บทความวิจัย

การตรวจจับกลุ่มเสี่ยงแฝงต่อความรุนแรงในครอบครัวแบบสองขั้นตอนด้วยแบบจำลองไฮบริดแบบไม่มีผู้สอน

วุฒิพงษ์ เขื่อนดิน สิทธิวรรต รอบรู้ และ ชนิดา แก้วเพชร*
สาขาวิชาระบบสารสนเทศและคอมพิวเตอร์ธุรกิจ คณะบริหารธุรกิจและเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 08 8674 2113 อีเมล: Chanida.k@rmutsb.ac.th DOI: 10.14416/j.kmutnb.2026.06.007
รับเมื่อ 15 ตุลาคม 2568 แก้ไขเมื่อ 2 มีนาคม 2569 ตอบรับเมื่อ 25 มีนาคม 2569 เผยแพร่ออนไลน์ 19 มิถุนายน 2569

บทคัดย่อ

การตรวจจับกลุ่มเสี่ยงแฝงของความรุนแรงในครอบครัวเป็นงานที่ท้าทาย เนื่องจากข้อมูลที่มีอยู่มักไม่ได้ระบุป้ายกำกับ ซึ่งงานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอแบบจำลองไฮบริดสองขั้นตอนแบบไม่มีผู้สอน ออกแบบมาเพื่อตรวจจับและระบุกลุ่มเสี่ยงสูงแฝง โดยบูรณาการจุดแข็งของแบบจำลองส่วนผสมเกาส์เซียน (Gaussian Mixture Model; GMM) ออโตเอ็นโค้ดเดอร์ (Autoencoder; AE) และตัวประกอบค่าผิดปกติเฉพาะที่ (Local Outlier Factor; LOF) ในขั้นตอนแรกแบบจำลองจะแยกกลุ่มเสี่ยงสูงอย่างชัดเจน (59 รายการ จากทั้งหมด 1,161 รายการ) ออกจากชุดข้อมูลทั้งหมด ขั้นตอนที่สองใช้รายการข้อมูลที่เหลือ 1,102 รายการ เพื่อระบุกลุ่มเสี่ยงสูงแฝงผ่านกลยุทธ์การผสานผลลัพธ์ด้วยตรรกะและ (AND) ตรรกะหรือ (OR) และการถ่วงน้ำหนัก (Weighted Fusion) คุณภาพเชิงโครงสร้างของการจัดกลุ่มถูกประเมินโดยใช้ดัชนีซิลูเอท (Silhouette Index) ดัชนีคาลินสกี้ฮาราบาสซ์ (Calinski–Harabasz Index; CH) และดัชนีเดวีส์-โบลดิน (Davies–Bouldin Index; DBI) พร้อมทั้งตรวจสอบความแตกต่างทางสถิติด้วยการทดสอบครัสคัล-วัลลิส (Kruskal–Wallis test) และการทดสอบไคสแควร์ (Chi square test) สำหรับตัวแปรเชิงหมวดหมู่ ผลการทดลองแสดงให้เห็นว่าแบบจำลอง GAAL ที่เสนอ (GMM + AE + LOF พร้อมตรรกะและ : AND) ให้ค่าคะแนนดัชนีซิลูเอทสูงสุดและค่าดัชนีเดวีส์-โบลดินต่ำสุด (2.8478) ซึ่งบ่งชี้ถึงความชัดเจนของคลัสเตอร์และเสถียรภาพเชิงโครงสร้างที่เหนือกว่าการแสดงภาพด้วย เทคนิคการลดมิติข้อมูลแบบที-เอสเอ็นอี (t-distributed Stochastic Neighbor Embedding; t-SNE) ยืนยันการแยกตัวของกลุ่มอย่างชัดเจนและสอดคล้องกับตัวชี้วัดเชิงปริมาณ นอกจากนี้พบว่าตัวแปรอายุ (Age) มีความแตกต่างอย่างมีนัยสำคัญระหว่างคลัสเตอร์ตามการทดสอบครัสคัล-วัลลิส ผลการวิจัยสะท้อนถึงคุณภาพเชิงโครงสร้างและความสามารถในการตรวจจับของวิธีการแบบผสมผสานในการค้นหารูปแบบความเสี่ยงที่ซ่อนอยู่จากข้อมูลทางสังคมที่ไม่ได้ติดป้ายกำกับ

คำสำคัญ: การตรวจหาความผิดปกติ ความรุนแรงในครอบครัว การเรียนรู้ของเครื่อง แบบจำลองไฮบริด กลุ่มเสี่ยงแฝง

การอ้างอิงบทความ: วุฒิพงษ์ เขื่อนดิน, สิทธิวรรต รอบรู้ และ ชนิดา แก้วเพชร, "การตรวจจับกลุ่มเสี่ยงแฝงต่อความรุนแรงในครอบครัวแบบสองขั้นตอนด้วยแบบจำลองไฮบริดแบบไม่มีผู้สอน," *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 36, ฉบับที่ 3, หน้า 1–16, ก.ค.–ก.ย. 2569, เลขที่บทความ 263-088233, doi: 10.14416/j.kmutnb.2026.06.007.

Research Article

Two-Stage Unsupervised Hybrid Model for Silent Risk Detection in Domestic Violence

Wutthiphong Khuandin, Sittiwat Robroo and Chanida Kaewphet*
Major of Information System and Business Computer, Faculty of Business Administration and Information Technology, Rajamangala University of Technology Suvarnabhumi, Suphanburi, Thailand

* Corresponding Author, Tel. 08 8674 2113, E-mail: Chanida.k@rmutsb.ac.th
DOI: 10.14416/j.kmutnb.2026.06.007
Received 15 October2025; Revised 2 March 2026; Accepted 25 March 2026; Published online: 19 June 2026

Abstract

Detecting latent risk groups for domestic violence is a challenging task because the available data are typically unlabeled. This study presents a two-stage hybrid unsupervised model designed to detect and identify latent high-risk groups by integrating the strengths of the Gaussian Mixture Model (GMM), Autoencoder (AE), and Local Outlier Factor (LOF). In the first stage, the model isolates clearly high-risk cases (59 out of 1,161 records) from the entire dataset. The second stage utilizes the remaining 1,102 records to identify latent high-risk groups through logical fusion strategies, including AND, OR, and weighted fusion. The clustering quality and structural validity are evaluated using the Silhouette Index, Calinski–Harabasz Index (CH), and Davies–Bouldin Index (DBI), along with statistical validation using the Kruskal–Wallis test and the chi-square test for categorical variables. The results indicate that the proposed GAAL model (GMM + AE + LOF with AND logic) achieves the highest Silhouette score and the lowest DBI value (2.8478), indicating superior cluster clarity and structural stability. The t-SNE visualization further confirms the distinct separation between the identified groups, consistent with the quantitative metrics. In addition, the variable Age is found to differ significantly among clusters based on the Kruskal–Wallis test. These findings demonstrate the structural quality and detection capability of the proposed hybrid approach in uncovering hidden risk patterns from unlabeled social data.

Keywords: Anomaly Detection, Domestic Violence, Machine Learning, Hybrid Model, Latent Risk

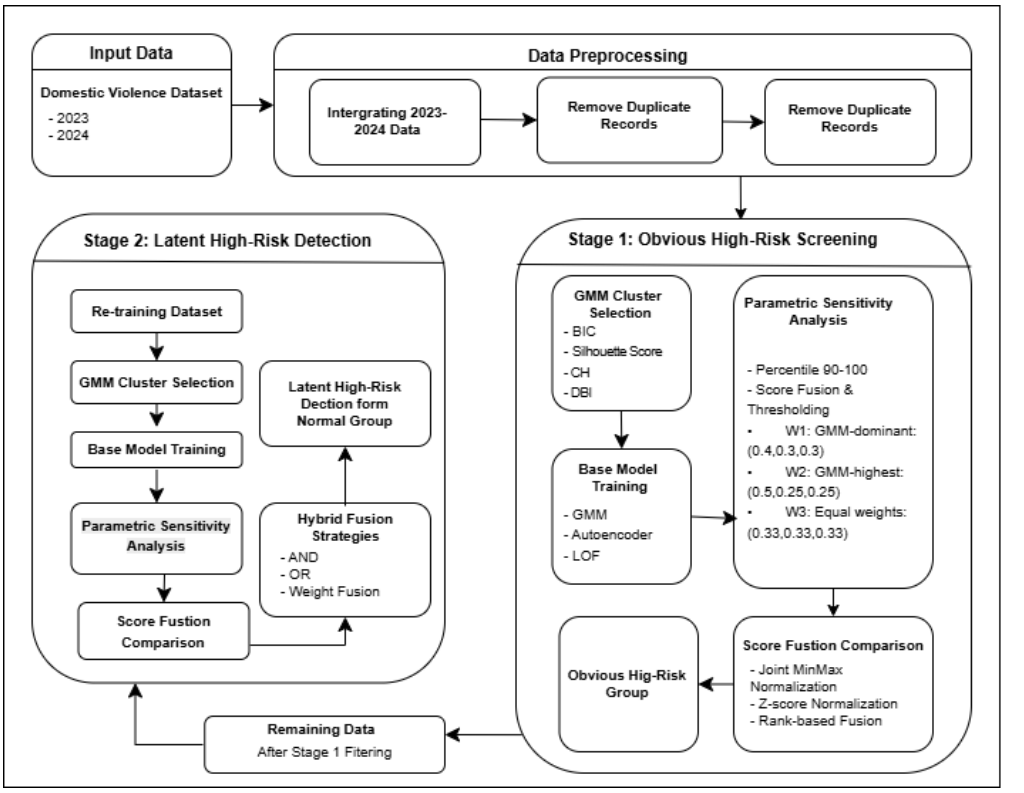
Please cite this article as: W. Khuandin, S. Robroo, and C. Kaewphet, "Two-stage unsupervised hybrid model for silent risk detection in domestic violence," *The Journal of KMUTNB*, vol. 36, no. 3, pp. 1–16, Jul.–Sep. 2026 (in Thai), Art. no. 263-088233, doi: 10.14416/j.kmutnb.2026.06.007.

