

การศึกษาระบบแนะนำสถานที่ท่องเที่ยวในกรุงเทพฯ และปริมณฑล
สำหรับนักท่องเที่ยวต่างชาติ

A Study of Tourist Attractions Recommendation System in Bangkok
for Foreign Tourist

ภัทร สุตรสุวรรณ¹ เดือนเพ็ญ ธีรรณวิวัฒน์² พาชิตชนัด ศิริพานิช³

Phatara Sutarasuwan¹, Duanpen Teerawanviwat², Pachitjanut Siripanich³

Received April 01, 2021 & Revise June 03, 2021 & Accepted September 30, 2021

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบแนะนำสถานที่ท่องเที่ยวตามรสนิยมสำหรับนักท่องเที่ยวต่างชาติในกรุงเทพฯ และปริมณฑลด้วยการเปรียบเทียบประสิทธิภาพของระบบแนะนำที่พัฒนาจากอัลกอริทึมด้านการแนะนำ ดังต่อไปนี้ (1) วิธีแนะนำตามความนิยม (2) วิธีพิจารณาข้อมูลจากเนื้อหาของสิ่งของ (3) วิธีพิจารณาการกรอกร่วมกัน (4) วิธีการแบบผสม โดยใช้ข้อมูลประวัติการท่องเที่ยว หรือการให้คะแนนของนักท่องเที่ยวต่างชาติที่ดึงมาจากเว็บไซต์ Tripadvisor ที่เป็นเว็บไซต์เกี่ยวกับการท่องเที่ยวที่ใหญ่ที่สุดในโลก ข้อมูลที่ได้จะถูกแบ่งเป็นข้อมูลสำหรับทำตัวแบบ (Training data) เป็น 90% และข้อมูลสำหรับทดสอบ (Testing data) เป็น 10% ซึ่งข้อมูลที่อยู่ในกลุ่มทดสอบจะเป็นข้อมูลสถานที่ท่องเที่ยวล่าสุดที่ผู้ใช้แต่ละคนเคยไปในอดีต จากนั้นวัดประสิทธิภาพของแต่ละอัลกอริทึมด้วยค่าเฉลี่ยของค่าเฉลี่ยความแม่นยำ (MAP) และค่าเฉลี่ยของค่าเฉลี่ยความระลึกได้ (MAR) ในขอบเขตการแนะนำ (k) ตั้งแต่ 1 ถึง 10 ลำดับ ผลการศึกษาพบว่าวิธีการแบบผสมโดยใช้วิธีการเฉลี่ยถ่วงน้ำหนัก (Weighted) มีผลลัพธ์ที่ดีที่สุดในทุกวิธีการทั้งหมด โดยมีค่า

¹ นักศึกษามหาบัณฑิต หลักสูตรการวิเคราะห์ธุรกิจและวิทยาการข้อมูล คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

Master of Science Program in Business Analytics and Data Science, Graduate School of Applied Statistics, National Institute of Development Administration

² รองศาสตราจารย์ หลักสูตรการวิเคราะห์ธุรกิจและวิทยาการข้อมูล คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

Associate Professor, Department of Business Analytics and Data Science, Graduate School of Applied Statistics, National Institute of Development Administration

³ รองศาสตราจารย์ หลักสูตรการวิเคราะห์ธุรกิจและวิทยาการข้อมูล คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

Associate Professor, Department of Business Analytics and Data Science, Graduate School of Applied Statistics, National Institute of Development Administration

* Corresponding author. E-mail: Phatara.sut@stu.nida.ac.th



MAP@10 และค่า MAR @10 เท่ากับ 0.7283 และ 0.9659 รองลงมาจะเป็นวิธีการพิจารณาข้อมูลจากเนื้อหาของสิ่งของ ซึ่งเป็นวิธีการที่มีประสิทธิภาพการทำงานแบบเดียวสูงที่สุด โดยมีค่า MAP@10 และค่า MAR@10 เท่ากับ 0.6991 และ 0.9654 ตามลำดับ ในทางตรงกันข้ามวิธีพิจารณาการกรอกร่วมกันทั้งสามแบบเป็นวิธีที่มีประสิทธิภาพน้อยที่สุด

คำสำคัญ: ระบบแนะนำสถานที่ท่องเที่ยว วิธีพิจารณาการกรอกร่วมกัน วิธีพิจารณาข้อมูลจากเนื้อหาของสิ่งของ วิธีแนะนำแบบผสม

Abstract

The objective of this research is to develop a personalized recommendation system for foreigners in Bangkok Metropolitan Region by comparing the performance of recommendation system that developed from (1) Popularity based (2) Content-based (3) Collaborative filtering (4) Hybrid model by using travel historical data of foreign tourists that scraped from Tripadvisor, which was the world's largest travel site. The data was split to training data for 90% and testing data for 10%, which data in the testing set was the latest attractions that each user visited in the past. Then evaluated a performance of each algorithms with mean average precision (MAP) and mean average recall (MAR) in range of recommend list(k) from 1 to 10. Experimental results demonstrate that Hybrid model, which using the weighted average method given the best results that has MAP@10 and MAR@10 values of 0.7283 and 0.9659, followed by Content-based which was the best independent algorithm, has MAP@10 and MAR@10 values of 0.6991 and 0.9654, respectively. In the other hand, Collaborative filtering given the worst results.

Keywords: Tourist attractions recommender system, Collaborative filtering, Content-based, Hybrid model

ความเป็นมาและความสำคัญของปัญหา

ภาคการท่องเที่ยวของประเทศไทยเป็นส่วนที่สร้างรายได้ให้กับประเทศไทยอย่างมหาศาล โดยสัดส่วนรายได้ด้านการท่องเที่ยวในผลิตภัณฑ์มวลรวมของประเทศมีมากกว่าร้อยละ 21.6 เป็นจำนวนเงินกว่า 3 พันล้านบาทต่อปี และมีแนวโน้มเพิ่มขึ้นในช่วงเวลาที่ผ่านมา (Knoema, 2020) อีกทั้งแต่ละปีจะมีนักท่องเที่ยวเดินทางมาเที่ยวในประเทศไทยกว่าปีละ 40 ล้านคน สร้างรายได้ให้กับประเทศมากกว่าปีละ 2



พันล้านบาท กล่าวได้ว่า 2 ใน 3 ของรายได้ด้านการท่องเที่ยวของประเทศไทยเป็นรายได้ที่มาจากนักท่องเที่ยวต่างชาติ (พชรพจน์ นันทรามาศ, กิตติพงษ์ เรือนทิพย์, และจรรุวรรณ เหล่าสัมฤทธิ์, 2563) (Nantaramat, P., Rueanthip, K. & Laosamrit, J., 2020) แสดงให้เห็นว่านักท่องเที่ยวต่างชาติเป็นส่วนสำคัญในการสร้างรายได้ให้แก่ประเทศไทย และเป็นส่วนสำคัญสำหรับเศรษฐกิจของประเทศไทย

การวางแผนท่องเที่ยวของนักท่องเที่ยวต่างชาติโดยปกติจะทำการวางแผนล่วงหน้าก่อนการเดินทางเป็นระยะเวลานาน และหาข้อมูลสถานที่ท่องเที่ยวต่าง ๆ จากการค้นหาข้อมูลในเว็บไซต์รีวิวสถานที่ท่องเที่ยวบนอินเทอร์เน็ตที่มีข้อมูลเป็นจำนวนมาก ซึ่งบางส่วนเป็นข้อมูลที่เป็นเท็จ หรือบรรยายสรรพคุณเกินจริง ทำให้ไม่สะดวกสบายในการรวบรวม และพิจารณาข้อมูลเนื่องจากต้องใช้เวลาในการค้นหา อีกทั้งในปัจจุบันเว็บไซต์ที่มีการแนะนำสถานที่ท่องเที่ยวส่วนใหญ่ มักจะใช้ระบบแนะนำที่ไม่มีประสิทธิภาพ โดยมักจะแนะนำเพียงสถานที่ท่องเที่ยวยอดนิยม และไม่ได้พิจารณาถึงรสนิยมส่วนบุคคล ทำให้โอกาสที่นักท่องเที่ยวต่างชาติจะกลับมาเที่ยวในสถานที่เดิมนั้นมีน้อย เพราะได้รับข้อมูลสถานที่ท่องเที่ยวที่จำกัด เกิดเป็นความไม่หลากหลายของการท่องเที่ยว รวมถึงสถานที่ท่องเที่ยวที่เพิ่งจัดตั้งขึ้นมาใหม่จะไม่มีโอกาสเป็นที่รู้จัก ดังนั้นระบบแนะนำสถานที่ท่องเที่ยวที่มีประสิทธิภาพจึงมีความสำคัญ เนื่องจากการเพิ่มความสะดวกสบาย และประหยัดเวลาในการหาข้อมูลให้กับนักท่องเที่ยวต่างชาติ ทั้งยังเป็นตัวช่วยในการตัดสินใจเลือกสถานที่ที่จะเดินทางไป และเพิ่มโอกาสให้กับสถานที่ท่องเที่ยวที่มีคุณภาพ แต่ยังไม่เป็นที่รู้จักให้นักท่องเที่ยวต่างชาติเดินทางไปเที่ยวมากขึ้น ซึ่งจะช่วยให้สถานที่ท่องเที่ยวดังกล่าวเป็นที่รู้จักมากขึ้น

จากข้อมูลข้างต้นทำให้ผู้วิจัยเห็นถึงความสำคัญของปัญหาดังกล่าว ผู้วิจัยจึงต้องการศึกษา และพัฒนาระบบแนะนำ ซึ่งเป็นอัลกอริทึมเบื้องหลังในการแนะนำสถานที่ท่องเที่ยวให้กับนักท่องเที่ยวต่างชาติบนเว็บไซต์ และแอปพลิเคชันที่มีประสิทธิภาพมากที่สุด โดยการนำข้อมูลประวัติการท่องเที่ยวของนักท่องเที่ยวต่างชาติมาให้อัลกอริทึมด้านการแนะนำรูปแบบต่าง ๆ เรียนรู้ จากนั้นจึงจับรูปแบบจนกระทั่งสามารถแนะนำสถานที่ท่องเที่ยวที่นักท่องเที่ยวแต่ละคนมีโอกาสชอบมากที่สุด ทั้งนี้ผู้วิจัยเลือกศึกษาเฉพาะสถานที่ท่องเที่ยวในกรุงเทพฯ และปริมณฑล เนื่องจากเป็นจังหวัดที่มีนักท่องเที่ยวเข้ามาเที่ยวมากที่สุดในประเทศไทย และสร้างรายได้ให้กับประเทศไทยมากที่สุด (ฉิชาภัทร สุรวัฒนานนท์, 2562) (Surawattananon, N., 2019)

