วารสารวิชาการพระจอมเกล้าพระนครเหนือ ปีที่ 31, ฉบับที่ 3 ก.ค.–ก.ย. 2564
The Journal of KMUTNB, Vol. 31, No. 3, Jul.–Sep. 2021

บทความวิจัย

การสร้างตัวแบบเพื่อทำนายผลสัมฤทธิ์ทางการศึกษาจากพฤติกรรมการใช้สมาร์ตโฟนด้วยเทคนิคการทำเหมืองข้อมูล

วันทนี ประจวบศุภกิจ*
ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
ธัญญาภรณ์ บุญยัง และ สุกัญญา บุญศรี
คณะครุศาสตร์อุตสาหกรรม มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 08 1649 6846 อีเมล: Wanthanee.p@fitm.kmutnb.ac.th
DOI: 10.14416/j.kmutnb.2021.05.017
รับเมื่อ 20 เมษายน 2563 แก้ไขเมื่อ 5 มิถุนายน 2563 ตอบรับเมื่อ 9 มิถุนายน 2563 เผยแพร่ออนไลน์ 24 พฤษภาคม 2564
© 2021 King Mongkut's University of Technology North Bangkok. All Rights Reserved.

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาตัวแบบด้วยเทคนิคการทำเหมืองข้อมูลเพื่อนำมาใช้ในการทำนายผลสัมฤทธิ์ทางการศึกษาของนักศึกษาในหลักสูตรคอมพิวเตอร์จากพฤติกรรมการใช้สมาร์ตโฟน สำหรับการเก็บรวบรวมข้อมูลตัวอย่างใช้แบบสอบถามแบบออนไลน์เพื่อเก็บข้อมูลจากนักศึกษาระดับปริญญาตรี หลักสูตรทางคอมพิวเตอร์ จากมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปราจีนบุรี และมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี จำนวนทั้งสิ้น 623 ตัวอย่าง และ 19 ปัจจัย ซึ่งชุดข้อมูลที่ได้เป็นชุดข้อมูลที่มีจำนวนตัวอย่างในแต่ละคลาสไม่สมดุลกัน ดังนั้นผู้วิจัยจึงเลือกเพิ่มข้อมูลตัวอย่างในคลาสน้อยด้วยขั้นตอนวิธี SMOTE และเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลเพื่อสร้างตัวแบบทั้งสิ้น 5 ขั้นตอนวิธี ประกอบไปด้วย ขั้นตอนวิธีต้นไม้ตัดสินใจ ขั้นตอนวิธีเพื่อนบ้านใกล้เคียง ขั้นตอนวิธีโครงข่ายประสาทเทียมแบบหลายชั้น ขั้นตอนวิธีการเรียนรู้แบบเบย์อย่างง่าย และขั้นตอนวิธีการสุ่มป่าไม้ โดยใช้เกณฑ์ในการประเมินประสิทธิภาพตัวแบบ 4 ตัว คือ ค่าความถูกต้อง ค่าความแม่นยำ ค่าความระลึก และค่าประสิทธิภาพ ซึ่งผลการทดลองพบว่า ขั้นตอนวิธีการสุ่มป่าไม้ มีความสามารถในการเรียนรู้ได้แม่นยำกว่าขั้นตอนวิธีอื่นๆ ในทุกเกณฑ์ที่วัดประสิทธิภาพบนชุดข้อมูลได้ถูกปรับจำนวนตัวอย่างเรียบร้อยแล้ว โดยมีค่าความถูกต้องเท่ากับร้อยละ 78.04 นอกจากนั้นจากการศึกษาพบว่า ปัจจัยพฤติกรรมส่วนบุคคลและพฤติกรรมการเล่นสมาร์ตโฟนที่ต่างกันจะมีผลต่อผลสัมฤทธิ์ทางการเรียนที่แตกต่างกันด้วย

คำสำคัญ: พฤติกรรมการใช้สมาร์ตโฟน ผลสัมฤทธิ์ทางการศึกษา การจำแนกข้อมูล เหมืองข้อมูล การสุ่มป่าไม้

การอ้างอิงบทความ: วันทนี ประจวบศุภกิจ, ธัญญาภรณ์ บุญยัง และ สุกัญญา บุญศรี, "การสร้างตัวแบบเพื่อทำนายผลสัมฤทธิ์ทางการศึกษาจากพฤติกรรมการใช้สมาร์ตโฟนด้วยเทคนิคการทำเหมืองข้อมูล," *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 31, ฉบับที่ 3, หน้า 550–560, ก.ค.–ก.ย. 2564.

Research Article

A Construction Model for Predicting Student's Academic Achievement by Smartphone Usage Behaviors Using Data Mining Techniques

Wanthanee Prachuabsupakij*
Department of Information Technology, Faculty of Industrial and Technology Management, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
Thanyaporn Boonyoung and Sukanya Boonsri
Faculty of Technical Education, Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand

* Corresponding Author, Tel. 08 1649 6846 E-mail: wanthanee.p@fitm.kmutnb.ac.th DOI: 10.14416/j.kmutnb.2021.05.017
Received 20 April 2020; Revised 5 June 2020; Accepted 9 June 2020; Published online: 24 May 2021
© 2021 King Mongkut's University of Technology North Bangkok. All Rights Reserved.

Abstract

The research aims to develop a model using data mining techniques in order to predict students' academic achievement by smartphone usage behaviors. Data collection was carried out through an online survey. The samples contain undergraduate students undertaking Computer Science subjects from these institutions: King Mongkut's University of Technology North Bangkok, Prachinburi Campus and Rajamangala University of Technology Thanyaburi. This dataset contains 623 instances and 19 attributes, which is considered an imbalanced dataset. Therefore, we generated synthetic instances of the minority class using the SMOTE algorithm. The performance comparison of five classifications was conducted to build the model. The feature sets comprise Decision Trees, K-nearest Neighbor, Naïve Bayes, Multi-layer Perceptron and Randomforest using four performances, including Accuracy, Precision, Recall, and F-measure. The result showed that Randomforest obtained the highest values among all measures on balanced dataset. It obtained the accuracy of 78.04%. Moreover, the study results showed that different individual factors and smartphone usage behaviors significantly affect different levels of student achievement.

Keywords: Smartphone Usage Behaviors, Student's Academic Achievement, Classification, Data Mining Random Forest

Please cite this article as: W. Prachuabsupakij, T. Boonyoung, and S. Boonsri, "A construction model for predicting student's academic achievement by smartphone usage behaviors using data mining techniques," *The Journal of KMUTNB*, vol. 31, no. 3, pp. 550–560, Jul.–Sep. 2021 (in Thai).

