

การพัฒนาตัวแบบพยากรณ์การรักษาฐานลูกค้า
ด้วยเทคนิคการเรียนรู้ของเครื่อง: กรณีศึกษาธุรกิจแฟรนไชส์จตุรัสส่งพัสดุ
A Development of Predictive Model for Customer Retention using
Machine Learning Techniques: A Case Study of a Parcel Service Franchise
จามรกุล เหล่าเกียรติกุล¹ และวุดมินท์ บุญโพธิ์²
Jamornkul Laokietkul¹ and Wuttinun Boonpho²

¹ อาจารย์ประจำหลักสูตรระบบสารสนเทศเพื่อธุรกิจดิจิทัล วิทยาลัยบริหารธุรกิจนวัตกรรมและการบัญชี มหาวิทยาลัยธุรกิจบัณฑิต

² กรรมการผู้จัดการ บริษัท หู คีต โซลูชั่น จำกัด

¹ Lecturer of Information System for Digital Business, College of Innovative Business and Accountancy, Dhurakij Pundit University

² Managing Director, two kit solution Company Limited.

E-mail: Jamornkul.lao@dpu.ac.th, wuttinun101@hotmail.com.

Received: Feb. 28,2026; Revised: Apr. 23,2026; Accepted: May. 1,2026

บทคัดย่อ

การศึกษาวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) พัฒนาตัวแบบพยากรณ์ข้อมูลด้วยเทคนิคการเรียนรู้ของเครื่องที่สามารถทำนายการกลับมาใช้บริการ 2) เปรียบเทียบประสิทธิภาพของตัวแบบที่สร้างด้วยเทคนิคการเรียนรู้ของเครื่อง และ 3) วิเคราะห์ปัจจัยสำคัญที่ส่งผลต่อการรักษาฐานลูกค้าในธุรกิจแฟรนไชส์จตุรัสส่งพัสดุ ในส่วนของวิธีดำเนินการวิจัยได้อ้างอิงตามกระบวนการมาตรฐานการทำเหมืองข้อมูล โดยใช้ข้อมูลรายการธุรกรรม (เดือนสิงหาคม 2568 ถึง มกราคม 2569) ข้อมูลถูกจัดเตรียมด้วยการสกัดคุณลักษณะตามกรอบแนวคิดการวิเคราะห์พฤติกรรมลูกค้าด้วย ปัจจัยด้านความใหม่ ความถี่ และ มูลค่าทางการเงิน ร่วมกับปัจจัยการใช้บริการเก็บเงินปลายทาง และนำเทคนิคการสังเคราะห์ข้อมูลรูปแบบตารางด้วยโครงข่ายเชิงปริภูมิ มาใช้ในการปรับสมดุลของข้อมูลในชุดการเรียนรู้ เพื่อยกระดับประสิทธิภาพของตัวแบบพยากรณ์ ก่อนดำเนินการเปรียบเทียบประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่อง คือ เทคนิคการถดถอยโลจิสติก เทคนิคป่าสุ่ม เทคนิคเอ็กซ์จีบูสต์ และ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน ผลการวิจัยตอบวัตถุประสงค์ดังนี้ 1) ผลการพัฒนาตัวแบบพยากรณ์สำเร็จ ได้ตัวแบบที่สามารถจัดความไม่สมดุลในการเรียนรู้ของตัวแบบได้ 2) ผลการเปรียบเทียบพบว่า ตัวแบบเทคนิคเอ็กซ์จีบูสต์ มีประสิทธิภาพในการพยากรณ์สูงที่สุด โดยให้ค่าพื้นที่ใต้กราฟ เท่ากับ 0.8462 ซึ่งมีความเหมาะสมที่สุดในการนำไปประยุกต์ใช้เพื่อพยากรณ์การกลับมาของลูกค้า 3) ผลการวิเคราะห์ปัจจัยสำคัญพบว่า ปัจจัยด้านความใหม่ของการใช้บริการ เป็นตัวแปรที่มีอิทธิพลต่อการพยากรณ์มากที่สุด ตามด้วยสถานะการใช้บริการเก็บเงินปลายทางผลลัพธ์จากงานวิจัยนี้บ่งชี้ว่าการใช้ชุดข้อมูลจากธุรกิจขนาดเล็กสร้างตัวแบบพยากรณ์ สามารถนำตัวแบบนี้ไปใช้เป็นเครื่องมือสนับสนุนการตัดสินใจเพื่อวางแผนกลยุทธ์การรักษาฐานลูกค้าเชิงรุกเพิ่มอัตราการใช้บริการซ้ำ และสร้างความได้เปรียบทางการแข่งขันในธุรกิจขนส่งได้อย่างมีประสิทธิภาพ

คำสำคัญ: การรักษาฐานลูกค้า การเรียนรู้ของเครื่อง การสังเคราะห์ข้อมูลรูปแบบตารางด้วยโครงข่ายเชิงปริภูมิ การวิเคราะห์พฤติกรรมลูกค้า เอ็กซ์จีบูสต์

Abstract

This research aims to 1) develop a predictive model using machine learning techniques to forecast customer retention, 2) compare the performance of various machine learning models, and 3) analyze key determinants influencing customer re-usage behavior in a parcel delivery franchise. The research methodology was conducted based on the Cross-Industry Standard Process for Data Mining (CRISP-DM). The study utilized empirical transaction data from August 2025 to January 2026. Features extraction was based on the Recency, Frequency, and Monetary (RFM) analysis framework, supplemented by the Cash on Delivery (COD) service. The Tabular Generative Adversarial Networks (CTGAN) was employed to synthesize data and mitigate class imbalance, enhancing model performance. Subsequently, four machine learning algorithms as Logistic Regression, Random Forest, XGBoost, and Support Vector Machine (SVM) were evaluated. The findings successfully addressed the research objectives as follows: 1) The predictive model was successfully developed, effectively mitigating the learning bias caused by data imbalance. 2) The comparative analysis demonstrated that the XGBoost model achieved the highest predictive performance, yielding an Area Under the ROC Curve (AUC-ROC) of 0.8462, thereby proving to be the most suitable algorithm for predicting customer return. 3) The feature importance analysis indicated that the Recency of service usage was the most influential predictor, followed by the COD service usage status. The results suggest that leveraging datasets from small businesses to construct predictive models can serve as an effective decision-support tool. Furthermore, this enables businesses to formulate proactive customer retention strategies, increase re-usage rates, and establish a competitive advantage in the logistics industry.

Keywords: Customer Retention, Machine Learning, Data Synthesis, CTGAN, RFM Analysis, XGBoost

บทนำ

ปัจจุบันอุตสาหกรรมโลจิสติกส์และการขนส่งพัสดุในประเทศไทยมีการเติบโตขยายตัวอย่างต่อเนื่อง ตามการเติบโตของพาณิชย์อิเล็กทรอนิกส์ (e-Commerce) ส่งผลให้เกิดการแข่งขันที่รุนแรงทั้งในด้านราคาและการบริการ ธุรกิจแฟรนไชส์จัดส่งพัสดุ (Parcel Service Franchise) จึงเผชิญกับความท้าทายในการรักษาความได้เปรียบทางการแข่งขัน การรักษาฐานลูกค้าเดิม (Customer Retention) จึงกลายเป็นกลยุทธ์สำคัญในเชิงต้นทุนและให้ผลตอบแทนที่ยั่งยืนกว่าการแสวงหาลูกค้าใหม่เพียงอย่างเดียว อย่างไรก็ตาม พฤติกรรมการตัดสินใจกลับมาใช้บริการของลูกค้ามีความซับซ้อนและเปลี่ยนแปลงไปตามปัจจัยต่างๆ เช่น ความรวดเร็วในการให้บริการ ความถี่ในการส่งและรูปแบบการทำธุรกรรม (ณรรฐคุณ วิรุฬห์ศรี และคณะ, 2565) ดังนั้นการวิเคราะห์ข้อมูลธุรกรรม (Transaction Data) เพื่อพยากรณ์ความน่าจะเป็นในการกลับมาใช้บริการ จึงเป็นเครื่องมือสำคัญที่ช่วยให้ผู้ประกอบการรายย่อยสามารถจำแนกกลุ่มลูกค้าตามระดับความเสี่ยงที่จะหายไปของลูกค้า และสามารถนำมาประยุกต์ใช้เพื่อวางแผนกลยุทธ์การตลาดเชิงรุก (Proactive Marketing) ได้อย่างตรงจุด (Sunarya et al., 2024) หากแต่ปัญหาสำคัญที่พบในข้อมูลธุรกิจจริงคือ "ความไม่สมดุลของข้อมูล" (Class Imbalance) เนื่องจากจำนวนลูกค้าที่ไม่ได้กลับมาใช้บริการเป็นประจำมักมีสัดส่วนที่สูงกว่ากลุ่มลูกค้าประจำอย่างมีนัยสำคัญ ส่งผลให้ตัวแบบพยากรณ์พื้นฐานเกิดความลำเอียง (Bias) และอาจไม่มีประสิทธิภาพเท่าที่ควรในระบูกลุ่มลูกค้าเป้าหมาย

จากการทบทวนวรรณกรรมที่เกี่ยวข้อง พบว่ามีงานวิจัยที่ศึกษาวิเคราะห์พฤติกรรมของลูกค้าในธุรกิจต่างๆ อย่างหลากหลาย ดังเช่น งานวิจัยของ ณรรธคุณ วิรุฬห์ศรี และคณะ (2565); กาญจนา ฤทธิรพพงศ์ และปิยมาศ จิตตระ (2561) ที่เป็นเพียงการวิจัยเชิงพรรณนาเพื่อศึกษาวิเคราะห์พฤติกรรมลูกค้าเพื่อจัดกลุ่มเป้าหมายปัจจุบันเท่านั้น มิได้มีการต่อยอดสู่การวิจัยเชิงการพยากรณ์เพื่อคาดการณ์เหตุการณ์ที่น่าจะเกิดขึ้นในอนาคต แม้ว่าจะมีการนำเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) มาประยุกต์ใช้ในวิเคราะห์พฤติกรรมลูกค้าได้รับความนิยมอย่างแพร่หลาย เนื่องจากมีจุดเด่นด้านความสามารถในการจัดการกับข้อมูลที่มีความซับซ้อนสูง ตัวอย่างเช่น งานวิจัยของ Sunarya et al. (2024); Monisha et al. (2020) ที่ประยุกต์ใช้อัลกอริทึมที่หลากหลาย เช่น เทคนิคป่าสุ่ม ในการทำนายพฤติกรรม การซื้อซ้ำและการยกเลิกใช้บริการของลูกค้า รวมถึงการใช้เทคนิค เอกซ์จีบูสต์ และ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน ที่มีประสิทธิภาพสูงในการวิเคราะห์ข้อมูลและจำแนกกลุ่มเป้าหมาย ก็ตาม แต่งานวิจัยส่วนใหญ่ยังคงมุ่งเน้นไปที่การวิเคราะห์ข้อมูลใหญ่ ไม่พบการวิจัยในบริบทการใช้ข้อมูลในธุรกิจขนาดเล็ก ที่มีข้อจำกัดสำคัญคือปริมาณข้อมูลในธุรกิจที่อาจส่งผลต่อความไม่สมดุลของข้อมูล ซึ่งเมื่อพิจารณาในประเด็นการจัดการปัญหาข้อมูลที่มีความไม่สมดุลนี้ พบงานวิจัยของ Xu et al. (2019) ที่ได้คิดค้นเทคนิคการสังเคราะห์ข้อมูล เป็นนวัตกรรมในกลุ่มการเรียนรู้ในเชิงลึก ในการสังเคราะห์ข้อมูลรูปแบบตารางด้วยโครงข่ายเชิงปริภูมิ (Conditional Tabular Generative Adversarial Networks : CTGAN); Moon et al. (2020) ที่นำเทคนิคการสังเคราะห์ข้อมูลดังกล่าว มาประยุกต์ใช้ด้วย ซึ่งวิธีนี้มีจุดเด่นและเหมาะสมต่อการศึกษาคือ สามารถสร้างข้อมูลสังเคราะห์รูปแบบตารางที่รักษารูปแบบและความสัมพันธ์ของข้อมูลจริงเอาไว้ได้อย่างสมจริง อีกทั้งยังช่วยแก้ปัญหาอคติในการเรียนรู้ของตัวแบบ (Model Bias) ป้องกันไม่ให้ตัวแบบจดจำแต่ข้อมูลกลุ่มใหญ่ และช่วยยกระดับความแม่นยำของตัวแบบพยากรณ์ จึงมีความเหมาะสมอย่างยิ่งที่จะนำมาประยุกต์ใช้กับบริบทของกลุ่มธุรกิจขนาดเล็กที่มีข้อจำกัดด้านปริมาณและมีสัดส่วนข้อมูลไม่สมดุล

