

การเปรียบเทียบแบบจำลองการกรองข้อมูลสำหรับระบบแนะนำหลักสูตร ระดับปริญญาตรีภายใต้ข้อจำกัดความเบาบางของข้อมูล A Comparison of Data Filtering Models for Undergraduate Program Recommendation under Data Sparsity Constraints

วรพจน์ วรณช¹ อเนก พุทธิเดช² และ สุวิทย์ สมสุภาพรุ่งยศ^{1*}

Worapoj Woranuch¹ Anek Putthidet² and Suwit Somsuphaprunyos^{1*}

^{1*}สาขาวิชาวิทยาการจัดการนวัตกรรมดิจิทัล คณะบริหารธุรกิจและเทคโนโลยีสารสนเทศ

มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ จังหวัดพระนครศรีอยุธยา 13000

²สาขาวิชาวิทยาการข้อมูล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ
จังหวัดพระนครศรีอยุธยา 13000

^{1*}Program of Digital Innovation Management Program, Faculty of Business Administration and
Information Technology, Rajamangala University of Technology Suvarnabhumi,
Phra Nakhon Si Ayutthaya Province, 13000

²Program of Data Science Program, Faculty of Science and Technology, Rajamangala University of
Technology Suvarnabhumi, Phra Nakhon Si Ayutthaya Province, 13000

*Corresponding author E-mail: suwit.s@rmutsb.ac.th

Received 10 May 2026, Accepted 8 July 2026, Published 2 October 2026

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองการกรองข้อมูลสำหรับระบบแนะนำหลักสูตรระดับปริญญาตรี ภายใต้ข้อจำกัดความเบาบางของข้อมูล โดยเปรียบเทียบแบบจำลองระบบแนะนำจำนวน 7 แบบจำลอง ซึ่งจัดเป็น 3 วิธีหลัก ได้แก่ 1) Content-Based Filtering ประกอบด้วย Content-Based Filtering ด้วยแนวทาง Probability-Based Matching 2) Collaborative Filtering ประกอบด้วย User-Based Collaborative Filtering, Item-Based Collaborative Filtering และ Singular Value Decomposition (SVD) และ 3) Hybrid Filtering ประกอบด้วย Hybrid (User-Based CF, CB), Hybrid (Item-Based CF, CB) และ Hybrid (SVD, CB) กลุ่มตัวอย่างคือนักเรียนระดับมัธยมศึกษาตอนปลาย จำนวน 417 คน โดยคัดเลือกแบบเจาะจงและให้เลือกหลักสูตรที่สนใจจริง 5 อันดับจากหลักสูตรระดับปริญญาตรี 49 หลักสูตร จากนั้นแปลงอันดับเป็นคะแนนเพื่อสร้างเมทริกซ์ผู้ใช้-หลักสูตรขนาด 417×49 ซึ่งมีความเบาบางของข้อมูลประมาณร้อยละ 89.8 การประเมินผลใช้แนวทาง Top-N Recommendation ที่ค่า $K = 1, 3, 5$ และ 10 ด้วยตัวชี้วัด Hit Rate, Precision, Recall, F1-score, MAP และ MAR ผลการวิจัยพบว่า Hybrid (SVD, CB) ที่ค่า $\alpha = 0.80$ ให้ผลดีที่สุดโดยรวม โดยเฉพาะที่ $K=5$ มีค่า Hit Rate@5 = 1.0000, Precision@5 = 0.8590, Recall@5 = 0.9153, F1@5 = 0.8863, MAP@5 = 0.9089 และ MAR@5 = 0.5367 ผลการวิจัยชี้ให้เห็นว่าการผสมผสาน ระหว่าง Model – Based Collaborative Filtering แบบ SVD Model และ Content-Based Filtering สามารถเพิ่มประสิทธิภาพของระบบแนะนำหลักสูตรได้ภายใต้ความเบาบางของชุดข้อมูล

คำสำคัญ: แบบจำลองการกรองข้อมูล ระบบแนะนำ การประเมินแบบจัดอันดับ ความเบาบางของข้อมูล

ABSTRACT

This study aimed to compare the performance of data filtering models for an undergraduate program recommender system under data sparsity constraints. Seven recommendation models were compared and categorized into three main approaches: 1) Content-Based Filtering, consisting of Content-Based Filtering using the Probability-Based Matching approach; 2) Collaborative Filtering, consisting of User-Based Collaborative Filtering, Item-Based Collaborative Filtering, and Singular Value Decomposition (SVD); and 3) Hybrid Filtering, consisting of Hybrid (User-Based CF, CB), Hybrid (Item-Based CF, CB), and Hybrid (SVD, CB). The sample consisted of 417 high school students selected through purposive sampling. Participants were asked to choose their top five genuinely preferred undergraduate programs from a total of 49 programs. The ranked preferences were then transformed into scores to establish a 417×49 user-program matrix with approximately 89.8% data sparsity. The evaluation utilized the Top-N Recommendation approach at $K=1, 3, 5$, and 10 using Hit Rate, Precision, Recall, F1-score, Mean Average Precision (MAP), and Mean Average Recall (MAR) as performance metrics. The results indicated that the Hybrid (SVD, CB) model with $\alpha=0.80$ achieved the best overall performance. In particular, at $K=5$, the model obtained $\text{Hit Rate}@5 = 1.0000$, $\text{Precision}@5 = 0.8590$, $\text{Recall}@5 = 0.9153$, $\text{F1}@5 = 0.8863$, $\text{MAP}@5 = 0.9089$, and $\text{MAR}@5 = 0.5367$. The findings suggest that integrating the SVD-based Model-Based Collaborative Filtering model with Content-Based Filtering can significantly improve the performance of undergraduate program recommender systems under sparse data constraints.

Keywords: Filtering Model, Recommendation System, Top-N Evaluation, Data Sparsity

บทนำ

การเลือกศึกษาต่อในระดับปริญญาตรีเป็นกระบวนการตัดสินใจที่มีความซับซ้อน เนื่องจากผู้เรียนต้องพิจารณาปัจจัยหลายด้านร่วมกัน เช่น ความถนัด ความสนใจ ผลการเรียนรู้ เป้าหมายอาชีพ รายละเอียดหลักสูตร และโอกาสในการประกอบอาชีพ ขณะเดียวกัน สถาบันอุดมศึกษามีหลักสูตรที่หลากหลายมากขึ้น ส่งผลให้นักเรียนระดับมัธยมศึกษาตอนปลายต้องเผชิญกับข้อมูลจำนวนมากในการคัดกรอง เปรียบเทียบ และตัดสินใจเลือกหลักสูตรที่เหมาะสมกับตนเอง ระบบแนะนำจึงเป็นแนวทางหนึ่งที่สามารถช่วยลดภาระจากภาวะข้อมูลล้นหลาม และนำเสนอรายการที่สอดคล้องกับความสนใจหรือคุณลักษณะของผู้ใช้แต่ละรายได้อย่างเป็นระบบ (Adomavicius and Tuzhilin, 2005; Isinkaye et al., 2015) มหาวิทยาลัยราชภัฏพระนครศรีอยุธยาเปิดสอนหลักสูตรระดับปริญญาตรีจำนวน 49 หลักสูตร ครอบคลุม 4 คณะ ได้แก่ คณะครุศาสตร์ คณะวิทยาการจัดการ คณะวิทยาศาสตร์และเทคโนโลยี และคณะมนุษยศาสตร์และสังคมศาสตร์ ความหลากหลายของหลักสูตรดังกล่าวเป็นประโยชน์ต่อผู้เรียนในด้านการเลือกทางการศึกษา แต่ในขณะเดียวกันก็เพิ่มความซับซ้อนของการตัดสินใจ โดยเฉพาะเมื่อต้องเชื่อมโยงข้อมูลหลักสูตรกับคุณลักษณะของผู้เรียนแต่ละราย เช่น สายการเรียนรู้ ผลการเรียนรู้ เฉลี่ยสะสม วิชาที่ถนัด และทักษะที่สนใจ ดังนั้น การพัฒนาระบบที่สามารถจัดอันดับหลักสูตรที่เหมาะสมกับผู้เรียนจึงมีความสำคัญต่อการสนับสนุนการแนะแนวการศึกษาต่อ และอาจช่วยให้ผู้เรียนได้รับข้อมูลที่ตรงกับความต้องการของตนเองมากขึ้น

ในบริบทของประเทศไทย ระบบแนะนำได้รับการศึกษาและประยุกต์ใช้ในหลายสาขา เช่น การพาณิชย์ อิเล็กทรอนิกส์ ภาพยนตร์ การท่องเที่ยว ห้องสมุดออนไลน์ และการศึกษา โดยงานสำรวจงานวิจัยระบบแนะนำในประเทศไทยระหว่าง พ.ศ. 2550–2560 พบว่ามีการนำแบบจำลองระบบแนะนำไปใช้กับข้อมูลหลายประเภท และมีการใช้แบบจำลองการกรองข้อมูลหลายรูปแบบเพื่อเพิ่มความเหมาะสมของผลลัพธ์ (สมเพ็ชร, 2562) นอกจากนี้ งานวิจัยด้านระบบแนะนำห้องสมุดออนไลน์ในประเทศไทยยังแสดงให้เห็นว่าการใช้ Collaborative Filtering ร่วมกับข้อมูลส่วนบุคคลของผู้ใช้สามารถช่วยให้

