การพยากรณ์แนวโน้มการเลือกสาขาของผู้สมัครเข้าศึกษาต่อระดับปริญญาตรี โดยใช้เทคนิคเหมืองข้อมูล

The Prediction of Bachelor Admission Trends in the University Using Data Mining Techniques

สุธารัตน์ ชาวนาฟาง (Sutarat Chaonafang)*

Received: September 30, 2023
Revised: December 25, 2023
Accepted: December 28, 2023

*ผู้นิพนธ์ประสานงาน: สุธารัตน์ ชาวนาฟาง (Sutarat Chaonafang) อีเมล: sutarat@webmail.npru.ac.th

DOI:10.14416/j.it.2025.v1.001

บทคัดย่อ

ปัญหาจำนวนนักศึกษาลดลงส่งผลกระทบต่อหลายมหาวิทยาลัยทั่วโลกทั้งในปัจจุบันและอนาคต และปัจจุบันมหาวิทยาลัยมีจำนวนเพิ่มขึ้นทั้งมหาวิทยาลัยภาครัฐและภาคเอกชน แต่อัตราลงทะเบียนเรียนในมหาวิทยาลัยลดลง ทำให้เกิดการแข่งขันระหว่างมหาวิทยาลัยในการรับสมัครนักศึกษา งานวิจัยนี้มีวัตถุประสงค์เพื่อพยากรณ์แนวโน้มการเลือกสาขาของผู้สมัครเข้าศึกษาต่อระดับปริญญาตรี และเพื่อใช้ในการตัดสินใจการจัดสรรทรัพยากรและพัฒนาหลักสูตรของมหาวิทยาลัย โดยใช้เทคนิคเหมืองข้อมูล แบบจำลองที่ใช้ในการพยากรณ์ประกอบด้วย 3 รูปแบบ ได้แก่ นาอีฟเบย์ การถดถอยโลจิสติกส์ และต้นไม้ตัดสินใจ การวิจัยครั้งนี้ใช้ชุดข้อมูลจากระบบสารสนเทศงานทะเบียนของสำนักส่งเสริมวิชาการและงานทะเบียน มหาวิทยาลัยราชภัฏนครปฐม เป็นชุดข้อมูลนักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏนครปฐม ตั้งแต่ปีการศึกษา พ.ศ. 2557 - พ.ศ. 2561 จำนวนทั้งสิ้น 89,847 รายการ ผลการปรียบเทียบประสิทธิภาพการจำแนกข้อมูลแบบจำลองของการพยากรณ์ด้วยวิธีการตรวจสอบแบบไขว้ พบว่า ผลการวัดประสิทธิภาพแบบจำลองต้นแบบที่ดีที่สุด คือ การถดถอยโลจิสติกส์ โดยให้ค่าความถูกต้อง 61.3% งานวิจัยนี้สามารถนำไปประยุกต์ใช้ในการคาดการณ์แนวโน้มการเลือกสาขาของผู้สมัครเข้าศึกษาต่อระดับปริญญาตรี การตัดสินใจเชิงกลยุทธ์ และการจัดสรรทรัพยากรของมหาวิทยาลัย

คำสำคัญ: เทคนิคเหมืองข้อมูล การพยากรณ์ นาอีฟเบย์ การถดถอยโลจิสติกส์ ต้นไม้ตัดสินใจ

Abstract

The declining number of students has become a pressing issue, impacting many universities worldwide, both in the present and the future. Currently, there is an increase in the number of public and private universities. However, university enrollment rates are decreasing, leading to intensified competition among universities for student admissions. The objective of this research is to predict fields of study trends in Bachelor degree of University applicants and to use this information for decision-making, resource allocation, and university curriculum development by using Data Mining Technique. The predictive models in this study comprise 3 formats: Naïve Bayes, Logistic Regression and Decision Tree. The dataset used in this research is sourced from the information system of the Academic Promotion and Registration Division at Nakhon Pathom Rajabhat University, encompassing undergraduate student records from the academic years 2014 to 2018, totaling 89,847 records. The results of comparing the performance of data classification models using K-Fold Cross Validation revealed that the best-performing prototype model is Logistic Regression, with an accuracy of 61.3%. This research can be applied to predict trends in prospective undergraduate

** สาขาวิศวกรรมซอฟต์แวร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม*

** Software Engineering Department, Faculty of Science and Technology, Nakhon Pathom Rajabhat University*

students' selection of academic majors, inform strategic decision-making, and facilitate resource allocation within the university.

Keywords: Data Mining, Prediction, Naïve Bayes, Logistic Regression, Decision Tree.

1. บทนำ

วิกฤตการศึกษาได้ส่งกระทบต่อหลายมหาวิทยาลัยทั่วโลก ในแง่ของจำนวนนักศึกษาที่ลดลง [1] เนื่องจากการเปลี่ยนแปลงทางประชากรศาสตร์ครั้งใหญ่ในศตวรรษนี้ [2] อัตราการเติบโตของประชากรลดลง [3] ประกอบกับปัจจุบันมหาวิทยาลัยมีจำนวนเพิ่มขึ้นและเติบโตอย่างรวดเร็วทั้งมหาวิทยาลัยภาครัฐและภาคเอกชน แต่อัตราการลงทะเบียนเรียนในมหาวิทยาลัยลดลง เมื่อการลงทะเบียนเรียนลดลงทำให้เกิดการแข่งขันระหว่างมหาวิทยาลัยในการรับสมัครนักศึกษาสูงขึ้น เพื่อแย่งชิงจำนวนนักศึกษาตามที่ตั้งเป้าหมายไว้ มหาวิทยาลัยได้เปลี่ยนรูปแบบการดำเนินงานเป็นเชิงพาณิชย์เพื่อดึงดูดนักศึกษา มีการแข่งขันทางการตลาดเพื่อสร้างจุดเด่นและดึงดูดให้นักศึกษาเข้ามาสมัครเรียน เกิดหลักสูตรการเรียนการสอนรูปแบบใหม่ โครงการรับตรง โควต้าพิเศษ และทุนการศึกษาต่าง ๆ ทำให้นักศึกษามีทางเลือกเพิ่มมากขึ้นในการตัดสินใจเข้าศึกษาต่อ แต่ส่งผลต่อคุณภาพการศึกษา เนื่องจากมหาวิทยาลัยต้องลดมาตรฐานการรับเข้าเรียน โดยมีความเข้มงวดในการรับสมัครลดลง [4] - [6]

จากปัญหาดังกล่าว ส่งผลกระทบต่อมหาวิทยาลัยทั้งในปัจจุบันและอาจจะเป็นปัญหาระยาวที่กำลังจะเกิดขึ้นในอนาคต เพราะความสามารถในการอยู่รอดและความยั่งยืนของมหาวิทยาลัยขึ้นอยู่กับจำนวนนักศึกษา เนื่องจากรายได้ของมหาวิทยาลัยส่วนหนึ่งมาจากค่าธรรมเนียมการศึกษา รายได้ส่วนนี้สามารถใช้ในการดำเนินงานต่าง ๆ ภายในมหาวิทยาลัย เช่น การพัฒนาและปรับปรุงหลักสูตร การเพิ่มศักยภาพของอาจารย์และนักศึกษา การจัดสรรทรัพยากร เป็นต้น จำนวนนักศึกษาจึงมีความสำคัญมาก เพราะยิ่งนักศึกษามีจำนวนมากเท่าไร มหาวิทยาลัยจะได้รับรายได้มากขึ้น