ด้วยเหตุนี้ ผู้วิจัยจึงมีแนวคิดในการบูรณาการแนวคิดการวิเคราะห์พฤติกรรมลูกค้า (RFM Analysis) ซึ่งเป็นเครื่องมือสำคัญในการสกัดคุณลักษณะเชิงลึกที่สะท้อนพฤติกรรมของลูกค้า เข้ากับเทคนิคการสังเคราะห์ข้อมูล CTGAN เพื่อขจัดข้อจำกัดด้านความไม่สมดุลของข้อมูล (Class Imbalance) การผสมผสานทฤษฎีเชิงพฤติกรรมศาสตร์เข้ากับวิทยาการข้อมูลเชิงลึกในครั้งนี้ ไม่เพียงเป็นการลดความลำเอียงของตัวแบบ (Model Bias) และยกระดับขีดความสามารถของอัลกอริทึมการเรียนรู้ของเครื่องให้สามารถเรียนรู้และจำแนกรูปแบบพฤติกรรมที่ซับซ้อนได้อย่างแม่นยำยิ่งขึ้นเท่านั้น ผลลัพธ์ที่ได้จากการวิจัยนี้ ยังเป็นการสร้างองค์ความรู้ใหม่ที่พิสูจน์ให้เห็นว่า การนำเทคโนโลยีสารสนเทศมาประยุกต์ใช้อย่างเป็นระบบ สามารถเปลี่ยนข้อมูลธุรกรรมระดับรายย่อยให้กลายเป็นเครื่องมือเชิงกลยุทธ์ ช่วยให้ผู้ประกอบการสามารถวางแผนการตลาดเชิงรุกเพื่อรักษาลูกค้า และสร้างความได้เปรียบทางการแข่งขันในอุตสาหกรรมขนส่งพัสดุได้อย่างเป็นรูปธรรม

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาตัวแบบพยากรณ์ข้อมูลด้วยเทคนิคการเรียนรู้ของเครื่องที่สามารถทำนายการกลับมาใช้บริการ
2. เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบที่สร้างด้วยอัลกอริทึมการเรียนรู้ของเครื่อง
3. เพื่อวิเคราะห์ปัจจัยสำคัญที่ส่งผลต่อการรักษาลูกค้าในธุรกิจแฟรนไชส์จัดรับส่งพัสดุ

แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

ในการสร้างตัวแบบพยากรณ์การรักษาลูกค้า กรณีศึกษาธุรกิจแฟรนไชส์จัดรับส่งพัสดุ ในครั้งนี้เป็นการบูรณาการทฤษฎีเชิงพฤติกรรมศาสตร์เข้ากับศาสตร์วิทยาการข้อมูล ซึ่งผู้วิจัยได้ทบทวนทฤษฎีและวรรณกรรมที่เกี่ยวข้อง ได้แก่

1) เทคนิคการวิเคราะห์พฤติกรรมลูกค้า

เทคนิคการวิเคราะห์พฤติกรรมลูกค้า (Recency Frequency Monetary Analysis: RFM) คือ เทคนิคด้านการตลาดในการวิเคราะห์ข้อมูลจัดกลุ่มลูกค้าตามพฤติกรรม การซื้อสินค้าหรือการใช้บริการ (Customer Segmentation) (ชาญชววัฒน์ ภัคศิริ, 2566) ซึ่งเป็นวิธีที่ช่วยประเมินความจงรักภักดีของลูกค้า โดยพิจารณาจากปัจจัยด้านระยะเวลาล่าสุดที่มีปฏิสัมพันธ์ ความถี่ในการปฏิสัมพันธ์ และ มูลค่าการทำธุรกรรม (Rizkyanto & Gaol, 2023) โดยพิจารณาจากพฤติกรรมการซื้อหรือการใช้บริการในอดีต ผ่านตัวชี้วัดสำคัญ 3 องค์ประกอบหลัก อันได้แก่ (1) Recency: R คือ ระยะเวลาที่ไม่ได้มาใช้บริการ คำนวณได้จากระยะห่างของเวลาที่ใช้บริการครั้งล่าสุดกับครั้งก่อนหน้า ซึ่งยิ่งค่าระยะห่างน้อย หมายถึง ยิ่งมีโอกาสกลับมาสูง (2) Frequency: F คือ ความถี่ในการใช้บริการ ที่บ่งบอกถึงความจงรักภักดี (Loyalty) และ (3) Monetary: M คือ มูลค่ารวมที่ลูกค้าใช้จ่าย สะท้อนถึงมูลค่าตลอดช่วงชีวิตของลูกค้า (Customer Lifetime Value) (กาญจนา หุทธรพงษ์ และปิยมาศ จิตตระ, 2561) โดยทั้ง 3 องค์ประกอบหลักดังกล่าวของเทคนิคการวิเคราะห์พฤติกรรมลูกค้านี้ ช่วยให้ธุรกิจเข้าใจพฤติกรรมลูกค้าได้ชัดเจนขึ้นและสามารถวางกลยุทธ์การตลาดที่เหมาะสมกับลูกค้าแต่ละกลุ่มได้อย่างมีประสิทธิภาพ (ปิ่นดาเรีย สุทรวราภาส และคณะ, 2563) ทั้งนี้ ในการประยุกต์ใช้ RFM ในการทำเหมืองข้อมูลนั้น เป็นการนำเอาองค์ประกอบทั้ง 3 มาใช้เป็น "คุณลักษณะพื้นฐาน" (Baseline Features) ที่มีนัยสำคัญสูงในการพยากรณ์พฤติกรรมในอนาคต โดยเฉพาะในธุรกิจบริการที่มีการทำธุรกรรมบ่อยครั้ง

ในขณะที่เมื่อพิจารณาการรักษาฐานลูกค้า (Customer Retention) ที่หมายถึง ความสามารถของธุรกิจในการรักษาความสัมพันธ์กับลูกค้าเดิมในช่วงเวลาที่กำหนด ซึ่งจากการศึกษาแนวคิด และงานวิจัยที่เกี่ยวข้องในอดีต ยืนยันว่าการรักษาลูกค้าเดิมมีต้นทุนที่ต่ำกว่าการแสวงหาลูกค้าใหม่ (Customer Acquisition) 5-25 เท่า (Ang & Buttle, 2006) โดยปัจจัยหลักที่ส่งผลต่อการซื้อซ้ำ (Repurchase Intention) ในธุรกิจบริการโลจิสติกส์ ประกอบด้วย ความสะดวกของสถานที่ (Convenience), ความรวดเร็วในการให้บริการ และความเชื่อมั่นในตราสินค้า (ณรรฐคุณ วิรุฬห์ศรี และคณะ, 2565)

2) การทำเหมืองข้อมูลจำแนกประเภท

การทำเหมืองข้อมูลจำแนกประเภท (Classification Mining) จัดเป็นการวิเคราะห์ข้อมูลแบบมีข้อมูลผู้สอน (Supervise Training) ที่มีวัตถุประสงค์เพื่อวิเคราะห์และจำแนกชุดข้อมูลออกเป็นกลุ่มได้ตามความต้องการ ในส่วนกระบวนการทำเหมืองข้อมูลจำแนกประเภทนั้น จะแบ่งชุดข้อมูลเป็น 2 ส่วน ได้แก่ ชุดข้อมูลสำหรับการเรียนรู้ (Training Data Set) และ ชุดข้อมูลสำหรับการทดสอบ (Testing Data Set) นำมาเข้าสู่ขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ที่เหมาะสม เพื่อให้ได้ตัวแบบในการจำแนกกลุ่มข้อมูล และทำการทดสอบประเมินเพื่อวัดประสิทธิภาพของตัวแบบที่ได้ตามลำดับ (จามรกุล เหล่าเกียรติกุล, 2558)

โดยในการวิจัยครั้งนี้เป็นทำเหมืองข้อมูลจำแนกประเภทเพื่อมุ่งสร้างตัวแบบสำหรับพยากรณ์ความน่าจะเป็นในการกลับมาใช้บริการของลูกค้า ที่เปรียบเทียบเทคนิคการเรียนรู้ของเครื่อง 4 เทคนิค ได้แก่ เทคนิคการถดถอยโลจิสติก (Logistic Regression) เทคนิคป่าสุ่ม (Random Forest) เทคนิคเอกซ์จีบูสต์ (XGBoost) และ เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) โดยสรุปลักษณะสำคัญในลักษณะ จุดเด่น และข้อจำกัดเป็นตารางเปรียบเทียบเทคนิคการเรียนรู้ของเครื่อง ทั้ง 4 วิธี ได้ดังนี้

ตารางที่ 1 เปรียบเทียบเทคนิคการเรียนรู้ของเครื่องทั้ง 4 วิธี

ลักษณะ	Logistic Regression	Random Forest	Gradient Boosting	SVM
จุดเด่น	เป็นรูปแบบที่เข้าใจง่าย, ทำงานได้รวดเร็ว เหมาะกับข้อมูลที่มีรูปแบบความสัมพันธ์เชิงเส้น (Sunarya et al., 2024)	มีความแม่นยำสูง ช่วยลดปัญหา Overfitting สามารถบ่งชี้ความสำคัญของคุณลักษณะข้อมูล (ตัวแปร) ได้	มีการเรียนรู้เร็ว มีประสิทธิภาพสูง และลดปัญหาการ Overfitting	เหมาะกับการวิเคราะห์ข้อมูลที่มีหลายมิติ สามารถจัดการข้อมูลไม่เป็นเชิงเส้นได้ดี
ข้อจำกัด	ไม่เหมาะสมกับข้อมูลที่มีความซับซ้อน มีความอ่อนไหวต่อค่าผิดปกติ (Outliers) (Sunarya et al., 2024)	ใช้ทรัพยากรระบบสูง ตีความผลลัพธ์ได้ยากกว่าเทคนิคพื้นฐานอื่นๆ (Meric & Ozer, 2022; Sunarya et al., 2024)	มีความซับซ้อนยุ่งยากในการปรับตั้งค่าต่างๆ เพื่อใช้งาน (Korstanje, 2021)	ไม่เหมาะสมกับข้อมูลที่มีขนาดใหญ่ (Big Data) และไม่เหมาะกับข้อมูลที่มีความผิดพลาด (Noisy data) (Gundogdu, 2022; วิราภานต์ กิตติวรกุล และคณะ, 2565)

จากการศึกษาและเปรียบเทียบคุณลักษณะของเทคนิคการเรียนรู้ของเครื่อง ในการทำเหมืองข้อมูลจำแนกประเภท จะเห็นได้ว่าแต่ละวิธีการมีจุดเด่นและข้อจำกัดแตกต่างกัน ทั้งนี้ขึ้นอยู่กับสภาพการณ์ของปัญหา และลักษณะของข้อมูลอันจะส่งผลต่อความเหมาะสมในการเลือกใช้เทคนิคเพื่อนำไปประยุกต์ใช้ในการแก้ปัญหา

