



ระบบแนะนำสินค้าอาหารโดยใช้ระบบแนะนำแบบผสมผสาน

FOOD RECOMMENDATION SYSTEM USING A HYBRID RECOMMENDATION METHOD



นิภาภรณ์ พันธนาม

บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

2563

ระบบแนะนำสินค้าอาหารโดยใช้ระบบแนะนำแบบผสมผสาน



สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2563
ลิขสิทธิ์ของมหาวิทยาลัยศรีนครินทรวิโรฒ

FOOD RECOMMENDATION SYSTEM USING A HYBRID RECOMMENDATION METHOD



NIPAPORN PANNAM

A Master's Project Submitted in Partial Fulfillment of the Requirements

for the Degree of MASTER OF SCIENCE

(Data Science)

Faculty of Science, Srinakharinwirot University

2020

Copyright of Srinakharinwirot University

สารนิพนธ์

เรื่อง

ระบบแนะนำสินค้าอาหารโดยใช้ระบบแนะนำแบบผสมผสาน

ของ

นิภาภรณ์ พันธุ์นาม

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)

คณบดีบัณฑิตวิทยาลัย

คณะกรรมการสอบปากเปล่าสารนิพนธ์

ที่ปรึกษาหลัก

ประธาน

(ผู้ช่วยศาสตราจารย์ ดร.วราภรณ์ วิทยานนท์)

(ผู้ช่วยศาสตราจารย์ ดร.อัศรา ประโยชน์)

กรรมการ

(อาจารย์ ดร.ศิริสรรพ เหล่าหะเกียรติ)

ชื่อเรื่อง	ระบบแนะนำสินค้าอาหารโดยใช้ระบบแนะนำแบบผสมผสาน
ผู้วิจัย	นิภาภรณ์ พันธุ์นาม
ปริญญา	วิทยาศาสตร์มหาบัณฑิต
ปีการศึกษา	2563
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. วราภรณ์ วิทยานนท์

การเพิ่มยอดขายและนำเสนอสินค้าให้ตรงตามความต้องการของผู้บริโภคเป็นช่องทางที่ทำให้เพิ่มรายได้ให้กับผู้ประกอบการและเพิ่มความพึงพอใจให้กับผู้บริโภคที่มีต่อสินค้ามากยิ่งขึ้น วัตถุประสงค์ของงานวิจัยนี้คือการพัฒนาแบบจำลองการแนะนำสินค้าอาหารมาช่วยในการแนะนำสินค้า โดยใช้เทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering) ซึ่งใช้หลักการ ความคล้ายคลึงแบบโคไซน์ และเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) โดยนำโลภาริชของ Surprise มาสร้างแบบจำลองโดยใช้ อัลกอริทึม ได้แก่ SVD, NMF, Baseline และ KNN การประเมินผลความแม่นยำของแบบจำลองนั้นดำเนินการโดยเปรียบเทียบค่า RMSE และ MAE ผลการทดลองพบว่าค่า RMSE และ MAE ที่มีค่าต่ำสุดและให้ผลลัพธ์ที่ดีที่สุดคืออัลกอริทึม SVD ที่ได้ค่า RMSE เท่ากับ 1.2528 และ MAE เท่ากับ 0.9376 ในส่วนของการทำงานแบบผสมผสานได้นำเทคนิคการกรองแบบอิงเนื้อหา และ เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมมาทำงานร่วมกัน ซึ่งใช้ข้อดีของทั้ง 2 เทคนิคนั้นคือนำแบบจำลองของเทคนิคการกรองแบบอิงเนื้อหา ที่มีการทำความเข้าใจของโคไซน์และแบบจำลองของเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมที่ใช้อัลกอริทึม SVD มาทำงานร่วมกันทำให้เกิดการแนะนำรายการสินค้าที่เหมาะสม

คำสำคัญ : ระบบแนะนำ, การกรองแบบอิงเนื้อหา, การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม, ระบบแนะนำแบบผสมผสาน

Title	FOOD RECOMMENDATION SYSTEM USING A HYBRID RECOMMENDATION METHOD
Author	NIPAPORN PANNAM
Degree	MASTER OF SCIENCE
Academic Year	2020
Thesis Advisor	Asst. Prof. Waraporn Viyanon , Ph.D.

The increasing in both sales and products offered to meet the needs of consumers is a channel that increases income for entrepreneurs and increases product satisfaction among consumers. The objective of this research is to develop a food recommendation model to assist making product recommendations using Content-Based Filtering techniques, which uses the principle of Cosine Similarity and Collaborative Filtering, using Surprise libraries to build a model using algorithms such as SVD, NMF, Baseline and KNN. The evaluation of model accuracy was performed by comparing RMSE and MAE values. The results showed that the SVD algorithm obtained the lowest RMSE and MAE values of 1.2528 and 0.9376, respectively. Then, another model was built that combines the techniques, Content-Based Filtering and Collaborative Filtering which takes the advantages of both techniques to more suitable product recommendations.

Keyword : Recommendation System, Content-Based Filtering, Collaborative Filtering, Hybrid Recommendation System

กิตติกรรมประกาศ

สารนิพนธ์เล่มนี้สำเร็จผลได้ด้วยความช่วยเหลือสนับสนุนในการทำสารนิพนธ์เล่มนี้ให้ประสบความสำเร็จ ทางผู้วิจัยขอกราบขอบพระคุณ ผศ.ดร.วราภรณ์ วิทยานนท์ อาจารย์ที่ปรึกษาที่เมตตาให้คำแนะนำและช่วยเหลือในหลาย ๆ ด้านในการทำสารนิพนธ์ตลอดจนจบ

ขอกราบขอบพระคุณ ประธาน คณะกรรมการการสอบสารนิพนธ์และอาจารย์ทุกท่านที่ให้คำแนะนำในแก้ไขปัญหาและชี้แนะแนวทางในการทำวิจัยให้ผลลัพธ์ออกมาดีขึ้น

ขอกราบขอบพระคุณ ผศ.ดร.นุรีย์ วิวัฒน์วัฒนาและอาจารย์ทุกท่านในสาขาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ที่ได้มอบความรู้และแนะนำสิ่งต่าง ๆ ในการศึกษาให้แก่ชีวิต

ขอกราบขอบพระคุณทางบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ในการสนับสนุนทุนการนำเสนอผลงานวิจัยของบัณฑิตศึกษา

นิภาภรณ์ พันธุ์นาม

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ	ช
สารบัญตาราง.....	ญ
สารบัญรูปภาพ	ฎ
บทที่ 1 บทนำ.....	1
1.1 ความสำคัญและความเป็นมาของงานวิจัย.....	1
1.2 ความมุ่งหมายของงานวิจัย	2
1.3 ขอบเขตงานวิจัย	3
1.4 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย	3
บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง.....	4
2.1 ระบบแนะนำ (Recommendation System)	4
2.2 การกรองแบบอิงเนื้อหา (Content-Based Filtering).....	6
2.3 การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering)	9
2.3.1 ข้อมูลที่ได้จากการใช้งานของผู้ใช้โดยตรง (Explicit rating)	9
2.3.2 ข้อมูลทางอ้อมของผู้ใช้งาน (Implicit rating)	10
2.3.3 การพิจารณาแบบจดจำ (Memory-based)	12
2.3.4 การพิจารณาแบบจำลอง (Model-based).....	13
2.3.4.1 การแยกตัวประกอบเมทริกซ์ (Matrix Factorization).....	16
2.3.5 ผสมผสาน (Hybrid).....	18

2.4 ระบบแนะนำแบบผสมผสาน (Hybrid Recommendation System)	18
2.5 วิธีการวัดความคล้ายคลึง	19
2.5.1 Term Frequency-Inverse Document Frequency.....	19
2.5.2 ความคล้ายคลึงแบบโคไซน์ (Cosine Similarity)	20
2.6 โลบราวี Surprise	22
2.7 การวัดประสิทธิภาพของแบบจำลอง.....	22
2.7.1 ค่าความคลาดเคลื่อนกำลังสอง (Root Mean Square Error)	22
2.7.2 ค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Mean Absolute Error)	23
2.8 อัลกอริทึมในการสร้างแบบจำลองเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม	23
2.8.1 K-Nearest Neighbor (K-NN)	23
2.8.2 Singular Value Decomposition (SVD).....	25
2.8.3 Non-negative Matrix Factorization (NMF).....	26
2.9 งานวิจัยที่เกี่ยวข้อง	27
บทที่ 3 การดำเนินการวิจัย	34
3.1 กระบวนการสร้างแบบจำลอง	34
3.2 แผนการดำเนินการวิจัย.....	36
3.3 ข้อมูลที่นำมาใช้ในงานวิจัย.....	36
3.4 การสำรวจข้อมูล	36
3.5 การเตรียมข้อมูล.....	38
3.6 การสร้างแบบจำลองข้อมูล.....	40
3.6.1 เทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering).....	40
3.6.2 การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering).....	41
3.6.3 ระบบแนะนำแบบผสมผสาน (Hybrid Recommendation System).....	42

3.7 การประเมินประสิทธิภาพแบบจำลอง	43
3.7.1 เทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering).....	43
3.7.2 การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering).....	43
3.7.3 การทำงานร่วมกันแบบผสมผสาน.....	43
3.8 สรุปผลการดำเนินงาน	43
บทที่ 4 ผลการดำเนินงานวิจัย.....	45
4.1 เทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering)	45
4.2 การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering)	49
4.3 ระบบแนะนำแบบผสมผสาน (Hybrid Recommendation System)	52
บทที่ 5 สรุปผลการวิจัย อภิปรายผลและข้อเสนอแนะ.....	53
5.1 สรุปผลการวิจัย	53
5.2 อภิปรายผลการวิจัย.....	54
5.2.1 เทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering).....	54
5.2.2 เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering)	54
5.2.3 ระบบแนะนำแบบผสมผสาน (Hybrid Recommendation System).....	55
5.3 ข้อเสนอแนะ	55
บรรณานุกรม	56
ภาคผนวก ก	57
ภาคผนวก ข.....	60
ภาคผนวก ค	62
ประวัติผู้เขียน.....	64

สารบัญตาราง

	หน้า
ตาราง 1 แผนการดำเนินการวิจัย	36
ตาราง 2 โครงสร้างข้อมูลสินค้า (product.csv)	37
ตาราง 3 โครงสร้างข้อมูลความคิดเห็น (reviews.csv).....	37
ตาราง 4 ตัวอย่างข้อมูลในไฟล์ product.csv.....	39
ตาราง 5 ตัวอย่างข้อมูลในไฟล์ reviews.csv	39
ตาราง 6 ผลลัพธ์การกำหนดค่าเกณฑ์ขั้นต่ำตั้งแต่ 0.05-0.95 เมนู Berry Sweet Mascarpone	46
ตาราง 7 ผลลัพธ์การกำหนดค่าเกณฑ์ขั้นต่ำตั้งแต่ 0.05-0.95 เมนู Delights Vanilla Bean.....	47
ตาราง 8 การวัดผลประสิทธิภาพของแบบจำลองด้วยค่าของ RMSE และ MAE	49
ตาราง 9 ค่าพารามิเตอร์ในการสร้างแบบจำลองอัลกอริทึม SVD	50
ตาราง 10 ค่าพารามิเตอร์ในการสร้างแบบจำลองอัลกอริทึม NMF	51
ตาราง 11 ค่าพารามิเตอร์ในการสร้างแบบจำลองอัลกอริทึม Baseline	51
ตาราง 12 ค่าพารามิเตอร์ในการสร้างแบบจำลองอัลกอริทึม K-NN	51

สารบัญรูปภาพ

	หน้า
ภาพประกอบ 1 ตัวอย่างของการทำงานของระบบแนะนำในชีวิตประจำวัน	2
ภาพประกอบ 2 ระบบคำแนะนำของ Netflix	5
ภาพประกอบ 3 วิธีการสำหรับระบบแนะนำ	5
ภาพประกอบ 4 ตัวอย่างของเทคนิคการกรองแบบอิงเนื้อหา	6
ภาพประกอบ 5 ตัวอย่างของคำแนะนำภาพยนตร์สำหรับผู้ใช้งาน Netflix	7
ภาพประกอบ 6 การสร้าง User Vector	7
ภาพประกอบ 7 เวกเตอร์รายการสำหรับภาพยนตร์	8
ภาพประกอบ 8 การคูณเวกเตอร์ 2 มิติ	8
ภาพประกอบ 9 explicit และ implicit feedback	9
ภาพประกอบ 10 ตัวอย่างการให้คะแนนจากผู้ใช้งาน	10
ภาพประกอบ 11 ตัวอย่างข้อมูลทางอ้อมของผู้ใช้งาน	10
ภาพประกอบ 12 ตัวอย่างการแนะนำแบบ User-based	11
ภาพประกอบ 13 ตัวอย่างการแนะนำแบบ Item-based	11
ภาพประกอบ 14 เทคนิคการกรองข้อมูลแบบพึ่งพาผู้เข้าร่วมแต่ละประเภท	12
ภาพประกอบ 15 ความสัมพันธ์ระหว่างผู้ใช้งาน	13
ภาพประกอบ 16 ความสัมพันธ์ระหว่างสินค้า	13
ภาพประกอบ 17 เวกเตอร์แต่ละผู้ใช้งานที่ให้ความสำคัญกับคุณลักษณะของโรงแรม	14
ภาพประกอบ 18 คุณลักษณะของโรงแรม	15
ภาพประกอบ 19 ตัวอย่างการคาดเดาผู้ใช้งาน B จะมีความชื่นชอบโรงแรมใด	15
ภาพประกอบ 20 เมทริกซ์ของผู้ใช้งานและการให้คะแนน	16
ภาพประกอบ 21 ตัวอย่างการแยกตัวประกอบของเมทริกซ์	17

ภาพประกอบ 22	การหาเมทริกซ์ของผู้ใช้งาน X และเมทริกซ์ของสินค้า Y	18
ภาพประกอบ 23	ตัวอย่างของระบบแนะนำแบบผสมผสาน	19
ภาพประกอบ 24	การหาความคล้ายคลึงแบบโคไซน์	21
ภาพประกอบ 25	การประยุกต์ใช้การหาความคล้ายคลึงแบบโคไซน์	21
ภาพประกอบ 26	K-Nearest Neighbor	24
ภาพประกอบ 27	เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมโดยใช้อัลกอริทึม K-NN	24
ภาพประกอบ 28	ตัวอย่างการแปลงเมทริกซ์	25
ภาพประกอบ 29	ตัวอย่างการลดมิติของวิธี Non-negative Matrix Factorization	26
ภาพประกอบ 30	Non-Negative Matrix สำหรับเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม	27
ภาพประกอบ 31	กราฟย่อยของเครือข่าย Bayesian แบบผสมผสาน	28
ภาพประกอบ 32	ขั้นตอนระบบการแนะนำหลักสูตรแบบผสมผสาน	30
ภาพประกอบ 33	ระบบแนะนำอาหารแบบผสมผสาน	31
ภาพประกอบ 34	แดชบอร์ดของระบบแนะนำ	32
ภาพประกอบ 35	โครงสร้างของระบบแนะนำแบบผสมผสาน	33
ภาพประกอบ 36	กระบวนการทำงานของการสร้างแบบจำลองแต่ละประเภท	35
ภาพประกอบ 37	การทำงานของระบบแนะนำสินค้า	38
ภาพประกอบ 38	การสร้างแบบจำลองของเทคนิคการกรองแบบอิงเนื้อหา	40
ภาพประกอบ 39	ตัวอย่างค่าเวกเตอร์วิธีการของ TF-IDF ที่ได้จากคอดัชนี description	40
ภาพประกอบ 40	ตัวอย่างโค้ดการทำ TF-IDF เพื่อนำมาสร้างความคล้ายคลึงแบบโคไซน์	41
ภาพประกอบ 41	การสร้างแบบจำลองของการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม	41
ภาพประกอบ 42	การสร้างแบบจำลองของระบบแนะนำแบบผสมผสาน	42
ภาพประกอบ 43	ตัวอย่างชื่อเมนูที่ได้จากระบบแนะนำด้วยเทคนิคการกรองแบบอิงเนื้อหา	45

ภาพประกอบ 44 การประเมินประสิทธิภาพแบบจำลองด้วยค่า RMSE และ MAE ของอัลกอริทึม SVD, NMF, Baseline และ KNN..... 50

ภาพประกอบ 45 ตัวอย่างผลลัพธ์ชื่อเมนูที่ได้จากการแนะนำด้วยวิธีระบบแนะนำแบบผสมผสาน 52

ภาพประกอบ 46 การประเมินประสิทธิภาพแบบจำลองด้วยค่า RMSE และ MAE ของอัลกอริทึม SVD..... 54



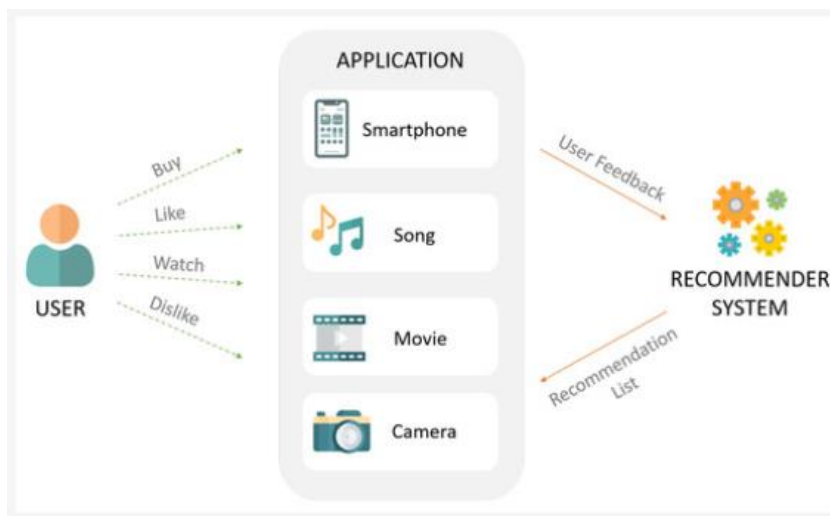
บทที่ 1

บทนำ

1.1 ความสำคัญและความเป็นมาของงานวิจัย

ปัจจุบันธุรกิจผลิตภัณฑ์ประเภทไอศกรีมได้เติบโตและแพร่หลายเป็นอย่างมาก ซึ่งมีหลากหลายแบรนด์และรสชาติ ที่พบเห็นได้ตามร้านค้าหรือห้างสรรพสินค้าทั่วไป เพื่อเป็นการเพิ่มยอดขายให้กับผู้ประกอบการและนำเสนอผลิตภัณฑ์ที่มีจำนวนมากให้ตรงตามความต้องการลูกค้าและช่วยให้ลูกค้าตัดสินใจซื้อผลิตภัณฑ์ได้ง่ายขึ้น ระบบแนะนำจึงเป็นวิธีที่จะช่วยให้ผู้ประกอบการสามารถเพิ่มรายได้และช่วยให้ผู้บริโภคได้รับสินค้าตรงตามความชื่นชอบหรือสนใจ

ระบบแนะนำ (Recommendation System) คือระบบที่ทำการแนะนำสินค้าหรือบริการ ช่วยเพิ่มความสะดวกให้กับผู้ใช้งานในการเลือกซื้อสินค้าและบริการ สามารถเพิ่มยอดขายให้กับทางร้านค้าโดยระบบจะแนะนำสินค้าที่คาดว่าจะผู้ใช้ให้ความสนใจ ระบบแนะนำจึงมีความสำคัญในการกระตุ้นยอดขายซึ่งถูกนำมาใช้อย่างแพร่หลายบนเว็บไซต์หรือแอปพลิเคชัน จากภาพประกอบ 1 โดยจะเก็บข้อมูลหลากหลายรูปแบบของผู้ใช้งานเพื่อนำไปประมวลผลในการสร้างแบบจำลอง โดยจะใช้อัลกอริทึมต่าง ๆ และนำข้อมูลที่ได้จากการเก็บตัวอย่างข้อมูลการใช้งาน เช่น เวลาที่ผู้ใช้งานกดดูภาพยนตร์เรื่องนั้น ภาพยนตร์เป็นประเภทไหนหรือความนิยมของหนังเท่าไร ระบบจะทำการแนะนำหนังที่ผู้ใช้อาจจะชื่นชอบมาให้โดยดูจากพฤติกรรมการดูและเทียบกับผู้ใช้รายอื่น ๆ ประกอบด้วย ในส่วนของเทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering) คือ การแนะนำสินค้าหรือบริการที่ได้มาจากการพิจารณาความคล้ายคลึงของพฤติกรรมจากประวัติการใช้งานของผู้ใช้งาน โดยจะดูจากลักษณะของสินค้าที่จะทำการแนะนำและจะทำการแนะนำสินค้าที่มีลักษณะใกล้เคียงที่ผู้ใช้งานเคยใช้หรือชื่นชอบในครั้งถัดไป สำหรับเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้งานร่วม (Collaborative Filtering) คือการนำการนำข้อมูลจากผู้ใช้งานรายอื่น ๆ มาช่วยในการคาดเดาว่าผู้ใช้งานรายนี้จะชื่นชอบสินค้าอะไร เช่น ผู้ใช้งาน A และ B ชอบดูหนังประเภทเดียวกัน ถ้าผู้ใช้งาน B ชื่นชอบดูหนังแนววิทยาศาสตร์ ระบบจะทำการแนะนำหนังแนววิทยาศาสตร์ให้กับผู้ใช้งาน A ด้วยเช่นกัน หรือแม้กระทั่งเว็บไซต์ Medium ซึ่งเป็นฟรีบล็อก ที่จะใช้ข้อมูลจากประวัติการอ่านและค้นหาบทความของผู้ใช้งาน ซึ่งก่อนหน้านี้ที่ผู้ใช้จะมีการใช้งานหรือมีการเก็บข้อมูลหมวดหมู่ที่สนใจไว้ก่อนและมีการเก็บระยะเวลาในการอ่านแต่ละบทความ เพื่อเพิ่มความแม่นยำในการส่งบทความไปให้ตรงใจมากที่สุด และนำไปสู่การสร้างเป็นแบบจำลองเพื่อช่วยในการเพิ่มประสิทธิภาพของระบบแนะนำ