1. บทนำ

ความรุนแรงในครอบครัวเป็นหนึ่งในปัญหาสังคมที่ส่งผลกระทบอย่างกว้างขวางทั้งต่อความมั่นคงของสังคมและคุณภาพชีวิตของประชาชนในมิติสุขภาพจิตและเศรษฐกิจ [1], [2] ในระดับโลกพบว่าสตรีกว่า 840 ล้านคน เคยประสบความรุนแรงจากคู่ครองหรือทางเพศตลอดช่วงชีวิต [3] นอกจากนี้รายงานของสำนักงานว่าด้วยยาเสพติดและอาชญากรรมแห่งสหประชาชาติ (United Nations Office on Drugs and Crime; UNODC) และหน่วยงานเพื่อการส่งเสริมความเสมอภาคระหว่างเพศ และพลังสตรีแห่งสหประชาชาติ (UN Women) ระบุว่าใน พ.ศ. 2566 มีสตรีและเด็กหญิงถูกสังหารโดยคู่ครองหรือสมาชิกในครอบครัวมากกว่า 51,100 ราย [4] และต้นทุนทางเศรษฐกิจจากปัญหาดังกล่าวคิดเป็นสัดส่วนร้อยละ 1.2–3.7 ของผลิตภัณฑ์มวลรวมในประเทศ (Gross Domestic Product; GDP) ของประเทศกำลังพัฒนา [5] สำหรับประเทศไทยรายงานประจำปี พ.ศ. 2566 ระบุว่ามีผู้ถูกกระทำความรุนแรงในครอบครัวที่ได้รับการช่วยเหลือ จำนวน 2,312 ราย และใน พ.ศ. 2567 ตัวเลขดังกล่าวเพิ่มขึ้นเป็น 4,833 ราย หรือเฉลี่ยวันละ 42 ราย [6], [7] สะท้อนถึงแนวโน้มที่เพิ่มขึ้นอย่างต่อเนื่อง และรุนแรงยิ่งขึ้น ขณะที่รายงานอย่างเป็นทางการภายใต้พระราชบัญญัติคุ้มครองผู้ถูกกระทำด้วยความรุนแรงในครอบครัว พ.ศ. 2550 [8] ชี้ให้เห็นถึงความจำเป็นเร่งด่วนในการพัฒนาเครื่องมือเชิงเทคโนโลยีเพื่อสนับสนุนการเฝ้าระวัง และคัดกรองกลุ่มเสี่ยงล่วงหน้า และการแก้ไขปัญหาดังกล่าวต้องอาศัยทั้งกลไกเชิงชุมชน และการวิเคราะห์ข้อมูลเชิงดิจิทัลควบคู่กัน [9], [10]

งานวิจัยที่ผ่านมาด้านการตรวจจับความเสี่ยงความรุนแรงในครอบครัวด้วยปัญญาประดิษฐ์แบ่งได้เป็น 3 กลุ่ม ได้แก่ 1) วิธีการแบบเดี่ยวที่ต้องใช้ข้อมูลที่มีป้ายกำกับ (Label) ซึ่งหมายถึงสถานะที่ผู้เชี่ยวชาญระบุไว้ล่วงหน้าว่าแต่ละรายการข้อมูลเป็นกลุ่มเสี่ยงหรือกลุ่มปกติ เช่น บันทึกคดีที่ผ่านการร้องเรียนแล้ว ซึ่งพบว่ามีข้อจำกัดด้านความแม่นยำเมื่อข้อมูลไม่มีป้ายกำกับ และหายากในบริบทสังคม [11], [12] 2) วิธีการแบบเดี่ยวที่ไม่ต้องใช้ป้ายกำกับ (Unsupervised) ได้แก่ แบบจำลองส่วนผสมเกาส์เซียน (Gaussian Mixture Model; GMM) [13], ออโตเอ็นโค้ดเดอร์ (Autoencoder; AE) [14], และ ตัวประกอบค่าผิดปกติเฉพาะที่ (Local Outlier Factor; LOF) [15] ซึ่งแต่ละวิธีมีจุดแข็งเฉพาะด้าน แต่มีข้อจำกัดเมื่อใช้เพียงลำพังในข้อมูลมิติสูง และ 3) วิธีการแบบผสม (Hybrid/Ensemble) ที่รวมจุดแข็งของหลายอัลกอริทึมเพื่อเพิ่มความเสถียรของผลลัพธ์ [16], [17] รวมถึงการประยุกต์ใช้เชิงนโยบายด้วยการวิเคราะห์กลุ่มแฝง (Latent Class Analysis) เพื่อระบุกลุ่มเสี่ยงในบริบทที่หลากหลาย [18] อย่างไรก็ตามงานวิจัยที่ผ่านมายังขาดการออกแบบกระบวนการแบบสองขั้นตอนที่แยกกลุ่มเสี่ยงชัดเจนออกจากกลุ่มเสี่ยงแฝงอย่างเป็นระบบในข้อมูลสังคมที่ไม่มีป้ายกำกับโดยเฉพาะ

งานวิจัยนี้จึงนำเสนอแบบจำลองไฮบริดแบบไม่มีผู้สอนสองขั้นตอน (Two-Stage Unsupervised Hybrid Model) โดยเลือกใช้อัลกอริทึมหลัก 3 รูปแบบ ที่มีหลักการแตกต่างกัน ได้แก่ 1) แบบจำลองส่วนผสมเกาส์เซียน ซึ่งเป็นแบบจำลองเชิงสถิติ (Probabilistic Model) ที่ประมาณความหนาแน่นของข้อมูลเพื่อตรวจจับความผิดปกติในเชิงโครงสร้างภาพรวม [13] 2) ออโตเอ็นโค้ดเดอร์ ซึ่งเป็นโครงข่ายประสาทเทียมเชิงลึกที่เรียนรู้คุณลักษณะเชิงซ่อน (Latent Representation) เพื่อระบุความผิดปกติทางโครงสร้าง [14] และ 3) ตัวประกอบค่าผิดปกติเฉพาะที่ ซึ่งเป็นแบบจำลองเชิงความหนาแน่นเฉพาะที่ที่ตรวจจับข้อมูลที่แตกต่างจากกลุ่มรอบข้าง [15] โดยขั้นตอนแรกคัดกรองกลุ่มเสี่ยงสูงชัดเจนออกจากข้อมูลทั้งหมดผ่านกลยุทธ์การผสานผลลัพธ์ (Fusion Logic) สามรูปแบบ ได้แก่ ตรรกะและ (AND) ตรรกะหรือ (OR) และการถ่วงน้ำหนัก (Weighted Fusion) และขั้นตอนที่สองค้นหากลุ่มเสี่ยงแฝงจากข้อมูลที่เหลือ แนวทางนี้แตกต่างจากการรวมแบบจำลอง (Ensemble) ทั่วไปในเชิงแนวคิด กล่าวคือการรวมแบบจำลองทั่วไปจะประมวลชุดข้อมูลเดียวกันพร้อมกันทุกแบบจำลองในขั้นตอนเดียว แต่แนวทางที่นำเสนอใช้การกรองแบบลำดับขั้น (Sequential Filtering) วิธีนี้ช่วยลดการรบกวนของข้อมูลผิดปกติที่ชัดเจนต่อการตรวจจับกลุ่มเสี่ยงแฝงที่มีขนาดเล็กกว่า



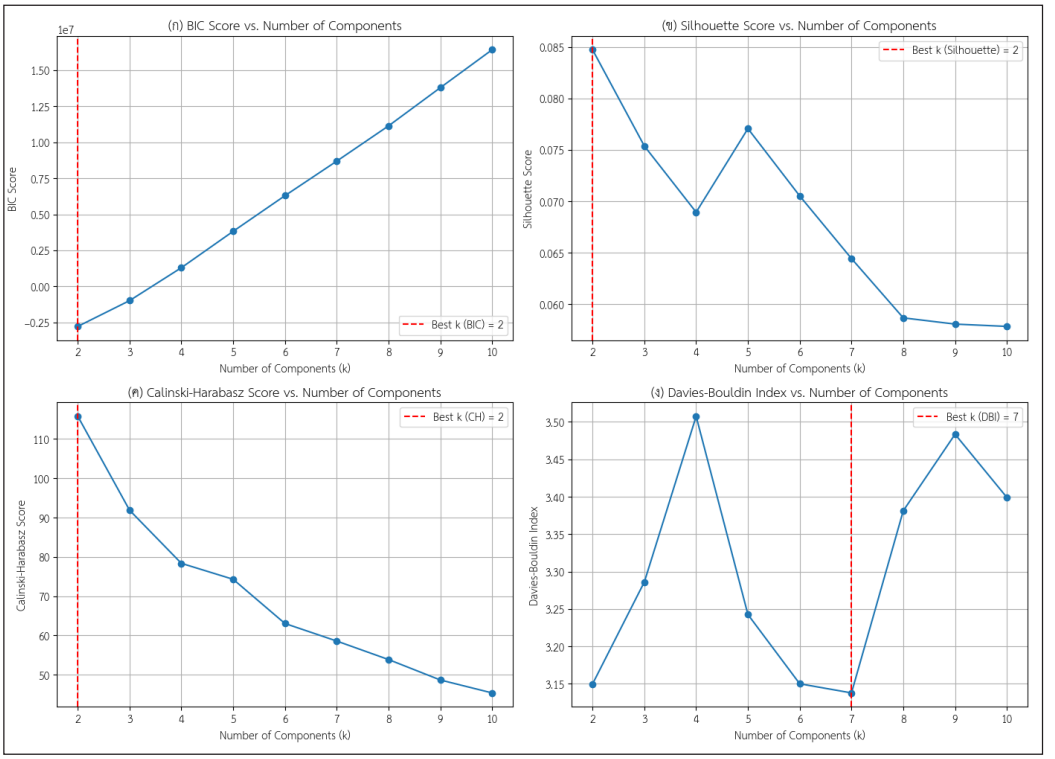
รูปที่ 1 ภาพรวมกระบวนการดำเนินงานวิจัยแบบสองขั้นตอน

2. วิธีการทดลอง

กระบวนการดำเนินงานวิจัยประกอบด้วยสองขั้นตอนหลัก ได้แก่ ขั้นตอนที่ 1 การคัดกรองกลุ่มเสี่ยงสูงที่ชัดเจน และขั้นตอนที่ 2 การตรวจจับกลุ่มเสี่ยงแฝง ดังแสดงในรูปที่ 1

