

การศึกษาการแนะนำสินค้าในร้านค้าปลีกแบบออฟไลน์โดยวิธีการกรองแบบร่วมกัน กรณีศึกษาบริษัทเอกชนแห่งหนึ่ง

จิตรานุช โล่ห์ซึ่งชัยฤทธิ์*

คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

เอกรัฐ รัฐกาญจน์

คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

*Correspondence: jitranut.loc@stu.nida.ac.th

doi: 10.14456/jisb.2021.14

วันที่รับบทความ: 10 พ.ย. 2564

วันแก้ไขบทความ: 27 พ.ย. 2564

วันที่ตอบรับบทความ: 9 ธ.ค. 2564

บทคัดย่อ

ความหลากหลายของสินค้าที่ลูกค้าซื้อมีความสำคัญสำหรับผู้ค้าปลีก เมื่อลูกค้าซื้อสินค้าชนิดใดชนิดหนึ่งเป็นจำนวนมาก ลูกค้าอาจจะได้ประโยชน์ เช่น ส่วนลดเพิ่มเติม ในทางกลับกันอาจเป็นผลเสียต่อผู้ค้าปลีกเนื่องจากมีสินค้าเพียงบางชนิดที่มีอัตราการหมุนเวียนของสินค้าสูง ด้วยเหตุนี้ผู้ค้าปลีกรายใหญ่จึงพยายามดึงดูดลูกค้าให้ซื้อสินค้าหลากหลายรูปแบบ อย่างไรก็ตามการแนะนำสินค้าสำหรับร้านค้าปลีกรายย่อยแบบออฟไลน์นั้นค่อนข้างเป็นไปได้ยากเนื่องจากไม่มีข้อมูลเกี่ยวกับความชอบของลูกค้า งานวิจัยนี้ใช้ข้อมูลการซื้อของร้านค้าปลีกแบบออฟไลน์เป็นเวลาสี่ปีเพื่อหาการให้เรตติ้งความชอบจากพฤติกรรมการซื้อและสร้างระบบแนะนำสินค้าโดยใช้วิธีการกรองแบบร่วมกันสองวิธีคือ วิธีการกรองข้อมูลร่วมแบบจดจำ โดยอัลกอริทึม K-Nearest Neighbor และวิธีการกรองข้อมูลร่วมแบบจำลอง โดยการใช้วิธี SVD matrix Factorization เปรียบเทียบประสิทธิภาพของอัลกอริทึมด้วยค่า Root mean square error (RMSE) โดยตัวแบบได้ถูกกำหนดค่าพารามิเตอร์ที่เหมาะสมที่สุดไว้ จากผลการวิจัยพบว่า วิธีการกรองข้อมูลร่วมแบบจดจำที่พิจารณาผู้ใช้เป็นหลักเป็นวิธีที่มีประสิทธิภาพที่สุดเมื่อนำไปจัดอันดับการแนะนำและวัดผลด้วยค่า Mean precision average และเมื่อแนะนำรายการสินค้ามากขึ้นแล้วระดับความแม่นยำจะลดลงเรื่อยๆ

คำสำคัญ: ระบบแนะนำ; การกรองแบบร่วมกัน; คำปลีกแบบออฟไลน์; คัดกรองผู้ใช้; คัดกรองสิ่งของ

A Study of Product Recommendation in Offline Retail Store by using Collaborative Filtering: A Case Study of a Private Company

Jitranut Lochingchairit*

Faculty of Applied Statistics, National Institute of Development Administration

Ekarat Rattagan

Faculty of Applied Statistics, National Institute of Development Administration

*Correspondence: Jitranut.loc@stu.nida.ac.th

doi: 10.14456/jisb.2021.14

Received: 10 Nov 2021

Revised: 27 Nov 2021

Accepted: 9 Dec 2021

Abstract

Diversity of purchased products is important for retailers. When a customer purchases a specific product in large volume, the customer might get some benefit, such as more discounts. In contrast, this might be disadvantage for retailers because only some products have high inventory turnover. Therefore, many large retailers try to motivate their customers to purchase various products. However, product recommendation in offline retail store is quite challenging because of the unavailability of information regarding customers' preferences. This study uses four years of purchase transaction data to compute implicit rating of customers from frequency of purchasing. The data was analyzed by employs two methods of Collaborative Filtering which are Memory-based filtering with K-Nearest Neighbor and Model-based filtering with SVD Matrix Factorization. The results of analysis were compared by using Root mean square error (RMSE) with the appropriate hyper parameter. The result shows that Memory-based filtering with K-Nearest Neighbor algorithm is the most efficient. The results of employing the recommend model indicate that the more items are introduced, the accuracy level will gradually decrease.

Keyword: Recommender system; Collaborative filtering; Offline retail store; User-based filtering; Item-based filtering

1. บทนำ

1.1 ความสำคัญและที่มาของปัญหาการวิจัย

ระบบแนะนำเป็นระบบที่ใช้การเก็บรวบรวมสารสนเทศจากความพึงพอใจของผู้ใช้ในตัวสินค้าและบริการที่สนใจ เพื่อทำนายสิ่งที่ผู้ใช้สนใจในตัวสินค้าหรือบริการจากข้อมูลที่มีความสัมพันธ์กัน ระบบนี้ทำให้บริษัทต่างๆ สามารถพัฒนากลยุทธ์ทางการตลาด ดึงดูดลูกค้าได้มากขึ้น และเพิ่มยอดขาย ดังนั้นหลายบริษัทจึงพยายามใช้ระบบแนะนำเพื่อประโยชน์ทางธุรกิจ ระบบแนะนำได้ถูกนำไปใช้ในหลากหลายอุตสาหกรรม เช่น เพลง ภาพยนตร์ หนังสือ การท่องเที่ยว เว็บไซต์พาณิชย์อิเล็กทรอนิกส์ (E-commerce) ห้องสมุดอิเล็กทรอนิกส์ และระบบสื่อการสอนออนไลน์ (E-learning)

ถึงแม้ว่าระบบแนะนำจะถูกนำไปใช้อย่างแพร่หลายในตลาดอีคอมเมิร์ซ แต่ก็สามารถนำระบบแนะนำไปใช้กับร้านค้าปลีกแบบดั้งเดิมได้เช่นกัน ซึ่งระบบแนะนำเป็นส่วนหนึ่งในการทำการตลาดแบบเฉพาะบุคคล (Personalization) เป็นกลยุทธ์ทางการตลาดที่สำคัญในการเพิ่มการรักษาลูกค้าและความภักดีต่อแบรนด์ (Pratama et al., 2020) การตลาดแบบเฉพาะบุคคลเป็นวิธีการที่นำเสนอโฆษณาและข้อเสนอที่เฉพาะเจาะจงรายบุคคลแก่ลูกค้าแต่ละรายเพื่อสร้างความสนใจและความเกี่ยวข้องของลูกค้าแต่ละบุคคล โดยใช้ข้อมูลรายละเอียดและความชอบของลูกค้าแต่ละราย ข้อมูลเหล่านี้สามารถบ่งบอกความต้องการของลูกค้าแต่ละรายอย่างเจาะจงได้ ซึ่งร้านค้าปลีกสามารถดำเนินกลยุทธ์การตลาดแบบเฉพาะบุคคลได้โดยการใช้ระบบแนะนำ ที่ผ่านมาอีคอมเมิร์ซได้ใช้ระบบคำแนะนำที่มีประโยชน์มากมาย เช่น การเพิ่มระดับการโต้ตอบกับลูกค้า การเพิ่มยอดขาย ความหลากหลายของสินค้าที่ขาย ความพึงพอใจหรือความภักดีของลูกค้า และยังเข้าใจความต้องการของลูกค้าได้ดีขึ้น ซึ่งคาดว่าในร้านค้าปลีกแบบดั้งเดิมหรือแบบออฟไลน์จะได้รับประโยชน์ดังกล่าวเช่นกัน