ดังนั้นทางรอดจากวิกฤตทางการศึกษาในยุคที่จำนวนนักศึกษาลดลง เพื่อให้มหาวิทยาลัยอยู่รอดในระยะยาว มหาวิทยาลัยควรเตรียมการเพื่อรองรับการเปลี่ยนแปลงในอนาคต ต้องมีการปรับตัวการวางแผนในการดำเนินงาน และสร้างกลยุทธ์เพื่อสร้างความได้เปรียบทางการแข่งขัน มหาวิทยาลัยควรประยุกต์ใช้เทคนิคเหมืองข้อมูลในด้านการศึกษา ซึ่งเป็นเทคนิคที่ได้รับความสนใจอย่างมากในช่วงไม่กี่ปีที่ผ่านมา [7] โดยใช้เทคนิคเหมืองข้อมูลเพื่อพยากรณ์แนวโน้มการเลือกสาขาของผู้สมัครเข้าศึกษาต่อระดับปริญญาตรี เนื่องจากการพยากรณ์ว่านักศึกษามีแนวโน้มเลือกสาขาใดในระดับปริญญาตรีถือเป็นสิ่งสำคัญ เป็นความต้องการที่สำคัญและเร่งด่วนในมหาวิทยาลัยส่วนใหญ่ เนื่องจากขั้นตอนการรับเข้าเรียนเป็นขั้นตอนที่สำคัญขั้นตอนหนึ่งในมหาวิทยาลัย และเป็นขั้นตอนที่ใช้เวลานาน [8] มหาวิทยาลัยส่วนใหญ่กำลังเผชิญกับความท้าทายในการวิเคราะห์ข้อมูลขนาดใหญ่แทนวิธีทางสถิติแบบเดิม [9] เพราะการประยุกต์ใช้เทคนิคเหมืองข้อมูลในด้านการศึกษาช่วยให้มหาวิทยาลัยได้เปรียบทางการแข่งขัน

จุดประสงค์ของงานวิจัยนี้เพื่อพัฒนาแบบจำลองการพยากรณ์แนวโน้มการเลือกสาขาของผู้สมัครเข้าศึกษาต่อระดับปริญญาตรี และเพื่อใช้ในการตัดสินใจและจัดสรรทรัพยากรและพัฒนาหลักสูตรของมหาวิทยาลัย การเปรียบเทียบประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning Algorithms) ที่ใช้สำหรับการพยากรณ์ ได้แก่ เทคนิคนาอีฟเบย์ เทคนิคการถดถอยโลจิสติกส์ และเทคนิคต้นไม้ตัดสินใจ โดยชุดข้อมูลที่ใช้ในการศึกษานี้ได้มาจากระบบสารสนเทศงานทะเบียนของสำนักส่งเสริมวิชาการและงานทะเบียน มหาวิทยาลัยราชภัฏนครปฐม เป็นชุดข้อมูลนักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏนครปฐม ตั้งแต่ปีการศึกษา พ.ศ. 2557 - พ.ศ. 2561

ผลลัพธ์ของงานวิจัยนี้ ช่วยให้มหาวิทยาลัยสามารถคาดการณ์แนวโน้มการเลือกสาขาของผู้สมัครเข้าศึกษาต่อระดับปริญญาตรี ช่วยคาดการณ์จำนวนหลักสูตรในอนาคตที่ตอบสนองต่อความต้องการของผู้เรียน การวางแผนเชิงกลยุทธ์สำหรับพัฒนาหลักสูตรใหม่ การวิเคราะห์และตัดสินใจการรับเข้าเรียน นอกจากนั้นยังสามารถวิเคราะห์แนวโน้มและผลลัพธ์รายได้ของมหาวิทยาลัย ซึ่งมีความสำคัญต่อการตัดสินใจวางแผนกลยุทธ์และจัดสรรทรัพยากรในอนาคตได้อย่างแม่นยำ ตลอดจนเพิ่มศักยภาพในการบริหารงานและนักศึกษา [10] - [12] โดยผู้ใช้สามารถดูผลลัพธ์การรายงานแนวโน้มการเลือกสาขาของผู้สมัครเข้าศึกษาต่อระดับปริญญาตรีผ่านเว็บแอปพลิเคชัน

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

เทคนิคเหมืองข้อมูล เป็นกระบวนการค้นหาความรู้จากข้อมูลจำนวนมาก เพื่อค้นหาความสัมพันธ์และรูปแบบที่เป็นประโยชน์ ประเภทการทำเหมืองข้อมูล เช่น การจำแนกประเภทข้อมูล (Classification) การจัดกลุ่มข้อมูล (Clustering) เป็นต้น [13] ส่วนการทำเหมืองข้อมูลด้านการศึกษา คือ การดึงข้อมูลที่เป็นประโยชน์จากชุดข้อมูลที่สร้างขึ้นในขอบเขตเกี่ยวกับด้านการศึกษา เพื่อการวิเคราะห์เชิงคาดการณ์ต่าง ๆ สามารถนำไปใช้ในการพัฒนาและปรับปรุงคุณภาพทางการศึกษา [14] เช่น การพยากรณ์การสมัครเรียน ผลการเรียน การลาออก และศักยภาพในการสำเร็จการศึกษา เป็นต้น

อนุกรมเวลา (Time Series) คือ ลำดับของตัวแปรที่มีระยะห่างเท่ากัน เป็นชุดของข้อมูลที่เก็บรวบรวมและเกิดขึ้นตามลำดับเวลา ณ ช่วงเวลาใดเวลาหนึ่ง (t) ช่วงเวลานี้อาจจะเป็นหน่วยเวลา เช่น นาที ชั่วโมง และปี เป็นต้น [15] ชุดข้อมูลจะถูกกำหนดด้วยเวลา มีความต่อเนื่องและเพิ่มขึ้นภายใต้เวลา จึงจำเป็นต้องเก็บข้อมูลอย่างละเอียดและต่อเนื่องตามระยะเวลาที่ต้องการ เช่น การสังเกตรายชั่วโมงเกี่ยวกับกระบวนการผลิตทางวิทยาศาสตร์ ลำดับของสินค้าที่ขายในซุปเปอร์มาร์เก็ต เป็นต้น การวิเคราะห์อนุกรมเวลาจึงมีความสำคัญมากในด้านสถิติ [16]