2.1 แหล่งข้อมูลและการเตรียมข้อมูล

ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลทุติยภูมิ (Secondary Data) ซึ่งเป็นข้อมูลที่ถูกรวบรวม และเผยแพร่โดยหน่วยงานภาครัฐ ผู้วิจัยไม่ได้เก็บรวบรวมข้อมูลเอง เช่น สถิติทางราชการหรือฐานข้อมูลสาธารณะ โดยงานวิจัยนี้ใช้ชุดข้อมูลผู้กระทำความรุนแรงในครอบครัวจำแนกตามปัญหาหรือสาเหตุรายประเด็น ประจำปี พ.ศ. 2566 และ พ.ศ. 2567 ซึ่งจัดทำโดยกรมกิจการสตรีและสถาบันครอบครัว กระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ และเผยแพร่ผ่านระบบบัญชีข้อมูลภาครัฐ บนเว็บไซต์ gdcatalog.m-society.go.th [19], [20] ข้อมูลดังกล่าวเป็นข้อมูลเชิงสถิติแบบสรุป (Aggregate Data) ที่ไม่ระบุตัวบุคคล และสอดคล้องกับพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. 2562 (PDPA) ชุดข้อมูลประกอบด้วยตัวแปร 3 กลุ่มหลัก ได้แก่ 1) ตัวแปรเชิงประชากรศาสตร์ เช่น อายุ เพศ และอาชีพ 2) ตัวแปรด้านสาเหตุของความรุนแรง เช่น ประเภทปัญหา และแรงจูงใจ และ 3) ตัวแปรด้านบริบทความรุนแรง เช่น ลักษณะความสัมพันธ์ และภูมิภาค ในขั้นตอนการเตรียมข้อมูลเบื้องต้น ผู้วิจัยได้รวบรวมไฟล์ข้อมูลในรูปแบบ CSV จำนวน 2 ชุด จาก พ.ศ. 2566 และ พ.ศ. 2567 ซึ่งมีโครงสร้างตัวแปรหลักเหมือนกันแต่มีความแตกต่างของบางคอลัมน์ย่อย จึงได้ดำเนินการตรวจสอบความสอดคล้องของโครงสร้างตัวแปร (Schema Alignment) และปรับตัวแปรให้เป็นรูปแบบเดียวกัน (Feature Harmonization) ก่อนทำการรวมข้อมูล (Data Integration) ทำให้ได้ชุดข้อมูลตั้งต้นรวมทั้งสิ้น

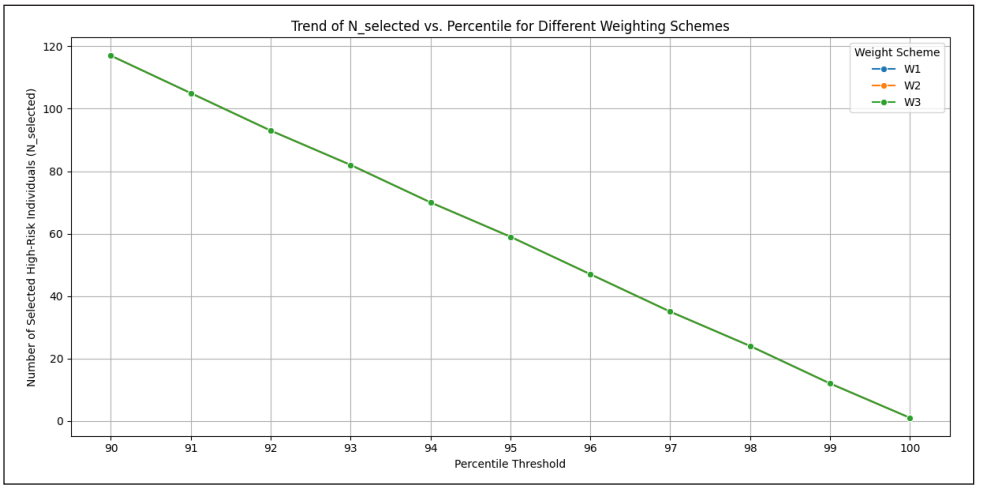


รูปที่ 2 การเปรียบเทียบค่าดัชนีประเมินคลัสเตอร์ของแบบจำลองส่วนผสมเกาส์เซียนขั้นตอนที่ 1 การคัดกรองกลุ่มเสี่ยงสูงที่ชัดเจน

1,161 รายการ ภายหลังการรวมข้อมูล จึงดำเนินกระบวนการเตรียมข้อมูล (Preprocessing) โดยแปลงตัวแปรเชิงหมวดหมู่ให้เป็นข้อมูลเชิงตัวเลขด้วยเทคนิคการเข้ารหัสแบบวันฮอต (One-hot encoding) และปรับสเกลตัวแปรเชิงตัวเลขให้เป็นมาตรฐาน (Standardization) ด้วยฟังก์ชัน StandardScaler เพื่อลดอคติจากความแตกต่างของช่วงข้อมูล ส่งผลให้ได้ชุดข้อมูลที่พร้อมสำหรับการวิเคราะห์ซึ่งมีมิติ (1,161×869) โดย 869 คือ จำนวนตัวแปร (Features) ที่เกิดจากการขยายตัวแปรเชิงหมวดหมู่ภายหลังการเข้ารหัสแบบวันฮอต ทั้งนี้จำนวนรายการ 1,161 รายการ เป็นข้อมูลตั้งต้นก่อนการคัดกรองกลุ่มเสี่ยงสูงในขั้นตอนที่ 1 ของแบบจำลองสองขั้นตอน

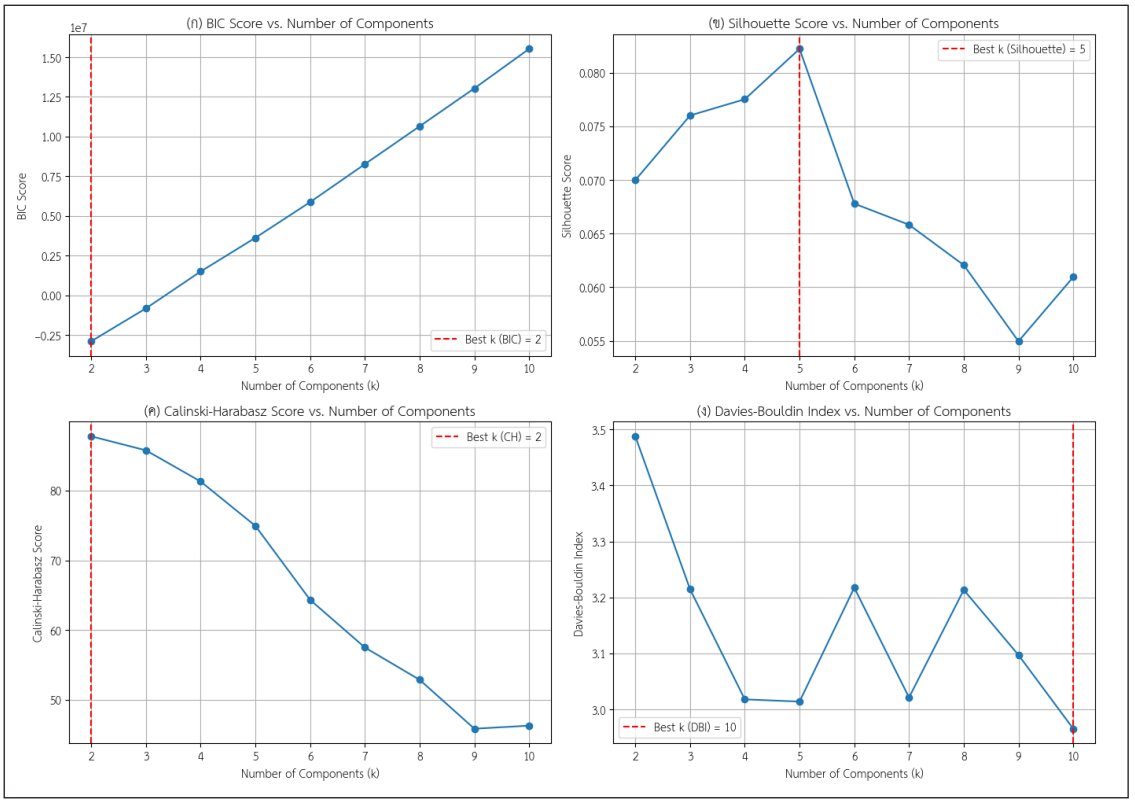
2.2 ขั้นตอนที่ 1 การคัดกรองกลุ่มเสี่ยงสูงที่ชัดเจน

ในขั้นตอนนี้ใช้ชุดข้อมูลที่ผ่านการเตรียมไว้แล้วจำนวน 1,161 รายการ โดยประกอบด้วยกระบวนการย่อย ได้แก่ 1) การประเมินค่าเกณฑ์ข้อมูลแบบเบย์เซียน (Bayesian Information Criterion; BIC) ร่วมกับดัชนีซิลูเอท (Silhouette Index) ดัชนีคาลินสกี้-ฮาราบาสซ์ (Calinski–Harabasz Index; CH) และดัชนีเดวีส์-โบลดิน (Davies–Bouldin Index; DBI) โดยดัชนีซิลูเอทวัดความกระชับและการแยกตัวของกลุ่ม (ค่ายิ่งสูงยิ่งดี) [21] ดัชนีคาลินสกี้-ฮาราบาสซ์วัดอัตราส่วนระหว่างความแปรปรวนระหว่างกลุ่มต่อภายในกลุ่ม (ค่ายิ่งสูงยิ่งดี) [22] และดัชนีเดวีส์-โบลดินวัดความคล้ายคลึงเฉลี่ยระหว่างกลุ่ม (ค่ายิ่งต่ำยิ่งดี) [23] ทั้งสามดัชนีเป็นตัวชี้วัดเชิงโครงสร้างภายใน (Internal Validity Index) ซึ่งมีข้อจำกัดด้านความไวต่อการกระจายของข้อมูล เพื่อระบุจำนวนคลัสเตอร์ (k) ที่เหมาะสมที่สุดสำหรับการใช้งานในแบบจำลองส่วนผสมเกาส์เซียน ผลการประเมินค่าดัชนีทั้งสี่แสดงดังรูปที่ 2 ซึ่งบ่งชี้ว่าจำนวนคลัสเตอร์ที่เหมาะสมที่สุดคือ $k = 2$ ตามเกณฑ์การประเมินค่าเกณฑ์ข้อมูลแบบเบย์เซียน ดัชนีซิลูเอท และดัชนีคาลินสกี้-ฮาราบาสซ์ แม้ดัชนีเดวีส์-โบลดินจะมีค่าต่ำสุดที่ $k = 7$ แต่



