การเปรียบเทียบแบบจำลองเพื่อทำนายสาขาวิชาชีพของนักศึกษา
Classification Model Comparison for Predicting Professional Fields

พิชญสินี กิจวัฒนาถาวร (Phichayasini Kitwatthanathawon)

Received: November 30, 2023
Revised: August 22, 2024
Accepted: August 27, 2024

*ผู้นิพนธ์ประสานงาน: พิชญสินี กิจวัฒนาถาวร (Phichayasini Kitwatthanathawon) อีเมล: pichak@g.sut.ac.th

DOI:10.14416/j.it.2026.v1.002

บทคัดย่อ

ปัจจุบันผู้ที่ศึกษาในสำนักวิชาศาสตร์และศิลป์ดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารี จำเป็นจะต้องเลือกสาขาวิชาชีพใดสาขาหนึ่งได้แก่ เทคโนโลยีดิจิทัล และนิเทศศาสตร์ดิจิทัล ซึ่งการเลือกสาขาวิชาชีพของนักศึกษาถือเป็นเรื่องสำคัญมาก เนื่องจากส่งผลโดยตรงต่อชุดวิชาที่ต้องเรียนและเส้นทางอาชีพในอนาคต งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อสร้างและเปรียบเทียบประสิทธิภาพของแบบจำลองทำนายสาขาวิชาชีพของนักศึกษา โดยได้นำข้อมูลจากแบบสอบถามความคิดเห็นของนักศึกษาในสำนักวิชาศาสตร์และศิลป์ดิจิทัลมาวิเคราะห์ด้วยเทคนิคการจำแนกประเภท ซึ่งเป็นเทคนิคหนึ่งของการทำเหมืองข้อมูล และสร้างแบบจำลองทำนายสาขาวิชาชีพของนักศึกษาด้วย 5 ขั้นตอนวิธี ได้แก่ ต้นไม้ตัดสินใจ นาอีฟเบย์ วันอาร์ ซัพพอร์ตเวคเตอร์แมชชีน และเพื่อนบ้านใกล้ที่สุด จากนั้นจึงที่ได้มาประเมินประสิทธิภาพด้วยค่าความแม่นยำ ค่าความระลึก ค่าความถูกต้อง และค่าถ่วงดุลร่วมกับการทดสอบแบบไขว้จำนวน 10 20 และ 30 ส่วน ผลการทดลองพบว่าวิธีสร้างแบบจำลองที่มีค่าความถูกต้องมากที่สุดเท่ากับ 89.6% คือ นาอีฟเบย์ ซึ่งแบ่งชุดข้อมูลทดสอบออกเป็น 20 ส่วน วิธีซัพพอร์ตเวคเตอร์แมชชีนสามารถสร้างแบบจำลองที่มีค่าความแม่นยำมากที่สุดเท่ากับ 82.1% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 30 ส่วน และวิธีต้นไม้ตัดสินใจสามารถสร้างแบบจำลองที่มีค่าความระลึกและค่าถ่วงดุลมากที่สุดเท่ากับ 83.3% และ 81.5% ตามลำดับ เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 10 ส่วน

คำสำคัญ: สาขาวิชาชีพ การเลือกสาขาวิชา การทำนายเหมืองข้อมูล การจำแนกประเภท

Abstract

Currently, peoples who study at the Institute of Digital Arts and Science, Suranaree University of Technology, have to select one of the professional fields between Digital Technology and Digital Communication. The professional field was very important because it will have a direct effect on course modules to be studied and a future career path. The purpose of this research was to construct and compare the classification performance of professional fields prediction model by collecting and analyzing data from the student opinion questionnaire. The classification technique which is part of a data mining approach was applied with 5 algorithms, e.g. Decision Tree, Naïve Bayes, OneR, Support Vector Machines, and K-Nearest Neighbors. The performance of the classification model was evaluated and compared by Precision, Recall, Accuracy, and F-Measure with 10-folds, 20-folds, and 30-folds cross-validation. The evaluation results indicated that (1) the classification model obtained from the Naïve Bayes algorithm achieved the highest Accuracy at 89.6%, using 20-folds cross-validation; (2) the classification model derived from the Support Vector Machines algorithm achieved the highest Precision at 89.6%, using 30-folds cross-validation; and (3) the classification model obtained from the Decision Tree algorithm achieved the highest Recall and F-measure at 83.3% and 82.5%, respectively, using 10-folds cross-validation.

Keywords: Professional Fields, Major Selection, Prediction, Data Mining, Classification.

*สำนักวิชาศาสตร์และศิลป์ดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารี
*Institute of Digital Arts and Science, Suranaree University of Technology.

1. บทนำ

ตั้งแต่อดีตจนถึงปัจจุบันการศึกษานั้นถือว่าเป็นพื้นฐานที่สำคัญที่สุดเพราะเป็นรากฐานของการพัฒนาในทุก ๆ ด้าน นอกจากจะทำให้มีความรู้และพัฒนาตนเองได้แล้ว ยังเป็นตัวกำหนดทิศทางของประเทศได้อีกด้วยโดยเฉพาะการศึกษาในระดับอุดมศึกษาเพราะเป็นการมุ่งพัฒนาผู้เรียนให้สามารถเป็นผู้นำเพื่อรองรับการเปลี่ยนแปลงทางเศรษฐกิจและสังคมได้อย่างมีประสิทธิภาพ ประเทศใดที่มีการส่งเสริมการศึกษาที่ถูกทิศทาง รวมถึงการวางแผนที่มีคุณภาพย่อมมีทรัพยากรมนุษย์ที่มีประสิทธิภาพ และทำให้ประเทศนั้น ๆ มีต้นทุนที่เป็นทรัพยากรมนุษย์ที่มีศักยภาพในการทำให้ประเทศก้าวไกล ในขณะเดียวกันการเลือกสาขาที่จะเข้าศึกษาต่อก็เป็นสิ่งที่สำคัญเช่นกัน การเลือกเรียนสาขาที่ตนชอบหรือถนัดนั้นมักส่งผลดีต่อหน้าที่การงานในอนาคตเป็นอย่างมาก [1]

สำนักวิชาศาสตร์และศิลป์ดิจิทัล (Institute of Digital Arts and Science: DIGITECH) มหาวิทยาลัยเทคโนโลยีสุรนารี เป็นสำนักวิชาที่จัดการเรียนสอนทางด้านเทคโนโลยีดิจิทัล มีหลักสูตรที่สอนวิทยาการด้านเทคโนโลยีดิจิทัล (Digital Technology) และหลักสูตรวิทยาการด้านนิเทศศาสตร์ดิจิทัล (Digital Communication) ซึ่งสอดคล้องกับความต้องการของภาคธุรกิจและอุตสาหกรรมในปัจจุบัน ตามนโยบาย Thailand 4.0 ของรัฐบาลและแนวโน้มสำคัญของโลก [2]

การเลือกสาขาวิชาของนักศึกษาถือเป็นเรื่องที่สำคัญเป็นอย่างมากเพราะมีผลโดยตรงต่อเส้นทางอาชีพในอนาคต และอาจส่งผลต่อชีวิตของผู้เรียนเนื่องด้วยนักศึกษายังขาดประสบการณ์ ไม่ทราบถึงความต้องการที่แน่นอนและอาจรู้จักแต่ละสาขาวิชาไม่มากพอใช้ความรู้สึกชอบในการตัดสินใจเข้าศึกษาตามเพื่อนหรือความคิดเห็นจากผู้ปกครอง หรือมีปัญหาด้านฐานะทางครอบครัว ซึ่งอาจส่งผลทำให้เกิดการเสียเวลาและโอกาสเมื่อพบว่าตนเองไม่เหมาะสมกับสาขาวิชาที่เลือกเรียน [3] จากเหตุผลดังกล่าวทำให้ผู้วิจัยได้เห็นความสำคัญของการพัฒนาแบบจำลองที่จะช่วยในการทำนายการเลือกสาขาวิชาชีพที่จะเข้าศึกษาของนักศึกษาระดับปริญญาตรี งานวิจัยนี้จึงได้ประยุกต์ใช้เทคนิคการจำแนกประเภท (Classification) อันเป็นเทคนิคหนึ่งที่สำคัญในการทำเหมืองข้อมูล (Data Mining) มาใช้ในการสร้างแบบจำลองการทำนายด้วยขั้นตอนวิธี (Algorithm) ได้แก่ ต้นไม้ตัดสินใจ (Decision Tree) นาอีฟเบย์ (Naive Bayes) วันอาร์ (OneR) ซัพพอร์ตเวคเตอร์แมชชีน (Support Vector Machines) และเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors) เพื่อช่วยในการวิเคราะห์ปัจจัยที่เกี่ยวข้องและทำนายสาขาวิชาชีพของนักศึกษาในอนาคต

2. วัตถุประสงค์

เพื่อสร้างและเปรียบเทียบประสิทธิภาพของแบบจำลองทำนายสาขาวิชาชีพของนักศึกษา

3. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

3.1 การทำเหมืองข้อมูล (Data Mining)

เป็นกระบวนการในการค้นหารูปแบบ แนวทาง ความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลขนาดใหญ่ โดยใช้การเรียนรู้เรื่องของเครื่อง สถิติ ระบบฐานข้อมูล และการรู้จำแบบ โดยการทำเหมืองข้อมูลนั้นเป็นขั้นตอนวิธีการในการการค้นหาความรู้ในฐานข้อมูล (Knowledge Discovery In Databases: KDD) หรือในอีกนิยามหนึ่งการทำเหมืองข้อมูล คือ กระบวนการที่กระทำกับข้อมูลโดยจะนำข้อมูลที่มีอยู่มาวิเคราะห์แล้วดึงความรู้หรือสิ่งสำคัญออกมาเพื่อใช้ในการวิเคราะห์หรือทำนายสิ่งต่าง ๆ [4] ขั้นตอนวิธีในการวิเคราะห์ข้อมูลมีหลากหลาย แสดงตัวอย่างดังนี้