ระบบแนะนำข้อมูลมีความสอดคล้องกับความต้องการของผู้ใช้มากขึ้น (วรณูช และวงกต, 2560) แนวคิดดังกล่าวสะท้อนให้เห็นว่าระบบแนะนำสามารถนำมาประยุกต์ใช้กับงานสนับสนุนการตัดสินใจด้านการศึกษาได้ โดยเฉพาะในกรณีที่ผู้ใช้ต้องเลือกจากรายการทางเลือกจำนวนมาก

แบบจำลองระบบแนะนำสามารถจำแนกได้หลายกลุ่ม ได้แก่ Content-Based Filtering, Collaborative Filtering และ Hybrid Filtering โดย Content-Based Filtering เป็นวิธีที่ใช้คุณลักษณะของผู้ใช้หรือรายการในการสร้างคำแนะนำ ส่วน Collaborative Filtering ใช้รูปแบบพฤติกรรมหรือความสนใจร่วมของผู้ใช้ในการคาดการณ์รายการที่เหมาะสม (Adomavicius and Tuzhilin, 2005; Lops et al., 2011; Isinkaye et al., 2015) อย่างไรก็ตาม Collaborative Filtering แบบดั้งเดิมอาจมีข้อจำกัดเมื่อข้อมูลมีความเบาบางสูง เนื่องจากการคำนวณความคล้ายคลึงระหว่างผู้ใช้หรือระหว่างรายการต้องอาศัยข้อมูลที่มีการให้คะแนนหรือเลือกซ้ำร่วมกันในระดับหนึ่ง ขณะที่ Matrix Factorization เช่น Singular Value Decomposition (SVD) เป็นแนวทางในกลุ่ม Model-Based Collaborative Filtering ที่สามารถเรียนรู้ความสัมพันธ์แฝงระหว่างผู้ใช้และรายการจากเมทริกซ์ข้อมูลได้ (Koren et al., 2009; Aggarwal, 2016) นอกจากการใช้แบบจำลองเดี่ยวแล้ว Hybrid Filtering เป็นแนวทางที่ผสานแบบจำลองระบบแนะนำมากกว่าหนึ่งวิธีเข้าด้วยกัน เพื่อลดข้อจำกัดของแต่ละวิธีและเพิ่มประสิทธิภาพของผลลัพธ์ (Burke, 2002) แนวทางดังกล่าวเหมาะกับบริบทที่ข้อมูลมีทั้งข้อมูลพฤติกรรมของผู้ใช้และข้อมูลคุณลักษณะประกอบ เช่น งานแนะนำหลักสูตรที่สามารถใช้ทั้งข้อมูลการเลือกหลักสูตรของผู้เรียนและข้อมูลประวัติหรือความสนใจของผู้เรียนร่วมกันได้ ในบริบทด้านการศึกษา มีงานวิจัยที่นำระบบแนะนำไปใช้กับการแนะนำรายวิชาและการเลือกเรียนในระบบออนไลน์ เช่น Sakboonyarat and Tantatsanawong (2022) ได้พัฒนาระบบแนะนำรายวิชาในสภาพแวดล้อม Massive Open Online Courses (MOOCs) โดยใช้เทคนิค Big Data ร่วมกับ Deep Neural Network และ Long Short-Term Memory (LSTM) บนสถาปัตยกรรมที่รองรับการประมวลผลข้อมูลขนาดใหญ่ ผลการทดลองรายงานค่า Precision, Accuracy และ Recall ในระดับสูง รวมถึงช่วยเพิ่มประสิทธิภาพด้านเวลาในการฝึกแบบจำลอง ขณะที่ Herrera Rivera (2020) ได้พัฒนาระบบแนะนำรายวิชาเลือกสำหรับนักศึกษาวิศวกรรมศาสตร์โดยใช้แนวทาง Hybrid Recommender System ที่ผสาน Content-Based Filtering ซึ่งใช้ Natural Language Processing และ TF-IDF กับ Collaborative Filtering และนำผลลัพธ์ไปประมวลผลร่วมกับแบบจำลองจำแนกประเภท เช่น k-NN และ XGBoost โดยรายงานค่า MAP-k และ Precision อยู่ในระดับสูง งานวิจัยดังกล่าวสะท้อนให้เห็นว่าการผสมผสานเทคนิคหลายวิธีสามารถเพิ่มประสิทธิภาพของระบบแนะนำในบริบทการศึกษาได้

เหตุผลที่งานวิจัยนี้เลือกพัฒนาแบบจำลอง Hybrid Filtering โดยผสาน Content-Based Filtering เข้ากับ User-Based CF, Item-Based CF และ SVD เนื่องจากข้อมูลการเลือกหลักสูตรของผู้เรียนมีลักษณะเบาบางสูง ผู้เรียนแต่ละรายเลือกหลักสูตรเพียง 5 รายการจากทั้งหมด 49 หลักสูตร ทำให้แบบจำลอง Collaborative Filtering ที่อาศัยความคล้ายคลึงจากพฤติกรรมในการเลือกหลักสูตรอาจมีข้อมูลร่วมกันไม่เพียงพอ ขณะที่ Content-Based Filtering สามารถใช้ข้อมูลคุณลักษณะของผู้เรียน เช่น สายการเรียน ผลการเรียนเฉลี่ยสะสม วิชาที่ถนัด และทักษะที่สนใจเพื่อช่วยเสริมการคำนวณคะแนนความเหมาะสมของหลักสูตร ดังนั้น การผสาน CB กับ UBCF, IBCF และ SVD จึงมีเป้าหมายเพื่อเปรียบเทียบว่าแนวทางการรวมข้อมูลเชิงคุณลักษณะกับข้อมูลพฤติกรรมรูปแบบใดสามารถลดผลกระทบจากความเบาบางของข้อมูลและเพิ่มประสิทธิภาพของการจัดอันดับหลักสูตรได้ดีที่สุด

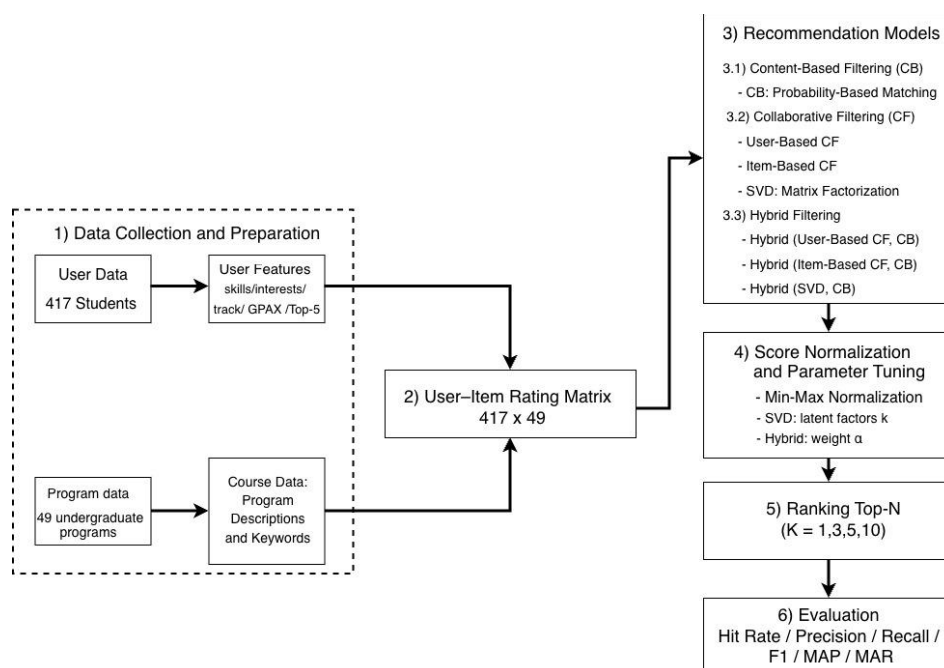
ข้อมูลการเลือกหลักสูตรในงานวิจัยนี้มีลักษณะเฉพาะ กล่าวคือ ผู้เรียนแต่ละรายเลือกหลักสูตรที่สนใจจริงเพียง 5 อันดับ จากหลักสูตรทั้งหมด 49 หลักสูตร ของมหาวิทยาลัยราชภัฏพระนครศรีอยุธยา ทำให้เมทริกซ์ผู้ใช้-หลักสูตรมีข้อมูลที่ไม่น่าเป็นศูนย์จำนวนมากน้อยและมีความเบาบางสูง ลักษณะดังกล่าวอาจส่งผลกระทบต่อประสิทธิภาพของวิธี Collaborative Filtering แบบดั้งเดิม โดยเฉพาะวิธีที่อาศัยความคล้ายคลึงระหว่างผู้ใช้ เนื่องจากข้อมูลที่ซ้อนทับกันระหว่างผู้ใช้อาจมีจำนวนจำกัด (Aggarwal, 2016) ดังนั้น การประเมินระบบแนะนำในงานวิจัยนี้จึงควรใช้ตัวชี้วัดที่เหมาะสมกับงานจัดอันดับ เช่น Hit Rate, Precision, Recall, F1-score, Mean Average Precision และ Mean Average Recall มากกว่าการพิจารณาค่าความถูกต้องแบบการจำแนกประเภทเพียงอย่างเดียว (Herlocker et al., 2004; Aggarwal, 2016)

