผลลัพธ์ได้อย่างถูกต้องโดยรวม นอกจากนี้ยังให้ค่าความแม่นยำ (Precision) ร้อยละ 90.77 และค่าความสามารถในการตรวจจับ (Recall) ร้อยละ 88.59 ซึ่งบ่งชี้ถึงความแม่นยำในการทำนายกลุ่มนักศึกษาที่พ้นสภาพ และความสามารถในการตรวจพบกลุ่มนักศึกษาที่มีความเสี่ยงได้อย่างครอบคลุม โดยค่าความถ่วงดุล (F1-Score) ซึ่งเป็นค่ากลางระหว่าง ค่าความแม่นยำ (Precision) และค่าความสามารถในการตรวจจับ (Recall) อยู่ที่ร้อยละ 89.67 แสดงให้เห็นถึงความสมดุลของแบบจำลองในด้านความแม่นยำและความครอบคลุม ในขณะที่แบบจำลอง Decision Tree แม้จะมีค่าความแม่นยำโดยรวมรองลงมา คือ ร้อยละ 96.62 แต่มีความโดดเด่นด้านค่าความแม่นยำ (Precision) ที่สูงถึงร้อยละ 91.47 อย่างไรก็ตาม ค่าความสามารถในการตรวจจับ (Recall) ค่อนข้างต่ำ ร้อยละ 70.87 สะท้อนว่าแบบจำลองอาจไม่สามารถตรวจพบกลุ่มนักศึกษาที่พ้นสภาพได้ครบถ้วนเท่าที่ควร สำหรับแบบจำลอง KNN พบว่ามีค่าความสามารถในการตรวจจับ (Recall) สูงถึงร้อยละ 84.68 ซึ่งแสดงถึงความสามารถในการตรวจจับนักศึกษาที่พ้นสภาพได้ดี แต่มีค่าความแม่นยำ (Precision) เพียงร้อยละ 20.52 และค่าความถ่วงดุล (F1-Score) ต่ำ ร้อยละ 33.04 จึงไม่เหมาะสมสำหรับการนำไปใช้ในเชิงปฏิบัติ เนื่องจากมีแนวโน้มในการพยากรณ์ผิดพลาดในกลุ่มที่ไม่พ้นสภาพในระดับสูง โดยสรุป แบบจำลอง Naïve Bayes แสดงให้เห็นถึงประสิทธิภาพโดยรวมที่ดีที่สุด ทั้งในด้านความแม่นยำ ความสมดุลของการ

จำแนก และมีความเหมาะสมในการนำไปใช้ในบริบทของการพยากรณ์ความเสี่ยงของนักศึกษาระดับปริญญาตรี คณะบริหารธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรีต่อไป

2. การวิเคราะห์เมทริกซ์ความยุ่งเหยิง (Confusion Matrix)

การประเมินประสิทธิภาพของแบบจำลอง Naïve Bayes ด้วยเมทริกซ์ความยุ่งเหยิง (Confusion Matrix) ช่วยให้สามารถวิเคราะห์ผลการพยากรณ์ในแต่ละกลุ่มได้อย่างชัดเจน และจำแนกความถูกต้องของแบบจำลองในมิติต่างๆ ดังแสดงในตารางที่ 3

ตารางที่ 3 เมทริกซ์ความยุ่งเหยิง (Confusion Matrix) สำหรับแบบจำลอง Naïve Bayes

	ทำนายว่า “พ้นสภาพ”	ทำนายว่า “ไม่พ้นสภาพ”
จริง: พ้นสภาพ	295 (True Positive - TP)	38 (False Negative - FN)
จริง: ไม่พ้นสภาพ	30 (False Positive - FP)	3,161 (True Negative - TN)

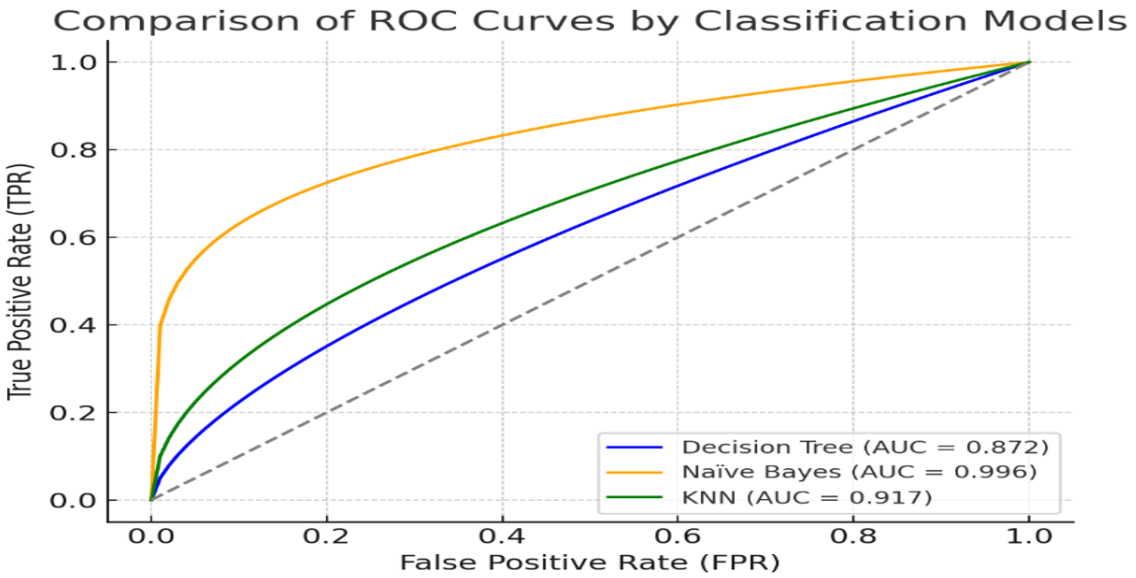
จากตารางข้างต้นสามารถสรุปผลการจำแนกของแบบจำลองได้ดังนี้