รูปที่ 3 แนวโน้มกลุ่มเสี่ยงสูงตามระดับเปอร์เซ็นไทล์

เนื่องจากงานวิจัยนี้กำหนดโครงสร้างเชิงทฤษฎีเป็นสองกลุ่ม จึงเลือกใช้ $k = 2$ ตามเสียงข้างมากของดัชนี และหลักการวิเคราะห์ประเภทความเสี่ยงในบริบทนโยบาย [24] 2) หลังจากได้จำนวนคลัสเตอร์ที่เหมาะสมแล้วจากนั้นเริ่มฝึกสอนแบบจำลองเพื่อหาค่าผิดปกติ กำหนดให้แบบจำลองส่วนผสมเกาส์เซียนใช้ค่า k จากข้อ 1) ออโตเอ็นโค้ดเดอร์ใช้โครงสร้างชั้นซ่อน 16-8-16 เพื่อคำนวณคะแนนความผิดปกติ และตัวประกอบค่าผิดปกติเฉพาะที่ใช้การประเมินความหนาแน่นเฉพาะที่ 3) ขั้นตอนนี้ทำการวิเคราะห์ความไวของพารามิเตอร์ (Grid Search Sensitivity Analysis) เพื่อระบุค่าน้ำหนักและระดับเปอร์เซ็นไทล์ที่เหมาะสมที่สุดสำหรับคุณภาพการจัดกลุ่มเสี่ยงสูงที่ชัดเจน โดยทดสอบเปอร์เซ็นไทล์ (Percentile) ตั้งแต่ 90–100 ร่วมกับสมมติฐานค่าน้ำหนัก 3 แบบ ได้แก่ W1 (0.4:0.3:0.3) ซึ่งให้ความสำคัญกับแบบจำลองส่วนผสมเกาส์เซียนมากกว่าเล็กน้อย W2 (0.5:0.25:0.25) ให้แบบจำลองส่วนผสมเกาส์เซียนมีน้ำหนักมากที่สุด และ W3 (0.33:0.33:0.33) ซึ่งให้น้ำหนักเท่ากันทั้งสามแบบจำลอง 4) ทำการคำนวณแนวโน้มของจำนวนบุคคลที่ถูกจัดเป็นกลุ่มเสี่ยงสูงที่ชัดเจน ภายใต้แต่ละระดับเปอร์เซ็นไทล์ ดังแสดงในรูปที่ 3 พบว่าแนวโน้มของจำนวนกลุ่มเสี่ยงสูงชัดเจนมีค่าลดลงอย่างต่อเนื่องเมื่อค่าเปอร์เซ็นไทล์เพิ่มขึ้น ซึ่งเป็นลักษณะที่คาดได้ตามหลักสถิติ เนื่องจากการกำหนดเปอร์เซ็นไทล์ที่สูงขึ้นหมายถึงการตั้งเกณฑ์ตัดสินที่เข้มงวดมากขึ้น ทำให้มีเพียงข้อมูลที่มีคะแนนความผิดปกติสูงสุดเท่านั้นที่ถูกจัดเป็นกลุ่มเสี่ยง [15] โดยทั้งสามสมมติฐานค่าน้ำหนักให้แนวโน้มที่ใกล้เคียงกัน แสดงถึงความเสถียรของแบบจำลองต่อการเปลี่ยนแปลงพารามิเตอร์ 5) จากนั้นได้ทำการเปรียบเทียบวิธีการผสานคะแนน (Score Fusion) 3 รูปแบบ ได้แก่ การปรับช่วงข้อมูลแบบต่ำสุด-สูงสุดร่วมกัน (Joint MinMax Normalization) การปรับมาตรฐานด้วยค่าซี (Z-score Normalization) และการผสานผลลัพธ์ตามลำดับ (Rank-based Fusion) โดยประเมินคุณภาพการจัดกลุ่มของแต่ละวิธีด้วยค่าดัชนีซิลูเอท ภายใต้สมมติฐานค่าน้ำหนัก W1–W3 ผลการวิเคราะห์พบว่าการผสานผลลัพธ์ตามลำดับภายใต้ค่าน้ำหนัก W3 (0.33:0.33:0.33) ให้ค่าดัชนีซิลูเอทสูงสุด (0.5946) ซึ่งสะท้อนถึงความชัดเจนของโครงสร้างคลัสเตอร์ที่ดีที่สุด ดังนั้นจึงเลือกใช้ระดับเปอร์เซ็นไทล์ที่ 95 และค่าน้ำหนัก W3 เป็นพารามิเตอร์อ้างอิงในการกรองข้อมูลเบื้องต้น โดยได้กลุ่มเสี่ยงสูงที่ชัดเจน จำนวน 59 รายการจากข้อมูลทั้งหมด 1,161 รายการ เหลือข้อมูลสำหรับการฝึกซ้ำ (Re-training) จำนวน 1,102 รายการ การเลือก $k = 2$ ยังสอดคล้องกับแนวคิดเชิงทฤษฎีของงานวิจัย ซึ่งกำหนดให้ข้อมูลถูกจัดกลุ่มเชิงโครงสร้างเป็นสองกลุ่ม ได้แก่ กลุ่มเสี่ยงและกลุ่มปกติ ซึ่งเป็นโครงสร้างพื้นฐานที่สอดคล้องกับการ



รูปที่ 4 การเปรียบเทียบค่าดัชนีประเมินคลัสเตอร์ของ GMM ขั้นตอนที่ 2 การคัดกรองกลุ่มเสี่ยงแฝงจากกลุ่มปกติ

วิเคราะห์ประเภทความเสี่ยงในบริบทนโยบาย [24]

2.3 ขั้นตอนที่ 2 การตรวจจับกลุ่มเสี่ยงสูงแฝงจากกลุ่มปกติ

ขั้นตอนนี้เป็นการดำเนินการซ้ำตามแนวทางของขั้นตอนที่ 1 โดยใช้ข้อมูลที่เหลือจำนวน 1,102 รายการ (จากเดิม 1,161 รายการ หลังตัดกลุ่มเสี่ยงสูงที่ชัดเจน จำนวน 59 รายการออก) ซึ่งผ่านการแปลงให้อยู่ในรูปแบบที่เหมาะสมต่อการเรียนรู้ของเครื่อง โดยแปลงตัวแปรเชิงหมวดหมู่เป็นข้อมูลเชิงตัวเลขด้วยการเข้ารหัสแบบวันฮอต และปรับสเกลตัวแปรเชิงตัวเลขให้เป็นมาตรฐาน ด้วยฟังก์ชัน StandardScaler เพื่อป้องกันอคติจากขนาดของข้อมูล ผลลัพธ์คือ ชุดข้อมูลที่พร้อมสำหรับการวิเคราะห์ ซึ่งมีมิติ (1,102×869) จากนั้นดำเนินการดังนี้ 1) หาจำนวนคลัสเตอร์ที่เหมาะสมของแบบจำลองส่วนผสมเกาส์เซียนทำการประเมินค่าตัวชี้วัดเกณฑ์ข้อมูลแบบเบย์เซียน ดัชนีซิลูเอท ดัชนีคาลินสกี้-ฮาราบาสซ์ และดัชนีเดวีส์-โบลดินเพื่อระบุจำนวนคลัสเตอร์ที่เหมาะสมที่สุดของแบบจำลองส่วนผสมเกาส์เซียนสำหรับชุดข้อมูลใหม่ ผลการประเมินดังแสดงในรูปที่ 4 พบว่าเกณฑ์ข้อมูลแบบเบย์เซียนและดัชนีคาลินสกี้-ฮาราบาสซ์ชี้ว่า $k = 2$ เป็นค่าที่เหมาะสมที่สุด แม้ดัชนีซิลูเอทจะมีค่าสูงสุดที่ $k = 5$ และดัชนีเดวีส์-โบลดินต่ำสุดที่ $k = 10$ แต่เนื่องจากงานวิจัยนี้กำหนดโครงสร้างเชิงทฤษฎีเป็นสองกลุ่ม ได้แก่ กลุ่มเสี่ยงและกลุ่มปกติ จึงเลือกใช้ $k = 2$ ตามหลักการวิเคราะห์ประเภทความเสี่ยงในบริบทนโยบาย [24] 2) สร้างแบบจำลองพื้นฐานและการวิเคราะห์ความไวของพารามิเตอร์เพื่อระบุค่าน้ำหนัก และระดับเปอร์เซ็นไทล์ที่เหมาะสมที่สุด โดยทดสอบค่าเปอร์เซ็นไทล์ 90–100 ร่วมกับสมมติฐานค่าน้ำหนัก 3 แบบ ได้แก่ W1 (0.4:0.3:0.3), W2 (0.5:0.25:0.25) และ W3 (0.33:0.33:0.33) เช่นเดียวกับขั้นตอนที่ 1 ทำการประเมินค่าตัวชี้วัดเกณฑ์ข้อมูลแบบเบย์เซียน