ร้านค้าปลีกแบบดั้งเดิมแตกต่างจากอีคอมเมิร์ซหลายประการ ประการแรกคือร้านค้าปลีกแบบดั้งเดิมยังคงมีหน้าร้านจริงสำหรับการจัดเก็บและการจัดวางสินค้า ทำให้เกิดค่าใช้จ่ายสำหรับการจัดเก็บสินค้าที่มีอัตราการหมุนเวียนของสินค้าต่ำ (Low inventory turnover) ในทางกลับกันความหลากหลายในความต้องการของลูกค้ากลับเพิ่มขึ้นเรื่อยๆ ผู้ค้าปลีกต้องสามารถระบุความต้องการของลูกค้าได้อย่างถูกต้องและนำเสนอผลิตภัณฑ์ที่หลากหลาย เพื่อหมุนเวียนสินค้าในคลังให้เคลื่อนไหวอยู่เสมอ ซึ่งแตกต่างจากการขายแบบอีคอมเมิร์ซที่ไม่จำเป็นต้องมีคลังสินค้า ข้อแตกต่างประการที่สองคือร้านค้าปลีกออฟไลน์สามารถเข้าถึงประวัติการซื้อและสังเกตพฤติกรรมการซื้อของลูกค้าได้ยาก ในบางกรณีร้านค้าออฟไลน์เก็บธุรกรรมโดยไม่รู้ว่ามีผู้ซื้อเป็นใคร ซึ่งแตกต่างจากร้านค้าออนไลน์ซึ่งข้อมูลประจำตัวของลูกค้าและกิจกรรมจะถูกบันทึกไว้ในระบบ อีกทั้งร้านค้าปลีกออฟไลน์เก็บความคิดเห็นเกี่ยวกับสินค้าที่ลูกค้าซื้อได้ยาก ผลที่ตามมาคือการทำนายของระบบแนะนำสินค้ามักจะไม่แม่นยำต่ำเมื่อใช้ระบบแนะนำสำหรับสินค้าที่ไม่ค่อยได้รับเรตติ้ง (Walter et al., 2012) ซึ่งแตกต่างจากธุรกิจอีคอมเมิร์ซที่ระบบจะขอให้ลูกค้าให้เรตติ้งสินค้า การให้เรตติ้งดังกล่าวเป็นสิ่งสำคัญสำหรับการศึกษาระบบแนะนำ เนื่องจากเรตติ้งของลูกค้ากลายเป็นแหล่งข้อมูลสามารถแนะนำสินค้าไปยังลูกค้ารายอื่นได้

จากข้อมูลข้างต้นผู้วิจัยจึงต้องการศึกษาอัลกอริทึมที่จะช่วยร้านค้าปลีกแบบออฟไลน์ให้สามารถนำข้อมูลมาประยุกต์ใช้สร้างระบบแนะนำสินค้า โดยผู้วิจัยจะใช้ข้อมูลประวัติการซื้อขายตั้งแต่ปี 2560 ถึง 2564 จำนวน 38,282 รายการจากระบบจัดการหน้าร้านของบริษัทจำหน่ายอุปกรณ์ไฟฟ้าและแสงสว่างแห่งหนึ่งเป็นกรณีศึกษา เพื่อนำอัลกอริทึมไปประยุกต์ใช้ สร้างความสามารถในการแข่งขัน (Competitiveness advantage) และการสร้างความสัมพันธ์กับลูกค้า (Customer relationship) ซึ่งสามารถนำผลลัพธ์ของอัลกอริทึมที่เหมาะสมไปประยุกต์ใช้จริงกับร้านค้าปลีกแบบออฟไลน์อื่นๆ ได้อีกด้วย

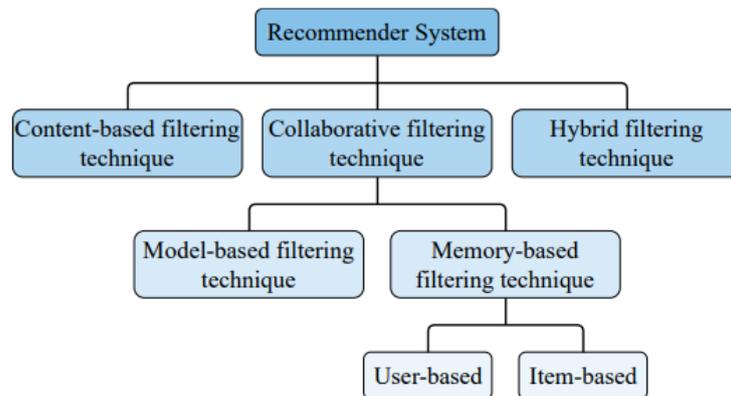
1.2 วัตถุประสงค์

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาตัวแบบสำหรับการแนะนำสินค้า โดยใช้เทคนิคการเรียนรู้เครื่องจักรและเพื่อหาตัวแบบที่เหมาะสมที่สุด โดยเปรียบเทียบประสิทธิภาพของแต่ละอัลกอริทึมในการแนะนำสินค้าสำหรับลูกค้าแต่ละราย เพื่อนำเสนอแนวทางการให้ผู้ประกอบการสามารถนำระบบแนะนำสินค้าไปประยุกต์ใช้ในธุรกิจค้าปลีกแบบออฟไลน์ เพื่อส่งเสริมการขายสินค้า ทำการตลาดแบบเฉพาะบุคคล สร้างความสัมพันธ์กับลูกค้า เพิ่มความสามารถในการแข่งขันและเพิ่มโอกาสในการขายสินค้าอื่นๆ

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ระบบแนะนำ (Recommendation system)

ระบบแนะนำเป็นระบบที่ถูกสร้างขึ้นมาเพื่อช่วยเหลือผู้ใช้ในการคัดกรองสิ่งที่สนใจ หรือชื่นชอบจากสิ่งต่างๆ ที่มีอยู่เป็นจำนวนมาก มีบทบาทสำคัญในการสร้างรายการแนะนำข้อมูลที่คาดเดาว่าผู้ใช้ชื่นชอบหรือให้ความสนใจจากฐานข้อมูลที่มีขนาดใหญ่ ซึ่งช่วยให้ผู้ใช้สามารถเข้าถึงข้อมูลสินค้าหรือบริการที่ชื่นชอบได้อย่างมีประสิทธิภาพ และสามารถเพิ่มผลกำไรและรักษฐานลูกค้าให้กับระบบธุรกิจพาณิชย์อิเล็กทรอนิกส์ (โกเมศ อัมพวัน และคณะ, 2562) อาทิเช่น ระบบแนะนำภาพยนตร์ (Netflix, IMDb) ระบบแนะนำเพลง (Spotify) และระบบแนะนำวิดีโอ (Youtube) โดยวิธีการที่นิยมใช้ในการสร้างรายการแนะนำข้อมูลสามารถแบ่งได้ 3 วิธี ได้แก่ วิธีการกรองเนื้อหา (Content based filtering) วิธีการกรองแบบร่วมกัน (Collaborative filtering) และวิธีการกรองแบบผสม (Hybrid filtering) ดังภาพที่ 1 โดยมีรายละเอียดดังนี้



ภาพที่ 1 ระบบแนะนำประกอบไปด้วยขั้นตอนวิธีการต่าง ๆ

2.1.1 วิธีการกรองเนื้อหา (Content based filtering)