3) เทคนิคการสังเคราะห์ข้อมูลรูปแบบตารางด้วยโครงข่ายเชิงปริภูมิ

เทคนิคการสังเคราะห์ข้อมูลรูปแบบตารางด้วยโครงข่ายเชิงปริภูมิ (Conditional Tabular Generative Adversarial Networks : CTGAN) ที่ถูกคิดค้นพัฒนาขึ้นและเผยแพร่ในปี 2019 (Xu et al., 2019) จัดเป็นเทคนิคการเรียนรู้เชิงลึก (Deep Learning) สำหรับการสร้างข้อมูลสังเคราะห์แบบตาราง (Synthetic Tabular Data) ที่มีความสมจริงและมีคุณภาพสูง มักประกอบด้วยตัวแปรที่หลากหลาย มีความซับซ้อนและมีชนิดข้อมูลแบบผสม (Mixed data types) ทั้งข้อมูลตัวเลขต่อเนื่องและข้อมูลประเภทจัดกลุ่ม เทคนิคการสังเคราะห์ข้อมูล CTGAN จึงถูกพัฒนาขึ้นเพื่อจัดการกับปัญหานี้โดยเฉพาะ โดยมีรากฐานพัฒนาต่อยอดมาจากสถาปัตยกรรม การสังเคราะห์ข้อมูลแบบโครงข่าย (Generative Adversarial Networks: GAN) (Kim & Seok, 2024) เทคนิค CTGAN ทำงานโดยการบูรณาการ 2 ส่วน ได้แก่ 1) ตัวสร้างข้อมูล (Generator) ที่ทำหน้าที่เรียนรู้รูปแบบข้อมูลและสร้างข้อมูลจำลองขึ้นมา และ 2) ตัวจำแนก (Discriminator) ที่คอยทำหน้าที่จำแนก ว่าข้อมูลนั้นเป็นข้อมูลจริงหรือข้อมูลจำลอง และกระบวนการเรียนรู้สังเคราะห์ข้อมูลจะสิ้นสุดเมื่อตัวสร้างสามารถสร้างข้อมูลที่เสมือนจริง จนตัวจำแนกไม่สามารถแยกแยะได้

อาจกล่าวได้ว่าเทคนิคการสังเคราะห์ข้อมูล CTGAN นี้ จัดเป็นกลไกที่สร้างขึ้นเพื่อแก้ปัญหาข้อมูลหมวดหมู่ที่ไม่สมดุล (Phanbua et al., 2025) โดยใช้การสุ่มตัวอย่างแบบมีเงื่อนไข เพื่อเพิ่มประสิทธิภาพให้กับตัวแบบ ในกรณีที่มีชุดข้อมูลอย่างจำกัด หรือต้องการจำลองตัวอย่างข้อมูลของเหตุการณ์ที่เกิดขึ้นได้ยาก (Rare events) เป็นการสร้างความสมดุลของชุดข้อมูล โดยกำหนดให้ตัวแบบมีโอกาสได้เรียนรู้และสร้างข้อมูลของหมวดหมู่ชุดข้อมูลที่เป็น "กลุ่มน้อย" ในสัดส่วนที่เหมาะสม เพื่อป้องกันไม่ให้เกิดตัวแบบจดจำแต่ข้อมูลกลุ่มใหญ่เพียงอย่างเดียว นอกจากนี้ การเรียนรู้การกระจายตัวของข้อมูลด้วย CTGAN ยังช่วยรักษาความสัมพันธ์ (Correlation) ระหว่างคุณลักษณะต่างๆ ของข้อมูลไว้ได้อย่างครบถ้วน เทคนิคการสังเคราะห์ข้อมูลนี้ จึงถูกนำมาประยุกต์ใช้ในการเพิ่มปริมาณข้อมูล (Data Augmentation) ถือเป็นเทคโนโลยีขั้นสูงที่ช่วยให้นักวิจัยสามารถจำลองข้อมูลรูปแบบตารางที่มีความซับซ้อนออกมาได้อย่างแม่นยำ และสามารถนำไปใช้งานทดแทนข้อมูลจริงได้อย่างมีประสิทธิภาพ (Sinchana et al., 2025)

4) งานวิจัยที่เกี่ยวข้อง

จากการค้นคว้าทบทวนงานวิจัยที่เกี่ยวข้อง ในด้านการวิเคราะห์พฤติกรรมลูกค้าหรือผู้ใช้บริการ โดยใช้เทคนิคการวิเคราะห์พฤติกรรมลูกค้า (Recency Frequency Monetary Analysis: RFM) ได้แก่ งานวิจัยของกาญจนา หลุณรชพงศ์ และปิยมาศ จิตตระ (2561) ที่ประยุกต์วิธีการวิเคราะห์ RFM ในการแบ่งกลุ่มผู้ใช้งานเครื่องพิมพ์และเครื่องถ่ายเอกสารในองค์กร กรณีศึกษาสำนักวิชาสารสนเทศศาสตร์ มหาวิทยาลัยวลัยลักษณ์ ซึ่งพบว่าสามารถจำแนกกลุ่มพฤติกรรมผู้ใช้งานได้ 14 กลุ่ม หรือ งานวิจัยของ ปันดารีย์ สุนทรวราภาส และคณะ (2563) ที่ประยุกต์ใช้การวิเคราะห์แบบ RFM ในการแบ่งกลุ่มผู้ใช้งานตามพฤติกรรมการยืมทรัพยากรสารสนเทศของสำนักทรัพยากรการเรียนรู้คุณหญิงหลง อรรถกระวีสุนทร มหาวิทยาลัยสงขลานครินทร์ ซึ่งผลจากการแบ่งกลุ่มข้อมูลสามารถแบ่งได้ 18 กลุ่ม ซึ่งผลจากการศึกษาทั้ง 2 งานวิจัยข้างต้นแสดงให้เห็นว่า ปัจจัยทั้ง 3 ของเทคนิคการวิเคราะห์พฤติกรรมลูกค้า อันได้แก่ Recency Frequency และ Monetary เป็นปัจจัยสำคัญที่สามารถบ่งชี้พฤติกรรม และสามารถใช้เป็นองค์ประกอบในการจำแนกข้อมูลได้ ดังนั้น จึงสมควรนำมาเป็นองค์ประกอบของข้อมูลในการสร้างตัวแบบเพื่อพยากรณ์ความน่าจะเป็นในการกลับมาใช้บริการของลูกค้า

ในด้านงานวิจัยในการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่อง เพื่อจำแนกกลุ่มข้อมูล และ การวิเคราะห์พฤติกรรมลูกค้า พบนงานวิจัยของ Sunaraya (2024) ที่ศึกษาเปรียบเทียบตัวแบบ Logistic Regression และ Random Forest ในการทำนายพฤติกรรมลูกค้าอีคอมเมิร์ซ เพื่อทำนายการยกเลิกใช้บริการของลูกค้า (Customer Churn)

ซึ่งผลการวิจัยพบว่า เทคนิค Random Forest มีประสิทธิภาพการทำนายที่ดีกว่า ในขณะที่ งานวิจัยของชาญวัฒน์ ภาคศิษฐ์ (2566) ที่ศึกษาพฤติกรรมลูกค้าในธุรกิจอีคอมเมิร์ซ เพื่อจำแนกกลุ่มลูกค้าออกเป็น 3 กลุ่ม โดยใช้ RFM ร่วมกับเทคนิค K-Means และ DBSCAN Clustering ซึ่งผลการวิจัยพบว่า K-Means ให้ผลการทำนายจำแนกกลุ่มที่มีประสิทธิภาพดีกว่า

จากการศึกษาแนวคิด ทฤษฎี และ งานวิจัยข้างต้น แม้ว่าจะมีงานวิจัยที่ศึกษาพฤติกรรมลูกค้าในบริบทของธุรกิจขนาดใหญ่ที่มีข้อมูลสมบูรณ์ ในลักษณะภาพรวม แต่ยังไม่พบการศึกษาเฉพาะเจาะจงในระดับผู้ประกอบการรายย่อย และมักหยุดอยู่เพียงการใช้ RFM เพื่อวิเคราะห์ข้อมูลในอดีต แนวคิดของการวิจัยนี้จึงได้นำเสนอ การบูรณาการเทคนิคการวิเคราะห์พฤติกรรมด้านการตลาดเข้ากับศาสตร์ด้านวิทยาการข้อมูล เพื่อยกระดับผู้ประกอบการธุรกิจรายย่อยและต่อยอดการวิเคราะห์ข้อมูลเป็นตัวแบบเพื่อพยากรณ์ข้อมูล โดยประยุกต์ใช้ตัวแปร RFM ให้เป็นคุณลักษณะนำเข้าป้อนสู่อัลกอริทึมพยากรณ์ (Predictive Model) แทนการจัดกลุ่มข้อมูลแบบเดิม และประยุกต์ใช้เทคนิคการสร้างเครื่องจำลอง CTGAN ซึ่งเป็นนวัตกรรมการเรียนรู้เชิงลึกเข้ามาสังเคราะห์และปรับสมดุลข้อมูล ซึ่งเป็นข้อจำกัดสำคัญของธุรกิจขนาดเล็ก โดยเชื่อมั่นว่าการบูรณาการเทคนิคเหล่านี้ช่วยสร้างตัวแบบที่สามารถระบุกลุ่มลูกค้าเสี่ยงที่จะสูญหายได้ และเป็นเครื่องมือให้ผู้ประกอบการใช้ประเมินสถานการณ์เพื่อทำกลยุทธ์การตลาดเชิงรุก (Proactive Marketing) เช่น การมอบส่วนลดพิเศษติดตามลูกค้ากลุ่มสำคัญได้อย่างทันท่วงที แทนการบริหารงานแบบใช้สัญชาตญาณเพียงอย่างเดียว

ทั้งนี้ผู้วิจัยจึงได้สรุปเปรียบเทียบความแตกต่างและจุดเด่นของงานวิจัย เทียบกับงานวิจัยที่เกี่ยวข้องในอดีต ดังแสดงในตารางที่ 2

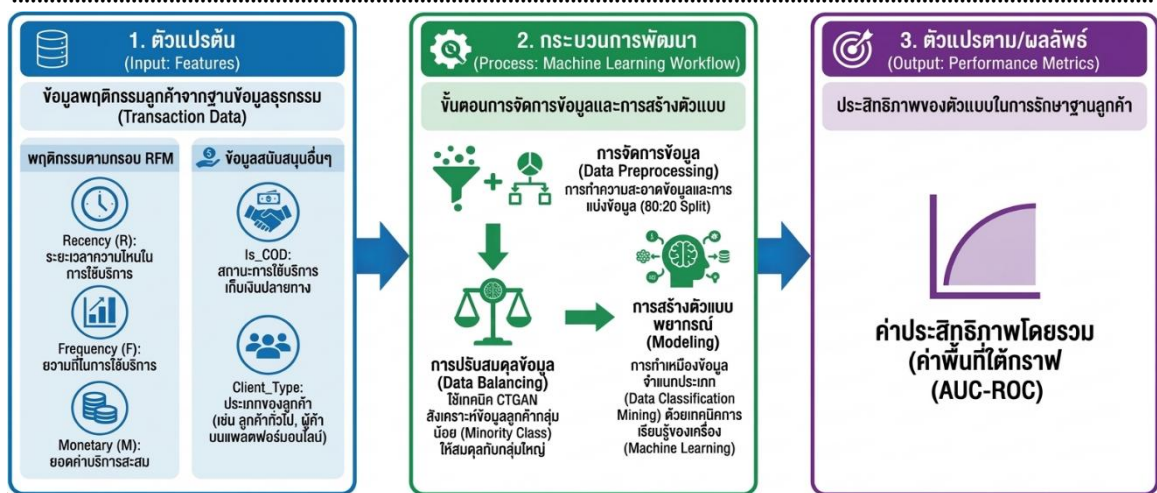
ตารางที่ 2 เปรียบเทียบความแตกต่างของงานวิจัย

งานวิจัย / ปีที่ตีพิมพ์	บริบทธุรกิจ รายย่อย	ประยุกต์ใช้ เทคนิคการวิเคราะห์ พฤติกรรมลูกค้า (RFM)	การทำเหมืองข้อมูล และสร้างตัวแบบพยากรณ์ด้วย เทคนิคการเรียนรู้ของเครื่อง	การจัดการข้อมูลไม่สมดุล เทคนิคการสังเคราะห์ ข้อมูล (CTGAN)
กาญจนา หฤธรรมพงศ์ และ ปิยะมาศ จิตตระ. (2561)	✗	✓	✗	✗
ปิ่นดาเรีย สุนทรวราภาส และคณะ (2563)	✗	✓	✗	✗
ชาญวัฒน์ ภัคศิริ (2566)	✗	✓	✗	✗
Sunarya et al. (2024)	✗	✗	✓	✗
งานวิจัยนี้	✓	✓	✓	✓