จากการทบทวนงานวิจัยที่เกี่ยวข้อง พบว่างานวิจัยด้านระบบแนะนำได้รับการศึกษาอย่างแพร่หลายทั้งในต่างประเทศและประเทศไทย แต่การเปรียบเทียบประสิทธิภาพของแบบจำลองการกรองข้อมูลหลายรูปแบบสำหรับระบบแนะนำหลักสูตรระดับปริญญาตรีในบริบทของมหาวิทยาลัยไทยยังมีอยู่อย่างจำกัด งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อเปรียบเทียบ

ประสิทธิภาพของแบบจำลองการกรองข้อมูลสำหรับระบบแนะนำหลักสูตรระดับปริญญาตรี ภายใต้ข้อจำกัดความเบาบางของข้อมูล โดยเปรียบเทียบแบบจำลองระบบแนะนำจำนวน 7 แบบจำลอง ซึ่งจัดเป็น 3 วิธีหลัก ได้แก่ 1) Content-Based Filtering ประกอบด้วย Content-Based Filtering ด้วยแนวทาง Probability-Based Matching 2) Collaborative Filtering ประกอบด้วย User-Based Collaborative Filtering, Item-Based Collaborative Filtering และ Singular Value Decomposition (SVD) และ 3) Hybrid Filtering ประกอบด้วย Hybrid (User-Based CF, CB), Hybrid (Item-Based CF, CB) และ Hybrid (SVD, CB) การประเมินผลใช้แนวทาง Top-N Recommendation ที่ค่า $K = 1, 3, 5$ และ 10 เพื่อระบุแบบจำลองที่ให้ผลดีที่สุดภายใต้ชุดข้อมูลและเงื่อนไขการทดลองของงานวิจัยนี้

วิธีการวิจัย

การวิจัยครั้งนี้ออกแบบขั้นตอนการดำเนินงาน 6 ขั้นตอน ตั้งแต่การเก็บรวบรวมและเตรียมข้อมูล การสร้าง User-Item Rating Matrix การพัฒนาแบบจำลองระบบแนะนำ การปรับสเกลคะแนนและการปรับค่าพารามิเตอร์ของแบบจำลอง การแบ่งข้อมูลและการจัดอันดับรายการแนะนำแบบ Top-N Recommendation และการประเมินประสิทธิภาพของแบบจำลอง ขั้นตอนโดยรวมของการทดลองสรุปได้ดังภาพที่ 1



ภาพที่ 1 ขั้นตอนการทดลองของระบบแนะนำหลักสูตรระดับปริญญาตรี

1. การเก็บรวบรวมและเตรียมข้อมูล

ข้อมูลที่ใช้ในการวิจัยได้จากแบบสอบถามนักเรียนระดับมัธยมศึกษาตอนปลาย ชั้นมัธยมศึกษาปีที่ 5-6 จำนวน 417 คน ที่เข้าร่วมกิจกรรมแนะแนวการศึกษาต่อของมหาวิทยาลัยราชภัฏพระนครศรีอยุธยา โดยใช้วิธีการคัดเลือกแบบเจาะจง ผู้ตอบแบบสอบถามเลือกหลักสูตรที่สนใจจริงจำนวน 5 อันดับ จากหลักสูตรระดับปริญญาตรีทั้งหมด 49 หลักสูตร ครอบคลุม 4 คณะ ได้แก่ คณะครุศาสตร์ คณะวิทยาการจัดการ คณะวิทยาศาสตร์และเทคโนโลยี และคณะมนุษยศาสตร์และสังคมศาสตร์

ข้อมูลที่น่ามาใช้ประกอบด้วย 2 ส่วน ได้แก่ ข้อมูลคุณลักษณะของผู้เรียน เช่น สายการเรียน ผลการเรียนเฉลี่ยสะสม วิชาที่ถนัด และทักษะที่สนใจ และข้อมูลการเลือกหลักสูตรที่ผู้เรียนสนใจจริง 5 อันดับ ซึ่งใช้เป็นข้อมูลพฤติกรรมความสนใจของผู้เรียนและใช้เป็นชุดรายการอ้างอิงสำหรับการประเมินผล

ข้อมูลคุณลักษณะของผู้เรียนถูกเตรียมให้อยู่ในรูปแบบที่สามารถนำไปใช้ในการคำนวณคะแนนของแบบจำลองได้ โดยตัวแปรสายการเรียนถูกจัดเป็นตัวแปรเชิงกลุ่ม เช่น วิทยาศาสตร์-คณิตศาสตร์ ศิลป์-ภาษา หรือสายอื่น ๆ ผลการเรียนเฉลี่ยสะสมถูกจัดเป็นช่วงระดับคะแนนหรือใช้เป็นตัวแปรตามรูปแบบข้อมูลที่เก็บได้ ส่วนวิชาที่ถนัดและทักษะที่สนใจถูกแปลงเป็นข้อมูลแบบหลายตัวเลือก เพื่อใช้คำนวณความสัมพันธ์ระหว่างคุณลักษณะของผู้เรียนกับหลักสูตรที่ถูกเลือกจริง ทั้งนี้ การแปลงข้อมูลดังกล่าวมีวัตถุประสงค์เพื่อให้แบบจำลอง Content-Based Filtering สามารถคำนวณความน่าจะเป็นของหลักสูตรภายใต้ค่าคุณลักษณะของผู้เรียนแต่ละรายได้อย่างเป็นระบบ

ตารางที่ 1 การเตรียมข้อมูลและการแปลงข้อมูลคุณลักษณะของผู้เรียนสำหรับการสร้างแบบจำลองระบบแนะนำ

ตัวแปร	ลักษณะข้อมูลเดิม	วิธีเตรียมข้อมูล
สายการเรียน	ข้อมูลเชิงกลุ่ม	จัดกลุ่มตามประเภทสายการเรียน
ผลการเรียนเฉลี่ยสะสม	ข้อมูลตัวเลข	ใช้ค่าตัวเลขหรือจัดเป็นช่วง GPA
วิชาที่ถนัด	เลือกได้หนึ่งหรือหลายรายการ	แปลงเป็นกลุ่มคุณลักษณะของผู้เรียน
ทักษะที่สนใจ	เลือกได้หนึ่งหรือหลายรายการ	แปลงเป็นกลุ่มทักษะเพื่อใช้คำนวณร่วมกับหลักสูตร
หลักสูตรที่สนใจ	อันดับ 1-5	แปลงเป็นคะแนน 5 4 3 2 1

2. การสร้างเมทริกซ์ผู้ใช้-หลักสูตร

ข้อมูลการเลือกหลักสูตรถูกแปลงเป็นคะแนนเชิงตัวเลขตามลำดับความสนใจ โดยกำหนดให้อันดับที่ 1 เท่ากับ 5 คะแนน อันดับที่ 2 เท่ากับ 4 คะแนน อันดับที่ 3 เท่ากับ 3 คะแนน อันดับที่ 4 เท่ากับ 2 คะแนน อันดับที่ 5 เท่ากับ 1 คะแนน และหลักสูตรที่ไม่ได้ถูกเลือกกำหนดเป็น 0 คะแนน จากนั้นสร้างเมทริกซ์ผู้ใช้-หลักสูตรขนาด 417×49 เพื่อใช้เป็นข้อมูลนำเข้าสำหรับแบบจำลองระบบแนะนำ

เนื่องจากผู้เรียนแต่ละรายเลือกหลักสูตรเพียง 5 รายการจากหลักสูตรทั้งหมด 49 หลักสูตร เมทริกซ์ผู้ใช้-หลักสูตรจึงมีค่าที่ไม่เป็นศูนย์จำนวน 2,085 ค่า จากจำนวนช่องข้อมูลทั้งหมด 20,433 ช่อง หรือมีความเบาบางของข้อมูลประมาณร้อยละ 89.8 ดังสมการที่ (1)

$$Sparsity = \frac{20,433 - 2,085}{20,433} \times 100 \approx 89.8 \quad (1)$$

ลักษณะข้อมูลดังกล่าวเป็นเหตุผลสำคัญที่งานวิจัยนี้เปรียบเทียบทั้งแบบจำลอง Collaborative Filtering แบบดั้งเดิม แบบจำลอง SVD และแบบจำลองแบบผสมผสาน เพื่อประเมินว่าแบบจำลองใดเหมาะสมกับข้อมูลความสนใจที่มีลักษณะเบาบางสูง