ดัชนีซิลูเอท ดัชนีคาลินสกี้-ฮาราบาสซ์ และดัชนีเดวีส์-โบลดิน เพื่อระบุจำนวนคลัสเตอร์ที่เหมาะสมที่สุดของแบบจำลองส่วนผสมเกาส์เซียนสำหรับชุดข้อมูลใหม่ ผลการประเมินดังแสดงในรูปที่ 4 พบว่าเกณฑ์ข้อมูลแบบเบย์เซียน และดัชนีคาลินสกี้-ฮาราบาสซ์ ชี้ว่า k = 2 เป็นค่าที่เหมาะสมที่สุด แม้ดัชนีซิลูเอทจะมีค่าสูงสุดที่ k = 5 และดัชนีเดวีส์-โบลดินต่ำสุดที่ k = 10 แต่เนื่องจากงานวิจัยนี้กำหนดโครงสร้างเชิงทฤษฎีเป็นสองกลุ่ม ได้แก่ กลุ่มเสี่ยง และกลุ่มปกติ จึงเลือกใช้ k = 2 ตามหลักการวิเคราะห์ประเภทความเสี่ยงในบริบทนโยบาย [24] 2) สร้างแบบจำลองพื้นฐานและการวิเคราะห์ความไวของพารามิเตอร์เพื่อระบุค่าน้ำหนัก และระดับเปอร์เซ็นไทล์ที่เหมาะสมที่สุด โดยทดสอบค่าเปอร์เซ็นไทล์ 90–100 ร่วมกับสมมติฐานค่าน้ำหนัก 3 แบบ ได้แก่ W1 (0.4:0.3:0.3), W2 (0.5:0.25:0.25) และ W3 (0.33:0.33:0.33) เช่นเดียวกับขั้นตอนที่ 1 3) การวิเคราะห์แนวโน้ม และการเปรียบเทียบวิธีผสานคะแนน ทำการคำนวณแนวโน้มของจำนวนบุคคลที่ถูกจัดเป็นกลุ่มเสี่ยงสูงแฝงภายใต้ระดับเปอร์เซ็นไทล์ที่แตกต่างกัน และเปรียบเทียบคุณภาพการจัดกลุ่มของวิธีการผสานคะแนนทั้ง 3 รูปแบบ ได้แก่ การปรับช่วงข้อมูลแบบต่ำสุด-สูงสุดร่วมกัน การปรับมาตรฐานด้วยค่าซี และการผสานผลลัพธ์ตามลำดับ เพื่อประเมินคุณภาพเชิงโครงสร้างของการจัดกลุ่มด้วยค่าดัชนีซิลูเอท ผลการทดสอบพบว่าการปรับช่วงข้อมูลแบบต่ำสุด-สูงสุดร่วมกันภายใต้ค่าน้ำหนัก W2 (0.5:0.25:0.25) ให้ดัชนีซิลูเอทสูงสุด (0.6510) ซึ่งสะท้อนถึงโครงสร้างการจัดกลุ่มที่ชัดเจน และมีความเสถียรมากที่สุดจึงเลือกใช้เปอร์เซ็นไทล์ที่ 95 และค่าน้ำหนัก W2 เป็นพารามิเตอร์อ้างอิงสำหรับคุณภาพการจัดกลุ่มเสี่ยงสูงที่แฝงอยู่จากข้อมูล 1,102 รายการ ที่เหลือหลังการคัดกรองในขั้นตอนแรก 4) การสร้างแบบจำลองไฮบริดจากค่าพารามิเตอร์ที่เหมาะสมดังกล่าว ทำการสร้างแบบจำลองพื้นฐานใหม่ตามค่าน้ำหนักที่กำหนด และพัฒนาแบบจำลองไฮบริดด้วยสามกลยุทธ์การผสานผล ได้แก่ ตรรกะและตรรกะหรือและการผสานแบบถ่วงน้ำหนัก เพื่อใช้ในคุณภาพการจัดกลุ่มเสี่ยงสูงแฝงขั้นสุดท้าย และประเมินคุณภาพของการจัดกลุ่มตรวจสอบความเหมาะสมของโครงสร้างกลุ่มในแบบจำลอง โดยการใช้เกณฑ์วัดคุณภาพเชิงโครงสร้างของการจัดกลุ่ม (Internal Clustering Validity Indices) จำนวน 3 ดัชนี ได้แก่ 1) ดัชนีซิลูเอท ซึ่งใช้วัดระดับความชัดเจนของการแยกกลุ่ม โดยพิจารณาความใกล้ชิดของข้อมูลภายในกลุ่มเทียบกับความห่างจากกลุ่มอื่น ค่าที่สูงแสดงถึงโครงสร้างกลุ่มที่ชัดเจน 2) ดัชนีคาลินสกี้-ฮาราบาสซ์ ซึ่งวัดอัตราส่วนระหว่างความแปรปรวนระหว่างกลุ่มต่อความแปรปรวนภายในกลุ่ม โดยค่าที่สูงสะท้อนถึงการกระจายกลุ่มที่เหมาะสม และ 3) ดัชนีเดวีส์-โบลดิน ซึ่งใช้วัดความคล้ายคลึงระหว่างกลุ่ม โดยค่าที่ต่ำแสดงถึงการแยกกลุ่มที่มีคุณภาพดีกว่า ทั้งนี้ ดัชนีทั้งสามเป็นตัวชี้วัดภายใน (Internal Validation) ที่เหมาะสมกับข้อมูลที่ไม่มีป้ายกำกับ (Unlabeled Data) จึงถูกใช้ร่วมกันเพื่อประเมินความเหมาะสมของโครงสร้างกลุ่มในแบบจำลองแบบไม่มีผู้สอน โดยค่าที่ต่ำกว่าจะบ่งชี้ถึงโครงสร้างที่มีคุณภาพดีกว่าเพื่อใช้เปรียบเทียบคุณภาพการจัดกลุ่มของแต่ละแนวทาง ทั้งนี้ ดัชนีซิลูเอทที่ยอมรับได้ในเชิงโครงสร้างควรมีค่ามากกว่า 0 โดยค่าที่สูงกว่า 0.5 บ่งชี้ถึงโครงสร้างกลุ่มที่มีความชัดเจนในระดับดี [25] ตามแนวทางการประเมินทางสถิติของ Field [26] นอกจากนี้การแสดงผลด้วยเทคนิคการลดมิติข้อมูลแบบที-เอสเอ็นอี (t-distributed Stochastic Neighbor Embedding; t-SNE) [27] ช่วยให้สามารถมองเห็นการกระจายตัวของข้อมูลในมิติต่ำ (2 มิติ) เพื่อยืนยันการแยกกลุ่มในเชิงภาพ อย่างไรก็ตามตัวชี้วัดทั้งสามเป็นดัชนีเชิงโครงสร้างภายใน (Internal Validity Indices) ซึ่งมีความไวต่อการกระจายของข้อมูลและจำนวนกลุ่ม จึงควรพิจารณาร่วมกันมากกว่าการใช้ดัชนีใดดัชนีหนึ่งเพียงลำพัง นอกจากนี้เพื่อควบคุมความเสี่ยงของการสรุปผลบวกลวงที่อาจเกิดจากการทดสอบหลายครั้ง ผู้วิจัยได้ใช้วิธีการควบคุมอัตราการค้นพบที่ผิดพลาด (False Discovery Rate; FDR) ตามแนวทางของ Benjamini และ Hochberg [28] ซึ่งตระหนักว่าผลลัพธ์ที่ได้จากดัชนีเชิงโครงสร้างภายในนั้นไม่มีข้อมูลป้ายกำกับภายนอก (External Labels) มาใช้ยืนยัน ดังนั้นเพื่อเพิ่มความน่าเชื่อถือผู้วิจัยจึงได้ดำเนินการวิเคราะห์เสถียรภาพโดยใช้วิธีการสุ่มตัวอย่างซ้ำ (Resampling-based Stability

Analysis) จำนวน 20 รอบ โดยในแต่ละรอบจะสุ่มเลือกข้อมูลร้อยละ 85 ของชุดข้อมูลทั้งหมด และตรวจสอบความสอดคล้องของคุณภาพการจัดกลุ่มเสี่ยงทั้งสองขั้นตอนผ่านค่าความถี่ของเสถียรภาพ (Stability Frequency) เพื่อลดความเสี่ยงของการสรุปผลที่เกิดจากความบังเอิญของชุดข้อมูล [29] ผลลัพธ์นี้ยืนยันว่าวิธีการปรับช่วงข้อมูลแบบต่ำสุด-สูงสุดร่วมกันและการรวมแบบผลรวมถ่วงน้ำหนัก (Joint MinMax + Weighted Sum) สามารถเพิ่มคุณภาพเชิงโครงสร้างของการจัดกลุ่มในการตรวจจับกลุ่มเสี่ยงแฝงได้ดีกว่าวิธีการผสานคะแนนแบบอื่น โดยยังคงความเสถียรของโครงสร้างกลุ่มได้ดีภายใต้จำนวนคลัสเตอร์คือ $k = 2$ ที่ได้จากแบบจำลองส่วนผสมเกาส์เซียน

3. ผลการทดลองและอภิปรายผลการทดลอง

3.1 ผลคุณภาพการจัดกลุ่มเสี่ยงสูงที่ชัดเจน

ผลคุณภาพการจัดกลุ่มเสี่ยงสูงที่ชัดเจนในขั้นตอนที่ 1 พบว่าจากข้อมูลทั้งหมด 1,161 รายการ มีจำนวน 59 รายการ (ร้อยละ 5.08) ที่ถูกจัดให้อยู่ในกลุ่มเสี่ยงสูงชัดเจน โดยอ้างอิงจากเกณฑ์เปอร์เซ็นไทล์ที่ 95 และค่าน้ำหนักแบบ W3 (0.33:0.33:0.33) ซึ่งได้จากการวิเคราะห์ความไวของพารามิเตอร์ และการผสานคะแนนในขั้นตอนก่อนหน้า ก่อนประเมินคุณภาพการจัดกลุ่ม ข้อมูลได้ผ่านการตรวจสอบการแจกแจงแบบปกติตามแนวทางของ Ghasemi และ Zahediasl [30] ผลการทดสอบชาร์ปิโร-วิลค์ (Shapiro-Wilk Test) พบว่าข้อมูลไม่เป็นไปตามการแจกแจงแบบปกติ จึงเลือกใช้การทดสอบครัสคัล-วัลลิส (Kruskal–Wallis test) ซึ่งเป็นสถิติแบบสถิตินอนพารามิเตอร์ (Non-parametric statistics) ที่เหมาะสมตามแนวทางของ Mishra และคณะ [31] โดยมีวัตถุประสงค์สองประการ ได้แก่ 1) วิเคราะห์ลักษณะเฉพาะของแต่ละกลุ่มในเชิงเนื้อหา และ 2) ยืนยันความเหมาะสมของการใช้เกณฑ์เชิงเปอร์เซ็นไทล์ในการระบุค่าความเบี่ยงเบน โดยไม่ได้มีวัตถุประสงค์เพื่อยืนยันผลการจัดกลุ่มของแบบจำลองโดยตรง โดยตั้งสมมติฐานดังนี้ H_0: ตัวแปรนั้นมีการกระจายตัวไม่แตกต่างกันอย่างมีนัยสำคัญระหว่างกลุ่มเสี่ยงสูงชัดเจน กลุ่มเสี่ยงสูงแฝง และกลุ่มปกติ และ H_1: ตัวแปรนั้นมีการกระจายตัวแตกต่างกันอย่างมีนัยสำคัญอย่างน้อยหนึ่งกลุ่ม หลังจากนั้นกรองข้อมูลออกเหลือสำหรับใช้ฝึกในขั้นตอนที่ 2 จำนวน 1,102 รายการ