- นาอีฟเบย์ (Naive Bayes)

นาอีฟเบย์หรือทฤษฎีของเบย์ (Bayes' Theorem) ถูกพัฒนาขึ้นโดยโทมัส เบย์ (Thomas Bayes) ใช้หลักการของความน่าจะเป็นมาพัฒนาเป็นทฤษฎีบทดังกล่าว [5] โดยเรียนรู้ปัญหาที่เกิดขึ้นเพื่อนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่ด้วยขั้นตอนวิธีที่ไม่ซับซ้อน หลักการของนาอีฟเบย์ได้ใช้การคำนวณหาความน่าจะเป็นในการทำนายผล ซึ่งเป็นวิธีหนึ่งในการแก้ปัญหาแบบจำแนกประเภทที่สามารถคาดการณ์ผลลัพธ์ได้ โดยจะทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ ขั้นตอนวิธีนี้เหมาะกับกรณีของชุดข้อมูลตัวอย่างที่มีจำนวนมาก และคุณสมบัติของตัวอย่างนั้นไม่ขึ้นต่อกัน ขั้นตอนวิธีนี้จะมีประสิทธิภาพมากกับชุดข้อมูลขนาดเล็ก

- ต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจเป็นการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบต้นไม้ [6] ประกอบไปด้วย 1) โหนด (Node) ในแต่ละโหนดนั้นจะแสดงถึงการตัดสินใจบนข้อมูลของคุณสมบัติต่าง ๆ 2) กิ่ง (Branch) เป็นค่าหรือผลลัพธ์ที่ได้จากการทดสอบ ซึ่งเป็นส่วนเชื่อมต่อระหว่างโหนด

และแสดงคุณลักษณะที่เป็นไปได้ทั้งหมดของแต่ละโหนด 3) โหนดปลาย (Leaf Node) เป็นโหนดที่อยู่ชั้นล่างสุดของต้นไม้ตัดสินใจ แสดงถึงกลุ่มของข้อมูลหรือผลลัพธ์ที่ได้จากเงื่อนไขการตัดสินใจ ต้นไม้การตัดสินใจเป็นที่นิยมเพราะง่ายต่อการเข้าใจและตีความ สามารถจัดการได้ทั้งข้อมูลเชิงหมวดหมู่และตัวเลข และสามารถสร้างแบบจำลองจากความสัมพันธ์ที่ซับซ้อนได้

- เพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbors: K-NN)

เพื่อนบ้านใกล้ที่สุดเป็นขั้นตอนวิธีในการเรียนรู้ของเครื่อง (Machine Learning) ที่ใช้ในงานจำแนกประเภท (Classification) และการทำนายค่าตัวเลข (Regression) เป็นวิธีการแบ่งคลาสสำหรับใช้จัดหมวดหมู่ข้อมูลตามข้อมูลที่อยู่ใกล้ที่สุด ทำงานโดยการหาข้อมูลจำนวน K รายการที่ใกล้เคียงที่สุดกับข้อมูลนั้น ๆ ในชุดข้อมูลเรียนรู้ (Training Data) โดยใช้ฟังก์ชั่นวัดระยะห่างระหว่างข้อมูล เช่น ระยะห่างแบบยูคลิเดียน (Euclidean Distance) แมนฮัตตัน (Manhattan Distance) หรือมินกอว์สกิ (Minkowski Distance) แล้วคำนวณค่าเฉลี่ยหรือค่าโหวตของประเภทของข้อมูลนั้น ๆ เพื่อใช้ในการทำนายหรือจำแนกประเภทของข้อมูลที่มาใหม่ [7] หากข้อมูลที่กำลังสนใจนั้นอยู่ใกล้ข้อมูลใดมากที่สุด ระบบจะให้คำตอบเป็นเหมือนคำตอบของข้อมูลที่อยู่ใกล้ที่สุดนั้น ซึ่งลักษณะการทำงานแบบนี้ไม่ได้ใช้ชุดข้อมูลเรียนรู้ในการสร้างแบบจำลอง แต่จะใช้ข้อมูลที่อยู่ใกล้ที่สุดนี้มาสร้างตัวแบบจำลองเลย ขั้นตอนวิธีเพื่อนบ้านใกล้ที่สุดเป็นขั้นตอนวิธีที่ง่ายต่อการเข้าใจและการนำไปใช้งาน แต่อาจมีปัญหาในการจัดแบ่งกลุ่มข้อมูลที่มีความซับซ้อน

- ซัพพอร์ตเวคเตอร์แมชชีน (Support Vector Machines: SVM)

ซัพพอร์ตเวคเตอร์แมชชีนเป็นขั้นตอนวิธีแบบมีผู้สอน (Supervised Learning) โดยใช้การค้นหาเวคเตอร์ (Vector) ที่เป็นของจริงหรือเป็นไปได้สูงสุดมาช่วยแก้ปัญหาการจำแนกข้อมูล [8] โดยอาศัยหลักการของการหาสัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกป้อนเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ ซึ่งเน้นไปยังเส้นแบ่งแยกแยะกลุ่มข้อมูลได้ดีที่สุด ซัพพอร์ตเวคเตอร์แมชชีนมีความสามารถในการจัดแบ่งข้อมูลที่ซับซ้อนและไม่เป็นเชิงเส้นได้ดี ด้วยวิธีการหาเวคเตอร์ที่เป็นไปได้สูงสุดที่สามารถแบ่งกลุ่มข้อมูลได้อย่างชัดเจนและมีความแม่นยำสูงที่สุด (Margin Maximization)

- วันอาร์ (OneR)

วันอาร์เป็นขั้นตอนวิธีที่ทำความเข้าใจถึงวิธีการสร้างกฎได้ง่าย และใช้งานง่าย โดยใช้หลักการสร้างกฎการจำแนกประเภทด้วยการเลือกแอททริบิวต์ที่มีค่าความผิดพลาดน้อยที่สุดเพียงแอททริบิวต์เดียวมาเป็นแบบจำลองทำนายคลาสของข้อมูล ซึ่งทำให้ได้รับกฎการจำแนกประเภทข้อมูลที่มีจำนวนน้อยมาก [9]

3.2 กระบวนการวิเคราะห์ข้อมูลด้วยคริปส์-ดีเอ็ม (CRISP DM)

เป็นกระบวนการมาตรฐานที่ใช้สำหรับการทำเหมืองข้อมูล มีไว้เพื่อทำการวิเคราะห์และนำไปใช้ประโยชน์ทางธุรกิจ [10] มีขั้นตอนดังภาพที่ 1

ภาพที่ 1 *กระบวนการวิเคราะห์ข้อมูลด้วยคริปส์-ดีเอ็ม*

จากภาพที่ 1 แสดงกระบวนการวิเคราะห์ข้อมูลด้วยคริสป์ - ดีเอ็ม ประกอบด้วย 6 ขั้นตอน ได้แก่

ขั้นตอนที่ 1 การทำความเข้าใจธุรกิจ (Business Understanding) คือการทำความเข้าใจเกี่ยวกับความต้องการทางธุรกิจหรือองค์กรที่เป็นเป้าหมาย โดยการสังเกตเกี่ยวกับปัจจัยทางด้านธุรกิจต่าง ๆ จากนั้นนำมาแปลงผลให้อยู่ในรูปแบบที่สามารถนำมาวิเคราะห์ข้อมูลและวางแผนการดำเนินงาน

ขั้นตอนที่ 2 การทำความเข้าใจข้อมูล (Data Understanding) ในขั้นตอนนี้จะเริ่มต้นด้วยการตรวจสอบข้อมูลที่ได้จัดเก็บรวบรวมมา จากนั้นนำข้อมูลมาทำความเข้าใจเพื่อเลือกข้อมูลที่เหมาะสมกับการวิเคราะห์

ขั้นตอนที่ 3 การเตรียมข้อมูล (Data Preparation) คือ การแปลงรูปข้อมูลให้พร้อมที่จะนำเข้าสู่การสร้างแบบจำลอง

ขั้นตอนที่ 4 การสร้างแบบจำลอง (Modeling) ในขั้นตอนนี้เป็นสร้างแบบจำลองโดยใช้ขั้นตอนวิธีที่หลากหลาย เพื่อหาแบบจำลองที่สามารถนำไปใช้แก้ปัญหาได้

ขั้นตอนที่ 5 การวัดประสิทธิภาพของแบบจำลอง (Evaluation) เป็นการวัดประสิทธิภาพของแบบจำลองเพื่อเตรียมความพร้อมก่อนการนำไปใช้งาน

ขั้นตอนที่ 6 การนำแบบจำลองไปใช้งานจริง (Deployment) เป็นการนำแบบจำลองที่ได้ผลลัพธ์ที่ดีที่สุด และผ่านการวัดประสิทธิภาพแล้วมาใช้งานจริง เพื่อวิเคราะห์และแก้ไขปัญหา

3.3 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการทำเหมืองข้อมูลทางการศึกษาในปัจจุบันมีหลากหลายด้าน รวมถึงการทำนายผลการเรียนของนักเรียน การปรับเปลี่ยนประสบการณ์การเรียนรู้ในแบบเฉพาะบุคคล การระบุนักเรียนที่มีความเสี่ยง การเพิ่มประสิทธิภาพการออกแบบหลักสูตร และการวิเคราะห์พฤติกรรมของนักเรียน มีรายละเอียดดังนี้

