ระบบแนะนำตามบริบทสำหรับธุรกิจวัสดุก่อสร้างและอาคาร

Contextual Recommender Systems for Building and Construction Materials Business

สุทธิรัตน์ เกลี้ยงเกลา

Sutthirat Kliangklao

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา
วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์
มหาวิทยาลัยสงขลานครินทร์

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Engineering in Computer Engineering
Prince of Songkla University
2565

ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์

ระบบแนะนำตามบริบทสำหรับธุรกิจวัสดุก่อสร้างและอาคาร

Contextual Recommender Systems for Building and Construction Materials Business

สุทธิรัตน์ เกลี้ยงเกลา

Sutthirat Kliangklao

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา
วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์
มหาวิทยาลัยสงขลานครินทร์

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Engineering in Computer Engineering
Prince of Songkla University
2565

ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์

ชื่อวิทยานิพนธ์ ระบบแนะนำตามบริบทสำหรับธุรกิจวัสดุก่อสร้างและอาคาร
ผู้เขียน นายสุทธิรัตน์ เกลี้ยงเกลา
สาขาวิชา วิศวกรรมคอมพิวเตอร์

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

.....
(ผู้ช่วยศาสตราจารย์ ดร. นิคม สุวรรณวร)

คณะกรรมการสอบ

.....ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. ธเนศ เคารพาพงศ์)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. นิคม สุวรรณวร)

.....กรรมการ
(รองศาสตราจารย์ ดร. มนตรี กาญจนะเดชะ)

.....กรรมการ
(ดร. อนันต์ ชกสุริวงค์)

.....กรรมการ
(รองศาสตราจารย์ ดร. วัฒนพงศ์ เกิดทองมี)

บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษา ตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์

.....
(ผู้ช่วยศาสตราจารย์ ดร. เถกิง วงศ์ศิริโชติ)
รักษาการแทนคณบดีบัณฑิตวิทยาลัย

ขอรับรองว่า ผลงานวิจัยนี้มาจากการศึกษาวิจัยของนักศึกษาเอง และได้แสดงความขอบคุณบุคคลที่มีส่วนช่วยเหลือแล้ว

ลงชื่อ .....
(ผู้ช่วยศาสตราจารย์ ดร. นิคม สุวรรณวร)
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

ลงชื่อ .....
(นายสุทธิรัตน์ เกลี้ยงเกลา)
นักศึกษา

ข้าพเจ้าขอรับรองว่า ผลงานวิจัยนี้ไม่เคยเป็นส่วนหนึ่งในการอนุมัติปริญญาในระดับใดมาก่อน และไม่ได้ถูกใช้ในการยื่นขออนุมัติปริญญาในขณะนี้

ลงชื่อ .....
(นายสุทธิรัตน์ เกลี้ยงเกลา)
นักศึกษา

ชื่อวิทยานิพนธ์ ระบบแนะนำตามบริบทสำหรับธุรกิจวัสดุก่อสร้างและอาคาร
ผู้เขียน นายสุทธิรัตน์ เกลี้ยงเกลา
สาขาวิชา วิศวกรรมคอมพิวเตอร์
ปีการศึกษา 2565

บทคัดย่อ

ปัจจุบันระบบการแนะนำเป็นหนึ่งในเทคโนโลยีสำคัญที่สนับสนุนสำหรับอีคอมเมิร์ซ โดยมีจุดมุ่งหมายเพื่อแนะนำสินค้าหรือบริการที่ตรงตามความต้องการของผู้ซื้อ เพื่อเพิ่มยอดขายสำหรับธุรกิจ ในการวิจัยนี้เรามุ่งเน้นไปที่การพัฒนาระบบแนะนำสำหรับธุรกิจวัสดุก่อสร้าง ธุรกิจวัสดุก่อสร้างเป็นธุรกิจที่จำหน่ายวัสดุก่อสร้างและอุปกรณ์ที่เกี่ยวข้อง เช่น สินค้าโครงสร้าง อุปกรณ์เครื่องมือ และอื่นๆ สำหรับลูกค้าที่มาซื้อผลิตภัณฑ์จะเป็นช่างก่อสร้างหรือลูกค้าที่ต้องการปรับปรุงบ้าน ระบบแนะนำสินค้าในธุรกิจนี้จะแนะนำสินค้าที่สามารถนำไปใช้ในอาชีพได้ โดยทั่วไป ระบบจะแนะนำผลิตภัณฑ์ที่คล้ายกับที่ซื้อ แต่ไม่คำนึงถึงบริบทหรืออาชีพของลูกค้า โดยในงานวิจัยนี้ เราได้นำเสนอการสร้างแบบจำลองข้อมูลการรับรู้บริบท เพื่อที่จะพัฒนาระบบแนะนำสำหรับธุรกิจวัสดุก่อสร้าง

คำสำคัญ : พาณิชย์อิเล็กทรอนิกส์, ระบบแนะนำ, การรับรู้บริบท, การแบ่งกลุ่มข้อมูลแบบเคมีน, ระบบประมวลผลภาษาไทย

Thesis Title Contextual Recommender Systems for Building and Construction Materials Business
Author Mr. Sutthirat Kliangklao
Major Program Computer Engineering
Academic Year 2022

ABSTRACT

Nowadays, the recommendation system is one of the most important supported technologies to e-commerce that aims for recommending the products or services to be purchased, to increase sales. In this work, the focus on the recommendation system for the building materials business. Building materials business is a business that sales construction related materials and equipment, such as, structural goods, tools supplies, etc. For customers who come to buy products will builder professionally or customers who want to improve their homes. Products recommendation system in this business will recommend products that can be used in profession. Generally, system recommends products that are like the ones purchased but regardless of context or profession of the customer. In this paper, we propose a context awareness data modeling to specialize the recommendation system aiming for the building materials business.

Keywords: E-Commerce, Recommendation, Context-Awareness, K-means, Thai NLP

กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จลุล่วงได้ด้วยดี ต้องขอขอบพระคุณการสนับสนุนของอาจารย์ที่ปรึกษางานวิจัย ผศ.ดร.นิคม สุวรรณวร เป็นอย่างสูง ซึ่งได้ให้ทั้งคำปรึกษา ความรู้ กำลังใจ การช่วยเหลือต่างๆ ตลอดการวิจัยนี้

ขอขอบพระคุณ โครงการสนับสนุนการพัฒนาเทคโนโลยีของอุตสาหกรรมไทย (ITAP) ภายใต้การสนับสนุนของฝ่ายพัฒนาเทคโนโลยีอุตสาหกรรม ศูนย์บริการจัดการเทคโนโลยี สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (สวทช.) ที่กรุณาสนับสนุน ทุนการศึกษาแก่ผู้วิจัยตลอดการศึกษา

ขอขอบพระคุณบริษัท วนาวัฒน์ วัสดุ จำกัดที่ได้ให้การสนับสนุนในการเก็บข้อมูลเพื่อนำไปใช้ในการทดสอบผล และให้ความช่วยเหลือในด้านต่างๆ

สุดท้ายนี้ขอขอบพระคุณอาจารย์ทุกท่านที่ได้ให้ความรู้ คำแนะนำ และช่วยขัดเกลางานวิจัยนี้ให้ประสบผลสำเร็จลุล่วงทุกท่าน ขอขอบคุณบิดา มารดา ที่คอยให้กำลังใจและสนับสนุนตลอดการศึกษา หากงานวิจัยนี้มีข้อผิดพลาดประการใดขออภัยมา ณ ที่นี้ด้วย

สุทธิรัตน์ เกลี้ยงเกลา

สารบัญ

หน้า

สารบัญ (8)
รายการตาราง (10)
รายการภาพประกอบ (11)
รายการภาพประกอบ(ต่อ) (12)
บทที่ 1 1
บทนำ 1
1.1 ที่มาและความสำคัญของงานวิจัย 1
1.2 วัตถุประสงค์ของงานวิจัย 4
1.3 ขอบเขตของการวิจัย 4
1.4 ขั้นตอนและวิธีดำเนินงาน 5
1.5 ประโยชน์ที่คาดว่าจะได้รับ 5
บทที่ 2 7
ทฤษฎีและหลักการที่เกี่ยวข้อง 7
2.1 หลักการและทฤษฎีที่เกี่ยวข้อง 7
2.2 ทฤษฎีที่เกี่ยวข้องในด้านเทคนิค 16
2.3 งานวิจัยที่เกี่ยวข้อง 19
บทที่ 3 23
วิธีการดำเนินวิจัย 23
3.1 กระบวนการออกแบบและพัฒนาระบบขายสินค้าออนไลน์ 23
3.1.1 ภาพรวมของระบบ 24
3.1.2 การวิเคราะห์ข้อมูลด้านการไหลของข้อมูล 25
3.3 กระบวนการออกแบบและพัฒนาระบบแนะนำ 27
3.3.1 การสร้างโมเดลของข้อมูล 28
3.3.2 การสร้างโมเดลการกรองข้อมูล 30
บทที่ 4 34
ผลการวิจัย 34
4.1 ผลการออกแบบ พัฒนาและติดตั้งระบบขายสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ 34
4.1.1 ระบบขายสินค้าออนไลน์ (เดิม) 34

สารบัญ (ต่อ)

หน้า

4.1.2 ระบบขายสินค้าออนไลน์ที่ออกแบบพัฒนาขึ้นใหม่ 37
4.1.3 สรุปผลการดำเนินการ 40
4.2 ผลการวิจัยและพัฒนาระบบแนะนำสินค้าแบบการรับรู้บริบทสำหรับอีคอมเมิร์ซ 41
4.2.1 ผลการจัดเก็บข้อมูลพฤติกรรมของลูกค้า 40
4.2.2 โมเดลข้อมูล Data model 42
4.2.3 โมเดลการกรอง Filtering model 49
4.2.4 ผลลัพธ์การพัฒนาระบบแนะนำสินค้า 51
4.2.5 ผลการประเมินระบบแนะนำสินค้า 52
4.2.6 ข้อมูลยอดขายสินค้า 53
บทที่ 5 54
สรุปผลวิจัย และข้อเสนอแนะ 54
5.1 สรุปผลการวิจัย 54
5.2 ข้อจำกัดของการวิจัย 55
5.3 ข้อเสนอแนะ 55
บรรณานุกรม 56
ภาคผนวก 58
ภาคผนวก ก 59
ตัวอย่างแบบสอบถามเพื่อประเมินความพึงพอใจในการใช้งานระบบแนะนำตามบริบทสำหรับธุรกิจวัสดุก่อสร้างและอาคาร 59
ภาคผนวก ข 63
ผลงานตีพิมพ์เผยแพร่จากวิทยานิพนธ์ 63
ประวัติผู้เขียน 71

รายการตาราง

หน้า

ตารางที่ 2-1 การเปรียบเทียบข้อดีข้อเสียของ CMS 15
ตางรางที่ 3-1 ตัวอย่างของ COUNT MATRIX ของข้อมูลโปรไฟล์ผู้ใช้ 30
ตารางที่ 3-2 ตัวอย่างของ COUNT MATRIX ของข้อมูลโปรไฟล์รายการ 30
ตารางที่ 3-3 ตัวอย่างของ COUNT MATRIX ของข้อมูลโปรไฟล์บริบท 31
ตารางที่ 3-4 ตัวอย่างของ TF-IDF ของข้อมูลโปรไฟล์ผู้ใช้ 31
ตารางที่ 3-5 ตัวอย่างของ TF-IDF ของข้อมูลโปรไฟล์รายการ 32
ตารางที่ 3-6 ตัวอย่างของ TF-IDF ของข้อมูลโปรไฟล์บริบท 32
ตารางที่ 3-7 ตัวอย่างของข้อมูลค่าความคล้ายคลึง 33
ตารางที่ 4-1 ข้อมูลประวัติการซื้อ 41
ตารางที่ 4-2 ข้อมูลประวัติการคลิกสินค้า 41
ตารางที่ 4-3 ตัวอย่างข้อมูลสินค้าสำหรับวิจัยและพัฒนาโมเดลข้อมูล Data model 42
ตารางที่ 4-4 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 1 44
ตารางที่ 4-5 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 2 45
ตารางที่ 4-6 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 3 45
ตารางที่ 4-7 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 4 46
ตารางที่ 4-8 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 5 46
ตารางที่ 4-9 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 6 47
ตารางที่ 4-10 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 7 47
ตารางที่ 4-11 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 8 48
ตารางที่ 4-12 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 9 48
ตารางที่ 4-13 ข้อมูล User Profile ประวัติการซื้อตัวอย่างที่ 1 49
ตารางที่ 4-14 ข้อมูล User Profile ประวัติการคลิกสินค้าตัวอย่างที่ 1 50
ตารางที่ 4-15 ผลการเลือกกลุ่มจากโมเดลข้อมูลของผู้ใช้ตัวอย่างที่ 1 50
ตารางที่ 4-16 ผลการคำนวณหาค่าความคล้ายคลึงของผู้ใช้ตัวอย่างที่ 1 เรียงจากมากไปน้อย จำนวน 10 รายการ 50

รายการภาพประกอบ

หน้า

รูปที่ 1-1 คลังจัดเก็บบริษัท วนาวัฒน์วัสดุ จำกัด 1
รูปที่ 1-2 ลักษณะภายในคลัง INV (ซ้าย) และคลัง Sale (ขวา) 1
รูปที่ 1-3 Conceptual Design ระบบขายสินค้าออนไลน์ บริษัท วนาวัฒน์วัสดุ จำกัด 2
รูปที่ 1-4 ขั้นตอนการทำงานของระบบแนะนำสินค้า 3
รูปที่ 2-1 วิธีการกรองเนื้อหา (Content-based Filtering) 8
รูปที่ 2-2 วิธีการกรองข้อมูลร่วม (Collaborative Filtering) 9
รูปที่ 2-3 การทำงานของ User-based filtering กับ Item-based filtering 10
รูปที่ 2-4 วิธีการแบบผสมผสาน (Hybrid Filtering) 10
รูปที่ 2-5 ตัวอย่างการจัดกลุ่มด้วยวิธี K-mean Clustering 16
รูปที่ 2-6 แนวคิดหลักของระบบคำแนะนำ 19
รูปที่ 2-7 รูปแบบความคล้ายคลึงกันของผลิตภัณฑ์ตามข้อความ 20
รูปที่ 2-8 กระบวนการจัดกลุ่มสำหรับข้อมูลผลิตภัณฑ์อีคอมเมิร์ซ 21
รูปที่ 2-9 รูปแบบของจัดกลุ่มด้วย K=3 22
รูปที่ 3-1 กระบวนการทำงานของระบบขายสินค้าออนไลน์ 24
รูปที่ 3-2 แผนภาพบริบท (Context Diagram) 25
รูปที่ 3-3 Data Flow Diagram Level 1 ระบบขายออนไลน์ 26
รูปที่ 3-4 แนวคิดของระบบผู้แนะนำตามบริบท 27
รูปที่ 3-5 แนวคิดการสร้างโมเดลของข้อมูล 28
รูปที่ 3-6 กระบวนการประมวลผลข้อมูลล่วงหน้า 28
รูปที่ 3-7 กระบวนการสร้างคุณลักษณะ (Feature Extraction) 29
รูปที่ 4-1 หน้าหลักของระบบ 34
รูปที่ 4-2 เมนูหมวดหมู่สินค้าของระบบ 34
รูปที่ 4-3 หน้ารายการสินค้าของระบบ 35
รูปที่ 4-4 หน้ารายละเอียดสินค้าของระบบ 35
รูปที่ 4-5 หน้าตะกร้าสินค้าของระบบ 36
รูปที่ 4-6 หน้าเข้าสู่ระบบ 36
รูปที่ 4-7 หน้าสมัครสมาชิกของระบบ 37

รายการภาพประกอบ (ต่อ)

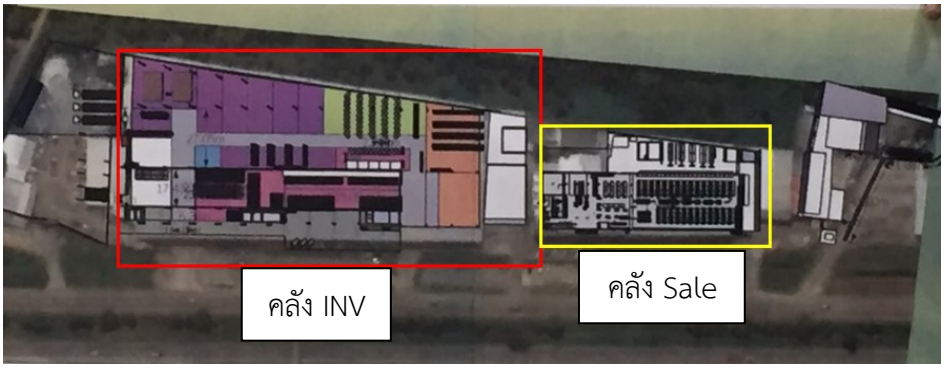
หน้า

รูปที่ 4-8 หน้าหลักของระบบ 37
รูปที่ 4-9 เมนูหมวดหมู่สินค้าของระบบ 38
รูปที่ 4-10 เมนูหมวดหมู่สินค้าย่อยของระบบ 38
รูปที่ 4-11 หน้ารายการสินค้าของระบบ 39
รูปที่ 4-12 หน้ารายละเอียดสินค้าของระบบ 39
รูปที่ 4-13 หน้าตะกร้าสินค้าของระบบ 40
รูปที่ 4-14 หน้าเข้าสู่ระบบและลงทะเบียนของระบบ 40
รูปที่ 4-15 ผลการคำนวณหาค่า K สำหรับสร้าง Data model 44
รูปที่ 4-16 กราฟแสดงจำนวนสินค้าในแต่ละกลุ่มจากโมเดลข้อมูล (Data model) 49
รูปที่ 4-17 ผลการแนะนำสินค้าตัวอย่างที่ 1 เรียงจากมากไปน้อย จำนวน 10 รายการ 51
รูปที่ 4-18 ผลการสร้างระบบแนะนำแบบ 3 มิติ สำหรับลูกค้าไม่ Login 51
รูปที่ 4-19 ผลการสร้างระบบแนะนำแบบ 3 มิติ สำหรับลูกค้า Login 52
รูปที่ 4-20 การเปรียบเทียบระหว่างการสร้างแบบจำลองข้อมูล 2 มิติและ 3 มิติ 52
รูปที่ 4-21 ข้อมูลยอดขายสินค้าก่อนการติดตั้งระบบแนะนำสินค้า 53
รูปที่ 4-22 ข้อมูลยอดขายสินค้าก่อนการติดตั้งระบบแนะนำสินค้า 53

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของงานวิจัย

บริษัท วนาวัฒน์วัสดุ จำกัด ดำเนินธุรกิจเกี่ยวกับการซื้อมาขายไปของวัสดุที่เกี่ยวข้องกับการก่อสร้าง เช่น เหล็กเส้น เหล็กแผ่นและเหล็กรูปพรรณต่าง ๆ มีสินค้ามากกว่า 30,000 รายการ โดยกลุ่มลูกค้าจะมีหลากหลาย ตั้งแต่ ลูกค้าทั่วไป ลูกค้าโครงการ โครงการของรัฐบาล เป็นต้น การดำเนินงานของบริษัทฯจะเป็นการดำเนินงานเกี่ยวกับระบบคลังสินค้า คือ การซื้อสินค้ามาจัดเก็บเพื่อรอการจำหน่าย โดยทางบริษัทฯจะแบ่งกระบวนการจัดเก็บเป็น 2 ส่วนด้วยกัน คือ ส่วนของคลังสินค้า (คลัง INV) จัดเก็บสินค้าก่อสร้างพื้นฐาน และส่วนของการขายหน้าร้าน (คลัง Sale) จัดเก็บสินค้าจำพวกฮาร์ดแวร์ เครื่องมือช่าง ประมาณ 80% เป็นต้น ดังแสดงในรูปที่ 1-1 และรูปที่ 1-2



