

การวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์สำหรับธุรกิจอีคอมเมิร์ซด้วยเทคนิคเหมืองข้อมูลแบบไฮบริด

Real-Time Customer Behavior Analysis for E-Commerce Businesses

Using Hybrid Data Mining Techniques

อัคราภรณ์ นาชัยทอง^{1*} วุฒิชัย อินทร์แก้ว²

¹คณะบริหารธุรกิจและการบัญชี มหาวิทยาลัยราชภัฏร้อยเอ็ด ²คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏสงขลา

Atcharaporn Nachaithong^{1*} Wuttichai Inkaew²

¹Faculty of Business Administration and Accountancy, Roi Et Rajabhat University ²Faculty of Management Sciences, Songkhla Rajabhat University

*Corresponding Author E-mail: atcharaporn_ju@reru.ac.th

(วันที่รับบทความ: 17 กุมภาพันธ์ 2568 วันที่แก้ไขบทความ: 17 มิถุนายน 2568 วันที่ตอบรับบทความ: 14 กรกฎาคม 2568)

บทคัดย่อ

การศึกษางานวิจัยในครั้งนี้ มีวัตถุประสงค์เพื่อ 1) พัฒนาโมเดลการวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์ โดยใช้การผสมผสานของเทคนิคเหมืองข้อมูล และ 2) เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลแบบไฮบริด กับเทคนิคการวิเคราะห์ข้อมูลแบบดั้งเดิม ดำเนินการเก็บรวบรวมข้อมูล โดยมีการสกัดและดึงข้อมูลจากเว็บไซต์ (Web Scraping) และทำการเชื่อมต่อโปรแกรมประยุกต์ (API) เพื่อดึงข้อมูลจากแพลตฟอร์มอีคอมเมิร์ซและโซเชียลมีเดีย คือ ช้อปปี้ (Shopee) ลาซาด้า (Lazada) และแอมะซอน (Amazon) ที่เกี่ยวกับพฤติกรรมการซื้อของลูกค้า เช่น รีวิวสินค้า ความคิดเห็นของลูกค้า และการบริการ การประเมินผลลัพธ์ของโมเดลใช้ตัวชี้วัดมาตรฐาน คือ ความถูกต้องความแม่นยำ ความระลึก และความถี่ ความเร็วในการประมวลผลข้อมูล และความสามารถในการรองรับข้อมูลขนาดใหญ่ ตามลำดับ

ผลการวิจัยพบว่า เทคนิคเหมืองข้อมูลแบบไฮบริดแสดงให้เห็นถึงความสามารถในการเพิ่มประสิทธิภาพการวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์ได้อย่างมีนัยสำคัญ โดยโมเดลการแบ่งกลุ่ม (Clustering) ร่วมกับการจำแนกประเภท (Classification) มีค่าความถูกต้องสูงสุดร้อยละ 92.50 และความเร็วในการประมวลผล พบว่า การแบ่งกลุ่มร่วมกับการจำแนกประเภทมีความเร็วสูงสุด 1.80 วินาที และการเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลแบบไฮบริดกับเทคนิคการวิเคราะห์ข้อมูลแบบดั้งเดิม ผลการวิจัยชี้ให้เห็นว่า โมเดลที่ใช้เทคนิคเหมืองข้อมูลแบบไฮบริดมีความถูกต้องและประสิทธิภาพสูงกว่าการใช้เทคนิคเดี่ยว โดยเฉพาะ เค-มีนส์ (K-Means) ร่วมกับต้นไม้สุ่ม (Random Forest) ที่ให้ความถูกต้องร้อยละ 92.50 ผลการวิจัยนี้สามารถนำไปใช้ในการพัฒนาโมเดลการวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์ เพื่อช่วยธุรกิจในการตัดสินใจ และวางกลยุทธ์การตลาดที่มีประสิทธิภาพ โดยสามารถนำไปปรับใช้ในระบบการพาณิชย์อิเล็กทรอนิกส์ และโซเชียลมีเดียเพื่อเพิ่มความพึงพอใจและการมีส่วนร่วมของลูกค้า

คำสำคัญ: การวิเคราะห์พฤติกรรมลูกค้า, เรียลไทม์, ธุรกิจอีคอมเมิร์ซ, เทคนิคเหมืองข้อมูลแบบไฮบริด

Abstract

This research aimed to 1) develop a real-time customer behavior analysis model by integrating data mining techniques, and 2) compare the performance of hybrid data mining techniques with traditional data analysis methods. Data collection was carried out using Web Scraping and APIs to extract information from e-commerce and social media platforms such as Shopee, Lazada, and Amazon, focusing on customer purchasing behavior, including product reviews, customer comments, and service feedback. The evaluation of the model's performance was conducted using standard metrics, including accuracy, precision, recall, F1-score, processing speed and scalability to handle large datasets.

The research findings indicate that the hybrid data mining technique significantly enhances the real-time analysis of customer behavior. The model combining Clustering and Classification achieved the highest accuracy of 92.50%, with a processing speed of up to 1.8 seconds. When comparing the performance of the hybrid data mining technique with traditional data analysis methods, the results show that the hybrid model outperforms single-technique approaches, particularly the combination of K-Means and Random Forest, which also achieved an accuracy of 92.50%. These findings can be applied to develop real-time customer behavior analysis models to assist businesses in decision-making and effective marketing strategy formulation. Additionally, the model can be utilized in e-commerce systems and social media platforms to enhance customer satisfaction and engagement.

Keywords: Customer Behavior Analysis, Real-Time, E-Commerce Business, Hybrid Data Mining Techniques

บทนำ

ในยุคดิจิทัลที่ธุรกิจ “การพาณิชย์อิเล็กทรอนิกส์ หรืออีคอมเมิร์ซ” เติบโตอย่างรวดเร็ว การทำความเข้าใจพฤติกรรมของลูกค้าถือเป็นปัจจัยสำคัญในการเพิ่มขีดความสามารถทางการแข่งขันและสร้างความพึงพอใจให้กับผู้บริโภค ด้วยความก้าวหน้าของเทคโนโลยีการประมวลผลข้อมูล ปัจจุบันธุรกิจสามารถติดตามและวิเคราะห์พฤติกรรมของลูกค้าได้แบบเรียลไทม์ (Real-Time Customer Behavior Analysis) ซึ่งช่วยให้สามารถปรับวิธีการวางแผนทางธุรกิจได้อย่างทันท่วงที (Madhumita et al. 2024) อย่างไรก็ตามความท้าทายที่สำคัญ คือการจัดการข้อมูลจำนวนมากที่เกิดขึ้นตลอดเวลา รวมถึงการสกัดข้อมูลเชิงลึกที่มีประโยชน์ต่อการดำเนินธุรกิจ การใช้เทคนิคเหมืองข้อมูลแบบไฮบริด หรือเทคนิคเหมืองข้อมูลแบบผสมผสานจึงเป็นทางเลือกที่มีประสิทธิภาพ เนื่องจากช่วยให้สามารถรวบรวม วิเคราะห์ และพยากรณ์พฤติกรรมลูกค้าได้อย่างแม่นยำมากขึ้น

ในปัจจุบัน แม้ว่าการวิเคราะห์ข้อมูลแบบเรียลไทม์จะมีประโยชน์อย่างมาก (เอกรัฐ รัชกาญจน์, 2564) แต่ก็ยังมีข้อจำกัดและอุปสรรคที่ทำให้ธุรกิจหลายแห่งไม่สามารถนำไปใช้งานได้เต็มที่ประสิทธิภาพ ปัญหาสำคัญคือการประมวลผลข้อมูลที่มีโครงสร้างแตกต่างกัน เช่น ข้อมูลคำสั่งซื้อ ข้อมูลรีวิวลินค้า (Unstructured Data) และข้อมูลพฤติกรรมการใช้งานเว็บไซต์ ซึ่งจำเป็นต้องใช้เทคนิคเหมืองข้อมูลที่สามารถจัดการกับข้อมูลหลายประเภทได้อย่างเหมาะสม (สมัคร ชัยสงวน, 2561) นอกจากนี้ ความเร็วในการประมวลผลก็เป็นอีกปัจจัยหนึ่งที่ต้องคำนึงถึง เนื่องจากข้อมูลแบบเรียลไทม์ต้องการการวิเคราะห์ที่รวดเร็วเพื่อนำไปใช้ในการตัดสินใจทางธุรกิจได้ทันที งานวิจัยนี้มุ่งเน้นการใช้เทคนิคเหมืองข้อมูลแบบไฮบริดที่ผสมผสานเทคนิคการเรียนรู้ของเครื่อง (Machine Learning) การทำเหมืองข้อมูลแบบดั้งเดิม (Traditional Data Mining) และการวิเคราะห์เชิงสถิติ (Statistical Analysis) เพื่อสร้างแนวทางการวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์ที่มีประสิทธิภาพ (Li, 2023) เพื่อให้การศึกษานี้สามารถนำไปใช้ประโยชน์ได้จริง งานวิจัยนี้มีวัตถุประสงค์หลักในการพัฒนา โมเดลการวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์สำหรับธุรกิจอีคอมเมิร์ซ โดยจะศึกษาการนำเทคนิคการทำเหมืองข้อมูลแบบไฮบริด (Hybrid Data Mining) มาประยุกต์ใช้กับข้อมูลพฤติกรรมลูกค้า รวมถึงเปรียบเทียบประสิทธิภาพของเทคนิคการวิเคราะห์ข้อมูลแบบต่างๆ เพื่อระบุวิธีการที่เหมาะสมที่สุด นอกจากนี้ยังมีการศึกษาปัจจัยที่ส่งผลต่อความแม่นยำของการวิเคราะห์ข้อมูลแบบเรียลไทม์ และวิธีการลดระยะเวลาในการประมวลผลข้อมูลเพื่อให้สามารถตอบสนองต่อการเปลี่ยนแปลงของตลาดได้อย่างรวดเร็ว (Lei, Guo & Zheng, 2024) อย่างไรก็ตาม แม้งานวิจัยจำนวนมากจะได้เสนอแนวทางการวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์โดยใช้เทคนิคเดียว เช่น การเรียนรู้ด้วยเครื่อง หรือ การเรียนรู้เชิงลึก แต่ยังไม่มีการผสมผสานเทคนิคต่าง ๆ เข้าด้วยกันในรูปแบบไฮบริด เพื่อตอบโจทย์การจัดการกับข้อมูลที่หลากหลายได้อย่างมีประสิทธิภาพ ซึ่งจุดนี้สะท้อนให้เห็นถึงช่องว่างทางวิชาการที่งานวิจัยนี้จึงมีจุดเด่น โดยการออกแบบโมเดลที่ผสานจุดแข็งของเทคนิคต่าง ๆ เพื่อให้สามารถวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์ได้อย่างแม่นยำและรวดเร็ว

