การเปรียบเทียบประสิทธิภาพเทคนิคเหมืองข้อมูลเพื่อสร้างตัวแบบพยากรณ์ความต้องการของการใช้ผลิตภัณฑ์เครื่องสำอางที่มีส่วนผสมของว่านหางจระเข้ สำหรับวิสาหกิจชุมชนขนาดกลางและขนาดย่อมในพื้นที่จังหวัดประจวบคีรีขันธ์

ศิริเรือง พัฒน์ช่วย[1*]
มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์
ตำบลหนองแก อำเภอหัวหิน จังหวัดประจวบคีรีขันธ์ 77110
ณพรรธนนท์ ทองปาน[2]
มหาวิทยาลัยราชภัฏมหาสารคาม
80 ถนนนครสวรรค์ ตำบลตลาด อำเภอเมือง จังหวัดมหาสารคาม 44000
และ รุจิรา จุลภักดิ์[3]
มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก
122 ถนนวิภาวดีรังสิต แขวงดินแดง เขตดินแดง กรุงเทพมหานคร 10400

Received: 10 September 2022; Revised: 31 March 2023; Accepted: 19 June 2023

บทคัดย่อ

ผลิตภัณฑ์เครื่องสำอางที่มีส่วนผสมของว่านหางจระเข้มีผลเป็นอย่างมากในการตัดสินใจซื้อของผู้บริโภค ผู้วิจัยจึงสนใจสร้างโมเดลสำหรับการพยากรณ์ความต้องการใช้ผลิตภัณฑ์เครื่องสำอางที่มีส่วนผสมของว่านหางจระเข้สำหรับวิสาหกิจชุมชนขนาดกลางและขนาดย่อม วัตถุประสงค์ของงานวิจัยนี้ คือ 1.) เพื่อเลือกปัจจัยที่ส่งผลต่อความต้องการของการใช้ผลิตภัณฑ์เครื่องสำอางที่มีส่วนผสมของว่านหางจระเข้ 2.) เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการพยากรณ์ความต้องการของการใช้ผลิตภัณฑ์เครื่องสำอางที่มีส่วนผสมของว่านหางจระเข้ ด้วยวิธี NaïveBayes , ANN, J48 และ REPTree และใช้วิธีการ K-fold Cross Validation ในการแบ่งชุดข้อมูลเป็นชุดข้อมูลสำหรับการการเรียนรู้และเป็นชุดทดสอบ จากการเปรียบเทียบประสิทธิภาพพบว่าการใช้วิธี ANN ร่วมกับวิธีการเลือกคุณลักษณะด้วยวิธี InfoGainAttributeEval โดยให้ค่าความแม่นยำมากสุดคิดเป็นร้อยละ 93.54

คำสำคัญ: เหมืองข้อมูล, การจำแนกข้อมูล, เครื่องสำอางว่านหางจระเข้

* Corresponding author. E-mail: siriruang.pha@rmutr.ac.th

[1] คณะอุตสาหกรรมและเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์

[2] คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม

[3] คณะบริหารธุรกิจและเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีราชมงคลตะวันออก

A Comparison of the Efficiency of Data Mining Techniques to Create a Model for Forecasting the use of Cosmetic Products Containing Aloe Vera for Small and Medium-Sized Community Enterprises in Prachuap Khiri Khan Province

Siriruang Phatchuay[1*]
Rajamangala University of Technology Rattanakosin,
Nong Kae, Hua Hin , Prachuap Khiri Khan 77110
Naphattanon Thongpan[2]
Rajabhat Mahasarakham University
80 Nakhon Sawan Road, Talat, Mueang, Maha Sarakham Province 44000
and Rujira Jullapak[3]
Rajamangala University of Technology Tawan-ok
122 Vibhavadi Rangsit Road, Din Daeng, Din Daeng, Bangkok 10400

Received: 10 September 2022; Revised: 31 March 2023; Accepted: 19 June 2023

Abstract

The Cosmetic products containing aloe vera have a huge impact on consumers' purchasing decisions. Therefore, the researcher is interested in creating a model for forecasting the demand for cosmetic products containing aloe vera for small and medium-sized community enterprises. The objectives of this research were 1.) To select factors affecting the need for cosmetic products containing aloe vera. and 2.) To compare the efficiency of the algorithm for forecasting the need for cosmetic products containing aloe vera using NaïveBayes, ANN, J48 and REPTree methods and using the K-fold Cross Validation method. The data is a learning and testing dataset. From the comparison of efficiency, it was found that the ANN method was used together with the feature selection method. The InfoGainAttributeEval with the highest accuracy accounted for 93.54%.

Keywords: data mining, classification, aloe vera cosmetics

* Corresponding author. E-mail: siriruang.pha@rmutr.ac.th

[1] Faculty of Industry and Technology, Rajamangala University of Technology Rattanakosin

[2] Faculty of Information Technology, Rajabhat Mahasarakham University

[3] Faculty of Business Administration and Information Technology, Rajamangala University of Technology Tawan-ok

1. บทนำ

ว่านหางจระเข้ (Aloe Vera) เป็นพืชสมุนไพรธรรมชาติที่มีส่วนผสมในผลิตภัณฑ์ดูแลผิวมากที่สุดชนิดหนึ่ง มีสรรพคุณและคุณค่าในเรื่องของความงาม ยกตัวอย่างเช่น 1) บำรุงเส้นผม 2) บำรุงผิวหน้า 3) บำรุงผิวกาย 4) เพิ่มความชุ่มชื้นในผิวหน้าและผิวกาย และ 5) บรรเทาอาการปวดแสบปวดร้อนจากแผลพุพองและแผล นอกจากนี้ผู้ผลิตเครื่องสำอางจากธรรมชาติจึงได้นำเอาว่านหางจระเข้มาแปรรูปผสมผสานเข้ากับสูตรเฉพาะของตัวเองเพื่อให้ได้มาซึ่งอรรถประโยชน์รอบด้านแก่ผู้บริโภค โดยจัดจำหน่ายเครื่องสำอางจากธรรมชาติและสมุนไพร ผ่านช่องทางการจัดจำหน่าย ใน 4 กลุ่มดังนี้ [1] คือ