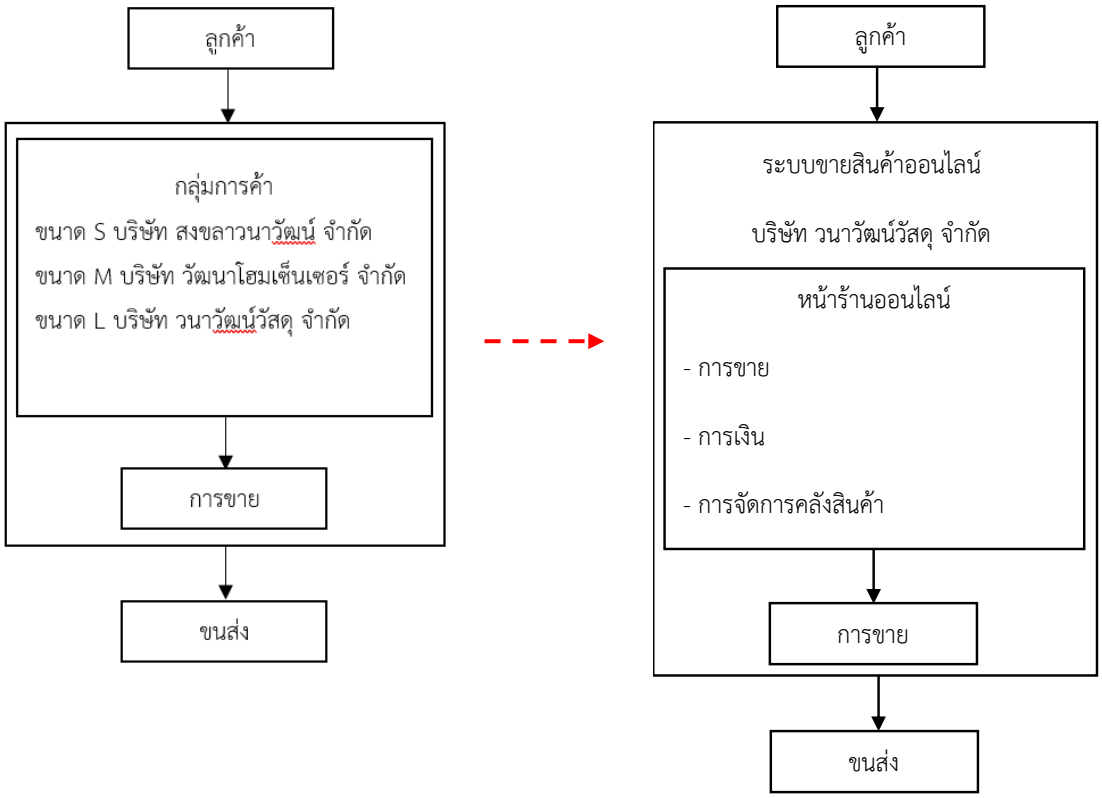
รูปที่ 1-1 คลังจัดเก็บบริษัท วนาวัฒน์วัสดุ จำกัด

รูปที่ 1-2 ลักษณะภายในคลัง INV (ซ้าย) และคลัง Sale (ขวา)

ในส่วนของข้อมูลลูกค้าที่มีเป็นจำนวนมากในปัจจุบัน ทำให้ยากที่จะนำเสนอสินค้าและบริการให้กับลูกค้าที่สนใจในสินค้าและบริการนั้น ๆ ทางบริษัทฯจึงมีความต้องการที่จะดำเนินธุรกิจในรูปแบบออนไลน์ เนื่องจากสามารถนำเสนอสินค้าและเข้าถึงกลุ่มลูกค้าในปัจจุบันได้มากกว่า แต่การที่จะได้สินค้าที่ต้องการนั้นไม่ใช่เรื่องง่าย เนื่องจากข้อมูลของสินค้ากระจัดกระจายหลากหลาย จึงจำเป็นต้องมีระบบสำหรับแนะนำสินค้าเพื่อให้ลูกค้าสามารถเลือกซื้อสินค้าได้ตรงกับความต้องการ โดยสามารถเลือกซื้อสินค้าผ่านเว็บไซต์หรือแอปพลิเคชันบนมือถือได้ เพื่อง่ายต่อการใช้งาน

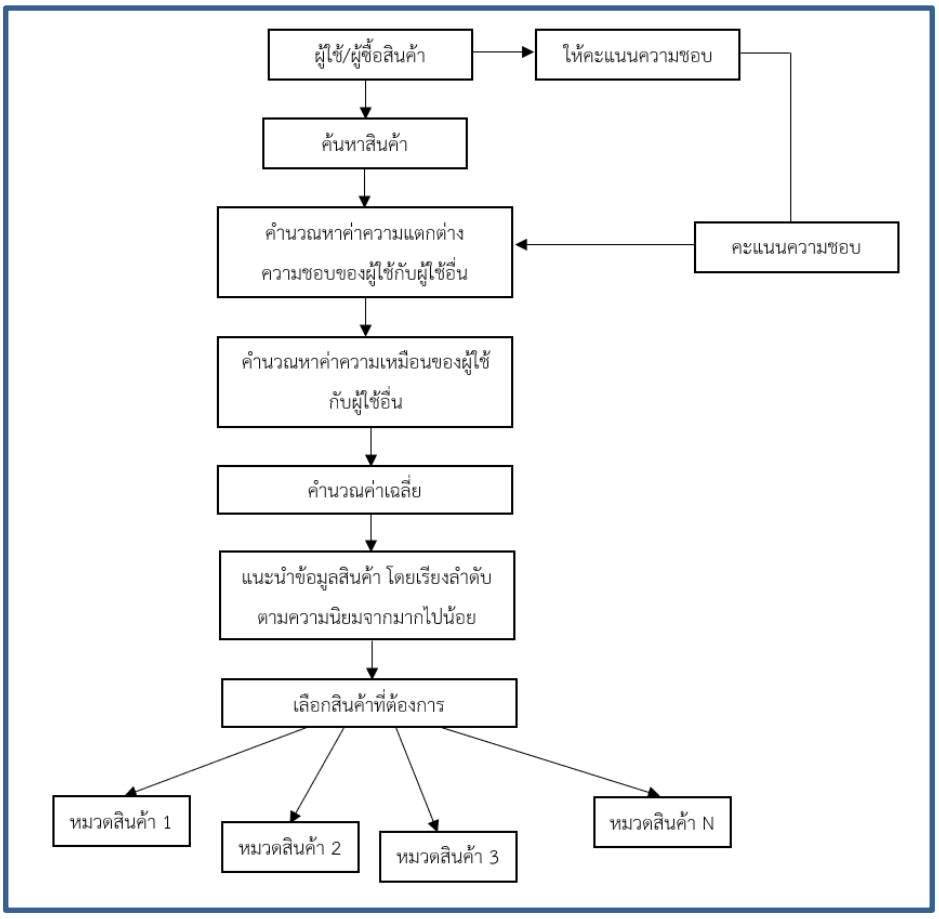
ระบบสนับสนุนการซื้อขายสินค้าในรูปแบบออนไลน์จะช่วยส่งเสริมให้บริษัทฯสามารถดำเนินธุรกิจออนไลน์ตามแนมโน้มของธุรกิจสมัยใหม่ โดยจะนำเทคโนโลยีระบบแนะนำสินค้ามาร่วมใช้งานเพื่อที่จะดูลักษณะของสินค้าที่จะแนะนำ และแนะนำสิ่งที่มีลักษณะหรือมีคำอธิบายคล้ายกับโปรไฟล์ของผู้ซื้อ รวมถึงลักษณะของสิ่งที่ผู้ซื้อเคยใช้หรือเคยชอบ เพื่อเพิ่มโอกาสที่ผู้ซื้อจะซื้อสินค้ามากขึ้น โดยระบบสามารถทำงานเชื่อมโยงกันกับทุกฝ่ายในบริษัทฯ เช่น ฝ่ายบุคคล ฝ่ายการตลาด ฝ่ายการเงิน ฝ่ายคลังสินค้า เป็นต้น เพื่อให้ง่ายต่อการบริหารจัดการ

แนวคิดในการพัฒนาระบบ



รูปที่ 1-3 Conceptual Design ระบบขายสินค้าออนไลน์ บริษัท วนาวัฒน์วัสดุ จำกัด

ระบบแนะนำสินค้า



รูปที่ 1-4 ขั้นตอนการทำงานของระบบแนะนำสินค้า

หลักการทำงานระบบแนะนำสินค้า

1. เมื่อผู้ใช้เข้าสู่ระบบ จะกรอกข้อมูลสินค้าที่ต้องการค้นหาหรือให้คะแนนความชอบที่มีต่อสินค้านั้น ๆ
2. เมื่อผู้ใช้ให้คะแนนความชอบ ระบบจะนำคะแนนไปเก็บไว้ และสร้างเป็นโปรไฟล์ของผู้ใช้
3. เมื่อผู้ใช้ค้นหาสินค้าที่ต้องการจะซื้อ ระบบจะทำการหาค่าความแตกต่างความชอบของผู้ใช้กับผู้ใช้อื่นที่มีในระบบ
4. คำนวณหาค่าความเหมือนของผู้ใช้กับผู้ใช้อื่นที่มีอยู่ในระบบ
5. คำนวณค่าเฉลี่ยความชอบของผู้ใช้และผู้ใช้อื่นที่มีอยู่ในระบบ
6. ระบบจะแนะนำสินค้าโดยเรียงลำดับตามความนิยมจากมากไปน้อยและแยกเป็นหมวดหมู่สินค้า

แนวทางดำเนินงาน

1. ศึกษากระบวนการจำหน่ายสินค้าของบริษัท วนาวัฒน์วัสดุ จำกัด
2. ศึกษาฟังก์ชันการทำงานของเว็บไซต์และแอปพลิเคชันที่ใช้ในการซื้อขายสินค้าออนไลน์
3. ออกแบบ พัฒนาและติดตั้งระบบขายสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัด
4. วิจัยและพัฒนาระบบแนะนำสินค้า
5. วิเคราะห์ข้อมูลที่ได้จากระบบขายสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัด สรุปผลการดำเนินงาน

1.2 วัตถุประสงค์ในการวิจัย

1.2.1 เพื่อออกแบบ พัฒนาและติดตั้งระบบขายสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัด

1.2.2 เพื่อวิจัยและพัฒนาระบบแนะนำสินค้า (Recommendation System) สนับสนุนการขายสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัด

1.2.3 เพื่อปรับปรุงระบบขายสินค้าออนไลน์ ให้สนับสนุนการแนะนำสินค้า

1.2.4 เพื่อพัฒนางานวิจัยที่เกี่ยวข้องกับภาคอุตสาหกรรมโดยนักศึกษาปริญญาโท

1.3 ขอบเขตของการวิจัย

1.3.1 ศึกษาและพัฒนาระบบขายสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัด โดยมีคุณสมบัติประกอบด้วย

- สามารถบริหารจัดการในส่วนของการขายได้
- สามารถบริหารจัดการในส่วนของการเงินได้
- สามารถบริหารจัดการในส่วนของคลังสินค้าได้

1.3.2 วิจัยและพัฒนาระบบแนะนำสินค้า โดยมีคุณสมบัติ ประกอบด้วย

- สร้างโมเดลข้อมูล (Data model) สำหรับเลือกกลุ่มสินค้า โดยใช้ข้อมูลสินค้า (Product Profile)
- พัฒนาโมเดลการกรอง (Filtering model) 3 มิติในการแนะนำสินค้า โดยใช้ข้อมูลโปรไฟล์ลูกค้า (User Profile) ข้อมูลโมเดลข้อมูล (Context Profile) และข้อมูลรายการ (Item Profile)

1.4 ขั้นตอนและวิธีดำเนินงาน

1.4.1 กิจกรรม/ขั้นตอนการดำเนินงานปีที่ 1

ระบบขายสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัด มีขั้นตอนดังนี้

1. ลงพื้นที่วินิจฉัย
2. ออกแบบ พัฒนาและติดตั้งระบบขายสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัด
3. วิจัยและพัฒนาระบบแนะนำสินค้า (Recommendation System) สนับสนุนการขายสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัด
4. การทดสอบและวัดประสิทธิภาพการทำงานของระบบซื้อขายสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัด
5. จัดทำผลงานตีพิมพ์ทางวิชาการ
6. จัดทำรายงานและวิทยานิพนธ์

1.4.2 กิจกรรม/ขั้นตอนการดำเนินงานปีที่ 2

ระบบแนะนำสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัด มีขั้นตอนดังนี้

1. ออกแบบและพัฒนาระบบจัดเก็บข้อมูลพฤติกรรมของลูกค้า จากระบบขายสินค้าออนไลน์
2. จัดเตรียมข้อมูล Dataset เพื่อการออกแบบโมเดล
3. วิจัยและพัฒนา Data model สำหรับระบบแนะนำสินค้า
4. วิจัยและพัฒนา Filtering model สำหรับระบบแนะนำสินค้า
5. ออกแบบและพัฒนาระบบแนะนำสินค้า
6. ทดสอบและวัดประสิทธิภาพการทำงานของระบบแนะนำสินค้า
7. สรุปผลการทำงานของระบบแนะนำสินค้า
8. จัดทำผลงานตีพิมพ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ได้ระบบซื้อขายสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัด ที่มีลักษณะเฉพาะสามารถแนะนำสินค้าได้อย่างมีประสิทธิภาพ

1.5.2 ได้ระบบแนะนำสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัด ที่มีลักษณะเฉพาะสามารถแนะนำสินค้าได้อย่างมีประสิทธิภาพ

1.5.3 สามารถวัดประสิทธิภาพการทำงานของระบบซื้อขายสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัด ได้

1.5.4 สามารถวัดประสิทธิภาพการทำงานของระบบแนะนำสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัด ได้

1.5.5 เพิ่มทางเลือกในการซื้อสินค้าออนไลน์ให้กับผู้ซื้อ

1.5.6 เพิ่มช่องทางในการขายสินค้าให้กับผู้ประกอบการ

1.5.7 ผู้ซื้อจะได้รับการแนะนำสินค้าที่ตรงกับความต้องการของตนมากขึ้น

บทที่ 2
ทฤษฎีและหลักการที่เกี่ยวข้อง

ในงานวิจัยนี้เราได้แบ่งการดำเนินงานวิจัยออกเป็น 2 ขั้นตอนหลัก คือการวิจัยเกี่ยวกับการสร้างระบบแนะนำ และการสร้างโมเดลของข้อมูล ซึ่งสามารถแบ่งทฤษฎีที่เกี่ยวข้องได้ตามหัวข้อดังนี้

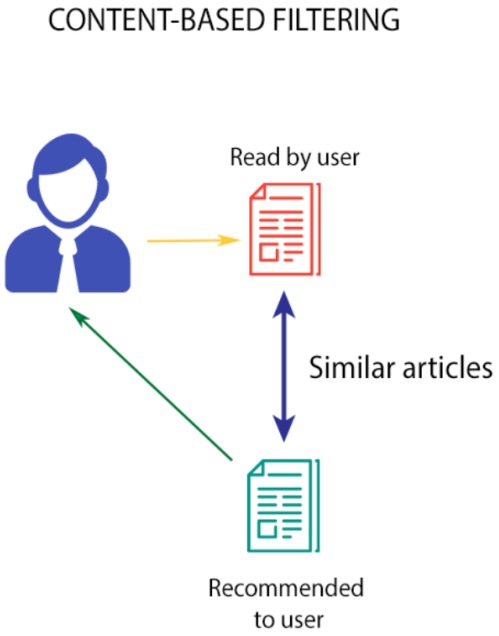
2.1 หลักการและทฤษฎีที่เกี่ยวข้อง

2.1.1 ระบบแนะนำ (Recommender System)

ระบบแนะนำ (Recommender System) เป็นระบบที่ถูกสร้างขึ้นโดยใช้เทคนิคการค้นหาความรู้จากข้อมูลที่มีไปวิเคราะห์เพื่อแนะนำสินค้าให้กับผู้ใช้ ซึ่งระบบสามารถวิเคราะห์ได้ว่าสินค้านั้นเหมาะสมกับผู้ใช้หรือไม่ เพื่อช่วยเพิ่มโอกาสที่ผู้ใช้จะซื้อสินค้ามากยิ่งขึ้น โดยจะนำข้อมูลได้แก่ ประวัติผู้ใช้ กิจกรรมต่างๆที่เกิดจากผู้ใช้ นำมาวิเคราะห์เพื่อหารายการสินค้าที่ผู้ใช้ชอบหรือสนใจ (วรากร ประดิษฐ์, ปราลี มณีรัตน์ และนิเวศ จิระวิชิตชัย, 2564)

2.1.1.1 วิธีการกรองเนื้อหา (Content-based Filtering)

วิธีการกรองเนื้อหาเป็นวิธีการสร้างรายการแนะนำข้อมูล โดยให้ความสำคัญกับคุณลักษณะของข้อมูลและสิ่งของเป็นหลัก โดยพิจารณาจากคุณสมบัติของสิ่งของที่ผู้ใช้เคยชอบในอดีตและประวัติการซื้อ จากนั้นระบบจะหาความสัมพันธ์ระหว่างสิ่งของกับผู้ใช้แล้วแนะนำสิ่งของที่มีคุณสมบัติตรงกับสิ่งที่ผู้ใช้เคยชอบในอดีต หากสิ่งของนั้นมีค่าใกล้เคียงกับคุณสมบัติและลักษณะตรงตามความชอบของผู้ใช้ ระบบจึงแนะนำข้อมูลเหล่านั้นให้กับผู้ใช้ (วรากร ประดิษฐ์ และคณะ, 2564) กระบวนการแสดงดังภาพประกอบที่ 2-1

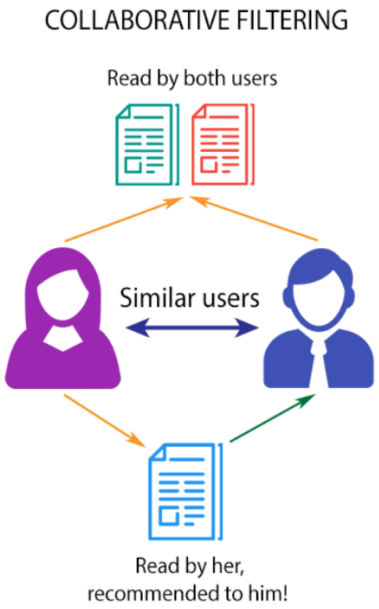


รูปที่ 2-1 วิธีการกรองเนื้อหา (Content-based Filtering)

ที่มา: M. Mohamed, M. Khafagy and H. Ibrahim. (2019)

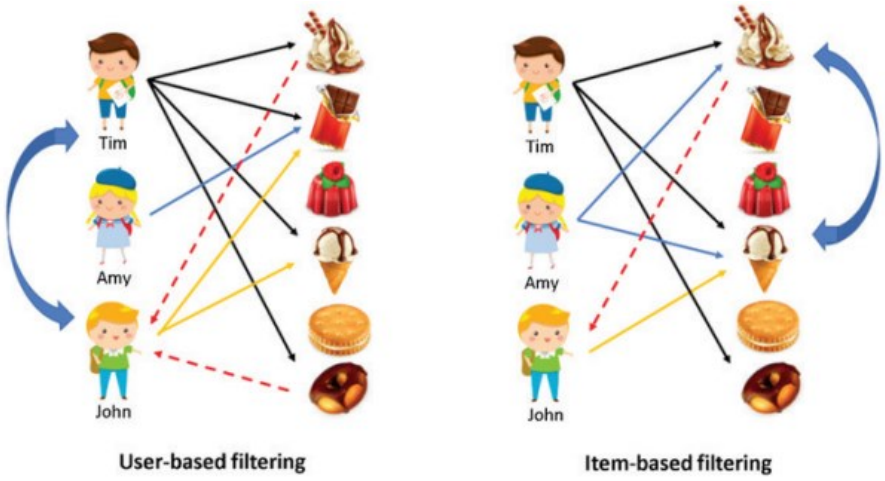
2.1.1.2 วิธีการกรองข้อมูลร่วม (Collaborative Filtering)

วิธีการกรองข้อมูลร่วมเป็นวิธีการสร้างรายการแนะนำข้อมูลที่นิยมมากที่สุด โดยพิจารณาจากประวัติการค้นหาของผู้ใช้คนอื่นที่มีพฤติกรรมคล้ายกัน ซึ่งเคยค้นหาและทำการประเมินหรือให้ค่าคะแนนความนิยมกับข้อมูลนั้นๆ ไว้ก่อนหน้านี้แล้ว วิธีการกรองร่วมเป็นการพิจารณาความคล้ายคลึงกันของผู้ใช้งานเป้าหมายกับผู้ใช้งานอื่นจากการให้คะแนนสินค้า (วรากร ประดิษฐ์ และคณะ, 2564) กระบวนการแสดงดังภาพประกอบที่ 2-2



รูปที่ 2-2 วิธีการกรองข้อมูลร่วม (Collaborative Filtering) ที่มา: M. Mohamed, M. Khafagy and H. Ibrahim. (2019)

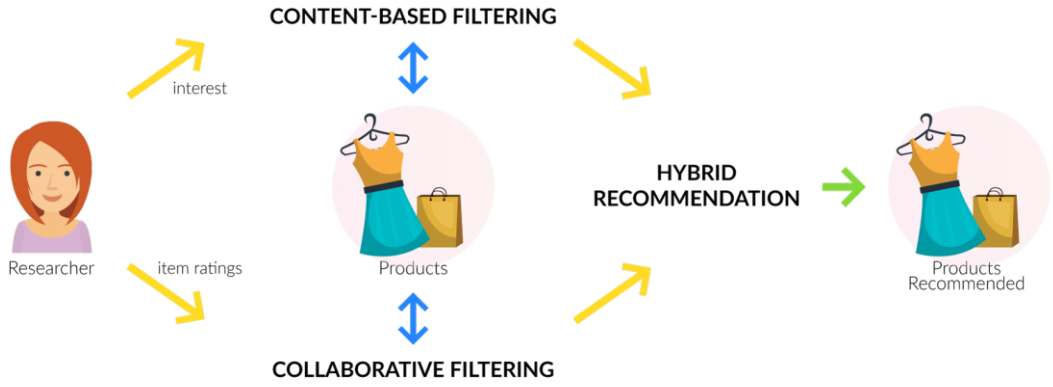
จากภาพประกอบที่ 2-2 วิธีการกรองแบบ Collaborative Filtering จำเป็นต้องอาศัยข้อมูลการให้คะแนนความชอบ (Rating) สินค้าเป็นพื้นฐานในการแนะนำโดยสามารถแบ่งเป็น 2 รูปแบบ ได้แก่ (1) User-based filtering เป็นการคำนวณหาความล้ายคลึงของผู้ใช้ทั้งหมด โดยเริ่มจากการจัดเรียงตามคะแนนของสินค้า และกรองเฉพาะผู้ใช้ที่มีค่าสมาชิกที่ใกล้เคียง (K-nearest neighbor) เพื่อทำนายรายการสินค้านั้น และแนะนำให้กับผู้ใช้ และ (2) item-based filtering วิธีนี้จะคล้ายกับ User-based filtering โดยจะเป็นการคำนวณหาความคล้ายคลึงกันระหว่างสินค้าเพื่อแนะนำสินค้าที่คล้ายกัน