นอกจากนี้ ยังพบว่างานวิจัยที่ผ่านมายังขาดการประเมินผลเชิงเปรียบเทียบระหว่างเทคนิคไฮบริดกับเทคนิคแบบเดี่ยวในบริบทของธุรกิจอีคอมเมิร์ซ ซึ่งเป็นบริบทที่มีข้อมูลขนาดใหญ่ เปลี่ยนแปลงตลอดเวลา และต้องการการตอบสนองทันทีงานวิจัยนี้จึงมีจุดเด่น

ที่ต่อยอดจากงานก่อนหน้าโดยเน้นการพัฒนาโมเดลที่สามารถประมวลผลข้อมูลขนาดใหญ่ในเวลาจริง พร้อมทั้งเปรียบเทียบกับเทคนิคดั้งเดิม เพื่อให้เห็นถึงประสิทธิภาพและความเหมาะสมของการใช้เทคนิคเหมืองข้อมูลแบบไฮบริดอย่างเป็นระบบและครบวงจรในเชิงปฏิบัติ งานวิจัยนี้มีเป้าหมายเพื่อสร้างเครื่องมือวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์ที่สามารถนำไปประยุกต์ใช้ได้จริงในธุรกิจอีคอมเมิร์ซ โดยเฉพาะในด้านการจำแนกกลุ่มลูกค้าแบบอัตโนมัติ การคาดการณ์แนวโน้มการซื้อ และการปรับกลยุทธ์การตลาดแบบทันเวลา ซึ่งจะช่วยเพิ่มอัตราการเปลี่ยนแปลง (Conversion Rate) และเพิ่มความพึงพอใจของลูกค้าได้อย่างมีประสิทธิภาพด้วยเหตุนี้งานวิจัยนี้จึงมีวัตถุประสงค์หลักเพื่อพัฒนาโมเดลการวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์ โดยใช้เทคนิคเหมืองข้อมูลแบบไฮบริดที่ผสมผสานเทคนิคต่าง ๆ เช่น การจัดกลุ่มข้อมูล การจำแนกประเภท และการวิเคราะห์อนุกรมเวลา (Time Series Analysis) และมุ่งเน้นให้ผลลัพธ์สามารถนำไปใช้งานจริงในธุรกิจอีคอมเมิร์ซ เพื่อให้สามารถตอบสนองต่อลูกค้าในเวลาจริงอย่างมีประสิทธิภาพและแม่นยำ

การวิจัยนี้เป็นประโยชน์อย่างมากทั้งในด้านการวิชาการและเชิงธุรกิจ โดยในเชิงวิชาการ ผลลัพธ์จากงานวิจัยนี้จะช่วยเติมเต็มองค์ความรู้เกี่ยวกับการใช้เทคนิคเหมืองข้อมูลแบบไฮบริดในการวิเคราะห์ข้อมูลแบบเรียลไทม์ ส่วนในเชิงธุรกิจ การพัฒนาโมเดลที่สามารถนำไปใช้งานจริงจะช่วยให้ธุรกิจอีคอมเมิร์ซสามารถเข้าใจพฤติกรรมลูกค้าได้อย่างลึกซึ้ง และสามารถออกแบบกลยุทธ์ทางการตลาดที่มีประสิทธิภาพมากขึ้น ทั้งนี้ การใช้เทคนิคเหมืองข้อมูลแบบไฮบริดไม่เพียงแต่ช่วยให้ธุรกิจสามารถวิเคราะห์พฤติกรรมลูกค้าได้อย่างแม่นยำ แต่ยังช่วยให้สามารถคาดการณ์แนวโน้มของตลาดและตอบสนองความต้องการของลูกค้าได้อย่างทันทั่วถึง ซึ่งเป็นปัจจัยสำคัญที่ช่วยให้ธุรกิจสามารถรักษาความสามารถในการแข่งขันในตลาดดิจิทัลที่เปลี่ยนแปลงอย่างรวดเร็ว

วัตถุประสงค์การวิจัย

1. เพื่อพัฒนาโมเดลการวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์ โดยใช้การผสมผสานของเทคนิคเหมืองข้อมูล
2. เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลแบบไฮบริด กับเทคนิคการวิเคราะห์ข้อมูลแบบดั้งเดิม

บททวนวรรณกรรม

ผู้วิจัยได้มีการศึกษาแนวคิด ทฤษฎีต่าง ๆ และบททวนวรรณกรรมที่เกี่ยวข้อง เพื่อให้ผู้วิจัยได้รับความรู้พื้นฐานในการทำวิจัยในครั้งนี้ ซึ่งผู้วิจัยได้พัฒนารอบแนวคิดการวิจัยดังนี้

1. แนวคิดการวิเคราะห์พฤติกรรมลูกค้า (Customer Behavior Analysis) เป็นกระบวนการศึกษาพฤติกรรมของผู้บริโภคผ่านข้อมูลที่ได้รับจากแหล่งต่าง ๆ เพื่อทำความเข้าใจปัจจัยที่มีอิทธิพลต่อการตัดสินใจซื้อสินค้าและบริการ งานวิจัยในด้านนี้ได้รับความสนใจอย่างมากจากนักวิชาการและภาคธุรกิจ เนื่องจากข้อมูลเชิงพฤติกรรมของลูกค้าสามารถนำไปใช้พัฒนากลยุทธ์ทางการตลาดและปรับปรุงประสบการณ์ของลูกค้าให้เหมาะสมมากขึ้น งานวิจัยในปัจจุบันมุ่งเน้นไปที่การใช้เทคโนโลยีที่ก้าวหน้า เช่น การวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data Analytics) การเรียนรู้ด้วยเครื่อง และการเรียนรู้เชิงลึก เพื่อวิเคราะห์พฤติกรรมลูกค้าอย่างมีประสิทธิภาพตามแนวคิดของ (Solomon, 2020)

การวิเคราะห์พฤติกรรมลูกค้าเป็นกระบวนการศึกษาการรับรู้ ตัดสินใจ และพฤติกรรมที่เกี่ยวข้องกับการซื้อสินค้าและบริการ ซึ่งได้รับอิทธิพลจากปัจจัยหลายประการ เช่น แรงจูงใจส่วนบุคคล ปัจจัยทางสังคม และอิทธิพลของแบรนด์ งานวิจัยของ Abbasimehr & Shabani (2021) เน้นว่าแนวทางการวิเคราะห์พฤติกรรมลูกค้าสามารถแบ่งออกเป็นสองแนวทางหลัก ได้แก่ พฤติกรรมเชิงปริมาณ (Quantitative Behavior) ซึ่งอาศัยข้อมูลธุรกรรม การซื้อสินค้า และอัตราการมีส่วนร่วมของลูกค้า และพฤติกรรมเชิงคุณภาพ (Qualitative Behavior) ซึ่งเน้นที่ความคิดเห็น ความรู้สึก และปฏิสัมพันธ์ของลูกค้ากับแบรนด์ (Burney, Arnold-Saritepe & McCann, 2023) จากการศึกษาโดย Afzal et al. (2024) พบว่าการใช้การแบ่งกลุ่มลูกค้า (Customer segmentation) สามารถช่วยให้ธุรกิจเข้าใจพฤติกรรมของลูกค้าในกลุ่มต่างๆ และสามารถออกแบบกลยุทธ์ที่ตอบสนองความต้องการของแต่ละกลุ่มได้ดีขึ้น นอกจากนี้ยังเสนอโมเดลการวิเคราะห์การเดินทางของลูกค้า (Customer Journey Analysis) ที่ช่วยให้นักการตลาดเข้าใจเส้นทางของลูกค้าตั้งแต่การรับรู้แบรนด์จนถึงการตัดสินใจซื้อ