ภาพประกอบ 1 ตัวอย่างของการทำงานของระบบแนะนำในชีวิตประจำวัน
ที่มา: <https://www.mdpi.com/2076-3417/10/16/5510>

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองระบบแนะนำสินค้าให้กับลูกค้าโดยใช้เทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering) และเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) โดยนำจุดเด่นของทั้ง 2 เทคนิคมาใช้งานร่วมกันแบบผสมผสาน (Hybrid Recommendation System) เพื่อช่วยเพิ่มประสิทธิภาพของระบบแนะนำสินค้าซึ่งจะนำจุดเด่นของแต่ละเทคนิคไปปรับปรุงจุดด้อยของอีกเทคนิค เนื่องจากในส่วนของเทคนิคการกรองแบบอิงเนื้อหา จะมีจุดด้อยคือได้รับการแนะนำสินค้าที่มีความหลากหลายค่อนข้างน้อย ทำให้ลูกค้าไม่ได้รับการแนะนำสินค้าประเภทอื่นที่นอกเหนือจากที่เคยใช้งาน และทางด้านเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม การคำนวณนั้นจะดูจากข้อมูลความชอบของผู้ใช้ที่มีต่อสินค้าทั้งหมด เช่น ระดับคะแนนความพึงพอใจ แต่ทั้งนี้ผู้ใช้งานก็ไม่ได้ใช้สินค้าภายในร้านทั้งหมด

1.2 ความมุ่งหมายของงานวิจัย

ในการวิจัยครั้งนี้ผู้วิจัยได้ตั้งความมุ่งหมายไว้ดังนี้

1.2.1 พัฒนาระบบแนะนำ โดยใช้วิธีการเทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering) และเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) มาแนะนำให้ผู้บริโภคสามารถตัดสินใจเลือกซื้อสินค้าได้ง่ายขึ้นต่อความต้องการ ซึ่งสามารถเพิ่มยอดขายให้กับทางร้านค้าได้

1.2.2 พัฒนาระบบและเปรียบเทียบเทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering) และเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) โดยนำทั้ง 2 เทคนิคมาทำงานร่วมกันแบบผสม (Hybrid Recommendation System) เพื่อเพิ่มประสิทธิภาพการทำงานของโมเดล และลดข้อจำกัดของแต่ละเทคนิค

1.3 ขอบเขตงานวิจัย

งานวิจัยนี้นำข้อมูล Ice Cream Dataset มาทำการศึกษาข้อมูลและสร้างระบบแนะนำในการแนะนำสินค้าให้กับลูกค้า โดยใช้เทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering) และเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) ซึ่งไฟล์ข้อมูลมีจำนวนทั้งหมด 2 ไฟล์ ได้แก่ ชุดข้อมูลไฟล์ products.csv และ ชุดข้อมูลไฟล์ reviews.csv โดยใช้เทคนิคการกรองแบบอิงเนื้อหา และ เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมมาทำงานร่วมกันแบบผสมผสานเพื่อให้ระบบคำแนะนำมีประสิทธิภาพมากขึ้น

1.4 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

1. นำระบบแนะนำสินค้าไปประยุกต์ใช้ในการเพิ่มยอดขายสินค้าในรายการอื่นได้
2. สามารถใช้วิธีการเทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering) และเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) เพิ่มยอดขายให้กับร้านค้า และเพิ่มความพึงพอใจให้กับลูกค้า ในกรณีที่ร้านแต่ละร้านมีสินค้าชนิดเดียวกันหรือคล้ายคลึงกัน ลูกค้าจะใช้เกณฑ์อะไรในการตัดสินใจว่าจะเลือกซื้อสินค้าร้านใด โดยจะใช้ระบบแนะนำสินค้ามาช่วยให้ลูกค้าสามารถตัดสินใจเลือกซื้อสินค้าได้ง่ายขึ้น

บทที่ 2

วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

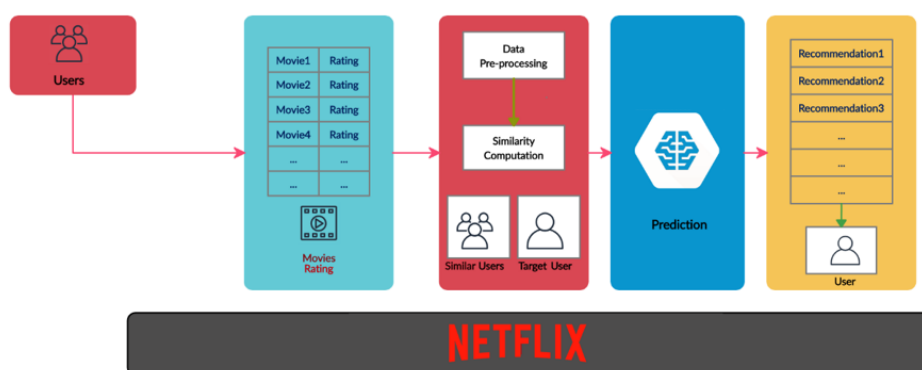
การวิจัยครั้งนี้ทางผู้วิจัยได้ทำการศึกษาบทความทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการพัฒนาระบบแนะนำสินค้า(Recommendation System) โดยใช้เทคนิคการกรองแบบอิงเนื้อหา(Content-Based Filtering) และเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) มาทำงานร่วมกันแบบผสมผสาน จึงได้นำเสนอหัวข้อดังนี้

1. ระบบแนะนำสินค้า (Recommendation System)
2. การกรองแบบอิงเนื้อหา (Content-Based Filtering)
3. การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering)
4. ระบบแนะนำแบบผสมผสาน (Hybrid Recommendation System)
5. วิธีการวัดความคล้ายคลึง
6. โลบลารี Surprise
7. การวัดประสิทธิภาพของแบบจำลอง
8. อัลกอริทึมในการสร้างแบบจำลองเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม
9. งานวิจัยที่เกี่ยวข้อง

2.1 ระบบแนะนำ (Recommendation System)

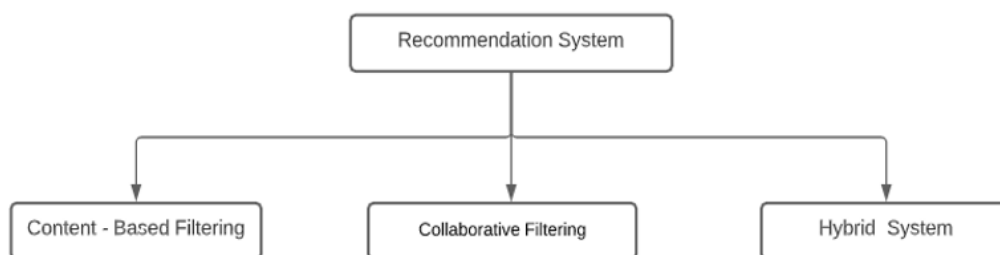
ระบบแนะนำ คือระบบที่มีไว้เพื่อแนะนำสินค้าหรือบริการให้กับผู้ใช้งาน ซึ่งจะช่วยแนะนำสินค้าให้กับผู้ใช้งานในการเลือกซื้อสินค้าและบริการต่าง ๆ ระบบจะทำหน้าที่ช่วยในการแนะนำสินค้าโดยอ้างอิงจากข้อมูลและประวัติการใช้งานของผู้ใช้บริการ ระบบจะทำการเรียนรู้จากพฤติกรรมประวัติการใช้งาน เพื่อที่จะทำการแนะนำสินค้าให้ตรงตามความต้องการของผู้ใช้งาน ทำให้เกิดประโยชน์ในด้านการใช้งานโดยที่ผู้ใช้งานจะไม่เสียเวลาในการค้นหาสินค้าหรือบริการ ซึ่งระบบแนะนำได้ถูกนำมาใช้งานเพื่อแนะนำสินค้าและบริการอย่างมาก จากภาพประกอบ 2 Netflix ได้มีการนำระบบแนะนำเข้ามาแนะนำประเภทของหนังให้กับผู้ใช้งาน ในกรณีที่ผู้ใช้งาน A ชอบดูหนังประเภทระทึกขวัญ ระบบจะทำการเรียนรู้พฤติกรรมโดยใช้ข้อมูลจากประวัติการดูหนังของผู้ใช้งาน A และแนะนำเฉพาะหนังประเภทระทึกขวัญในเรื่องอื่น ๆ ให้กับผู้ใช้งาน A ด้วยเช่นกัน หรือถ้าผู้ใช้งาน A และผู้ใช้งาน B มีความชื่นชอบหนังประเภทสารคดีเหมือนกัน เมื่อผู้ใช้งาน A เข้าไปดูหนังประเภทอาชญากรรม ระบบของหนังก็จะแนะนำหนังประเภทอาชญากรรมให้กับผู้ใช้งาน B ด้วย จากภาพประกอบ 2 คือขั้นตอนการทำงานของระบบ

คำแนะนำที่แสดงการทำงานร่วมกันของผู้ใช้งานเกี่ยวกับการให้คะแนนของภาพยนตร์หรือรายการต่าง ๆ ผู้ใช้งานใหม่จะได้รับคำแนะนำตามคำแนะนำของผู้ใช้งานที่มีอยู่ โดยถ้าระบบแนะนำสินค้าสามารถแนะนำได้ตรงตามความต้องการของผู้ใช้งาน ก็จะช่วยให้สร้างรายได้เพิ่มขึ้น ซึ่งทางผู้วิจัยเลือกที่จะพัฒนาระบบแนะนำสินค้าให้กับผู้ใช้งานโดยใช้วิธีการดังภาพประกอบ 3



ภาพประกอบ 2 ระบบคำแนะนำของ Netflix

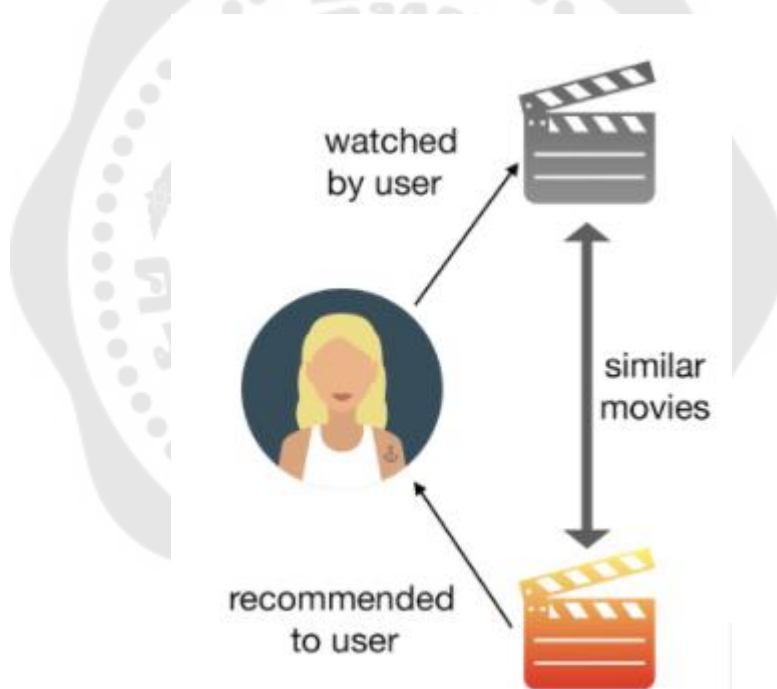
ที่มา: <https://pub.towardsai.net/recommendation-system-in-depth-tutorial-with-python-for-netflix-using-collaborative-filtering-533ff8a0e444>



ภาพประกอบ 3 วิธีการสำหรับระบบแนะนำ

2.2 การกรองแบบอิงเนื้อหา (Content-Based Filtering)

เทคนิคการกรองแบบอิงเนื้อหา คือการแนะนำสินค้าหรือบริการที่ได้มาจากการพิจารณาความคล้ายคลึงของพฤติกรรมจากประวัติการใช้งานของผู้ใช้งาน โดยจะดูจากลักษณะของสินค้าที่จะแนะนำและจะทำการแนะนำสินค้าที่มีลักษณะใกล้เคียงกับผู้ใช้งานเคยใช้หรือชื่นชอบ จากภาพประกอบ 4 ระบบจะทำการแนะนำเนื้อหาของหนังสือที่มีความคล้ายคลึงกับหนังสือที่ผู้ใช้งานเคยดูมาก่อนหน้านั้น ซึ่งข้อดีของเทคนิคการกรองแบบอิงเนื้อหา นั้น คือ ไม่ต้องใช้ข้อมูลจำนวนมากในการประมวลผลเพื่อแนะนำสินค้า และการแนะนำสินค้าใหม่ที่ยังไม่มีข้อมูลสามารถทำได้เพราะจะพิจารณาจากความคล้ายคลึงกับลักษณะสินค้าประเภทเดิม ส่วนข้อเสียของเทคนิคการกรองแบบอิงเนื้อหา จะไม่สามารถแนะนำสินค้าที่มีความแตกต่างจากสินค้าประเภทที่ผู้ใช้งานเคยซื้อหรือใช้บริการ ทำให้ผู้ใช้งานไม่ได้รับการแนะนำสินค้าที่มีความหลากหลาย



ภาพประกอบ 4 ตัวอย่างของเทคนิคการกรองแบบอิงเนื้อหา

ที่มา: <https://towardsdatascience.com/how-to-build-from-scratch-a-content-based-movie-recommender-with-natural-language-processing-25ad400eb243>

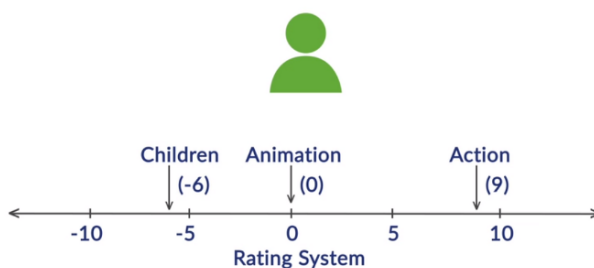


Movies	Reviews Given	Rating
Mission Impossible	✓	Good
James Bond	✓	Good
Toy Story	✓	Bad

ภาพประกอบ 5 ตัวอย่างของคำแนะนำภาพยนตร์สำหรับผู้ใช้งาน Netflix

ที่มา: <https://medium.com/@bindhubalu/content-based-recommender-system-4db1b3de03e7>

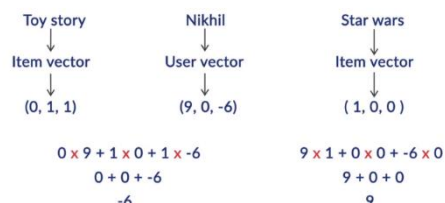
จากภาพประกอบ 5 ตัวอย่างของคำแนะนำภาพยนตร์สำหรับผู้ใช้งานใน Netflix โดยใช้ชื่อโปรไฟล์ Nikhil ในกรณีที่ Nikhil ให้คะแนนที่ดีกับหนังเรื่อง Mission Impossible และ James Bond ที่ถูกจัดเป็นประเภทหนังบู๊ และให้คะแนนต่ำกับหนังเรื่อง Toy Story ที่ถูกจัดเป็นประเภทหนังสำหรับเด็ก ก็จะสามารถทำการสร้างเวกเตอร์ของผู้ใช้งานสำหรับ Nikhil ตามการให้คะแนน 3 อันดับ ดังภาพประกอบ 6 ในระดับคะแนน -10 ถึง 10 เนื่องจาก Nikhil ชอบภาพยนตร์ประเภทหนังบู๊ จึงได้กำหนดค่า 9 ให้กับ ประเภทหนังบู๊ และ Nikhil ยังไม่ได้ดูภาพยนตร์แอนิเมชัน จึงกำหนด 0 ให้กับประเภทแอนิเมชัน และเนื่องจาก Nikhil ได้ให้คำวิจารณ์ที่ไม่ดีสำหรับ ภาพยนตร์ที่ประเภทเด็กจึงกำหนด -6 ให้กับประเภทหนังสำหรับเด็ก



ภาพประกอบ 6 การสร้าง User Vector

ที่มา: <https://medium.com/@bindhubalu/content-based-recommender-system-4db1b3de03e7>

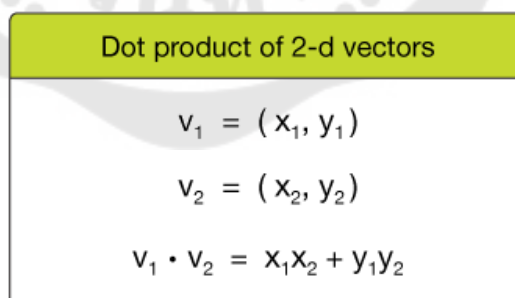
ดังนั้นเวกเตอร์ของ Nikhil จึงมีค่าเท่ากับ (9, 0, -6) ตามลำดับ (ประเภทหนังบู๊, ประเภทแอนิเมชัน, ประเภทหนังสำหรับเด็ก) เวกเตอร์สำหรับภาพยนตร์เรื่อง Toy Story คือ (0, 1, 1) และภาพยนตร์เรื่อง Star Wars คือ (1, 0, 0) ตามลำดับ ตามภาพประกอบ 7



ภาพประกอบ 7 เวกเตอร์รายการสำหรับภาพยนตร์

ที่มา: <https://medium.com/@bindhubalu/content-based-recommender-system-4db1b3de03e7>

จากภาพประกอบ 8 การคูณเวกเตอร์ 2 มิติสองตัว คือ เวกเตอร์ของรายการ และเวกเตอร์ของผู้ใช้งาน ดังนั้นการคูณเวกเตอร์ของ Toy Story คือ -6 และ Star Wars คือ 9 จึงได้แนะนำหนังเรื่อง Star Wars ให้กับ Nikhil ซึ่งตรงกับความชอบของ Nikhil ที่ชอบดูหนังประเภทหนังบู๊และไม่ชอบภาพยนตร์สำหรับเด็ก ในลักษณะที่คล้ายกันนั้นสามารถคำนวณการคูณเวกเตอร์ของเวกเตอร์รายการทั้งหมดของภาพยนตร์ทั้งหมดในร้านค้าและแนะนำภาพยนตร์ 10 อันดับแรกให้กับ Nikhil

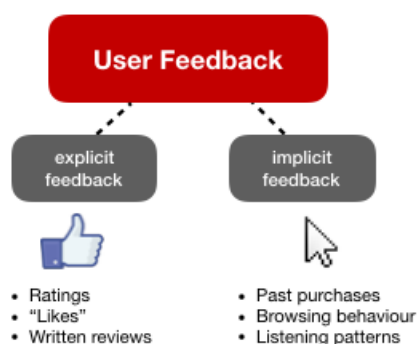


ภาพประกอบ 8 การคูณเวกเตอร์ 2 มิติ

ที่มา: <https://medium.com/@bindhubalu/content-based-recommender-system-4db1b3de03e7>

2.3 การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering)

การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม คือ การแนะนำสินค้าหรือบริการ โดยพิจารณาจากความคล้ายคลึงกันกับผู้ใช้งานรายอื่นมาประกอบคำแนะนำด้วย โดยนำข้อมูลจากผู้ใช้งานรายอื่นมาทำการคาดเดาว่าผู้ใช้งานรายนี้จะมีความชื่นชอบสินค้าประเภทใด ซึ่งจะใช้หลักการของ The Wisdom of the Clouds คือฐานข้อมูลจะต้องมีค่า rating ของสินค้าหรือบริการของผู้ใช้งานแต่ละคนที่ผ่านมา จากภาพประกอบ 9 จะมีการแบ่ง rating ออกเป็น 2 ประเภท ได้แก่ 1. ข้อมูลที่ได้จากการใช้งานของผู้ใช้โดยตรง (Explicit rating) 2. ข้อมูลทางอ้อมของผู้ใช้งาน (Implicit rating)



ภาพประกอบ 9 explicit และ implicit feedback

ที่มา: <https://medium.com/@topspinj/recommender-systems-101-bcbdfbe1e6e7>

2.3.1 ข้อมูลที่ได้จากการใช้งานของผู้ใช้โดยตรง (Explicit rating)

ข้อมูลที่ได้จากการใช้งานของผู้ใช้โดยตรง เช่น การให้คะแนนความพึงพอใจระดับคะแนนตั้งแต่ 0 – 5 คะแนน เช่น ผู้ใช้งาน A ให้คะแนนรีวิวสินค้า B ด้วยคะแนน 8/10 คะแนน แสดงว่าผู้ใช้งาน A นั้นมีความชื่นชอบสินค้า B อยู่พอสมควร หรือถ้ามีการให้คะแนนสินค้า C ด้วยคะแนนเพียงแค่ 1/10 คะแนน ก็จะพอทราบได้ว่าผู้ใช้ A นั้นไม่ชื่นชอบสินค้า C ข้อดี คือ ผู้ใช้งานสามารถบอกได้โดยตรงว่าชอบหรือไม่ชอบอะไรและสามารถบอกระดับความชอบ จากน้อยไปมากได้ ส่วนข้อเสียที่ผู้ใช้งานส่วนใหญ่ไม่ได้ซื้อหรือใช้สินค้าทั้งหมดภายในร้าน โดยเฉพาะถ้าเว็บไซต์ของมีสินค้าจำนวนมากหรือถ้าผู้ใช้งานได้ลองใช้สินค้าแล้ว แต่ในบางทีก็อาจไม่ได้เข้าทำการให้คะแนนนี้ ทำให้ข้อมูลในส่วนของ rating นี้มีจำนวนน้อยมาก ซึ่งอาจจะส่งผลให้เกิดปัญหาในเรื่องของการ sparse และ scalability จากภาพประกอบ 10 จะเป็นตัวอย่างการให้คะแนนจากผู้ใช้งาน