1. บทนำ

ปัจจุบันสมาร์ตโฟนจัดว่าเป็นสิ่งสำคัญในการดำเนินชีวิตประจำวันของประชาชนในทุกเพศทุกวัย และอาจจะถือว่าเป็นสิ่งประดิษฐ์ที่อำนวยความสะดวกในการใช้ชีวิตประจำวันของหลายๆ คน และในงานวิจัยของจุฑามาศ [1] พบว่า ผู้ใช้สมาร์ตโฟนในประเทศไทยกลุ่มวัยรุ่นมีสัดส่วนการใช้สมาร์ตโฟนสูงสุดเป็นกลุ่มที่มีแนวโน้มในการใช้สมาร์ตโฟนมากขึ้น นอกจากนี้ นักศึกษาในระดับอุดมศึกษาทุกคนมีสมาร์ตโฟนเพื่อตอบสนองความต้องการ บางส่วนนำสมาร์ตโฟนมาใช้ในเวลาเรียนทั้งในฐานะที่เป็นเครื่องมือในการเรียนรู้ หรือแม้แต่นำมาใช้เพื่อวัตถุประสงค์ส่วนตัว จากความนิยมในการใช้สมาร์ตโฟนในข้างต้นทำให้เกิดพฤติกรรมการใช้สมาร์ตโฟนที่ไม่ถูกต้องในกลุ่มนักศึกษาเป็นจำนวนมาก ตัวอย่างเช่น พฤติกรรมการนอนดึก อ่อนเพลีย และที่สำคัญพฤติกรรมเหล่านี้อาจจะนำไปสู่การเกิดปัญหาการเรียนตกต่ำ ซึ่งสอดคล้องกับการศึกษาของภลดา [2] พบว่า หากนักศึกษามีการใช้สมาร์ตโฟนในขณะเรียนจะทำให้ขาดสมาธิ เนื่องจากมีการแอบใช้งานอินเทอร์เน็ตในขณะที่มีการเรียนการสอนโดยให้เหตุผลว่า ส่วนใหญ่จะเล่นในคาบเวลาที่น่าเบื่อ เนื้อหาไม่น่าสนใจ หรือเป็นคาบวิชาที่ส่วนใหญ่อาจารย์จะสอนแบบบรรยาย อีกทั้งนักศึกษาโดยส่วนใหญ่ยังเป็นผู้ที่จัดสรรเวลาไม่เหมาะสมเนื่องจากพฤติกรรมการเล่นสมาร์ตโฟนนี้เองจนทำให้จัดสรรเวลาไม่ถูกต้องว่าเมื่อไหร่ควรจะเล่นสมาร์ตโฟน จนทำให้ส่งผลกระทบต่อการศึกษา จากปัญหาข้างต้นประกอบกับแนวคิดทฤษฎีทางสังคมวิทยาดิจิทัล (Digital Sociology) [3] ได้เสนอว่า ควรมีการศึกษาวิเคราะห์ประเด็นที่เกี่ยวกับปฏิสัมพันธ์ระหว่างคนกับสื่อเทคโนโลยีสมัยใหม่ เช่น วัตถุประสงค์ของการใช้อุปกรณ์สื่อเทคโนโลยีสมัยใหม่ การนำอุปกรณ์มาใช้ในสถานที่ต่างๆ ในชีวิตประจำวัน ระยะเวลาในการใช้อุปกรณ์เหล่านั้น

จากปัญหางานวิจัยข้างต้นพบว่า ปัญหาการใช้สมาร์ตโฟนส่งผลกระทบต่อคุณภาพการศึกษาในปัจจุบันอย่างเห็นได้ชัด ซึ่งหากมีการศึกษาและวิเคราะห์ประเด็นที่เกี่ยวกับพฤติกรรมการใช้สมาร์ตโฟนที่ส่งผลต่อการเรียนของนักศึกษาเพื่อให้ได้ความสัมพันธ์ของปัจจัยที่ส่งผลกระทบต่อการเรียนรู้ของนักศึกษา โดยเฉพาะอย่างยิ่งกับนักศึกษาในหลักสูตรทางคอมพิวเตอร์ซึ่งเป็นหลักสูตรที่ต้องอาศัยการสืบค้นข้อมูลหรือการหาความรู้เพิ่มเติมผ่านแอปพลิเคชันต่างๆ หรือจากเสิร์ชเอนจิน (Search Engine) และด้วยตัวหลักสูตรเองจะต้องมีเนื้อหาเกี่ยวกับการพัฒนาแอปพลิเคชัน การพัฒนาเกม หรือพัฒนาเครื่องมือต่างๆ ที่เกี่ยวข้องกับดิจิทัล

ดังนั้น งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อสร้างตัวแบบที่สามารถนำมาใช้ในการทำนายผลสัมฤทธิ์ทางการเรียนของนักศึกษาจากพฤติกรรมการใช้สมาร์ตโฟนโดยใช้เทคนิคการทำเหมืองข้อมูล โดยผู้วิจัยเลือกใช้เทคนิคการจำแนกข้อมูลโดยเปรียบเทียบตัวจำแนกทั้งสิ้น 5 ตัวจำแนก ประกอบไปด้วย ขั้นตอนวิธีต้นไม้ตัดสินใจ (Decision Trees) ขั้นตอนวิธีเพื่อนบ้านใกล้เคียงที่สุด (K-nearest Neighbor) ขั้นตอนวิธีโครงข่ายประสาทเทียมแบบหลายชั้น (Multi-layer Perceptron) ขั้นตอนวิธีการเรียนรู้แบบเบย์อย่างง่าย (Naïve Bayes) และขั้นตอนวิธีการสุ่มป่าไม้ (Random Forest) โดยใช้เกณฑ์ในการประเมินประสิทธิภาพตัวแบบ 4 ตัว คือ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าประสิทธิภาพโดยรวม (F-measure) เพื่อให้ได้ตัวแบบที่มีประสิทธิภาพสูงสุด รวมถึงการดำเนินการวิเคราะห์ปัจจัยที่เกี่ยวกับการเล่นสมาร์ตโฟนที่ได้จากตัวแบบที่ส่งผลกระทบต่อการผลสัมฤทธิ์ทางการเรียน เพื่อให้ได้องค์ความรู้ที่สามารถนำไปใช้ประโยชน์ได้ในลำดับต่อไป

การสุ่มป่าไม้ [4] หลักการของการสุ่มป่าไม้ คือ สร้างตัวแบบจากต้นไม้ตัดสินใจหลายๆ ตัวแบบย่อยๆ โดยแต่ละตัวแบบจะได้รับชุดข้อมูลที่ไม่เหมือนกัน (Random Sampling) โดยข้อมูลย่อยจะถูกแบ่งออกมาเป็น n ชุดที่ไม่เหมือนกันตามจำนวนต้นไม้ตัดสินใจ ขั้นตอนวิธีนี้จะสร้างตัวแบบเพื่อทำการพยากรณ์คำตอบของตนเองด้วยการโหวต ผลโหวตที่มากที่สุดจะถูกเลือกมาใช้ระบุสถานะของคลาส

ขั้นตอนวิธีการปรับข้อมูลตัวอย่างด้วยขั้นตอนวิธี SMOTE นำเสนอโดย Chawla [5] ซึ่งได้เพิ่มข้อมูลตัวอย่างของคลาสบวกหรือ Minority Class โดยการสร้างตัวอย่าง