รูปที่ 2-3 การทำงานของ User-based filtering กับ Item-based filtering
ที่มา: วรากร ประดิษฐ์ และคณะ (2564)

2.1.1.3 วิธีการแบบผสมผสาน (Hybrid Filtering)

วิธีการแบบผสมผสานเป็นการที่ผสมผสานวิธีการข้างต้นเข้าด้วยกัน เช่น การใช้วิธีการกรองแบบเนื้อหาร่วมกับวิธีการกรองแบบร่วม หรือการนำวิธีการแก้ไขปัญหาจากที่อื่นที่ใช้วิธีการแก้ไขปัญหาตั้งแต่ 2 เทคนิคขึ้นไปมาสร้างการแนะนำข้อมูลหรือสินค้าให้แก่ผู้ใช้งาน เพื่อลดข้อเสียและเสริมข้อดีของทั้ง 2 เทคนิค เพื่อแก้ไขปัญหาที่อาจจะเกิดขึ้นจากผู้ใช้ใหม่ที่ไม่เคยมีข้อมูลในระบบ (Cold-start problem) และจำนวนของช่องข้อมูลที่ว่าง (Sparsity problem) (วรากร ประดิษฐ์ และคณะ, 2564)



รูปที่ 2-4 วิธีการแบบผสมผสาน (Hybrid Filtering)
ที่มา: วรากร ประดิษฐ์ และคณะ (2564)

2.1.2 การหาค่าความคล้ายคลึง (Similarity measure)

2.1.2.1 Euclidean Distance

วิธีการวัดระยะทางแบบยุคลิดคือพื้นฐานของการวัดค่าความคล้ายคลึง และความแตกต่างระหว่างเวกเตอร์ x และ y โดยใช้สมการดังนี้

$$\boldsymbol{d(x,y) = \sqrt{\sum_i^n (x_i - y_i)^2}}$$

โดยที่ $\boldsymbol{x_i}$ และ $\boldsymbol{y_i}$ คือคะแนนการให้คะแนนของรายการที่กำหนดโดยผู้ใช้ที่แตกต่าง กันสองคนสำหรับรายการเดียวกัน

$\boldsymbol{n}$ คือจำนวนของรายการที่ให้คะแนนโดยทั่วไป

กล่าวคือระยะทางแบบยุคลิด คือรากที่สองของผลรวมของผลต่างกำลังสองระหว่างองค์ประกอบที่สอดคล้องกันของเวกเตอร์ 2 ตัว สำหรับสูตรใช้ค่า X และ Y โดยไม่มีการปรับเปลี่ยนสำหรับความแตกต่างของมาตราส่วน ระยะทางแบบยุคลิดเหมาะสำหรับข้อมูลที่วัดในระดับเดียวกันเท่านั้น

2.1.2.2 Cosine Similarity

วิธีการหาค่าความคล้ายคลึงกันของโคไซน์เป็นวิธีที่ใช้หาค่าคล้ายคลึงกันระหว่าง 2 อินสแตนซ์ a และ b ในการดึงข้อมูลที่วัตถุอยู่ในรูปของเวกเตอร์ xa และเวกเตอร์ xb และคำนวณค่าโคไซน์เวกเตอร์ (Cosine Vector) หรือปริภูมิเวกเตอร์ (Vector Space) ที่คล้ายคลึงกันเพื่อหาค่าระยะห่างระหว่างเวกเตอร์ โดยใช้สมการดังนี้

$$\frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

โดยที่ $\boldsymbol{A_i}$ คือคะแนนของผู้ใช้ A

$\boldsymbol{B_i}$ คือคะแนนของผู้ใช้ B

$\boldsymbol{n}$ คือจำนวนรายการที่ได้รับการจัดอันดับทั่วไป

2.1.2.3 Jaccard coefficient

สัมประสิทธิ์ความคล้ายคลึงกันของ Jaccard หรือเรียกอีกอย่างว่าสัมประสิทธิ์ความคล้ายคลึงกันของ Tanimoto เป็นสถิติที่ใช้สำหรับวัดความคล้ายคลึงด้วยจุดตัดที่หารด้วยการรวมกันของวัตถุสำหรับเอกสารข้อความ สัมประสิทธิ์ความคล้ายคลึงกันของ Jaccard จะเปรียบเทียบน้ำหนักรวมของคำที่ใช้ร่วมกันกับน้ำหนักรวมของคำที่มีอยู่ในเอกสารทั้ง 2 ฉบับแต่ไม่ใช่คำที่ใช้ร่วมกัน โดยใช้สมการดังนี้

$$SIM_J(\overrightarrow{t_a}, \overrightarrow{t_b}) = \frac{\overrightarrow{t_a} \cdot \overrightarrow{t_b}}{|\overrightarrow{t_a}|^2 + |\overrightarrow{t_b}|^2 - \overrightarrow{t_a} \cdot \overrightarrow{t_b}}$$

โดยที่ $\boldsymbol{t_a}$ และ $\boldsymbol{t_b}$ คือการจัดอันดับของผู้ใช้ A และผู้ใช้ B สำหรับการรวมรายการระหว่างผู้ใช้

สัมประสิทธิ์ความคล้ายคลึงกันของ Jaccard เป็นมาตรวัดความคล้ายคลึงและมีค่าอยู่ระหว่าง 0 ถึง 1 โดยค่าที่เท่ากับ 1 หมายถึงวัตถุทั้งสองมีความคล้ายคลึงกัน และค่าเท่ากับ 0 หมายถึงวัตถุทั้งสองไม่มีความคล้ายคลึงกัน การวัดระยะทางที่สอดคล้องกันคือ DJ = 1 – SIMJ และเราจะใช้ DJ แทนในการทดลองครั้งต่อไป

2.1.3 พาณิชย์อิเล็กทรอนิกส์ (E-Commerce)

พาณิชย์อิเล็กทรอนิกส์ (วรินทร์พิพัชร วัชรพงษ์เกษม, 2560) เป็นการดำเนินธุรกรรมทางการค้าผ่านอิเล็กทรอนิกส์ ได้แก่การซื้อสินค้า การขายสินค้า การจัดส่งสินค้า การแลกเปลี่ยนสินค้าหรือบริการ และการสารสนเทศผ่านอินเทอร์เน็ต ซึ่งอาจมองเฉพาะคำว่า Commerce เป็นความหมายหลักทั้งหมดของการทำธุรกรรมทางการค้า (Transaction) และหากใช้คำนี้เป็นคำนิยามหลัก จะทำให้ความหมายของพาณิชย์อิเล็กทรอนิกส์ถูกจำกัดความหมายให้เล็กลงไปอีก ดังนั้นหนังสือหรือเอกสารต่างๆมักจะใช้คำว่า ธุรกรรมอิเล็กทรอนิกส์ เนื่องจากครอบคลุมความหมายที่กว้างกว่าพาณิชย์อิเล็กทรอนิกส์

ประเภทของพาณิชย์อิเล็กทรอนิกส์

ปัจจุบันเทคโนโลยีพาณิชย์อิเล็กทรอนิกส์ได้ถูกนำมาประยุกต์ใช้งานหลายประเภท ประกอบด้วยธุรกิจพาณิชย์อิเล็กทรอนิกส์แบบหวังผลกำไร และธุรกิจพาณิชย์อิเล็กทรอนิกส์แบบไม่หวังผลกำไร โดยประเภทของพาณิชย์อิเล็กทรอนิกส์มีดังต่อไปนี้

1) Business-to-Consumer หรือ B2C เป็นการดำเนินธุรกรรมระหว่างผู้ประกอบการกับผู้บริโภค โดยผู้ประกอบการเป็นผู้ขายสินค้า และผู้บริโภคคือลูกค้าซึ่งเป็นผู้ซื้อสินค้า พาณิชย์อิเล็กทรอนิกส์ประเภท B2C หรือสามารถเรียกอีกอย่างได้ว่า Electronic Retailing หรือ e-Tailing ซึ่งเป็นวิธีการขายตรงด้วยการซื้อสินค้าหรือบริการผ่านร้านค้าอิเล็กทรอนิกส์ เช่น การสั่งซื้อสินค้าผ่านเว็บไซต์โดยเว็บไซต์ดังกล่าวจะมีการเตรียมสินค้าที่หลายหมวดหมู่สำหรับให้ลูกค้าคลิกเพื่อเลือกชมค้นหา และเลือกซื้อได้ตามใจชอบ สิ่งสำคัญของพาณิชย์อิเล็กทรอนิกส์ประเภท B2C คือ ผู้บริโภคสามารถดำเนินการซื้อสินค้าหรือบริการกับทางร้านค้าได้โดยตรงไม่ผ่านคนกลาง ทำให้ราคาสินค้าหรือบริการมีราคาถูกกว่าท้องตลาดทั่วไป

2) Business-to-Business หรือ B2B เป็นการดำเนินธุรกรรมระหว่างผู้ประกอบการกับผู้ประกอบการด้วยกัน ซึ่งทั้งสองสามารถเป็นได้ทั้งผู้ขายและผู้ซื้อ การดำเนินธุรกรรมแบบ B2B จะส่งผลทำให้เกิดคู่ค้าทางธุรกิจที่ช่วยสร้างความสัมพันธ์ที่ดีร่วมกันได้ เนื่องจากการดำเนินธุรกิจในปัจจุบันจะเติบโตได้นั้นจำเป็นต้องพึ่งพาอาศัยซึ่งกันและกัน

3) Consumer-to-Consumer หรือ C2C เป็นการดำเนินธุรกรรมระหว่างผู้บริโภคกับผู้บริโภคด้วยกันเอง การดำเนินการซื้อขายสินค้าด้วยวิธีนี้มีวัตถุประสงค์ที่แตกต่างกันไม่ว่าจะเป็นการซื้อขายสินค้ามือสองหรือการแลกเปลี่ยนสินค้า ซึ่งหากทั้งผู้ซื้อและผู้ขายมีความพอใจในสินค้าก็จะตกลงซื้อขายกันเอง จากนั้นก็ทำการนัดสถานที่เพื่อชำระเงินหรือเพื่อแลกเปลี่ยนสินค้า

4) Consumer-to-Business หรือ C2B เป็นการดำเนินธุรกรรมระหว่างผู้บริโภคกับผู้ประกอบการ โดยที่ผู้บริโภคกลับมีสถานะเป็นผู้ค้า และมีหน้าที่ในการต่อรองเพื่อตั้งราคาสินค้า ซึ่งผู้ประกอบการจะเป็นตัวกลางนำเสนอราคาของลูกค้าส่งให้กับผู้ขายพิจารณาว่าสามารถจำหน่ายในราคานี้ได้หรือไม่

5) Business-to-Employee หรือ B2E เป็นการดำเนินธุรกิจอิเล็กทรอนิกส์สำหรับปรับปรุงแก้ไขในด้านการปฏิบัติงานให้ดีขึ้น ซึ่งจะนำมาใช้สำหรับการแลกเปลี่ยนข่าวสารและสารสนเทศภายในองค์กร ซึ่งทำให้พนักงานสามารถรับทราบข่าวสารเหล่านั้นและนำไปปรับปรุงแก้ไขได้ นอกจากนี้ยังช่วยลดปริมาณงานทางด้านเอกสาร โดยจะนำเอกสารอิเล็กทรอนิกส์มาใช้งานแทนซึ่งมีส่วนช่วยลดปริมาณค่าใช้จ่ายได้เป็นอย่างดี

6) e-Government เป็นการดำเนินธุรกรรมที่ภาครัฐนำสื่ออิเล็กทรอนิกส์มาบริการให้แก่ประชาชนเพื่อแก้ไขปรับปรุงในด้านการบริการทั่วไป โดยงานเหล่านี้จัดเป็นส่วนหน้าร้านหรือเรียกอีกอย่างหนึ่งว่า Front Office และอีกทั้งยังสามารถนำไปใช้เพื่อบริหารงานของภาครัฐเองได้ด้วย ซึ่งจัดเป็นส่วนหลังร้านหรือเรียกอีกอย่างหนึ่งว่า Back Office สำหรับการบริการให้แก่ประชาชน โดยรูปแบบการให้บริการประชาชนสามารถเข้ามาใช้บริการได้ ซึ่งสิ่งเหล่านี้เป็นผลดีที่เกิดขึ้นระหว่างภาครัฐกับประชาชน เช่นการบริการให้ประชาชนสามารถยื่นแบบการชำระภาษีผ่านระบบอินเทอร์เน็ตได้โดยไม่จำเป็นต้องเดินทางมายังสำนักงานของกรมสรรพากร ซึ่งนอกจากช่วยอำนวยความสะดวกแล้วยังช่วยให้ประชาชนประหยัดค่าใช้จ่ายด้วย

ประโยชน์ของพาณิชย์อิเล็กทรอนิกส์

1) ประโยชน์ต่อบุคคล คือทำให้ลูกค้าสามารถดำเนินธุรกรรมต่างๆ ได้ 24 ชั่วโมง และสามารถรับรู้ข้อมูลเกี่ยวกับสินค้าและบริการได้ในเวลาที่รวดเร็ว สามารถเลือกซื้อสินค้าที่ตรงตามความต้องการมากที่สุด สามารถแสดงความคิดเห็นกับลูกค้ารายอื่นได้ และทำให้เกิดการดำเนินงานภายในโซ่มูลค่าหรือ Value Chain Integration

2) ประโยชน์ต่อองค์กรธุรกิจ คือช่วยในด้านการขยายตลาดในระดับประเทศและระโลกได้อย่างรวดเร็วขึ้น และยังช่วยลดต้นทุนในการบริการลูกค้าได้จำนวนมาก ลดต้นทุนในด้านเอกสารเกี่ยวกับข้อง ลดต้นทุนทางด้านการสื่อสารโทรคมนาคมโดยใช้การดำเนินธุรกรรมผ่านอินเทอร์เน็ตแทน และยังเพิ่มโอกาสในการแข่งขันสำหรับบริษัทขนาดเล็กได้อีกด้วย

3) ประโยชน์ต่อสังคม คือช่วยลดปริมาณลดการจราจรและลดปัญหามลพิษทางอากาศในการเดินทางให้น้อยลงเนื่องจากไม่ต้องเดินทางไปซื้อสินค้า และยังช่วยส่งเสริมโอกาสในการขายสินค้าและบริการให้กับคนในหลายระดับ

4) ประโยชน์ต่อระบบเศรษฐกิจ คือเพิ่มโอกาสทางการตลาดทั้งในประเทศและต่างประเทศสำหรับสินค้า SMEs ช่วยลดบทบาทของพ่อค้าคนกลางทำให้ประชาชนสามารถเข้าถึงสินค้าและบริการได้โดยตรง และยังช่วยลดช่องว่างในการเข้าถึงสินค้าของผู้คนในชุมชนให้ดียิ่งขึ้น

2.1.4 โปรแกรมประยุกต์สำหรับสร้างและบริหารจัดการเว็บไซต์

CMS (ปาลิตา แสงศิริ, 2561) ย่อมาจาก Content Management System เป็นระบบที่ใช้ในการสร้างและบริหารจัดการข้อมูลของเว็บไซต์แบบสำเร็จรูป เช่น ข้อมูลตัวอักษร รูปภาพ หรือวิดีโอ ผู้ใช้ไม่จำเป็นต้องมีความรู้ในด้านการเขียนหรือพัฒนาเว็บไซต์ก็สามารถสร้างเว็บไซต์ได้

ข้อดีของ CMS

1) ผู้ใช้งานไม่จำเป็นต้องมีความรู้ในด้านการสร้างเว็บไซต์ก็สามารถใช้งานได้
2) พัฒนาเว็บได้อย่างรวดเร็ว
3) สามารถบริหารจัดการ และดูแลเว็บไซต์ได้ง่ายเพราะมีเครื่องมืออำนวยความสะดวกครบถ้วนอยู่แล้ว

ข้อเสียของ CMS

1) สำหรับผู้ที่ไม่ชำนาญในด้านคอมพิวเตอร์ จะมีความยุ่งยากในการติดตั้งกับเว็บเซิร์ฟเวอร์
2) หากผู้ใช้ต้องการพัฒนารูปแบบของเว็บ จะต้องมีความรู้พอสมควรเนื่องจาก CMS เป็นการนำระบบหลายระบบมารวมเข้าด้วยกัน

2.1.5.1 เปรียบเทียบข้อดีข้อเสียของ CMS

ตารางที่ 2-1 การเปรียบเทียบข้อดีข้อเสียของระบบ CMS

CMS	ข้อดี	ข้อเสีย
WordPress	1. ใช้งานง่ายสำหรับมือใหม่ 2. มีลูกเล่นและ Module เสริมเยอะ 3. ใช้พื้นที่ในการติดตั้งน้อย	1. หากต้องการทำเว็บขนาดใหญ่มีเนื้อหาหลายชนิดอาจไม่รองรับ 2. การแก้ไข Template ทำได้ค่อนข้างยาก
Drupal	1. Module, Plugin ฟรีเกือบทั้งหมด 2. ระบบป้องกันความปลอดภัย	1. มี Template ให้เลือกน้อย 2. ใช้งานค่อนข้างยากสำหรับผู้เริ่มใช้
Moodle	1. มีความยืดหยุ่นในการปรับแต่งแก้ไข 2. ใช้งานค่อนข้างง่ายสำหรับผู้เริ่มใช้	1. ยังมีผู้พัฒนาน้อย 2. มี Template ให้เลือกน้อย

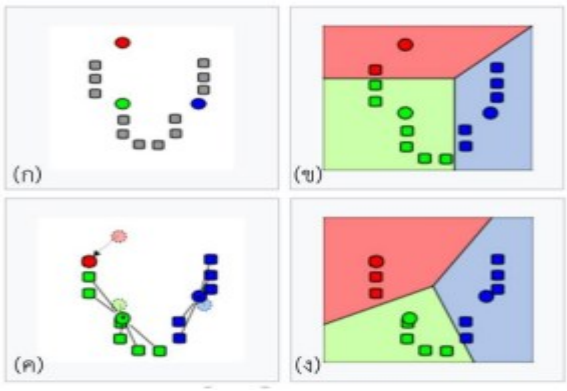
	3. รองรับมาตรฐาน E-learning เหมาะกับการทำสื่อแบบ E-learning	
October CMS	1. ใช้งานค่อนข้างง่ายและรวดเร็วในการสร้างเว็บไซต์ 2. มีเครื่องมืออำนวยความสะดวกครบถ้วน	1. หากต้องการทำเว็บขนาดใหญ่มีเนื้อหาหลายชนิดอาจไม่รองรับ 2. มี Template ให้เลือกน้อย

2.2 ทฤษฎีที่เกี่ยวข้องด้านเทคนิค

2.2.1 การแบ่งกลุ่มข้อมูลแบบเคมีน (K-means Clustering)

การแบ่งกลุ่มข้อมูลแบบเคมีน (สุชาวดี ปลั่งศรี, 2562) เป็นขั้นตอนวิธีการจัดกลุ่มประเภท Unsupervised Machine Learning โดยใช้วิธีการแบ่งกั้น (Partitioning) ซึ่งระเบียนข้อมูลจะถูกแบ่งกั้นเป็นกลุ่มที่ไม่มีสมาชิกร่วมกัน โดยใช้การกั้นระหว่างกลุ่มด้วยระยะห่างแบบ Euclidean distance วิธีการแบ่งกั้นนั้นจะกำหนดให้ข้อมูล n ระเบียน แบ่งเป็น K กลุ่มที่ไม่มีสมาชิกร่วมกัน วิธีการจัดกลุ่มโดยใช้วิธีแบ่งกั้นมีขั้นตอนดังนี้

1) แบ่งกลุ่มข้อมูลเป็น k กลุ่มที่ไม่ใช่เซตว่าง
2) คำนวณจุดกึ่งกลาง(centroid) ของกลุ่ม โดยใช้ค่าเฉลี่ยเลขคณิต(mean)
3) สำหรับแต่ละระเบียน นำระเบียนเทียบกับจุดกึ่งกลาง เพื่อกำหนดกลุ่มให้กับระเบียน โดยเลือกระยะจากระเบียนไปจุดกึ่งกลางที่ใกล้ที่สุด
4) วนซ้ำจนกระทั่งไม่มีการเปลี่ยนกลุ่มของระเบียน หรือครบจำนวนรอบสูงสุดที่กำหนดไว้