กรอบแนวคิดในการวิจัย

งานวิจัยนี้ เป็นการศึกษาการใช้ประโยชน์ข้อมูลในธุรกิจขนาดเล็กเพื่อพยากรณ์ความน่าจะเป็นในการรักษาสถานลูกค้า โดยมีขอบเขตของการวิจัย ใช้ข้อมูลรายการธุรกรรมจริงจากร้านแฟรนไชส์จตุรัสแห่งหนึ่ง ย้อนหลังระหว่างเดือนสิงหาคม 2568 ถึง มกราคม 2569 รวมระยะเวลา 6 เดือน คัดเลือกเฉพาะคุณลักษณะข้อมูลที่เกี่ยวข้องกับการจัดส่งพัสดุเฉพาะที่จำเป็น โดยไม่ระบุและไม่เชื่อมโยงเกี่ยวข้องกับข้อมูลส่วนบุคคล และประยุกต์เทคนิคการวิเคราะห์พฤติกรรมลูกค้า 3 ปัจจัย ได้แก่ ปัจจัยด้านความใหม่ (Recency) ปัจจัยด้านความถี่ (Frequency) และปัจจัยด้านมูลค่าทางการเงิน (Monetary) มาพิจารณาร่วมกับ ข้อมูลประเภทของลูกค้า และรูปแบบการชำระค่าบริการ นำมาใช้เป็นตัวแปรสำหรับพยากรณ์ และ กำหนดตัวแปรเป้าหมาย คือ สถานะของการกลับมาใช้บริการ โดยกำหนดเป็นช่วงระยะเวลา 30 วัน (โดยกำหนดให้ 0 = ไม่กลับมา, 1 = กลับมา)

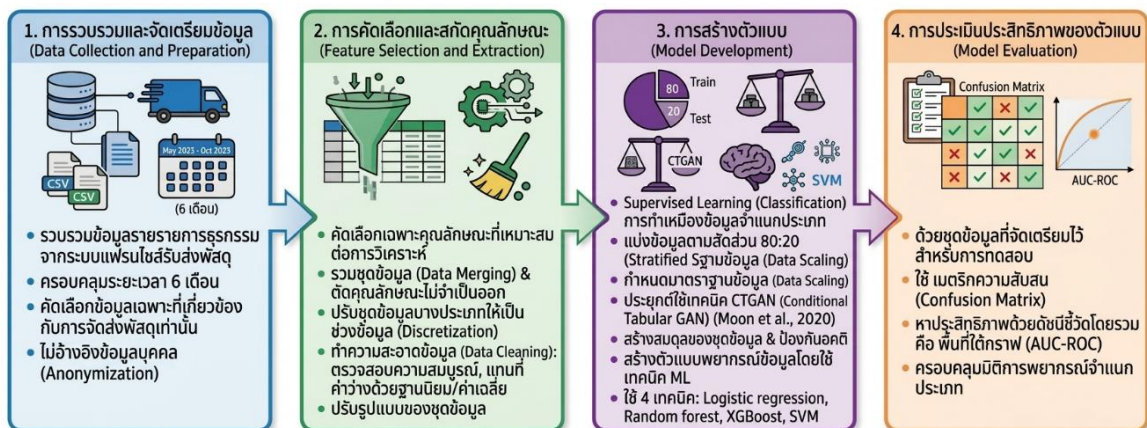
ในส่วนของกรอบแนวคิด (Conceptual Framework) ของการวิจัยครั้งนี้ เป็นการมุ่งวิเคราะห์ทำเหมืองข้อมูลจำแนกประเภท ด้วยเทคนิคการเรียนรู้ของเครื่อง เพื่อสร้างตัวแบบพยากรณ์ (Predictive Model) และแก้ไขปัญหาความสมดุลของข้อมูลในข้อมูลธุรกิจจริง ด้วยการผสมผสานเทคนิค CTGAN เข้าไปในส่วนของกระบวนการ เพื่อลดข้อจำกัดของข้อมูลที่ไม่สมดุลให้มีประสิทธิภาพที่แม่นยำกว่าการพัฒนาตัวแบบ โดยมีข้อมูลพฤติกรรมของลูกค้าในอดีตจากการประยุกต์ใช้เทคนิคการวิเคราะห์พฤติกรรมลูกค้า (RFM) และ การใช้บริการเก็บเงินปลายทาง รวมถึง ประเภทของลูกค้า เป็น ตัวแปรต้น (Input Features) และ ประเมินผลลัพธ์จากตัวแบบพยากรณ์ ด้วยค่าดัชนีวัดประสิทธิภาพของตัวแบบโดยรวม คือ ค่าพื้นที่ใต้กราฟ (AUC-ROC) เป็นตัวแปรตาม (Output Performance Metrics) จากกรอบแนวคิดข้างต้น จึงสร้างเป็นแผนภาพเพื่อแสดงกรอบแนวคิดของงานวิจัย ดังรูปที่ 1



รูปที่ 1 แผนภาพเพื่อแสดงกรอบแนวคิดการวิจัย

วิธีดำเนินการวิจัย

การวิจัยในครั้งนี้ ผู้วิจัยดำเนินการโดยประยุกต์หลักการวิธีการทำเหมืองข้อมูล อ้างอิงตามกระบวนการมาตรฐาน CRISP-DM (Cross-Industry Standard Process for Data Mining) ซึ่งเป็นขั้นตอนมาตรฐาน ดังปรากฏตามหลักการทฤษฎีในการทำเหมืองข้อมูล (จามรกุล เหล่าเกียรติกุล, 2558) โดยเริ่มจากการทำความเข้าใจในธุรกิจแฟรนไชส์รับส่งพัสดุ ที่มีความต้องการในการใช้ประโยชน์ข้อมูลการใช้บริการของลูกค้าธุรกิจแฟรนไชส์รับส่งพัสดุ เพื่อคาดการณ์การกลับมาใช้บริการของลูกค้า และดำเนินการวิจัย โดยแบ่งเป็นขั้นตอน การรวบรวมและจัดเตรียมข้อมูลให้อยู่ในรูปแบบที่เหมาะสม จากนั้นจึงทำการคัดเลือกและสกัดคุณลักษณะ นำมาทำการสร้างตัวแบบสำหรับพยากรณ์ข้อมูล และการประเมินประสิทธิภาพของตัวแบบ ทั้งนี้เพื่อความเข้าใจจึงสร้างแผนภาพกราฟิกแสดงขั้นตอนการดำเนินการวิจัย ดังแสดงในรูปที่ 2



รูปที่ 2 แผนภาพกราฟิกเพื่อแสดงขั้นตอนการดำเนินการวิจัย

1) การรวบรวมและจัดเตรียมข้อมูล

ในส่วนกระบวนการรวบรวมและจัดเตรียมข้อมูล (Data Collection and Preparation) ถือเป็นขั้นตอนสำคัญในการวิจัย เพื่อให้ได้ชุดข้อมูลที่ถูกต้อง แม่นยำ และพร้อมสำหรับกระบวนการเรียนรู้ของเครื่อง โดยทำการรวบรวมข้อมูลรายการธุรกรรม (Transaction Data)จากระบบบริหารจัดการร้านแฟรนไชส์รับส่งพัสดุ โดยครอบคลุมระยะเวลา 6 เดือน (สิงหาคม 2568 - มกราคม 2569) ซึ่งข้อมูลถูกรวบรวมและจัดเก็บในรูปแบบไฟล์ CSV รายเดือน ทำการคัดเลือกใช้คุณลักษณะข้อมูลเฉพาะที่เกี่ยวข้องกับการจัดส่งพัสดุนั้น โดยไม่อ้างอิงเกี่ยวข้องกับข้อมูลบุคคล โดยมีรายละเอียดชุดข้อมูลรายการธุรกรรมในตารางที่ 2 และมีรายการข้อมูลรวม 6 เดือน รวมทั้งสิ้น 6,011 ระเบียบข้อมูล

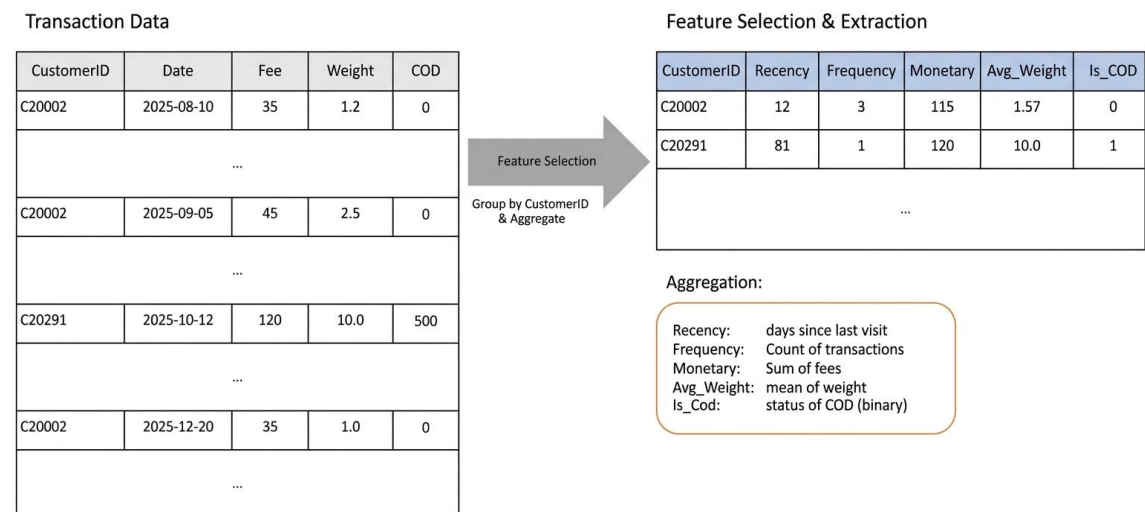
ตารางที่ 3 รายละเอียดคุณลักษณะที่คัดเลือกจากชุดข้อมูลรายการธุรกรรมของแฟรนไชส์รับส่งพัสดุ

ลำดับ	ชื่อคอลัมน์ (คุณลักษณะ)	รายละเอียดการจัดเก็บ
1.	รหัสผู้ส่ง	หมายเลขประจำตัวผู้ส่ง เพื่อใช้ในการจัดกลุ่มพัสดุ
2.	เวลารับพัสดุ	วันที่ และ เวลา ที่รับพัสดุเข้าระบบ
3.	เลขพัสดุ	เลขประจำพัสดุ
4.	ประเภทลูกค้า	ลูกค้าทั่วไป / ลูกค้าแพลตฟอร์มออนไลน์
5.	น้ำหนัก	น้ำหนักพัสดุ
6.	ขนาด	ขนาดพัสดุ กว้างxยาวxสูง
7.	COD	การเรียกเก็บเงินปลายทาง
8.	รูปแบบการชำระเงิน	การชำระเงินโดยผู้ส่ง/ผู้รับ

