ตัวแบบการจำแนกประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์ตามคุณลักษณะข้อมูลพื้นฐานของการประกอบการธุรกิจเพื่อเพิ่มประสิทธิภาพสำหรับการพยากรณ์ประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์ในประเทศไทย

A CLASSIFICATION MODEL FOR E-COMMERCE BUSINESS-BASED ON BASIC DATA ATTRIBUTES OF BUSINESS ENTREPRENEURSHIP TO OPTIMIZE THE PREDICTION OF E-COMMERCE BUSINESS TYPE IN THAILAND

อุณนดาทร มูลเพ็ญ[1], สุนันวดี พละศักดิ์[2*]

Unnadathorn Moonpen[1], Sunanvadee Palasak[2*]

[1, 2] อาจารย์ ดร. คณะบริหารธุรกิจและการบัญชี มหาวิทยาลัยราชภัฏร้อยเอ็ด จังหวัดร้อยเอ็ด 45120 ประเทศไทย*

[1,2]Lecturer Dr., Faculty of Business Administration and Accountancy, Roi Et Rajabhat University, Roi Et Province, 45120, Thailand*

E-mail address(Corresponding author): [2*]Sunanvade.baac@reru.ac.th; (Author):[1]Nada_Moon2523@hotmail.com

---

รับบทความ : 11 มิถุนายน 2566 / ปรับแก้ไข : 13 กันยายน 2566 / ตอบรับบทความ : 21 กันยายน 2566

Received : 11 June 2023 / Revised : 13 September 2023 / Accepted : 21 September 2023

DOI : 10.14456/nrru-rdi.2023.44

ABSTRACT

The preparation for market change in e-commerce businesses to enable them to compete effectively and move towards a better direction has led to this research. The research objectives were: 1) to present the best classification model of e-commerce business-based on basic data attributes operations; 2) to analyze attributes that affect the optimization suitable for predictive e-commerce businesses; 3) to compare the effectiveness of the classification model with the feature analysis model using t-test and ANOVA statistics. The sample was 400 corporations engaged in e-commerce businesses in Thailand, selected using a content-valid questionnaire with a reliability index of 0.91, and questionnaire reliability scores from 0.72 to 0.98. Data was collected via postal mail, and data analysis employed data mining as a forecasting. The findings revealed that: 1) The support vector machine model is the best for predictive e-commerce business type in Thailand with an accuracy of 80.97% 2) The best model for the optimization of the classification is the decision tree model with an accuracy of 80.04% and 3) The features that influenced the effectiveness of prediction included total business assets, business operation period, total current employees, average annual business income, and current job positions Moreover, 4) the decision tree model and the support vector machine model demonstrated no significant difference in effectiveness and can be used interchangeably. The research indicates that the models are capable of accurate and precise predictions, and users can correctly rank the importance of data attributes, which is essential for the success of e-commerce business types.

Keywords : Classification model, E-commerce business, Optimize, Predictive

บทคัดย่อ

การเตรียมความพร้อมต่อการเปลี่ยนแปลงทางการตลาดของธุรกิจให้สามารถก้าวข้ามการแข่งขันและเป็นไปในทิศทางที่ดีขึ้น นำมาสู่งานวิจัยครั้งนี้ ซึ่งมีวัตถุประสงค์การวิจัยเพื่อ 1) เสนอตัวแบบการจำแนกประเภทที่ดีที่สุดของธุรกิจพาณิชย์อิเล็กทรอนิกส์ตามคุณลักษณะข้อมูลพื้นฐานของการประกอบการธุรกิจพาณิชย์อิเล็กทรอนิกส์ 2) วิเคราะห์คุณลักษณะที่ส่งผลต่อการเพิ่มประสิทธิภาพที่เหมาะสมสำหรับการพยากรณ์ประเภทธุรกิจพาณิชย์อิเล็กทรอนิกส์ และ 3) เปรียบเทียบประสิทธิภาพความแตกต่างระหว่างตัวแบบการจำแนกประเภทกับตัวแบบวิเคราะห์คุณลักษณะที่ส่งผลต่อการเพิ่มประสิทธิภาพของตัวแบบด้วยสถิติที-เทส และสถิติอโนวา กลุ่มตัวอย่างคือนิติบุคคลที่ประกอบธุรกิจพาณิชย์อิเล็กทรอนิกส์ในประเทศไทย จำนวน 400 ราย ดำเนินการเก็บรวบรวมข้อมูลโดยใช้แบบสอบถาม ที่มีค่าความเที่ยงตรงเชิงเนื้อหาเท่ากับ 0.91 และค่าความเชื่อมั่นของแบบสอบถาม อยู่ระหว่าง 0.72-0.98 ด้วยการส่งทางไปรษณีย์ และใช้เทคนิคการจำแนกประเภท ในการวิเคราะห์ข้อมูลสำหรับการพยากรณ์ที่เป็นวิธีของการทำ

เหมืองข้อมูล ผลวิจัยพบว่า 1) ตัวแบบซัพพอร์ตเวกเตอร์แมชชีนดีที่สุดสำหรับการพยากรณ์ประเภทธุรกิจพาณิชย์อิเล็กทรอนิกส์ในประเทศไทย มีค่าความถูกต้อง 80.97% 2) ตัวแบบที่ดีที่สุดสำหรับการเพิ่มประสิทธิภาพการจำแนกประเภท คือ ตัวแบบต้นไม้การตัดสินใจ มีค่าความถูกต้อง 80.04% คุณลักษณะที่ส่งผลต่อการเพิ่มประสิทธิภาพการพยากรณ์มี 5 คุณลักษณะ คือ สินทรัพย์รวมของธุรกิจ (Asset), ระยะเวลาดำเนินธุรกิจ (Age), จำนวนพนักงานในปัจจุบันทั้งหมด (N_Em), รายได้ของธุรกิจเฉลี่ยต่อปี (IncomperY) และตำแหน่งงานปัจจุบัน (Position) และ 3) ตัวแบบต้นไม้การตัดสินใจและตัวแบบซัพพอร์ตเวกเตอร์แมชชีนมีประสิทธิภาพไม่แตกต่างกันสามารถใช้แทนกันได้ ผลวิจัยชี้ให้เห็นว่าตัวแบบสามารถพยากรณ์ได้อย่างถูกต้องและแม่นยำ และผู้ใช้สามารถจัดลำดับความสำคัญของคุณลักษณะข้อมูลได้อย่างถูกต้องเหมาะสมส่งผลต่อความสำเร็จของประเภทธุรกิจพาณิชย์อิเล็กทรอนิกส์
คำสำคัญ : ตัวแบบการจำแนกประเภท, ธุรกิจพาณิชย์อิเล็กทรอนิกส์, การเพิ่มประสิทธิภาพ, การพยากรณ์

บทนำ

จากการที่ผู้ประกอบการขยายช่องทางการขายสินค้าและบริการผ่านช่องทางออนไลน์มากขึ้น โดยพบว่าในปี 2565 ผู้ประกอบการธุรกิจพาณิชย์อิเล็กทรอนิกส์ (E-commerce) ประเทศไทยมีสถิติการจดทะเบียนธุรกิจมีประมาณ 1,214 ราย เพิ่มจากช่วงเวลาเดียวกันของปี 2564 คิดเป็นร้อยละ 12.51 และมีมูลค่าการจดทะเบียนจัดตั้งธุรกิจใหม่เพิ่มขึ้นคิดเป็นร้อยละ 13.83 ทั้งนี้ธุรกิจพาณิชย์อิเล็กทรอนิกส์ยังมีรายได้เพิ่มสูงขึ้นอย่างต่อเนื่อง (Electronic Transactions Development Agency, 2022) โดยเฉพาะหลังจากมีการระบาดของเชื้อ Covid-19 พบว่า ประชาชนจำนวนมากมีการเปลี่ยนพฤติกรรมการซื้อสินค้าผ่านช่องทางออนไลน์ แทนการซื้อสินค้าที่ร้านค้าและห้างสรรพสินค้า ส่งผลให้มูลค่าของธุรกิจพาณิชย์อิเล็กทรอนิกส์มีแนวโน้มเติบโตแบบก้าวกระโดด และมีโอกาสที่การเปลี่ยนแปลงพฤติกรรมดังกล่าวอาจเป็นการเปลี่ยนแปลงระยะยาว ระดับของการแพร่กระจายของธุรกิจพาณิชย์อิเล็กทรอนิกส์ประเภทธุรกิจกับธุรกิจจะสูงขึ้นในประเทศที่มีกรอบสถาบันระดับชาติสนับสนุน (Alsaad et al., 2021) ปัจจุบันภาครัฐและผู้ประกอบธุรกิจพาณิชย์อิเล็กทรอนิกส์ไม่มีตัวแบบช่วยในการวิเคราะห์การเปลี่ยนแปลงพฤติกรรมของผู้บริโภคที่ส่งผลต่อการตัดสินใจเลือกจดทะเบียนประเภทธุรกิจพาณิชย์อิเล็กทรอนิกส์ จึงทำให้เกิดการประเมินภาวะตลาดผิดพลาดหรือเกิดผลข้างเคียงที่ไม่พึงประสงค์ได้ อีกทั้งในปัจจุบันร้านค้าส่วนใหญ่นิยมขายสินค้าผ่านทั้งช่องทางธุรกิจพาณิชย์อิเล็กทรอนิกส์ในหลากหลายรูปแบบ (Wiwattananukul, 2021) ได้แก่ พาณิชย์อิเล็กทรอนิกส์แบบธุรกิจกับธุรกิจ (B2B) พาณิชย์อิเล็กทรอนิกส์แบบธุรกิจกับผู้บริโภค (B2C) พาณิชย์อิเล็กทรอนิกส์แบบผู้บริโภคกับผู้บริโภค (C2C) และพาณิชย์อิเล็กทรอนิกส์แบบธุรกิจผู้กับภาครัฐ(B2G) เป็นต้น จากการเติบโตของธุรกิจพาณิชย์อิเล็กทรอนิกส์ที่รวดเร็วและพฤติกรรมของผู้บริโภคมีการเปลี่ยนแปลงรวดเร็วนั้น ภาครัฐและผู้ประกอบธุรกิจพาณิชย์อิเล็กทรอนิกส์จำเป็นจะต้องมีความรู้ความเข้าใจผู้บริโภคให้มากที่สุด และจำเป็นจะต้องนำเทคโนโลยีเข้ามาช่วยในการคาดการณ์ความต้องการในอนาคตให้แม่นยำ การพัฒนาของประเทศที่พัฒนาแล้วและกำลังพัฒนาพาณิชย์อิเล็กทรอนิกส์ได้เปลี่ยนวิธีการแบ่งปันทางธุรกิจข้อมูลทั่วโลก (Xuhua et al., 2019) ในขณะเดียวกันการสร้างตราสินค้าด้วยการตระหนักรู้และการเสริมความแข็งแกร่งด้านการบริการเทคโนโลยีเป็นข้อได้เปรียบและโอกาสสำหรับการพัฒนาพาณิชย์อิเล็กทรอนิกส์ (Wang, Jing, & Lyu, 2022) จากการทบทวนวรรณกรรมทั้งในและต่างประเทศพบว่า การพัฒนาพาณิชย์อิเล็กทรอนิกส์ในชนบทของประเทศจีนมีความสำคัญอย่างยิ่งในการส่งเสริมการเปลี่ยนแปลงของอุตสาหกรรมการเกษตรสามารถกระตุ้นศักยภาพของเศรษฐกิจในชนบทและเสริมรายได้ให้แก่เกษตรกรนำไปสู่การเปลี่ยนแปลงที่สำคัญมาสู่ประเทศจีน (Guo et al., 2021) การพัฒนาพาณิชย์อิเล็กทรอนิกส์ในชนบทต่อการเกษตรนั้น