3. แบบจำลองระบบแนะนำที่ใช้ในการทดลอง

งานวิจัยนี้เปรียบเทียบแบบจำลองระบบแนะนำหลักสูตรจำนวน 7 แบบจำลอง โดยจัดลำดับการนำเสนอเป็น 3 วิธีหลัก ได้แก่ 1) Content-Based Filtering จำนวน 1 แบบจำลอง คือ Content-Based Filtering ด้วยแนวทาง Probability-Based Matching 2) Collaborative Filtering จำนวน 3 แบบจำลอง ได้แก่ User-Based Collaborative Filtering, Item-Based Collaborative Filtering และ Singular Value Decomposition (SVD) ซึ่งเป็นแนวทางแบบ Model-Based Collaborative Filtering และ 3) Hybrid Filtering จำนวน 3 แบบจำลอง ได้แก่ Hybrid (User-Based CF, CB), Hybrid

(Item-Based CF, CB) และ Hybrid (SVD, CB) การจัดกลุ่มดังกล่าวมีวัตถุประสงค์เพื่อให้เห็นความสัมพันธ์ระหว่างวิธีหลักและแบบจำลองย่อยที่ใช้ในการทดลองอย่างชัดเจน และให้สอดคล้องกับกรอบการวิจัยในภาพที่ 1

Content-Based Filtering ใช้ข้อมูลคุณลักษณะของผู้เรียนเพื่อคำนวณความเหมาะสมระหว่างผู้เรียนกับหลักสูตร โดยใช้แนวทาง Probability-Based Matching จากข้อมูลจริงของผู้เรียนและหลักสูตรที่ถูกเลือก ตัวแปรคุณลักษณะที่ใช้ประกอบด้วยสายการเรียน ผลการเรียนเฉลี่ยสะสม วิชาที่ถนัด และทักษะที่สนใจ จากนั้นคำนวณความน่าจะเป็นของหลักสูตรภายใต้ค่าคุณลักษณะของผู้เรียน และรวมคะแนนตามน้ำหนักของแต่ละคุณลักษณะ ดังสมการที่ (2)

$$Score_{CB}(u, i) = \sum_f w_f \times P(i | value_f(u)) \quad (2)$$

โดยที่ $Score_{CB}(u, i)$ คือคะแนนของหลักสูตร i สำหรับผู้ใช้ u , i คือคุณลักษณะของผู้เรียน, w_f คือน้ำหนักของคุณลักษณะ และ $P(i | value_f(u))$ คือความน่าจะเป็นที่หลักสูตร i จะถูกเลือกเมื่อผู้เรียนมีค่าคุณลักษณะ f ตามที่กำหนด

Collaborative Filtering ใช้ข้อมูลพฤติกรรมในการเลือกหลักสูตรของผู้เรียนเพื่อสร้างคำแนะนำ โดย User-Based Collaborative Filtering ประเมินค่าความสนใจของผู้ใช้เป้าหมายจากกลุ่มผู้ใช้ที่มีรูปแบบการเลือกหลักสูตรคล้ายกัน ส่วน Item-Based Collaborative Filtering สร้างคำแนะนำจากความคล้ายคลึงระหว่างหลักสูตรที่มีรูปแบบการถูกเลือกร่วมกัน ในการคำนวณความคล้ายคลึง ผู้วิจัยใช้ Adjusted Cosine Similarity เพื่อลดผลกระทบจากรูปแบบคะแนนที่เกิดจากการแปลงอันดับเป็นคะแนน และเพื่อให้เหมาะสมกับข้อมูลที่มีลักษณะเป็น sparse matrix

SVD เป็นวิธีในกลุ่ม Model-Based Collaborative Filtering ที่ใช้การแยกเมทริกซ์ผู้ใช้-หลักสูตรเพื่อเรียนรู้ความสัมพันธ์แฝงระหว่างผู้เรียนกับหลักสูตร ดังสมการที่ (3)

$$R \approx UV^T \quad (3)$$

โดยที่ R คือเมทริกซ์ผู้ใช้-หลักสูตร, U คือเมทริกซ์ Latent Factors ของผู้ใช้, Σ คือเมทริกซ์ค่าเอกฐาน และ V^T คือเมทริกซ์ Latent Factors ของหลักสูตร หลังจากได้เมทริกซ์ประมาณค่าแล้ว ระบบนำคะแนนทำนายมาจัดอันดับหลักสูตรสำหรับผู้เรียนแต่ละราย

4. การปรับสเกลคะแนนและการปรับค่าพารามิเตอร์ของแบบจำลอง

ภายหลังจากการกำหนดแบบจำลองระบบแนะนำทั้ง 7 แบบจำลองแล้ว ผู้วิจัยดำเนินการปรับสเกลคะแนนและปรับค่าพารามิเตอร์ของแบบจำลอง เพื่อให้การเปรียบเทียบประสิทธิภาพของแบบจำลองเป็นไปอย่างเป็นระบบ โดยในกรณีของแบบจำลอง Hybrid Filtering ทั้ง 3 รูปแบบ ได้แก่ Hybrid (User-Based CF, CB), Hybrid (Item-Based CF, CB) และ Hybrid (SVD, CB) ผู้วิจัยนำคะแนนจากแบบจำลอง Collaborative Filtering หรือ SVD มารวมกับคะแนนจาก Content-Based Filtering ภายหลังจากการปรับสเกลคะแนนด้วยวิธี Min-Max Normalization เพื่อให้คะแนนจากแต่ละแบบจำลองอยู่ในช่วงเดียวกันก่อนนำมารวมกัน จากนั้นจึงคำนวณคะแนนแบบผสมผสานตามค่าน้ำหนัก α สมการที่ (4)

$$Score_{Hybrid}(u, i) = \alpha \times CF_{norm}(u, i) + (1 - \alpha) \times CB_{norm}(u, i) \quad (4)$$

โดยที่ $CF_{norm}(u, i)$ คือคะแนนจาก User-Based CF, Item-Based CF หรือ SVD หลังการปรับสเกล สำหรับผู้ใช้ u และหลักสูตร i ส่วน $CB_{norm}(u, i)$ คือคะแนนจาก Content-Based Filtering หลังการปรับสเกล และ α คือน้ำหนักของคะแนนฝั่ง Collaborative Filtering หรือ SVD ก่อนการเปรียบเทียบประสิทธิภาพของแบบจำลอง ผู้วิจัยได้ดำเนินการปรับค่าพารามิเตอร์ของแบบจำลองที่สำคัญ ได้แก่ จำนวน Latent Factors ของแบบจำลอง SVD และค่าน้ำหนัก α ของแบบจำลอง Weighted Hybrid Filtering โดยใช้ค่า Precision@5 เป็นเกณฑ์หลักในการคัดเลือกค่าที่เหมาะสม เนื่องจากข้อมูลรายการอ้างอิงของผู้เรียนแต่ละรายประกอบด้วยหลักสูตรที่เลือกจริงจำนวน 5 อันดับ ดังนั้น ค่า $K=5$

จึงสอดคล้องกับจำนวนรายการอ้างอิงที่ใช้ในการประเมิน และสามารถสะท้อนความสามารถของแบบจำลองในการแนะนำหลักสูตรที่ตรงกับความสนใจของผู้เรียนภายในจำนวนรายการที่เท่ากับข้อมูลจริงได้อย่างเหมาะสม อย่างไรก็ตาม เพื่อไม่ให้เกิดการประเมินผลจำกัดอยู่เฉพาะค่า $K=5$ งานวิจัยนี้จึงประเมินผลเพิ่มเติมที่ $K=1, 3$ และ 10 เพื่อพิจารณาแนวโน้มของประสิทธิภาพเมื่อจำนวนรายการแนะนำเปลี่ยนแปลงไป สำหรับแบบจำลอง SVD ผู้วิจัยทดลองจำนวน Latent Factors หรือค่า k เท่ากับ $5, 10, 15, 20, 25$ และ 30 ผลการทดลองพบว่า SVD ให้ค่า Precision@5 สูงสุดเมื่อกำหนดค่า $k=30$ จึงเลือกใช้ค่านี้ในการเปรียบเทียบประสิทธิภาพกับแบบจำลองอื่นในขั้นตอนถัดไป ส่วนแบบจำลอง Weighted Hybrid Filtering ผู้วิจัยทดลองค่า α ตั้งแต่ 0.00 ถึง 1.00 โดยเพิ่มครั้งละ 0.05 เพื่อหาสัดส่วนการผสมคะแนนที่เหมาะสมระหว่างคะแนนจาก Collaborative Filtering หรือ SVD กับคะแนนจาก Content-Based Filtering ผลการทดลองพบว่า Hybrid (SVD, CB) ให้ผลดีที่สุดที่ $\alpha = 0.80$, Hybrid (Item-Based CF, CB) ให้ผลดีที่สุดที่ $\alpha = 0.40$ และ Hybrid (User-Based CF, CB) ให้ผลดีที่สุดที่ $\alpha = 0.10$ ทั้งนี้ งานวิจัยนี้ใช้ K ตัวพิมพ์ใหญ่ หมายถึง จำนวนรายการที่ระบบแนะนำในการประเมินแบบ Top-N Recommendation ส่วน k ตัวพิมพ์เล็ก หมายถึง จำนวน Latent Factors ของ SVD และ α หมายถึง ค่าน้ำหนักในการรวมคะแนนของแบบจำลอง Weighted Hybrid Filtering เพื่อป้องกันความสับสนในการตีความผลการทดลอง

