



สมาคมนวัตกรรมและเทคโนโลยีสร้างสรรค์ ร่วมกับ มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ

การประชุมวิชาการระดับชาติเชิงสร้างสรรค์ ครั้งที่ 9

21-23
AUG.
2024

2024
CreTech

Creative
Technology &
Artificial Intelligence
at Cape Racha Hotel, Sriracha, Chon Buri

ระบบแนะนำประกาศขายอสังหาริมทรัพย์แบบไฮบริด
โดยใช้ขั้นตอนวิธีเพื่อนบ้านที่ใกล้ที่สุดและการแยกค่าเอกฐาน
Hybrid Property Post Recommendation System
using K-Nearest Neighbors and Singular Value Decomposition

สมิตรา วงักดี¹, ชัชชญา อินณรงค์¹, สารทูล ขุนสนธิ¹,

วิธวินท์ สุธุทธิ¹, และ วรินทร์ วัฒนพรพรหม^{1*}

Sumitra Wongpakdee¹, Chatchaya Innarong¹, Saratool Khunsanit¹,

Wittawin Susutti¹, and Warin Wattanapornprom^{1*}

¹สาขาวิทยาการคอมพิวเตอร์ประยุกต์ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ประเทศไทย

¹Applied Computer Science Program, Department of Mathematics, Faculty of Science,

King Mongkut's University of Technology Thonburi, Thailand

*Tel: 66-954956245, E-mail: warin.wat@kmutt.ac.th

บทคัดย่อภาษาไทย

การเปลี่ยนแปลงอย่างรวดเร็วในอุตสาหกรรมอสังหาริมทรัพย์เน้นย้ำความจำเป็นในการมีระบบแนะนำขั้นสูงที่สามารถช่วยผู้ใช้ในการค้นหาข้อมูลอสังหาริมทรัพย์จำนวนมาก บทความนี้นำเสนอระบบแนะนำโพสต์อสังหาริมทรัพย์แบบไฮบริดที่ออกแบบมาเพื่อการใช้งานร่วมกับแพลตฟอร์มเช่น TerraBKK โดยใช้เทคนิคขั้นตอนวิธีเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors หรือ KNN) และการแยกค่าเอกฐาน (Singular Value Decomposition หรือ SVD) ในการมอบคำแนะนำอสังหาริมทรัพย์ที่ปรับให้เหมาะกับแต่ละบุคคล ระบบไฮบริดนี้รวมจุดแข็งของการกรองตามเนื้อหาและการกรองแบบร่วมมือกันเพื่อเพิ่มความแม่นยำและความเกี่ยวข้องของคำแนะนำ พร้อมทั้งแก้ไขปัญหาความหายากของข้อมูลและปัญหาการเริ่มต้นใช้งาน การดำเนินการประกอบด้วยการใช้แบบจำลองสองแบบ โดยผลลัพธ์จากทั้ง KNN และ SVD จะถูกนำมารวมกันด้วยค่าเฉลี่ยถ่วงน้ำหนักเพื่อเพิ่มคุณภาพของคำแนะนำ ประสิทธิภาพของระบบนี้ถูกแสดงให้เห็นโดยใช้ข้อมูลอสังหาริมทรัพย์ที่ได้รับจากบริษัท เทอร์รา มีเดีย แอนด์ คอนซัลติง จำกัด แสดงให้เห็นถึงศักยภาพในการเพิ่มการมีส่วนร่วมและการตัดสินใจของผู้ใช้ในตลาดอสังหาริมทรัพย์

คำสำคัญ : ระบบแนะนำแบบไฮบริด อสังหาริมทรัพย์ ขั้นตอนวิธีเพื่อนบ้านที่ใกล้ที่สุด การแยกค่าเอกฐาน

ABSTRACT

The rapid evolution of the real estate industry has highlighted the need for advanced recommendation systems that can assist users in navigating vast amounts of property data. This paper introduces a Hybrid Property Post Recommendation System, specifically designed for integration with platforms like TerraBKK.



The system employs both K-Nearest Neighbors (KNN) and Singular Value Decomposition (SVD) techniques to deliver personalized property recommendations. By combining the strengths of content-based and collaborative filtering, this hybrid system enhances the relevance and accuracy of recommendations, addressing challenges such as data sparsity and the cold-start problem. The implementation includes a dual-model approach where results from both KNN and SVD are synthesized using a weighted average to optimize recommendation quality. The system's effectiveness is demonstrated using real estate data provided by Terra Media and Consulting Co., Ltd., showcasing its potential to significantly improve user engagement and decision-making in the real estate market.

Keyword: Hybrid Recommendation System; Real Estate; K-Nearest Neighbors; Singular Value Decomposition

1. บทนำ

ในยุคที่อุตสาหกรรมอสังหาริมทรัพย์มีการเปลี่ยนแปลงอย่างรวดเร็ว แพลตฟอร์มดิจิทัลเช่น TerraBKK ได้กลายเป็นส่วนสำคัญในการเชื่อมต่อผู้ซื้อและผู้ขายกับข้อมูลอสังหาริมทรัพย์มากมาย แม้ว่าการเข้าถึงข้อมูลจำนวนมากจะเป็นประโยชน์ แต่ก็อาจทำให้ผู้ใช้งานสับสนและยากต่อการค้นหาทรัพย์สินที่ตรงตามความต้องการของตนเอง เพื่อเพิ่มประสบการณ์ของผู้ใช้งานและช่วยในการตัดสินใจ ระบบแนะนำโพสต์อสังหาริมทรัพย์แบบไฮบริดได้ถูกพัฒนาขึ้นมา โดยระบบนี้ออกแบบให้ผสานเข้ากับแพลตฟอร์มเช่น TerraBKK ได้อย่างลงตัว ใช้เทคนิคการเรียนรู้ของเครื่องขั้นสูง เช่น K-Nearest Neighbors (KNN) และ Singular Value Decomposition (SVD) เพื่อให้คำแนะนำอสังหาริมทรัพย์ที่มีความเฉพาะเจาะจงสำหรับแต่ละบุคคล

TerraBKK มีชื่อเสียงในด้านการให้บริการแพลตฟอร์มดิจิทัลที่ครอบคลุมในภาคอสังหาริมทรัพย์ โดยนำเสนอรายการอสังหาริมทรัพย์ที่หลากหลายและบริการที่เกี่ยวข้องต่างๆ การรวมระบบแนะนำแบบไฮบริดเข้ากับ TerraBKK ช่วยให้ผู้ใช้ได้รับประสบการณ์การค้นหาทรัพย์สินที่แม่นยำและปรับแต่งได้ ระบบนี้ผสมผสานจุดแข็งของการกรองตามเนื้อหาและการกรองแบบร่วมมือกัน สามารถ

จัดการกับข้อมูลที่ซับซ้อนและหลากหลายที่เป็นลักษณะของอสังหาริมทรัพย์ได้อย่างมีประสิทธิภาพ ทำให้มั่นใจได้ว่าคำแนะนำที่ได้รับไม่เพียงแต่ตรงกับความต้องการ แต่ยังสอดคล้องกับความต้องการเฉพาะของผู้ใช้แต่ละราย

การผสานระบบแนะนำนี้เข้ากับแพลตฟอร์มของ TerraBKK ยังช่วยแก้ปัญหาทั่วไป เช่น ปัญหาการเริ่มต้นเย็นและปัญหาความเร็ว โดยเสนอวิธีการที่สามารถขยายขนาดได้และปรับตัวได้ซึ่งเติบโตไปพร้อมกับฐานผู้ใช้ที่เพิ่มขึ้น นอกจากนี้ ด้วยการเตรียมอินเทอร์เฟซผู้ใช้ที่ใช้งานง่ายและเป็นมิตร ระบบยังช่วยเพิ่มการมีส่วนร่วมและความพึงพอใจของลูกค้า ซึ่งนำไปสู่การรักษาลูกค้าและความภักดีที่เพิ่มขึ้น การบูรณาการเทคโนโลยีขั้นสูงนี้ช่วยให้ TerraBKK เป็นผู้นำในตลาดอสังหาริมทรัพย์ดิจิทัล โดยนำเสนอเครื่องมือที่ทันสมัยเพื่อปรับปรุงประสบการณ์ผู้ใช้และทำให้กระบวนการค้นหาทรัพย์สินมีประสิทธิภาพยิ่งขึ้น

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ระบบแนะนำ (Recommender System)