เพื่อให้การตีความผลการจัดกลุ่มมีความชัดเจนเชิงเนื้อหา ผู้วิจัยได้ดำเนินการวิเคราะห์ลักษณะเฉพาะของแต่ละกลุ่ม (Profile Analysis) ดังแสดงในตารางที่ 1 โดยเปรียบเทียบ 3 กลุ่ม ได้แก่ กลุ่มปกติ (n = 1,046) กลุ่มเสี่ยงสูงชัดเจน (n = 59) และกลุ่มเสี่ยงสูงแฝง (n = 56) จากผลการวิเคราะห์สถิติเชิงพรรณนาพบว่าอายุเฉลี่ยของกลุ่มปกติ (Mean = 40.11, SD = 17.40) สูงกว่ากลุ่มเสี่ยงสูงชัดเจน (Mean = 22.78, SD = 15.74) และกลุ่มเสี่ยงสูงแฝง (Mean = 21.36, SD = 13.54) อย่างชัดเจน ผลการทดสอบครัสคัล-วัลลิส พบว่าอายุมีความแตกต่างระหว่างกลุ่มอย่างมีนัยสำคัญทางสถิติ (H = 105.07, p-value < 0.001) นอกจากนี้ การวิเคราะห์ตัวแปรเชิงหมวดหมู่ด้วยการทดสอบไคสแควร์ (Chi-square Test) พบว่าปัจจัยเชิงพฤติกรรมและบริบทหลายตัวแปร เช่น Alcohol, Drug, Rage, Health Problem, Mental Problem และ Economics Stress มีความแตกต่างระหว่างกลุ่มอย่างมีนัยสำคัญทางสถิติ (p-value < 0.05) ซึ่งบ่งชี้ว่ากลุ่มเสี่ยงมีลักษณะเชิงบริบทที่แตกต่างจากกลุ่มปกติอย่างมีนัยสำคัญ และสอดคล้องกับโครงสร้างการจัดกลุ่มของแบบจำลองไฮบริดแบบสองขั้นตอน

3.2 ผลคุณภาพการจัดกลุ่มเสี่ยงแฝงจากกลุ่มปกติ

ผลการวิเคราะห์จำนวนคลัสเตอร์ของแบบจำลองส่วนผสมเกาส์เซียนสำหรับข้อมูลขั้นตอนที่ 2 ดังแสดงในรูปที่ 4 พบว่าเกณฑ์ข้อมูลแบบเบย์เซียน และดัชนีคาลินสกี้-ฮาราบาสซ์ชี้ว่า k = 2 เป็นค่าที่เหมาะสมที่สุด แม้ดัชนีซิลูเอทจะมีค่าสูงสุดที่ k = 5 และดัชนีเดวีส์-โบลดินต่ำสุดที่ k = 10 แต่เนื่องจากงานวิจัยนี้กำหนดโครงสร้างเชิงทฤษฎีเป็นสองกลุ่ม ได้แก่ กลุ่มเสี่ยงและกลุ่มปกติ จึงเลือกใช้ k = 2 ตามหลักการวิเคราะห์ประเภทความเสี่ยงในบริบทนโยบาย [24] จากกระบวนการวิเคราะห์ความไวและการเปรียบเทียบวิธีผสานคะแนนในขั้นตอนที่ 2 ได้ข้อสรุปว่าการปรับช่วงข้อมูลแบบต่ำสุด-สูงสุดร่วมกัน ภายใต้ค่าน้ำหนัก W2 ที่ระดับ

ตารางที่ 1 การวิเคราะห์ลักษณะเฉพาะของกลุ่มตัวอย่าง (Profile Analysis) จำแนกตามกลุ่มความเสี่ยง

ตัวแปร	ปกติ (n = 1,046)	เสี่ยงชัด (n = 59)	เสี่ยงแฝง (n = 56)	p-value
อายุ (Mean±SD)	40.11±17.40	22.78±15.74	21.36±13.54	<0.001
Alcohol (Yes, %)	31.36	30.51	48.21	0.030
Drug (Yes, %)	35.85	45.76	55.36	0.005
Rage (Yes, %)	27.06	52.54	48.21	<0.001
Health Problem (Yes, %)	1.34	33.90	7.14	<0.001
Mental Problem (Yes, %)	14.34	47.46	14.29	<0.001
Gambling Addict (Yes, %)	1.82	6.78	5.36	0.011
Economics Stress (Yes, %)	11.76	28.81	23.21	<0.001

เปอร์เซ็นไทล์ 95 เป็นพารามิเตอร์ที่ให้คุณภาพการจัดกลุ่มเสี่ยงแฝงดีที่สุด โดยผ่านการควบคุมอัตราการค้นพบที่ผิดพลาด เพื่อความน่าเชื่อถือของผล [28]

3.3 การเปรียบเทียบแบบจำลองไฮบริดกับแบบจำลองพื้นฐาน

จากการประเมินคุณภาพการจัดกลุ่มของแบบจำลองโดยใช้ดัชนีซิลูเอท ดัชนีคาลินสกี้-ฮาราบาสซ์ และค่าดัชนีเดวีส์-โบลดิน เพื่อวัดคุณภาพเชิงโครงสร้างตามแนวทางของ Field [26] ดังตารางที่ 2 พบว่าแบบจำลอง GAAL แสดงคุณภาพการจัดกลุ่มสูงสุด โดยให้ค่าดัชนีซิลูเอทเท่ากับ 0.0825 และค่าดัชนีเดวีส์-โบลดินเท่ากับ 2.8478 ซึ่งเป็นค่าต่ำที่สุดในบรรดาแบบจำลองทั้งหมด แสดงถึงโครงสร้างการจัดกลุ่มที่มีความชัดเจนและความกระชับภายในกลุ่มที่ดีกว่าแบบจำลองอื่น ทั้งนี้ ค่าดัชนีซิลูเอท ในช่วง 0.0–0.2 ถือเป็นลักษณะที่พบได้ทั่วไปในข้อมูลมิติสูงที่ไม่มีป้ายกำกับ [25] ซึ่งสอดคล้องกับงานวิจัยด้านการตรวจจับความผิดปกติในข้อมูลเชิงสังคมที่รายงานค่าดัชนีซิลูเอท ในช่วงใกล้เคียงกัน [21] การพิจารณาจึงควรเปรียบเทียบเชิงสัมพัทธ์ระหว่างแบบจำลองในชุดข้อมูลเดียวกันมากกว่าการใช้ค่าสัมบูรณ์เป็นเกณฑ์ตัดสิน

เมื่อเปรียบเทียบกับแบบจำลองมาตรฐาน เช่น ไอโซเลชัน-ฟอเรสต์ (Isolation Forest) และวันคลาส เอสวีเอ็ม (One-Class SVM) ดังแสดงในตารางที่ 3 พบว่าแบบจำลองทั้งสองมีค่าดัชนีเดวีส์-โบลดิน สูงกว่าอย่างมีนัยสำคัญ (5.3229 และ 8.4708 ตามลำดับ) สะท้อนให้เห็นว่าโครงสร้างกลุ่มมีความซ้ำซ้อนและไม่ชัดเจนเท่ากับแบบจำลองไฮบริด GAAL ผลการเปรียบเทียบชี้ให้เห็นว่าแนวทางแบบไฮบริดที่ผสานจุดแข็งของทั้งสามอัลกอริทึมสามารถรักษาสมดุลระหว่างความเข้มงวดของการตรวจจับและความเสถียรของโครงสร้างกลุ่มได้ดีกว่าแบบจำลองเดี่ยว ซึ่งสอดคล้องกับแนวทางของงานวิจัยที่ยืนยันศักยภาพของการประยุกต์ใช้แบบจำลองไฮบริดในการวิเคราะห์ข้อมูลเชิงสังคม [24], [32] และหลักการบูรณาการองค์ประกอบหลายส่วนเพื่อลดข้อจำกัดของแต่ละกลไก โดยเฉพาะงานของ Tuntitippawan และ Asawarungsaengkul [33] ที่บูรณาการกลไกหน่วยความจำ (Memory Component) เข้ากับอัลกอริทึมเชิงฝูงอัจฉริยะ (Artificial Bee Colony) เพื่อเพิ่มความสามารถในการค้นหาคำตอบที่เหมาะสมในปัญหาการจัดเส้นทาง ซึ่งแสดงให้เห็นว่าการผสานองค์ประกอบที่มีจุดแข็งต่างกันสามารถเอาชนะข้อจำกัดของแบบจำลองเดี่ยวได้ แม้วัตถุประสงค์การประยุกต์ใช้จะแตกต่างกัน กล่าวคืองานดังกล่าวมุ่งแก้ปัญหาการจัดเส้นทางเชิงวิศวกรรม ในขณะที่งานวิจัยนี้มุ่งตรวจจับกลุ่มเสี่ยงเชิงสังคม แต่หลักการพื้นฐานของการรวมจุดแข็งเพื่อลดจุดอ่อนเป็นแนวคิดร่วมกัน

นอกจากนี้เพื่อให้เข้าใจโครงสร้างของกลุ่มข้อมูลที่จำแนกได้อย่างชัดเจนยิ่งขึ้น ได้มีการแสดงผลเชิงภาพด้วยเทคนิคการลดมิติข้อมูลแบบที-เอสเอ็นอี ซึ่งช่วยให้มองเห็นการกระจายตัวของข้อมูลในมิติต่ำ (2 มิติ) โดยพิจารณา

ตารางที่ 2 การเปรียบเทียบคุณภาพการจัดกลุ่มของแบบจำลองไฮบริดกับแบบจำลองพื้นฐาน

แบบจำลอง	ชื่อย่อ	วิธีรวม	Sil	CH	DBI
GMM	GMM	-	0.0702	87.9922	3.4828
AE	AE	-	0.1005	10.0067	4.8877
LOF	LOF	-	0.0655	6.4513	5.9416
GMM+AE	GAAE	AND	0.0731	7.9626	2.9938
GMM+AE	GAEOR	OR	0.0523	63.1402	4.0316
GMM+AE	GAEWS	Weighted Score	0.0731	7.9626	2.9938
GMM+AE	GAEW	Weighted	0.0731	7.9626	2.9938
GMM+LOF	GALOF	AND	0.0532	5.2782	3.5928
GMM+LOF	GALOFOR	OR	0.0605	74.5537	3.7171
GMM+LOF	GALOFWS	Weighted Score	0.0532	5.2782	3.5928
GMM+LOF	GALOFW	Weighted	0.0532	5.2782	3.5928
AE+LOF	AELOF	AND	0.0992	3.4572	4.6417
AE+LOF	AELOFOR	OR	0.0800	11.1368	5.8660
AE+LOF	AELOFWS	Weighted Score	0.0992	3.4572	4.6417
AE+LOF	AELOFW	Weighted	0.0992	3.4572	4.6417
GMM+AE+LOF	GAAL	AND	0.0825	2.8043	2.8478
GMM+AE+LOF	GAALOR	OR	0.0467	57.3455	4.1621
GMM+AE+LOF	GAALWS	Weighted Score	0.0825	2.8043	2.8478
GMM+AE+LOF	GAALW	Weighted	0.0610	9.5454	3.5139