เทคนิคนาอีฟเบย์ เป็นอัลกอริทึมที่ไม่ซับซ้อนจึงเป็นที่นิยมใช้สำหรับการพยากรณ์ มีประสิทธิภาพในการคำนวณสูง มีความแม่นยำในการจำแนกประเภทที่ดี [17] เป็นอัลกอริทึมที่ถูกสร้างขึ้นด้วยหลักความน่าจะเป็นแบบมีเงื่อนไข โดยคุณลักษณะทั้งหมดเป็นอิสระต่อกัน [18] และคำนวณความน่าจะเป็นภายหลังจากการพิจารณาจากความน่าจะเป็นก่อนหน้า [19] ดังสมการที่ 1

$$P(B \mid A) = \frac{P(A \mid B) \times P(B)}{P(A)} \qquad (1)$$

P (B | A) คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ B เมื่อเกิดเหตุการณ์ A ขึ้นก่อน

P (A | B) คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ A เมื่อเกิดเหตุการณ์ B ขึ้นก่อน

P (A) คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ A

P (B) คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ B

เทคนิคการถดถอยโลจิสติกส์ เป็นแบบจำลองเชิงเส้นทั่วไป (Generalized Linear Model) [20] เป็นอัลกอริทึมการเรียนรู้ของเครื่องแบบมีผู้สอนใช้ในการทำนายตัวแปรตามแบบหมวดหมู่ โดยสามารถใช้ชุดข้อมูลแบบต่อเนื่องและไม่ต่อเนื่องสำหรับประมาณค่าน่าความจะเป็นและจัดประเภทข้อมูลใหม่ [21] มีแนวคิดที่แตกต่างจากนาอีฟเบย์ที่มีคุณลักษณะเป็นอิสระต่อกัน แต่การถดถอยโลจิสติกส์จะวิเคราะห์คุณลักษณะโดยมีคุณสมบัติร่วมกัน [22]

เทคนิคต้นไม้ตัดสินใจ เป็นอัลกอริทึมที่นิยมใช้ในการพยากรณ์สำหรับการจำแนกหมวดหมู่ข้อมูล สามารถวิเคราะห์ข้อมูลที่มีความซับซ้อนในรูปแบบที่ง่าย โดยเป็นอัลกอริทึมแผนผังการตัดสินใจที่มีลักษณะเป็นโครงสร้างข้อมูลชนิดลำดับชั้น [23] ประกอบด้วยโหนดที่อยู่บนสุดแสดงถึงคุณลักษณะ และกิ่งก้านของโหนดที่แสดงถึงค่าที่เป็นไปได้ โดยผู้วิจัยนำอัลกอริทึมแบบ ID3 (Iterative Dichotomiser 3) มาใช้ในการทดลอง ซึ่งเป็นอัลกอริทึมพื้นฐานที่ใช้ในการสร้างแผนผังการตัดสินใจจากชุดข้อมูล โดยใช้ทฤษฎีสารสนเทศ (Information Theory) เพื่อกำหนดคุณลักษณะที่มีข้อมูลมากที่สุด [24] และสร้างต้นไม้จากบนลงล่างโดยไม่มีการย้อนกลับ ข้อมูลที่ได้นั้นใช้เพื่อเลือกคุณลักษณะที่มีประโยชน์ที่สุดสำหรับการจำแนกประเภท [25]

การตรวจสอบแบบไขว้ (K-Fold Cross Validation) เป็นวิธีการทดสอบประสิทธิภาพแบบจำลองที่ได้รับความนิยมโดยวิธีการกำหนดส่วนของข้อมูล ข้อมูลจะถูกแบ่งออกเป็นข้อมูลการฝึก (Training Data) และข้อมูลการทดสอบ (Testing Data) การแบ่งข้อมูลออกเป็นหลาย ๆ ส่วนนี้เรียกว่า k-cross validation โดยข้อมูลจะถูกแบ่งออกเป็น k ส่วน ข้อมูลแต่ละส่วนจะเท่ากัน ประกอบด้วย k-1 fold เป็นข้อมูลการฝึก และ 1-fold เป็นข้อมูลการทดสอบ [13]

คอนฟิวชันเมทริกซ์ (Confusion Matrix) เป็นเครื่องมือที่ใช้สำหรับประเมินผลลัพธ์ของแบบจำลองการพยากรณ์ โดยต้องวัดประสิทธิภาพของแบบจำลองก่อนนำไปใช้ ซึ่งถือเป็นหนึ่งขั้นตอนที่มีความสำคัญในการสร้างแบบจำลอง เพื่อประเมินว่าอัลกอริทึมที่ใช้นั้นมีประสิทธิภาพ [21] โดยนำผลลัพธ์ที่ได้จากแบบจำลองมาสร้างตารางคอนฟิวชันเมทริกซ์ และประเมินแบบจำลองด้วยตัวชี้วัดค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าประสิทธิภาพโดยรวม (F1 Score) [26] ดังตารางที่ 1

ตารางที่ 1 *คอนฟิวชันเมทริกซ์*

		Prediction	
		Positive	**Negative**
Actual	**Positive**	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

จากตารางที่ 1 ค่าที่ได้จากการทำนาย มีรายละเอียดดังนี้

ค่า TP คือ จำนวนข้อมูลที่จำแนกมีผลการทำนาย เป็นบวก มีค่าเป็นบวก

ค่า TN คือ จำนวนข้อมูลที่จำแนกมีผลการทำนาย เป็นลบ มีค่าเป็นลบ

ค่า FP คือ จำนวนข้อมูลที่จำแนกผิดมีผลการทำนายเป็นบวก มีค่าเป็นลบ

ค่า FN คือ จำนวนข้อมูลที่จำแนกผิดมีผลการทำนาย เป็นบวก มีค่าเป็นบวก

2.2 งานวิจัยที่เกี่ยวข้อง

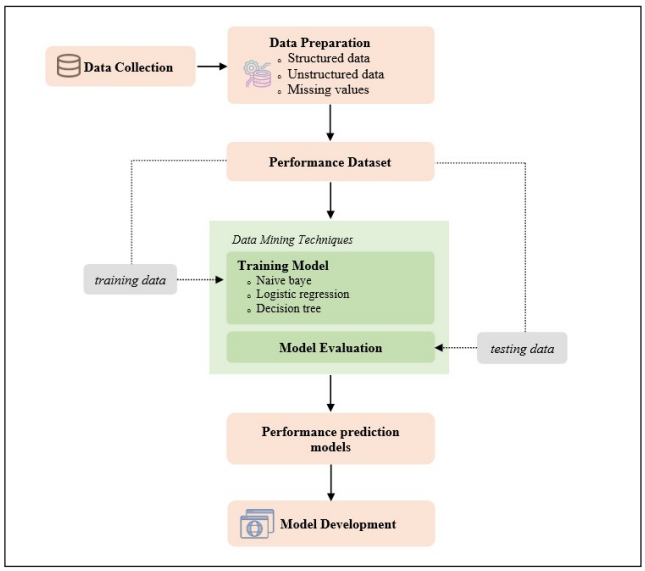
Terada Pinyo [27] ได้ศึกษาเกี่ยวกับการรายงานผลการวิเคราะห์ถดถอยโลจิสติกส์สำหรับงานวิจัย พบว่า การวิเคราะห์ถดถอยโลจิสติกส์เป็นกระบวนการวิจัยที่นิยมใช้สำหรับงานวิจัยเพื่อทำนายว่าเหตุการณ์หนึ่ง ๆ จะเกิดขึ้นหรือไม่ เกิดขึ้นได้มากน้อยเพียงใด โดยการคัดเลือกตัวแปรทำนายที่เหมาะสมเพื่อทำให้เปอร์เซ็นต์ความถูกต้องของการทำนายมีค่าสูงสุด