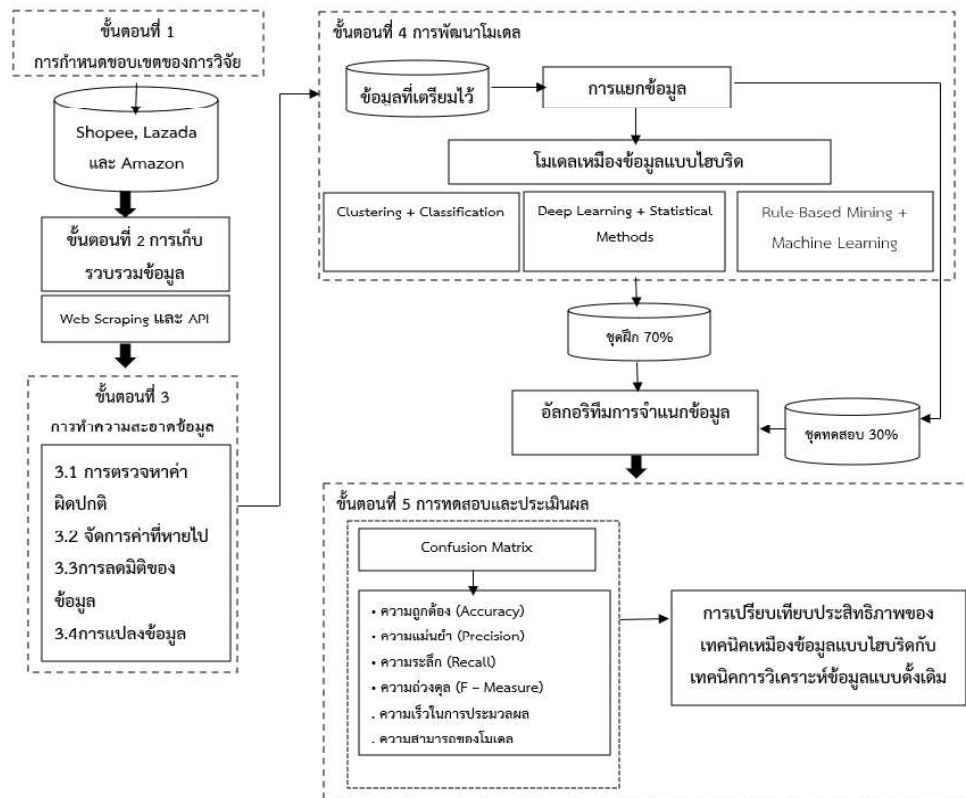
2. แนวคิดการวิเคราะห์แบบเรียลไทม์ (Real-Time Analytics) เป็นกระบวนการประมวลผลและวิเคราะห์ข้อมูลในขณะที่ข้อมูลถูกสร้างขึ้น เพื่อให้สามารถนำผลลัพธ์ไปใช้ตัดสินใจได้อย่างทันทั่วทั้งที่ เทคโนโลยีนี้มีความสำคัญอย่างยิ่งในยุคดิจิทัล เนื่องจากข้อมูลที่เกิดขึ้นแบบไดนามิก เช่น การทำธุรกรรมออนไลน์ การโต้ตอบบนสื่อสังคมออนไลน์ และข้อมูลจากเซ็นเซอร์ อินเทอร์เน็ตในทุกสรรพสิ่ง ต้องการการวิเคราะห์ที่รวดเร็วและแม่นยำเพื่อสร้างมูลค่าให้กับธุรกิจและองค์กร งานวิจัยในด้านนี้ได้รับความสนใจจากทั้งภาคธุรกิจและวิชาการ โดยเน้นไปที่การพัฒนาแพลตฟอร์มการวิเคราะห์ที่สามารถจัดการกับข้อมูลขนาดใหญ่ (Big Data) และข้อมูลที่มีการเปลี่ยนแปลงอย่างต่อเนื่อง (Alam et al. 2024) แนวคิดของการวิเคราะห์แบบเรียลไทม์ได้รับการพัฒนามาจากเทคนิคการวิเคราะห์ข้อมูลแบบดั้งเดิมที่ใช้การประมวลผลแบบแบตช์ซึ่งมีข้อจำกัดในเรื่องของความเร็วในการประมวลผลซึ่งอธิบายว่าการเปลี่ยนแปลงจากการประมวลผลแบบแบตช์ไปสู่การประมวลผลแบบสตรีมช่วยให้สามารถวิเคราะห์ข้อมูลขณะที่มันถูกสร้างขึ้นได้ทันที และส่งผลให้การตัดสินใจมีประสิทธิภาพมากขึ้น (Wolniak, 2023) ยังระบุว่าการวิเคราะห์ข้อมูลแบบเรียลไทม์ (Real-Time Analytics) เป็นองค์ประกอบสำคัญของการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data Analytics) ซึ่งช่วยให้ธุรกิจสามารถตรวจจับแนวโน้ม คาดการณ์พฤติกรรมของลูกค้า และดำเนินการตอบสนองต่อเหตุการณ์สำคัญได้ในเวลาจริง ในขณะที่การวิเคราะห์แบบเรียลไทม์มีข้อได้เปรียบในเรื่องของความเร็ว แต่ก็มีข้อท้าทายหลายประการ โดยเฉพาะด้านการจัดเก็บและประมวลผลข้อมูลที่มีปริมาณมาก (Scalability) ความถูกต้องของข้อมูล (Data Accuracy) และความปลอดภัยของข้อมูล (Data Security) เพื่อให้สามารถประมวลผลข้อมูลขนาดใหญ่ได้อย่างมีประสิทธิภาพ

3. แนวคิดธุรกิจอีคอมเมิร์ซ (E-Commerce) เป็นหนึ่งในรูปแบบธุรกิจที่มีการเติบโตอย่างรวดเร็วในยุคดิจิทัล ด้วยการพัฒนาเทคโนโลยีอินเทอร์เน็ต ความก้าวหน้าของเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence: AI) และการเปลี่ยนแปลงพฤติกรรมของผู้บริโภคที่มุ่งเน้นความสะดวกสบายมากขึ้น ธุรกิจอีคอมเมิร์ซสามารถแบ่งออกเป็นหลายประเภท เช่น ธุรกิจถึงผู้บริโภค (B2C) ธุรกิจต่อธุรกิจ (B2B) ผู้บริโภคถึงผู้บริโภค (C2C) และจากผู้บริโภคสู่ธุรกิจ (C2B) (Chaffey & Smith, 2022) งานวิจัยในด้านอีคอมเมิร์ซครอบคลุมตั้งแต่ปัจจัยที่ส่งผลต่อพฤติกรรมผู้บริโภค กลยุทธ์ทางการตลาดดิจิทัล การใช้เทคโนโลยีในการเพิ่มประสิทธิภาพการดำเนินธุรกิจ และแนวโน้มในอนาคตของอุตสาหกรรม ธุรกิจอีคอมเมิร์ซได้รับการศึกษาจากหลากหลายมิติ มีนักวิจัยหลายคนได้เน้นไปที่ปัจจัยที่มีอิทธิพลต่อความสำเร็จของธุรกิจอีคอมเมิร์ซ (Laudon & Traver, 2021) ระบุปัจจัยที่สำคัญในการดำเนินธุรกิจอีคอมเมิร์ซ ได้แก่ โครงสร้างพื้นฐานทางเทคโนโลยี ระบบโลจิสติกส์ การตลาดดิจิทัล และความไว้วางใจของลูกค้าซึ่งเป็นปัจจัยสำคัญที่ช่วยเพิ่มอัตราการแปลง (Conversion Rate) และสร้างความภักดีของลูกค้า ในเชิงพฤติกรรมผู้บริโภค (Baranauskas & Raišienė, 2021) ได้ศึกษาถึงกระบวนการตัดสินใจของลูกค้าและพบว่า การตัดสินใจซื้อสินค้าผ่านแพลตฟอร์มอีคอมเมิร์ซได้รับอิทธิพลจากปัจจัยหลัก 3 ประการ ได้แก่ 1) ราคาและโปรโมชั่น (Pricing & Promotions) 2) ความน่าเชื่อถือของเว็บไซต์ (Website Trustworthiness) และ 3) บริการหลังการขาย (After-sales Service) และยังชี้ให้เห็นว่าปัจจัยด้านสังคมและคำแนะนำจากผู้ใช้ (Social Influence & User Reviews) มีบทบาทสำคัญในการกระตุ้นให้ลูกค้าตัดสินใจซื้อด้วยเช่นกัน

4. แนวคิดเทคนิคเหมืองข้อมูลแบบไฮบริด (Hybrid Data Mining Techniques) เป็นแนวทางที่ได้รับความนิยมอย่างมากเนื่องจากความซับซ้อนของข้อมูลที่เพิ่มขึ้นและความต้องการผลลัพธ์ที่มีความแม่นยำและรวดเร็วมากขึ้น การทำเหมืองข้อมูลแบบไฮบริดเป็นการผสมผสานเทคนิคเหมืองข้อมูลหลายประเภท เช่น การเรียนรู้ด้วยเครื่อง การเรียนรู้เชิงลึก การวิเคราะห์ทางสถิติ (Statistical Analysis) การขุดตามกฎเกณฑ์ (Rule-Based Mining) และอัลกอริทึมเชิงวิวัฒนาการ (Evolutionary Algorithms) เพื่อให้สามารถสกัดข้อมูลเชิงลึกได้อย่างมีประสิทธิภาพสูงสุด งานวิจัยในด้านนี้ครอบคลุมทั้งการพัฒนาโมเดลใหม่ ๆ การปรับปรุงกระบวนการทำเหมืองข้อมูล และการนำไปใช้ในภาคธุรกิจและอุตสาหกรรมต่างๆ การทำเหมืองข้อมูลแบบไฮบริดเกิดจากข้อจำกัดของเทคนิคเหมืองข้อมูลแบบดั้งเดิมที่มีประสิทธิภาพต่ำเมื่อต้องประมวลผลข้อมูลที่มีโครงสร้างซับซ้อน Seyyedabbasi et al. (2021) อธิบายว่าเทคนิคเหมืองข้อมูลแบบไฮบริดสามารถแบ่งออกเป็น 3 รูปแบบ ได้แก่ 1) การรวมหลายอัลกอริทึมเข้าด้วยกัน (Hybrid Algorithm Combination) เช่น การใช้ต้นไม้สุ่ม (Random Forest) ร่วมกับโครงข่ายประสาทเทียม (Neural Networks) 2) การรวมหลายกระบวนการเหมืองข้อมูล (Hybrid Process Combination) เช่น การใช้วิธีการจัดกลุ่ม (Clustering) ก่อนทำการจำแนกประเภท (Classification) และ 3) การรวมหลายแหล่งข้อมูล (Hybrid Data Combination) เช่น การผสมผสานข้อมูลที่มีโครงสร้างและข้อมูลที่ไม่มี