สังเคราะห์ (Synthetic Samples) จากการสร้าง Feature Space และใช้ขั้นตอนวิธี K-Nearest Neighbor ในการหาเพื่อนบ้านใกล้เคียงตามจำนวน k ของข้อมูลตัวอย่างคลาสบวก x หลังจากนั้นจะสุ่มข้อมูลตัวอย่าง 1 ตัว คือ y ที่อยู่ในกลุ่ม k แล้วหาระยะห่างระหว่าง x และ y โดย SMOTE จะสุ่มค่า Gap ขึ้นมาระหว่าง 0 กับ 1 เพื่อนำค่า Gap นี้มาสร้างข้อมูลตัวอย่างสังเคราะห์ โดยนำมาคูณกับค่าระยะห่างแล้วบวกกับ x

จากการสำรวจงานวิจัยที่เกี่ยวข้องพบว่า มีการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลมาใช้เกี่ยวกับผลสัมฤทธิ์ทางการศึกษาและรูปแบบของนักศึกษา เช่น งานวิจัยของจีระนันต์ และคณะ [6] ได้วิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพนักศึกษาที่มีผลการเรียนปกติโดยใช้ต้นไม้ตัดสินใจ ซึ่งผลที่ได้จากการวิจัยพบว่า ตัวแบบให้ค่าความถูกต้องที่ร้อยละ 97.46 อย่างไรก็ตาม งานวิจัยนี้อาศัยข้อมูลจากระบบงานลงทะเบียนเพียงอย่างเดียวเท่านั้น เช่น เพศ รายได้ของผู้ปกครอง เกรดทุนการศึกษา ซึ่งปัจจัยที่ส่งผลกระทบต่อการออกกลางคันของนักศึกษานั้น อาจจะมีปัจจัยด้านพฤติกรรมของตัวนักศึกษาเองเข้ามามีส่วนเกี่ยวข้อง ดังนั้นจึงควรมีการทำแบบสอบถามหรือสัมภาษณ์พฤติกรรมของนักศึกษาที่อาจจะมีผลต่อการออกกลางคันเพิ่มเติมด้วย

นอกจากนี้ยังพบว่า มีงานวิจัยของ Zhang และคณะ [7] ที่ได้นำเทคนิคการทำเหมืองข้อมูลมาใช้ในการทำนายผลสัมฤทธิ์ทางการเรียนของนักศึกษาใช้ปัจจัยในการทำนายที่เกี่ยวข้องกับเกรด พฤติกรรมการบริโภคและการเข้าห้องสมุด เช่น จำนวนครั้งการยืมหนังสือห้องสมุด และจำนวนครั้งการเข้าห้องสมุด โดยเปรียบเทียบขั้นตอนวิธีในการจำแนกทั้งหมด 4 ตัวจำแนก คือ ขั้นตอนวิธีโครงข่ายประสาทเทียมแบบหลายชั้น ขั้นตอนวิธีการเรียนรู้เบย์อย่างง่าย ขั้นตอนวิธีต้นไม้ตัดสินใจ และขั้นตอนวิธีซัพพอร์ตเวกเตอร์แม่ชชีน ผลการเปรียบเทียบพบว่า ขั้นตอนวิธีโครงข่ายประสาทเทียมแบบหลายชั้นให้ค่าประสิทธิภาพในการจำแนกที่ดีที่สุดเท่ากับร้อยละ 62.04

อีกงานวิจัยหนึ่งที่มีการนำเทคนิคการทำเหมืองข้อมูลมาวิเคราะห์หารูปแบบการเรียนรู้ของนักศึกษาต่อการจัดทำปริญญานิพนธ์ของฐิติมา [8] โดยผู้วิจัยได้ยึดหลักทฤษฎีของเดวิด คอล์บ มาศึกษากับนักศึกษาระดับปริญญาตรี สาขาการจัดการเทคโนโลยีการผลิตและสารสนเทศได้เลือกวิธีการวิเคราะห์ด้วยรูปแบบ Rule Based Classification ด้วยวิธี Decision Table, Jrip, PART และเลือกวิธีวิเคราะห์ผลด้วยรูปแบบ Decision Tree Classification ด้วยวิธี LMT, J48, Random Tree จากการวิเคราะห์ผลทั้งหมดพบว่ารูปแบบของ Decision Tree ให้ค่าความถูกต้องสูงสุด (100%) ทำนายรูปแบบการเรียนรู้เป็นแบบเอกนัย (Converged) ส่วนรูปแบบของ Rule Based ด้วยวิธี PART ให้ค่าความถูกต้องสูงสุด (84.12%)

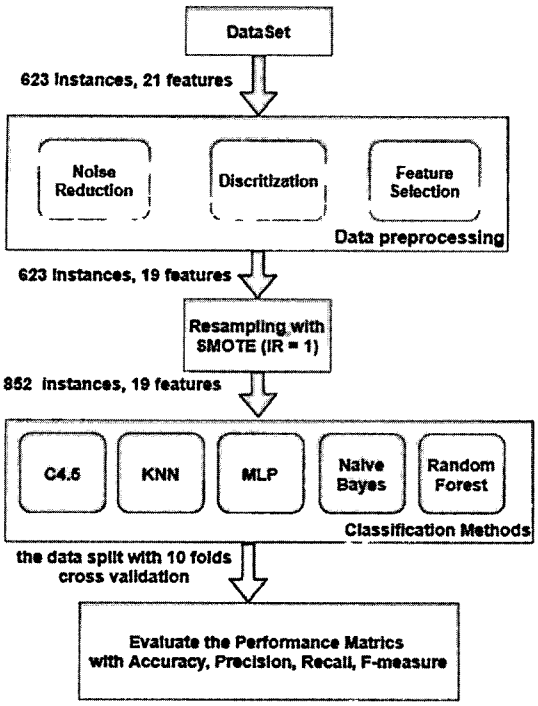
จากงานวิจัยที่สำรวจพบว่า มีหลายงานวิจัยที่นำเทคนิคการทำเหมืองข้อมูลมาใช้ในการทำนายผลสัมฤทธิ์ทางการศึกษา ซึ่งแต่ละงานมีความแตกต่างกันในเรื่องของปัจจัยและขั้นตอนวิธีที่เลือกใช้ อีกทั้งพบว่า ไม่มีการใช้เทคนิคดังกล่าวกับพฤติกรรมการใช้สมาร์ตโฟนที่ส่งผลกระทบต่อผลสัมฤทธิ์ทางการเรียน ดังนั้น ผู้วิจัยจึงมีแนวคิดในการสร้างตัวแบบเพื่อนำไปใช้ให้เกิดประโยชน์ในอนาคต

2. วัสดุ อุปกรณ์และวิธีการวิจัย

ขั้นตอนในการดำเนินงานวิจัยมี 4 ขั้นตอน แสดงดังรูปที่ 1 ดังนี้

2.1 การรวบรวมข้อมูล

ในการรวบรวมข้อมูลนั้น ผู้วิจัยได้รวบรวมข้อมูลจากแบบสอบถามทั้งหมด 4 ปีการศึกษา จากนักศึกษาหลักสูตรที่เกี่ยวกับคอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปราจีนบุรี (KMUTNB) และมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี (RMUTT) ตั้งแต่ปีการศึกษา 2558 ถึง 2562 มีจำนวนนักศึกษาที่ตอบแบบสอบถามทั้งสิ้น 623 คน 19 แอททริบิวต์ จำนวน 7 หลักสูตร คือ เทคโนโลยีสารสนเทศ (KMUTNB) เทคโนโลยีสารสนเทศเพื่ออุตสาหกรรม (KMUTNB) คอมพิวเตอร์ธุรกิจ (RMUTT) คอมพิวเตอร์ศึกษา (RMUTT) วิศวกรรมคอมพิวเตอร์ (RMUTT) เทคโนโลยีสารสนเทศ (RMUTT) และเทคโนโลยี



