Journal of Information and Learning, (2023), 34(3), 91-103
Journal homepage: https://so04.tci-thaijo.org/index.php/jil

Research Article

Prediction of Passing Probability of Students in Higher Education Level Using Machine Learning Algorithm

การทำนายโอกาสการเลื่อนชั้นของนักศึกษาในระดับอุดมศึกษา ด้วยอัลกอริทึมการเรียนรู้ของเครื่อง

Napaphat Wannatrong[1*], Aphisit Sangsiwi[1], and Jarumas Sangsavang[2]

ณปภัช วรรณตรง[1*], อภิสิทธิ์ สังสีวี[1], และ จารุมาศ แสงสว่าง[2]

[1]*Computer Science Program, Faculty of Science, Buriram Rajabhat University*

[1]*สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์*

[2]*Statistics and Information Science Program, Faculty of Science, Buriram Rajabhat University*

[2]*สาขาวิชาสถิติและวิทยาการสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์*

*Corresponding author: wannatrong@hotmail.com

Received August 16, 2023 ▮ Revised October 9, 2023 ▮ Accepted November 16, 2023 ▮ Published December 28, 2023

Abstract

The purposes of this research were: 1) to develop a model to predict the probability of first-year students in the Computer Science program passing to the second year, and 2) to implement an application predicting the probability of first-year students in the Computer Science program passing to the second year.

The data were collected using a questionnaire administered to first to fourth-year students in the Computer Science program, totaling 125 individuals, at Buriram Rajabhat University. Data were also collected from the Office of Academic Promotion and Registration of Buriram Rajabhat University. Imbalanced data were adjusted using SMOTE. The model was created using three algorithmic techniques, including logistic regression, random forest, and decision tree. The Efficiency of each algorithm was measured using accuracy and root mean square error (RMSE). The Results of the measurement of the models' efficiency revealed that the model utilizing logistic regression had the highest prediction efficiency.

With regard to the results of the development of an application for predicting the probability of passing to the second year of the first-year computer science students, the most efficient model was deployed into a web application used for prediction using Python and Django Framework with PostgreSQL as the database. Moreover, the results of an analysis of users' satisfaction with the web application showed that the users' satisfaction was at the highest level with a mean score of 4.59 and a standard deviation of 0.55.

Keywords: predictive modeling, machine learning, web application, logistic regression, random forest, decision tree

บทคัดย่อ

งานวิจัยนี้นำเสนอการทำนายข้อมูลโอกาสที่นักศึกษาชั้นปีที่ 1 จะได้เลื่อนชั้นขึ้นชั้นปีที่ 2 โดยใช้การเรียนรู้ของเครื่อง ในบริบทการศึกษาระดับอุดมศึกษา มีวัตถุประสงค์การวิจัย 1) เพื่อสร้างแบบจำลองทำนายโอกาสการได้ขึ้นชั้นปี 2 ของนักศึกษาชั้นปีที่ 1 สาขาวิชาวิทยาการคอมพิวเตอร์ และ 2) เพื่อพัฒนาเว็บแอปพลิเคชันทำนายโอกาสการได้ขึ้นชั้นปี 2 ของนักศึกษาชั้นปีที่ 1 สาขาวิชาวิทยาการคอมพิวเตอร์

ผลการสร้างแบบจำลองทำนายโอกาสการได้ขึ้นชั้นปี 2 ของนักศึกษาชั้นปีที่ 1 สาขาวิชาวิทยาการคอมพิวเตอร์ เก็บรวบรวมข้อมูลโดยการใช้แบบสอบถาม นักศึกษาสาขาวิชาวิทยาการคอมพิวเตอร์ชั้นปีที่ 1-4 มหาวิทยาลัยราชภัฏบุรีรัมย์จำนวน 125 คน และรวบรวบข้อมูลจากสำนักวิชาการและงานทะเบียนของมหาวิทยาลัยราชภัฏบุรีรัมย์ ปรับข้อมูลที่ไม่สมดุลด้วยเทคนิค SMOTE และสร้างแบบจำลองจากอัลกอริทึม 3 แบบ ได้แก่ เทคนิคการวิเคราะห์การถดถอยโลจีสติก เทคนิคป่าไม้ตัดสินใจ และเทคนิคต้นไม้ตัดสินใจ วัดประสิทธิภาพแต่ละอัลกอริทึมด้วยค่าความถูกต้อง และค่าความคลาดเคลื่อนของเครื่องมือ ผลการวัดประสิทธิภาพของแบบจำลอง พบว่า แบบจำลองที่ใช้เทคนิคการวิเคราะห์การถดถอยโลจีสติกมีประสิทธิภาพในการทำนายที่สูงที่สุด

ผลการพัฒนาเว็บแอปพลิเคชันทำนายโอกาสการได้ขึ้นชั้นปี 2 ของนักศึกษาชั้นปีที่ 1 สาขาวิชาวิทยาการคอมพิวเตอร์ นำแบบจำลองที่มีประสิทธิภาพสูงสุดมาพัฒนาในรูปแบบเว็บแอปพลิเคชันที่ใช้ในการทำนายด้วยภาษา Python และ Django Framework และใช้โปรแกรม Postgresql เป็นฐานข้อมูล ผลการวิเคราะห์ความพึงพอใจของผู้ใช้ที่มีต่อเว็บแอปพลิเคชัน พบว่า ผู้ใช้มีความพึงพอใจในระดับมากที่สุด มีค่าเฉลี่ยอยู่ที่ 4.59 และค่าเบี่ยงเบนมาตรฐาน 0.55

คำสำคัญ: แบบจำลองทำนาย, การเรียนรู้ของเครื่อง, เว็บแอปพลิเคชัน, การวิเคราะห์การถดถอยโลจีสติก, ป่าไม้ตัดสินใจ, ต้นไม้ตัดสินใจ

บทนำ (Introduction)

ปัญหาทั่วไปอย่างหนึ่งที่นักศึกษาระดับอุดมศึกษาต้องเผชิญในช่วงปีแรกคือความท้าทายในการปรับตัวให้เข้ากับการเรียนในระดับอุดมศึกษา ความท้าทายเหล่านี้มักเกิดจากความยากลำบากในการปรับตัวเข้ากับการเรียนในมหาวิทยาลัย และการขาดพื้นฐานการเรียนรู้ที่จำเป็น สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ เป็นสาขาวิชาที่เปิดสอนระดับปริญญาด้านคอมพิวเตอร์ สาขาวิชาวิทยาการคอมพิวเตอร์ มักมีปัญหาเกี่ยวกับการพ้นสภาพของนักศึกษาช่วงปีที่ 1 ที่จะขึ้นปีที่ 2 เป็นอย่างมาก เมื่อเทียบอัตราส่วนกับการพ้นสภาพของนักศึกษาในชั้นปีอื่น ๆ เนื่องจากนักศึกษาอาจมีปัญหา เช่น ไม่สามารถปรับตัวในการเรียนในระดับมหาวิทยาลัยได้ หรือไม่มีพื้นฐานความรู้ที่จำเป็นในการศึกษา เป็นต้น

การเรียนรู้ของเครื่อง (Machine learning) เป็นศาสตร์แขนงหนึ่งที่ใช้สร้างในการเรียนรู้ให้กับเครื่องโดยอาศัยวิธีการทางคณิตศาสตร์และสถิติในการสร้างแบบจำลอง (Model) จากข้อมูลที่มีอยู่ เพื่อใช้ในการทำนายผลและตัดสินใจ (Bancha, 2021, p. 11) การเรียนรู้ของเครื่อง คือ การใช้ข้อมูลที่มีอยู่เพื่อพัฒนาแบบจำลองที่สามารถใช้ในการคาดการณ์ผลลัพธ์ต่าง ๆ สำหรับข้อมูลใหม่ (Grus, 2019, p. 164) ได้มีการนำการเรียนรู้ของเครื่องมาใช้ทำนายเกี่ยวกับนักศึกษา อาทิ Hussain and Khan (2021) นำการเรียนรู้ของเครื่องมาใช้ในการทำนายประสิทธิภาพของนักเรียนระดับมัธยม Li and Liu (2021) ใช้การเรียนรู้ของเครื่องประเภทการเรียนรู้เชิงลึก (Deep learning) ทำนายประสิทธิภาพนักศึกษาระดับอุดมศึกษา เพื่อใช้ในการวางแผนในระดับการศึกษาที่สูงขึ้น Kepan et al. (2018) ได้ใช้เหมืองข้อมูลในการทำนายนักศึกษาระดับปริญญาตรีที่ออกกลางคัน Thaweechat et al. (2022) ได้ศึกษาเรื่องระบบทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล ซึ่งงานวิจัยทั้งหมดส่วนใหญ่จะพบว่าการทำนายด้วยการเรียนรู้ของเครื่องค่อนข้างให้ความแม่นยำในการทำนาย

นอกจากนี้ หลาย ๆ งานวิจัยยังนำแบบจำลองที่ทำนายมาพัฒนาต่อในรูปแบบเว็บแอปพลิเคชัน Clitan et al. (2021, pp. 70-74) พัฒนาเว็บแอปพลิเคชันซึ่งช่วยให้ผู้ใช้สามารถวิเคราะห์วิวัฒนาการของการระบาดใหญ่ของไวรัสโคโรนา 2019 ในประเทศโรมาเนียในรูปแบบกราฟิก และการทำนายวิวัฒนาการของผู้ติดเชื้อ Prasetyo et al. (2021, pp. 49-59) ได้สร้างระบบการทำนายโรคเบาหวาน โดยใช้อัลกอริทึม XGBoost ผ่านเว็บแอปพลิเคชัน Rekha et al. (2021, pp. 3770-3783) ได้พัฒนาเว็บแอปพลิเคชัน เพื่อคาดการณ์การใช้งานโดยใช้อัลกอริทึมการถดถอยเชิงเส้นที่มีตัวแปรหลายตัวเพื่อทำนายรูปแบบการใช้งานเพื่อความชัดเจนที่ดีขึ้น และการตัดสินใจทางการเงินที่ดีขึ้นในด้านต่าง ๆ เช่น งบประมาณ การตลาด ฯลฯ