โครงสร้าง ซึ่งแสดงให้เห็นว่าเทคนิคไฮบริดช่วยเพิ่มความถูกต้อง (Accuracy) ความแม่นยำ (Precision) ความระลึก (Recall) และความถ่วงดุล (F1-Score) ได้ดีกว่าเทคนิคเดี่ยว เช่น ต้นไม้ตัดสินใจ (Decision Trees) หรือซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines)

จากการทบทวนวรรณกรรมข้างต้น การวิจัยเรื่อง การวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์สำหรับธุรกิจอีคอมเมิร์ซด้วยเทคนิคเหมืองข้อมูลแบบไฮบริด มีเป้าหมายเพื่อพัฒนาโมเดลการวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์โดยใช้เทคนิคเหมืองข้อมูลแบบไฮบริด และเปรียบเทียบประสิทธิภาพของโมเดลกับเทคนิคการวิเคราะห์ข้อมูลแบบดั้งเดิม ผู้วิจัยได้พัฒนารอบแนวคิด ดังนี้



ภาพที่ 1 รอบแนวคิดการวิจัย

วิธีดำเนินการวิจัย

งานวิจัยเรื่อง การวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์สำหรับธุรกิจอีคอมเมิร์ซด้วยเทคนิคเหมืองข้อมูลแบบไฮบริดเป็นงานวิจัยเชิงประยุกต์ (Applied Research) และงานวิจัยเชิงทดลอง (Experimental Research) โดยมีวิธีการดำเนินการวิจัย 5 ขั้นตอนหลัก ดังต่อไปนี้

ขั้นตอนที่ 1 การกำหนดขอบเขตของการวิจัย ในขั้นตอนแรกของการวิจัย ผู้วิจัยได้ดำเนินการกำหนดขอบเขตของการศึกษาอย่างชัดเจน เพื่อให้แน่ใจว่าข้อมูลที่ใช้และกระบวนการวิเคราะห์จะสอดคล้องกับวัตถุประสงค์ของงานวิจัยอย่างแท้จริง โดยมีเป้าหมายหลักเพื่อให้สามารถพัฒนาโมเดลการวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์ที่มีความถูกต้องและนำไปใช้ได้จริงในภาคธุรกิจ ข้อมูลที่นำมาวิเคราะห์ถูกเลือกจากแหล่งข้อมูลที่เกี่ยวข้องกับพฤติกรรมการซื้อของลูกค้าในบริบทของธุรกิจอีคอมเมิร์ซ ซึ่งประกอบด้วยแพลตฟอร์มที่มีความนิยมและมีฐานลูกค้าขนาดใหญ่ ได้แก่ ซอปปี้ ลาซาด้า และแอมะซอน ทั้งสามแพลตฟอร์มนี้มีความสามารถในการให้ข้อมูลที่ครอบคลุมเกี่ยวกับการตัดสินใจซื้อ เช่น รีวิวสินค้า ความคิดเห็นของลูกค้า และพฤติกรรมการเลือกซื้อในหมวดหมู่ต่างๆ นอกจากข้อมูลจากแพลตฟอร์มอีคอมเมิร์ซแล้ว ผู้วิจัยยังใช้ข้อมูลจากสื่อสังคมออนไลน์ เช่น ข้อความแสดงความคิดเห็น

บนเว็บไซต์ผลิตภัณฑ์หรือโพสต์ที่เกี่ยวข้องกับประสบการณ์ของลูกค้า ซึ่งช่วยเสริมความเข้าใจในเชิงคุณภาพของพฤติกรรมผู้บริโภค ได้อย่างลึกซึ้งมากขึ้น อาทิเช่น ใช้งานตีราคาคุณค่า ส่งของไว้มาก เป็นต้น (Chung et al., 2022)

ข้อมูลทั้งหมดที่ใช้ในการศึกษานี้ครอบคลุมทั้ง ข้อมูลที่มีโครงสร้าง เช่น หมายเลขคำสั่งซื้อ หมวดหมู่สินค้า คะแนนเรตติ้ง ข้อมูลไม่มีโครงสร้าง เช่น ข้อความรีวิวสินค้า รูปภาพประกอบความคิดเห็น และข้อมูลกึ่งโครงสร้าง (Semi-structured Data) เช่น คำค้น (Keywords) แท็กสินค้า หรือข้อมูล HTML ที่จัดเก็บผ่านการสกรีนและดึงข้อมูลจากเว็บไซต์

ขั้นตอนที่ 2 การเก็บรวบรวมข้อมูล ในการเก็บรวบรวมข้อมูลครั้งนี้ ผู้วิจัยได้ใช้วิธีการที่หลากหลายเพื่อให้ได้ข้อมูลที่ครอบคลุม พฤติกรรมผู้บริโภคบนแพลตฟอร์มดิจิทัล โดยเลือกใช้เทคนิคการสกัดและดึงข้อมูลจากเว็บไซต์ และ API ในการดึงข้อมูลจากแพลตฟอร์ม อีคอมเมิร์ซ ได้แก่ ซอปปี้ ลาซาด้า และแอมะซอน รวมถึงแพลตฟอร์มโซเชียลมีเดีย เช่น เพจบุ๊กซึ่งมีข้อมูลรีวิวสินค้า และความคิดเห็นของผู้ใช้งานจำนวนมาก ทั้งนี้เพื่อให้การรวบรวมข้อมูลเป็นไปอย่างมีประสิทธิภาพในเชิงเวลาและปริมาณ ผู้วิจัยยังได้ใช้เทคนิค Streaming Data Processing โดยใช้เครื่องมือ เช่น Apache Kafka, Apache Flink และ Spark Streaming สำหรับการประมวลผล ข้อมูลแบบไหลเข้า (Real-time) ซึ่งช่วยให้สามารถจัดการกับข้อมูลที่ถูกรวบรวมขึ้นอย่างต่อเนื่องบนโลกออนไลน์ได้อย่างมีประสิทธิภาพ (Whang & Lee, 2020)

ข้อมูลที่ได้จากกระบวนการดังกล่าวถูกจัดเก็บไว้ในระบบฐานข้อมูลแบบ NoSQL ได้แก่ MongoDB และ Google Big Query ซึ่งสามารถรองรับข้อมูลขนาดใหญ่และไม่มีโครงสร้าง เช่น ข้อความรีวิว รูปภาพ และข้อมูลหมวดหมู่ที่หลากหลายได้อย่างยืดหยุ่น จากกระบวนการนี้ ได้ข้อมูลทั้งหมดจำนวน 98,910 เรคคอร์ด ครอบคลุม 59 แอททริบิวต์ ที่สะท้อนพฤติกรรมการซื้อ และความคิดเห็นของลูกค้า ซึ่งถือเป็นฐานข้อมูลสำคัญที่จะนำไปใช้ในขั้นตอนการวิเคราะห์เพื่อพัฒนาโมเดลพฤติกรรมลูกค้าแบบเรียลไทม์ในลำดับต่อไป

ขั้นตอนที่ 3 การทำความสะอาดข้อมูล เพื่อให้ข้อมูลที่นำมาใช้วิเคราะห์มีความถูกต้องและสมบูรณ์ที่สุด ผู้วิจัยเริ่มจากการตรวจสอบความผิดปกติของข้อมูลด้วยกระบวนการ Outlier Detection โดยใช้เทคนิคเชิงสถิติเช่น Z-Score และ IQR (Interquartile Range) เพื่อระบุค่าที่อยู่นอกช่วงข้อมูลปกติอย่างชัดเจน จากนั้นดำเนินการ จัดการค่าที่หายไป (Handling Missing Values) ด้วยวิธีที่เหมาะสมตามลักษณะข้อมูล ได้แก่ การแทนที่ด้วยค่าเฉลี่ย (Mean Imputation) หรือค่าสื่อกลาง (Median Imputation) ในกรณีของข้อมูลเชิงตัวเลข และการใช้ค่าที่พบบ่อยที่สุด (Mode) สำหรับข้อมูลประเภทจำแนก ทั้งนี้เพื่อให้ลดผลกระทบจากข้อมูลที่ไม่สมบูรณ์ต่อความแม่นยำของโมเดล (Côté et al. 2024)