Selvaprabu Jeganathan และคณะ [28] ได้ศึกษาเกี่ยวกับการพยากรณ์การรับเข้าเรียนระบบบัณฑิตศึกษา โดยใช้เทคนิคการจำแนกประเภท ได้แก่ การถดถอยโลจิสติกส์ นาอีฟเบย์ และต้นไม้ตัดสินใจ ซึ่งการใช้เทคนิคเหมืองข้อมูลสามารถช่วยในการตัดสินใจเชิงกลยุทธ์ การพยากรณ์การรับเข้าศึกษาต่อ กระบวนการรับสมัคร การปรับปรุงผลการเรียนของนักศึกษา เป็นต้น จากการทดลองพบว่าการถดถอยโลจิสติกส์มีประสิทธิภาพสูงสุดในการพยากรณ์

Pornthep Rojanavasu [1] ได้วิเคราะห์ข้อมูลด้านการศึกษาโดยใช้กฎของความสัมพันธ์ (Association Rule) การจำแนกประเภทข้อมูล และต้นไม้ตัดสินใจด้วยวิธี ID3 เพื่อช่วยในการวางแผนการรับเข้าเรียน การตัดสินใจเกี่ยวกับเกรดเฉลี่ยและข้อมูลนักศึกษาที่สำเร็จการศึกษาเพื่อทำนายหลังสำเร็จการศึกษา พบว่า เทคนิคเหมืองข้อมูลสามารถช่วยกระบวนการรับเข้าเรียน โดยมีความสัมพันธ์อย่างมีนัยสำคัญระหว่างภูมิภาค ชื่อสาขา คณะและจังหวัด ผลลัพธ์นี้ช่วยตอบสนองต่อความต้องการเข้าศึกษาต่อ และสามารถทำนายการมีงานของนักศึกษา โดยสาขามีความสำคัญต่อการประกอบอาชีพในอนาคตของนักศึกษา

3. วิธีดำเนินการวิจัย

ขั้นตอนของงานวิจัยได้ดำเนินการตามกรอบแนวคิด CRISP-DM โดยเริ่มจากการเก็บรวบรวมข้อมูลและศึกษาความสัมพันธ์ข้อมูล แปลงข้อมูลให้อยู่ในรูปแบบ CSV หลังจากนั้นจัดเตรียมข้อมูลเพื่อให้เหมาะสมกับแบบจำลอง [29] สร้างแบบจำลอง การประเมินผลและการนำไปใช้ โดยสรุปวิธีดำเนินการวิจัย ดังภาพที่ 1



ภาพที่ 1 *กรอบแนวคิดงานวิจัย*

3.1 การรวบรวมข้อมูล (Data Collection)

มหาวิทยาลัยได้จัดเก็บข้อมูลเกี่ยวกับนักศึกษาทั้งหมดลงในระบบอิเล็กทรอนิกส์ โดยชุดข้อมูลที่ใช้ในการศึกษานี้ได้มาจากระบบสารสนเทศงานทะเบียนของสำนักส่งเสริมวิชาการและงานทะเบียน มหาวิทยาลัยราชภัฏนครปฐม เป็นชุดข้อมูลนักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏนครปฐม ตั้งแต่ปีการศึกษา พ.ศ. 2557 - พ.ศ. 2561 จำนวน 89,847 รายการ โดยแบ่งชุดข้อมูลออกเป็น 3 ส่วน ได้แก่ ข้อมูลด้านการสมัครเรียน ข้อมูลด้านนักศึกษา และข้อมูลผู้สำเร็จการศึกษา

ตัวอย่างชุดข้อมูล ได้แก่ ข้อมูลวุฒิการศึกษาและผลการศึกษาเดิมของผู้สมัคร ข้อมูลการสมัครเรียน ข้อมูลสาขา

ที่ผู้สมัครเลือก ข้อมูลผลการสมัครเรียน ข้อมูลผลการศึกษา หน่วยกิตที่ลงทะเบียนเรียน ข้อมูลสถานะของนักศึกษา และข้อมูลสถานะการทำงานของบัณฑิต เป็นต้น โดยสรุปคุณลักษณะจากข้อมูลนักศึกษา ดังตารางที่ 2

ตารางที่ 2 *คุณลักษณะจากข้อมูลนักศึกษา*

Table	Field Name	Description
Student_data	STUDENTID	Student identification number.
	STUDENTSEX	Female, Male
	BIRTHDATE	Date of Birth.
	ENTRYDEGREE	entry degree' Student.
	HOMEPROVINCEID	Province student address.
	STUDENTSTATUS	Student status Id.
	ADMITACADYEAR	Year'student entry.
	LEVELID	Number of level.
	FACULTYID	Number of faculty.
	PROGRAMCODE	Number of program.
	DEGREEID	Number of degree.
	ADMITDATE	Date'student entry.
	FINISHDATE	Date'student out/finish.
	COUNTTERNSTD	Count semester student enroll.
Graduate_data	APPROVEYEAR	Year of Student graduate.
	STUDENTID	Number of student.
	WORKINGSTATUS	Work status Id
	OFFICEPROVINCEID	Province graduate office.
	WORKINGSALARY	Income per mo graduate.
	WORKINGPOSITION	Work position graduate.
Applicant_data	ACADYEAR	Year of applicant.
	APPLICANTID	Number of applicant.
	QUOTACODE	Number of program applicant.
Applicant_data	BLOODGROUP	Blood of applicant.
	BIRTHDATE	Date of birth applicant.
	BIRTHPROVINCEID	Province of birth applicant.
	HOMEPROVINCEID	Province of address applicant.
	ENTRYDEGREEDES	Degree of applicant.
Level	LEVELID	Number of level.
	LEVELNAME	Name of level.
Degree	DEGREEID	Number of degree.
	DEGREENAME	Name of degree.
Faculty	FACULTYID	Number of faculty.
	FACULTYNAME	Name of faculty.
Program	PROGRAMID	Number of program.
	PROGRAMNAME	Name of program.
Province	PROVINCEID	Number of province.
	PROVINCENAME	Name of province.
Quota	QUOTAID	Number of Quotadetails.
	QUOTADETAILS	Name of Quotadetails.