จากที่กล่าวมาข้างต้นผู้วิจัยจึงจะนำเทคนิคการเรียนรู้ของเครื่องมาประยุกต์ใช้ในการทำนายโอกาสที่นักศึกษาชั้นปีที่ 1 จะได้ขึ้นชั้นปีที่ 2 เพื่อจะทำนายเบื้องต้นได้ว่านักศึกษาชั้นปีที่ 1 คนใดมีแนวโน้มไม่ได้ขึ้นชั้นปีที่ 2 สาขาวิชาวิทยาการคอมพิวเตอร์จะได้วางแผนดูแลนักศึกษาที่มีแนวโน้มมีปัญหาเป็นกรณีพิเศษเพื่อช่วยเพิ่มโอกาสในการได้ขึ้นชั้นปีที่ 2 ลดการออกมากขึ้น โดยการสร้างแบบจำลองทำนายนี้ได้เก็บรวบรวมข้อมูลนักศึกษาจากนักศึกษาชั้นปีที่ 1-4 โดยการใช้แบบสอบถาม และรวบรวมข้อมูลจากสำนักวิชาการและงานทะเบียนของมหาวิทยาลัยราชภัฏบุรีรัมย์ เลือกเฉพาะนักศึกษาสาขาวิชาวิทยาการคอมพิวเตอร์ชั้นปีที่ 1-4 ที่ใช้หลักสูตรวิทยาการคอมพิวเตอร์ หลักสูตรปรับปรุง 2561 ปรับข้อมูลที่ไม่สมดุลด้วยการทำข้อมูลให้สมดุลแบบ SMOTE (Synthetic minority oversampling technique: SMOTE) ซึ่งเป็นวิธีที่มีประสิทธิภาพในการเพิ่มข้อมูลวิธีหนึ่งสำหรับข้อมูลไม่สมดุล (Imbalanced data) คือ ข้อมูลที่มีจำนวนข้อมูลในกลุ่มหนึ่งมากกว่าจำนวนข้อมูลของอีกกลุ่มหนึ่งเป็นจำนวนมาก (Ishaq et al., 2021; Elreedy & Atiya, 2019; Akarachantachote & Panitsupakamol, 2019) และเพื่อให้ได้แบบจำลองที่มีประสิทธิภาพสูงสุดที่ใช้ในการทำนาย ผู้วิจัยจึงได้เลือกอัลกอริทึมที่งานวิจัยต่าง ๆ นิยมใช้ 3 แบบ และนำมาเปรียบเทียบประสิทธิภาพโดยอัลกอริทึมที่นำมาเปรียบเทียบ ดังนี้

เทคนิคการวิเคราะห์การถดถอยโลจีสติก (Logistic regression) ซึ่งเป็นอัลกอริทึมที่จัดอยู่ในประเภทการเรียนรู้แบบมีผู้สอน (Supervised learning) สำหรับการทำนายผลแบบการจำแนก (Classification) โดยข้อมูลตัวอย่างแต่ละรายการจะมีคอลัมน์ที่เป็นผลลัพธ์หรือตัวแปรที่สามารถจำแนกประเภทหรือแบ่งได้เป็น 2 กลุ่ม (Binary classes) เป็นข้อมูลไม่ต่อเนื่อง เช่น Yes/No, True/False หรือ Success/Fail เป็นต้น ส่วนคอลัมน์ที่เป็นคุณลักษณะหรือตัวแปรอิสระอาจมีตั้งแต่ 1 คอลัมน์ขึ้นไป (Bancha, 2021, p. 152; Saraubon, 2020, p. 459) โดยได้มีการวิจัยหลายงานได้นำการถดถอยโลจีสติกมาใช้เป็นเทคนิคในการทำนาย อาทิ Fathiya and Sadath (2021) ใช้ในการทำนายโอกาสการเข้าศึกษามหาวิทยาลัยของนักเรียน Atik et al. (2021) ใช้เทคนิคการวิเคราะห์การถดถอยโลจีสติกในการทำนายเจตคติต่อสิ่งแวดล้อมของนักเรียนระดับชั้นมัธยมศึกษา Chacha et al. (2019) ใช้เทคนิคการวิเคราะห์การถดถอยโลจีสติกในการทำนายการออกกลางคันของนักศึกษา และ Wu et al. (2023) ใช้เทคนิคการวิเคราะห์การถดถอยโลจีสติกในการทำนายข้อเสนอการรับเข้าเรียนกับความน่าจะเป็นในการลงทะเบียนของนักศึกษา

เทคนิคต้นไม้ตัดสินใจ (Decision tree) เป็นอัลกอริทึมที่มีลักษณะสร้างลำดับการตัดสินใจเป็นแผนภูมิที่มีลักษณะโครงสร้างคล้ายกับต้นไม้แบบกลับหัว นิยมใช้งานกันอย่างแพร่หลาย เนื่องจากมีหลักการที่ไม่ยุ่งยาก สามารถเรียนรู้ได้ง่ายและรวดเร็ว สามารถทำนายผลได้ทั้งแบบจำแนกข้อมูลและแบบการถดถอย (Regression) ทั้งนี้ข้อมูลที่จะทำนายผลด้วยเทคนิคต้นไม้ตัดสินใจนั้น คอลัมน์ที่เป็นคุณลักษณะหรือตัวแปรอิสระควรเป็นข้อมูลที่สามารถจัดกลุ่มได้ ส่วนตัวแปรตามหรือผลลัพธ์อาจเป็นแบบจำแนกข้อมูลหรือแบบการถดถอย (Bancha, 2021, pp. 247-248) โดยได้มีการวิจัยหลายงานได้นำเทคนิคป่าไม้ตัดสินใจ มาใช้เป็นเทคนิคในการทำนาย อาทิ Gotardo (2019) ใช้เทคนิคต้นไม้ตัดสินใจในการทำนายประสิทธิภาพของนักศึกษา Khakata et al. (2019) ใช้เทคนิคต้นไม้ตัดสินใจในการทำนายประสิทธิภาพของนักศึกษาบนสภาพแวดล้อมที่เป็นสื่อกลางทางอินเทอร์เน็ต และ Thaweechat (2022) ใช้เทคนิคต้นไม้ตัดสินใจในการทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์

เทคนิคป่าไม้ตัดสินใจ (Random forest) เป็นอัลกอริทึมที่มีลักษณะคล้ายเทคนิคต้นไม้ตัดสินใจ แต่จะสุ่มข้อมูล (Instance) ไปสร้างเป็นต้นไม้หลาย ๆ ต้น แต่ละต้นเรียกว่าเซตย่อย (Subset) เหมือนป่าที่มีต้นไม้จำนวนมาก ๆ ซึ่งแต่ละต้นจะมีรูปแบบสุ่มไม่เหมือนกัน โดยในขั้นตอนการใช้งานทำนาย แต่ละต้นก็ทำการทำนายแล้วเลือกผลที่ได้จากการโหวตที่มากที่สุดมาเป็นผลลัพธ์ (Saraubon, 2020, p. 452) โดยได้มีการวิจัยหลายงานได้นำเทคนิคป่าไม้ตัดสินใจ มาใช้เป็นเทคนิคในการทำนาย อาทิ Behr et al. (2020) ใช้ในการทำนายล่วงหน้าของการออกจากมหาวิทยาลัยกลางคัน Dien et al. (2021) ใช้เทคนิคป่าไม้ตัดสินใจในการทำนายประสิทธิภาพของนักศึกษามหาวิทยาลัยสหสาขาวิชาชีพ และ Beaulac and Rosenthal (2019) ใช้เทคนิคป่าไม้ตัดสินใจในการทำนายความสำเร็จทางวิชาการและวิชาเอกของนักศึกษามหาวิทยาลัย

จากนั้นนำแบบจำลองที่สร้างโดยอัลกอริทึมทั้ง 3 แบบมาวัดประสิทธิภาพแบบจำลองด้วยการหาค่าความถูกต้อง (Accuracy) ซึ่งเป็นการตรวจสอบว่าผลลัพธ์ที่แบบจำลองทำนายออกมา (Predicted value) มีค่าตรงกับคำตอบจริง (Actual value) คิดเป็นสัดส่วนร้อยละเท่าใด ซึ่งเป็นตัววัดที่นิยมใช้ในการจำแนกข้อมูล (Pacharawongsakda, 2020, p. 112) และค่าความคลาดเคลื่อนของเครื่องมือวัด (Root mean square error: RMSE) เป็นการคำนวณค่าความคลาดเคลื่อนของเครื่องมือวัด เป็นการหาค่าความผิดพลาดหรือความแตกต่างโดยเฉลี่ยระหว่างข้อมูลจริงกับผลการทำนาย (Bancha, 2021, p. 129) จากนั้นนำอัลกอริทึมที่มีประสิทธิภาพดีที่สุดไปเพื่อพัฒนาเว็บแอปพลิเคชันที่ใช้ในการทำนายโอกาสการได้ขึ้นชั้นปีที่ 2 ของนักศึกษาชั้นปีที่ 1 สาขาวิชาวิทยาการคอมพิวเตอร์

วัตถุประสงค์การวิจัย (Objectives)

1. เพื่อสร้างแบบจำลองทำนายโอกาสการได้ขึ้นชั้นปี 2 ของนักศึกษาชั้นปีที่ 1 สาขาวิชาวิทยาการคอมพิวเตอร์
2. เพื่อพัฒนาเว็บแอปพลิเคชันทำนายโอกาสการได้ขึ้นชั้นปี 2 ของนักศึกษาชั้นปีที่ 1 สาขาวิชาวิทยาการคอมพิวเตอร์

กรอบแนวคิดการวิจัย (Conceptual framework)