รูปที่ 1 ขั้นตอนวิธีการดำเนินงานวิจัย

สารสนเทศทางการศึกษา (RMUTT)

แต่อย่างไรก็ตาม การจะวัดผลสัมฤทธิ์ทางการเรียนโดยดูจากคะแนน GPA นั้น อาจจะไม่เหมาะสมเนื่องจากนักศึกษาอยู่ต่างมหาวิทยาลัย ดังนั้น ผู้วิจัยจึงได้ดำเนินการจัดทำแบบทดสอบความรู้ทางคอมพิวเตอร์ที่ผ่านการประเมินจากผู้เชี่ยวชาญ โดยพิจารณาหัวข้อที่จะใช้ในการทดสอบให้ครอบคลุมเนื้อหาสาระตามผู้เชี่ยวชาญเสนอแนะ หลังจากนั้นหาความตรงของแบบทดสอบโดยการตรวจสอบความตรงของเนื้อหาจากผู้เชี่ยวชาญ (Item Objective Congruence; IOC) จำนวน 3 ท่าน โดยเลือกข้อสอบที่มีค่า IOC ตั้งแต่ 0.50 ขึ้นไป จำนวน 40 ข้อ โดยมีผลการวิเคราะห์หาความสัมพันธ์ระหว่างตัวแปรผลสัมฤทธิ์ทางการเรียนจากแบบสอบถาม GPA (ปัจจุบัน) GPA (ก่อนเข้ามหาวิทยาลัย) และผลคะแนนสอบ (SCORE) จากแบบทดสอบดังตารางที่ 1

จากตารางที่ 1 พบว่า คะแนนสอบทักษะทางคอมพิวเตอร์ของนักศึกษาสาขาเทคโนโลยีสารสนเทศมีค่าเฉลี่ย 21.41 คะแนน จาก 40 คะแนน และส่วนใหญ่มีเกรดเฉลี่ยปัจจุบันและก่อนเข้ามหาวิทยาลัยเท่ากับ 3.06 และ 3.05 ตามลำดับ เมื่อพิจารณาความสัมพันธ์ของตัวแปรทั้งหมดพบว่า เกรดเฉลี่ยปัจจุบันกับเกรดเฉลี่ยก่อนเข้ามหาวิทยาลัยมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.699 รองลงมาคือ คะแนนสอบกับเกรดเฉลี่ยปัจจุบันมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.327 และคะแนนสอบกับเกรดเฉลี่ยก่อนเข้ามหาวิทยาลัยมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.255 ตามลำดับ

ตารางที่ 1 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันของตัวแปร คะแนนสอบ เกรดเฉลี่ยปัจจุบัน และเกรดเฉลี่ยก่อนเข้ามหาวิทยาลัย

ตัวแปร	คะแนนสอบ (SCORE)	GPA (ปัจจุบัน)	GPA (ก่อนเข้ามหาวิทยาลัย)
คะแนนสอบ (SCORE)	-		
GPA (ปัจจุบัน)	0.327**	-	
GPA (ก่อนเข้ามหาวิทยาลัย)	0.255**	0.699**	-
ค่าเฉลี่ย (M)	21.41	3.06	3.05
ส่วนเบี่ยงเบนมาตรฐาน (SD)	5.45	0.53	0.52

** $p < 0.01$

จากความสัมพันธ์ดังกล่าวผู้วิจัยจึงนำคะแนนสอบจากแบบทดสอบ (SCORE) มาเป็นตัวแปรในการทำนายผลสัมฤทธิ์ทางการเรียน (Class) ของนักศึกษา เนื่องจากตัวแปรทั้งสามตัวมีความสัมพันธ์กันในเชิงสถิติที่ระดับ 0.01 จึงสามารถนำตัวแปรใดตัวแปรหนึ่งมาเป็นตัวแปรในการทำนายผลสัมฤทธิ์ทางการเรียนได้ โดยในครั้งนี้ผู้วิจัยเลือกตัวแปรคะแนนสอบจากแบบทดสอบ เนื่องจากว่าเป็นตัวแปร

ที่วัดผลสัมฤทธิ์ในเรื่องทักษะทางคอมพิวเตอร์โดยเฉพาะ อีกทั้งกลุ่มตัวอย่างทั้งหมดใช้แบบทดสอบชุดเดียวกัน ในขณะที่ตัวแปร GPA เป็นตัวแปรที่วัดผลสัมฤทธิ์รวมหลายๆ ด้าน และมีความแตกต่างกันในแต่ละมหาวิทยาลัย ดังนั้นในงานวิจัยนี้จึงเลือกใช้คะแนนสอบเป็นคำตอบในการวัดผลสัมฤทธิ์ทางการศึกษาของนักศึกษาจากทั้งสองมหาวิทยาลัย

2.2 ขั้นตอนการจัดเตรียมและแปลงข้อมูล

2.2.1 การตรวจสอบความถูกต้องของข้อมูล โดยผู้วิจัยพบว่า จากข้อมูลที่รวบรวมมาได้ทั้งหมด มีข้อมูลนักศึกษาจำนวน 1 คน ที่กรอกเกรดเฉลี่ยปัจจุบันเกินเกณฑ์ที่กำหนด (Noise Value) ผู้วิจัยจึงได้ดำเนินการแทนที่ค่านั้น ด้วยการหาค่าคำตอบที่ถูกต้องมาใส่แทนโดยตรวจสอบข้อมูลจากระบบข้อมูลงานทะเบียนนักศึกษา

2.2.2 การแบ่งข้อมูลเป็นช่วง (Discretization) ผู้วิจัยได้แบ่งคะแนนทดสอบออกเป็น 2 ประเภท ผลที่ได้จากการแบ่งมีดังนี้

ช่วงที่ 1 L = คะแนนต่ำกว่าค่ากลาง คือ คะแนนต่ำกว่า 20 คะแนน จำนวน 197 คน โดยให้กลุ่มนี้คือ นักเรียนกลุ่มที่มีแนวโน้มผลการเรียนที่ควรเฝ้าระวัง

ช่วงที่ 2 H = คะแนนสูงกว่าค่ากลาง คือ คะแนนตั้งแต่ 20 คะแนนขึ้นไป ซึ่งจำนวน 426 คน โดยให้กลุ่มนี้คือ นักเรียนกลุ่มที่มีแนวโน้มผลการเรียนดี

และแบ่งช่วงของเกรดเฉลี่ยก่อนเข้ามหาวิทยาลัยเป็น 2 ช่วง ให้ความกว้างของแต่ละช่วงเท่ากันดังนี้