- การทำนายผลการเรียนของนักเรียน

งานวิจัยในด้านนี้นิยมใช้ขั้นตอนวิธี เช่น ต้นไม้ตัดสินใจ โครงข่ายประสาทเทียม และเครือข่ายแบบเบย์ เพื่อทำนายผลการเรียนในอนาคตของนักเรียนจากข้อมูลในอดีต การทำนายนี้ทำให้อาจารย์เข้าช่วยเหลือนักเรียนที่อาจประสบปัญหาได้ตั้งแต่เริ่มต้น [11] - [15] ตัวอย่างเช่น งานวิจัยของนาคิโปวา และคณะ [13] ได้พัฒนา และนำขั้นตอนวิธีนาอีฟเบย์มาใช้ในการทำนายผลการเรียนตามคุณลักษณะของนักเรียน ผลการวิจัยยืนยันว่าการใช้ขั้นตอนวิธีนาอีฟเบย์ในบริบทนี้สามารถทำนายผลการเรียนได้อย่างมีประสิทธิภาพ และใช้งานได้จริงด้วยค่าความถูกต้อง 85% งานวิจัยของอัล-บารัค และคณะ [14] ได้ทำนายเกรดเฉลี่ยสุดท้ายของนักเรียน โดยพิจารณาจากผลการเรียนในรายวิชาบังคับที่เรียนก่อนหน้า ผู้วิจัยได้รวบรวมข้อมูลเกรดเฉลี่ยสุดท้ายและเกรดในทุกรายวิชาจากใบรับรองผลการเรียนของนักเรียน หลังจากการเตรียมข้อมูลแล้ว ผู้วิจัยได้ใช้ขั้นตอนวิธีต้นไม้ตัดสินใจ (J48) เพื่อค้นหากฎการจำแนก ผลลัพธ์ของการวิจัยทำให้สามารถระบุรายวิชาที่สำคัญที่สุดในแผนการศึกษาของนักเรียนได้ โดยพิจารณาจากผลการเรียนในรายวิชาบังคับ และงานวิจัยของชายานและซาเนน [15] วิเคราะห์แบบจำลองการทำนายผลการเรียนของนักเรียนจากพฤติกรรมการเรียนออนไลน์ โดยจัดเก็บข้อมูลจากระบบการจัดการการเรียนรู้ (Moodle LMS) ผู้วิจัยใช้ขั้นตอนวิธีต้นไม้ตัดสินใจ (J48 และ ID3) สร้างแบบจำลองที่สามารถเพิ่มผลสัมฤทธิ์ทางการเรียนของผู้เรียนได้ ผลการวิจัยพบว่าเกรดเฉลี่ยก่อนหน้า คะแนนสอบกลางภาค จำนวนการเข้าดูสื่อการสอน การคลิก และระยะเวลาการใช้งาน มีผลอย่างยิ่งต่อผลการเรียนของนักเรียน และผู้สอนจำเป็นต้องให้ความสำคัญกับเรื่องนี้มากขึ้น

- การระบุนักเรียนที่มีความเสี่ยง

งานวิจัยในด้านนี้มีวัตถุประสงค์เพื่อระบุนักเรียนที่มีความเสี่ยงที่จะออกจากการศึกษาโดยการวิเคราะห์รูปแบบการเข้าเรียน คะแนน และการมีส่วนร่วม ผลลัพธ์ของงานวิจัยในด้านนี้ทำให้สถาบันการศึกษาสามารถระบุนักเรียนที่มีความเสี่ยงได้ก่อนและสามารถให้การสนับสนุนที่ตรงเป้าหมายแก่นักศึกษาเหล่านี้ได้ [16], [17]

- การเสริมสร้างการออกแบบหลักสูตร

งานวิจัยในด้านนี้มีวัตถุประสงค์เพื่อวิเคราะห์ข้อมูลผลการเรียนของนักเรียน เพื่อให้ทราบว่าหลักสูตรต้องปรับปรุงในด้านใด ผลลัพธ์ของงานวิจัยในด้านนี้สามารถนำไปวางกลยุทธ์และพัฒนาสื่อการสอนให้มีประสิทธิภาพมากขึ้น [18], [19]

- การวิเคราะห์พฤติกรรมของนักเรียน

งานวิจัยในด้านนี้จะศึกษารูปแบบพฤติกรรมของนักเรียน รวมถึงการมีปฏิสัมพันธ์กับระบบการจัดการการเรียนรู้ และการมีส่วนร่วมแบบออนไลน์ เพื่อหาความสัมพันธ์ระหว่างพฤติกรรมของนักเรียนกับความสำเร็จทางวิชาการ [3], [20], [21]

- การปรับประสบการณ์การเรียนรู้ส่วนบุคคล

งานวิจัยในด้านนี้มุ่งเน้นการปรับแต่งประสบการณ์การเรียนรู้ให้ตรงกับความต้องการของนักเรียนแต่ละคน มีการนำเทคนิคต่าง ๆ มาประยุกต์ใช้ เช่น การจัดกลุ่ม และระบบการแนะนำ เพื่อระบุเส้นทางการเรียนรู้และทรัพยากรที่มีประสิทธิภาพสูงสุดสำหรับนักเรียนแต่ละคน [22] - [26] ตัวอย่างเช่น งานวิจัยของชอมธารี และคณะ [24] ได้พัฒนาเว็บแอปพลิเคชันที่ช่วยให้นักเรียนระดับมัธยมศึกษาสามารถเลือกเรียนหลักสูตรเพื่อประกอบอาชีพของตนเองได้ ผู้วิจัยได้ใช้ขั้นตอนวิธีสมการถดถอย ต้นไม้ตัดสินใจ และการสุ่มป่าไม้ ระบบนี้จะแนะนำทางเลือกอาชีพให้กับนักเรียนโดยพิจารณาจากลักษณะบุคลิกภาพ ความสนใจ และความสามารถในการเรียนหลักสูตรนั้น งานวิจัยของปานดีย์

และโมริยา [25] ได้ทำนายอาชีพที่นักศึกษาวิศวกรรมศาสตร์สามารถเลือกได้หลังจากสำเร็จการศึกษาโดยใช้เทคนิคการจำแนกด้วยการเรียนรู้ของเครื่องที่ได้รับความนิยมสูงสุด 6 เทคนิค ผลการทดลองพบว่า ขั้นตอนวิธีการลงคะแนนเสียง (Voting) มีความแม่นยำสูงที่สุด รองลงมาคือ ป่าไม่แบบสุ่ม กราเดียนดีเซนท์ และแผนผังการตัดสินใจ และงานวิจัยของทิวารี และเจน [26] ได้เปรียบเทียบขั้นตอนวิธีการเรียนรู้ของเครื่องยอดนิยม 6 แบบ โดยการทำนายโอกาสทางอาชีพของนักเรียน ซึ่งพิจารณาจากผลสัมฤทธิ์ทางการเรียน กิจกรรมนอกหลักสูตร และข้อมูลประชากร ผลการวิจัยสรุปว่าการเรียนรู้ของเครื่องอาจเปลี่ยนแปลงการศึกษาและแนะนำการประยุกต์ใช้เชิงกลยุทธ์เพื่อปรับปรุงคำแนะนำด้านอาชีพให้ดียิ่งขึ้น