Figure 1

Conceptual Framework

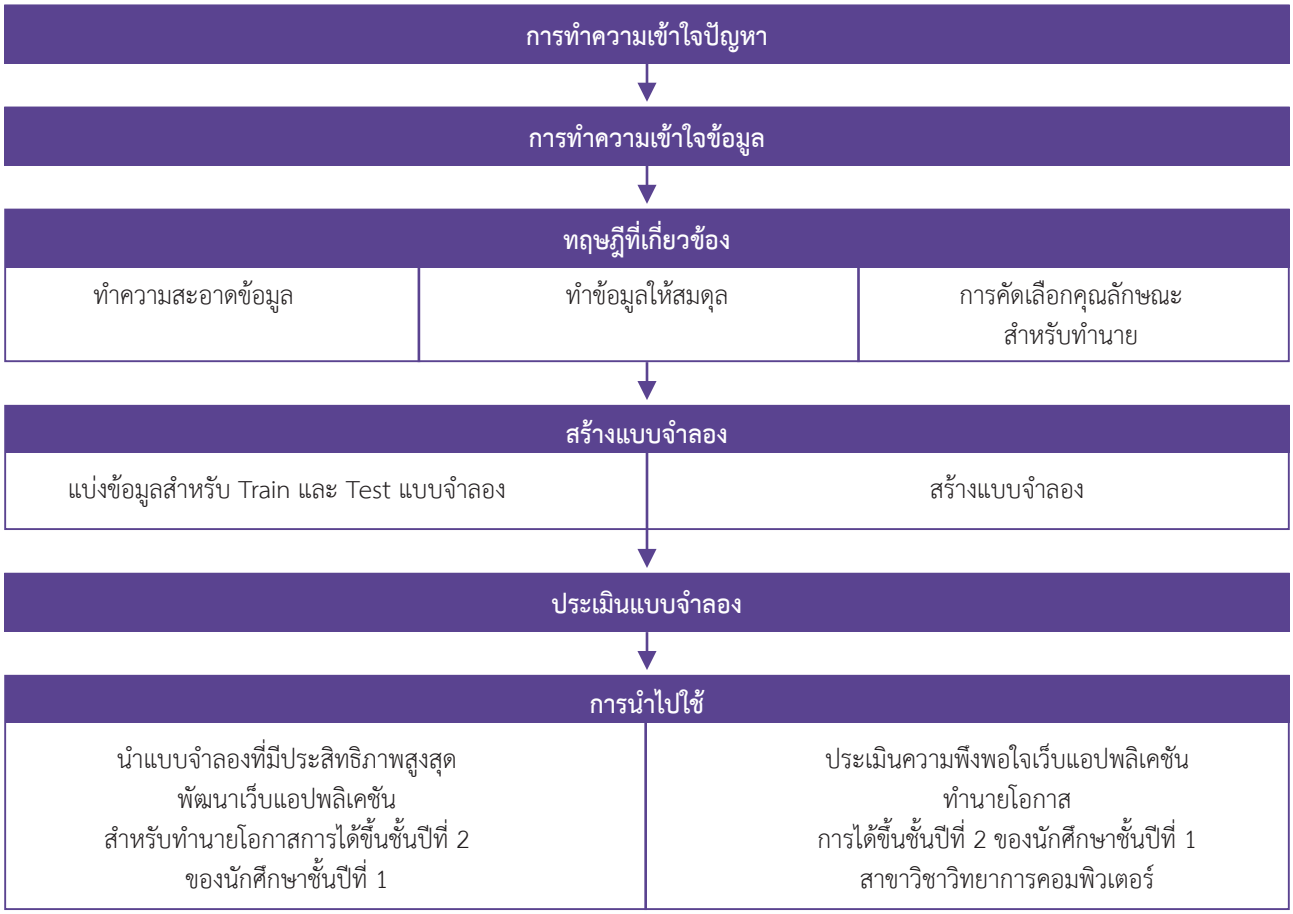
กรอบแนวคิดการวิจัย

นำข้อมูลนักศึกษาสาขาวิชาวิทยาการคอมพิวเตอร์ ชั้นปีที่ 1-4 ที่ใช้หลักสูตรวิทยาการคอมพิวเตอร์ หลักสูตรปรับปรุง 2561 และทฤษฎีที่เกี่ยวข้อง ได้แก่ วิทยาศาสตร์ข้อมูล การทำข้อมูลให้สมดุล และการเรียนรู้ของเครื่อง มาใช้ในการสร้างแบบจำลองที่ใช้การทำนายโอกาสที่มีประสิทธิภาพดีที่สุด จากนั้นเมื่อได้แบบจำลองที่ดีที่สุดแล้ว นำแบบจำลองและแนวทางทฤษฎีเว็บแอปพลิเคชันมาพัฒนาเว็บแอปพลิเคชันทำนายโอกาสการได้ขึ้นชั้นปีที่ 2 ของนักศึกษาชั้นปีที่ 1 สาขาวิชาวิทยาการคอมพิวเตอร์ เพื่อใช้ในการทำนาย และขั้นตอนสุดท้ายนำไปศึกษาความพึงพอใจผู้ใช้ที่มีต่อเว็บแอปพลิเคชันทำนายโอกาสการได้ขึ้นชั้นปีที่ 2 ของนักศึกษาชั้นปีที่ 1 สาขาวิชาวิทยาการคอมพิวเตอร์

วิธีดำเนินการวิจัย (Methodology)

ผู้วิจัยดำเนินการวิจัย มีขั้นตอนวิธีในการดำเนินการวิจัย ดัง Figure 2

Figure 2
Methodology
วิธีดำเนินการวิจัย



วิธีดำเนินการวิจัยดำเนินการตามขั้นตอน CRISP-DM process (Pacharawongsakda, 2020) โดยมีรายละเอียดการดำเนินการดังนี้

1. การทำความเข้าใจปัญหา (Business understanding) เป็นขั้นตอนการทำความเข้าใจปัญหา โดยขั้นตอนนี้ผู้วิจัยได้ศึกษาปัญหาจากผู้ที่เกี่ยวข้อง ได้แก่ นักศึกษา และอาจารย์ประจำสาขาวิชาวิทยาการคอมพิวเตอร์

2. การทำความเข้าใจข้อมูล (Data understanding) ผู้วิจัยเก็บรวบรวมข้อมูลโดยการใช้แบบสอบถาม โดยเลือกเฉพาะนักศึกษาสาขาวิชาวิทยาการคอมพิวเตอร์ชั้นปีที่ 1-4 ที่ใช้หลักสูตรวิทยาการคอมพิวเตอร์ หลักสูตรปรับปรุง 2561 และรวบรวมข้อมูลจากสำนักส่งเสริมวิชาการและงานทะเบียนของมหาวิทยาลัยราชภัฏบุรีรัมย์ เพื่อจะได้นำข้อมูลมาใช้ในการสร้างแบบจำลอง โดยมีประชากรและกลุ่มตัวอย่าง ดังนี้

2.1 ประชากร ได้แก่ นักศึกษาสาขาวิชาวิทยาการคอมพิวเตอร์ชั้นปีที่ 1-4 ที่ใช้หลักสูตรวิทยาการคอมพิวเตอร์ หลักสูตรปรับปรุง 2561 จำนวน 170 คน

2.2 กลุ่มตัวอย่าง ได้แก่ นักศึกษาสาขาวิชาวิทยาการ

คอมพิวเตอร์ชั้นปีที่ 1-4 ที่ใช้หลักสูตรวิทยาการคอมพิวเตอร์ หลักสูตรปรับปรุง 2561 โดยทำการเลือกกลุ่มตัวอย่างแบบเจาะจง (Purposive sampling) จำนวน 125 คน

3. การเตรียมข้อมูล (Data preparation) แบ่งรายละเอียดขั้นตอนการดำเนินงาน ได้ดังนี้

3.1 ทำความสะอาดข้อมูล (Data cleaning) ด้วยการตรวจสอบความถูกต้องของข้อมูล เติมข้อมูลที่ขาดให้สมบูรณ์ และแปลงข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปใช้ในการสร้างแบบจำลองได้

3.2 ทำข้อมูลให้สมดุล (Class data balancing) เนื่องจากกลุ่มข้อมูลผู้ไม่ได้ขึ้นชั้นปีที่ 2 สำหรับสร้างแบบจำลองมีจำนวนแตกต่างจากกลุ่มข้อมูลผู้ได้ขึ้นชั้นปีที่ 2 เป็นจำนวนมาก ผู้วิจัยจึงได้ทำการทำข้อมูลให้สมดุลแบบ SMOTE ซึ่งเป็นวิธีที่มีประสิทธิภาพในการเพิ่มข้อมูลวิธีหนึ่ง (Ishaq et al., 2021; Elreedy & Atiya, 2019; Akarachantachote & Panitsupakamol, 2019) โดยในการทำข้อมูลสมดุลจะใช้การสุ่มชื่อว่าการสุ่มตัวอย่างเพิ่ม (Over-sampling) กลุ่มส่วนน้อย วิธี SMOTE คำนวณดังนี้

$$x_S = x_i + u \cdot (\hat{x}_i - x_i)$$

โดยที่ x_S แทน ข้อมูลที่สังเคราะห์ใหม่
x_i แทน ข้อมูลเดิมที่สุ่ม
$\hat{x}_i$ แทน ข้อมูลที่เป็นเพื่อนบ้าน x_i
u แทน ค่าสุ่มที่อยู่ระหว่าง 0 – 1

3.3 การคัดเลือกคุณลักษณะสำหรับทำนาย (Feature selection) การคัดเลือกคุณลักษณะสำหรับทำนายมีการเลือก 2 ขั้นตอน คือ 1) เลือกคุณลักษณะจากการตั้งสมมติฐานและจากการศึกษางานวิจัยที่เกี่ยวข้อง และ 2) เลือกจากการใช้เทคนิคต้นไม้ตัดสินใจ (Decision tree)