ระบบแนะนำ คือเครื่องมือชุดคำสั่งที่ใช้เพื่อแนะนำสิ่งของแก่ผู้ใช้ตามความต้องการ โดยคำแนะนำนี้จะเชื่อมโยงกับกระบวนการตัดสินใจต่าง ๆ โดยปกติแล้วระบบแนะนำจะมุ่งเน้นไปที่ประเภทของรายการสิ่งของที่เฉพาะเจาะจง การออกแบบส่วนประสานกับผู้ใช้ และกลวิธี

หลักที่ใช้แนะนำนั้นจะได้รับการปรับแต่งทั้งหมดเพื่อสร้างคำแนะนำที่มีประโยชน์และมีประสิทธิภาพสำหรับรายการสิ่งของนั้น ๆ โดยเฉพาะ ซึ่งระบบแนะนำนี้มีความท้าทายในการพัฒนาดังนี้ [2]

1) การกระจายของข้อมูล (Data Sparsity) ระบบแนะนำเชิงพาณิชย์จำนวนมากใช้ชุดข้อมูลขนาดใหญ่ ดังนั้นเมทริกซ์ที่ได้จากข้อมูลผู้ใช้และสิ่งของ (User – Item Matrix) ที่ใช้ในการคัดกรองจะมีขนาดใหญ่และกระจายมาก ประสิทธิภาพของกระบวนการแนะนำอาจจะลดลง และสาเหตุทำให้เกิด Cold Start Problem ซึ่งก็คือปัญหาที่ไม่สามารถแนะนำสิ่งของกับผู้ใช้ใหม่ได้เนื่องจากไม่สามารถตรวจสอบการตั้งค่าของผู้ใช้ที่มีพฤติกรรมที่ใกล้เคียงกัน ในระบบแนะนำแบบ Collaborative Filtering การแนะนำรายการสิ่งของขึ้นอยู่กับค่าในอดีตของผู้ใช้ ดังนั้นผู้ใช้ใหม่จะต้องให้คะแนนจำนวนรายการสิ่งของที่เพียงพอเพื่อให้ระบบสามารถตรวจสอบการตั้งค่าได้อย่างถูกต้องและสามารถรับคำแนะนำที่ดีได้

2) ภาวะปรับขนาดได้ (Scalability) ขั้นตอนวิธีของ Collaborative Filtering แบบเดิมจะมีปัญหาในด้านนี้เนื่องจากจำนวนผู้ใช้และรายการสิ่งของเพิ่มขึ้น ยกตัวอย่างเช่น มีลูกค้า 10,000,000 คน คิดเป็น $O(M)$ และสิ่งของ 1,000,000 ชิ้น คิดเป็น $O(N)$ กับความซับซ้อนของขั้นตอนวิธี (Complexity of Algorithm) เป็น n ซึ่งมีขนาดใหญ่เกินไป เนื่องจากระบบแนะนำจะต้องตอบสนองความต้องการของผู้ใช้โดยทันทีแม้หลังการซื้อ จึงต้องการความสามารถในการปรับขนาดได้ที่มากขึ้น

3) ความหลากหลาย (Diversity) ระบบแนะนำควรจะสร้างความหลากหลาย เพราะระบบแนะนำช่วยให้พบผลิตภัณฑ์หรือสิ่งของใหม่ ๆ แต่ขั้นตอนวิธีบางอย่างอาจลดความหลากหลายเสียเองโดยไม่ได้ตั้งใจ ระบบแนะนำแนะนำรายการสิ่งของยอดนิยมและคะแนนสูงซึ่งเป็นที่ยอมรับของผู้ใช้โดยเฉพาะ สิ่งเหล่านี้จะทำให้ความถูกต้องแม่นยำลดลง

จึงต้องพัฒนาแนวทางใหม่ ๆ อย่างระบบแนะนำแบบไฮบริดเพื่อช่วยเพิ่มประสิทธิภาพของกระบวนการแนะนำ

4) ช่องโหว่ในการโจมตี (Vulnerability to Attack) ความปลอดภัยเป็นสิ่งสำคัญมากในระบบใดก็ตามที่ใช้งานบนเว็บไซต์ ซึ่งระบบแนะนำก็มีโอกาสได้รับอันตรายจากการโจมตี ที่อาจเป็นการพยายามยับยั้งหรือส่งเสริมสิ่งของบางอย่าง เป็นความท้าทายที่สำคัญที่ผู้พัฒนาระบบต้องเผชิญ

โดยทั่วไปแล้วระบบแนะนำจะแนะนำรายการสิ่งของแก่ผู้ใช้ที่อาจสนใจ ที่ผ่านมามีการแยกระบบแนะนำจำแนกตามประเภทของการกรองออกเป็น 4 ประเภท ได้แก่

2.1.1 Collaborative Filtering เป็นแนวทางที่ใช้กันอย่างแพร่หลายมากที่สุดในพัฒนาระบบแนะนำ แนวทางนี้จะแนะนำสำหรับผู้ใช้แต่ละคนโดยเปรียบเทียบกับความต้องการของผู้ใช้รายอื่นที่ให้คะแนนผลิตภัณฑ์หรือสิ่งของในลักษณะเดียวกับผู้ใช้ที่ใช้งานอยู่ แนวทางนี้มีความสำคัญมากแม้ว่าจะถูกนำไปใช้ร่วมกับแนวทางอื่น ๆ เช่น Content-Based แนวทางนี้เกิดจากการรวบรวมข้อมูลจำนวนมากซึ่งขึ้นอยู่กับพฤติกรรมของผู้ใช้ ความชอบหรือรสนิยมของผู้ใช้โดยใช้ความคล้ายคลึงกับผู้ใช้รายอื่น ไม่ได้ขึ้นอยู่กับความสามารถในการสกัดข้อความของเครื่องจึงสามารถแนะนำรายการสิ่งของผสมอย่างถูกต้องซึ่งเป็นข้อดีและประโยชน์หลักๆ ของแนวทาง Collaborative Filtering รายการสิ่งของที่ได้จากระบบแนะนำประเภทนี้จะมาจากการประเมินของผู้ใช้กลุ่มใหญ่

2.1.2 Content-Based Filtering เป็นแนวทางที่การแนะนำรายการสิ่งของจะขึ้นอยู่กับตัวเลือกรายการสิ่งของที่ผ่านมาของผู้ใช้ คำอธิบายของผลิตภัณฑ์หรือสิ่งของและข้อมูลแสดงลักษณะเฉพาะของผู้ใช้มีบทบาทสำคัญในแนวทางนี้ ขั้นตอนวิธีของแนวทางนี้จะพยายามแนะนำสิ่งของที่อิงตามค่าจำนวนที่ใกล้เคียง รายการสิ่งของที่ตรงที่สุดจะถูกแนะนำโดยการเปรียบเทียบรายการสิ่งของร่วมกับ



2024 CreTech

**Creative
Technology &
Artificial Intelligence**
at Cape Racha Hotel, Sriracha, Chon Buri

รายการที่ผู้ใช้ให้คะแนนก่อนหน้านี้ ขั้นตอนวิธีที่ใช้กันอย่างแพร่หลายมากที่สุดในแนวทางนี้คือ Term Frequency-Inverse Document Frequency (TF-IDF) โดยใช้สำหรับการสร้างโปรไฟล์ผู้ใช้ที่เน้นไปที่ข้อมูล 2 ประเภท คือ รูปแบบการตั้งค่าของผู้ใช้และบันทึกประวัติการเข้าใช้ของผู้ใช้ โดยทั่วไปวิธีการเหล่านี้จะใช้โปรไฟล์สิ่งของเพื่อตรวจสอบคุณสมบัติของสิ่งของในระบบ การสร้างโปรไฟล์ตามเนื้อหาของผู้ใช้ทำได้โดยใช้เวกเตอร์แบบถ่วงน้ำหนัก (Weighted Vector) ของคุณสมบัติรายการสิ่งของ โดยค่าน้ำหนักแสดงความสำคัญของแต่ละคุณสมบัติที่มีต่อผู้ใช้ เมื่อทราบคุณสมบัติของรายการสิ่งของและโปรไฟล์ของผู้ใช้หน้าที่สำคัญของระบบแนะนำแบบ Content-Based คือ การกำหนดว่าผู้ใช้จะชอบรายการใดรายการหนึ่งหรือไม่ ซึ่งจะได้คำตอบโดยใช้วิธีการศึกษาสำนึก (Heuristic Methods) หรือขั้นตอนวิธีการจำแนก เช่น การจำแนกเชิงเส้น วิธีการความน่าจะเป็น เป็นต้น ระบบแนะนำแนวทางนี้มีข้อดีคือ ให้อิสระแก่ผู้ใช้ผ่านการให้คะแนนโดยผู้ใช้อเองเพื่อสร้างโปรไฟล์การตั้งค่าของตนเอง ผู้ใช้สามารถรู้ได้ว่าระบบทำงานอย่างไร และสามารถแนะนำรายการสิ่งของใหม่ที่ยังไม่เคยได้คะแนนจากผู้ใช้คนใด ซึ่งจะเป็นประโยชน์สำหรับผู้ใช้ใหม่ แต่ยังมีข้อจำกัดหลายประการ ได้แก่ สร้างคุณสมบัติของรายการสิ่งของได้ยากในบางบริเวณ แนะนำแต่เพียงรายการประเภทเดียว เพราะปัญหา Overspecialization Problem การรับข้อเสนอแนะจากผู้ใช้ทำได้ยาก เนื่องจากโดยทั่วไปผู้ใช้ไม่ได้จัดรายการ จึงไม่สามารถระบุได้ว่าคำแนะนำนั้นถูกต้องหรือไม่