สำหรับงานวิจัยที่เกี่ยวกับการทำเหมืองข้อมูลทางการศึกษาในประเทศไทยที่มุ่งเน้นด้านการวิเคราะห์พฤติกรรมของนักเรียน เช่น รัชฎา เทพประสิทธิ์ และจรัญ แสนราช [3] ได้ทำการวิเคราะห์ปัจจัยที่มีผลต่อการเลือกเข้าศึกษาต่อในสาขาวิชาจำนวน 15 สาขาของคณะครุศาสตร์ มหาวิทยาลัยราชภัฏเชียงราย โดยเก็บข้อมูลตัวแปรทั้งหมด 9 ตัวแปร ของนักศึกษาระดับปริญญาตรีจากงานส่งเสริมวิชาการและงานทะเบียน ในช่วงปีการศึกษา พ.ศ. 2556 - 2560 จำนวน 3,867 คน นำมาวิเคราะห์ความสัมพันธ์ของตัวแปรโดยใช้กระบวนการเหมืองข้อมูลด้วยเทคนิคการจำแนกประเภทแบบวิธีต้นไม้ตัดสินใจ ผลวิจัยพบว่าปัจจัยที่มีผลต่อการเลือกสาขาวิชา ได้แก่ แผนการเรียนก่อนเข้าศึกษา และเพศ ผลการประเมินประสิทธิภาพของแบบจำลองมีค่าความแม่นยำถึง 72.5% ซึ่งถือว่าเป็นแบบจำลองที่มีประสิทธิภาพและมีความน่าเชื่อถือ งานวิจัยของภัทรสุดา จารุธีรพันธุ์ [27] ได้ศึกษาปัจจัยที่มีผลต่อการเลือกเรียนสาขาวิชาในระดับปริญญาตรีของคณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏเลย โดยใช้ข้อมูลจากนักศึกษาปีที่ 1 จำนวน 678 คน จากแบบสอบถามเกี่ยวกับปัจจัยที่มีผลต่อการเลือกเรียนของนักศึกษาคณะวิทยาการจัดการ โดยใช้สถิติที่ใช้ในการวิเคราะห์ข้อมูล คือ ค่าร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และการวิเคราะห์ค่าความแปรปรวนแอนโนวา ผลการวิจัยพบว่า ปัจจัยที่มีผลต่อการเลือกเรียนสาขาวิชา คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏเลย โดยเฉลี่ยอยู่ในระดับมาก และพบว่านักศึกษาได้ให้ความสำคัญปัจจัยทั้ง 5 ด้าน โดยมีด้านคุณลักษณะของสาขาวิชา ด้านการเผยแพร่และประชาสัมพันธ์ ด้านสถานที่ด้านการมีอิทธิพลต่อการตัดสินใจเลือกเรียน และมีด้านค่าใช้จ่ายกับตัวแปรสาขาวิชาที่เลือกเรียน ซึ่งมีความสัมพันธ์กันที่ระดับนัยสำคัญทางสถิติที่ระดับ 0.05 งานวิจัยของประเทือง วงษ์ทอง และคณะ [28] ได้สร้างและเปรียบเทียบประสิทธิภาพแบบจำลองความสัมพันธ์ของข้อมูล โดยใช้ขั้นตอนวิธีโครงข่ายประสาทเทียม นาอีฟเบย์ และต้นไม้ตัดสินใจ โดยเก็บรวบรวมข้อมูลจากผู้ประกอบอาชีพด้านคอมพิวเตอร์จำนวน 395 คน จากนั้นวิเคราะห์ด้วยวิธีแบ่งข้อมูล 5 ส่วน และวัดค่าความถูกต้อง และความแม่นยำ พบว่าแบบจำลองที่มีประสิทธิภาพมากที่สุด คือ แบบจำลองที่ได้จากขั้นตอนวิธีนาอีฟย์เบย์ มีค่าความถูกต้องเท่ากับ 79.64% และค่าความแม่นยำเท่ากับ 74.04% รองลงมาคือขั้นตอนวิธีโครงข่ายประสาทเทียม มีค่าความถูกต้องเท่ากับ 77.82% และค่าความแม่นยำเท่ากับ 73.60% และขั้นตอนวิธีต้นไม้ตัดสินใจมีค่าความถูกต้องเท่ากับ 54.69% และค่าความแม่นยำเท่ากับ 38.72% ตามลำดับ และงานวิจัยของวรารินทร์ ปัญญาวงษ์ [29] ได้ศึกษาปัจจัยที่มีผลต่อการเลือกสาขาเรียนต่อระดับปริญญาตรี และสร้างแบบจำลองสำหรับพยากรณ์การสำเร็จการศึกษา โดยรวบรวมข้อมูลนิสิตจากคณะสถาปัตยกรรมศาสตร์ ในปีการศึกษา 2557 - 2559 จำนวน 929 รายการ โดยนำขั้นตอนวิธีวันอาร์มาสร้างแบบจำลองในการพยากรณ์ และวัดประสิทธิภาพแบบจำลองด้วยค่าความถูกต้องและค่าระลึก ซึ่งแบบจำลองดังกล่าวให้ค่าความถูกต้องและค่าระลึกสูงใกล้เคียงกัน คือ 91% และ 91.1% ตามลำดับ

ตัวอย่างงานวิจัยที่เกี่ยวข้องกับการทำเหมืองข้อมูลทางการศึกษาด้านการทำนายผลการเรียนของนักเรียน เช่น ธนะวัฒน์ วรรณประภา [30] ได้สร้างแบบจำลองการคัดกรองผู้ที่มีแนวโน้มสัมฤทธิ์ผลในการเรียนปริญญาตรีด้วยการทำเหมืองข้อมูล โดยเก็บข้อมูลการสำเร็จการศึกษาของบัณฑิตจากระบบกองทะเบียนและประมวลผลการศึกษา จำนวน 738 คน จำแนกเป็น 30 แอททริบิวต์ ใช้ขั้นตอนวิธีต้นไม้ตัดสินใจ ซัพพอร์ตเวคเตอร์แมชชีน และการเรียนรู้เชิงลึก นอกจากนี้ยังมีการทดลองแบ่งสัดส่วนชุดข้อมูลเรียนรู้ และข้อมูลทดสอบเป็น 70:30 75:25 และ 80:20 ตามลำดับ ผลการศึกษาพบว่าขั้นตอนวิธีการเรียนรู้เชิงลึกให้ค่าความถูกต้องสูงสุดเท่ากับ 85.94% เมื่อแบ่งชุดข้อมูลเรียนรู้และข้อมูลทดสอบในอัตราส่วน 80:20 และงานวิจัยของจิราภรณ์ เจริญยิ่ง [31] ได้ทำเหมืองข้อมูลทางการศึกษาด้วยการจำแนกประเภทข้อมูล

เพื่อหาความสัมพันธ์ของตัวแปร และเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีต้นไม้ตัดสินใจ การสุ่มป่าไม้ นาอีฟเบย์ และเพื่อนบ้านใกล้ที่สุด โดยใช้ข้อมูลของนักเรียนระดับมัธยมศึกษา จำนวน 649 รายการ แบ่งเป็น 31 แอตทริบิวต์ ผลการวิจัยพบว่าขั้นตอนวิธีต้นไม้ตัดสินใจ การสุ่มป่าไม้ และนาอีฟเบย์ มีค่าความถูกต้องเท่ากับ 91.54% ส่วน ขั้นตอนวิธีเพื่อนบ้านใกล้ที่สุดมีค่าความถูกต้องเพียง 82.31%

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการทำเหมืองข้อมูลทางการศึกษาพบว่าส่วนใหญ่มุ่งเน้นที่การนำเทคนิคการเรียนรู้ของเครื่องมาใช้ทำนายผลการเรียนของนักเรียนโดยใช้ขั้นตอนวิธีที่หลากหลาย ซึ่งสามารถสรุปได้ดังตารางที่ 1

ตารางที่ 1 *เปรียบเทียบงานวิจัยที่เกี่ยวกับการทำเหมืองข้อมูลทางการศึกษา*

งานวิจัยirch	กลุ่มของงานวิจัย*	ขั้นตอนวิธี										
		การวิเคราะห์ทางสถิติ	การถดถอยโลจิสติก	เกาเดียบูทติ้ง	ต้นไม้ตัดสินใจ	ซัพพอร์ตเวกเตอร์แมชชีน	วันอาร์	นาอีฟเบย์	โครงข่ายประสาทเทียม	เพื่อนบ้านใกล้ที่สุด	การสุ่มป่าไม้	การเรียนรู้เชิงลึก
[3]	B				✓							
[13]	P		✓	✓	✓			✓	✓		✓	
[14]	P				✓							
[15]	P				✓							
[24]	L			✓	✓							
[25]	L		✓		✓						✓	
[26]	L			✓	✓	✓			✓	✓	✓	
[27]	B	✓										
[28]	B				✓			✓	✓			
[29]	B						✓					
[30]	P				✓	✓						✓
[31]	P				✓			✓		✓	✓	
งานนี้	L				✓	✓	✓	✓		✓		

* P = การทำนายผลการเรียนของนักเรียน, I = การระบุนักเรียนที่มีความเสี่ยง, C = การเสริมสร้างการออกแบบหลักสูตร, B = การวิเคราะห์พฤติกรรมของนักเรียน, L = การปรับประสบการณ์การเรียนรู้ส่วนบุคคล

จากการศึกษางานวิจัยที่เกี่ยวข้องสรุปได้ว่า การวิจัยเกี่ยวกับการทำเหมืองข้อมูลทางการศึกษาส่วนใหญ่ใช้ขั้นตอนวิธีต้นไม้ตัดสินใจ และนาอีฟเบย์ ซึ่งให้ค่าประสิทธิภาพการพยากรณ์สูง และงานส่วนใหญ่มุ่งเน้นไปที่การวิเคราะห์พฤติกรรมของนักเรียน และการทำนายผลการเรียนของนักเรียน งานวิจัยยังมีข้อจำกัดในเรื่องการนำลักษณะทางจิตวิทยาหรือลักษณะบุคลิกภาพมาวิเคราะห์ร่วมด้วย ซึ่งลักษณะดังกล่าวจะทำให้แบบจำลองการทำนายอาชีพที่ครอบคลุมมากขึ้น นอกจากนี้ การระบุคุณลักษณะที่ส่งผลต่อการทำนายสาขาวิชาชีพนั้นมีความซับซ้อนและมีความเฉพาะเจาะจงตามบริบทของงาน ด้วยเหตุนี้จึงจำเป็นต้องมีวิธีการที่ซับซ้อนมากขึ้นในการรวมแหล่งข้อมูลที่มีความหลากหลาย เช่น ข้อมูลทางการศึกษา ประวัติการทำงาน และความสนใจส่วนบุคคล ดังนั้น การบูรณาการความสนใจส่วนบุคคล และการทดลองที่หลากหลายจึงเป็นวิธีหนึ่งที่ทำให้แบบจำลองการเรียนรู้ของเครื่องมีความแม่นยำ ยุติธรรม และมีประโยชน์ในการนำไปใช้งานจริง

งานวิจัยนี้จึงนำขั้นตอนวิธีต้นไม้ตัดสินใจและนาอีฟเบย์มาใช้ อีกทั้งยังนำขั้นตอนวิธีอื่นอันเป็นที่นิยม ได้แก่ วันอาร์ เพื่อนบ้านใกล้ที่สุด และซัพพอร์ตเวกเตอร์แมชชีนมาสร้างแบบจำลองและเปรียบเทียบประสิทธิภาพการทำนายสาขาวิชาชีพของนักศึกษาสำนักวิชาศาสตร์และศิลป์ดิจิทัล เพื่อหาขั้นตอนวิธีที่เหมาะสมที่สุดในการจำแนกสาขาวิชาชีพของนักศึกษา ตลอดจนวิเคราะห์ปัจจัยที่ส่งผลต่อการเลือกสาขาวิชาชีพ เพื่อเป็นเครื่องมือช่วยในการเลือกสาขาให้กับผู้ที่สนใจเข้าศึกษาต่อไป