ปรากฏอยู่ในการจัดหาช่องทางการตลาดใหม่สำหรับสินค้าเกษตร เป็นต้น (Huang et al., 2022) การสร้างตัวแบบการพยากรณ์โดยใช้เทคโนโลยีสำหรับการวิเคราะห์ข้อมูลเชิงลึกเพื่อคาดการณ์แนวโน้มที่จะเกิดขึ้นในอนา คต จึงมีความสำคัญโดยเฉพาะการจำแนกประเภทร้านค้าออนไลน์ตามระดับของกิจกรรมธุรกิจจากข้อมูลที่มีอยู่จำนวนมาก มาทำการวิเคราะห์เชิงลึกและการปรับปรุงให้เกิดความเหมาะสมอย่างต่อเนื่อง เพื่อเสนอตัวแบบที่มีความถูกต้องแม่นยำสำหรับการพยากรณ์ (Wakil et al., 2020)

ดังนั้น การนำเสนอตัวแบบการจำแนกประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์ตามคุณลักษณะข้อมูลพื้นฐานของการประกอบการธุรกิจเพื่อเพิ่มประสิทธิภาพสำหรับการพยากรณ์ประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์ในประเทศไทย เป็นการพัฒนาและวิเคราะห์คุณลักษณะที่ส่งผลต่อการเพิ่มประสิทธิภาพของตัวแบบการจำแนกประเภทที่เหมาะสมสำหรับการนำไปพยากรณ์ประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์

วัตถุประสงค์การวิจัย

1. เพื่อเสนอตัวแบบการจำแนกประเภทที่ดีที่สุดของธุรกิจพาณิชย์อิเล็กทรอนิกส์ตามคุณลักษณะข้อมูลพื้นฐานของการประกอบการธุรกิจพาณิชย์อิเล็กทรอนิกส์
2. เพื่อวิเคราะห์คุณลักษณะที่ส่งผลต่อการเพิ่มประสิทธิภาพของตัวแบบการจำแนกประเภทที่เหมาะสมสำหรับการนำไปพยากรณ์ประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์
3. เพื่อเปรียบเทียบประสิทธิภาพความแตกต่างระหว่างตัวแบบการจำแนกประเภทกับตัวแบบวิเคราะห์คุณลักษณะที่ส่งผลต่อการเพิ่มประสิทธิภาพของตัวแบบด้วยสถิติที-เทส และสถิติอโนวา

ประโยชน์การวิจัย

1. เสนอตัวแบบการจำแนกประเภทที่ดีที่สุดของธุรกิจพาณิชย์อิเล็กทรอนิกส์ตามคุณลักษณะข้อมูลพื้นฐานของการประกอบการธุรกิจพาณิชย์อิเล็กทรอนิกส์
2. มีตัวแบบการจำแนกประเภทที่สามารถเพิ่มประสิทธิภาพการพยากรณ์ประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์ให้มีความถูกต้องและแม่นยำได้
3. ผู้ให้บริการมีตัวช่วยสนับสนุนการตัดสินใจการแบ่งประเภทของธุรกิจที่สอดคล้องกับการดำเนินงานของธุรกิจได้อย่างถูกต้องและแม่นยำ

การทบทวนวรรณกรรม

การทำเหมืองข้อมูล (Data mining)

การทำเหมืองข้อมูล คือ วิธีการค้นหารูปแบบใหม่ในเชิงลึกจากฐานข้อมูลแบบอัตโนมัติให้สามารถพยากรณ์ข้อมูลในอนาคตที่มีความถูกต้องน่าสนใจและใช้ให้เกิดประโยชน์ การทำเหมืองข้อมูลส่วนใหญ่เกี่ยวข้องกับการสร้างแบบจำลองหรือตัวแบบที่เหมาะสมกับสถานการณ์สามารถส่งผลลัพธ์ที่น่าสนใจที่จะเกิดขึ้นคือ การค้นพบความรู้ในฐานข้อมูล (Knowledge discovery in databases : KDD) ที่ซ่อนอยู่ในข้อมูลในอดีตมาคาดการณ์แนวโน้มที่จะเกิดขึ้นในอนาคต (Theeramunkong, 2017) การทำเหมืองข้อมูลมีหลายวิธีด้วยกันที่เป็นที่นิยม ได้แก่ การจำแนกประเภท การจัดกลุ่ม กฎความสัมพันธ์ เป็นต้น โดยงานวิจัยนี้ใช้หนึ่งในกระบวนการทำเหมืองข้อมูลที่ได้รับความนิยม คือ การจำแนกประเภท ซึ่งเป็นเทคนิคที่ใช้การเรียนรู้จากข้อมูลที่มีผลเฉลยหรือมีการจัดประเภทของข้อมูลใส่ไว้ในป้ายวัตถุ (Moonpen, Mungsing, & Banditwattanawong, 2020)

การจำแนกประเภท (Classification)

การจำแนกประเภท คือ การค้นหาวัตถุของคลาสเกี่ยวข้องกับการตัดสินใจของมนุษย์ในกิจกรรมการเรียนรู้ที่อาศัยผลเฉลยจากข้อมูลในอดีตมาสร้างการพยากรณ์แนวโน้มในอนาคต โดยงานวิจัยนี้จะทำการทดลองกับขั้นตอนวิธีการจำแนกประเภทที่ใช้กันอย่างแพร่หลายในปัจจุบัน (Mitchell, 1997) จำนวน 4 ตัวแบบ ดังต่อไปนี้

นาอีฟเบย์ (Naïve baysain) คือ หลักความน่าจะเป็นจากสมมติฐาน โดยที่ตัวแบบจะทำการเรียนรู้เพิ่มด้วยการคำนวณสมมติฐานเพื่อพยากรณ์คลาสให้ถูกต้องตามความน่าจะเป็นมากที่สุดของทุกสมมติฐาน ดังสมการ (1) แสดงถึงขั้นตอนการคำนวณที่อาศัยหลักความน่าจะเป็นมาสร้างตัวแบบการจำแนกประเภท ซึ่งนาอีฟเบย์เป็นเทคนิคการทำเหมืองข้อมูล ประเภทเทคนิคการจำแนกประเภทที่อาศัยการเรียนรู้จากข้อมูลในอดีตที่มีผลเฉลย และส่งผลต่อการวัดประสิทธิภาพความถูกต้อง ความแม่นยำ ประสิทธิภาพที่ได้สามารถนำไปใช้เปรียบเทียบกับตัวแบบสำหรับตัดสินใจเลือกวิธีการพยากรณ์ (Theeramunkong, 2017)

$$c(x) = \underset{c_i \in C}{argmax} P(c_i)P(a_1(x)|c_i) \dots P(a_n(x)|c_i) \quad (1)$$

โดยที่ $c(x)$ คือ ความเป็นไปได้ทั้งหมดของวัตถุ c (x) คือ ตัวแปรไม่ใช่วัตถุป้าย c คือ เซตของวัตถุ c_i คือ ความเป็นไปได้ของวัตถุ c ตำแหน่งที่ i a_n คือ จำนวนคุณลักษณะทั้งหมดมีค่าตั้งแต่ 1 ถึง n P คือ ค่าคงที่

ซัพพอร์ทเวกเตอร์แมชชีน (Support vector machine) คือ หลักการหาระนาบการตัดสินใจในการแบ่งเขตข้อมูลออกเป็นสองส่วนในการหาค่าสนับสนุนเวกเตอร์สำหรับตัดสินใจโดยมีวัตถุประสงค์ที่จะพยายามลดความผิดพลาดการพยากรณ์ นำเข้าข้อมูลสำหรับการสอน $\{\mathrm{x_i y_i}\}, \mathrm{i} = 1, \dots, \mathrm{n},\ \mathrm{x_i} \in \mathrm{R^d}, \mathrm{y_i} \in \{-1,1\}$ สมการ (2)

$$\text{Minimize} \quad \frac{1}{2}||\mathrm{w}||^2$$
$$\text{Subject to} \quad y_i(w^T x_i + b) - 1 \geq 0 \ \text{ for all } i \quad (2)$$