หลังจากได้ข้อมูลที่สะอาดและครบถ้วนแล้ว ผู้วิจัยจึงดำเนินการ ลดมิติของข้อมูล (Dimensionality Reduction) โดยใช้เทคนิคการวิเคราะห์องค์ประกอบ (Principal Component Analysis: PCA) ซึ่งเป็นวิธีการวิเคราะห์องค์ประกอบหลัก เพื่อช่วยลดจำนวนตัวแปรโดยไม่สูญเสียสาระสำคัญของข้อมูล โดยจากข้อมูลทั้งหมด 59 แอททริบิวต์หลังการใช้เทคนิคการวิเคราะห์องค์ประกอบสามารถลดเหลือ 18 องค์ประกอบหลัก (Principal Components) โดยยังสามารถอธิบายความแปรปรวนของข้อมูลได้มากกว่าร้อยละ 95 สำหรับข้อมูลข้อความ (Text Data) ได้มีการใช้เทคนิค Text Encoding โดยแปลงข้อความเป็นค่าตัวเลขผ่านการเข้ารหัส เช่น TF-IDF (Term Frequency-Inverse Document Frequency) และเทคนิคเวกเตอร์เชิงความหมาย เช่น Word Embeddings ได้แก่ Word2Vec และ BERT ซึ่งช่วยให้สามารถนำไปใช้ในการประมวลผลและวิเคราะห์โดยโมเดลได้อย่างมีประสิทธิภาพ กระบวนการเหล่านี้ได้รับการตั้งค่าและประเมินผลอย่างเป็นระบบ เพื่อให้สามารถทดลองซ้ำและปรับใช้ในกรณีศึกษาที่คล้ายคลึงกันได้อย่างถูกต้อง

ขั้นตอนที่ 4 การพัฒนาโมเดลเหมืองข้อมูลแบบไฮบริด (Hybrid Data Mining Model Development) หลังจากที่ข้อมูลได้รับการเตรียมไว้อย่างสมบูรณ์ในขั้นตอนก่อนหน้านี้ ขั้นตอนถัดมาคือการพัฒนาโมเดลเหมืองข้อมูลแบบไฮบริด (Hybrid Data Mining Model) เพื่อใช้ในการวิเคราะห์พฤติกรรมของลูกค้า โดยโมเดลจะถูกพัฒนาโดยอิงจาก ข้อมูลที่ผ่านการแปลงและทำความสะอาดแล้วทั้งหมดซึ่งจะถูกแบ่งออกเป็น 2 ชุด ได้แก่ ชุดฝึก (Training Set) ร้อยละ 70 และ ชุดทดสอบ (Test Set) ร้อยละ 30 เพื่อใช้ในการประเมินประสิทธิภาพของโมเดลในขั้นตอนต่อไป

ในขั้นตอนการพัฒนาโมเดล ผู้วิจัยออกแบบกระบวนการวิเคราะห์แบบไฮบริด (Hybrid Approach) การผสมผสานเทคนิคที่มีจุดเด่นแตกต่างกันเพื่อเพิ่มความแม่นยำในการทำนายพฤติกรรมลูกค้า โดยทำการปรับจูนพารามิเตอร์ ดังแสดงในตารางที่ 1

ตารางที่ 1 แนวทาง Hybrid Approach ในการพัฒนาโมเดล

Hybrid Approach	เทคนิคการปรับปรุงพารามิเตอร์	วัตถุประสงค์
Clustering ร่วมกับ Classification	K-Means (k=5), DBSCAN ($\epsilon=0.4$, min_samples=10), Decision Tree (gini, max_depth=10), Random Forest (n_estimators=100, max_depth=15)	แบ่งกลุ่มลูกค้าเบื้องต้น แล้วทำนายความถี่ในการซื้อ หรือความภักดี
Deep Learning ร่วมกับ Statistical Methods	LSTM (units=128, dropout=0.2, epochs=30, batch_size=64), Transformer (heads=8, d_model=512), Logistic Regression (l2, lbfgs), Bayesian Models	วิเคราะห์พฤติกรรมตามเวลา และประเมินโอกาสในการซื้อ
Rule-Based Mining ร่วมกับ Machine Learning	Apriori (min_support=0.03, min_confidence=0.5), FP-Growth, Gradient Boosting (lr=0.1, n_estimators=200), XGBoost (max_depth=6, lr=0.05, n_estimators=150)	ค้นหากฎความสัมพันธ์และเพิ่มความแม่นยำในการคาดการณ์

จากตารางที่ 1 แนวทาง Hybrid Approach ในการพัฒนาโมเดลมีความสามารถสูงในเรื่องของการทำนายและจัดการกับข้อมูลที่ไม่เป็นเชิงเส้นและซับซ้อนสูง ทั้งสามรูปแบบของโมเดลได้รับการประเมินผลด้วยชุดข้อมูลเดียวกันเพื่อเปรียบเทียบประสิทธิภาพอย่างเป็นธรรม

ขั้นตอนที่ 5 การทดสอบและประเมินผล (Model Evaluation & Comparison) ในขั้นตอนการทดสอบและประเมินผลโมเดล ผู้วิจัยได้แบ่งชุดข้อมูลออกเป็น 2 ส่วน ได้แก่ ส่วนที่ 1 คือ Training Set ร้อยละ 70 และส่วนที่ 2 คือ Test Set ร้อยละ 30 โดยใช้การแบ่งแบบสุ่ม (Random Split) เพื่อให้มั่นใจว่าข้อมูลทั้งสองชุดมีการกระจายตัวของคุณลักษณะอย่างสม่ำเสมอ นอกจากนี้ยังได้ใช้เทคนิค k-fold cross-validation (k=5) เพื่อประเมินความเสถียรและความแม่นยำของโมเดลในรอบการทดลองที่ต่างกัน ช่วยลดโอกาสของการ overfitting และเพิ่มความน่าเชื่อถือของผลลัพธ์ที่ได้ โมเดลแบบการเรียนรู้แบบมีผู้สอน (Supervised Learning) ทั้งหมดจะถูกวัดผลโดยใช้เกณฑ์มาตรฐาน ได้แก่ Accuracy, Precision, Recall และ F1-Score เพื่อประเมินทั้งความถูกต้องโดยรวมและความสามารถในการแยกแยะระหว่างคลาสอย่างแม่นยำ (Varoquaux & Colliot, 2023)

ในกรณีของโมเดลที่เน้นการประมวลผลแบบเรียลไทม์ เช่น โมเดลที่ใช้เทคนิคสตรีมมิ่ง หรือการเรียนรู้เชิงลึก ร่วมกับ การวิเคราะห์อนุกรมเวลา จะถูกประเมินเพิ่มเติมโดยใช้ความเร็วในการประมวลผล (Processing Speed) และเวลาแฝง (Latency) ซึ่งวัดระยะเวลาตอบสนองต่อข้อมูลที่เข้ามาแบบต่อเนื่อง รวมถึง Scalability ของระบบผ่านการทดสอบกับชุดข้อมูลขนาดต่าง ๆ เพื่อดูว่าระบบสามารถขยายขนาดได้โดยไม่สูญเสียประสิทธิภาพ สุดท้าย ผู้วิจัยได้นำผลลัพธ์ของ Hybrid Data Mining Model มาเปรียบเทียบกับโมเดลแบบดั้งเดิม เช่น ต้นไม้ตัดสินใจ เบย์แบบง่าย และการถดถอยโลจิสติก (Logistic Regression) ซึ่งใช้พารามิเตอร์เดียวกันในการฝึกฝนและทดสอบ เพื่อพิจารณาว่าแนวทางแบบไฮบริดมีความได้เปรียบทั้งในด้านความแม่นยำ ความเร็ว และความสามารถในการปรับใช้กับสถานการณ์จริงหรือไม่

ผลการวิจัย

1. ผลการศึกษาและพัฒนาโมเดลการวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์ พบว่าการใช้เทคนิคเหมืองข้อมูลแบบไฮบริดสามารถช่วยเพิ่มประสิทธิภาพของการวิเคราะห์โมเดลที่พัฒนาประกอบด้วยสามแนวทางหลัก ดังตารางที่ 2

ตารางที่ 2 ผลการพัฒนาโมเดลการวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์

เทคนิค	จุดเด่นของโมเดล	ข้อจำกัด	ความถูกต้อง ของโมเดล	ความเร็ว ในการ ประมวลผล (วินาที)
Clustering ร่วมกับ Classification (K-Means ร่วมกับ Random Forest)	สามารถจำแนกลูกค้าออกเป็นกลุ่ม ได้อย่างแม่นยำ และสามารถ พยากรณ์พฤติกรรมของลูกค้าแต่ละ กลุ่มได้ดีขึ้น	ต้องใช้ทรัพยากร คอมพิวเตอร์สูงเมื่อมีข้อมูลจำนวน มาก	92.50	1.80
Deep Learning ร่วมกับ Statistical Methods (LSTM ร่วมกับ Logistic Regression)	สามารถวิเคราะห์ข้อมูลพฤติกรรม ของลูกค้าในลักษณะ Time-Series และมีประสิทธิภาพสูงในการ คาดการณ์	ต้องใช้ชุดข้อมูลที่มีการเรียงลำดับ ทางเวลาอย่างเหมาะสม เพื่อให้ได้ ผลลัพธ์ที่ดีที่สุด	89.70	2.10
Rule-Based Mining ร่วมกับ Machine Learning (Apriori ร่วมกับ XGBoost)	สามารถสร้างกฎเชื่อมโยงที่แม่นยำ ระหว่างพฤติกรรมซื้อของลูกค้า และช่วยให้การพยากรณ์พฤติกรรมมี ความแม่นยำสูงขึ้น	ต้องปรับแต่งพารามิเตอร์ของกฎ เชื่อมโยงให้เหมาะสม มิฉะนั้นอาจ เกิด Overfitting	88.30	2.40

