

การแบ่งกลุ่มผู้บริโภคเนื้อหาจากการคัดเลือกคุณลักษณะที่มีความสำคัญ ต่อการยกเลิกบริการด้วยการจัดกลุ่มแบบผสมผสาน

จักรินทร์ สันติรัตนภักดี^{1*} รัญเกล้า เอื้อวงศ์กุล²

^{1,2}สาขาวิชาเทคโนโลยีธุรกิจดิจิทัล คณะบริหารธุรกิจ มหาวิทยาลัยวงษ์ชวลิตกุล, นครราชสีมา, ประเทศไทย

*อีเมลผู้ประพันธ์บรรณกิจ: chakkarin_san@vu.ac.th

รับต้นฉบับ: 8 มกราคม 2569; รับบทความฉบับแก้ไข: 21 กรกฎาคม 2569; ตอบรับบทความ: 22 กรกฎาคม 2569

เผยแพร่ออนไลน์: 28 สิงหาคม 2569

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาตัวแบบการแบ่งกลุ่มผู้บริโภคเนื้อหาจากการคัดเลือกคุณลักษณะที่มีความสำคัญต่อการยกเลิกบริการด้วยเทคนิคการจัดกลุ่มแบบผสมผสานที่มุ่งเติมเต็มช่องว่างทางวิชาการของวิธีวิจัยดั้งเดิมที่มักพึ่งพาเพียงข้อมูลด้านประชากรศาสตร์หรือการใช้เทคนิคการจัดกลุ่มเพียงรูปแบบเดียว ซึ่งมีข้อจำกัดในการจัดการกับข้อมูลที่มีมิติสูง มีความหนาแน่นแตกต่างกัน หรือมีโครงสร้างความสัมพันธ์แบบไม่เป็นเชิงเส้น โดยรวบรวมข้อมูลจากผู้ให้บริการยูทูปฟรีเมียมในพื้นที่กรุงเทพฯ และปริมาณทลจำนวน 10,983 รายการ กระบวนการวิเคราะห์ข้อมูลเริ่มต้นจากการเตรียมข้อมูลและการวิเคราะห์สหสัมพันธ์ ก่อนจะคัดเลือกคุณลักษณะสำคัญแบบทีละขั้นด้วยโครงข่ายประสาทเทียมร่วมกับ XGBoost และ LightGBM จากนั้นสร้างตัวแปรเชิงอนุพันธ์ใหม่เพื่อเพิ่มมิติความเสี่ยงเชิงพฤติกรรม และแบ่งกลุ่มผู้บริโภคด้วยโมเดล GMM, HDBSCAN และ Spectral clustering โดยประเมินประสิทธิภาพด้วยค่า Silhouette score, DBI และ CHI เพื่อระบุกลุ่มเป้าหมายที่มีความเสี่ยงต่อการยกเลิกบริการ ผลการวิจัยพบว่าตัวแปรที่ให้อรรถประโยชน์สูงสุดในการพยากรณ์ประกอบด้วยระยะเวลาใช้งาน รูปแบบสัญญา ชั่วโมงรับชมต่อวัน จำนวนวิดีโอที่รับชม จำนวนช่องที่ติดตาม ปัญหาการชำระเงิน และการติดต่อฝ่ายสนับสนุนที่มีความสอดคล้องอย่างมีนัยสำคัญกับความภักดีต่อแบรนด์และแนวคิดเส้นทางลูกค้า ซึ่งระบุว่าประสบการณ์เชิงบวกและความคุ้นชินในกระบวนการเป็นสมาชิกช่วยเพิ่มต้นทุนการย้ายบริการและลดความเสี่ยงในการยกเลิกบริการ นอกจากนี้ ยังสะท้อนถึงความคาดหวังต่อการกู้คืนบริการ โดยชี้ให้เห็นว่าปัญหาที่เกิดขึ้นนั้นส่งผลโดยตรงต่อความไม่พึงพอใจและแรงกระตุ้นใจเลิกใช้บริการ ร่วมกับตัวแปรเชิงอนุพันธ์ ได้แก่ WatchTime_per_Content, Subscribed_Ratio, SupportTickets_per_Period และ Churn_Risk สำหรับผลการแบ่งกลุ่มลูกค้า พบว่า HDBSCAN เป็นโมเดลที่มีความสมดุลที่สุดในการแยกกลุ่ม (cluster) และตรวจจับกลุ่มผิดปกติ โดยมีค่า DBI ต่ำสุดที่ 1.782 ขณะที่ Spectral clustering สามารถระบุกลุ่มเสี่ยงสูงสุด ซึ่งมีค่า Churn_Risk เฉลี่ยสูงถึง 0.699 ผลลัพธ์จากงานวิจัยนี้ช่วยสร้างองค์ความรู้ใหม่ในการประยุกต์ใช้อัลกอริทึมการเรียนรู้แบบรวม เพื่อคัดเลือกคุณลักษณะสำคัญร่วมกับการจัดกลุ่มแบบผสมผสาน เช่นเดียวกับการนำเสนอตัวแปรเชิงอนุพันธ์ที่ปรับแต่งให้สอดคล้องกับพฤติกรรมเชิงลึกที่ช่วยเพิ่มความละเอียดในการจำแนกกลุ่มผู้บริโภคเนื้อหาเกินกว่าข้อมูลพื้นฐานและการขยายขอบเขตองค์ความรู้สู่บริบทแพลตฟอร์มสตรีมมิง โดยบูรณาการปัจจัยเฉพาะด้านเนื้อหา ซึ่งมักไม่ค่อยได้รับการวิเคราะห์อย่างเข้มข้นในการยกเลิกบริการ ส่งผลให้ได้ตัวแบบที่สามารถระบุกลุ่มเสี่ยงและสนับสนุนการกำหนดกลยุทธ์รักษาลูกค้าเชิงรุกได้อย่างแม่นยำและเป็นรูปธรรม

คำสำคัญ: การแบ่งกลุ่มผู้บริโภค การจัดกลุ่มข้อมูล คุณลักษณะที่มีความสำคัญ การยกเลิกบริการ ยูทูปฟรีเมียม

Content Consumer Segmentation from Importance Feature Selection for Service Cancellation Using Hybrid Clustering

Chakkarin Santirattanaphakdi^{1*} Ratklao Eurvongkul²

^{1,2}*Department of Digital Business Technology, Faculty of Business Administration, Vongchavalitkul University,
Nakhon Ratchasima, Thailand*

*Corresponding author e-mail: chakkarin_san@vu.ac.th

Received: 8 January 2026; Revised: 21 July 2026; Accepted: 22 July 2026

Published online: 28 August 2026

Abstract

This research aims to develop a consumer segmentation model for churn prediction using hybrid clustering techniques and feature selection, addressing limitations in traditional methodologies that rely on demographics or single clustering methods, which often struggle with high-dimensional and non-linear data. The analysis involved 10,983 YouTube Premium users in the Bangkok metropolitan area. The methodology integrates stepwise feature selection via ANN, XGBoost, and LightGBM, alongside the creation of derived behavioral risk variables. Consumer segmentation was executed using GMM, HDBSCAN, and Spectral clustering, with performance assessed through Silhouette scores, DBI, and CHI. The results identify key predictors such as tenure (Period), contract type, watch time, content count, subscription count, payment issues, and support tickets. These findings align with brand loyalty and customer journey theories, where positive engagement increases switching costs, while issues with service recovery accelerate churn decisions. Derived features—including WatchTime_per_Content, Subscribed_Ratio, SupportTickets_per_Period, and Churn_Risk—further enhanced classification granularity. Analysis showed that HDBSCAN provided the most balanced clustering (DBI = 1.782), while Spectral Clustering identified the highest-risk group (average Churn_Risk = 0.699). This research contributes to the field by synthesizing ensemble learning and hybrid clustering, expanding churn analysis to the streaming platform context, and providing a robust framework for proactive customer retention.

Keywords: Consumer segmentation, Data clustering, Feature importance, Service cancellation, YouTube premium

บทนำ

ในยุคดิจิทัลที่การแข่งขันของธุรกิจบนแพลตฟอร์มออนไลน์ทวีความรุนแรงมากขึ้น ประกอบกับพฤติกรรมผู้บริโภคเนื้อหาที่เปลี่ยนไปอย่างรวดเร็ว ส่งผลให้การรักษาลูกค้าเดิมไว้จึงกลายเป็นความท้าทายของทุกธุรกิจ โดยเฉพาะบริการสตรีมมิงวิดีโอแบบสมัครสมาชิก ซึ่งผู้ใช้งานสามารถยกเลิกบริการได้ตลอดเวลา ผู้ให้บริการจึงจำเป็นต้องเปลี่ยนผ่านกระบวนการบริหารจัดการเชิงรับมาสู่การดำเนินกลยุทธ์เชิงรุกด้วยการใช้กระบวนการเทคโนโลยี เพื่อรักษาอัตราการคงอยู่ของลูกค้าอย่างยั่งยืน

การแบ่งกลุ่มผู้บริโภคเป็นแนวทางสำคัญในการจำแนกกลุ่มที่มีลักษณะหรือพฤติกรรมใกล้เคียงกัน (Siahaan & Sianipar, 2023) เพื่อสนับสนุนการกำหนดกลยุทธ์เฉพาะบุคคล และการสื่อสารทางการตลาดที่เหมาะสม (Qin et al., 2022) โดยเฉพาะในกรณีของผู้ใช้บริการยูทูปที่มีพฤติกรรมการบริโภคเนื้อหาความบันเทิงที่แตกต่างอย่างมีนัยสำคัญเมื่อเทียบกับบริการสตรีมมิงวิดีโออื่น ๆ ที่ผู้ใช้งานมักมีแนวโน้มรับชมเนื้อหาอย่างตั้งใจและเป็นระบบ ในทางกลับกัน ผู้ใช้งานบนยูทูปอาจแสดงพฤติกรรมการบริโภคเนื้อหาที่มีลักษณะเป็นการรับชมวิดีโอแบบสุ่มและไม่ต่อเนื่อง (passive consumption) ที่มีส่วนร่วมเพียงเล็กน้อยหรือแทบไม่มีปฏิสัมพันธ์ใด ๆ กับเนื้อหา (Khan, 2017) ซึ่งสะท้อนถึงความซับซ้อน และรูปแบบการใช้งานที่ยากต่อการจำแนกด้วยวิธีการแบ่งกลุ่มแบบดั้งเดิม

ในปี พ.ศ. 2568 พบว่ามีจำนวนผู้ใช้งานยูทูปสูงถึง 2,700 ล้านคน โดยมีผู้ชมมากกว่า 122 ล้านคนเข้าสู่ระบบในแต่ละวัน เพื่อรับชมและมีส่วนร่วมกับเนื้อหาต่าง ๆ อย่างต่อเนื่อง ขณะเดียวกัน ความต้องการที่เพิ่มสูงขึ้นของผู้ใช้ที่ต้องการรับชมวิดีโอแบบไม่มีโฆษณาคันส่งผลให้บริการแบบพรีเมียมของยูทูป (YouTube Premium) ขยายตัวอย่างรวดเร็ว โดยมีจำนวนสมาชิกมากกว่า 125 ล้านคน (Wetherby, 2026) ซึ่งชี้ให้เห็นถึงการเปลี่ยนแปลงพฤติกรรมการบริโภคสื่อดิจิทัลในปัจจุบัน โดยในบริบทของประเทศไทยถือเป็นตลาดที่มีรายได้จากผู้ให้บริการแบบจ่ายค่าสมัครสมาชิก (Subscription Video On Demand: SVOD) ใหญ่ที่สุดในเอเชียตะวันออกเฉียงใต้ แพลตฟอร์มโซเชียลวิดีโอและวิดีโอสั้นอย่าง YouTube และ TikTok ครองส่วนแบ่งการรับชมสูงถึงร้อยละ 96 ขณะที่วิดีโอพรีเมียมตามความต้องการ (Premium Video On Demand: PVOD) มีส่วนแบ่งเพียงร้อยละ 4 (Asia Video Industry Association, 2025) ซึ่งสวนทางกับที่ผู้ชมชาวไทยร้อยละ 77 เลือกใช้บริการแบบมีโฆษณามากกว่าการจ่ายค่าสมาชิกเพียงอย่างเดียว (Magnite, 2024) แสดงถึงโอกาสในการขยายฐานลูกค้าผ่านโมเดลการตลาดที่แม่นยำ

อย่างไรก็ดี มีผู้ใช้งานไม่น้อยที่ตัดสินใจยกเลิกบริการยูทูปพรีเมียมจากหลายสาเหตุ เช่น ราคาสมัครสมาชิกเพิ่มขึ้นหรือรู้สึกว่าคุณค่าของบริการไม่คุ้มค่าที่จ่ายไป (SBI, 2025) แม้ว่าจะงานวิจัยด้านการแบ่งกลุ่มผู้บริโภคและการวิเคราะห์พฤติกรรมการยกเลิกบริการจะมีจำนวนมาก แต่ยังมีช่องว่างสำคัญหลายประการ เนื่องจากงานส่วนใหญ่มีก้ำกัยเพียงข้อมูลประชากรศาสตร์หรือใช้เทคนิคการจัดกลุ่มเพียงวิธีเดียว ซึ่งไม่สามารถจัดการกับข้อมูลที่มีมิติสูง หรือโครงสร้างที่ไม่เป็นเชิงเส้นได้อย่างจำกัด จึงอาจไม่สามารถสะท้อนความซับซ้อนของพฤติกรรมผู้บริโภคยุคใหม่ได้อย่างเพียงพอ ดังนั้นการผสมผสานเทคนิคการแบ่งกลุ่มที่หลากหลายเข้าด้วยกันจึงมีความจำเป็น เพื่อให้สามารถตีความพฤติกรรมผู้ใช้ได้อย่างลึกซึ้งและสนับสนุนมาตรการรักษาลูกค้าเชิงปฏิบัติได้จริง

งานวิจัยนี้มุ่งเน้นการคัดเลือกลักษณะที่มีความสำคัญด้วยอัลกอริทึมการเรียนรู้แบบรวม เพื่อปิดช่องว่างทางเทคนิคและเพิ่มความแม่นยำในการจำแนกพฤติกรรมการเสี่ยงเชิงลึก ก่อนจะนำมาจัดกลุ่มผู้ใช้ที่ยกเลิกบริการยูทูปพรีเมียมแบบผสมผสาน โดยมีวัตถุประสงค์เพื่อสร้างโมเดลการแบ่งกลุ่มที่สะท้อนพฤติกรรมเชิงลึก ประกอบด้วย Gaussian Mixture Models (GMM), Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) และ Spectral clustering ร่วมกัน การผสมผสานเทคนิคเหล่านี้จึงสามารถเพิ่มความแม่นยำในการจำแนกกลุ่มผู้ใช้ที่ยกเลิกบริการที่มีพฤติกรรมซับซ้อนได้ครอบคลุมยิ่งขึ้น อีกทั้งสามารถนำไปใช้ในการออกแบบกลยุทธ์เชิงธุรกิจอย่างมีประสิทธิภาพ ตลอดจนขยายองค์ความรู้จากงานวิจัยในอดีตที่ผ่านมา โดยการบูรณาการปัจจัยเฉพาะทางด้านเนื้อหา เช่น ปริมาณการบริโภคและระดับความผูกพันเชิงพฤติกรรม เข้าสู่กรอบการวิเคราะห์การยกเลิกบริการ ซึ่งเป็นมิติที่ยังไม่ได้รับการศึกษาอย่างเข้มข้นในบริบทของแพลตฟอร์มสตรีมมิง นอกจากนี้ การนำเสนอตัวแปรเชิงอนุพันธ์ใหม่ยังเป็นการประยุกต์ใช้ทฤษฎีความกักตื้อต่อแบรนด์และเส้นทางลูกค้าให้สามารถอธิบายการตัดสินใจของผู้บริโภคในยุคดิจิทัลได้อย่างเป็นรูปธรรม ตลอดจนสามารถนำโมเดลที่พัฒนาขึ้นไปประยุกต์ใช้ตรวจจับสัญญาณความเสี่ยงของผู้ใช้บริการรายบุคคล เพื่อให้ธุรกิจสามารถดำเนินมาตรการรักษาลูกค้าได้อย่างทันท่วงทีก่อนเกิดการยกเลิกบริการจริง อีกทั้งสามารถใช้เป็นแนวทางในการวิเคราะห์ข้อมูลบนชุดข้อมูลพฤติกรรมขนาดใหญ่ที่มีความซับซ้อนสูงในอุตสาหกรรมดิจิทัลคอนเทนต์ต่อไป

บททวนวรรณกรรม

การทบทวนวรรณกรรมที่เกี่ยวข้องกับงานวิจัยนี้ โดยครอบคลุมทั้งแนวคิด ทฤษฎี และผลงานวิจัยที่สำคัญทั้งในและต่างประเทศ เพื่อออกแบบวิธีดำเนินการวิจัยให้มีความเหมาะสมและสอดคล้องกับวัตถุประสงค์ของการศึกษา มีรายละเอียดดังนี้

1. การแบ่งกลุ่มผู้บริโภค (Consumer Segmentation)

การแบ่งกลุ่มผู้บริโภคถือเป็นกระบวนการที่ช่วยให้ธุรกิจสามารถทำความเข้าใจความแตกต่างระหว่างลูกค้าแต่ละกลุ่มได้อย่างเป็นระบบ โดยอาศัยการแบ่งกลุ่มลูกค้าออกเป็นกลุ่มย่อยที่มีลักษณะร่วมกันในด้านต่าง ๆ เช่น ลักษณะทางประชากรศาสตร์ ภูมิศาสตร์ พฤติกรรม หรือจิตวิทยา (Kotler & Armstrong, 2018) เช่นเดียวกับกลุ่มผู้บริโภคเนื้อหา (content consumer) ที่มีลักษณะการบริโภคที่แตกต่างหลากหลาย ทั้งในด้านความถี่ ระยะเวลา ประเภทของเนื้อหาที่เลือกสรร ตลอดจนปัจจัยทางจิตวิทยาและปัจจัยทางสังคมที่ส่งผลต่อพฤติกรรมการตัดสินใจ ดังนั้นผู้ให้บริการจึงจำเป็นต้องเข้าใจพฤติกรรมการบริโภคของสมาชิกแต่ละกลุ่มอย่างลึกซึ้ง เพื่อออกแบบกลยุทธ์การรักษาฐานสมาชิกและลดอัตราการยกเลิกบริการอย่างมีประสิทธิภาพ