2) การคัดเลือกและสกัดคุณลักษณะ

ทำการคัดเลือกและสกัดคุณลักษณะ (Feature Selection and Extraction) จากข้อมูลที่ได้ทำการจัดเตรียม นำมาทำการคัดเลือกเฉพาะคอลัมน์ หรือ คุณลักษณะข้อมูล ที่เหมาะสมต่อการวิเคราะห์ข้อมูล ดำเนินการรวมชุดข้อมูล (Data Merging) โดยใช้โครงสร้างฐานข้อมูลที่มีคุณลักษณะ (Attributes) เดียวกัน เพื่อสร้างชุดข้อมูลรวม (Global Dataset) โดยพิจารณาจัดกลุ่มพัสดุโดยใช้ข้อมูลรหัสลูกค้า (CustomerID) เป็นคีย์หลัก ในการเชื่อมโยงพฤติกรรมลูกค้าข้ามช่วงเวลา ดังรูปที่ 2 และตัดคุณลักษณะที่ไม่จำเป็นออก ทำการปรับชุดข้อมูลบางประเภทให้เป็นช่วงข้อมูล (Discretization) เพื่อให้ง่ายต่อการวิเคราะห์ จากนั้นทำความสะอาดข้อมูล (Data Cleaning) โดยดำเนินการตรวจสอบความสมบูรณ์ของข้อมูล (Data Integrity) ทั้งการตรวจสอบค่าว่าง และแทนที่ด้วยฐานนิยม หรือ ค่าเฉลี่ยตามความเหมาะสมของประเภทข้อมูล และทำการปรับรูปแบบของชุดข้อมูลเพื่อความเหมาะสมต่อการสร้างตัวแบบ เช่น ปรับข้อมูลน้ำหนักเฉลี่ยของพัสดุที่ลูกค้านำมาส่ง ที่บ่งบอกถึงประเภทสินค้า การแปลงข้อมูลประเภทบัญชีลูกค้าให้จำแนกเป็น ลูกค้าแพลตฟอร์มร้านค้าออนไลน์ และ ลูกค้าทั่วไป ทำการแปลงข้อมูลสถานะการชำระเงินปลายทางไปเป็น Binary คือ 0 และ 1 เพื่อให้ง่ายต่อการวิเคราะห์ข้อมูล

จากนั้นทำการสกัดคุณลักษณะที่เกี่ยวข้อง (Feature Engineering) โดยในกระบวนการวิจัยนี้ ผู้วิจัยได้เปลี่ยนหน่วยของข้อมูลจาก "รายการธุรกรรมพัสดุ" (Transactional Level) ให้กลายเป็น "คุณลักษณะรายบุคคล" (Customer Level) เพื่อให้สอดคล้องกับเป้าหมายการพยากรณ์พฤติกรรมลูกค้า โดยประยุกต์ใช้กรอบแนวคิด RFM ซึ่งเป็นมาตรฐานในการวิเคราะห์พฤติกรรมลูกค้า สกัดเป็นคุณลักษณะเพิ่มเติม 3 คุณลักษณะ เพื่อสร้างเป็นคุณลักษณะเชิงพฤติกรรมที่สะท้อนถึงความผูกพันของลูกค้าต่อธุรกิจแฟรนไชส์รับส่งพัสดุ และ ใช้วิธีการ Time-Slicing ในการกำหนดตัวแปรตาม (Dependent Variable) อันเป็นตัวแปรเป้าหมายคือ สถานะการกลับมาใช้บริการ โดยกำหนดให้ 1 หมายถึง การกลับมาใช้บริการซ้ำ (ภายในระยะเวลา 30 วัน) และ 0 หมายถึง ไม่กลับมาใช้บริการซ้ำ



รูปที่ 3 ภาพแสดงตัวอย่างการจัดเตรียมข้อมูลให้อยู่ในรูปแบบที่เหมาะสมต่อการวิเคราะห์ข้อมูล

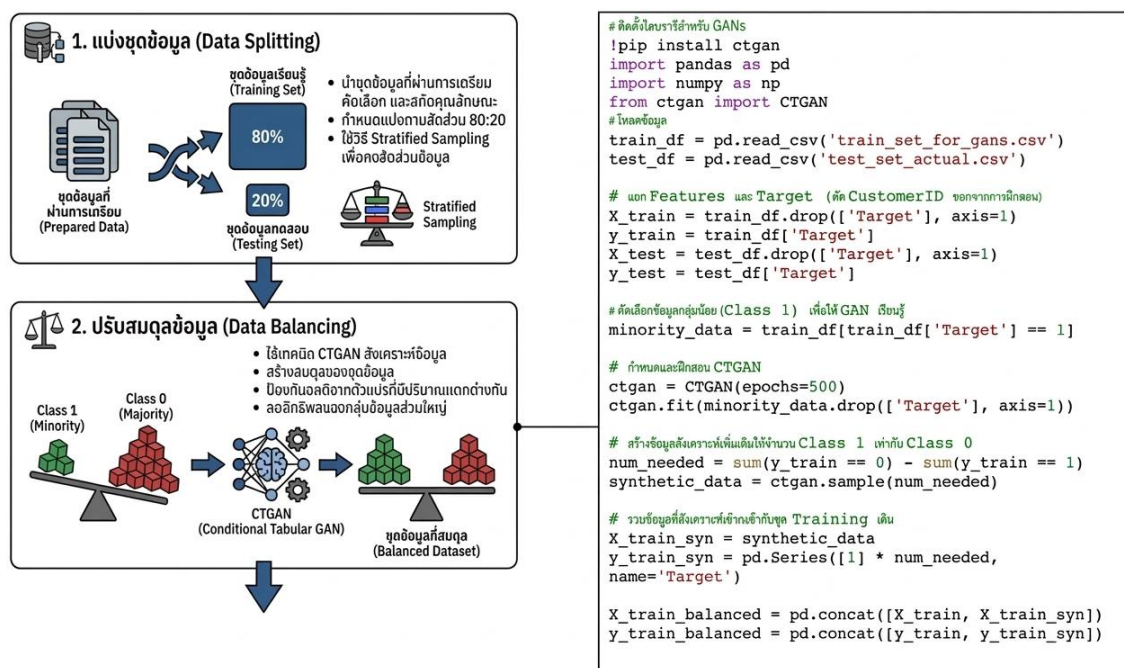
จากการคัดเลือกคุณลักษณะและสกัดคุณลักษณะเพิ่มเติมข้างต้น ทำให้ได้คุณลักษณะของชุดข้อมูลที่ผ่านขั้นตอนการจัดเตรียมข้อมูล ดังปรากฏในตารางที่ 4

ตารางที่ 4 คุณลักษณะข้อมูลเพิ่มเติมที่ได้จากสกัดคุณลักษณะที่เกี่ยวข้อง (Feature Engineering)

ชื่อคุณลักษณะ (Features)	ประเภทข้อมูล	คำอธิบาย
Recency	Integer	จำนวนวันนับจากครั้งล่าสุดที่มาใช้บริการ
Frequency	Integer	จำนวนครั้งรวมที่มาใช้บริการในรอบ 5 เดือน
Monetary	Float	ยอดค่าใช้จ่ายสะสมทั้งหมดของลูกค้า
Avg_Weight	Float	น้ำหนักพัสดุเฉลี่ยต่อการส่ง 1 ครั้ง
Is_COD_User	Binary	สถานะการใช้บริการเก็บเงินปลายทาง (0, 1)
Client_Type	Categorical	ประเภทลูกค้า (General_Client: ลูกค้าทั่วไป General_Client, Platform Client: ลูกค้าแพลตฟอร์ม Shopee, Tiktok, Lazada)
Target (Y)	Binary	สถานะการกลับมาใช้บริการซ้ำภายใน 30 วัน

3) การสร้างตัวแบบ

สำหรับขั้นตอนการสร้างตัวแบบ (Model Development) ของการวิจัยนี้ เป็นการทำให้ข้อมูลจำแนกประเภทแบบมีผู้สอน (Supervised Learning) ซึ่งต้องใช้ชุดข้อมูลสำหรับการเรียนรู้ในการสร้างตัวแบบ จากขั้นตอนการจัดเตรียมข้อมูล การคัดเลือก และ สกัดคุณลักษณะข้อมูล ในขั้นตอนก่อนหน้า นำมากำหนดชุดข้อมูลแบ่งเป็น 2 ส่วน ตามมาตรฐานข้อมูล (Data Scaling) โดยกำหนดชุดข้อมูลตามสัดส่วน 80:20 ด้วยวิธี Stratified Sampling โดยปัญหาที่พบคือความไม่สมดุลของข้อมูล ซึ่งกลุ่มลูกค้าที่กลับมาใช้บริการซ้ำ มีจำนวนแตกต่างจากกลุ่มที่ไม่กลับมาอย่างมาก ผู้วิจัยจึงได้ประยุกต์ใช้เทคนิคการสังเคราะห์ข้อมูล CTGAN ซึ่งถูกออกแบบมาเพื่อสังเคราะห์ข้อมูลเพื่อแก้ไขปัญหาคือ ขจัดปัญหาความลำเอียงของตัวแบบ (Xu et al., 2019) นำมาใช้เพื่อสร้างข้อมูลสังเคราะห์ในการคงสัดส่วนของลูกค้ากลุ่มที่ "กลับมา" และ "ไม่กลับมา" ให้ใกล้เคียงกันในทั้งสองชุด โดยนำเสนอขั้นตอนเป็นแผนภาพกราฟิก ดังรูปที่ 4



รูปที่ 4 แผนภาพกราฟิกแสดงขั้นตอนการปรับสมดุลข้อมูลด้วยเทคนิค CTGAN และ ชุดคำสั่ง Python

ทั้งนี้การใช้เทคนิค CTGAN เป็นการสร้างสมดุลของชุดข้อมูล เพื่อเป็นการป้องกันไม่ให้ตัวแปรที่มีปริมาณข้อมูลแตกต่างกันมีอิทธิพลเหนือตัวแบบมากเกินไป ถือเป็นการป้องกันอคติในการเรียนรู้ของตัวแบบ ทำให้ได้ ชุดข้อมูลสำหรับการเรียนรู้ (Training Data Set) และชุดข้อมูลสำหรับการทดสอบ (Testing Data Set) ดังตารางที่ 5

ตารางที่ 5 ชุดข้อมูลสำหรับการสร้างตัวแบบ

ชุดข้อมูล	สำหรับเรียนรู้ (Training Data Set)		สำหรับทดสอบ (Testing Data set)
	ก่อนปรับสมดุลข้อมูล	หลังปรับสมดุลด้วย CTGAN	
กลุ่ม "ไม่กลับมา"	466	466	117
กลุ่ม "กลับมา"	69	466	17
รวม	535 ระเบียบ	932 ระเบียบ	134 ระเบียบ

หลังจากทำการปรับสมดุลข้อมูลแล้ว จากนั้นจึงนำมาสร้างแบบตัวแบบพยากรณ์ข้อมูลโดยใช้เทคนิคการเรียนรู้ของเครื่อง โดยใช้เทคนิคการเรียนรู้ของเครื่อง 4 เทคนิค ได้แก่ Logistic regression Random forest XGBoost และ SVM

4) ทำการประเมินประสิทธิภาพของตัวแบบ

ในการประเมินประสิทธิภาพของตัวแบบ (Model Evaluation) ด้วยชุดข้อมูลที่จัดเตรียมไว้สำหรับการทดสอบตัวแบบ ผู้วิจัยทำการหาประสิทธิภาพตัวแบบด้วยค่าดัชนีชี้วัดประสิทธิภาพโดยรวม คือ ค่าพื้นที่ใต้กราฟ (AUC-ROC: Area Under the Curve - Receiver Operating Characteristic Curve) ซึ่งเป็นค่าดัชนีบ่งชี้ถึงขีดความสามารถในการจำแนกประเภทของตัวแบบที่ได้รับการยอมรับมากกว่า ค่าความถูกต้อง (Accuracy) เนื่องจากการบูรณาการการพิจารณาจาก ค่าความแม่นยำ (Precision) และ ค่าความระลึก (Recall) ณ ระดับเกณฑ์ตัดสินใจ (Threshold) ต่าง ๆ เป็นสำคัญ รวมถึง ค่าเอฟวัน (F1-Score) เพื่อนำมาพิจารณาความสัมพันธ์ เป็นตัวชี้วัดที่ประเมินประสิทธิภาพในภาพรวมของทุก ๆ ระดับเกณฑ์ตัดสินใจ (Classification thresholds) สะท้อนความสามารถของตัวแบบในการจัดอันดับ (Rank) การพยากรณ์ ทำนายผล ได้ดีในภาพรวม แสดงให้เห็นถึงการแลกเปลี่ยน (Trade-off) ระหว่าง