วัตถุประสงค์ของการวิจัย

เพื่อพัฒนาระบบแนะนำสถานที่ท่องเที่ยวตามรสนิยม (Personalize recommendation) สำหรับนักท่องเที่ยวต่างชาติ โดยการเปรียบเทียบระหว่างอัลกอริทึมที่พิจารณาข้อมูลจากเนื้อหาของสิ่งของ (Content-based), การพิจารณาการกรองร่วมกัน (Collaborative filtering), การแนะนำตามความนิยม (Popularity based), และวิธีการแบบผสม (Hybrid model)

บททวนวรรณกรรม



1. แนวคิดเกี่ยวกับระบบแนะนำ

ระบบแนะนำเป็นเครื่องมือคัดกรองที่ช่วยให้สามารถค้นหา และสกัดเฉพาะข้อมูลที่น่าสนใจออกมาจากข้อมูลขนาดใหญ่ หรือตัวเลือกอื่น ๆ จำนวนมาก ซึ่งเป็นการลดอุปสรรค และความสับสนในการค้นหาข้อมูล ตัวอย่างเช่น ในโซเชียลมีเดีย (Social media) มีผู้ใช้งานเป็นจำนวนมาก ทำให้ยากที่จะหาคนที่ต้องการสร้างความสัมพันธ์ด้วยพบได้ ระบบแนะนำจึงเข้ามามีบทบาทอย่างมากในการแนะนำบุคคลที่ผู้ใช้งานอาจต้องการสร้างความสัมพันธ์ด้วย เช่น ระบบแนะนำเพื่อนของ Facebook (Pourheidari, Mollashahi, Vassileva and Deters, 2018) ซึ่งนิยามข้างต้นได้ตรงกับงานวิจัยของ (Das, Sahoo, and Datta, 2017) ที่ศึกษาเกี่ยวกับระบบแนะนำ และเทคนิคที่ใช้ในการสร้างระบบแนะนำ โดยผู้วิจัยกลุ่มนี้ได้ให้คำนิยามของระบบแนะนำว่าเป็นระบบที่ช่วยให้ผู้ใช้สามารถหาสิ่งที่ตัวเองสนใจได้โดยไม่ต้องค้นหาจากสิ่งของทั้งหมด และแบ่งระบบแนะนำออกเป็น 5 แบบ ดังนี้

1.1 วิธีพิจารณาข้อมูลจากเนื้อหาของสิ่งของ หรือ Content-based approach (CB) เป็นวิธีที่ต้องใช้ข้อมูลเกี่ยวกับคุณลักษณะของวัตถุ หรือสิ่งที่จะแนะนำ จากนั้นจะแนะนำสิ่งที่มีลักษณะเหมือนกัน หรือใกล้เคียงกันกับวัตถุที่ผู้ใช้เคยใช้ และชื่นชอบในอดีตให้

1.2 วิธีพิจารณาการกรองร่วมกัน หรือ Collaborative filtering เป็นวิธีที่แนะนำโดยพิจารณาจากข้อมูลผู้ใช้เป้าหมาย และผู้อื่น ๆ ในอดีต ซึ่งวิธีการนี้สามารถทำได้หลายรูปแบบ แต่วิธีที่เป็นที่นิยมมากที่สุดจะมี 3 วิธีหลัก ๆ ได้แก่ (1) User-based collaborative filtering วิธีนี้จะแนะนำวัตถุที่ผู้ใช้เป้าหมายไม่เคยใช้จากวัตถุทั้งหมดที่ผู้อื่นที่มีรสนิยมใกล้เคียงกันเคยใช้ให้กับผู้ใช้งานเป้าหมาย ซึ่งวิธีการหาผู้ใช้ที่มีรสนิยมเหมือนกันโดยปกติจะต้องใช้ข้อมูลการให้คะแนน (Rating) ของผู้ใช้งานทั้งหมดเพื่อให้อัลกอริทึมทำการเรียนรู้ และหาค่าความสัมพันธ์ (Correlation) ของผู้ใช้เป้าหมายกับผู้อื่น ๆ ทั้งหมด ยิ่งค่า Correlation

ของผู้ใช้งานเป้าหมายกับผู้อื่นมากเท่าใด แสดงว่ารสนิยมของผู้ใช้คู่ดังกล่าวนี้มีความใกล้เคียงกันมากขึ้นเท่านั้น (2) Item-based collaborative filtering วิธีการนี้จะแนะนำวัตถุที่มีลักษณะใกล้เคียงกับสิ่งที่ผู้ใช้เป้าหมายเคยใช้ในอดีตให้ โดยวิธีการหาวัตถุที่มีความเหมือนกันก็จะใช้ค่า Correlation เช่นเดียวกับ User-based (3) Singular value decomposition วิธีการนี้จะนำหลักการการแยกองค์ประกอบของเมทริกซ์มาใช้ โดยพยายามจะแยกเมทริกซ์หลักออกเป็นเมทริกซ์ย่อย ๆ 3 ส่วน จากนั้นจะใช้ Optimization algorithm อย่าง Gradient descent ในการหาค่าพารามิเตอร์ที่เหมาะสมกับเมทริกซ์นั้น ๆ และนำผลลัพธ์ที่ได้มาหาค่าคะแนนทำนายที่จะใช้ในการแนะนำวัตถุต่าง ๆ ให้กับผู้ใช้เป้าหมาย

1.3 วิธีพิจารณาโดยใช้ข้อมูลส่วนตัว หรือ Demographic filtering เป็นวิธีการที่แนะนำโดยพิจารณาข้อมูลส่วนตัวของบุคคล ตัวอย่างเช่น เพศ, อายุ, อาชีพ, ที่อยู่, หรือสัญชาติ เป็นต้น แต่วิธีการนี้ไม่เป็นที่นิยมมากนัก เนื่องจากการมีกฎหมายคุ้มครองข้อมูลส่วนบุคคล (Privacy policy) ทำให้เข้าถึงข้อมูลได้ยาก



1.4 วิธีการใช้ความรู้ที่ได้จากผู้ให้โดยตรง หรือ Knowledge-based recommender เป็นวิธีการที่จะสอบถามความชอบของผู้ใช้โดยตรงว่าให้คุณค่ากับคุณลักษณะใดของสินค้า เช่น ราคาสินค้า, ความนิยมของสินค้า, สีของวัตถุ หรือความใหม่ของสินค้า เป็นต้น

1.5 วิธีแบบผสม หรือ Hybrid approach เป็นวิธีการที่รวมผลลัพธ์ของวิธีการอื่น ๆ เข้าด้วยกันตั้งแต่ 2 วิธีขึ้นไป เป็นการนำข้อดีของแต่ละวิธีมารวมกัน และลบจุดอ่อนของบางวิธีลง โดยปกติจะให้ผลลัพธ์ที่ดีกว่าการใช้อัลกอริทึมสำหรับแนะนำทั่วไป แต่ข้อเสียของวิธีการนี้ คือ ไม่สามารถแก้ปัญหาผู้เริ่มต้นใหม่ในระบบ (Cold-start problem) ได้ทั้งหมด และมีกระบวนการทำงานที่ซับซ้อน

ในปัจจุบันระบบแนะนำสามารถนำไปประยุกต์ใช้กับการแนะนำสิ่งต่าง ๆ นอกจากสินค้าได้มากมาย ไม่ว่าจะเป็นภาพยนตร์, เพลง, หรือคลิปวิดีโอ เป็นต้น โดยในภาคการท่องเที่ยวเองก็เริ่มมีการนำระบบแนะนำมาประยุกต์ใช้เพื่อแนะนำสถานที่ท่องเที่ยวต่าง ๆ ให้แก่ผู้บริโภค ดังเช่นงานวิจัยของ Chujai, Singthongchai, Yasaga, Surattgara, and Buranakutti (2020) ที่สร้างระบบแนะนำสถานที่ท่องเที่ยวด้วยวิธีการ Collaborative filtering โดยใช้วิธีการเก็บข้อมูลจากแบบสอบถามกับนักท่องเที่ยวจำนวน 100 คน ผลลัพธ์ที่ได้มีค่าความแม่นยำ (Precision) เท่ากับ 76.40%

2. แนวคิดการแยกองค์ประกอบของเมทริกซ์

การแยกองค์ประกอบของเมทริกซ์ หรือ Matrix factorization เป็นวิธีการที่ใช้ในการแยกเมทริกซ์หลักที่มีขนาดใหญ่ออกเป็นเมทริกซ์ย่อยที่มีขนาดเล็กลง (นัชชา ถิระสาโรช และวิโรจน์ อรุณมานะกุล, 2560) (Tirasaroj, N. & Arunmanakul, V., 2017) โดยในเรื่องของระบบแนะนำจะมีการนำเทคนิคที่เรียกว่า Singular value decomposition (SVD) ซึ่งหลักการของวิธีการนี้ คือ ต้องการแยกเมทริกซ์หลักออกเป็นเมทริกซ์ย่อย 3 ส่วน ที่ประกอบไปด้วย เมทริกซ์ที่เก็บคุณลักษณะของผู้ใช้, เมทริกซ์ของปัจจัยแฝง, และเมทริกซ์ที่เก็บคุณลักษณะของวัตถุที่จะแนะนำ จากนั้นจะพยายามปรับค่าภายในเมทริกซ์ทั้ง 3 ส่วนด้วย Gradient descent จนกว่าจะทำให้เมื่อ Dot product เมทริกซ์ทั้ง 3 เข้าด้วยกันแล้วจะทำให้ค่ากลับไปใกล้เคียงกับเมทริกซ์หลักมากที่สุด ผลลัพธ์ที่ได้จะทำให้ข้อมูลสูญหาย (Missing value) ในเมทริกซ์หลักถูกเติมเต็ม และสามารถนำไปใช้ในการแนะนำต่อไป (Sarwar, Karypis, Konstan, and Riedl, 2001)