2. คุณลักษณะที่มีความสำคัญ (Feature Importance)

คุณลักษณะที่มีความสำคัญคือระดับความสำคัญของตัวแปรอิสระหรือคุณลักษณะที่ส่งผลต่อการพยากรณ์ผลลัพธ์ของโมเดล ปัจจุบันมีหลายวิธีการที่สามารถใช้ประเมินความสำคัญของคุณลักษณะ สามารถจำแนกได้เป็น 4 วิธี ประกอบด้วย 1) การคำนวณความสำคัญที่มาจากโมเดลโดยตรง (model-based importance) เพื่อวัดว่า การใช้คุณลักษณะแต่ละตัวนั้นช่วยลดความไม่แน่นอนได้มากเพียงใด 2) การสลับค่าของคุณลักษณะหนึ่ง ๆ แบบสุ่มแล้ววัดว่าความแม่นยำของโมเดลลดลงมากน้อยเพียงใด (permutation importance) หากความแม่นยำลดลงมากย่อมแสดงว่าคุณลักษณะนั้นมีความสำคัญสูง 3) เทคนิค Shapley additive explanations (SHAP) ที่อธิบายการมีส่วนร่วมของคุณลักษณะในเชิงทฤษฎีเกม (Lundberg & Lee, 2017) และ 4) เทคนิค Local Interpretable Model-agnostic Explanations (LIME) ซึ่งเน้นการอธิบายผลเฉพาะจุดข้อมูลที่สนใจ (Ribeiro et al., 2016)

3. อัลกอริทึมการเรียนรู้แบบรวม (Ensemble Learning)

การเรียนรู้แบบรวมเป็นหนึ่งในแนวคิดสำคัญในสาขาการเรียนรู้ของเครื่อง โดยอาศัยหลักการรวมผลลัพธ์จากโมเดลพื้นฐานหลายตัว (base learners) ด้วยวิธีการที่เหมาะสม เพื่อช่วยเพิ่มความถูกต้อง ลดความผิดพลาด และทำให้ผลการพยากรณ์ข้อมูลมีเสถียรภาพมากขึ้น (Rokach, 2019) แบ่งเป็น 4 ประเภท ได้แก่ 1) Bagging (bootstrap aggregating) ด้วยแนวคิดการสร้างโมเดลหลายตัวจากการสุ่มตัวอย่างข้อมูล (sampling with replacement) แล้วรวมผลการพยากรณ์ โดยมีจุดเด่นที่การลดความแปรปรวน จึงทำให้ผลลัพธ์เสถียรมากขึ้น 2) Boosting ที่สร้างโมเดลทีละตัว โดยให้โมเดลใหม่โฟกัสกับตัวอย่างที่โมเดลก่อนหน้าพยากรณ์ผลลัพธ์ผิดพลาด ด้วยจุดเด่นที่สามารถลดความลำเอียง (bias) จึงได้รับการยอมรับว่าสามารถสร้างโมเดลที่มีประสิทธิภาพสูง จึงเป็นหนึ่งในวิธีที่ได้รับความนิยมในปัจจุบัน 3) Stacking (stacked generalization) ที่รวมผลลัพธ์ของโมเดลฐานหลายตัว โดยใช้โมเดลอีกตัวหนึ่งมาเรียนรู้จากผลลัพธ์ของโมเดลฐานทีละตัว ซึ่งช่วยให้เกิดการผสมผสานจุดแข็งของโมเดลที่ต่างกัน และ 4) Voting จากการใช้โมเดลหลายตัวมาพยากรณ์แล้วรวมผลโหวตเสียง แบ่งเป็นแบบเสียงข้างมาก (hard voting) และแบบใช้ค่าเฉลี่ยของความน่าจะเป็น (soft voting)

4. การยกเลิกบริการ (Service Cancellation)

การยกเลิกบริการ หมายถึง พฤติกรรมของลูกค้าหรือผู้ใช้บริการที่ตัดสินใจเลิกใช้บริการ (customer churn) หรือหยุดการเป็นสมาชิก (unsubscribe) ซึ่งสามารถเกิดขึ้นได้ทั้งในรูปแบบการยกเลิกแบบสมัครสมาชิก (subscription-based services) หรือการเลิกใช้บริการแบบรายครั้ง (one-time service) ในยุคดิจิทัล ปัญหาการยกเลิกบริการยังซับซ้อนมากขึ้น โดยเฉพาะในธุรกิจวิดีโอสตรีมมิ่ง ที่ผู้ใช้สามารถยกเลิกบริการหรือเปลี่ยนบริการได้อย่างรวดเร็ว (Corcoran, 2023) โดยสาเหตุสำคัญคือเรื่องของค่าใช้จ่ายในการสมัครสมาชิกที่อาจมีราคาสูงเกินไปหรือใช้งานไม่บ่อยพอที่จะคุ้มค่างบเงินที่จ่าย อีกหนึ่งสาเหตุคือบริการขาดความคุ้มค่า หากผู้ใช้ไม่สนใจเนื้อหาที่แพลตฟอร์มนำเสนออีกมีแนวโน้มที่จะยกเลิกการสมัคร ทั้งนี้ หากการมีทางเลือกอื่นที่สามารถเข้าถึงเนื้อหาได้ในราคาถูกกว่า

หรืออาจไม่มีค่าใช้จ่ายเลย ไม่ว่าจะเป็นแพลตฟอร์มคู่แข่งหรือแม้แต่การเข้าถึงผ่านช่องทางที่ผิดกฎหมาย ก็มีส่วนสำคัญที่ผลักดันให้เกิดการยกเลิกบริการเช่นกัน อีกทั้งยังมีพฤติกรรมการณ์เปลี่ยนผู้ให้บริการบ่อยครั้ง โดยยกเลิกบริการเพื่อสมัคร บริการอื่นที่ตอบสนองความต้องการได้ดีกว่า (Statista Research Department, 2024)

5. ยูทูบและยูทูบพรีเมียม (YouTube and YouTube Premium)

ยูทูบเป็นแพลตฟอร์มวิดีโอออนไลน์ที่ได้รับความนิยมระดับโลกด้วยการเปิดโอกาสให้ผู้ใช้งานสามารถอัปโหลด รับชม แบ่งปัน และมีปฏิสัมพันธ์กับวิดีโอได้หลากหลายประเภท อย่างไรก็ตาม ยังมีข้อจำกัดบางประการ เช่น โฆษณาระหว่างการรับชม เพื่อแก้ไขข้อจำกัดเหล่านี้จึงมีบริการยูทูบพรีเมียมเป็นบริการสมัครสมาชิก (YouTube, n.d.) โดยมีค่าธรรมเนียมรายเดือน สำหรับรับชมวิดีโอโดยไม่มีโฆษณา การเล่นวิดีโอเบื้องหลัง การดาวน์โหลดวิดีโอเพื่อดูแบบออฟไลน์ และสิทธิ์เข้าถึงยูทูบมิวสิกพรีเมียม ซึ่งปัจจุบันมี 4 แผนหลักคือแผนสมาชิกทั่วไป (individual plan) ทั้งแบบรายเดือนและรายปี แผนแบบครอบครัว (family plan) แผนนักเรียนนักศึกษา (student plan) และแผนประหยัด (premium lite plan)

6. ความภักดีต่อแบรนด์ (Brand Loyalty)

ความเชื่อมโยงเชิงบวกที่ลูกค้ามีต่อสินค้าและบริการสามารถก่อให้เกิดความภักดีต่อแบรนด์ ซึ่งมีความสำคัญต่อการบริการเสริมเนื่องจากช่วยสร้างความสัมพันธ์ที่ยาวนานขึ้นระหว่างธุรกิจและลูกค้าที่ต้องการใช้บริการอย่างต่อเนื่อง (Oliver, 1999)

บริการเสริมมีแนวทางหลากหลายในการเปลี่ยนลูกค้าทั่วไปให้กลายเป็นลูกค้าที่มีความภักดี หนึ่งในแนวทางดังกล่าวคือการทำให้แพลตฟอร์มมีความเป็นเฉพาะบุคคล (personalization) เพื่อให้แพลตฟอร์มตอบสนองความต้องการของลูกค้า (Palomba, 2022) โดยสามารถเลือกประเภทของเนื้อหาหรือช่วงเวลาที่น่าสนใจ และจะปรับการนำเสนอให้สอดคล้องกับความชอบเฉพาะบุคคล (Gomez-Urbe & Hunt, 2016) ส่งผลให้ลูกค้ารับรู้คุณค่า คุณภาพ ความเป็นประโยชน์ และความพึงพอใจที่เพิ่มขึ้นจากการใช้บริการ และเมื่อเวลาผ่านไปย่อมก่อให้เกิดความจงรักภักดีต่อแบรนด์ (Palomba, 2022) ทั้งนี้ยังมีการระบุว่าความพึงพอใจ ความผูกพัน และผลกำไรมีความสัมพันธ์กันอย่างชัดเจน แต่ไม่ได้หมายความว่าธุรกิจควรให้บริการลูกค้าทุกกลุ่มในระดับเท่าเทียมกัน (Hallowell, 1996)

ธุรกิจจึงมุ่งสร้างความสัมพันธ์ระยะยาวกับลูกค้าที่มีความภักดี เพื่อนำไปสู่ผลกำไรที่เพิ่มขึ้น (Grönroos, 1989) โดยลูกค้าที่มีความภักดีจึงมีความสำคัญอย่างยิ่งต่อโมเดลธุรกิจแบบสมัครสมาชิก เนื่องจากมีความเสี่ยงต่อการยกเลิกการใช้บริการต่ำกว่าลูกค้าทั่วไป และยังช่วยลดต้นทุนในการได้มาซึ่งลูกค้าใหม่ อีกทั้งการสร้างผลกำไรจากลูกค้าที่มีความภักดีจำเป็นต้องอาศัยระดับความภักดีที่มากพอเพื่อให้เกิดผลเชิงบวกต่อความสามารถในการทำกำไรของลูกค้าแต่ละราย (Helgesen, 2006) ดังนั้นหากธุรกิจสามารถพัฒนากลยุทธ์หรือข้อเสนอเฉพาะสำหรับลูกค้าที่มีความภักดี แม้การดำเนินการดังกล่าวต้องใช้ทรัพยากร แต่หากดำเนินการได้อย่างมีประสิทธิภาพก็สามารถเพิ่มความสามารถในการทำกำไรของธุรกิจในระยะยาวได้ (Lin & Bowman, 2022)

7. เส้นทางลูกค้า (Customer Journey)

เส้นทางลูกค้า หมายถึง ชุดของประสบการณ์ที่ลูกค้าได้รับผ่านจุดสัมผัส (touchpoints) ต่าง ๆ ของธุรกิจตลอดช่วงเวลาในการปฏิสัมพันธ์กับแบรนด์ (Lemon & Verhoef, 2016) หากประสบการณ์จากจุดสัมผัสเหล่านี้เป็นเชิงบวก ก็จะส่งเสริมให้ลูกค้าใช้บริการต่อเนื่องและทำให้เส้นทางของลูกค้ายาวนานขึ้น (Bansal, 2023) ในบริบทของบริการเสริมนี้สามารถจำแนกออกเป็น 3 ระยะ ได้แก่ ระยะก่อนการซื้อ (pre-purchase stage) เริ่มจากการรับรู้ถึงแบรนด์ (brand awareness) และเข้าสู่กระบวนการเปลี่ยนผ่านกรวยลูกค้า (customer funnel) สำหรับบริการเสริม ช่วงนี้รวมถึงเสนอการทดลองใช้ฟรี เพื่อเปิดโอกาสให้ลูกค้าสำรวจบริการก่อนตัดสินใจสมัครสมาชิกแบบชำระเงิน (Gupta & Singharia, 2021) แต่การทดลองใช้ฟรีอาจส่งผลย้อนกลับต่อธุรกิจ หากลูกค้าใช้ช่วงเวลานี้เพื่อรับชมเนื้อหาแบบต่อเนื่องจนจบ (binge-watching) และยกเลิกบริการทันทีหลังสิ้นสุดช่วงทดลอง (Godinho de Matos & Ferreira, 2020) เมื่อเข้าสู่ระยะการซื้อหรือการเป็นสมาชิก (purchase/subscription stage) ลูกค้าจะเริ่มบริโภคเนื้อหาและสร้างทัศนคติที่มีต่อบริการโดยเปรียบเทียบความคาดหวังกับประสบการณ์ใช้งานจริง รวมถึงคุณค่าที่รับรู้จากบริการ ในช่วงนี้ธุรกิจ

มักมุ่งใช้กลยุทธ์เพื่อป้องกันการยกเลิกบริการ (proactive churn management) เช่น การนำเสนอเนื้อหาหลัก เพื่อรักษาความสนใจ หรือใช้ระบบแนะนำเนื้อหาที่ปรับตามความชอบเฉพาะบุคคล เพื่อเสริมความผูกพันกับแบรนด์ และระยะหลังการซื้อหรือหลังการเลิกสมาชิก (post-purchase stage) ซึ่งถือเป็นจุดสิ้นสุดของเส้นทางลูกค้า เว้นแต่ว่าธุรกิจจะประสบความสำเร็จในการดึงลูกค้ากลับมาใช้งานอีกครั้ง (win-back) (Kumar et al., 2015) ในระยะนี้ธุรกิจจะดำเนินการรวบรวมข้อมูลเกี่ยวกับสาเหตุของการเลิกใช้บริการเพื่อใช้ในการพยากรณ์การยกเลิกในอนาคต รวมถึงการเสนอส่วนลดหรือข้อเสนอพิเศษ เพื่อกระตุ้นให้ลูกค้ากลับมาสมัครสมาชิกอีกครั้ง

8. การจัดกลุ่มข้อมูล (Data Clustering)

การจัดกลุ่มข้อมูล เป็นกระบวนการสำคัญของการเรียนรู้แบบไม่มีผู้สอน (unsupervised learning) มีเป้าหมายในการค้นหา รูปแบบหรือโครงสร้างที่ซ่อนอยู่ในข้อมูลจำนวนมาก งานวิจัยนี้คัดเลือกเทคนิคการจัดกลุ่มที่สำคัญ 3 วิธี ได้แก่

8.1 *Gaussian Mixture Model (GMM)* เป็นวิธีการจัดกลุ่มข้อมูลเชิงสถิติจากการผสมเชิงเส้นของการแจกแจงแบบปกติหลายตัว (McLachlan & Peel, 2000) โดยแต่ละการแจกแจงแทนหนึ่งกลุ่มในเชิงนัย ทำให้สามารถจัดการกับข้อมูลที่มีรูปร่างของกลุ่มซับซ้อน กันได้ดีกว่าวิธีที่อาศัยเพียงระยะทางเชิงเรขาคณิต เนื่องจากสามารถจัดกลุ่มที่มีรูปร่างแตกต่างกันได้ตามโครงสร้างของเมทริกซ์ความแปรปรวนร่วม (covariance matrix) อย่างไรก็ตาม ข้อจำกัดสำคัญคือต้องกำหนดจำนวนกลุ่มล่วงหน้า และความไวต่อค่าเริ่มต้นของพารามิเตอร์ ซึ่งอาจไม่เหมาะสมกับข้อมูลที่มีความซับซ้อนหรือไม่เป็นไปตามสมมติฐานเชิงสถิติ

8.2 *Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN)* เป็นการต่อยอดจากการจัดกลุ่มตามความหนาแน่น (DBSCAN) (Campello et al., 2013) บนแนวคิดว่าจุดข้อมูลที่อยู่ในพื้นที่ที่มีความหนาแน่นสูงควรถูกจัดให้อยู่ในกลุ่มเดียวกัน ขณะที่จุดที่อยู่ในพื้นที่ความหนาแน่นต่ำหรือโดดเดี่ยวจะถูกพิจารณาเป็นสัญญาณรบกวน (noise) แต่ก็มีข้อจำกัดในเรื่องการปรับค่าพารามิเตอร์ เช่น ค่าขนาดกลุ่มที่ต่ำที่สุด (minimum cluster size) และจำนวนตัวอย่างขั้นต่ำ (minimum samples) ซึ่งส่งผลโดยตรงต่อประสิทธิภาพของผลลัพธ์

8.3 *Spectral Clustering* ที่อาศัยทฤษฎีกราฟและการแยกค่าเฉพาะ (eigenvalue) (Shi & Malik, 2000) จึงมีความสามารถสูงในการแยกกลุ่มข้อมูลที่มีโครงสร้างไม่เป็นเชิงเส้น (non-linear manifolds) เช่น รูปวงกลมหรือเกลียว ซึ่งวิธีที่อิงระยะทางยูคลิด (Euclidean) เชิงเส้นไม่สามารถทำได้ อย่างไรก็ตาม ความซับซ้อนเชิงการคำนวณ โดยเฉพาะการหาค่าเอกลักษณ์ของเมทริกซ์ขนาดใหญ่ เป็นข้อจำกัดสำคัญที่ต้องพิจารณา