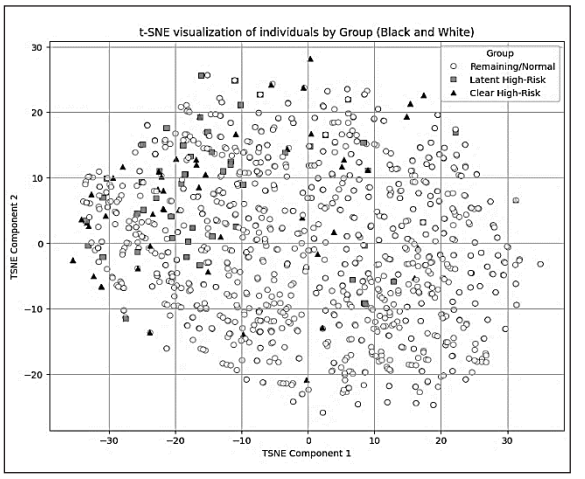
ตารางที่ 3 การเปรียบเทียบแบบจำลองไฮบริดกับแบบจำลองพื้นฐาน

แบบจำลอง	จำนวนกลุ่มเสี่ยงแฝง	Silhouette	CH	DBI
GAAL	56	0.0825	2.8043	2.8478
Isolation Forest	57	0.0650	8.0910	5.3229
One-Class SVM	56	0.0907	3.4425	8.4708

จากความไม่ทับซ้อนของกลุ่มข้อมูลในพื้นที่ 2 มิติ เป็นหลักฐานเชิงภาพประกอบการตีความ [27] พบว่ากลุ่มเสี่ยงมีการกระจุกตัวในบริเวณเฉพาะของปริภูมิแฝง (Latent Space) ซึ่งหมายถึงพื้นที่เชิงนามธรรมที่ออโตเอ็นโค้ดเดอร์สร้างขึ้นเพื่อแสดงโครงสร้างข้อมูลในมิติที่ต่ำลง ในขณะที่กลุ่มปกติกระจายตัวกว้างกว่า แสดงให้เห็นว่าแบบจำลองไฮบริดสามารถสร้างขอบเขตการจัดกลุ่มที่มีความแตกต่างเชิงโครงสร้างภายในข้อมูลได้จริง

3.4 การอภิปรายผลการทดลอง

ผลการทดลองชี้ให้เห็นว่าแบบจำลองไฮบริด GAAL ให้คุณภาพการจัดกลุ่มเสี่ยงแฝงสูงสุด โดยการผสานจุดแข็งของแบบจำลองส่วนผสมเกาส์เซียนที่ตรวจจับความผิดปกติเชิงสถิติจากการกระจายตัวของข้อมูลในภาพรวม ออโตเอ็นโค้ดเดอร์ที่เรียนรู้โครงสร้างเชิงลึกเพื่อระบุความเบี่ยงเบนเชิงรูปแบบ และตัวประกอบค่าผิดปกติเฉพาะที่ที่วิเคราะห์ความหนาแน่นเฉพาะที่สำหรับการระบุค่าผิดปกติในระดับจุลภาค



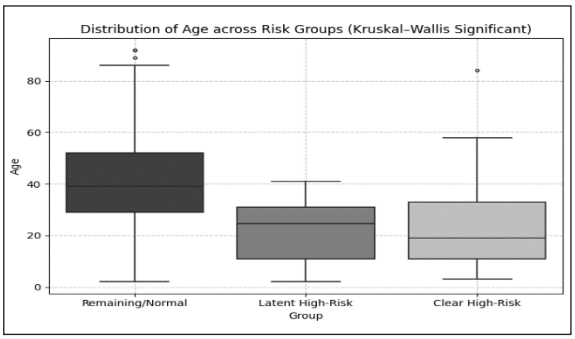
รูปที่ 5 การกระจายตัวของกลุ่มด้วยเทคนิคการลดมิติข้อมูลแบบที-เอสเอ็นอี

ทำให้สามารถวิเคราะห์ความผิดปกติได้ครอบคลุมหลายมิติพร้อมกัน การใช้ตรรกะและเพิ่มความเข้มงวดในการคัดกรองและลดผลบวกเทียมได้อย่างมีนัยสำคัญ แม้ค่าดัชนีซิลูเอทจะอยู่ในช่วงที่พบได้ทั่วไปสำหรับข้อมูลมิติสูงที่ไม่มีป้ายกำกับ แต่ค่าดัชนีเดวีส์-โบลดินที่ต่ำกว่า ไอโซเลชันฟอเรสต์และวันคลาส เอสวีเอ็ม อย่างเด่นชัด สะท้อนว่าโครงสร้างการจัดกลุ่มของ GAAL มีความชัดเจนและเสถียรกว่าอย่างมีนัยสำคัญ ทั้งนี้ แม้การทดสอบครัสคัล-วัลลิส จะเป็นสถิตินอนพารามิเตอร์ (Non-parametric Statistics) ที่เหมาะสมกับข้อมูลที่ไม่เป็นไปตามการแจกแจงปกติ แต่มีข้อจำกัดสำคัญที่ควรพิจารณา กล่าวคือวิธีการนี้มีกำลังทางสถิติ (Statistical Power) ต่ำกว่าการทดสอบแบบพาราเมตริก (Parametric Test) เมื่อข้อมูลเข้าใกล้การแจกแจงปกติ อีกทั้งทดสอบเพียงความแตกต่างของการกระจายตัวโดยรวมระหว่างกลุ่ม โดยไม่ระบุว่ากลุ่มใดแตกต่างจากกลุ่มใด ดังนั้นผลการทดสอบในงานวิจัยนี้จึงควรตีความเป็นหลักฐานเสริมประกอบกับดัชนีเชิงโครงสร้างภายในไม่ใช่เป็นการยืนยันผลการจัดกลุ่มแบบเบ็ดเสร็จ [31]

การแสดงผลด้วยเทคนิคการลดมิติข้อมูลแบบที-เอสเอ็นอี ดังแสดงในรูปที่ 5 ยืนยันเชิงภาพถึงความสามารถของแบบจำลองในการจัดกลุ่มเสี่ยงพบว่ากลุ่มปกติ (Remaining/Normal) (วงกลม) มีการกระจายตัวหนาแน่นบริเวณศูนย์กลาง ขณะที่กลุ่มเสี่ยงสูงแฝง (Latent High-Risk) (สี่เหลี่ยม) และกลุ่มเสี่ยงสูงที่ชัดเจน (Clear High-Risk) (สามเหลี่ยม) ปรากฏแยกออกจากกลุ่มหลักอย่างชัดเจน โดยเฉพาะกลุ่มเสี่ยงแฝงที่อยู่บริเวณขอบของกลุ่มใหญ่ สะท้อนลักษณะการเปลี่ยนผ่านเชิงพฤติกรรมที่สอดคล้องกับแนวคิดการวิเคราะห์เชิงแฝงในบริบทความรุนแรงในครอบครัว [18] ผลดังกล่าวยืนยันว่าแนวทางแบบสองขั้นตอนช่วยให้กระบวนการค้นหากลุ่มเสี่ยงแฝงที่มีลักษณะซับซ้อนมีความแม่นยำและเสถียรกว่าการใช้แบบจำลองเดี่ยว ซึ่งสอดคล้องกับงานวิจัยที่ยืนยันว่าการนำปัญญาประดิษฐ์ (Artificial Intelligence; AI) และการวิเคราะห์ข้อมูลเชิงพฤติกรรมมาประยุกต์ใช้ร่วมกับตัวแปรด้านสุขภาพจิต ปัจจัยทางสังคม สิ่งแวดล้อม และการจัดการข้อมูลขนาดใหญ่สามารถยกระดับความสามารถของระบบป้องกันความรุนแรงได้อย่างมีนัยสำคัญ [10], [12], [34], [35] และมีคุณค่าต่อการพัฒนากลไกเฝ้าระวังเชิงดิจิทัลในระดับชุมชนต่อไป [9], [33]

4. สรุปผลการทดลอง

ผลการทดลองแสดงให้เห็นว่าแบบจำลองไฮบริดแบบไม่มีผู้สอนภายใต้แนวทางแบบจำลองไฮบริดแบบสองขั้นตอน ซึ่งผสานการทำงานของ แบบจำลองส่วนผสมเกาส์เซียน ออโตเอ็นโค้ดเดอร์ และตัวประกอบค่าผิดปกติเฉพาะที่สามารถเพิ่มคุณภาพเชิงโครงสร้างของการจัดกลุ่มและความชัดเจนของการแยกคลัสเตอร์ของกลุ่มเสี่ยงต่อความรุนแรงในครอบครัวได้อย่างมีนัยสำคัญ โดยเฉพาะในขั้นตอนที่สอง ซึ่งมุ่งเน้นการตรวจจับกลุ่มเสี่ยงสูงแฝง แบบจำลอง GAAL ให้ค่าดัชนีซิลูเอทสูงสุดและค่า DBI ต่ำสุด แสดงถึงความชัดเจนของโครงสร้างกลุ่มและความเสถียรของโครงสร้างการจัดกลุ่มที่เหนือกว่าแบบจำลองเดี่ยวและแบบจำลองมาตรฐาน เช่น ไอโซเลชันฟอเรสต์ และวันคลาส เอสวีเอ็ม แสดงในตารางที่ 3 ผลการวิเคราะห์ด้วยการทดสอบครัสคัล-วัลลิส พบว่าตัวแปรอายุ (Age) มีความแตกต่างระหว่างกลุ่มอย่างมีนัยสำคัญทางสถิติ (p-value < 0.05) ดังแสดงในรูปที่ 6 ซึ่งยืนยันว่าเป็นปัจจัยสำคัญในคุณภาพ



รูปที่ 6 การกระจายของอายุระหว่างกลุ่มความเสี่ยง

การจัดกลุ่มเสี่ยงโดยกลุ่มเสี่ยงสูงแฝงและกลุ่มเสี่ยงสูงชัดเจนมีอายุต่ำกว่ากลุ่มปกติอย่างมีนัยสำคัญ