ภาพประกอบ 10 ตัวอย่างการให้คะแนนจากผู้ใช้งาน

ที่มา: <https://www.medium.com/@sinart.t/recommendation-system-แบบสรุปเอาเอง-ce6246f49754>

2.3.2 ข้อมูลทางอ้อมของผู้ใช้งาน (Implicit rating)

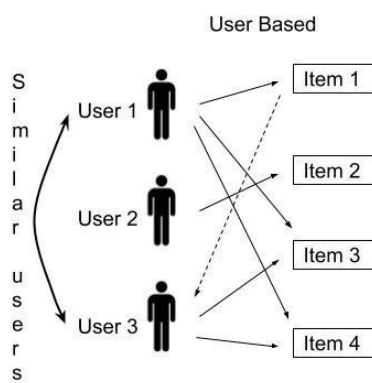
ข้อมูลทางอ้อมของผู้ใช้งาน คือ การคาดเดาจากการใช้งานของผู้ใช้ที่มีความชื่นชอบหรือสนใจในสินค้า เช่น การคลิกเข้าไปดูสินค้าชิ้นนั้น หรือการตัดสินใจที่จะเลือกซื้อสินค้าชิ้นนั้น ตามภาพประกอบ 11 ข้อมูลทางอ้อมของผู้ใช้งานจะมีปริมาณข้อมูลที่เยอะมากกว่าข้อมูลที่ได้จากการใช้งานของผู้ใช้โดยตรง เมื่อมีผู้ใช้งานเข้ามาใหม่สามารถเก็บข้อมูลได้ทันทีเมื่อผู้ใช้งานทำการคลิกเข้าชมสินค้า ข้อดีคือมีปริมาณข้อมูลมากเมื่อเทียบกับ Explicit rating สามารถใช้ได้กับผู้ใช้ใหม่ที่เข้ามาได้ ถึงแม้จะยังไม่เคยมีประวัติการใช้งานมาก่อน ส่วนข้อเสียนั้นเดาได้เพียงแค่ว่าผู้ใช้งานชอบอะไร แต่ไม่มีผลตอบรับกลับมา เช่น หากผู้ใช้งานไม่ได้คลิกสินค้ารายการ A ทำให้ไม่สามารถบอกได้ว่าผู้ใช้งานรายนี้ไม่ได้ชอบสินค้า A แต่อาจจะเป็นเพราะว่าผู้ใช้งานไม่เห็นสินค้า A



ภาพประกอบ 11 ตัวอย่างข้อมูลทางอ้อมของผู้ใช้งาน

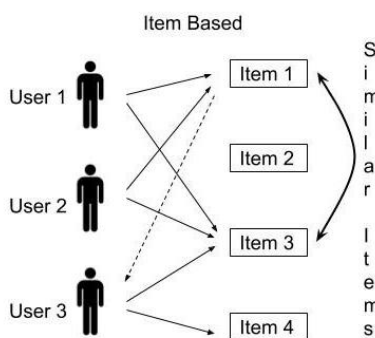
ที่มา: <http://blog.pandata.co/how-does-spotify-know-what-music-i/>

ข้อมูล rating สามารถนำมาใช้ในการแนะนำสินค้าได้ 2 วิธี ได้แก่ 1. User-Based คือ ผู้ใช้งานที่ถูกจัดอยู่ในกลุ่มประเภทเดียวกัน ซึ่งจะมีความชื่นชอบสินค้าชนิดเดียวกัน จาก ภาพประกอบ 12 ถ้าผู้ใช้ 1 และ ผู้ใช้ 3 ชอบสินค้ารายการ 3 และ 4 เหมือนกัน ในกรณีที่ ผู้ใช้ 1 สั่งสินค้ารายการ 1 เพิ่มเติม ระบบแนะนำก็จะทำการแนะนำรายการสินค้า 1 ให้กับ ผู้ใช้ 3 ด้วย เช่นกัน 2. Item-Based คือ สินค้าที่มีลักษณะใกล้เคียงกัน เช่น ถ้าสินค้ารายการ 1 และ 3 มีความ คล้ายคลึงกัน ในกรณีที่ ผู้ใช้ 1 ตัดสินใจเลือกซื้อสินค้ารายการ 1 โอกาสที่ ผู้ใช้ 3 จะตัดสินใจเลือก ซื้อสินค้ารายการ 1 ก็มีเช่นกัน โดยดูตัวอย่างได้จากภาพประกอบ 13



ภาพประกอบ 12 ตัวอย่างการแนะนำแบบ User-based

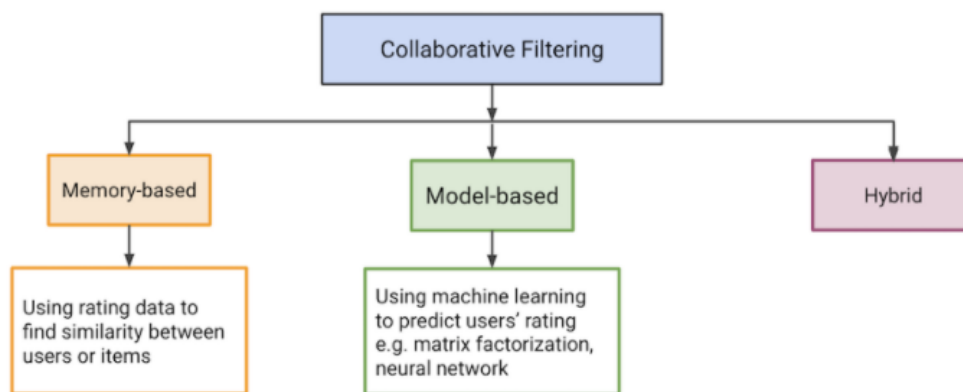
ที่มา: <https://levelup.gitconnected.com/the-mathematics-of-recommendation-systems-e8922a50bdea>



ภาพประกอบ 13 ตัวอย่างการแนะนำแบบ Item-based

ที่มา: <https://levelup.gitconnected.com/the-mathematics-of-recommendation-systems-e8922a50bdea>

เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม มีการนำข้อมูลจากผู้ใช้งานรายอื่นมาช่วยในการแนะนำว่าผู้ใช้งานรายนี้จะชื่นชอบสินค้าประเภทใด โดยรูปแบบของเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม จะมีทั้งหมด 3 ประเภท ตามภาพประกอบ 14 ซึ่งงานวิจัยนี้จะมีการสร้างเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมจากประเภทของ Model-based

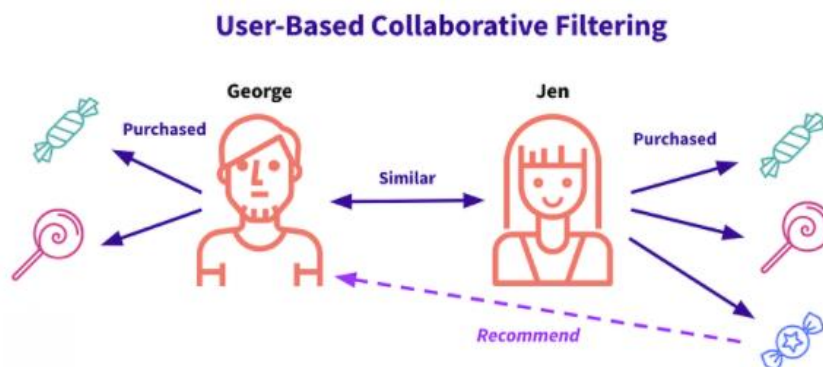


ภาพประกอบ 14 เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมแต่ละประเภท

ที่มา: <https://www.medium.com/@sinart.t/recommendation-system-แบบสรุปเอาเอง-ce6246f49754>

2.3.3 การพิจารณาแบบจดจำ (Memory-based)

ดูจากข้อมูลและหาความสัมพันธ์ระหว่างผู้ใช้งานกับสินค้า การใช้ Nearest Neighbor ในการคาดเดา rating จากผู้ใช้งาน (user-based) หรือ สินค้า (item-based) ที่มีความใกล้เคียงกัน จากภาพประกอบ 15 ข้อมูลจากผู้ใช้งาน (user-based) ถ้าคุณ George และ คุณ Jen มีการใช้งานที่คล้ายคลึงกัน ในกรณีที่คุณ Jen ทำการสั่งซื้อสินค้าประเภทอื่นเพิ่มเติม ระบบแนะนำก็จะทำการแนะนำสินค้าชิ้นนี้ให้กับคุณ George ด้วยเช่นกัน และจากภาพประกอบ 16 ข้อมูลจากสินค้า เช่น ถ้าสินค้าเบอร์ 1 มีความคล้ายคลึงกับสินค้าเบอร์ 2 ในกรณีที่ผู้ใช้งานเลือกซื้อสินค้าเบอร์ 1 ก็อาจจะซื้อสินค้าเบอร์ 2 ด้วย



ภาพประกอบ 15 ความสัมพันธ์ระหว่างผู้ใช้งาน

ที่มา: <https://blog.dataiku.com/recommendation-engines-how-they-work-in-plain-english>








ภาพประกอบ 16 ความสัมพันธ์ระหว่างสินค้า

ที่มา: <https://blog.dataiku.com/recommendation-engines-how-they-work-in-plain-english>

2.3.4 การพิจารณาแบบจำลอง (Model-based)

มีการใช้ Machine Learning เพื่อหา user embedding และ item embedding มาทำนาย rating ที่ผู้ใช้งานจะให้คะแนนกับสินค้า Model-based ส่วนใหญ่จะอยู่ในรูปแบบการใช้แบบจำลองหา user weight และ item weight เพื่อให้สามารถประเมินความชื่นชอบของผู้ใช้งานได้ จากภาพประกอบ 17 ตัวอย่างเว็บไซต์การจองโรงแรม ในการเลือกโรงแรมนั้นผู้ใช้งานมักจะมี






เกณฑ์ในการตัดสินใจเลือกโรงแรมที่ต่างกัน เช่น บางคนให้ความสำคัญกับตำแหน่งที่ตั้งโรงแรม หรือบางคนอาจจะให้ความสำคัญของรีวิวที่พิกมากกว่า เป็นต้น ผู้ใช้งานแต่ละคนจะมีเวกเตอร์ที่บ่งบอกว่าให้ความสำคัญกับคุณลักษณะใดของโรงแรมเท่าไรบ้าง เช่น user weights คือ การบอกลัดส่วนตัว หรือการให้ weight ว่าผู้ใช้งานให้ star rating และ location ที่ลัดส่วนตัวเท่าไร เช่น ผู้ใช้งาน B ให้ความสำคัญกับ location ของโรงแรมมากกว่า star rating หรือผู้ใช้งาน E นั้นให้ความสำคัญกับ star rating กับ location ที่เท่ากัน

		Star rating	location
A		0.3	0.7
B		0.1	0.9
C		0.8	0.2
D		0.15	0.85
E		0.5	0.5

ภาพประกอบ 17 เวกเตอร์แต่ละผู้ใช้งานที่ให้ความสำคัญกับคุณลักษณะของโรงแรม

ที่มา: <https://www.medium.com/@sinart.t/recommendation-system-แบบสรุปเอาเอง-ce6246f49754>

hotel features คือ คุณลักษณะของโรงแรมแต่ละโรงแรมว่ามีคะแนนในแต่ละด้านเป็นอย่างไร จากภาพประกอบ 18 ตัวเลขในแต่ละช่อง คือ คะแนนของแต่ละหมวดหมู่ เช่น โรงแรมหมายเลข 5 มีคะแนนของ star rating 10 คะแนน หรือโรงแรมหมายเลข 10 นั้นมีคะแนนในส่วน of star rating อยู่ที่ 2 คะแนน ซึ่งไม่ดีมากนัก แต่ว่ามีคะแนนในส่วน of location ที่ดี

	1	2	3	4	5
					
Star rating	2.0	7.0	3.0	8.0	10.0
location	9.0	6.0	8.5	7.5	7.0

ภาพประกอบ 18 คุณลักษณะของโรงแรม

ที่มา: <https://www.medium.com/@sinart.t/recommendation-system-แบบสรุปเอาเอง-ce6246f49754>

จากข้อมูลเบื้องต้นนั้นสามารถทำการคาดเดาได้ว่าผู้ใช้งานแต่ละคนนั้นจะมีความชื่นชอบโรงแรมใดบ้าง เช่น ผู้ใช้งาน B มีความชื่นชอบโรงแรมหมายเลข 1 เนื่องจากผู้ใช้งาน B ให้ความสำคัญกับ location และโรงแรมหมายเลข 1 ก็เป็นโรงแรมที่มี location ดีมาก ส่วนผู้ใช้งาน C อาจจะมีความชื่นชอบโรงแรมหมายเลข 5 เนื่องจากผู้ใช้งาน C ชอบโรงแรมที่มีคะแนน star rating ที่สูง ซึ่งการคาดเดาลักษณะแบบนี้สามารถนำ weight ของ users มาคูณกับคะแนนด้านต่าง ๆ ของโรงแรมได้ เช่น เมื่อนำข้อมูลผู้ใช้งาน B มาทำการคาดเดาว่าในบรรดาทั้ง 5 โรงแรมนั้น ผู้ใช้งาน B จะมีความชื่นชอบโรงแรมใดมากที่สุด และนำ weight vector ของผู้ใช้งาน B มาคูณกับ hotel features ตามภาพประกอบ 19



ภาพประกอบ 19 ตัวอย่างการคาดเดาผู้ใช้งาน B จะมีความชื่นชอบโรงแรมใด

ที่มา: <https://www.medium.com/@sinart.t/recommendation-system-แบบสรุปเอาเอง-ce6246f49754>

2.3.4.1 การแยกตัวประกอบเมทริกซ์ (Matrix Factorization)

Matrix Factorization คือ การแยกตัวประกอบของเมทริกซ์ให้ได้เมทริกซ์ย่อยลงไป โดยที่ผลคูณของเมทริกซ์ย่อยนั้นจะได้เท่ากับเป็นเมทริกซ์ดั้งเดิม โดยทั่วๆไปจะเป็นเมทริกซ์ของ rating หรือ การให้คะแนน rating ของแต่ละผู้ใช้งาน แต่ละสินค้านั้นจะทำการแยกเมทริกซ์ rating นี้ออกเป็น 2 ส่วน ได้แก่ ส่วนที่เป็นเมทริกซ์ของผู้ใช้งาน X และเมทริกซ์ของสินค้า Y โดยเมื่อนำเมทริกซ์ทั้ง X และ Y มาคูณกันก็จะได้กลับมาเป็นเมทริกซ์ rating ดั้งเดิม

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	0	3	0	3	0
User 2	4	0	0	2	0
User 3	0	0	3	0	0
User 4	3	0	4	0	3
User 5	4	3	0	4	0

ภาพประกอบ 20 เมทริกซ์ของผู้ใช้งานและการให้คะแนน

ที่มา: <http://katbailey.github.io/post/matrix-factorization-with-tensorflow/>

จากภาพประกอบ 20 คือเมทริกซ์ของการให้คะแนนที่มีผู้ใช้งานเป็นแถวและรายการเป็นคอลัมน์ ซึ่งการแยกตัวประกอบเมทริกซ์ของเทคนิคการกรองข้อมูลแบบพืงพาผู้ร่วมนั้นจะมีเมทริกซ์ 2 เมทริกซ์ คือ users และ items โดยที่เมทริกซ์ของ user กับ item จะมี ลักษณะที่จะบ่งบอกถึงความชอบหรือความใกล้เคียง และเมื่อนำเมทริกซ์มาคูณกันแล้วจะได้ความชอบของ user ที่มีต่อ item นั้น ตามภาพประกอบ 21

	Feature 1	Feature 2
User 1	?	?
User 2	?	?
User 3	?	?
User 4	?	?
User 5	?	?

X

	Item 1	Item 2	Item 3	Item 4	Item 5
Feature 1	?	?	?	?	?
Feature 2	?	?	?	?	?






=

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	0?	3	0?	3	0?
User 2	4	0?	0?	2	0?
User 3	0?	0?	3	0?	0?
User 4	3	0?	4	0?	3
User 5	4	3	0?	4	0?

ภาพประกอบ 21 ตัวอย่างการแยกตัวประกอบของเมทริกซ์

ที่มา: <http://katbailey.github.io/post/matrix-factorization-with-tensorflow/>

จากภาพประกอบ 22 ในการเรียนรู้จะพยายามหาเมทริกซ์ของผู้ใช้งาน X และเมทริกซ์ของสินค้า Y เพื่อให้ได้ผลคูณของ X และ Y นั้นให้มีความใกล้เคียงกับเมทริกซ์ rating ของจริงให้ได้มากที่สุด โดยวิธีการนั้นก็สามารใช้ SVD, PCA หรือ Gradient Descent ในส่วนของข้อดีนั้นการใช้ user และ item embedding ทำให้แบบจำลองสามารถเรียนรู้หากการแทนข้อมูลที่เหมาะสมได้และยังช่วยลดเรื่องมิติของข้อมูลได้อีกด้วย ในส่วนข้อเสียยังคงมีปัญหสำหรับผู้ใช้งานใหม่หรือสินค้าใหม่ เช่น ในกรณีที่จะทำระบบแนะนำโรงแรมให้ผู้ใช้งาน จะไม่สามารถใช้ข้อมูลอื่น ๆ ที่ผู้ใช้งานจะใช้ในการค้นหาที่พัก ไม่ว่าจะป็นจำนวนคนเข้าพัก เมืองที่ผู้ใช้ค้นหา หรือจำนวนวันเข้าพักมาร่วมใช้ในการทำนายได้

	1	2	3	4	5
					
A	6.90	6.00	6.80		8.00
B	8.30	6.00	8.00	7.40	7.00
C		7.00		8.00	9.50
D	8.00	6.00	7.75		7.50
E		6.50	5.50	7.75	

Users' ratings

\approx

Users (X)	Items (Y)				
	1	2	3	4	5
A					
B					
C					
D					
E					

\times

ภาพประกอบ 22 การหาเมทริกซ์ของผู้ใช้งาน X และเมทริกซ์ของสินค้า Y

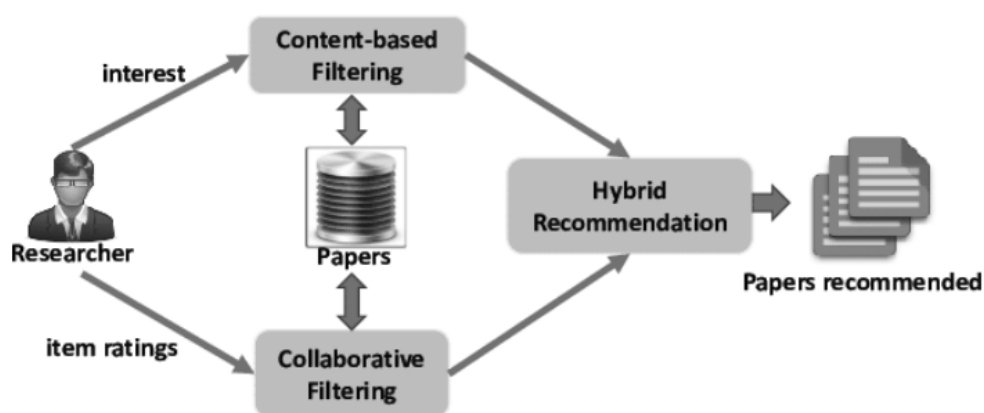
ที่มา: <https://www.medium.com/@sinart.t/recommendation-system-แบบสรุปเอาเอง-ce6246f49754>

2.3.5 ผสมผสาน (Hybrid)

การใช้เทคนิคของทั้งวิธีการของ Memory-based และ Model-based มาผสมกัน เพื่อเพิ่มประสิทธิภาพให้กับแบบจำลองโดยนำข้อดีของทั้ง 2 วิธีการนั้นมาใช้และทำให้ลดปัญหาที่เกิดขึ้นจากแบบจำลองแบบปกติ แต่ก็ต้องแลกมาด้วยความซับซ้อนที่มากขึ้นในการสร้างแบบจำลอง

2.4 ระบบแนะนำแบบผสมผสาน (Hybrid Recommendation System)

ระบบแนะนำแบบผสมผสาน คือการใช้เทคนิคของระบบแนะนำสินค้ามากกว่า 2 เทคนิคขึ้นไป จากภาพประกอบ 23 ได้นำเทคนิคการกรองแบบอิงเนื้อหา และเทคนิคการกรองข้อมูลแบบพึ่งพาผู้เข้าร่วม มาทำงานรวมกันเพื่อให้ระบบแนะนำมีประสิทธิภาพการทำงานที่ดีขึ้น โดยจะนำจุดเด่นของเทคนิคหนึ่งไปปรับปรุงจุดด้อยของอีกเทคนิคหนึ่ง เช่น เมื่อมีสินค้าใหม่เข้ามาและยังไม่มีประวัติการให้คะแนน ทำให้ไม่สามารถแนะนำด้วยเทคนิคการกรองข้อมูลแบบพึ่งพาผู้เข้าร่วมได้ ดังนั้นจึงนำวิธีการเทคนิคการกรองแบบอิงเนื้อหา มาช่วยสร้างโอกาสในการแนะนำสินค้าที่เข้ามาใหม่



ภาพประกอบ 23 ตัวอย่างของระบบแนะนำแบบผสมผสาน

ที่มา: https://www.researchgate.net/figure/A-hybrid-paper-recommendation-system_fig5_330077673

2.5 วิธีการวัดความคล้ายคลึง

2.5.1 Term Frequency-Inverse Document Frequency

Term Frequency-Inverse Document Frequency (TF-IDF) คือ วิธีการบ่งบอกน้ำหนักของคำเฉพาะ ใช้เพื่อเปรียบเทียบความเหมือนกันของคำสองคำ โดยวัดจากค่าของ Term Frequency และ Inverse Document Frequency ดังสมการ (3) ซึ่งนำ Term Frequency คือ จำนวนครั้งที่แต่ละ คำที่ปรากฏขึ้นในแต่ละข้อความและนำมาหารด้วยจำนวนคำทั้งหมดในข้อความนั้น ดังสมการ (1) จากนั้นนำมาคูณกับ Inverse Document Frequency คือจำนวนของเอกสารทั้งหมดและหารด้วยจำนวนเอกสารที่แต่ละคำปรากฏอยู่แล้วทำการ take log เข้าไป ดังสมการ (2)

$$TF(t, d) = \frac{\text{number of times term}(t) \text{ appears in document}(d)}{\text{total number of terms in document}(d)} \quad (1)$$