1) กลุ่มที่จำหน่ายโดยการตั้งเคาน์เตอร์ในห้างสรรพสินค้า โดยภาพรวมกลุ่มนี้จะเป็นกลุ่มเครื่องสำอางจากธรรมชาติและสมุนไพรที่มีราคาสูง เป็นยี่ห้อนำเข้าจากต่างประเทศและมีภาพลักษณ์ที่เป็นสากลเน้นความเชื่อถือในตัวสินค้า 2) กลุ่มที่จำหน่ายโดยการตั้งร้านของตนเองโดยเฉพาะ โดยร้านค้าเหล่านี้ส่วนใหญ่มักอยู่ในห้างสรรพสินค้า กลุ่มนี้จะเป็นเครื่องสำอางจากธรรมชาติและสมุนไพรที่มีราคารองลงมาจากกลุ่มแรกมีทั้งยี่ห้อต่างประเทศและของไทย 3) กลุ่มที่ลูกค้าเลือกซื้อเองตามซุปเปอร์มาร์เก็ต เป็นช่องทางการจำหน่ายที่เล็ก มุ่งเน้นตลาด ระดับล่างเป็นหลัก สินค้ามีราคาไม่สูงนักผู้บริโภคสามารถเลือกเองโดยการอ่านคุณสมบัติและวิธีการใช้จาก บรรจุภัณฑ์ และ 4) กลุ่มที่ใช้วิธีขายตรงโดยผ่านพนักงานขายโดยตรง

ผลิตภัณฑ์เครื่องสำอาง ที่มีส่วนประกอบจากธรรมชาติ ถือเป็นปัจจัยหลักของกลุ่มผู้บริโภค การเลือกซื้อของผู้บริโภค ส่วนใหญ่เน้นไปที่ ผลิตภัณฑ์ที่มีคุณภาพ มีการจัดโปรโมชั่น มีการลดราคาสินค้า ซึ่งทำให้ผู้บริโภคจะรู้สึกถึงความคุ้มค่ากับสินค้า อีกทั้งกระแสการรีวิวสินค้าผ่านโซเชียลมีเดีย, YouTube, บล็อกเกอร์, อินสตาแกรม ฯลฯ ส่งผลให้ผู้บริโภคเน้นการซื้อสินค้าตามผู้มีอิทธิพลในโลกออนไลน์ (Influencer) โดยปัจจุบันวิธีดังกล่าวพิสูจน์แล้วว่าได้ผลอย่างมีประสิทธิภาพสูงสุด จึงเป็นที่นิยมในการใช้เพื่อเพิ่มยอดขายสินค้า ดังนั้นผู้ผลิตเครื่องสำอางจากธรรมชาติและสมุนไพรของไทยต่างพยายามคิดค้นพัฒนา ผลิตภัณฑ์ใหม่ๆ เพื่อตอบสนองความต้องการของผู้บริโภคที่เปลี่ยนแปลงตลอดเวลา จะทำให้สามารถขยายตลาดเครื่องสำอางจากธรรมชาติและสมุนไพรให้ใหญ่ขึ้น เนื่องจากเป็นการดึงดูดลูกค้าที่ไม่สนใจให้มาบริโภคสินค้านั้นได้หรือจะส่งเสริมการจัดกิจกรรมร่วมกันเพื่อเพิ่มส่วนแบ่งการขายเมื่อถึงระดับที่ตั้งไว้

จากข้อมูลที่กล่าวมาข้างต้นผู้วิจัย พบว่าคุณลักษณะ หรือปัจจัยความต้องการของการใช้ผลิตภัณฑ์เครื่องสำอางที่มีส่วนผสมของว่านหางจระเข้ มีผลอย่างมากในการตัดสินใจเลือกใช้ผลิตภัณฑ์ หรือ ตัดสินใจซื้อผลิตภัณฑ์เครื่องสำอางที่มีส่วนผสมของว่านหางจระเข้ ผู้วิจัยจึงมีแนวคิดในการการประยุกต์ใช้เทคนิคเหมืองข้อมูลเพื่อพยากรณ์ความต้องการของการใช้ผลิตภัณฑ์ โดยวัตถุประสงค์ของงานวิจัยนี้ เพื่อเลือกปัจจัยที่ส่งผลต่อความต้องการของการใช้ผลิตภัณฑ์เครื่องสำอางที่มีส่วนผสมของว่านหางจระเข้ และเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการพยากรณ์ความต้องการของการใช้ผลิตภัณฑ์เครื่องสำอางที่มีส่วนผสมของว่านหางจระเข้ โดยใช้เทคนิคในการสร้างแบบจำลองโมเดล คือ NaïveBayes , ANN, J48 และ REPTree ผู้วิจัยได้ใช้วิธีการ K-fold Cross Validation ในการแบ่งชุดข้อมูลเป็นชุดข้อมูลสำหรับการการเรียนรู้และเป็นชุดทดสอบ และวัดประสิทธิภาพการพยากรณ์ของแบบจำลองด้วยค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall)

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 งานวิจัยที่เกี่ยวข้อง

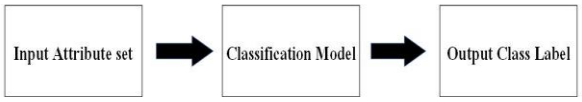
จากในอดีตที่นิยมนำว่านหางจระเข้มาใช้ในการรักษาแผลและเห็นผล จึงมีนักวิจัยจำนวนมาก ได้นำว่านหางจระเข้มาทำการทดลองเพื่อหาประสิทธิภาพของว่านหางจระเข้

ประกายดาว ศรีมุษิกโพธิ์[2] ได้ทดสอบการใช้วุ้นว่านหางจระเข้รักษาแผลฉีกขาดและโรคผิวหนัง ชนิดขี้เรื้อนเปียกในสุนัข พบว่า สามารถเร่งการหายของแผลฉีกขาดของสุนัข และทำให้การติดเชื้อแทรกซ้อนลดลง

สมฤดี สังขาวและคณะ[3] ได้นำสบู่ว่านหางจระเข้ผสมน้ำผึ้งเพื่อหาปริมาณวิตามินซี หาความเป็นกรด เป็นด่าง และทดสอบฤทธิ์การต้านอนุมูลอิสระ พบว่าสบู่ว่านหาง