วิธีการกรองเนื้อหาเป็นวิธีการสร้างรายการแนะนำข้อมูล โดยพิจารณาจากความคล้ายคลึงของเนื้อหาหรือคุณสมบัติต่างๆ ของข้อมูลที่คล้ายกันกับข้อมูลของผู้ใช้ในอดีต วิธีการนี้จะให้ความสนใจกับเนื้อหาของข้อมูลเป็นสำคัญ วิธีการกรองเนื้อหาที่นิยมในการสร้างรายการแนะนำ คือ วิธีการหาค่าน้ำหนักของคำสำคัญ หรือ Term frequency/Inverse document frequency (TF-IDF) เป็นวิธีการทางสถิติในการสืบค้นข้อมูลโดยพิจารณาคำสำคัญของเอกสาร (Keyword)

2.1.2 วิธีการกรองแบบร่วมกัน (Collaborative filtering)

วิธีการกรองแบบร่วมกันเป็นวิธีการสร้างรายการแนะนำข้อมูลที่นิยมมากที่สุด โดยวิธีนี้จะนิยมใช้กับระบบแนะนำที่มีการให้คะแนนความชอบ (Rating) ต่อสิ่งของด้วย โดยระบบจะสร้าง User profile เก็บข้อมูลของผู้ใช้แต่ละรายเพื่อใช้ในการคำนวณความคล้ายคลึง แต่อย่างไรก็ตามข้อเสียของวิธีการนี้คือ (1) เมื่อเกิดเหตุการณ์ที่มีผู้ใช้ใหม่ หรือข้อมูล

ใหม่ ผู้ใช้ใหม่ที่เข้าสู่ระบบครั้งแรกจะยังไม่ได้รับการแนะนำและข้อมูลใหม่ก็ยังไม่ถูกแนะนำเช่นกัน เนื่องจากไม่มีข้อมูลในอดีต ทำให้เกิดปัญหา Cold start problem (2) เมื่อข้อมูลการให้เรตติ้งในฐานะข้อมูลมีจำนวนน้อย ทำให้เกิดปัญหาข้อมูลมีปริมาณน้อย (Data sparsity problem) ซึ่งอาจทำให้ผลของระบบแนะนำมีความคลาดเคลื่อน

โดยวิธีการกรองแบบร่วมกันแบ่งออกเป็น 2 ประเภท คือ (1) วิธีการกรองข้อมูลร่วมแบบจำลอง (Model-based filtering) เป็นวิธีการนำข้อมูลของผู้ใช้ทั้งหมดที่มีในฐานะข้อมูล มาสร้างเป็นแบบจำลองสำหรับสร้างรายการแนะนำให้กับผู้ใช้ที่มีลักษณะคล้ายคลึงกัน (2) วิธีการกรองข้อมูลร่วมแบบจดจำ (Memory-based filtering) เป็นวิธีการนำข้อมูลของผู้ใช้แต่ละรายมาเปรียบเทียบกับมีความคล้ายคลึงกับข้อมูลผู้ใช้คนอื่น ๆ ที่มีอยู่ในฐานะข้อมูล ซึ่งถ้าพิจารณาจากความคล้ายคลึงหรือใกล้เคียงกันระหว่างผู้ใช้ในฐานะข้อมูล จะเรียกว่าวิธีการกรองแบบร่วมกันโดยพิจารณาผู้ใช้เป็นหลัก (User-based Collaborative filtering) แต่ถ้าพิจารณาความคล้ายคลึงหรือใกล้เคียงกันระหว่างข้อมูลในฐานะข้อมูล จะเรียกว่าวิธีการกรองข้อมูลร่วมโดยพิจารณาข้อมูลเป็นหลัก (Item-based collaborative filtering)

2.1.3 วิธีการกรองแบบผสม (Hybrid filtering)

วิธีการกรองแบบผสมผสานเป็นการรวมข้อดีของทั้งวิธีการกรองเนื้อหาและวิธีการกรองแบบร่วมกันไว้ด้วยกัน ซึ่งสามารถช่วยในการหลีกเลี่ยงข้อจำกัดของทั้งสองวิธีการได้ ทำให้สามารถสร้างรายการแนะนำสินค้าและคาดว่าผู้ใช้นั้นชอบได้อย่างมีประสิทธิภาพ อย่างไรก็ตามวิธีการนี้ต้องใช้ข้อมูลของสินค้าจำนวนมากในการสร้างรายการแนะนำ อาจทำให้มีการใช้ทรัพยากรและมีความซับซ้อนในการแนะนำที่สูงขึ้น ซึ่งอาจไม่เหมาะสมกับระบบธุรกิจพาณิชย์อิเล็กทรอนิกส์ที่ไม่มีการจัดเก็บคุณสมบัติของข้อมูล

2.2 ประเภทของการกรอง

2.2.1 วิธีการกรองข้อมูลร่วมแบบจำลอง (Model-based)

วิธีการกรองข้อมูลร่วมแบบจำลอง (Model-based) เป็นวิธีการกรองที่พิจารณาการแตกค่าแบบเอกฐาน (Singular value decomposition หรือ SVD) เป็นวิธีหนึ่งใน Matrix factorization โดยวิธีนี้จะใช้เพื่อหา Latent semantic indexing (LSI) ซึ่งใช้ในการหาความสัมพันธ์ภายในข้อมูลและเพื่อลดมิติของข้อมูล โดยวิธี SVD จะสร้างเมตริกซ์ที่มีขนาด $m \times n$ มิติกับอันดับ r โดยสามารถแยกองค์ประกอบได้ดังนี้

$$Z = U \Sigma V^T \quad (1)$$

โดย

U แทนเมตริกซ์ขนาด $m \times r$ โดยที่คอลัมน์เป็นโอเกนเวกเตอร์ตั้งฉาก (Orthogonal eigenvectors) ของ ZZ^T

Σ แทนเมตริกซ์ขนาด $r \times r$ โดยค่าในแนวเส้นทแยงมุมเท่านั้นที่มีค่าไม่เป็นศูนย์แต่เป็นค่า Eigen value ของ $Z^T Z$

V แทนเมตริกซ์ขนาด $n \times r$ โดยมีมิติที่เป็นคอลัมน์เป็นโอเกนเวกเตอร์ตั้งฉากของ $Z^T Z$

2.2.2 วิธีการกรองข้อมูลร่วมแบบจดจำ (Memory-based) ประกอบด้วยวิธีการดังนี้

(1) วิธีการกรองข้อมูลร่วมโดยพิจารณาผู้ใช้เป็นหลัก (User-based)

วิธีนี้จะพิจารณาความชอบของผู้ใช้ในการเลือกบริโภคหรือบริการข้อมูลที่คล้ายกันกับผู้ใช้คนอื่นในฐานะข้อมูล เช่น การรับชม ความชอบ การเลือกซื้อ จากนั้นจะทำการเลือกผู้ใช้ที่มีพฤติกรรมใกล้เคียงกับผู้ใช้เป้าหมายมากที่สุดแล้วทำการสร้างการทำนายเรตติ้งของข้อมูลต่างๆ โดยพิจารณาจากผู้ใช้ที่มีพฤติกรรมที่ใกล้เคียงกันต่อจากนั้นจะสร้างรายการแนะนำข้อมูลที่คาดว่าผู้ใช้จะชื่นชอบโดยเรียงลำดับค่าการทำนายจากมากไปหาน้อยตามลำดับ

(2) วิธีการกรองข้อมูลร่วมโดยพิจารณาข้อมูลเป็นหลัก (Item-based)

เป็นวิธีการพิจารณาความชอบของข้อมูลที่ใช้ทำการเลือกบริโภคที่คล้ายกันกับข้อมูลอื่นในฐานข้อมูล จากนั้นพิจารณาข้อมูลที่มีลักษณะในการเลือกบริโภคที่คล้ายกันมากที่สุดแล้วสร้างรายการแนะนำข้อมูล โดยพิจารณาจากค่าการทำนายเรตติ้งของข้อมูลข้างเคียงจากมากไปหาน้อยตามลำดับ