เมื่อ t = จำนวนครั้งที่แต่ละคำที่ปรากฏขึ้นในแต่ละข้อความ

d = จำนวนคำทั้งหมดในข้อความนั้น

$$IDF(t, D) = \log\left(\frac{\text{total number of document}(D)}{\text{number of documents with the term}(t)\text{in it}}\right) \quad (2)$$

เมื่อ D = จำนวนของเอกสารทั้งหมด
 t = จำนวนเอกสารที่แต่ละคำปรากฏ

$$TF(t, d) \times IDF(t, D) \quad (3)$$

2.5.2 ความคล้ายคลึงแบบโคไซน์ (Cosine Similarity)

ความคล้ายคลึงแบบโคไซน์ คือการหาความเหมือนของข้อมูล โดยใช้ค่าโคไซน์จากการวัดค่าข้อมูลเป็นลายลักษณ์อักษรในเทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering) จะถูกใช้เพื่อวัดความเหมือนระหว่างเวกเตอร์ของน้ำหนัก TF-IDF ค่าโคไซน์จะมีค่าอยู่ระหว่าง 0 ถึง 1 ยกตัวอย่างเช่น ค่า 1 หมายถึง ข้อมูล A และ ข้อมูล B มีลักษณะข้อมูลที่เหมือนกัน และ ส่วนของค่า 0 นั้นหมายถึงข้อมูล A และข้อมูล B มีลักษณะข้อมูลที่ไม่เหมือนกัน ดังสมการ (4) โดยมีที่มาจากกฎสามเหลี่ยมคือ $\cos(\theta) = \text{ชิด/ฉาก}$ โดยที่ A และ B คือรายการที่ต่างกัน ตามภาพประกอบ 24

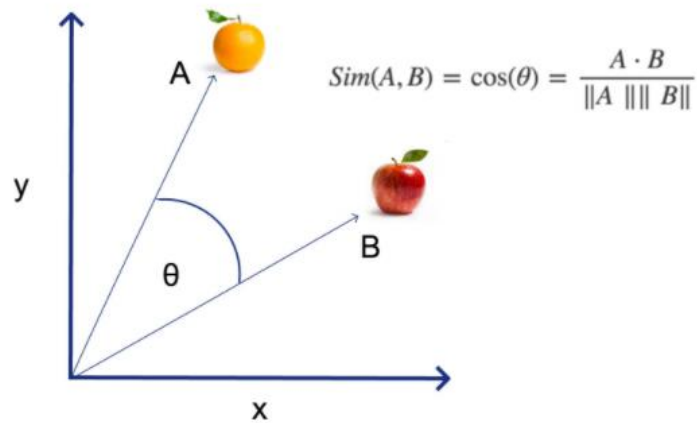
$$\text{similarity} = \cos\theta = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (4)$$

เมื่อ A = รายการ A

B = รายการ B

n = จำนวนคู่ของรายการ A และ รายการ B ทั้งหมด

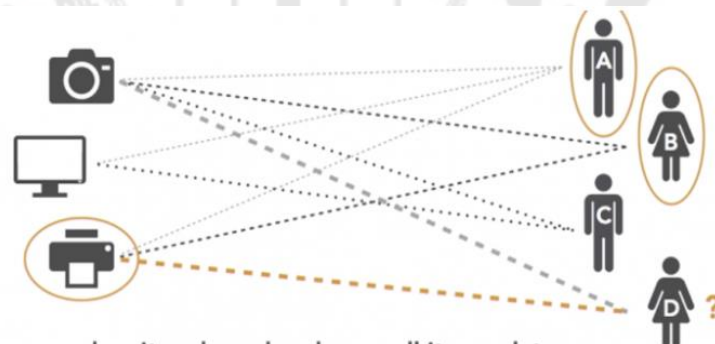
Cosine Similarity



ภาพประกอบ 24 การหาความคล้ายคลึงแบบโคไซน์

ที่มา: <https://www.linkedin.com/pulse/cosine-similarity-classification-michael-lin>

จากภาพประกอบ 25 การประยุกต์ใช้การหาความคล้ายคลึงแบบโคไซน์ การทำระบบแนะนำเพื่อดูว่าผู้ใช้งานส่วนใหญ่เมื่อซื้อสินค้า A แล้วจะมีโอกาสซื้อสินค้า B หรือ C เท่าไร



Recommend an item based on how well it correlates with other items with respect to user ratings

ภาพประกอบ 25 การประยุกต์ใช้การหาความคล้ายคลึงแบบโคไซน์

ที่มา: <https://www.softnix.co.th/2019/05/29/similarity-ความเหมือนที่แตกต่าง>

2.6 ไลบรารี Surprise

Surprise (Hug) เป็นส่วนหนึ่งของ scikit-learn ซึ่งเป็น library ใน python สำหรับนำมาวิเคราะห์และสร้างระบบแนะนำที่นำมาใช้จัดการกับข้อมูลที่มี rating โดย Surprise ถูกออกแบบมาเพื่อจุดประสงค์ดังนี้

1. ให้ผู้ใช้งานจัดการและควบคุมการทดลองได้ทั้งหมด จึงให้ความสำคัญกับการสร้างเอกสารเป็นอย่างมากเพื่อให้ผู้ใช้มีความชัดเจนในทุกรายละเอียดของอัลกอริทึม
 2. ช่วยลดขั้นตอนที่ยุ่งยากในการจัดการกับข้อมูล ซึ่งผู้ใช้สามารถใช้อุปกรณ์ได้หลากหลาย
 3. มีอัลกอริทึมใช้ทำนายและมาตรวจวัดความคล้ายคลึงกันที่พร้อมใช้ให้หลากหลาย
 4. ช่วยให้การสร้างอัลกอริทึมได้ง่ายขึ้น
 5. มีเครื่องมือในการวัดผล วิเคราะห์ และเปรียบเทียบประสิทธิภาพของอัลกอริทึม
- โดยชื่อ Surprise นั้น ย่อมาจาก Simple Python Recommendation System Engine ซึ่ง surprise ไม่รองรับการสร้างระบบแบบ Content-Based Information

2.7 การวัดประสิทธิภาพของแบบจำลอง

2.7.1 ค่าความคลาดเคลื่อนกำลังสอง (Root Mean Square Error)

Root Mean Square Error (RMSE) คือ วิธีการวัดความคลาดเคลื่อนจากค่าที่ทำนายจากแบบจำลอง ในการวัดค่าประเมินผลความแม่นยำของแบบจำลอง ยิ่งค่า RMSE ที่ได้มีค่าน้อยแสดงว่าแบบจำลองสามารถประมาณค่าได้มีความใกล้เคียงกับความจริง และถ้าหากมีค่าเท่ากับ 0 แสดงว่าไม่เกิดความคลาดเคลื่อนในแบบจำลองนั้น ดังสมการ (5)

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{Y}_i - Y_i)^2}{n}} \quad (5)$$

เมื่อ \hat{Y} = ผลลัพธ์ที่ได้จากการทำนาย

Y_i = ค่าผลลัพธ์ที่ถูกต้อง

n = จำนวนข้อมูลทั้งหมด

2.7.2 ค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Mean Absolute Error)

Mean Absolute Error (MAE) คือ วิธีการวัดค่าคลาดเคลื่อน ในการวัดค่าประเมินผล ความแม่นยำของแบบจำลอง หากผลลัพธ์ของ MAE ที่ได้มีค่าน้อยแสดงว่าแบบจำลองสามารถ ประมาณค่าได้มีความใกล้เคียงกับความจริง และถ้าหากมีค่าเท่ากับ 0 แสดงว่าไม่เกิดความคลาดเคลื่อนในแบบจำลอง ดังสมการ (6)

$$MAE = \frac{1}{n} \sum_{i=1}^n |X_i - X| \quad (6)$$

เมื่อ X_i = ผลการทำนายของแบบจำลอง

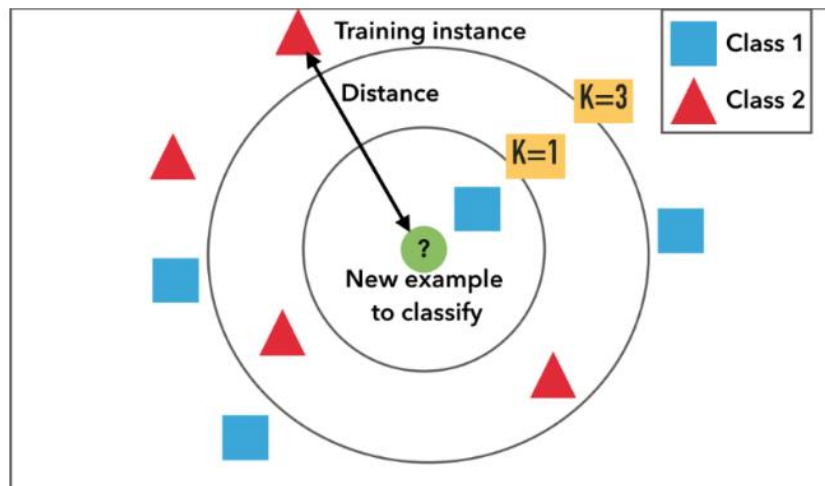
X = ค่าจริงของผลลัพธ์ข้อมูล

n = จำนวนข้อมูลทั้งหมด

2.8 อัลกอริทึมในการสร้างแบบจำลองเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม

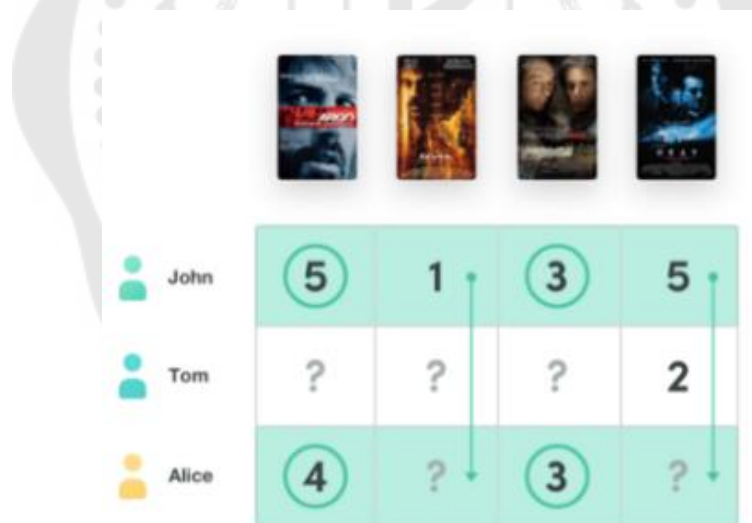
2.8.1 K-Nearest Neighbor (K-NN)

K-Nearest Neighbor คือ การคำนวณหาเพื่อนบ้านที่ใกล้ที่สุด เป็นวิธีการแบ่งคลาส สำหรับใช้จัดหมวดหมู่ข้อมูล (Classification) โดยใช้หลักการเปรียบเทียบข้อมูลที่สนใจกับข้อมูลอื่นว่ามีความคล้ายคลึงมากน้อยเพียงใด หากข้อมูลที่กำลังสนใจอยู่ใกล้ข้อมูลใดมากที่สุด ระบบจะให้คำตอบเหมือนคำตอบของข้อมูลที่อยู่ใกล้ที่สุด จากภาพประกอบ 26 โดยนำระยะทางที่ได้จากสมาชิกในข้อมูลและเลือกสมาชิกที่มีระยะทาง (Distance) ใกล้เคียงที่สุด โดยวัดระยะทางแบบระยะห่างยูคลิด (Euclidean Distance) การวัดระยะทางระหว่างสองจุด ถ้ามีระยะห่างกันมากแสดงว่าวัตถุมีความคล้ายกันน้อย แต่ถ้าระยะห่างมีค่าน้อยแสดงว่ามีความคล้ายคลึงกันมาก ตัวอย่างจากภาพประกอบ 27 ในเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม โดยการหารายการที่คล้าย ๆ กัน k ชิ้น โดยใช้การหาความคล้ายคลึงแบบโคไซน์ ในการวัดความใกล้เคียงของรายการ จากตัวอย่าง Alice มีลักษณะการให้คะแนนของหนังคล้ายกับ John แปลว่าหนังบางเรื่องที่ John เคยดูแล้ว แต่ Alice ยังไม่เคยดู Alice อาจจะให้คะแนนหนังเรื่องนั้น ๆ คล้ายกับ John



ภาพประกอบ 26 K-Nearest Neighbor

ที่มา: <https://towardsdatascience.com/prototyping-a-recommender-system-step-by-step-part-1-knn-item-based-collaborative-filtering-637969614ea>



ภาพประกอบ 27 เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมโดยใช้อัลกอริทึม K-NN

ที่มา: <https://medium.com/@nutorbitx/machine-learning-for-recommendation-system-part-1-d82caa77b82d>

2.8.2 Singular Value Decomposition (SVD)

Singular Value Decomposition (SVD) คือ วิธีแยกตัวประกอบที่มีความนิยม โดย SVD จะทำการแปลง เมทริกซ์ขนาดใหญ่ให้ออกมาเป็น 3 เมทริกซ์ขนาดเล็กกว่าที่คูณกันแล้วได้เท่ากับเมทริกซ์ต้นฉบับ ซึ่งทั้ง 3 เมทริกซ์ใหม่ที่ได้ออกมานั้นจะมีคุณสมบัติพิเศษบางอย่างทำให้สามารถนำมาใช้งานวิเคราะห์ข้อมูลได้ดีขึ้น ดังสมการ (7)

$$M = U\Sigma V^* \quad (7)$$

เมื่อ M = เมทริกซ์เดิม

U = เมทริกซ์ขนาด $m \times m$

Σ = เมทริกซ์ขนาด $m \times n$

V = เมทริกซ์ขนาด $n \times n$

จากภาพประกอบ 28 วิธีการ Singular Value Decomposition นี้เป็นหนึ่งในวิธีการที่ทางผู้วิจัยได้นำมาสร้างเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม และเป็นวิธีที่ผู้วิจัยได้ทำการเลือกไปสร้างเป็นระบบแนะนำแบบผสมผสานต่อไป

The diagram shows the decomposition of matrix M into matrices U , Σ , and V^* . Below this, it shows that $U U^* = I_m$ and $V V^* = I_n$, where I_m and I_n are identity matrices of size m and n respectively.

$$\begin{array}{cccc}
 \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} &
 \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} &
 \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} &
 \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} \\
 \mathbf{M} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^* & & & \\
 m \times n & m \times m & m \times n & n \times n
 \end{array}$$

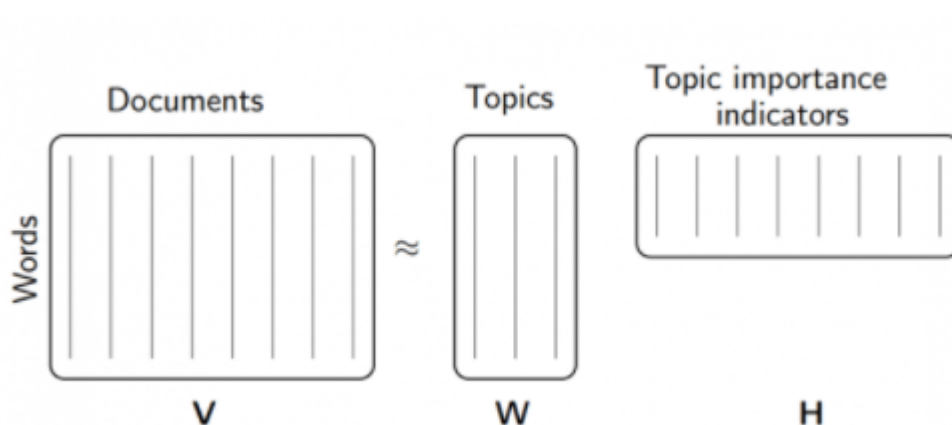
$$\begin{array}{ccc}
 \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} &
 \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} &
 \begin{array}{|c|c|c|c|} \hline 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \\ \hline \end{array} \\
 \mathbf{U} \mathbf{U}^* = \mathbf{I}_m & & & \\
 \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} &
 \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} &
 \begin{array}{|c|c|c|c|} \hline 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \\ \hline \end{array} \\
 \mathbf{V} \mathbf{V}^* = \mathbf{I}_n & & &
 \end{array}$$

ภาพประกอบ 28 ตัวอย่างการแปลงเมทริกซ์

ที่มา: <https://www.bualabs.com/archives/2971/lsa-latent-semantic-analysis-text-classification-singular-value-decomposition-svd-non-negative-matrix-factorization-nmf-nlp-ep-4/>

2.8.3 Non-negative Matrix Factorization (NMF)

Non-negative Matrix Factorization (NMF) ทำหน้าที่ลดมิติของข้อมูลโดยที่ข้อมูลจะไม่มีค่าเป็นลบ ผลลัพธ์จะได้เมทริกซ์ที่เป็นค่าบวกเท่านั้น ดังภาพประกอบ 29 ข้อดีของ NMF คือ ใช้งานสะดวกและทำงานได้อย่างรวดเร็วแต่มีข้อจำกัดคือ ผลลัพธ์เป็นค่าประมาณทำให้ไม่สามารถรวมกลับเป็น เมทริกซ์ต้นฉบับเหมือนเดิมได้เหมือน SVD



ภาพประกอบ 29 ตัวอย่างการลดมิติของวิธี Non-negative Matrix Factorization

ที่มา: <https://www.bualabs.com/archives/2971/lsa-latent-semantic-analysis-text-classification-singular-value-decomposition-svd-non-negative-matrix-factorization-nmf-nlp-ep-4/>

ตัวอย่างการใช้ Non-negative Matrix ในแอปพลิเคชัน Spotify (TITIPATA, 2016) เริ่มต้นจากการเก็บข้อมูลของผู้ใช้งาน โดยผู้ใช้งาน U มีขนาด n คน และมีเซตของเพลงซึ่งมีขนาด m เพลง นอกจากนั้นถ้ามีประวัติการฟังเพลงของสัปดาห์ที่ผ่านมาของแต่ละผู้ใช้งานว่าฟังเพลงแต่ละเพลงไปจำนวนเท่าไรบ้าง สามารถเขียนจำนวนการฟังเพลงของผู้ใช้งานทั้งหมดในรูปแบบของเมทริกซ์ โดยให้ R ซึ่งมีขนาดเท่ากับขนาดของผู้ฟังทั้งหมดและจำนวนเพลง หรือว่า $n \times m$ ในแต่ละตำแหน่งของเมทริกซ์จะใช้สัญลักษณ์ r_{ij} แทนจำนวนครั้งที่ผู้ใช้งาน i ฟังเพลง j ส่วนเพลงที่ผู้ใช้งานไม่เคยฟังนั้น จะไม่มีการเก็บข้อมูลในส่วนนี้ ถัดมาคือการหาค่าเมทริกซ์อันดับต่ำที่จะบรรยายทั้งผู้ใช้งานและเพลงที่ผู้ใช้งานฟัง ดังสมการ (8)

$$\hat{R} = P x Q^T \approx R \quad (8)$$

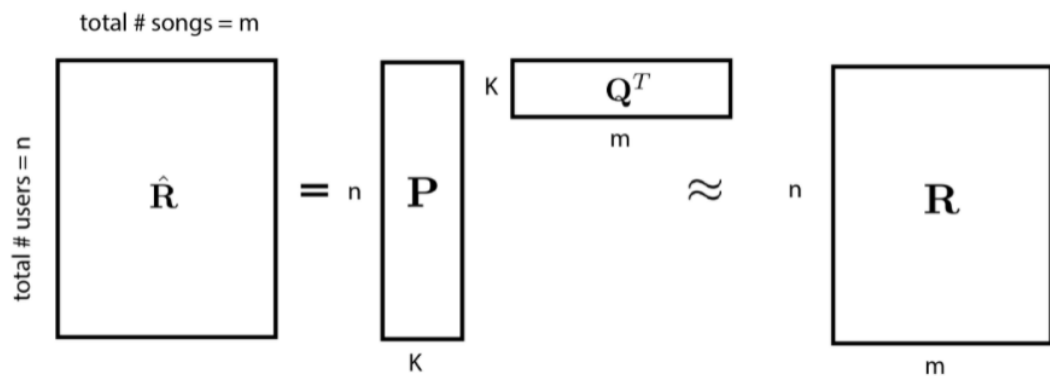
เมื่อ \hat{R} = ค่าเมทริกซ์อันดับต่ำ

P = เมทริกซ์ขนาด $n \times k$

Q^T = เมทริกซ์ขนาด $k \times m$

R = เมทริกซ์ของผู้ใช้งานทั้งหมด

จากภาพประกอบที่ 30 แสดงให้เห็นว่า P มีขนาด $n \times k$ และ Q^T มีขนาด $k \times m$ สามารถประมาณเมทริกซ์ R ด้วย $\hat{R} = P x Q^T$ จะพบว่าเมทริกซ์ P มีจำนวนแถวเท่ากับจำนวนผู้ใช้งาน เมทริกซ์ในแต่ละแถวสามารถบอกได้ว่าผู้ใช้งานคนใดมีลักษณะการฟังเพลงที่คล้าย ๆ กัน นอกจากนั้นสามารถนำแถวหนึ่งจากเมทริกซ์ P มาคูณกับเมทริกซ์ Q จะบอกได้ว่าผู้ใช้งานควร จะฟังเพลงอะไรเพิ่มเติมนอกเหนือจากที่เคยฟังมา