5. การแบ่งข้อมูลและการจัดอันดับรายการแนะนำแบบ Top-N Recommendation

งานวิจัยนี้ใช้แนวทาง Full-Ranking Protocol สำหรับการประเมินแบบ Top-N Recommendation โดยกำหนดให้หลักสูตรที่ผู้เรียนเลือกจริงทั้ง 5 อันดับเป็นชุดรายการอ้างอิง จากนั้นให้แบบจำลองจัดอันดับหลักสูตรระดับปริญญาตรีทั้งหมด 49 หลักสูตรสำหรับผู้เรียนแต่ละราย และพิจารณาว่าหลักสูตรที่อยู่ในลำดับ Top-K ตรงกับรายการหลักสูตรที่ผู้เรียนเลือกจริงหรือไม่ โดยกำหนดค่า $K=1, 3, 5$ และ 10 การประเมินลักษณะนี้มีเป้าหมายเพื่อพิจารณาความสามารถของแบบจำลองในการนำหลักสูตรที่เกี่ยวข้องกับความสนใจของผู้เรียนขึ้นมาอยู่ในลำดับต้นของรายการแนะนำ ทั้งนี้ งานวิจัยได้ใช้การตรวจสอบผลด้วย 5-Fold Cross Validation ร่วมด้วย เพื่อประเมินความสม่ำเสมอของผลการทดลองภายใต้การแบ่งข้อมูลเป็นหลายรอบ

6. การประเมินประสิทธิภาพของแบบจำลอง

งานวิจัยนี้ใช้หลักสูตรที่ผู้เรียนเลือกจริงทั้ง 5 อันดับเป็นชุดรายการอ้างอิงสำหรับการประเมินผล โดยกำหนดค่า $K=1, 3, 5$ และ 10 ตัวชี้วัดที่ใช้ในการประเมินประกอบด้วย Hit Rate@K, Precision@K, Recall@K, F1@K, Mean Average Precision@K (MAP@K) และ Mean Average Recall@K (MAR@K) โดย Hit Rate@K ใช้วัดว่ารายการแนะนำ Top-K มีหลักสูตรที่ตรงกับชุดรายการอ้างอิงอย่างน้อย 1 รายการหรือไม่ Precision@K ใช้วัดสัดส่วนของรายการที่แนะนำแล้วตรงกับชุดรายการอ้างอิง Recall@K ใช้วัดสัดส่วนของรายการอ้างอิงที่ระบบค้นพบ F1@K ใช้วัดสมมูลระหว่าง Precision@K และ Recall@K ส่วน MAP@K และ MAR@K ใช้ประเมินคุณภาพของลำดับรายการแนะนำ จากรูปแบบการวิจัยแบบ Top-N Recommendation ภายใต้ข้อมูลที่มีความเบาบางสูง การใช้เพียงค่า Accuracy จะไม่สะท้อนประสิทธิภาพที่แท้จริงเนื่องจากไม่คำนึงถึงลำดับ จึงมีการวัดประสิทธิภาพ MAP@K และ MAR@K เพื่อประเมินคุณภาพเชิงลำดับของรายการแนะนำ

ผลการวิจัย และวิจารณ์ผลการวิจัย

ผลการวิจัย

ผลของงานวิจัยนี้นำเสนอเพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองการกรองข้อมูลสำหรับระบบแนะนำหลักสูตรระดับปริญญาตรี โดยพิจารณาทั้งผลการปรับค่าพารามิเตอร์ ผลการเปรียบเทียบประสิทธิภาพของแบบจำลอง และการอภิปรายผลภายใต้บริบทของข้อมูลที่มีความเบาบางสูง

1. ผลการปรับค่าพารามิเตอร์ของแบบจำลองก่อนการเปรียบเทียบประสิทธิภาพของแบบจำลอง ผู้วิจัยได้ปรับค่าพารามิเตอร์ของแบบจำลองที่สำคัญ ได้แก่ จำนวน Latent Factors ของ SVD และค่าน้ำหนัก α ของ Weighted Hybrid

Filtering โดยใช้ค่า Precision@5 เป็นเกณฑ์หลักในการเลือกค่าที่เหมาะสม เนื่องจากข้อมูลรายการอ้างอิงของผู้เรียนแต่ละรายประกอบด้วยหลักสูตรที่เลือกจริงจำนวน 5 อันดับ ผลการทดลองพบว่า SVD ให้ค่า Precision@5 สูงสุดเมื่อกำหนดจำนวน Latent Factors เท่ากับ $k=30$ ดังนั้นจึงเลือกค่า $k=30$ สำหรับการเปรียบเทียบแบบจำลองในขั้นตอนถัดไป สำหรับ Weighted Hybrid Filtering พบว่า Hybrid (SVD, CB) ให้ผลดีที่สุดที่ $\alpha = 0.80$, Hybrid (Item-Based CF, CB) ให้ผลดีที่สุดที่ $\alpha = 0.40$ และ Hybrid (User-Based CF, CB) ให้ผลดีที่สุดที่ $\alpha = 0.10$ ผลดังกล่าวสะท้อนว่าการผสมผสานคะแนนจาก SVD กับ Content-Based Filtering ในสัดส่วนที่ให้น้ำหนักกับ SVD สูงกว่าสามารถเพิ่มประสิทธิภาพของระบบแนะนำได้ดีที่สุดภายใต้ชุดข้อมูลนี้

ตารางที่ 2 ผลการปรับค่าพารามิเตอร์ของแบบจำลองโดยใช้ Precision@5 เป็นเกณฑ์

Model	Parameter	Tested values	Best value / setting	Selection criterion
SVD	Latent factors (k)	5 10 15 20 25 30	30	Precision@5
Hybrid (User-Based CF, CB)	Weight (α)	0.00–1.00, step 0.05	0.10	Precision@5
Hybrid (Item-Based CF, CB)	Weight (α)	0.00–1.00, step 0.05	0.40	Precision@5
Hybrid (SVD, CB)	Weight (α)	0.00–1.00, step 0.05	0.80	Precision@5

2. ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองการประเมินผลใช้แนวทาง Top-N Recommendation ที่ค่า $K=1, 3, 5$ และ 10 โดยใช้ตัวชี้วัด Hit Rate@K, Precision@K, Recall@K, F1@K, MAP@K และ MAR@K เพื่อพิจารณาทั้งความถูกต้อง ความครอบคลุม และคุณภาพของลำดับรายการแนะนำ เมื่อพิจารณาผลการทดลองที่ค่า $K=5$ ซึ่งสอดคล้องกับรายการหลักสูตรที่ผู้เรียนเลือกจริงจำนวน 5 อันดับ พบว่า Weighted Hybrid Filtering แบบ Hybrid (SVD, CB) ที่ $\alpha = 0.80$ ให้ผลดีที่สุดในการรวม โดยมีค่า Hit Rate@5 = 1.0000, Precision@5 = 0.8590, Recall@5 = 0.9153, F1@5 = 0.8863, MAP@5 = 0.9089 และ MAR@5 = 0.5367 รองลงมาคือ SVD ที่ $k=30$ ขณะที่ User-Based Collaborative Filtering ให้ผลต่ำที่สุดในหลายตัวชี้วัด รายละเอียดผลการเปรียบเทียบเชิงตัวเลขที่ค่า $K=5$ สรุปได้ดังตารางที่ 3

ตารางที่ 3 ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองที่ค่า $K = 5$

Method	Hit Rate@5	Precision@5	Recall@5	F1@5	MAP@5	MAR@5
Hybrid (SVD, $\alpha = 0.80$)	1.0000	0.8590	0.9153	0.8863	0.9089	0.5367
SVD ($k=30$)	1.0000	0.8547	0.9110	0.8820	0.9040	0.5326
Hybrid (IB, $\alpha = 0.40$)	1.0000	0.6499	0.7005	0.6742	0.6456	0.3463
Item-Based CF	0.9832	0.6120	0.6485	0.6297	0.5307	0.3032
Hybrid (UB, $\alpha = 0.10$)	0.9329	0.3765	0.4227	0.3982	0.3396	0.1777
Content-Based	0.9113	0.3751	0.4218	0.3971	0.3296	0.1802
User-Based CF	0.9209	0.2926	0.3020	0.2972	0.1664	0.0874