รูปที่ 2-5 ตัวอย่างการจัดกลุ่มด้วยวิธี K-mean Clustering

จากรูปที่ 2-5 ขั้นตอนจะเริ่มจากรูป ก เลือกค่าเฉลี่ยเริ่มต้น k (ในกรณีนี้ k=3) โดยการสุ่มจากข้อมูล หลังจากนั้นรูป ข จะทำการสร้างกลุ่ม k กลุ่ม โดยคำนวณระยะห่างของข้อมูล รูป ค จุดกึ่งกลางของแต่ละกลุ่มจะถูกกำหนดเป็นค่าเฉลี่ยค่าใหม่ และสุดท้ายรูป ง ทำซ้ำขั้นตอน ข และ ค จนกระทั่งค่ากลางของแต่ละกลุ่มไม่เปลี่ยนแปลง

2.2.1.1 ข้อมูลที่ใช้ประกอบการจัดกลุ่ม

1) ตัวแปรที่ใช้ในการรวมกลุ่ม
2) จำนวนกลุ่มเป้าหมายที่ต้องการ
3) วิธีการวนซ้ำ (iterate) หรือจำแนกข้อมูล (classify)
4) วิธีการเลือกค่ากึ่งกลาง

โดยตัวแปรที่ใช้ในวิธีจัดกลุ่มแบบ K-means Clustering จะต้องเป็นตัวแปรเชิงปริมาณ คือเป็นสเกลอันตรภาค (Interval Scale) หรือสเกลอัตราส่วน (Ration Scale) ซึ่งจะไม่สามารถใช้กับข้อมูลความถี่ หรือ Binary ได้

2.2.1.2 การคำนวณระยะห่างของข้อมูล

การคำนวณระยะห่างของข้อมูลจะใช้วิธีการคำนวณแบบยุคลิด โดยขั้นตอนการคำนวณจะเริ่มจากหาค่ากลางเป็นค่าตั้งต้นก่อน 1 ชุด จากนั้นดำเนินการจัดกลุ่มข้อมูลโดยคำนวณจากระยะห่างจากศูนย์กลาง และต่อมาให้ดำเนินการคำนวณหาค่ากลางอีก 1 ชุดหนึ่งขึ้นมา โดยจะมีการเปลี่ยนแปลงข้อมูลในกลุ่มโดยคำนวณจากระยะห่างจากศูนย์กลางค่าใหม่ ดำเนินการซ้ำจนกระทั่งค่ากลางของกลุ่มไม่มีการเปลี่ยนแปลง

วิธีการพิจารณากลุ่มจะขึ้นอยู่กับความคล้ายคลึงกันของข้อมูลโดยจะคำนวณระยะห่างระหว่างแต่ละคู่ ซึ่งวิธีการคำนวณระยะห่างแบบยุคลิด สามารถคำนวณได้จากสมการดังนี้

$$\boldsymbol{d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}}$$

p และ q คือความยาวของส่วนของเส้นตรง p q ถ้า p = (p1, p2, ..., p_n) และ q = (q1,q2, ..., q_n) และ n คือจำนวนคุณลักษระของข้อมูลแต่ละชุด โดยคุณสมบัติของระยะห่าง คือ

1) ระยะห่างเป็นบวกเสมอ
2) ระยะห่างจากจุดหนึ่งไปยังจุดตัวเอง คือศูนย์

3) ระยะห่างจาก p ถึง q จะเท่ากับ q ถึง p เสมอ
4) ระยะห่างจากจุด p ถึง q ไม่สามารถมากกว่าผลรวมของระยะห่างจาก p ไปยังจุด r และระยะห่างจาก q ถึง p

2.2.2 Term Frequency Inverse Document Frequency (TF-IDF)

TF-IDF (ทัศนีย์ อุทัยสุริ, 2556) เป็นวิธีการถ่วงน้ำหนักที่นิยมใช้กันมากในการทำงานด้าน Information Retrieval และ Text mining ซึ่งเป็นวิธีการทางด้านสถิติที่ใช้ในการวิเคราะห์หาค่าความสำคัญของคำต่อเอกสาร โดยค่าความสำคัญของคำเป็นสัดส่วนโดยตรงกับจำนวนครั้งที่คำคำนั้นปรากฏในเอกสารนั้น แต่ความสำคัญถูกลดเนื่องจากความถี่ของคำนั้นในกลุ่มเอกสารทั้งหมด โดยมีสูตรในการคำนวณดังนี้

หาค่า TF (Term Frequency) ได้จากสมการ $\boldsymbol{tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}}$

หาค่า IDF (Inverse Document Frequency) ได้จากสมการ $\boldsymbol{idf_{i,j} = \log \frac{|D|}{|\{d_j : t_j \in d_j\}|}}$

หาค่า TF-IDF ได้จากสมการ $\boldsymbol{tfidf_{i,j} = tf_{i,j} \times idf_i}$

เมื่อ TF คือจำนวนครั้งที่คำคำนั้นปรากฏอยู่ในเอกสารที่สนใจ โดยจากสูตร

$\boldsymbol{n_{i,j}}$ คือจำนวนการปรากฏของคำคำนั้นในเอกสารที่ $\boldsymbol{d_j}$

$\boldsymbol{\sum_k n_{k,j}}$ คือจำนวนการปรากฏของทุกคำในเอกสารที่ $\boldsymbol{d_j}$

IDF คือการวัดค่าความสำคัญของคำ เพื่อลบคำที่พบในหลายๆเอกสาร ซึ่งคำนวณหาได้จากการนำจำนวนเอกสารทั้งหมดในกลุ่มเอกสาร มาหารด้วยจำนวนเอกสารที่พบคำที่สนใจ

$\boldsymbol{|D|}$ คือจำนวนเอกสารทั้งหมดในกลุ่มเอกสาร

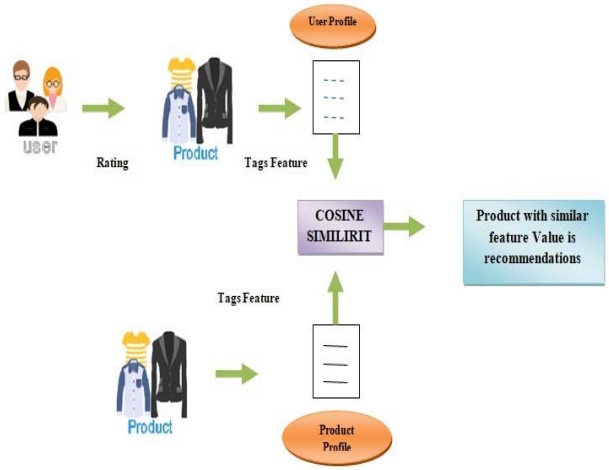
$\boldsymbol{|\{d_j : t_j \in d_j\}|}$ คือจำนวนเอกสารที่คำคำนั้นปรากฏ

ค่าน้ำหนัก TF-IDF จะมีค่าสูงเมื่อค่า Term Frequency มีค่าสูง และ Document Frequency มีค่าต่ำ

2.3 งานวิจัยที่เกี่ยวข้อง

2.3.1 ระบบแนะนำ (Recommendation System)

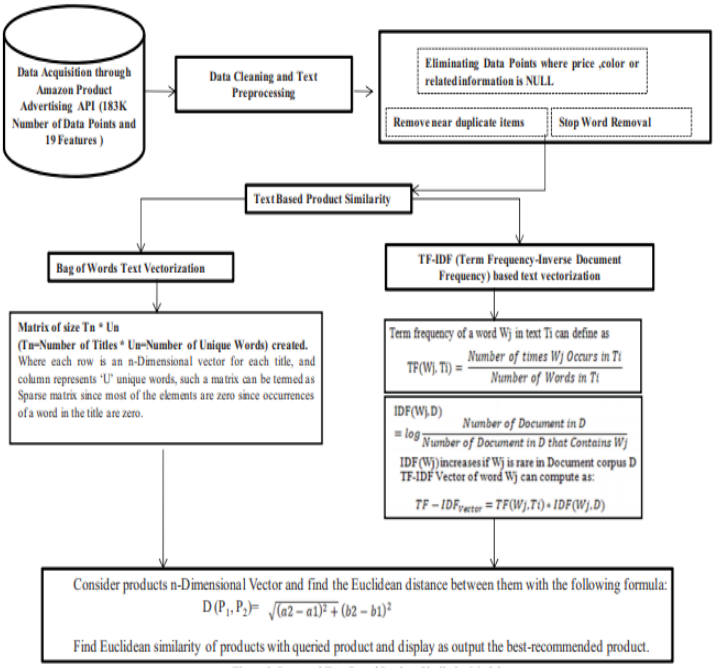
Yadav และคณะ [8] ได้เสนอระบบแนะนำแบบไฮบริด ที่ใช้ข้อมูลโปรไฟล์ผู้ใช้และโปรไฟล์ผลิตภัณฑ์ ได้แก่รายละเอียดของผลิตภัณฑ์ โดยข้อมูล Tags Feature จะถูกสร้างขึ้นโดยอัตโนมัติโดยใช้วิธี Text TF-IDF (term frequency-inverse document frequency) เป็นองค์ประกอบของการกรองตามเนื้อหา แท็กผลิตภัณฑ์ถูกรวมเข้ากับโปรไฟล์ผู้ใช้โดยใช้วิธีการคล้ายคลึงกันของโคไซน์ มีการใช้กันอย่างแพร่หลายในการกรองการทำงานร่วมกัน วิธีการกรองทั้งสองนี้ถูกใช้ร่วมกันเพื่อเพิ่มความแม่นยำของระบบการแนะนำดังภาพประกอบที่ 2-6 แนวคิดหลักของระบบคำแนะนำ



รูปที่ 2-6 แนวคิดหลักของระบบคำแนะนำ

ผลจากการทดลองรายการสินค้าที่แนะนำมีความคล้ายคลึงกับคำอธิบายผลิตภัณฑ์ และการตั้งค่าโปรไฟล์ผู้ใช้ที่มีค่าความแม่นยำเฉลี่ยอยู่ที่ 69.7% และมีค่า recall 73.63%

Shrivastava และคณะ [9] ได้เสนอแนวทางการแนะนำผลิตภัณฑ์โดยใช้แบบจำลองการเรียนรู้ตามความคล้ายคลึงกันของข้อความ การใช้เทคนิค Bag of Words (BOW) และ Term Frequency-Inverse Document Frequency (TF-IDF) ซึ่งมีขั้นตอนการดำเนินการ ดังภาพประกอบที่ 2-7

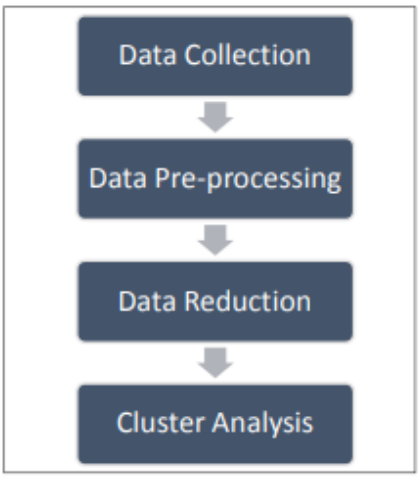


รูปที่ 2-7 รูปแบบความคล้ายคลึงกันของผลิตภัณฑ์ตามข้อความ

คำอธิบายของผลิตภัณฑ์ที่แปลงเป็นเวกเตอร์ n มิติ และสามารถวัดความคล้ายคลึงโดยใช้วิธี Euclidean similarity ระหว่างเวกเตอร์ n มิติของผลิตภัณฑ์ที่สืบค้นกับผลิตภัณฑ์อื่นๆ ความคล้ายคลึงกันของผลิตภัณฑ์แบบข้อความโดยใช้เทคนิคการสร้างเวกเตอร์ข้อความมีประโยชน์มากในการดำเนินการแนะนำผลิตภัณฑ์ตามเนื้อหาและแนะนำรายการที่คล้ายคลึงกันให้กับผู้ใช้ สามารถใช้ในแอปพลิเคชันอีคอมเมิร์ซต่างๆ เนื่องจากแอปพลิเคชันเหล่านี้มีคำอธิบายของผลิตภัณฑ์อยู่เป็นจำนวนมาก Bag of words และ TF-IDF สร้างเวกเตอร์ n- มิติของคำอธิบายที่เป็นข้อความที่แตกต่างกันของผลิตภัณฑ์ ซึ่งจะนำไปสู่การแนะนำผลิตภัณฑ์ที่ดีขึ้น ผลการทดลองและส่วนการวิเคราะห์อธิบายอย่างชัดเจนว่าแบบจำลองที่เสนอสำหรับผลิตภัณฑ์ค้นหาความคล้ายคลึงกันของผลิตภัณฑ์กับผลิตภัณฑ์ที่สอบถามและแสดงผลเป็นผลิตภัณฑ์ที่แนะนำที่ดีที่สุด

2.3.2 ระบบจัดกลุ่มรายการสินค้า

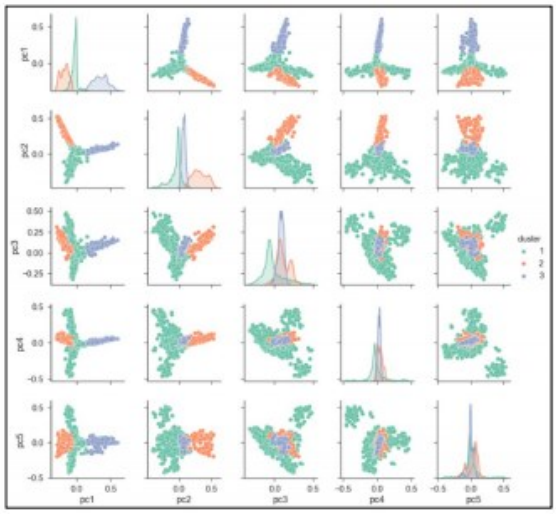
Mathivanan และคณะ [10] ได้เสนอแนวทางการใช้อัลกอริทึมการจัดกลุ่ม K-Means ในการจัดหมวดหมู่ผลิตภัณฑ์อีคอมเมิร์ซขนาดใหญ่ มีหลายขั้นตอนที่เกี่ยวข้องกับการจัดกลุ่มข้อมูลตามที่แสดงดังภาพประกอบที่ 2-8 เป็นขั้นตอนพื้นฐานในการวิจัยที่เกี่ยวข้องกับการทำเหมืองข้อความ



รูปที่ 2-8 กระบวนการจัดกลุ่มสำหรับข้อมูลผลิตภัณฑ์อีคอมเมิร์ซ

จากรูปที่ 2-8 การเก็บรวบรวมข้อมูลได้รวบรวมข้อมูลผลิตภัณฑ์จากเว็บไซต์ร้านค้าออนไลน์แสดงรายละเอียดของคลังข้อมูลที่รวบรวม มีข้อมูลผลิตภัณฑ์จำนวน 56176 รายการ และมีข้อมูล Features จำนวน 1596 รายการ Features คือการนับจำนวนคำที่รวมอยู่ในคลังข้อมูล หลังจากใช้ Principal component analysis (PCA) มิติของข้อมูลลดลงเหลือ 1234 รายการ องค์ประกอบหลักเป็นคุณลักษณะสำหรับการจัดกลุ่มชุดข้อมูล จากนั้นการประมวลผลข้อมูลล่วงหน้าเป็นขั้นตอนสำคัญในการกำหนดมาตรฐานข้อมูลบางส่วนและอยู่ในรูปแบบที่เหมาะสม การลดข้อมูลประกอบด้วยการดึงข้อมูลคุณลักษณะและขั้นตอนการเลือกคุณลักษณะที่ใช้ Term Frequency-Inverse Document Frequency (TFIDF) สำหรับการดึงข้อมูล Feature extraction ในขณะเดียวกัน จะใช้การวิเคราะห์องค์ประกอบหลัก PCA เพื่อทำการเลือกคุณสมบัติ จากนั้น ส่วนประกอบหลักที่เลือกจะถูกใช้เป็น

ข้อมูลสำหรับการจัดกลุ่มโดยใช้อัลกอริทึมการจัดกลุ่ม K-Means ผลการทดลองแสดงให้เห็นว่าข้อมูลสามารถจัดกลุ่มได้เป็น 3 กลุ่ม โดยมีลักษณะเฉพาะแตกต่างกันดังแสดงในภาพประกอบที่ 2-9



รูปที่ 2-9 รูปแบบของจัดกลุ่มด้วย K=3

บทที่ 3
วิธีการดำเนินการวิจัย

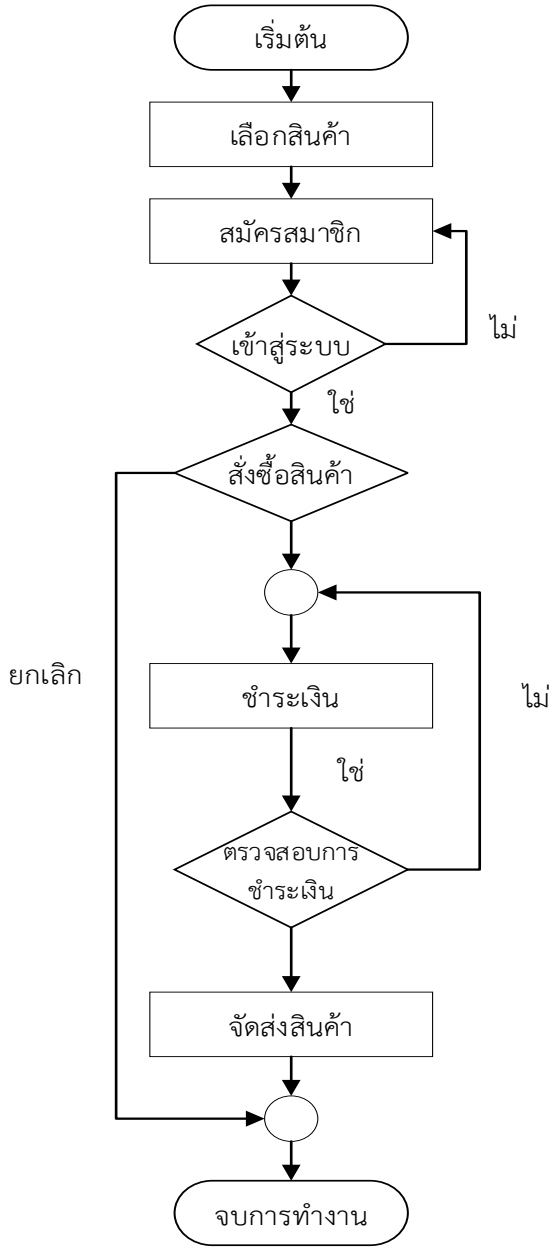
การดำเนินโครงการครั้งนี้เป็นการวิจัยและพัฒนาระบบแนะนำสินค้าแบบการรับรู้บริบทสำหรับอีคอมเมิร์ซ บริษัท วนาวัฒน์วัสดุ จำกัด ซึ่งมีวัตถุประสงค์เพื่อออกแบบ วิจัย พัฒนาและติดตั้งระบบแนะนำสินค้าพร้อมด้วยระบบขายสินค้าออนไลน์ ให้แก่ บริษัท วนาวัฒน์วัสดุ จำกัด โดยการดำเนินการประกอบด้วย 2 ขั้นตอนดังนี้ 1.กระบวนการออกแบบและพัฒนาระบบขายสินค้าออนไลน์ และ 2.กระบวนการออกแบบและพัฒนาระบบแนะนำ

3.1 กระบวนการออกแบบและพัฒนาระบบขายสินค้าออนไลน์

กระบวนการออกแบบและพัฒนาระบบขายสินค้าออนไลน์มีวัตถุประสงค์ขั้นต้น เพื่อออกแบบ พัฒนาและติดตั้งระบบขายสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัด ซึ่งมีฟังก์ชันการทำงานดังนี้

1. ระบบสมาชิก
2. ระบบสั่งซื้อสินค้าออนไลน์
3. ระบบจัดการข้อมูลสินค้า
4. ระบบจัดการข้อมูลลูกค้า
5. ระบบการรายงานสรุปยอดขาย
6. ระบบยกเลิกออเดอร์
7. ระบบคูปอง
8. ระบบเปรียบเทียบราคาสินค้า
9. ระบบขนส่ง
10. ระบบคืนสินค้า
11. ระบบชำระเงิน