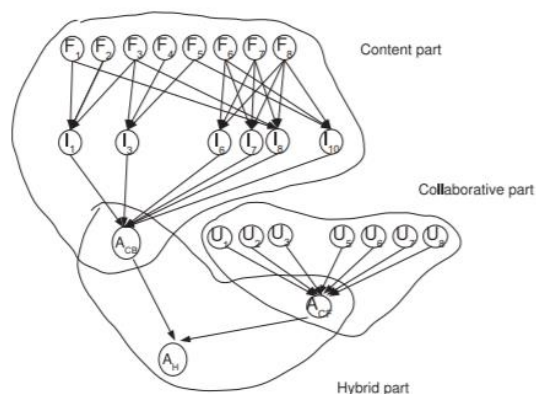
ภาพประกอบ 30 Non-Negative Matrix สำหรับเทคนิคการกรองข้อมูลแบบฟังพาผู้ใช้ร่วม
ที่มา: <https://tupleblog.github.io/spotify/>

2.9 งานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้ได้ทำการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องกับเทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering) และเทคนิคการกรองข้อมูลแบบฟังพาผู้ใช้ร่วม (Collaborative Filtering) โดยงานวิจัยที่เกี่ยวข้องมีรายละเอียดดังนี้

(1) บทความวิจัยเรื่อง Combining content-based and collaborative recommendations A hybrid approach based on Bayesian networks (de Campos, Fernández-Luna, Huete, & Rueda-Morales, 2010)

ระบบแนะนำช่วยให้ผู้ใช้งานสามารถเข้าถึงผลิตภัณฑ์หรือหัวข้อต่าง ๆ ที่อาจไม่ได้คำนึงถึง เพราะมีข้อมูลที่มีอยู่มากในอินเทอร์เน็ต ซึ่งการทำระบบแนะนำ ที่นิยมใช้จะมี 2 เทคนิค คือ เทคนิคการกรองแบบอิงเนื้อหา และ เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้งาน ซึ่งแต่ละวิธีการจะมีข้อดีและข้อเสียแตกต่างกันไป ในบางครั้งสามารถนำทั้ง 2 เทคนิคนี้มาใช้งานร่วมกันได้ เพื่อพัฒนาคุณภาพของการทำระบบแนะนำ โดยเครือข่ายของ Bayesian มีการใช้งานอย่างกว้างขวางและได้ผลลัพธ์ที่ดีในการนำมาใช้กับปัญหาที่มีความไม่แน่นอนสูง ซึ่งในส่วนของระบบแนะนำจึงมีความน่าสนใจมากที่จะนำหลักการของ Bayesian มาใช้เพื่อให้เกิดประโยชน์ งานวิจัยนี้นำเสนอเกี่ยวกับโมเดลของเครือข่ายแบบ Bayesian ที่นำมาใช้เพื่อจัดการกับปัญหาของระบบแนะนำแบบผสมผสาน ที่มีการใช้ทั้งเทคนิคการกรองแบบอิงเนื้อหา และ เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้งาน ร่วมกันซึ่งผู้วิจัยจะปรับใช้ให้โมเดลมีประสิทธิภาพในการประมาณการแจกแจงความน่าจะเป็น เพื่อให้โมเดลสามารถอนุมานความน่าจะเป็นได้ จากภาพประกอบ 31 ข้อมูลของผู้ใช้งานจะถูกใช้เพื่อทำนายว่า ผู้ใช้งานจะมีการให้คะแนนสินค้าแต่ละอย่างเท่าไร ในระบบแนะนำแบบผสมผสานของผู้วิจัยได้ใช้ทั้งวิธีของ เทคนิคการกรองแบบอิงเนื้อหา และ เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้งาน มาสร้างเป็นระบบแนะนำแบบผสมผสานที่สมบูรณ์มากยิ่งขึ้น ซึ่งข้อมูลที่ใช้ในนั้นจะมาจากข้อมูลของ Movie-Lens และ IMDB



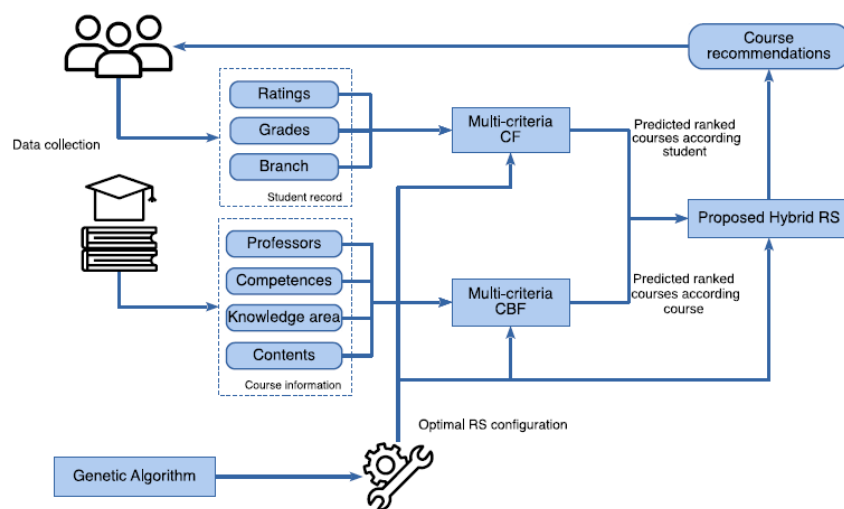
ภาพประกอบ 31 กราฟย่อยของเครือข่าย Bayesian แบบผสมผสาน
ที่มา : (de Campos, Fernández-Luna, Huete, & Rueda-Morales, 2010)

(2) บทความวิจัยเรื่อง Generating Items Recommendations by Fusing Content and User-Item based Collaborative Filtering (Tewari, 2019)

ปัจจุบันอีคอมเมิร์ซได้แพร่กระจายไปทั่วโลก ร้านค้าอิเล็กทรอนิกส์สามารถมีสินค้าได้หลายร้อยหรือหลายพันรายการโดยไม่มีขอบเขต ซึ่งข้อมูลผลิตภัณฑ์เหล่านี้ทั้งหมดมีอยู่บนอินเทอร์เน็ต อีกทั้งยังมีข้อมูลของลูกค้าเป็นจำนวนมาก ระบบแนะนำจะค้นหาความสนใจของผู้ใช้งาน และแนะนำรายการที่ตรงกับความต้องการของผู้ใช้งานมากที่สุด เทคนิคที่ใช้ในการแนะนำสินค้าคือเทคนิคการกรองแบบอิงเนื้อหาและ เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม แนะนำรายการที่มีลักษณะคล้ายคลึงกับสินค้าที่ผู้ใช้งานเคยใช้ในอดีต ในส่วนของเทคนิค เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม สร้างกลุ่มผู้ใช้ที่คล้ายกันและแนะนำรายการให้กับผู้ใช้งาน จุดประสงค์หลักของงานวิจัยนี้คือเพื่อลดจำนวนของสินค้าที่จะนำมาแนะนำลง เพื่อช่วยในการจัดวางสินค้าที่มีความสัมพันธ์กันให้ดียิ่งขึ้น ซึ่งผลการทดลองแสดงให้เห็นว่างานที่ทำขึ้นนั้น ดีกว่าวิธีการแนะนำแบบทั่วไป

(3) บทความวิจัยเรื่อง Helping university students to choose elective courses by using a hybrid multi-criteria recommendation system with genetic optimization. (Esteban, Zafra, & Romero, 2020)

การศึกษาในระดับมหาวิทยาลัยมักจะมีหลักสูตรรายวิชาเลือกวิชาต่าง ๆ ซึ่งจะถูกลีอกจากนักเรียนจำนวนมากในมหาวิทยาลัย นักเรียนจะใช้เวลาจำนวนมากในการค้นหาข้อมูลเกี่ยวกับหลักสูตรที่แตกต่างกันเพื่อการตัดสินใจเลือกรายวิชาให้เหมาะสมกับตัวเองมากที่สุด ในการตัดสินใจของนักเรียนนั้นอาจจะไม่ดีพอเนื่องจากนักเรียนมีข้อมูลไม่เพียงพอ ดังนั้นโดยทั่วไปนักเรียนอาจได้รับอิทธิพลจากความคิดเห็นของนักศึกษาคนอื่น หรืออาจจะมี ความชอบและความสนใจของนักเรียนมาเป็นเกณฑ์ในการพิจารณาด้วย วิธีในการแก้ไขปัญหานี้คือการใช้ระบบการแนะนำ (Recommendation System) ที่ช่วยให้นักเรียนตัดสินใจได้ดี โดยปรับตามความต้องการเฉพาะของตนเอง เทคนิคในการทำระบบแนะนำที่นิยมใช้คือ เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมและ เทคนิคการกรองแบบอิงเนื้อหา จากภาพประกอบ 32 คือขั้นตอนของระบบการแนะนำหลักสูตรแบบผสมผสาน โดยเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม แนะนำรายการที่อิงตามการให้คะแนนของผู้ใช้ที่คล้ายกันในขณะที่ เทคนิคการกรองแบบอิงเนื้อหา แนะนำรายการที่อ้างอิงเนื้อหาของรายการที่คล้ายกันในโปรไฟล์ผู้ใช้ โดยจะนำทั้งสองอัลกอริทึมนี้มาทำการแนะนำแบบผสมผสานเพื่อเพิ่มประสิทธิภาพในการแนะนำหลักสูตรรายวิชาเรียนให้กับนักเรียนได้ดีขึ้น

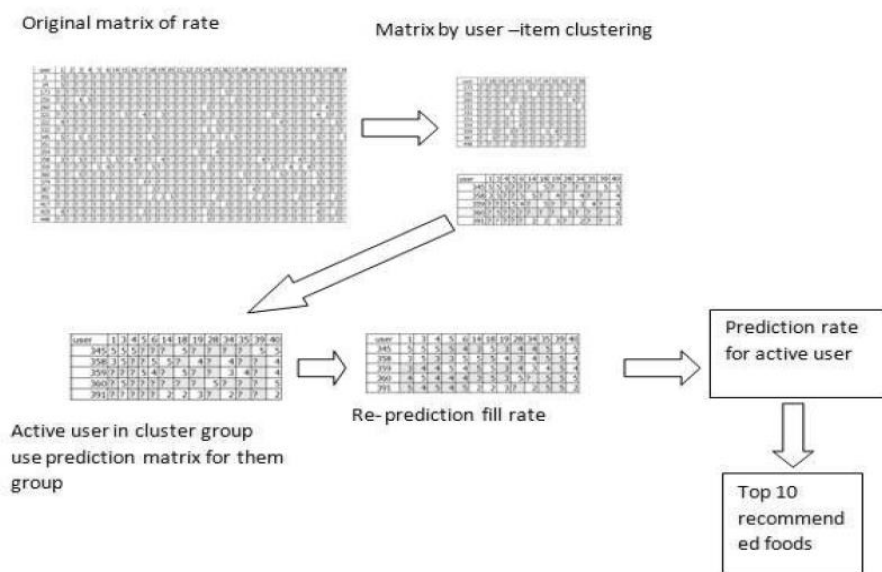


ภาพประกอบ 32 ขั้นตอนระบบการแนะนำหลักสูตรแบบผสมผสาน
ที่มา : (Esteban et al., 2020)

(4) บทความวิจัยเรื่อง A healthy food recommendation system by combining clustering technology with the weighted slope one predictor (Bundasak, 2017)

อาหารมีความสำคัญต่อการดำรงชีวิตมากขึ้นและอาหารที่มีความหลากหลาย สามารถเลือกได้ ดังนั้นเมื่อต้องการรับประทานอาหารเช้าเป็นเรื่องยากที่จะตัดสินใจว่าจะรับประทานอาหารเช้าในแต่ละมือของแต่ละวัน โดยเฉพาะอย่างยิ่งผู้ที่เป็นโรคหรือผู้ที่มีความเสี่ยงต่อการเกิดโรค การรับประทานอาหารเช้าที่ดีต่อสุขภาพจึงนับเป็นสิ่งที่สำคัญอย่างมาก วัตถุประสงค์ในการทำวิจัยเพื่อพัฒนาระบบแนะนำที่มุ่งเน้นไปที่สุขภาพของผู้ใช้งาน จากภาพประกอบ 33 ขั้นตอนของกระบวนการวิจัยแบ่งออกเป็น 5 ขั้นตอนดังนี้ 1. เพื่อจัดกลุ่มผู้ใช้ในการกรองร่วมกันและค้นหาผู้ใช้ที่มีลักษณะคล้ายกัน 2. สร้างเมตริกซ์ข้อมูลกลุ่มผู้ใช้งานแบบคลัสเตอร์โดยการค้นหาค่าที่หายไปด้วยอัลกอริทึมความชันแบบถ่วงน้ำหนัก 3. ในการเลือกค่าของเมตริกซ์ทำนาย rate ของรายการต่าง ๆ จะขึ้นอยู่กับอัลกอริทึมของ weighted slope one 4. พิจารณาผลลัพธ์ของรายการที่ทำกราคาดการณ์เพื่อค้นหาสิบอันดับแรกของรายการไอเท็ม 5. ติดตามและประเมินผลเพื่อเปรียบเทียบค่าที่คาดการณ์โดยอัลกอริทึมแบบถ่วงน้ำหนักแบบถ่วงน้ำหนักบนระบบคำแนะนำการกรองแบบทำงานร่วมกันทั่วไปและระบบผู้แนะนำแบบไฮบริดซึ่งรวมเทคนิคการทำคลัสเตอร์ในการกรองและการเติมร่วมกัน การใช้ระบบที่แนะนำร่วมกับการชุดข้อมูลโดยการวิเคราะห์ข้อมูลจากการจัดเก็บ

จนถึงความพึงพอใจของอาหารและพฤติกรรมการกิน การศึกษาพฤติกรรมที่คาดว่าจะจะเป็นปัจจัยในการก่อโรคของโรคเป็นสิ่งสำคัญที่ต้องเน้นว่าคนส่วนใหญ่กำลังเป็นอยู่ และโรคเหล่านี้อาจส่งผลกระทบต่ออาหารเมื่อผู้คนเหล่านั้นรับประทานเข้าไป



ภาพประกอบ 33 ระบบแนะนำอาหารแบบผสมผสาน

ที่มา : (Bundasak, 2017)

(5) บทความวิจัยเรื่อง Combining Content-based and Collaborative Filtering for Personalized Sports News Recommendations (Philip Lenhart, 2016)

ในการนำเสนอข่าวกีฬา ผู้อ่านมักจะมาพร้อมกับอารมณ์ที่แนบแน่นกับทีมกีฬาหรือผู้เล่น นอกจากนี้ความสนใจในหัวข้อของข่าวนั้นสามารถเปลี่ยนแปลงได้อย่างรวดเร็ว เช่น มีการแข่งขันกีฬาที่สำคัญเกิดขึ้นในงานนี้ ได้นำเสนอระบบผู้แนะนำข่าวกีฬาแบบไฮบริดที่รวมคำแนะนำตามเนื้อหาเข้ากับการกรองแบบร่วมมือ มีพัฒนาแดชบอร์ดผู้แนะนำและรวมเข้ากับเว็บไซต์ Sport1.de จากภาพประกอบ 20 การพัฒนาวิดเจ็ตแดชบอร์ดซึ่งสามารถรวมเข้าด้วยกันได้ ในเว็บไซต์ที่มีอยู่เพื่อให้คำแนะนำข่าวกีฬาส่วนบุคคล โดยแสดงภาพหน้าจอบริเวณปัจจุบันของแดชบอร์ดผู้แนะนำของเรา มีการนำเสนอคำแนะนำแก่รายการในครั้งเดียว เมื่อเลื่อนเมาส์ไปเหนือบทความหนึ่งผู้ใช้งานสามารถอ่าน ("Ansehen") หรือปฏิเสธคำแนะนำ ("Entfernen") ในการศึกษาผู้ใช้งานได้มีการประเมินวิธีแก้ปัญหาซึ่ง ผลลัพธ์แสดงให้เห็นว่าแนวทางที่เป็นฐานเนื้อหาล้วนให้คำแนะนำข่าวสาร

ที่ถูกต้องและผู้ใช้งานยืนยันว่าแดชบอร์ดของระบบแนะนำ มีความสามารถในการใช้งานสูง อย่างไรก็ตามองค์ประกอบการรองรับการทำงานร่วมกันของแนวทางผสมผสานมีความจำเป็นเพื่อเพิ่มความหลากหลายของระบบแนะนำ

IHR SPORT1 DASHBOARD

IHRE SPORT1 ARTIKEL-EMPFEHLUNGEN




Guardiola zieht es in die Innenstadt

Pep Guardiola wird in Manchester wohl eine Wohnung in der Innenstadt beziehen. Einen vornehmen Wohnort, wie ihn etwa Jose Mourinho bevorzugt, will Guardiola meiden. [mehr >](#)



Ronaldo leidet an Schlafproblemen

Sein verlorkester EM-Start und der verschossene Elfmeter gegen Österreich machen Portugals Superstar Cristiano Ronaldo schwer zu schaffen. Sein Trainer vertraut ihm blind. [mehr >](#)



Das erwartet Sie heute bei der EM

Die Entscheidung in der Gruppe B steht an, die Tabellensituation verspricht Hochspannung. England braucht einen Punkt, Wales wohl einen Sieg. SPORT1 hat die Infos. [mehr >](#)



Ronaldo-Selfie ruft UEFA auf den Plan

Erneut geraten bei der EM in Frankreich drei Teams wegen ihrer Anhänger ins Visier der UEFA: Auch das Selfie eines Portugal-Fans mit Ronaldo ruft ein Verfahren hervor. [mehr >](#)




Die Gruppe B im SPORT1-Check

England hat die vielversprechendste Mannschaft seit Jahrzehnten und ist Favorit in EM-Gruppe B. Russland droht gegen Wales und die Slowakei ein peinliches Aus. [mehr >](#)




Bierhoff: Löw mit jedem Turnier abgeklärter

Selbst ein unvorethafter Griff in die Hose kann Joachim Löw nicht aus dem Konzept bringen. Dem Teammanager gefällt die Coolness und Ruhe des Bundestrainers. [mehr >](#)



Rätsel Ronaldo: Nur das Selfie sitzt

Verschossener Elfmeter, nicht gegebenes Abseitstor: Cristiano Ronaldo erlebt gegen Österreich einen gebrauchten Tag. Seine wohl größte Szene hat er nach dem Spiel. [mehr >](#)



Del Bosque denkt schon ans Finale

Mit dem Sieg gegen die Türkei legt Spanien einen perfekten Start in die EM hin. Trainer Vicente del Bosque spricht bereits jetzt vom Finale. [mehr >](#)



Nordirland rechnet sich Chancen aus

Nordirland zeigt sich vor dem abschließenden Gruppenspiel gegen Deutschland durchaus mutig. Der Kultsong über Will Grigg beginnt die Spieler zu nerven. [mehr >](#)

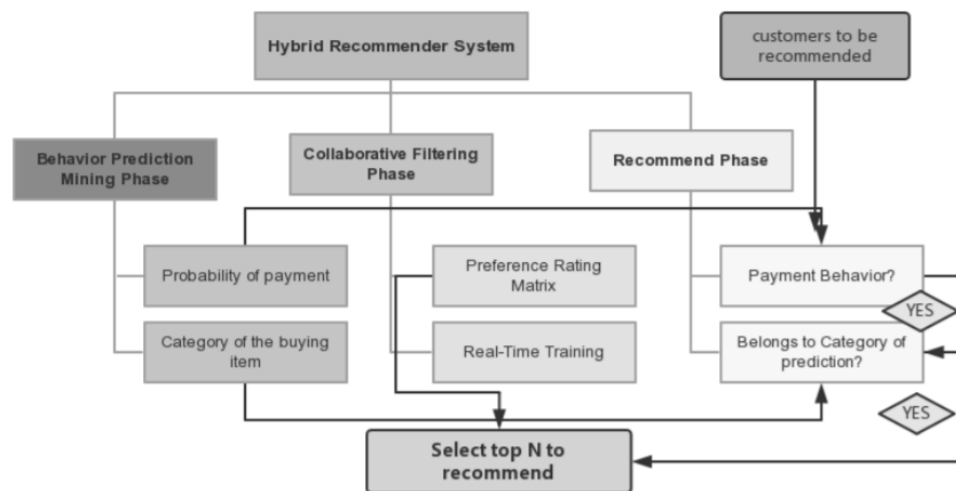
ภาพประกอบ 34 แดชบอร์ดของระบบแนะนำ

ที่มา : (Philip Lenhart, 2016)

(6) บทความวิจัยเรื่อง HYBRID RECOMMENDER SYSTEM BASED ON PERSONAL BEHAVIOR MINING(Zhiyuan Fang, 2016)

ระบบแนะนำเป็นที่รู้จักอย่างแพร่หลายจากเว็บไซต์อีคอมเมิร์ซ และส่วนใหญ่จะเป็นในรูปแบบของแบบจำลอง อัลกอริทึมที่ใช้แนะนำความชอบส่วนบุคคล ที่รวมถึงการแนะนำแบบพึ่งพาผู้ใช้ร่วมจาก item-based ที่ใช้ใน Amazon การแนะนำแบบพึ่งพาผู้ใช้ร่วมจาก matrix factorization based ที่ใช้กับ Netflix และวิธีอื่นอีกมากมาย ในบทความนี้ผู้วิจัยจะทำการรวม

แบบจำลองแบบเดิม ๆ กับวิธีแบบ pattern extraction ซึ่งใช้ข้อมูล desensitized mobile transaction record ที่ได้มาจาก T-mal ของ Alibaba จากภาพประกอบ 35 จะทำการบันทึกการทำธุรกรรมซึ่งจัดทำโดย T-mall, Alibaba ซึ่งมีข้อมูลการซื้อในอดีตของผู้ใช้งาน โดยจะแบ่งพฤติกรรมที่แตกต่างกันของผู้บริโภคออกเป็น 4 ประเภทได้แก่ คลิก เลือกร หยิบใส่รถเข็นและชำระเงิน เพื่อสร้างระบบแนะนำแบบผสมผสานที่แนะนำได้อย่างหลากหลาย sequential pattern mining มีจุดประสงค์เพื่อหาความถี่ของ รูปแบบตามลำดับใน ลำดับของฐานข้อมูล และจะนำไปใช้ในแบบจำลองแบบผสมผสานเพื่อใช้ทำนาย พฤติกรรมการจ่ายเงินของลูกค้า และมีส่วนในการเพิ่มความถูกต้องของแบบจำลอง