ช่วงที่ 1 Low คือ ค่าเกรดเฉลี่ยก่อนเข้ามหาวิทยาลัยต่ำกว่า 2.50 จำนวน 279 คน

ช่วงที่ 2 High คือ ค่าเกรดเฉลี่ยก่อนเข้ามหาวิทยาลัยตั้งแต่ 2.50 ขึ้นไป จำนวน 344 คน

2.2.3 การคัดเลือกคุณลักษณะสำคัญ ผู้วิจัยได้เลือกใช้เทคนิค Feature Selection ด้วยวิธีการที่หลากหลาย เช่น Infomation Gain และ Relief เพื่อคัดเลือกคุณลักษณะสำคัญ ผลปรากฏว่าทั้ง 19 แอททริบิวต์ มีความสำคัญที่จะทำให้จำแนกประเภทได้ถูกต้อง ดังนั้น จึงทำให้ไม่ต้องตัดแอททริบิวต์ใดๆ ออกจากชุดข้อมูลเลย โดยชุดข้อมูลมีทั้งสิ้น 623 รายการ 19 แอททริบิวต์ ดังรายละเอียดดังตารางที่ 1

2.2.4 การปรับจำนวนข้อมูลตัวอย่างในแต่ละคลาส จากชุดข้อมูลฝึกที่ได้พบว่า ข้อมูลคลาสคำเฉลยซึ่งก็คือคะแนนทดสอบ มีจำนวนข้อมูลนักศึกษาที่มีได้มีผลการเรียนที่ควรเฝ้าระวัง 197 คน คิดเป็นร้อยละ 32 ของข้อมูลทั้งหมด ส่วนข้อมูลนักศึกษาที่มีแนวโน้มผลการเรียนดีจำนวน 426 คน คิดเป็นร้อยละ 68 ของข้อมูล ซึ่งมีความไม่สมดุลของคลาสคำเฉลย คิดเป็นอัตราความไม่สมดุล (Imbalance Ratio; IR) ระหว่างคลาสผลการเรียนมีแนวโน้มที่ดี (Majority Class; คลาสลบ) กับคลาสผลการเรียนที่ควรเฝ้าระวัง (Minority class; คลาสบวก) เท่ากับ 2.16 โดยอัตราความไม่สมดุลหรือ IR คือ สัดส่วนของคลาสที่มีจำนวนมากที่สุดกับจำนวนของคลาสที่มีจำนวนน้อยที่สุด (426/197 = 2.16) ซึ่งจากความไม่สมดุลนี้ส่งผลให้ตัวจำแนกพื้นฐานจำแนกข้อมูลได้ดีในคลาสผลการเรียนมีแนวโน้มที่ดี ในทางตรงกันข้ามจะจำแนกข้อมูลคลาสผลการเรียนที่ควรเฝ้าระวังได้ไม่ถูกต้อง

ดังนั้นผู้วิจัยจึงได้ดำเนินการปรับข้อมูลตัวอย่างให้มีการกระจายตัวของข้อมูลตัวอย่างในแต่ละคลาสให้มีความสมดุลกันโดยใช้ขั้นตอนวิธี SMOTE เพื่อให้มีจำนวนตัวอย่างในแต่ละคลาสที่เท่ากันด้วยเหตุผลที่ว่าขั้นตอนวิธี SMOTE เป็นขั้นตอนวิธีที่สร้างตัวอย่างสังเคราะห์แทนการเพิ่มข้อมูลตัวอย่างแบบซ้ำๆ ซึ่งผลการทดลองจากงานวิจัยจำนวนมากพบว่า การนำ SMOTE มาปรับจำนวนข้อมูลตัวอย่างทำให้ประสิทธิภาพในการจำแนกเพิ่มขึ้น และไม่ทำให้เกิดปัญหา Overfitting โดยผู้วิจัยได้กำหนดอัตราความไม่สมดุลให้มีค่าเท่ากับ 1 (IR = 1) ดังนั้นชุดข้อมูลหลังการปรับความสมดุลจึงมีจำนวนข้อมูลตัวอย่างทั้งสิ้น 852 ตัวอย่าง และมีจำนวนแอททริบิวต์เท่ากับ 19 แอททริบิวต์

2.3 สร้างตัวแบบสำหรับการจำแนกประเภทข้อมูล

โดยผู้วิจัยได้ดำเนินการทดลองโดยใช้ขั้นตอนวิธีประเภทจำแนกข้อมูล (Classification) 5 ตัวจำแนก ซึ่งผู้วิจัยได้กำหนดค่าพารามิเตอร์ของแต่ละวิธีดังนี้

1) ต้นไม้ตัดสินใจ (C4.5 หรือ J48 ใน Weka) กำหนดค่า Confident Factor เท่ากับ 0.25

2) ขั้นตอนวิธีเพื่อนบ้านใกล้ที่สุด (KNN) กำหนดค่า K เท่ากับ 7 เนื่องจากผู้วิจัยได้ทำการทดสอบแล้วพบว่า ค่า 7 คือค่าที่มีประสิทธิภาพที่ดีที่สุดของ KNN และวิธีการวัดระยะทางแบบยุคลิด (Euclidean Distance)