3.1.1 ภาพรวมของระบบ

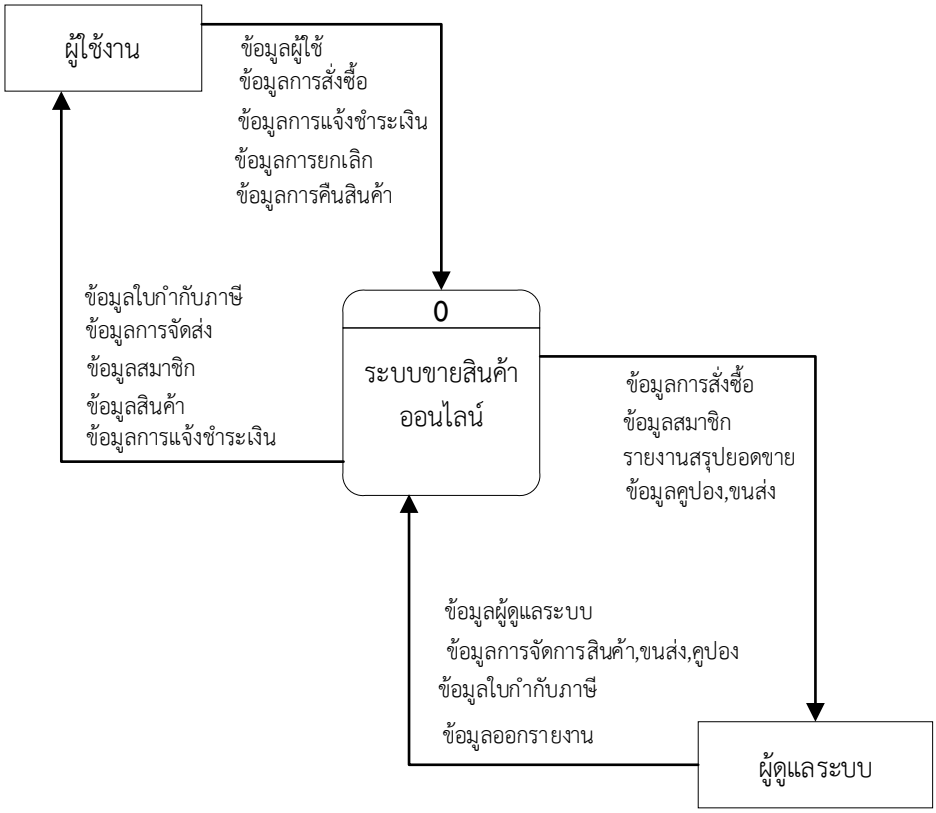


รูปที่ 3-1 กระบวนการทำงานของระบบขายสินค้าออนไลน์

จากรูปที่ 3-1 กระบวนการทำงานของระบบขายสินค้าออนไลน์เมื่อผู้ใช้งานเข้ามายังระบบขายสินค้าออนไลน์เพื่อทำการสั่งซื้อสินค้า จากนั้นก่อนจะดำเนินการชำระเงินผู้ใช้งานจะต้องทำการลงทะเบียนเพื่อเข้าสู่ระบบขายสินค้าออนไลน์ เมื่อทำการชำระเงินแล้วระบบจะเก็บข้อมูลเข้าสู่ระบบเพื่อรอให้ทางร้านค้ายืนยันการสั่งซื้อและดำเนินการจัดส่งสินค้า

3.1.2 การวิเคราะห์ข้อมูลด้านการไหลของข้อมูล

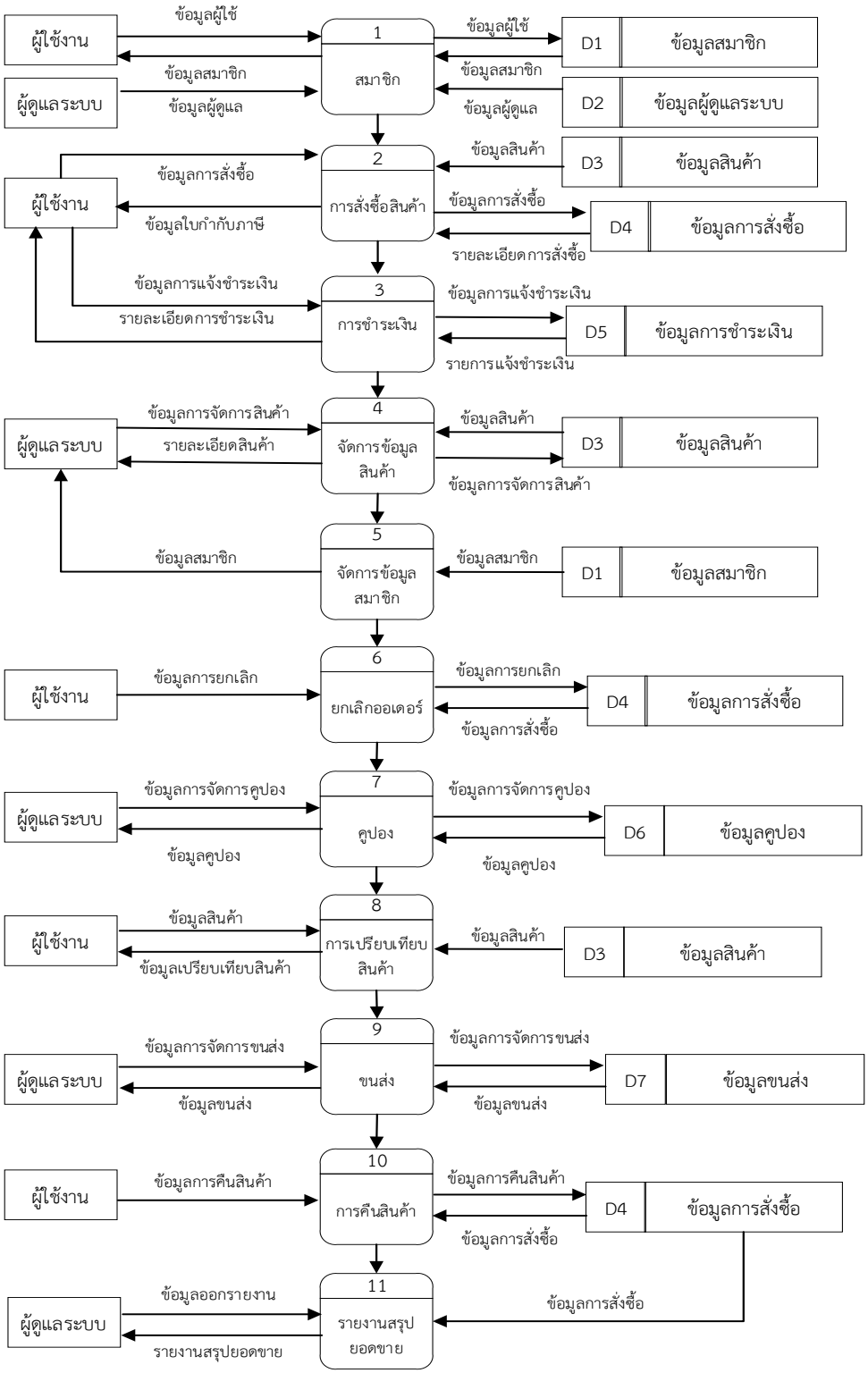
1) การออกแบบขั้นตอนและการทำงานของระบบ Context Diagram ของระบบขายสินค้าออนไลน์



รูปที่ 3-2 แผนภาพบริบท (Context Diagram)

จากรูปที่ 3-2 เป็นการไหลของกระแสข้อมูลในระดับ 0 ของระบบขายสินค้าออนไลน์ ซึ่งมีผู้เกี่ยวข้องกับระบบได้แก่ ผู้ใช้งาน และผู้ดูแลระบบ โดยผู้ใช้งานจะได้รับข้อมูลสินค้า ข้อมูลสมาชิก ข้อมูลการจัดส่งสินค้า และข้อมูลใบกำกับภาษีจากระบบ และผู้ใช้จะต้องส่งข้อมูลผู้ใช้ ข้อมูลการสั่งซื้อ และข้อมูลการชำระเงินเพื่อให้ระบบเก็บข้อมูล ในส่วนของผู้ดูแลระบบจะได้ข้อมูลผู้ใช้งาน ข้อมูลการสั่งซื้อ และข้อมูลรายงานการสั่งซื้อจากระบบ และผู้ดูแลระบบจะต้องส่งข้อมูลผู้ดูแลระบบ ข้อมูลการจัดส่งสินค้า และข้อมูลใบกำกับภาษีมายังระบบเพื่อเก็บข้อมูล

2) Data Flow Diagram Level 1

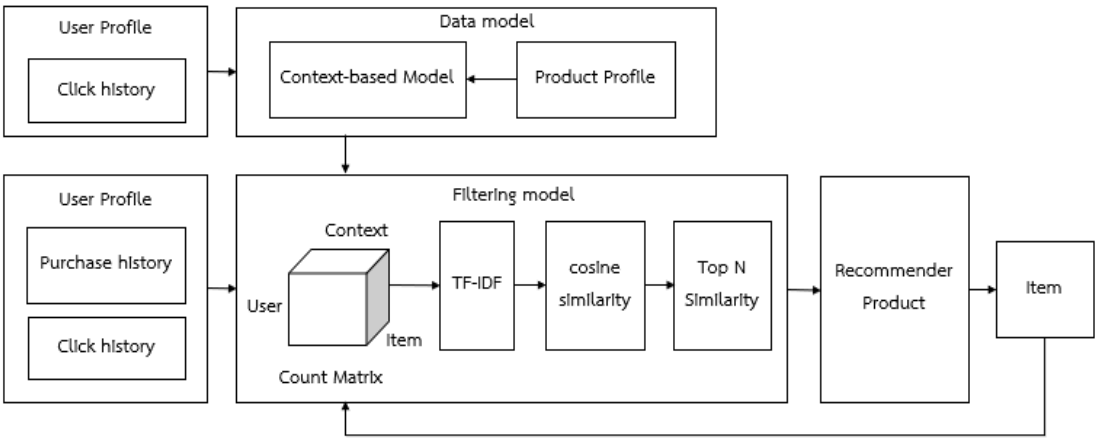


รูปที่ 3-3 Data Flow Diagram Level 1 ระบบขายออนไลน์

จากภาพที่ 3-3 เป็น Data Flow Diagram Level 1 แสดงขั้นตอนการทำงานของระบบขายออนไลน์ ประกอบด้วย 11 ระบบ ได้แก่ ระบบสมาชิก ระบบสั่งซื้อสินค้าออนไลน์ ระบบชำระเงิน ระบบจัดการข้อมูลสินค้า ระบบจัดการข้อมูลลูกค้า ระบบยกเลิกออเดอร์ ระบบคูปอง ระบบเปรียบเทียบราคาสินค้า ระบบขนส่ง ระบบคืนสินค้า และระบบการรายงานสรุปยอดขาย

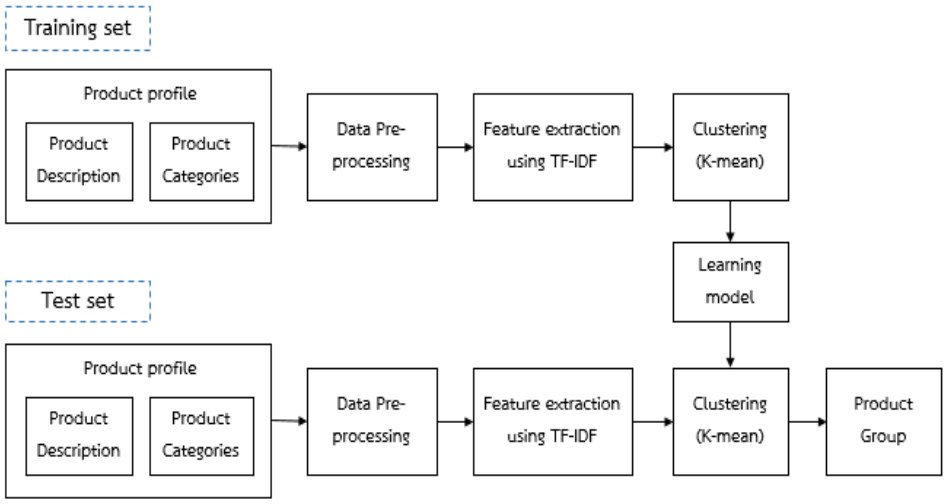
3.2 กระบวนการออกแบบและพัฒนาระบบแนะนำ

ขั้นตอนการออกแบบและพัฒนาระบบแนะนำประกอบด้วย 2 ขั้นตอน ได้แก่ขั้นตอนการสร้างแบบโมเดลของข้อมูล (Data model) ซึ่งจะใช้ข้อมูลโปรไฟล์ผลิตภัณฑ์ (Product Profile) ในการจัดกลุ่มสินค้าตามบริบท และข้อมูลพฤติกรรมการคลิกของผู้ใช้ซึ่งนำมาใช้เลือกกลุ่มสินค้าที่จัดได้ หลังจากนั้นขั้นตอนต่อมาคือการสร้างโมเดลข้อมูล (Filtering model) ซึ่งจะใช้ข้อมูลโปรไฟล์ผู้ใช้ (User Profile) ข้อมูลประวัติการแนะนำ (Item Profile) และข้อมูลการเลือกกลุ่มสินค้า (Context Profile) ที่ได้จากโมเดลของข้อมูล เพื่อนำมาสร้างเมทริกซ์ในขั้นตอน Count Matrix จากนั้นนำค่าที่ได้ไปหาค่าความถี่ของคำโดยใช้อัลกอริทึม term frequency-inverse document frequency (TF-IDF) และใช้วิธี Cosine similarity เพื่อหาค่าความคล้ายคลึงจากค่า TF-IDF และขั้นตอนสุดท้ายคือจัดอันดับรายการที่มีค่าความคล้ายคลึงสูงสุดเพื่อใช้ในการแนะนำ ขั้นตอนการดำเนินการแสดงดังรูปที่ 3-4



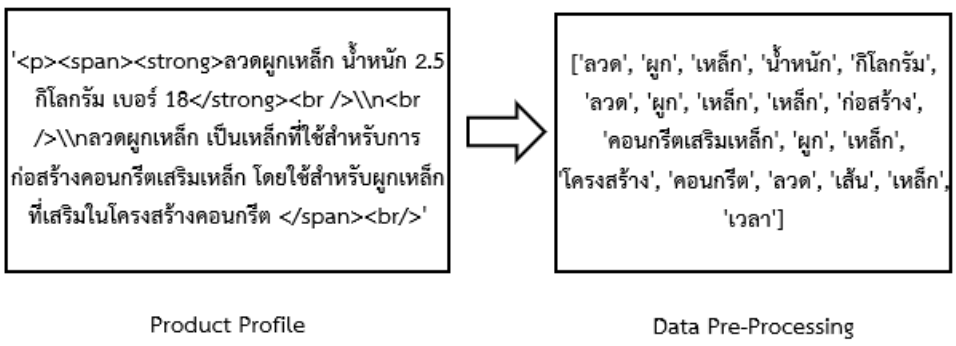
รูปที่ 3-4 แนวคิดของระบบผู้แนะนำตามบริบท

3.2.1 การสร้างโมเดลของข้อมูล



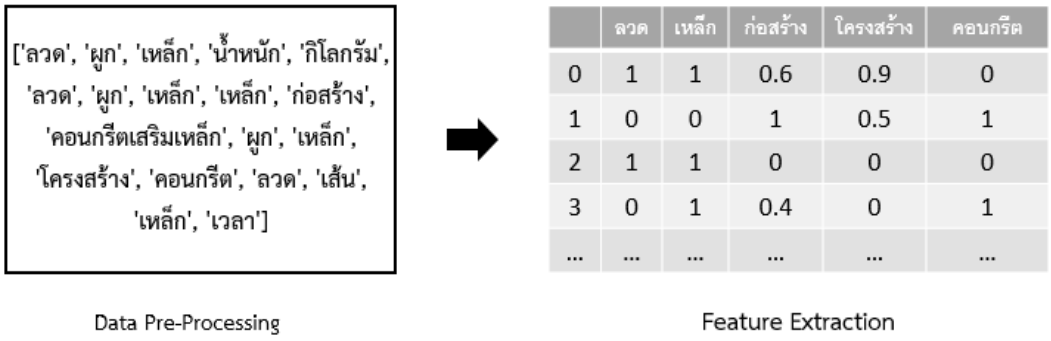
รูปที่ 3-5 แนวคิดการสร้างโมเดลของข้อมูล

จากภาพประกอบที่ 3-5 ขั้นตอนการสร้างโมเดลของข้อมูลใช้ข้อมูลโปรไฟล์ผลิตภัณฑ์ (Product Profile) ได้แก่ คำอธิบายและประเภทของผลิตภัณฑ์ที่สามารถนำมาใช้เป็นข้อมูลบริบทของผลิตภัณฑ์สำหรับใช้ในการจัดกลุ่ม ขั้นตอนเริ่มต้นจากการประมวลผลข้อมูลล่วงหน้า (Data Pre-processing) โดยใช้เทคโนโลยีประมวลผลภาษาธรรมชาติ NLP หรือ Natural Language Processing ในงานวิจัยได้ใช้ PyThaiNLP คือไลบรารี Python สำหรับงานด้านการประมวลผลข้อมูลภาษาไทย โดยกระบวนการแสดงดังรูปที่ 3-6



รูปที่ 3-6 กระบวนการประมวลผลข้อมูลล่วงหน้า

จากนั้นนำข้อมูลจากกระบวนการประมวลผลข้อมูลล่วงหน้า (Data Pre-processing) มาสร้างคุณลักษณะ (Feature Extraction) และเลือกคุณลักษณะโดยใช้วิธีการ Term Frequency-Inverse Document Frequency (TFIDF) ซึ่งจะหาค่าโดยใช้สมการที่ (3-1) สำหรับการดึงข้อมูลคุณลักษณะกระบวนการแสดงดังรูปที่ 3-7 จากนั้นข้อมูลคุณลักษณะ (Feature Extraction) จะถูกใช้สำหรับการจัดกลุ่มสินค้าโดยใช้วิธีการจัดกลุ่มแบบ K-Means



	ลวด	เหล็ก	ก่อสร้าง	โครงสร้าง	คอนกรีต
0	1	1	0.6	0.9	0
1	0	0	1	0.5	1
2	1	1	0	0	0
3	0	1	0.4	0	1
...	...	...	...	...	...

รูปที่ 3-7 กระบวนการสร้างคุณลักษณะ (Feature Extraction)

$$\boldsymbol{TFIDF_{d,t} = FREQ_{t,d}\left(1 + \log\left(\frac{\mathbf{N}}{DFREQ_t}\right)\right)} \quad (3\text{-}1)$$

โดยที่ $\boldsymbol{FREQ_{t,d}}$ = จำนวนคำ t ในผลคูณ d

$\mathbf{N}$ = จำนวนสินค้าทั้งหมด

$\boldsymbol{DFREQ}$ = จำนวนสินค้าที่มีคำว่า t ปรากฏขึ้น

3.2.2 การสร้างโมเดลการกรองข้อมูล

1) Count Matrix

กระบวนการ Count Matrix เป็นขั้นตอนแรกในการนับจำนวนคำหรือ Keyword ของข้อมูล 3 รายการคือ ข้อมูลโปรไฟล์ผู้ใช้ (User Profile) ได้แก่ ข้อมูลประวัติการซื้อและประวัติการคลิก ข้อมูลประวัติการแนะนำสินค้า (Item Profile) และข้อมูลการเลือกกลุ่ม (Context Profile) จากแบบจำลองข้อมูล 3 มิติ ซึ่งตัวอย่างการนับจำนวนคำหรือ Keyword แสดงดังตารางที่ 3-1, 3-2 และ 3-3

ตางรางที่ 3-1 ตัวอย่างของ COUNT MATRIX ของข้อมูลโปรไฟล์ผู้ใช้

User	โซฟา	ฉนวน	เลื่อย	เก้าอี้	ที่นั่ง
Product A	2	0	0	3	2
Product B	1	0	0	2	2
Product C	0	0	0	0	0
Product D	0	0	0	0	0

ตารางที่ 3-2 ตัวอย่างของ COUNT MATRIX ของข้อมูลโปรไฟล์รายการ

Item	โซฟา	ฉนวน	เลื่อย	เก้าอี้	ที่นั่ง
Product A	1	0	0	0	2
Product B	0	0	0	0	0
Product C	1	0	0	0	2
Product D	0	0	0	0	0

ตารางที่ 3-3 ตัวอย่างของ COUNT MATRIX ของข้อมูลโปรไฟล์บริบท

Context	โซฟา	ฉนวน	เลื่อย	เก้าอี้	ที่นั่ง
Product 1	3	0	0	3	2
Product 2	2	0	0	2	2
Product 3	2	0	0	0	2
Product 4	0	0	0	0	0

2) การหาค่าความถี่ของคำโดยใช้วิธี Term Frequency Inverse Document Frequency (TF-IDF)

การคำนวณหาค่าความถี่ของคำที่ปรากฏ (TF-IDF) จะใช้ข้อมูลจากกระบวนการ Count Matrix และใช้สมการที่ (3-1) ในการคำนวณ ซึ่งในตารางที่ 3-4, 3-5 และ 3-6 แสดงตัวอย่างของค่าความถี่ของคำโดยใช้วิธี TF-IDF

ตารางที่ 3-4 ตัวอย่างของ TF-IDF ของข้อมูลโปรไฟล์ผู้ใช้

User	โซฟา	ฉนวน	เลื่อย	เก้าอี้	ที่นั่ง
Product A	0.38	0	0	0.56	0.38
Product B	0.26	0	0	0.52	0.52
Product C	0	0	0	0	0
Product D	0	0	0	0	0

ตารางที่ 3-5 ตัวอย่างของ TF-IDF ของข้อมูลโปรไฟล์รายการ

Item	โซฟา	ฉนวน	เลื่อย	เก้าอี้	ที่นั่ง
Product A	0.43	0	0	0	0.87
Product B	0	0	0	0	0
Product C	0.43	0	0	0	0.87
Product D	0	0	0	0	0

ตารางที่ 3-6 ตัวอย่างของ TF-IDF ของข้อมูลโปรไฟล์บริบท

Context	โซฟา	ฉนวน	เลื่อย	เก้าอี้	ที่นั่ง
Product 1	0.42	0	0	0.49	0.28
Product 2	0.27	0	0	0.43	0.37
Product 3	0.56	0	0	0	0.56
Product 4	0	0	0	0	0