1. นักศึกษาที่พ้นสภาพจริง และแบบจำลองสามารถพยากรณ์ได้อย่างถูกต้อง (TP) มีจำนวน 295 คน
2. นักศึกษาที่พ้นสภาพจริง แต่แบบจำลองพยากรณ์ผิดว่าไม่พ้นสภาพ (FN) จำนวน 38 คน
3. นักศึกษาที่ไม่พ้นสภาพจริง แต่แบบจำลองพยากรณ์ผิดว่าเป็นผู้ที่มีความเสี่ยง (FP) จำนวน 30 คน
4. นักศึกษาที่ไม่พ้นสภาพจริง และแบบจำลองพยากรณ์ได้อย่างถูกต้อง (TN) จำนวน 3,161 คน

ผลการวิเคราะห์แสดงให้เห็นว่า แบบจำลอง Naïve Bayes มีความสามารถในการจำแนกโดยรวมได้อย่างมีประสิทธิภาพ อย่างไรก็ตาม กลุ่มที่ควรให้ความสำคัญเป็นพิเศษคือกลุ่ม False Negative (FN) ซึ่งในที่นี้มีจำนวน 38 คน หมายถึงนักศึกษาที่พ้นสภาพจริง แต่แบบจำลองไม่สามารถตรวจจับได้ ส่งผลให้อาจารย์ที่ปรึกษาหรือผู้ที่เกี่ยวข้องอาจไม่สามารถให้ความช่วยเหลือหรือดำเนินการป้องกันได้ทันเวลา แม้ว่าแบบจำลองจะมีค่าความถูกต้อง (Accuracy) อยู่ในระดับสูงถึงร้อยละ 98.07 และมีค่าความสามารถในการตรวจจับ (Recall) สำหรับกลุ่มนักศึกษาที่พ้นสภาพเท่ากับร้อยละ 88.59 ซึ่งถือว่าอยู่ในเกณฑ์ที่ดี แต่การมีจำนวนนักศึกษาที่หลุดรอดจากการตรวจจับยังคงเป็นความท้าทายในการประยุกต์ใช้โมเดลในสถานการณ์จริง ดังนั้นแนวทางในการเพิ่มประสิทธิภาพของระบบควรให้ความสำคัญกับการเพิ่มค่าความไว (Sensitivity) หรือค่าความสามารถในการตรวจจับ (Recall) สำหรับกลุ่มนักศึกษาที่พ้นสภาพ พร้อมกับการลดจำนวน False Negative ให้เหลือน้อยที่สุดเท่าที่จะเป็นไปได้ เพื่อยกระดับความแม่นยำของระบบแจ้งเตือนล่วงหน้าและเพิ่มศักยภาพในการให้คำปรึกษาเชิงรุกได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

3. การวิเคราะห์ด้วย ROC Curve และค่า AUC

ในการประเมินประสิทธิภาพของแบบจำลองการพยากรณ์ งานวิจัยนี้ได้นำกราฟ ROC (Receiver Operating Characteristic Curve) และค่า AUC (Area Under the Curve) มาใช้เป็นตัวชี้วัดเพื่อวิเคราะห์ความสามารถของแบบจำลองในการแยกแยะระหว่างนักศึกษาที่พ้นสภาพกับนักศึกษาที่ไม่พ้นสภาพ โดยข้อดีของการใช้ ROC และ AUC คือสามารถประเมินผลได้โดยไม่ขึ้นกับค่าตัดสิน (threshold) แบบเฉพาะเจาะจง ทำให้ได้ภาพรวมของศักยภาพของแบบจำลองในการจำแนกได้อย่างแม่นยำ ดังแสดงในแผนภาพที่ 3 และตารางที่ 4



แผนภาพที่ 3 แผนภาพเปรียบเทียบ ROC Curve ของ Decision Tree, Naïve Bayes และ KNN

ตารางที่ 4 เปรียบเทียบค่า AUC ที่ได้จากกราฟ ROC ของแบบจำลองต่างๆ

แบบจำลอง	ค่า AUC ที่ได้จากกราฟ	ระดับประสิทธิภาพ
Decision Tree	0.872	ดี (Good)
Naïve Bayes	0.996	ดีเยี่ยม (Excellent)
k-Nearest Neighbors (KNN)	0.917	ดีเยี่ยม (Excellent)

จากผลการวิเคราะห์พื้นที่ใต้กราฟ ROC พบว่าแบบจำลองแต่ละชนิดมีระดับความสามารถในการจำแนกกลุ่มเป้าหมายที่แตกต่างกัน โดยสามารถอธิบายได้ดังนี้

1. Naïve Bayes ให้ค่า AUC สูงที่สุดที่ 0.996 ซึ่งอยู่ในช่วง 0.90 - 1.00 จัดอยู่ในระดับ “ดีเยี่ยม” (Excellent) ตามเกณฑ์ของ Hosmer et al. (2013) แสดงให้เห็นถึงความสามารถในการแยกแยะกลุ่มนักศึกษาที่พ้นสภาพและไม่พ้นสภาพได้อย่างแม่นยำ และมีความเหมาะสมสูงในการนำไปใช้งานจริง