(3) Pearson's correlation similarity

Pearson's correlation เป็นค่าความคล้ายคลึงที่มักใช้กับระบบแนะนำที่ใช้วิธีการกรองข้อมูลร่วมแบบจดจำ (Memory-based collaborative filtering) ซึ่งเป็นการหาค่าความคล้ายคลึงระหว่างผู้ใช้ u และ v โดยกำหนดให้ r_{ui} เป็นอันดับความสนใจของสินค้า i และผู้ใช้ u และ \bar{r}_u คือค่าเฉลี่ยของอันดับความสนใจทั้งหมดของผู้ใช้ u ที่มีการให้อันดับสินค้าเหมือนกันของผู้ใช้ (co-rate) และ n คือจำนวนสินค้าที่มีการให้อันดับความสนใจเหมือนกันของผู้ใช้ (co-rate) ดังนี้

$$sim(u, v) = Pearson(u, v) = \frac{\sum_{i \in I_u \cap I_v} (r_{ui} - \bar{r}_u) \cdot (r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I_u \cap I_v} (r_{ui} - \bar{r}_u)^2} \cdot \sqrt{\sum_{i \in I_u \cap I_v} (r_{vi} - \bar{r}_v)^2}} \quad (2)$$

(4) ค่าความคล้ายคลึงแบบโคซายน์ (Cosine similarity)

Cosine similarity เป็นตัวชี้วัดความคล้ายระหว่าง 2 เวกเตอร์ ซึ่งจะถูกวัดด้วยค่าโคซายน์ (Cos) ของมุมระหว่าง 2 เวกเตอร์ และบ่งบอกว่า 2 เวกเตอร์ชี้ไปในทิศทางเดียวกันหรือไม่ โดยการหาค่าความคล้ายคลึงเชิงมุมกำหนดให้ r_{ui} เป็นอันดับความสนใจของสินค้า i โดยผู้ใช้ u และ n คือจำนวนของสินค้าที่มีการให้อันดับความสนใจเหมือนกันของผู้ใช้ (co-rate) ดังนี้

$$sim(u, v) = \frac{\sum_{i \in I_{uv}} r_{ui} \cdot r_{vi}}{\sqrt{\sum_{i \in I_{uv}} r_{ui}^2} \cdot \sqrt{\sum_{i \in I_{uv}} r_{vi}^2}} \quad (3)$$

(5) วิธีเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor หรือ KNN)

เป็นวิธีการในการจัดแบ่งคลาสโดยจะตัดสินใจว่าคลาสใดที่จะแทนเงื่อนไข หรือกรณีใหม่ๆ ได้บ้าง เป็นวิธีการหนึ่งสำหรับแก้ปัญหาประมาณค่าฟังก์ชันนอเนพาราเมตริกสำหรับการจำแนกของกลุ่มข้อมูลที่ไม่เป็นรูปร่างที่ดี หรือข้อมูลที่กระจัดกระจาย โดยทำการตรวจสอบจำนวนบางจำนวนของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุดเท่ากับจำนวน k ที่ต้องการ โดยการหาระยะทางที่ใกล้ที่สุด

(6) การทำนายเรตติ้ง (Prediction)

การทำนายเรตติ้งที่ผู้ใช้จะให้คะแนนความชอบต่อสินค้าสามารถคำนวณได้ 3 วิธีดังนี้

- 1) KNNBasic** คือการให้เรตติ้งโดยประมาณนั้นเป็นค่าเฉลี่ยถ่วงน้ำหนักของการให้เรตติ้งที่ผู้ชมมอบให้กับรายการที่คล้ายกัน ถ่วงน้ำหนักโดยความคล้ายคลึงกัน ดังนี้

$$\hat{r}_{ui} = \frac{\sum_{v \in N_i^k(u)} sim(u, v) \cdot r_{vi}}{\sum_{v \in N_i^k(u)} sim(u, v)} \quad (4)$$

2) **KNNWithMeans** คือการปรับสูตร KNNBasic ด้วยเรตติ้งเฉลี่ยของผู้ใช้ ดังนี้

$$\hat{r}_{ui} = \mu_u + \frac{\sum_{v \in N_i^k(u)} \text{sim}(u,v) \cdot (r_{vi} - \mu_v)}{\sum_{v \in N_i^k(u)} \text{sim}(u,v)} \quad (5)$$

3) **KNNWithZScore** คือการปรับด้วยค่าเบี่ยงเบนมาตรฐานของการให้เรตติ้งของผู้ใช้ ดังนี้

$$\hat{r}_{ui} = \mu_u + \sigma_u \frac{\sum_{v \in N_i^k(u)} \text{sim}(u,v) \cdot (r_{vi} - \mu_v) / \sigma_u}{\sum_{v \in N_i^k(u)} \text{sim}(u,v)} \quad (6)$$

2.3 วิธีการประเมินประสิทธิภาพ

ในการประเมินประสิทธิภาพของวิธีการกรองสามารถพิจารณาจากค่าต่างๆ ดังนี้

2.3.1 **ค่าความแม่นยำ (Precision)** คืออัตราส่วนระหว่างสินค้าที่แนะนำซึ่งสัมพันธ์กับผู้ใช้กับจำนวนสินค้าที่แนะนำทั้งหมด ดังนี้

$$\text{Precision} = \frac{\text{Correctly recommended items}}{\text{Total recommended items}} \quad (7)$$

โดย

Correctly recommended items แทนจำนวนสินค้าที่สัมพันธ์กับลูกค้าจากสินค้าที่แนะนำ

Total recommended items แทนจำนวนสินค้าที่แนะนำทั้งหมด

2.3.2 **ค่าความแม่นยำเฉลี่ย (Average precision หรือ AP)** คือ ผลรวมของความแม่นยำของอันดับต่างๆ สำหรับสินค้าที่แนะนำทั้งหมด

$$\text{Average Precision@N} = \frac{1}{m} \sum_{k=1}^N (\text{Precision@k if } k^{\text{th}} \text{ item was relevant}) \quad (8)$$

โดย

Precision@k แทนความแม่นยำที่อันดับ k

m แทนจำนวนครั้งที่ระบบแนะนำสัมพันธ์กับผู้ใช้

N แทนจำนวนสินค้าที่แนะนำทั้งหมด

2.3.3 ค่าเฉลี่ยของความแม่นยำเฉลี่ย (Mean average precision หรือ MAP) คือ ผลรวมของค่าเฉลี่ยความแม่นยำของสินค้าที่แนะนำทั้งหมด ดังนี้

$$\text{Mean Average Precision}@N = \frac{1}{N} \sum_{i=1}^N (\text{Average Precision}@k)_i \quad (9)$$

2.3.4 ค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean absolute error หรือ MAE) คือ ผลรวมของค่าเฉลี่ยความคลาดเคลื่อนของสินค้าที่แนะนำทั้งหมด ดังนี้

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

2.3.5 รากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root mean square error หรือ RMSE) คือ ผลรวมของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสินค้าที่แนะนำทั้งหมด ดังนี้