3. แนวคิดการรวมผลลัพธ์ใน Hybrid model

ในการทำระบบแนะนำด้วยวิธีการแบบผสมจะต้องมีขั้นตอนในการรวมผลลัพธ์ของแต่ละเทคนิคเข้าด้วยกัน วิธีการรวมผลลัพธ์ดังกล่าวสามารถแบ่งได้เป็น 7 วิธี ได้แก่ (1) Weighted (2) Switching (3) Mixed (4) Feature combination (5) Cascade (6) Feature augmentation (7) Meta-level (Bruke, 2002) โดยในงานวิจัยครั้งนี้จะใช้วิธีการ Weighted ซึ่งเป็นการเฉลี่ยถ่วงน้ำหนักของผลลัพธ์จากวิธีต่าง ๆ เป็นหลัก เนื่องจากเป็นวิธีการที่เหมาะสมกับลักษณะของผลลัพธ์ที่ได้จากแต่ละอัลกอริทึม

4. แนวคิดการวัดประสิทธิภาพระบบแนะนำ

การวัดประสิทธิภาพของระบบแนะนำมีวิธีการที่หลากหลายขึ้นอยู่กับลักษณะของข้อมูล และตัวอัลกอริทึมที่ใช้ในการสร้าง แต่โดยทั่วไปแล้ววิธีการวัดผลจะแบ่งออกเป็น 2 รูปแบบ ได้แก่



4.1 การวัดผลแบบออนไลน์ วิธีการนี้เป็นการนำระบบแนะนำที่สร้างเสร็จแล้วไปทดลองใช้กับผู้ใช้จริง เช่น การติดตั้งระบบแนะนำลงบนเว็บไซต์ และวัดจำนวนการซื้อสินค้าของลูกค้าที่ได้รับการแนะนำ

4.2 การวัดผลแบบออฟไลน์ วิธีการนี้จะเป็นการทดสอบประสิทธิภาพด้วยการดูค่าสถิติต่าง ๆ ซึ่งปกติจะต้องแบ่งข้อมูลเป็น 2 ส่วน ส่วนแรกสำหรับให้อัลกอริทึมทำการเรียนรู้ (Training data) และส่วนที่สองเป็นข้อมูลสำหรับทดสอบ (Testing data) จากนั้นจะนำข้อมูลในส่วนแรกไปสร้างระบบแนะนำ และทดสอบประสิทธิภาพด้วยข้อมูลส่วนที่สอง (Shani and Gunawardana, 2011) โดยในงานวิจัยนี้จะใช้การวัดผลแบบออฟไลน์ด้วยวิธีการทดสอบกับคะแนนทำนาย ซึ่งมีรายละเอียดดังต่อไปนี้

วิธีการทดสอบคะแนนทำนาย

เป็นการวัดผลของระบบแนะนำโดยดูจากความแม่นยำของระบบว่าสามารถแนะนำวัตถุที่ตรงกับข้อมูลสำหรับทดสอบได้มากแค่ไหน ทั้งนี้ผู้วิจัยมีสมมติฐานว่าข้อมูลใน Testing data เป็นข้อมูลวัตถุที่ผู้ใช้งานจะซื้อในอนาคต ค่าความแม่นยำดังกล่าวสามารถคำนวณได้ดังสมการที่ 2.1 และ 2.2

$$Precision = \frac{TP}{TP+FP} \quad (2.1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2.2)$$

เมื่อ TP แทนจำนวนสิ่งที่ระบบแนะนำได้ตรงกับสิ่งที่ผู้ใช้เป้าหมายชอบ

FP แทนจำนวนสิ่งที่ระบบแนะนำได้ไม่ตรงกับสิ่งที่ผู้ใช้เป้าหมายชอบ

FN แทน จำนวนวัตถุที่ผู้ใช้เป้าหมายชอบ แต่ระบบไม่ได้แนะนำ

ทั้งนี้การวัดผลโดยใช้เพียงค่า Precision และ Recall จะทำให้ไม่สามารถระบุได้ว่าระบบแนะนำสามารถแนะนำวัตถุที่ตรงกับสิ่งที่อยู่ใน Testing data ได้ตั้งแต่การแนะนำลำดับที่เท่าใด ดังนั้นเพื่อให้สามารถนำผลของลำดับในการแนะนำมาพิจารณาด้วยได้ งานวิจัยในครั้งนี้จึงใช้ค่าเฉลี่ยความแม่นยำ (Average precision) กับค่าเฉลี่ยของความแม่นยำเฉลี่ย (Mean average precision) และค่าเฉลี่ยความระลึกได้ (Average recall) กับค่าเฉลี่ยของค่าเฉลี่ยความระลึกได้ (Mean average recall) ในการวัดประสิทธิภาพของระบบ (Sawtelle, 2020) (Taifi, 2020) โดยมีสมการดังต่อไปนี้

$$AP@K = \frac{1}{M} \sum_{k=1}^K P(k) \cdot rel(k) \quad (2.3)$$

$$AR@K = \frac{1}{M} \sum_{k=1}^K R(k) \cdot rel(k) \quad (2.4)$$

$$MAP@K = \frac{1}{|U|} \sum_{k=1}^K U|(AP@k)_u \quad (2.5)$$

$$MAR@K = \frac{1}{|U|} \sum_{k=1}^K U|(AR@k)_u \quad (2.6)$$

เมื่อ $P(k) \cdot rel(k)$ แทนค่า Precision ที่ระบบแนะนำได้ตรงกับสิ่งที่ผู้ใช้เป้าหมายชอบ

$R(k) \cdot rel(k)$ แทนค่า Recall ที่ระบบแนะนำได้ตรงกับสิ่งที่ผู้ใช้เป้าหมายชอบ

M แทนจำนวนครั้งที่ระบบแนะนำได้ตรงกับสิ่งที่ผู้ใช้เป้าหมายชอบ



$AP@k$	แทนค่า Average Precision ของการแนะนำ k ลำดับ
$AR@k$	แทนค่า Average Recall ของการแนะนำ k ลำดับ
k	แทนจำนวนลำดับการแนะนำ
U	แทนจำนวนผู้ใช้งานทั้งหมด

ขอบเขตงานวิจัย

1. ศึกษาเฉพาะข้อมูลประวัติการเที่ยวของนักท่องเที่ยวต่างชาติที่เคยมาเที่ยวในกรุงเทพฯ และปริมณฑล และมีประวัติในเว็บไซต์ Tripadvisor (ที่มา: <https://www.tripadvisor.com/Tourism-g293916-Bangkok-Vacations.html> สืบค้นเมื่อ 16 พฤศจิกายน พ.ศ. 2563)
2. ข้อมูลที่นำมาศึกษาอยู่ในช่วงระหว่างเดือนตุลาคม พ.ศ. 2553 – มีนาคม พ.ศ. 2563 และศึกษาเฉพาะสถานที่ท่องเที่ยว 50 แห่งที่มีจำนวนการรีวิวสูงที่สุดในกรุงเทพฯ และปริมณฑล
3. “ระบบแนะนำ” ของงานวิจัยนี้หมายถึง อัลกอริทึมสำหรับแนะนำสถานที่ท่องเที่ยวเบื้องหลังของระบบแนะนำแบบสมบูร์ณบนเว็บไซต์ และแอปพลิเคชัน ตัวอย่างเช่น วิธีพิจารณาข้อมูลจากเนื้อหาของสิ่งของ (Content-based) หรือวิธีการพิจารณาการกรองร่วมกัน (Collaborative filtering)

วิธีดำเนินการวิจัย

การเก็บรวบรวมข้อมูล

ข้อมูลที่ใช้ในงานวิจัยนี้เป็นข้อมูลทุติยภูมิ (Secondary data) ที่นำมาจากการทำ Web Scraping ในเว็บไซต์ Tripadvisor ซึ่งเป็นเว็บไซต์รีวิวสถานที่ท่องเที่ยวที่มีความน่าเชื่อถือ และมีข้อมูลเกี่ยวกับสถานที่ท่องเที่ยวทั่วโลก ผู้วิจัยเลือกสถานที่ท่องเที่ยวในกรุงเทพฯ ที่มีการรีวิวมากที่สุด 50 อันดับแรกจากเว็บไซต์ดังกล่าวมาใช้ในงานวิจัยนี้ โดยเป็นข้อมูลที่อยู่ในช่วงระหว่างเดือนตุลาคม พ.ศ. 2553 – มีนาคม พ.ศ. 2563 ทั้งนี้ข้อมูลที่ได้จะมีทั้งหมด 2 ส่วน โดยมีรายละเอียดดังต่อไปนี้

1. ข้อมูลลักษณะ และประเภทของสถานที่ท่องเที่ยว ข้อมูลส่วนนี้เป็นข้อมูลเกี่ยวกับคุณลักษณะต่าง ๆ และประเภทของสถานที่ท่องเที่ยวในกรุงเทพฯ และปริมณฑลทั้ง 50 แห่ง ประกอบไปด้วย รหัสเฉพาะของสถานที่ท่องเที่ยว (Location ID), ชื่อสถานที่ท่องเที่ยว (Location), ประเภท และหมวดหมู่ของสถานที่ท่องเที่ยว (Type), และภาพรวมของสถานที่ท่องเที่ยว (Overview)
2. ข้อมูลประวัติการท่องเที่ยวของนักท่องเที่ยวต่างชาติที่เคยมาเที่ยวในประเทศไทย ข้อมูลส่วนนี้เป็นข้อมูลประวัติต่าง ๆ ของนักท่องเที่ยวต่างชาติที่เคยเข้ามาเที่ยวในสถานที่ท่องเที่ยวในกรุงเทพฯ และปริมณฑลทั้ง 50 แห่ง จากเว็บไซต์รีวิวสถานที่ท่องเที่ยว www.tripadvisor.com จากนั้นจะจัดเก็บข้อมูลที่ได้รับมาในรูปแบบไฟล์ประเภท Comma separated value (CSV) โดยข้อมูลที่นำมาใช้ในงานวิจัยนี้ประกอบไปด้วย เดือนที่ผู้ใช้เดินทางมาเที่ยว, ปีที่ผู้ใช้เดินทางมาเที่ยว, ชื่อสถานที่ท่องเที่ยวที่ไปเที่ยว, ชื่อผู้ใช้งาน, คะแนน