4. สร้างแบบจำลอง (Modeling) แบ่งรายละเอียดขั้นตอนการดำเนินงาน ได้ดังนี้

4.1 แบ่งข้อมูลสำหรับสอนแบบจำลองและทดสอบแบบจำลอง (Spilt data for train and test model) โดยใช้อัตราส่วนการแบ่ง 80:20 ได้แก่ จำนวน 80% มาสร้างแบบจำลองทำนายโอกาสที่นักศึกษาชั้นปีที่ 1 จะได้ขึ้นชั้นปีที่ 2 และข้อมูลจำนวน 20% ที่เหลือใช้การวัดค่าความถูกต้องของแบบจำลอง

4.2 สร้างแบบจำลอง (Model building) ในงานวิจัยนี้ได้ใช้อัลกอริทึมในการสร้างแบบจำลองสำหรับทำนาย ได้แก่ เทคนิคการวิเคราะห์การถดถอยโลจีสติก เทคนิคต้นไม้ตัดสินใจ และเทคนิคป่าไม้ตัดสินใจ โดยจะมีการคัดเลือกแบบจำลองที่มีประสิทธิภาพสูงสุดที่จะใช้ในการทำนายโอกาสได้ขึ้นชั้นปีที่ 2 ของนักศึกษา ข้อมูลทำนายผลจะเป็นสถานะของนักศึกษา (ผ่าน/ไม่ผ่าน)

5. การประเมิน (Evaluation) วัดประสิทธิภาพแบบจำลองด้วยการหาค่าความถูกต้องและค่าความคลาดเคลื่อนของเครื่องมือวัด

6. การนำไปใช้ (Deployment) นำแบบจำลองที่มีค่าความถูกต้องสูงสุดมาพัฒนาในรูปแบบเว็บแอปพลิเคชันที่ใช้ในการทำนายโอกาสได้ขึ้นชั้นปี 2 ของนักศึกษา และประเมินความพึงพอใจผู้ใช้ โดยใช้สถิติ ได้แก่ ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน จากนั้นนำผลที่ได้เทียบกับเกณฑ์การประเมิน (Srisa-ard, 2013) ดังนี้

ค่าเฉลี่ย 4.51-5.00 หมายความว่า ระดับมากที่สุด
ค่าเฉลี่ย 3.51-4.50 หมายความว่า ระดับมาก
ค่าเฉลี่ย 2.51-3.50 หมายความว่า ระดับปานกลาง
ค่าเฉลี่ย 1.51-2.50 หมายความว่า ระดับน้อย
ค่าเฉลี่ย 1.00-1.50 หมายความว่า ระดับน้อยที่สุด

ผลการวิจัย (Results)

โดยมีรายละเอียดการดำเนินการ ดังนี้

1. ผลการทำความเข้าใจปัญหา

ผลจากการทำความเข้าใจปัญหา โดยการเก็บรวบรวมข้อมูลจากการสัมภาษณ์นักศึกษาและคณาจารย์ สาขาวิชาวิทยาการคอมพิวเตอร์ พบว่า สาเหตุที่นักศึกษาสาขาวิชาวิทยาการคอมพิวเตอร์ชั้นปีที่ 1 ต้องออกจากมหาวิทยาลัยไม่ได้ขึ้นชั้นปีที่ 2 ได้แก่ เกรดไม่ถึงทำให้ถูกให้ออก หรือลาออก ซึ่งมาจากสาเหตุหลายประการ โดยสาเหตุที่พบมากที่สุด คือ ไม่สามารถปรับตัวจากการเรียนระดับมัธยมศึกษาเข้าสู่การเรียนระดับอุดมศึกษาได้ และค้นพบว่าไม่ใช่สาขาที่ตนถนัด

2. ผลการทำความเข้าใจข้อมูล

ผู้วิจัยเก็บรวบรวมข้อมูล โดยการใช้แบบสอบถามโดยเลือกเฉพาะนักศึกษาสาขาวิชาวิทยาการคอมพิวเตอร์ชั้นปีที่ 1-4 ที่ใช้หลักสูตรวิทยาการคอมพิวเตอร์ หลักสูตรปรับปรุง 2561 ซึ่งมีจำนวน 125 คน และรวบรวมข้อมูลจากสำนักส่งเสริมวิชาการและงานทะเบียนของมหาวิทยาลัยราชภัฏบุรีรัมย์ เพื่อจะนำข้อมูลมาใช้ในการสร้างแบบจำลอง โดยข้อมูลที่เก็บรวบรวมประกอบด้วย 1) รหัสนักศึกษา 2) ชั้นปี 3) สถาบันที่จบการศึกษา 4) แผนการเรียน 5) ผลการเรียนตลอดหลักสูตรที่ใช้สมัครเข้าเรียนตอนปี 1 6) ผลการเรียนกลุ่มรายวิชาคณิตศาสตร์ 7) ผลการเรียนกลุ่มรายวิชาวิทยาศาสตร์ 8) ผลการเรียนกลุ่มรายวิชาภาษาต่างประเทศ 9) รอบที่มาสมัครเรียนสาขาวิทยาการคอมพิวเตอร์ 10) เกรดในชั้นปีที่ 1 ภาคการศึกษาที่ 1 11) เคยแข่งทักษะด้านคอมพิวเตอร์ไหม 12) เคยอบรมเกี่ยวกับคอมพิวเตอร์ไหม 13) เคยเรียนเขียนโปรแกรมไหม 14) เคยอบรมเขียนโปแกรมไหม 15) มีทักษะด้านอื่น ๆ ไหม 16) สนใจอยากพัฒนาทักษะด้านใดเพิ่ม 17) รายได้ของครอบครัวต่อเดือน 18) สถานภาพครอบครัว

19) รู้จักสาขาวิชาวิทยาการคอมพิวเตอร์จากช่องทางใด และ 20) ทำไมถึงเลือกเรียนสาขานี้

3. ผลการเตรียมข้อมูล

3.1 ผลการทำความสะอาดข้อมูล ทำความสะอาดข้อมูลด้วยการตรวจสอบความถูกต้องของข้อมูล ได้แก่ ข้อมูลเกรดเฉลี่ยรวมระดับมัธยมตอนปลาย และเกรดในชั้นปีที่ 1 ภาคการศึกษาที่ 1 โดยเทียบกับข้อมูลเกรดที่ได้จากสำนักส่งเสริมวิชาการและงานทะเบียนของมหาวิทยาลัยราชภัฏบุรีรัมย์ เติมข้อมูลที่ขาดให้สมบูรณ์ และแปลงข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปสร้างแบบจำลองได้ ด้วยการใช้ python library ได้แก่ pandas และ NumPy

3.2 ผลการทำข้อมูลให้สมดุล ในการทำนายที่มีความถูกต้องนั้นข้อมูลส่วนของคลาสที่ต้องการต้องมีจำนวนใกล้เคียงหรือเท่ากัน เพื่อให้มีความถูกต้องของแบบจำลองที่ใช้ในการทำนายมากที่สุด ผู้วิจัยจึงได้ทำการทำข้อมูลให้สมดุล โดยได้ผลการทำข้อมูลสมดุลดัง Table 1

Table 1

The Number of Data Before Data Balancing and the Number of Data After Data Balancing
จำนวนข้อมูลก่อนทำสมดุลข้อมูล และจำนวนข้อมูลหลังทำสมดุลข้อมูล

คลาส	จำนวนข้อมูล	
	ข้อมูลที่ยังไม่ทำสมดุลข้อมูล (Imbalanced data)	ข้อมูลหลังจากทำสมดุลข้อมูล (Balanced data)
ผ่าน	119 คน	119 คน
ไม่ผ่าน	6 คน	119 คน
รวม	125 คน	238 คน

3.3 ผลการคัดเลือกคุณลักษณะสำหรับทำนาย ผลคุณลักษณะที่ใช้ในการสร้างแบบจำลอง ได้แก่ ข้อมูลเกรดเฉลี่ยรวมในระดับมัธยมตอนปลาย แผนการเรียน ขนาดโรงเรียน สถานะครอบครัว ประสบการณ์เขียนโปรแกรม และข้อมูลอื่น ๆ มาพิจารณาสำหรับการได้ขึ้นชั้นปีที่ 2 สาขาวิชาวิทยาการคอมพิวเตอร์ และใช้ข้อมูลในการสร้างแบบจำลองด้วยเทคนิคต้นไม้ตัดสินใจได้ผลดัง Figure 3