ตารางที่ 2 แอททริบิวต์ที่ใช้ในการสร้างตัวแบบ

ที่	ชื่อข้อมูล	ตัวอย่าง
1.	เพศ	ชาย หญิง
2.	ชั้นปีที่กำลังศึกษา	1, 2, 3, 4, 5
3.	สาขาวิชา/หลักสูตร	IT, ComEd, ITI, ComSci, ComB, ITEd, ComEng
4.	GPA ก่อนเข้าศึกษาในมหาวิทยาลัย	Low, High
5.	ค่าบริการโทรศัพท์มือถือต่อเดือน	น้อยกว่า 300 บาท ตั้งแต่ 300 บาทแต่ไม่เกิน 600 บาท ตั้งแต่ 600 บาทแต่ไม่เกิน 900 บาท มากกว่า 900 บาทขึ้นไป
6.	การเริ่มใช้งานสมาร์ตโฟน	น้อยกว่า 3 ปี ตั้งแต่ 3 ปี แต่ไม่เกิน 5 ปี ตั้งแต่ 5 ปีแต่ไม่เกิน 8 ปี มากกว่า 8 ปี
7.	จำนวนชั่วโมงของการใช้สมาร์ตโฟนต่อวัน	น้อยกว่า 3 ชั่วโมง ตั้งแต่ 3 ชั่วโมงแต่ไม่เกิน 5 ชั่วโมง ตั้งแต่ 5 ชั่วโมงแต่ไม่เกิน 7 ชั่วโมง ตั้งแต่ 7 ชั่วโมงขึ้นไป
8.	ช่วงเวลาที่ใช้สมาร์ตโฟนบ่อยที่สุด	06.01–12.00 น., 12.01–18.00 น., 18.01–24.00 น., 24.01–06.00 น.
9.	สถานที่ที่ใช้สมาร์ตโฟนมากที่สุด	หอพัก ห้อง Lab บ้าน รอบๆ อาคารเรียน ร้านอาหาร/ร้านเครื่องดื่ม
10.	ใช้สมาร์ตโฟนเพื่อค้นคว้าข้อมูล และทำกิจกรรมที่อาจารย์มอบหมายในห้องเรียน	ไม่เคย บางครั้ง ทุกครั้ง
11.	ใช้สมาร์ตโฟนเพื่อค้นคว้าเนื้อหาที่ไม่เข้าใจเพิ่มเติมนอกห้องเรียน	ไม่เคย บางครั้ง ทุกครั้ง
12.	ใช้สมาร์ตโฟนเพื่อถ่ายรูป Power Point ของอาจารย์แทนการจัดบันทึก	ไม่เคย บางครั้ง ทุกครั้ง
13.	ใช้สมาร์ตโฟนเพื่อติดต่อ/หรือสอบถามเกี่ยวกับเรื่องการเรียน กับเพื่อน	ไม่เคย บางครั้ง ทุกครั้ง
14.	ระยะเวลาการแอบใช้สมาร์ตโฟนในห้องเรียน	มากกว่าครึ่งหนึ่งของเวลาเรียน ครึ่งหนึ่งของเวลาเรียน น้อยกว่าครึ่งหนึ่งของเวลาเรียน
15.	วัตถุประสงค์หลักที่แอบใช้สมาร์ตโฟนในห้องเรียน	โทรเข้า-ออก รับส่ง E-mail ใช้สื่อสังคมออนไลน์ เล่นอินเทอร์เน็ต ฟังเพลง ภาพยนตร์ เล่นเกม
16.	ช่วงเวลาที่แอบใช้สมาร์ตโฟนในห้องเรียน	ช่วงเวลาที่เพื่อนนำเสนอหน้าห้องเรียน ช่วงเวลาที่อาจารย์ให้นักศึกษาทำงานกลุ่ม ช่วงเวลาที่อาจารย์สอนแบบบรรยาย ช่วงเวลาที่อาจารย์สอนปฏิบัติ
17.	ใช้สมาร์ตโฟนสำหรับนำเสนอผลงาน/งานที่ได้รับมอบหมายหน้าชั้นเรียน	ไม่เคย บางครั้ง ทุกครั้ง
18.	ใช้สมาร์ตโฟนเพื่อติดต่อ/สอบถามอาจารย์ผู้สอนเกี่ยวกับเรื่องการเรียน	ไม่เคย บางครั้ง ทุกครั้ง
19.	คะแนนทดสอบที่ได้ในปัจจุบัน (Class)	L, H

3) ขั้นตอนวิธีโครงข่ายประสาทเทียมแบบหลายชั้น (MLP) กำหนดค่า Learning Rate เท่ากับ 0.3 และ Momentum เท่ากับ 0.2

4) ขั้นตอนวิธีการเรียนรู้เบย์อย่างง่าย ไม่ได้มีการกำหนดค่าพารามิเตอร์เพิ่มเติม

5) ขั้นตอนวิธีการสุ่มป่าไม้ ได้กำหนดค่า Number Iterations เท่ากับ 70 เนื่องจากผู้วิจัยได้ทำการทดสอบแล้วพบว่า ค่าจำนวนรอบเท่ากับ 70 คือ ค่าที่มีประสิทธิภาพที่ดีที่สุดของวิธีการนี้

2.4 การทดสอบความถูกต้องของตัวแบบ

ผู้วิจัยเลือกรูปแบบการทดสอบความถูกต้องแบบ 10-fold Cross Validation โดยใช้เกณฑ์ในการประเมินประสิทธิภาพตัวแบบ 4 ตัว ดังแสดงในตารางที่ 3

ตารางที่ 3 ตัววัดประสิทธิภาพของการจำแนกคลาส

มาตรวัด	สมการ
ค่าความถูกต้อง (Accuracy)	$\frac{(TP+TN)}{(TP+TN+FP+FN)}$
ค่าความแม่นยำ (Precision)	$\frac{TP}{(TP+FP)}$
ค่าความระลึก (Recall)	$\frac{TP}{(TP+FN)}$
ค่าประสิทธิภาพโดยรวม (F-Measure)	$\frac{(2*Recall*Precision)}{(Recall+Precision)}$

3. ผลการทดลอง

3.1 ผลการเปรียบเทียบประสิทธิภาพของตัวจำแนก

ผู้วิจัยแสดงผลการทดลอง 2 ส่วน คือ การเปรียบเทียบประสิทธิภาพของตัวจำแนกกับชุดข้อมูลต้นฉบับ และการเปรียบเทียบประสิทธิภาพของตัวจำแนกกับชุดข้อมูลที่ผ่านการปรับความสมดุลของจำนวนตัวอย่างในแต่ละคลาสเรียบร้อยแล้วแสดงดังตารางที่ 4 และตารางที่ 5 ตารางที่ 4 แสดงผลการทดลองจากชุดข้อมูลต้นฉบับซึ่งมีจำนวนทั้งสิ้น 623 ตัวอย่าง พบว่า ขั้นตอนวิธี Naivebayes ให้ประสิทธิภาพในการจำแนกได้ดีที่สุดในชุดข้อมูลต้นฉบับในทุกมาตรวัดประสิทธิภาพ โดยให้ค่าความถูกต้องโดยรวมเท่ากับร้อยละ 74.76 และให้ค่าเฉลี่ยของค่าความแม่นยำ ค่าความระลึก และค่าประสิทธิภาพโดยรวมเท่ากับ 0.746, 0.748 และ 0.747 ตามลำดับ แต่เมื่อพิจารณาที่การจำแนกในแต่ละคลาสแล้วพบว่า คลาสที่มีแนวโน้มผลการเรียนที่ควรเฝ้าระวังหรือ L มีค่าประสิทธิภาพที่ต่ำกว่าคลาสอีกคลาสหนึ่งพอสมควร ซึ่งเป็นไปในแนวทางเดียวกันกับทุกขั้นตอนวิธี หากพิจารณาที่ขั้นตอนวิธี Naivebayes จะเห็นว่าค่าความแม่นยำของคลาส L มีค่าเท่ากับ 0.682 ในขณะที่คลาส H มีค่าเท่ากับ 0.784 สำหรับค่าความระลึกและค่าประสิทธิภาพโดยรวมของคลาส L นั้น ก็เป็นไปในแนวทางเดียวกันคือต่ำกว่าคลาส H โดยเฉพาะอย่างยิ่งค่าความระลึกจะเห็นว่าคลาส L มีค่าเท่ากับ 0.660 ในขณะที่คลาส H นั้น มีค่าเท่ากับ 0.804

ตารางที่ 4 ผลการเปรียบเทียบประสิทธิภาพของตัวจำแนกกับชุดข้อมูลเบื้องต้น

Agorithms / Classes	Precision			Recall			F-measure			Accuracy
	H	L	Macro Avg.	H	L	Macro Avg.	H	L	Macro Avg.	Overall
Random Forest	0.763	0.681	0.731	0.819	0.602	0.735	0.790	0.639	0.731	73.48%
C4.5	0.771	0.671	0.732	0.804	0.627	0.735	0.787	0.648	0.733	73.48%
KNN	0.727	0.637	0.692	0.809	0.525	0.698	0.766	0.575	0.692	69.81%
MLP	0.749	0.621	0.699	0.767	0.598	0.701	0.758	0.610	0.700	70.13%
Naivebayes	0.787	0.682	0.746	0.804	0.660	0.748	0.795	0.671	0.747	74.76%