จระเข้ผสมน้ำผึ้งมีฤทธิ์การต้านอนุมูลอิสระ 32.07% ปริมาณวิตามินซี 18.06% และค่าความเป็นด่าง เท่ากับ 9.23% และจากการประเมินความพึงพอใจของผู้ใช้ที่มีต่อสบู่ว่านหางจระเข้ผสมน้ำผึ้งพบว่าอยู่ในระดับมาก

2.2 การเรียนรู้ของเครื่อง (Machine Learning) คือการเรียนรู้ของเครื่องจากตัวอย่างหรือประสบการณ์ที่ได้รับจากการฝึกฝน สามารถแบ่งประเภทของการเรียนรู้ได้ 2 แบบ คือ การเรียนรู้แบบมีผู้สอน (Supervised learning) และการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) การเรียนรู้แบบมีผู้สอน



รูปที่ 1 แบบจำลองจำแนกประเภทข้อมูล

2.3 การคัดเลือกคุณลักษณะของข้อมูล

การคัดเลือกคุณลักษณะของข้อมูล จะเป็นวิธีการวิเคราะห์ปัจจัยโดยการคัดเลือกจากแอททริบิวต์ที่สามารถเป็นตัวแทนของกลุ่มแอททริบิวต์เพื่อลดจำนวนแอททริบิวต์ในการพยากรณ์[4] แบ่งออกได้เป็น 2 วิธี ได้แก่ การกรอง (Filter Approach) และการควบรวม (Wrapper Approach)

2.3.1 Info Gain Attribute Evaluation คือการแบ่งข้อมูลด้วยการคำนวณหาค่า Gain ของแต่ละลักษณะข้อมูล

เป็นการคำนวณหาค่าน้ำหนักในการเลือกคุณลักษณะที่สำคัญโดยมีวิธีการคำนวณโดยเริ่มต้นจากการคำนวณหาค่าเอนโทรปีกลุ่มทั้งหมด ดังสมการ (1)

$$H(X) = -\sum_{i=1}^{n} P(x_i)\ \log_b P(x_i) \qquad (1)$$

โดย $P(x_i)$ คือความน่าจะเป็นที่เหตุการณ์ x_i จะเกิดขึ้น โดยมีทั้งหมด n เหตุการณ์

ค่าเอนโทรปีที่ได้นำมาใช้ในการคำนวณหาค่า Information Gain(IG) เพื่อใช้สำหรับการเลือกลักษณะข้อมูลที่ดีที่สุด จากเซตของลักษณะข้อมูลทั้งหมด $A(a \epsilon A)$ ในเซตตัวอย่างทั้งหมดของ S ดังสมการที่ 2

$$IG(S,a) = H(S) - \sum_{v \epsilon Values(a)} \frac{|S_v|}{S} H(S_v) \qquad (2)$$

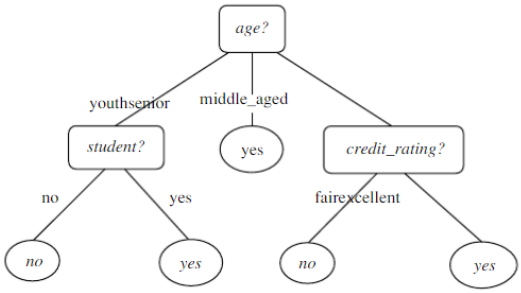
โดย V หมายถึง ค่าของลักษณะข้อมูล $Values(a)$ โดยเครื่องหมาย $|S|$ หมายถึงจำนวนสมาชิกของเซต S

2.3.2 Correlation based Feature Selection (CFS) คือ การพิจารณาหาความสัมพันธ์ของกลุ่มคุณลักษณะที่ได้รับการประเมินค่าจากความสามารถในการคาดการณ์ ซึ่งคุณลักษณะที่ไม่เกี่ยวข้องกันจะให้ประสิทธิภาพในการจำแนกที่ต่ำ[5]

2.4 วิธีการจำแนกข้อมูล

การจำแนกประเภทข้อมูล (Data Classification) เป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยเป็นการสร้างโมเดลจำแนกประเภทเพื่อทำนายกลุ่มของข้อมูลใหม่ ซึ่งคำตอบที่เป็นประเภทค่าต่าง ๆ เหล่านี้ในการวิเคราะห์ข้อมูลเรียกว่าคลาส (Class) หรือ ลาเบล (Label)[6]

2.4.1 วิธีต้นไม้ตัดสินใจ (Decision Tree) เป็นเทคนิคหนึ่งของ Classification[7,8] โดยใช้ตัวแบบทางคณิตศาสตร์ที่ใช้จำแนกประเภทข้อมูลที่มีความแม่นยำสูงแต่ประสิทธิภาพของตัวแบบขึ้นอยู่กับคุณลักษณะของข้อมูลที่นำมาใช้ ซึ่งค่าของคุณลักษณะเหล่านี้จะอยู่ในรูปของค่าที่ไม่ต่อเนื่อง (Discrete Value) เช่น เพศ สาขาวิชา เป็นต้น ตัวอย่างการนำต้นไม้ตัดสินใจไปใช้ เช่น วงการแพทย์ การวิเคราะห์ทางการเงิน การวิเคราะห์ทางการตลาด[9] และงานวิจัยนี้ได้ใช้อัลกอริทึมชนิด J48 ซึ่งพัฒนามาจาก ID3 สามารถใช้ได้กับข้อมูลแบบไม่ต่อเนื่องและแบบต่อเนื่อง ต่างจาก ID3 ที่ใช้ได้เพียงข้อมูลแบบไม่ต่อเนื่องเท่านั้น[10] และ J48 เป็นอัลกอริทึมในการสร้างต้นไม้ตัดสินใจจากกลุ่มของข้อมูลฝึกสอนโดยใช้ความถูกต้องของแต่ละคุณลักษณะของข้อมูล เพื่อใช้เป็นการตัดสินใจแบ่งกลุ่มข้อมูลกลุ่มย่อย ๆ โดยพิจารณาจากค่าความแตกต่างใน Entropy[11]ผลลัพธ์จากการเลือกคุณลักษณะสำหรับแบ่งกลุ่มข้อมูล ด้วยค่า Normalized information gain ที่สูงสุดนั้นคือการสร้างการตัดสินใจ[12]