ภาพประกอบ 35 โครงสร้างของระบบแนะนำแบบผสมผสาน

ที่มา : (Zhiyuan Fang, 2016)

บทที่ 3

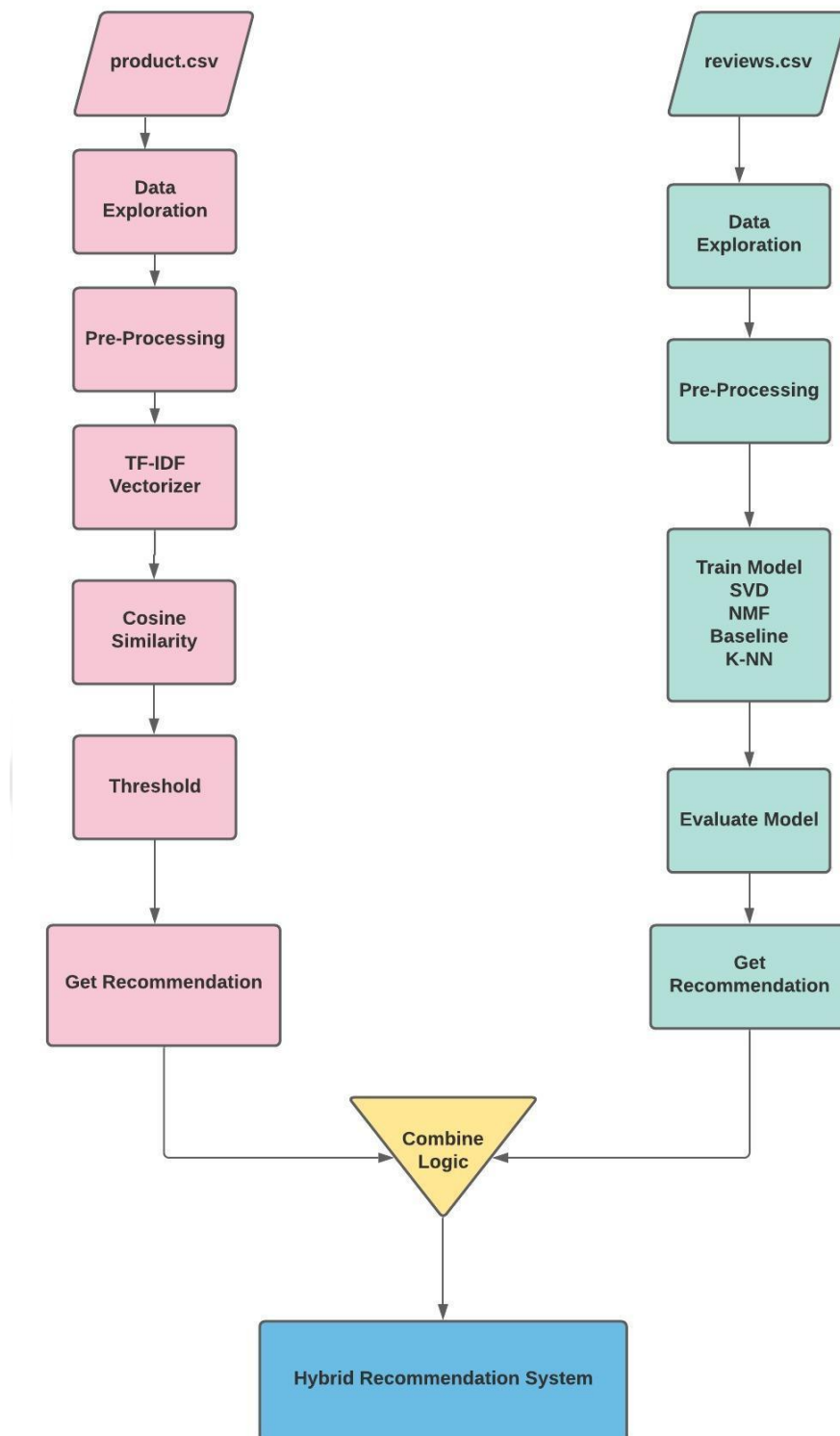
การดำเนินการวิจัย

ในการดำเนินการวิจัย ผู้วิจัยได้ดำเนินการดังต่อไปนี้

1. กระบวนการสร้างแบบจำลอง
2. แผนการดำเนินการวิจัย
3. ข้อมูลที่นำมาใช้ในงานวิจัย
4. การสำรวจข้อมูล
5. การเตรียมข้อมูล
6. อัลกอริทึมที่ใช้ในการทำนาย
7. ประเมินประสิทธิภาพแบบจำลอง
8. สรุปผลการดำเนินงาน

3.1 กระบวนการสร้างแบบจำลอง

ในการสร้างระบบแนะนำ ผู้วิจัยจะทำการสร้างแบบจำลองทั้งหมด 3 ประเภท ได้แก่ 1. เทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering) ที่ใช้ชุดข้อมูลจากไฟล์ product.csv มาใช้สร้างแบบจำลอง โดยจะมีการตรวจสอบความสมบูรณ์ของข้อมูลเบื้องต้น ได้มีการใช้หลักการของ TF-IDF และ การหาความคล้ายคลึงของโคไซน์ เพื่อสร้างเป็นคำแนะนำให้กับลูกค้า 2. การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) ที่ใช้ชุดข้อมูลจากไฟล์ reviews.csv มาสร้างเป็นแบบจำลอง โดยใช้วิธีการแบบ Model-based ในการสร้างแบบจำลอง จาก 4 อัลกอริทึม ได้แก่ SVD, NMF, Baseline และ K-NN เพื่อมาเปรียบเทียบประสิทธิภาพ โดยมีเมตริกซ์ในการวัดประสิทธิภาพคือค่า RMSE และ MAE ซึ่งแบบจำลองจากทั้ง 4 อัลกอริทึมนั้นจะสามารถสร้างเป็นคำแนะนำให้กับลูกค้าได้ทั้งหมด 3. ระบบแนะนำแบบผสมผสาน (Hybrid Recommendation System) คือการนำทั้ง 2 วิธีการก่อนหน้านี้มาทำงานร่วมกัน โดยในส่วนของ การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม จะเลือกอัลกอริทึมที่มีความน่าสนใจ และให้ผลลัพธ์ที่ดีที่สุดมา ใช้ร่วมกับ เทคนิคการกรองแบบอิงเนื้อหา และทำการแนะนำสินค้าให้กับลูกค้า โดยขั้นตอน ทั้งหมดของการสร้างแบบจำลองทั้ง 3 ประเภทสามารถดูรายละเอียดได้จากภาพประกอบ 36



ภาพประกอบ 36 กระบวนการทำงานของการสร้างแบบจำลองแต่ละประเภท

3.2 แผนการดำเนินการวิจัย

ตาราง 1 แผนการดำเนินการวิจัย

แผนการดำเนินการวิจัย	2563				
	สค.	กย.	ตค.	พย.	ธค.
1 ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง					
2 ศึกษาข้อมูลของชุดข้อมูล					
3 การสำรวจข้อมูล					
4 การเตรียมข้อมูล					
5 ดำเนินการทดลอง					
6 สรุปผลการทดลอง					

3.3 ข้อมูลที่นำมาใช้ในงานวิจัย

การสร้างแบบจำลองใช้ชุดข้อมูล Ice Cream Dataset (Pond, 2020) ซึ่งประกอบด้วยข้อมูลของสินค้าและรีวิวการให้คะแนนของไอศกรีม 4 ยี่ห้อ ได้แก่ Ben & Jerry's, Haagen-Dazs, Breyers และ Talenti ทั้งหมดจำนวน 241 รสชาติ ในรูปแบบของไฟล์ csv จำนวน 2 ไฟล์ ได้แก่ ชุดข้อมูล products.csv คือชุดข้อมูลที่อธิบายเกี่ยวกับชื่อรสของไอศกรีม คำอธิบายของรสชาติ ระดับคะแนนรีวิวจาก 1-5 คะแนน และส่วนผสมของไอศกรีม โดยมีข้อมูลทั้งหมด 241 แถว ส่วนชุดข้อมูล reviews.csv ประกอบด้วยข้อมูลรีวิวของสินค้าจำนวนทั้งหมด 21,674 แถว

3.4 การสำรวจข้อมูล

การวิจัยครั้งนี้ทางผู้วิจัยได้นำข้อมูล Ice Cream Dataset จาก Kaggle โดยใช้ไฟล์ข้อมูลจำนวน 2 ชุด ได้แก่ 1. ชุดข้อมูล product.csv ประกอบด้วยข้อมูลจำนวน 7 คอลัมน์ ดังตาราง 2 ทางผู้วิจัยได้นำมาทดสอบโดยใช้เทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering) โดยใช้ข้อมูลทั้งหมด 3 คอลัมน์ ได้แก่ รหัสสินค้า (key) ชื่อสินค้า (name) คำโฆษณาสินค้า (description) มาทำการวิเคราะห์ในการแนะนำสินค้าให้กับลูกค้า 2. ชุดข้อมูล reviews.csv ประกอบด้วยข้อมูลจำนวน 7 คอลัมน์ ดังตาราง 3 ชุดข้อมูลนี้ทางผู้วิจัยได้นำมาทดสอบโดยใช้ใน ส่วนเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้งาน (Collaborative Filtering) โดยใช้ข้อมูลทั้งหมด 3 คอลัมน์ ได้แก่ รหัสสินค้า (key) ชื่อผู้ใช้งาน (author) การให้คะแนนจากผู้ใช้งานจาก 1-5 คะแนน (rating)

ตาราง 2 โครงสร้างข้อมูลสินค้า (product.csv)

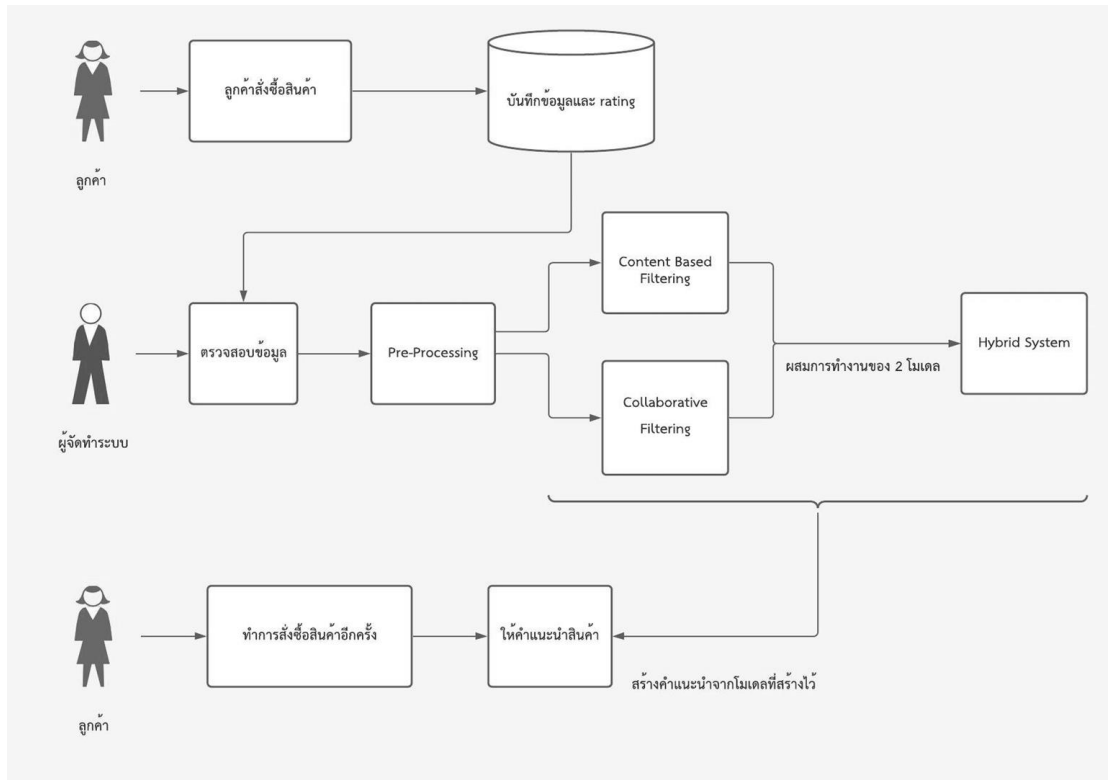
ลำดับ	ข้อมูลตัวแปร(Variable)	คำอธิบายข้อมูล(Description)
1	key	รหัสสินค้า
2	name	ชื่อสินค้า
3	subhead	คำบรรยายไอศกรีม
4	description	คำโฆษณาสินค้า
5	rating	การให้คะแนนจากผู้ใช้งาน จาก 1-5 คะแนน
6	rating count	จำนวนผู้ให้คะแนน
7	ingredients	ส่วนประกอบไอศกรีม

ตาราง 3 โครงสร้างข้อมูลความคิดเห็น (reviews.csv)

ลำดับ	ข้อมูลตัวแปร(Variable)	คำอธิบายข้อมูล(Description)
1	key	รหัสสินค้า
2	author	ชื่อผู้ใช้งาน
3	date	วัน เดือน ปี ที่ให้คะแนนรีวิว
4	star	การให้คะแนนจากผู้ใช้งาน จาก 1-5 คะแนน
5	title	ข้อความรีวิว
6	helpful yes	จำนวนผู้ที่ชื่นชอบการให้รีวิว
7	helpful no	จำนวนผู้ที่ไม่ชื่นชอบการให้รีวิว

ระบบเริ่มเก็บข้อมูลจากการที่ลูกค้าทำการสั่งซื้อสินค้า ซึ่งเก็บข้อมูลที่สำคัญ ได้แก่ ชื่อลูกค้าผู้ใช้งาน ประเภทสินค้า ระดับคะแนนความพึงพอใจที่ได้ทำการให้คะแนน ข้อมูลเหล่านี้จะถูกนำมาสำรวจและวิเคราะห์ข้อมูลเบื้องต้น เพื่อหาความสัมพันธ์ของข้อมูล เช่น เมนูที่มี rating ดีที่สุดของแต่ละแบรนด์ เป็นต้น และในขั้นตอนการทำ pre-processing จะมีการทำข้อมูลให้พร้อมใช้งานมากขึ้น เช่น กำจัด missing value หรือลบคอลัมน์ที่ไม่จำเป็นต่อการสร้างแบบจำลองออก จากนั้นนำข้อมูลที่พร้อมใช้มาทำการสร้างแบบจำลอง โดยจะใช้เทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering) เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering)

และการทำงานร่วมกันแบบผสมผสาน (Hybrid Recommendation System) ระบบจะทำการแนะนำสินค้าให้กับผู้ใช้งานเมื่อทำการสั่งซื้อสินค้าอีกครั้ง ดังภาพประกอบ 37



ภาพประกอบ 37 การทำงานของระบบแนะนำสินค้า

3.5 การเตรียมข้อมูล

ข้อมูลไฟล์ product.csv นำมาใช้สำหรับเทคนิคการกรองแบบอิงเนื้อหา โดยเริ่มต้นนั้นได้เลือกคอลัมน์ที่เกี่ยวข้องกับสินค้ามาใช้ในการวิเคราะห์ ได้แก่ 1) key 2) name 3) description ซึ่ง description ของสินค้านั้นมีรายละเอียดของสินค้าที่สามารถนำมาใช้วัดความใกล้เคียงของสินค้า เช่น salty, fudge และ chocolate เป็นต้น ทำให้ทราบรายละเอียดถึงสินค้านั้น ๆ จึงสามารถนำมาเปรียบเทียบความคล้ายคลึงของสินค้าได้ ตัวอย่างข้อมูล product.csv ดังตาราง 4 และสำหรับข้อมูล reviews.csv นั้นได้นำคอลัมน์ 1) key 2) name 3) star ซึ่งคอลัมน์ star จะเป็นตัวกำหนดระดับความชื่นชอบของลูกค้าที่สามารถนำมาใช้ในการวิเคราะห์สำหรับเทคนิคการกรองข้อมูลแบบพึ่งพาผู้เข้าร่วม ตัวอย่างข้อมูล review.csv แสดงในตาราง 5

ตาราง 4 ตัวอย่างข้อมูลในไฟล์ product.csv

Key	Name	Description
0_bj	Salted Caramel Core	brownies, you'll be in total control of your own ice cream destiny.
1_bj	Netflix & Chill'd	There's something for everyone to watch on Netflix & flavors for everyone to enjoy from Ben & Jerry's, so we've teamed up to bring you a chillaxing new creation that's certain to satisfy any sweet or salty snack craving. It's a flavorful world, and everyone is invited to grab a spoon.
3_bj	Chip Happens	Sometimes chip happens and everything's a mess, but we Nailed It! with this chip-filled limited batch. When smooth chocolate ice cream meets fudge chips & salty swirls, they pack a serious one-two crunch. The best part? There won't be anything left to clean up.
4_bj	Cannoli	As a Limited Batch that captured the rapture of the classic Sicilian dessert, our Cannoli captivated fans like no other Cannoli could. Now that it's a full-time flavor, you and your Cannoli can re-capture the rapture all over again.
5_bj	Gimme'S more!	It's a gimme: there's always room for s'more. And we're pretty sure you'll want to make s'more room in your freezer for this shmallowy-rich, graham-good-n-chocolatey concoction now that it's full time flavor!

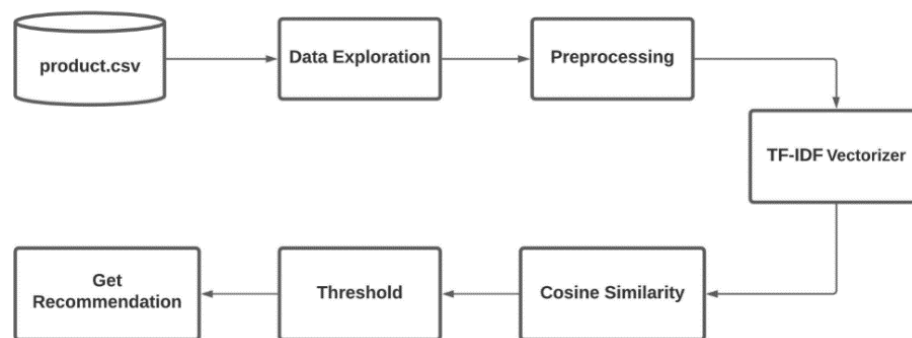
ตาราง 5 ตัวอย่างข้อมูลในไฟล์ reviews.csv

key	author	stars
0_bj	llovebennjerry	3
0_bj	Sweettooth909	5
0_bj	LaTanga71	3
0_bj	chicago220	5
0_bj	Kassidyk	1

3.6 การสร้างแบบจำลองข้อมูล

3.6.1 เทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering)

การสร้างแบบจำลองเทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering) จากภาพประกอบ 38 เริ่มต้นจากการนำชุดข้อมูล products.csv มาทำการสำรวจและวิเคราะห์ข้อมูล เพื่อเตรียมข้อมูลให้พร้อมที่จะนำไปใช้งาน และทำการแตกเวกเตอร์ของคำโดยใช้ TF-IDF ซึ่งวิธีการของ TF-IDF นั้นจะเป็นวิธีในการวัดน้ำหนักของคำแต่ละคำ ดังตัวอย่างภาพประกอบ 39 โดยคอลัมน์คือคำต่างๆทั้งหมดที่อยู่คอลัมน์ของ description ซึ่งจะได้คำที่เป็นค่าตัวเลขเพื่อนำมาหาค่าความคล้ายคลึงกัน โดยใช้โคไซน์และช่วยในการแปลงข้อความเป็นตัวเลขเพื่อใส่ลงในเวกเตอร์ของคำ และนำมาสร้างค่าความคล้ายคลึงแบบโคไซน์จากเวกเตอร์ที่ได้ จากภาพประกอบ 40 ซึ่งเป็นตัวอย่างโค้ดในกระบวนกรนี้ หลังจากนั้นจึงได้มีการสร้างคำแนะนำจากค่าความคล้ายคลึงแบบโคไซน์ที่ได้สูงที่สุด 10 อันดับ โดยดูเพิ่มเติมได้จากภาคผนวก ก



ภาพประกอบ 38 การสร้างแบบจำลองของเทคนิคการกรองแบบอิงเนื้อหา

	100	110	130	140	150	1866	1984	1985	260	270	...	wrapped	wrong	www	year	years	yes	yolks	york	yurt	zest
0	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0
1	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0
2	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0
3	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0
4	0.000000	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0
...
236	0.046036	0.0	0.0	0.0	0.072637	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.074215	0.0	0.0	0.0	0.0	0.0
237	0.051342	0.0	0.0	0.0	0.081008	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.082768	0.0	0.0	0.0	0.0	0.0
238	0.072325	0.0	0.0	0.0	0.114115	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.116595	0.0	0.0	0.0	0.0	0.0
239	0.047840	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0
240	0.062835	0.0	0.0	0.0	0.099142	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.101297	0.0	0.0	0.0	0.0	0.0