จากตารางที่ 2 ผลการพัฒนาโมเดลด้วยเทคนิคเหมืองข้อมูลแบบไฮบริด สามารถสรุปได้ว่า การใช้เทคนิค Clustering ร่วมกับ Classification (K-Means ร่วมกับ Random Forest) มีความแม่นยำสูงสุดร้อยละ 92.50 เมื่อเทียบกับเทคนิคอื่น ๆ และสามารถแบ่งกลุ่มลูกค้าได้อย่างมีประสิทธิภาพ ซึ่งช่วยให้การพยากรณ์พฤติกรรมลูกค้าเป็นไปอย่างแม่นยำมากขึ้น อย่างไรก็ตามเทคนิคนี้ต้องใช้ทรัพยากรคอมพิวเตอร์สูงเมื่อมีข้อมูลจำนวนมาก สำหรับเทคนิค Deep Learning ร่วมกับ Statistical Methods (LSTM ร่วมกับ Logistic Regression) มีความแม่นยำอยู่ที่ร้อยละ 89.70 ซึ่งเหมาะกับข้อมูลพฤติกรรมลูกค้าที่มีลำดับทางเวลา (Time-Series) โดยสามารถช่วยในการคาดการณ์แนวโน้มของพฤติกรรมลูกค้าได้อย่างดี แต่ข้อจำกัดของเทคนิคนี้ คือจำเป็นต้องใช้ชุดข้อมูลที่มีการจัดลำดับเวลาอย่างเหมาะสม มิฉะนั้นโมเดลอาจให้ผลลัพธ์ที่ไม่แม่นยำ อีกทั้งเทคนิคที่สาม Rule-Based Mining ร่วมกับ Machine Learning (Apriori ร่วมกับ XGBoost) มีความแม่นยำอยู่ที่ร้อยละ 88.30 โดยเน้นการสร้างกฎเชื่อมโยงที่สามารถระบุพฤติกรรมการซื้อของลูกค้า เช่น พฤติกรรมการซื้อ การคลิกสินค้า ซึ่งสามารถช่วยให้การพยากรณ์แนวโน้มการซื้อของลูกค้าแม่นยำขึ้น ข้อจำกัดของเทคนิคนี้คือจำเป็นต้องปรับแต่งพารามิเตอร์ของกฎเชื่อมโยงให้เหมาะสมเพื่อหลีกเลี่ยงปัญหา Overfitting

ในมุมมองของความเร็วในการประมวลผล พบว่า Clustering ร่วมกับ Classification มีความเร็วสูงสุด (1.80 วินาที) รองลงมาคือ Deep Learning ร่วมกับ Statistical Methods (2.10 วินาที) และสุดท้าย Rule-Based Mining ร่วมกับ Machine Learning (2.40 วินาที) ซึ่งแสดงให้เห็นว่าแม้เทคนิค Deep Learning และ Rule-Based Mining จะให้ผลลัพธ์ที่แม่นยำในระดับที่ดี แต่ต้องแลกมาด้วยเวลาในการประมวลผลที่มากขึ้น ในการวิเคราะห์ในครั้งนี้ ผู้วิจัยใช้คอมพิวเตอร์ Acer Intel(R) Core(TM) i3-N305 1.80 GHz Ram 8.00 GB

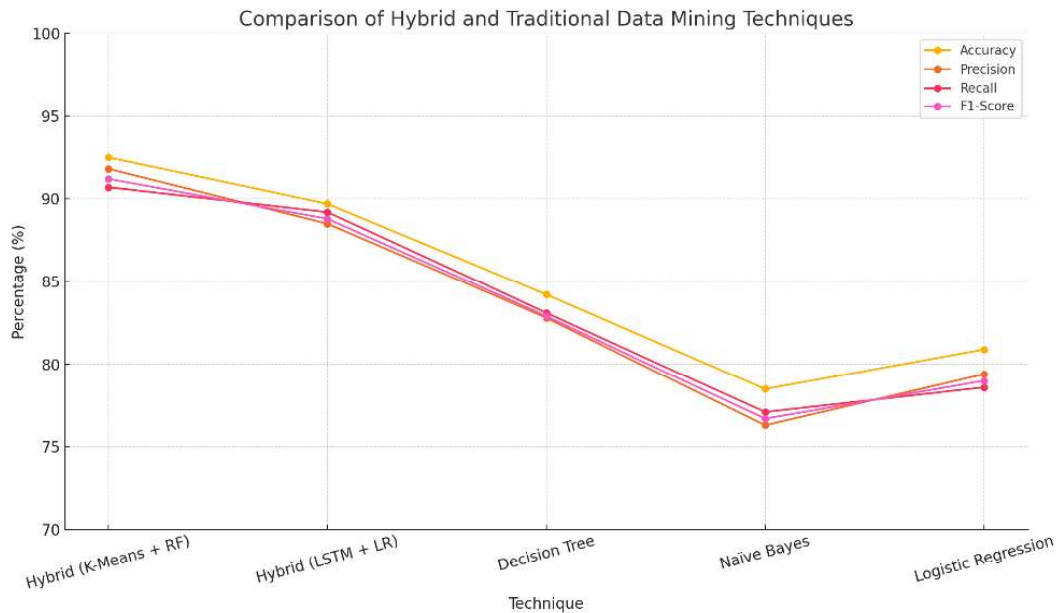
2. จากการเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลแบบไฮบริดกับเทคนิคการวิเคราะห์ข้อมูลแบบดั้งเดิม การประเมินผลลัพธ์ของโมเดลโดยใช้ตัวชี้วัดมาตรฐาน ซึ่งมีผลการเปรียบเทียบดังตารางที่ 3

ตารางที่ 3 ผลการเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลแบบไฮบริดกับเทคนิคการวิเคราะห์ข้อมูลแบบดั้งเดิม

เทคนิค	ความถูกต้อง	ความแม่นยำ	ความระลึก	ความถ่วงดุล	ความเร็วในการประมวลผล (วินาที)	ความสามารถของโมเดล
Hybrid Data Mining (K-Means ร่วมกับ Random Forest)	92.50	91.81	90.70	91.20	1.80	สูง
Hybrid Data Mining (LSTM ร่วมกับ Logistic Regression)	89.70	88.50	89.20	88.80	2.10	สูง
Traditional Decision Tree	84.20	82.80	83.10	82.90	3.4	ปานกลาง
Naïve Bayes	78.50	76.30	77.10	76.70	2.90	ปานกลาง
Logistic Regression	80.90	79.40	78.60	79.00	2.70	ปานกลาง

จากตารางที่ 3 ผู้วิจัยได้คัดเลือกเทคนิค Hybrid จำนวน 2 รูปแบบหลัก ได้แก่ Hybrid Data Mining (K-Means ร่วมกับ Random Forest) และ Hybrid Data Mining (LSTM ร่วมกับ Logistic Regression) เพื่อเปรียบเทียบกับเทคนิคแบบดั้งเดิม ได้แก่ Decision Tree, Naïve Bayes และ Logistic Regression เนื่องจากทั้งสองเทคนิค Hybrid ที่เลือกมีลักษณะโครงสร้างที่แตกต่างกัน กล่าวคือ แนวทางแรกผสมผสานการจัดกลุ่ม (Clustering) กับการจำแนกประเภท (Classification) ในขณะที่แนวทางที่สองผสมผสานการวิเคราะห์เชิงลำดับเวลา (Time Series) เข้ากับวิธีทางสถิติ เพื่อให้การเปรียบเทียบสะท้อนถึงประสิทธิภาพของ Hybrid Approach ในมิติที่หลากหลาย ในการประเมินประสิทธิภาพโมเดล เพื่อให้ผลการประเมินมีความน่าเชื่อถือ ผู้วิจัยได้แบ่งชุดข้อมูลออกเป็น ชุดฝึกสอน (Training Set) และ ชุดทดสอบ (Testing Set) ในอัตราส่วน 70 และ 30 โดยการสุ่มแบบ Stratified Sampling เพื่อรักษาสัดส่วนของคลาสข้อมูล และนอกจากนี้ ได้มีการใช้เทคนิค k-fold cross-validation (k = 5) เพื่อประเมินค่าประสิทธิภาพในเบื้องต้น และนำค่าเฉลี่ยจากการทดลองแต่ละรอบมาใช้ประกอบในการวิจัย ทั้งนี้ค่าประสิทธิภาพที่แสดงในตารางที่ 3 มาจากค่าเฉลี่ยผลลัพธ์ของ k-fold cross-validation จำนวน 5 รอบ เพื่อให้มั่นใจว่าโมเดลไม่มี bias จากการแบ่งข้อมูลเพียงครั้งเดียว และสามารถสะท้อนความแม่นยำของโมเดลได้อย่างแท้จริงทั้งนี้ ความเร็วในการประมวลผลถูกวัดจาก ระยะเวลาเฉลี่ยต่อรอบของการประเมินด้วยสเปกเครื่องคอมพิวเตอร์ที่ระบุไว้ท้ายตารางที่ 2