ตารางที่ 5 ผลการเปรียบเทียบประสิทธิภาพของตัวจำแนกกับชุดข้อมูลที่ปรับจำนวนตัวอย่างด้วย SMOTE

Agorithms / Classes	Precision			Recall			F-measure			Accuracy
	H	L	Macro Avg.	H	L	Macro Avg.	H	L	Macro Avg.	Overall
Random Forest +SMOTE	0.796	0.767	0.781	0.754	0.807	0.780	0.774	0.786	0.780	78.04%
C4.5+SMOTE	0.752	0.739	0.745	0.730	0.760	0.745	0.741	0.749	0.745	74.51%
KNN+SMOTE	0.797	0.682	0.739	0.605	0.846	0.725	0.688	0.755	0.721	72.55%
MLP+SMOTE	0.742	0.725	0.734	0.715	0.733	0.752	0.728	0.738	0.733	73.33%
Naivebayes+ SMOTE	0.760	0.767	0.763	0.770	0.757	0.763	0.765	0.762	0.763	76.33%

สำหรับผลการทดลองจากชุดข้อมูลที่ผ่านการปรับความสมดุลด้วย SMOTE มีจำนวนทั้งสิ้น 852 ตัวอย่าง แบ่งเป็นจำนวนตัวอย่างของคลาส H จำนวน 426 ตัวอย่าง และจำนวนตัวอย่างของคลาส L จำนวน 426 ตัวอย่าง แสดงดังตารางที่ 5 พบว่า ขั้นตอนวิธี Random Forest+SMOTE นั้น ให้ค่าความถูกต้องได้สูงสุดที่ร้อยละ 78.04 ในขณะที่ขั้นตอนวิธี Naivebayes+SMOTE ให้ค่าความถูกต้องรองลงมาเป็นอันดับที่สองด้วยค่าความถูกต้องเท่ากับร้อยละ 76.33 และเมื่อพิจารณาในแต่ละคลาสในแง่ของค่าความแม่นยำพบว่าขั้นตอนวิธี Random Forest+SMOTE และ Naivebayes + SMOTE ให้ค่าความแม่นยำของคลาส L ได้เท่ากันคือ 0.767 ในขณะที่คลาส H ขั้นตอนวิธี KNN+SMOTE ให้ค่าความเที่ยงตรงที่ดีที่สุดเท่ากับ 0.797 ส่วนค่าความระลึกของคลาส L นั้น ขั้นตอนวิธี KNN+SMOTE ให้ค่าระลึกที่ดีที่สุดเท่ากับ 0.846 ส่วนคลาส H จะเป็นขั้นตอนวิธี Random Forest +SMOTE ให้ค่าความระลึกที่ดีที่สุดเท่ากับ 0.754

3.2 ผลการวิเคราะห์ปัจจัยพฤติกรรมการเล่นสมาร์ตโฟนที่ส่งผลกระทบต่อผลสัมฤทธิ์ทางการเรียน

อย่างไรก็ตาม ถึงแม้ว่าตัวแบบที่ได้จากขั้นตอนวิธี Random Forest จะให้ค่าประสิทธิภาพที่ดีที่สุด แต่ด้วยขั้นตอนวิธีดังกล่าวไม่สามารถที่จะอธิบายความสัมพันธ์ของปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนของนักศึกษาในหลักสูตรคอมพิวเตอร์ได้ ดังนั้น ผู้วิจัยจึงได้หาความสัมพันธ์ของปัจจัยด้วยขั้นตอนวิธี C4.5 ซึ่งให้ค่าประสิทธิภาพโดยรวมใกล้เคียงกับ Random Forest มีรายละเอียดดังนี้

3.2.1 มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปราจีนบุรี มีความสัมพันธ์ของปัจจัยหลักๆ ดังนี้

นักศึกษาที่มีเกรดเฉลี่ยก่อนเข้ามหาวิทยาลัยต่ำกว่า 2.50 จะมีแนวโน้มที่มีผลการเรียนที่ควรเฝ้าระวัง

ในขณะที่ถ้านักศึกษามีเกรดก่อนเข้ามหาวิทยาลัย 2.50 ขึ้นไปนั้น ต้องพิจารณาเพศประกอบ โดยถ้าเป็นเพศชาย จะมีแนวโน้มที่มีผลการเรียนดี ถ้าเป็นเพศหญิง ปัจจัยปริมาณการใช้สมาร์ตโฟนต่อวันจะมีผลต่อการจำแนกดังนี้

- ใช้สมาร์ตโฟนน้อยว่า 3 ชั่วโมงต่อวัน จะมีแนวโน้มที่มีผลการเรียนที่ไม่ดี
- ใช้สมาร์ตโฟนระหว่าง 3–5 ชั่วโมงต่อวัน จะมีผลการเรียนที่ควรเฝ้าระวัง
- ใช้สมาร์ตโฟนระหว่าง 5–7 ชั่วโมงต่อวัน แล้วใช้สมาร์ตโฟนในการถ่ายภาพ Power Point หรือเอกสารบนกระดานบางครั้งจะมีแนวโน้มที่มีผลการเรียนดี แต่ถ้าถ่ายเสมอจะมีผลการเรียนที่ควรเฝ้าระวัง
- ใช้สมาร์ตโฟนมากกว่า 7 ชั่วโมงต่อวัน ถ้าเป็นปี 1 จะมีแนวโน้มที่มีผลการเรียนที่ควรเฝ้าระวัง แต่ถ้าเป็นปี 2 ขึ้นไปจะมีแนวโน้มที่มีผลการเรียนดี

3.2.2 มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี มีความสัมพันธ์ของปัจจัยหลักๆ ดังนี้

ถ้าเป็นนักศึกษาเพศหญิง มีแนวโน้มที่มีผลการเรียนที่ควรเฝ้าระวังในขณะที่ถ้าเป็นเพศชาย และมีการใช้สมาร์ตโฟนในการถ่ายภาพ Power Point เสมอมีแนวโน้มที่จะมีผลการเรียนที่ควรเฝ้าระวัง แต่ถ้าใช้บางครั้งและเป็นปี 1 ก็ยังมีแนวโน้มที่มีผลการเรียนที่ควรเฝ้าระวังเช่นกัน แต่ถ้าเป็นปี 2 ขึ้นไปแล้วมีเกรดก่อนเข้าศึกษาอยู่ในระดับดีก็จะมีแนวโน้มที่มีผลการเรียนดี แต่ในทางตรงกันข้ามถ้ามีเกรดก่อนเข้าศึกษาต่ำกว่า 2.50 จะมีผลการเรียนที่ควรเฝ้าระวัง