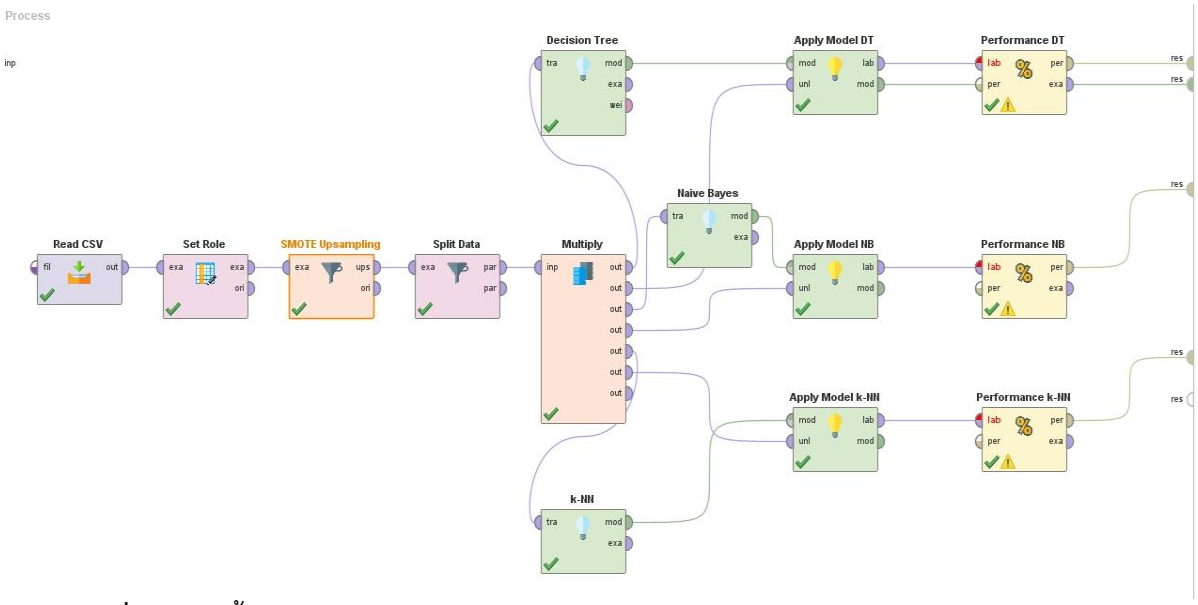
2. k-Nearest Neighbors (KNN) มีค่า AUC เท่ากับ 0.917 อยู่ในระดับ “ดีเยี่ยม” (Excellent) เช่นกัน สะท้อนถึงความสามารถในการจำแนกกลุ่มเสี่ยงได้อย่างมีประสิทธิภาพ อย่างไรก็ตาม อัลกอริทึมนี้อาจมีข้อจำกัดด้านประสิทธิภาพเชิงเวลาเมื่อใช้กับชุดข้อมูลขนาดใหญ่หรือซับซ้อน

3. Decision Tree มีค่า AUC เท่ากับ 0.872 ซึ่งอยู่ในช่วง 0.80 - 0.89 จัดอยู่ในระดับ “ดี” (Good) สะท้อนถึงความสามารถในการจำแนกในระดับที่น่าเชื่อถือ จุดแข็งของแบบจำลองนี้ คือ ความสามารถในการตีความได้ง่าย สามารถแสดงผลในรูปแบบเงื่อนไข (if–then rules) ซึ่งเหมาะสมสำหรับผู้ใช้งานที่ไม่มีความเชี่ยวชาญด้านเทคนิค โดยเฉพาะการให้คำปรึกษาเชิงรุก

สรุป โดยภาพรวม แบบจำลอง Naïve Bayes และ KNN แสดงศักยภาพสูงในการจำแนกกลุ่มนักศึกษาที่มีความเสี่ยง โดยเฉพาะ Naïve Bayes ที่มีค่า AUC สูงสุดและมีความสม่ำเสมอของผลลัพธ์อย่างเด่นชัด ขณะที่แบบจำลอง Decision Tree แม้จะมีค่า AUC ต่ำกว่าเล็กน้อย แต่ยังคงมีความเหมาะสมอย่างยิ่งสำหรับการใช้งานจริง เนื่องจากสามารถสื่อสารผลการวิเคราะห์ในเชิงเหตุผลได้อย่างเข้าใจง่าย และสามารถนำไปสู่การตัดสินใจที่มีประสิทธิภาพได้ทันที

4. ปัญหาข้อมูลไม่สมดุล (Imbalanced Data)

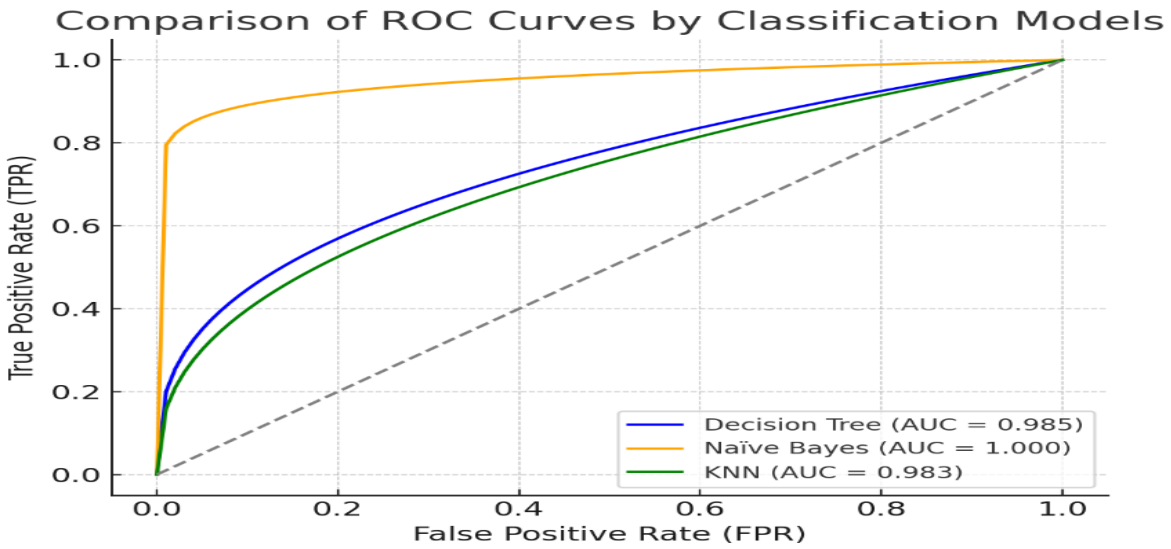
จากการวิเคราะห์ข้อมูลเบื้องต้นพบว่า กลุ่มนักศึกษาที่พ้นสภาพมีจำนวนน้อยกว่ากลุ่มที่ไม่พ้นสภาพอย่างมีนัยสำคัญ กล่าวคือ จากจำนวนนักศึกษาทั้งหมด 5,034 คน มีนักศึกษาที่พ้นสภาพเพียง 475 คน คิดเป็นร้อยละ 9.44 ของกลุ่มตัวอย่างทั้งหมด ส่งผลให้ชุดข้อมูลมีลักษณะไม่สมดุล (Imbalanced Data) ซึ่งอาจกระทบต่อประสิทธิภาพของแบบจำลองการจำแนก โดยเฉพาะอย่างยิ่งการเรียนรู้ของอัลกอริทึมที่มีแนวโน้มโน้มเอียงไปยังกลุ่มที่มีจำนวนมากกว่า เพื่อแก้ไขปัญหาดังกล่าว งานวิจัยนี้ได้ประยุกต์ใช้เทคนิค SMOTE (Synthetic Minority Over-sampling Technique) ในการสร้างข้อมูลตัวอย่างสังเคราะห์จากกลุ่มนักศึกษาที่พ้นสภาพ ส่งผลให้เกิดความสมดุลระหว่างกลุ่ม “พ้นสภาพ” และ “ไม่พ้นสภาพ” มากขึ้น ซึ่งช่วยเสริมประสิทธิภาพของแบบจำลองในการเรียนรู้และจำแนกกลุ่มเป้าหมายได้อย่างเท่าเทียม ดังแผนภาพที่ 4