4. วิธีดำเนินการวิจัย

การสร้างแบบจำลองทำนายการเลือกสาขาวิชาชีพของนักศึกษาสำนักวิชาศาสตร์และศิลป์ดิจิทัลนี้ ได้นำกระบวนการวิเคราะห์ข้อมูลด้วยคริสป์-ดีเอ็ม ซึ่งเป็นกระบวนการมาตรฐานที่ใช้สำหรับการทำเหมืองข้อมูลมาประยุกต์ใช้ให้เหมาะสมตามบริบทของงานวิจัยดังนี้

4.1 การทำความเข้าใจปัญหา

การทำความเข้าใจปัญหาด้านการเลือกสาขาวิชาชีพของนักศึกษา เพื่อศึกษาปัญหา รวบรวมข้อมูลที่เกี่ยวข้อง รวมถึงกำหนดเป้าหมายของงานวิจัยในการวิเคราะห์ปัจจัยที่มีผลต่อการเลือกกลุ่มวิชาชีพของนักศึกษาแต่ละคน

4.2 การทำความเข้าใจข้อมูล

การทำความเข้าใจข้อมูล เป็นขั้นตอนการรวบรวมข้อมูลส่วนตัวและความสนใจของนักศึกษาแต่ละคนโดยใช้แบบสอบถาม ซึ่งประกอบด้วยข้อคำถามที่ผ่านการตรวจสอบและประเมินความสอดคล้อง (Index of Consistency: IOC) จากผู้เชี่ยวชาญ 3 ท่านแล้ว จำนวน 15 ข้อ จำแนกเป็นปัจจัยทางด้านบุคคล ปัจจัยทางด้านการเงิน และปัจจัยด้านการประกอบอาชีพ ดังแสดงในตารางที่ 2 มีผู้ตอบแบบสอบถามทั้งสิ้น 310 ราย จากนั้นจึงนำข้อมูลดังกล่าวไปสร้างแบบจำลองด้วยขั้นตอนวิธีนาอีฟเบย์ ต้นไม้ตัดสินใจ เพื่อนบ้านใกล้ที่สุด และซัพพอร์ตเวคเตอร์แมชชีน

4.3 การเตรียมข้อมูลก่อนการวิเคราะห์

ในขั้นนี้เป็นการสำรวจและตรวจสอบข้อมูลที่รวบรวมได้จากขั้นตอนที่แล้ว เพื่อจัดเตรียมข้อมูลก่อนนำไปวิเคราะห์และสร้างแบบจำลอง มีรายละเอียดดังนี้

4.3.1 การปรับแต่งโปรแกรมให้เหมาะสมกับชุดข้อมูล งานวิจัยนี้ได้นำเข้าข้อมูลมาวิเคราะห์บนโปรแกรมเวก้า (Weka) เวอร์ชั่น 3.8.5 แต่เนื่องจากข้อมูลนำเข้าดังกล่าวเป็นภาษาไทย จึงจำเป็นต้องปรับแต่งโปรแกรมในส่วนของการเข้ารหัสข้อมูล (File Encoding) ให้รองรับภาษาไทยก่อน

4.3.2 การแปลงช่วงข้อมูลตัวเลข (Discretization) เพื่อเปลี่ยนประเภทของข้อมูลทุกแอตทริบิวต์ในปัจจัยด้านการประกอบอาชีพจากข้อมูลเชิงตัวเลข (Numeric) ให้เป็นเชิงหมวดหมู่ (Nominal)

4.3.3 การทำความสะอาดข้อมูล ตรวจสอบ และแก้ไขข้อมูลที่ผิดพลาด โดยการลบข้อมูลระเบียนที่มีการระบุข้อมูลไม่ชัดเจนหรือขาดหาย ขจัดช่องว่างที่ปะปนอยู่กับข้อมูล และแก้ไขคำผิด

ผลที่ได้จากการเตรียมข้อมูลทำให้ได้ชุดข้อมูลที่พร้อมสำหรับการนำไปสร้างแบบจำลอง จำนวน 226 ระเบียน ประกอบด้วย 15 แอตทริบิวต์ ดังแสดงในตารางที่ 2

4.4 การสร้างแบบจำลอง (Modeling)

เป็นการนำชุดข้อมูลเรียนรู้ (Training Data) มาหารูปแบบความสัมพันธ์ภายในชุดข้อมูลตามขั้นตอนวิธีที่เป็นมาตรฐาน เพื่อให้ได้ผลลัพธ์ที่ดีที่สุดในการทำนายสาขาวิชาชีพของนักศึกษา ได้แก่ ต้นไม้ตัดสินใจ นาอีฟเบย์ วันอาร์ เพื่อนบ้านใกล้ที่สุด และซัพพอร์ตเวคเตอร์แมชชีน งานวิจัยนี้ได้ใช้โปรแกรมเวก้า (Weka) เวอร์ชั่น 3.8.5 เป็นเครื่องมือ

ตารางที่ 2 รายละเอียดข้อมูลนำเข้า

ที่	กลุ่มข้อคำถาม	คำอธิบายข้อมูล	ประเภทข้อมูล	หมายเหตุ
1	ปัจจัยทางด้านบุคคล	เพศ	ข้อมูลเชิงหมวดหมู่	ชาย, หญิง, อื่น ๆ
2	ปัจจัยทางด้านการเงิน	รายได้ครอบครัวต่อเดือน	ข้อมูลเชิงตัวเลข	
3	ปัจจัยด้านอาชีพ 1	หางานทำได้ง่ายเมื่อสำเร็จการศึกษาในกลุ่มวิชาชีพนั้น	ข้อมูลเชิงหมวดหมู่	สนใจมากที่สุด, มาก, ปานกลาง, น้อย, น้อยที่สุด
4	ปัจจัยด้านอาชีพ 2	มีโอกาสได้รับค่าตอบแทนสูงหากทำงานในกลุ่มวิชาชีพนั้น	ข้อมูลเชิงหมวดหมู่	สนใจมากที่สุด, มาก, ปานกลาง, น้อย, น้อยที่สุด
5	ปัจจัยด้านอาชีพ 3	เป็นกลุ่มวิชาชีพที่สามารถหารายได้พิเศษได้หลายทาง	ข้อมูลเชิงหมวดหมู่	สนใจมากที่สุด, มาก, ปานกลาง, น้อย, น้อยที่สุด
6	ปัจจัยด้านอาชีพ 4	มีแนวโน้มความต้องการในตลาดแรงงานสูงขึ้นในกลุ่มวิชาชีพนั้น	ข้อมูลเชิงหมวดหมู่	สนใจมากที่สุด, มาก, ปานกลาง, น้อย, น้อยที่สุด
7	ปัจจัยด้านอาชีพ 5	เป็นกลุ่มวิชาชีพที่องค์กรให้ความสำคัญเป็นอย่างมาก	ข้อมูลเชิงหมวดหมู่	สนใจมากที่สุด, มาก, ปานกลาง, น้อย, น้อยที่สุด
8	ปัจจัยด้านอาชีพ 6	มีตำแหน่งงานรองรับทั้งในภาครัฐบาลและเอกชน	ข้อมูลเชิงหมวดหมู่	สนใจมากที่สุด, มาก, ปานกลาง, น้อย, น้อยที่สุด
9	ปัจจัยด้านอาชีพ 7	เป็นกลุ่มวิชาชีพที่สอดคล้องกับสภาวะเศรษฐกิจปัจจุบันและอนาคต	ข้อมูลเชิงหมวดหมู่	สนใจมากที่สุด, มาก, ปานกลาง, น้อย, น้อยที่สุด
10	ปัจจัยด้านอาชีพ 8	ความรู้ในกลุ่มวิชาชีพนั้นสามารถนำไปประยุกต์ใช้งานได้เกือบทุกอาชีพ	ข้อมูลเชิงหมวดหมู่	สนใจมากที่สุด, มาก, ปานกลาง, น้อย, น้อยที่สุด
11	ปัจจัยด้านอาชีพ 9	เป็นกลุ่มวิชาชีพที่น่าจะสามารถสร้างความเจริญก้าวหน้าได้อย่างรวดเร็ว	ข้อมูลเชิงหมวดหมู่	สนใจมากที่สุด, มาก, ปานกลาง, น้อย, น้อยที่สุด
12	ปัจจัยด้านอาชีพ 10	เป็นกลุ่มวิชาชีพที่สามารถสร้างความมั่นคงในชีวิตได้	ข้อมูลเชิงหมวดหมู่	สนใจมากที่สุด, มาก, ปานกลาง, น้อย, น้อยที่สุด

***ตารางที่ 2** (ต่อ)*

ที่	กลุ่มข้อคำถาม	คำอธิบายข้อมูล	ประเภทข้อมูล	หมายเหตุ
13	บัจจัยด้านอาชีพ 11	เป็นกลุ่มวิชาชีพที่มีบรรยากาศและสิ่งแวดล้อมในการทำงานอยู่ภายในสำนักงาน	ข้อมูลเชิงหมวดหมู่	สนใจมากที่สุด, มาก, ปานกลาง, น้อย, น้อยที่สุด
14	บัจจัยด้านอาชีพ 12	เป็นกลุ่มวิชาชีพที่เปิดโอกาสให้ได้ใช้สติปัญญาและความคิดอยู่ตลอดเวลา	ข้อมูลเชิงหมวดหมู่	สนใจมากที่สุด, มาก, ปานกลาง, น้อย, น้อยที่สุด
15	กลุ่มวิชาชีพ	สาขาวิชาชีพ	ข้อมูลเชิงหมวดหมู่	DT – เทคโนโลยีดิจิทัล DC – นิเทศศาสตร์ดิจิทัล