จากตารางที่ 3 พบว่า Hybrid (SVD, CB) ที่ $\alpha = 0.80$ ให้ผลดีที่สุดโดยรวม โดยมีค่า Precision@5 = 0.8590, Recall@5 = 0.9153, F1@5 = 0.8863, MAP@5 = 0.9089 และ MAR@5 = 0.5367 สูงกว่าแบบจำลองอื่น ขณะที่ SVD ($k=30$) ให้ผลรองลงมาและมีค่าตัวชี้วัดใกล้เคียงกับ Hybrid (SVD, CB) ในหลายมิติ ได้แก่ Hit Rate@5 เท่ากันที่ 1.0000, Precision@5 ต่างกัน 0.0043, Recall@5 ต่างกัน 0.0043, F1@5 ต่างกัน 0.0043, MAP@5 ต่างกัน 0.0049 และ MAR@5 ต่างกัน 0.0041 ผลดังกล่าวแสดงให้เห็นว่า SVD เป็นแบบจำลองเดี่ยวที่มีประสิทธิภาพสูงภายใต้ข้อมูลที่มีความเบาบาง แต่การผสมผสานกับ Content-Based Filtering ในรูปแบบ Hybrid ยังช่วยเพิ่มค่าตัวชี้วัดได้เล็กน้อยในทุกมิติการประเมินที่ค่า $K=5$

เมื่อพิจารณาผลในภาพรวมทุกค่า K พบว่า Hybrid (SVD, CB) และ SVD ให้ค่าตัวชี้วัดสูงกว่าแบบจำลองอื่นอย่างต่อเนื่อง โดยเฉพาะในด้าน Precision, Recall, F1 และ MAP ขณะที่เมื่อค่า K เพิ่มขึ้น ค่า Recall และ MAR มีแนวโน้มเพิ่มขึ้น เนื่องจากระบบมีโอกาสค้นพบหลักสูตรที่ผู้เรียนสนใจจริงได้มากขึ้น ส่วนค่า Precision มีแนวโน้มลดลง เนื่องจากจำนวนรายการแนะนำเพิ่มขึ้นและมีโอกาสรวมรายการที่ไม่ตรงกับความต้องการจริงเข้ามามากขึ้น ลักษณะดังกล่าวเป็นไปตามธรรมชาติของการประเมินแบบ Top-N Recommendation แนวโน้มของผลการทดลองเมื่อค่า K เปลี่ยนแปลงสรุปได้ดังตารางที่ 4 และ ตารางที่ 5

แม้ว่า Hybrid (SVD, CB) จะให้ค่าตัวชี้วัดสูงที่สุดในภาพรวม แต่ค่า MAR@5 เท่ากับ 0.5367 ยังอยู่ในระดับปานกลางเมื่อเทียบกับค่า Recall@5 ที่เท่ากับ 0.9153 ประเด็นนี้สะท้อนว่าแบบจำลองสามารถค้นพบหลักสูตรที่ตรงกับความต้องการของผู้เรียนได้ค่อนข้างมากภายในรายการแนะนำ 5 อันดับแรก แต่ลำดับของหลักสูตรที่ตรงกับรายการอ้างอิงอาจยังไม่ได้อยู่ในตำแหน่งต้นสุดครบทุกกรณี กล่าวคือ ระบบมีความสามารถในการนำหลักสูตรที่เกี่ยวข้องเข้าสู่ Top-5 ได้ดี แต่ยังมีข้อจำกัดด้านคุณภาพของการจัดอันดับภายใน Top-5 ดังนั้น ค่า MAR@5 จึงช่วยชี้ให้เห็นว่าการพัฒนาระบบในขั้นต่อไปควรให้ความสำคัญกับการปรับปรุงกลไกการเรียงลำดับรายการแนะนำ เพื่อให้หลักสูตรที่เกี่ยวข้องมากที่สุดถูกจัดให้อยู่ในอันดับต้นมากขึ้น

ตารางที่ 4 ผลการเปรียบเทียบค่า Precision, Recall และ F1-score จำแนกตามแบบจำลองและค่า K

Method	Precision				Recall				F1-score			
	K1	K3	K5	K10	K1	K3	K5	K10	K1	K3	K5	K10
Hybrid(SVD, $\alpha=0.80$)	1.0000	0.9552	0.8590	0.4614	0.2399	0.6271	0.9153	0.9797	0.3870	0.7571	0.8863	0.6273
SVD(k=30)	1.0000	0.9552	0.8547	0.4640	0.2399	0.6271	0.9110	0.9851	0.3870	0.7571	0.8820	0.6309
Hybrid(Item-Based CF, $\alpha=0.40$)	0.9257	0.7970	0.6499	0.4060	0.2248	0.5257	0.7005	0.8676	0.3618	0.6335	0.6742	0.5532
Item-Based CF	0.6906	0.7050	0.6120	0.4329	0.1408	0.4439	0.6485	0.9216	0.2339	0.5448	0.6297	0.5890
Hybrid(User-Based CF, $\alpha=0.10$)	0.6571	0.4708	0.3765	0.2676	0.1639	0.3245	0.4227	0.5851	0.2623	0.3842	0.3982	0.3673
Content-Based	0.5803	0.4516	0.3751	0.2758	0.1478	0.3149	0.4218	0.6014	0.2357	0.3711	0.3971	0.3782
User-Based CF	0.2206	0.3269	0.2926	0.1885	0.0458	0.2032	0.3020	0.3880	0.0759	0.2507	0.2972	0.2537

ตารางที่ 5 ผลการเปรียบเทียบค่า MAP, MAR และ Hit Rate จำแนกตามแบบจำลองและค่า K

Method	MAP				MAR				Hit Rate			
	K1	K3	K5	K10	K1	K3	K5	K10	K1	K3	K5	K10
Hybrid(SVD, $\alpha=0.80$)	1.0000	0.9851	0.9089	0.9558	0.2399	0.4291	0.5367	0.5997	1.0000	1.0000	1.0000	1.0000
SVD(k=30)	1.0000	0.9851	0.9040	0.9572	0.2399	0.4291	0.5326	0.6051	1.0000	1.0000	1.0000	1.0000
Hybrid(Item-Based CF, $\alpha=0.40$)	0.9257	0.7927	0.6456	0.7391	0.2248	0.3308	0.3463	0.4898	0.9257	1.0000	1.0000	1.0000
Item-Based CF	0.6906	0.6306	0.5307	0.6773	0.1408	0.2570	0.3032	0.5423	0.6906	0.9616	0.9832	1.0000
Hybrid(User-Based CF, $\alpha=0.10$)	0.6571	0.4408	0.3396	0.4021	0.1639	0.1803	0.1777	0.2754	0.6571	0.8609	0.9329	0.9880
Content-Based	0.5803	0.4119	0.3296	0.4025	0.1478	0.1753	0.1802	0.2923	0.5803	0.8393	0.9113	0.9784
User-Based CF	0.2206	0.2057	0.1664	0.1921	0.0458	0.0828	0.0874	0.1263	0.2206	0.7746	0.9209	0.9544

วิจารณ์ผลการวิจัย

ผลการวิจัยพบว่าแบบจำลอง Hybrid (SVD, CB) ให้ประสิทธิภาพดีที่สุดโดยรวม โดยเฉพาะค่า Precision@5, Recall@5, F1@5 และ MAP@5 ผลดังกล่าวสะท้อนให้เห็นว่าการผสาน SVD กับ Content-Based Filtering สามารถช่วยลดข้อจำกัดของการใช้แบบจำลองเดี่ยวได้ ซึ่งสอดคล้องกับแนวคิดของ Hybrid Recommender Systems ที่มุ่งรวมจุดเด่นของหลายวิธีเพื่อเพิ่มคุณภาพของคำแนะนำ (Burke, 2002) และสอดคล้องกับงานทบทวนสมัยใหม่ที่ชี้ว่าระบบแนะนำแบบผสมผสานมีความยืดหยุ่นและสามารถปรับใช้ได้ดีในบริบทที่ข้อมูลมีความซับซ้อน (Sabiri et al., 2025) เมื่อเปรียบเทียบกับ

แบบจำลองเดี่ยว พบว่า SVD ให้ผลรองลงมาและมีประสิทธิภาพสูงกว่า User-Based Collaborative Filtering และ Item-Based Collaborative Filtering ทั้งนี้เนื่องจาก SVD สามารถเรียนรู้ความสัมพันธ์แฝงระหว่างผู้เรียนกับหลักสูตรจากเมทริกซ์ที่มีความเบาบางได้ดีกว่าวิธีที่อาศัยการคำนวณความคล้ายคลึงโดยตรง (Koren et al., 2009; Aggarwal, 2016) ประเด็นนี้สอดคล้องกับงานทบทวนล่าสุดที่ระบุว่าความเบาบางของข้อมูลเป็นปัญหาสำคัญที่ส่งผลต่อความแม่นยำ ความครอบคลุม และความสามารถในการสร้างคำแนะนำของระบบแนะนำ (Gunathilaka et al., 2025)