9. งานวิจัยที่เกี่ยวข้อง

การจัดกลุ่มข้อมูลเป็นกระบวนการพื้นฐานในงานเรียนรู้ของเครื่องที่มุ่งแยกกลุ่มข้อมูลออกจากกัน โดยไม่ต้องพึ่งพายก่ากับที่ถูกกำหนดไว้ล่วงหน้า งานวิจัยในปัจจุบันมีการนำเทคนิคการจัดกลุ่มข้อมูลแบบไม่มีโครงสร้างเข้ามาจัดการกับข้อมูลทางธุรกิจที่ซับซ้อน ด้วยเทคนิคและวิธีการที่ต่างกัน ซึ่งจากการปริทัศน์วรรณกรรมและงานวิจัยที่เกี่ยวข้อง พบว่า วิธีการที่ได้รับความนิยมมากที่สุดคือการแบ่งกลุ่มแบบแบ่งส่วน (Partitioning methods) เช่น K-means (MacQueen, 1967) และ K-medoids (Kaufman & Rousseeuw, 1990) ซึ่งถูกนำมาใช้แพร่หลายในงานวิเคราะห์พฤติกรรมผู้บริโภคและการแบ่งกลุ่มตลาด เนื่องจากสามารถทำงานได้รวดเร็วกับข้อมูลจำนวนมาก และให้ผลลัพธ์ที่ง่ายต่อการทำความเข้าใจ โดยงานวิจัยของ Kumar and Shah (2006) แสดงให้เห็นถึงประสิทธิภาพของอัลกอริทึม K-means ในการจัดกลุ่มลูกค้าปลีก (retail customers) อย่างไรก็ตาม วิธีนี้มักให้ผลลัพธ์ที่แม่นยำก็ต่อเมื่อข้อมูลมีการกระจายตัวเป็นทรงกลมหรือสมมาตร

หากข้อมูลมีรูปร่างที่ซับซ้อนหรือมีค่าสุดโต่งผลลัพธ์จะไม่เสถียรนัก สำหรับข้อมูลที่มีโครงสร้างเป็นลำดับขั้นหรือต้องการมุมมองหลายระดับ มักใช้การจัดกลุ่มแบบเชิงลำดับขั้น (hierarchical clustering) โดยการสร้างเดนไดรแกรม (dendrogram) ช่วยให้เห็นการรวมและแยกของกลุ่มได้อย่างยืดหยุ่น จึงถูกนำมาประยุกต์ใช้ในการแบ่งกลุ่มลูกค้าตามระดับการให้บริการ เพื่อวิเคราะห์ภาพรวมและรายละเอียดเชิงลึก (Everitt et al., 2011) เนื่องจากไม่จำเป็นต้องกำหนดจำนวนกลุ่มล่วงหน้า แต่จะสร้างเดนไดรแกรม

หรือโครงสร้างที่มีลักษณะคล้ายต้นไม้ (tree-like structure) ที่ช่วยให้สามารถมองเห็นความสัมพันธ์ระหว่างลูกค้าในหลายระดับของความคล้ายคลึงกัน (Wedel & Kamakura, 2000) อย่างไรก็ตาม วิธีนี้จำเป็นต้องใช้ทรัพยากรการประมวลผลในระดับสูง

สำหรับข้อมูลที่มีขนาดใหญ่ อัลกอริทึมแบบอิงความหนาแน่น (density-based methods) มีบทบาทสำคัญ ในการวิเคราะห์ข้อมูลเชิงพื้นที่ หรือข้อมูลภาพถ่ายดาวเทียม ด้วยเทคนิคอย่าง DBSCAN ที่สามารถจัดกลุ่มที่มีรูปร่างไม่สม่ำเสมอได้อย่างมีประสิทธิภาพ และยังสามารถตรวจจับค่าสุดโต่งได้ในเวลาเดียวกัน (Ester et al., 1996) ต่อมาอัลกอริทึม OPTICS ได้ถูกพัฒนาขึ้นเพื่อแก้ข้อจำกัดด้านการตั้งค่าพารามิเตอร์เพียงค่าเดียว ทำให้เหมาะสมกับข้อมูลที่มีความหนาแน่นไม่สม่ำเสมอ (Ankerst et al., 1999) ก่อนจะพัฒนาเป็น HDBSCAN ที่ได้รับความนิยมอย่างสูง เนื่องจากสามารถค้นหากลุ่มที่มีความหนาแน่นแตกต่างกันได้ และให้ผลลัพธ์ที่เสถียรมากกว่าในการใช้งานจริง (Campello et al., 2013) เช่น การตรวจจับกลุ่มลูกค้าที่มีพฤติกรรมเฉพาะกลุ่ม (McInnes et al., 2017)

ในมุมมองของการสร้างแบบจำลองความน่าจะเป็นนั้น GMM ถือเป็นอีกเทคนิคที่สำคัญ จากแนวคิดการสมมติว่าข้อมูลเกิดจากการผสมของการแจกแจงปกติหลายชุด (Dempster et al., 1977) ทำให้ได้ผลลัพธ์ที่อนุญาตให้จุดข้อมูลหนึ่ง ๆ สามารถเป็นสมาชิกของหลายกลุ่มพร้อมกันได้ (soft clustering) ซึ่งมีประโยชน์อย่างมากในงานที่ต้องการจัดกลุ่มแบบซ้อนทับกัน เช่น งานวิจัยของ Gupta and Ghosh (2006) นำเสนอการแบ่งกลุ่มลูกค้าทางการเงิน โดยชี้ให้เห็นว่ามีความสามารถในการจัดการกับรูปร่างกลุ่มที่ไม่เป็นระเบียบ (irregular cluster shapes) และความแตกต่างของพฤติกรรมสำหรับข้อมูลที่มีโครงสร้างซับซ้อน

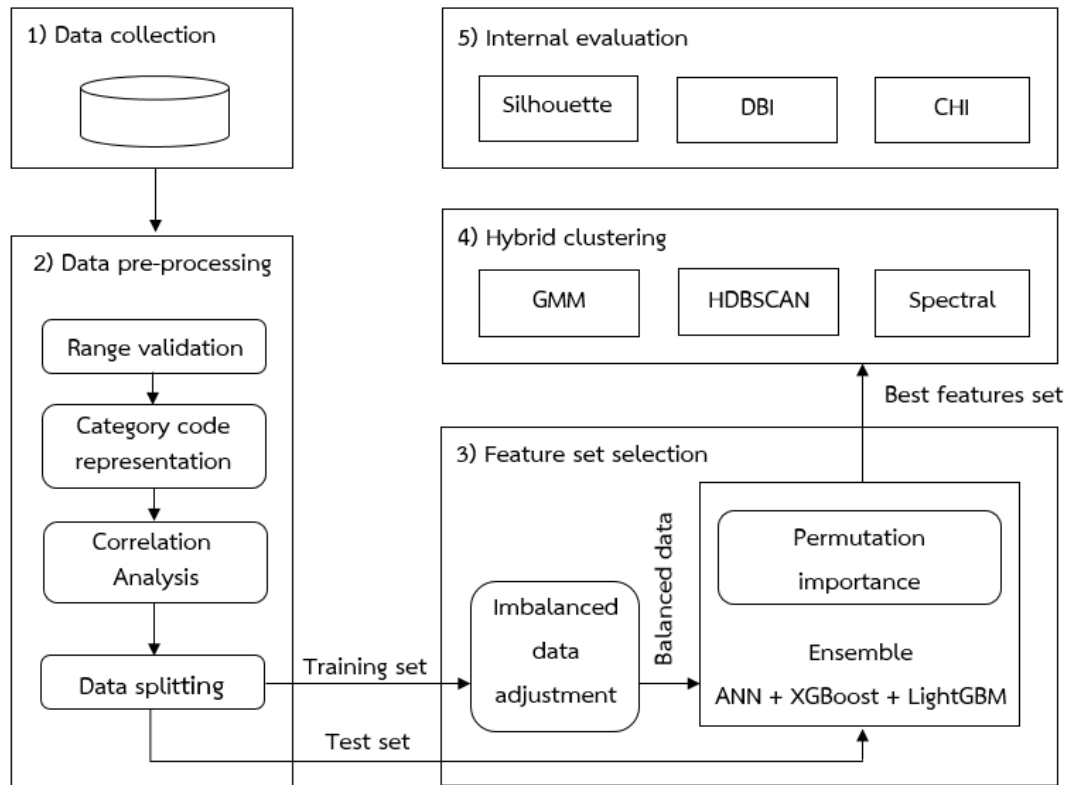
ในเชิงกราฟนั้นเทคนิค Spectral clustering กลายเป็นเครื่องมือที่เหมาะสม เนื่องจากสามารถใช้คุณสมบัติของกราฟ Laplacian เพื่อแบ่งกลุ่มโหนดที่เชื่อมโยงกันหนาแน่นจากข้อมูลในเครือข่ายสังคมออนไลน์หรือโครงสร้างข้อมูลในองค์กร (Ng et al., 2001) ทำให้สามารถตรวจจับความสัมพันธ์ที่ซับซ้อนระหว่างลูกค้าในการทำธุรกรรมทางการเงินที่มีการเปลี่ยนแปลงอย่างต่อเนื่อง (Von Luxburg, 2007) ในทำนองเดียวกัน วิธีการตรวจจับชุมชน (community detection) เช่น Louvain and Leiden ก็ถูกนำมาใช้เพื่อระบุกลุ่มย่อยในเครือข่ายทางธุรกิจ (as cited in Traag et al., 2019) หรือความสัมพันธ์ระหว่างผู้ใช้งานกับผู้จัดจำหน่ายสินค้า (Blondel et al., 2008) อีกเทคนิคหนึ่งที่ถูกใช้มากขึ้นคือการจัดกลุ่มแบบคลุมเครือ (fuzzy clustering) โดยเฉพาะการใช้เทคนิค Fuzzy C-means ซึ่งเหมาะกับกรณีที่ข้อมูลมีความซ้อนทับกันสูง พฤติกรรมผู้บริโภคบางอย่างที่อาจอยู่ได้มากกว่าหนึ่งกลุ่ม ผลลัพธ์ในลักษณะนี้ช่วยให้ธุรกิจออกแบบกลยุทธ์การตลาดที่ยืดหยุ่นต่อผู้บริโภคมากขึ้น นอกจากนี้ยังมีการพัฒนา Subspace clustering เพื่อจัดการกับข้อมูลมิติสูง ตัวอย่างเช่น ข้อมูลธุรกรรมทางการเงินที่แต่ละกลุ่มอาจมีความโดดเด่นเฉพาะในบางมิติ โดยเทคนิคอย่าง CLIQUE ถูกนำมาใช้เพื่อระบุ Subspace ที่มีความหมาย ทำให้สามารถค้นพบความสัมพันธ์ที่อาจสูญหายไปหากใช้วิธีแบบดั้งเดิม (Agrawal et al., 1998)

เมื่อธุรกิจต้องการจัดการกับข้อมูลที่มีการเปลี่ยนแปลงตลอดเวลาอย่างข้อมูลสตรีมมิงจากธุรกรรมออนไลน์หรือ IoT (Internet of Thing) วิธีการจัดกลุ่มข้อมูลแบบสตรีมมิง (streaming clustering) CluStream หรือ DenStream เป็นอาทิ ถูกนำมาใช้เพื่อปรับปรุงเพื่อจัดกลุ่มอย่างต่อเนื่องโดยไม่ต้องจัดเก็บข้อมูลทั้งหมดในหน่วยความจำ (Aggarwal et al., 2003) ทำให้สามารถตรวจสอบพฤติกรรมผู้ใช้งานได้แบบเรียลไทม์ ในที่สุดเมื่อข้อมูลมีความซับซ้อนสูงอย่างข้อมูลภาพหรือข้อความ ส่งผลให้การจัดกลุ่มเชิงลึก (deep clustering) ได้รับความสนใจเพิ่มมากขึ้น โดยเทคนิคอย่าง DEC (Deep Embedding Clustering) ที่ใช้โครงข่ายประสาทเทียมเชิงลึกเพื่อเรียนรู้การเป็นตัวแทนของข้อมูลก่อนทำการจัดกลุ่ม ซึ่งช่วยให้ได้ข้อมูลแบบฝัง (embedding) ที่เหมาะสมและสามารถเพิ่มความแม่นยำในการจัดกลุ่มได้มากขึ้น (Xie et al., 2016) โดยเฉพาะในงานที่ต้องการจัดการกับคุณลักษณะที่มีความซับซ้อนสูง

จะเห็นได้ว่าแต่ละเทคนิคต่างก็มีจุดเด่นและข้อจำกัดที่แตกต่างกัน โดยเทคนิคต่าง ๆ ก็ถูกพัฒนาขึ้นเพื่อตอบโจทย์ข้อมูลที่มีลักษณะแตกต่างกัน ทั้งเชิงรูปทรง ความหนาแน่น มิติของข้อมูล ตลอดจนการไหลเข้าของข้อมูล ซึ่งการเลือกใช้อัลกอริทึมจึงต้องพิจารณาจากลักษณะข้อมูลและเป้าหมายทางธุรกิจเป็นหลัก จากที่กล่าวมาสะท้อนให้เห็นว่าการจัดกลุ่มข้อมูลมิได้มีคำตอบที่ตายตัว แต่เป็นการปรับใช้ให้เหมาะสมกับสถานการณ์จริง เพื่อให้ได้ผลลัพธ์ที่มีความแม่นยำและสะท้อนพฤติกรรมผู้บริโภคได้รอบด้านยิ่งขึ้น ดังนั้นผลลัพธ์ของงานวิจัยนี้จึงเป็นการสังเคราะห์โมเดลเชิงคณิตศาสตร์ขั้นสูงเข้ากับทฤษฎีการแบ่งกลุ่มผู้บริโภคยุคใหม่ โดยมุ่งแก้ข้อจำกัดของระเบียบวิธีวิจัยดั้งเดิมเพื่อสร้างตัวแบบที่มีความแม่นยำสูงขึ้น ซึ่งการออกแบบการศึกษาในลักษณะนี้ไม่เพียงแต่ช่วยเพิ่มความแม่นยำทางสถิติ แต่ยังเป็นการสร้างกรอบการวิเคราะห์ที่สามารถอธิบายถึงพลวัตของพฤติกรรมรายปลีกบริการในอุตสาหกรรมสตรีมมิงได้อย่างรอบด้านกว่าการใช้เพียงเทคนิคใดเทคนิคหนึ่งตามแนวทางเดิม

วิธีดำเนินการวิจัย

การพัฒนาตัวแบบการแบ่งกลุ่มผู้บริโภคเนื้อหาจากการคัดเลือกคุณลักษณะที่มีความสำคัญต่อการยกเลิกบริการด้วยการจัดกลุ่มแบบผสมผสาน ประกอบด้วย 5 ขั้นตอน ได้แก่ การสร้างชุดข้อมูล การเตรียมข้อมูลก่อนการประมวลผล การคัดเลือกชุดคุณลักษณะ การจัดกลุ่มแบบผสมผสาน และการประเมินผลภายใน ดังรูปที่ 1



รูปที่ 1: ตัวแบบการแบ่งกลุ่มผู้บริโภคเนื้อหาจากการคัดเลือกคุณลักษณะที่มีความสำคัญต่อการยกเลิกบริการด้วยการจัดกลุ่มแบบผสมผสาน

1. การสร้างชุดข้อมูล (Data Collection)

งานวิจัยนี้กำหนดประชากรคือกลุ่มประชากรสัญชาติไทยตามข้อมูลเดือนธันวาคม พ.ศ. 2567 ที่อาศัยอยู่ในพื้นที่กรุงเทพมหานคร และปริมณฑลที่ประกอบไปด้วย 5 จังหวัด ได้แก่ นนทบุรี นครปฐม ปทุมธานี สมุทรปราการ และสมุทรสาคร จำนวนทั้งสิ้น 10,732,879 ราย (The Bureau of Registration Administration, 2025) ซึ่งถือว่าเป็นกลุ่มที่มีโอกาสเข้าถึงบริการยูทูปพรีเมียม มากที่สุด เนื่องจากเป็นศูนย์กลางทางเศรษฐกิจ โครงสร้างพื้นฐาน และการพัฒนาประเทศ เก็บรวบรวมข้อมูลโดยใช้วิธีการวิจัยแบบสำรวจผ่านแบบสอบถามออนไลน์ที่เปิดเป็นสาธารณะ ตั้งแต่เดือนมกราคมถึงกันยายน พ.ศ. 2568 รวมทั้งสิ้น 9 เดือน แบ่งตัวแปรออกเป็น 4 หมวดหมู่ ประกอบด้วย 1) ข้อมูลตัวเลขแบบจำนวนเต็ม (integer numerical) ที่สามารถนับได้เป็นตัวเลข และมีค่าเป็นจำนวนเต็ม 2) ข้อมูลหมวดหมู่แบบไม่จัดลำดับ (nominal categorical) เป็นข้อมูลที่ถูกระบุประเภทหรือกลุ่ม โดยไม่มีการเรียงลำดับหรือเปรียบเทียบกัน 3) ข้อมูลหมวดหมู่แบบจัดลำดับ (ordinal categorical) เป็นข้อมูลแบบลำดับขั้น สามารถเปรียบเทียบมากกว่าหรือน้อยกว่าได้ แต่ไม่สามารถระบุระยะห่างเชิงปริมาณระหว่างลำดับได้อย่างชัดเจน และ 4) ข้อมูลหมวดหมู่แบบไบนารี (binary categorical) ที่มีเพียงสองทางเลือกเท่านั้น โดยแสดงสถานะของการมีหรือไม่มี หรือการเกิดขึ้นหรือไม่เกิดขึ้นของเหตุการณ์ ดังตารางที่ 1