3.2 การเตรียมข้อมูล (Data Preparation)

ขั้นตอนการเตรียมชุดข้อมูลเพื่อให้ได้ชุดข้อมูลที่เป็นมาตรฐานและเหมาะสมกับอัลกอริทึม เนื่องจากชุดข้อมูลที่รวบรวมจากฐานข้อมูลประกอบด้วยคุณลักษณะจำนวนมาก ข้อมูลที่ไม่สอดคล้องกัน ข้อมูลไม่สมบูรณ์หรือข้อมูลที่มีค่าว่าง ในการประมวลผลจึงต้องจัดการกับค่าข้อมูลที่มีว่าง และทำข้อมูลให้เป็นมาตรฐาน เนื่องจากอัลกอริทึมสามารถประมวลผลและได้ค่าที่แม่นยำมาจากการนำเข้าชุดข้อมูลที่ดี โดยการเตรียมข้อมูลมีดังนี้

3.2.1 การล้างข้อมูล (Data Cleaning) และการจัดการข้อมูลที่มีค่าว่าง (Missing Values) ในการตัดทิ้งรายการที่มีข้อมูลสูญหาย โดยเลือกข้อมูลที่ได้มาจากฐานข้อมูลว่าฟิลด์ใดที่มีค่าว่างเป็นจำนวนมากกว่า 50% ขึ้นไป จะใช้การตัดคอลัมน์ออก เพื่อไม่นำมาใช้ในการประมวลผลข้อมูล หลังจากนั้น

จึงทำการตรวจสอบข้อมูลแต่ละรายการอีกครั้ง หากข้อมูลที่ได้มามีเนื้อหาภายในไม่ถึง 80% ซึ่งส่วนใหญ่เกิดจากนักศึกษาที่มาสมัครเรียน แต่ไม่มาเรียนจึงทำให้ไม่พบข้อมูลของนักศึกษา เช่น ข้อมูลส่วนตัว ผลการเรียน เป็นต้น จึงจำเป็นต้องลบข้อมูลดังกล่าวออก เพื่อไม่ให้ข้อมูลที่ว่างมีจำนวนมากเกินความจำเป็น หลังจากนั้นจึงทำการเติมหรือแทนที่ค่าที่ขาดหาย เพิ่มเติมข้อมูลที่ผิดพลาด ซึ่งอาจเกิดจากการกรอกข้อมูลในระบบที่ผิดพลาด เช่น วัน/เดือน/ปีเกิด จาก พ.ศ. เป็น ค.ศ. ส่งผลให้ระบบคำนวณอายุผิดพลาด จึงจำเป็นต้องแทนที่ด้วยค่าเฉลี่ยของอายุที่ได้จากการคำนวณในส่วนที่ไม่เกิดข้อผิดพลาด แล้วนำกลับมาใส่ในข้อมูลส่วนที่ข้อผิดพลาด เพื่อทำให้การคำนวนค่าต่าง ๆ ที่ออกมาภายหลังไม่คลาดเคลื่อนจากค่าที่ไม่เกิดข้อผิดพลาดมากเกินไป

นอกจากนั้นยังตัดข้อมูลที่สามารถระบุตัวตนของนักศึกษาออก เช่น ชื่อ-นามสกุล ที่อยู่ รหัสนักศึกษา เลขประจำตัวประชาชน เบอร์โทรศัพท์ อีเมลหรือข้อมูลใด ๆ ที่สามารถระบุตัวบุคคลหรือสิ่งที่เกี่ยวข้องกับข้อมูลส่วนตัวของนักศึกษาได้จะถูกตัดออก

3.2.2 การแปลงข้อมูล (Data Transformation) เป็นกระบวนการแปลงข้อมูล และจัดโครงสร้างข้อมูลที่ซับซ้อนเพื่อให้ประมวลผลข้อมูลได้ง่ายขึ้น

การเตรียมข้อมูลจะนำชุดข้อมูลที่รวบรวมได้มาสร้างเป็นอนุกรมเวลาโดยเรียงตามลำดับปี ระดับการศึกษา เกรดเฉลี่ย และสาขาวิชา เป็นต้น ตั้งแต่ปีการศึกษา พ.ศ. 2557 - พ.ศ. 2561 ดังภาพที่ 2

ACADYEAR	ENTRYDEGREELEVELNAME	GPA	PROGRAMNAMEEN	STUDENTSEX
2556	HiSch	Good	Finance and Banking	F
2557	Grad	Ex	Physics	M
2557	VCM	Good	Public Health	F
2557	VCM	Good	Public Health	F
2557	VCM	Good	Public Health	F
2557	VCM	Good	Public Health	F
2557	VCM	None	Public Health	F
2557	VCA	Good	Public Health	M
2557	Grad	Good	Public Health	F
2557	VCM	Good	Public Health	F
2557	VCM	Good	Public Health	F
2557	VCM	Ex	Public Health	M
2557	VCM	Good	Public Health	F
2557	VCM	Good	Public Health	F
2557	VCM	Ex	Public Health	F
2557	VCM	Ex	Public Health	F
2557	VCM	Ex	Public Health	F
2557	VCM	Ex	Public Health	F
2557	VCM	Ex	Public Health	F
2557	VCM	Ex	Public Health	F
2557	HiSch	Good	Public Health	F

ภาพที่ 2 *ข้อมูลอนุกรมเวลาที่สร้างจากชุดข้อมูล*

3.3 การสร้างแบบจำลอง (Model)

แบบจำลองที่ใช้ในการพยากรณ์แนวโน้มการเลือกสาขาของผู้สมัครเข้าศึกษาต่อระดับปริญญาตรี โดยใช้เทคนิคการจำแนกประเภท 3 เทคนิค ได้แก่ นาอีฟเบย์ การถดถอยโลจิสติกส์ และต้นไม้ตัดสินใจ ซึ่งแต่ละเทคนิคมีทั้งส่งผลทางตรงและทางอ้อมต่อการพยากรณ์ที่แตกต่างกัน โดยการสร้างแบบจำลองใช้ซอฟต์แวร์ RapidMiner ช่วงการทดสอบเทคนิคนั้นซอฟต์แวร์จะปรับแต่งค่าที่ดีที่สุดและสรุปผลอย่างอัตโนมัติด้วยโหมด Turbo Prep

3.4 การประเมินผล (Evaluation)

การทดสอบประสิทธิภาพของแบบจำลองด้วยวิธี 10-fold cross validation โดยแบ่งข้อมูลออกเป็น 10 ส่วน แต่ละส่วนมีจำนวนข้อมูลเท่ากัน ข้อมูลจำนวน 9 ส่วน ใช้สำหรับการฝึกแบบจำลอง และข้อมูลชุดที่เหลือใช้สำหรับทดสอบ ซึ่งจะทำวนซ้ำไปจนครบ 10 ครั้ง

การประเมินผลลัพธ์ประสิทธิภาพของแบบจำลองการพยากรณ์โดยใช้คอนฟิวชันเมทริกซ์ ประกอบด้วยตัวชี้วัด ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าประสิทธิภาพโดยรวม โดยมีรายละเอียดดังนี้

ค่าความถูกต้อง คือ ประสิทธิผลของแบบจำลองที่มีเปอร์เซ็นต์ของจำนวนการทำนายที่ถูกต้องทั้งหมด ซึ่งเป็นจำนวนข้อมูลที่ทำนายถูกของคลาสดังสมการที่ 2