ผลประโยชน์จากการทายถูก (True Positive) และค่าเสียโอกาสจากการทายผิด (False Positive) พื้นที่ใต้กราฟ (AUC) ที่กว้างกว่าจึงเป็นตัวบ่งชี้ทางสถิติที่แข็งแกร่งว่า ตัวแบบนั้นสามารถสร้างผลประโยชน์ได้สูงสุดภายใต้ความเสี่ยงที่ต่ำที่สุด เมื่อเทียบกับตัวแบบอื่น ซึ่งการพิจารณาค่าพื้นที่ใต้กราฟ สามารถตีความหมายได้ ดังตารางที่ 6 ดังนี้

ตารางที่ 6 การตีความพื้นที่ใต้กราฟ (AUC-ROC) (ประยุกต์จาก Hosmer et al., 2013)

ค่าพื้นที่ใต้กราฟ (AUC-ROC)	ความหมาย
1.0	ตัวแบบมีประสิทธิภาพสมบูรณ์ ในการแยกแยะคลาส (เหมาะสมในการนำไปใช้พยากรณ์)
0.8 – 0.9	ตัวแบบมีประสิทธิภาพในระดับดีเยี่ยม
0.7 – 0.8	ตัวแบบมีประสิทธิภาพในระดับพอใช้ถึงดี
0.5	ตัวแบบไม่มีความสามารถในการพยากรณ์ (เหมือนการสุ่มเดา)

ในการดำเนินการประเมินประสิทธิภาพของตัวแบบ ดำเนินการต่อเนื่องจากการสร้างตัวแบบและทำการประเมินวัดค่าประสิทธิภาพ ของตัวแบบทั้ง 4 เทคนิค

ผลการวิจัย

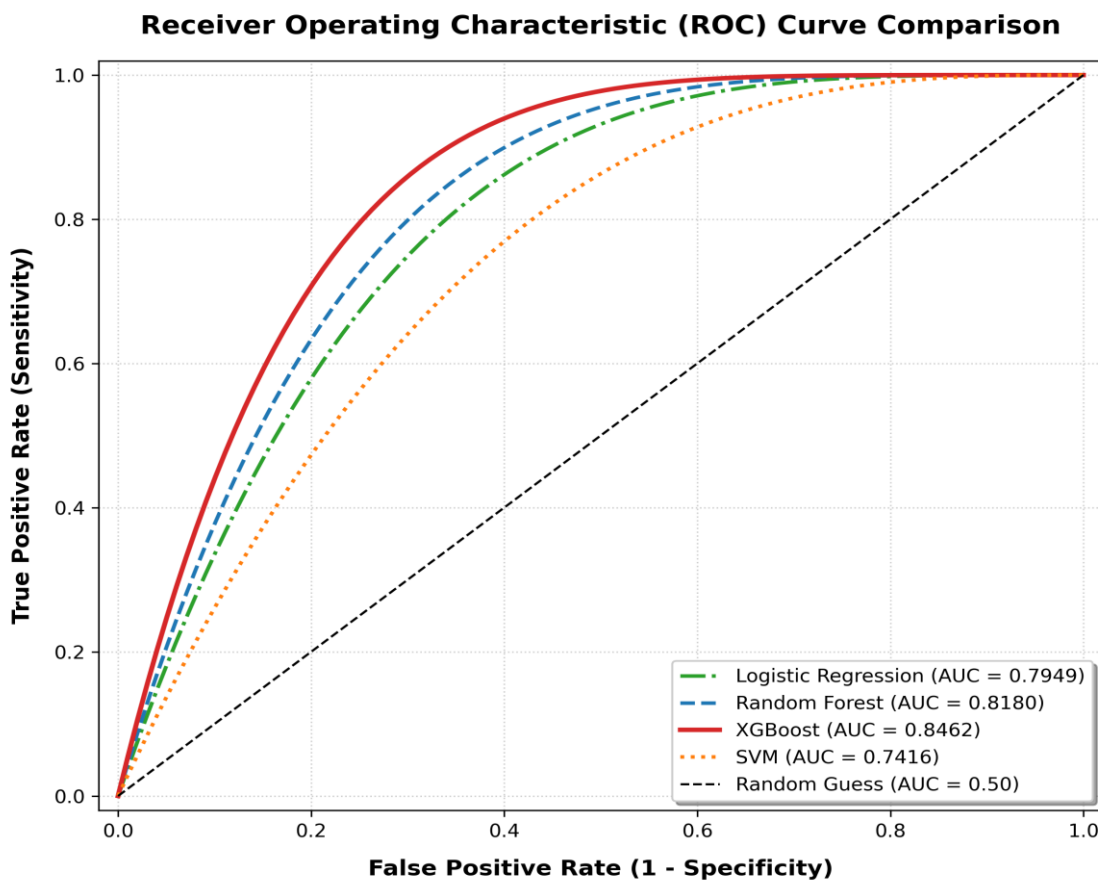
ในการวิจัยครั้งนี้ มีวัตถุประสงค์ เพื่อพัฒนาตัวแบบพยากรณ์ข้อมูลด้วยเทคนิคการเรียนรู้ของเครื่องที่สามารถทำนายการกลับมาใช้บริการ เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบที่สร้างด้วยอัลกอริทึมการเรียนรู้ของเครื่อง และเพื่อวิเคราะห์ปัจจัยสำคัญที่ส่งผลต่อการรักษาลูกค้าในธุรกิจแฟรนไชส์จตุรัสหงส์ โดยมีการละเอียดผลวิจัยดังนี้

1) ผลการพัฒนาตัวแบบพยากรณ์ข้อมูลด้วยเทคนิคการเรียนรู้ของเครื่อง จากการรวบรวมและจัดเตรียมข้อมูลรายการธุรกรรมของธุรกิจแฟรนไชส์จตุรัสหงส์ ผู้วิจัยได้ดำเนินการสกัดคุณลักษณะ ตามกรอบแนวคิด RFM ร่วมกับคุณลักษณะการใช้บริการเก็บเงินปลายทาง และประเภทลูกค้า เพื่อเปลี่ยนข้อมูลระดับธุรกรรมให้เป็นคุณลักษณะรายบุคคล จากนั้นได้ดำเนินการขจัดปัญหาความไม่สมดุลของข้อมูล ด้วยเทคนิคการสังเคราะห์ข้อมูล CTGAN เพื่อป้องกันอคติในการเรียนรู้ของตัวแบบ ก่อนจะนำชุดข้อมูลที่ผ่านการปรับสมดุลแล้ว ไปใช้สร้างตัวแบบพยากรณ์การกลับมาใช้บริการด้วยเทคนิคการเรียนรู้ของเครื่อง 4 เทคนิค ได้แก่ Logistic Regression Random Forest XGBoost and SVM

2) ผลการเปรียบเทียบประสิทธิภาพของตัวแบบที่สร้างด้วยเทคนิคการเรียนรู้ของเครื่อง ผู้วิจัยทำการเปรียบเทียบประสิทธิภาพของตัวแบบการเรียนรู้ของเครื่อง จำนวน 4 เทคนิค ได้แก่ Logistic Regression Random Forest XGBoost และ SVM ในการพยากรณ์พฤติกรรมกลับมาใช้บริการซ้ำของลูกค้า (ราย 30 วัน) ซึ่งผลประสิทธิภาพของตัวแบบสรุปได้ดังตารางที่ 7 โดยมีค่าดัชนีชี้วัดประสิทธิภาพโดยภาพรวม คือ ค่าพื้นที่ใต้กราฟ (AUC-ROC) ดังปรากฏในรูปที่ 5

ตารางที่ 7 ผลจากการสร้างตัวแบบด้วยเทคนิคการเรียนรู้ของเครื่อง ทั้ง 4 เทคนิค

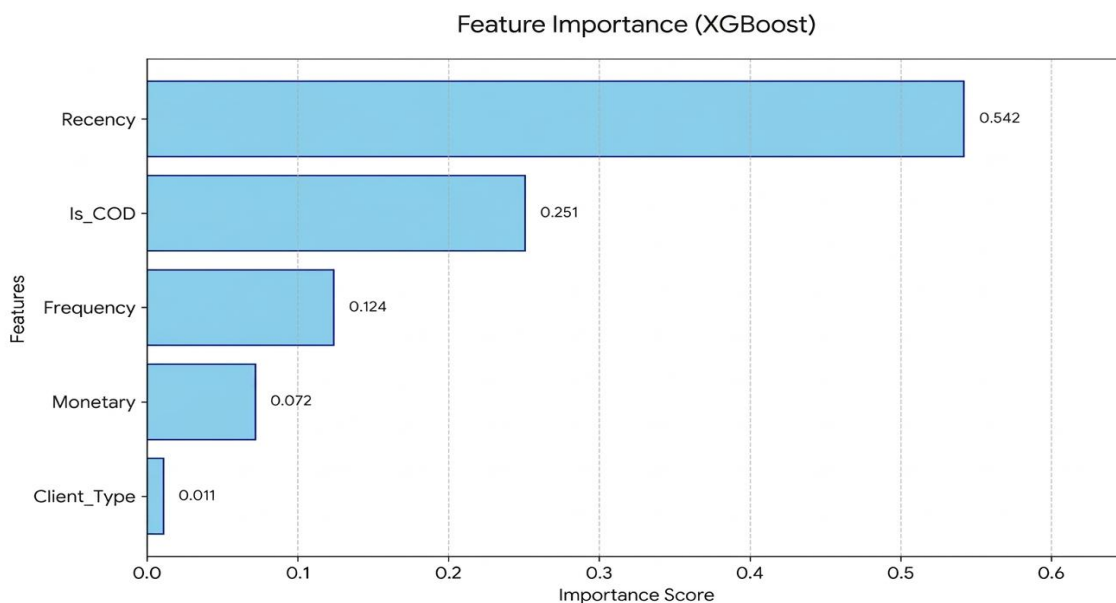
ตัวแบบ (Model)	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	79.10%	35.14%	76.47%	48.15%	0.7959
Random Forest	85.82%	42.86%	35.29%	38.71%	0.8180
XGBoost	85.82%	43.75%	52.94%	42.42%	<u>0.8462</u>
SVM	88.06%	66.67%	11.76%	21.05%	0.7416



รูปที่ 5 กราฟแสดงการเปรียบเทียบค่าพื้นที่ใต้กราฟ (AUC-ROC) ของตัวแบบทั้ง 4 เทคนิค

ผลจากการประเมินเปรียบเทียบประสิทธิภาพของตัวแบบทั้ง 4 เทคนิค ด้วยค่าดัชนีพื้นที่ใต้กราฟ (AUC-ROC) ซึ่งเป็นดัชนีชี้วัดขีดความสามารถในการจำแนกกลุ่มลูกค้า พบว่า แบบจำลอง XGBoost ให้ประสิทธิภาพในการพยากรณ์สูงที่สุด โดยมีค่า AUC-ROC เท่ากับ 0.8462 ซึ่งอยู่ในระดับประสิทธิภาพที่ "ดี" รองลงมาคือ Random Forest (0.8180), Logistic Regression (0.7959) และ SVM (0.7416) ตามลำดับ นอกจากนี้ แม้ว่า SVM ให้ค่าความถูกต้อง (Accuracy) และ ค่าความแม่นยำ (Precision) สูงสุดที่ 88.06% และ 66.67% แต่กลับมีค่าความระลึก (Recall) ต่ำที่สุดที่ 11.76% ซึ่งบ่งชี้ถึงข้อจำกัดในการตรวจจับกลุ่มลูกค้าที่จะกลับมาใช้บริการจริงได้เพียงส่วนน้อย ในขณะที่ XGBoost กลับให้ค่าที่ 52.94% และเมื่อพิจารณาค่า F1-Score ยังมีค่าที่ 42.42% ทำให้เมื่อพิจารณาค่า AUC-ROC แสดงให้เห็นถึงประสิทธิภาพด้านความสมดุลระหว่างความแม่นยำและการตรวจจับกลุ่มเป้าหมายได้ครอบคลุมกว่าเทคนิคอื่นๆ ดังนั้น XGBoost จึงเป็นตัวแบบที่มีความเหมาะสมที่สุดในการนำไปประยุกต์ใช้เป็นระบบสนับสนุนการตัดสินใจเพื่อพยากรณ์การกลับมาของลูกค้า