ตารางที่ 1: การจัดกลุ่มตัวแปรตามประเภทข้อมูล

ชนิดตัวแปร	ตัวแปร	ความหมาย	ค่าที่เป็นไปได้
Numeric (Integer)	Age	อายุ	จำนวนเต็ม (ปี)
	Period	ระยะเวลาใช้งาน	จำนวนเต็ม (เดือน)
	WatchTime	ชั่วโมงรับชมต่อวัน	จำนวนเต็ม (ชั่วโมง)
	ContentCount	จำนวนวิดีโอที่รับชมต่อวัน	จำนวนเต็ม (รายการ)
	LoginFrequency	ความถี่การเข้าสู่ระบบ	จำนวนเต็ม (ครั้งต่อวัน)
	Playlists	จำนวนเพลย์ลิสต์ที่สร้าง/ติดตาม	จำนวนเต็ม (เพลย์ลิสต์)
	Subscribed	จำนวนช่องที่สมัครรับข้อมูล	จำนวนเต็ม (ช่อง)
	ServiceChanges	จำนวนครั้งที่เปลี่ยนแปลงบริการ	จำนวนเต็ม (ครั้ง)
	PaymentIssues	จำนวนครั้งที่ปัญหาการชำระเงิน	จำนวนเต็ม (ครั้ง)
	SupportTickets	จำนวนครั้งที่ติดต่อฝ่ายสนับสนุน	จำนวนเต็ม (ครั้ง)
Categorical (Nominal)	Sex	เพศ	ชาย, หญิง
	Place	สถานที่หลักในการใช้บริการ	บ้าน, ที่ทำงาน, ระหว่างเดินทาง, สถานที่สาธารณะ
	TopContent	ประเภทเนื้อหาที่รับชมบ่อย	เสียง, ภาพและเสียง
	TopViewContent	ประเภทเนื้อหาหลักที่รับชม	เกม, เพลง, ข่าวสาร, กีฬา, พอดแคสต์, ภาพยนตร์/ทีวี, ความงาม/แฟชั่น, การศึกษา, สยองขวัญ, อื่น ๆ
	PaymentMethod	วิธีการชำระเงิน	บัตรเครดิต, Google Play, AIS, True Money
	TypeSubscription	รูปแบบการสมัครสมาชิก	แผนส่วนบุคคล, แผนครอบครัว, แผนนักเรียน/นักศึกษา
	AcquisitionSource	แหล่งที่มาของลูกค้า	การโฆษณา, การแนะนำ, การทดลองใช้ฟรี
Categorical (Ordinal)	InternetService	ประเภทอินเทอร์เน็ต	Wi-Fi, 4G/5G
	Married	สถานะสมรส	โสด, สมรส, หม้าย/หย่าร้าง
	Contract	ระยะเวลาสัญญา	รายเดือน, รายปี
Categorical (Binary)	Child	มีบุตร	ไม่มีบุตร, มีบุตร
	TechSupport	เคยรับบริการแก้ปัญหาหรือไม่	ไม่เคย, เคย
	StreamingTV	มีบริการสตรีมมิงทีวีหรือไม่	ไม่มี, มี
	DigitalTV	มีบริการดิจิทัลทีวีหรือไม่	ไม่มี, มี
	Cancellation*	การยกเลิกบริการ	ไม่ยกเลิก, ยกเลิก

หมายเหตุ: * ตัวแปรตาม (dependent variable)

2. การเตรียมข้อมูลก่อนการประมวลผล (Data Pre-processing)

การเตรียมข้อมูลก่อนการประมวลผลนั้นมีวัตถุประสงค์สำคัญเพื่อให้ข้อมูลมีความถูกต้อง สมบูรณ์ และอยู่ในรูปแบบที่เหมาะสมต่อการวิเคราะห์ผล ประกอบด้วย 4 ขั้นตอน ได้แก่ การตรวจสอบช่วงของตัวเลข การแทนค่ารหัสหมวดหมู่ การตรวจสอบความสอดคล้องเชิงตรรกะ และการแบ่งชุดข้อมูล มีรายละเอียดดังนี้

2.1 การตรวจสอบช่วงของตัวเลข (Range Validation) แม้การเก็บข้อมูลด้วยแบบสอบถามนั้นจะช่วยลดปัญหาข้อมูลที่สูญหาย (missing value) และปัญหาข้อมูลผิดรูปแบบได้ แต่การให้ผู้ใช้ระบุข้อมูลจำนวนเต็ม เช่น อายุ หรือระยะเวลาใช้งาน อาจเสี่ยงต่อค่าที่ผิดปกติ (outlier) อันเกิดจากความตั้งใจหรือไม่ตั้งใจก็ ดังนั้นจึงต้องพิจารณาค่าที่ไม่สอดคล้องกับความเป็นจริง เพื่อให้ข้อมูลมีความสมบูรณ์และสอดคล้องกันมากที่สุด ทั้งนี้หากตรวจพบค่าที่นอกช่วงตัวเลขจะทำเครื่องหมายแยกออกมาเป็น “ผิดรูปแบบ”

2.2 การแทนค่ารหัสหมวดหมู่ (Categorical Code Representation) เนื่องจากข้อมูลที่อยู่ในรูปของจำนวนเต็ม สามารถนำไปใช้ในการวิเคราะห์และประมวลผลได้โดยตรง โดยไม่จำเป็นต้องมีการแปลงรหัสเพิ่มเติม เนื่องจากตัวเลขเหล่านี้มีความหมายเชิงปริมาณ และสามารถคำนวณทางสถิติได้ทันที ในทางกลับกัน ตัวแปรเชิงหมวดหมู่ จำเป็นต้องทำการแปลงให้อยู่ในรูปตัวเลขเพื่อความสะดวกในการประมวลผลเชิงคำนวณ

2.3 การวิเคราะห์สหสัมพันธ์ (Correlation Analysis) เป็นกระบวนการทางสถิติที่ใช้ในการตรวจสอบความสัมพันธ์เชิงเส้นระหว่างตัวแปรเชิงปริมาณ เพื่อระบุว่าตัวแปรหนึ่งมีแนวโน้มเปลี่ยนแปลงไปในทิศทางเดียวกันหรือตรงข้ามกับอีกตัวแปรหนึ่งมากน้อยเพียงใด นอกจากนี้ การวิเคราะห์ความสัมพันธ์ยังมีบทบาทสำคัญในการคัดเลือกคุณลักษณะที่มีนัยสำคัญต่อตัวแปรตามหรือตัวแปรเป้าหมาย เนื่องจากสามารถลดการถูกรบกวนจากตัวแปรที่ไม่เกี่ยวข้อง อีกทั้งยังช่วยตรวจสอบปัญหาการเกิดความสัมพันธ์เชิงเส้นพหุ (multicollinearity) ระหว่างตัวแปรอิสระที่สูงเกินไป ซึ่งอาจส่งผลกระทบต่อประสิทธิภาพในการเรียนรู้ของโมเดล

2.4 การแบ่งชุดข้อมูล (Data Splitting) คือการแยกชุดข้อมูลต้นฉบับออกเป็นส่วยย่อยเพื่อใช้ในการฝึกฝนและประเมินประสิทธิภาพ นอกจากนี้ ยังช่วยลดความเสี่ยงต่อการเกิดภาวะที่โมเดลไม่สามารถจับความสัมพันธ์ของข้อมูลได้อย่างมีประสิทธิภาพ (underfitting) ตลอดจนส่งผลให้โมเดลสามารถพยากรณ์ผลลัพธ์กับข้อมูลที่ไม่เคยเห็นมาก่อนได้อย่างแม่นยำมากยิ่งขึ้น

3. การคัดเลือกชุดคุณลักษณะ (Feature Selection)

เนื่องจากข้อมูลที่นำมาใช้ในการวิเคราะห์มีความไม่สมดุลของคลาส อาจก่อให้เกิดอคติในกระบวนการเรียนรู้ของเครื่องจนอาจนำไปสู่ความคลาดเคลื่อนในการพยากรณ์ จึงจำเป็นต้องมีการปรับค่าพารามิเตอร์ เพื่อช่วยถ่วงน้ำหนักให้ความผิดพลาดในคลาสย่อยบริการให้มีความสำคัญมากขึ้นระหว่างฝึกฝนโมเดลด้วยการคัดเลือกตัวแปรทีละขั้น (stepwise feature elimination) แบบคัดเลือกย้อนกลับ (backward elimination) ด้วยอัลกอริทึมการเรียนรู้แบบรวม ประกอบด้วยโมเดลโครงข่ายประสาทเทียม (neural network model) ร่วมกับโมเดล XGBoost ที่เป็นอัลกอริทึมเพิ่มประสิทธิภาพแบบไล่ระดับ (gradient boosting algorithm) (Chen & Guestrin, 2016) บนพื้นฐานของต้นไม้ตัดสินใจ เช่นเดียวกับ LightGBM ที่ถูกออกแบบให้ทำงานได้เร็วกว่าโมเดลอื่น ๆ ในกลุ่มเดียวกัน (Ke et al., 2017) โดยเริ่มจากการใช้ตัวแปรทั้งหมดและค่อย ๆ ลบตัวแปรที่มีความสำคัญน้อยที่สุดทีละตัว จนกว่าจะได้ชุดตัวแปรคุณลักษณะที่ให้ผลลัพธ์ดีที่สุดเมื่อวัดผลจากค่าประสิทธิภาพโดยรวมจากกระบวนการฝึกฝนแบบหลายรอบ (iterative feature selection) เพื่อให้การคัดเลือกตัวแปรที่มีผลกระทบสูงในการพยากรณ์อย่างมีนัยสำคัญ

4. การจัดกลุ่มแบบผสมผสาน (Hybrid Clustering)

เมื่อได้ชุดคุณลักษณะที่ดีที่สุดแล้ว ขั้นตอนถัดไปคือการนำข้อมูลดังกล่าวมาประยุกต์ใช้ในการจัดกลุ่มผู้บริโภค โดยมีเป้าหมายเพื่อแบ่งกลุ่มผู้บริโภคที่มีลักษณะพฤติกรรมคล้ายคลึงกัน การจัดกลุ่มนี้ช่วยให้สามารถทำความเข้าใจรูปแบบพฤติกรรมเชิงลึกของผู้บริโภคที่มีแนวโน้มยกเลิกบริการได้อย่างแม่นยำ อีกทั้งยังเป็นฐานข้อมูลสำคัญที่สนับสนุนการออกแบบกลยุทธ์เชิงธุรกิจที่ตรงเป้าหมาย

งานวิจัยนี้เลือกใช้แนวทางการจัดกลุ่มแบบผสมผสาน เพื่อเพิ่มความแม่นยำและความเสถียรของผลการวิเคราะห์ และลดข้อจำกัดของการใช้วิธีใดวิธีหนึ่งเพียงลำพัง โดยอาศัยจุดแข็งของเทคนิคที่แตกต่างกันมาเสริมซึ่งกันและกันในการจัดการกับข้อมูลที่มีความซับซ้อนและหลากหลายมิติได้อย่างครอบคลุม ประกอบด้วย 3 โมเดล ได้แก่ 1) GMM ที่เหมาะกับข้อมูลที่ซ้อนทับกัน และสามารถคำนวณความน่าจะเป็นของสมาชิกในแต่ละกลุ่มได้อย่างยืดหยุ่น 2) HDBSCAN ซึ่งสามารถตรวจจับรูปร่างกลุ่มที่ไม่เป็นเชิงเส้นและระบุข้อมูลค่าสุดโต่งได้อย่างมีประสิทธิภาพ และ 3) Spectral clustering ที่อาศัยโครงสร้างกราฟเพื่อค้นหาความสัมพันธ์เชิงลึก การบูรณาการวิธีการทั้งสามทำให้ผลลัพธ์การจัดกลุ่มมีความครอบคลุมทั้งในเชิงความน่าจะเป็น ความหนาแน่น และโครงสร้างเครือข่าย ส่งผลให้การวิเคราะห์มีความแข็งแกร่งและเชื่อถือได้ ก่อนจะแสดงผลแบบภาพข้อมูล (data visualization) เพื่อสำรวจและประเมินผลของการจัดกลุ่มแต่ละเทคนิคอย่างชัดเจน

5. การประเมินผลภายใน (Internal Evaluation)

การประเมินผลภายในเป็นวิธีที่ใช้ตรวจสอบประสิทธิภาพของแบบจำลองการจัดกลุ่ม โดยไม่จำเป็นต้องอาศัยข้อมูลป้ายกำกับจริง (ground truth) แต่จะประเมินจากโครงสร้างและความสัมพันธ์เชิงระยะทางระหว่างข้อมูลภายในแบบจำลองเองประกอบด้วยค่า Silhouette score, DBI และ CHI ซึ่งสะท้อนคุณภาพของการจัดกลุ่มที่มีความกะชับ (compactness) และสามารถแยกความแตกต่างระหว่างกลุ่ม (group separation) ได้อย่างมีประสิทธิภาพ มีรายละเอียดดังนี้

5.1 คำ *Silhouette score* สำหรับประเมินความเหมาะสมของการจัดกลุ่ม โดยพิจารณาระยะห่างของจุดข้อมูลภายในกลุ่มเดียวกันเมื่อเทียบกับระยะทางระหว่างกลุ่มอื่น ซึ่งใช้เป็นตัวชี้วัดโดยรวมของคุณภาพการจัดกลุ่ม

5.2 คำ *Davies-Bouldin Index (DBI)* สำหรับใช้ประเมินประสิทธิภาพของการจัดกลุ่ม โดยพิจารณาจากความสัมพันธ์ระหว่างการกระจายตัวภายในกลุ่ม (intra-cluster dispersion) และระยะห่างระหว่างกลุ่ม (inter-cluster distance) ดังนั้น DBI ที่มีค่าต่ำแสดงถึงการจัดกลุ่มที่มีคุณภาพดี เนื่องจากการกระจายภายในกลุ่มมีระดับต่ำและมีความแตกต่างระหว่างกลุ่มในระดับสูง

5.3 คำ *Calinski-Harabasz Index (CHI)* หรือ Variance ratio criterion ใช้วัดอัตราส่วนระหว่างความแปรปรวนระหว่างกลุ่ม (between-cluster dispersion) กับความแปรปรวนภายในกลุ่ม (within-cluster dispersion) ดังนั้น CHI ที่มีค่าสูงบ่งบอกถึงการจัดกลุ่มที่มีความหนาแน่นภายในกลุ่มสูงและมีการแยกกลุ่มกันอย่างชัดเจน

ผลการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาตัวแบบการแบ่งกลุ่มผู้บริโภคเนื้อหาจากการคัดเลือกคุณลักษณะที่มีความสำคัญต่อการยกเลิกบริการด้วยการจัดกลุ่มแบบผสมผสาน แบ่งผลการวิจัยออกเป็น 5 ส่วน มีรายละเอียดดังนี้

1. ผลการสร้างชุดข้อมูล

งานวิจัยนี้รวบรวมข้อมูลจากผู้ให้บริการยูทูบพรีเมียมในพื้นที่กรุงเทพฯ และปริมณฑลด้วยวิธีการวิจัยแบบการสำรวจผ่านทางแบบสอบถามออนไลน์ที่เปิดเป็นสาธารณะ เมื่อตรวจสอบความถูกต้องของข้อมูล พบว่า มีข้อมูลทั้งสิ้น 10,983 รายการที่ตรงตามข้อกำหนดในการวิจัย จากนั้นจึงนำเข้าสู่การวิเคราะห์ข้อมูลเบื้องต้น ดังตารางที่ 2 ประกอบด้วยค่าน้อยที่สุด ค่ามากที่สุด ค่าเฉลี่ย และค่ามัธยฐานสำหรับข้อมูลที่อยู่ในรูปของจำนวนเต็ม และตารางที่ 3 การแจกแจงความถี่ สำหรับตัวแปรเชิงหมวดหมู่

ตารางที่ 2: ค่าน้อยที่สุด ค่ามากที่สุด ค่าเฉลี่ย และค่ามัธยฐาน สำหรับข้อมูลที่อยู่ในรูปของจำนวนเต็ม

ตัวแปร	ความหมาย	ค่าน้อยที่สุด	ค่ามากที่สุด	ค่าเฉลี่ย	ค่ามัธยฐาน
Age	อายุ (ปี)	18	70	43.61	43
Period	ระยะเวลาใช้งาน (เดือน)	1	120	47.80	44
WatchTime	ชั่วโมงรับชมต่อวัน (ชั่วโมง)	0	12	5.26	5
ContentCount	จำนวนวิดีโอที่รับชมต่อวัน (รายการ)	0	50	17.61	14
LoginFrequency	ความถี่การเข้าสู่ระบบ (ครั้งต่อวัน)	0	30	15.15	15
Playlists	จำนวนเพลย์ลิสต์ที่สร้าง/ติดตาม (เพลย์ลิสต์)	0	20	10.10	10
Subscribed	จำนวนช่องที่สมัครรับข้อมูล (ช่อง)	0	100	37.34	33
ServiceChanges	จำนวนครั้งที่เปลี่ยนแปลงบริการ (ครั้ง)	0	5	2.51	3
PaymentIssues	จำนวนครั้งที่ปัญหาการชำระเงิน (ครั้ง)	0	5	2.50	2
SupportTickets	จำนวนครั้งที่ติดต่อฝ่ายสนับสนุน (ครั้ง)	0	10	4.99	5

ตารางที่ 3: การแจกแจงความถี่ สำหรับตัวแปรเชิงหมวดหมู่

หมวดหมู่	ตัวแปร	ความหมาย	การแจกแจงคลาส	
Categorical (Nominal)	Sex	เพศ	ชาย	4,527
			หญิง	6,456
	Place	สถานที่หลักในการใช้บริการ	บ้าน	4,743
			ที่ทำงาน	2,706
			ระหว่างเดินทาง	1,738
			สถานที่สาธารณะ	1,796

ตารางที่ 3: การแจกแจงความถี่ สำหรับตัวแปรเชิงหมวดหมู่ (ต่อ)