รูปที่ 2 ตัวอย่างต้นไม้ตัดสินใจ

และต้นไม้ตัดสินใจแบบ Reduced error pruning (REP Tree) คือเทคนิคที่ใช้ Regression tree logic และสร้าง Tree หลายๆ ต้นที่แตกต่างกัน หลังจากนั้นก็เลือกที่ดีที่สุดจาก Tree ที่สร้างทั้งหมดมาเป็นตัวแทนของ Tree ทั้งหมด ในการตัดกิ่ง ใช้ค่า Mean square error ในการพยากรณ์เป็นพื้นฐานการวัด REP Tree เป็น Decision Tree ที่มีการเรียนรู้และสร้างแบบ จำลองอย่างรวดเร็วบนพื้นฐานของ Information gain หรือ Reducing the variance และตัดกิ่งโดยใช้การลดข้อผิดพลาด ในการตัดแต่ใช้ได้เฉพาะตัวแปรที่เป็นตัวเลขเท่านั้น มีนักวิจัยจำนวนมากได้นำเทคนิค REP Tree มาใช้ในการจำแนก [13]

2.4.2 โครงข่ายประสาทเทียม (Artificial Neural Network, ANN) แบบจำลองทางคณิตศาสตร์สำหรับการประมวลผลสารสนเทศ เพื่อจำลองการทำงานของโครงข่ายประสาทในสมองมนุษย์ ให้มีความสามารถในการเรียนรู้ การจดจำรูปแบบ (Pattern Recognition) และการอุปมานความรู้ (Knowledge Deduction) เช่นเดียวกับความสามารถที่มีในสมองของมนุษย์ โดยสามารถนำมาใช้ในการพยากรณ์ (Forecasting) ซึ่งเป็นการคาดคะเนลักษณะต่าง ๆ หรือเป็นการประเมินสิ่งที่จะเกิดขึ้นในอนาคตด้วยการคาดการณ์ล่วงหน้า โดยใช้ข้อมูลที่ผ่านมาในอดีต การพยากรณ์จึงมีบทบาทสำคัญในทุก ๆ ภาคส่วนที่จะช่วยให้ได้ข้อมูลในอนาคต เพื่อประกอบการวางแผนสนับสนุนการตัดสินใจในเรื่องต่างๆ และมัลติเลเยอร์เพอร์เซปตรอนหลายชั้น (Multilayer perceptron) โดยกำหนดค่าอัตราการเรียนรู้ (Learning rate) เป็น 0.1 ค่าโมเมนตัม (Momentum) เป็น 0.9 จำนวนรอบการสอน (Training time) 20,000 รอบ[14]ซึ่งในการวิจัยครั้งนี้ใช้อัลกอริทึมของวิธีโครงข่ายประสาทเทียมชนิดเพอร์เซปตรอนหลายชั้นที่มีชั้นซ่อน (Hidden layer) 1 ชั้น แม้ว่าโครงสร้างโครงข่ายประสาทเทียมที่ซับซ้อนสามารถมีชั้นซ่อนมากกว่า 1ชั้น แต่ในทางปฏิบัตินั้นการกำหนดชั้นซ่อน 1 ชั้น ก็เพียงพอต่อการวิเคราะห์ข้อมูล

2.4.3 นาอีฟเบย์ (Naïve Bayes) เป็นเทคนิคการเรียนรู้ที่อาศัยหลักการความน่าจะเป็น (Probability) ตามทฤษฎีของเบย์ (Bayes's theorem) ซึ่งมีอัลกอริทึมที่ไม่ซับซ้อน เป็นขั้นตอนวิธีในการจำแนกข้อมูล โดยการเรียนรู้ปัญหาที่เกิดขึ้น เพื่อนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่ หลักการของนาอีฟเบย์ใช้หลักการของความน่าจะเป็น โดยมีสมมติฐานว่าปริมาณของความสนใจขึ้นอยู่กับการกระจายความน่าจะเป็น (Probability Distribution)[15] เป็นเทคนิคในการแก้ปัญหาแบบจำแนกประเภทที่สามารถคาดการณ์ผลลัพธ์ได้ โดยทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ เหมาะกับกรณีของเซตตัวอย่างที่มีจำนวนมากและคุณลักษณะ (Attribute) ของตัวอย่างไม่ขึ้นต่อกัน โดยกำหนดให้ความน่าจะเป็นของข้อมูลเท่ากับสมการ

2.5 การเปรียบเทียบเพื่อหาประสิทธิภาพการจำแนกประเภท

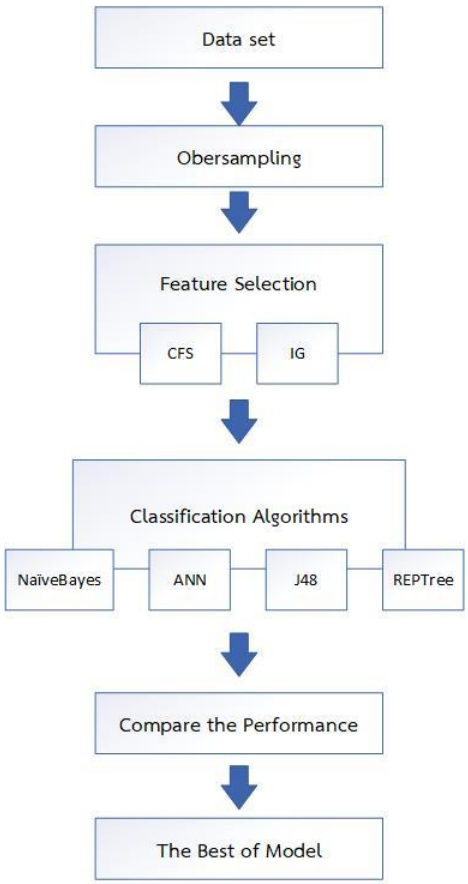
การเปรียบเทียบตัวแบบพยากรณ์และวิธีเลือกคุณลักษณะของข้อมูล ในงานวิจัยนี้ใช้การทดสอบแบบไขว้ทบ (K - fold Cross Validation) แบบ 10 ส่วนแล้วทำการทดสอบเพื่อประเมินประสิทธิภาพ[16] โดยวัดค่าความถูกต้องของการจำแนกข้อมูล (Accuracy) ดังสมการ (3)

$$Accuracy = \frac{number\ of\ matched\ labels}{number\ of\ samples} \text{x} 100\% \qquad (3)$$