ส่วนเทคนิค Hybrid ที่ใช้ Rule-Based Mining แม้จะมีจุดเด่นในการค้นหารูปแบบเชิงความสัมพันธ์แต่ไม่เหมาะสมสำหรับการนำมาเปรียบเทียบกับเทคนิคที่ใช้ในการทำนายแบบการเรียนรู้แบบมีผู้สอน เนื่องจากไม่ได้มุ่งเน้นการทำนายเชิงพฤติกรรมในระดับบุคคลอีก ทั้งยังมีข้อจำกัดในการจัดการกับข้อมูลที่มีจำนวนมิติมากและเปลี่ยนแปลงต่อเนื่อง จึงไม่สามารถให้ผลลัพธ์ที่สอดคล้องกับเป้าหมายของการวิจัยในด้านการวิเคราะห์ข้อมูลแบบเรียลไทม์ ด้วยเหตุนี้ ผู้วิจัยจึงเลือกไม่นำเทคนิคแบบ Rule-Based มาอยู่ในตารางเปรียบเทียบหลัก แต่ยังคงพิจารณาในลักษณะเสริมสำหรับการวิเคราะห์รูปแบบ (Pattern Discovery) แยกต่างหากจากผลการวิจัยผู้วิจัยขอเสนอผลการเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลแบบไฮบริดกับเทคนิคการวิเคราะห์ข้อมูลแบบดั้งเดิมในรูปแบบกราฟ ดังภาพที่ 2



ภาพที่ 2 ผลการเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูล

จากภาพที่ 2 สามารถสรุปได้ว่า โมเดลของเทคนิคเหมืองข้อมูลแบบไฮบริดมีความถูกต้องและประสิทธิภาพสูงกว่าการใช้เทคนิคแบบเดี่ยว โดยเฉพาะ K-Means ร่วมกับ Random Forest มีค่าความถูกต้องร้อยละ 92.50 และ LSTM ร่วมกับ Logistic Regression มีค่าความถูกต้องร้อยละ 89.70 ที่ให้ผลลัพธ์ที่ดีในแง่ของค่าความถูกต้อง และความถ่วงดุล นอกจากนี้ ความเร็วในการประมวลผลของโมเดลเทคนิคเหมืองข้อมูลแบบไฮบริดยังเร็วกว่าเทคนิคดั้งเดิมเนื่องจากสามารถทำการประมวลผลแบบขนาน และการวิเคราะห์สตรีมมิ่งเพื่อลดเวลาในการประมวลผลข้อมูล จากผลการวิเคราะห์เชิงเปรียบเทียบ พบว่า

1. เหมืองข้อมูลแบบไฮบริดสามารถรวมข้อดีของแต่ละเทคนิคเข้าไว้ด้วยกัน ทำให้สามารถวิเคราะห์ข้อมูลได้ทั้งเชิงโครงสร้างและไม่มีโครงสร้าง
2. การใช้การเรียนรู้เชิงลึก ร่วมกับการเรียนรู้ด้วยเครื่องทำให้สามารถวิเคราะห์ข้อมูลได้แม่นยำขึ้นโดยเฉพาะเมื่อข้อมูลมีความซับซ้อนและมีความเปลี่ยนแปลงแบบไดนามิก
3. เทคนิคแบบดั้งเดิม เช่น ต้นไม้ตัดสินใจ และเบย์แบบง่ายอาจให้ผลลัพธ์ที่รวดเร็วกว่าในบางกรณี แต่มีข้อจำกัดด้านความแม่นยำและความสามารถในการรองรับข้อมูลขนาดใหญ่

สรุปและอภิปรายผลการวิจัย

1. ผลการพัฒนาโมเดลโดยใช้เทคนิคเหมืองข้อมูลแบบไฮบริดแสดงให้เห็นว่า สามารถเพิ่มประสิทธิภาพในการวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์ได้อย่างมีนัยสำคัญ โดยโมเดลการจัดกลุ่มร่วมกับการจำแนกประเภทมีค่าความถูกต้องสูงสุดร้อยละ 92.50 ซึ่งเหมาะสำหรับการจำแนกกลุ่มลูกค้าได้อย่างแม่นยำ ขณะที่การเรียนรู้เชิงลึก มีค่าความถูกต้อง ร้อยละ 89.70 และ Rule-Based Mining มีค่าความถูกต้อง ร้อยละ 88.30 ซึ่งมีความเหมาะสมในการพยากรณ์แนวโน้มพฤติกรรมของลูกค้า โมเดลเหล่านี้สามารถนำไปประยุกต์ใช้ในกลยุทธ์การตลาดดิจิทัลและการบริหารจัดการลูกค้าของธุรกิจอีคอมเมิร์ซได้อย่างมีประสิทธิภาพ ผลการศึกษานี้สอดคล้องกับงานวิจัยของ Sreenivasa & Nirmala (2019) งานวิจัยนี้นำเสนอโมเดลการแนะนำสินค้าแบบไฮบริดที่เน้นข้อมูลตามตำแหน่ง (Hybrid Location-Centric Prediction: HLCP) ซึ่งสามารถเรียนรู้พฤติกรรมของผู้ใช้งานทั้งในระยะสั้นและระยะยาวได้อย่างมีประสิทธิภาพ โดยคำนึงถึงลักษณะพฤติกรรมที่เปลี่ยนแปลงตามบริบท เช่น สถานที่หรือสภาพอากาศ ผลการทดลองแสดงให้เห็นว่า HLCP มีประสิทธิภาพสูงกว่าวิธีเดิมในการทำนายความต้องการของผู้ใช้ โดยเฉพาะในด้านค่า Mean Reciprocal Rank (MRR) และ Hit

Rate (HR) โดยให้ค่าความแม่นยำร้อยละ 94.35 นอกจากนี้ งานวิจัยของ Li & Lei (2022) ได้นำเสนอโมเดลแบบไฮบริดและการรวมข้อมูลสำหรับการวิเคราะห์ข้อมูลขนาดใหญ่ของอีคอมเมิร์ซ โดยโมเดลการแนะนำสินค้าในธุรกิจอีคอมเมิร์ซด้วยข้อมูลขนาดใหญ่ โดยอิงจากเทคโนโลยีการหลอมรวมข้อมูล (Data Fusion) และอินเทอร์เน็ตในทุกสรรพสิ่ง (Internet of Things) โดยมีการเลือกใช้ระบบฝังตัว (Embedded System) ในการพัฒนาแพลตฟอร์มอีคอมเมิร์ซ ซึ่งทำหน้าที่รวบรวม ส่งผ่าน และกรองข้อมูลจากแหล่งข้อมูลที่หลากหลายซึ่งมีอัตราการเรียกคืน (Recall Rate) ของโมเดลแพลตฟอร์มอีคอมเมิร์ซที่พัฒนาขึ้นสูงกว่าโมเดลที่ใช้วิธีจับคู่คำ (Word Matching) ถึงร้อยละ 5 ซึ่งแสดงให้เห็นว่าแนวทางที่นำเสนอมีประสิทธิภาพในการตอบสนองความต้องการของผู้ใช้ได้ดีกว่า

2. การเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลแบบไฮบริดกับเทคนิคการวิเคราะห์ข้อมูลแบบดั้งเดิม ผลการศึกษาชี้ให้เห็นว่า โมเดลที่ใช้เทคนิคเหมืองข้อมูลแบบไฮบริดมีความแม่นยำและประสิทธิภาพสูงกว่าการใช้เทคนิคเดี่ยว โดยเฉพาะ K-Means ร่วมกับ Random Forest ที่ให้ค่าความแม่นยำ ร้อยละ 92.50 และ LSTM ร่วมกับ Logistic Regression ที่ให้ค่าความแม่นยำ ร้อยละ 89.70 นอกจากนี้ โมเดลไฮบริดสามารถประมวลผลได้เร็วกว่า เนื่องจากอาศัย การประมวลผลแบบขนาน (Parallel Processing) และการวิเคราะห์สตรีมมิ่ง (Streaming Analytics) เพื่อลดระยะเวลาในการประมวลผลข้อมูล ผลลัพธ์ดังกล่าว สอดคล้องกับงานวิจัยของ Anand & Hughes (1998) ซึ่งได้นำเสนอแนวทางการชุดข้อมูลแบบไฮบริด เพื่อแก้ไขข้อจำกัดของกรอบแนวคิดแบบแยกส่วน โดยแบ่งประเภทของระบบไฮบริดตามระดับปฏิสัมพันธ์ขององค์ประกอบภายในระบบ และเสนอแนวทางในการพัฒนาระบบที่มีประสิทธิภาพสูงขึ้น งานวิจัยดังกล่าวยังได้นำเสนอ การปรับปรุงอัลกอริทึม k-Nearest Neighbors ผ่านการวัดระยะทางด้วยสถิติ การผสานกับเครือข่ายประสาทเทียม และการเชื่อมโยงกับอัลกอริทึมทางพันธุกรรมเพื่อหาหน้าหนักของแอตทริบิวต์ ซึ่งช่วยเพิ่มประสิทธิภาพของ k-NN ในการจัดการเป้าหมายแบบถดถอย (Regression) เมื่อเปรียบเทียบกับแนวทางแบบดั้งเดิม นอกจากนี้งานวิจัยของ Chhabra et al. (2023) ได้ทำการวิเคราะห์เชิงเปรียบเทียบเกี่ยวกับการประเมินผลการปฏิบัติงานของเทคนิคการชุดข้อมูลโดยใช้แนวทางการเรียนรู้เชิงลึกแบบไฮบริด ซึ่งได้ทำการตรวจสอบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่องมาตรฐาน เช่น Support Vector Machine (SVM), Random Forest (RF), J48, Neural Networks และ Naïve Bayes รวมถึงแบบจำลองการเรียนรู้เชิงลึก เช่น Recurrent Neural Networks (RNN), RNN-LSTM (Sigmoid, Tanh และ ReLU) ผลการทดลองแสดงให้เห็นว่า RNN-LSTM (ReLU) มีอัตราความถูกต้องสูงสุด ร้อยละ 97 ซึ่งยืนยันว่าแนวทางการเรียนรู้เชิงลึกแบบไฮบริดสามารถเพิ่มความแม่นยำในการจำแนกประเภทข้อมูลที่มีความซับซ้อนได้อย่างมีประสิทธิภาพ และงานวิจัยของ Bakirarar, Cosgun & Elhan (2024) ได้ศึกษาความเป็นไปได้ของ แนวทางการชุดข้อมูลแบบไฮบริด ผ่านการผสมผสานเทคนิคการชุดข้อมูลหลายรูปแบบเพื่อเพิ่มประสิทธิภาพและลดข้อจำกัดของแต่ละเทคนิค โดยใช้ ชุดข้อมูลจำลองและชุดข้อมูลโรคตับอักเสบจาก UCI ผลการทดลองแสดงให้เห็นว่า ค่า Matthew's Correlation Coefficient (MCC) เพิ่มขึ้นเมื่อขนาดตัวอย่างและความสัมพันธ์ระหว่างตัวแปรอิสระสูงขึ้น อีกทั้งยังพบว่า โมเดลไฮบริดสามารถเพิ่มความแม่นยำของการจำแนกประเภทได้ดีกว่าแนวทางการชุดข้อมูลแบบเดี่ยว