หมวดหมู่	ตัวแปร	ความหมาย	การแจกแจงคลาส	
Categorical (Nominal) (ต่อ)	TopContent	ประเภทเนื้อหาที่รับชมบ่อย	เสียง	3,518
			ภาพและเสียง	7,465
	TopViewContent	ประเภทเนื้อหาหลักที่รับชม	เกม	1,165
			เพลง	1,657
			ข่าวสาร	1,925
			กีฬา	1,109
			พอดแคสต์	688
			ภาพยนตร์/ทีวี	746
			ความงาม/แฟชั่น	522
			การศึกษา	1,069
			สยองขวัญ	1,063
			อื่น ๆ	1,039
	PaymentMethod	วิธีการชำระเงิน	บัตรเครดิต	2,761
			Google Play	2,919
			AIS	2,922
			True Money	2,381
	TypeSubscription	รูปแบบการสมัครสมาชิก	แผนส่วนบุคคล	6,682
			แผนครอบครัว	1,678
			แผนนักเรียน/นักศึกษา	2,623
	AcquisitionSource	แหล่งที่มาลูกค้า	การโฆษณา	4,667
			การแนะนำ	1,644
			การทดลองใช้ฟรี	4,672
	InternetService	ประเภทอินเทอร์เน็ต	Wi-Fi	7,598
			4G/5G	3,385
Categorical (Ordinal)	Married	สถานะสมรส	โสด	6,023
			สมรส	2,998
			หม้าย/หย่าร้าง	1,962
	Contract	ประเภทของสัญญา	รายเดือน	6,627
			รายปี	4,356
Categorical (Binary)	Child	มีบุตร	ไม่มีบุตร	6,001
			มีบุตร	4,982
	TechSupport	เคยรับบริการแก้ปัญหาหรือไม่	ไม่เคย	6,049
			เคย	4,934
	StreamingTV	มีบริการสตรีมมิงทีวีหรือไม่	ไม่มี	8,550
			มี	2,433
	DigitalTV	มีบริการดิจิทัลทีวีหรือไม่	ไม่มี	3,560
			มี	7,423
	Cancellation*	ยกเลิกใช้บริการหรือไม่	ไม่ยกเลิก	7,526
			ยกเลิก	3,457

หมายเหตุ: * ตัวแปรตาม

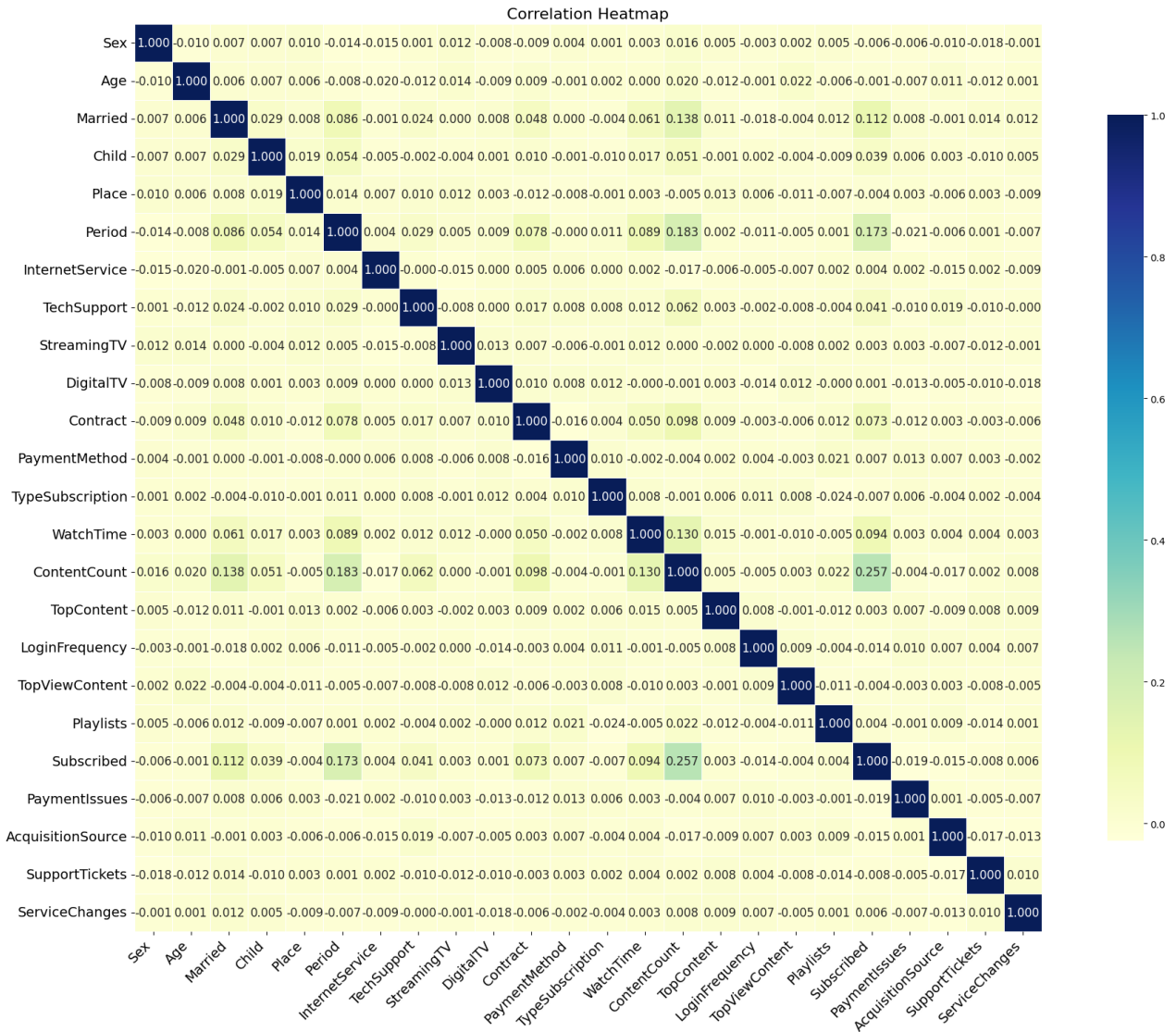
2. ผลการเตรียมข้อมูลก่อนการประมวลผล

เนื่องจากข้อมูลดิบอาจมีความไม่สมบูรณ์ เช่น ข้อมูลผิดรูปแบบหรือค่าที่ผิดปกติ รวมถึงค่าที่ไม่สอดคล้องกับความเป็นจริง จึงต้องผ่านขั้นตอนการเตรียมข้อมูลก่อนการประมวลผล เพื่อให้ข้อมูลมีความสมบูรณ์และสอดคล้องกันมากที่สุด ประกอบด้วย

2.1 ผลการตรวจสอบช่วงของตัวเลข สำหรับข้อมูลที่อยู่ในรูปของจำนวนเต็มจากการกำหนดขอบเขตที่เป็นไปได้ เช่น ตัวแปรอายุ (age) ถูกกำหนดให้มีค่าอยู่ระหว่าง 10–99 ปี ขณะที่ตัวแปรเวลาการรับชม (WatchTime) กำหนดให้อยู่ระหว่าง 0–24 ชั่วโมงต่อวัน นอกจากนี้ ตัวแปรที่เป็นหน่วยนับ เช่น จำนวนวิดีโอที่รับชมต่อวัน (ContentCount) จะกำหนดช่วงให้อยู่ที่ระหว่าง 0–500 รายการ เพื่อสะท้อนขอบเขตการใช้งานจริง ส่วนตัวแปรเชิงปริมาณอื่น ๆ เช่น ความถี่ในการเข้าสู่ระบบ (LoginFrequency) ระยะเวลาใช้งาน (period) จำนวนเพลย์ลิสต์ที่สร้าง/ติดตาม (playlists) จำนวนช่องที่สมัครรับข้อมูล (subscribed) รวมถึงจำนวนครั้งที่เปลี่ยนแปลงบริการ จำนวนครั้งที่มีการชำระเงิน และจำนวนครั้งที่ติดต่อฝ่ายสนับสนุน ได้แก่ ServiceChanges, PaymentIssues และ SupportTickets ตามลำดับ โดยทั้งหมดต้องมีค่าเป็นจำนวนเต็มที่ไม่ติดลบ (≥ 0) ทั้งนี้ หากตรวจพบค่าที่นอกช่วงตัวเลขจะพิจารณาตัดข้อมูลรายการดังกล่าวออกไป

2.2 ผลการแทนค่ารหัสหมวดหมู่ เพื่อให้ข้อมูลอยู่ในรูปตัวเลขที่เหมาะสมสำหรับการประมวลผล กำหนดให้ตัวแปรเชิงหมวดหมู่แบบไม่จัดลำดับ (nominal categorical) เช่น เพศ (sex) สถานที่หลักในการใช้บริการ (place) ประเภทเนื้อหาที่รับชมบ่อย (TopContent) ประเภทเนื้อหาหลักที่รับชม (TopViewContent) วิธีการชำระเงิน (PaymentMethod) รูปแบบการสมัครสมาชิก (TypeSubscription) แหล่งที่มาของลูกค้า (AcquisitionSource) และประเภทอินเทอร์เน็ต (InternetService) จะใช้การเข้ารหัสแบบ one-hot encoding เพื่อหลีกเลี่ยงการตีความลำดับที่ผิดพลาด ขณะที่ข้อมูลหมวดหมู่แบบจัดลำดับ (ordinal categorical) เช่น สถานะสมรส (married) และประเภทของสัญญา (contract) ซึ่งจะแทนค่าเป็นลำดับตัวเลข เช่น 0 = โสด, 1 = สมรส, 2 = หม้าย/หย่าร้าง หรือ 0 = รายเดือน, 1 = รายปี ตามลำดับ เพื่อสะท้อนถึงลำดับเชิงตรรกะของตัวแปร และข้อมูลหมวดหมู่แบบไบนารี เช่น มีบุตร (child) การรับบริการแก้ปัญหา (TechSupport) การมีบริการสตรีมมิงทีวี (StreamingTV) และการมีบริการดิจิทัลทีวี (DigitalTV) สามารถแทนค่าเป็น 0 และ 1 ได้โดยตรง เช่น มี/ไม่มี หรือเคย/ไม่เคย เป็นต้น ดังนั้นการยกเลิกใช้บริการ (cancellation) หากยังคงใช้บริการอยู่ในปัจจุบันจะกำหนดค่าเป็น 0 (No) และหากเคยใช้บริการ แต่ปัจจุบันไม่ได้ใช้บริการแล้วจะกำหนดค่าเป็น 1 (Yes)

2.3 ผลการวิเคราะห์สหสัมพันธ์ จากการวิเคราะห์ความสัมพันธ์ของตัวแปรเชิงปริมาณทั้งหมดในชุดข้อมูล โดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน (Pearson correlation coefficient) พบว่า ตัวแปรส่วนใหญ่มีความสัมพันธ์เชิงเส้นในระดับต่ำถึงปานกลาง ทั้งในทิศทางบวกและทิศทางลบ ซึ่งสะท้อนให้เห็นถึงการกระจายตัวของคุณลักษณะที่ค่อนข้างเป็นอิสระต่อกัน ดังรูปที่ 2 แผนภาพความหนาแน่นของสหสัมพันธ์ แสดงให้เห็นค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรเชิงตัวเลขทั้งหมด ยกเว้นตัวแปร Cancellation ที่ถูกตัดออกจากการวิเคราะห์ เพื่อป้องกันการรบกวนของตัวแปรเป้าหมายในการประเมินความสัมพันธ์ระหว่างตัวแปรอื่น ๆ ผลการวิเคราะห์ พบว่า ไม่มีคู่ของตัวแปรใดที่มีค่าสัมประสิทธิ์สหสัมพันธ์เกิน 0.9 ซึ่งหมายความว่าไม่มีปัญหาความสัมพันธ์เชิงเส้นพหุที่รุนแรงระหว่างตัวแปรแต่ละตัว จึงสามารถคงตัวแปรทั้งหมดไว้สำหรับการสร้างแบบจำลองเชิงพยากรณ์ได้อย่างเหมาะสม นอกจากนี้ ยังพบว่า บางคู่ของตัวแปร เช่น Period กับ ContentCount ($r = 0.183$) และ ContentCount กับ Subscribed ($r = 0.257$) มีความสัมพันธ์เชิงบวกในระดับต่ำถึงปานกลาง ซึ่งอาจสะท้อนถึงความสัมพันธ์ของพฤติกรรมการใช้งานบริการในเชิงเนื้อหาและการสมัครสมาชิก ทั้งนี้ความสัมพันธ์เหล่านี้ควรได้รับการพิจารณาเพิ่มเติมในขั้นตอนการคัดเลือกคุณลักษณะ เพื่อให้แบบจำลองแม่นยำและเสถียรยิ่งขึ้น



รูปที่ 2: แผนภาพแสดงความหนาแน่นของสหสัมพันธ์

2.4 ผลการแบ่งชุดข้อมูล จากการแบ่งชุดข้อมูลจำนวน 10,983 รายการออกเป็นชุดข้อมูลฝึกฝน (training set) สำหรับเรียนรู้รูปแบบและคุณลักษณะสำคัญของข้อมูลร้อยละ 80 และร้อยละ 20 สำหรับชุดข้อมูลทดสอบ (test set) ในการประเมินความสามารถของโมเดล คิดเป็น 8,786 และ 2,197 รายการตามลำดับ เมื่อพิจารณาการแบ่งคลาสแบบ Stratify เพื่อรักษาสัดส่วนเดิมของคลาสพบว่า ผู้ใช้ที่ไม่ยกเลิกบริการและผู้ใช้ที่ยกเลิกบริการภายในชุดข้อมูลฝึกฝนมีจำนวนประมาณ 6,018 และ 2,768 ตัวอย่าง และในชุดข้อมูลทดสอบมีจำนวนประมาณ 1,508 และ 691 ตัวอย่าง ตามลำดับ

3. ผลการคัดเลือกชุดคุณลักษณะ

งานวิจัยนี้ให้ความสำคัญกับคลาสที่มีจำนวนตัวอย่างน้อย จึงกำหนดค่า *scale_pos_weight* เท่ากับ 2.18 จากจำนวนการยกเลิกใช้บริการคือไม่ยกเลิกเท่ากับ 7,526 รายการ และยกเลิกเท่ากับ 3,457 รายการ เนื่องจากในชุดข้อมูลฝึกฝนนั้นมีจำนวนตัวอย่างไม่ยกเลิกบริการมากกว่าจำนวนตัวอย่างที่ยกเลิกบริการถึงประมาณ 2.18 เท่า คิดเป็นสัดส่วนร้อยละ 68.53 และ 31.47 ตามลำดับ

การพยากรณ์การยกเลิกบริการยูทูปพรีเมียมของผู้บริโภคเนื้อหาด้วยการคัดเลือกตัวแปรที่ละชั้นแบบคัดเลือกย้อนกลับ โดยใช้โมเดลโครงข่ายประสาทเทียม ร่วมกับโมเดล XGBoost และ LightGBM พบว่า ผลการประเมินรอบเริ่มต้นที่มีตัวแปรทั้งหมด 24 ตัว โมเดลมีค่าความถูกต้อง (accuracy) เท่ากับ 0.8421 ค่าความแม่นยำ (precision) เท่ากับ 0.6954 ค่าการเรียกคืน (recall) เท่ากับ

0.8873 ค่าประสิทธิภาพโดยรวม (F1-score) เท่ากับ 0.7797 และค่า AUC เท่ากับ 0.9268 เมื่อทำการลดจำนวนตัวแปรลงทีละรอบพบว่า ค่าประสิทธิภาพโดยรวมของโมเดลมีแนวโน้มเพิ่มขึ้นอย่างค่อยเป็นค่อยไป จนถึงรอบที่ 18 ซึ่งมีตัวแปรเหลือ 7 ตัว และได้ค่าประสิทธิภาพโดยรวมสูงสุดที่ 0.7862 ก่อนที่จะลดลงเมื่อลบตัวแปรสำคัญเพิ่มเติม

เมื่อพิจารณาในรายละเอียดพบว่า ชุดคุณลักษณะที่สามารถรักษาประสิทธิภาพสูงสุดของโมเดลได้ประกอบด้วย 7 ตัวแปร ได้แก่ Period, Contract, WatchTime, ContentCount, Subscribed, PaymentIssues และ SupportTickets ซึ่งสะท้อนถึงความสามารถในการจำแนกผู้ยกเลิกและผู้ไม่ยกเลิกบริการอย่างสมดุล รายละเอียดดังตารางที่ 4

ตารางที่ 4: ชุดคุณลักษณะและค่าความสำคัญที่สามารถรักษาประสิทธิภาพสูงสุดของโมเดล

ลำดับ	คุณลักษณะ	ค่าความสำคัญ
1	ContentCount	0.2003
2	Subscribed	0.1509
3	Period	0.1376
4	PaymentIssues	0.1177
5	WatchTime	0.1163
6	SupportTickets	0.0665
7	Contract	0.0628

จากตารางที่ 4 พบว่า ContentCount เป็นตัวแปรที่มีความสำคัญสูงสุด ซึ่งสะท้อนว่าจำนวนเนื้อหาที่ผู้ใช้งานเข้าถึงมีผลต่อความเสี่ยงในการยกเลิกบริการอย่างเด่นชัด ผู้ใช้ที่เข้าถึงเนื้อหาบ่อยๆจะมีแนวโน้มที่จะยกเลิกบริการมากกว่า ในลำดับถัดมาคือ Subscribed ซึ่งแสดงให้เห็นถึงความสัมพันธ์ระหว่างการสมัครสมาชิกและความต่อเนื่องในการใช้งานบ่งชี้ว่าผู้ที่สมัครสมาชิกเต็มรูปแบบ มักมีความเสี่ยงต่อการยกเลิกบริการต่ำกว่า เช่นเดียวกับ Period ที่บ่งบอกวาระระยะเวลาการใช้งานก็เป็นอีกตัวแปรที่มีผลต่อความต่อเนื่องในการใช้งาน ยิ่งผู้ใช้มีระยะเวลาการใช้งานนานจะส่งผลให้ความเป็นไปได้ในการยกเลิกบริการจะลดลง นอกจากนั้น PaymentIssues แสดงให้เห็นถึงบทบาทของปัญหาการชำระเงินที่ยุ่งยากและซับซ้อน อาจทำให้ลูกค้าตัดสินใจยกเลิกบริการ ส่วน WatchTime นั้นสะท้อนระดับการมีส่วนร่วมของผู้ใช้งานกับแพลตฟอร์ม ซึ่งเป็นตัวชี้วัดสำคัญต่อพฤติกรรมบริการ ยกเลิกบริการ ตลอดจนตัวแปรที่แม้จะมีความสำคัญน้อยกว่า แต่ยังคงมีส่วนสนับสนุนการพยากรณ์ ได้แก่ SupportTickets จากการติดต่อฝ่ายสนับสนุนที่ชี้ถึงประสิทธิภาพการให้บริการและปัญหาที่ลูกค้าพบในการใช้งาน ส่วน Contract บ่งชี้ถึงลักษณะของสัญญาที่แตกต่างกันนั้นมีผลต่อความมั่นคงในการใช้งานของลูกค้า