3. วิธีดำเนินการวิจัย

งานวิจัยนี้ได้ใช้เทคนิคการจำแนกประเภทข้อมูล (Classification) จำนวน 4 วิธี ได้แก่ NaïveBayes, ANN, J48 และ REPTree ซึ่งทุกเทคนิคทำการเปรียบเทียบตัวแบบด้วยการทดสอบประสิทธิภาพด้วยวิธี 10 - Cross-validation Test เพื่อหาประสิทธิภาพของการจำแนกข้อมูลที่เหมาะสม แล้วได้ตัวแบบการพยากรณ์ที่มีประสิทธิภาพมากยิ่งขึ้น

ในการวิจัยครั้งนี้ผู้วิจัยได้แบ่งขั้นตอนการดำเนินการออกเป็น 6 ขั้นตอน ดังรูปที่ 3



รูปที่ 3 แสดงขั้นตอนการทำงาน

3.1 ข้อมูลที่ใช้ในการวิจัย

งานวิจัยนี้ได้ทำการเก็บรวบรวมข้อมูลการใช้ผลิตภัณฑ์เครื่องสำอางที่มีส่วนผสมของว่านหางจระเข้ ซึ่งเก็บรวบรวมออนไลน์จากทั้งหมด 6 ภูมิภาค คือ ภาคเหนือ ภาคกลาง ภาคตะวันออกเฉียงเหนือ ภาคตะวันออก ภาคตะวันตก และภาคใต้ ประกอบด้วยกลุ่มตัวอย่างจำนวน 1,034 ราย รายเอียดของชุดข้อมูลดังตารางที่ 1

ตารางที่ 1 รายละเอียดของข้อมูลที่ใช้ในการวิจัย

No.	Attribute	Description
1	Sex	เพศ
2	Age	อายุ
3	Status	สถานภาพ
4	Career	อาชีพ
5	Degree	ระดับการศึกษา
6	Income	รายได้ต่อเดือน
7	Region	ภูมิภาค
8	Product_type	ประเภทผลิตภัณฑ์ที่ซื้อ
9	Product_used	ผลิตภัณฑ์ที่เคยใช้
10	Product_buy	ผลิตภัณฑ์ที่จะซื้อซ้ำ
11	Choose_buy	ซื้อผลิตภัณฑ์จากที่ใด
12	Product_reason	เหตุผลในการเลือกซื้อ
13	Product_price	ราคาผลิตภัณฑ์ที่ซื้อ
14	Product_marketing	การส่งเสริมการขาย
15	Time_periad	ระยะเวลาซื้อครั้งต่อไป
16	Repeat_purchases	ซื้อสินค้าซ้ำอีกหรือไม่

เนื่องจากข้อมูลที่ได้เก็บรวมรวมไว้ มีความไม่สมดุลกันของข้อมูลทั้งสองกลุ่ม ซึ่งข้อมูลที่รวบรวมมาได้ แบ่งเป็น 2 กลุ่ม คือ กลุ่มที่ 1 กลุ่มกลับมาซื้อซ้ำ มีอยู่จำนวน 960 ราย กลุ่มที่ 2 กลุ่มที่ไม่กลับมาซื้อซ้ำ มีอยู่จำนวน 74 ราย ดังตารางที่ 2 ซึ่งหากไม่ได้มีการปรับความสมดุลให้กับชุดข้อมูลก่อนนำไปใช้งาน จะส่งผลการให้การเรียนรู้ของเครื่องเอียนเอียงไปในทางของกลุ่มข้อมูลที่มีปริมาณมาก ซึ่งส่งผลต่อความแม่นยำของโมเดล ดังนั้นผู้วิจัยจึงได้แก้ไขปัญหาความไม่สมดุลกันของข้อมูล[17,18] ด้วยการใช้วิธีการสุ่มเกิน (Oversampling) ซึ่งเป็นการเพิ่มจำนวนข้อมูลที่อยู่ในกลุ่มข้อมูลที่มีจำนวนน้อย ให้เพิ่มขึ้นใกล้เคียงกับจำนวนของกลุ่มข้อมูลที่มีปริมาณมาก ซึ่งหลังจากการปรับความสมดุลของข้อมูลแล้ว คลาสไม่ซื้อซ้ำจะเป็น 960 เท่ากัน

ตารางที่ 2 ปริมาณข้อมูลของแต่ละคลาส

คลาส	จำนวนเรคคอร์ด	อัตราส่วน
ซื้อซ้ำ	960	92.84%
ไม่ซื้อซ้ำ	74	7.15%
รวม	1,034	100%

ขั้นตอนการสร้างตัวแบบพยากรณ์นั้น นำข้อมูลที่ผ่านการปรับความสมดุลกันของข้อมูลเรียบร้อยแล้ว ไปผ่านกระบวนการคัดเลือกคุณลักษณะ (Feature Selection) [19,20] ด้วยวิธี CfsSubsetEval และ InfoGainAttributeEval การคัดเลือกคุณลักษณะด้วยวิธี CfsSubsetEval จะลดจำนวนคุณลักษณะข้อมูลเหลือ 6 คุณลักษณะ และวิธี InfoGainAttributeEval เหลือคุณลักษณะข้อมูล 7 คุณลักษณะ ดังตารางที่ 3

ตารางที่ 3 แสดงข้อมูลที่ผ่านการคัดเลือกคุณลักษณะ

ข้อมูลที่ผ่านการคัดเลือกคุณลักษณะ (Feature Selection) ด้วยวิธี CfsSubsetEval และ InfoGain AttributeEval	
CfsSubsetEval	ระดับการศึกษา ภูมิภาค ผลิตภัณฑ์ที่เคยใช้ ผลิตภัณฑ์ที่จะซื้อซ้ำ ซื้อผลิตภัณฑ์จากที่ใด เหตุผลในการเลือกซื้อ ประเภทผลิตภัณฑ์ที่ซื้อ
InfoGainAttributeEval	ระดับการศึกษา ภูมิภาค ผลิตภัณฑ์ที่เคยใช้ ผลิตภัณฑ์ที่จะซื้อซ้ำ ซื้อผลิตภัณฑ์จากที่ใด เหตุผลในการเลือกซื้อ