2.1.3 Demographic Filtering กำหนดขึ้นจากข้อมูลแสดงลักษณะเฉพาะของประชากรผู้ใช้ คำแนะนำที่ได้จะขึ้นอยู่กับข้อมูลที่ผู้ใช้ให้มาที่ถูกพิจารณาว่าใกล้เคียงกันตามพารามิเตอร์ทางประชากร เช่น สัญชาติ อายุ เพศ เป็นต้น

2.1.4 Hybrid Filtering หรือ แนวทางแบบไฮบริดเป็นการรวมกันของแนวทางในการพัฒนาระบบ

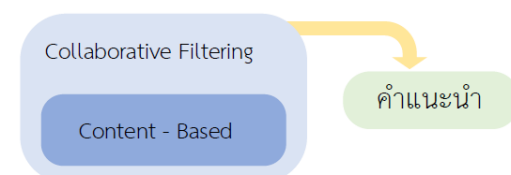
แนะนำมากกว่าหนึ่งแนวทาง แนวทางแบบไฮบริดถูกนำมาใช้เพื่อแก้ปัญหาที่พบโดยทั่วไปของแนวทางการทำระบบแนะนำข้างต้น เช่น Cold Start Problem, Overspecialization Problem และ Sparsity Problem เป็นต้น แนวทางแบบไฮบริดมีประสิทธิภาพมากกว่า Collaborative Filtering หรือ Content-Based ในหลายกรณีเพื่อปรับปรุงข้อเสียของแต่ละแนวทาง แนวทางไฮบริดสามารถทำได้หลายวิธี เช่น

1) พัฒนาระบบแนะนำในแนวทาง Collaborative Filtering และ Content-Based แยกกันแล้วนำผลการแนะนำมารวมกันดังรูปที่ 1



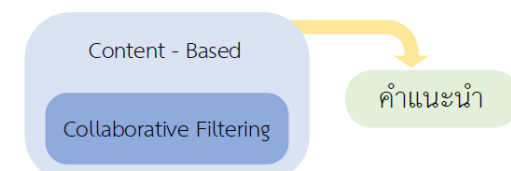
รูปที่ 1 ลักษณะของการพัฒนาระบบ Collaborative Filtering และ Content-Based แยกกัน แล้วนำผลการแนะนำมารวมกัน

2) นำเอาคุณสมบัติบางอย่างที่ดีของแนวทาง Content-Based เข้ามาปรับปรุงประสิทธิภาพด้วยแนวทาง Collaborative Filtering ในรูปที่ 2



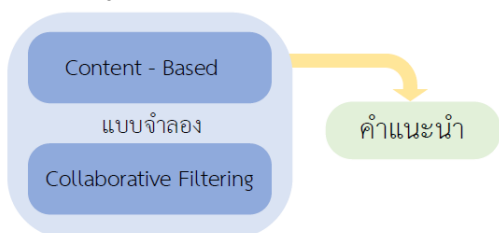
รูปที่ 2 ลักษณะของแนะนำจากการกรองสองชั้นโดยใช้ Collaborative Filtering กับ Content-Based

3) นำเอาคุณสมบัติบางอย่างที่ดีของแนวทาง Collaborative Filtering เข้ามารวมกับแนวทาง Content-Based ดังรูปที่ 3



รูปที่ 3 ลักษณะของการกรองสองชั้นโดยใช้คำตอบของ Content-Based กรองอีกชั้นด้วย Collaborative Filtering เพื่อพัฒนาประสิทธิภาพของระบบ

4) สร้างแบบจำลองที่นำทั้งแนวทาง Collaborative Filtering และ Content-Based มาใช้ร่วมกัน ในรูปที่ 4



รูปที่ 4 ลักษณะของการพัฒนาโดยสร้างแบบจำลองแล้วนำคุณสมบัติบางอย่างที่ดีของ Collaborative Filtering และ Content-Based

ในช่วงหลายปีที่ผ่านมา ระบบการแนะนำอสังหาริมทรัพย์ได้เห็นความก้าวหน้าที่สำคัญ โดยเฉพาะในการแก้ปัญหาเช่น ความเบาบางของข้อมูล (data sparsity) และปัญหาการเริ่มต้นใหม่ (cold-start problem) มีการใช้วิธีการที่หลากหลาย รวมถึง K-Nearest Neighbors (KNN) และ Singular Value Decomposition (SVD) เพื่อเพิ่มความแม่นยำและความเกี่ยวข้องของคำแนะนำ

2.1.5 K-Nearest Neighbors (KNN) พบทั้งในรูปแบบที่ใช้อิงจากผู้ใช้และอิงจากรายการสินค้ามีการใช้อย่างแพร่หลาย Wang และคณะ [7] ได้ปรับปรุงมาตรการความคล้ายคลึงของ Pearson ใน KNN ที่อิงจากผู้ใช้ (UKNN) เพื่อสะท้อนถึงความชอบของผู้ใช้ที่มีรสนิยมคล้ายกันได้ดียิ่งขึ้น โดยเฉพาะในบริบทของคำแนะนำด้านที่อยู่อาศัย การปรับปรุงนี้มีวัตถุประสงค์เพื่อเพิ่มความแม่นยำในการคำนวณความคล้ายคลึงของผู้ใช้ Yu และคณะ [8] ในทางกลับกัน ได้ใช้ KNN ที่อิงจากรายการสินค้า (IKNN) เป็นฐานในการประเมินประสิทธิภาพของอัลกอริทึมคำแนะนำที่เพิ่มการพิจารณาปัจจัยทางภูมิศาสตร์ วิธีการนี้มีความสำคัญในการเน้นความสำคัญของปัจจัยทางภูมิศาสตร์

ในการเลือกอสังหาริมทรัพย์ เช่นเดียวกัน Ojokoh และคณะ [9] ได้พัฒนาโมเดลแบบไฮบริดที่รวม IKNN กับระบบที่ใช้กฎฟัซซี เพื่อจัดอันดับทรัพย์สิน โดยมุ่งเน้นการแก้ไขกระบวนการตัดสินใจที่ซับซ้อนในด้านอสังหาริมทรัพย์ นอกจากนี้ Chonwiharnphan คณะ [10] ได้ใช้ KNN ในการสร้างบันทึกผู้ใช้ที่สมจริงสำหรับรายการอสังหาริมทรัพย์ใหม่ผ่านโมเดลที่ใช้ GAN เพื่อต่อสู้กับปัญหาการเริ่มต้นใหม่โดยการจำลองปฏิสัมพันธ์ของผู้ใช้ที่เป็นไปได้

2.1.6 Singular Value Decomposition (SVD) ได้มีบทบาทสำคัญในการพัฒนาระบบคำแนะนำ Yu และคณะไม่เพียงแต่สำรวจการปรับปรุงด้านความใกล้ชิดทางภูมิศาสตร์ แต่ยังเปรียบเทียบวิธีการของพวกเขากับอัลกอริทึมพื้นฐานอื่น ๆ เช่น Weighted Regularized Matrix Factorization (WRMF) และ Probabilistic Matrix Factorization (PMF) การเปรียบเทียบนี้ได้เน้นถึงความสามารถของ SVD ในการลดมิติและค้นหาปัจจัยแฝงในการโต้ตอบระหว่างผู้ใช้และรายการสินค้า Paterek [11] ได้สร้างผลงานพื้นฐานให้กับ SVD โดยการปรับปรุงรูปแบบที่ได้รับการควบคุม ซึ่งมีความสำคัญต่อระบบการกรองร่วม (collaborative filtering) การปรับปรุงเหล่านี้มีความเกี่ยวข้องโดยเฉพาะสำหรับการจัดการชุดข้อมูลขนาดใหญ่ที่มีข้อมูลเบาบางซึ่งเป็นลักษณะทั่วไปของแพลตฟอร์มอสังหาริมทรัพย์ออนไลน์