4. ผลการจัดกลุ่มแบบผสมผสาน

หลังได้ชุดคุณลักษณะที่มีความสำคัญแล้ว จะนำตัวแปรทั้ง 7 ตัว มาสร้างเป็นตัวแปรเชิงอนุพันธ์ (derived features) ประกอบด้วย 4 ตัวแปร ดังนี้ 1) การใช้งานเนื้อหาต่อจำนวนเนื้อหา (WatchTime_per_Content) จากการหารค่า WatchTime ด้วย ContentCount + $1e-6$ เพื่อป้องกันการหารด้วยศูนย์ ตัวแปรนี้สะท้อนถึงระดับความสนใจและการมีส่วนร่วมของผู้บริโภคต่อเนื้อหาที่เข้าถึง ยิ่งค่าสูงหมายถึงผู้บริโภคใช้เวลาเข้าถึงเนื้อหาแม้จำนวนเนื้อหาน้อย 2) อัตราการสมัครบริการเสริม (Subscribed_Ratio) จากการหารค่า Subscribed ด้วย Period + $1e-6$ เพื่อวัดอัตราการสมัครบริการเสริมต่อช่วงระยะเวลาการเป็นสมาชิก ซึ่งช่วยสะท้อนพฤติกรรมการใช้บริการเสริมที่ต่อเนื่องและความผูกพันของผู้บริโภค 3) การร้องเรียนต่อช่วงเวลา (SupportTickets_per_Period) จากการหารค่า SupportTickets ด้วย Period + $1e-6$ เพื่อแสดงความถี่ในการร้องเรียนหรือขอความช่วยเหลือต่อช่วงเวลาการเป็นสมาชิก ตัวแปรนี้ช่วยระบุกลุ่มผู้บริโภคที่อาจมีปัญหาในการใช้บริการบ่อย ซึ่งเป็นสัญญาณของความไม่พึงพอใจหรือความเสี่ยงต่อการยกเลิกบริการ และ 4) ความเสี่ยงของการยกเลิกบริการ (Churn_Risk) มีค่าตั้งแต่ 0 ถึง 1 ซึ่งมาจากการสร้างโมเดล Gradient boosting classifier โดยใช้ตัวแปรหลัก 7 ตัวเป็นอินพุต และตัวแปร Cancellation เป็นตัวแปรตอบสนอง หลังจากฝึกฝนโมเดลแล้วจะได้ค่าความน่าจะเป็น

เป็นที่ลูกค้าจะยกเลิกบริการในอนาคต ดังนั้นเมื่อรวมตัวแปรเชิงอนุพันธ์กับตัวแปรหลักจึงทำให้มีตัวแปรทั้งหมด 11 ตัวแปร ซึ่งจะช่วยให้มีการวิเคราะห์ที่สะท้อนพฤติกรรมและความเสี่ยงเชิงลึก ซึ่งไม่สามารถสังเกตได้จากตัวแปรต้นเพียงอย่างเดียว ก่อนนำมาจัดกลุ่มผู้บริโภคที่มีลักษณะพฤติกรรมและความเสี่ยงต่อการยกเลิกบริการแตกต่างกัน ดังตารางที่ 5 โดยแบ่งความเสี่ยงออกเป็น 3 ระดับ ได้แก่ ระดับความเสี่ยงต่ำ (low risk) มีค่าเฉลี่ยความเสี่ยงการยกเลิกบริการตั้งแต่ 0.00–0.20 ซึ่งสะท้อนถึงกลุ่มผู้ใช้บริการที่มีความภักดีสูงและมีพฤติกรรมการใช้งานที่สม่ำเสมอ ระดับความเสี่ยงปานกลาง (medium risk) มีค่าเฉลี่ยความเสี่ยงการยกเลิกบริการมากกว่า 0.20–0.50 แสดงถึงกลุ่มผู้ใช้บริการที่เริ่มแสดงสัญญาณความผูกพันที่ลดลง หรืออยู่ในกลุ่มที่ต้องเฝ้าระวัง และระดับความเสี่ยงสูง (high risk) มีค่าเฉลี่ยความเสี่ยงการยกเลิกบริการมากกว่า 0.50 ขึ้นไป เป็นกลุ่มผู้ใช้บริการที่มีความน่าจะเป็นที่จะยกเลิกบริการสูงกว่าความน่าจะเป็นที่จะใช้บริการต่อ ซึ่งจำเป็นต้องเข้าสู่มาตรการรักษาลูกค้าอย่างเร่งด่วน

ตารางที่ 5: การจำแนกกลุ่มลูกค้าที่มีความเสี่ยงสูง ปานกลาง และต่ำต่อการยกเลิกบริการ

คุณลักษณะ	GMM			HDBSCAN			Spectral Clustering		
	0	1	2	-1	0	1	0	1	2
Count	9,343	114	1,526	5,650	3,895	1,438	3,358	4,476	3,149
mean Churn_Risk	0.294	0.033	0.445	0.105	0.590	0.377	0.699	0.200	0.061
Risk level	Medium	Low	Medium	Low	High	Medium	High	Low	Low
mean Period	54.182	65.982	7.341	54.675	39.669	42.782	35.800	50.000	57.455
mean Contract	1.396	1.518	1.389	1.516	1.000	2.000	1.000	1.973	1.000
mean WatchTime	5.274	6.158	5.123	5.909	4.575	4.579	4.503	5.475	5.768
mean ContentCount	17.891	0.000	17.194	24.011	10.774	10.960	10.371	19.041	23.290
mean Subscribed	36.799	49.325	39.726	48.097	25.913	25.990	25.379	39.846	46.516
mean PaymentIssues	2.504	2.281	2.515	2.523	2.529	2.354	2.868	2.473	2.157
mean SupportTickets	4.857	5.158	5.799	5.120	4.851	4.858	5.647	5.005	4.269
mean WatchTime/Content	0.691	6157894	27523	131681	0.792	0.798	0.803	163092	4446
mean Subscribed Ratio	0.850	1.025	9.016	3.019	0.942	0.759	1.584	2.564	1.595
mean SupportTickets/Period	0.118	0.113	1.334	0.400	0.175	0.147	0.323	0.360	0.145

ผลการจัดกลุ่มด้วยเทคนิค GMM พบว่า ระบุผู้บริโภคออกเป็น 3 กลุ่มหลัก โดย Cluster 0 มีจำนวนสมาชิกมากที่สุดถึง 9,343 รายการ มีค่า Churn_Risk ในระดับปานกลาง เท่ากับ 0.294 ระยะเวลาการเป็นสมาชิกค่อนข้างนาน ประมาณ 54.182 เดือน และ WatchTime_per_Content เท่ากับ 0.691 และ SupportTickets_per_Period เท่ากับ 0.118 ถือว่าความเสี่ยงอยู่ในระดับต่ำ แสดงถึงกลุ่มผู้บริโภคที่ใช้งานปกติมีความผูกพันกับบริการในระดับปานกลาง ในทางกลับกัน Cluster 1 เป็นกลุ่มเล็กสุด เท่ากับ 114 รายการ แต่มีค่า Churn_Risk ต่ำมากที่ 0.033 ซึ่งระยะเวลาการเป็นสมาชิกยาวที่สุดคือประมาณ 65.982 เดือน แต่ ContentCount เป็น 0 คือแทบไม่เข้าถึงเนื้อหา อีกทั้งมีค่า Subscribed_Ratio สูงและ WatchTime_per_Content สูงมาก เท่ากับ 6157894.146 สะท้อนผู้บริโภคเฉพาะกลุ่มที่สมัครบริการเสริมมาก แต่ความเสี่ยงต่อการยกเลิกต่ำ นอกจากนั้น Cluster 2 มีสมาชิก 1,526 ราย โดยมี Churn_Risk ระดับสูงที่ 0.445 ระยะเวลาการเป็นสมาชิกสั้นประมาณ 7.341 เดือน แต่ค่า WatchTime_per_Content สูงเท่ากับ 27523.571 และ Subscribed_Ratio สูงเท่ากับ 9.016 แสดงถึงผู้บริโภคใหม่แต่มีการใช้งานแบบเฉพาะกิจและเสี่ยงต่อการยกเลิกสูง

สำหรับการจัดกลุ่มด้วยเทคนิค HDBSCAN สามารถระบุกลุ่มผู้บริโภคที่มีความหนาแน่นสูงและมีค่าสุดโต่งได้ โดย Cluster -1 (outlier) มีสมาชิก 5,650 รายการ มีค่า Churn_Risk ต่ำเพียง 0.105 WatchTime_per_Content สูงมากเท่ากับ 131681.981 และ Subscribed_Ratio สูงที่ 3.019 แสดงถึงกลุ่มที่มีลักษณะเฉพาะไม่เป็นกลุ่มชัดเจน แต่ความเสี่ยงต่อการยกเลิกต่ำ ในทางกลับกัน Cluster 0 มี 3,895 รายการ มีค่า Churn_Risk สูงถึง 0.590 WatchTime_per_Content ต่ำที่ 0.792 แสดงถึงผู้บริโภคที่ใช้งานน้อย

และเสี่ยงต่อการยกเลิก นอกจากนั้น Cluster 1 มีสมาชิก 1,438 รายการ แต่มี Churn_Risk ระดับปานกลางที่ 0.377 ระยะเวลาการเป็นสมาชิกคือ 42.782 และ ContentCount ต่ำ แสดงถึงกลุ่มผู้บริโภคที่ใช้งานปานกลางและเสี่ยงต่อการยกเลิกระดับกลาง

ผลการจัดกลุ่มด้วยเทคนิค Spectral clustering แสดงถึงรูปแบบความสัมพันธ์ไม่เป็นเชิงเส้นของผู้บริโภค โดย Cluster 0 นั้นมีสมาชิกจำนวน 3,358 รายการ มีค่า Churn_Risk สูงสุดอยู่ที่ 0.699 ระยะเวลาการเป็นสมาชิกสั้น ประมาณ 35.800 เดือน WatchTime_per_Content ปานกลางที่ 0.803 เป็นกลุ่มเสี่ยงสูงที่ควรให้ความสนใจ โดยที่ Cluster 1 มีจำนวนสมาชิก 4,476 รายการมีค่า Churn_Risk ต่ำ อยู่ที่ประมาณ 0.200 แต่ค่า WatchTime_per_Content สูงมาก ถึง 163092.697 แสดงถึงผู้บริโภคที่ใช้เวลาเข้าถึงเนื้อหาเยอะ แต่เสี่ยงต่อการยกเลิก นอกจากนั้น Cluster 2 มีจำนวนสมาชิก 3,149 รายการ มีค่า Churn_Risk ต่ำสุดเพียง 0.061 ระยะเวลาการเป็นสมาชิกนานประมาณ 57.455 เดือน และค่า WatchTime_per_Content อยู่ในระดับปานกลางเท่ากับ 4446.433 แสดงถึงผู้บริโภคที่ใช้งานต่อเนื่องและมีความเสี่ยงอยู่ในระดับต่ำ

5. ผลการประเมินผลภายใน

การประเมินประสิทธิภาพการจัดกลุ่มลูกค้าทั้ง 3 เทคนิค ได้แก่ GMM, HDBSCAN และ Spectral clustering ด้วยมาตรฐาน 3 ตัว ประกอบด้วย Silhouette score, DBI และ CHI รายละเอียดดังตารางที่ 6 ผลการประเมิน พบว่า การจัดกลุ่มด้วยเทคนิค HDBSCAN เป็นโมเดลที่มีความสมดุลทั้งในด้านความชัดเจนของแต่ละกลุ่ม และความแตกต่างระหว่างกลุ่ม ขณะที่การจัดกลุ่มด้วยเทคนิค GMM มีความสามารถในการแยกกลุ่มในระดับปานกลาง ส่วนการจัดกลุ่มด้วย Spectral clustering แม้จะมีค่า CHI สูง แต่ค่าความชัดเจนและการแยกตัวของจำนวนกลุ่มยังต่ำกว่าโมเดลอื่น

ตารางที่ 6: ผลการประเมินประสิทธิภาพของเทคนิคการจัดกลุ่มลูกค้าทั้ง 3 รูปแบบ

เทคนิคการจัดกลุ่ม	Silhouette score	DBI	CHI
GMM	0.217	1.856	962.676
HDBSCAN	0.217	1.782	1228.063
Spectral clustering	0.116	2.396	1248.606

จากตารางที่ 6 เมื่อพิจารณาในรายละเอียด พบว่า การจัดกลุ่มด้วยเทคนิค GMM และ HDBSCAN มีค่า Silhouette score เท่ากันที่ 0.217 ซึ่งบ่งบอกถึงความสามารถในการแยกกลุ่มในระดับปานกลาง ขณะที่เทคนิค Spectral clustering มีค่าเพียง 0.116 แสดงถึงความแตกต่างระหว่างกลุ่มน้อยกว่า และความชัดเจนในการจัดกลุ่มต่ำกว่า ขณะเดียวกัน เมื่อพิจารณาค่า DBI จากการจัดกลุ่มด้วยเทคนิค HDBSCAN มีค่าต่ำสุดเท่ากับ 1.782 แสดงถึงความสามารถในการจัดกลุ่มที่มีความแตกต่างและชัดเจนที่สุด ในขณะที่การจัดกลุ่มด้วยเทคนิค Spectral clustering มีค่า DBI สูงสุด 2.396 บ่งบอกถึงแต่ละกลุ่มมีความทับซ้อนกันบางส่วน อย่างไรก็ตาม การจัดกลุ่มด้วย Spectral clustering มีค่า CHI สูงสุดเท่ากับ 1248.606 ตามมาด้วยเทคนิค HDBSCAN และ GMM เท่ากับ 1228.063 และ 962.676 ตามลำดับ สะท้อนถึงความสามารถในการแยกตัวอย่างในกลุ่มได้อย่างมีโครงสร้าง แม้จะมีค่า Silhouette score ของการจัดกลุ่มด้วย Spectral clustering จะต่ำ แต่ค่า CHI สูง ซึ่งบ่งบอกว่าความหนาแน่นภายในกลุ่มค่อนข้างดี

ผลการจัดกลุ่มแบบผสมผสานในงานวิจัยนี้ทำให้สามารถเข้าถึงมุมมองที่หลากหลาย ไม่ว่าจะเป็นการตรวจจับผู้ที่มีพฤติกรรมผิดปกติ การระบุกลุ่มผู้ใช้ที่มีความเสี่ยงสูง หรือการตีความความน่าจะเป็นของการเป็นสมาชิกกลุ่มต่าง ๆ ได้อย่างยืดหยุ่น นอกจากนี้ผลลัพธ์ยังสามารถนำไปใช้เชิงธุรกิจผ่านการออกแบบกลยุทธ์รักษาลูกค้าแบบเฉพาะบุคคล (personalized retention strategy) ซึ่งช่วยในการจัดกลุ่มโครงสร้างที่ซ่อนอยู่ในพฤติกรรมการใช้บริการและเพิ่มความแม่นยำในการพยากรณ์การยกเลิกบริการได้อย่างมีนัยสำคัญ

อภิปรายและสรุปผล

ผลการคัดเลือกคุณลักษณะด้วยการคัดเลือกตัวแปรแบบทีละขั้นร่วมกับโมเดลโครงข่ายประสาทเทียม XGBoost และ LightGBM พบว่าชุดตัวแปรที่ให้ประสิทธิภาพสูงสุดประกอบด้วย 7 ตัวแปร ได้แก่ ContentCount, Subscribed, Period, PaymentIssues, WatchTime, SupportTickets และ Contract เมื่อพิจารณาในรายละเอียดของแต่ละตัวแปร พบว่า มีงานวิจัยที่เกี่ยวข้องกับการพยากรณ์การยกเลิกบริการที่มีความสอดคล้องกัน มีรายละเอียดดังนี้

1) ContentCount สะท้อนพฤติกรรมการบริโภคเนื้อหาของผู้ใช้ผ่านจำนวนเนื้อหาที่มีการรับชมในช่วงเวลาหนึ่ง ซึ่งเป็นตัวแทนของระดับความถี่ในการใช้งาน (usage frequency) งานวิจัยด้านการยกเลิกบริการจำนวนมากให้ความสำคัญกับตัวแปรลักษณะนี้ในฐานะปัจจัยเชิงพฤติกรรม เช่น Fujio et al. (2022) ใช้ตัวแปร PercChangeMinutes เพื่อสะท้อนความเข้มข้นในการใช้งาน ขณะที่ Xu et al. (2023) ใช้โมเดล RFM ที่มีองค์ประกอบ Frequency เป็นหนึ่งในตัวบ่งชี้หลักของความสัมพันธ์ระหว่างผู้ใช้และผู้ให้บริการ ซึ่งสะท้อนว่าการลดลงของปริมาณการบริโภคเนื้อหาอาจเป็นสัญญาณเชิงพฤติกรรมที่นำไปสู่ความเสี่ยงของการยกเลิกบริการ

2) Subscribed ที่แสดงถึงพฤติกรรมการสมัครใช้บริการเสริมหรือการมีส่วนร่วมกับแพ็คเกจหรือฟีเจอร์พิเศษเพิ่มเติมจากบริการพื้นฐาน ซึ่งเป็นตัวแปรที่สะท้อนความผูกพันและมูลค่าลูกค้าเชิงโครงสร้าง งานวิจัยของ Xu et al. (2023) แสดงให้เห็นว่าบริการเสริม เช่น OnlineSecurity และ TechSupport มีอิทธิพลต่อการยกเลิกบริการ ขณะที่งานของ Kirgiz et al. (2024) พบว่า Gamified loyalty participation มีผลลดความเสี่ยงการยกเลิกอย่างมีนัยสำคัญ ซึ่งสะท้อนมุมมองร่วมกันว่าการมีส่วนร่วมกับการบริการเสริมช่วยเพิ่มต้นทุนการย้ายบริการ (switching cost) และความภักดี (loyalty) ของลูกค้า ส่งผลให้ความเสี่ยงในการยกเลิกลดลง