ผลการวิจัยยังสอดคล้องกับงานวิจัยด้านระบบแนะนำในบริบทการศึกษา ซึ่งพบว่าการเปรียบเทียบหลายแนวทาง เช่น User-Based, Item-Based และ Hybrid Recommendation มีความสำคัญต่อการเลือกแบบจำลองที่เหมาะสมกับข้อมูลและเป้าหมายการใช้งาน (Lahoud et al., 2023) รวมถึงงานระบบแนะนำรายวิชาที่พบว่าแนวทางแบบผสมผสานสามารถช่วยสนับสนุนการเลือกเรียนของผู้เรียนได้อย่างมีประสิทธิภาพมากขึ้น (Zhu and Wang, 2022) นอกจากนี้ งานวิจัยในประเทศไทยยังพบว่าแนวทางแบบผสมผสานด้วยวิธีถ่วงน้ำหนักสามารถให้ผลดีกว่าวิธีเดียวในบางบริบทของระบบแนะนำ (ภัทร และคณะ, 2564) อย่างไรก็ตาม ผลการวิจัยนี้ควรตีความภายใต้บริบทของชุดข้อมูลที่ใช้ในการทดลอง เนื่องจากกลุ่มตัวอย่างเป็นนักเรียนที่เข้าร่วมกิจกรรมแนะนำของมหาวิทยาลัย และข้อมูลความสนใจมาจากการเลือกหลักสูตรเพียง 5 อันดับจากทั้งหมด 49 หลักสูตร ส่งผลให้เมทริกซ์ผู้ใช้-หลักสูตรมีความเบาบางสูง ดังนั้น การนำแบบจำลองไปใช้จริงควรทดสอบกับข้อมูลที่หลากหลายมากขึ้น เช่น หลายปีการศึกษา หลายพื้นที่ หรือกลุ่มผู้เรียนที่ยังไม่ตัดสินใจเลือกสถาบัน เพื่อประเมินความสามารถของแบบจำลองในบริบทที่กว้างขึ้น ผลที่ User-Based Collaborative Filtering และ Item-Based Collaborative Filtering ให้ค่าตัวชี้วัดต่ำกว่า SVD และ Hybrid (SVD, CB) สามารถอธิบายได้จากลักษณะของข้อมูลที่มีความเบาบางสูง เนื่องจากผู้เรียนแต่ละรายเลือกหลักสูตรเพียง 5 รายการจากทั้งหมด 49 หลักสูตร ทำให้เมทริกซ์ผู้ใช้-หลักสูตรมีข้อมูลที่ไม่เป็นศูนย์จำนวนน้อย ข้อมูลที่ซ้อนทับกันระหว่างผู้ใช้และข้อมูลการถูกเลือกร่วมกันระหว่างหลักสูตร จึงมีจำกัด ส่งผลให้การคำนวณความคล้ายคลึงของ User-Based และ Item-Based Collaborative Filtering ทำได้ไม่สมบูรณ์เท่าที่ควร ขณะที่ SVD สามารถเรียนรู้ความสัมพันธ์แฝงจากโครงสร้างของเมทริกซ์ได้ดีกว่า และเมื่อผสมกับ Content-Based Filtering ในรูปแบบ Hybrid (SVD, CB) จึงช่วยลดผลกระทบจากความเบาบางของข้อมูลและให้ประสิทธิภาพการแนะนำสูงกว่าแบบจำลองเดี่ยวในชุดข้อมูลนี้

สำหรับการเลือกค่า α ของแบบจำลอง Hybrid แต่ละรูปแบบ พบว่าค่า α ที่เหมาะสมแตกต่างกัน ได้แก่ Hybrid (User-Based CF, CB) มีค่า $\alpha = 0.10$, Hybrid (Item-Based CF, CB) มีค่า $\alpha = 0.40$ และ Hybrid (SVD, CB) มีค่า $\alpha = 0.80$ ความแตกต่างดังกล่าวไม่ได้เกิดจากการกำหนดค่าน้ำหนักโดยอำเภอใจ แต่เป็นผลจากกระบวนการปรับค่าพารามิเตอร์ภายใต้ช่วงการทดลองเดียวกัน คือ α ตั้งแต่ 0.00 ถึง 1.00 โดยเพิ่มครั้งละ 0.05 และใช้ Precision@5 เป็นเกณฑ์ในการคัดเลือกเหมือนกันทุกแบบจำลอง ดังนั้น ค่า α ที่แตกต่างกันจึงสะท้อนระดับการมีส่วนร่วมของคะแนนจากแบบจำลอง Collaborative Filtering หรือ SVD เมื่อผสมกับคะแนนจาก Content-Based Filtering ในแต่ละกรณี กล่าวคือ Hybrid (SVD, CB) ให้ผลดีที่สุดเมื่อให้น้ำหนักกับ SVD สูงกว่า CB แสดงว่า SVD สามารถเรียนรู้โครงสร้างแฝงของเมทริกซ์ผู้ใช้-หลักสูตรได้ดีภายใต้ข้อมูลชุดนี้ ขณะที่ Hybrid (User-Based CF, CB) ให้ผลดีที่สุดเมื่อให้น้ำหนักกับ CB มากกว่า สะท้อนว่า User-Based CF ได้รับผลกระทบจากความเบาบางของข้อมูลมากกว่า จึงต้องอาศัยข้อมูลคุณลักษณะของผู้เรียนจาก CB มาช่วยเสริมมากขึ้น อย่างไรก็ตาม การเลือกค่า α ที่เหมาะสมเฉพาะแต่ละแบบจำลองอาจทำให้ผลการเปรียบเทียบสะท้อนประสิทธิภาพภายใต้พารามิเตอร์ที่ดีที่สุดของแต่ละวิธี มากกว่าการเปรียบเทียบภายใต้ค่าน้ำหนักเดียวกันทุกแบบจำลอง ดังนั้น งานวิจัยในอนาคตอาจพิจารณาเพิ่มการวิเคราะห์ความไวของค่า α หรือเปรียบเทียบผลภายใต้ค่า α เดียวกัน เพื่อยืนยันความแตกต่างของประสิทธิภาพระหว่างแบบจำลองให้ชัดเจนยิ่งขึ้น

สำหรับการเปรียบเทียบระหว่าง Hybrid (SVD, CB) และ SVD พบว่าค่าตัวชี้วัดหลายรายการมีความใกล้เคียงกัน เช่น Precision@5, Recall@5, F1@5, MAP@5 และ MAR@5 ดังนั้น แม้ Hybrid (SVD, CB) จะให้ค่าตัวชี้วัดสูงกว่าในเชิงตัวเลข แต่การสรุปความแตกต่างของประสิทธิภาพระหว่างแบบจำลองควรพิจารณาด้วยความระมัดระวัง งานวิจัยในอนาคตควรเพิ่มการทดสอบนัยสำคัญทางสถิติ เช่น Paired t-test หรือ Signed-rank test โดยใช้ค่าตัวชี้วัดรายผู้เรียนหรือรายรอบการทดลอง เพื่อยืนยันว่าความแตกต่างของค่าตัวชี้วัดระหว่างแบบจำลองมีนัยสำคัญทางสถิติหรือไม่

โดยสรุปแบบจำลองที่ผสมผสานข้อมูลพฤติกรรมในการเลือกหลักสูตรกับข้อมูลเชิงคุณลักษณะของผู้เรียนสามารถให้ผลการจัดอันดับที่ดีกว่าแบบจำลองเดี่ยวในชุดข้อมูลนี้ โดยเฉพาะในบริบทที่ข้อมูลมีความเบาบางสูง ทั้งนี้ การพัฒนาต่อยอดควรให้ความสำคัญกับการเพิ่มความหลากหลายของข้อมูล การตรวจสอบค่าพารามิเตอร์ และการออกแบบระบบให้สามารถอธิบายเหตุผลของคำแนะนำได้ เพื่อเพิ่มความน่าเชื่อถือและความเหมาะสมต่อการใช้งานจริง

อย่างไรก็ตาม ผลการวิจัยนี้ควรตีความภายใต้ข้อจำกัดของบริบทข้อมูล เนื่องจากชุดข้อมูลที่ใช้ในการทดลองมาจากหลักสูตรระดับปริญญาตรีจำนวน 49 หลักสูตรของมหาวิทยาลัยราชภัฏพระนครศรีอยุธยาเพียงแห่งเดียว และกลุ่มตัวอย่างเป็นนักเรียนที่เข้าร่วมกิจกรรมแนะแนวการศึกษาต่อของมหาวิทยาลัย ดังนั้น ผลลัพธ์ของแบบจำลองอาจสะท้อนรูปแบบความสนใจของผู้เรียนและโครงสร้างหลักสูตรเฉพาะบริบทของสถาบันดังกล่าว ยังไม่สามารถสรุปได้โดยตรงว่าแบบจำลองจะให้ประสิทธิภาพในระดับเดียวกันเมื่อนำไปใช้กับมหาวิทยาลัยอื่นที่มีจำนวนหลักสูตร โครงสร้างคณะ ลักษณะผู้เรียน หรือพฤติกรรมในการเลือกหลักสูตรแตกต่างกัน งานวิจัยในอนาคตจึงควรมีการทดสอบข้ามสถาบัน หรือ Cross-Institutional Validation โดยนำข้อมูลจากหลายมหาวิทยาลัยมาทดลองร่วมกัน หรือฝึกแบบจำลองจากสถาบันหนึ่งแล้วทดสอบกับอีกสถาบันหนึ่ง เพื่อประเมินความสามารถในการนำแบบจำลองไปใช้จริงในบริบทที่กว้างขึ้น