ภาพประกอบ 39 ตัวอย่างค่าเวกเตอร์วิธีการของ TF-IDF ที่ได้จากคอลัมน์ description

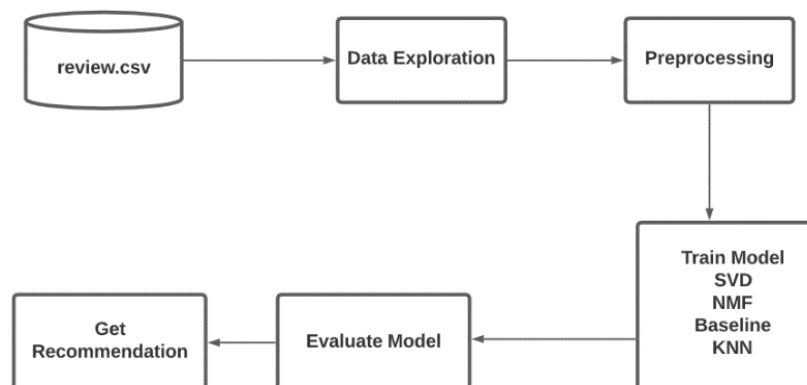
```
tfidf = TfidfVectorizer(stop_words='english')
tfidf_matrix = tfidf.fit_transform(product_df['description'])

cosine_sim = linear_kernel(tfidf_matrix, tfidf_matrix)|
```

ภาพประกอบ 40 ตัวอย่างโค้ดการทำ TF-IDF เพื่อนำมาสร้างเวกเตอร์ความคล้ายคลึงแบบโคไซน์

3.6.2 การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering)

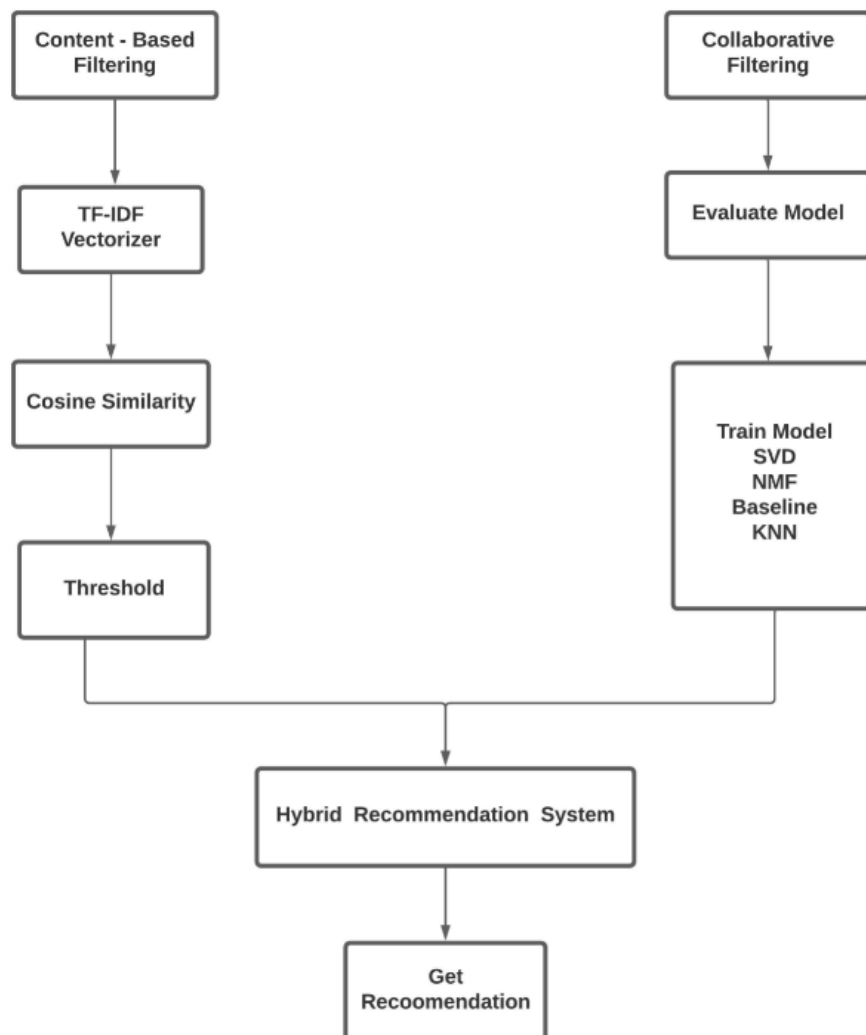
การสร้างแบบจำลองการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) จากภาพประกอบ 41 เริ่มต้นจากการนำชุดข้อมูล reviews.csv ทำการสำรวจ และวิเคราะห์ข้อมูล เช่น การกรองชื่อแบรนด์ไอศกรีมที่ไม่ซ้ำกัน หรือการลบคอลัมน์ที่ไม่ได้ใช้ในส่วนของการสร้างแบบจำลอง เพื่อเตรียมข้อมูลให้พร้อมที่จะนำไปใช้งาน หลังจากนั้นจะนำไลบรารีของ Scikit Learn ชื่อว่า Surprise ซึ่งเป็นไลบรารีสำหรับสร้างแบบจำลองการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม โดยเฉพาะ มาทำการสร้างแบบจำลอง 4 อัลกอริทึม ได้แก่ SVD, NMF, Baseline และ KNN โดยมีการแบ่งสัดส่วนของข้อมูลเป็น train 80 % และ test 20% และเปรียบเทียบประสิทธิภาพการทำงานของแต่ละอัลกอริทึม ในการสร้างแบบจำลองจะเลือกใช้คอลัมน์ author, key และ star ในการนำมาสร้างแบบจำลอง ดูเพิ่มเติมได้จากภาคผนวก ข โดยประเมินประสิทธิภาพของแบบจำลองจากค่า RMSE และ MAE ซึ่งทุกแบบจำลองสามารถนำมาสร้างคำแนะนำให้กับลูกค้าได้



ภาพประกอบ 41 การสร้างแบบจำลองของการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม

3.6.3 ระบบแนะนำแบบผสมผสาน (Hybrid Recommendation System)

ระบบแนะนำแบบผสมผสาน ได้นำเทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering) และเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) มาทำงานร่วมกัน โดยเริ่มต้นใช้ เทคนิคการกรองแบบอิงเนื้อหาทำการเลือกชื่อเมนูไอศกรีมที่มีความสัมพันธ์มากที่สุดจากการหาค่าความคล้ายคลึงของโคไซน์ และนำผลลัพธ์ที่ได้มาประมวลผลกับแบบจำลองของเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมที่ให้ผลลัพธ์การประเมินประสิทธิภาพของแบบจำลองจากค่า RMSE และ MAE ที่ดีที่สุด ดังภาพประกอบ 42 โดยดูรายละเอียดเพิ่มเติมได้จากภาคผนวก ค



ภาพประกอบ 42 การสร้างแบบจำลองของระบบแนะนำแบบผสมผสาน

3.7 การประเมินประสิทธิภาพแบบจำลอง

3.7.1 เทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering)

ในส่วนของเทคนิคการกรองแบบอิงเนื้อหาที่ได้มีการหาความคล้ายคลึงแบบโคไซน์ เมื่อผู้วิจัยได้ทำการสำรวจค่าความคล้ายคลึงแบบโคไซน์จากผลลัพธ์ที่ได้จึงมีการกำหนดเกณฑ์ขั้นต่ำในการแนะนำขึ้น

3.7.2 การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering)

การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมที่ได้มีการสร้างแบบจำลองจาก 4 อัลกอริทึมได้แก่ SVD, NMF, Baseline และ KNN ซึ่งมีการวัดผลด้วยค่า RSME และ MAE จะสามารถดูได้ดังตาราง 6 ซึ่งอัลกอริทึมที่ให้ผลลัพธ์ที่ดีที่สุดจะมีอยู่ 2 อัลกอริทึมคือ SVD และ Baseline ซึ่งมีค่าของเมตริกซ์แต่ละตัวใกล้เคียงกันมาก

3.7.3 การทำงานร่วมกันแบบผสมผสาน

สำหรับการวัดผลของระบบแนะนำแบบผสมผสานนั้น จะไม่ได้มีการวัดผลของตัวแบบจำลองชนิดนี้ที่ชัดเจน เนื่องจากระบบแนะนำแบบผสมผสานมีการใช้งานการทำงานของกรกรองแบบอิงเนื้อหาที่ไม่มีการวัดผลที่ชัดเจน มีแค่การสร้างเกณฑ์ขั้นต่ำขึ้นมาเท่านั้น ทำให้หากต้องการที่จะดูประสิทธิภาพของระบบแนะนำแบบผสมผสานที่ได้นั้น จะสามารถอิงได้จากประสิทธิภาพของการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมที่ทางผู้วิจัยได้ใช้อัลกอริทึม SVD มาใช้ในการสร้างระบบแนะนำแบบผสมผสาน เพียงอย่างเดียวเท่านั้น แต่ก็จะไม่สะท้อนประสิทธิภาพทั้งหมดของแบบจำลอง

3.8 สรุปผลการดำเนินงาน

เทคนิคการกรองแบบอิงเนื้อหาที่สร้างขึ้นมานั้น จะทำการสร้างคำแนะนำจากค่าความคล้ายคลึงของโคไซน์สูงสุด 10 อันดับ โดยวิธีการนี้ไม่ได้มีการวัดประสิทธิภาพที่ชัดเจน เพราะอิงจากค่าความคล้ายคลึงของโคไซน์เพียงค่าเดียว ส่วนการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม จะมีการพัฒนาแบบจำลองขึ้นมาจาก 4 อัลกอริทึม ได้แก่ KNN, SVD, Baseline และ NMF โดยใช้ไลบรารี Surprise ในการสร้างแบบจำลองซึ่งมีค่า RMSE และ MAE ให้ใช้ในการวัดประสิทธิภาพของแบบจำลองเช่นกัน ส่วนระบบแนะนำแบบผสมผสานจะใช้วิธีการการทำงานของก่อนหน้าทั้งเทคนิคการกรองแบบอิงเนื้อหา และการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมที่สร้างมาก่อนหน้านี้ ซึ่งเป็นระบบแนะนำแบบผสมผสานจะมีการใช้ค่าความคล้ายคลึงของโคไซน์มาใช้เพื่อเลือกเฉพาะบาง

เมนูที่มีค่าความคล้ายคลึงสูงสุด และนำเมนูเหล่านั้นมาทำงานร่วมกันกับแบบจำลองที่ได้จากการ
กรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมเพื่อสร้างคำแนะนำได้สูงสุด 10 อันดับ



บทที่ 4

ผลการดำเนินงานวิจัย

ในการวิจัยเพื่อศึกษาและพัฒนาระบบแนะนำสินค้าประเภทไอศกรีม จากการการทดลองครั้งนี้ทางผู้วิจัยได้ใช้เว็บจูปิเตอร์เน็ตเวิร์คสำหรับสร้างแบบจำลองการทำนาย 3 แบบ ซึ่งได้แก่

1. การกรองแบบอิงเนื้อหา (Content-Based Filtering)
2. การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering)
3. ระบบการแนะนำแบบผสมผสาน (Hybrid Recommendation System)

4.1 เทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering)

ผู้วิจัยได้ใช้วิธีการ TF-IDF เพื่อนำคำในคอลัมน์ description ของสินค้ามาสร้างเป็น feature vector เพื่อนำคำใน description เหล่านี้มาคำนวณหาค่าความคล้ายคลึงแบบโคไซน์ โดยผู้วิจัยมีการตรวจสอบค่า threshold ในช่วง 0.05 – 0.95 จึงพบว่าเมนูไอศกรีมส่วนใหญ่จะได้คำแนะนำออกมาก็ต่อเมื่อมีการตั้งค่าเกณฑ์ขั้นต่ำอยู่ที่ช่วง 0.15 ตามตัวอย่างตาราง 6 เมนู Berry Sweet Mascarpone ซึ่งจะมีผลลัพธ์ของการแนะนำรายการออกมาก็ต่อเมื่อต้องกำหนดเกณฑ์ขั้นต่ำไว้ที่ 0.15 เท่านั้น และตาราง 7 เมนู Delights Vanilla Bean ซึ่งก็แสดงให้เห็นว่ามีบางเมนูที่มีค่าความคล้ายคลึงสูงกว่า 0.85 เช่นกันแต่ก็เป็นส่วนน้อยจากที่ผู้วิจัยได้สำรวจมา จึงได้เลือกเกณฑ์ขั้นต่ำไว้ที่ 0.15 เพื่อรองรับกับเมนูที่เหลือ ซึ่งผลลัพธ์ที่ออกมานั้นจะลิสต์รายการแนะนำได้สูงสุดมา 10 อันดับ โดยดูเพิ่มเติมได้จากภาคผนวก ก

```
get_recommendations('Berry Sweet Mascarpone')
```

	name	sim_value
0	Cookies and Cream Ice Cream	0.178228
1	Peanut Butter World®	0.150411
2	Salted Caramel Almond	0.150411

ภาพประกอบ 43 ตัวอย่างชื่อเมนูที่ได้จากระบบแนะนำด้วยเทคนิคการกรองแบบอิงเนื้อหา

จากภาพประกอบ 43 แสดงให้เห็นถึงผลลัพธ์ของไอศกรีมรสชาติ Berry Sweet Mascarpone ระบบจะแนะนำสินค้าที่มีความใกล้เคียงกับไอศกรีมรสชาติดังกล่าว ซึ่งได้มาจากการนำ description ของสินค้ามาทำการหาค่าความคล้ายคลึงแบบโคไซน์ โดยอิงจากเกณฑ์ขั้นต่ำที่กำหนดไว้ก่อนหน้านี้

ตาราง 6 ผลลัพธ์การกำหนดค่าเกณฑ์ขั้นต่ำตั้งแต่ 0.05-0.95 เมนู Berry Sweet Mascarpone

Threshold	Name	Similarity
0.15	Cookies and Cream Ice Cream	0.178228
	Peanut Butter World	0.150411
	Salted Caramel Almond	0.150411
0.10	Cookies and Cream Ice Cream	0.178228
	Peanut Butter World®	0.150411
	Salted Caramel Almond	0.150411
	Wake & " No Bake " Cookie Dough Core	0.121098
	Coffee Almond Crunch Ice Cream Bar	0.115688
	Sweet Like Sugar Cookie Dough Core	0.114189
	Non-Dairy OREO® Cookies & Cream	0.104464
	Red, White & Blueberry	0.100273
0.05	Cookies and Cream Ice Cream	0.178228
	Peanut Butter World®	0.150411
	Salted Caramel Almond	0.150411
	Wake & " No Bake " Cookie Dough Core	0.121098
	Coffee Almond Crunch Ice Cream Bar	0.115688
	Sweet Like Sugar Cookie Dough Core	0.114189
	Non-Dairy OREO® Cookies & Cream	0.104464
	Red, White & Blueberry	0.100273
	Vanilla Blackberry Chocolate TRIO CRISPY LAYERS	0.098113
	Milk & Cookies	0.096126

ตาราง 7 ผลลัพธ์การกำหนดค่าเกณฑ์ขั้นต่ำตั้งแต่ 0.05-0.95 เมนู Delights Vanilla Bean

Threshold	name	similarity
0.85	Delights Cookies & Cream	0.871479
	Delights Creamy Chocolate	0.859610
0.80	Delights Cookies & Cream	0.871479
	Delights Creamy Chocolate	0.859610
0.75	Delights Cookies & Cream	0.871479
	Delights Creamy Chocolate	0.859610
0.70	Delights Cookies & Cream	0.871479
	Delights Creamy Chocolate	0.859610
	Delights Mint Chip	0.724395
0.65	Delights Cookies & Cream	0.871479
	Delights Creamy Chocolate	0.859610
	Delights Mint Chip	0.724395
0.60	Delights Cookies & Cream	0.871479
	Delights Creamy Chocolate	0.859610
	Delights Mint Chip	0.724395
0.55	Delights Cookies & Cream	0.871479
	Delights Creamy Chocolate	0.859610
	Delights Mint Chip	0.724395
0.50	Delights Cookies & Cream	0.871479
	Delights Creamy Chocolate	0.859610
	Delights Mint Chip	0.724395
0.45	Delights Cookies & Cream	0.871479
	Delights Creamy Chocolate	0.859610
	Delights Mint Chip	0.724395
0.40	Delights Cookies & Cream	0.871479
	Delights Creamy Chocolate	0.859610
	Delights Mint Chip	0.724395
0.35	Delights Cookies & Cream	0.871479
	Delights Creamy Chocolate	0.859610
	Delights Mint Chip	0.724395

ตาราง 7 (ต่อ)

Threshold	name	similarity
0.30	Delights Cookies & Cream	0.871479
	Delights Creamy Chocolate	0.859610
	Delights Mint Chip	0.724395
	Vanilla Chocolate	0.326861
0.25	Delights Cookies & Cream	0.871479
	Delights Creamy Chocolate	0.859610
	Delights Mint Chip	0.724395
	Vanilla Chocolate	0.326861
	Natural Vanilla	0.291182
	Chocolate	0.289863
	Vanilla Chocolate Strawberry	0.287427
	Natural Vanilla Snack Cups 10ct	0.263089
0.20	Delights Cookies & Cream	0.871479
	Delights Creamy Chocolate	0.859610
	Delights Mint Chip	0.724395
	Vanilla Chocolate	0.326861
	Natural Vanilla	0.291182
	Chocolate	0.289863
	Vanilla Chocolate Strawberry	0.287427
	Natural Vanilla Snack Cups 10ct	0.263089
	Homemade Vanilla	0.244359
Cherry Vanilla	0.237817	
0.15	Delights Cookies & Cream	0.871479
	Delights Creamy Chocolate	0.859610
	Delights Mint Chip	0.724395
	Vanilla Chocolate	0.326861
	Natural Vanilla	0.291182
	Chocolate	0.289863
	Vanilla Chocolate Strawberry	0.287427
	Natural Vanilla Snack Cups 10ct	0.263089
	Homemade Vanilla	0.244359
	Cherry Vanilla	0.237817

ตาราง 7 (ต่อ)

Threshold	name	similarity
0.05	Delights Cookies & Cream	0.871479
	Delights Creamy Chocolate	0.859610
	Delights Mint Chip	0.724395
	Vanilla Chocolate	0.326861
	Natural Vanilla	0.291182
	Chocolate	0.289863
	Vanilla Chocolate Strawberry	0.287427
	Natural Vanilla Snack Cups 10ct	0.263089
	Homemade Vanilla	0.244359
	Cherry Vanilla	0.237817

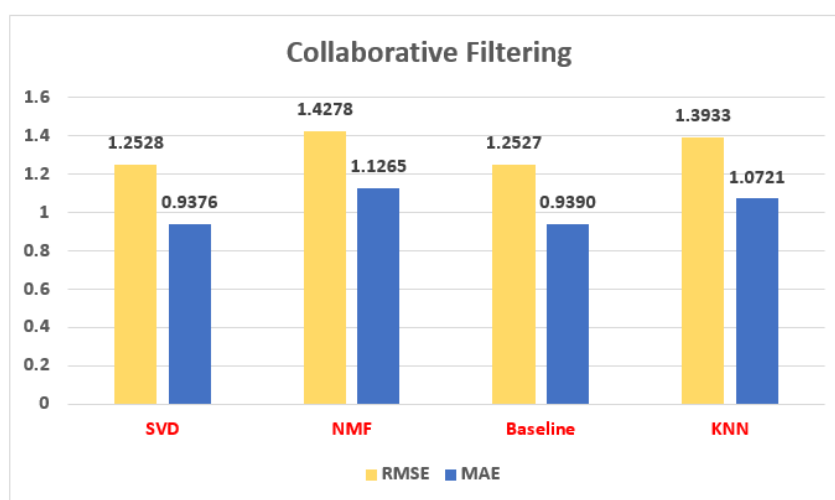
4.2 การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering)

สำหรับเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมใช้ไลบรารี Surprise มาช่วยในการสร้างแบบจำลอง ซึ่งใช้ 4 อัลกอริทึมมาเปรียบเทียบ ได้แก่ SVD, NMF, Baseline และ KNN โดยพิจารณาจากค่า rating โดยกำหนดค่าพารามิเตอร์ดังตาราง 9 ถึง ตาราง 12 ซึ่งใช้ค่า default ของแต่ละแบบจำลอง และทั้ง 4 อัลกอริทึมก็นำมาวัดประสิทธิภาพด้วยค่า RMSE และ MAE จากตาราง 8 จะเห็นได้ว่า ถ้าค่า Loss Function ถ้ามีความใกล้เคียงกับ 0 มากเท่าไรแสดงว่าแบบจำลองสามารถทำนายได้มีค่าความคาดเคลื่อนน้อย จากผลลัพธ์ที่ดีที่สุดก็คืออัลกอริทึมของ SVD โดยใช้ในการวัดผลจาก RMSE 1.2528 และ MAE 0.9376

ตาราง 8 การวัดผลประสิทธิภาพของแบบจำลองด้วยค่าของ RMSE และ MAE

	RMSE	MAE
SVD	1.2528	0.9376
NMF	1.4278	1.1265
Baseline	1.2527	0.9390
KNN	1.3933	1.0721

จากภาพประกอบ 44 ผลลัพธ์จากการเปรียบเทียบประสิทธิภาพของแบบจำลองทั้ง 4 อัลกอริทึมนั้น จะสังเกตเห็นมีอัลกอริทึมให้ผลลัพธ์ที่ดีที่สุดคือ SVD ที่ให้ค่า RMSE และ MSE เท่ากับ 1.2528 และ 0.9376 ตามลำดับ และ Baseline ที่ให้ค่า RMSE และ MSE เท่ากับ 1.2527 และ 0.9390 ตามลำดับ ซึ่งมีผลลัพธ์ที่แทบจะไม่มี ความแตกต่างกัน ดังนั้นผู้วิจัยจึงได้เลือกวิธีการที่ใช้อัลกอริทึมของ SVD เนื่องจากเป็นวิธีการที่มีความนิยมสูง และได้รางวัลจาก Netflix มาแล้วซึ่งทำให้ผู้วิจัยมีความสนใจในการนำมาใช้มากกว่า



ภาพประกอบ 44 การประเมินประสิทธิภาพแบบจำลองด้วยค่า RMSE และ MAE ของอัลกอริทึม SVD, NMF, Baseline และ KNN