3) Period ที่กำหนดระยะเวลาการใช้บริการสะสมของผู้ใช้ ซึ่งเทียบเท่ากับตัวแปร Tenure ที่ปรากฏในงานวิจัยด้านการยกเลิกบริการแทบทั้งหมด เช่น Fujio et al. (2022), Poudel et al. (2024), Wagh et al. (2024), Wei (2024) และ Xu et al. (2023) สอดคล้องในเชิงทฤษฎีว่าระยะเวลาที่ลูกค้าอยู่ในระบบนานขึ้นมักสะท้อนความคุ้นชินหรือคุณค่าทางประสบการณ์ที่เพิ่มขึ้น ซึ่งช่วยลดความน่าจะเป็นของการยกเลิกบริการ ในขณะที่ระยะเวลานั้นเป็นสัญญาณของความเสี่ยงสูงเนื่องจากต้นทุนการย้ายบริการยังต่ำอยู่

4) PaymentIssues บ่งชี้ถึงปัญหาหรือข้อขัดข้องที่เกี่ยวข้องกับการชำระเงินหรือการเรียกเก็บเงิน ซึ่งเป็นปัจจัยสำคัญในด้านประสบการณ์ลูกค้า งานวิจัยของ Wei (2024) พบว่าปัจจัยด้านราคา เช่น MonthlyCharges และ TotalCharges เป็นคุณลักษณะที่มีน้ำหนักเชิงทำนายสูงสุดในการพยากรณ์การยกเลิกบริการ ซึ่งชี้ให้เห็นถึงความสำคัญของภาระด้านค่าใช้จ่าย ขณะที่ PaymentIssues ในบริบทบริการแบบสมัครสมาชิกยังสัมพันธ์กับความไม่พึงพอใจเชิงอารมณ์ (affective dissatisfaction) ที่อาจเร่งการตัดสินใจยกเลิก โดยเฉพาะเมื่อเกิดร่วมกับความไม่โปร่งใสของราคาค่าบริการ

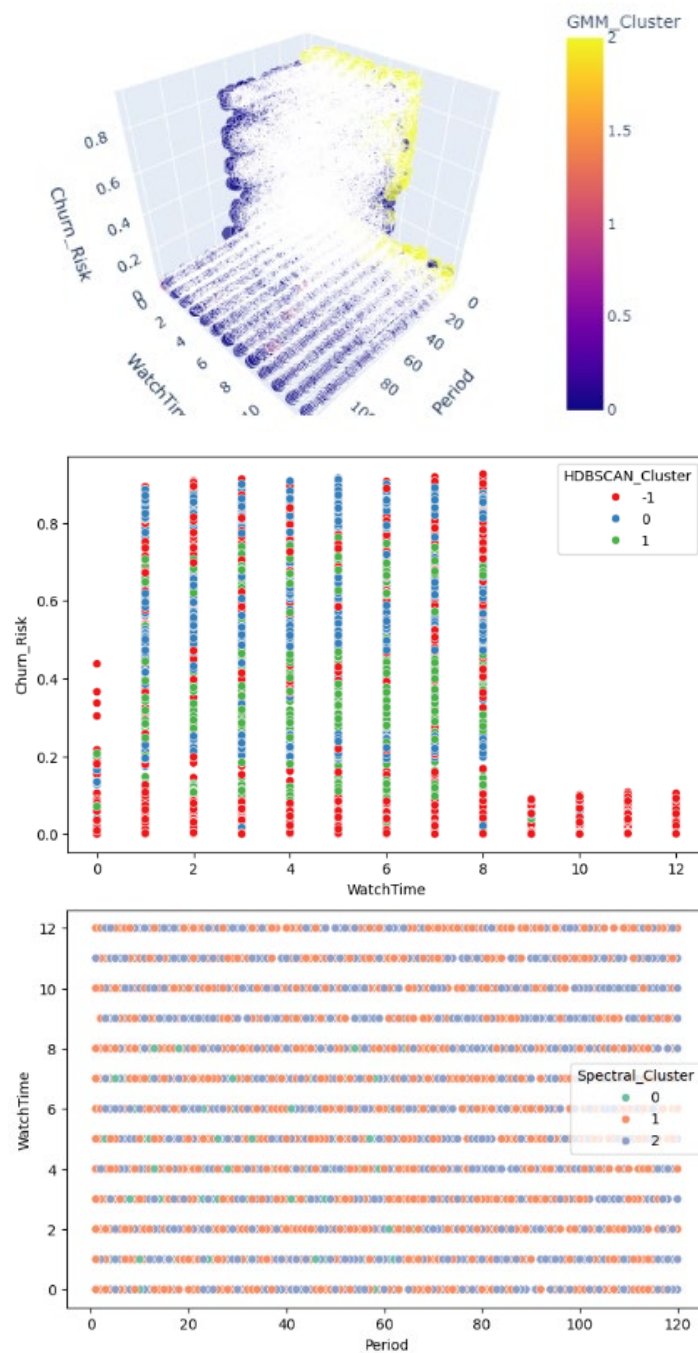
5) WatchTime เป็นตัวแทนของระดับความสัมพันธ์เชิงการใช้งาน (engagement) ซึ่งสะท้อนถึงความสนใจ และความพึงพอใจเชิงเนื้อหา งานวิจัยของ Fujio et al. (2022) ใช้ตัวแปรเกี่ยวกับการเปลี่ยนแปลงช่วงเวลากการใช้งาน (PercChangeMinutes) เพื่อวัดความเข้มข้นของการใช้บริการ และสัมพันธ์กับการยกเลิกบริการอย่างมีนัยสำคัญ นอกจากนี้ในเชิงโมเดลพฤติกรรมผู้ใช้ WatchTime ยังอยู่ในชุดตัวแปรประเภท Behavioral indicator ซึ่งเป็นปัจจัยที่ใช้ในงานแยกส่วนผู้ใช้ (segmentation) เช่น RFM ของ Xu et al. (2023) ที่ส่วน Recency และ Frequency มีความใกล้เคียงกันเชิงความหมาย จึงถือว่ามีความสอดคล้องทั้งเชิงทฤษฎีและเชิงประจักษ์

6) SupportTickets บ่งชี้ถึงจำนวนเหตุการณ์ที่ผู้ใช้ติดต่อเพื่อขอความช่วยเหลือหรือแก้ไขปัญหา งานวิจัยของ Kirgiz et al. (2024) พบว่า Call Center Interactions เป็นหนึ่งในปัจจัยที่มีอิทธิพลต่อการทำนายการยกเลิกบริการมากที่สุด ซึ่งสะท้อนว่าปริมาณการติดต่อที่สูงอาจบ่งชี้ถึงระดับความไม่พึงพอใจที่สะสมขึ้น ส่งผลให้เกิดการยกเลิกบริการในเชิงพฤติกรรม นอกจากนี้ยังสอดคล้องกับทฤษฎีด้าน Service failure – recovery expectation ซึ่งระบุว่าหากการแก้ไขปัญหาที่ไม่เพียงพอสามารถนำไปสู่การยกเลิกบริการได้

7) Contract สะท้อนความผูกพันเชิงโครงสร้างระหว่างผู้ใช้และผู้ให้บริการ ซึ่งเป็นหนึ่งในตัวแปรที่ปรากฏอย่างสม่ำเสมอในงานวิจัยด้านการยกเลิกบริการ เช่น Poudel et al. (2024), Wagh et al. (2024), Wei (2024) และ Xu et al. (2023) ซึ่งล้วนสรุปตรงกันว่า Contract type มีบทบาทสำคัญในการพยากรณ์การยกเลิกบริการ โดยผู้ใช้ที่อยู่ในสัญญารายเดือนมีความเสี่ยงสูงกว่าผู้ใช้ที่ผูกพันในระยะยาว เนื่องจากต้นทุนในการยกเลิกต่ำกว่า และไม่มีข้อจำกัดด้านค่าปรับ (commitment) ตามสัญญา

ผลการศึกษาสนับสนุนข้อค้นพบจากงานวิจัยก่อนหน้านี้ในประเด็นที่ว่าปัจจัยเกี่ยวกับรูปแบบสัญญา ระยะเวลาการใช้บริการ และพฤติกรรมการมีส่วนร่วมกับการ เป็นตัวชี้วัดสำคัญของแนวโน้มการยกเลิกบริการสมาชิก อย่างไรก็ตาม งานวิจัยนี้ขยายขอบเขตขององค์ความรู้ไปสู่บริบทแพลตฟอร์มสตรีมมิง โดยเพิ่มปัจจัยเชิงพฤติกรรมด้านเนื้อหา ได้แก่ ชั่วโมงรับชม จำนวนวิดีโอที่บริโภค และจำนวนช่องที่ติดตาม ซึ่งเป็นปัจจัยที่ไม่ได้ปรากฏอย่างเด่นชัดในงานวิจัยก่อนหน้านี้ ทำให้ผลการจัดกลุ่มมีความจำเพาะและเหมาะสมกับบริบทของผู้บริโภคเนื้อหาแบบสมัครสมาชิกมากขึ้น

นอกจากนั้น เมื่อนำชุดตัวแปรที่ให้ประสิทธิภาพสูงสุดร่วมกับตัวแปรเชิงอนุพันธ์ที่สร้างขึ้นใหม่ ได้แก่ WatchTime_per_Content, Subscribed_Ratio, SupportTickets_per_Period และ Churn_Risk ผลการจัดกลุ่มลูกค้าด้วยเทคนิค GMM, HDBSCAN และ Spectral clustering ดังรูปที่ 3



รูปที่ 3: ผลการจัดกลุ่มลูกค้าด้วย GMM, HDBSCAN และ Spectral clustering

จากรูปที่ 3 เมื่อพิจารณาในรายละเอียดด้วยตัวชี้วัดคุณภาพการจัดกลุ่ม พบว่า ผลลัพธ์ของแต่ละเทคนิคสะท้อนความแตกต่างทั้งจำนวนสมาชิกเฉลี่ยในแต่ละกลุ่ม ความเสี่ยงในการยกเลิกบริการ (Churn_Risk) และพฤติกรรมการใช้งาน มีรายละเอียดดังนี้

1. การจัดกลุ่มด้วยเทคนิค GMM แสดงให้เห็นถึงความสามารถในการจัดกลุ่มลูกค้าตามความน่าจะเป็นของการยกเลิกบริการได้อย่างมีประสิทธิภาพ โดยผลการวิเคราะห์เชิงพฤติกรรมในมิติของตัวแปร Period, WatchTime และ Churn_Risk พบว่า ลูกค้าที่มีระยะเวลาการเป็นสมาชิกสั้น และระยะเวลาการรับชมต่ำมักอยู่ในกลุ่มที่มีความเสี่ยงสูงต่อการยกเลิกบริการ ขณะที่กลุ่มที่มีการเป็นสมาชิกยาวนานและมีพฤติกรรมรับชมในระดับปานกลางมีแนวโน้มอยู่ในกลุ่มที่มีความเสี่ยงต่ำ โดยแบ่งออกเป็น 3 กลุ่ม ได้แก่ Cluster 0 มีขนาดใหญ่ที่สุดเป็นจำนวน 9,343 รายการ มีค่า Churn_Risk เฉลี่ย 0.294 อยู่ในระดับปานกลาง และมีค่า Contract เฉลี่ย 1.396 สะท้อนถึงรูปแบบสัญญาระยะยาวและความต่อเนื่อง ผลการประเมินคุณภาพการจัดกลุ่ม พบว่า ค่า Silhouette score เท่ากับ 0.217 ค่า DBI เท่ากับ 1.856 และค่า CHI เท่ากับ 962.676 ซึ่งแสดงถึงความสามารถในการจัดกลุ่มลูกค้าในระดับปานกลาง ทั้งนี้ ผลการจัดกลุ่มยังสะท้อนความแตกต่างระหว่างกลุ่มที่มีความเสี่ยงต่ำที่สุด (Cluster 1) และกลุ่มที่มีความเสี่ยงสูงที่สุด (Cluster 2) ได้อย่างชัดเจน แสดงให้เห็นว่าเทคนิค GMM สามารถจับลักษณะเชิงพฤติกรรมได้อย่างมีนัยสำคัญ แนวทางเชิงกลยุทธ์จากการจัดกลุ่มด้วยเทคนิค GMM คือการจัดลำดับความสำคัญของลูกค้าโดยใช้ระดับความเสี่ยงเป็นตัวชี้วัด เช่น การออกแบบมาตรการเชิงรุกเพื่อรักษาลูกค้าในกลุ่มที่ Churn_Risk สูงด้วยการนำเสนอสิทธิพิเศษเฉพาะบุคคล หรือการสื่อสารเชิงรุกเพื่อกระตุ้นการใช้งาน นอกจากนี้ ความแตกต่างเชิงพฤติกรรมระหว่างกลุ่มยังสามารถใช้พยากรณ์พฤติกรรมที่ลดการมีส่วนร่วมของลูกค้าก่อนการยกเลิกบริการจริง

2. การจัดกลุ่มด้วยเทคนิค HDBSCAN แสดงประสิทธิภาพในการจัดกลุ่มข้อมูลที่มีลักษณะไม่สมมาตรและสามารถตรวจจับจุดข้อมูลที่มีพฤติกรรมผิดปกติได้อย่างแม่นยำ โดยแบ่งออกเป็น 3 กลุ่มหลัก ประกอบด้วย กลุ่ม Cluster -1 มีสมาชิกจำนวน 5,650 คน โดย Cluster 0 มีความเสี่ยงต่อการยกเลิกบริการสูงที่สุด ซึ่งมีค่า Churn_Risk เฉลี่ย 0.590 ขณะที่กลุ่มอื่นมีความเสี่ยงต่ำถึงปานกลาง การวิเคราะห์เชิงลึกของตัวแปรอนุพันธ์ เช่น WatchTime_per_Content, Subscribed_Ratio และ SupportTickets_per_Period พบว่า มีความแตกต่างเชิงพฤติกรรมอย่างมีนัยสำคัญระหว่างแต่ละกลุ่ม โดยเฉพาะในกลุ่มที่มีลักษณะการใช้งานที่ไม่ปกติและมีแนวโน้มยกเลิกบริการสูง ผลการประเมินคุณภาพของการจัดกลุ่ม พบว่า ค่า Silhouette score เท่ากับ 0.217 ค่า DBI เท่ากับ 1.782 และค่า CHI เท่ากับ 1,228.063 ซึ่งสะท้อนถึงประสิทธิภาพในการตรวจจับกลุ่มข้อมูลที่แตกต่างจากกลุ่มหลักอย่างชัดเจน ซึ่งเหมาะสำหรับใช้เป็นเครื่องมือระบุลูกค้ากลุ่มที่มีลักษณะพิเศษที่อาจมีพฤติกรรมเบี่ยงเบนจากกลุ่มหลัก เช่น ลูกค้าที่กำลังลดระยะเวลาการใช้งานอย่างรวดเร็ว หรือมีปัญหาในการชำระเงินบ่อยครั้ง จึงสามารถนำผลลัพธ์ไปใช้ในการจัดลำดับความสำคัญของการติดตามลูกค้าแบบเฉพาะราย เช่น การมอบสิทธิทดลองใช้งานใหม่ หรือการสื่อสารเชิงรุกผ่านช่องทางแบบเฉพาะเจาะจง นอกจากนี้ เมื่อพิจารณาข้อมูลจากตัวแปรเชิงอนุพันธ์ เช่น WatchTime_per_Content, Subscribed_Ratio และ SupportTickets_per_Period ยังสามารถนำไปใช้สร้างตัวชี้วัดพฤติกรรมความเสี่ยงเชิงรุก เพื่อตรวจจับความเปลี่ยนแปลงการใช้บริการของลูกค้าแบบเรียลไทม์ได้อีกด้วย

3. การจัดกลุ่มด้วยเทคนิค Spectral clustering สามารถตรวจจับความสัมพันธ์เชิงซับซ้อนและไม่เชิงเส้นของข้อมูลลูกค้าได้อย่างละเอียด โดยแบ่งลูกค้าออกเป็น 3 กลุ่มหลัก ซึ่งกลุ่มที่มีค่า Period และ WatchTime สูงมักเป็นลูกค้าที่มีความจงรักภักดี และคงอยู่กับธุรกิจในระยะยาว ขณะที่กลุ่มที่มีค่าทั้งสองต่ำนั้นมีความเสี่ยงต่อการยกเลิกบริการสูง โดย Cluster 0 เป็นกลุ่มที่มีค่า Churn_Risk เฉลี่ยสูงที่สุดเท่ากับ 0.699 และ Cluster 2 เป็นกลุ่มที่มีความเสี่ยงต่ำที่สุดเท่ากับ 0.061 จากการวิเคราะห์ตัวแปรเชิงอนุพันธ์ เช่น WatchTime_per_Content และ Subscribed_Ratio พบว่า กลุ่มที่มีความเสี่ยงต่ำมีพฤติกรรมการใช้งานที่สม่ำเสมอและมีอัตราการสมัครรับเนื้อหาสูง ในขณะที่กลุ่มที่มีความเสี่ยงสูงแสดงรูปแบบการใช้งานไม่ต่อเนื่อง อย่างไรก็ตาม ผลการประเมินคุณภาพการจัดกลุ่มด้วยค่า Silhouette score เท่ากับ 0.116 ค่า DBI เท่ากับ 2.396 และค่า CHI เท่ากับ 1,248.606 ซึ่งบ่งชี้ว่า แม้การจัดกลุ่มด้วยเทคนิค Spectral clustering จะมีความหนาแน่นภายในกลุ่มที่ดีจากค่า CHI ที่ค่อนข้างสูง แต่การแยกกลุ่มนั้นยังไม่ชัดเจนและมีความทับซ้อนกันบางส่วนระหว่างกลุ่ม ซึ่งสอดคล้องกับลักษณะของข้อมูลของผู้ใช้บริการจริงที่มีความซับซ้อนเชิงโครงสร้างและไม่เป็นเชิงเส้น แนวทางเชิงกลยุทธ์ที่ได้คือการออกแบบแคมเปญการตลาดแบบเฉพาะกลุ่ม เช่น การแนะนำเนื้อหาหรือแพ็คเกจที่สอดคล้องกับรูปแบบการใช้งานจริงของลูกค้าแต่ละกลุ่ม รวมถึงการให้กลุ่มลูกค้าที่มีความเสี่ยงต่ำเป็นกลุ่มทดลองผลิตภัณฑ์หรือบริการใหม่ (pilot users) เพื่อทดสอบการตอบสนองของตลาดก่อนขยายผลในวงกว้าง