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \qquad (2)$$

ค่าความแม่นยำ คือ เปอร์เซ็นต์ของการสังเกตเชิงบวกที่ถูกต้อง เป็นการทำนายผลการพยากรณ์ว่าทำนายถูกกี่เปอร์เซ็นต์ ดังสมการที่ 3

$$\text{Precision} = TP / (TP + FP) \qquad (3)$$

ค่าความระลึก คือ เปอร์เซ็นต์ของผลบวกที่ทำนายได้อย่างถูกต้อง มีค่าเป็นบวกทั้งหมด ดังสมการที่ 4

$$\text{Recall} = TP / (TP + FN) \qquad (4)$$

ค่าประสิทธิภาพโดยรวม คือ ค่าเฉลี่ยของความแม่นยำและค่าความระลึก ดังสมการที่ 5

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \qquad (5)$$

3.5 การนำไปใช้งาน (Deployment)

จากขั้นตอนก่อนหน้านี้ทำให้ได้แบบจำลองที่เหมาะสมที่สุดเพื่อใช้เป็นตัวแบบในการพยากรณ์แนวโน้มการเลือกสาขาของผู้สมัครเข้าศึกษาต่อระดับปริญญาตรี โดยนำแบบจำลองที่ได้ไปพัฒนาในรูปแบบของเว็บแอปพลิเคชัน เพื่อให้ผู้ใช้งานง่ายและสะดวก โดยผู้วิจัยได้พัฒนาเว็บแอปพลิเคชันด้วยภาษา PHP และใช้ฐานข้อมูล MySQL และสร้าง Graph Report จากไลบรารี่ของ CanvasJS เพื่อเพิ่มความสวยงาม

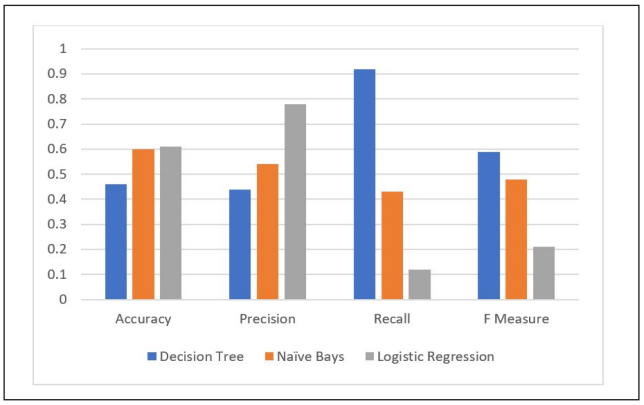
4. ผลการดำเนินงาน

ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ทั้ง 3 เทคนิค เพื่อหาแบบจำลองที่มีความถูกต้องแม่นยำมากที่สุด และนำมาสร้างตัวแบบการพยากรณ์ จากการเปรียบเทียบพบว่า เทคนิคนาอีฟเบย์ มีค่าความถูกต้อง 60.0% ค่าความแม่นยำ 54.1% ค่าความระลึก 42.5% และค่าประสิทธิภาพโดยรวม 47.6% เทคนิคการถดถอยโลจิสติกส์ มีค่าความถูกต้อง 61.3% ค่าความแม่นยำ 77.5% ค่าความระลึก 12.1% และค่าประสิทธิภาพโดยรวม 20.9% และเทคนิคต้นไม้ตัดสินใจ มีค่าความถูกต้อง 45.7% ค่าความแม่นยำ 43.5% ค่าความระลึก 92.1% และค่าประสิทธิภาพโดยรวม 59.1% ดังตารางที่ 3

ตารางที่ 3 *ประสิทธิภาพของแบบจำลองการพยากรณ์*

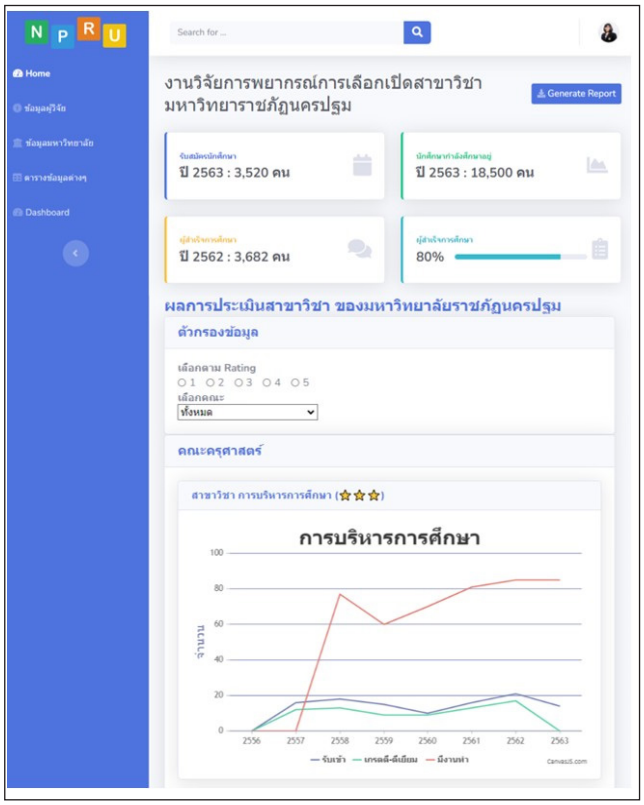
Model	Accuracy	Precision	Recall	F Measure
Naïve Bays	60.0%	54.1%	42.5%	47.6%
Logistic Regression	61.3%	77.5%	12.1%	20.9%
Decision Tree	45.7%	43.5%	92.1%	59.1%

จากตารางที่ 3 เทคนิคการถดถอยโลจิสติกส์เป็นแบบจำลองที่มีประสิทธิภาพดีที่สุดที่จะให้เปอร์เซ็นต์ของความถูกต้องในการพยากรณ์ที่มีค่าสูงสุดหากเปรียบเทียบกับแบบจำลองอื่น ๆ โดยการเลือกคุณลักษณะ (Feature Selection) ที่ใช้ในการพยากรณ์นั้นจะนำค่าต่าง ๆ จากการวัดประสิทธิภาพมาเลือกว่าเทคนิคใดมีค่าประสิทธิภาพมากที่สุดจึงใช้เทคนิคนั้น ดังภาพที่ 3



ภาพที่ 3 *ผลการเปรียบประสิทธิภาพของแบบจำลอง*

หลังจากทดสอบประสิทธิภาพของแบบจำลองที่ใช้ในการพยากรณ์จะได้แบบจำลองต้นแบบที่ดีสุดเพื่อนำไปพัฒนาระบบ เพื่อให้ผู้ใช้สามารถดูผลลัพธ์ผ่านเว็บแอปพลิเคชัน ดังภาพที่ 4 เป็นหน้าจอหลักของระบบการพยากรณ์ที่แสดงข้อมูลจำนวนการรับสมัครนักศึกษา จำนวนนักศึกษาที่กำลังศึกษาอยู่ จำนวนผู้สำเร็จการศึกษา จำนวนการมีงานทำ และแผนภูมิแสดงผลการประเมินสาขาวิชาของมหาวิทยาลัยราชภัฏนครปฐม



ภาพที่ 4 *หน้าจอระบบพยากรณ์*

เว็บแอปพลิเคชันจะแสดงแผนภูมิเปรียบเทียบข้อมูลของนักศึกษาของแต่ละสาขาและเรียงลำดับตามปีการศึกษา และสามารถจัดลำดับคะแนนความนิยม เพื่อให้ทราบว่าสาขาใดได้รับความนิยมในการเข้าศึกษาต่อ ซึ่งการให้ระดับคะแนนมีเกณฑ์ดังนี้