4. อภิปรายผลและสรุป

ผลจากการเปรียบเทียบประสิทธิภาพของตัวจำแนกข้อมูลกับข้อมูลต้นฉบับพบว่า ขั้นตอนวิธี Naivebayes ให้ค่าประสิทธิภาพในการจำแนกได้ดีที่สุดในทุกมาตรวัดประสิทธิภาพ สำหรับข้อมูลที่มีการปรับจำนวนตัวอย่างในแต่ละคลาสด้วย SMOTE พบว่า ขั้นตอนวิธี Random Forest ให้ค่าประสิทธิภาพดีที่สุดในทุกมาตรวัดประสิทธิภาพเช่นกัน และเมื่อวิเคราะห์ดูในแต่ละคลาสจะเห็นได้ว่าการจำแนกข้อมูลบนคลาส L จะมีค่าประสิทธิภาพที่ดีขึ้นในทุกมาตรวัดประสิทธิภาพ แต่สำหรับคลาส H นั้น พบว่า เมื่อมีการปรับข้อมูลตัวอย่างให้มีความสมดุลกันแล้วทุกขั้นตอนวิธีมีค่าความระลึกและค่าประสิทธิภาพโดยรวมที่ลดลงกว่าการจำแนกบนข้อมูลต้นฉบับเฉลี่ยประมาณ 0. 1 ซึ่งสอดคล้องกับงานวิจัยที่มีการปรับความสมดุลของข้อมูลตัวอย่างหลายๆ งานที่คลาสที่มีการเพิ่มจำนวนข้อมูลตัวอย่างจะมีประสิทธิภาพดีขึ้นอย่างมากในขณะเดียวกันคลาสที่มีจำนวนมากส่วนใหญ่แล้วจะจำแนกได้ถูกต้องน้อยลงเล็กน้อย [9]–[11] แต่อย่างไรก็ตาม เมื่อพิจารณาในภาพรวม (ค่าเฉลี่ย) ในทุกขั้นตอนวิธีเมื่อมีการปรับข้อมูลตัวอย่างแล้วส่งผลให้จำแนกข้อมูลได้ถูกต้องมากยิ่งขึ้น ดังนั้น ขั้นตอนวิธี Random Forest จึงถูกนำมาใช้เป็นตัวแบบในการทำนายข้อมูลในงานวิจัยนี้

นอกจากนี้ผู้วิจัยจึงได้อธิบายความสัมพันธ์ของปัจจัยในรูปแบบของกฎด้วยขั้นตอนวิธีต้นไม้ตัดสินใจ ซึ่งพบว่ามีกฎบางส่วนที่แสดงความสัมพันธ์ของปัจจัยพฤติกรรมกับผลสัมฤทธิ์ทางการศึกษาโดยไม่ขึ้นกับหลักสูตรการศึกษาคือ ถ้านักศึกษามีเกรดก่อนเข้ามหาวิทยาลัยในระดับดี และอยู่ในชั้นปีสูงเช่นปี 3 หรือปี 4 แต่เล่นสมาร์ตโฟนในช่วงดึกก็จะไม่ส่งผลกระทบต่อการเรียน แต่ในทางตรงกันข้ามหากเป็นนักศึกษาปี 1 และมีพฤติกรรมการใช้สมาร์ตโฟนในช่วงดึกคือ 18.01–24.00 น. อีกทั้งมีวัตถุประสงค์ในการใช้สมาร์ตโฟนในห้องเรียนเพื่อเล่นโซเชียลมีเดียนั้น นักศึกษาเหล่านี้จะมีแนวโน้มที่มีผลการเรียนที่ไม่ดี ซึ่งสอดคล้องกับงานวิจัยของภลดา [2] ซึ่งบอกว่านักศึกษาอาจจะเป็นผู้ที่จัดสรรเวลาไม่เหมาะสมเนื่องจากพฤติกรรมการเล่นสมาร์ตโฟนนี้เองจนทำให้จัดสรรเวลาไม่ถูกต้อง โดยเฉพาะอย่างยิ่งถ้าเป็นนักศึกษาชั้นปีที่ 1 ซึ่งต้องมีการปรับตัวหลายๆ ด้าน จนทำให้ส่งผลกระทบต่อการศึกษา

อย่างไรก็ตาม ความถูกต้องของการวิเคราะห์ข้อมูลโดยใช้เทคนิคการทำเหมืองข้อมูลนั้น ขึ้นอยู่กับข้อมูลที่รวบรวมได้ ดังนั้นในอนาคตอาจจะมีการเพิ่มเติมจำนวนข้อมูลของนักศึกษาจากมหาวิทยาลัยอื่นๆ เพื่อให้ตัวแบบที่ได้มีประสิทธิภาพและใช้ได้ครอบคลุมมากยิ่งขึ้น

5. กิตติกรรมประกาศ

งานวิจัยครั้งนี้ได้รับทุนสนับสนุนการวิจัยจากสำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยี รวมถึงได้รับความอนุเคราะห์เป็นอย่างดีจากคณะศึกษาศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี และคณะเทคโนโลยีและการจัดการอุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ วิทยาเขตปราจีนบุรี ที่ให้ใช้ทรัพยากรต่างๆ ในการจัดทำงานวิจัยจนสำเร็จ

เอกสารอ้างอิง

[1] C. Kitisri, R. Nokham, and K. Phetcharat, "A smartphone using behavior and health status perception of nursing students " *Community Health Development Quarterly Khon Kaen University,* vol. 5, no. 5, pp. 19–34, 2018 (in Thai).

[2] P. Wongsafu, "Impacts of Internet using through a mobile phone of higher education students in Chiangmai province," M.S. thesis, Art in Communications, Majoe University, 2014 (in Thai).

[3] D. Lupton. (2012, August). *Digital Sociology: An Introduction.* [Online]. Available:https://ses.library.usyd.edu.au/bitstream/handle/2123/8621/Digital?sequence=2

[4] L. Breiman, "Random forests," *Machine Learning,* vol. 45, no. 1, pp. 5–32, 2001.

[5] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal Artificial Intelligence Research,* vol. 16, no. 1, pp. 321–357, 2002.

[6] J. Chareonrat, "Analysis on factors affecting normal-grade student dismissal using decision tree," *SNRU Journal of Science and Technology,* vol. 8, no. 2, pp. 256–267, 2016 (in Thai).

[7] X. Zhang, R. Xue, B. Liu, W. Lu, and Y. Zhang, "Grade prediction of student academic performance with multiple classification models," in *Proceedings 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD),* 2018. pp. 1086–1090.

[8] T. Chuangchai, "The student's learning style analysis on student's senior project by using data mining," *Journal of Graduate Studies Valaya Alongkorn Rajabhat University,* vol. 10, no. 2, pp. 53–62, 2016 (in Thai).

[9] A. Sonak, R. Patankar, and N. Pise, "A new approach for handling imbalanced dataset using ANN and genetic algorithm," in *Proceedings 2016 International Conference on Communication and Signal Processing (ICCSP),* 2016, pp. 1987–1990.

[10] H. Guoxun, H. Hui, and W. Wenyuan, "An over-sampling expert system for learing from Imbalanced data sets," in *Proceedings 2005 International Conference on Neural Networks and Brain,* 2005, pp. 537–541.

[11] A. Gosain and S. Sardana, "Handling class imbalance problem using oversampling techniques: A review," in *Proceedings 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI),* 2017, pp. 79–85.