3. วิธีการดำเนินการวิจัย

การพัฒนากระบวนแนะนำประกาศขายอสังหาริมทรัพย์ในครั้งนี้ใช้ข้อมูลจากเว็บไซต์ ของบริษัท เทอร์รา มีเดีย แอนด์ คอนซัลติง จำกัด คือ www.terrakk.com เท่านั้น ซึ่งเป็นข้อมูลที่ไม่ได้เผยแพร่สู่สาธารณะ โดยใช้ชุดข้อมูลทั้งหมด 2 ชุด ประกอบด้วย

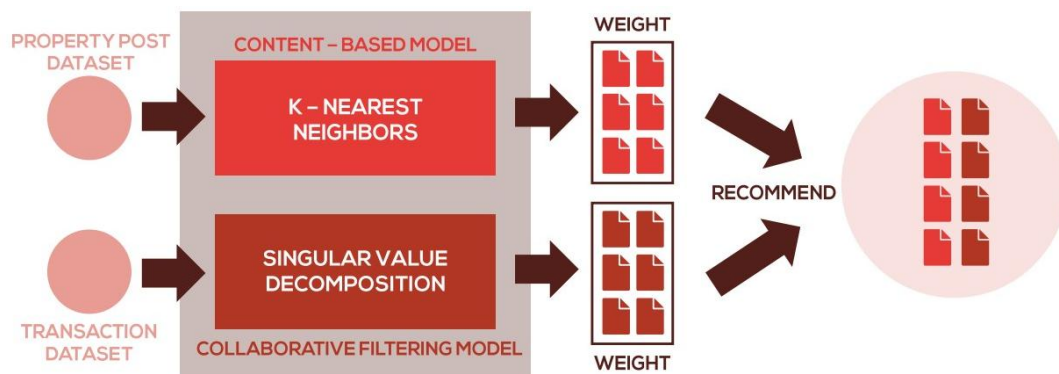
1. ข้อมูลประกาศขายอสังหาริมทรัพย์ที่อยู่ในช่วง 1 – 2 ปีที่ผ่านมา จำนวน 100,000 ประกาศ ขนาด 54 MB

2. ข้อมูลการเข้าชมหน้าประกาศขายอสังหาริมทรัพย์
ของผู้ใช้ 500,000 คน ในช่วงวันที่ 30 กรกฎาคม พ.ศ. 2563 –
31 สิงหาคม พ.ศ. 2563 จำนวน 600,000 ครั้ง ขนาด 51.9 MB

3.1 การพัฒนาแบบจำลองแบบไฮบริด

งานวิจัยนี้จะเลือกใช้โครงสร้างสองแบบจำลองซึ่งจะ
ทำงานแยกออกจากกัน แบบจำลองแบบที่หนึ่งจะเป็นแบบ
Content – Based ซึ่งจะใช้ขั้นตอนวิธีเพื่อนบ้านที่ใกล้ที่สุดใน
การแนะนำ และแบบจำลองที่สองจะเป็นแบบ Collaborative
Filtering ซึ่งจะใช้การแยกค่าเอกฐานในการแนะนำ จากนั้นจะนำ
ผลลัพธ์ที่ได้จากทั้งสองแบบจำลองมาคำนวณกับค่าน้ำหนักแบบ
เฉลี่ยแล้วจึงแนะนำชุดประกาศขายอสังหาริมทรัพย์ดังรูปที่ 5

แม้ว่าในการพัฒนาระบบแนะนำจะมีเทคโนโลยีที่มีความ
หลากหลายในมิติอื่นๆ มากกว่า เช่น การประมวลผล
ภาษาธรรมชาติ แต่เทคโนโลยีเหล่านั้นใช้ทรัพยากรมากกว่ามาก
ทำให้หน่วยประมวลผลและหน่วยความจำที่เป็นสภาพแวดล้อม
ของคณะผู้จัดทำไม่สามารถฝึกสอนและนำไปใช้ได้ การพัฒนา
ระบบแนะนำทั้งหมดจะพัฒนาโดยใช้ภาษา Python ใช้โปรแกรม
Spyder และ Jupyter Notebook บนโน้ตบุ๊กคอมพิวเตอร์ ซีพียู
Intel CORE i7-9750H 9th Gen หน่วยความจำ (RAM) ขนาด
16 GB



รูปที่ 5 โครงสร้างและการทำงานของระบบแนะนำประกาศขายอสังหาริมทรัพย์แบบไฮบริด

3.1.1 ขั้นตอนวิธีการเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbors หรือ KNN)

เป็นขั้นตอนวิธีรูปแบบหนึ่งของการเรียนรู้ของ
เครื่องแบบมีผู้สอน (Supervised Learning) ที่ใช้สำหรับการ
จำแนกประเภท (Classification) และการถดถอย (Regression)
[3] ใช้หลักการข้อมูลที่มีลักษณะคล้ายกันจะรวมอยู่ใกล้กัน
เปรียบเทียบความใกล้เคียงระหว่างข้อมูลที่สนใจกับข้อมูลอื่น ๆ
ที่อยู่ใกล้เคียง โดยมีข้อมูลนำเข้าประกอบด้วยข้อมูลพื้นที่

คุณลักษณะ (Feature Space) รูปแบบที่มีลักษณะใกล้เคียงกัน
ข้อมูลออกนั้นขึ้นอยู่กับคำแนะนำขั้นตอนวิธีนี้ไปใช้

ขั้นตอนวิธีเพื่อนบ้านที่ใกล้ที่สุดทำงานโดยมีข้อมูล
นำเข้าเป็นเวกเตอร์ที่อยู่ในพื้นที่คุณลักษณะหลายมิติ ซึ่งในแต่
ละเวกเตอร์จะมีคลาสกำกับ ขั้นตอนการฝึกสอนของขั้นตอนวิธีนี้
มีเพียงการจัดเก็บเวกเตอร์คุณลักษณะและคลาสที่กำกับข้อมูล
นั้น ๆ ในขั้นตอนการจำแนกประเภท K เป็นค่าที่ผู้ใช้ต้องกำหนด
เอง ข้อมูลทดสอบหรือเวกเตอร์ที่ไม่ได้ระบุคลาสหรือประเภทจะ



ถูกกำหนดประเภทด้วยการวัดระยะห่างกับข้อมูลรอบข้างแล้ว กำหนดให้เข้ากับข้อมูลรอบข้างมากที่สุด หรือก็คือจัดให้เป็นประเภทเดียวกับข้อมูลรอบข้างที่อยู่ใกล้ที่สุด ซึ่งการหาระยะห่างนั้น โดยปกติการวัดระยะห่างสำหรับตัวแปรต่อเนื่องมักจะใช้ระยะทางแบบยูคลิด (Euclidean Distance) ส่วนตัวแปรไม่ต่อเนื่องสามารถใช้การวัดแบบอื่น ๆ ได้ เช่น ระยะทางแบบแฮมมิง (Hamming Distance) ข้อเสียของการจัดประเภทโดยอิงข้อมูลข้างมากจะเกิดขึ้นเมื่อการกระจายของข้อมูลมีความเอนเอียง ข้อมูลประเภทที่มีจำนวนมากกว่าจะมีอิทธิพลต่อการทำนายผลลัพธ์ในการจัดประเภทของข้อมูลใหม่ เนื่องจากข้อมูลใหม่มีโอกาสพบข้อมูลประเภทนั้นมากกว่าประเภทอื่น ๆ เพราะมีจำนวนมาก [4] วิธีการแก้ปัญหานี้คือการกำหนดค่าน้ำหนักในการจำแนกโดยคิดจากระยะห่างจากข้อมูลที่ทดสอบจนถึงข้อมูลรอบข้างที่ใกล้ที่สุด คลาสหรือค่าการถดถอยของข้อมูลแต่ละตำแหน่งที่ใกล้ที่สุด K ตำแหน่งจะนำมาคูณด้วยค่าน้ำหนักตามสัดส่วนกับค่าผกผันของระยะห่างระหว่างจุดนั้นกับข้อมูลตำแหน่งที่ทดสอบ สำหรับการเลือกค่า K ที่ดีที่สุดนั้นขึ้นอยู่กับข้อมูล โดยทั่วไปค่า K ที่สูงขึ้นจะลดผลกระทบจากการรบกวนในการจำแนกประเภท แต่จะทำให้ขอบเขตระหว่างแต่ละคลาสน้อยลง ค่า K ที่ดีสามารถเลือกได้โดยใช้เทคนิคฮิวริสติกต่าง ๆ ค่าความแม่นยำ (Accuracy) ของขั้นตอนวิธีเพื่อนบ้านที่ใกล้ที่สุดจะลดลงอย่างมากหากคุณสมบัติไม่สอดคล้องกับความสัมพันธ์ การรบกวน หรือสเกลของคุณลักษณะไม่สอดคล้องกับค่าสำคัญ