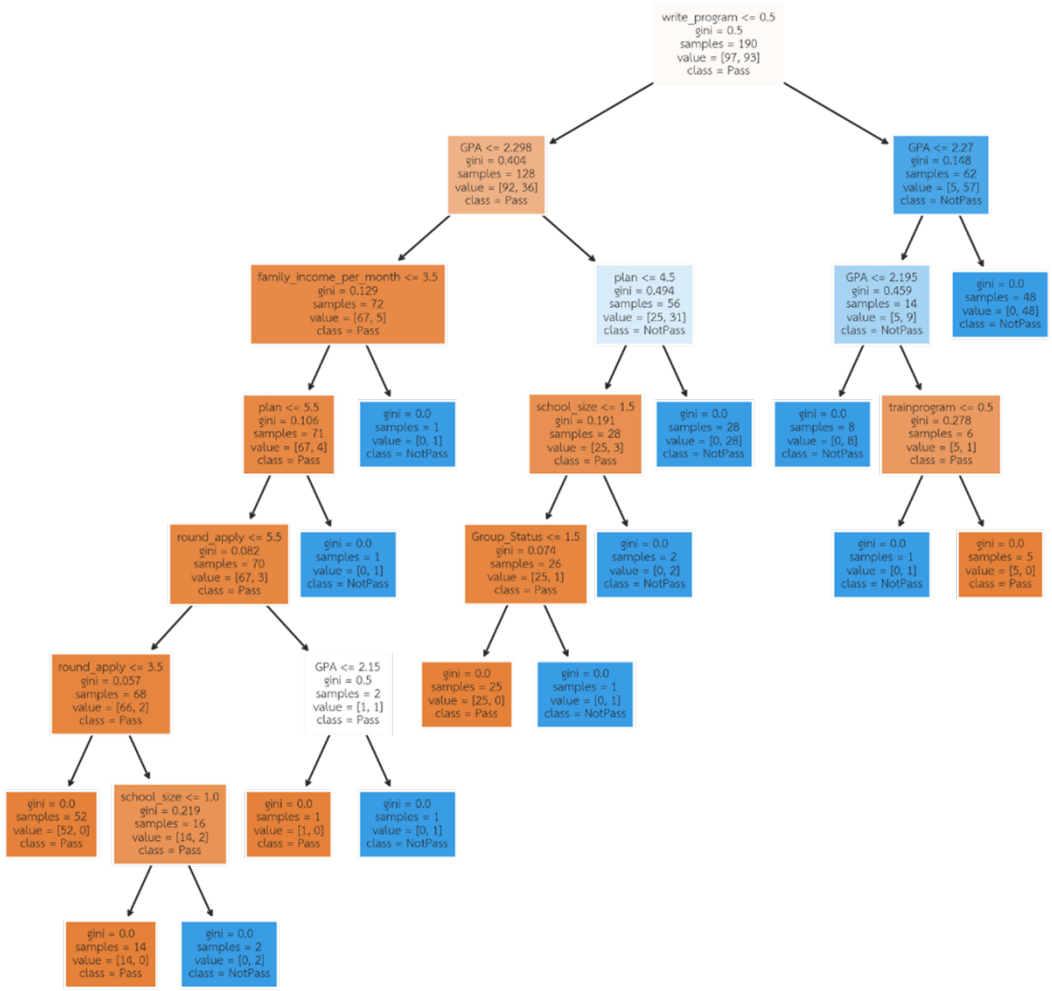
โดยใช้ค่าความสำคัญจากเทคนิคต้นไม้ตัดสินใจที่มีผลดัง Figure 3 ในการเลือกคุณลักษณะที่ใช้ในการทำนายได้ผลดัง Table 2

Table 2

Table of Importance Values of Features Influencing Data Prediction From Decision Tree Technique
ตารางค่าความสำคัญของคุณลักษณะที่มีผลในการทำนายข้อมูลจากเทคนิคต้นไม้ตัดสินใจ

คุณลักษณะ (Feature)	ค่าความสำคัญ (Importance values of features)
เคยเขียนโปรแกรมหรือไม่	0.358205
แผนการเรียน	0.254090
เกรดเฉลี่ยรวมในชั้นมัธยม	0.234765
สถานะครอบครัว	0.056416
ขนาดโรงเรียน	0.047389
รายได้ของครอบครัวต่อเดือน	0.018495
เคยอบรมเขียนโปรแกรมหรือไม่	0.017552
รอบที่สมัครเรียน	0.013089

Figure 3
Model Created Using Decision Tree Technique
แบบจำลองที่สร้างด้วยเทคนิคต้นไม้ตัดสินใจ



จาก Table 2 จึงได้คุณลักษณะที่จะนำไปใช้ในการสร้างแบบจำลอง ได้แก่ เคยเขียนโปรแกรมหรือไม่ แผนการเรียนและเกรดเฉลี่ยรวมในระดับมัธยมตอนปลาย เนื่องจากมีค่าความสำคัญสูง คือ 0.358205, 0.254090 และ 0.234765 ตามลำดับ

4. ผลการสร้างแบบจำลอง

4.1 ผลการแบ่งข้อมูลสำหรับสอนแบบจำลองและทดสอบแบบจำลอง แบ่งข้อมูลสำหรับสอนแบบจำลองและทดสอบแบบจำลองโดยใช้อัตราส่วนการแบ่ง 80:20 ได้แก่ จำนวน 80% มาสร้างแบบจำลองทำนายโอกาสที่นักศึกษาชั้นปีที่ 1 จะได้ขึ้นชั้นปีที่ 2 และข้อมูลจำนวน 20% ที่เหลือใช้การวัดค่าความถูกต้องของแบบจำลอง

4.2 ผลการสร้างแบบจำลอง ในงานวิจัยนี้ได้ใช้อัลกอริทึมในการสร้างแบบจำลองสำหรับทำนาย ได้แก่ เทคนิคการวิเคราะห์การถดถอยโลจีสติก เทคนิคต้นไม้ตัดสินใจ และเทคนิคป่าไม้ตัดสินใจ ซึ่งข้อมูลที่นำเข้าฝึกสอนและทดสอบจะมี 3 คุณลักษณะ ได้แก่ เคยเขียนโปรแกรมหรือไม่ แผนการเรียนและเกรดเฉลี่ยรวมในชั้นมัธยม ส่วนข้อมูลทำนายผลจะเป็นสถานะของนักศึกษา (ผ่าน/ไม่ผ่าน) โดยเรียกใช้อัลกอริทึมทั้ง 3 แบบจาก library Scikit-learn และใช้ Google Colab เป็นเครื่องมือในการสร้างแบบจำลอง

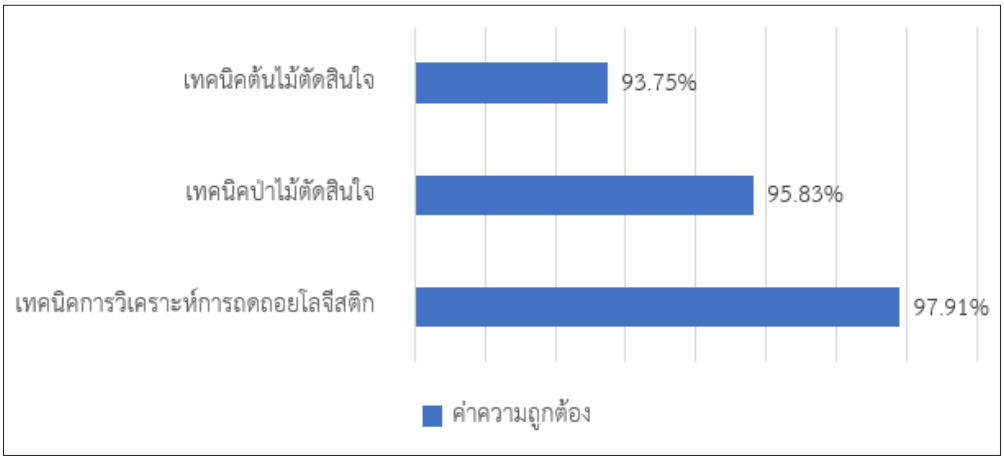
5. ผลการวัดประสิทธิภาพแบบจำลอง

วัดประสิทธิภาพแบบจำลองด้วยการหาค่าความถูกต้องและค่าความคลาดเคลื่อนของเครื่องมือวัด หลังจากที่แบ่งข้อมูลออกเป็น 2 ส่วนและใช้แบบจำลองที่สร้างได้จากข้อมูลจำนวน 80% มาทำนายโอกาสที่นักศึกษาชั้นปีที่ 1 จะได้ขึ้นชั้นปีที่ 2 ข้อมูลจำนวน 20% ที่เหลือใช้การวัดค่าความถูกต้อง จากการนำข้อมูลนักศึกษาสาขาวิชาวิทยาการคอมพิวเตอร์จำนวน 125 คน ไปทดสอบกับแบบจำลองทำนาย 3 แบบ คือ เทคนิคการวิเคราะห์การถดถอยโลจีสติก เทคนิคต้นไม้ตัดสินใจ และเทคนิคป่าไม้ตัดสินใจ ผลการทดสอบค่าความถูกต้องของแบบจำลองทำนาย พบว่า แบบจำลองทำนาย

ที่ใช้เทคนิคการวิเคราะห์การถดถอยโลจีสติก มีค่าความถูกต้องในการทำนายที่สูงสุด ตามด้วยเทคนิคป่าไม้ตัดสินใจ และเทคนิคต้นไม้ตัดสินใจ ที่ร้อยละ 97.91, 95.83 และ 93.75 ตามลำดับ และได้ทำการเปรียบเทียบค่าความคลาดเคลื่อนของเครื่องมือวัด พบว่า เทคนิคการวิเคราะห์การถดถอยโลจีสติกมีค่าน้อยสุด ตามด้วยเทคนิคป่าไม้ตัดสินใจ และเทคนิคต้นไม้ตัดสินใจเท่ากับ 0.14, 0.20 และ 0.25 ตามลำดับ โดยแสดงในรูปแบบแผนภูมิได้ดัง Figure 4

Figure 4

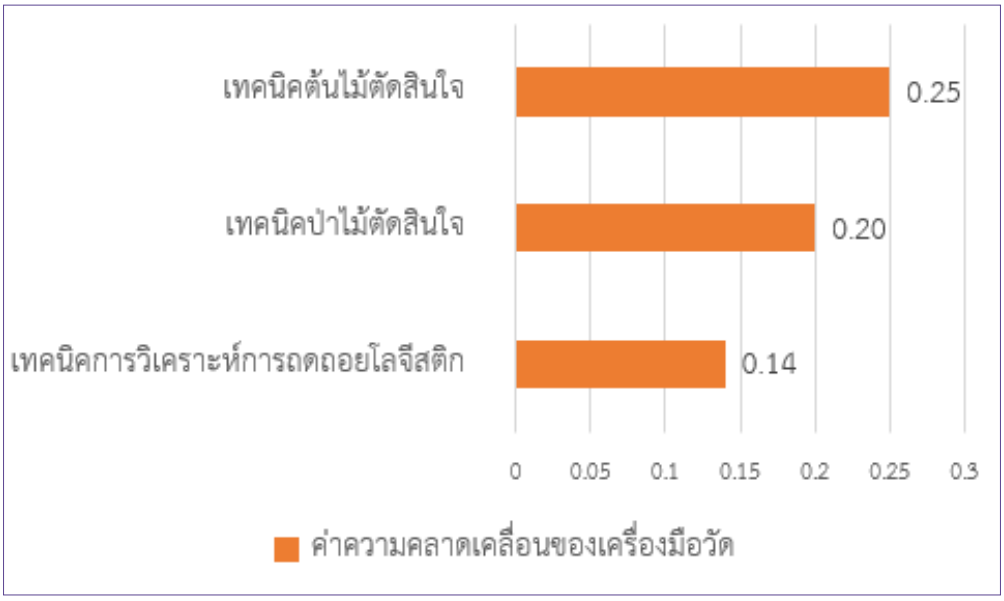
Comparison of Prediction Accuracy Values From Models Created by Different Algorithms
เปรียบเทียบค่าความถูกต้องการทำนายข้อมูลของแบบจำลองแต่ละอัลกอริทึม