โดยที่ w คือ ค่าระยะห่างยุคลิด b คือ ค่าสัมประสิทธิ์, x_i คือ ตัวเลขโดยอ้างอิงเวกเตอร์ด้วยมิติของ d และ y คือ คลาสไบนารีของตัวเลข -1 หรือ 1 T คือ จำนวนข้อมูลทดสอบทั้งหมด

นิวรอลเน็ต (Neural net) คือ วิธีการเรียนรู้จากตัวอย่างต้นแบบแล้วฝึกให้ระบบได้คิดแก้ปัญหาที่กว้างขึ้นได้ โครงสร้างของนิวรอลเน็ตจะประกอบด้วยโหนดสำหรับนำข้อมูลเข้า,นำข้อมูลออก และการประมวลผล หลักการทำงานจะอาศัยการส่งผ่านโหนดตามระดับชั้นเพื่อการพยากรณ์โหนดผลลัพธ์ที่ต้องการ ดังสมการ (3)

$$\mathrm{y_i} = \int\left(\sum_{\mathrm{i=0}}^{\mathrm{N}} \mathrm{w_{ij}} \mathrm{x_i} + \mathrm{\theta_j}\right) \quad (3)$$

โดยที่ $\mathrm{y_i}$ คือ ตัวแปรผลการพยากรณ์ $\mathrm{x_i}$ คือ ตัวแปรข้อมูลนำเข้า N คือ จำนวนตัวแปรนำเข้าทั้งหมด $\mathrm{w_{ij}}$ คือ ค่านำหนัก $\mathrm{\theta_j}$ คือ ค่าการพยากรณ์ผิดพลาด

ต้นไม้การตัดสินใจ (Decision tree) คือ ลักษณะโครงสร้างลำดับชั้นในการตัดสินใจแต่ละระดับชั้นประกอบด้วยสัญลักษณ์อยู่ 3 สัญลักษณ์ คือ 1) โหนดรากเป็นโหนดตัวกลางที่อยู่บนสุดแสดงออกชัดเจนถึงบรรทัดฐานการตัดสินใจ 2) เส้นที่แสดงถึงการเชื่อมโยงการตัดสินใจ และ3) ส่วนที่อยู่ล่างสุดเรียกว่าโหนดใบจะแสดงถึงกลุ่มของข้อมูลที่จะการพยากรณ์ไปยังโหนดรากของต้นไม้ตัดสินใจ การทำงานของต้นไม้การตัดสินใจจะต้องเลือกคุณสมบัติชุดทดลองในส่วนที่ดีที่สุดเพื่อที่จะจัดวางโหนดรากสำหรับแตกแยกจำนวนสาขาย่อยสาขาอื่นที่มีค่าความเป็นไปได้ ดังสมการ (4)

Information gain :

$$\mathrm{InfoGain(T, A_i)} = \mathrm{Info(T)} - \sum_{\mathrm{k=1}}^{n_i} \frac{|\mathrm{T_{ij}}|}{|\mathrm{T}|} I\mathrm{nfo}\left(\mathrm{T_{ij}}\right) \quad (4)$$
$$\mathrm{Info(T)} = \sum_{k=1}^{m}(-1) \cdot p(c_k, T) \cdot \log_2\left(p(c_k, T)\right)$$

$$\text{Info}(\mathrm{T_{ij}}) = \sum_{k=1}^{m}(-1)\cdot p(c_k, \mathrm{T_{ij}})\cdot \log_2\left(p(c_k, \mathrm{T_{ij}})\right)$$

$$\text{where } p(c_k, T) = \frac{|\mathrm{T_k}|}{|\mathrm{T}|} \text{ and } p(c_k, \mathrm{T_{ij}}) = \frac{|\mathrm{T_{ijk}}|}{|\mathrm{T_{ij}}|}$$

Gain Ratio :

$$\text{GainRatio}(\mathrm{T}, \mathrm{A_i}) = \frac{\text{infoGain}(\mathrm{T},\mathrm{A_i})}{\text{SplitInfo}(\mathrm{T},\mathrm{A_i})}$$

$$\text{SplitInfo}(\mathrm{T}, \mathrm{A_i}) = \sum_{k=1}^{m}(-1)\cdot \mathrm{p}(\mathrm{T_{ij}})\cdot \log_2\left(\mathrm{p}(\mathrm{T_{ij}})\right)$$

$$\text{where } \mathrm{p}(\mathrm{T_{ij}}) = \frac{|\mathrm{T_{ij}}|}{|\mathrm{T}|}$$

โดยที่ $\mathrm{A_i}$ คือ โหนดที่ใช้เลือกเพื่อแตกแยกสาขา, $\mathrm{C_k}$ คือ การจัดแบ่งหมวดหมู่จำนวน k หมวดหมู่, T คือ การแตกแยกก่อนการทดลองชุดข้อมูล, $\mathrm{T_k}$ คือ ค่าคงที่ของชุดข้อมูลด้วยการจัดหมวดหมู่ $\mathrm{C_k}$ในชุดข้อมูล T, $\mathrm{T_{ij}}$ คือ ชุดข้อมูลย่อยที่มีการแตกแยกหลังจากที่ทำการทดลองชุดข้อมูล, $\mathrm{T_{ijk}}$ คือ ค่าคงที่ของชุดข้อมูลด้วยการจัดหมวดหมู่ $\mathrm{C_k}$ในชุดข้อมูลย่อย $\mathrm{T_{ij}}$, $|\mathrm{T}|$ คือ ตัวเลขผลรวมของค่าคงที่ก่อนการแตกแยกในชุดข้อมูลทดลอง, $|\mathrm{T_k}|$ คือ ค่าคงที่การจัดหมวดหมู่ k ของตัวเลขในชุดข้อมูล T, $|\mathrm{T_{ij}}|$ คือ ค่าคงที่ของตัวเลขในชุดข้อมูลย่อย, $|\mathrm{T_{ijk}}|$ คือ คือ ค่าคงที่การจัดหมวดหมู่ k ของตัวเลขในชุดข้อมูลย่อย $\mathrm{T_{ij}}$ กำหนดให้ $\mathrm{A_i} = \mathrm{a_{ij}}$

พาณิชย์อิเล็กทรอนิกส์

พาณิชย์อิเล็กทรอนิกส์ (Electronic commerce : E-commerce) หมายถึง การทำธุรกิจทางอินเทอร์เน็ต ได้รับการยอมรับอย่างกว้างขวางทั่วโลก (Taher, 2021) การแลกเปลี่ยนข้อมูลองค์กรและการดำเนินงานการทำธุรกรรมผ่านระบบโทรคมนาคม การใช้ประโยชน์จากคอมพิวเตอร์ อินเทอร์เน็ต และเทคโนโลยีซอฟต์แวร์ที่ใช้ร่วมกันระหว่างผู้บริโภค ผู้ให้บริการ พนักงาน หรือชุมชน

การแบ่งประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์ (Kalakota & Whinston, 1996) มี 4 ประเภทหลัก ดังนี้

1. ธุรกิจกับธุรกิจ (Business to business : B2B) เป็นธุรกิจที่เน้นการขายสินค้าหรือบริการให้กับบริษัทผู้อื่น ที่เรียกว่า B2B เป็นการทำการค้าขายระหว่างธุรกิจกับธุรกิจด้วยกัน

2. ธุรกิจกับผู้บริโภค (Business to consumer : B2C) หรือ ผู้บริโภคกับธุรกิจ (Consumer to business : C2B)

B2C คือ ประเภทของธุรกรรมการค้าที่ธุรกิจขายผลิตภัณฑ์หรือบริการแก่ผู้บริโภค ซึ่งผู้ผลิตหรือผู้ค้าขายผลิตภัณฑ์ของตนให้กับผู้บริโภคผ่านทางอินเทอร์เน็ต ตรงกันข้ามกับรูปแบบธุรกิจกับผู้บริโภคแบบดั้งเดิม

C2B คือ ธุรกิจสามารถดึงคุณค่าหรือกำไรจากความเต็มใจของผู้บริโภคในการตั้งชื่อเป็นเจ้าของราคา หรือให้ข้อมูลหรือการตลาดแก่บริษัทหรือธุรกิจ

3. ผู้บริโภคถึงผู้บริโภค (Consumer to consumer : C2C) เป็นรูปแบบธุรกิจที่อำนวยความสะดวกในการทำธุรกรรมของสินค้าหรือบริการระหว่างลูกค้ากับลูกค้าด้วยกันเอง หรือผู้ซื้อกับผู้ซื้อ เป็นต้น

4. ธุรกิจกับภาครัฐ (Business to government : B2G) เป็นธุรกิจการบริหารการค้าของประเทศเพื่อเน้นการบริหารการจัดการที่ดีของรัฐบาล ตัวอย่างเช่น การเผยแพร่ข้อมูลเพื่อการเปิดประมูลผ่านทางเครือข่ายเกี่ยวกับการจัดซื้อจัดจ้างของรัฐ การจดทะเบียนการค้า การสืบค้นเครื่องหมายการค้า หรือสิทธิบัตรผ่านทางเครือข่าย เป็นต้น

การวิเคราะห์คุณลักษณะ (Feature analysis)