ข้อเสนอแนะ

ข้อเสนอแนะเพื่อนำผลการวิจัยไปใช้

1. ผลการวิจัยสามารถนำไปใช้เพื่อปรับปรุง ระบบวิเคราะห์พฤติกรรมลูกค้าแบบเรียลไทม์ ในแพลตฟอร์มอีคอมเมิร์ซ โดยเฉพาะการจำแนกกลุ่มลูกค้า การคาดการณ์อัตราการเลิกใช้บริการ และการพัฒนาระบบแนะนำสินค้า (Personalized Recommendation System) เพื่อเพิ่มประสิทธิภาพของกลยุทธ์ทางการตลาด

2. แนวทาง Hybrid Data Mining ที่พัฒนาขึ้นนี้ สามารถนำไปใช้กับ Big Data Analytics เพื่อช่วยให้ธุรกิจที่ต้องจัดการข้อมูลขนาดใหญ่สามารถประมวลผลได้เร็วขึ้น ลดข้อผิดพลาด และเพิ่มความแม่นยำในการวิเคราะห์ข้อมูล โดยเฉพาะในอุตสาหกรรมที่ต้องการการคาดการณ์แนวโน้มของลูกค้าแบบเรียลไทม์

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. ควรมีการศึกษาการนำโมเดลเหมืองข้อมูลแบบไฮบริด ไปใช้กับข้อมูลจากอุตสาหกรรมอื่น ๆ เช่น การเงิน สุขภาพ การตรวจจับการฉ้อโกง (Fraud Detection) เพื่อประเมินว่าโมเดลที่พัฒนามีความสามารถในการทำงานได้ดีในบริบทที่แตกต่างกัน

2. ควรมีการพัฒนา โมเดลที่สามารถลดระยะเวลาในการประมวลผล ให้รองรับข้อมูลขนาดใหญ่และสามารถทำงานได้อย่าง Real-Time Streaming Analytics โดยอาจใช้เทคนิคเพิ่มเติม เช่น Federated Learning, Edge Computing หรือ Quantum Computing เพื่อเพิ่มประสิทธิภาพของการวิเคราะห์

เอกสารอ้างอิง

- สมัคร ชัยสงวน. (2561). การพัฒนาระบบวิเคราะห์ความรู้สึกแบบเรียลไทม์ของนักศึกษาบนเฟซบุ๊กโดยใช้ตัวจำแนกข้อมูลนาอ์ฟเบย์ สำหรับภาษาไทย. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต มหาวิทยาลัยสุโขทัยธรรมาธิราช.
- เอกรัฐ รัฐกาญจน์. (2564). การเปรียบเทียบอัลกอริทึมสำหรับการวิเคราะห์ข้อมูลแบบเรียลไทม์. วารสารสถาบันเทคโนโลยีไทย-ญี่ปุ่น: วิศวกรรมศาสตร์และเทคโนโลยี, 9(1), 1-8.
- Abbasimehr, H., & Shabani, M. (2021). A new methodology for customer behavior analysis using time series clustering: A case study on a bank's customers. *Kybernetes*, 50(2), 221-242.
- Alam, M. A., Nabil, A. R., Mintoo, A. A., & Islam, A. (2024). Real-time analytics in streaming big data: Techniques and applications. *Journal of Science and Engineering Research*, 1(01), 104-122.
- Afzal, A., Khan, L., Hussain, M. Z., Hasan, M. Z., Mustafa, M., Khalid, A., Awan, R., Ashraf, F., Khan, Z. A., & Javaid, A. (2024). Customer Segmentation Using Hierarchical Clustering. In *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)* (pp. 1-6). Pune, India. <https://ieeexplore.ieee.org/document/10543349>
- Anand, S. S., & Hughes, J. G. (1998). Hybrid data mining systems: the next generation. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 13-24). Springer Berlin Heidelberg.
- Bakirarar, B., Cosgun, E., & Elhan, A. H. (2024). Hybrid model approach in data mining. *Communications in Statistics-Simulation and Computation*, 53(10), 4998-5007.
- Baranauskas, G., & Raišienė, A. G. (2021). Reflections on the customer decision-making process in the digital insurance platforms: an empirical study of the baltic market. *Applied Sciences*, 11(18). <https://doi.org/10.3390/app11188524>
- Burney, V., Arnold-Saritepe, A., & McCann, C. M. (2023). Rethinking the place of qualitative methods in behavior analysis. *Perspectives on Behavior Science*, 46(1), 185-200.
- Chaffey, D., & Smith, P. R. (2022). *Digital marketing excellence: planning, optimizing and integrating online marketing*. Routledge.
- Chhabra, G. S., William, P., Lanke, G. R., Jain, K., Lakshmi, T. V., & Varshney, N. (2023). Comparative analysis of data mining based performance evaluation using hybrid deep learning approach. In *International Conference on Mobile Radio Communications & 5G Networks* (pp. 607-621). Springer Nature Singapore.
- Chung, K., Quach, S., & Thaichon, P. (2022). Introduction to artificial intelligence (AI): Definition and scope of AI. In *Artificial Intelligence for Marketing Management* (pp. 3-17). Routledge.
- Côté, P. O., Nikanjam, A., Ahmed, N., Humeniuk, D., & Khomh, F. (2024). Data Cleaning and Machine Learning: A Systematic Literature Review. *Automated Software Engineering*, 31(2). <https://doi.org/10.1007/s10515-024-00453-w>
- Laudon, K. C., & Traver, C. G. (2021). *E-commerce 2020-2021: business, technology, society*. Pearson Education.
- Lei, L., Guo, X., & Zheng, R. (2024). An innovative dynamic anomaly detection method based on hybrid data mining technology for building energy Consumption. *Energy and Buildings*, 319, 114-559.

- Li, B., & Lei, Q. (2022). Hybrid IoT and data fusion model for e-commerce big data analysis. *Wireless Communications and Mobile Computing*, 2022(1). <https://doi.org/10.1155/2022/2292321>
- Li, H. (2023). Research on Hybrid Data mining methods for social media sentiment analysis. In *Proceedings of the 2023 4th International Conference on Big Data Economy and Information Management* (pp. 204-209). ACM.
- Madhumita, G., Das, T., Das, S., Khatri, E., Ravisankar, P., & Hemachandu, P. (2024). IoT and AI for real-time customer behavior analysis in digital banking. In *International Conference on Recent Trends in Computer Science and Technology (ICRTCST) 2024 5th* (pp. 198-203). Jamshedpur, India: RVS College of Engineering and Technology. <https://ieeexplore.ieee.org/document/10578458>
- Seyyedabbasi, A., Aliyev, R., Kiani, F., Gulle, M. U., Basyildiz, H., & Shah, M. A. (2021). Hybrid algorithms based on combining reinforcement learning and metaheuristic methods to solve global optimization problems. *Knowledge-Based Systems*, 223. <https://doi.org/10.1145/3659211.36592>
- Solomon, M. R. (2020). *Consumer Behavior: Buying, Having and Being*. Pearson Education.
- Sreenivasa, B. R., & Nirmala, C. R. (2019). Hybrid location-centric e-commerce recommendation model using dynamic behavioral Traits of Customer. *Iran Journal of Computer Science*, 2(3), 179-188.
- Varoquaux, G., & Colliot, O. (2023). Evaluating machine learning models and their diagnostic value. In *Machine Learning for Brain Disorders* (pp. 601-630). Humana Press.
- Whang, S. E., & Lee, J. G. (2020). Data collection and quality challenges for deep Learning. *Proceedings of the VLDB Endowment*, 13(12), 3429-3432.
- Wolniak, R. (2023). Functioning of Real-Time Analytics in Business. *Scientific Papers of Silesian University of Technology Organization and Management Series Silesian University of Technology*, 172, 659-677.