แผนภาพที่ 4 แสดงขั้นตอนการใช้เทคนิค SMOTE (Synthetic Minority Over-sampling Technique)

ตารางที่ 5 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลอง (หลังจัดการข้อมูลไม่สมดุล)

เทคนิคเหมืองข้อมูล	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
Decision Tree	97.45	97.97	96.90	97.43	0.985
Naïve Bayes	94.91	100.00	89.82	94.63	1.000
k-Nearest Neighbors (KNN)	81.95	74.57	96.96	84.31	0.983



แผนภาพที่ 5 แผนภาพเปรียบเทียบ ROC Curve ของ Decision Tree, Naïve Bayes และ KNN ที่ได้จากการจัดการข้อมูลไม่สมดุลด้วยเทคนิค SMOTE

ผลการประเมินประสิทธิภาพของแบบจำลองหลังการจัดการข้อมูลไม่สมดุลแสดงในตารางที่ 5 พบว่าแบบจำลอง Naïve Bayes ให้ค่า AUC สูงสุดเท่ากับ 1.000 สะท้อนถึงความสามารถในการจำแนกกลุ่มนักศึกษาได้อย่างแม่นยำโดยไม่มีข้อผิดพลาด ขณะที่แบบจำลอง Decision Tree แสดงค่าความถูกต้อง (Accuracy) สูงสุดที่ร้อยละ 97.45 และมีค่า AUC อยู่ที่ 0.985 ซึ่งอยู่ในระดับ “ดีเยี่ยม” (Excellent) ตามเกณฑ์ของ Hosmer et al. (2013) นอกจากนี้ แบบจำลอง Decision Tree ยังมีข้อได้เปรียบด้านโครงสร้างแบบกฎเชิงตรรกะ (if–then rules) ที่สามารถตีความและนำไปใช้ในการสื่อสารผลลัพธ์กับผู้ใช้งานที่ไม่เชี่ยวชาญด้านเทคนิคได้อย่างมีประสิทธิภาพ จึงเหมาะสมต่อการนำไปประยุกต์ใช้ในระบบเฝ้าระวังและระบบแจ้งเตือนล่วงหน้าในบริบทของการบริหารจัดการด้านการศึกษาต่อไป

ตารางที่ 6 เมทริกซ์ความยุ่งเหยิง (Confusion Matrix) สำหรับแบบจำลอง Naïve Bayes

	ทำนายว่า “พ้นสภาพ” YES	ทำนายว่า “ไม่พ้นสภาพ” NO
จริง: พ้นสภาพ YES	2,866 (True Positive - TP)	325 (False Negative - FN)
จริง: ไม่พ้นสภาพ NO	0 (False Positive - FP)	3,191 (True Negative - TN)

ตารางที่ 7 เมทริกซ์ความยุ่งเหยิง (Confusion Matrix) สำหรับแบบจำลอง Decision Tree

	ทำนายว่า “พ้นสภาพ” YES	ทำนายว่า “ไม่พ้นสภาพ” NO
จริง: พ้นสภาพ YES	3,092 (True Positive - TP)	99 (False Negative - FN)
จริง: ไม่พ้นสภาพ NO	64 (False Positive - FP)	3,127 (True Negative - TN)

จากการวิเคราะห์เมทริกซ์ความยุ่งเหยิง (Confusion Matrix) ของแบบจำลอง Naïve Bayes และ Decision Tree หลังการจัดการข้อมูลไม่สมดุล พบว่าแบบจำลองทั้งสองสามารถจำแนกนักศึกษาที่พ้นสภาพและไม่พ้นสภาพได้อย่างมีประสิทธิภาพ แต่มีข้อแตกต่างในเชิงคุณภาพของผลการพยากรณ์ แบบจำลอง Naïve Bayes มีจุดเด่นในการจำแนกนักศึกษาที่ ไม่พ้นสภาพจริง ได้อย่างแม่นยำ โดยไม่พบกรณี False Positive เลย อย่างไรก็