และใช้ขั้นตอนวิธีเจสี่สิบแปด (J48) นาอีฟเบย์ (Naive Bayes) วันอาร์ (OneR) ไอบีเค (IBK) และเอสเอ็มโอ (SMO) ตามลำดับ โดยใช้ค่าพารามิเตอร์ของแต่ละขั้นตอนวิธีตามการตั้งค่าเริ่มต้นของโปรแกรมเวก้า ชุดข้อมูลสำหรับการสร้างแบบจำลองเป็นชุดข้อมูลที่พร้อมสำหรับการวิเคราะห์ จำนวน 226 ระเบียน แบ่งออกเป็นชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ (Test Data) งานวิจัยนี้ได้ใช้วิธีการทดสอบแบบไขว้ (Cross-Validation) โดยแบ่งข้อมูลออกเป็น 10 ส่วน 20 ส่วน และ 30 ส่วน ซึ่งในรอบแรกจะนำข้อมูลส่วนแรกเป็นชุดข้อมูลทดสอบ แล้วส่วนที่เหลือเป็นชุดข้อมูลเรียนรู้เพื่อนำไปสร้างแบบจำลอง ทำเช่นนี้ไปเรื่อย ๆ โดยในแต่ละรอบจะสลับส่วนกันเป็นชุดข้อมูลทดสอบจนกว่าจะครบทุกส่วน ทั้งนี้ เนื่องจากการแบ่งส่วนข้อมูลดังกล่าวเป็นการแบ่งกลุ่มที่นิยมใช้เป็นมาตรฐาน และมีจำนวนเพิ่มขึ้นในปริมาณเท่ากัน

4.5 การวัดประสิทธิภาพของแบบจำลอง (Model Evaluation)

งานวิจัยนี้ได้ใช้ชุดข้อมูลทดสอบมาทำการทดสอบแบบไขว้ โดยแบ่งข้อมูลออกเป็น 10 ส่วน 20 ส่วน และ 30 ส่วน เพื่อคำนวณค่าประสิทธิภาพของแบบจำลองที่ใช้ทำนายสาขาวิชาชีพของนักศึกษา โดยจะคัดเลือกแบบจำลองที่ให้ค่าประสิทธิภาพสูงสุด ซึ่งค่าประสิทธิภาพดังกล่าวคำนวณจากตารางประเมินผลลัพธ์ (Confusion Matrix) ของการทำนายสาขาวิชาชีพของนักศึกษา (Prediction) เปรียบเทียบกับสาขาวิชาชีพจริงที่นักศึกษาเลือกเรียน (Actual) ซึ่งสามารถจำแนกประเภทของข้อมูลที่ใช้เพื่อประเมินผลตามตารางที่ 2 ได้แก่

1) ข้อมูลที่แบบจำลองทำนายว่านักศึกษาจะเลือกสาขาวิชาชีพด้านเทคโนโลยีดิจิทัล (DT) และค่าคำตอบจริงของนักศึกษาได้เลือกเรียนในสาขาวิชาดังกล่าว (True Positive: TP)

2) ข้อมูลที่แบบจำลองทำนายว่านักศึกษาจะเลือกสาขาวิชาชีพด้านเทคโนโลยีดิจิทัล (DT) แต่ค่าคำตอบจริงของนักศึกษาไม่ได้เลือกเรียนในสาขาวิชาดังกล่าว (False Positive: FP)

3) ข้อมูลที่แบบจำลองทำนายว่านักศึกษาจะไม่เลือกสาขาวิชาชีพด้านเทคโนโลยีดิจิทัล (DT) และค่าคำตอบจริงของนักศึกษาก็ไม่ได้เลือกเรียนในสาขาวิชาดังกล่าว (True Negative: TN)

4) ข้อมูลที่แบบจำลองทำนายว่านักศึกษาจะไม่เลือกสาขาวิชาชีพด้านเทคโนโลยีดิจิทัล (DT) แต่ค่าคำตอบจริงของนักศึกษาได้เลือกเรียนในสาขาวิชาดังกล่าว (False Negative: FN)

***ตารางที่ 3** ตารางประเมินผลลัพธ์ (Confusion Matrix)*

		ค่าทำนาย	
		DT - เทคโนโลยีดิจิทัล	DC - นิเทศศาสตร์ดิจิทัล
ค่าจริง	DT - เทคโนโลยีดิจิทัล	TP	FN
	DC - นิเทศศาสตร์ดิจิทัล	FP	TN

ค่าจากตารางประเมินผลลัพธ์ สามารถนำมาคำนวณค่าประสิทธิภาพของแบบจำลอง ได้แก่ ค่าความถูกต้อง (Accuracy) ค่าความระลึก (Recall) ค่าความแม่นยำ (Precision) และค่าถ่วงดุล (F-measure) ดังแสดงในสมการที่ 1 - 4 [32]

$$\text{Accuracy} = \frac{TP + TN}{(TP + TN + FP + FN)} \times 100\% \quad (1)$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \times 100\% \quad (2)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \times 100\% \quad (3)$$

$$\text{F-measure} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (4)$$

5. ผลการดำเนินงาน

งานวิจัยนี้ได้ใช้ชุดข้อมูลจากแบบสอบถามความคิดเห็นของนักศึกษาสำนักวิชาศาสตร์และศิลป์ดิจิทัล จำนวน 226 ระเบียน

ประกอบด้วย 15 แอตทริบิวต์ ใช้เทคนิคการทำเหมืองข้อมูลแบบจำแนกประเภทสร้างแบบจำลองด้วยขั้นตอนวิธีต้นไม้ตัดสินใจ นาอีฟเบย์ วันอาร์ ซัพพอร์ตเวคเตอร์แมชชีน และเพื่อนบ้านใกล้ที่สุด โดยแบ่งชุดข้อมูลทดสอบแบบไขว้ออกเป็น 10 ส่วน 20 ส่วน และ 30 ส่วน แล้วนำแบบจำลองที่ได้มาประเมินประสิทธิภาพด้วยค่าความแม่นยำ ค่าความระลึก ค่าความถูกต้อง และค่าถ่วงดุล ผลการวิเคราะห์ประสิทธิภาพของแบบจำลองได้ผลดังตารางที่ 3 และตารางที่ 4

ตารางที่ 4 *ผลการคำนวณค่าความแม่นยำ และค่าความระลึก*

ขั้นตอนวิธี	ค่าความแม่นยำ			ค่าความระลึก		
	10 ส่วน	**20 ส่วน**	**30 ส่วน**	**10 ส่วน**	**20 ส่วน**	**30 ส่วน**
ต้นไม้ตัดสินใจ	**79.8%**	75.0%	74.9%	**83.3%**	70.0%	80.8%
นาอีฟเบย์	80.0%	**81.0%**	80.2%	77.5%	**80.0%**	79.2%
วันอาร์	79.3%	**79.6%**	79.4%	65.0%	66.7%	**67.5%**
ซัพพอร์ตเวคเตอร์แมชชีน	81.5%	81.4%	**82.1%**	**73.3%**	71.7%	**73.3%**
เพื่อนบ้านใกล้ที่สุด	**69.1%**	69.0%	67.5%	64.2%	**67.1%**	60.0%

จากตารางที่ 4 จะเห็นได้ว่า ค่าความแม่นยำของแบบจำลองจากขั้นตอนวิธีซัพพอร์ตเวคเตอร์แมชชีนมีค่าความแม่นยำมากที่สุดเท่ากับ 82.1% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 30 ส่วน รองลงมาคือขั้นตอนวิธีนาอีฟเบย์มีค่าความแม่นยำ 81.0% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 20 ส่วน ขั้นตอนวิธีต้นไม้ตัดสินใจค่าความแม่นยำ 79.8% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 10 ส่วน ขั้นตอนวิธีวันอาร์มีค่าความแม่นยำ 79.6% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 20 ส่วน และขั้นตอนวิธีเพื่อนบ้านใกล้ที่สุดมีค่าความแม่นยำน้อยที่สุดเท่ากับ 69.1% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 10 ส่วน

ในส่วนของค่าความระลึกจะเห็นได้ว่า ค่าความระลึกของแบบจำลองจากขั้นตอนวิธีต้นไม้ตัดสินใจมีค่าความระลึกมากที่สุดเท่ากับ 83.3% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 10 ส่วน รองลงมาคือขั้นตอนวิธีนาอีฟเบย์มีค่าความระลึก 80.0% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 20 ส่วน ขั้นตอนวิธีซัพพอร์ตเวคเตอร์แมชชีนมีค่าความระลึก 73.3% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 10 และ 30 ส่วน ขั้นตอนวิธีวันอาร์มีค่าความระลึกอยู่ที่ 67.5% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 30 ส่วน และขั้นตอนวิธีเพื่อนบ้านใกล้ที่สุดมีค่าความระลึกน้อยที่สุดเท่ากับ 67.1% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 20 ส่วน

จากตารางที่ 4 จะเห็นได้ว่า ค่าความถูกต้องของแบบจำลองจากขั้นตอนวิธีนาอีฟเบย์มีค่าความถูกต้องมากที่สุดเท่ากับ 89.6% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 20 ส่วน รองลงมาคือขั้นตอนวิธีต้นไม้ตัดสินใจและซัพพอร์ตเวคเตอร์แมชชีน ซึ่งมีค่าความถูกต้องเท่ากันคือ 89.3% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 10 และ 30 ส่วน ตามลำดับ ขั้นตอนวิธีวันอาร์มีค่าความถูกต้อง 87.0% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 30 ส่วน และขั้นตอนวิธีเพื่อนบ้านใกล้ที่สุดมีค่าความถูกต้องน้อยที่สุดเท่ากับ 80.6% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 10 และ 20 ส่วน