3) ผลการวิเคราะห์ปัจจัยสำคัญที่ส่งผลต่อการรักษาลูกค้า จากการวิเคราะห์ความสำคัญของคุณลักษณะข้อมูล (Feature Importance) ที่ส่งผลต่อการพยากรณ์การกลับมาใช้บริการของตัวแบบ พบว่า ปัจจัยด้านความใหม่ของการใช้บริการ (Recency) เป็นตัวแปรที่มีอิทธิพลต่อการพยากรณ์มากที่สุด ซึ่งสะท้อนให้เห็นว่าพฤติกรรมการส่งพัสดุในช่วงเวลาล่าสุดมีนัยสำคัญและสัมพันธ์โดยตรงกับแนวโน้มที่จะเป็นในรักษาลูกค้า รองลงมาคือ สถานะการใช้บริการเก็บเงินปลายทาง (COD) ซึ่งถือเป็นกลุ่มลูกค้าที่เป็นฐานรายได้หลักของธุรกิจ นอกจากนี้ ปัจจัยด้านความถี่ในการใช้บริการ (Frequency) ยังสามารถสะท้อนถึงขนาดของธุรกิจลูกค้า โดยพบว่ากลุ่มลูกค้าแพลตฟอร์มออนไลน์ที่มีความถี่ในการส่งพัสดุอย่างสม่ำเสมอ มีโอกาสที่จะกลับมาใช้บริการซ้ำสูงกว่าลูกค้าทั่วไปอย่างมีนัยสำคัญ ดังแสดงในรูปที่ 6



รูปที่ 6 กราฟแสดงผลการวิเคราะห์ปัจจัยสำคัญที่ส่งผลต่อการรักษาลูกค้า

สรุปผลและอภิปรายผล

จากผลการวิจัยในครั้งนี้ สามารถสรุปและอภิปรายผลสอดคล้องตามวัตถุประสงค์ ได้ดังนี้

จากการพัฒนาตัวแบบสำหรับพยากรณ์การรักษาลูกค้าชี้ให้เห็นว่า ข้อมูลรายการธุรกรรมการส่งพัสดุ โดยบูรณาการเทคนิคการเรียนรู้ของเครื่องเข้ากับตัวแปร RFM สามารถนำมาวิเคราะห์สร้างตัวแบบสำหรับพยากรณ์ความน่าจะเป็นในการกลับมาใช้บริการ ซึ่งสอดคล้องกับงานวิจัยของชาญวัฒน์ ภัคศิษฐ์ (2566) ที่ใช้ RFM มาวิเคราะห์จัดกลุ่มข้อมูลด้วยเทคนิคการเรียนรู้ของเครื่องได้ หากแต่ในงานวิจัยนี้มีข้อแตกต่างคือแก้ไขข้อจำกัดด้านสมดุลข้อมูลที่นำมาใช้ในการวิเคราะห์ข้อมูล โดยประยุกต์เทคนิคการสังเคราะห์ข้อมูล CTGAN ของ Xu et al. (2019) สำหรับปรับสมดุลข้อมูล ร่วมกับเทคนิคการเรียนรู้ของเครื่องในการสร้างตัวแบบพยากรณ์ ทำให้สามารถสร้างตัวแบบพยากรณ์ ช่วยให้ผู้ประกอบการธุรกิจได้รับส่งพัสดุรายย่อย สามารถพยากรณ์พฤติกรรมลูกค้าได้ ซึ่งสอดคล้องกับงานวิจัยของ Sinchana et al. (2025) ที่ยืนยันว่า CTGAN มีความสอดคล้องกับข้อมูลจริงสูงมาก และ Phanbua et al. (2025) ที่ใช้ CTGAN ในการสังเคราะห์ข้อมูล โดยต่างยืนยันว่า CTGAN มีส่วนช่วยในการปรับสมดุลข้อมูลในรูปแบบตารางได้อย่างมีนัยสำคัญ

จากการเปรียบเทียบประสิทธิภาพของตัวแบบที่สร้างด้วยอัลกอริทึมการเรียนรู้ของเครื่อง มีข้อค้นพบที่สำคัญคือ ตัวแบบการเรียนรู้ของเครื่องจากเทคนิค XGBoost มีประสิทธิภาพในการพยากรณ์สูงที่สุด (ค่าประสิทธิภาพโดยรวมคิดเป็นร้อยละ 85%) ผลการวิจัยนี้สอดคล้องกับแนวคิดของ Korstanje (2021) และ งานของ ปารรณนา ด้านก่อโพธิ์ (2567) ที่ระบุว่าเทคนิค XGBoost มีประสิทธิภาพสูงในการจัดการข้อมูลที่มีความซับซ้อนและเรียนรู้จากความผิดพลาดของตัวแบบก่อนหน้าได้ดี รวมถึงสอดคล้องกับงานวิจัยของ Sunarya et al. (2024); Monisha et al. (2020) ที่ยืนยันว่าอัลกอริทึมในกลุ่ม การเรียนรู้หลากหลาย (Ensemble Learning) มีความแม่นยำสูงในการทำนายพฤติกรรมการซื้อขายและการยกเลิกใช้บริการของลูกค้า นอกจากนี้ ความสำเร็จในการพยากรณ์ชุดข้อมูลขนาดเล็กยังเป็นผลมาจากการประยุกต์ใช้เทคนิคการสังเคราะห์ข้อมูล CTGAN ซึ่งผลการวิจัยได้ยืนยันและสอดคล้องกับงานวิจัยของ Moon et al. (2020) รวมถึง (Kim & Seok (2024); Phanbua et al. (2025) ที่พบว่า CTGAN สามารถสร้างสมดุลให้กับข้อมูลแบบ

ตารางและช่วยลดอคติ ในการเรียนรู้ของตัวแบบได้อย่างมีประสิทธิภาพ ทำให้ตัวแบบที่ได้มีความเหมาะสมที่สุดสำหรับการนำไปประยุกต์ใช้เป็นตัวแบบสำหรับในระบบสนับสนุนการตัดสินใจเพื่อพยากรณ์การกลับมาของลูกค้า

ในประเด็นการวิเคราะห์ปัจจัยสำคัญที่ส่งผลต่อการรักษฐานลูกค้า จากการวิเคราะห์ความสำคัญของตัวแปรจากเทคนิคการวิเคราะห์พฤติกรรมลูกค้า พบว่า ปัจจัยด้านความใหม่ (Recency) เป็นตัวแปรที่มีอิทธิพลมากที่สุด ซึ่งสะท้อนให้เห็นว่าปัจจัย RFM และตัวแปรการใช้บริการเก็บเงินปลายทาง (COD) ที่นำมาใช้ในการวิเคราะห์ข้อมูลนั้นสามารถนำมาพยากรณ์การกลับมาของลูกค้าได้ในระดับ "ดี" ผลการวิจัยนี้สอดคล้องกับงานวิจัยของกาญจนา หลุณรพพงศ์ และปิยมาศ จิตตระ (2561) และงานของ ปันดารีย์ สุนทรวราภาส และคณะ (2563) ที่พบว่าปัจจัยด้านความใหม่ ปัจจัยด้านความถี่ และปัจจัยด้านมูลค่า สามารถจำแนกกลุ่มผู้ใช้งานได้ และสอดคล้องกับงานวิจัยของ Monisha et al. (2020) ที่พบว่าการประยุกต์ใช้แบบจำลอง RFM ร่วมกับเทคนิคการเรียนรู้ของเครื่อง สามารถทำนายความตั้งใจในการซื้อซ้ำได้อย่างมีประสิทธิภาพ โดยสอดคล้องกับทฤษฎีด้านการตลาดที่ว่า ลูกค้าที่เพิ่งมาใช้บริการไม่นานมีโอกาสที่จะยังมีความต้องการส่งพัสดุอย่างต่อเนื่อง มากกว่าลูกค้าที่หายไถ่ไปนานแล้ว นอกจากนี้ ปัจจัยด้านความถี่ในการใช้บริการ (Frequency) ยังสะท้อนถึงขนาดของธุรกิจลูกค้า หากเป็นลูกค้าแพลตฟอร์มออนไลน์ ที่มีการส่งพัสดุสัปดาห์ โอกาสที่จะกลับมาใช้บริการซ้ำจะสูงกว่าลูกค้าทั่วไปเกือบ 5 เท่า ซึ่งการค้นพบนี้ช่วยสนับสนุนแนวคิดและสอดคล้องกับงานวิจัยของชาญวัฒน์ ภัคศิริ (2566) ที่ระบุว่า ความถี่ในการเข้าใช้บริการเป็นดัชนีสำคัญที่บ่งบอกถึงระดับ ความจงรักภักดี (Loyalty) ของลูกค้าที่มีต่อธุรกิจ

ในส่วนของประโยชน์ของการวิจัยครั้งนี้ ในเชิงธุรกิจ ผลจากการวิจัยในครั้งนี้พิสูจน์ให้เห็นว่าข้อมูลในธุรกิจขนาดเล็ก ก็สามารถสร้างมูลค่าให้แก่เจ้าของธุรกิจแฟรนไชส์จุดรับส่งพัสดุได้ โดยสามารถใช้ผลความน่าจะเป็น (Probability Score) จากตัวแบบที่ได้ นำมาคาดการณ์ พยากรณ์ พฤติกรรมการกลับมาใช้บริการ เพื่อนำมาใช้ประโยชน์ เช่น สร้างระบบแจ้งเตือนอัตโนมัติ เมื่อลูกค้าชั้นดีมีความน่าจะเป็นที่จะกลับมาลดต่ำลงเกินเกณฑ์ที่กำหนด (เช่น ต่ำกว่า 50%) เพื่อให้สาขาทำการติดต่อหรือมอบโปรโมชั่นพิเศษก่อนที่ลูกค้าจะเปลี่ยนใจไปใช้บริการที่อื่น เป็นต้น รวมถึงแนวทางการใช้ประโยชน์ข้อมูลครั้งนี้จะสามารถนำไปประยุกต์ใช้กับธุรกิจขนาดเล็กอื่น ๆ ทั้งนี้เพื่อสามารถนำไปเพิ่มมูลค่าแก่ธุรกิจได้ต่อไป

ในส่วนของประโยชน์ในเชิงวิชาการ การวิจัยนี้สร้างประโยชน์ในด้านการบูรณาการข้ามศาสตร์ ระหว่างทฤษฎีพฤติกรรมศาสตร์ด้านการตลาดและวิทยาการข้อมูลขั้นสูง โดยผลการศึกษาให้หลักฐานเชิงประจักษ์ว่าการประยุกต์ใช้นวัตกรรมการสังเคราะห์ข้อมูลด้วยเทคนิค CTGAN นั้นสามารถแก้ไขปัญหาค่าความไม่สมดุลของข้อมูล และลดความลำเอียงของตัวแบบ ซึ่งเป็นอุปสรรคสำคัญในการวิเคราะห์ชุดข้อมูลธุรกิจขนาดเล็กได้อย่างมีนัยสำคัญ นอกจากนี้ ผลการวิเคราะห์ลำดับความสำคัญของตัวแปร ยังช่วยยืนยันแนวคิดพฤติกรรมของลูกค้าว่า ปัจจัยด้านความใหม่ เป็นตัวแปรพยากรณ์หลักที่มีอิทธิพลต่อการกลับมาใช้บริการซ้ำ การค้นพบนี้ จึงเป็นการต่อยอดความสมบูรณ์ของทฤษฎีการตลาดดั้งเดิมด้วยระเบียบวิธีวิจัยทางวิทยาการข้อมูลสมัยใหม่ และสามารถใช้เป็นบรรทัดฐานสำหรับการศึกษาวิจัยด้านการทำเหมืองข้อมูลในบริบทของธุรกิจรายย่อยและอุตสาหกรรมอื่นที่เกี่ยวข้องต่อไป