(Rating) ที่อยู่ระหว่างช่วง 1-5 คะแนน, รีวิวที่ผู้ใช้ให้กับสถานที่ท่องเที่ยว, หัวข้อการรีวิว, ประเภทการเดินทาง, ที่อยู่ของผู้ใช้งาน, ชื่อผู้เข้าใช้งาน (Username), และวันที่เขียนรีวิว จำนวนทั้งสิ้น 149,475 ธุรกรรม โดยเป็นข้อมูลจากผู้ใช้งานทั้งหมด 77,985 คน รายชื่อสถานที่ท่องเที่ยวที่ใช้ในการศึกษาครั้งนี้เรียงลำดับตามจำนวนการรีวิวแสดงดังตารางที่ 1

ตารางที่ 1 รายชื่อสถานที่ท่องเที่ยวในกรุงเทพฯ และปริมณฑลเรียงตามจำนวนการรีวิว

1. Wat Phra Chetuphon	18. Safari World	35. Wat Benchamabophit
2. Chatuchak weekend market	19. Platinum Fashion Mall	36. Museum of Contemporary Art
3. Temple of Dawn (Wat Arun)	20. Siam Niramit Show	37. IconSiam
4. The Grand Palace	21. SEA LIFE Bangkok Ocean World	38. Dream World
5. Jim Thompson house	22. Patpong Night Market	39. Wat Suthat
6. MBK Center	23. Erawan Shrine	40. The National Museum & Wang Na Palace
7. Temple of the Emerald Buddha	24. Madame Tussauds Bangkok	41. Art in Paradise Bangkok
8. KhaoSan Road	25. Baiyoke Sky Tower	42. Siriraj Medical Museum
9. ASIATIQUE The Riverfront	26. Sukhumvit	43. Siam AMAZING Park
10. Siam Paragon	27. Snake Farm (Queen Saovabha Memorial Institute)	44. Rajadamnern Thai Boxing Stadium
11. Temple of the Golden Buddha	28. Pratunam Market	45. Pantip Plaza
12. Lumpini park	29. Calypso Cabaret	46. Flow House Bangkok
13. Chinatown - Bangkok	30. Emporium Bangkok and	47. Benjasiri Park



(Yaowarat Road)	EmQuartier Bangkok	
14. Terminal 21	31. Soi Cowboy	48. King Power Mahanakhon
15. Central World	32. Taling Chan Floating Market	49. Samutprakan Crocodile Farm and Zoo
16. Damnoen Saduak Floating Market	33. Train Night Market Ratchada	50. Wat Bowonniwet Vihara
17. Chao Phraya River	34. Bangkok Art and Culture Center	

การจัดการข้อมูล

ในการสร้างระบบแนะนำจากอัลกอริทึมด้านการแนะนำต่าง ๆ จะต้องมีการจัดการกับข้อมูลเบื้องต้นให้เหมาะสมกับการนำมาใช้ ซึ่งข้อมูลบางส่วนที่ได้มาจากการทำ Web scraping นั้นยังไม่เหมาะสมจะนำมาสร้างระบบแนะนำ ผู้วิจัยจึงต้องมีการจัดการกับข้อมูลเพื่อให้สามารถนำไปสร้างระบบแนะนำแบบต่าง ๆ ในงานวิจัยครั้งนี้ได้ รวมถึงต้องมีการเตรียมข้อมูลเพื่อใช้สำหรับการทดสอบประสิทธิภาพของระบบแนะนำแบบออนไลน์ โดยมีรายละเอียดดังต่อไปนี้

1) ลบข้อมูลผู้ใช้ที่มีการให้คะแนนเพียงแค่สถานที่เดียวออก เพื่อลดการเกิดค่า Missing values ในเมทริกซ์ที่ใช้การสร้างระบบแนะนำ รวมถึงลดผลของ Sparse matrix ของ Collaborative filtering

2) ตรวจสอบข้อมูลที่ซ้ำกัน (Duplicate) และลบข้อมูลทั้งหมดที่ไม่มีชื่อผู้ใช้งานออก เนื่องจากเป็นข้อมูลที่ไม่สามารถระบุได้ว่าเป็นข้อมูลของใคร

3) นำข้อมูลที่ผ่านกระบวนการข้างต้นแล้วไปแบ่งเป็นข้อมูลสำหรับสร้างตัวแบบ (Training data) และข้อมูลสำหรับทดสอบ (Testing data) ในสัดส่วน 90% และ 10% ตามลำดับ โดยใช้ Library ของ Sklearn

ที่ชื่อว่า train_test_split ทำการสุ่มชื่อผู้ใช้งานให้ได้ตามสัดส่วนที่กำหนดก่อน จากนั้นทำการจัดระเบียบให้ข้อมูลที่อยู่ในชุดทดสอบให้เป็นข้อมูลสถานที่ท่องเที่ยวล่าสุดที่นักท่องเที่ยวแต่ละคนเดินทางไป ซึ่งเมื่อทำการแบ่งข้อมูลแล้วจะทำให้มีข้อมูล Training data จำนวนทั้งสิ้น 91,650 รายการ มีจำนวนผู้ใช้งานทั้งหมด 31,584 คน และข้อมูล Testing data จำนวน 10,006 รายการ มีผู้ใช้งานทั้งหมด 8,661 คน

ขั้นตอนการสร้างระบบแนะนำ

นำข้อมูลที่ผ่านการกระบวนการจัดการข้อมูลมาสร้างเป็นระบบแนะนำด้วยอัลกอริทึมด้านการแนะนำที่แนะนำตามรสนิยมของผู้ใช้ (Personalize recommendation algorithms) ซึ่งประกอบไปด้วยอัลกอริทึมดังต่อไปนี้



1.) Popularity based อัลกอริทึมนี้เป็นวิธีการที่เว็บไซต์รีวิวสถานที่ท่องเที่ยวนิยมใช้ในปัจจุบัน โดยจะเป็นการใช้ค่าเฉลี่ยถ่วงน้ำหนักระหว่างค่าคะแนนเฉลี่ยของสถานที่ท่องเที่ยวแต่ละแห่ง และจำนวนครั้งเฉลี่ยที่นักท่องเที่ยวเดินทางไปเที่ยวในสถานที่ที่เที่ยวดังกล่าวของข้อมูลการรีวิว ซึ่งระบบแนะนำที่สร้างจากอัลกอริทึมนี้มักจะแนะนำสถานที่ท่องเที่ยวยอดนิยมชุดเดิมให้กับผู้ใช้งานทุกคนตลอด ดังนั้นระบบแนะนำนี้จึงเป็นวิธีการที่ง่ายแต่ไม่มีประสิทธิภาพ ผู้วิจัยจึงใช้ Popularity based เป็นอัลกอริทึมตั้งต้น (Baseline) ในการเปรียบเทียบประสิทธิภาพของอัลกอริทึมอื่น ๆ โดยมีขั้นตอนดังต่อไปนี้

1.1 นำข้อมูลการให้คะแนน (Rating) จากส่วนที่เป็นประวัติการท่องเที่ยวของนักท่องเที่ยวต่างชาติมาหาจำนวนครั้งที่นักท่องเที่ยวมีการเดินทางไป (Count) และคะแนนเฉลี่ยที่นักท่องเที่ยวได้ให้กับสถานที่ท่องเที่ยว (Average rating) ของสถานที่ท่องเที่ยวแต่ละแห่ง จากนั้นคำนวณหาค่าเฉลี่ยของ Average rating ของทุกสถานที่ท่องเที่ยว (C)

1.2 คำนวณหา Percentile ที่ 50 ของค่าผลรวมจำนวนการให้คะแนนของสถานที่ท่องเที่ยวแต่ละแห่ง (m) จากนั้นจะลบสถานที่ท่องเที่ยวที่มีผลรวมการให้คะแนนต่ำกว่า m ออก เพื่อให้เหลือเพียงสถานที่ท่องเที่ยวที่มีจำนวนการเดินทางไปสูง ๆ

1.3 นำค่าทั้งหมดที่ได้คำนวณมาคำนวณกับด้วยสมการ Weighted rating (WR) ซึ่งเป็นสมการที่ใช้ถ่วงน้ำหนักระหว่างจำนวนครั้งสถานที่ท่องเที่ยวแต่ละแห่งมีนักท่องเที่ยวเดินทางไป และคะแนนเฉลี่ยที่สถานที่ท่องเที่ยวนั้นได้รับ โดยมีสมการที่ (3.1)

$$WR = \left(\frac{v}{(v+m)} \times R \right) + \left(\frac{m}{(v+m)} \times C \right) \quad (3.1)$$

โดยที่ V แทนจำนวนครั้งที่มีการคะแนนของแต่ละสถานที่

m แทน Percentile ที่ 50 ของผลรวมการให้คะแนนของแต่ละสถานที่

R แทนค่าเฉลี่ยของคะแนนของแต่ละสถานที่ C แทนค่าเฉลี่ยของค่าเฉลี่ยการให้คะแนนของทุกสถานที่

1.4 ผลลัพธ์ที่ได้จะออกมาในรูปของค่าคะแนน (Score) ของแต่ละสถานที่ ซึ่งเป็นค่าที่แสดงถึงความนิยมของสถานที่ท่องเที่ยว จากนั้นจะนำค่า Score มาเรียงลำดับจากมากไปน้อย และ Popularity based จะแนะนำสถานที่ท่องเที่ยวที่มีค่า Score สูงมากที่สุดให้กับผู้ใช้งานเป้าหมายก่อน