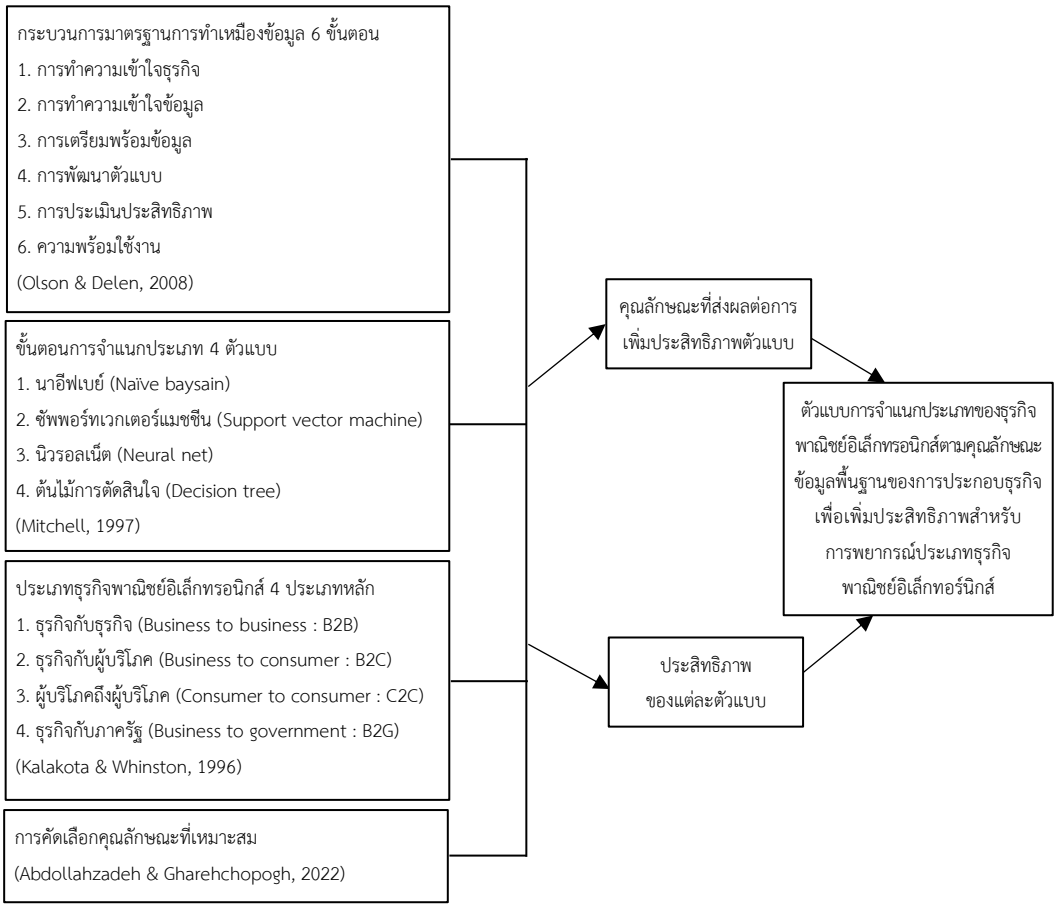
การวิเคราะห์คุณลักษณะ คือ วิธีการประเมินและจัดเรียงลักษณะประจำทีละรายการ เพื่อการค้นหาและเลือกลักษณะประจำที่ระบุตัวประเมินที่อยู่สูงกว่าจุดตัดที่เลือก และเพิ่มประสิทธิภาพถูกต้องแม่นยำของการจัดกลุ่ม ซึ่งการวิจัยนี้ ใช้การวิเคราะห์ลักษณะประจำ 1 วิธี ดังนี้

การหาค่าที่เหมาะสมที่สุดด้วยการคัดเลือกคุณลักษณะ (Feature selection optimization) คือ ขั้นตอนวิธีที่สำคัญในการทำเหมืองข้อมูลที่สามารถการเพิ่มประสิทธิภาพความแม่นยำ ช่วยลดระยะเวลาสำหรับการประมวลผลและการคัดเลือกคุณลักษณะที่เหมาะสม โดยใช้จำนวนขั้นต่ำของคุณลักษณะชุดข้อมูลประกอบด้วยคุณลักษณะ D และ L ของอินสแตน และคุณลักษณะของชุดหลักคือ Fet หลักการทำงานคือ การเลือกคุณสมบัติ d (d ≤ D) จากชุด Fet เพื่อเพิ่มฟังก์ชันวัตถุประสงค์ H การเลือกคุณสมบัติ X ⊆ Fet ชุดย่อยเพื่อให้ได้ความแม่นยำสูงสุดในการจำแนกประเภทสตริงไบนารีใช้เพื่อแสดงวิธีแก้ปัญหา X งานวิจัยนี้ใช้การหาค่าที่เหมาะสมที่สุดด้วยการคัดเลือกคุณลักษณะ (Abdollahzadeh & Gharehchopogh, 2022) ดังสมการ (5)

$$x = (x_1, x_2, \dots, x_3), x \in \{0,1\} \quad (5)$$

โดยที่ X คือ คุณลักษณะทั้งหมด D คือ จำนวนคุณลักษณะที่มากที่สุด x_j =1 คือ คุณลักษณะที่ถูกเลือก, x_j =0 คือ คุณลักษณะที่ไม่ถูกเลือก, j^{th} คือ จำนวนคุณลักษณะที่ถูกเลือก

กรอบแนวคิดในการวิจัย



ภาพ 1 กรอบแนวคิดในการวิจัย

สมมติฐานการวิจัย

1. การวิเคราะห์คุณลักษณ์ร่วมกับขั้นตอนการจำแนกประเภทจะสามารถเพิ่มประสิทธิภาพการพยากรณ์ประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์ให้มีความถูกต้องและแม่นยำได้หรือไม่

2. ตัวแบบการจำแนกประเภทที่มีประสิทธิภาพแตกต่างกันจะสามารถใช้แทนกันได้หรือไม่

โดยมีสมมติฐานทางสถิติ ดังนี้

สมมติฐานที่ 1 (H_0) คือ ยอมรับสมมติฐานตัวแบบใช้แทนกันได้

สมมติฐานที่ 2 (H_1) คือ ยอมรับสมมติฐานตัวแบบใช้แทนกันไม่ได้

คำถามนำวิจัย

1. การวิเคราะห์คุณลักษณะจะส่งผลต่อการเพิ่มประสิทธิภาพของตัวแบบการจำแนกประเภทที่เหมาะสมสำหรับการนำไปพยากรณ์ประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์ให้มีความถูกต้องและแม่นยำเพิ่มขึ้นได้หรือไม่

2. ประสิทธิภาพของตัวแบบการจำแนกประเภทแต่ละตัวแบบที่แตกต่างกันสามารถใช้แทนกันได้หรือไม่

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยและพัฒนา การวิจัยประยุกต์ และการวิจัยเชิงทดลอง สำหรับการวิเคราะห์ข้อมูลเชิงลึกและการใช้ประโยชน์จากข้อมูลที่มีประสิทธิภาพ มีวิธีดำเนินการวิจัยดังต่อไปนี้

ประชากรและกลุ่มตัวอย่าง

ประชากรในการวิจัย คือ นิติบุคคลที่ประกอบธุรกิจพาณิชย์อิเล็กทรอนิกส์ในประเทศไทย จำนวน 2,134 ราย (Department of Business Development, 2021) และใช้สูตรของทาโร่ยามาเน่ (Yamane, 1967) สำหรับคำนวณหากลุ่มตัวอย่าง จำนวน 337 ราย ทั้งนี้คณะวิจัยส่งแบบสอบถามทางไปรษณีย์ 1,685 ฉบับ ได้รับการตอบการตอบกลับมา 445 ราย และเมื่อตรวจสอบความสมบูรณ์ของแบบสอบถามที่สามารถนำมาคำนวณได้อยู่ที่ 400 ฉบับ ดังนั้นจึงได้นำข้อมูลที่สมบูรณ์มาใช้เป็นกลุ่มตัวอย่างของการวิจัยครั้งนี้

เครื่องมือการวิจัย

เครื่องมือที่ใช้ในการวิจัย คือ แบบสอบถามที่ผ่านการตรวจสอบการวัดความเที่ยงตรง โดยการพิจารณาจากผู้เชี่ยวชาญ 5 คน กำหนดเกณฑ์การให้คะแนน คือ +1 คะแนน หมายถึงแน่ใจว่าสอดคล้อง 0 คะแนน หมายถึงไม่แน่ใจว่าสอดคล้อง และ -1 คะแนน หมายถึงแน่ใจว่าไม่สอดคล้อง นำผลคะแนนที่ได้มาคำนวณหาค่าดัชนีความสอดคล้องกับวัตถุประสงค์ (The index of item-objective congruence : IOC) (Tirakanun, 2008) ได้เท่ากับ 0.91 ซึ่งค่าความเที่ยงตรงเชิงเนื้อหาอยู่ในเกณฑ์ดี และคำนวณหาค่าความเชื่อมั่นด้วยการหาค่าสัมประสิทธิ์ Cronbach's alpha พบว่ามีค่าอยู่ระหว่าง 0.72-0.98 ซึ่งค่าความเชื่อมั่นอยู่ในเกณฑ์ดี (Pinkum & Kidrakarn, 2023) และดำเนินการตามขั้นตอนการทำเหมืองข้อมูลตามมาตรฐาน CRISP-DM ใน 6 ขั้นตอน (Olson & Delen, 2008) ดังนี้