ตารางที่ 5 *ผลการคำนวณค่าความถูกต้อง และค่าถ่วงดุล*

ขั้นตอนวิธี	ค่าความถูกต้อง			ค่าถ่วงดุล		
	10 ส่วน	**20 ส่วน**	**30 ส่วน**	**10 ส่วน**	**20 ส่วน**	**30 ส่วน**
ต้นไม้ตัดสินใจ	**89.3%**	84.8%	85.4%	**81.5%**	72.4%	77.7%
นาอีฟเบย์	88.7%	**89.6%**	89.0%	78.7%	**80.5%**	79.7%
วันอาร์	86.7%	86.4%	**87.0%**	71.2%	72.3%	**72.8%**
ซัพพอร์ตเวคเตอร์แมชชีน	88.3%	88.7%	**89.3%**	77.1%	76.1%	**77.4%**
เพื่อนบ้านใกล้ที่สุด	**80.6%**	**80.6%**	80.0%	**66.5%**	65.0%	63.4%

ในส่วนของค่าถ่วงดุลจะเห็นได้ว่า ค่าถ่วงดุลของแบบจำลองจากขั้นตอนวิธีต้นไม้ตัดสินใจมีค่าถ่วงดุลมากที่สุดเท่ากับ 81.5% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 10 ส่วน รองลงมาคือขั้นตอนวิธีนาอีฟเบย์มีค่าถ่วงดุล 80.5% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 20 ส่วน ขั้นตอนวิธีซัพพอร์ตเวคเตอร์แมชชีนมีค่าถ่วงดุล 77.4% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 30 ส่วน ขั้นตอนวิธีวันอาร์มีค่าถ่วงดุลอยู่ที่ 72.8% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 30 ส่วน และขั้นตอนวิธีเพื่อนบ้านใกล้ที่สุดมีค่าถ่วงดุลน้อยที่สุดเท่ากับ 66.5% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 10 ส่วน

6. อภิปรายผล

เมื่อพิจารณาค่าประสิทธิภาพที่ดีที่สุดของทุกขั้นตอนวิธี สามารถสรุปได้ว่า แบบจำลองจากขั้นตอนวิธีต้นไม้ตัดสินใจมีค่าประสิทธิภาพที่ดีกว่าขั้นตอนวิธีอื่น โดยมีค่าความแม่นยำ 79.8% ซึ่งบ่งบอกถึงความสามารถในการทำนายสาขาวิชาวิชาชีพของนักศึกษาด้วยความแม่นยำสูง ค่าความระลึกที่มากที่สุดเท่ากับ 83.3% หมายความว่า แบบจำลองสามารถระบุ 83.3% ของนักเรียนที่เลือกสาขาวิชาชีพด้านเทคโนโลยีดิจิทัล (DT) ได้อย่างถูกต้อง ค่าความถูกต้อง 89.3% หมายความว่า การทำนายนักเรียนที่เลือกสาขาวิชาชีพด้านเทคโนโลยีดิจิทัล (DT) โดยใช้อัลกอริทึมนี้มีความถูกต้องโดยรวมที่ 89.3% และค่าถ่วงดุล 81.5% บ่งบอกถึงความสมดุลที่ดีระหว่างค่าความแม่นยำและค่าความระลึก ค่าประสิทธิภาพเหล่านี้มีความสำคัญในการระบุนักเรียนที่ต้องการความช่วยเหลือเพิ่มเติมในการแนะนำอาชีพ ดังนั้น งานวิจัยนี้จึงเลือกแบบจำลองจากขั้นตอนวิธีต้นไม้ตัดสินใจมาใช้ทำนายการเลือกสาขาวิชาชีพของนักศึกษา เนื่องจากแบบจำลองดังกล่าวสามารถทำนายสาขาอาชีพของนักเรียนได้อย่างแม่นยำ และสามารถระบุนักเรียนที่เลือกสาขาวิชาชีพของเทคโนโลยีดิจิทัล (DT) ได้ดี นอกจากนี้ ขั้นตอนวิธีต้นไม้ตัดสินใจยังมีความเหมาะสมกับข้อมูลเชิงหมวดหมู่ สามารถสร้างแบบจำลองที่มีโครงสร้างที่เข้าใจได้ง่ายและมีความสัมพันธ์ของข้อมูลที่ซับซ้อน ตลอดจนมีประสิทธิภาพกับชุดข้อมูลขนาดเล็ก ซึ่งเหมาะกับบริบทของงานวิจัยนี้

การแบ่งชุดข้อมูลทดสอบแบบไขว้จำนวน 30 ส่วน ส่งผลให้ค่าความถูกต้องและค่าความแม่นยำมีค่าสูง ในขณะที่การแบ่งชุดข้อมูลทดสอบแบบไขว้จำนวน 10 และ 20 ส่วน ส่งผลให้ค่าความระลึกและความถ่วงดุลมีค่าสูง ตามลำดับ และเมื่อพิจารณาค่าประสิทธิภาพโดยเฉลี่ยของทุกขั้นตอนวิธี พบว่า ไม่ว่าจะแบ่งชุดข้อมูลทดสอบแบบไขว้จำนวน 10 20 หรือ 30 ส่วน ต่างล้วนให้ค่าประสิทธิภาพที่ใกล้เคียงกันมาก โดยจะมีความแตกต่างกันประมาณ 1% แต่การแบ่งชุดข้อมูลทดสอบแบบไขว้จำนวน 10 ส่วนจะให้ค่าประสิทธิภาพสูงกว่าการแบ่งเป็น 20 และ 30 ส่วน

นอกจากนี้ เมื่อพิจารณากฎที่ได้จากแบบจำลองวิธีต้นไม้ตัดสินใจซึ่งเป็นแบบจำลองที่เข้าใจง่ายและให้ค่าความระลึกและค่าถ่วงดุลมากที่สุด แบบจำลองดังกล่าวได้แสดงให้เห็นความสัมพันธ์ของปัจจัยที่สำคัญจากแบบสอบถามที่นักศึกษาใช้ในการเลือกสาขาวิชาชีพ ได้แก่ นักศึกษาสาขาวิชาเทคโนโลยีดิจิทัลให้ความสนใจในประเด็นภาวะเศรษฐกิจปัจจุบันและมีแนวโน้มในการเลือกอาชีพที่มีตำแหน่งรองรับทั้งในภาครัฐบาลและเอกชนมากกว่าสาขาวิชานิเทศศาสตร์ดิจิทัล นอกจากนี้ นักศึกษาสาขาวิชาเทคโนโลยีดิจิทัลยังมีความสนใจอย่างมากที่จะเลือกกลุ่มวิชาชีพที่สามารถหางานทำได้ง่ายเมื่อสำเร็จการศึกษาและได้รับค่าตอบแทนสูง ส่วนนักศึกษาสาขาวิชานิเทศศาสตร์ดิจิทัลจะมีความสนใจในการเลือกกลุ่มวิชาชีพเช่นเดียวกับสาขาวิชาเทคโนโลยีดิจิทัลในประเด็นที่มีตำแหน่งรองรับทั้งในภาครัฐบาลและเอกชน สามารถหางานทำได้ง่ายเมื่อสำเร็จการศึกษา และได้ทำงานในกลุ่มวิชาชีพที่ได้รับค่าตอบแทนสูง แต่นักศึกษาสาขาวิชานิเทศศาสตร์ดิจิทัลจะเลือกกลุ่มวิชาชีพที่องค์กรให้ความสำคัญมากกว่าและให้ความสนใจกลุ่มวิชาชีพที่มีบรรยากาศและสิ่งแวดล้อมในการทำงานอยู่ภายในสำนักงานมากกว่านักศึกษาสาขาวิชาเทคโนโลยีดิจิทัล จากกฎที่ได้กล่าวข้างต้นสามารถนำไปประกอบการตัดสินใจแก่นักศึกษาที่ยังไม่ได้เลือกกลุ่มอาชีพหรือกำลังตัดสินใจในการเลือกเรียนสาขาวิชาชีพในอนาคตได้

7. สรุป

งานวิจัยนี้ได้สร้างและเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายการเลือกสาขาวิชาชีพของสำนักวิชาศาสตร์และศิลป์ดิจิทัลโดยใช้เทคนิคการจำแนกประเภท ซึ่งเป็นเทคนิคหนึ่งของการทำเหมืองข้อมูล โดยสร้างแบบจำลองด้วยขั้นตอนวิธีต้นไม้ตัดสินใจ นาอีฟเบย์ วันอาร์ ซัพพอร์ตเวคเตอร์แมชชีน และเพื่อนบ้านใกล้ที่สุด และทดลองแบ่งชุดข้อมูลทดสอบแบบไขว้ออกเป็น 10 20 และ 30 ส่วน จากนั้นจึงนำแบบจำลองที่ได้มาประเมินประสิทธิภาพด้วยค่าความแม่นยำ ค่าความระลึก ค่าความถูกต้อง และค่าถ่วงดุล ผลการวิเคราะห์ประสิทธิภาพของแบบจำลองพบว่า วิธีนาอีฟเบย์สามารถสร้างแบบจำลองที่มีค่าความถูกต้องมากที่สุดเท่ากับ 89.6% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 20 ส่วน วิธีซัพพอร์ตเวคเตอร์แมชชีนสามารถสร้างแบบจำลองที่มีค่าความแม่นยำมากที่สุดเท่ากับ 82.1% เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 30 ส่วน และวิธีต้นไม้ตัดสินใจสามารถสร้างแบบจำลองที่มีค่าความระลึกและค่าถ่วงดุลมากที่สุดเท่ากับ 83.3% และ 81.5% ตามลำดับ เมื่อแบ่งชุดข้อมูลทดสอบแบบไขว้เป็น 10 ส่วน ดังนั้น งานวิจัยนี้จึงเลือกใช้แบบจำลองที่สร้างจากขั้นตอนวิธีต้นไม้ตัดสินใจ