2) Content-based อัลกอริทึมนี้จะใช้ข้อมูลส่วนคุณลักษณะ และประเภทของสถานที่ท่องเที่ยวเป็นหลัก ซึ่งจำเป็นต้องใช้ข้อมูลสถานที่ท่องเที่ยวที่ผู้ใช้งานเป้าหมายเคยไปมาในอดีต แต่ไม่ต้องใช้ข้อมูลประวัติการให้คะแนนของผู้ใช้คนอื่นมาพิจารณาด้วย โดย Content-based จะแนะนำสถานที่ท่องเที่ยวที่มีลักษณะและประเภทใกล้เคียงกับสถานที่ท่องเที่ยวข้างต้นมากที่สุดให้ และมีขั้นตอนการทำได้ต่อไปนี้

2.1 นำข้อมูลเกี่ยวกับภาพรวมของสถานที่ท่องเที่ยว (Overview) กับคุณลักษณะ และประเภท (Type) ของสถานที่ท่องเที่ยวมาทำ preprocessing ด้วยการ ใช้ TF-IDF Vectorizer เพื่อสกัดเอาคำสำคัญ (Keyword) ที่อยู่ในข้อความทั้งหมดออกมา จากนั้นเก็บในรูปแบบของ Bag of words



2.2 ทำการสร้างเมทริกซ์ที่มีแถว (Row) เป็นรายชื่อสถานที่ท่องเที่ยวทั้ง 50 แห่ง และมีคอลัมน์ (Column) เป็นค่าทั้งหมดที่อยู่ใน Bag of words ซึ่งค่าในเมทริกซ์จะมีค่าเป็น 0 และ 1 โดยจะเป็น 1 ก็ต่อเมื่อในสถานที่ในแถวนั้นมีค่าที่อยู่ใน Bag of words ในคอลัมน์นั้น และเป็น 0 ถ้าไม่มี

2.3 นำเมทริกซ์ดังกล่าวมาหาค่าความเหมือนกันของสถานที่ท่องเที่ยวที่ผู้ใช้งานเคยไปมา กับสถานที่ท่องเที่ยวอื่น ๆ ทั้งหมด ด้วยการใช้ Cosine similarity ดังสมการที่ 3.2 ซึ่งเป็นเทคนิคที่ใช้ในการหาค่าความเหมือนของเวกเตอร์เป็นคู่ ในที่นี้เวกเตอร์ที่ใช้จะเป็นเวกเตอร์ของสถานที่ท่องเที่ยวที่ผู้ใช้เคยไป (A_i) เปรียบเทียบกับเวกเตอร์ของสถานที่ท่องเที่ยวอื่น ๆ ทั้งหมด (B_i) ค่าที่ได้ออกมาจะเป็นค่าความเหมือน (Similarity score) ของทุก ๆ สถานที่

$$SIM = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (3.2)$$

2.4 จากนั้นทำขั้นตอนที่ 3 ใหม่จนครบทุกสถานที่ที่ผู้ใช้งานดังกล่าวเคยให้คะแนนมา ถ้าหากผู้ใช้งานคนที่มีประวัติการท่องเที่ยวในสถานที่ท่องเที่ยวหลายแห่ง ให้นำค่า Similarity score ของการคำนวณแต่ละรอบมาหาค่าเฉลี่ยแล้วนำค่าเฉลี่ย Similarity score มาเรียงลำดับจากมากไปน้อย สุดท้าย Content-based จะแนะนำสถานที่ที่มีค่า Similarity score สูงที่สุดให้กับผู้ใช้งานเป้าหมาย

3) User-based Collaborative filtering และ Item-based Collaborative filtering อัลกอริทึมนี้จะใช้ข้อมูลประวัติการให้คะแนนสถานที่ท่องเที่ยวที่ผู้ใช้เคยไปมาในอดีต และข้อมูลของผู้ใช้อื่นมาพิจารณา ซึ่งในกรณีที่เป็น Item-based ตัวอัลกอริทึมจะใช้ K-Nearest neighbors ไปหาสถานที่ท่องเที่ยวที่มีลักษณะการถูกให้คะแนนคล้ายกับสถานที่ท่องเที่ยวที่ผู้ใช้เป้าหมายเคยไปมาที่สุดมาพิจารณาหาสถานที่ท่องเที่ยวที่ผู้ใช้เป้าหมายน่าจะชอบให้ ส่วนในกรณี User-based จะเป็นการหาผู้ใช้ที่มีลักษณะการให้คะแนนคล้ายกันมากที่สุดมาพิจารณาแทน โดยมีขั้นตอนการดำเนินงานต่อไปนี้

3.1 นำข้อมูลประวัติการท่องเที่ยวของผู้ใช้มาสร้าง Pivot table ที่มีแถว (Row) เป็นรายชื่อสถานที่ท่องเที่ยวทั้ง 50 แห่ง และมีคอลัมน์ (Column) เป็นชื่อผู้เข้าใช้งาน (Username) ทั้งหมด 77,985 คอลัมน์ โดยที่ค่าภายในจะเป็นคะแนน (Rating) ที่ผู้ใช้แต่ละคนให้กับสถานที่ท่องเที่ยวที่เคยไปเที่ยวมาทั้งหมด

3.2 ใช้ K-Nearest neighbors หาค่าความเหมือนกันของข้อมูล โดยในกรณีของ Item-based จะใช้เวกเตอร์การให้คะแนนของสถานที่ท่องเที่ยวที่ผู้ใช้งานเป้าหมายเคยให้คะแนนไว้สูงๆ จาก Pivot table ในขั้นตอนแรกมาหาค่าความเหมือนด้วย Pearson correlation กับสถานที่ท่องเที่ยวอื่น ๆ ทั้งหมด ส่วนในกรณีของ User-based จะใช้เวกเตอร์การให้คะแนนของผู้ใช้แต่ละคน ซึ่งมาจากการ Transpose Pivot table ในขั้นตอนที่ 1 มาหาค่าความเหมือนของข้อมูลด้วย Pearson correlation แทน ผลที่ได้จะเป็นค่า Similarity score ที่บอกว่าข้อมูลมีความเหมือนกันมากน้อยแค่ไหน หากค่า Similarity score มีค่าที่สูงจะหมายความว่าข้อมูลมีความเหมือนกันมาก

3.3 จากนั้นนำค่า Similarity score ที่ได้ไปคำนวณหาค่าคะแนนทำนาย (Predicted rating) ของสถานที่ท่องเที่ยวที่ผู้ใช้งานไม่เคยไปมาก่อน ซึ่ง Item-based จะคำนวณจากสถานที่ที่มีค่า Similarity



score สูง ส่วน User-based จะคำนวณจากผู้ใช้งานอื่นที่มีค่า Similarity score สูงแทน สุดท้ายนำผลลัพธ์ที่ได้มาเรียงลำดับจากมากไปหาน้อย และแนะนำสถานที่ที่มีค่า Predicted rating สูงมากที่สุดให้กับผู้ใช้เป้าหมาย

4) Singular value decomposition อัลกอริทึมนี้เป็นเป็นส่วนย่อยของ Collaborative filtering ซึ่งใช้หลักการของ Matrix factorization ในการหาค่า Predicted rating โดยมีขั้นตอนการดำเนินงานดังนี้

4.1 นำข้อมูลประวัติการท่องเที่ยวของผู้ใช้มาสร้าง Pivot table ที่มีแถว (Row) เป็นรายชื่อสถานที่ท่องเที่ยวทั้ง 50 แห่ง และมีคอลัมน์ (Column) เป็นชื่อผู้เข้าใช้งาน (Username) ทั้งหมด 77,985 คอลัมน์ โดยที่ค่าภายในจะเป็นคะแนน (Rating) ที่ผู้ใช้แต่ละคนให้กับสถานที่ท่องเที่ยวที่เคยไปเที่ยวมาทั้งหมด

4.2 ทำการแยกองค์ประกอบของ Pivot table ออกเป็น 3 ส่วน ด้วย Matrix factorization โดยที่ส่วนที่ 1 จะเป็นเมทริกซ์ที่แสดงถึงลักษณะของผู้ใช้งาน (U) เมทริกซ์ที่ 2 จะเป็นเมทริกซ์ที่แสดงถึงปัจจัยแฝง หรือ Latent factor ของวัตถุ (Σ) และเมทริกซ์ที่ 3 จะเป็นเมทริกซ์ที่แสดงคุณลักษณะของสถานที่ท่องเที่ยว (V^T)

4.3 หลังจากนั้นจะใช้ Stochastic Gradient Descent ในการปรับค่าภายในเมทริกซ์ทั้ง 3 ที่แยกออกมา จนกว่าจะทำให้เมื่อ Dot product เมทริกซ์ทั้ง 3 แล้ว เมทริกซ์ที่ออกมาจะมีค่าภายในใกล้เคียงกับเมทริกซ์ตั้งต้น ซึ่งก็คือ Pivot table ในขั้นตอนแรกมากที่สุด จากนั้นเมื่อได้เมทริกซ์ที่มีค่าใกล้เคียงกับเมทริกซ์ตั้งต้นที่สุดแล้ว ค่าในเมทริกซ์ใหม่ที่ได้จะมีค่าของ Rating ที่ผู้ใช้งานไม่ได้มีการให้มาก่อนด้วย ทำให้สามารถทราบค่า Predicted rating ของผู้ใช้งานทุกคนที่ให้กับสถานที่ท่องเที่ยวที่ไม่เคยไปได้ สุดท้าย SVD จะแนะนำสถานที่ท่องเที่ยวที่มีค่า Predicted rating สูงที่สุดให้กับผู้ใช้เป้าหมาย