3.1.2 การแยกค่าเอกฐาน (Singular Value Decomposition หรือ SVD)

คือเทคนิคการแยกตัวประกอบเมทริกซ์ (Matrix Factorization) รูปแบบหนึ่งที่มาจากพีชคณิตเชิงเส้น (Linear Algebra) ที่เริ่มได้รับความนิยมในแวดวงวิทยาศาสตร์ข้อมูลและการเรียนรู้ของเครื่องจากการนำมาประยุกต์ใช้พัฒนาระบบแนะนำ [5] โดยทั่วไปมักจะนำมาใช้เป็นเทคนิคในการลดมิติของ

ข้อมูลในการเรียนรู้ของเครื่อง ซึ่งจะลดจำนวนคุณลักษณะของชุดข้อมูลโดยลดขนาดมิติช่องว่างจาก N มิติให้เหลือ K มิติ โดยที่ K มีค่าน้อยกว่า N ในการพัฒนาระบบแนะนำ การแยกค่าเอกฐานจะถูกนำมาใช้ในแนวทาง Collaborative Filtering ใช้โครงสร้างเมทริกซ์โดยที่แถวแสดงถึงผู้ใช้และคอลัมน์แสดงถึงสิ่งของ ค่าข้อมูลภายในเมทริกซ์จะเป็นค่าคะแนนของสิ่งของโดยผู้ใช้เป็นผู้ป้อนเข้าไปในระบบ การแยกค่าเอกฐานจะแยกตัวประกอบของเมทริกซ์ออกเป็น 3 เมทริกซ์ดังนี้

$$A = USV^T$$

โดยที่ A คือ Utility Matrix ขนาด $m \times n$ และ U คือ Orthogonal Left Singular Matrix เมทริกซ์ขนาด $m \times r$ ที่แสดงถึงความสัมพันธ์ระหว่างผู้ใช้และปัจจัยแฝง ส่วน S คือ เมทริกซ์ทแยงมุมขนาด $r \times r$ ที่แสดงถึงลักษณะของปัจจัยแฝงและ V^T คือ Diagonal Right Singular Matrix ขนาด $r \times n$ ที่แสดงถึงความคล้ายคลึงกันระหว่างสิ่งของและปัจจัยแฝง

ปัจจัยแฝงในที่นี้คือลักษณะของสิ่งของ [6]

กำหนดให้สิ่งของแต่ละอย่างแสดงด้วย x_i และผู้ใช้แต่ละคนเป็น y_u การให้คะแนนสิ่งของที่คาดหวังโดยผู้ใช้งานสามารถแทนด้วย

$$Y_{ui} = x_i^T y_u$$

เป็นรูปแบบของการแยกตัวประกอบในการแยกค่าเอกฐาน และสามารถหาได้ในลักษณะของผลต่างค่าความผิดพลาดกำลังสองระหว่างผลคูณเชิงสเกลาร์ของเวกเตอร์กับค่าคะแนนที่อยู่ในเมทริกซ์สิ่งของและผู้ใช้ สามารถแสดงได้ดังนี้

$$\text{Min}(x, y) \sum_{(u,i) \in K} (r_{ui} - x_i^T y_u)^2$$

2024
CreTechCreative
Technology &
Artificial Intelligence
at Cape Racha Hotel, Sriracha, Chon Buri

เพื่อให้แบบจำลองทำงานได้ดีและไม่เกิดภาวะ Overfitting ในข้อมูลฝึกสอน จึงต้องมีการกำหนดเงื่อนไข เพื่อเป็นค่าปรับในสมการข้างต้น

$$\text{Min}(x, y) \sum_{(u,i) \in K} (r_{ui} - x_t^T y_u)^2 + \lambda(\|x_i\|^2 + \|y_u\|^2)$$

เพื่อลดข้อผิดพลาดระหว่างค่าที่แบบจำลองทำนายได้กับค่าจริง ขั้นตอนวิธีจึงใช้ค่าความลำเอียง (Bias) โดยกำหนดให้ค่าของผู้ใช้กับสิ่งของ เป็น (u, i) ให้ค่าเฉลี่ยคะแนนของสิ่งของทั้งหมดเป็น μ ให้ค่าเฉลี่ยคะแนนของสิ่งของ i ลบด้วย μ เป็น b_i และให้ค่าเฉลี่ยคะแนนของสิ่งของที่ให้โดยผู้ใช้ u ลบด้วย μ เป็น b_u สมการสุดท้ายที่เพิ่มค่าความเอนเอียงจะได้

$$\text{Min}(x, y, b_i, b_u) \sum_{(u,i) \in K} (r_{ui} - x_t^T y_u - \mu - b_i - b_u)^2 + \lambda(\|x_i\|^2 + \|y_u\|^2 + b_i^2 + b_u^2)$$

สมการและค่าข้างต้นเป็นองค์ประกอบหลักของขั้นตอนวิธีที่ใช้กับการแยกค่าเอกลักษณ์ที่ใช้สำหรับระบบแนะนำ

3.2 การพัฒนา Application Programming Interface (API)

เนื่องด้วยระบบแนะนำนี้จะถูกนำไปใช้งานบนเว็บไซต์และทางบริษัทต้องการระบบแนะนำที่สามารถนำไปใช้ได้หลากหลายเครื่องมือ ทั้งคอมพิวเตอร์หรือโทรศัพท์เคลื่อนที่ คณะผู้จัดทำจึงเลือกเขียน Web Service ด้วยภาษา Python เช่นเดียวกับขั้นตอนการพัฒนาแบบจำลองและใช้ Django Framework ซึ่งเป็นชุดคำสั่งที่ใช้ในการเขียนเว็บไซต์ในฝั่ง

Back-End เพื่อให้เกิดความสะดวกในการเรียกใช้งานแบบจำลองที่ออกแบบและสร้างขึ้น

3.3 การเตรียมข้อมูลสำหรับการฝึกสอน

เนื่องจากระบบแนะนำใช้แบบจำลองสองแบบจำลอง คือแบบจำลองแบบ Content – Based และแบบจำลองแบบ Collaborative Filtering ซึ่งแบบจำลองแต่ละแบบใช้ข้อมูลที่มีลักษณะและบริบทต่างกัน จึงจำเป็นต้องเตรียมชุดข้อมูล 2 ชุดสำหรับแต่ละแบบจำลอง ดังนี้

3.3.1 การเตรียมข้อมูลสำหรับแบบจำลองแบบ Content - Based

ชุดข้อมูลที่ใช้สำหรับการฝึกสอนแบบจำลองแบบ Content – Based จะเป็นชุดข้อมูลของประกาศขายอสังหาริมทรัพย์ คณะผู้จัดทำจึงประสานงานกับผู้ที่เกี่ยวข้องและดูแลเรื่องข้อมูลประกาศขายอสังหาริมทรัพย์ที่อยู่บนเว็บไซต์ของบริษัทเพื่อช่วยเหลือในการดึงข้อมูลจากฐานข้อมูลเพื่อนำมาใช้ฝึกสอน เมื่อได้ชุดข้อมูลประกาศขายอสังหาริมทรัพย์แล้ว คณะผู้จัดทำจะทำการสำรวจและวิเคราะห์ข้อมูลประกาศขายอสังหาริมทรัพย์ที่ได้ว่าเป็นข้อมูลในลักษณะใดประกอบกับการแก้ไขข้อมูลที่ไม่ถูกต้อง

พบว่าภายในชุดข้อมูลมีข้อมูลแบบจัดประเภทเป็นจำนวนมาก แต่ข้อมูลที่มีนัยยะสำคัญหรือมีผลต่อการตัดสินใจเลือกซื้อหรือเช่าอสังหาริมทรัพย์มีจำนวนน้อยเกินไปเมื่อเทียบกับข้อมูลทั้งหมด จึงได้ทำการสร้างข้อมูลคุณสมบัติของประกาศขายอสังหาริมทรัพย์เพิ่มเติม 4 คุณสมบัติ ได้แก่