ข้อเสนอแนะ

ข้อเสนอแนะจากการวิจัย

จากข้อค้นพบจากงานวิจัยการพัฒนาตัวแบบพยากรณ์การรักษฐานลูกค้าในครั้งนี้ ซึ่งค้นพบว่าตัวแบบ XGBoost และปัจจัยด้าน Recency รวมถึง COD มีความสำคัญสูงสุด ผู้วิจัยจึงมีข้อเสนอแนะเชิงปฏิบัติสำหรับผู้ประกอบการและผู้ที่เกี่ยวข้อง ในประเด็นสำคัญ กล่าวคือ ในประเด็นด้านการประยุกต์ใช้ตัวแบบเพื่อการตลาดเชิงรุก (Proactive Marketing) ผู้ประกอบการสามารถนำตัวแบบพยากรณ์ที่พัฒนาขึ้นนี้ไปประยุกต์ใช้เป็นเครื่องมือ

ในการรักษาลูกค้า โดยมุ่งเน้นไปที่ลูกค้ากลุ่มที่ตัวแบบพยากรณ์ว่า "มีโอกาสไม่กลับมาใช้บริการซ้ำ" (Predicted Churn) เพื่อสร้างโอกาส ในการทำการตลาดเชิงรุก เช่นการให้ส่วนลด หรือสิทธิพิเศษ เพื่อดึงดูดให้กลับมาใช้บริการ อีกทั้งยังสามารถใช้ประโยชน์ข้อมูลจากการพยากรณ์ในการสำรวจข้อมูล ความพึงพอใจ เพื่อรับทราบปัญหา หรือ อุปสรรคในการใช้บริการ เพื่อนำข้อมูลมาปรับปรุงคุณภาพการบริการต่อไป รวมไปถึงการให้ความสำคัญกับกลุ่มลูกค้า COD ในฐานะฐานรายได้หลัก ซึ่งจากผลการวิจัยยืนยันว่าสถานะการใช้บริการเก็บเงินปลายทาง (COD) เป็นหนึ่ง ปัจจัยในการรักษาลูกค้า ผู้ประกอบการควรมีกลยุทธ์รักษาลูกค้ากลุ่มนี้เป็นพิเศษ เช่น การจัดสิทธิประโยชน์ระดับ สมาชิก (Tiered Loyalty Program) มอบสิทธิพิเศษสำหรับผู้ค้าออนไลน์ที่มียอดการใช้บริการเก็บเงินปลายทางอย่างสม่ำเสมอ ซึ่งถือเป็นการต่อยอดใช้ประโยชน์จากตัวแบบพยากรณ์ในการรักษาลูกค้าได้

นอกจากนี้ ในประเด็นปัจจัยค่าการติดตามค่าความใหม่ของการบริการ (Recency Monitoring) สามารถต่อยอดผลจากการพัฒนาตัวแบบพยากรณ์รักษาลูกค้านี้ นำไปต่อยอดในการจัดทำระบบ Dashboard อย่างง่าย เพื่อติดตามค่า Recency ของลูกค้าแต่ละราย ซึ่งหากพบว่าลูกค้าไม่ได้มาใช้บริการเกินกว่า 7-14 วัน (ตามพฤติกรรมปกติของแต่ละกลุ่ม) ระบบควรมีการแจ้งเตือน (Alert) เพื่อให้พนักงานหน้าร้านหรือเจ้าของสาขาเตรียมแผนการสื่อสารกับลูกค้าได้ทันทั่วถึง เพื่อเป็นการรักษาลูกค้าไว้ต่อไป

ข้อเสนอแนะในการวิจัยครั้งต่อไป

สำหรับการต่อยอดการศึกษา วิจัย ไปสู่การวิจัยในครั้งต่อไป หากมีการบูรณาการต่อยอดผลการศึกษาจากการวิจัยนี้ เข้ากับซอฟต์แวร์การจัดการจัดการจัดส่งพัสดุหน้าร้าน (Parcel Shop Management System) ของเจ้าของแฟรนไชส์หลัก เพื่อให้มีฟังก์ชันการวิเคราะห์ และรายงาน กลุ่มลูกค้าเสี่ยง ได้ จะทำให้เจ้าของกิจการแฟรนไชส์ธุรกิจจัดส่งพัสดुरายย่อย ได้มีเครื่องมือช่วยในการรักษาลูกค้า อันจะเป็นการเพิ่มประสิทธิภาพในการรักษาลูกค้าให้แก่เจ้าของธุรกิจ และเพิ่มศักยภาพแข่งขันให้แก่ธุรกิจรายย่อยได้ นอกจากนี้ แม้ว่าจะมีการประยุกต์ใช้เทคนิคการสังเคราะห์ข้อมูลในการปรับสมดุลข้อมูล แต่ด้วยข้อจำกัดของขนาดชุดข้อมูลทดสอบ หากมีการเก็บรวบรวมข้อมูลของธุรกิจในระยะเวลายาวนานมากขึ้น เชื่อว่าจะสามารถพัฒนาประสิทธิภาพในการเรียนรู้ และจำแนกประเภทได้ดียิ่งขึ้น รวมถึงการศึกษาแนวทางเทคนิคการทำเหมืองข้อมูลจำแนกประเภทด้วยเทคนิคการเรียนรู้ของเครื่อง วิธีการอื่นๆ โดยอาจได้แนวทางในการเพิ่มประสิทธิภาพของตัวแบบในการรักษาลูกค้าต่อไป

เอกสารอ้างอิง

- กาญจนา หฤพรพงษ์ และ ปิยมาศ จิตตระ. (2561). การวิเคราะห์แบบ RFM ในการแบ่งกลุ่มผู้ใช้งานเครื่องพิมพ์และเครื่องถ่ายเอกสารในองค์กร กรณีศึกษาสำนักวิชาสารสนเทศศาสตร์ มหาวิทยาลัยวลัยลักษณ์. *วารสารวิชาการการจัดการเทคโนโลยีสารสนเทศและนวัตกรรม*, 5(1), 21-29.
- จามรกุล เหล่าเกียรติกุล. (2558). *เหมืองข้อมูลเบื้องต้น : Introduction to Data Mining*. กรุงเทพฯ : แดเน็กซ์ อินเทอร์เน็ตปอเรชั่น.
- ชาญวิวัฒน์ ภัคศิรี. (2566). *การพัฒนากลยุทธ์ทางการตลาดโดยใช้วิทยาศาสตร์ข้อมูลในการจัดกลุ่มลูกค้าตามพฤติกรรมการซื้อสินค้า* [การค้นคว้าอิสระปริญญาโทมหาบัณฑิต, มหาวิทยาลัยธรรมศาสตร์]. วิทยาลัยนวัตกรรม. https://digital.library.tu.ac.th/tu_dc/frontend/Info/item/dc:313603
- ณรรฐคุณ วิรุฬห์ศรี, ลัทพล โชครัตน์ประภา, ณัฐนิชา ศรีสมาน และพรทิพย์ เดชพิชัย. (2565). การวิเคราะห์แบ่งกลุ่มลูกค้าโดยใช้พฤติกรรมซื้อเชิงลึก: กรณีศึกษา บริษัทผู้ผลิตอาหารสัตว์เลี้ยงแห่งหนึ่ง. *วารสารวิทยาศาสตร์ลาดกระบัง*, 31(1), 103-120.

- ปณดารีย์ สุนทรวราภาส, น้ำทิพย์ ตระกูลเมธี, และ สุรีนา มะตาหยง. (2563). การวิเคราะห์แบบ RFM ในการแบ่งกลุ่มผู้ใช้งานตามพฤติกรรมการเยี่ยมชมทรัพยากรสารสนเทศ ของสำนักทรัพยากรการเรียนรู้คุณหญิงหลงอรรถกระวีสุนทร มหาวิทยาลัยสงขลานครินทร์. *วารสารบรรณศาสตร์ มศว*, 13(1), 30-45.
- ปรารถนา ด้านก่อโพธิ์. (2567). *การพยากรณ์ยอดขายโดยใช้อัลกอริทึม XGBoost และ TimesFM* [วิทยานิพนธ์ปริญญามหาบัณฑิต, จุฬาลงกรณ์มหาวิทยาลัย].
https://www.cp.eng.chula.ac.th/~prabhas/thesis/prathana_complete_2024.pdf
- วิราภานต์ กิตติบรรกุล, ศรายุทธ นนท์ศิริ, และ พิชิตชัย คำอินทร์. (2565). การเปรียบเทียบประสิทธิภาพเทคนิคการเรียนรู้ของเครื่องสำหรับการบำรุงรักษาเชิงคาดการณ์ของเครื่องยนต์อากาศยาน. *วารสารวิชาการสมาคมสถาบันอุดมศึกษาเอกชนแห่งประเทศไทย (ฉบับวิทยาศาสตร์และเทคโนโลยี)*, 11(1), 16-29.
- Ang, L., & Buttle, F. (2006). Customer retention management processes: A survey of Australian companies. *European Journal of Marketing*, 40(1/2), 83-99.
<https://doi.org/10.1108/03090560610637329>
- Gundogdu, S. (2022). Hepatitis C Disease Detection Based on PCA-SVM Model. *Hittite Journal of Science and Engineering*, 9(2), 111-116.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression (3rd ed.)*. John Wiley & Sons.
- Kim, J., & Seok, J. (2024). ctGAN: Combined transformation of gene expression and survival data with generative adversarial network. *Briefings in Bioinformatics*, 25(4), bbae325.
<https://doi.org/10.1093/bib/bbae325>
- Korstanje, J. (2021). Gradient Boosting with XGBoost and LightGBM. In *Advanced Forecasting with Python*. (pp. 203-221) Apress. Berkeley, CA. https://doi.org/10.1007/978-1-4842-7150-6_15.
- Meric, E., & Ozer, C. (2022). Symptom Based Health Status Prediction via Decision Tree, KNN, XGBoost, LDA, SVM, and Random Forest. In *Proceedings of International Conference on Computing Intelligence and Data Analytics*, Koceli, Turkey, 193-207.
- Monisha, A. S., Suresh Kumar, N., & Sreeramulu, M. (2020). Customer Segmentation and Repurchase Intention using RFM Model and Machine Learning. *International Journal of Engineering and Advanced Technology*, 9(4), 2142-2146.
- Moon, Jaeuk & Jung, Seungwon & Park, Sungwoo & Hwang, Eenjun. (2020). Conditional Tabular GAN-Based Two-Stage Data Generation Scheme for Short-Term Load Forecasting. *IEEE Access*. 8. 205327-205339. <https://doi.org/10.1109/ACCESS.2020.3037063>.
- Phanbua, P., Arwatchananukul, S., Hristov, G., & Temdee, P. (2025). CTGAN-augmented ensemble learning models for classifying dementia and heart failure. *Inventions*, 10(6), 101.
<https://doi.org/10.3390/inventions10060101>
- Rizkyanto, H., & Gaol, F. L. (2023). Customer segmentation of personal credit using Recency, Frequency, Monetary (RFM) and K-means on financial industry. *International Journal of Advanced Computer Science and Applications*, 14(4), 312-318.
<https://doi.org/10.14569/IJACSA.2023.0140417>
- Sinchana, K. C., Anthraper, M. G., Sanjaykumar, K., Kumari, S., & D., U. (2025). Synthetic data generation using CTGAN with agentic workflows and retrieval-augmented generation. *Proceedings of the 5th International Conference on AI Research (ICAIR 2025)*, 5(1), 472-480.
<https://doi.org/10.34190/icair.5.1.4280>

- Sunarya, P.A., Rahardja, U., Chen, S.C. , Choi, K., & Ku, C. T. (2024) Deciphering Digital Social Dynamics: A Comparative Study of Logistic Regression and Random Forest in Predicting e-Commerce Customer Behavior. *Journal of Applied Data Sciences*, 5(1), 100-113.
<https://doi.org/10.47738/jads.v5i1.155>
- Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Modeling tabular data using conditional GAN. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (Article No. 659, pp. 7335–7345). Curran Associates Inc.