1. การทำความเข้าใจธุรกิจ ปัจจุบันเทคโนโลยีทำให้มนุษย์ติดต่อสื่อสารระหว่างกันผ่านบริการช่องทางอินเตอร์เน็ตกันมากขึ้น เกิดเป็นข้อมูลจำนวนมากที่เป็นพลวัตและซับซ้อนยากต่อการพยากรณ์ จึงมีความจำเป็นยิ่งที่ต้องบูรณาการศาสตร์หลายมิติร่วมกับศาสตร์การพยากรณ์ความต้องการ เพื่อการพัฒนาธุรกิจของประเทศไทยให้มีการบริหารจัดการข้อมูลเชิงลึกและนำข้อมูลไปใช้ประโยชน์ให้เต็มที่และมีประสิทธิภาพ (Moonpen, Mungsing, & Banditwattanawong, 2020) การระบุและจำแนกประเภทคุณลักษณะต่าง ๆ ของขั้นตอนการพัฒนาพาณิชย์อิเล็กทรอนิกส์ตามประเภทของธุรกิจ ได้แก่ พาณิชย์อิเล็กทรอนิกส์แบบธุรกิจกับธุรกิจ (B2B) พาณิชย์อิเล็กทรอนิกส์แบบธุรกิจกับผู้บริโภค (B2C) พาณิชย์อิเล็กทรอนิกส์แบบผู้บริโภคกับผู้บริโภค (C2C) และพาณิชย์อิเล็กทรอนิกส์แบบธุรกิจผู้กับภาครัฐ (B2G) เป็นต้น ที่มีอยู่เพื่อช่วยพัฒนาการวางแผนความต้องการของตลาดของแต่ละประเภทธุรกิจสำหรับอนาคตได้

2. การทำความเข้าใจข้อมูล โดยใช้ข้อมูลคุณลักษณะพื้นฐานนิติบุคคลที่ประกอบธุรกิจพาณิชย์อิเล็กทรอนิกส์นิติบุคคลที่ประกอบธุรกิจพาณิชย์อิเล็กทรอนิกส์จากการสอบถามกลุ่มตัวอย่าง จำนวน 400 ราย โดยแต่ละระเบียนมีคุณลักษณะของข้อมูล จำนวน 13 คุณลักษณะ ประกอบด้วย เพศ อายุเจ้าของธุรกิจ ระดับการศึกษา ประสบการณ์การทำงานในธุรกิจพาณิชย์อิเล็กทรอนิกส์ รายได้เฉลี่ยต่อเดือน ตำแหน่งงานปัจจุบัน ทุนจดทะเบียนของธุรกิจ รายได้ของธุรกิจเฉลี่ยต่อปี ประเภทของการประกอบธุรกิจ จำนวนพนักงานในปัจจุบันทั้งหมด (คน) ระยะเวลาในการดำเนินธุรกิจ สินทรัพย์รวมของธุรกิจ (รวมอสังหาริมทรัพย์) และประเภทธุรกิจพาณิชย์อิเล็กทรอนิกส์ ซึ่งเป็นตัวแปรเป้าหมายสำหรับการสร้างตัวแบบการจำแนกประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์

3. การเตรียมพร้อมข้อมูล โดยแปลงข้อมูลให้อยู่ในมาตรฐานเดียวกันด้วยการกำหนดเป็นจำนวนเต็ม (Integer) คุณลักษณะเป้าหมายที่จะพยากรณ์ทำการประยุกต์เป็นป้ายเชิงวัตถุ (Class label) การกำจัดข้อมูลที่ค่าหายไป (Missing value) ออก และมีการกำหนดค่าคุณลักษณะข้อมูล ดังแสดงในตาราง 1

ตาราง 1 การเตรียมข้อมูล

ชื่อตัวแปร	คำอธิบาย	ชนิดตัวแปร	ค่าตัวแปร
Ec_Type	ประเภทธุรกิจพาณิชย์อิเล็กทรอนิกส์	นามบัญญัติ	(B2C) (B2B) (B2G) (C2C) (B2C, B2B) (B2B, B2G) (B2C, B2G) (B2C, C2C) (B2B, C2C) (B2C, B2B, B2G) (B2C, B2B, C2C) (B2B, B2G, C2C)
Gender	เพศ	ตัวเลข	[1] ชาย, [2] หญิง
Age.Person	อายุเจ้าของธุรกิจ	ตัวเลข	[1] น้อยกว่า 30 ปี, [2], อายุ30-40 ปี, [3] อายุน้อยกว่า 41-50 ปี, [4] อายุ มากกว่า 50 ปี
Edu	ระดับการศึกษา	ตัวเลข	[1] มัธยมศึกษาหรือต่ำกว่า, [2] ปวช./ปวส./อนุปริญญา, [3] ปริญญาตรี, [4] ปริญญาโท, [5] ปริญญาเอก, [6] อื่นๆ โปรดระบุ.....
Exp	ประสบการณ์การทำงานในธุรกิจพาณิชย์อิเล็กทรอนิกส์	ตัวเลข	[1] น้อยกว่า 1 ปี, [2] 1-5 ปี, [3] 6-10 ปี [4], มากกว่า 10 ปี
Income.P	รายได้เฉลี่ยต่อเดือน	ตัวเลข	[1] น้อยกว่า 25,000 บาท, [2] 25,000-50,000 บาท, [3] 50,001-100,000 บาท, [4] มากกว่า 100,000 บาท
Position	ตำแหน่งงานปัจจุบัน	ตัวเลข	[1] เจ้าของธุรกิจ, [2] ประธานเจ้าหน้าที่บริหาร/กรรมการผู้จัดการ/ผู้จัดการทั่วไป, [3] อื่นๆ โปรดระบุ .....
Capital	ทุนจดทะเบียนของธุรกิจ	ตัวเลข	[1] น้อยกว่า 500,000 บาท, [2] 500,000-1,000,000 บาท [3] 1,000,001-10,000,000 บาท, [4] มากกว่า 10,000,000 บาท
IncomeperY	รายได้ของธุรกิจเฉลี่ยต่อปี	ตัวเลข	[1] น้อยกว่า 1,000,000 บาท, [2]1,000,000-25,000,000 บาท, [3] 25,000,001-50,000,000 บาท, [4] 50,000,001-75,000,000 บาท, [5] 75,000,001-100,000,000 บาท, [6] มากกว่า 100,000,000 บาท
Type	ประเภทของการประกอบธุรกิจ	ตัวเลข	[1] กิจการเจ้าของคนเดียว, [2] ห้างหุ้นส่วนสามัญ/ห้างหุ้นส่วนจำกัด, [3] บริษัทจำกัด, [4] บริษัทมหาชนจำกัด, [5] อื่นๆ โปรดระบุ ...
N_Em	จำนวนพนักงานในปัจจุบันทั้งหมด (คน)	จำนวนนับ	ระบุ.....(จำนวนนับ)
Age	ระยะเวลาในการดำเนินธุรกิจ	ตัวเลข	[1] น้อยกว่า 1 ปี, [2] 1-5 ปี, [3] 6-10 ปี, [4] มากกว่า 10 ปี
Asset	สินทรัพย์รวมของธุรกิจ (รวมอสังหาริมทรัพย์)	ตัวเลข	[1] น้อยกว่า 1,000,000 บาท, [2] 1,000,000-25,000,000 บาท, [3] 25,000,001-50,000,000 บาท, [4] 50,000,001-75,000,000 บาท, [5] 75,000,001-100,000,000 บาท, [6] มากกว่า 100,000,000 บาท

ทั้งนี้ เครื่องมือที่ใช้สำหรับการวิเคราะห์ข้อมูล ใช้โปรแกรมแรพิดไมเนอร์ (RapidMiner) รุ่นที่ 9.0 เพื่อการพัฒนากระบวนการขั้นตอนวิธีการจำแนกประเภทร่วมกับขั้นตอนวิธีการเพิ่มประสิทธิภาพ (RapidMiner, 2023)

4. การพัฒนาตัวแบบ ทำการพัฒนาตัวแบบการจำแนกประเภทร่วมกับขั้นตอนวิธีการเพิ่มประสิทธิภาพเพื่อการวิเคราะห์หาคุณลักษณะของข้อมูลที่ส่งผลต่อการเพิ่มประสิทธิภาพตัวแบบการจำแนกประเภทที่ดีที่สุด และเปรียบเทียบความแตกต่างประสิทธิภาพของตัวแบบ ทั้ง 4 ตัวแบบ ประกอบด้วยขั้นตอนดังต่อไปนี้

4.1 เลือกขั้นตอนวิธีการจำแนกประเภท ในหัวข้อนี้เราทำการเลือกขั้นตอนวิธีการจำแนกประเภทเพื่อใช้สร้างตัวแบบการจำแนกประเภทที่ดีที่สุด จำนวน 4 ตัวแบบ คือ นาอีฟเบย์ นิวรอลเน็ต ซัพพอร์ทเวกเตอร์แมชชีน และต้นไม้การตัดสินใจ และกำหนดค่าพารามิเตอร์ที่ทดสอบความถูกต้อง ความแม่นยำ และความครบถ้วนที่มีค่าสูงที่สุด ดังแสดงในตาราง 2

ตาราง 2 การปรับค่าพารามิเตอร์สำหรับการสร้างตัวแบบการจำแนกประเภท

ตัวแบบ	ค่าพารามิเตอร์
นาอีฟเบย์ (Naïve Bayes)	laplace correction=True
นิวรอนเน็ต(Neural Network)	hidden layers=2, training cycles=500, learning rate=0.01, momentum=0.9, shuffle=True, normalize=True, use local random seed=False, local random seed=2500
ซัพพอร์ทเวกเตอร์แมชชีน (Support Vector Machine)	kernel type=polynomial, kernel degree=3.0, kernel cache=200, C=0.0, convergence epsilon=0.01, max iterations=100000, scale=True, L pos=1.0, L neg=1.0
ต้นไม้การตัดสินใจ (Decision Tree)	criterion=accuracy, maximal depth=20, apply pruning=True, confidence=0.1, apply prepruning=True, minimal gain=0.01, minimal leaf size=2, minimal size for split=4, number of prepruning alternatives=3

4.2 การเพิ่มประสิทธิภาพของตัวแบบการจำแนกประเภทที่เหมาะสมด้วยขั้นตอนวิธีการคัดเลือกค่าที่เหมาะสมที่สุด (Optimize selection) และกำหนดค่าพารามิเตอร์ ดังแสดงในตาราง 3