ทำการเปรียบเทียบค่าความคลาดเคลื่อนของเครื่องมือวัด พบว่า เทคนิคการวิเคราะห์การถดถอยโลจีสติกมีค่าน้อยสุด ตามด้วยเทคนิคป่าไม้ตัดสินใจ และเทคนิคต้นไม้ตัดสินใจ เท่ากับ 0.14, 0.24 และ 0.25 ตามลำดับ โดยแสดงในรูปแบบแผนภูมิได้ดัง Figure 5

Figure 5

Comparison of tool measurement errors from models created by different algorithms
เปรียบเทียบค่าความคลาดเคลื่อนของเครื่องมือวัดแบบจำลองแต่ละอัลกอริทึม



จากข้างต้นการเปรียบเทียบประสิทธิภาพของตัวทำนาย ทั้ง 3 แบบ ด้วยการใช้ข้อมูลนักศึกษาสาขาวิชาวิทยาการคอมพิวเตอร์พบว่าแบบจำลองที่ให้ค่าความถูกต้องสูงที่สุด คือ เทคนิคการวิเคราะห์การถดถอยโลจีสติกด้วยค่าความถูกต้อง 97.91% และค่าความคลาดเคลื่อนของเครื่องมือวัดเท่ากับ 0.14 ซึ่งเป็นค่าที่น้อยที่สุด เมื่อเปรียบเทียบกับอัลกอริทึมอื่น ๆ รองลงมาตามลำดับ คือ เทคนิคต้นไม้ตัดสินใจด้วยค่าความถูกต้อง 95.83% และค่าความคลาดเคลื่อนของเครื่องมือวัดเท่ากับ 0.25 และเทคนิคป่าไม้ตัดสินใจด้วยค่า ความถูกต้อง 93.75% ค่าความคลาดเคลื่อนของเครื่องมือวัดเท่ากับ 0.20

ดังนั้น เมื่อพิจารณาค่าความถูกต้องจากชุดข้อมูลเดียวกันแล้วและค่าทั้ง 2 ค่าแล้วที่มีค่าความถูกต้องสูงสุดและค่าความคลาดเคลื่อนของเครื่องมือวัดมีค่าน้อยที่สุด พบว่าอัลกอริทึมผู้วิจัยเลือกที่จะนำมาประยุกต์ใช้ในการพัฒนาเว็บแอปพลิเคชันการทำนายข้อมูลโอกาสที่นักศึกษาชั้นปีที่ 1 จะได้ขึ้นชั้นปีที่ 2 คือ เทคนิคการวิเคราะห์การถดถอยโลจีสติก ค่าจากการวัดประสิทธิภาพ ได้แก่ ค่าความถูกต้อง ที่ 97.91% และค่าความคลาดเคลื่อนของเครื่องมือวัด เท่ากับ 0.25

6. การนำไปใช้

ผลการพัฒนาเว็บแอปพลิเคชันที่ใช้ในการทำนายได้เว็บแอปพลิเคชันที่ใช้ในการทำนายดัง Figure 6-7

Figure 6

Screen for Selecting the Student Cohort to Predict Data
หน้าจอเลือกรุ่นนักศึกษาที่จะทำนายข้อมูล

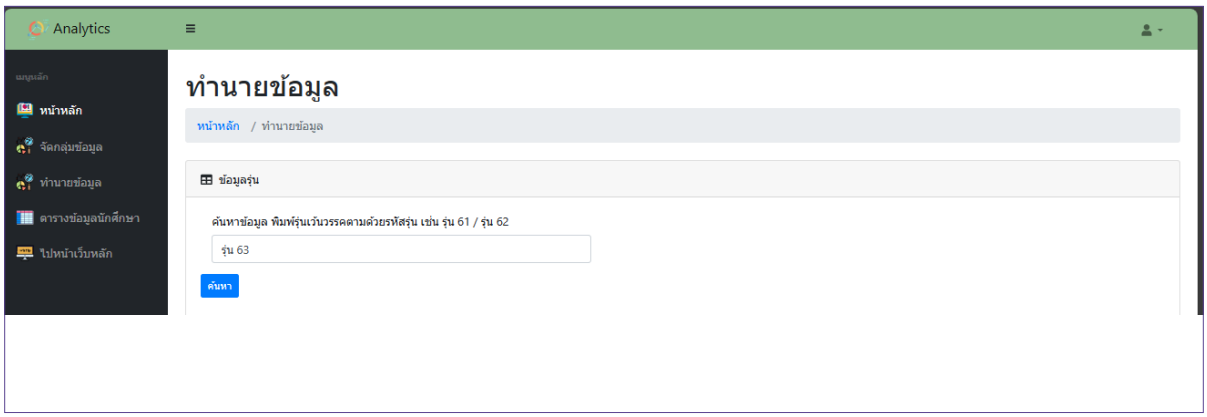


Figure 6 หน้าจอเลือกรุ่นนักศึกษาที่จะทำนายข้อมูล ในส่วนนี้จะมีการกรอกค้นหาเลือกรุ่นของนักศึกษามาเพื่อทำนายข้อมูลโอกาสที่นักศึกษาชั้นปีที่ 1 จะได้ขึ้นชั้นปีที่ 2

Figure 7

Screen Displaying Student Data Before Prediction

หน้าจอแสดงข้อมูลนักศึกษาก่อนทำนาย

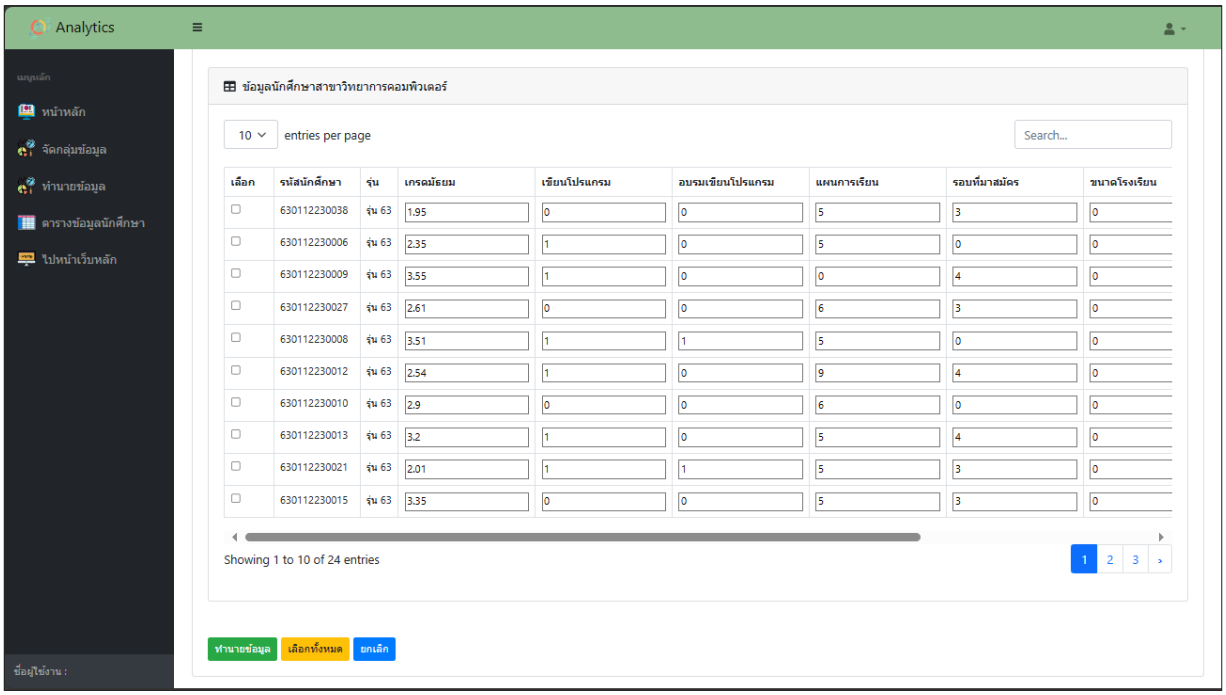


Figure 7 หน้าจอแสดงข้อมูลนักศึกษาก่อนทำนาย ในส่วนนี้จะแสดงข้อมูลนักศึกษาแต่ละคนก่อนกดปุ่มทำนาย โอกาสที่นักศึกษาชั้นปีที่ 1 จะได้ขึ้นชั้นปีที่ 2