3) การหาค่าความคล้ายคลึงโดยใช้วิธี Cosine similarity

การหาค่าความคล้ายคลึงโดยใช้วิธี Cosine similarity เป็นขั้นตอนเพื่อกำหนดความคล้ายคลึงกันระหว่างข้อมูลสินค้าจากโปรไฟล์บริบท (Context Profile) และข้อมูลสินค้าจากโปรไฟล์ผู้ใช้ (User Profile) และโปรไฟล์รายการ (Item Profile) การคำนวณหาค่าความคล้ายคลึงจะใช้สมการที่ (3-2) ในการคำนวณ

$$\boldsymbol{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \qquad (3\text{-}2)$$

การคำนวณความคล้ายคลึงจะดำเนินการ 2 ครั้งโดยใช้ค่า TF-IDF ครั้งแรก A เป็นเวกเตอร์ของ TF-IDF ของโปรไฟล์ผู้ใช้ และ B เป็นเวกเตอร์ของ TF-IDF ของโปรไฟล์บริบท จากนั้น ในครั้งที่สอง A เป็นเวกเตอร์ของ TF-IDF ของโปรไฟล์รายการ และ B เป็นเวกเตอร์ของ TF-IDF ของโปรไฟล์บริบท ซึ่งในตารางที่ 3-7 แสดงตัวอย่างของข้อมูลค่าความคล้ายคลึงที่ได้จากการคำนวณ

ตารางที่ 3-7 ตัวอย่างของข้อมูลค่าความคล้ายคลึง

	User				Item			
Context	Product A	Product B	Product C	Product D	Product A	Product B	Product C	Product D
Product 1	0.99	0.93	0	0	0.62	0	0.62	0
Product 2	0.99	099	0	0	0.72	0	0.72	0
Product 3	0.69	0.71	0	0	0.95	0	0.95	0
Product 4	0	0	0	0	0	0	0	0

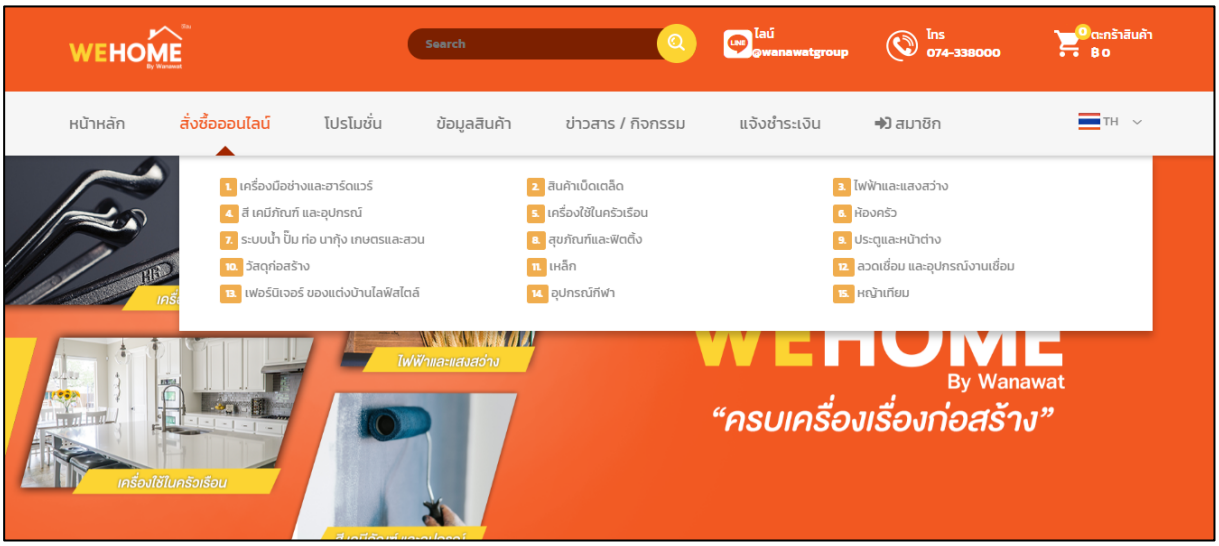
บทที่ 4
ผลการวิจัย

ผลการศึกษางานวิจัยในครั้งนี้ได้นำเสนอในรูปแบบรูปภาพและตารางประกอบคำอธิบาย โดยแบ่งผลการศึกษางานวิจัยดังนี้

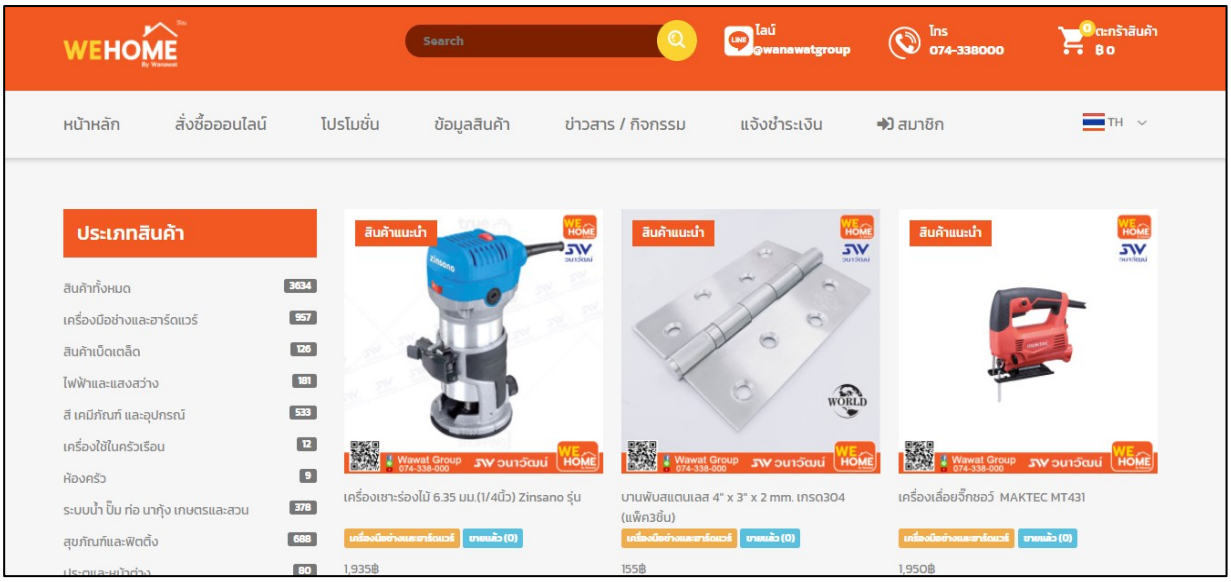
4.1 ผลการออกแบบ พัฒนาและติดตั้งระบบขายสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ

4.1.1 ระบบขายสินค้าออนไลน์ (เดิม)

รูปที่ 4-1 หน้าหลักของระบบ

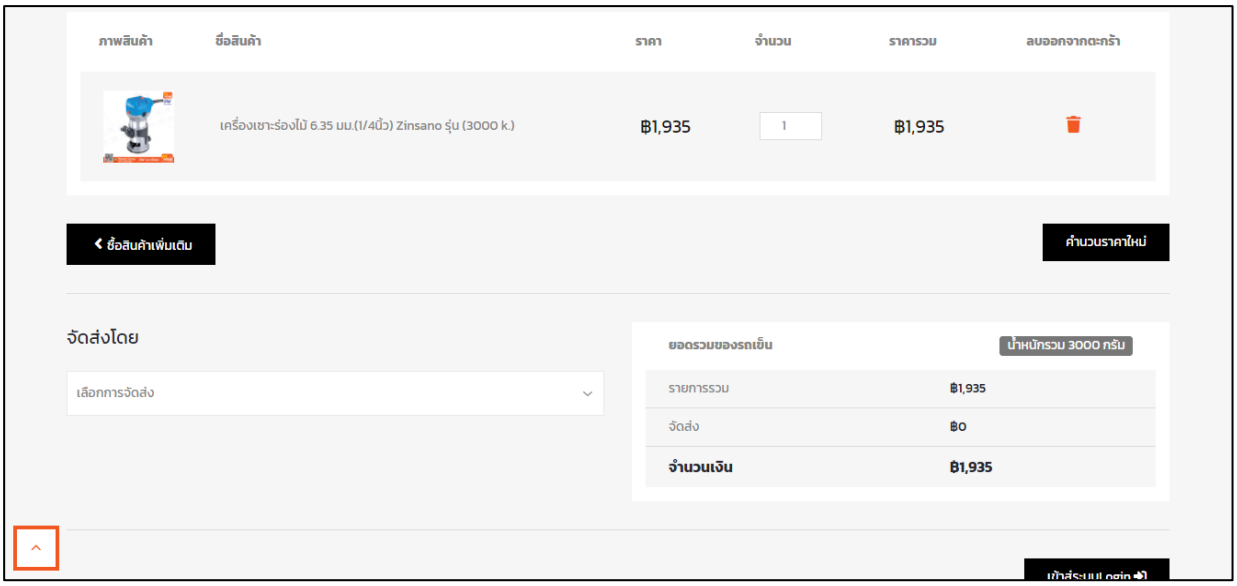


รูปที่ 4-2 เมนูหมวดหมู่สินค้าของระบบ

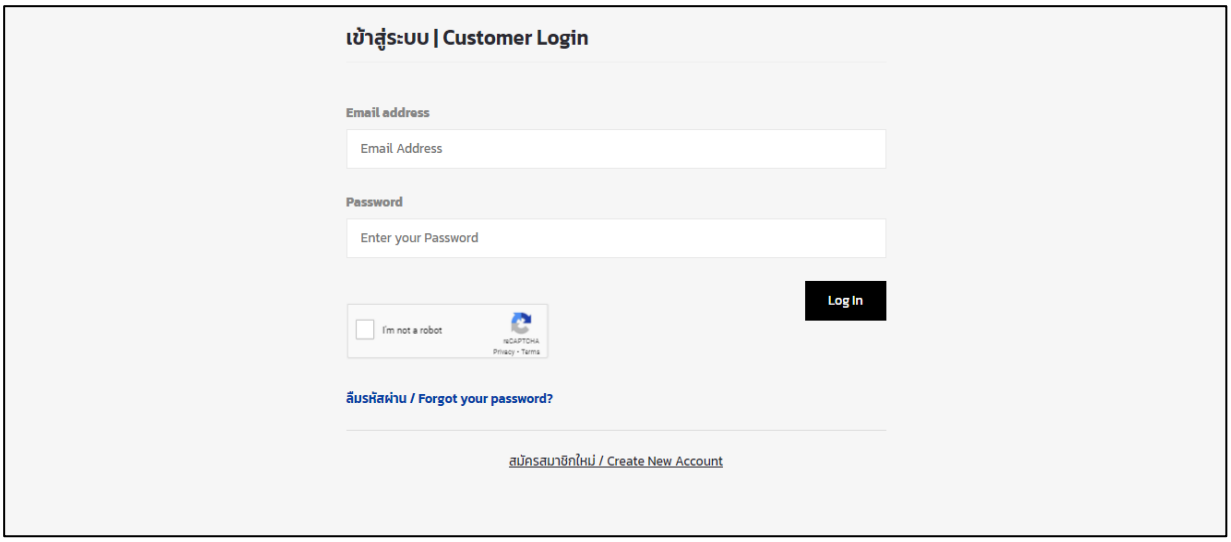


รูปที่ 4-3 หน้ารายการสินค้าของระบบ

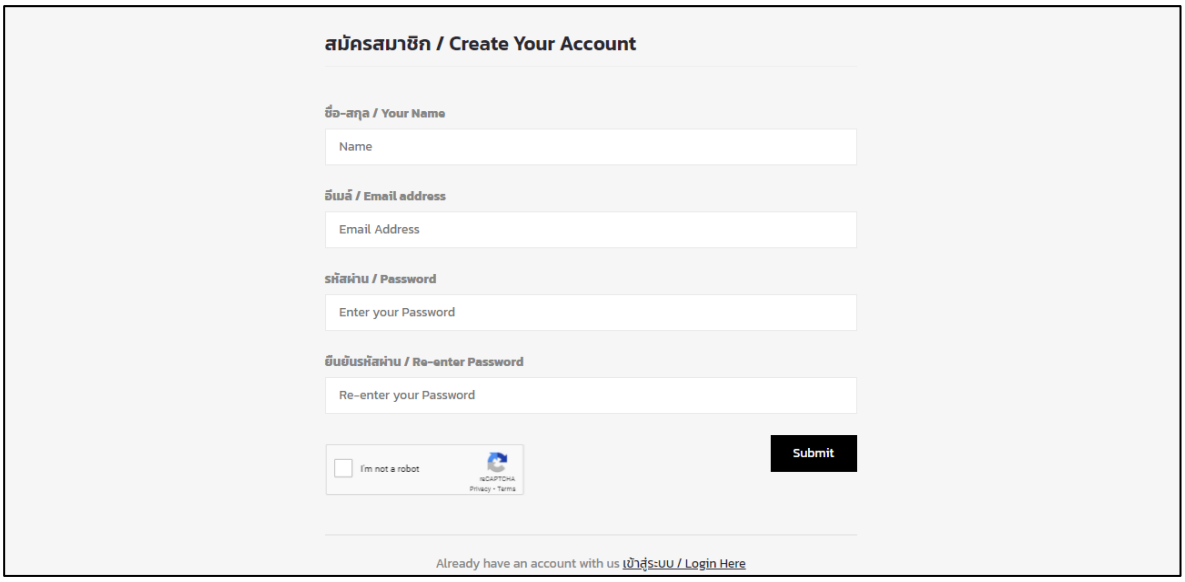
รูปที่ 4-4 หน้ารายละเอียดสินค้าของระบบ



รูปที่ 4-5 หน้าตะกร้าสินค้าของระบบ

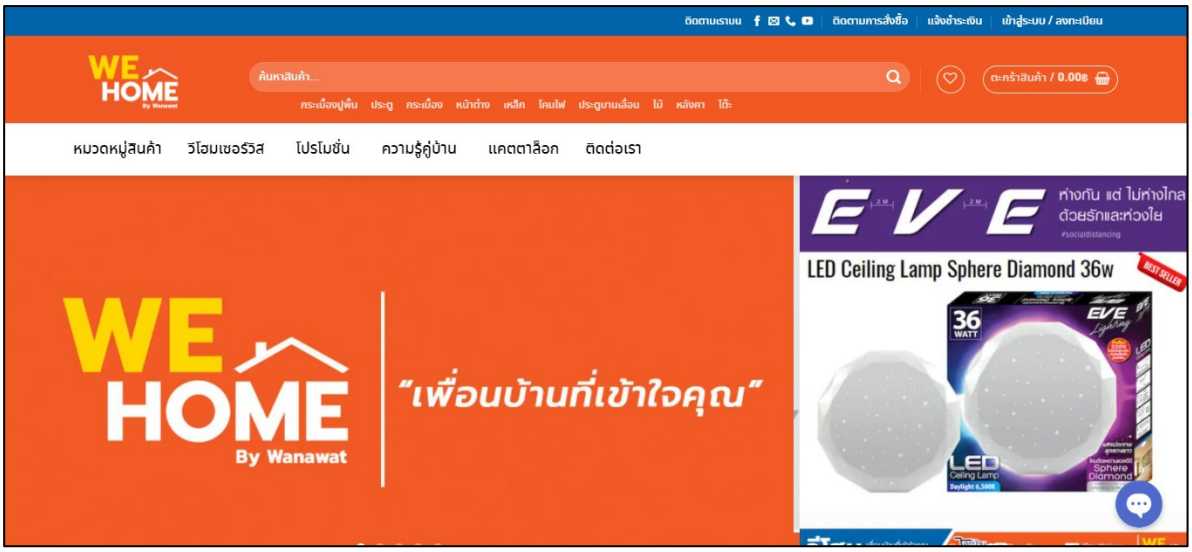


รูปที่ 4-6 หน้าเข้าสู่ระบบ

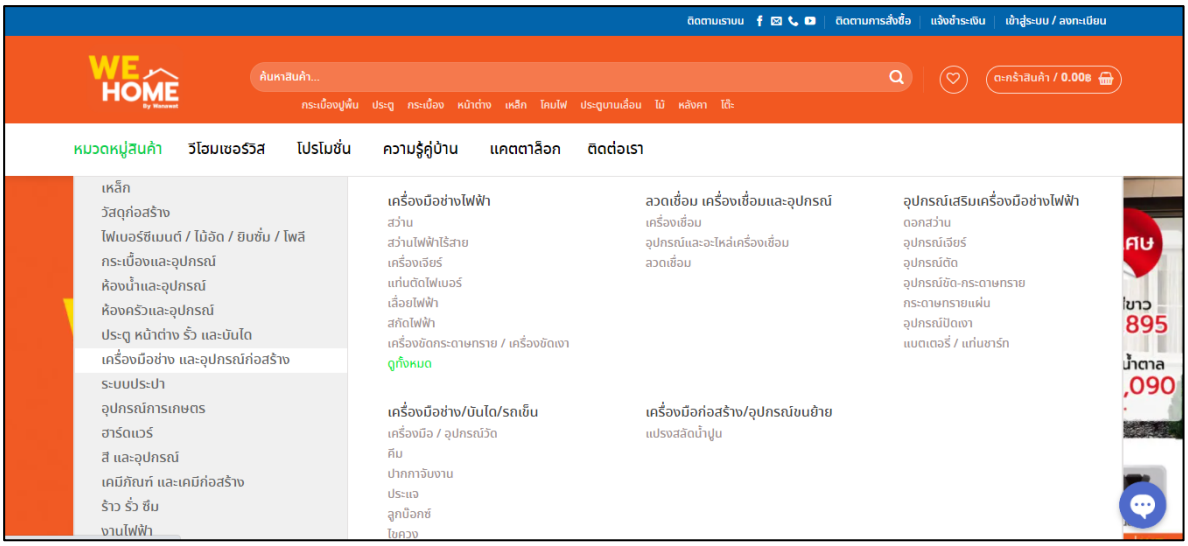


รูปที่ 4-7 หน้าสมัครสมาชิกของระบบ

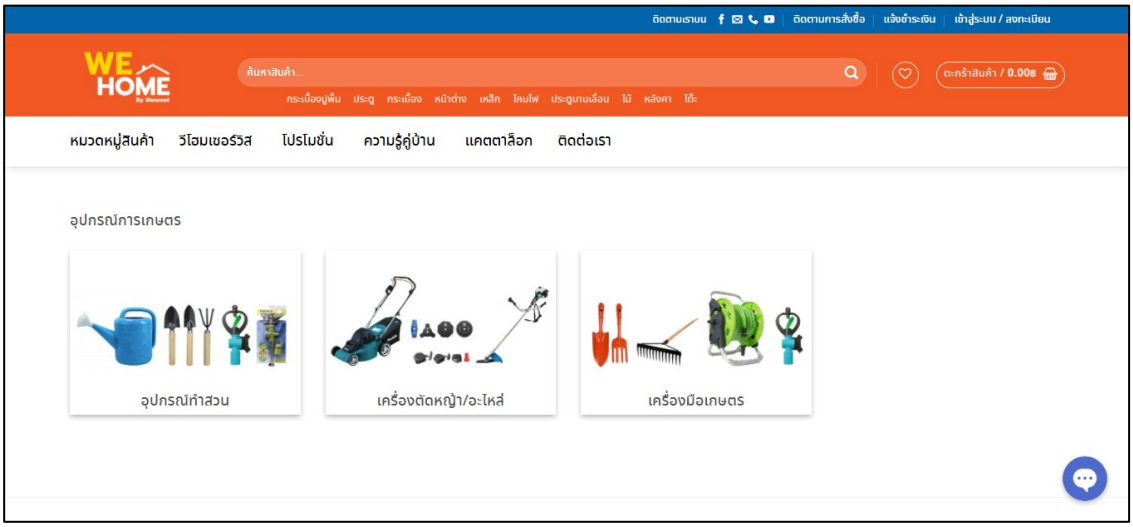
4.1.2 ระบบขายสินค้าออนไลน์ที่ออกแบบพัฒนาขึ้นใหม่