5) Hybrid model อัลกอริทึมนี้จะรวมผลลัพธ์ของอัลกอริทึมด้านการแนะนำอื่นตั้งแต่ 2 อัลกอริทึมขึ้นไป ซึ่งก็คือการรวมค่า Similarity score หรือ Predicted rating ของระบบแนะนำอื่น ๆ เข้าด้วยกัน โดย Hybrid model สามารถให้ผลลัพธ์ที่แตกต่างไปตามประเภทของข้อมูล, ลักษณะของสิ่งที่จะแนะนำ และวิธีการรวมผลลัพธ์ ในการวิจัยครั้งนี้ผู้วิจัยทำการรวมผลลัพธ์ของ Content-based, User-based collaborative filtering, Item-based collaborative filtering, และ Singular value decomposition ทั้งหมด 4 อัลกอริทึมด้วยวิธีการเฉลี่ยถ่วงน้ำหนัก (Weighted) ค่าที่ได้จากการรวมผลจะเป็นผลลัพธ์สุดท้าย (Final Score) ที่ใช้ในการพิจารณาสถานที่ท่องเที่ยวที่จะแนะนำ และจะแนะนำสถานที่ท่องเที่ยวที่มี Final Score สูงที่สุด สำหรับการรวมผลลัพธ์ผู้วิจัยจะทำการหาค่าน้ำหนักที่เหมาะสม (Optimization) ด้วยการใช้ Gradient descent โดยจะเริ่มจากการกำหนดค่าน้ำหนักของทุกอัลกอริทึมให้เท่ากันก่อน จากนั้นจะนำค่าน้ำหนักดังกล่าวไปสร้างระบบแนะนำแบบ Hybrid model แล้วนำไปทดสอบประสิทธิภาพด้วยค่า Mean average precision (MAP) และ Mean average recall (MAR) กับข้อมูลใน Testing set แล้วจะทำการปรับค่าน้ำหนักของแต่ละอัลกอริทึมใหม่โดยใช้ Gradient descent ทำการปรับค่าจนกว่าค่า MAP และ MAR ของ Hybrid model สูงมากที่สุด และไม่มีเปลี่ยนแปลง

การทดสอบประสิทธิภาพของระบบแนะนำ



หลังจากพัฒนาระบบแนะนำสถานที่ท่องเที่ยวด้วยอัลกอริทึมทั้ง 5 แบบข้างต้นแล้ว ผู้วิจัยจะวัดประสิทธิภาพของระบบที่พัฒนามาข้างต้นด้วยค่า MAP และ MAR ซึ่งสามารถคำนวณได้จากการหาค่าเฉลี่ยของ Average precision (AP) และ Average recall (AR) โดยที่ค่า AP และ AR สามารถหาได้จากการหาค่าเฉลี่ยของ Precision และ Recall ของระบบ ทั้งนี้จะพิจารณาเฉพาะลำดับที่มีการแนะนำได้ตรงกับข้อมูลใน Testing set เท่านั้น ดังนั้นถ้าหากระบบแนะนำสามารถแนะนำสถานที่ท่องเที่ยวได้ตรงกับข้อมูลใน Testing set ตั้งแต่การแนะนำลำดับต้น ๆ จะทำให้ค่า AP และ AR รวมถึงค่า MAP และ MAR มีค่าสูง ซึ่งค่า MAP และ MAR เป็นค่าที่บอกว่าโดยเฉลี่ยแล้วประสิทธิภาพของระบบแนะนำดีแค่ไหน

ผลการวิจัย

ผลลัพธ์ของการทดสอบ และเปรียบเทียบประสิทธิภาพระบบแนะนำที่สร้างจากอัลกอริทึมสำหรับแนะนำแต่ละแบบแสดงในตารางที่ 2 ซึ่งแสดงความแม่นยำในการแนะนำสถานที่ท่องเที่ยวของอัลกอริทึมแบบต่าง ๆ จากค่า MAP และ MAR โดยเมื่อพิจารณาจากความแม่นยำในการแนะนำ 10 อันดับ พบว่า

1. ระบบแนะนำแบบ Hybrid model เป็นระบบแนะนำที่มีประสิทธิภาพในการแนะนำสถานที่ท่องเที่ยวในกรุงเทพฯ และปริมณฑลให้กับนักท่องเที่ยวต่างชาติมากที่สุด โดยมีค่า MAP@10 เท่ากับ 0.7282 หมายความว่าในการแนะนำ 10 อันดับแรก โดยเฉลี่ยอัลกอริทึมสามารถแนะนำได้ตรงกับสถานที่ที่ผู้ใช้เป้าหมายจะไปในอนาคตตาม Testing data ได้ 72.87% และ MAR@10 เท่ากับ 0.9659 หมายความว่า ในบรรดาสถานที่ท่องเที่ยวที่อยู่ใน Testing data ทั้งหมด โดยเฉลี่ยอัลกอริทึมสามารถแนะนำได้ตรง 96.59% ซึ่งค่าทั้งสองสูงมากที่สุดในบรรดาอัลกอริทึมทั้งหมด ทั้งนี้อาจจะเป็นเพราะ ระบบแนะนำแบบ Hybrid model ได้มีการเลือกส่วนของแต่ละอัลกอริทึมมารวมกันด้วยการหาค่าน้ำหนักที่เหมาะสมมากที่สุดจาก Gradient descent ทำให้โอกาสในการแนะนำผิดพลาด

2. ระบบแนะนำแบบ Content-based เป็นระบบแนะนำที่มีประสิทธิภาพรองลงมา โดยมีค่า MAP@10 เท่ากับ 0.6991 หมายความว่าในการแนะนำ 10 อันดับแรก โดยเฉลี่ยอัลกอริทึมสามารถแนะนำได้ตรงกับสถานที่ที่ผู้ใช้เป้าหมายจะไปในอนาคตตาม Testing data ได้ 69.91% และมีค่า MAR@10 เท่ากับ 0.9654 หมายความว่าในบรรดาสถานที่ท่องเที่ยวที่อยู่ใน Testing data ทั้งหมด โดยเฉลี่ยอัลกอริทึมสามารถแนะนำได้ตรง 96.54% ทั้งนี้อาจเป็นเพราะ จำนวนสิ่งที่จะแนะนำมีจำนวนไม่มาก และสถานที่ท่องเที่ยวในกรุงเทพฯ และปริมณฑลไม่มีความหลากหลายมากนัก จึงทำให้อัลกอริทึมสามารถจับรูปแบบการแนะนำโดยดูจากประวัติการเที่ยวของผู้ใช้เป้าหมายได้ดี

3. ระบบแนะนำแบบ Collaborative filtering ทั้ง 3 แบบไม่ว่าจะเป็น User-based, Item-based, และ SVD ให้ผลลัพธ์ไม่แตกต่างกันมากนัก โดยมีค่า MAP@10 เท่ากับ 0.1108, 0.0938, 0.1020 และ MAR@10 เท่ากับ 0.2854, 0.2427, 0.1513 ตามลำดับ แสดงให้เห็นว่าระบบแนะนำแบบ Collaborative filtering



ทุกรูปแบบให้ผลลัพธ์ที่ไม่ดี อาจเป็นเพราะ ข้อมูลการท่องเที่ยวของนักท่องเที่ยวต่างชาตินั้นมีไม่มากพอ ทำให้เมทริกซ์ที่ใช้ในอัลกอริทึมมีค่าว่างมาก อัลกอริทึมจึงไม่สามารถหาค่าคะแนนทำนายที่ดีได้

ทั้งนี้เมื่อพิจารณาค่า MAP และMAR ของทุกระบบเทียบกับ Popularity based ซึ่งเป็น Baseline

Model	Popularity based		Content based		User-based CF		Item-based CF		Singular value decomposition (SVD)		Hybrid model	
	MAP	MAR	MAP	MAR	MAP	MAR	MAP	MAR	MAP	MAR	MAP	MAR
1	0.088	0.079	0.591	0.551	0.058	0.052	0.050	0.045	0.088	0.080	0.591	0.557
2	0.128	0.152	0.630	0.622	0.073	0.077	0.061	0.064	0.091	0.084	0.677	0.724
3	<u>0.156</u>	<u>0.231</u>	<u>0.670</u>	<u>0.741</u>	<u>0.080</u>	<u>0.097</u>	<u>0.068</u>	<u>0.088</u>	<u>0.092</u>	<u>0.086</u>	<u>0.705</u>	<u>0.807</u>
4	0.165	0.263	0.675	0.771	0.087	0.120	0.073	0.101	0.092	0.087	0.718	0.865
5	0.168	0.276	0.688	0.849	0.092	0.146	0.078	0.122	0.093	0.092	0.723	0.895
6	0.176	0.321	0.693	0.880	0.097	0.175	0.081	0.144	0.096	0.106	0.726	0.919
7	0.183	0.371	0.702	0.958	0.101	0.202	0.085	0.167	0.099	0.129	0.731	0.958
8	0.193	0.448	0.700	0.963	0.105	0.229	0.089	0.195	0.100	0.134	0.729	0.964
9	0.196	0.477	0.700	0.965	0.108	0.255	0.091	0.221	0.101	0.139	0.728	0.965
10	0.199	0.500	0.699	0.965	0.111	0.285	0.094	0.243	0.102	0.151	0.728	0.966

ของการวิจัยในครั้งนี้ จะเห็นได้ว่าทั้ง Hybrid model และContent-based มีผลลัพธ์ที่ดีกว่าอย่างเห็นได้ชัด ส่วน Collaborative filtering ทั้ง 3 แบบให้ผลลัพธ์ที่ไม่แตกต่างกับ Popularity based อีกทั้งผลลัพธ์ที่ได้ยังแสดงให้เห็นว่าค่า MAP และMAR ของลำดับการแนะนำทุกตัวจะมีการเปลี่ยนแปลงไปน้อยมากตั้งแต่การแนะนำ 3 ลำดับขึ้นไป ดังนั้น ลำดับในการแนะนำที่เหมาะสมสำหรับแนะนำสถานที่ท่องเที่ยวในกรุงเทพฯและปริมณฑลควรแนะนำเพียงแค่ 3 ลำดับแรก เนื่องจากถึงจะเพิ่มลำดับในการแนะนำก็ไม่ทำให้ประสิทธิภาพการแนะนำดีขึ้น และเป็นการเปลืองพื้นที่ในการแนะนำบนหน้าเว็บไซต์ หรือแอปพลิเคชัน รวมถึงช่วยลดการขั้นตอนการแสดงผลของอัลกอริทึม ทำให้สามารถทำงานได้เร็วขึ้น