สรุปผลการวิจัย

ผลการวิจัยพบว่า แบบจำลอง Weighted Hybrid Filtering แบบ Hybrid (SVD, CB) ที่กำหนดค่าน้ำหนัก $\alpha = 0.80$ ให้ค่าตัวชี้วัดหลักสูงกว่าแบบจำลองอื่นภายใต้ชุดข้อมูลและเงื่อนไขการทดลองนี้ โดยเฉพาะที่ค่า $K=5$ ซึ่งสอดคล้องกับจำนวนหลักสูตรที่ผู้เรียนเลือกจริง แบบจำลองดังกล่าวให้ค่า Precision@5, Recall@5, F1@5 และ MAP@5 สูงกว่าแบบจำลองอื่น แสดงให้เห็นว่าการผสมผสาน SVD กับ Content-Based Filtering สามารถช่วยเพิ่มประสิทธิภาพในการจัดอันดับหลักสูตรได้ดีกว่าการใช้แบบจำลองเดี่ยว การเปรียบเทียบแบบจำลองระบบแนะนำในหลายรูปแบบภายใต้บริบทด้านการศึกษา โดยใช้ข้อมูลจริงจากนักเรียนระดับมัธยมศึกษาตอนปลาย และใช้ข้อมูลหลักสูตรระดับปริญญาตรีของมหาวิทยาลัยจำนวน 49 หลักสูตรในการทดลอง นอกจากนี้ งานวิจัยยังประเมินประสิทธิภาพของแบบจำลองหลายมิติ ทั้งด้านความถูกต้อง ความครอบคลุม และคุณภาพของลำดับรายการแนะนำ ผ่านตัวชี้วัด Hit Rate, Precision, Recall, F1-score, MAP และ MAR ทำให้สามารถพิจารณาประสิทธิภาพของระบบแนะนำได้รอบด้านมากกว่าการใช้ค่าความถูกต้องเพียงตัวเดียว ผลการวิจัยสะท้อนให้เห็นว่ากรณีที่ข้อมูลความสนใจของผู้เรียนมีลักษณะเบาบางสูง การใช้แบบจำลองแบบผสมผสานที่นำข้อมูลพฤติกรรมในการเลือกหลักสูตรและข้อมูลเชิงคุณลักษณะของผู้เรียนมาประกอบกัน สามารถช่วยเพิ่มคุณภาพของการแนะนำหลักสูตรได้อย่างมีประสิทธิภาพ งานวิจัยนี้จึงสามารถใช้เป็นแนวทางในการพัฒนาระบบสนับสนุนการแนะแนวหลักสูตรระดับปริญญาตรีที่เหมาะสมกับความสนใจและคุณลักษณะของผู้เรียนแต่ละราย ทั้งนี้ การนำไปใช้จริงควรทดสอบกับข้อมูลที่หลากหลายมากขึ้น และควรพัฒนาระบบให้สามารถอธิบายเหตุผลของคำแนะนำได้ เพื่อเพิ่มความน่าเชื่อถือและความเหมาะสมต่อการใช้งานในบริบทการศึกษาไทยต่อไป

ข้อเสนอแนะ

1. ควรนำแบบจำลองที่ให้ผลดีที่สุดไปพัฒนาต่อยอดเป็นระบบสนับสนุนการแนะแนวหลักสูตรระดับปริญญาตรี โดยออกแบบให้ผู้เรียนสามารถรอกข้อมูลความสนใจ ความถนัด ผลการเรียนรู้ และทักษะที่เกี่ยวข้อง เพื่อให้ระบบจัดอันดับหลักสูตรที่เหมาะสมกับผู้เรียนแต่ละรายได้อย่างเป็นระบบ
2. ควรเพิ่มตัวชี้วัดด้านอื่นนอกเหนือจากความแม่นยำของการแนะนำ เช่น ความครอบคลุมของรายการแนะนำ (Coverage), ความหลากหลายของคำแนะนำ (Diversity) เพื่อประเมินว่าระบบสามารถแนะนำหลักสูตรได้ครอบคลุมหลากหลาย และไม่จำกัดอยู่เฉพาะหลักสูตรที่ได้รับความนิยมสูงเท่านั้น ซึ่งจะช่วยให้ระบบแนะนำมีความเหมาะสมต่อการใช้งานจริงมากขึ้น

3. ควรขยายการทดสอบแบบจำลองไปยังข้อมูลหลักสูตรและกลุ่มผู้เรียนจากหลายสถาบัน เพื่อดำเนินการทดสอบข้ามสถาบัน (Cross-Institutional Validation) และประเมินว่าแบบจำลองที่พัฒนาขึ้นสามารถนำไปประยุกต์ใช้ในบริบทของมหาวิทยาลัยที่มีโครงสร้างหลักสูตรและกลุ่มผู้เรียนแตกต่างกันได้หรือไม่

เอกสารอ้างอิง

- ภัทร สุตรสุวรรณ, เดือนเพ็ญ ชีวรรณวิวัฒน์, และพาจิตชนัด ศิริพานิช. (2564). การศึกษาระบบแนะนำสถานที่ท่องเที่ยวใน กรุงเทพฯ และปริมณฑลสำหรับนักท่องเที่ยวต่างชาติ. *วารสารสมาคมนักวิจัย*, 26(3), 121–139.
- สมเพ็ชร จุลลาบุตดี. (2562). การสำรวจงานวิจัยระบบแนะนำในประเทศไทย พ.ศ. 2550–2560. *วารสารสารสนเทศศาสตร์*, 37(3), 95–122.
- วรรณช ศรีพลัง, และวงกต ศรีอุไร. (2560). การพัฒนาระบบแนะนำข้อมูลสำหรับห้องสมุดออนไลน์โดยใช้วิธีการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมและข้อมูลส่วนบุคคล. *วารสารมหาวิทยาลัยศรีนครินทรวิโรฒ สาขาวิทยาศาสตร์และเทคโนโลยี*, 9(18), 150–164.
- Adomavicius, G., and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749. doi: 10.1109/TKDE.2005.99
- Aggarwal, C. C. (2016). **Recommender Systems: The Textbook**. Cham, Switzerland: Springer. doi: 10.1007/978-3-319-29659-3
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331–370. doi: 10.1023/A:1021240730564
- Gunathilaka, T. M. A. U., Manage, P. D., Zhang, J., Li, Y., and Kelly, W. (2025). Addressing sparse data challenges in recommendation systems: A systematic review of rating estimation using sparse rating data and profile enrichment techniques. *Intelligent Systems with Applications*, 25, 200474. doi: 10.1016/j.iswa.2024.200474
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1), 5–53. doi: 10.1145/963770.963772
- Herrera, R. J. E. (2020). A hybrid recommender system to enrollment for elective subjects in engineering students using classification algorithms. *International Journal of Advanced Computer Science and Applications*, 11(7), 400–406. doi: 10.14569/IJACSA.2020.0110752
- Isinkaye, F. O., Folajimi, Y. O., and Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3), 261–273. doi: 10.1016/j.eij.2015.06.005
- Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30–37. doi: 10.1109/MC.2009.263
- Lahoud, C., Moussa, S., Obeid, C., El Khoury, H., and Champin, P.-A. (2023). A comparative analysis of different recommender systems for university major and career domain guidance. *Education and Information Technologies*, 28, 8733–8759. doi: 10.1007/s10639-022-11541-3
- Lops, P., de Gemmis, M., and Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor (Eds.), **Recommender Systems Handbook** (pp. 73–105). New York, NY: Springer. doi: 10.1007/978-0-387-85820-3_3

- Sabiri, B., Khtira, A., El Asri, B., and Rhanoui, M. (2025). Hybrid quality-based recommender systems: A systematic literature review. **Journal of Imaging**, 11(1), 12. doi: 10.3390/jimaging11010012
- Sakboonyarat, S., and Tantatsanawong, P. (2022). Applied big data technique and deep learning for massive open online courses (MOOCs) recommendation system. **ECTI-CIT Transactions**, 16(4), 436–447. doi: 10.37936/ecti-cit.2022164.245873
- Zhu, L., and Wang, B. (2022). Course selection recommendation based on hybrid recommendation algorithms. In **Proceedings of the International Conference on Modern Education and Information Management** (pp. 476–482). doi: 10.2991/978-94-6463-044-2_60