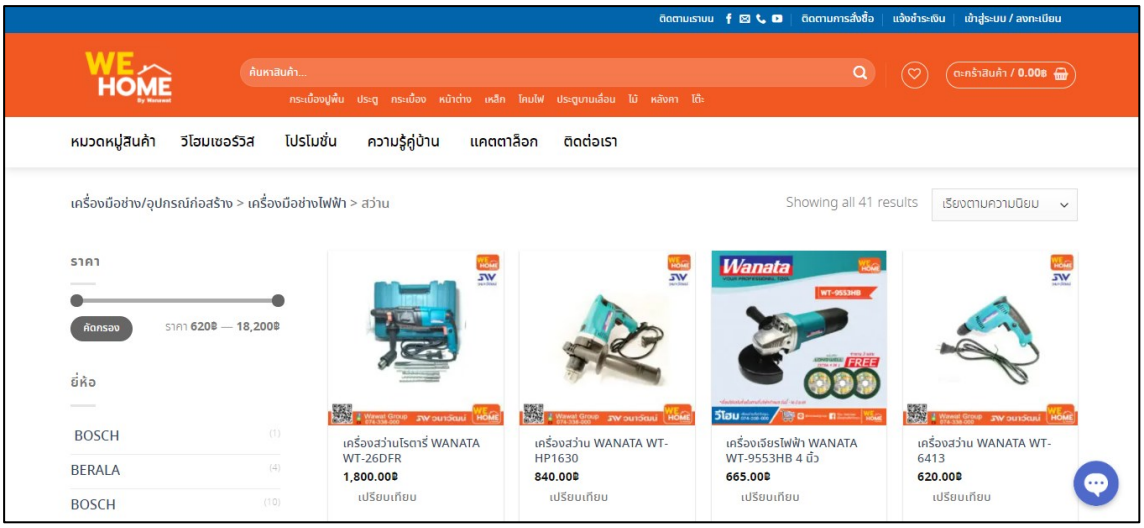
รูปที่ 4-8 หน้าหลักของระบบ



รูปที่ 4-9 เมนูหมวดหมู่สินค้าของระบบ

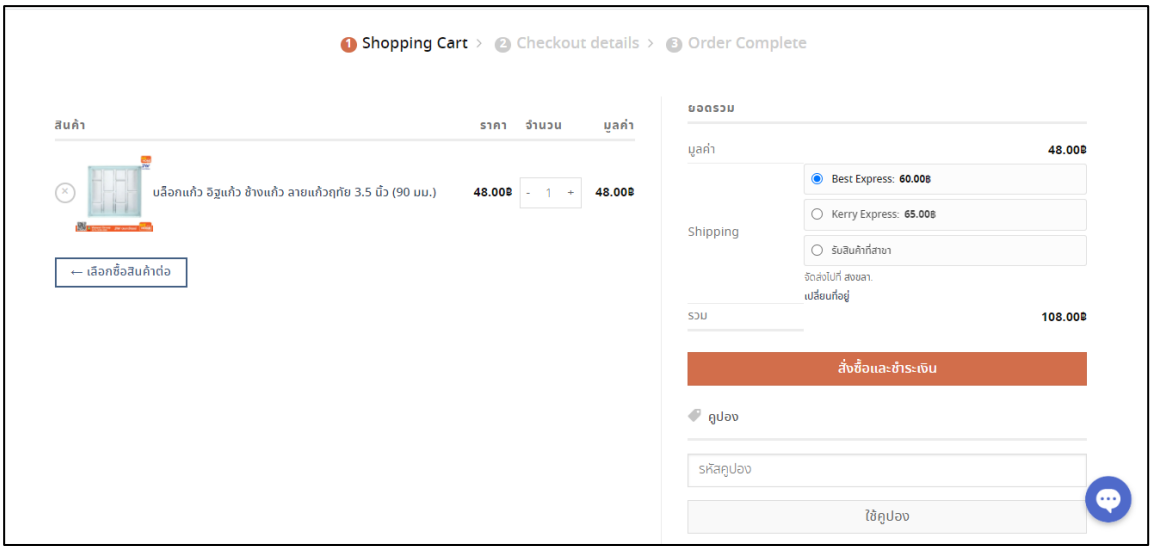


รูปที่ 4-10 เมนูหมวดหมู่สินค้าย่อยของระบบ

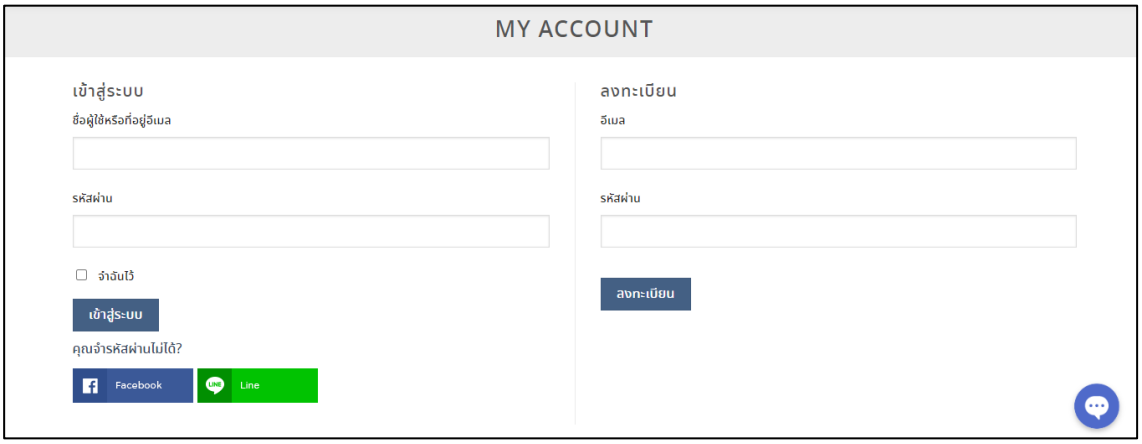


รูปที่ 4-11 หน้ารายการสินค้าของระบบ

รูปที่ 4-12 หน้ารายละเอียดสินค้าของระบบ



รูปที่ 4-13 หน้าตะกร้าสินค้าของระบบ



รูปที่ 4-14 หน้าเข้าสู่ระบบและลงทะเบียนของระบบ

4.2 ผลการวิจัยและพัฒนาระบบแนะนำสินค้าแบบการรับรู้บริบทสำหรับอีคอมเมิร์ซ

โครงการวิจัยระบบแนะนำสินค้าแบบการรับรู้บริบทสำหรับอีคอมเมิร์ซบริษัท วนาวัฒน์วัสดุ จำกัด ปีที่ 2 มีวัตถุประสงค์เพื่อวิจัยและพัฒนาระบบแนะนำสินค้า ซึ่งมีกระบวนการดำเนินการดังนี้

4.2.1 ผลการจัดเก็บข้อมูลพฤติกรรมของลูกค้า

การออกแบบและพัฒนาระบบจัดเก็บข้อมูลพฤติกรรมของลูกค้า ผู้พัฒนาดำเนินการเขียนโค้ดเพื่อทำการดึงค่าข้อมูลพฤติกรรมของลูกค้าจาก Plugin ที่ทำการติดตั้งบนระบบ Wordpress ของบริษัท ได้แก่ Plugin WooCommerce และ Recently Viewed product YITH

จากนั้นนำข้อมูลที่ได้เก็บยัง Database เพื่อที่จะนำข้อมูลไปใช้กับระบบแนะนำสินค้าที่พัฒนาขึ้น แสดงตัวอย่างข้อมูลในตารางที่ 4-1 และ 4-2

ตารางที่ 4-1 ข้อมูลประวัติการซื้อ

order_item_id	order_id	customer_id	product_id	date_created	product_qty
492	35983	159	35239	2022-08-03 19:24:50	1
493	35983	159	16828	2022-08-03 19:24:50	1
494	35983	159	35248	2022-08-03 19:24:50	1
496	35984	159	23448	2022-08-03 19:28:35	1
497	35984	159	23453	2022-08-03 19:28:35	1
499	36270	163	24467	2022-08-18 12:05:05	1
502	38369	159	18413	2022-12-09 16:20:55	1
504	38657	206	35257	2023-01-03 22:02:18	1
506	39099	220	18123	2023-01-29 23:38:18	1
509	39234	224	32695	2023-02-05 17:53:44	1
509	39234	224	32702	2023-02-05 17:53:44	1

ตารางที่ 4-2 ข้อมูลประวัติการคลิกสินค้า

cilck_item_id	customer_id	product_id	time_stamp
194	159	16922|16921|16923|23840|30778|30777|18129|18021|18017|18016	2023-02-23 12:50:23
195	125	18112|18113|18114|18119|18110|18130|30778|30777|18110|18110	2023-02-15 15:23:07
196	143	|20679|30034|20681|20585|32989|29119|33001|32999|18021|18017	2022-12-21 15:55:54
199	163	24467|24467|24467|24467|24467|24467|24467|24467|24467|24467	2022-08-18 12:11:45
200	170	16809|16809|16809|16809|16809|16809|16809|16809|16809|16809	2022-09-11 9:57:13
201	174	24537|24537|24537|24537|24537	2022-09-30 12:10:35

202	179	|23572|23457|23572|23457|23572|23457|23572|23457|23572|23457|23572	2022-10-05 11:53:11
203	182	17069|17069|17069|17069|17069|17069|17069|17069|17069|17069	2022-10-15 14:08:39
204	183	23988|23988|23988|23988|23988|23988|23988|23988|23988|23988	2022-10-17 21:39:36
205	224	32702|32695	2023-02-06 12:03:25

4.2.2 โมเดลข้อมูล (data model)

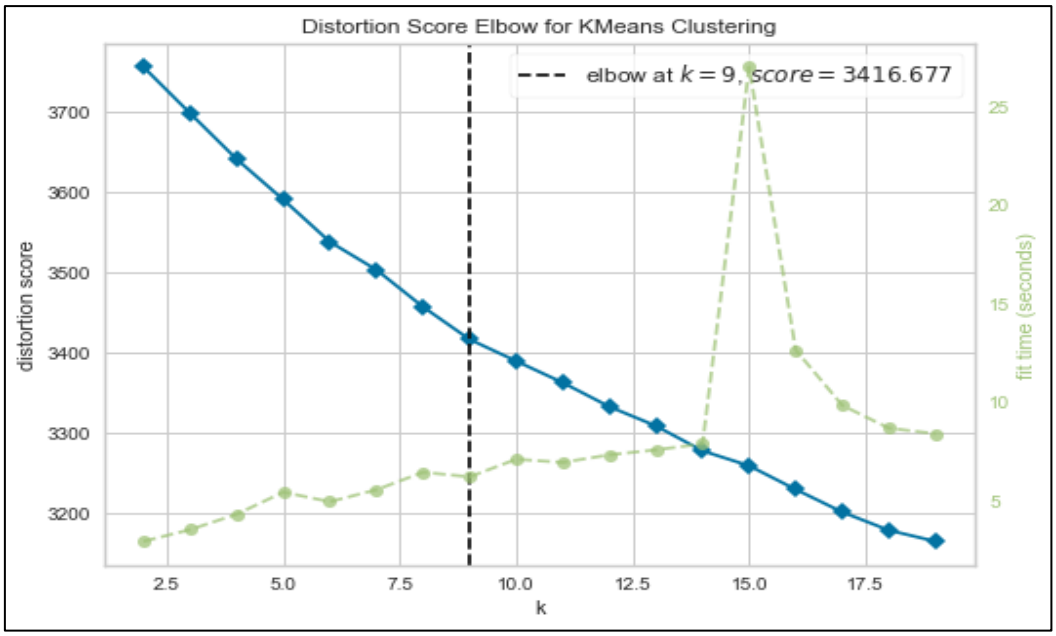
การวิจัยและพัฒนาโมเดลข้อมูล Data model สำหรับระบบแนะนำสินค้าได้ใช้ข้อมูลสินค้าจำนวน 4000 รายการ ประกอบด้วย ชื่อสินค้า รายละเอียดสินค้า และประเภทสินค้า ดังแสดงตัวอย่างในตารางที่ 4-3

ตารางที่ 4-3 ตัวอย่างข้อมูลสินค้าสำหรับวิจัยและพัฒนาโมเดลข้อมูล Data model

SKU	Name	Description	Categories
GSEL060002001	ลวดผูกเหล็ก ลวดดำ น้ำหนัก 2.3 กิโลกรัม เบอร์ 18	ลวดผูกเหล็ก ลวดดำ น้ำหนัก 2.3 กิโลกรัม เบอร์ 18 \nลวดผูกเหล็ก เป็นเหล็กที่ใช้สำหรับการก่อสร้างคอนกรีตเสริมเหล็ก โดยใช้สำหรับผูกเหล็กที่เสริมในโครงสร้างคอนกรีตเข้าด้วยกัน \nเป็นลวดเส้นเล็กๆที่ช่วยยึดเหล็กให้อยู่ทรง เมื่อเวลาเทคอนกรีต ได้น้ำหนัก และได้แรงตามที่ต้องการ หรือที่คำนวณไว้ \nลวดผูกเหล็กใช้ผูกเหล็กปลอกในเสา หรือในเหล็กเสริมพื้นคอนกรีต \nเหล็กชนิดนี้ใช้กันมากในการก่อสร้างคอนกรีตเสริมเหล็ก \nน้ำหนักลวดมีค่าบวกลบประมาณ 0.1-0.2	เหล็ก > ลวด > ลวดผูกเหล็ก
GRFL120102002	ฉนวนกันความร้อน Stay Cool - Premium ขนาด 3 นิ้ว ตราช้าง SCG	ฉนวนกันความร้อน Stay Cool - Premium ขนาด 3 นิ้ว ตราช้าง \n <span	วัสดุก่อสร้าง > ฉนวน / อุปกรณ์ติดตั้งฝ้าและผนัง > ฉนวนกันความร้อนบนฝ้า

		style="font-size: 16px;"> ฉนวนกันความร้อน เอสซีจี รุ่น STAY COOL ผลิตจากใยแก้ว ออกแบบมาสำหรับติดตั้งบริเวณเหนือฝ้าเพดาน ทั้งแบบฉาบเรียบและแบบ ที-บาร์ เพื่อป้องกัน ความร้อนที่มาทางโถงหลังคา ผ่านฝ้าเพดานลงมาสู่ ภายในบ้าน \n-ค่าต้านทานความร้อนรวมของระบบหลังคา รุ่น 6 นิ้ว = Rt 38.0 , รุ่น 3 นิ้ว = Rt 27.0 \n-หุ้มรอบด้านด้วยแผ่นอลูมิเนียมฟอยล์เสริมแรง \n• เนื้อฉนวนเขียวหนาสม่ำเสมอทั่วทั้งแผ่น \n• เนื้อฉนวนหุ้มรอบด้านด้วยแผ่นอลูมิเนียม \n-เนื้อฉนวนเขียว ได้รับการรับรองจากกลุ่ม ผู้เชี่ยวชาญด้านสุขภาพและความปลอดภัยของวัสดุประเภทเส้นใย \n-ผลิตจากแก้วรีไซเคิล100% และผลิตด้วย กระบวนการที่ส่งผลกระทบต่อสิ่งแวดล้อมน้อยที่สุด \n• มีสาร HydroProtecTM ที่ช่วยลดการอุ้มน้ำได้ถึง 10 เท่า ไม่ดูดซับน้ำ อีกทั้งสามารถกันความชื้น ทำให้เนื้อฉนวนคงประสิทธิภาพได้อย่างยาวนาน \n• อายุการใช้งานยาวนานกว่า 10 ปี \n• ได้รับการทดสอบว่าเป็นวัสดุที่ไม่ลามไฟ ตามมาตรฐาน ASTM E84, BS476 </p>	

สำหรับการจัดกลุ่มสินค้าในการสร้าง Data model ได้ใช้วิธี Elbow method ในการหาค่า K ของจำนวนกลุ่มสินค้าซึ่งวิธีการ Elbow method เป็นวิธีที่ใช้สำหรับการวัดข้อผิดพลาด (Error measurement) โดยค่า Distortion Score คือ ผลรวมของระยะทางกำลังสองระหว่าง Object กับ Centroid และ Fit time คือ ระยะเวลาในการฝึกแบบจำลองการจัดกลุ่ม จะเห็นได้ว่าค่า K หลังจากคำนวณโดยใช้วิธีนี้มีค่าเท่ากับ 9 และ ค่า Score มีค่าเท่ากับ 3416.677 ซึ่งเป็นค่าที่ดีที่สุดจากการคำนวณ



รูปที่ 4-15 ผลการคำนวณหาค่า K สำหรับสร้าง Data model

ตารางที่ 4-4 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 1

SKU	Name	Cluster
GPAL023406002	ไขควงแฉก INDY หัวตอกได้ 4 นิ้ว #AB234	1
GPAL030512002	คีมย้ำตะปูรีเวท STANLEY #STHT69800-8	1
GPAL030529002	ใบเลื่อยวงเดือนตัดเหล็ก 4 นิ้ว PUMPKIN #38190	1
GPAL030612001	คีมย้ำรีเวท KENNEDY #KEN-569-3000K	1
GPAL030815003	คีมผูกลวด 8 นิ้ว TX-TP08 TEXAS BULL	1
GPAL031602001	คีมใช้ 3 อย่าง SOLO ด้ามแดง 8 นิ้ว #333	1
GPAL031604001	คีมปากแหลม SOLO ด้ามหนาแดง 6 นิ้ว #6026	1
GPAL037402001	คีมจับอ๊อก ALIBABA 600A	1
GPBL014901001	กบไสไม้ไฟฟ้า 3 นิ้ว ดีจริง Dongcheng #DMB03-82	1
GPBL034908001	กรรไกรตัดเหล็กไฟฟ้า 2.5-3.2mm. 620W ดีจริง Dongcheng #DJJ32	1

ตารางที่ 4-5 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 2

SKU	Name	Cluster
GWBL050811004	จระเข้ อะคลีลิก ซิลพลัส กันเชื้อรา 450กรัม สีน้ำตาล	2
GWBL070901004	เดพโก้ ทีทีบี พลัส กาวซีเมนต์ ปูกระเบื้องทั่วไป 20 กก.	2
GWBL070101002	ทีโอเอ กาวซีเมนต์ ซุปเปอร์ไทล์ 20 กก. แดง	2
GWBL041208002	เวเบอร์ ซีล เทป 10 ซม.x3 ม.	2
GWBL041208003	เวเบอร์ ซีล เทป 30 ซม.x3 ม.	2
GWBL051105001	ซิก้า maxtack กาวพลังตะปู 280 มล. SIKA	2
GWBL040801014	จระเข้ อีโค ซิลด์ซีเมนต์กันซึม 20กก.ส่วนผสมเดียว	2
GWAL071206002	กาวซีเมนต์ เวเบอร์ไทล์ โน สเตน 20 กก.	2
GWBL040812002	เวเบอร์ เทป ไฟเบอร์เมช 5 เทปตาข่ายไฟเบอร์ชนิดมีกาวในตัว	2
GWBL020804005	จระเข้ ฟลอร์ สมูท แพทช์ ซ่อมพื้น ปรับผิวคอนกรีตชนิดบาง 5 กก.	2

ตารางที่ 4-6 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 3

SKU	Name	Cluster
GCBL028005001	ชุดสายฉีดชำระ โครเมี่ยม A36 แกนทองเหลือง VEGARR	3
GCBL028005002	ชุดสายฉีดชำระ โครเมี่ยม A37 สายโครเมี่ยม VEGARR	3
GCBL028005005	ชุดสายฉีดชำระ โครเมี่ยม A44 สายโครเมี่ยม VEGARR	3
GCBL028005007	ชุดสายฉีดชำระ โครเมี่ยม 45 สายโครเมี่ยม VEGARR	3
GCBL028005009	ชุดสายฉีดชำระ โครเมี่ยม A46 แกนทองเหลือง สายสแตนเลส VEGARR	3
GCBL020105006	ชุดสายฉีดชำระ สแตนเลสชุบโครเมี่ยม A-4900-CH AMERICAN STANDARD	3
GCBL020105002	ชุดสายฉีดชำระ สแตนเลส A-4900-ST AMERICAN STANDARD	3
GCBL020105012	ชุดสายฉีดชำระ สแตนเลส A-5604-CH AMERICAN STANDARD	3
GCBL028005003	ชุดสายฉีดชำระ โครเมี่ยม B23S VEGARR	3
GCBL028005004	ชุดสายฉีดชำระ B23S-W ขาว สายแสตนเลส VEGARR	3

ตารางที่ 4-7 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 4

SKU	Name	Cluster
GCDL091702048	มือจับประตูทางเข้า มือจับหลิก นิกเกิ้ลด้าน 489.94.627 HAFELE	4
GCDL091702047	มือจับประตูทางเข้า มือจับหลิก นิกเกิ้ลด้าน 489.94.629 HAFELE	4
GCDL091702038	มือจับประตูทางเข้า หลอกทองแดง ยางน้ำมัน 489.94.927 HAFELE	4
GCDL091702158	มือจับฝังบานเลื่อน นิกเกิ้ลรมดำ 499.65.083 HAFELE	4
GCDL091702033	มือจับบานเลื่อน พร้อมระบบล็อค นิกเกิ้ลด้าน 499.65.092 HAFELE	4
GCDL091702014	มือจับบานเลื่อน พร้อมระบบล็อค ทองแดงรมดำ 499.65.100 HAFELE	4
GCDL091702161	มือจับฝังบานเลื่อน นิกเกิ้ลรมดำ 499.65.122 HAFELE	4
GCDL091702006	มือจับประตูสแตนเลส 499.68.011 HAFELE	4
GCDL091702005	มือจับประตูสแตนเลส 499.68.012 HAFELE	4
GCDL091702128	มือจับดึงสแตนเลส รูปตัวซี 499.68.030 HAFELE	4

ตารางที่ 4-8 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 5

SKU	Name	Cluster
GHCL131104031	ข้องอบาง 45 องศา PVC ท่อน้ำไทย 3 นิ้ว	5
GHCL131104032	ข้องอบาง 45 องศา PVC ท่อน้ำไทย 4 นิ้ว	5
GHCL131104033	ข้องอบาง 45 องศา PVC ท่อน้ำไทย 5 นิ้ว	5
GHCL131104034	ข้องอบาง 45 องศา PVC ท่อน้ำไทย 6 นิ้ว	5
GHCL131104035	ข้องอบาง 45 องศา PVC ท่อน้ำไทย 8 นิ้ว	5
GHCL131104036	ข้องอบาง 90 องศา PVC ท่อน้ำไทย 1 1/4 นิ้ว	5
GHCL131104037	ข้องอบาง 90 องศา PVC ท่อน้ำไทย 1 1/2 นิ้ว	5
GHCL131104038	ข้องอบาง 90 องศา PVC ท่อน้ำไทย 2 นิ้ว	5
GHCL131104039	ข้องอบาง 90 องศา PVC ท่อน้ำไทย 2 1/2 นิ้ว	5
GHCL131104040	ข้องอบาง 90 องศา PVC ท่อน้ำไทย 3 นิ้ว	5