ตาราง 3 การกำหนดค่าพารามิเตอร์การวิเคราะห์หาลักษณะประจำ

ขั้นตอนวิธี	การกำหนดค่าพารามิเตอร์
การคัดเลือกค่าที่เหมาะสมที่สุด (Optimize selection)	Selection Direction=forward, limit generations without improval=yes, generations without improval=10, limt number of generations=Yes, keep best=3, maximum number of generations=50, normalize weights=yes, use local random seed=2500, show population plotter=yes, plot generations=10, draw dominated=yes points

จากตาราง 3 ส่งผลต่อการเพิ่มประสิทธิภาพของตัวแบบการจำแนกประเภทให้มีค่าความถูกต้อง ความแม่นยำ ค่าความครบถ้วน และค่าเอฟเมชเชอร์ที่เพิ่มสูงขึ้น และมีจำนวนคุณลักษณะที่น้อยที่สุด เพื่อทดสอบการวัดประสิทธิภาพ การปรับค่าพารามิเตอร์ และเปรียบเทียบการเพิ่มประสิทธิภาพที่ดีที่สุด

4.3 การเปรียบเทียบความแตกต่างของตัวแบบด้วยสถิติทีเทส (T-test) และอโนวาเทส (ANOVA test) เป็นการประเมินประสิทธิภาพของตัวแบบการพยากรณ์ที่สร้างไว้แล้ว จำนวน 4 ตัวแบบ มาดำเนินการเปรียบเทียบกันทั้งหมด เนื่องจากผลการสร้างตัวแบบการจำแนกประเภทและผลการเพิ่มประสิทธิภาพของตัวแบบการจำแนกประเภทมีประสิทธิภาพใกล้เคียงกัน ดังนั้นการทำการปรียบเทียบประสิทธิภาพของตัวแบบด้วยสถิติ ANOVA และแบบรายคู่ (T-test) ในการวัดค่าเฉลี่ยประสิทธิภาพความแตกต่างของตัวแบบทั้งหมด จะให้ข้อสรุปว่าตัวแบบใดมีความแตกต่างกันจะไม่สามารถใช้แทนกันได้ แต่ตรงกันข้ามกับตัวแบบใดไม่มีความแตกต่างกันจะไม่สามารถใช้แทนกันได้

5. การประเมินประสิทธิภาพ

5.1 การตรวจสอบแบบไขว้ (Cross validation) เป็นวิธีการแยกข้อมูลออกเป็นสองส่วนคือ ข้อมูลใช้ในการสอน และข้อมูลที่ใช้สำหรับทดสอบ (RapidMiner, 2014) โดยแบ่งข้อมูลสำหรับการสอน 70% และสำหรับการทดสอบ 30% เพื่อวัดประสิทธิภาพของความถูกต้อง ความแม่นยำ ค่าความครบถ้วน และค่า F-measure ตาม (6) (7) (8)และ(9) ทั้งนี้ การสร้างตัวแบบการจำแนกประเภทต้องอาศัยการคำนวณหาค่าการพยากรณ์ที่ดีที่สุดตามสมการดังกล่าว จึงจะนำไปใช้ในการตัดสินใจเลือกตัวแบบการจำแนกประเภทที่ให้ค่ามากที่สุดจะถือว่าเป็นตัวแบบที่ดีที่สุดและเหมาะสมสำหรับการนำไปใช้พยากรณ์แนวโน้มที่จะเกิดขึ้นในอนาคตตามวัตถุประสงค์การวิจัย

$$\text{ความถูกต้อง} = \frac{TP}{\text{TP+FP}} \quad (6)$$

$$\text{ความแม่นยำ} = \frac{TP}{\text{TP+FN}} \quad (7)$$

$$\text{ความครบถ้วน} = \frac{TP+TN}{\text{TP+TN+FP+FN}} \quad (8)$$

$$\text{F-measure} = \frac{\text{2 x recall x precision}}{\text{(recall+precision)}} \quad (9)$$

โดยที่ **TP** คือ จำนวนที่ทำนายถูกจริงอยู่ในคลาสที่พิจารณา, **TN** คือ จำนวนที่ทำนายถูกไม่อยู่ในคลาสที่พิจารณา, **FP** คือ จำนวนที่ทำนายผิดแต่อยู่ในคลาสที่พิจารณา, และ **FN** คือ จำนวนที่ทำนายผิดและไม่อยู่ในคลาสที่พิจารณา ตัวแบบที่ดีจะต้องมีค่าความถูกต้อง ความแม่นยำ และความครบถ้วนสูง (Theeramunkong, 2017)

5.2 ทดสอบสมมติฐานเปรียบเทียบค่าเฉลี่ยประสิทธิภาพความแตกต่างของตัวแบบด้วยสถิติ ANOVA เป็นวิธีการทดสอบสมมติฐานจากการหาค่าเฉลี่ยของประสิทธิภาพความแตกต่างของตัวแบบ 4 ตัวแบบนั้นสามารถใช้แทนกันได้หรือไม่ งานวิจัยมีสมมติฐานดังนี้ สมมติฐานที่ 1 (H_0) คือ ยอมรับสมมติฐานตัวแบบใช้แทนกันได้, สมมติฐานที่ 2 (H_1) คือ ยอมรับสมมติฐานตัวแบบใช้แทนกันไม่ได้ ตาม (10)

$$F = \frac{\text{ค่าเฉลี่ยความแปรปรวนระหว่างกลุ่ม}}{\text{ค่าเฉลี่ยความแปรปรวนภายในกลุ่ม}} \quad (10)$$

โดยที่ F คือ ค่าเฉลี่ยระหว่างความแปรปรวน,ความน่าเชื่อถือ ≤ 0.05 คือ ยอมรับ H_0 ปฏิเสธ H_1 หรือความน่าเชื่อถือ ($p<0.05$) คือ ยอมรับ H_1 ปฏิเสธ H_0

6. ความพร้อมใช้งาน สามารถนำตัวแบบพยากรณ์ประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์และวิเคราะห์คุณลักษณะข้อมูลที่ส่งผลต่อการเพิ่มประสิทธิภาพของตัวแบบการจำแนกประเภทไปช่วยวางแผนการบริหารจัดการประเภทธุรกิจพาณิชย์อิเล็กทรอนิกส์ให้มีความเหมาะสม ทั้งนี้ยังช่วยวางแผนการทำการตลาดแข่งขันกับคู่แข่งได้อย่างรวดเร็วและแม่นยำ สามารถเพิ่มประสิทธิภาพการพยากรณ์โดยการมุ่งเน้นในการเลือกคุณลักษณะที่เหมาะสมที่สุด

ผลการวิจัย

ผลการสร้างตัวแบบการจำแนกประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์ตามคุณลักษณะข้อมูลพื้นฐานของการประกอบการธุรกิจที่มีประสิทธิภาพ แสดงผลลัพธ์ที่ได้แสดงดังตาราง 4

ตาราง 4 ประสิทธิภาพตัวแบบการจำแนกประเภทของข้อมูลธุรกิจพาณิชย์อิเล็กทรอนิกส์

ตัวแบบ	ความถูกต้อง	ความแม่นยำ	ความครบถ้วน	เอฟเมเซอร์ (F-Measure)
นาอีฟเบย์	46.63%	49.41%	58.15%	53.43%
นิวรอนเน็ต	73.78%	64.49%	59.11%	61.83%
ซัพพอร์ทเวกเตอร์แมชชีน	80.97%	94.58%	72.00%	81.76%
ต้นไม้การตัดสินใจ	71.70%	59.68%	58.06%	58.86%

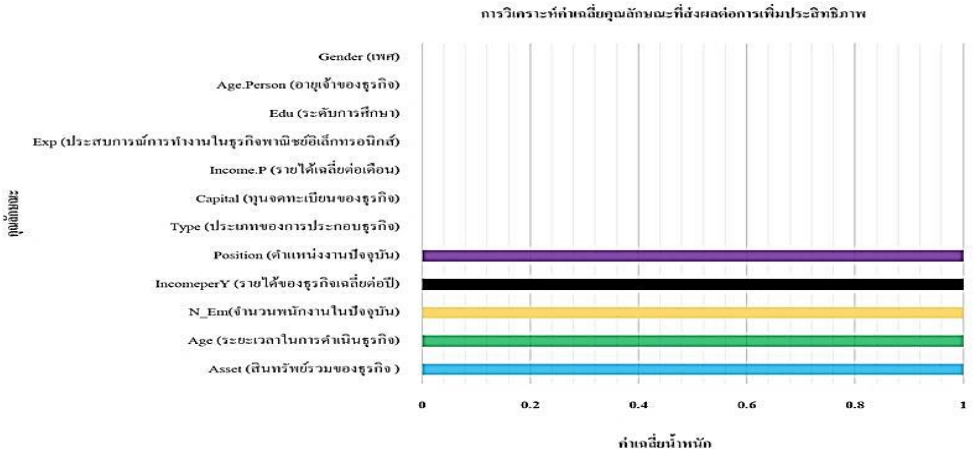
จากตาราง 4 พบว่า ขั้นตอนวิธีที่ดีที่สุดในการสร้างและเปรียบเทียบตัวแบบการจำแนกประเภท จำนวน 4 ตัวแบบ ซึ่งตัวแบบซัพพอร์ทเวกเตอร์แมชชีนดีที่สุด โดยวัดค่าประสิทธิภาพความถูกต้อง 80.97% ค่าความแม่นยำ 94.58% ค่าความครบถ้วน 72.00% และค่าเอฟเมเซอร์ 81.76%

ผลการวิเคราะห์คุณลักษณะที่ส่งผลต่อการเพิ่มประสิทธิภาพของตัวแบบการจำแนกประเภทที่เหมาะสมสำหรับการนำไปพยากรณ์ประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์ แสดงผลลัพธ์แสดงดังตาราง 5

ตาราง 5 ผลการเพิ่มประสิทธิภาพขั้นตอนวิธีการจำแนกประเภทด้วยขั้นตอนวิธีการคัดเลือกค่าที่เหมาะสมที่สุด