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (11)$$

2.4 งานวิจัยที่เกี่ยวข้อง

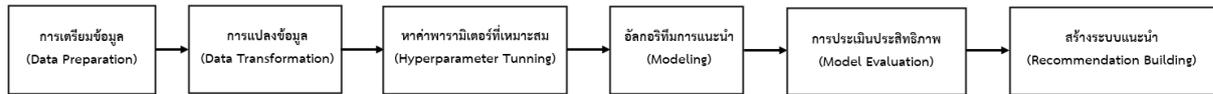
Pratama et al. (2020) ได้ศึกษาการแนะนำสินค้าในอุตสาหกรรมค้าปลีกแบบออนไลน์โดยใช้วิธีการกรองแบบร่วมกัน โดยใช้ข้อมูลจากร้านค้าปลีกแห่งหนึ่งในประเทศอินโดนีเซีย เพื่อสร้างรายการแนะนำสินค้าและหาวิธีที่ดีที่สุด ผลการศึกษาพบว่าวิธีการกรองข้อมูลรวมแบบจดจำมีประสิทธิภาพเหนือกว่าวิธีการกรองข้อมูลรวมแบบจำลอง และพบว่ายิ่งให้ตัวแบบเรียนรู้ข้อมูลมากเท่าใด ประสิทธิภาพของระบบการแนะนำก็จะยิ่งดีขึ้นเท่านั้น งานวิจัยนี้ยังพยายามทำ Customer segmentation เพื่อแก้ปัญหาข้อมูลมีปริมาณน้อย (Data sparsity problem) ได้แบ่ง Segment โดยใช้ RFM Model (Recency, Frequency, and Monetary value) ในการแบ่งกลุ่ม (Clustering) พบว่าการแบ่งกลุ่มลูกค้ามาช่วยไม่ได้ทำให้ประสิทธิภาพของตัวแบบดีขึ้น

Shaikh (2020) ได้ศึกษาเรื่องระบบแนะนำภาพยนตร์โดยใช้เทคนิคการแนะนำ 5 เทคนิค ได้แก่ KNNBaseline, KNNWithMeans User-based, KNNWithMeans Item-based, SVD, และ SVDpp และใช้ RMSE และ MAE ในการวัดประเมินประสิทธิภาพของแต่ละตัวแบบ ผลการศึกษาพบว่า SVDpp มีผลเรตติ้งความแม่นยำ RMSE ดีที่สุด และการทำ K-Fold cross validator ได้ผล RMSE ดีกว่าการทำ LOOCV cross validator

Herlocker et al. (2004) ได้ศึกษาเรื่องการตัดสินใจเลือกใช้วิธีการประเมินผลของวิธีการกรองแบบร่วมกันในระบบแนะนำ พบว่าตัววัดความแม่นยำเชิงคาดการณ์ (Predictive accuracy metrics) เช่น Mean absolute error และ Root mean square error ไม่เหมาะสมกับการทำนายผลลัพธ์ที่อยู่ในรูปแบบอันดับ (Ranking) ซึ่งผู้ใช้อาจสนใจเฉพาะข้อผิดพลาดในรายการที่มีอันดับสูง การเลือกตัวชี้วัดที่เหมาะสมนั้นเป็นความท้าทาย เนื่องจากมีตัวชี้วัดที่ถูกเผยแพร่ออกมา มีความหลากหลาย การตัดสินใจเลือกใช้จึงขึ้นอยู่กับการใช้งานของผู้ใช้ ประเภทของการวิเคราะห์และประเภทของชุดข้อมูล ทำให้ไม่มีเมตริกที่เป็นมาตรฐานในสาขานี้ นักวิจัยจึงยังคงแนะนำเมตริกใหม่ๆ ในการประเมินระบบต่อไปด้วยการใช้เมตริกการประเมินที่หลากหลาย จึงเป็นเรื่องยากที่จะเปรียบเทียบผลลัพธ์จากงานวิจัยหนึ่งกับผลลัพธ์ในงานวิจัยอื่น ด้วยเหตุนี้จึงเป็นเรื่องยากที่จะรวมสิ่งตีพิมพ์ที่หลากหลายเหล่านี้เข้ากับองค์ความรู้ที่สอดคล้องกันเกี่ยวกับคุณภาพของอัลกอริทึมของระบบแนะนำ

3. วิธีดำเนินการวิจัย

งานวิจัยนี้ได้นำเสนอระบบแนะนำสินค้าสำหรับร้านค้าปลีกแบบดั้งเดิมโดยวิธีการกรองแบบร่วมกัน ใช้ข้อมูลประวัติการขายจากบริษัทจำหน่ายอุปกรณ์ไฟฟ้าและแสงสว่างแห่งหนึ่ง โดยดึงข้อมูลจากโปรแกรมบัญชี Express ตั้งแต่เดือนมกราคม ปี 2560 ถึงเดือนเมษายน ปี 2564 โดยผู้วิจัยได้วางแผนการดำเนินงานวิจัยดังภาพที่ 2



ภาพที่ 2 ขั้นตอนการดำเนินงานวิจัย

3.1 การเตรียมข้อมูล

บริษัทนี้ดำเนินธุรกิจจัดจำหน่ายอุปกรณ์ไฟฟ้าและแสงสว่างให้กับลูกค้าปลีกรายย่อยและช่าง ไฟล์ข้อมูลจากโปรแกรมบัญชี Express ที่อยู่ในรูปแบบไฟล์ประเภท .DBF (dBASE file) จึงจำเป็นต้องแปลงไฟล์ให้อยู่ในรูปแบบ Semicolon separated value เพื่อให้สะดวกต่อการเตรียมข้อมูลและสร้างตัวแบบ ชุดข้อมูลประกอบด้วยรหัสสินค้า (STKCOD), ชื่อสินค้า (STKDES), เลขที่เอกสาร (DOCNUM), รหัสลูกค้า (CUSCOD), วันที่ (DOCDAT) มีข้อมูลธุรกรรมการขายทั้งหมด 118,105 รายการ ซึ่งการศึกษานี้จะใช้ข้อมูลธุรกรรมเฉพาะลูกค้าที่เป็นสมาชิก (ลูกค้าที่มีรหัสลูกค้า) จำนวน 358 คน เป็นธุรกรรมจำนวน 38,282 รายการมาวิเคราะห์ เนื่องจากสามารถระบุตัวตนของลูกค้าได้ทำให้ธุรกิจสามารถแนะนำสินค้ากับลูกค้าโดยตรง โดยมุ่งหวังว่าธุรกิจจะสามารถประยุกต์ระบบแนะนำนี้ในการสร้างความสัมพันธ์กับลูกค้าและการทำการตลาดแบบเฉพาะบุคคล นอกจากนี้ผู้วิจัยได้ตัดรายการสินค้าที่ซ้ำกันออก (Duplicate) เช่น หากสินค้าเป็นสินค้าชนิดเดียวกัน แต่สีต่างกัน หรือขนาดต่างกันก็ถือว่าเป็นสินค้าเดียวกัน รวมถึงนำรายการธุรกรรมที่ไม่เกี่ยวข้องออก เช่น ค่าแรงและค่าจัดส่งสินค้า เหลือสินค้าที่ถูกซื้อโดยลูกค้าที่เป็นสมาชิกจำนวน 458 ชนิด