จะเห็นได้ว่าในบริบทของแพลตฟอร์มสตรีมมิงนั้น พฤติกรรมการยกเลิกบริการมีลักษณะเฉพาะตัวและมีความซับซ้อน เนื่องจากพฤติกรรมการบริโภคเนื้อหาเป็นแบบรับชมเชิงรับและไม่ต่อเนื่อง โดยเฉพาะบนแพลตฟอร์มอย่างยูทูบที่ผู้ชมมักมีเนื้อหาแบบสลับแตกต่างจากบริการสตรีมมิงที่มีรูปแบบการรับชมเชิงตั้งใจและเป็นลำดับ ส่งผลให้การจัดกลุ่มผู้บริโภคด้วยวิธีการดั้งเดิมทำได้ยาก ขณะเดียวกัน ต้นทุนการเปลี่ยนแปลงบริการที่ค่อนข้างต่ำในยุคดิจิทัลทำให้ผู้ใช้สามารถยกเลิกสมาชิกหรือเปลี่ยนไปใช้บริการคู่แข่ง หรือแม้แต่ช่องทางที่ไม่ถูกลิขสิทธิ์ได้ตลอดเวลา หากทราบว่าบริการไม่มีความคุ้มค่า นอกจากนี้ ลักษณะข้อมูลพฤติกรรมผู้ชมยังมีมิติสูง อีกทั้งความหนาแน่นแตกต่างกัน และมีความสัมพันธ์แบบไม่เป็นเชิงเส้น ทำให้การวิเคราะห์ด้วยข้อมูลประชากรศาสตร์หรือเทคนิคการจัดกลุ่มเพียงวิธีเดียวไม่สามารถสะท้อนความซับซ้อนของพฤติกรรมได้อย่างเพียงพอ ดังนั้นแนวทางการจัดกลุ่มแบบผสมผสานรวมกับการคัดเลือกลักษณะสำคัญจึงมีความเหมาะสมอย่างยิ่ง โดยการใช้อัลกอริทึมการเรียนรู้แบบรวมช่วยระบุตัวแปรที่มีอิทธิพลต่อการยกเลิกบริการได้อย่างแม่นยำเมื่อทำงานกับข้อมูลที่ไม่สมดุล พร้อมทั้งชี้ให้เห็นสัญญาณเชิงพฤติกรรมสำคัญ เช่น จำนวนเนื้อหาที่รับชมและระยะเวลาใช้งาน ในขณะที่การจัดกลุ่มแบบผสมผสานยังสามารถช่วยจัดการกับข้อมูลที่มีโครงสร้างซับซ้อน การตรวจจับความซ้อนทับของกลุ่ม การระบุพฤติกรรมผิดปกติ และเปิดเผยความสัมพันธ์ที่ไม่เป็นเชิงเส้นของกลุ่มเสี่ยงได้อย่างมีประสิทธิภาพ อีกทั้งการสร้างตัวแปรเชิงอนุพันธ์ยังช่วยเพิ่มมิติการวิเคราะห์ที่สะท้อนความเสี่ยงเชิงลึก ทำให้สามารถทำความเข้าใจพฤติกรรมผู้ใช้และออกแบบกลยุทธ์รักษาลูกค้าได้อย่างแม่นยำและสอดคล้องกับพลวัตของอุตสาหกรรมสตรีมมิงและสื่อดิจิทัลในปัจจุบัน

ผลการวิจัยนี้ให้ข้อค้นพบที่มีนัยสำคัญต่อการบริหารจัดการแพลตฟอร์มสตรีมมิง โดยเฉพาะธุรกิจวิดีโอออนไลน์ที่ใช้โมเดลหารายได้จากการบริการสมาชิกแบบมีค่าใช้จ่าย ซึ่งนำแบบจำลองการคัดเลือกลักษณะสำคัญและการจัดกลุ่มแบบผสมผสานไปใช้เพื่อระบุผู้ใช้ที่มีความเสี่ยงต่อการยกเลิกบริการได้ล่วงหน้า ช่วยให้การออกแบบกลยุทธ์รักษาลูกค้าเชิงรุก และการสื่อสารการตลาดแบบเฉพาะบุคคลได้อย่างแม่นยำมากขึ้น ทั้งในการพัฒนาระบบแนะนำเนื้อหา การออกแบบแพ็คเกจราคา และการยกระดับประสบการณ์ผู้ใช้เพื่อลดอัตราการเลิกใช้บริการ ตลอดจนการตรวจจับกลุ่มผู้ใช้ที่มีพฤติกรรมผิดปกติหรือมีความเสี่ยงสูง ยังช่วยให้ผู้ให้บริการสามารถจัดสรรทรัพยากรด้านการบริการลูกค้าได้อย่างมีประสิทธิภาพและทันทั่วถึง นอกจากนี้ ในเชิงนโยบายผลการวิจัยชี้ให้เห็นถึงความจำเป็นของการกำกับดูแลแพลตฟอร์มดิจิทัลให้เกิดความเป็นธรรมและความโปร่งใสในการใช้ข้อมูลพฤติกรรมผู้บริโภค หน่วยงานที่มีหน้าที่กำกับดูแลสามารถนำแนวคิดการวิเคราะห์ข้อมูลเชิงลึกไปสนับสนุนการคุ้มครองผู้บริโภค เช่น การกำหนดแนวทางการใช้ข้อมูลส่วนบุคคลอย่างมีจริยธรรม การส่งเสริมความโปร่งใสของอัลกอริทึมแนะนำเนื้อหา และการสร้างมาตรการป้องกันการผูกขาดเชิงข้อมูล ในขณะเดียวกัน ภาครัฐสามารถใช้ข้อมูลเชิงพฤติกรรมจากแพลตฟอร์มสตรีมมิงเพื่อส่งเสริมเศรษฐกิจดิจิทัลและอุตสาหกรรมคอนเทนต์อย่างยั่งยืน รวมถึงสนับสนุนการแข่งขันที่เป็นธรรม และลดการเข้าถึงเนื้อหาผิดกฎหมาย ซึ่งจะช่วยสร้างระบบนิเวศดิจิทัลที่สมดุลระหว่างนวัตกรรมทางธุรกิจและการคุ้มครองสิทธิของผู้บริโภคของประเทศไทยต่อไป

ข้อเสนอแนะการวิจัยครั้งต่อไป

แม้ว่าผลลัพธ์จากงานวิจัยนี้จะสามารถจำแนกกลุ่มผู้บริโภคและระบุปัจจัยสำคัญที่สัมพันธ์กับการยกเลิกบริการได้อย่างมีประสิทธิภาพ แต่การศึกษาในอนาคตควรมุ่งขยายขอบเขตข้อมูลให้ครอบคลุมบริบททางภูมิศาสตร์และประชากรที่หลากหลายเพื่อเพิ่มความสามารถในการอธิบายทั่วไปของแบบจำลอง ควบคู่กับการบูรณาการข้อมูลเชิงลำดับเวลาเพื่อวิเคราะห์พลวัตพฤติกรรมผู้ชมตลอดวงจรชีวิตสมาชิก ซึ่งจะสามารถตรวจจับสัญญาณเตือนล่วงหน้าได้อย่างแม่นยำมากยิ่งขึ้น นอกจากนี้ การผสานตัวแปรเชิงจิตวิทยาและประสบการณ์ผู้ใช้ เช่น ความพึงพอใจที่รับรู้และคุณค่าที่รับรู้ จะช่วยเสริมมิติการอธิบายการตัดสินใจยกเลิกบริการให้ครอบคลุมมากขึ้น ขณะเดียวกัน การประยุกต์ใช้แนวคิด XAI (explainable AI) ที่เป็นแนวคิดปัญญาประดิษฐ์แบบอธิบายได้ เพื่อเพิ่มความโปร่งใสของแบบจำลอง และการออกแบบการทดลองเชิงปฏิบัติ เช่น A/B testing เพื่อประเมินประสิทธิผลของกลยุทธ์รักษาลูกค้าในสถานการณ์จริง จะช่วยเชื่อมโยงผลลัพธ์เชิงพยากรณ์สู่การใช้งานเชิงกลยุทธ์ได้อย่างเป็นรูปธรรม ในระยะยาว การพัฒนาระบบวิเคราะห์และแนะนำแบบเรียลไทม์ที่สามารถปรับตัวตามพฤติกรรมผู้ใช้จะเป็นทิศทางสำคัญในการยกระดับการบริหารความสัมพันธ์ลูกค้าและเพิ่มความยั่งยืนของธุรกิจบริการดิจิทัลต่อไปในอนาคต

REFERENCES

- Aggarwal, C. C., Han, J., Wang, J., & Yu, P. S. (2003). A framework for clustering evolving data streams. *Proceedings of the 29th International Conference on Very Large Data Bases*, 29, 81–92.
- Agrawal, R., Gehrke, J., Gunopulos, D., & Raghavan, P. (1998). Automatic subspace clustering of high dimensional data for data mining applications. *ACM SIGMOD Record*, 27(2), 94–105. <https://doi.org/10.1145/276305.276314>
- Ankerst, M., Breunig, M. M., Kriegel, H.-P., & Sander, J. (1999). OPTICS: Ordering points to identify the clustering structure. *ACM SIGMOD Record*, 28(2), 49–60.
- Asia Video Industry Association. (2025). *The Asia video industry report 2026*. <https://avia.org/wp-content/uploads/2025/12/Asia-Video-Industry-Report-2026.pdf>
- Bansal, K. (2023). Customer experience: Creating value through transforming customer journeys. *International Journal of Advances in Engineering and Management (IJAEM)*, 5(9), 715–725.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, Article P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), *Lecture notes in computer science: Vol. 7819. Advances in knowledge discovery and data mining, Part II* (pp. 160–172). Springer.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery.
- Corcoran, A. (2023). *Integrated digital marketing in practice*. Cambridge University Press.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1–38.
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD)* (pp. 226–231). AAAI Press.
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.
- Fujo, S. W., Subramanian, S., & Khder, M. A. (2022). Customer churn prediction in telecommunication industry using deep learning. *Information Sciences Letters*, 11(1), 185–198.
- Godinho de Matos, M., & Ferreira, P. (2020). The effect of binge-watching on the subscription of video on demand: Results from randomized experiments. *Information Systems Research*, 31(4), 1337–1360.
- Gomez-Urbe, C. A., & Hunt, N. (2016). The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems*, 6(4), Article 13. <https://doi.org/10.1145/2843948>
- Grönroos, C. (1989). Defining marketing: A market-oriented approach. *European Journal of Marketing*, 23(1), 52–60.
- Gupta, G., & Ghosh, J. (2006). Bregman bubble clustering: A robust, scalable framework for locating multiple, dense regions in data. *Proceedings of the Sixth IEEE International Conference on Data Mining* (pp. 232–243). IEEE. <https://doi.org/10.1109/ICDM.2006.32>
- Gupta, G., & Singharia, K. (2021). Consumption of OTT media streaming in COVID-19 lockdown: Insights from PLS analysis. *Vision*, 25(1), 36–46.
- Hallowell, R. (1996). The relationships of customer satisfaction, customer loyalty, and profitability: An empirical study. *International Journal of Service Industry Management*, 7(4), 27–42.
- Helgesen, Ø. (2006). Are loyal customers profitable? Customer satisfaction, customer (action) loyalty and customer profitability at the individual level. *Journal of Marketing Management*, 22(3–4), 245–266.
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons. <https://doi.org/10.1002/9780470316801>

- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)* (pp. 3149–3157). Curran Associates.
- Khan, M. L. (2017). Social media engagement: What motivates user participation and consumption on YouTube? *Computers in Human Behavior*, 66, 236–247. <https://doi.org/10.1016/j.chb.2016.09.024>
- Kirgiz, O. B., Kiygi-Calli, M., Cagliyor, S., & El Oraiby, M. (2024). Assessing the effectiveness of OTT services, branded apps, and gamified loyalty giveaways on mobile customer churn in the telecom industry: A machine-learning approach. *Telecommunications Policy*, 48(8), Article 102816. <https://doi.org/10.1016/j.telpol.2024.102816>
- Kotler, P., & Armstrong, G. (2018). *Principles of marketing* (17th global ed.). Pearson.
- Kumar, V., & Shah, D. (2009). Expanding the role of marketing: From customer equity to market capitalization. *Journal of Marketing*, 73(6), 119–136.
- Kumar, V., Bhagwat, Y., & Zhang, X. (2015). Regaining “lost” customers: The predictive power of first-lifetime behavior, the reason for defection, and the nature of the win-back offer. *Journal of Marketing*, 79(4), 34–55.
- Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, 80(6), 69–96.
- Lin, C., & Bowman, D. (2022). The impact of introducing a customer loyalty program on category sales and profitability. *Journal of Retailing and Consumer Services*, 64, Article 102769. <https://doi.org/10.1016/j.jretconser.2021.102769>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 4768–4777). Curran Associates.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam, & J. Neyman (Eds.), *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability: Vol. 1.* (pp. 281–297). University of California Press.
- Magnite. (2024, March 22). *Streaming TV’s new era: Embracing ad-supported streaming across screens in Thailand*. <https://magnite.showpad.com/share/PDZT19bbOii84SYUKdNbK>
- McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical den sity based clustering. *Journal of Open Source Software*, 2(11), Article 205. <https://doi.org/10.21105/joss.00205>
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. John Wiley & Sons. <https://doi.org/10.1002/0471721182>
- Ng, A. Y., Jordan, M. I., & Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. In T. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Proceedings of the 15th International Conference on Neural Information Processing Systems: Natural and Synthetic* (pp. 849–856). MIT Press.
- Oliver, R. L. (1999). Whence consumer loyalty? *Journal of Marketing*, 63(4_suppl1), 33–44.
- Palomba, A. (2022). Building OTT brand loyalty and brand equity: Impact of original series on OTT services. *Telematics and Informatics*, 66, Article 101733. <https://doi.org/10.1016/j.tele.2021.101733>
- Poudel, S. S., Pokharel, S., & Timilsina, M. (2024). Explaining customer churn prediction in telecom industry using tabular machine learning models. *Machine Learning with Applications*, 17, Article 100567. <https://doi.org/10.1016/j.mlwa.2024.100567>
- Qin, Z., Shuai, Q., Wang, G., Zhang, P., Cao, M., & Chen, M. (2022). *E-commerce: Concepts, principles, and application*. Springer Nature Singapore.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939778>
- Rokach, L. (2019). *Ensemble learning: Pattern classification using ensemble methods* (2nd ed.). World Scientific.
- SBI. (2025, January 6). *Tearing down YouTube Premium’s pricing strategy*. <https://sbigrowth.com/insights/youtube-premium-pricing>
- Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 888–905.

- Siahaan, V., & Sianipar, R. H. (2023). *Customer segmentation, clustering, and prediction with Python* (2nd ed). Balige Publishing.
- Statista Research Department. (2024, September 27). *Reasons for unsubscribing from subscription video-on-demand (SVoD) services in the next three months in Thailand as of June 2024, by age*. Statista. <https://www.statista.com/statistics/1261149/thailand-reasons-for-unsubscribing-from-subscription-video-on-demand-services-by-age/>
- The Bureau of Registration Administration. (2025). *Monthly population statistics of Thailand, December 2024* (in Thai). Department of Provincial Administration, Ministry of Interior. Retrieved July 21, 2026, from <https://stat.bora.dopa.go.th/stat/statnew/statMONTH/statmonth/#/mainpage>
- Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, 9, Article 5233.
- Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17, 395–416.
- Wagh, S. K., Andhale, A. A., Wagh, K. S., Pansare, J. R., Ambadekar, S. P., & Gawande, S. H. (2024). Customer churn prediction in telecom sector using machine learning techniques. *Results in Control and Optimization*, 14, Article 100342. <https://doi.org/10.1016/j.rico.2023.100342>
- Wedel, M., & Kamakura, W. A. (2000). *Market segmentation: Conceptual and methodological foundations* (2nd ed.). Kluwer Academic.
- Wei, Y. (2024). Telco customer churn prediction. *Highlights in Science, Engineering and Technology*, 92, 218–226.
- Wetherby, S. (2026, July 28) *How Many People Use YouTube in 2026? [Users Data]*. Limelight Digital. <https://www.limelightdigital.co.uk/youtube-statistics/>
- Xie, J., Girshick, R., & Farhadi, A. (2016). Unsupervised deep embedding for clustering analysis. In M. F. Balcan, & K. Q. Weinberger (Eds.), *Proceedings of the 33rd International Conference on Machine Learning (ICML): Vol. 48* (pp. 478–487). JMLR.org.
- Xu, T., Ma, Y., Ao, C., Qu, M., & Meng, X. (2023). A novel telecom customer churn analysis system based on RFM model and feature importance ranking. *Interdisciplinary Journal of Information, Knowledge, and Management*, 18, 719–737.
- YouTube. (n.d.). *YouTube Premium*. Retrieved July 21, 2026, from <https://www.youtube.com/premium>