ในการทดลองเพื่อทดสอบประสิทธิภาพของอัลกอริทึมในการจำแนกประเภท ผู้วิจัยได้แบ่งชุดข้อมูลออกเป็น 2 กลุ่ม คือ กลุ่มสำหรับการเรียนรู้และกลุ่มสำหรับการทดสอบ โดยเลือกใช้วิธี K-fold Cross Validation [21] ในการแบ่งชุดข้อมูลเพื่อให้ข้อมูล สำหรับข้อมูลชุดนี้ ได้กำหนดให้ K มีค่าเท่ากับ 10 ตัวอย่าง 10 fold Cross Validation แสดงไว้ในรูปที่ 4

ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3	ครั้งที่ 4	ครั้งที่ 5
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4	4	4	4
5	5	5	5	5
6	6	6	6	6
7	7	7	7	7
8	8	8	8	8
9	9	9	9	9
10	10	10	10	10

ครั้งที่ 6	ครั้งที่ 7	ครั้งที่ 8	ครั้งที่ 9	ครั้งที่ 10
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4	4	4	4
5	5	5	5	5
6	6	6	6	6
7	7	7	7	7
8	8	8	8	8
9	9	9	9	9
10	10	10	10	10

ข้อมูลทดสอบ ข้อมูลเรียนรู้

รูปที่ 4 วิธี 10 –fold cross validation

4. ผลการทดลอง

4.1 ผลการเปรียบเทียบประสิทธิภาพการจำแนกความต้องการของการใช้ผลิตภัณฑ์เครื่องสำอางที่มีส่วนผสมของว่านหางจระเข้

การสร้างตัวแบบพยากรณ์และคัดเลือกคุณลักษณะเฉพาะที่สำคัญ ได้ใช้เทคนิคการจำแนกประเภทข้อมูล แล้วเปรียบเทียบประสิทธิภาพอัลกอริทึม โดยมีการกำหนดค่า hyper parameter ของแต่ละวิธีดังนี้ วิธี NaïveBayes กำหนดค่า Hyperparameter ให้กับ Attribute Repeat_purchases ด้วยวิธีการ Bernoulli distribution ในการทำ Estimate probability เนื่องจากมีค่าความเป็นไปได้ของผลลัพธ์แค่ 2 อย่าง คือ ซื้อซ้ำกับไม่ซื้อซ้ำ ส่วน Attribute Degree ใช้วิธีการ Categorical Distribution เนื่องจากแบ่งข้อมูลออกเป็นช่วงและ Attribute อื่นๆ ที่เหลือใช้วิธีการ Multinomial distribution เนื่องจากข้อมูล

สามารถมีคำตอบได้มากกว่า 1 ค่า ,สำหรับวิธี J48 ที่ผ่านการคัดเลือกคุณสมบัติด้วยวิธี CfsSubsetEval จะกำหนดค่า Max-depth =7 เนื่องจากมี Attribute เหลืออยู่ 7 Attribute ส่วน J48 ที่ผ่านการคัดเลือกคุณสมบัติด้วยวิธี InfoGainAttributeEval จะกำหนดให้ Max-depth =6 , วิธี ANN ได้กำหนด Input layer และ Hidden layer ให้มีค่าเท่ากับจำนวนของ Attribute ดังนั้นข้อมูลที่ผ่านการคัดเลือกคุณสมบัติด้วยวิธี CfsSubsetEval จะกำหนดค่าให้กับ Input layer และ Hidden layer เท่ากับ 7 ส่วน Output layer เท่ากับ 2 ข้อมูลที่ใช้วิธีการ InfoGainAttributeEval ในการคัดเลือกคุณสมบัติได้กำหนดให้กับ Input layer และ Hidden layer เท่ากับ 6 ส่วน Output layer เท่ากับ 2 และใช้วิธี REPTree จากนั้นทำการทดสอบหาประสิทธิภาพในการจำแนกข้อมูลด้วยเทคนิค Cross-validation Test ซึ่งผลลัพธ์แสดงในตารางที่ 4

ตารางที่ 4 ผลการเปรียบเทียบค่าความถูกต้องของข้อมูลเมื่อถูกจำแนกประเภทด้วยอัลกอรึทึม ทั้ง 4 วิธี

Model	10 – fold Cross Validation Test
	Average Accuracy
NaïveBayes	81.33
ANN	86.30
J48	83.12
REPTree	83.81

จากตารางที่ 4 แสดงค่าความถูกต้องของข้อมูลเมื่อถูกจำแนกด้วย 4 วิธี ก่อนการนำข้อมูลไปคัดเลือกคุณสมบัติด้วยวิธี CfsSubsetEval และวิธี InfoGainAttributeEval ผลการเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูพบว่าเทคนิค ANN มีประสิทธิภาพในการจำแนกสูงสุด โดยมีค่าความถูกต้องร้อยละ 86.30 และรองลงวิธี REPTree มีค่าความถูกต้องที่ร้อยละ 83.81 และวิธี J48 และ NaïveBayes ร้อยละ 83.12 และ 81.33 ตามลำดับ

4.2 ผลการเปรียบเทียบการคัดเลือกคุณลักษณะของข้อมูลระหว่างวิธี CfsSubsetEval และ InfoGain AttributeEval

สำหรับข้อมูลการใช้ผลิตภัณฑ์เครื่องสำอางที่มีส่วนผสมของว่านหางจระเข้ ซึ่งเก็บรวบรวมกลุ่มตัวอย่างไว้จำนวน 1,034 ราย แต่เนื่องจากข้อมูลมีจำนวนทั้งสิ้น 16 attribute ส่งผลต่อประสิทธิภาพและความเร็วในการจำแนกข้อมูล ทางผู้วิจัยจึงหาวิธีการในการคัดเลือกคุณสมบัติที่สำคัญด้วยวิธี วิธี CfsSubsetEval และ InfoGain AttributeEval เพื่อให้ได้ Attribute ที่มีความสำคัญและลดจำนวน Attribute ลง แต่ยังคงมีประสิทธิภาพในการจำแนกไว้ ดังนั้นเมื่อนำข้อมูลมาคัดเลือกคุณลักษณะด้วยวิธี CfsSubsetEval พบว่าจำนวน Attribute ที่มีผลต่อประสิทธิภาพของการจำแนกมีอยู่จำนวน 7 คุณลักษณะ ได้แก่ ระดับการศึกษา ภูมิภาค ผลิตภัณฑ์ที่เคยใช้ ผลิตภัณฑ์ที่จะซื้อซ้ำ ซื้อผลิตภัณฑ์จากที่ใด เหตุผลในการเลือกซื้อประเภทผลิตภัณฑ์ที่ซื้อ และเมื่อใช้วิธี InfoGainAttributeEval พบว่าจำนวน Attribute ที่มีผลต่อประสิทธิภาพของการจำแนก มีอยู่ 6 คุณลักษณะ ได้แก่ ระดับการศึกษา ภูมิภาค ผลิตภัณฑ์ที่เคยใช้ ผลิตภัณฑ์ที่จะซื้อซ้ำ ซื้อผลิตภัณฑ์จากที่ใด เหตุผลในการเลือกซื้อ