3.2 การแปลงข้อมูล

วิธีการกรองแบบร่วมกันใช้ User-Item rating matrix ในการทำนาย ข้อมูลในเมทริกซ์นี้อยู่ในรูปแบบผู้ใช้ (User) ให้คะแนนความชอบ (Rating) กับสิ่งของ (Item) แต่การเก็บข้อมูลการให้เรตติ้งสินค้าจากการซื้อสินค้าในรูปแบบการขายปลีกแบบดั้งเดิมนั้นเป็นไปได้ยาก ประวัติการซื้อของลูกค้าจึงถูกนำมาใช้เพื่อแสดงถึงความคิดเห็นของลูกค้าที่มีต่อสิ่งที่พวกเขาซื้อซึ่งเหมาะกับการวิเคราะห์ลูกค้าที่สามารถระบุตัวตนได้และลูกค้าที่มาซื้อซ้ำ หากมีการซื้อรายการใดรายการหนึ่งซ้ำๆ เป็นการบ่งชี้ได้ว่าลูกค้ารายนั้นชอบรายการนั้นเพราะแสดงให้เห็นว่ายิ่งลูกค้าซื้อสินค้าบ่อย เรตติ้งที่ลูกค้าให้ต่อสินค้าชนิดนั้นก็ยิ่งสูงขึ้น ฉะนั้น User-item matrix ในงานวิจัยนี้เป็น Implicit rating ที่มาจากจำนวนครั้งที่ลูกค้าซื้อสินค้าครั้งหนึ่ง ข้อมูลนี้จะถูกแปลงให้อยู่ในรูปแบบเรตติ้ง 1 ถึง 5 โดยการใช้ Min-max scaling algorithm (Pratama et al., 2020) ดังสมการที่ 12

$$r_{ui} = \left[\left(\frac{f_{ui} - f_{min}}{f_{max} - f_{min}} \right) \times (r_{max} - r_{min}) \right] + r_{min} \quad (12)$$

โดย

r_{ui} แทนเรตติ้งที่แปลงให้อยู่ในพิสัย 1 (r_{min}) ถึง 5 (r_{max})

f_{ui} แทนจำนวนครั้งที่ซื้อ

f_{max} แทนจำนวนการซื้อที่มากที่สุดในสินค้าชนิดนั้นๆ

f_{min} แทนจำนวนการซื้อที่น้อยที่สุดในสินค้าชนิดนั้นๆ

ตัวอย่าง User-item matrix ก่อนการแปลงข้อมูลและหลังการแปลงข้อมูล แสดงในตารางที่ 1 และตารางที่ 2

ตารางที่ 1 ตัวอย่าง **User-Item Matrix** ก่อนการแปลงข้อมูลซึ่งเป็นข้อมูลจำนวนครั้งที่ลูกค้าซื้อสินค้าชนิดนั้นๆ

	Item_1	Item_2	Item_3	Item_4	Item_5
User_A	2	-	-	7	-
User_B	-	1	-	-	5
User_C	8	-	3	2	1
User_D	-	20	-	3	-
User_E	10	-	-	-	-

ตารางที่ 2 ตัวอย่าง **User-Item Matrix** หลังจากการแปลงข้อมูลเป็นเรตติ้ง 1 ถึง 5 และใช้ 0 แทนที่สินค้าที่ไม่เคยถูกซื้อโดยลูกค้ารายนั้นๆ

	Item_1	Item_2	Item_3	Item_4	Item_5
User_A	1	0	0	5	0
User_B	0	1	0	0	5
User_C	4	0	5	1	1
User_D	0	5	0	2	0
User_E	5	0	0	0	0

3.3 การหาค่าพารามิเตอร์ที่เหมาะสม

การคัดเลือกตัวแบบที่ดีที่สุดต้องมาจากการกำหนดค่าพารามิเตอร์ที่เหมาะสมด้วยเช่นกัน ซึ่งจะหาค่าพารามิเตอร์ที่เหมาะสมโดยการทำ Hyperparameter tuning ที่ทำให้ตัวแบบทำนายได้ค่าความผิดพลาดต่ำที่สุด ในที่นี้กำหนดให้ RMSE เป็นค่าความผิดพลาด ได้ผลลัพธ์การกำหนดค่าพารามิเตอร์ที่เหมาะสมที่สุดของแต่ละอัลกอริทึม ดังตารางที่ 3

ตารางที่ 3 ผลลัพธ์การกำหนดค่าพารามิเตอร์ที่เหมาะสมที่สุดของแต่ละอัลกอริทึม

Model	Parameter	Best Parameter	RMSE
SVD	Number of epoch	20	1.00935
	Learning rate	0.01	
	Regularization term	0.4	
KNNBasic	Method	Alternating Least Square	1.00208
	Number of K	10	
	Similarity	Pearson's correlation	
	Based	Item-based	
KNNWithMeans	Method	Alternating Least Square	1.00151
	Number of K	10	
	Similarity	Pearson's correlation	
	Based	Item-based	
KNNWithZScore	Method	Alternating Least Square	1.00206
	Number of K	10	
	Similarity	Pearson's correlation	
	Based	Item-based	

3.4 การพัฒนาตัวแบบ (Modeling)

งานวิจัยนี้ต้องการเปรียบเทียบวิธีการกรองแบบร่วมกัน 2 แบบคือ วิธีการกรองข้อมูลร่วมแบบจำลอง (Model-based filtering) โดยใช้ SVD และวิธีการกรองข้อมูลร่วมแบบจดจำ (Memory-based filtering) โดยใช้ KNN ซึ่งตัวแบบทั้ง SVD และ KNN ได้รับการพัฒนาโดยใช้ค่าพารามิเตอร์ที่เหมาะสมที่สุดที่พบในการทดลองนี้ (Hyperparameter tuning) ดังแสดงในตารางที่ 3

การแบ่งข้อมูลที่ใช้ในการทดสอบ (Split testing) 80:20 แบบ K-fold Cross-validation คือการแบ่งข้อมูลออกเป็น K ชุดเท่าๆ กันและทำการคำนวณค่า Error สำหรับ K รอบ โดยแต่ละรอบการคำนวณข้อมูลชุดหนึ่งจากข้อมูล K ชุดจะถูกเลือกออกมาเพื่อเป็นข้อมูลทดสอบ และข้อมูลอีก K-1 ชุดจะถูกใช้เป็นข้อมูลสำหรับการเรียนรู้

ตัวแบบจำลองที่นำไปใช้ในระบบแนะนำต้องเป็นตัวแบบที่มีประสิทธิภาพ โดยเปรียบเทียบตัวแบบจากค่า RMSE และ MAE หากอัลกอริทึมใดมีค่าความผิดพลาด (Error) ต่ำ จะจัดว่าเป็นอัลกอริทึมที่ใช้แนะนำสินค้าได้ดีที่สุด จากนั้นนำอัลกอริทึมที่มีประสิทธิภาพดีที่สุดมาใช้แนะนำสินค้า Top-N ชิ้น และวัดผลอัลกอริทึมอีกครั้งโดยใช้ค่า Mean average precision (MAP) ในการประเมินความแม่นยำของตัวแบบว่า อันดับสินค้า Top-N ที่ได้รับการแนะนำมานั้นมีความแม่นยำในระดับใด

4. ผลการศึกษา

ผลการศึกษาเรื่อง การแนะนำสินค้าในร้านจำหน่ายอุปกรณ์ไฟฟ้าและแสงสว่างโดยวิธีการกรองแบบร่วมกัน กรณีศึกษาบริษัทเอกชนแห่งหนึ่งแบ่งเป็น 2 ส่วน คือ ผลลัพธ์การเปรียบเทียบประสิทธิภาพของอัลกอริทึม และผลลัพธ์ความแม่นยำในการแนะนำ

4.1 ผลลัพธ์การเปรียบเทียบประสิทธิภาพของอัลกอริทึม

ตารางที่ 4 แสดงอัลกอริทึมวิธีการกรองแบบร่วมกันสำหรับระบบแนะนำทั้ง 7 อัลกอริทึมคือ SVD แบบคัดกรองผู้ใช้ แบบคัดกรองสิ่งของ แบบคัดกรองผู้ใช้ด้วยค่าเฉลี่ย แบบคัดกรองสิ่งของด้วยค่าเฉลี่ย แบบคัดกรองผู้ใช้ด้วยค่ามาตรฐาน และแบบคัดกรองสิ่งของด้วยค่ามาตรฐานพบว่า แบบคัดกรองสิ่งของด้วยค่าเฉลี่ยให้ค่า RMSE ต่ำที่สุด แบบคัดกรองสิ่งของด้วยค่ามาตรฐานให้ค่า MAE ต่ำที่สุด แบบคัดกรองผู้ใช้ใช้เวลา Fit time ต่ำที่สุด และ SVD ใช้เวลา Test time ต่ำที่สุด