1. ระยะทางจากอสังหาริมทรัพย์ถึงร้านสะดวกซื้อที่ใกล้ที่สุด
2. ระยะทางจากอสังหาริมทรัพย์ถึงศูนย์การค้าหรือห้างสรรพสินค้าที่ใกล้ที่สุด



3. ระยะทางจากอสังหาริมทรัพย์ถึงสถานศึกษาที่ใกล้ที่สุด และ

4. ระยะทางจากอสังหาริมทรัพย์ถึงสถานีรถไฟฟ้าหรือรถไฟฟ้ามหานครที่ใกล้ที่สุด

โดยใช้ ข้อมูลพิกัด (ละติจูดและลองจิจูด) ของอสังหาริมทรัพย์และข้อมูลพิกัดของสถานที่ที่คณะผู้จัดทำค้นหาเพิ่มเติมได้จาก map.longdo.com มาคำนวณหาระยะทางโดยใช้ Haversine Formula แล้วเพิ่มไปในชุดข้อมูล จากนั้นแก้ไขความผิดพลาดที่มีในชุดข้อมูล เช่น ค่าว่าง (null values) จัดการค่าที่ผิดปกติ (outlier) โดยใช้ค่าคะแนนมาตรฐาน (Z-score) หรือข้อมูลแถวที่ซ้ำกัน ตัดข้อมูลตัวอักษรและข้อมูลระบุเฉพาะ เช่น ID ของอสังหาริมทรัพย์ ตำบล อำเภอและจังหวัด ที่ไม่สามารถนำมาคำนวณได้ ข้อมูลแบบจำแนกประเภทที่แทนประเภทด้วยตัวเลข หากนำข้อมูลประเภทนี้ไปฝึกสอนอาจทำให้การแนะนำไม่มีประสิทธิภาพได้เนื่องจากตัวเลขที่มามีน้อยมีนัยสำคัญต่อการคำนวณและผลลัพธ์ คณะผู้จัดทำจึงใช้ One - Hot Encoding เปลี่ยนประเภทให้อยู่ในรูปตัวเลข 1 (เป็นประเภทนั้น) และ 0 (ไม่เป็นประเภทนั้น) จะได้ชุดข้อมูลที่สามารถนำไปฝึกสอนแบบจำลองแบบ Content - Based ได้

3.3.2 การเตรียมข้อมูลสำหรับแบบจำลองแบบ Collaborative Filtering

ชุดข้อมูลที่ใช้สำหรับการฝึกสอนแบบจำลองแบบ Collaborative Filtering จะเป็นชุดข้อมูลพฤติกรรมกรรมการเข้าชมหน้าประกาศขายอสังหาริมทรัพย์ของผู้ใช้หรือธุรกรรมกรรมการใช้งานเว็บไซต์แล้วทำการสำรวจและวิเคราะห์ พบว่าพฤติกรรมในการเข้าชมหน้าประกาศขายอสังหาริมทรัพย์แต่ละหน้าสามารถแบ่งการกระทำออกเป็น 3 ประเภท ได้แก่ เข้าชม (view) เข้าดูรายละเอียด (look information) และเข้าดูหมายเลขโทรศัพท์ของผู้ลงประกาศ (look tel) รวมอยู่ในบันทึกธุรกรรมหนึ่งครั้ง

จึงทำการเพิ่มข้อมูลธุรกรรมโดยเพิ่มจากประเภทการกระทำ คือจากการที่รวบรวมข้อมูลหนึ่งธุรกรรมซึ่งเป็นการบันทึกผู้ใช้แต่ละคนเข้าชมหน้าประกาศขายอสังหาริมทรัพย์แต่ละหน้าและบันทึกการเข้าดูรายละเอียดและหมายเลขโทรศัพท์ ออกเป็นหนึ่งธุรกรรมจะประกอบด้วยข้อมูลระบุผู้ใช้ หน้าประกาศขายอสังหาริมทรัพย์ ประเภทการกระทำ และวันเวลา จากนั้นนำมาแทนค่าประเภทด้วยค่าคะแนนที่เปรียบเสมือนระดับความสนใจที่มีต่อประกาศขายอสังหาริมทรัพย์ โดยให้คะแนนการเข้าชมหน้าประกาศขายอสังหาริมทรัพย์เป็น 1 คะแนนการเข้าดูรายละเอียดของประกาศขายอสังหาริมทรัพย์เป็น 2 และคะแนนการเข้าดูหมายเลขโทรศัพท์ของผู้ลงประกาศเป็น 3 แล้วนำมารวมกันเป็นข้อมูลความสัมพันธ์ของผู้ใช้และหน้าประกาศขายอสังหาริมทรัพย์ จากนั้นนำข้อมูลมาคัดเลือกเฉพาะข้อมูลธุรกรรมของผู้ใช้ที่เข้าชมหน้าประกาศขายอสังหาริมทรัพย์ตั้งแต่ 3 หน้าถึง 20 หน้า เพื่อให้สามารถระบุพฤติกรรมการใช้งานของผู้ใช้ได้และสามารถนำไปฝึกสอนแบบจำลองแบบ Collaborative Filtering ได้

3.4 การฝึกสอนและทดสอบระบบ

โครงสร้างระบบแนะนำประกาศขายอสังหาริมทรัพย์แบบไฮบริดจะประกอบด้วยแบบจำลอง 2 แบบจำลอง ประกอบด้วยแบบจำลองแบบ Content - Based ที่ใช้ขั้นตอนวิธีเพื่อนบ้านที่ใกล้ที่สุดและแบบจำลองแบบ Collaborative Filtering ที่ใช้ขั้นตอนวิธีการแยกค่าเอกลักษณ์

3.4.1 การฝึกสอนแบบจำลองแบบ Content - Based

ก่อนที่จะนำไปคำนวณในขั้นตอนวิธีเพื่อนบ้านที่ใกล้ที่สุดจะต้องนำค่าข้อมูลมาทำให้เป็นบรรทัดฐาน (Normalization) ด้วยการทำให้ค่าข้อมูลอยู่ในช่วงตั้งแต่ 0 - 1 เพื่อให้ค่าข้อมูลอยู่ในมาตราส่วนเดียวกันและลดความผิดพลาดของแบบจำลองที่จะเกิดขึ้น จากนั้นนำชุดค่าข้อมูลที่ผ่านมาการทำให้เป็นบรรทัดฐานแล้วมาคำนวณหาระยะทาง โดยในโครงงานนี้จะใช้ระยะทางแบบ

ยูคลิด (Euclidean Distance) ระบบแนะนำของแบบจำลองนี้ทำงานโดยการส่งค่าข้อมูลที่เป็นคุณสมบัติของประกาศขายอสังหาริมทรัพย์เข้าไปคำนวณระยะห่างกับค่าข้อมูลคุณสมบัติของประกาศขายอสังหาริมทรัพย์อื่น ๆ ทั้งหมดโดยรอบแล้วเรียงลำดับประกาศขายอสังหาริมทรัพย์ที่มีค่าระยะห่างจากน้อยที่สุดไปมากที่สุด แล้วเลือก k ประกาศ โดยที่ค่า k เป็นค่าคงที่ที่เราเลือกเอง

3.4.2 การฝึกสอนแบบจำลองแบบ Collaborative Filtering เริ่มจากนำข้อมูลธุรกรรมและข้อมูลประกาศขายอสังหาริมทรัพย์มารวมกันแล้ว pivot เป็นเมทริกซ์ความสัมพันธ์ของผู้ใช้และประกาศขายอสังหาริมทรัพย์ โดยที่ค่าข้อมูลภายในเป็นค่ารวมของค่าคะแนนการให้คะแนนของผู้ใช้แต่ละคนต่อประกาศขายอสังหาริมทรัพย์นั้น ๆ จากนั้นนำไปแยกตัวประกอบเมทริกซ์ด้วยการแยกค่าเอกฐาน จากนั้นจะได้เมทริกซ์หนึ่งนำมาค้นหาค่าสหสัมพันธ์ของแต่ละหน้าประกาศขายอสังหาริมทรัพย์ระบบแนะนำของแบบจำลองนี้ทำงานโดยส่งค่าข้อมูลประกาศอสังหาริมทรัพย์ เพื่อค้นหาประกาศขายอสังหาริมทรัพย์ที่มีค่าสหสัมพันธ์ต่อประกาศที่เลือกเข้าใกล้ 1 นั่นคือมีความแปรผันตรงต่อประกาศที่เลือกไม่ต่ำกว่าที่กำหนด ในที่นี้กำหนดให้เลือกประกาศขายอสังหาริมทรัพย์ที่มีค่าสหสัมพันธ์ไม่ต่ำกว่า 0.99 มาแนะนำ