ตาม แบบจำลองยังมีข้อจำกัดในการตรวจจับกลุ่มนักศึกษาที่ พ้นสภาพจริง ซึ่งส่งผลให้เกิด False Negative จำนวนมากถึง 325 คน อันอาจนำไปสู่ความล่าช้าในการให้ความช่วยเหลือเชิงรุก ในทางกลับกัน แบบจำลอง Decision Tree แสดงผลการพยากรณ์ที่สมดุลยิ่งกว่า โดยสามารถลดจำนวน False Negative เหลือเพียง 99 คน และให้ค่า True Positive สูงถึง 3,092 คน แม้จะมี False Positive อยู่บ้าง จำนวน 64 คน แต่ยังอยู่ในระดับที่ยอมรับได้ในทางปฏิบัติ ทั้งนี้ แบบจำลองยังมีข้อได้เปรียบด้านความสามารถในการตีความ ซึ่งเอื้อต่อการนำไปประยุกต์ใช้ในระบบแจ้งเตือนล่วงหน้าและการบริหารจัดการนักศึกษาที่มีความเสี่ยง จากผลการวิเคราะห์ข้างต้นจึงสามารถสรุปได้ว่า แบบจำลอง Decision Tree มีความเหมาะสมกว่าในการนำไปใช้จริง เนื่องจากให้ผลการพยากรณ์ที่แม่นยำ มีความสามารถในการตรวจจับกลุ่มเสี่ยงได้ดี และรองรับการใช้งานในเชิงปฏิบัติอย่างมีประสิทธิภาพ

5. ปัจจัยสำคัญที่ส่งผลต่อการพ้นสภาพของนักศึกษา

จากการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล สามารถระบุปัจจัยสำคัญที่มีอิทธิพลต่อการพ้นสภาพของนักศึกษาระดับปริญญาตรีได้อย่างชัดเจน โดยมีองค์ประกอบหลักที่ควรให้ความสำคัญ ดังต่อไปนี้

1. เกรดเฉลี่ยสะสม (Grade Point Average: GPA) นักศึกษาที่มีค่า GPA ต่ำกว่า 2.00 มีแนวโน้มสูงที่จะพ้นสภาพ โดยเกรดเฉลี่ยสะสมถือเป็นตัวชี้วัดภาพรวมของความสำเร็จทางการเรียนและเป็นดัชนีสำคัญในการคัดกรองความเสี่ยง

2. ผลการเรียนในรายวิชาหลัก การได้รับเกรด D+ หรือ F ในวิชาเอกหรือวิชาที่มีความซับซ้อน เช่น รายวิชาคำนวณ สะท้อนถึงความบกพร่องทางวิชาการในเนื้อหาหลักของหลักสูตร ซึ่งเป็นสัญญาณเตือนที่สำคัญ

3. ความถี่ในการลงทะเบียนเรียนซ้ำ นักศึกษาที่ลงทะเบียนเรียนรายวิชาเดียวกันมากกว่า 2 ครั้ง มีความเสี่ยงต่อการพ้นสภาพในระดับสูง เนื่องจากแสดงถึงปัญหาการทำความเข้าใจเนื้อหาหรือการจัดการการเรียนรู้ในระยะยาว

4. ผลสัมฤทธิ์ทางการศึกษาในปีแรกของนักศึกษา ผลการเรียนในชั้นปีที่ 1 มีความสัมพันธ์เชิงบวกกับความสำเร็จทางการศึกษาในระยะยาว โดยนักศึกษาที่มีผลสัมฤทธิ์ต่ำตั้งแต่ปีแรกมักเผชิญปัญหาในการปรับตัวและมีโอกาสพ้นสภาพสูงขึ้น

ผลการวิเคราะห์กฎการตัดสินใจจากแบบจำลอง Decision Tree พบว่า นักศึกษาที่มีเกรดเฉลี่ยสะสมต่ำกว่า 2.00 และเคยได้รับเกรด D หรือ F ในรายวิชาหลัก มีโอกาสพ้นสภาพสูงถึงร้อยละ 80 ในทางตรงกันข้าม นักศึกษาที่มีเกรดเฉลี่ยสะสม ตั้งแต่ 2.50 ขึ้นไป และไม่มีประวัติการเรียนซ้ำ มีโอกาสสำเร็จการศึกษาสูงถึงร้อยละ 90 นอกจากนี้ การประยุกต์ใช้แบบจำลอง Decision Tree ภายหลังการปรับสมดุลข้อมูลด้วยเทคนิค SMOTE ส่งผลให้ค่าความสามารถในการจำแนก (AUC) เพิ่มขึ้นจาก 0.872 เป็น 0.985 และค่าความสามารถในการตรวจจับ (Recall) เพิ่มขึ้นจากร้อยละ 70.87 เป็นร้อยละ 96.90 แสดงให้เห็นว่าแบบจำลองมีศักยภาพเพิ่มขึ้นอย่างมีนัยสำคัญในการระบุและจับกลุ่มนักศึกษาที่มีความเสี่ยงได้อย่างครอบคลุมและแม่นยำมากยิ่งขึ้น ดังนั้น แบบจำลอง Decision Tree จึงเหมาะสมอย่างยิ่งสำหรับการพัฒนาเป็นระบบเฝ้าระวังความเสี่ยงและระบบแจ้งเตือนล่วงหน้า เพื่อช่วยให้อาจารย์ที่ปรึกษาและผู้บริหารสามารถติดตามนักศึกษากลุ่มเสี่ยงได้อย่างทันท่วงที โดยเฉพาะเมื่อต้องการจัดลำดับความสำคัญของปัจจัยสำคัญ ได้แก่ เกรดเฉลี่ยสะสม ผลการเรียนในรายวิชาหลัก และประวัติการเรียนซ้ำ ซึ่งเป็นตัวชี้วัดสำคัญที่สามารถนำไปใช้