ตารางที่ 4 ตารางเปรียบเทียบค่า RMSE, MAE, Fit time และ Test time ของอัลกอริทึม วิธีการกรองแบบร่วมกันประเภทต่าง ๆ

Model	Train_RMSE	Test_RMSE	Train_MAE	Test_MAE	Fit time	Test time
SVD	1.0094	1.0094	0.9775	0.9775	7.0656	0.2387
KNNBasic User-based	0.9899	1.0062	0.9676	0.9767	1.6523	6.7061
KNNBasic Item-based	0.9822	1.0020	0.9625	0.9738	2.0280	8.8795
KNNWithMeans User-based	0.9887	1.0043	0.9666	0.9753	1.6717	6.8100
KNNWithMeans Item-based	0.9820	1.0018	0.9623	0.9735	2.0705	8.9040
KNNWithZScore User-based	0.9880	1.0039	0.9659	0.9749	1.6865	6.6305
KNNWithZScore Item-based	0.9816	1.0019	0.9617	0.9732	2.1351	8.9968

4.2 ผลลัพธ์ความแม่นยำในการแนะนำ

เมื่อผู้ใช้ใช้ระบบแนะนำอาจสนใจเฉพาะข้อผิดพลาดในรายการที่มีอันดับสูง ผลลัพธ์ของระบบแนะนำที่นำไปใช้จริง จะอยู่ในรูปแบบจัดอันดับ (Ranking) สินค้าที่ลูกค้าอาจจะชอบสูง (อันดับแรก) ไปยังสินค้าที่ลูกค้าอาจจะชอบรองลงมา (อันดับถัดมา) ในการทดสอบการแนะนำสินค้าอันดับ Top-N จะประเมินความแม่นยำจากค่า Mean average precision (MAP) โดยพิจารณาการแนะนำหนังสือจำนวน N เล่ม (Top-N recommendation) โดยในงานวิจัยนี้พิจารณาอันดับ N ที่น่าสนใจที่ 1 ถึง 15 จากตารางที่ 5 พบว่าเมื่อแนะนำรายการสินค้ามากขึ้นแล้วระดับความแม่นยำจะลดลงเรื่อยๆ

ตารางที่ 5 ตารางเปรียบเทียบค่า Mean Average Precision ตั้งแต่ Top 1 ถึง Top 15 ของแต่ละอัลกอริทึม

Top-N	SVD	KNNBasic User-User	KNNBasic Item-Item	KNNWith Means User-User	KNNWith Means Item-Item	KNNWith ZScore User-User	KNNWith ZScore Item-Item
1	0.047	0.246	0.202	0.242	0.217	0.249	0.211
2	0.053	0.230	0.186	0.222	0.198	0.215	0.196
3	0.056	0.214	0.179	0.211	0.181	0.200	0.185
4	0.051	0.207	0.170	0.197	0.169	0.182	0.170
5	0.055	0.194	0.164	0.188	0.166	0.171	0.166
6	0.056	0.182	0.158	0.181	0.161	0.168	0.156
7	0.054	0.176	0.152	0.173	0.152	0.157	0.149
8	0.049	0.169	0.148	0.163	0.147	0.151	0.144
9	0.054	0.161	0.144	0.158	0.144	0.142	0.138
10	0.053	0.159	0.139	0.150	0.139	0.137	0.134
11	0.051	0.149	0.133	0.148	0.134	0.132	0.133
12	0.053	0.146	0.132	0.141	0.129	0.127	0.128
13	0.053	0.141	0.128	0.139	0.124	0.124	0.121
14	0.054	0.135	0.122	0.133	0.121	0.120	0.118
15	0.054	0.130	0.119	0.127	0.118	0.116	0.115

จากผลการศึกษาอัลกอริทึมสำหรับแนะนำด้วยวิธีการกรองแบบร่วมกันด้วยข้อมูลการขายสินค้าของร้านค้าปลีกแบบออนไลน์ จำนวนอันดับ (Top-N) ที่เหมาะสมต่อการแนะนำคือ 5 อันดับ โดยพิจารณาจากค่า Mean average precision จากตารางพบว่าวิธีการกรองแบบร่วมกันโดยพิจารณาผู้ใช้เป็นหลักด้วยค่ามาตรฐาน (User-based collaborative filtering With Z-Score) มีค่า MAP@1 ดีที่สุดเท่ากับ 0.249 สำหรับ MAP@2 จนถึง สำหรับ MAP@5 นั้น วิธีการกรองแบบร่วมกันโดยพิจารณาผู้ใช้เป็นหลักแบบปกติ (User-based collaborative filtering) จะมีความแม่นยำสูงที่สุด เมื่อจำแนกลูกค้าเป็นกลุ่ม ๆ พบว่ากลุ่มลูกค้าประเภทโรงแรมและรีสอร์ทจะได้รับการแนะนำให้ซื้อคอมพิวเตอร์และหลอดเข็ม

5. สรุปและอภิปรายผลการศึกษา

งานวิจัยนี้เป็นการศึกษาเกี่ยวกับการสร้างระบบแนะนำสินค้าด้วยเทคนิคกรองแบบร่วมกันสำหรับธุรกิจค้าปลีกแบบออนไลน์ เพื่อหาอัลกอริทึมสำหรับการแนะนำที่เหมาะสมที่สุด ซึ่งผู้ประกอบการร้านค้าปลีกแบบออนไลน์, ผู้ประกอบการที่ให้บริการด้านเว็บไซต์อิเล็กทรอนิกส์ (E-commerce) ไปจนถึงห้างสรรพสินค้าสามารถนำไปเป็นทางเลือกในการประยุกต์ใช้ต่อไปได้ งานวิจัยนี้มีกระบวนการหลัก 7 ขั้นตอน ได้แก่ การเตรียมข้อมูล การแปลงข้อมูล การหาค่าพารามิเตอร์ที่เหมาะสมการสร้างตัวแบบ การประเมินประสิทธิภาพตัวแบบ และการสร้างระบบแนะนำ จากการศึกษาพบว่าวิธีการกรองข้อมูลร่วมแบบจดจำ (Memory-based filtering) มีความแม่นยำสูงกว่าวิธีการกรองข้อมูลร่วมแบบจำลอง (Model-based filtering) โดยใช้การแตกค่าแบบเอกฐาน (Singular value decomposition หรือ SVD) โดยวิธีการกรองข้อมูลร่วมแบบจำลอง ซึ่งกรองข้อมูลโดยพิจารณาผู้ใช้เป็นหลัก (User-based) เป็นวิธีที่ดีที่สุดในการสร้างระบบแนะนำสำหรับงานวิจัยนี้ เนื่องจากวิธีการกรองแบบร่วมกันโดยพิจารณาผู้ใช้เป็นหลัก มีความแม่นยำสูงกว่า