ตาราง 9 ค่าพารามิเตอร์ในการสร้างแบบจำลองอัลกอริทึม SVD

Algorithm	Parameter	Description
SVD	n_factors = 100	จำนวนของปัจจัย
	n_epochs = 20	จำนวนการทำซ้ำของ SGD
	biased = true	จะมีการใช้อคติในแบบจำลองหรือไม่
	init_mean = 0	ค่าเฉลี่ยของการแจกแจงปกติสำหรับการสร้างเวกเตอร์ตัว
	init_std_dev = 1	ค่าเบี่ยงเบนมาตรฐานของการแจกแจงปกติสำหรับการสร้างเวกเตอร์ตัวประกอบ
	lr_all = 0.005	อัตราการเรียนรู้สำหรับพารามิเตอร์ทั้งหมด
	reg_all = 0.02	เงื่อนไขการทำให้เป็นมาตรฐานสำหรับพารามิเตอร์ทั้งหมด

ตาราง 10 ค่าพารามิเตอร์ในการสร้างแบบจำลองอัลกอริทึม NMF

Algorithm	Parameter	Description
NMF	n_factors = 15	จำนวนของปัจจัย
	n_epochs = 50	จำนวนการทำซ้ำของ SGD
	reg_pu = 0.06	เงื่อนไขในการทำให้เป็นมาตรฐานสำหรับผู้ใช้
	reg_qi = 0.06	เงื่อนไขในการทำให้เป็นมาตรฐานสำหรับรายการ
	reg_bu = 0.02	เงื่อนไขในการทำให้เป็นมาตรฐานสำหรับ bu
	reg_bi = 0.02	เงื่อนไขในการทำให้เป็นมาตรฐานสำหรับ bi
	lr_bu = 0.005	อัตราการเรียนรู้สำหรับ bu
	lr_bi = 0.005	อัตราการเรียนรู้สำหรับ bi

ตาราง 11 ค่าพารามิเตอร์ในการสร้างแบบจำลองอัลกอริทึม Baseline

Algorithm	Parameter	Description
Baseline	bsl_options = { 'method': 'als', 'n_epochs': 5, 'reg_u': 12, 'reg_i': 5 }	วิธีการ Alternating Least Squares จำนวนการทำซ้ำของ ALS พารามิเตอร์การทำให้เป็นมาตรฐานสำหรับผู้ใช้ พารามิเตอร์การทำให้เป็นมาตรฐานสำหรับรายการต่างๆ

ตาราง 12 ค่าพารามิเตอร์ในการสร้างแบบจำลองอัลกอริทึม K-NN

Algorithm	Parameter	Description
KNN	k = 40	จำนวนเพื่อนบ้านที่ต้องคำนึงถึงสำหรับอัลกอริทึม (สูงสุด)
	min_k = 1	จำนวนเพื่อนบ้านขั้นต่ำที่ต้องคำนึงถึงสำหรับอัลกอริทึม หากมีเพื่อนบ้านไม่เพียงพอการรวมเพื่อนบ้านจะถูกตั้งค่าเป็นศูนย์

4.3 ระบบแนะนำแบบผสมผสาน (Hybrid Recommendation System)

ระบบแนะนำแบบผสมผสานจะผสมผสานการทำงานระหว่างเทคนิคการกรองแบบอิงเนื้อหา และเทคนิคการกรองข้อมูลแบบพึ่งพาผู้เข้าร่วมซึ่งประสิทธิภาพจะขึ้นอยู่กับวิธีที่นำมาใช้ผสมผสานร่วมกัน จึงไม่ได้มีการวัดประสิทธิภาพในส่วนของเทคนิคนี้ จากภาพประกอบ 45 จะแสดงตัวอย่างผลลัพธ์ของไอศกรีมที่ได้มาจากระบบแนะนำแบบผสมผสาน ซึ่งจะมีการทำงานในรูปแบบของการกรองแบบอิงเนื้อหาโดยการลิสต์อันดับเมนูแต่ละเมนูที่มีค่าความเหมือนมากที่สุด 25 อันดับ เพื่อไปทำงานกับวิธีของการกรองข้อมูลแบบพึ่งพาผู้เข้าร่วมต่อ และใช้แบบจำลองที่ได้จากการกรองข้อมูลแบบพึ่งพาผู้เข้าร่วมนั้นแนะนำผลลัพธ์ที่ดีที่สุด สูงสุด 10 อันดับจากรายการทั้งหมด 25 อันดับที่ได้มาก่อนหน้านี้

brand	key	name	description	rating	rating_count	est
bj	36_bj	Peanut Butter World®	What makes Ben & Jerry's so euphoric? Some say...	4.9	69	4.832473
hd	26_hd	Cookies and Cream Ice Cream	Pieces of rich, chocolaty cookies are dunked i...	4.6	23	4.631420
bj	42_bj	Salted Caramel Almond	What makes Ben & Jerry's so euphoric? Some say...	4.5	23	4.307090

ภาพประกอบ 45 ตัวอย่างผลลัพธ์ชื่อเมนูที่ได้จากการแนะนำด้วยวิธีระบบแนะนำแบบผสมผสาน

บทที่ 5

สรุปผลการวิจัย อภิปรายผลและข้อเสนอแนะ

ในการวิจัยเพื่อศึกษาและพัฒนาระบบแนะนำสินค้าประเภทไอศกรีม โดยใช้เทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering) และเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) มาทำงานร่วมกันแบบผสมผสาน (Hybrid Recommendation System) โดยมีการวัดประสิทธิภาพของเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) โดยสามารถแบ่งหัวข้อในการสรุปผลได้ดังต่อไปนี้

1. สรุปผลการวิจัย
2. อภิปรายผลการวิจัย
3. ข้อเสนอแนะ

5.1 สรุปผลการวิจัย

ในการวิจัยนี้เพื่อศึกษาและพัฒนาระบบแนะนำสินค้าประเภทไอศกรีม โดยใช้เทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering) และเทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) มาทำงานร่วมกันแบบผสมผสาน (Hybrid Recommendation System) โดยศึกษา 3 วิธีการดังนี้

วิธีที่ 1 การกรองแบบอิงเนื้อหาโดยใช้ทฤษฎีของ TF-IDF แล้วนำเวกเตอร์ของคำที่ได้ไปทำการหาค่าความคล้ายคลึงของโคไซน์ เพื่อนำมาเปรียบเทียบความสัมพันธ์ของคำแต่ละคำ และสร้างเป็นคำแนะนำออกมาให้ลูกค้าจากค่าความคล้ายคลึงที่ได้จากค่าความคล้ายคลึงของโคไซน์

วิธีที่ 2 การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม โดยจะทำการสร้างแบบจำลองขึ้นมาจาก 4 อัลกอริทึม ได้แก่ SVD, NMF, Baseline และ KNN และนำมาวัดประสิทธิภาพกันด้วยค่า RMSE และ MAE เพื่อเลือกอัลกอริทึมที่น่าสนใจที่สุดและนำไปสร้างระบบแบบผสมผสานต่อไป

วิธีที่ 3 ระบบแนะนำแบบผสมผสาน โดยจะผสมผสานการทำงานระหว่างการกรองแบบอิงเนื้อหาที่ได้สร้างมาจากวิธีก่อนหน้านี้ และแบบจำลองจากอัลกอริทึม SVD ที่ได้เลือกมาจากการวัดประสิทธิภาพของวิธีการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม เพื่อสร้างเป็นคำแนะนำให้กับผู้ใช้งาน

5.2 อภิปรายผลการวิจัย

จากการทดลองครั้งนี้ทางผู้วิจัยได้ใช้เว็บแอปพลิเคชันที่บุคคลสำหรับสร้างสร้างแบบจำลอง และการทำนายผลการทดลอง 3 เทคนิคได้แก่

5.2.1 เทคนิคการกรองแบบอิงเนื้อหา (Content-Based Filtering)

เทคนิคการกรองแบบอิงเนื้อหาใช้วิธีการ TF-IDF เข้ามาช่วยในการทำ Vectorization ของคำแต่ละคำในคอลัมน์ของ description และใช้ทฤษฎีของความคล้ายคลึงแบบโคไซน์เข้ามาช่วยในการคำนวณความสัมพันธ์ของคำแต่ละคำ โดยกำหนดเกณฑ์ขั้นต่ำของค่าความเหมือนอยู่ที่ 0.15 โดยได้มีการสำรวจข้อมูลจากเกณฑ์ระดับต่าง ๆ แล้วโดยมีค่าตั้งแต่ 0.05 – 0.95 ที่ผู้วิจัยได้ทำการสำรวจมา หลังจากนั้นจะเลือกชื่อเมนูไอศกรีมที่มีความสัมพันธ์มากที่สุดมาได้สูงสุด 10 อันดับ เพื่อทำการแนะนำให้กับลูกค้า โดยสาเหตุที่กำหนดเกณฑ์ไว้ค่อนข้างจะต่ำนั้นเนื่องจากค่าความเหมือนที่ได้จากวิธีการของความคล้ายคลึงแบบโคไซน์ของรายการเมนูส่วนใหญ่จะได้ค่าออกมาค่อนข้างจะต่ำ ซึ่งจะมีค่าอยู่ในช่วงซึ่งอาจจะเป็นปัญหาจากข้อมูลที่ใช้ ทำให้ค่าความเหมือนที่ออกมามีค่าน้อย จึงต้องกำหนดเกณฑ์ของค่าความเหมือนต่ำลงไปด้วย เพื่อให้สามารถทำการแนะนำผลลัพธ์ของรายการเมนูส่วนใหญ่ออกมาแสดงให้ลูกค้าได้

5.2.2 เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering)

เทคนิคการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม ทางผู้จัดทำได้ใช้ไลบรารีของ Scikit Learn ที่ชื่อว่า Surprise ในการสร้างแบบจำลองขึ้นมา 4 อัลกอริทึม ได้แก่ SVD, NMF, Baseline และ KNN โดยใช้คอลัมน์ author, key และ star ในการนำมาสร้างแบบจำลอง ซึ่งผู้จัดทำได้ใช้เมตริกซ์ สำหรับวัดผล ได้แก่ RMSE และ MAE เพื่อวัดผลและเปรียบเทียบประสิทธิภาพของแบบจำลองทั้ง 4 แบบจำลอง โดยถ้าค่า Loss Function มีความใกล้เคียงกับ 0 แสดงว่าโมเดลสามารถทำนายได้มีความคาดเคลื่อนน้อย ซึ่งโมเดลที่มีผลคะแนนโดยรวมดีที่สุดหรือมีค่าตัวเลขใกล้เคียงค่า 0 ที่สุดคืออัลกอริทึมของ SVD ซึ่ง ได้ค่า RMSE 1.2528 และ MAE 0.9376 ดังภาพประกอบ 46

```
RMSE: 1.2528
MAE: 0.9376
```

```
-----
Metrics - RMSE: 1.2527884944980345, MAE: 0.9375664299939616
```

ภาพประกอบ 46 การประเมินประสิทธิภาพแบบจำลองด้วยค่า RMSE และ MAE ของอัลกอริทึม

SVD

5.2.3 ระบบแนะนำแบบผสมผสาน (Hybrid Recommendation System)

ระบบแนะนำแบบผสมผสาน คือ เทคนิคที่ผสมผสานการทำงานระหว่างวิธีการแบบเทคนิคการกรองแบบอิงเนื้อหา และเทคนิคการกรองข้อมูลแบบพึ่งพาผู้เข้าร่วม โดยจะไม่มีผลลัพธ์ที่ชัดเจนเหมือนวิธีการกรองข้อมูลแบบพึ่งพาผู้เข้าร่วม เนื่องจากการวัดผลของการทำ ระบบแนะนำนั้นจะสามารถวัดได้ในกรณีที่ตัวแบบจำลองมีการทำนายค่า rating ซึ่งวิธีการของระบบแนะนำแบบผสมผสานนั้น ได้มีการนำการทำงานของเทคนิคการกรองแบบอิงเนื้อหา ที่มีการใช้ความคล้ายคลึงแบบโคไซน์ ที่ไม่ได้มีการทำนายค่าอะไรมาผสมในการทำงานของแบบจำลองด้วย ซึ่งถ้าต้องการวัดประสิทธิภาพของแบบจำลองระบบแนะนำแบบผสมผสาน นั้นจะสามารถอิงได้จากประสิทธิภาพของเทคนิคการกรองข้อมูลแบบพึ่งพาผู้เข้าร่วม เพียงอย่างเดียวเท่านั้น เพราะมีการวัดผลแบบจำลองที่ชัดเจนด้วยค่า RMSE และ MAE แต่ก็จะไม่มีการวัดผลที่ชัดเจนของตัวแบบจำลองเอง

5.3 ข้อเสนอแนะ

เนื่องจากการวิจัยครั้งนี้ในส่วนของ user ที่มีการให้คะแนนรีวิวหรือคอมเมนต์รีวิวเกี่ยวกับรสชาติไอศกรีมต่าง ๆ นั้นข้อมูลที่นำมาใช้ในการวิเคราะห์นั้น อาจจะไม่มีความหลากหลาย และความสมบูรณ์เพียงพอ โดยจะเห็นได้ชัดจากการสร้างระบบด้วยเทคนิคการกรองแบบอิงเนื้อหาซึ่งได้มีการกำหนดเกณฑ์ขั้นต่ำไว้ค่อนข้างน้อยมาก ถ้ามีโอกาสต่อยอดงานวิจัยในรูปแบบนี้ จะทำการคัดเลือกข้อมูลที่มีความสมบูรณ์ และหลากหลายมากยิ่งขึ้น ซึ่งอาจจะทำให้ให้ระบบแนะนำสินค้าที่สร้างขึ้นมามีประสิทธิภาพมากยิ่งขึ้น และในส่วนของ การสร้างแบบจำลองแบบระบบแนะนำแบบผสมผสานนั้น ทางผู้วิจัยอาจจะเลือกวิธีการอื่นที่ไม่มีการใช้ทฤษฎีของความคล้ายคลึงแบบโคไซน์ มาทำการสร้างแบบจำลองด้วย เพื่อที่จะสามารถวัดผลตัวแบบจำลองได้ชัดเจนมากยิ่งขึ้น

บรรณานุกรม

- de Campos, L. M., Fernández-Luna, J. M., Huete, J. F., & Rueda-Morales, M. A. (2010). Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks. *International Journal of Approximate Reasoning*, 51(7), 785-799.
- Esteban, A., Zafra, A., & Romero, C. (2020). Helping university students to choose elective courses by using a hybrid multi-criteria recommendation system with genetic optimization. *Knowledge-Based Systems*, 194.
- Hug, N. Surprise A Python scikit for recommender systems. <http://surpriselib.com/>
- Philip Lenhart, D. H. (2016). Combining Content-based and Collaborative Filtering for Personalized Sports News Recommendations.
- Pond, T. (2020). Ice Cream Dataset. <https://www.kaggle.com/tysonpo/ice-cream-dataset>
- Tewari, A. S. (2019). Generating Items Recommendations by Fusing Content and User-Item based Collaborative Filtering
- TITIPATA. (2016). Collaborative filtering พี่เจอรี่์การแนะนำเพลงของ Spotify. <https://tupleblog.github.io/spotify/>
- Zhiyuan Fang, L. Z., Kun Chen. (2016). HYBRID RECOMMENDER SYSTEM BASED ON PERSONAL BEHAVIOR MINING.



รายละเอียดโค้ดการทำ Content-Based Filtering

```

1. import pandas as pd
2. import numbers
3. import matplotlib.pyplot as plt
4. from sklearn.feature_extraction.text import TfidfVectorizer
5. from surprise import Reader, Dataset, SVD, BaselineOnly, NMF,
   accuracy, KNNBasic
6. from surprise.model_selection import train_test_split
7. from sklearn.metrics.pairwise import linear_kernel
8.
9. product_df = pd.read_csv('products.csv')
10. product_df.head()
11. print(product_df['brand'].drop_duplicates())
12.
13. sort_bj = product_df[product_df['brand'] ==
   'bj'].sort_values('rating_count', ascending=False)
14. top_ten_bj = sort_bj.head(10).sort_values('rating', ascending=False)
15.
16. sort_breyers = product_df[product_df['brand'] ==
   'breyers'].sort_values('rating_count', ascending=False)
17. top_ten_breyers = sort_breyers.head(10).sort_values('rating',
   ascending=False)
18.
19. sort_talenti = product_df[product_df['brand'] ==
   'talenti'].sort_values('rating_count', ascending=False)
20. top_ten_talenti = sort_talenti.head(10).sort_values('rating',
   ascending=False)
21.
22. sort_hd = product_df[product_df['brand'] ==
   'hd'].sort_values('rating_count', ascending=False)
23. top_ten_hd = sort_hd.head(10).sort_values('rating', ascending=False)
24.
25. top_ten_bj.head(10).plot(kind='bar', x='name', y='rating')
26. top_ten_talenti.head(10).plot(kind='bar', x='name', y='rating')
27. top_ten_breyers.head(10).plot(kind='bar', x='name', y='rating')
28. top_ten_hd.head(10).plot(kind='bar', x='name', y='rating')
29.
30.
31. product_df['description'] = product_df['description'].fillna('')
32.
33. tfidf = TfidfVectorizer(stop_words='english')
34. tfidf_matrix = tfidf.fit_transform(product_df['description'])
35. tfidf_matrix.shape
36. tfidf.get_feature_names()[100:110]
37.
38. cosine_sim = linear_kernel(tfidf_matrix, tfidf_matrix)
39. cosine_sim.shape
40. idx = product_df['name'].drop_duplicates()
41. indices = pd.Series(idx.index, index=idx)
42. indices[:10]
43.

```

```
44. def get_recommendations(name, cosine_sim=cosine_sim):
45.     try:
46.         idx = indices[name]
47.
48.         if idx.size == 0:
49.             print("cannot get recommendaion : please input a correct
ice cream name")
50.             return
51.
52.         sim_scores = list(enumerate(cosine_sim[idx]))
53.         sim_scores = sorted(sim_scores, key=lambda x: x[1],
reverse=True)
54.
55.         sim_scores = sim_scores[1:11]
56.         threshold_data = [i for i in sim_scores if i[1] > 0.15]
57.         print(threshold_data)
58.
59.         ic_indices = [i[0] for i in threshold_data]
60.         sim_value = [i[1] for i in threshold_data]
61.
62.         result = pd.DataFrame(product_df['name'].iloc[ic_indices])
63.         result['sim_value'] = sim_value
64.         result.reset_index(drop=True, inplace=True)
65.
66.         return result
67.     except:
68.         print("cannot get recommendaion : please input a correct ice
cream name")
69.
70. get_recommendations('Berry Sweet Mascarpone')
```



รายละเอียดโค้ดการทำ Collaborative Filtering

```

1. reviews_df.stars.value_counts().plot(kind='bar')
2. plt.show()
3.
4. reader = Reader(rating_scale=(1, 5))
5. dataset = Dataset.load_from_df(reviews_df[['author', 'key', 'stars']],
    reader)
6.
7. algos = []
8. def addAlgorithm(algo):
9.     algos.append(algo)
10.
11.
12. def train_and_evaluate():
13.     for algo in algos:
14.         algo.fit(train)
15.         predictions = algo.test(test)
16.         rmse = accuracy.rmse(predictions)
17.         mae = accuracy.mae(predictions)
18.         print('-----')
19.         print(f'        Metrics - RMSE: {rmse}, MAE: {mae}')
20.         print('-----')
21.
22. svd = SVD()
23. addAlgorithm(svd)
24.
25. nmf = NMF()
26. addAlgorithm(nmf)
27.
28. bsl_options = {'method': 'als',
29.                'n_epochs': 5,
30.                'reg_u': 12,
31.                'reg_i': 5
32.                }
33. als = BaselineOnly(bsl_options=bsl_options)
34. addAlgorithm(als)
35.
36. knn = KNNBasic(k=3, min_k=1)
37. addAlgorithm(knn)
38.
39. train_and_evaluate()
40.
41. def predict_top_ten_est(predictor , user , reviews):
42.     review_tmp = reviews[['key', 'brand', 'stars']]
43.     review_tmp['est'] = reviews['key'].drop_duplicates().apply(lambda x:
    predictor.predict(user, x, verbose=False).est)
44.     review_tmp = review_tmp.sort_values('est', ascending=False)
45.     return review_tmp.head(10)
46.
47. pred = nmf.predict('chicago220', '1_hd', verbose=True)
48. print(pred)
49.
50. predict_top_ten_est(svd, 'chicago220', reviews_df)
51.

```



รายละเอียดโค้ดการทำ Hybrid Recommendation System

```
1. def get_hybrid_recommendations(user , name, cosine_sim=cosine_sim):
2.     try:
3.         idx = indices[name]
4.         sim_scores = list(enumerate(cosine_sim[idx]))
5.         sim_scores = sorted(sim_scores, key=lambda x: x[1],
reverse=True)
6.         sim_scores = sim_scores[1:26]
7.
8.         ic_indices = [i[0] for i in sim_scores]
9.         product_from_cn = product_df.iloc[ic_indices]
10.        product_from_cn['est'] = product_from_cn['key'].apply(lambda
x: svd.predict(user, x).est)
11.        product_from_cn = product_from_cn.sort_values('est',
ascending=False)
12.        return product_from_cn.head(10)
13.    except:
14.        print("cannot get recommendaion : please input a correct ice
cream name")
15.
16. get_hybrid_recommendations('chicago220', 'Berry Sweet Mascarpone')
17.
```

ประวัติผู้เขียน

ชื่อ-สกุล	นิภาภรณ์ พันธุ์นาม
วัน เดือน ปี เกิด	27 กันยายน 2535
สถานที่เกิด	กรุงเทพมหานคร
วุฒิการศึกษา	พ.ศ. 2555 คณะวิทยาศาสตร์ สาขา ชีววิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ พ.ศ. 2562 วิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูล มหาวิทยาลัยศรีนครินทรวิโรฒ
ที่อยู่ปัจจุบัน	293/32 ซอยแยกราชฎา ถนน ราชปรารภ แขวง มักกะสัน เขต ราชเทวี กทม 10400