1 คะแนน หมายถึง ความนิยมน้อยที่สุด

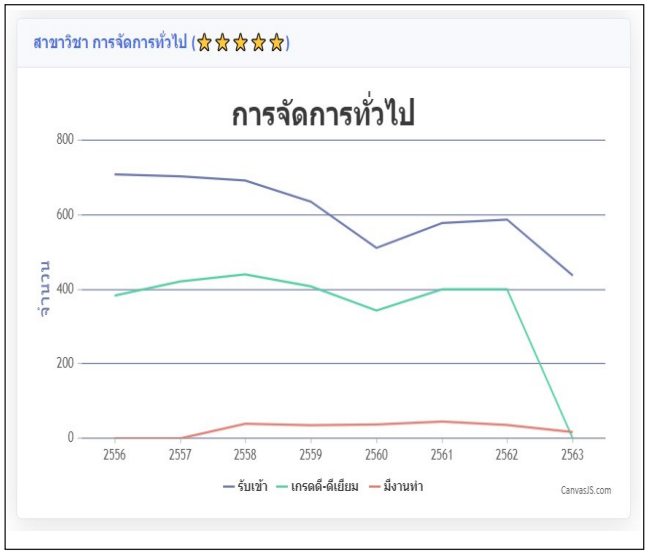
2 คะแนน หมายถึง ความนิยมน้อย

3 คะแนน หมายถึง ความนิยมปานกลาง

4 คะแนน หมายถึง ความนิยมมาก

5 คะแนน หมายถึง ความนิยมมากที่สุด

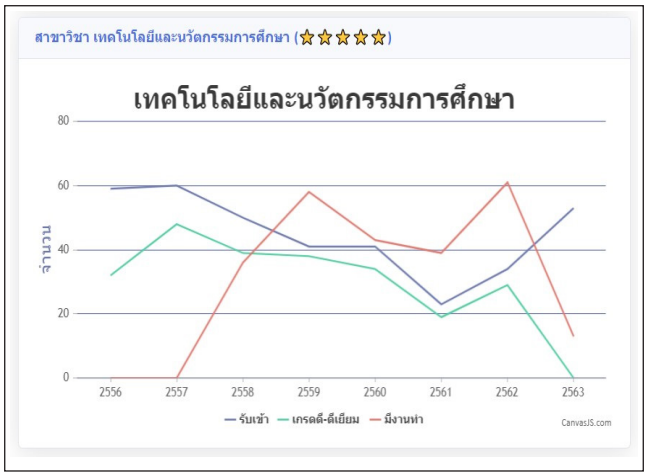
ตัวอย่างหน้าจอแสดงผลการประเมินสาขาที่ได้ระดับคะแนน 5 คะแนน โดยพิจารณาจากข้อมูลจำนวนนักศึกษาเข้าศึกษาต่อในแต่ละปีเป็นอันดับแรก ตามด้วยข้อมูลผลการเรียนของนักศึกษา และข้อมูลการมีงานทำของบัณฑิต ดังภาพที่ 5 สาขาการจัดการทั่วไปมียอดจำนวนนักศึกษาเข้าศึกษาต่อเป็นจำนวนมากอย่างต่อเนื่อง จึงได้ระดับคะแนน 5 คะแนน



ภาพที่ 5 หน้าจอแสดงผลการประเมินสาขาวิชาการจัดการทั่วไป

ตัวอย่างสาขาที่จำนวนนักศึกษาเข้าศึกษาต่อมีจำนวนน้อย แต่มีนักศึกษามีผลการเรียนดีและการมีงานทำของบัณฑิตจำนวนมาก ได้ระดับคะแนน 5 คะแนนเช่นกัน ดังภาพที่ 6

หน้าจอแสดงเปอร์เซ็นต์จำนวนของนักศึกษาของแต่ละคณะของมหาวิทยาลัยราชภัฏนครปฐม ดังภาพที่ 7



ภาพที่ 6 หน้าจอแสดงผลการประเมินสาขาวิชาเทคโนโลยีและนวัตกรรม

ภาพที่ 7 หน้าจอสรุปจำนวนนักศึกษาแยกตามคณะ

5. สรุป

การวิจัยเพื่อเปรียบเทียบประสิทธิภาพแบบจำลองการพยากรณ์แนวโน้มการเลือกสาขาของผู้สมัครเข้าศึกษาต่อระดับปริญญาตรี พบว่า แบบจำลองที่เหมาะสมที่สุดใช้ในการพยากรณ์คือ การถดถอยโลจิสติกส์ โดยมีประสิทธิภาพดีที่สุดที่จะให้เปอร์เซ็นต์ของความถูกต้องในการพยากรณ์ โดยมีค่าสูงสุดหากเปรียบเทียบกับแบบจำลองอื่น ๆ โดยมีค่าความถูกต้อง 61.3% การพยากรณ์แสดงให้เห็นถึงแนวโน้มจำนวนผู้สมัครเข้าศึกษาต่อระดับปริญญาตรีในปีการศึกษาถัดไป และต้นไม้การตัดสินใจให้ค่าความถูกต้องต่ำที่สุด 45.7%

ข้อเสนอแนะในการวิจัย 1) การเลือกปัจจัยด้านอื่น ๆ อาจจะส่งผลต่อการเลือกสาขาของผู้สมัครเข้าศึกษาต่อระดับปริญญาตรี เช่น ปัจจัยด้านสภาพทางครอบครัว

ข้อมูลพฤติกรรมของผู้สมัครเรียน ปัจจัยด้านหลักสูตร ข้อมูลโครงสร้างหลักสูตร ปัจจัยด้านอาจารย์ผู้สอน เป็นต้น 2) การเลือกเทคนิคการเรียนรู้ของเครื่องประเภทอื่น ๆ เพื่อนำมาเปรียบเทียบและวัดค่าความถูกต้อง เพื่อให้ได้อัลกอริทึมที่มีประสิทธิภาพมากที่สุด และ 3) ผลการวิจัยนี้สามารถนำไปใช้ประโยชน์ในการคาดการณ์แนวโน้มการเลือกสาขาของผู้สมัครเข้าศึกษาต่อระดับปริญญาตรีในปีการศึกษาถัดไปในอนาคต สามารถใช้ในการตัดสินใจและจัดสรรทรัพยากรและพัฒนาหลักสูตรของมหาวิทยาลัยได้อย่างมีประสิทธิภาพ

6. กิตติกรรมประกาศ

ขอขอบคุณ สำนักส่งเสริมวิชาการและงานทะเบียน มหาวิทยาลัยราชภัฏนครปฐม ที่ได้อนุญาตให้ใช้ชุดข้อมูลนักศึกษาระดับปริญญาตรีเพื่อนำมาใช้วิเคราะห์ในงานวิจัยนี้ และสถาบันวิจัยและพัฒนา มหาวิทยาลัยราชภัฏนครปฐม ที่ได้สนับสนุนทุนในการดำเนินการวิจัย

7. เอกสารอ้างอิง

[1] P. Rojanavasu. "Educational Data Analytics using Association Rule Mining and Classification." *2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON),* Thailand, pp. 142-145, 2019.