ตัวแบบ	ขั้นตอนวิธีการคัดเลือกค่าที่เหมาะสมที่สุด				
	ความถูกต้อง	ความแม่นยำ	ความครบถ้วน	เอฟเมเซอร์ (%เปลี่ยนแปลง)	จำนวนคุณลักษณะ
นาอีฟเบย์	50.13%	45.35%	40.17%	42.60% (-20.27%)	3
นิวรอนเน็ต	73.79%	60.30%	58.37%	59.31% (-4.07%)	10
ซัพพอร์ทเวกเตอร์แมชชีน	83.30%	95.63%	74.64%	83.84% (+2.54%)	10
ต้นไม้การตัดสินใจ	80.04%	72.29%	69.80%	71.02% (+20.66%)	5

จากตาราง 5 พบว่า ผลลัพธ์ของขั้นตอนวิธีการคัดเลือกค่าที่เหมาะสมที่สุดร่วมกับขั้นตอนวิธีการจำแนกประเภทที่ส่งผลต่อการเพิ่มประสิทธิภาพของตัวแบบการจำแนกประเภทที่ดีที่สุด คือ ตัวแบบต้นไม้การตัดสินใจดีที่สุด โดยมีค่าความถูกต้อง 80.04% ค่าความแม่นยำ 72.29% ค่าความครบถ้วน 69.80% และค่าเอฟเมเซอร์ 71.02% ซึ่งมีประสิทธิภาพเพิ่มขึ้นสูงที่สุด 20.66% พิจารณาจากเปอร์เซนต์เปลี่ยนแปลงของค่าเอฟเมเซอร์ 58.86% (ตาราง 4) เทียบกับค่าเอฟเมเซอร์ 71.02% (ตาราง 5) และผลการวิเคราะห์คุณลักษณะที่ส่งผลต่อการเพิ่มประสิทธิภาพของตัวแบบที่ดีที่สุด พบว่า มีคุณลักษณะที่น้อยที่สุด คือ จำนวน 5 คุณลักษณะ จากทั้งหมด 13 คุณลักษณะ แสดงดังภาพ 2

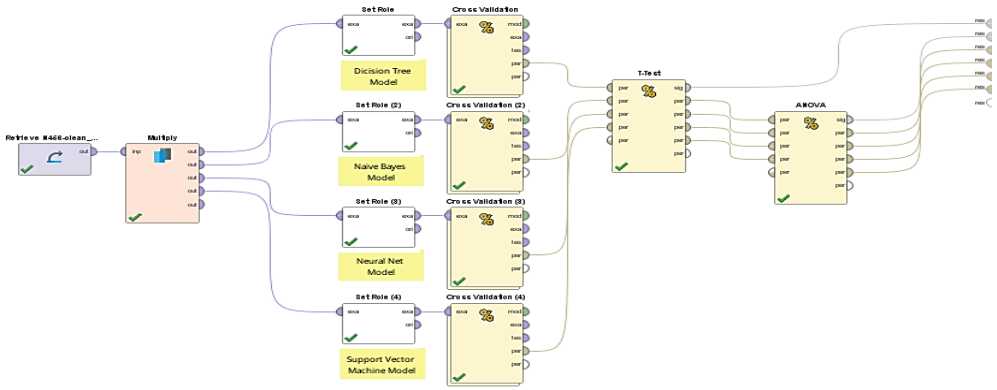


ภาพ 2 ค่าเฉลี่ยน้ำหนักที่เหมาะสมของคุณลักษณะที่เพิ่มประสิทธิภาพของตัวแบบการจำแนกประเภทที่ดีที่สุด

จากภาพ 2 แสดงผลค่าเฉลี่ยน้ำหนักของคุณลักษณะที่เหมาะสม โดยมีค่าอยู่ระหว่าง 0 และ 1 พบว่าค่าเฉลี่ยน้ำหนักของคุณลักษณะที่เหมาะสมที่สุดจะมีค่าเท่ากับ 1 หมายความว่า คุณลักษณะที่เหมาะสมที่สุดสำหรับการเพิ่มประสิทธิภาพตัวแบบการจำแนกประเภทการพยากรณ์ประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์ประกอบด้วย 5 คุณลักษณะ คือ Asset, Age, N_Em, IncomperY และ Position และค่าเฉลี่ยน้ำหนักที่เหมาะสม

ของคุณลักษณะที่มีค่าเท่ากับ 0 หมายความว่า คุณลักษณะประจำที่ไม่เหมาะสมไม่ส่งผลต่อการมีประสิทธิภาพที่เพิ่มขึ้น จะไม่แสดงแท่งกราฟ มีจำนวน 8 ลักษณะประจำ ตัวอย่างเช่น Size และCapital เป็นลักษณะประจำที่ไม่เหมาะสม ดังนั้น การจัดแบ่งประเภทธุรกิจพาณิชย์อิเล็กทรอนิกส์ไม่ต้องระบุ Size และCapital เพราะไม่มีผลต่อการเพิ่มประสิทธิภาพของการจำแนกสำหรับการพยากรณ์ประเภทธุรกิจพาณิชย์อิเล็กทรอนิกส์

ผลการเปรียบเทียบประสิทธิภาพความแตกต่างระหว่างตัวแบบการจำแนกประเภทกับตัวแบบวิเคราะห์คุณลักษณะที่ส่งผลต่อการเพิ่มประสิทธิภาพของตัวแบบการจำแนกประเภทที่เหมาะสมสำหรับการนำไปพยากรณ์ประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์ แสดงดังภาพ 3



ภาพ 3 การพัฒนาการเปรียบเทียบค่าเฉลี่ยประสิทธิภาพความแตกต่างของตัวแบบด้วยสถิติอโนวา

จากภาพ 3 แสดงผลลัพธ์ของกระบวนการพัฒนาวิธีการเปรียบเทียบค่าเฉลี่ยประสิทธิภาพความแตกต่างของตัวแบบด้วยสถิติอโนวาและทีเทส ด้วยการนำตัวแบบทั้งหมด 4 ตัวแบบ คือ นาอีฟเบย์ นิวรอนเน็ต ซัพพอร์ทเวกเตอร์แมชชีน และต้นไม้การตัดสินใจมาเปรียบเทียบประสิทธิภาพพร้อมกัน ได้ผลลัพธ์ดังตาราง 6

ตาราง 6 ผลการเปรียบเทียบค่าเฉลี่ยประสิทธิภาพความแตกต่างของตัวแบบด้วยสถิติ ANOVA

เปรียบเทียบค่าเฉลี่ยประสิทธิภาพของตัวแบบ	ผลรวมกำลังสองของค่าเบี่ยงเบน	จำนวนกลุ่มของตัวแบบ	ค่าเฉลี่ยความแปรปรวน	ค่าเฉลี่ยระหว่างความแปรปรวน (F)	ค่าความน่าเชื่อถือ (Prob)
ระหว่างตัวแบบ	0.475	3	0.158	69.173	0.000
ภายในตัวแบบ	0.071	31	0.002	-	-
รวม	0.454	34	-	-	-

จากตาราง 6 พบว่า มี 3 ตัวแบบที่มีประสิทธิภาพแตกต่างกันระหว่างตัวแบบ คือ นาอีฟเบย์ นิวรอนเน็ต ซัพพอร์ตเวกเตอร์แมชชีน และนาอีฟเบย์ นิวรอนเน็ต ต้นไม้การตัดสินใจ โดยมีค่าเฉลี่ยความแปรปรวน 0.158 ค่าเฉลี่ยระหว่างความแปรปรวน 69.173 และค่าความน่าเชื่อถือ 0.000 ซึ่งมีค่าน้อยกว่าค่าระดับนัยสำคัญคือ 0.05 แสดงว่ายอมรับสมมติฐานที่ 1 ชี้ให้เห็นว่า มี 2 ตัวแบบที่สามารถใช้แทนกันได้ เป็นจริง และผลลัพธ์คือ ตัวแบบซัพพอร์ตเวกเตอร์แมชชีน และตัวแบบต้นไม้การตัดสินใจ มีค่าเฉลี่ยประสิทธิภาพไม่แตกต่างกันสามารถใช้แทนกันได้ แสดงดังตาราง 7

ตาราง 7 ผลการเปรียบเทียบค่าเฉลี่ยประสิทธิภาพความแตกต่างของตัวแบบด้วยรายคู่ (T-test)

ตัวแบบ / ค่าเฉลี่ยประสิทธิภาพ		ต้นไม้การตัดสินใจ	นาอีฟเบย์	นิวรอนเน็ต	ซัพพอร์ทเวกเตอร์แมชชีน
		0.717+/-0.061	0.453 +/- 0.043	0.726 +/- 0.038	0.812 +/- 0.043
ต้นไม้การตัดสินใจ	0.717+/-0.061	-	0.000**	0.696	0.001*
นาอีฟเบย์	0.453 +/- 0.043	0.000**	-	0.000**	0.000**
นิวรอนเน็ต	0.726 +/- 0.038	0.696	0.000**	-	0.000**
ซัพพอร์ทเวกเตอร์แมชชีน	0.812 +/- 0.043	0.001*	0.000**	0.000**	-

* p<0.05, ** p<0.01

จากตาราง 7 พบว่า ตัวแบบเปรียบเทียบรายคู่ทั้ง 5 คู่ เป็นตัวแบบที่มีประสิทธิภาพไม่แตกต่างกัน ซึ่งมีค่าน้อยกว่าค่ามาตรฐานคือ 0.05 อย่างมีนัยสำคัญ จึงสามารถนำไปใช้สร้างตัวแบบการจำแนกประเภทสำหรับการพยากรณ์แทนกันได้ และมีตัวแบบเปรียบเทียบรายคู่ที่มีประสิทธิภาพแตกต่างกัน 1 คู่ คือ ต้นไม้การตัดสินใจและนิวรอนเน็ต โดยมีค่าเท่ากับ 0.696 ซึ่งมีค่ามากกว่าค่าระดับนัยสำคัญทางสถิติที่ยอมรับได้คือ 0.05 หมายความว่า ไม่สามารถนำตัวแบบต้นไม้การตัดสินใจไปใช้แทนตัวแบบนิวรอนเน็ตได้