สนับสนุนการวางแผนการให้คำปรึกษาเชิงรุกและลดอัตราการพ้นสภาพของนักศึกษาในระดับปริญญาตรีได้อย่างมีประสิทธิภาพ

5. สรุป อภิปรายผล และข้อเสนอแนะ

5.1 สรุป และอภิปรายผล

ผลการวิจัยพบว่าแบบจำลองการจำแนกประเภทด้วยเทคนิค Decision Tree สามารถจำแนกสถานภาพของนักศึกษาได้อย่างมีประสิทธิภาพ โดยมีค่าความแม่นยำสูงถึง ร้อยละ 97.45 ซึ่งสูงกว่า Naïve Bayes และ k-Nearest Neighbors (KNN) อย่างมีนัยสำคัญ จุดเด่นของโมเดล Decision Tree คือ ความสามารถในการสร้างกฎการตัดสินใจที่ตีความได้ง่าย และสามารถนำไปใช้งานเชิงบริหารจัดการได้อย่างมีประสิทธิภาพ ผลลัพธ์นี้สอดคล้องกับงานของ Dutt et al. (2017) และ Romero & Ventura (2010) ที่ระบุว่า Decision Tree เป็นเทคนิคที่เหมาะสมกับการวิเคราะห์ข้อมูลด้านการศึกษา เนื่องจากสามารถนำเสนอผลในรูปแบบที่เข้าใจได้ง่าย และมีความแม่นยำสูงในงานประเภทการจำแนกกลุ่มนักศึกษา

เมื่อพิจารณาในระดับตัวแปร พบว่า ปัจจัยที่มีอิทธิพลต่อการพ้นสภาพของนักศึกษา ได้แก่ เกรดเฉลี่ยสะสมต่ำกว่า 2.00 ผลการเรียนในรายวิชาหลัก (โดยเฉพาะวิชาเอกและวิชาคำนวณ) การลงทะเบียนเรียนซ้ำหลายครั้งในรายวิชาเดิม และผลสัมฤทธิ์ทางวิชาการในปีแรกของการศึกษา ซึ่งสะท้อนถึงปัญหาพื้นฐานด้านความรู้ ทักษะการเรียนรู้ และแรงจูงใจในการศึกษา ผลลัพธ์นี้สอดคล้องกับแนวคิดของ Tinto (2012), Robbins et al. (2004) และ Kuh et al. (2008) ที่กล่าวถึงผลกระทบของผลสัมฤทธิ์ทางวิชาการและความล้มเหลวในรายวิชาพื้นฐานต่อโอกาสการอยู่รอดในระบบการศึกษา นอกจากนี้ ยังตรงกับงานวิจัยในประเทศไทย เช่น ขวัญจิตร สงวนโรจน์ และภัคสุภาส์ จิตโกศลวณิชย์ (2564) ที่พบว่าผลการเรียนและการสนับสนุนเชิงวิชาการมีบทบาทสำคัญต่อความคงอยู่ของนักศึกษาในระดับอุดมศึกษา

กฎการตัดสินใจที่ได้จากแบบจำลองยังให้ภาพที่ชัดเจนเกี่ยวกับพฤติกรรมกลุ่มเสี่ยง เช่น นักศึกษาที่มีเกรดเฉลี่ยสะสม ต่ำกว่า 2.00 และได้เกรด D หรือ F ในรายวิชาหลัก มีโอกาสพ้นสภาพสูงถึง ร้อยละ 80 ในขณะที่นักศึกษาที่มีเกรดเฉลี่ยสะสม ตั้งแต่ 2.50 ขึ้นไป และไม่มีประวัติการเรียนซ้ำ มีแนวโน้มสำเร็จการศึกษาสูงกว่าร้อยละ 90 ข้อค้นพบนี้สามารถนำไปใช้เป็นเกณฑ์ในการออกแบบมาตรการคัดกรองและติดตามนักศึกษากลุ่มเสี่ยงอย่างเป็นระบบ ซึ่งสอดคล้องกับข้อเสนอของ Bean & Eaton (2000) ที่เน้นความสำคัญของข้อมูลเชิงพยากรณ์ในการออกแบบนโยบายสนับสนุนนักศึกษา

5.2 ข้อเสนอแนะในเชิงนโยบาย

จากผลการวิจัยในครั้งนี้ ผู้วิจัยเสนอแนวทางในการนำผลการวิเคราะห์ไปใช้ประโยชน์ในเชิงนโยบายและการบริหารจัดการระดับคณะ เพื่อป้องกันและลดโอกาสการพ้นสภาพของนักศึกษาอย่างเป็นระบบ ดังนี้

1. ควรมีการพัฒนา ระบบแจ้งเตือนล่วงหน้าสำหรับนักศึกษากลุ่มเสี่ยง โดยใช้ตัวแปรสำคัญที่ได้รับจากผลการวิเคราะห์ เช่น เกรดเฉลี่ยสะสม การลงทะเบียนเรียนซ้ำ และเกรดในรายวิชาหลัก มาเป็นเกณฑ์ในการประมวลผลเพื่อตรวจจับความเสี่ยงของนักศึกษาแบบอัตโนมัติ ระบบนี้ควรถูกเชื่อมโยงกับระบบสารสนเทศของคณะและแจ้งเตือนมายังอาจารย์ที่ปรึกษาและเจ้าหน้าที่งานทะเบียนและวัดผล เพื่อให้สามารถติดตามและให้ความช่วยเหลือได้อย่างทันท่วงที

2. ควรส่งเสริมการให้คำปรึกษาเชิงรุกและเฉพาะกลุ่ม โดยเฉพาะนักศึกษาที่เข้าข่ายกลุ่มเสี่ยงตามกฎการตัดสินใจของแบบจำลอง เช่น นักศึกษาที่มีเกรดเฉลี่ยสะสมต่ำ หรือเรียนซ้ำในรายวิชาสำคัญ ควรได้รับการ