โดยสรุปแบบจำลองที่พัฒนาขึ้นนี้สามารถประยุกต์ใช้เป็นเครื่องมือคัดกรองและประเมินความเสี่ยงได้อย่างมีความน่าเชื่อถือและเสถียร อีกทั้งยังมีศักยภาพในการต่อยอดสู่การพัฒนาระบบเฝ้าระวังเชิงพฤติกรรมในระดับชุมชนหรือประเทศเพื่อสนับสนุนการตัดสินใจเชิงนโยบายและการดำเนินมาตรการเชิงป้องกันอย่างทันท่วงที แนวทางดังกล่าวจึงเป็นพื้นฐานสำคัญสำหรับการประยุกต์ใช้ ปัญญาประดิษฐ์เพื่อสังคม (AI for Social Good) ในการลดและป้องกันปัญหาความรุนแรงในครอบครัวอย่างยั่งยืน อย่างไรก็ตาม การประเมินผลในงานวิจัยนี้พึ่งพาตัวชี้วัดเชิงโครงสร้างภายใน เป็นหลัก ซึ่งไม่สามารถยืนยันความถูกต้องเชิงเนื้อหาได้โดยตรง ดังนั้นการนำไปใช้จริงควรมีการตรวจสอบร่วมกับผู้เชี่ยวชาญด้านสังคมสงเคราะห์หรือจิตวิทยา

5. กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จได้ด้วยความอนุเคราะห์และการสนับสนุนอันทรงคุณค่าจากหลายภาคส่วน คณะผู้วิจัยขอขอบพระคุณกรมกิจการสตรีและสถาบันครอบครัว กระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ สำหรับการจัดทำและเผยแพร่ชุดข้อมูลผู้กระทำความรุนแรงในครอบครัวผ่านระบบบัญชีข้อมูลภาครัฐ ซึ่งเป็นข้อมูลตั้งต้นในการวิเคราะห์และพัฒนาระบบจำแนกกลุ่มเสี่ยงในงานนี้ อีกทั้งขอขอบคุณนักวิจัยและเจ้าของผลงานวิชาการทุกท่านที่เป็นแรงบันดาลใจและรากฐานทางองค์ความรู้ ตลอดจนวารสารวิชาการพระจอมเกล้าพระนครเหนือที่ให้เกียรติพิจารณาและมอบโอกาสเผยแพร่ผลงาน ซึ่งคณะผู้วิจัยหวังเป็นอย่างยิ่งว่างานวิจัยนี้จะเป็นประโยชน์ต่อการพัฒนากลยุทธ์ในการป้องกันและรับมือปัญหาความรุนแรงในครอบครัวต่อไป

เอกสารอ้างอิง

[1] World Health Organization. (2021). Violence against women: Intimate partner and sexual violence against women. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/violence-against-women

[2] A. T. Vazsonyi, D. J. Flannery, and M. DeLisi, Eds., *The Cambridge Handbook of Violent Behavior and Aggression*, 2nd ed. Cambridge, UK: Cambridge University Press, 2018.

[3] World Health Organization. (2025, Nov.). Lifetime toll: 840 million women faced partner or sexual violence. [Online]. Available: https://www.who.int/news/item/19-11-2025-lifetime-toll--840-million-women-faced-partner-or-sexual-violence

[4] United Nations Office on Drugs and Crime and UN Women. (2024). Femicides in 2023: Global estimates of intimate partner/family member femicides. [Online]. Available: https://www.unwomen.org/en/digital-library/publications/2024/11/femicides-in-2023-global-estimates-of-intimate-partner-family-member-femicides.

[5] N. Duvvury, C. Grown, and J. Redner, "Estimating the economic costs of violence against women," *Washington, DC: World Bank*, 2013.

[6] Department of Women's Affairs and Family Development. (2025). Domestic violence

situation report 2023. [Online]. Available: https://www.dwf.go.th/storage/91029/20c2b8ee-0623-43f1-b579-cd5e19664c70-document-16358.pdf (in Thai)

[7] Department of Women's Affairs and Family Development. (2025). Domestic violence situation, 2024. [Online]. Available: https://www.dwf.go.th/contents/71825 (in Thai).

[8] Department of Women's Affairs and Family Development. (2022). Report on the domestic violence situation in Thailand under section 17 of the domestic violence victim protection act B.E. 2550. [Online]. Available: https://dwf.go.th/contents/48156 (in Thai).

[9] P. Thooltham and M. Nakham, "The construction of social networks to tackle domestic violence: A case analysis of Huay Sam Pad Subdistrict, Thailand," *Humanities, Arts and Social Sciences Studies*, vol. 25, no. 1, pp. 54–66, 2025 (in Thai), doi: 10.69598/hasss.25.1.267448.

[10] F. Başaran and P. Duru, "Determining domestic violence against women using machine learning methods: The case of Türkiye," *Journal of Evaluation in Clinical Practice*, vol. 31, no. 3, Art. no. e14180, Apr. 2025, doi: 10.1111/jep.14180.

[11] K. de Boer, J. L. Mackelprang, M. Nedeljkovic, D. Meyer, and R. Iyer, "Using artificial intelligence to detect risk of family violence: Protocol for a systematic review and meta-analysis," *JMIR Research Protocols*, vol. 13, Dec. 2024, Art. no. e54966, doi: 10.2196/54966.

[12] M. C. Cruz-Mendoza, R. A. Melendez-Armenta, J. Canul-Reich, and J. Muñoz-Benítez, "Machine learning applied to improve prevention of, response to, and understanding of violence against women," *Informatics*, vol. 12, no. 2, 2025, Art. no. 40, doi: 10.3390/informatics12020040.

[13] A. Oluwasegun and J.-C. Jung, "A multivariate Gaussian mixture model for anomaly detection in transient current signature of control element drive mechanism," *Nuclear Engineering and Design*, vol. 402, 2023, Art. no. 112098, doi: 10.1016/j.nucengdes.2022.112098.

[14] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *Proceedings 2nd Workshop on Machine Learning for Sensory Data Analysis*, Gold Coast, Australia, 2014, pp. 4–11, doi: 10.1145/2689746.2689747.

[15] O. Alghushairy, R. Alsini, T. Soule, and X. Ma, "A review of local outlier factor algorithms for outlier detection in big data streams," *Big Data and Cognitive Computing*, vol. 5, no. 1, Jan. 2021, Art. no. 1, doi: 10.3390/bdcc5010001.

[16] N. Saini, V. B. Kasaragod, K. Prakasha, and A. K. Das, "A hybrid ensemble machine learning model for detecting APT attacks based on network behavior anomaly detection," *Concurrency and Computation: Practice and Experience*, vol. 35, 2023, Art. no. e7865, doi: 10.1002/cpe.7865.

[17] S. Nazat, W. Alayed, L. Li, and M. Abdallah, "Ensemble learning framework for anomaly detection in autonomous driving systems," *Sensors*, vol. 25, no. 16, 2025, Art. no. 5105, doi: 10.3390/s25165105.

[18] A. Restrepo, N. Montoya, and L. Zuluaga, "Typologies of intimate partner violence

against women in five Latin-American countries: A latent class analysis," *International Journal of Public Health*, vol. 67, 2022, Art. no. 1604000, doi: 10.3389/ijph.2022.1604000.

[19] Ministry of Social Development and Human Security. (2023). Domestic violence offenders classified by issues or causes. Government Data Catalog (MSDHS). [Online]. Available: https://gdcatalog.m-society.go.th/dataset/dwf050507_pb_dmv01_03 (in Thai).

[20] Ministry of Social Development and Human Security. (2024). Domestic violence offenders classified by issues or causes. *Government Data Catalog (MSDHS)*. [Online]. Available: https://gdcatalog.m-society.go.th/dataset/dwf050507_pb_dmv01_04 (in Thai).

[21] A. Dudek, "Silhouette index as clustering evaluation tool," in *Classification and Data Analysis: Theory and Applications, K. Jajuga, J. Batóg, and M. Walesiak, Eds. Cham*, Switzerland: Springer, 2020, pp. 19–33, doi: 10.1007/978-3-030-52348-0_2.

[22] P. Patel, B. Sivaiah, and R. Patel, "Approaches for finding optimal number of clusters using K-Means and agglomerative hierarchical clustering techniques," *2022 International Conference on Intelligent Controller and Computing for Smart Power, Hyderabad, India*, 2022, pp. 1–6, doi: 10.1109/ICICCSP53532.2022.9862439.

[23] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979, doi: 10.1109/TPAMI.1979.4766909.

[24] N. Senawong, S. Wichitchan, and O. Kumphon, "Large-scale data classification based on K-means clustering and deep learning," *Journal of King Mongkut's University of Technology North Bangkok*, vol. 32, no. 4, pp. 957–965, 2022 (in Thai), doi: 10.14416/j.kmutnb.2021.03.012.

[25] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.

[26] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, 5th ed. London, UK: SAGE Publications, 2021.

[27] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008

[28] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society: Series B*, vol. 57, no. 1, pp. 289–300, 1995, doi: 10.1111/j.2517-6161.1995.tb02031.x.

[29] A. Ben-Hur, A. Elisseeff, and I. Guyon, "A stability based method for discovering structure in clustered data," *Pacific Symposium on Biocomputing*, vol. 7, 2002, pp. 6–17, doi: 10.1142/9789812799623_0002.

[30] A. Ghasemi and S. Zahediasl, "Normality tests for statistical analysis: A guide for non-statisticians," *International Journal of Endocrinology and Metabolism*, vol. 10, no. 2, pp. 486–489, 2012, doi: 10.5812/ijem.3505.

[31] P. Mishra, C. M. Pandey, U. Singh, A. Gupta, C.

Sahu, and A. Keshri, "Descriptive statistics and normality tests for statistical data," *Annals of Cardiac Anaesthesia*, vol. 22, no. 1, pp. 67–72, Jan. 2019, doi: 10.4103/aca.ACA_157_18.

[32] T. Sutthison and P. Pienkhawsook, "Hybrid model for forecasting time series data of monthly household electrical distribution units in Thailand," *Journal of King Mongkut's University of Technology North Bangkok*, vol. 34, no. 2, pp. 1–18, 2024 (in Thai), doi: 10.14416/j.kmutnb.2024.02.003.

[33] N. Tuntitippawan and K. Asawarungsaengkul, "A memory integrated artificial bee colony algorithm with local search for vehicle routing problem with backhauls and time windows," *KMUTNB International Journal of Applied Science and Technology*, vol. 11, no. 2, pp. 85–92, 2018, doi: 10.14416/j.ijast.2018.03.001.

[34] P. Kaewkerd and P. Thammakorn, "Big data and business operations in the digital age," *Journal of Business and Industrial Development*, vol. 3, no. 1, pp. 92–99, 2023 (in Thai), doi: 10.14416/j.bid.2023.04.008.

[35] K. Allen, G. J. Melendez-Torres, T. Ford, C. Bonell, and V. Berry, "Parental domestic violence and abuse, mental ill-health, and substance misuse and the impact on child mental health: A secondary data analysis using the UK Millennium Cohort Study," *BMC Public Health*, vol. 24, 2024, Art. no. 2310, doi: 10.1186/s12889-024-19694-1.