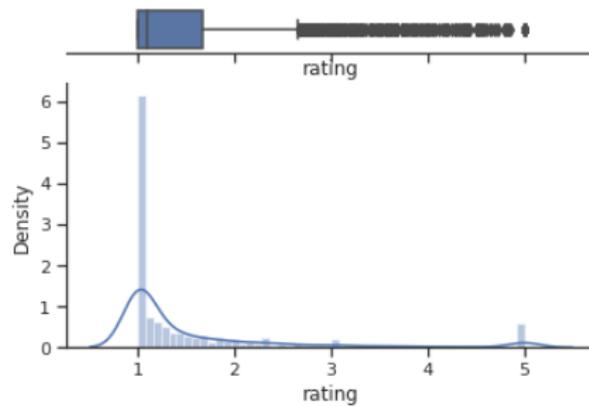
วิธีการกรองแบบร่วมกันโดยพิจารณาสิ่งของเป็นหลัก (Item-based) โดยพิจารณาจากค่า Mean precision ที่สูงที่สุดในอันดับที่ Top 1 ถึง Top 5 เนื่องจากตัววัดความแม่นยำเชิงคาดการณ์ (Predictive accuracy metrics) เช่น Mean absolute error (MAE) และ Root mean square error (RMSE) ไม่เหมาะสมกับการทำนายผลลัพธ์ที่อยู่ในรูปแบบอันดับ (Ranking) ซึ่งผู้ใช้อาจสนใจเฉพาะข้อผิดพลาดในรายการที่มีอันดับสูง

วิธีการกรองแบบร่วมกันโดยพิจารณาผู้ใช้เป็นหลักเป็นการหาความคล้ายคลึงของผู้ใช้จากการซื้อของลูกค้า โดยความคล้ายคลึงกันระหว่างลูกค้าสามารถวัดได้จากพฤติกรรมกรรมการซื้อสินค้า ซึ่งระบบจะแนะนำสินค้าที่ผู้ใช้เคยเลือกซื้อให้กับผู้ใช้ที่มีความคล้ายคลึงกันที่คำนวณได้จากอัลกอริทึม K-Nearest Neighbor บริษัทกรณีศึกษาดำเนินธุรกิจจัดจำหน่ายอุปกรณ์ไฟฟ้าและแสงสว่าง โดยลูกค้าประจำ (ลูกค้าที่มี Cuscod) ส่วนใหญ่เป็นกลุ่มผู้รับเหมาและช่างไฟฟ้าซึ่งมักจะซื้อสินค้าจำพวกวัสดุที่ใช้เมื่อสร้างบ้านหรือต่อเติมบ้านเหมือนกัน ยกตัวอย่างเช่น วัสดุที่ใช้ในการเดินสายไฟ เช่น ท่อร้อยสายมักจะถูกซื้อคู่กับสายไฟและข้อต่อ หรือ วัสดุที่ใช้ในการติดตั้งเต้ารับและสวิตช์ เช่น สวิตช์และเต้ารับ มักจะถูกซื้อคู่กับปลั๊กกลอยและหน้ากาก เมื่อลูกค้ามีพฤติกรรมกรรมการซื้อสินค้าคล้ายคลึงกันส่งผลให้ วิธีการกรองแบบร่วมกันโดยพิจารณาผู้ใช้เป็นหลัก เป็นวิธีที่เหมาะสมในการสร้างระบบแนะนำสินค้า

อย่างไรก็ตามในงานวิจัย Herlocker et al. (2004) พบว่าการตัดสินใจเลือกใช้วิธีการประเมินผลขึ้นอยู่กับการใช้งานของผู้ใช้ ประเภทของการวิเคราะห์และประเภทของชุดข้อมูล ทำให้ไม่มีเมตริกที่เป็นมาตรฐานในสาขานี้ จึงเป็นเรื่องยากที่จะเปรียบเทียบผลลัพธ์จากงานวิจัยหนึ่งกับผลลัพธ์ในงานวิจัยอื่น ซึ่งสำหรับงานวิจัยการสร้างระบบแนะนำสินค้าด้วยเทคนิคกรองแบบร่วมกันสำหรับธุรกิจค้าปลีกแบบออฟไลน์พบว่าวิธีการกรองแบบร่วมกันโดยพิจารณาผู้ใช้เป็นหลัก และวิธีการกรองแบบร่วมกันโดยพิจารณาสิ่งของเป็นหลัก มีค่าความแม่นยำและค่าความผิดพลาดใกล้เคียงกันมาก การใช้ชุดข้อมูลอื่นอาจให้ผลลัพธ์และประสิทธิภาพแตกต่างออกไปจากนี้

6. ปัญหาและข้อเสนอนะ

ปัญหาที่พบในงานวิจัยคือปัญหาข้อมูลมีปริมาณน้อย (Data sparsity problem) และเบาบาง เนื่องจากข้อมูลมีอยู่เพียง 5% จากข้อมูลทั้งหมด และจากข้อมูลที่มีอยู่พบว่าเราตั้งส่วนใหญ่เท่ากับ 1 ดังภาพที่ 3 ทำให้ตัวแบบมีความแม่นยำต่ำกว่าที่ควร แนวทางที่จะพัฒนาต่อไปในอนาคต ควรทดสอบด้วยชุดข้อมูลที่เหมาะสม เช่น ข้อมูลที่มีปริมาณมาก โดยการขยายระยะเวลาในการเก็บข้อมูลและพยายามระบุตัวตนลูกค้ารายย่อยให้อยู่ในระบบ เนื่องจากรายการธุรกรรมการขายจำนวน 67% เป็นของลูกค้าปลีกที่ไม่สามารถระบุตัวตนได้ งานวิจัยนี้ไม่สามารถนำข้อมูลส่วนนี้มาใช้ได้เนื่องจากไม่สามารถจัดเก็บข้อมูลการกลับมาซื้อซ้ำของลูกค้าได้ นอกจากนี้หากมีการเตรียมแคตตาล็อกข้อมูลสินค้าและการสร้าง Feature ที่เหมาะสมเพื่ออธิบายตัวสินค้า เช่น กลุ่ม/ประเภท ขนาด ราคา สี สไตล์ แบรินด์ หรือคำอธิบายตัวสินค้าแบบย่อ เป็นต้น ก็สามารถนำข้อมูลส่วนนี้มาใช้ในกับวิธีการกรองเนื้อหาได้ รวมถึงสามารถใช้วิธีการกรองแบบผสม (Hybrid filtering) ซึ่งมีหลายงานวิจัยพิสูจน์ว่ามีประสิทธิภาพที่สุด สุดท้ายแล้วงานวิจัยนี้ยังไม่สามารถทดสอบจริงได้กับบริษัทกรณีศึกษา เนื่องจากสถานการณ์การแพร่ระบาด COVID-19 เพราะหากนำไปทดสอบอาจจะได้ผลคลาดเคลื่อนได้ อย่างไรก็ตามระบบแนะนำนี้ควรนำไปทดสอบจริงในร้านค้าปลีกแบบดั้งเดิม โดยวัดผลจากอัตราการตอบสนองของลูกค้า (Hit rate) เทียบกับค่า MAP เพื่อประเมินประสิทธิภาพของระบบแนะนำ



ภาพที่ 3 Histogram และ Boxplot ของ Rating

บรรณานุกรม

- โกเมศ อัมพวัน, อนุชิต จิตพัฒนกุล, และ วราวุฒิ ฟ้าเจริญ. (2562). ระบบแนะนำหนังสืออ่านเพิ่มเติมเพื่อส่งเสริมการอ่านและการศึกษาด้อยตนเอง. คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา.
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1), 5-53.
- Pratama, B. Y., Budi, I., & Yuliawati, A. (2020). Product Recommendation in Offline Retail Industry by using Collaborative Filtering. *International Journal of Advanced Computer Science and Applications*, 11(9), 635-643.
- Shaikh, M. I. (2020). *Top-N Nearest Neighbourhood based Movie Recommendation System using different Recommendation Techniques* (Doctoral dissertation, Dublin, National College of Ireland).
- Walter, F. E., Battiston, S., Yildirim, M., & Schweitzer, F. (2012). Moving recommender systems from on-line commerce to retail stores. *Information systems and e-business management*, 10(3), 367-393.