ให้คำปรึกษาในด้านการวางแผนการเรียน การจัดการเวลา และการเสริมจุดอ่อนในรายวิชาพื้นฐานอย่างต่อเนื่อง โดยเน้นการดำเนินงานตั้งแต่ปีแรกของการศึกษา

3. คณะควรจัดกิจกรรมเสริมทักษะทางวิชาการ เช่น ค่ายติวเสริมหรือหลักสูตรเร่งรัดเฉพาะรายวิชาที่มีอัตราการสอบตกสูง เช่น คณิตศาสตร์ ภาษาอังกฤษ หรือสถิติ เพื่อเพิ่มโอกาสให้นักศึกษาเข้าใจบทเรียนและลดความจำเป็นในการลงทะเบียนเรียนซ้ำ

4. ผลการวิเคราะห์สามารถนำไปใช้เป็นข้อมูลประกอบการวางแผนทรัพยากรทางการศึกษาอย่างมีประสิทธิภาพ เช่น การจัดสรรจำนวนอาจารย์ที่ปรึกษาให้เพียงพอกับนักศึกษากลุ่มเสี่ยง การจัดระบบทุนการศึกษาให้ตรงกับความต้องการ หรือการจัดหลักสูตรเสริมเฉพาะกลุ่ม ซึ่งจะช่วยลดความเสี่ยงและเพิ่มอัตราการสำเร็จการศึกษาของนักศึกษาในระยะยาว

5.3 ข้อเสนอแนะสำหรับการวิจัยในครั้งต่อไป

จากข้อค้นพบและข้อจำกัดที่เกิดขึ้นในการวิจัยครั้งนี้ ผู้วิจัยเห็นว่ามีประเด็นสำคัญหลายประการที่ควรได้รับการศึกษาเพิ่มเติมในอนาคต เพื่อยกระดับความถูกต้องของแบบจำลอง เพิ่มความหลากหลายของข้อมูล และขยายโอกาสในการนำผลวิจัยไปประยุกต์ใช้ในบริบทที่กว้างขึ้น ดังนี้

1. ควรขยายขอบเขตของตัวแปรที่ใช้ในแบบจำลอง โดยเพิ่มเติมตัวแปรด้านจิตสังคมและพฤติกรรมของนักศึกษา เช่น ความถี่ในการเข้าเรียน การส่งงานตรงเวลา ความถี่ในการเข้ารับคำปรึกษา หรือแรงจูงใจในการเรียน รวมถึงความผูกพันกับสถาบัน เพื่อให้แบบจำลองสามารถสะท้อนพฤติกรรมของนักศึกษาได้ครอบคลุมมากยิ่งขึ้น การเก็บข้อมูลสามารถดำเนินการผ่านแบบสอบถามมาตรฐาน เช่น แบบวัดแรงจูงใจ (Motivated Strategies for Learning Questionnaire: MSLQ) หรือแบบวัดความผูกพันกับการเรียน และควรเชื่อมโยงข้อมูลกับระบบสารสนเทศที่มีอยู่ เช่น การเข้าใช้งานระบบ E-learning หรือการลงทะเบียนในระบบออนไลน์

2. ควรมีการพัฒนาเครื่องมือหรือระบบต้นแบบ (Prototype) ที่สามารถนำผลการวิเคราะห์ไปประยุกต์ใช้อย่างเป็นรูปธรรม เช่น การสร้าง Dashboard ติดตามสถานะนักศึกษาแบบรายบุคคล โดยใช้เครื่องมือพัฒนา Web-based Application หรือ Microsoft Power BI ซึ่งสามารถเชื่อมโยงข้อมูลจากระบบทะเบียนกลางและนำเสนอผลวิเคราะห์ในรูปแบบที่อาจารย์ที่ปรึกษาและเจ้าหน้าที่งานทะเบียนสามารถใช้งานได้จริงในระดับปฏิบัติ ทั้งนี้ เพื่อให้ผลการวิจัยไม่เพียงแต่เป็นข้อเสนอเชิงทฤษฎี แต่สามารถนำไปประยุกต์ใช้เพื่อการเฝ้าระวังและช่วยเหลือนักศึกษาได้อย่างมีประสิทธิภาพและยั่งยืนในบริบทของคณะหรือมหาวิทยาลัย

เอกสารอ้างอิง

ขวัญจิตร สงวนโรจน์ และภัคสุภาส์ จิตโกศลวณิชย์. (2564). การศึกษาเกี่ยวกับการคงอยู่และการออกกลางคันของนักศึกษาระดับปริญญาตรีภาคปกติ ของมหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา ระหว่างปีการศึกษา 2561 - 2563 สำหรับการจำแนกสภาพและเสนอแนวทางธำรงรักษานักศึกษา. **วารสารก้าวทันโลกวิทยาศาสตร์, 21**(2), 127-149.

จีระนันต์ เจริญรัตน์. (2559). การวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาที่มีผลการเรียนปกติโดยใช้ต้นไม้ตัดสินใจ. **SNRU Journal of Science and Technology, 8**(2), 256-267.

ชณิดาภา บุญประสม และจรัญ แสนราช. (2561). การวิเคราะห์การทำนายการลาออกกลางคันของนักศึกษาระดับปริญญาตรี โดยใช้เทคนิควิธีการทำเหมืองข้อมูล. **วารสารวิชาการครุศาสตร์อุตสาหกรรม พระจอมเกล้าพระนครเหนือ, 9**(1), 142-151.

ซอและ เกปัน พิมลพรรณ ลีลาภัทรพันธุ์ และอัจฉราพร ยกขุน. (2561). การวิเคราะห์ปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษาโดยใช้เทคนิคเหมืองข้อมูล กรณีศึกษา หลักสูตรวิทยาการคอมพิวเตอร์และหลักสูตรเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏยะลา. **วารสารสาขาวิทยาศาสตร์และเทคโนโลยี, 5**(4), 96-110.