Figure 8

Screen Displaying Prediction Results Prediction Results Display Screen

หน้าจอแสดงผลการทำนาย

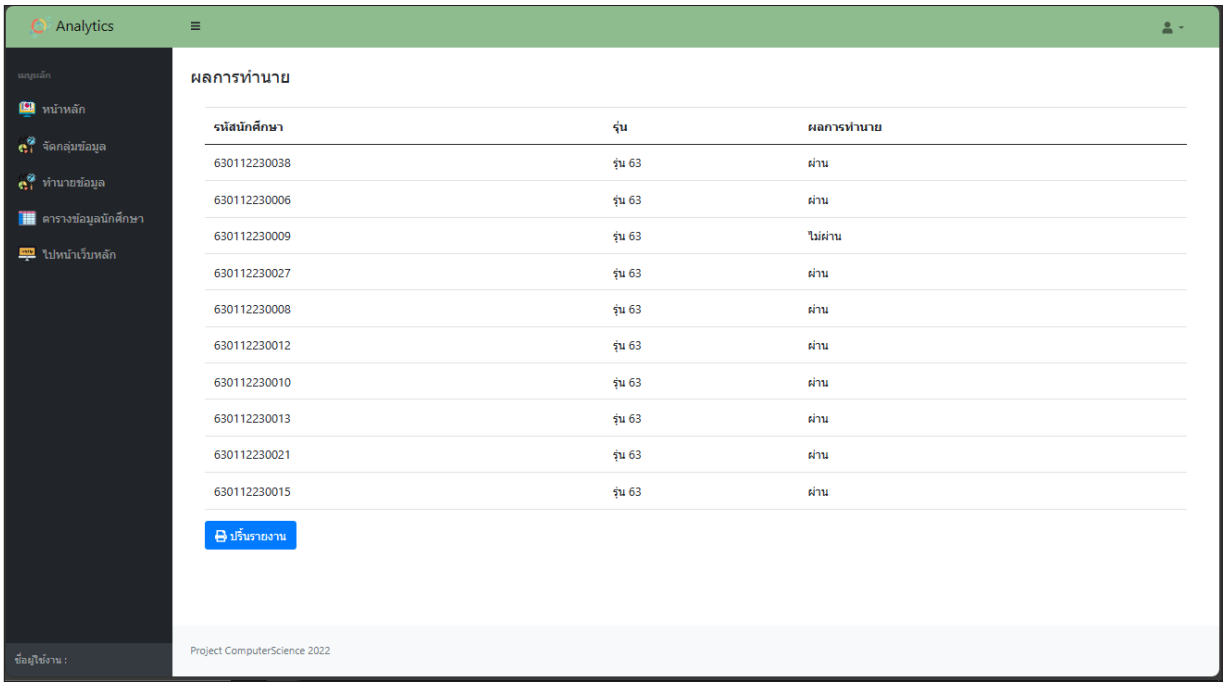


Figure 8 หน้าจอแสดงผลการทำนายในส่วนนี้จะแสดงผลการทำนายโอกาสที่นักศึกษาชั้นปีที่ 1 จะได้ขึ้นชั้นปีที่ 2

จากนั้นนำเว็บแอปพลิเคชันที่ใช้ในการทำนายข้อมูลโอกาสที่นักศึกษาชั้นปีที่ 1 จะได้ขึ้นชั้นปีที่ 2 ประเมินความพึงพอใจโดยผู้ใช้ ได้แก่ อาจารย์ในสาขาวิชาวิทยาการคอมพิวเตอร์ จำนวน 3 ท่าน ได้ผลดัง Table 3-4

Table 3

Analysis Results of User Satisfaction Towards System Performance and Data Security Maintenance in the System
ผลการวิเคราะห์ความพึงพอใจของผู้ใช้ที่มีต่อด้านการทำงานของระบบและด้านการรักษาความปลอดภัยของข้อมูลในระบบ

รายการ	*M*	*SD*	แปลผล
ด้านการทำงานของระบบ			
ความรวดเร็วในการประมวลผลของระบบ	5.00	0.00	มากที่สุด
ความถูกต้องของผลลัพธ์การทำนาย	4.40	0.82	มาก
ความถูกต้องในการจัดการข้อมูลนำเข้า	4.50	0.70	มาก
ความง่ายต่อการใช้งาน	4.10	0.90	มาก
โดยรวม	4.50	0.61	มาก
ด้านการรักษาความปลอดภัยของข้อมูลในระบบ			
การกำหนดรหัสผู้ใช้ และรหัสผ่านในการตรวจสอบผู้ใช้ระบบ	4.60	0.50	มากที่สุด
การตรวจสอบสิทธิ์ก่อนการใช้งานของผู้ใช้ระบบ	4.70	0.49	มากที่สุด
การควบคุมให้ใช้งานตามสิทธิ์ผู้ใช้ได้อย่างถูกต้อง	4.70	0.49	มากที่สุด
โดยรวม	4.67	0.49	มากที่สุด
เฉลี่ยรวม	4.59	0.55	มากที่สุด

Table 3 ผลการวิเคราะห์ความพึงพอใจของผู้ใช้ที่มีต่อด้านการทำงานของระบบ พบว่า ผู้ใช้มีความพึงพอใจระดับมาก ข้อที่มีค่าเฉลี่ยคะแนนสูงสุด คือ ความรวดเร็วในการประมวลผลของระบบ ความพึงพอใจของผู้ใช้ที่มีต่อด้านการรักษาความปลอดภัยของข้อมูลในระบบ พบว่า ผู้ใช้มีความพึงพอใจระดับมากที่สุด ข้อที่มีค่าเฉลี่ยคะแนนสูงสุด คือ การตรวจสอบสิทธิ์ก่อนการใช้งานของผู้ใช้ระบบ และการควบคุมให้ใช้งานตามสิทธิ์ผู้ใช้ได้อย่างถูกต้อง

โดยสรุปภาพรวมผลการวิเคราะห์ความพึงพอใจของผู้ใช้ที่มีต่อเว็บแอปพลิเคชันที่ใช้ในการทำนายข้อมูลโอกาสที่นักศึกษาชั้นปีที่ 1 จะได้ขึ้นชั้นปีที่ 2 เฉลี่ยรวมทั้งด้านการทำงานของระบบ และด้านการรักษาความปลอดภัยของข้อมูลในระบบ พบว่า ผู้ใช้มีความพึงพอใจระดับมากที่สุด มีค่าเฉลี่ยรวมอยู่ที่ 4.59 และค่าเบี่ยงเบนมาตรฐานรวม 0.55

อภิปรายผล (Discussions)

การเลือกคุณลักษณะสำคัญที่ใช้ในการทำนายในงานวิจัยนี้มาจากการพิจารณาค่าความสำคัญของคุณลักษณะที่ได้จากเทคนิคต้นไม้ตัดสินใจ โดยเลือกคุณลักษณะที่มีค่าความสำคัญ 3 อันดับที่สูงที่สุด โดยคุณลักษณะที่มีค่าความสำคัญอันดับหนึ่ง ได้แก่ เคยเขียนโปรแกรมหรือไม่ มีค่าความสำคัญ 0.358205 อันดับที่สอง แผนการเรียน มีค่าความสำคัญ 0.254090 และอันดับสุดท้าย เกรดเฉลี่ยรวมระดับชั้นมัธยมศึกษาตอนปลาย มีค่าความสำคัญ คือ 0.234765 โดยคุณลักษณะแผนการเรียนและเกรดเฉลี่ยรวมระดับชั้นมัธยมศึกษาตอนปลายเป็นคุณลักษณะที่สอดคล้องกับงานวิจัยของ Thaweechat et al. (2022, p. 1) ที่ได้ศึกษาเรื่อง ระบบทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล ผลการศึกษาพบว่า แบบจำลองการทำนายการพ้นสภาพของนักศึกษามีปัจจัยที่เกี่ยวข้อง 2 ปัจจัย คือ เกรดเฉลี่ยจากโรงเรียนเดิมและหลักสูตรที่จบจากโรงเรียนเดิม และสอดคล้องกับ Chacha et al. (2019, p. 1) ที่ได้ศึกษาเรื่อง เทคนิคการวิเคราะห์การถดถอยโลจีสติกในการทำนายการออกกลางคันของนักศึกษา พบว่า เกรดเฉลี่ยภาคการศึกษาแรกเป็นตัวแปรสำคัญในการทำนาย

งานวิจัยนี้ผู้วิจัยได้สร้างแบบจำลองจากอัลกอริทึม 3 แบบ ได้แก่ เทคนิคการวิเคราะห์การถดถอยโลจีสติก เทคนิคป่าไม้ตัดสินใจ และเทคนิคต้นไม้ตัดสินใจ และนำมาเปรียบเทียบประสิทธิภาพ ผลการวัดประสิทธิภาพ พบว่า แบบจำลองทำนายที่ใช้เทคนิคการวิเคราะห์การถดถอยโลจีสติกมีประสิทธิภาพในการทำนายที่สูงสุด ซึ่งสอดคล้องกับงานวิจัย Wu et al. (2023, p. 1) พบว่า การใช้อัลกอริทึมวิเคราะห์การถดถอยโลจีสติกมีความน่าจะเป็นที่จะช่วยให้การคัดเลือกนักเรียนได้อย่างมีประสิทธิภาพมากขึ้น และ Chacha et al. (2019, p. 330) ได้ใช้เทคนิคการวิเคราะห์การถดถอยโลจีสติกในการทำนายการออกกลางคันของนักศึกษาที่พบว่าให้ค่าความถูกต้องในระดับสูง

สาเหตุที่งานวิจัยที่ผู้วิจัยได้นำเสนอนี้ พบว่า เทคนิคการวิเคราะห์การถดถอยโลจีสติกมีประสิทธิภาพมากที่สุด เนื่องมาจากเทคนิคการวิเคราะห์การถดถอยโลจีสติกเหมาะกับปัญหาการจำแนกกลุ่มในการตัดสินใจ โดยเฉพาะการตัดสินใจแบบทวิภาค (Binary decision) ซึ่งงานวิจัยที่ผู้วิจัยได้นำเสนอนี้ก็มีลักษณะดังกล่าว คือ การตัดสินใจแบบทวิภาคเป็นการทำนายผ่านและไม่ผ่าน ซึ่งสอดคล้องกับทฤษฎีเทคนิคการวิเคราะห์การถดถอยโลจีสติกที่กล่าวว่า เป็นอัลกอริทึมสำหรับการทำนายผลแบบการจำแนกโดยข้อมูลตัวอย่างแต่ละรายการจะมีคอลัมน์ที่เป็นผลลัพธ์แบ่งได้เป็น 2 กลุ่ม (Binary classes) (Bancha, 2021. p. 152; Saraubon, 2020. p. 459)