ตารางที่ 2 ค่า MAP และMAR ของระบบแนะนำที่สร้างจากอัลกอริทึมด้านการแนะนำในงานวิจัยนี้

สรุปการวิจัยและอภิปรายผล

การศึกษาในครั้งนี้มีเป้าหมายในการพัฒนาระบบแนะนำซึ่งเป็นอัลกอริทึมเบื้องหลังที่ใช้ในการแนะนำสถานที่ท่องเที่ยวในกรุงเทพฯ และปริมณฑลให้กับนักท่องเที่ยวต่างชาติได้มีประสิทธิภาพมากที่สุด โดยผู้วิจัยได้เปรียบเทียบประสิทธิภาพการแนะนำของระบบแนะนำที่พัฒนามาจากอัลกอริทึมด้านการแนะนำต่าง ๆ ได้แก่ (1) Popularity based (2) Content-based (3) Collaborative filtering (4) Hybrid model ด้วยค่า MAP และ MAR ของระบบแนะนำแต่ละแบบ ผลการศึกษาสามารถพิจารณาได้ 2 ประเด็นสำคัญ ดังต่อไปนี้



1. ด้านประสิทธิภาพของระบบแนะนำแบบเดี่ยว พบว่า ระบบแนะนำที่สร้างจากอัลกอริทึม Content-based มีประสิทธิภาพมากที่สุดในการทำงานแบบเดี่ยว เนื่องจากวิธีการนี้จะนำข้อมูลลักษณะ และประเภทของสถานที่ท่องเที่ยวในอดีตของผู้ใช้เป้าหมายที่มีการให้คะแนนสูง ๆ ไปหาสถานที่ท่องเที่ยวที่มีคุณลักษณะ และประเภทคล้ายกับสถานที่ดังกล่าว ซึ่งลักษณะการท่องเที่ยวของนักท่องเที่ยวต่างชาติส่วนใหญ่จะเที่ยวในสถานที่ที่มีลักษณะใกล้เคียงกัน รวมถึงสถานที่ท่องเที่ยวในกรุงเทพฯ และปริมณฑลมีรูปแบบที่จำกัด จึงทำให้ Content-based มีประสิทธิภาพมาก ในทางตรงกันข้ามระบบแนะนำที่ใช้ข้อมูลสถานที่ท่องเที่ยวอื่นหรือผู้ใช้งานคนอื่นมาพิจารณาด้วยอย่าง Collaborative filtering กลับให้ผลลัพธ์ที่ไม่ดี เนื่องมาจากการที่นักท่องเที่ยวต่างชาติไม่ได้เดินทางไปเที่ยวต่างประเทศบ่อยมากนัก อีกทั้งเมื่อนักท่องเที่ยวต่างชาติเดินทางมาเที่ยวในกรุงเทพฯ และปริมณฑลจะมีเวลาในการท่องเที่ยวที่จำกัดทำให้ไม่สามารถเที่ยวในสถานที่ท่องเที่ยวได้หลากหลาย ทำให้เมตริกซ์ที่ใช้ในการคำนวณคะแนนทำนาย (Predicted rating) ของ Collaborative filtering มีข้อมูล Missing values ค่อนข้างมาก โดยเฉพาะสถานที่ท่องเที่ยวบางแห่งที่เพิ่งเปิด หรือยังไม่เป็นที่รู้จักมากนักในกลุ่มนักท่องเที่ยวต่างชาติ ทำให้มีข้อมูลที่นำไปให้อัลกอริทึมเรียนรู้ได้น้อย ส่งผลให้ระบบแนะนำสถานที่ท่องเที่ยวที่พัฒนาจากอัลกอริทึม Collaborative filtering ให้ผลลัพธ์ที่ไม่ดี

2. ด้านระบบแนะนำที่มีประสิทธิภาพที่สุด พบว่า ระบบแนะนำที่มีประสิทธิภาพมากที่สุดในการแนะนำสถานที่ท่องเที่ยวในงานวิจัยครั้งนี้ คือ ระบบแนะนำแบบ Hybrid model ที่รวมผลลัพธ์ด้วยวิธีการเฉลี่ยถ่วงน้ำหนัก (Weighted) เนื่องจากวิธีการนี้เป็นการรวมผลลัพธ์ค่า Similarity scores หรือคะแนนทำนาย (Predicted rating) ของระบบแนะนำที่สร้างจากอัลกอริทึม Content-based, User-based, Item-based, และ SVD ค่าที่ได้ออกมาจะเป็น Final score ซึ่งเป็นค่าที่ Hybrid model ใช้ในการพิจารณาสถานที่ท่องเที่ยวที่จะแนะนำให้กับผู้ใช้เป้าหมาย โดยจากการรวมผลลัพธ์ของ Content-based ทำให้ระบบแบบ Hybrid model สามารถแนะนำสถานที่ท่องเที่ยวได้ตรงกับคุณลักษณะของสถานที่ที่นักท่องเที่ยวต่างชาติแต่ละคนชอบขึ้นชอบ ทั้งยังช่วยลดปัญหาการมีข้อมูลน้อยที่เป็นปัญหาของ Collaborative filtering รวมถึงยังสามารถแนะนำสถานที่ท่องเที่ยวได้แปลกใหม่ และครอบคลุมมากกว่า Content-based เพราะ มีการนำ Collaborative filtering ทั้ง 3 แบบ ที่ใช้ข้อมูลของผู้ใช้งานอื่นมาพิจารณาด้วย ซึ่งผลดังกล่าวสอดคล้องกับงานวิจัยของ Badaro, Hajj, El-Hajj, & Nachman (2013) ซึ่งได้ทำการศึกษาระบบแนะนำแบบ Hybrid model ด้วยการรวมอัลกอริทึม User-based และ Item-based เข้าด้วยกัน ผลการศึกษาพบว่าการรวมผลลัพธ์ของทั้งสองอัลกอริทึมช่วยให้ความแม่นยำของการแนะนำเพิ่มสูงขึ้น อีกทั้งยังสอดคล้องกับงานวิจัยของ Li et al. (2019) ที่ได้ศึกษาระบบแนะนำสถานที่ท่องเที่ยวในประเทศจีนด้วยอัลกอริทึมด้านการแนะนำแบบต่าง ๆ โดยใช้ข้อมูลแบบสอบถามออนไลน์จำนวน 1000 ตัวอย่าง ผลการศึกษาพบว่าการใช้ Hybrid model ให้ผลลัพธ์การแนะนำที่ดีที่สุดเมื่อเปรียบเทียบกับการทำงานแบบเดี่ยว

3. ด้านการเลือกระบบแนะนำ พบว่า ในกรณีที่ต้องการสร้างระบบแนะนำสถานที่ท่องเที่ยวที่สมบูรณ์บนเว็บไซต์ หรือแอปพลิเคชัน ถ้าหากมีการเก็บข้อมูลประวัติการรีวิว หรือคะแนนความพึงพอใจในการท่องเที่ยวของนักท่องเที่ยวต่างชาติในปริมาณมาก การเลือกระบบแนะนำแบบ Hybrid model จะให้ผลลัพธ์ที่



ดีที่สุด เนื่องจากเมื่อมีข้อมูลให้อัลกอริทึมทำการเรียนรู้เยอะ ตัว Hybrid model จะทำการสกัดหารูปแบบของข้อมูลในการแนะนำผู้ใช้แต่ละคนได้ดี และให้ผลลัพธ์ที่แม่นยำ แต่ถ้าหากไม่มีข้อมูลประวัติการท่องเที่ยว หรือข้อมูลมีจำนวนน้อย ระบบแนะนำที่เหมาะสมที่สุดจะเป็น Content-based เนื่องจาก Content-based เป็นระบบแนะนำที่อ้างอิงจากระบบนิยมของผู้ใช้เป้าหมายเป็นหลัก ไม่จำเป็นต้องใช้ข้อมูลของผู้ใช้อื่นมาพิจารณา ทำให้ไม่ต้องใช้ข้อมูลในปริมาณมาก และไม่จำเป็นต้องใช้ข้อมูลการให้คะแนนสถานที่ท่องเที่ยว (Rating) ซึ่งถ้าเป็น Hybrid model จะเป็นจะต้องมีข้อมูลในส่วนนี้ถึงจะใช้ได้

องค์ความรู้ที่ได้จากการวิจัย

องค์ความรู้ที่ได้จากการศึกษาระบบแนะนำสถานที่ท่องเที่ยวสำหรับนักท่องเที่ยวต่างชาติ มีส่วนช่วยให้หน่วยงานที่เกี่ยวข้องด้านการท่องเที่ยวสามารถพิจารณาเลือกอัลกอริทึมที่จะใช้ในการแนะนำสถานที่ท่องเที่ยวบนเว็บไซต์หรือแอปพลิเคชันได้ดียิ่งขึ้น โดยพิจารณาจากปริมาณข้อมูลที่มี ถ้าหากมีข้อมูลน้อยการเลือก Content-based จะดีที่สุด แต่หากมีปริมาณข้อมูลเยอะการใช้ Hybrid model จะให้ประสิทธิภาพที่ดีที่สุด รวมถึงลำดับการแนะนำบนเว็บไซต์ควรแนะนำเพียง 3 ลำดับแรกในกรณีที่พื้นที่การแสดงผลมีจำกัด เนื่องจาก เป็นลำดับการแนะนำที่มีประสิทธิภาพสูงไม่แตกต่างกับการแนะนำทุกลำดับ

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้ประโยชน์

สำหรับการวิจัยในครั้งนี้ผลลัพธ์ที่ได้จะเป็นอัลกอริทึมเบื้องหลังของระบบแนะนำแบบสมบุรณ์ (Recommendation system) ซึ่งหน่วยงานที่เกี่ยวข้องกับด้านการท่องเที่ยวไม่ว่าจะเป็น การท่องเที่ยวแห่งประเทศไทย, กระทรวงการท่องเที่ยว และกีฬา, การท่องเที่ยวจังหวัด รวมถึงภาคเอกชนที่ต้องการสร้างเว็บไซต์เกี่ยวกับสถานที่ท่องเที่ยวในกรุงเทพฯ และปริมณฑลอาจนำไปใช้เพื่อเป็นแนวทางในการพิจารณาเพื่อเลือกอัลกอริทึม และสร้างเป็นระบบแนะนำแบบสมบุรณ์ที่สามารถแนะนำสถานที่ท่องเที่ยวในกรุงเทพฯ และปริมณฑลให้กับผู้ใช้งานบนเว็บไซต์ หรือแอปพลิเคชันของหน่วยงานดังกล่าวได้อย่างเหมาะสม และมีประสิทธิภาพที่สุด โดยมีขั้นตอนดังนี้