3.4.3 การแนะนำแบบไฮบริด เมื่อได้ชุดคำแนะนำประกาศขายอสังหาริมทรัพย์จากแต่ละแบบจำลองแล้วจะนำผลลัพธ์ของทั้งสองแบบจำลองมารวมกันโดยแยกค่าคะแนนที่ได้จากแต่ละแบบจำลองออกจากกัน หมายความว่า หากประกาศใดถูกแนะนำจากแบบจำลองหนึ่ง จะกำหนดให้ค่าคะแนนที่จะได้จากอีกแบบจำลองที่ไม่ได้แนะนำเป็น 0 เว้นแต่ประกาศนั้นจะถูกแนะนำทั้งสองแบบจำลองจะสามารถนำมารวมกันได้เลย จากนั้นจะใช้ค่าน้ำหนักแบบเฉลี่ยหรือถ่วงด้วยค่าคงที่ โดยจะ

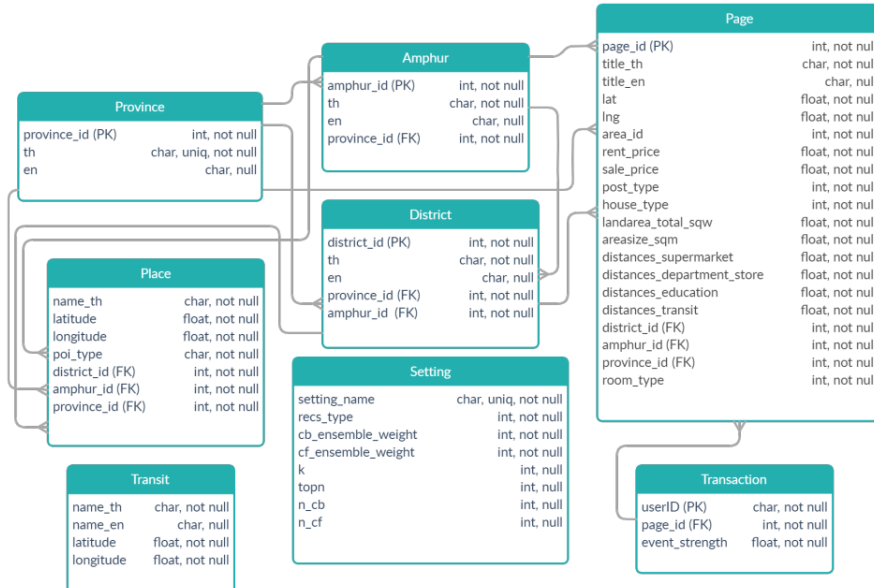
กำหนดค่าคงที่สองค่าสำหรับคูณกับแต่ละค่าคะแนนแล้วนำมารวมกันเป็นค่าคะแนนของระบบแนะนำ ตามสมการดังนี้

$$hb = (cb \times w_1) + (cf \times w_2)$$

โดยที่ hb คือค่าคะแนนของระบบไฮบริด cb คือค่าคะแนนจากแบบจำลองแบบ Content - Based cf คือค่าคะแนนจากแบบจำลองแบบ Collaborative Filtering w_1 คือ ค่าน้ำหนักแบบเฉลี่ยสำหรับแบบจำลองแบบ Content - Based และ w_2 คือ ค่าน้ำหนักแบบเฉลี่ยสำหรับแบบจำลองแบบ Collaborative Filtering ตามลำดับ จากนั้นจัดลำดับคะแนนจากมากไปหาน้อยแล้วแนะนำเป็นชุดประกาศขายอสังหาริมทรัพย์ คณะผู้จัดทำได้ออกแบบการแนะนำของระบบออกเป็น 3 ประเภท ได้แก่ การแนะนำแบบปกติที่ใช้ค่าน้ำหนักแบบเฉลี่ย การแนะนำแบบปกติที่ใช้ค่าน้ำหนักแบบเฉลี่ยที่ยกประกาศจากแบบจำลองแบบ Content - based ที่ใกล้เคียงที่สุด 3 อันดับแรกขึ้นก่อน และการแนะนำที่ไม่ใช้ค่าน้ำหนักแบบเฉลี่ยแต่กำหนดจำนวนประกาศที่ต้องการแนะนำในแต่ละแบบจำลองแทน

3.5 การพัฒนา Application Programming Interface (API)

เนื่องจากทางบริษัทต้องการระบบแนะนำที่สามารถใช้งานได้หลากหลายแพลตฟอร์ม คณะผู้จัดทำจึงเลือกที่จะพัฒนา API เป็น Web Service พัฒนาด้วยภาษา Python และใช้ Django Framework เป็นตัวช่วยในการสร้าง ในลักษณะของเว็บแอปพลิเคชันตัวอย่างสาธิตอย่างง่ายหรือ Demo และ API ที่จะเป็นส่วนที่จัดการข้อมูลและการแนะนำ เพื่อให้ระบบแนะนำทำงานด้วยความรวดเร็วมากขึ้นและสามารถจัดการข้อมูลได้ดียิ่งขึ้น จึงเริ่มต้นในฝั่ง Back - End ด้วยการออกแบบฐานข้อมูลเพื่อเป็นการวางโครงสร้างในการเข้าถึงข้อมูลของระบบแนะนำ ซึ่งมีโครงสร้างดังรูปที่ 6



รูปที่ 6 แบบจำลองความสัมพันธ์ของข้อมูลของระบบแนะนำประกาศขายอสังหาริมทรัพย์

จากนั้นจึงเริ่มพัฒนาในส่วนโครงสร้างและเชื่อมต่อกับฐานข้อมูลโดยใช้ MySQL เป็นระบบจัดการฐานข้อมูลเพื่อให้เกิดความสอดคล้องกับฐานข้อมูลจริงของบริษัท แล้วจึงจัดวางระบบแนะนำอยู่ในส่วนของ API โดยพัฒนา API แนะนำออกเป็น 3 แบบ ได้แก่

1) การแนะนำแบบค่าตั้งต้น เป็นการแนะนำประกาศขายอสังหาริมทรัพย์โดยค่าที่ใช้เป็นพารามิเตอร์ในการแนะนำ เช่น คำน้ำหนักแบบเฉลี่ยของแต่ละแบบจำลอง ประเภทการแนะนำของระบบ ค่า K ค่าจำนวนประกาศที่ต้องการแนะนำ และค่าจำนวนประกาศที่ได้จากแต่ละแบบจำลอง จะใช้ค่าตั้งต้นที่เป็นพื้นฐานที่ระบบตั้งให้ไว้แล้วในระบบ โดยที่ผู้ใช้ไม่ต้องตั้งค่าใด ๆ

2) การแนะนำแบบมีพารามิเตอร์ เป็นการแนะนำประกาศขายอสังหาริมทรัพย์ที่ผู้ใช้ต้องป้อนค่าพารามิเตอร์ที่ใช้ในการแนะนำประกาศเองทั้งหมด ได้แก่ คำน้ำหนักแบบเฉลี่ยของแต่ละแบบจำลอง ประเภทการแนะนำของระบบ ค่า K ค่า

จำนวนประกาศที่ต้องการแนะนำ และค่าจำนวนประกาศที่ได้จากแต่ละแบบจำลอง

3) การแนะนำแบบใช้ชุดการตั้งค่า เป็นการแนะนำประกาศขายอสังหาริมทรัพย์โดยค่าที่ใช้เป็นพารามิเตอร์ในการแนะนำจะถูกดึงมาจากชุดการตั้งค่าที่ถูกเก็บไว้ในฐานข้อมูลที่ผู้ใช้สามารถสร้างชุดการตั้งค่ารูปแบบต่าง ๆ ไว้เรียกใช้ได้ โดยระบุชื่อชุดการตั้งค่าแทนพารามิเตอร์

นอกจากนี้ คณะผู้จัดทำได้สร้าง REST API สำหรับจัดการในส่วนของการเข้าถึงข้อมูลที่ใช้ในระบบแนะนำได้ ได้แก่ ข้อมูลประกาศขายอสังหาริมทรัพย์ ข้อมูลธุรกรรม ข้อมูลที่อยู่ (ตำบล อำเภอ และจังหวัด) ข้อมูลสถานที่และข้อมูลสถานีรถไฟฟ้าทั้งในรูปแบบของข้อมูลเดี่ยว และการอัปโหลดไฟล์ชุดข้อมูลประเภท CSV เพื่อที่จะสามารถปรับข้อมูลได้ในอนาคตและปรับค่าคุณสมบัติของประกาศขายอสังหาริมทรัพย์ให้ตรงกับความจริงในปัจจุบัน