จากตารางที่ 5 เมื่อพิจารณาผลการประสิทธิภาพของการจำแนกการกลับมาซื้อผลิตภัณฑ์ซ้ำ ของกลุ่มตัวอย่างจาก 6 ภูมิภาค ด้วยการคัดเลือกคุณลักษณะด้วยวิธี CfsSubsetEval และวิธี InfoGainAttributeEval แล้วนำคุณลักษณะที่ได้จากแต่ละวิธีมาจำแนกการกลับมาซื้อซ้ำด้วยอัลกอรึทึมทั้ง 4 คือ วิธี NaïveBayes พบว่า

ตารางที่ 5 การวิเคราะห์การเปรียบเทียบการคัดเลือกคุณลักษณะ

Feature Selection	Attributes	Models	Average Accuracy
CfsSubsetEval	7	NaïveBayes	81.93
		ANN	80.32
		J48	81.93
		REPTree	82.58
InfoGain Attribute Eval	6	NaïveBayes	84.51
		ANN	93.54
		J48	81.93
		REPTree	84.51

การคัดเลือกคุณลักษณะด้วยวิธีการ CfsSubsetEval จะได้ Attribute ที่มีความสำคัญกับการจำแนกจำนวน 7 Attribute ซึ่งเมื่อนำมาจำแนกข้อมูลการกลับมาซื้อผลิตภัณฑ์ซ้ำ พบว่าค่าความถูกต้องของเทคนิค REPTree มีค่าความถูกต้องสูงสุดอยู่ที่ร้อยละ 82.58 รองลงมาคือ J48 และ NaïveBayes มีค่าความถูกต้องร้อยละ 81.93 เท่ากัน และวิธี ANN มีค่าความแม่นยำต่ำสุดร้อยละ 80.32

ส่วนการเลือกคุณลักษณะที่สำคัญด้วยวิธี InfoGainAttributeEval จะมี Attribute ที่มีความสำคัญต่อการจำแนกอยู่จำนวน 6 Attribute เมื่อนำมาจำแนกด้วยเทคนิคข้างต้น พบว่าเทคนิค ANN มีค่าความถูกต้องมากที่สุดร้อยละ 93.54 มีประสิทธิภาพดีที่สุด รองลงมาเป็นเทคนิค NaïveBayes และเทคนิค REPTree โดยมีค่าความถูกต้องร้อยละ 84.51 เท่ากันและเทคนิค J48 มีค่าความถูกต้องต่ำสุดร้อยละ 81.93

ดังนั้นการเลือกใช้วิธี InfoGainAttributeEval เพื่อการคัดเลือกคุณสมบัติที่สำคัญนั้น เมื่อเปรียบเทียบแล้วพบว่ามีประสิทธิภาพมากกว่าการเลือกใช้วิธีการ CfsSubsetEval กับทุกเทคนิคที่กล่าวมา ซึ่ง Attribute ที่แตกต่างกันและส่งผลต่อประสิทธิภาพของการจำแนกหรืออาจส่งผลให้เกิดปัญหา Over fit นั่นคือ Attribute ประเภทผลิตภัณฑ์ที่ซื้อ นั่นเอง

5. สรุปผลงานวิจัยและข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเลือกปัจจัยที่ส่งผลต่อความต้องการของการใช้ผลิตภัณฑ์เครื่องสำอางที่มีส่วนผสมของว่านหางจระเข้ และเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการพยากรณ์ความต้องการของการใช้ผลิตภัณฑ์เครื่องสำอางที่มีส่วนผสมของว่านหางจระเข้ ด้วยวิธี NaïveBayes , ANN, J48 และ REPTree และใช้วิธีการ K-fold Cross พบว่าเทคนิค ANN ใช้ร่วมกับการเลือกคุณลักษณะที่สำคัญด้วยวิธี InfoGainAttributeEval มีค่าความถูกต้องมากที่สุดร้อยละ 93.54 มีประสิทธิภาพดีที่สุด

และจากผลการเปรียบเทียบประสิทธิภาพในครั้งนี้จะเห็นได้ว่าค่าความถูกต้องที่ได้ร้อยละ 93.54 นั้นถือได้ว่ามีประสิทธิภาพสูงและเป็นประโยชน์แก่กลุ่มวิสาหกิจชุมชนเป็นอย่างมาก สามารถนำไปพัฒนาระบบพยากรณ์ความต้องการการใช้ผลิตภัณฑ์เครื่องสำอางที่มีส่วนผสมของว่านหางจระเข้ของกลุ่มวิสาหกิจชุมชน และเป็นแนวทางในการประชาสัมพันธ์การตลาดเฉพาะกลุ่ม เพื่อแนะนำการซื้อสินค้าผลิตภัณฑ์เครื่องสำอางที่มีส่วนผสมของว่านหางจระเข้ให้กลุ่มผู้บริโภคที่สนใจได้เข้าถึงสินค้าได้อย่างทั่วถึง

ข้อเสนอแนะในการวิจัยครั้งนี้ เพื่อให้การพยากรณ์ได้ผลการวิเคราะห์ข้อมูลที่มีความครอบคลุมมากขึ้น อาจใช้วิธีการจำแนกโดยใช้อัลกอริทึมประเภทอื่น ๆ เช่น วิธีต้นไม้ตัดสินใจอัลกอรีทึมแบบ Random forest และ Random tree และ วิธีซัพพอร์ตเวกเตอร์ แมชชีนใช้อัลกอริทึม SMO เป็นต้น

6. กิตติกรรมประกาศ

ขอขอบคุณมหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์ ที่สนับสนุนทุนวิจัยสัญญาเลขที่ A21/2563 ตลอดจนเครื่องมือ และสถานที่ ที่ทำให้งานวิจัยนี้สำเร็จลุล่วงด้วยดี

7. เอกสารอ้างอิง

[1] สินีนาถ เลิศไพรวัน สุมิตรา ศรีวิบูลย์ และ จันทร์จรัส ศรีสิริ, “รายงานวิจัยโครงการศึกษาและพัฒนาบรรจุภัณฑ์เครื่องสำอางสมุนไพรเพื่อการส่งออก,” กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ, 2552.