ผลการวิเคราะห์ความพึงพอใจของผู้ใช้ที่มีต่อเว็บแอปพลิเคชันที่ใช้ในการทำนายข้อมูลโอกาสที่นักศึกษาชั้นปีที่ 1 จะได้ขึ้นชั้นปีที่ 2 พบว่า ผู้ใช้มีความพึงพอใจต่อระดับมากที่สุด มีค่าเฉลี่ยรวมอยู่ที่ 4.59 และค่าเบี่ยงเบนมาตรฐานรวม 0.55 เนื่องจากความง่ายต่อการใช้งานซึ่งสอดคล้องกับงานวิจัยของ Rekha et al. (2021, p. 3770) ที่พบว่า การพัฒนาเว็บแอปพลิเคชันช่วยให้ระบบจัดการขนส่งสาธารณะมีประสิทธิภาพมากขึ้น และใช้งานง่ายขึ้นสำหรับประชาชนทั่วไป และ Clitan et al. (2021, p. 73) พบว่า เว็บแอปพลิเคชันช่วยทำให้การตีความข้อมูลจำนวนมากดูเข้าใจง่ายมากยิ่งขึ้น

ข้อเสนอแนะ (Recommendations)

1. ขั้นตอนการรวบรวมข้อมูลเป็นขั้นตอนที่สำคัญ เนื่องจากการสร้างแบบจำลองที่มีประสิทธิภาพส่วนหนึ่งขึ้นอยู่กับข้อมูลที่ใช้ในการสร้างแบบจำลอง สำหรับการวิจัยนี้การเก็บรวบรวมข้อมูลที่อาจพบปัญหาจากการเก็บข้อมูลนักศึกษาที่พ้นสภาพไปแล้ว ทำให้การรวบรวมข้อมูลค่อนข้างยาก นอกจากนี้ข้อมูลที่เก็บรวบรวมมาต้องตรวจสอบความถูกต้องอีกครั้ง เพราะผู้ให้ข้อมูลอาจให้ข้อมูลที่ไม่ถูกต้องโดยเฉพาะข้อมูลสำคัญ เช่น เกรดเฉลี่ยระดับมัธยมหรือเกรดในชั้นปีที่ 1 ภาคการศึกษาที่ 1 จึงอาจแก้ปัญหาด้วยการนำข้อมูลมาเปรียบเทียบจากแหล่งข้อมูลที่น่าเชื่อถือ เช่น ข้อมูลจากสำนักวิชาการและงานทะเบียนมหาวิทยาลัยราชภัฏบุรีรัมย์

2. เนื่องจากในการวิจัยนี้ใช้ข้อมูลเฉพาะนักศึกษาสาขาวิชาวิทยาการคอมพิวเตอร์ชั้นปีที่ 1-4 ที่ใช้หลักสูตรวิทยาการคอมพิวเตอร์ หลักสูตรปรับปรุง 2561 จึงอาจมีจำนวนข้อมูลค่อนข้างน้อย ทำให้ข้อมูลสำหรับใช้ในการเรียนรู้อาจไม่เพียงพอ ดังนั้น หากสามารถเพิ่มจำนวนข้อมูลอาจช่วยเพิ่มประสิทธิภาพในการทำนายของแบบจำลองให้ดียิ่งขึ้น

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

เนื่องจากงานวิจัยนี้ใช้อัลกอริทึม 3 แบบ ได้แก่ เทคนิคการวิเคราะห์การถดถอยโลจีสติก เทคนิคป่าไม้ตัดสินใจ และเทคนิคต้นไม้ตัดสินใจ แนวทางการพัฒนาระบบในอนาคตอาจใช้เทคนิคการเรียนรู้ของเครื่องอัลกอริทึมแบบอื่น ๆ เพื่อเป็นการพัฒนาแบบจำลองในการทำนายให้มีประสิทธิภาพที่สุด

กิตติกรรมประกาศ (Acknowledgements)

ผู้วิจัยขอขอบพระคุณสำนักส่งเสริมวิชาการและงานทะเบียนของมหาวิทยาลัยราชภัฏบุรีรัมย์ที่ให้ข้อมูลประกอบการทำวิจัย คณาจารย์สาขาวิชาวิทยาการคอมพิวเตอร์ที่ช่วยทดสอบการใช้งานเว็บแอปพลิเคชัน และนักศึกษาสาขาวิชาวิทยาการคอมพิวเตอร์ชั้นปีที่ 1-4 ที่ให้ความร่วมมือในการให้ข้อมูลสำหรับทำวิจัยครั้งนี้

เอกสารอ้างอิง (References)

Akarachantachote, N., & Panitsupakamol, D. (2019). *Comparison of imbalanced data problem solving for income classification of type I pharmacies entrepreneur.* The 9 th STOU National Research Conference, Nonthaburi, Thailand. http://sci-database.hcu.ac.th/science/file/rsID292_F1_20200122213859.pdf

Atik, D. A., Işildar, Y. G., & Erkoç, F. (2021). Prediction of secondary school students' environmental attitudes by a logistic regression model. *Environment, Development and Sustainability, 24*, 4355-4370. https://doi.org/10.1007/S10668-021-01618-3

Bancha, P. (2021). *Sā ngǭkān rīanrū samrap AI dūai Python machine learning* [Create learning for AI with Python machine learning]. Se-Ed.

Beaulac, C., & Rosenthal, S. J. (2019). predicting university students' academic success and major using Random Forests. *Research in Higher Education, 60*(7), 1048-1064. https://doi.org/10.1007/S11162-019-09546-Y

Behr, A., Giese, M., Herve, D., Teguim, K., & Theune, K. (2020). Early prediction of university dropouts – a random forest approach. *Journal of Economics and Statistics, 240*(6), 743-789. https://doi.org/10.1515/JBNST-2019-0006

Chacha, R. C. B., López, L. G. W., Guerrero, X. V. V., & Villacis, G. V. W. (2019, December 3-5). *Student dropout model based on logistic regression.* First International Conference, ICAT 2019, Quito, Ecuador. https://doi.org/10.1007/978-3-030-42520-3_26

Clitan, I., Puscasiu, A., Muresan, V., Unguresan L. M., & Abrudean, M., (2021). Web application for statistical tracking and predicting the evolution of active cases with the novel Coronavirus (SARS-CoV-2). *International Journal of Modeling and Optimization, 11*(3), 70-74. https://doi.org/10.7763/IJMO.2021.V11.780

Dien, T. T., Duy-Anh, L., Hong-Phat. N., Van-Tuan, N., Thanh-Chanh, T., Minh-Bang, L., Thanh-Hai, N., & Thai-Nghe, N. (2021, July 1-3). *Four grade levels-based models with Random Forest for student performance prediction at a multidisciplinary university.* Proceedings of the 15th International Conference on Complex, Intelligent and Software Intensive Systems (CISIS-2021), Asan, Korea. https://doi.org/10.1007/978-3-030-79725-6_1

Elreedy, D., & Atiya, F. A. (2019). A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. *Information Sciences, 505*, 32-64. https://doi.org/10.1016/j.ins.2019.07.070

Fathiya, H., & Sadath, L. (2021, March 17-18). *University admissions predictor using Logistic Regression.* 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Amity University Dubai, UAE46-51. https://doi.org/10.1109/ICCIKE51210.2021.9410717

Gotardo, A. M. (2019). Using decision tree algorithm to predict student performance. *Indian Journal of Science and Technology, 12*(5), 1-8. https://doi.org/10.17485/IJST/2019/V12I5/140987

Grus, J. (2019). *Data science from scratch* (2nd ed). Core Function.

Hussain, S., & Khan, Q. M. (2021). Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. *Annals of Data Science, 10*(3), 637-655. https://doi.org/10.1007/s40745-021-00341-0

Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, P. (2021). *Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. IEEE Access, 9*, 39707-39716. https://doi.org/10.1109/ACCESS.2021.3064084

Kepan, S., Leelapatarapun, P., & Yokkhun, A. (2018). *Analysis of factors influencing the dismissal of students using data mining techniques case study: Computer Science Program and Information Technology Program of Yala Rajabhat University* [Master's thesis, Yala Rajabhat University]. YRU Wisdom Bank. https://wb.yru.ac.th/handle/yru/3447

Khakata, E., Omwenga, O., & Msanjila. S. (2019). Student performance prediction on internet mediated environments using decision trees. *International Journal of Computer Applications, 181*(42), 1-9. https://doi.org/10.5120/IJCA2019918466

Li, S., & Liu, T. (2021). Performance prediction for higher education students using deep learning. *Hindawi Complexity, 2021*, 9958203. https://doi.org/10.1155/2021/9958203

Pacharawongsakda, E. (2020). *A little book of big data and machine learning.* IDC Premier.

Prasetyo, D. H., Hogantara, A. P., & Isnainiyah, N. I. (2021). A web-based diabetes prediction application using XGBoost algorithm. *Journal of Computing and Applied Informatics, 5*(2), 49-59. https://doi.org/10.32734/JOCAI.V5.I2-6290

Rekha, R., Abirami, A. P., Aishvarya, G., Akshaya, B., Annapoorna, K. A., & Sanchana, S. (2021). A web based application for tracking public transport and predicting usage. *International Journal of Aquatic Science, 12*(2), 3770-3783. https://www.journal-aquaticscience.com/article_135799.html

Saraubon, K. (2020). *Rīanrū data science læ AI: Machine learning dūai python* [Learn data science and AI: Machine learning with Python. Media Network.

Srisa-ard, B. (2013). *Kānwičhai būrangton* [Introduction to research] (9th ed.). Suweerivasarn.

Thaweechat, N., Pengprachan, O., Yathongchai, W., & Yathongchai, C. (2022). A prediction system for undergraduate student dropout at faculty of science, Buriram Rajabhat University using data mining techniques. *Science and Technology Buriram Rajabhat University, 4*(1), 47-60. https://ph02.tci-thaijo.org/index.php/scibru/article/view/242082

Wu, J., Lin, M., & Tsai, C. (2023). A predictive model that aligns admission offers with student enrollment probability. *Education Sciences, 13*(5), 1-13. https://doi.org/10.3390/educsci13050440