4. ผลการวิจัยและอภิปรายผล

ระบบแนะนำสามารถแนะนำประกาศขายอสังหาริมทรัพย์ที่มีความใกล้เคียงกับประกาศขายอสังหาริมทรัพย์ที่กำลังเข้าชมและมีความหลากหลาย

4.1 ผลการแนะนำประกาศขายอสังหาริมทรัพย์แบบค่าตั้งต้นและแบบมีพารามิเตอร์

ผู้ใช้สามารถใช้ระบบแนะนำได้โดยการเรียกใช้ API และส่งค่า id ของประกาศขายอสังหาริมทรัพย์ที่เข้าชมในปัจจุบันเพียงอย่างเดียว ผลลัพธ์

ผู้ใช้สามารถใช้ระบบแนะนำได้โดยการเรียกใช้ API และส่งค่า id ของประกาศขายอสังหาริมทรัพย์ที่เข้าชมในปัจจุบันประกอบกับค่าพารามิเตอร์ ได้แก่ ค่าน้ำหนักแบบเฉลี่ยสำหรับแบบจำลองแบบ Content – Based ค่าน้ำหนักแบบเฉลี่ยสำหรับแบบจำลองแบบ Collaborative Filtering ค่า K ค่าจำนวนประกาศขายอสังหาริมทรัพย์ที่ต้องการให้แนะนำ (topn) ค่าจำนวนประกาศขายอสังหาริมทรัพย์ที่ต้องการจากแบบจำลองแบบ Content – Based และค่าจำนวนประกาศขายอสังหาริมทรัพย์ที่ต้องการจากแบบจำลองแบบ Collaborative Filtering ตัวอย่างการป้อนค่าพารามิเตอร์

4.2 ผลการแนะนำประกาศขายอสังหาริมทรัพย์แบบค่าตั้งต้นและแบบมีพารามิเตอร์

ผู้ใช้สามารถใช้ระบบแนะนำได้โดยการเรียกใช้ API และส่งค่า id ของประกาศขายอสังหาริมทรัพย์ที่เข้าชมในปัจจุบัน ประกอบกับชื่อการตั้งค่าที่ต้องการใช้งานและถูกจัดเก็บไว้แล้วในฐานข้อมูล ตัวอย่างการป้อนค่าผลลัพธ์ที่ได้และตัวอย่างชุดการตั้งค่า

ผู้ใช้สามารถเข้าถึงและจัดการข้อมูลชุดการตั้งค่าได้ทั้งหมด ทั้งการเรียกดู เพิ่มชุดการตั้งค่า แก้ไขการตั้งค่า และลบการตั้งค่า ดังรูปที่ 7

Recommendation for 16149			
ให้เช่า Condo ใกล้รถไฟฟ้า MRT สถานี Regent Home 5 ถ.สุขาภิบาล 16 แขวงบ้านดิน อ.เมือง จ.นนทบุรี 11000.0 0.0			
page_id	title_th	rent_price	sale_price
760348	***ใกล้รถไฟฟ้าสถานี Regent Home 5 (10.0.0.0)	11000.0	0.0
557438	Ideo Ratchada-Huaykwang	13000.0	0.0
71887	Ideo Ratchada-Huaykwang	13000.0	0.0
67176	Ideo Ratchada-Huaykwang	12000.0	0.0
695435	คอนโดใหม่ ใกล้รถไฟฟ้า MRT สถานี Regent Home 5 (10.0.0.0)	8500.0	0.0

Setting Name	Type	CB Weight	CF Weight	K	Topn	n_cb	n_cf
setting_0	NORMAL RECOMMENDER	3.0	1.0	10	5	3	2

รูปที่ 7 ตัวอย่างเว็บแอปพลิเคชันสาคิดอย่างง่าย ส่วนแนะนำประกาศขายอสังหาริมทรัพย์แบบใช้ชุดการตั้งค่าหลังส่งค่า และ ตัวอย่างชุดการตั้งค่า

5. สรุปผลการวิจัย

โครงการนี้สามารถแนะนำประกาศขายอสังหาริมทรัพย์ที่ใกล้เคียงกับความต้องการของผู้ใช้ได้ และทางบริษัทสามารถเลือกใช้การแนะนำประเภทและรูปแบบต่าง ๆ ได้ตามความต้องการทั้งแบบค่าตั้งต้น แบบส่งค่าพารามิเตอร์และการใช้ชุดการตั้งค่า นอกจากนี้ยังสามารถบริหารจัดการการเข้าถึงข้อมูลที่ใช้ในระบบแนะนำได้อย่างอิสระและไม่มีปัญหา แต่ในส่วนของการเร็วในการทำงานของระบบแนะนำยังถือว่าค่อนข้างช้า เมื่อตรวจสอบพบว่า ส่วนที่ทำให้ระบบแนะนำค่อนข้างช้าคือส่วนของแบบจำลองแบบ Content – Based ที่ใช้ขั้นตอนวิธีเพื่อนบ้านที่ใกล้ที่สุด เนื่องจากมีขั้นตอนของการตรวจสอบค่าข้อมูลคุณสมบัติที่ขาดหายไปของประกาศขายอสังหาริมทรัพย์ จึงทำให้ต้องมีการจัดเตรียมข้อมูลและเข้าแบบจำลองทุกครั้งที่ใช้เรียกใช้งานระบบเพื่อให้ระบบแนะนำ



ประกาศได้ใกล้เคียงกับความต้องการของผู้ใช้มากที่สุดสุดท้ายคือในขั้นตอนการใช้งานจริงจำเป็นต้องตั้งค่าสภาพแวดล้อมให้พร้อมก่อนใช้งาน เช่น Library ที่ใช้ใน ระบบ การตั้งค่าระบบฐานข้อมูล การใช้งานครั้งแรกจำเป็นต้องเพิ่มข้อมูลเข้าในฐานข้อมูล และจำเป็นต้องเรียกใช้ API ที่จัดการข้อมูลประกาศและข้อมูลธุรกรรมทุกครั้งที่มีการเพิ่มประกาศ เปลี่ยนสถานะประกาศ และลบประกาศเพื่อให้ข้อมูลในระบบเป็นปัจจุบันที่สุด

6. อ้างอิง

- [1] TerraBKK. Real Estate Information and Services. Available from: <https://www.terrabkk.com>
- [2] Poonam B. T., R. M. Goudar, Sunita B., "Survey on Collaborative Filtering, Content-Based Filtering and Hybrid Recommendation System," International Journal of Computer Applications. 2015; 110(4): 31 – 36
- [3] Altman, Naomi S. (1992)., "An introduction to kernel and nearest-neighbor nonparametric regression," The American Statistician. 1992; 46(3): 175–185.
- [4] Coomans, D., & Massart, D. L. (1982)., "Alternative k-nearest neighbour rules in supervised pattern recognition," Analytica Chimica Acta. 1982; 136: 15–27.
- [5] Rounick P., Rajdeep C. (2020)., "Recommender System Using K-Nearest Neighbors and Singular Value Decomposition Algorithms: A Hybrid Approach," Progress in Computing, Analytics and Networking. Advances in Intelligent Systems and Computing. 2020; 497 – 504.
- [6] Vaibhav K., "Singular Value Decomposition (SVD) & Its Application In Recommender System" [Internet]. 2020. [Retrieved 2020 Aug 19]. Available from: <https://analyticsindiamag.com/singular-value-decomposition-svd-application-recommender-system/>
- [7] Wang L, Hu X, Wei J, Cui X. A collaborative filtering based personalized top-k recommender system for housing. In: Proceedings of the 2012 International Conference of Modern Computer Science and Applications. Wuhan, China: Springer; 2013. p. 461–6.
- [8] Yu Y, Wang C, Zhang L, Gao R, Wang H. Geographical proximity boosted recommendation algorithms for real estate. In: Proceedings of the International Conference on Web Information Systems Engineering. Amsterdam, Netherlands: Springer; 2018. p. 51–66.
- [9] Ojokoh B, Olayemi O, Babalola A, Eyo E. A User-Centric Housing Recommender System. Inf. Manag. Bus. Rev. 2018;10:17–24.
- [10] Chonwiharnphan P, Thienprapasith P, Chuangsuwanich E. Generating realistic users using generative adversarial network with recommendation-based embedding. IEEE Access. 2020;8:41384–93.
- [11] Paterek A. Improving regularized singular value decomposition for collaborative filtering. In: Proceedings of the KDD Cup and Workshop. San Jose, CA, USA; 2007. p. 5–8.