[2] W. L. Liu and W. McKibbin. "Macroeconomic Impacts of Global Demographic Change: The Case of Australia." *Asian Economic Papers,* Vol. 21, No. 3, pp. 78-111, 2022.

[3] D. Gu, K. Andreev, and M. E. Dupre. "Major Trends in Population Growth Around the World." *China CDC Weekly,* Vol. 3, No. 28, pp. 604 - 613, July, 2021.

[4] B. H. Kunaryo, M. Somantri, and A. Sofwan. "Applicant Mapping of Predicting Seriousness in Choosing Private University using Self Organizing Map (SOM) Algorithm." *International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE),* Indonesia, pp. 130-134, 2020.

[5] J. P. Wu, M. S. Lin, and C.L. Tsai. "A Predictive Model That Aligns Admission Offers with Student Enrollment Probability." *Education Sciences,* Vol. 13, No. 5, No. 440, April, 2023.

[6] R. Jeansuti, H. Jeansuti, S. Tanawutpornpinit, and S. Siriwatcharaworakun. "Approaches to Developing New Normal Teachers in Digital Literacy." *Journal of teacher professional development,* Vol. 2, No. 2, pp. 16-31, May-August, 2021.

[7] M. Yağcı. "Educational data mining: prediction of students' academic performance using machine learning algorithms." *Smart Learning Environments,* Vol. 9, No. 1, March, 2022.

[8] H. Bousnguar, L. Najdi, and A. Battou. "Forecasting approaches in a higher education setting." *Education and Information Technologies,* Vol. 27, pp. 1993-2011, 2021.

[9] H. A. Mengash. "Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems." *IEEE Access,* Vol. 8, pp. 55462-55470, 2020.

[10] H. Sabnani, M. More, P. Kudale, and S. Janrao. "Prediction of Student Enrolment Using Data Mining Techniques." *International Research Journal of Engineering and Technology (IRJET),* Vol. 5, No. 4, pp. 1830-1833, 2018.

[11] A. P.D. Cruz, Ma. L.B. Basallo, B.A.B. III, J. Aguilar, C.K.P. Calvo, J.C.T. Arroyo, and A.J.P. Delima. "Higher Education Institution (HEI) Enrollment Forecasting Using Data Mining Technique." *International Journal of Advanced Trends in Computer Science and Engineering,* Vol. 9, No. 2, pp. 2060-2064, 2020.

[12] I. Ognjanovic, D. Gasevic, and S. Dawson. "Using institutional data to predict student course selections in higher education." *The Internet and Higher Education,* Vol. 29, pp. 49-62, April, 2016.

[13] Muladi, U. Pujianto, and U. Qomaria. "Predicting high school graduates using Naïve Bayes in State University Entrance Selections." *International*

Conference on Vocational Education and Training (ICOVET), Indonesia, pp. 155-159, 2020.

[14] A. I. Adekitan and E. Noma-Osaghae. "Data mining approach to predicting the performance of first year student in a university using the admission requirements." *Education and Information Technologies,* Vol. 24, pp. 1527-1543, 2019.

[15] S. Vanitha and R. Jayahree. "A Prediction on Education Time Series Data Using Statical Machine Learning Model-An Experiment Analysis." *Journal of Theoretical and Applied Information Technology,* Vol. 100, No. 14, pp. 5189-5200, 2020.

[16] I. D. Yakubu and J. A. Awaab. "Assessing Students' Enrolment in Bolgatanga Polytechnic Using Time Series Analysis." *East African Scholars Journal of Engineering and Computer Sciences,* Vol. 2, No. 4, pp. 120-135, 2019.

[17] M. V. Amazona and A. A. Hernandez. "Modelling Student Performance Using Data Mining Techniques: Inputs for Academic Program Development." *ICCDE '19: Proceedings of the 2019 5th International Conference on Computing and Data Engineering,* pp. 36-40, 2019.

[18] E. N. Iqbal and R. Asif. "Analyzing Applicant's Pre-admission data and Predicting Applicant's Performance in Pre-Admission Test using Data Mining Techniques." *International Conference on Innovations in Computer Science & Software Engineering (ICONICS),* Karachi, Pakistan, pp. 1-11, 2022.

[19] G. Dewantoro and N. Ardisa. "A Decision Support System for Undergraduate Students Admissions using Educational Data Mining." *2020 7th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE),* Indonesia, pp. 105-109, 2020.

[20] Y. Zhao, Q. Xu, and M. Chen. "Predicting Student Performance in a Master of Data Science Program using Admissions Data." *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020),* pp. 325-333, 2020.

[21] S. Garmpis, M. Maragoudakis, and A. Garmpis. "Assisting Educational Analytics with AutoML Functionalities." *Computers,* Vol. 11, No. 6, pp. 97, 2022.

[22] B. Ujkani, D. Minkovska, and L. Stoyanova. "A Machine Learning Approach for Predicting Student Enrollment in the University." *International Scientific Conference Electronics (ET),* Sozopol, Bulgaria, pp. 1-4, 2021.

[23] W. Tenpipat and K. Akkarajitsakul. "Student Dropout Prediction: A KMUTT Case Study." *International Conference on Big Data Analytics and Practices (IBDAP),* Bangkok, Thailand, pp. 1-5, 2020.

[24] N. Bedregal-Alpaca, V. Cornejo-Aparicio, J, Zárate-Valderrama, and P. Yanque-Churo. "Classification Models for Determining Types of Academic Risk and Predicting Dropout in University Students." *International Journal of Advanced Computer Science and Applications (IJACSA),* Vol. 11, No. 1, pp. 266-272, 2020.

[25] M. Charlotte. "Modeling an Automated Student's Performance Predictor by Using Decision Tree." *East African Journal of Science and Technology,* Vol. 9, No. 1, pp. 121-137, 2019.

[26] M. Ursan and M. Bucos. "Predictive Analytics Models for Student Admission to Master Programs in Romania." *International Symposium on Electronics and Telecommunications (ISETC),* Timisoara, Romania, pp. 1-4, 2022.

[27] T. Pinyo. "The Study on the Report of the Logistic Regression Analysis Results in Research Work." *Humanities, Social Sciences and arts,* Vol. 12, No. 5, pp. 544-558, 2019.

[28] S. Jeganathan, S. Parthasarath, A.R. Lakshminarayanan, P. M. Kumarand, and K. A. Khan. "Predicting the Post Graduate Admissions using Classification

Techniques." *International Conference on Emerging Smart Computing and Informatics (ESCI) AISSMS Institute of Information Technology,* India, pp. 5-7, 2021.

[29] M. A. A. Walid, S. M. Masum Ahmed, and S. M. S. Sadique. "A Comparative Analysis of Machine Learning Models for Prediction of Passing Bachelor Admission Test in Life-Science Faculty of a Public University in Bangladesh." *IEEE Electric Power and Energy Conference (EPEC),* Edmonton, Canada, pp. 1-6, 2020.