นนทวัฒน์ ทวีชาติ อารยา เพ็งประจญ, วิไลรัตน์ ยาทองไชย และชูศักดิ์ ยาทองไชย. (2562). การวิเคราะห์ปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษาระดับปริญญาตรี ด้วยเทคนิคการทำเหมืองข้อมูล. **การประชุมวิชาการระดับชาติการจัดการเทคโนโลยีและนวัตกรรม ครั้งที่ 5 มหาวิทยาลัยราชภัฏมหาสารคาม** (น. 47-60). มหาวิทยาลัยราชภัฏมหาสารคาม.

บุษราภรณ์ มหัทธนชัย ครรชิต มาลัยวงศ์ เสมอแข สมหอม และณัฐิยา ตันตรานนท์. (2559). กฎความสัมพันธ์ของรายวิชาที่มีผลต่อการพ้นสภาพนักศึกษาโดยใช้อัลกอริทึมอพริโอริ. **การประชุมวิชาการระดับชาติมหาวิทยาลัยราชภัฏกำแพงเพชร ครั้งที่ 3** (น. 456-469). มหาวิทยาลัยราชภัฏกำแพงเพชร.

ปฏิพัทธ์ ปฤชานนท์ และวงกต ศรีอุไร. (2561). การประยุกต์ใช้กฎความสัมพันธ์เพื่อวิเคราะห์ความเสี่ยงการออกกลางคันของนักศึกษาสาขาเทคโนโลยีสารสนเทศ. **วารสารวิทยาศาสตร์และวิทยาศาสตร์ศึกษา, 1**(2), 123-133.

มหาวิทยาลัยเทคโนโนโลยีราชมงคลธัญบุรี. (2564). **ประกาศมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี เรื่อง เกณฑ์การวัดและประเมินผลการศึกษาระดับปริญญาตรี.** [ออนไลน์] ค้นเมื่อ 4 สิงหาคม 2565 จาก https://www.oreg.rmutt.ac.th/?wpfb_dl=1842

ระบบบริการการศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี. (2565). **ระบบบริการการศึกษา.** [ออนไลน์] ค้นเมื่อ 4 สิงหาคม 2565 จาก https://oreg.rmutt.ac.th/?page_id=14908

สิรินยา โมสิกะ. (2562). **การพัฒนาตัวจำแนกประเภทสถานะการสำเร็จการศึกษา จากกฎความสัมพันธ์ของกลุ่มวิชาพื้นฐานวิชาชีพ** [วิทยานิพนธ์ปริญญามหาบัณฑิต ไม่ได้ตีพิมพ์]. มหาวิทยาลัยทักษิณ.

สุกัญญา ทารส และทรงศักดิ์ ภูสีอ่อน. (2563). ปัจจัยจำแนกการออกกลางคันของนิสิตปริญญาตรี มหาวิทยาลัยมหาสารคาม. **วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม, 26**(1), 273-287.

สุรวัชร ศรีเปารยะ และสายชล สินสมบูรณ์ทอง. (2560). การเปรียบเทียบประสิทธิภาพวิธีการจำแนกกลุ่มการเป็นโรคไตเรื้อรัง: กรณีศึกษาโรงพยาบาลแห่งหนึ่งในประเทศอินเดีย. **วารสารวิทยาศาสตร์และเทคโนโลยี, 25**(5), 839-853.

อัจจิมา ปุ่นสุวรรณ และฐิมาพร เพชรแก้ว. (2564). การค้นหาปัจจัยที่ส่งผลต่อการพ้นสภาพกลางคันของนักศึกษาโดยใช้การค้นหากฎความสัมพันธ์. **วารสารมหาวิทยาลัยนราธิวาสราชนครินทร์ สาขามนุษยศาสตร์และสังคมศาสตร์, 8**(1), 112-136.

เอกวิจัย เมยไธสง ฉวีวรรณ สีสม และสุเทพ เมยไธสง. (2565). การพยากรณ์ผลการเรียนเพื่อวางแผนการลงทะเบียนเรียนของนักศึกษา สาขาวิชาวิทยาศาสตร์การกีฬาโดยใช้วิธีการเรียนรู้ของเครื่อง. **วารสารพุทธปรัชญาวิวัฒน์, 6**(2), 329-340.

Astin, A. W. (1999). Student involvement: A developmental theory for higher education. **Journal of College Student Development, 40**(5), 518-529.

Bean, J. P., & Eaton, S. B. (2000). The psychology underlying successful retention practices. **Journal of College Student Retention, 3**(1), 73-89.

Dutt, A., Ismail, M. A., & Herawan, T. (2017). A systematic review on educational data mining. **IEEE Access, 5**, 15991-16005.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). **Applied logistic regression** (3rd ed.). John Wiley & Sons.

Kuh, G. D., Cruce, T. M., Shoup, R., Kinzie, J., & Gonyea, R. M. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. **The Journal of Higher Education, 79**(5), 540-563.

Pascarella, E. T., & Terenzini, P. T. (2005). **How college affects students: A third decade of research** (2nd ed.). Jossey-Bass.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. **Psychological Bulletin, 130**(2), 261-288.

Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. **IEEE Transactions on Systems Man and Cybernetics Part C (Applications and Reviews), 40**(6), 601 - 618.

Tinto, V. (2012). **Leaving college: Rethinking the causes and cures of student attrition**. University of Chicago Press.

Tinto, V. (2017). Through the eyes of students. Journal of College Student Retention: Research, **Theory & Practice, 19**(3), 254-269.

Williamson, B. (2017). **Big data in education: The digital future of learning, policy and practice.** Sage Publications.

Yorke, M., & Longden, B. (2004). **Retention and student success in higher education**. McGraw-Hill Education.