เนื่องจากแบบจำลองดังกล่าวสามารถทำนายสาขาวิชาชีพของนักศึกษาได้อย่างถูกต้องตามผลการวิเคราะห์ประสิทธิภาพของแบบจำลอง

ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลในงานวิจัยครั้งนี้ คือแบบจำลองที่สามารถนำไปทำนายการเลือกสาขาวิชาชีพของนักศึกษาสำนักวิชาศาสตร์และศิลป์ดิจิทัลได้ด้วยค่าประสิทธิภาพสูงสุด ซึ่งจะถูกนำไปใช้พัฒนาเป็นส่วนหนึ่งของเว็บไซต์ระบบแนะนำสาขาวิชาชีพ เพื่อประกอบการตัดสินใจแก่ผู้ที่กำลังตัดสินใจเลือกสาขาวิชาชีพในอนาคต อีกทั้งยังสามารถนำไปเป็นแนวทางให้กับผู้ที่สนใจวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูลได้

8. ข้อเสนอแนะ/งานวิจัยในอนาคต

8.1 เพื่อให้แบบจำลองสามารถนำไปใช้ประโยชน์ได้มากขึ้น ในอนาคตอาจเพิ่มจำนวนแอตทริบิวต์และจำนวนระเบียนในชุดข้อมูลที่นำมาสร้างแบบจำลอง รวมถึงการจัดการกับข้อมูลที่ไม่สมดุล และการปรับแต่งค่าพารามิเตอร์เพื่อเพิ่มความถูกต้องของแบบจำลอง

8.2 เพิ่มเติมขั้นตอนวิธีสร้างแบบจำลองเพื่อให้เกิดความหลากหลายและได้ประสิทธิภาพมากยิ่งขึ้น

9. เอกสารอ้างอิง

[1] J. Suknoi. *Decision-Making to Further Study at a Higher Education Level of High School Students at Thasala Prasitsuksa School,* An Independent Study Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Business Administration in Finance and Banking, Ramkhamhaeng University, 2020.

[2] Institute of Digital Arts and Science, Suranaree University of Technology, *About the Institute of Digital Arts and Science*. Available Online at https://digitech.sut.ac.th/about-program.php, accessed on 29 January 2023.

[3] R. Thepprasit and C. Sanrach. "The Analysis of Factors Affecting Choosing a Major of Undergraduate Students of the Faculty of Education by Using Data Mining Technique." *Journal of Graduate Studies Valaya Alongkorn Rajabhat University,* Vol. 14, No. 1, pp. 134 - 146, January - April, 2020.

[4] C. H. A. Yu. *Data Mining and Exploration from Traditional Statistics to Modern Data Science,* 1st Ed., CRC Press, New York, 2022.

[5] A. Chutipascharoen and C. Sanrach. "A Comparison of the Efficiency of Algorithms and Feature Selection Methods for Predicting the Success of Personal Overseas Money Transfer." *KKU Research Journal of Humanities and Social Sciences (Graduate Studies),* Vol. 6, No. 3, pp. 105 - 113, September - December, 2018.

[6] J. Han, J. Pei, and H. Tong. *Data Mining: Concepts and Techniques,* Morgan Kaufmann, United States, 2022.

[7] S. Deegalla. *Nearest Neighbor Classification in High Dimensions,* Stockholm University, Sweden, 2024.

[8] D. A. Pisner and D. M. Schnyer. *Machine learning,* Academic Press, Massachusetts, United States, 2020.

[9] E. Kannan, S. Ravikumar, A. Anitha, S. Kumar, and M. Vijayasarathy. "Analyzing uncertainty in cardiotocogram data for the prediction of fetal risks based on machine learning techniques using rough set." *Journal of Ambient Intelligence and Humanized Computing,* pp. 1 - 13, 2021.

[10] F. M. Plumed, L. C. Ochando, C. Ferri, J. H. Orallo, M. Kull, N. Lachiche, M. J. Quintana, and P. A. Flach. "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories." *IEEE Transactions on Knowledge and Data Engineering,* Vol. 3, No. 8, pp. 3048 - 3061, 2021.

[11] M. Hussain, W. Zhu, W. Zhang, and S. M. Abidi. "Student engagement predictions in an e-learning system and their impact on student course assessment scores." *Computational Intelligence and Neuroscience,* Vol. 2018, No. 6, pp. 1 - 21, 2018.

[12] D. Solomon, T. Godwin and O. Yinka. "Predicting Students' Academic Performance using Artificial Neural Networks." *International Journal of Computer Applications,* Vol. 185, No. 19, pp. 1 - 7, 2023.

[13] V. Nakhipova, Y. Kerimbekov, Z. Umarova, L. Suleimenova, S. Botayeva, A. Ibashova, and N. Zhumatayev.

"Use of the Naive Bayes Classifier Algorithm in Machine Learning for Student Performance Prediction." *International Journal of Information and Education Technology,* Vol. 14, No. 1, pp. 92 - 98, 2024.

[14] M. A. A. Barrak and M. Al-Razgan. "Predicting Students Final GPA Using Decision Trees: A Case Study." *International Journal of Information and Education Technology,* Vol. 6, No. 7, pp. 528 - 533, 2016.

[15] P. Shayan and M. V. Zaanen. "Predicting Student Performance from Their Behavior in Learning Management Systems." *International Journal of Information and Education Technology,* Vol. 9, No. 5, pp. 337 - 341, 2019.

[16] M. Vaarma and H. Li. "Predicting student dropouts with machine learning: An empirical study in Finnish higher education." *Technology in Society,* Vol. 76, 2024. Article 102474.

[17] H. Karalar, C. Kapucu, and H. Gürüler. "Predicting students at risk of academic failure using ensemble model during pandemic in a distance learning system." *International Journal of Educational Technology in Higher Education,* Vol. 18, 2021. Article 63.

[18] G. Méndez, X. Ochoa, K. Chiluiza, and B. de Wever. "Curricular Design Analysis: A Data-Driven Perspective." *Journal of Learning Analytics,* Vol. 1, No. 3, pp. 84 - 119, 2014.

[19] D. Dennehy, K. Conboy, and J. Babu. "Adopting Learning Analytics to Inform Postgraduate Curriculum Design: Recommendations and Research Agenda." *Information Systems Frontiers,* Vol. 25, No. 4, pp. 1315 - 1331, 2023.

[20] R. A. P. Álvarez, J. M. Mahauad, K. Sharma, D. S. Opazo, and M. P. Sanagustín. "Characterizing Learners' Engagement in MOOCs: An Observational Case Study Using the NoteMyProgress Tool for Supporting Self-Regulation." *IEEE Transactions on Learning Technologies,* Vol. 13, No. 4, pp. 676 - 688, October-December, 2020.

[21] S. Chaveesuk, B. Khalid, M. Bsoul-Kopowska, E. Rostańska, and W. Chaiyasoonthorn. "Comparative analysis of variables that influence behavioral intention to use MOOCs." *PLoS ONE,* Vol. 17, No. 4, 2022, Article e0262037.

[22] A. Alamri, H. Alharbi, and R. Alturki. "Adaptive learning systems using machine learning algorithms for personalized education." *Journal of Educational Technology Development and Exchange,* Vol. 16, No. 2, pp. 123 - 140, 2023.

[23] S. Y. Chen and J-H. Wang. "Individual differences and personalized learning: a review and appraisal." *Universal Access in the Information Society,* Vol. 20, pp. 833 - 849, 2021.

[24] D. Chaudhary, H. Prajapati, R. Rathod, P. Patel, and R. Gurjwar. "Student Future Prediction Using Machine Learning." *International Journal of Scientific Research in Computer Science, Engineering and Information Technology,* Vol. 5, No. 2, pp. 1104 - 1108, 2019.

[25] A. Pandey and L. S. Maurya. "Prediction Of Undergraduate Students' Career Using Various Machine Learning And Ensemble Learning Algorithms." *Webology,* Vol. 18, No. 6, pp. 3506 - 3523, 2021.

[26] M. Tiwari and N. Jain. "Student Career Prediction Through Performance Evaluation of Machine Learning Algorithms." *Tuijin Jishu/Journal of Propulsion Technology,* Vol. 44, No. 6, pp. 8176 - 8191, 2023.

[27] P. Jaruteerapan. "Factors Influencing on Choosing Studying Programs in Bachelor's Degree Level in the Faculty of Management Science, Loei Rajabhat University." *Research and Development Journal, Loei Rajabhat University,* Vol. 10, No. 32, pp. 35 - 46, April - June, 2015.

[28] P. Wongthong, W. Kankaew, A. Kwankaew, and Y. Chomdeang. *Applied of Data Mining Techniques for Searching Characteristics of Computer Career.* Research Project, Rajamangala University of Technology Suvarnabhumi, 2019.

[29] W. Panyawong. "Utilization of Rule-Based Predicting Models Faculty of Architecture, Urban Design and Creative Arts Students at Mahasarakham University by Data Mining." *Journal of Architecture, Design, and Construction,* Vol. 2, No. 2, pp. 109 - 119, May - August, 2020.

[30] T. Wannaprapha. *Using Data Mining Techniques for Screening People Successfully in Undergraduate Studies of Educational Technology,* A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Engineering in Computer Engineering, Dhurakij Pundit University, 2021.

[31] J. Jareanying. *The prediction of student performance using data mining techniques with Rapid Miner,* A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Science, Srinakharinwirot University, 2020.

[32] S. Chitra and P. Srivaramangai. "Feature Selection Methods for Improving Classification Accuracy – A Comparative Study." *UGC Care Group I Listed Journal,* Vol. 10, No. 1, pp. 66 - 79, 2020.