1. เลือกอัลกอริทึมที่เหมาะสมกับข้อมูลที่มี โดยถ้าหากมีข้อมูลมากอัลกอริทึมแบบ Hybrid model จะให้ผลลัพธ์ที่ดี แต่ถ้าหากมีข้อมูลน้อยอัลกอริทึมแบบ Content-based จะให้ผลลัพธ์ที่ดีกว่า
2. นำข้อมูลประวัติการท่องเที่ยวของนักท่องเที่ยวมาให้อัลกอริทึมทำการเรียนรู้ ซึ่งถ้าหากหน่วยงานใดยังไม่มีเก็บข้อมูลความพึงพอใจ หรือการให้คะแนนของนักท่องเที่ยวให้ทำการเริ่มเก็บข้อมูลดังกล่าวก่อน
3. หลังจากอัลกอริทึมทำการเรียนรู้เสร็จให้นำโค้ดส่วนที่เสร็จแล้วไป Deploy ติดบนเว็บไซต์ หรือแอปพลิเคชันของหน่วยงานนั้น เมื่อมีนักท่องเที่ยวเข้ามาสมัครสมาชิก และใช้งานเว็บไซต์ดังกล่าว ตัวระบบนำที่ติดไว้จะทำการแนะนำสถานที่ท่องเที่ยวตามรสนิยมส่วนบุคคลของผู้ใช้แต่ละคน ทั้งนี้หน่วยงานที่มี



เว็บไซต์ที่สามารถแนะนำสถานที่ท่องเที่ยวได้อย่างมีประสิทธิภาพจะช่วยเพิ่มความสะดวกรวดสบาย และลดเวลาในการวางแผนการท่องเที่ยวให้กับนักท่องเที่ยวที่ต้องการมาเที่ยวในกรุงเทพฯ และปริมณฑล อีกทั้งยังสามารถแนะนำสถานที่ท่องเที่ยวได้หลากหลาย และตรงกับรสนิยมส่วนบุคคลของนักท่องเที่ยวแต่ละคนมากกว่าการแนะนำของเว็บไซต์ที่รวบรวมสถานที่ท่องเที่ยวในปัจจุบันที่แนะนำเพียงสถานที่ท่องเที่ยวยอดนิยมเท่านั้น

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

การวิจัยครั้งนี้เก็บรวบรวม และใช้ข้อมูลประวัติการเที่ยวของนักท่องเที่ยวต่างชาติจากเว็บไซต์ Tripadvisor ซึ่งเก็บมาเฉพาะข้อมูลของสถานที่ท่องเที่ยวในกรุงเทพฯ และปริมณฑลเท่านั้น ทำให้จำนวนข้อมูล และประเภทของสถานที่ท่องเที่ยวมีไม่มากนัก ดังนั้นสำหรับงานวิจัยในอนาคต สามารถเพิ่มขอบเขตการเก็บรวบรวมข้อมูลให้กว้างมากขึ้นด้วยการรวบรวมข้อมูลของจังหวัดอื่น ๆ นอกจากกรุงเทพฯ และปริมณฑล มาพิจารณาเพิ่มจะทำให้ประเภทของสถานที่ท่องเที่ยวมีความหลากหลายมากขึ้น และทำให้สามารถแนะนำสถานที่ท่องเที่ยวได้ละเอียดขึ้นอีกด้วย อีกทั้งเมื่อปริมาณข้อมูลมีจำนวนเพิ่มมากขึ้นจะทำให้สามารถใช้เทคนิคในการสร้างระบบแนะนำแบบอื่นที่ไม่สามารถใช้ในงานวิจัยนี้ได้ อันเนื่องมาจากขนาดของข้อมูลที่ไม่เพียงพอ เช่น วิธีการเรียนรู้แบบอัตโนมัติด้วยการเลียนแบบการทำงานของโครงข่ายประสาทของมนุษย์ (Deep learning) ด้วย Neural network แบบต่าง ๆ ที่เป็นวิธีการที่เหมาะสมกับข้อมูลขนาดใหญ่ รวมถึงการที่มีข้อมูลมากขึ้นยังทำให้สามารถใช้อัลกอริทึมสำหรับแบ่งกลุ่มข้อมูล (Clustering algorithm) มาแบ่งกลุ่มนักท่องเที่ยวต่างชาติตามรสนิยมการท่องเที่ยว ก่อนที่จะนำไปศึกษา และเปรียบเทียบประสิทธิภาพของระบบแนะนำแบบต่าง ๆ ซึ่งจะทำให้สามารถพัฒนาระบบแนะนำที่เหมาะสมกับรสนิยมของนักท่องเที่ยวแต่ละกลุ่มได้

เอกสารอ้างอิง

- ณิชาภัทร สุวัฒน์นนท์. (2562). การท่องเที่ยวกับบทบาทขับเคลื่อนเศรษฐกิจไทย ฮีโร่จำเป็นหรือ ฮีโร่ตัวจริง?. สืบค้นเมื่อ 17 สิงหาคม 2563, จาก https://www.bot.or.th/Thai/ResearchAndPublications/articles/Pages/Article_29Oct2019.aspx
- นัชชา ธีระสาโรช และวิโรจน์ อรุณมานะกุล. (2560). “การศึกษาการแยกนัยความหมายของ หัว ในภาษาไทย โดยใช้วิธีการวิเคราะห์ความหมายแบบแฝง,” (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ) คณะอักษรศาสตร์ มหาวิทยาลัยจุฬาลงกรณ์, กรุงเทพฯ
- พชรพจน์ นันทรามาศ, กิตติพงษ์ เรือนทิพย์, และจากรุวรรณ เหล่าสัมฤทธิ์. (2563). เจาะพฤติกรรมท่องเที่ยวใน New Normal: เมื่อโควิดทำชีวิตเปลี่ยน. สืบค้น 24 สิงหาคม 2563, จาก http://krungthai.com/download/economyresources/EconomyResourcesDownload_450เจาะพฤติกรรมท่องเที่ยวใน New Normal:เมื่อโควิดทำชีวิตเปลี่ยน_31_08_63.pdf
- Bruke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Model User-Adap Inter*, 12(4), 331-370
- Badaro, G., Hajj, H., El-Hajj, W., Nachman, L. (2013). A Hybrid Approach with Collaborative



- Filtering for Recommender Systems. 2013 9th International Wireless Communications and Mobile Computing Conference, IWCMC 2013. 349-354.
10.1109/IWCMC.2013.6583584.
- Chujai, P., Singthongchai, J., Yasaga, S., Surattgara, N., and Buranakutti, K. (2020). The Tourist Attractions Recommender System for Bangkok Thailand. *Recommender System for Bangkok Thailand, International Journal of Computer Theory and Engineering*, 12(1), 22-27
- Das, D., Sahoo, L., and Datta, S. (2017). A Survey on Recommendation System. *International Journal of Computer Applications*, 160(7), 6-10.
- Knoema. (2020). *Thailand Contribution Of Travel And Tourism To GDP (% Of GDP), 1995-2020*. Retrieved August 17, 2020, from <https://knoema.com/atlas/Thailand/topics/Tourism/Travel-and-Tourism-Total-Contribution-to-GDP/Contribution-of-travel-and-tourism-to-GDP-percent-of-GDP>
- Li, G., Hua, J., Yuan, T., Wu, J., Jiang, Z., Zhang, H., and Li, T. (2019). Novel Recommendation System for Tourist Spots Based on Hierarchical Sampling Statistics and SVD++. *Mathematical Problems in Engineering*, 19, 1-15.
- Pourheidari, V., Mollashahi, E. S., Vassileva J., and Deters, R. (2018). "Recommender System based on Extracted Data from Different Social Media. A Study of Twitter and LinkedIn. 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, 215-222.
- Sawtelle, S. (2020). *Mean Average Precision (MAP) For Recommender Systems*. Retrieved August 24, 2020, from <https://sdsawtelle.github.io/blog/output/mean-average-precision-MAP-for-recommender-systems.html>
- Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2001). Incremental singular value decomposition algorithms for highly scalable recommender systems. *Proceedings of the 10th international conference on World Wide Web*, 285-295.
- Shani, G., Gunawardana A. (2011) Evaluating Recommendation Systems. In: Ricci F., Rokach L., Shapira B., Kantor P. (eds) *Recommender Systems Handbook*. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-85820-3_8
- Taifi, M. (2020). *MRR vs MAP NDCG: Rank-Aware Evaluation Metrics and When to Use Them*. Retrieved August 17, 2020, from <https://medium.com/swlh/rank-aware-recsys-evaluation-metrics-5191bba16832>



Translate Thai References

- Nantaramat, P., Rueanthip, K. & Laosamrit, J. (2020). *Penetrating travel behavior in New Normal: when COVID changes life*. Retrieved August 24, 2020, from https://krungthai.com/Download/economyresources/EconomyResourcesDownload_450เจาะพฤติกรรมท่องเที่ยวใน_New_Normal_เมื่อโควิดทำชีวิตเปลี่ยน_31_08_63.pdf
- Surawattananon, N. (2019). *Tourism and its role in driving the Thai economy. Is it a necessary hero or a real hero?*. Retrieved August 17, 2020, from https://www.bot.or.th/ThaiResearchAndPublications/articles/Pages/Article_29Oct2019.aspx
- Tirasaroj, N. & Arunmanakul, V. (2017). *A Study of Word Sense Discrimination of /Hua4/ in Thai Using Latent Semantic Analysis*. (Doctoral dissertation) Faculty of Arts, Chulalongkorn University, Bangkok