อภิปรายผล

การสร้างตัวแบบการจำแนกประเภทที่ดีที่สุดสำหรับการพยากรณ์ประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์ในประเทศไทย คือ ตัวแบบซัพพอร์ทเวกเตอร์แมชชีน โดยมีค่าความถูกต้อง 80.97% ค่าความแม่นยำ 94.58% ค่าความครบถ้วน 72.00% และค่าเอฟเมเชอร์ 81.76% ซึ่งแตกต่างจากงานวิจัยของ Mirkovic et al. (2022) ที่พบว่าการถดถอยเชิงตรรกะมีประสิทธิภาพดีที่สุด ที่เป็นเช่นนั้นเนื่องจากงานวิจัยดังกล่าว ใช้จำนวนข้อมูลประวัติที่แตกต่างกันสำหรับวิเคราะห์คุณลักษณะข้อมูลและการเลือกใช้การจำแนกประเภทที่แตกต่างกัน จึงนำไปสู่ผลลัพธ์ของตัวแบบที่มีประสิทธิภาพดีที่สุดแตกต่างกัน สำหรับการคัดเลือกค่าที่เหมาะสมที่สุดร่วมกับการจำแนกประเภทที่ส่งผลต่อการเพิ่มประสิทธิภาพของตัวแบบการจำแนกประเภทที่ดีที่สุด คือ ตัวแบบต้นไม้การตัดสินใจดีที่สุด โดยมีค่าความถูกต้อง คือ 80.04% ค่าความแม่นยำ 72.29% ค่าความครบถ้วน 69.80% และค่าเอฟเมเชอร์ 71.02% ซึ่งมีประสิทธิภาพเพิ่มขึ้นสูงที่สุด 20.66% และผลการวิเคราะห์คุณลักษณะที่ส่งผลต่อการเพิ่มประสิทธิภาพของตัวแบบการจำแนกประเภทที่ดีที่สุดจำนวน 5 คุณลักษณะ คือ Asset, Age, N_Em, IncomperY และ Position สอดคล้องกับงานวิจัย Hayble-Gomes (2023) ที่ได้เสนอตัวแบบต้นไม้การตัดสินใจที่ดีที่สุด มีค่าความถูกต้อง คือ 98% และค่าความแม่นยำ คือ 97% ที่เป็นเช่นนั้นเพราะคุณลักษณะข้อมูลและค่าพารามิเตอร์ที่ใช้มีผลต่อการเพิ่มประสิทธิภาพของตัวแบบให้มีความเหมือนหรือแตกต่างกันไป งานวิจัยนี้ได้ค้นพบลำดับความสำคัญคุณลักษณะของขั้นตอนการพัฒนาอีคอมเมิร์ซในแต่ละประเภทธุรกิจ คือ B2B, B2C, B2G และ C2C ที่บ่งชี้ได้ว่าธุรกิจดังกล่าว มีการใช้คุณลักษณะในระบบอีคอมเมิร์ซเพิ่มประสิทธิภาพบริการให้ลูกค้าได้ และสร้างความไว้วางใจได้สำเร็จ (Bueno & Gallego, 2021) อย่างไรก็ตาม ในประเทศที่พัฒนาแล้วและกำลังพัฒนาพบว่า ระดับของการแพร่กระจายของธุรกิจพาณิชย์อิเล็กทรอนิกส์ประเภทธุรกิจกับธุรกิจจะสูงขึ้นในประเทศที่มีกรอบสถาบันระดับชาติสนับสนุน ส่วนการเปรียบเทียบตัวแบบรายคู่เพื่อค้นหาประสิทธิภาพตัวแบบแสดงให้เห็นว่าตัวแบบต้นไม้การตัดสินใจและตัวแบบซัพพอร์ตเวกเตอร์แมชชีนสามารถใช้แทนกันได้ และผู้ประกอบธุรกิจสามารถนำไปประยุกต์ใช้งานเพื่อช่วยสนับสนุนการตัดสินใจเกี่ยวกับการดำเนินธุรกิจได้ตรงเป้าหมาย

ข้อเสนอแนะ

ข้อเสนอแนะในการวิจัยครั้งต่อไป

ควรมีการเพิ่มตัวแบบรีเกตชั่น ตัวแบบเพื่อนบ้านที่ใกล้ที่สุด และตัวแบบการเพิ่มประสิทธิภาพ ด้วยการหาค่าที่เหมาะสมที่สุดแบบวิวัฒนาการ สำหรับการศึกษาเปรียบเทียบให้เห็นถึงประสิทธิภาพความแตกต่างได้ครอบคลุมมากยิ่งขึ้น และสามารถพัฒนาต่อยอดเพิ่มเป็นระบบการแนะนำประเภทของธุรกิจพาณิชย์อิเล็กทรอนิกส์ที่เหมาะสมอย่างอัตโนมัติสำหรับผู้ประกอบธุรกิจให้สามารถการเข้าถึงได้ทุกที่ทุกเวลา รวดเร็ว และสะดวกในการใช้งานยิ่งขึ้น

เอกสารอ้างอิง

Abdollahzadeh, B., & Gharehchopogh, F. S. (2022). A multi-objective optimization algorithm for feature selection problems. *Engineering with Computers, 38*(3), 1845-1863.

Bueno, S., & Gallego, M. D. (2021). eWOM in C2C platforms: Combining IAM and customer satisfaction to examine the impact on purchase intention. *Journal of Theoretical and Applied Electronic Commerce Research, 16*(5), 1612-1630.

Department of Business Development. (2021). *Legal entity information and financial statements.* Retrieved April 1, 2021, from https://datawarehouse.dbd.go.th/shearchJuristicInfo (In Thai)

Electronic Transactions Development Agency. (2022). *Value of e-Commerce Survey in Thailand 2021.* Retrieved February 2, 2022, from https://www.etda.or.th/th/Our-Service/Statistics-and-information.aspx (In Thai)

Guo, H., Liu, Y., Shi, X., & Chen, K. Z. (2021). The role of e-commerce in the urban food system under COVID-19: Lessons from China. *China Agricultural Economic Review, 13*(2), 436-455.

Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques*. N. P. : Morgan kaufmann.

Hayble-Gomes, E. (2023). The use of predictive modeling to identify relevant features for suspicious activity reporting. *Journal of Money Laundering Control, 26*(4), 806-830.

Huang, L., Huang, Y., Huang, R., Xie, G., & Cai, W. (2022). Factors influencing returning migrants' entrepreneurship intentions for rural E-commerce: an empirical investigation in China. *Sustainability, 14*(6), 3682.

Mirkovic, M., Lolic, T., Stefanovic, D., Anderla, A., & Gracanin, D. (2022). Customer Churn Prediction in B2B Non-Contractual Business Settings Using Invoice Data. *Applied Sciences*, *12*(10), 1-18.

Moonpen, U., Mungsing, S., & Banditwattanawong, T. (2020). Clustering Algorithm Optimization Model for Essential Attribute Analysis of Tour Package Forms Inbound Tourism Market in Thailand. *The Journal of King Mongkut's University of Technology North, 30*(4), 656-667. (In Thai)

Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques*. New York : Springer.

Pinkum, S., & Kidrakarn, K. (2023). An Investigation into Improving Service Quality Distribution and Perceived Value in the Passenger Loyalty of Low-Cost Airlines in Thailand. *Journal of Distribution Scicence (JDS), 21*(4), 21-33.

RapidMiner. (2014). *Rapid Miner 6.0 User Manual*. Retrieved June, 14, 2021, from https://docs.rapidminer.com/download/RapidMiner-v6-user-manual.pdf

Taher, G. (2021). E-Commerce: Advantages and Limitations. *International Journal of Academic Research in Accounting Finance and Management Sciences*, *11*(1), 153-165.

Theeramunkong, T. (2017). *Introduction to Concepts and Techniques in Data Mining and Application to Text Mining Second Edition.* Bangkok : Thammasat University Printing House. (In Thai)

Tirakanun, S. (2008). *Social science research methods: guidelines for practice* (7th ed.). Bangkok : Chulalongkorn University Printing House. (In Thai)

Wakil, K., Alyari, F., Ghasvari, M., Lesani, Z., & Rajabion, L. (2020). A new model for assessing the role of customer behavior history, product classification, and prices on the success of the recommender systems in e-commerce. *Kybernetes, 49*(5), 1325-1346.

Wang, Y., Jing, Z., & Lyu, J. (2022). Study on Sustainable Development Strategy of Rural E-Commerce in the Northeast of China—A Case Study of 11 Villages, 11 Towns and 4 Counties. *Sustainability, 14*(24), 1-21.

Wiwattananukul, B. (2021). *Get to know B2C e-Commerce business: business model, payment model and how to track online market conditions.* Retrieved May 29, 2023, from https://www.bot.or.th/content/dam/bot/documents/th/research-andpublications/research/stat-horizon-and-stat-in-focus/stat-horizon/B2C_eCommerce.pdf (In Thai)

Xuhua, H., Ocloo, C. E., Tsetse, E., Spio-Kwofie, A., & Musah Abdul-Aziz, I. (2019). The Moderating Role of Organizational Culture on Business to Business Electronic Commerce Adoption in Small and Medium Manufacturing Firms. *International Journal of Management Sciences and Business Research, 8*(3), 10-31.

Yamane, T. (1967). *Statistics: An Introductory Analysis* (2nd ed.). New York : Harper and Row.