[2] ประกายดาว ศรีมุษิกโพธิ์, “การใช้วุ้นว่านหางจระเข้รักษาแผลฉีกขาดและโรคผิวหนังชนิดขี้เรื้อนเปียกในสุนัข,” *วารสารเกษตร.*, ปีที่ 22, ฉบับที่ 3, น. 245–252, 2549.

[3] สมฤดี สังขาว, อทิตยา ขวัญวงส์, นงนุช ขอนทอง, เกศินี ใจดี, ณัฐฐินันท์ ชำรวย, อาภรณ์ พาชัย, มณฑา หมีไพรพฤกษ์ และ ณัฐภานี บัวดี, “ฤทธิ์การต้านอนุมูลอิสระปริมาณวิตามินซีและความพึงพอใจของสบู่ว่านหางจระเข้ผสมน้ำผึ้ง: สมุนไพรพญาไพร อำเภอเมือง จังหวัดกำแพงเพชร,” *วารสารวิทยาศาสตร์และเทคโนโลยี(สทวท.).*, ปีที่ 4, ฉบับที่ 1, น. 119-126, 2560.

[4] M. A. Hall and G. Holmes, “Benchmarking attribute selection techniques for discrete class data mining,” *IEEE Transactions on Knowledge and Data engineering.*, vol. 15, no. 6, pp. 1437-1447, 2003.

[5] เสกสรรค์ วิลัยลักษณ์, วิภา เจริญภัณฑารักษ์ และ ดวงดาว วิชาดากุล, “การใช้เทคนิคการทำเหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา,” *วารสาร Veridian E Journal ฯ สาขาวิทยาศาสตร์และเทคโนโลยี.*, ปีที่ 2, ฉบับที่ 2, น. 1-17, 2558.

[6] U. aiza and N. Ussiph, “Appraisal of the classification technique in data mining of student performance using J48 decision tree, K-Nearest neighbor and multilayer perceptron algorithms,” *International Journal of Computer Applications.*, vol. 179, no. 33, pp. 39-46, 2018.

[7] ณัฐวดี หงษ์บุญมี และ พงศ์นรินทร์ ศรรุ่ง, “การประยุกต์ใช้เทคนิคจำแนกข้อมูลแบบต้นไม้ตัดสินใจเพื่อการวินิจฉัยโรคในโคเบื้องต้นบนโทรศัพท์มือถือ,” *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี.*, ปีที่ 20, ฉบับที่ 1, น. 44-58, 2561.

[8] D. Verma and N. Mishra, “Analysis and Prediction of Breast Cancer And Diabetes Disease Datasets Using Data Mining Classification Techniques,” in *International Conference on Intelligent Sustainable Systems (ICISS)*, India, 2017.

[9] J. Han and M. Kamber, Data mining concepts and techniques, (2nd ed.), United States of America: Morgan Kaufman Publishers, 2006.

[10] รัชพล กลัดชื่น และ จรัญ แสนราช, “การเปรียบเทียบประสิทธิภาพอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อการทำนายผลสัมฤทธิ์ทางการเรียนของนักศึกษาระดับอาชีวศึกษา,” *วารสารวิจัยมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี.*, ปีที่ 17, ฉบับที่ 1, น. 1-10, 2561.

[11] ภรัณยา ปาลวิสุทธ์, “การเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุลโดยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อยสำหรับข้อมูลการเป็นโรคติดอินเทอร์เน็ต,” *วารสารเทคโนโลยีสารสนเทศ.*, ปีที่ 12, ฉบับที่ 1, น. 54-63, 2559.

[12] J. Han, M. Kamber and J. Pei, Data mining concepts and techniques, 3rd ed., United States of America: Morgan Kaufman Publishers, 2011.

[13] W. Nor, H. W. Mohamed, M. S. Najib, M. N. Salleh and A. Halim Omar, “A Comparative Study of Reduced Error Pruning Method in Decision Tree Algorithms,” in *IEEE International Conference on Control System, Computing and Engineering*, Malaysia, 2012.

[14] สายชล สินสมบูรณ์ทอง, “การเปรียบเทียบประสิทธิภาพในการทำนายผลความไม่สมดุลของ

ข้อมูลในการจำแนกด้วยเทคนิคการทำเหมืองข้อมูล," *วารสารวิทยาศาสตร์และเทคโนโลยี.,* ปีที่ 28, ฉบับที่ 3, น. 383-393, 2563.

[15] A. Navada, A. Ansari, S. Patil, S and B. A. Sonkamble, "Overview of Use of Decision Tree Algorithms In Machine Learning," in *IEEE Control and System Graduate Research Colloquium*, Malaysia, 2011.

[16] Z. Wang, J. Liu, G. Sun, J. Zhao, Z. Ding and X. Guan, "An Ensemble Classification Algorithm for Text Data Stream Based on Feature Selection and Topic Model," in *IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, China, 2020.

[17] 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.

[18] วิชญ์วิสิธ เกษรสิทธิ์, วิชิต หล่อจีระชุณห์กุล และ จิราวัลย์ จิตรถเวช, "การแก้ปัญหาข้อมูลไม่สมดุลของข้อมูลสำหรับการจำแนกผู้ป่วยโรคเบาหวาน," *KKU Research Journal (Graduate Studies)*., ปีที่ 18, ฉบับที่ 3, น. 11-21, 2560.

[19] Y. Dong, Q. Yue and M. Feng, "An efficient feature selection method for network video traffic classification," in *IEEE 17th International Conference on Communication Technology (ICCT)*, China, 2017.

[20] J. Bi, K. Zhang and X. Cheng, "Intrusion Detection Base on RBF Neural Network," in *International Symposium on Information Engineering and Electronic Commerce*, Ukraine, 2009.

[21] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence.,* Canada, 1995.