ตารางที่ 4-9 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 6

SKU	Name	Cluster
GCEL032706006	ตู้แขวนผนังบานเดี่ยว KING Platinum เซอร์คอน ขาว CU-PK-1Z-WT	6
GCEL032708003	ตู้แขวนเสริม KING Platinum เซอร์คอน ขาว CU-PKHZ-WT	6
GCEL032709003	ตู้แขวนเข้ามุม KING Platinum เซอร์คอน ขาว CU-PK-CZ-WT	6
GCEL032705009	ตู้แขวนผนังบานคู่ KING Platinum เซอร์คอน โอ๊คดำ CU-POA-2Z	6
GCEL032708004	ตู้แขวนเสริม KING เซอร์คอน โอ๊คดำ CU-PKHZ-OA	6
GCEL032708005	ช่องเสริมแนวนอน KING Platinum ขาว CU-PKS-WT	6
GCEL032708006	ช่องเสริมแนวตั้ง KING Platinum ขาว CU-PKSH-WT	6
GCEL032708007	ช่องเสริมแนวนอน KING Platinum โอ๊คดำ CU-PKS-OA	6
GCEL032708008	ช่องเสริมแนวตั้ง KING Platinum โอ๊คดำ CU-PKSH-OA	6
GCEL032702001	บานซิงค์เดี่ยว KING ซูเปอร์ลายเกล็ด LII ขาว APSSU-1L2-WT	6

ตารางที่ 4-10 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 7

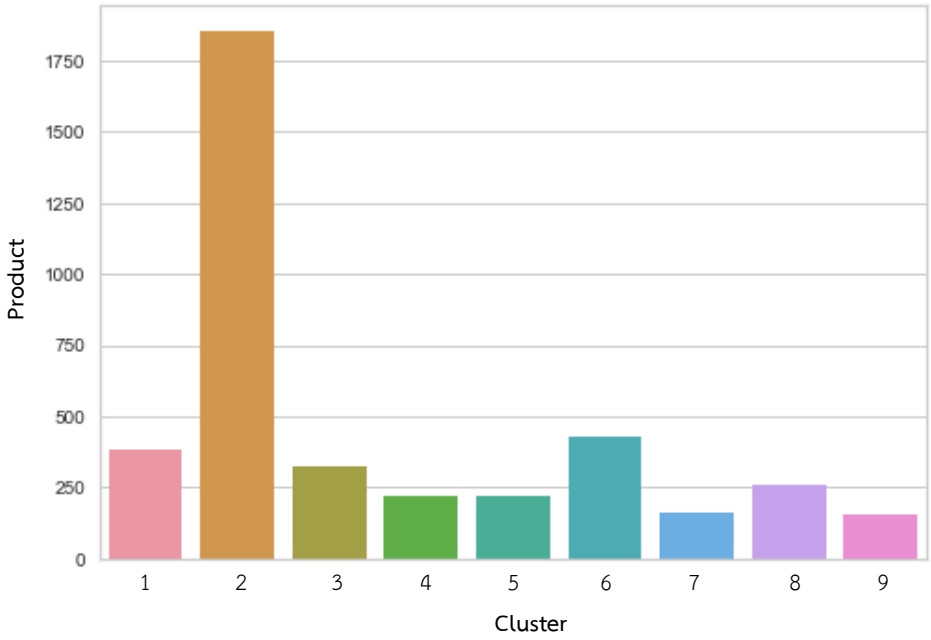
SKU	Name	Cluster
GPBL152903006	ดอกสว่าน คอนกรีต MEXCO 1/2 x 6	7
GPBL152903007	ดอกสว่าน คอนกรีต MEXCO 3/16 x 4	7
GPBL152903009	ดอกสว่าน คอนกรีต MEXCO 5/16 x 6	7
GPBL152903010	ดอกสว่าน คอนกรีต MEXCO 5/16 x 8	7
GPBL152903012	ดอกสว่าน คอนกรีต MEXCO 3/8 x 6	7
GPBL152903014	ดอกสว่าน คอนกรีต MEXCO 5/8 x 6	7
GPBL152903015	ดอกสว่าน คอนกรีต MEXCO 5/8 x 8	7
GPBL153002002	ดอกสว่าน FPB รุ่น JB 3/64	7
GPBL153002003	ดอกสว่าน FPB รุ่น JB 1/16	7
GPBL153002005	ดอกสว่าน FPB รุ่น JB 3/32	7

ตารางที่ 4-11 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 8

SKU	Name	Cluster
GWAL010110004	สีรองพื้นปูนเก่า TOA รุ่น 4SEASONS SUPER PRIMER 5 แกลลอน	8
GWAL010110006	สีทาฝ้า TOA รุ่น 4SEASONS โฟร์ซีซัน 5 แกลลอน #A8000 สีควันบุหรี่	8
GWAL010111031	สีรองพื้นปูนใหม่ TOA รุ่น Super Matex ซุปเปอร์เมเทค 5 แกลลอน	8
GWAL010163003	สีรองพื้นปูนอเนกประสงค์ TOA รุ่น HYDRO QUICK ไฮโดรควิก 15 ลิตร	8
GWAL010217004	สีสีรองพื้นปูนอเนกประสงค์ JOTUN รุ่น Ultra Primer อัลตร้าไพรเมอร์ 2.5 แกลลอน	8
GWAL010218003	สีรองพื้นปูนใหม่ JOTUN รุ่น Essence Easy Primer เอสเซ้น อีซี่ไพรเมอร์ 5 แกลลอน	8
GWAL010339003	สีทาฝ้าเพดาน Beger รุ่น Cool All Plus Interior คูลออลพลัส 1 แกลลอน #3599 สีควันบุหรี่	8
GWAL010339004	สีทาฝ้าเพดาน Beger รุ่น Cool All Plus Interior คูลออลพลัส 5 แกลลอน #3599 สีควันบุหรี่	8
GWAL010340002	สีรองพื้นปูนใหม่ Beger รุ่น Cool All Plus คูลออลพลัส 5 แกลลอน #6000 สีขาว	8
GWAL010343002	สีรองพื้นปูนใหม่ Beger รุ่น นาโนโปร 5 แกลลอน #E-9400	8

ตารางที่ 4-12 ตัวอย่างผลการจัดกลุ่มสินค้าด้วยโมเดลข้อมูลของกลุ่มที่ 9

SKU	Name	Cluster
GEBL011100026	หลอดไฟ จีนนี่ E27 11W วอร์มไวท์ PHILIPS	9
GEBL011100036	หลอดไฟ ทอร์นาโด E27 15W เดย์ไลท์ PHILIPS	9
GEBL011100067	หลอดไฟ TLD Super 18W/865 PHILIPS	9
GEBL011100112	หลอดไฟ BULB 6.5W วอร์มไวท์ PHILIPS	9
GEBL011100115	หลอดไฟนีออน ขั้วเขียว 36W เดย์ไลท์ PHILIPS	9
GEBL013401001	หลอดไฟ LED T8 ECO 22W เดย์ไลท์ 527249,569720 EVE	9
GEBL013401003	หลอดประหยัดไฟ มาตรฐาน 3U 18W E27 เดย์ไลท์ #500372 EVE	9
GEBL013401010	หลอดไฟ LED T8 TROY ขั้วอลู 9W เดย์ไลท์ 570795 EVE	9
GEBL013401019	หลอดไฟ LED อิโค่ ปิงปอง 3W E27 (แก้วขุ่น) วอร์มไวท์ 526785 EVE	9
GEBL013401021	หลอดไฟ LED TD 13W E27 วอร์มไวท์ 573345 EVE	9



รูปที่ 4-16 กราฟแสดงจำนวนสินค้าในแต่ละกลุ่มจากโมเดลข้อมูล (Data model)

จากรูปที่ 4-18 กราฟแสดงจำนวนสินค้าในแต่ละกลุ่มโดยใช้โมเดลข้อมูล (Data model) ในการจัดกลุ่มสินค้าจำนวน 4,000 รายการ แบ่งออกเป็น 9 กลุ่ม ได้แก่ กลุ่มที่ 1 จำนวน 381 รายการ, กลุ่มที่ 2 จำนวน 1853 รายการ, กลุ่มที่ 3 จำนวน 324 รายการ, กลุ่มที่ 4 จำนวน 222 รายการ, กลุ่มที่ 5 จำนวน 217 รายการ, กลุ่มที่ 6 จำนวน 425 รายการ, กลุ่มที่ 7 จำนวน 163 รายการ, กลุ่มที่ 8 จำนวน 259 รายการ, และกลุ่มที่ 9 จำนวน 156 รายการ.

4.2.3 โมเดลการกรอง (filtering model)

กระบวนการทำงานของโมเดลการกรองจะใช้ข้อมูลโปรไฟล์ผู้ใช้ (User Profile), ข้อมูลจากโมเดลข้อมูล (Context Profile) และข้อมูลประวัติการแนะนำ (Item Profile) เพื่อนำมาคำนวณหาค่า Count Matrix และ TF-IDF จากนั้นนำไปหาค่าความคล้ายคลึงของสินค้าในขั้นตอน Cosine similarity ก่อนที่จะส่งไปยัง Top N Similarity เพื่อเรียงรายการสินค้าที่จะแนะนำ

ตารางที่ 4-13 ข้อมูล User Profile ประวัติการซื้อตัวอย่างที่ 1

order_id	customer_id	product_id	date_created
39234	224	32695	2023-02-05 17:53:44
39234	224	32702	2023-02-05 17:53:44

ตารางที่ 4-14 ข้อมูล User Profile ประวัติการคลิกสินค้าตัวอย่างที่ 1

customer_id	product_id	time_stamp
224	32702|32695	2023-02-06 12:03:25

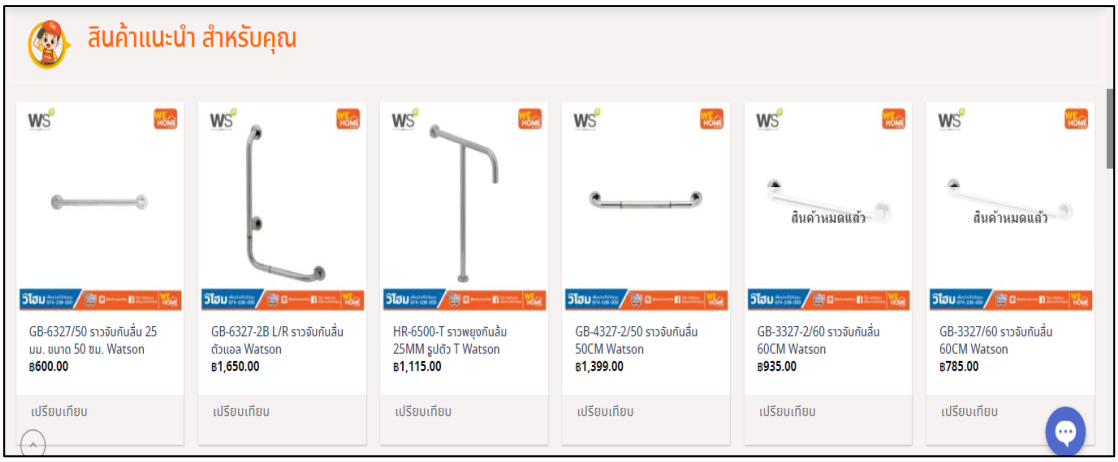
ตารางที่ 4-15 ผลการเลือกกลุ่มจากโมเดลข้อมูลของผู้ใช้ตัวอย่างที่ 1

product_id	SKU	product_id	Cluster
32702	GCBL065009031	GB-6327/50 ราวจับกันลื่น 25 มม. ขนาด 50 ซม. Watson	2
32695	GCBL065009020	GB-6327-2B L/R ราวจับกันลื่น ตัวแอล Watson	2
32700	GCBL065009028	HR-6500-T ราวพยุงกันล้ม 25MM รูปตัว T Watson	2
32691	GCBL065009011	GB-4327-2/50 ราวจับกันลื่น 50CM Watson	2
32690	GCBL065009010	GB-3327-2/60 ราวจับกันลื่น 60CM Watson	2

จากตารางที่ 4-15 ผลการเลือกกลุ่มจากโมเดลข้อมูลจะได้กลุ่มที่ 2 ข้อมูล Context Profile จะใช้ข้อมูลสินค้าจากกลุ่มที่ 2 มาใช้ในโมเดลการกรองข้อมูล (filtering model)

ตารางที่ 4-16 ผลการคำนวณหาค่าความคล้ายคลึงของผู้ใช้ตัวอย่างที่ 1 เรียงจากมากไปน้อย จำนวน 10 รายการ

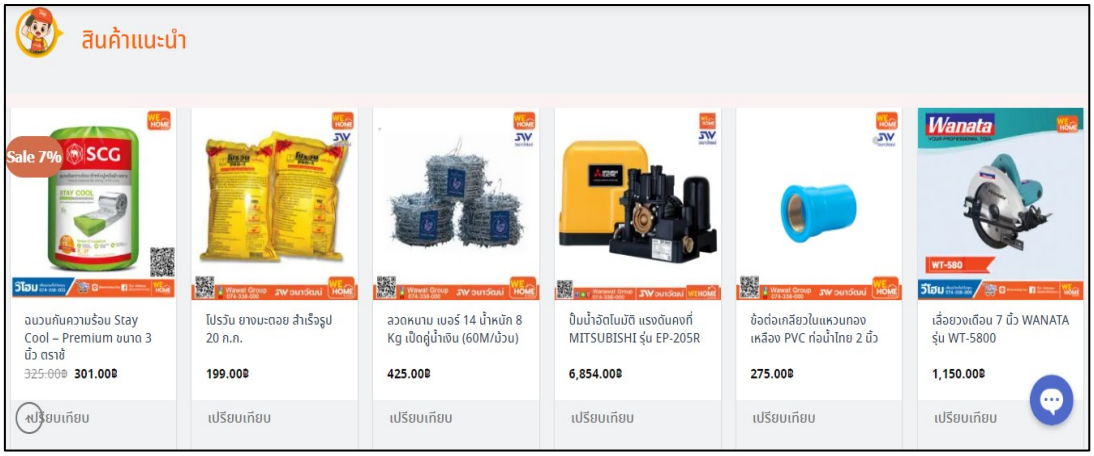
No	SKU	product_id	Similarity
1	GCBL065009031	GB-6327/50 ราวจับกันลื่น 25 มม. ขนาด 50 ซม. Watson	1.00
2	GCBL065009020	GB-6327-2B L/R ราวจับกันลื่น ตัวแอล Watson	1.00
3	GCBL065009028	HR-6500-T ราวพยุงกันล้ม 25MM รูปตัว T Watson	1.00
4	GCBL065009011	GB-4327-2/50 ราวจับกันลื่น 50CM Watson	1.00
5	GCBL065009010	GB-3327-2/60 ราวจับกันลื่น 60CM Watson	1.00
6	GCBL065009006	GB-3327/60 ราวจับกันลื่น 60CM Watson	1.00
7	GCBL065009019	GB-6327-2/60 ราวจับกันลื่น 60CM Watson	1.00
8	GCBL065009012	GB-4327-2/60 ราวจับกันลื่น 60CM Watson	1.00
9	GCBL065009008	GB-3327-2/35 ราวจับกันลื่น 35CM Watson	0.99
10	GCBL065009021	HR-1101/30 ราวจับกันลื่น 30CM ตัววี Watson	0.98



รูปที่ 4-17 ผลการแนะนำสินค้าตัวอย่างที่ 1 เรียงจากมากไปน้อย จำนวน 10 รายการ

4.2.4 ผลลัพธ์การพัฒนาระบบแนะนำสินค้า

การออกแบบการแสดงผลของระบบแนะนำบนระบบขายสินค้าออนไลน์เป็น 3 แบบ คือกรณีที่ลูกค้าไม่ Login เข้าระบบ และ Login แล้วแต่ไม่มีข้อมูลการซื้อหรือการคลิกดูสินค้าระบบจะแนะนำสินค้าที่ขายดีดังรูปที่ 4-18 แต่ในกรณีที่ลูกค้า Login เข้าระบบ ระบบจะแนะนำสินค้าตามบริบทโดยใช้ข้อมูลโปรไฟล์ผู้ใช้ (User Profile) ข้อมูลจากโมเดลข้อมูล (Context Profile) และข้อมูลรายการ (Item Profile) ซึ่งทำให้สินค้าที่แนะนำนั้นคล้ายคลึงกันกับสินค้าที่ผู้ใช้เคยซื้อหรือคลิกสินค้า



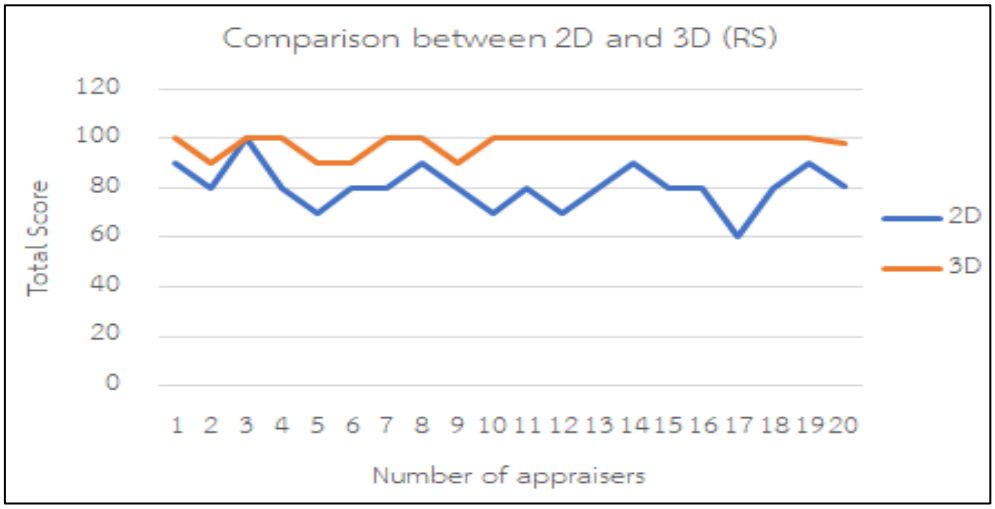
รูปที่ 4-18 ผลการสร้างระบบแนะนำแบบ 3 มิติ สำหรับลูกค้าไม่ Login

รูปที่ 4-19 ผลการสร้างระบบแนะนำแบบ 3 มิติ สำหรับลูกค้า Login

4.2.5 ผลการประเมินระบบแนะนำสินค้า

การประเมินประสิทธิภาพสำหรับระบบแนะนำสินค้าดำเนินการโดยการเปรียบเทียบระหว่างการสร้างแบบจำลองข้อมูล 2 มิติและ 3 มิติ โดยมีการประเมินผลดังนี้

1) ดำเนินการทดสอบกับสินค้าจำนวน 300 รายการ
2) ผู้ประเมินจำนวน 20 คน ได้แก่ พนังงาน IT และพนักงานการตลาดของบริษัท วนาวัฒน์วัสดุ จำกัด ทำการคลิกเลือกสินค้าที่สนใจจำนวน 5 รายการ จากนั้นระบบจะแนะนำสินค้าจำนวน 10 รายการ
3) ผู้ประเมินจำนวน 20 คน ให้คะแนนสินค้าที่ระบบแนะนำ โดยมีระดับคะแนนคือ 1 (ไม่ดี) ถึง 10 (ดี)
4) การประเมินผลสำหรับการสร้างแบบจำลองข้อมูลจะดำเนินการประเมินทั้ง 2 มิติและ 3 มิติ



รูปที่ 4-20 การเปรียบเทียบระหว่างการสร้างแบบจำลองข้อมูล 2 มิติและ 3 มิติ

รูปที่ 4-27 แสดงการเปรียบเทียบระหว่างการสร้างแบบจำลองข้อมูล 2 มิติและ 3 มิติ โดยผลคะแนนการสร้างแบบจำลองข้อมูล 3 มิติ คือ 97.8 คะแนน ซึ่งดีกว่า 2 มิติ ซึ่งได้คะแนนอยู่ที่ 80.5 คะแนน คะแนนต่ำสุดคือ 60.00 คะแนนสำหรับ 2 มิติ และสูงสุดคือ 100.00 สำหรับ 3 มิติ

4.2.6 ข้อมูลยอดขายสินค้า

สำหรับข้อมูลยอดขายสินค้าบนระบบขายสินค้าออนไลน์ ซึ่งระบบได้มีการเปิดให้ลูกค้าได้ใช้งานก่อนที่จะติดตั้งระบบแนะนำสินค้าที่วิจัยและพัฒนาขึ้น โดยก่อนการติดตั้งระบบมียอดขายสินค้าไม่เกิน 10 รายการดังรูปที่ 4-21 และเมื่อได้ทำการติดตั้งระบบแนะนำสินค้าแล้วทำให้ยอดขายสินค้าเพิ่มขึ้นเป็น 28 รายการ และมีรายการที่ลูกค้ากลับมาซื้อสินค้าซ้ำอีก 1 รายการ ดังรูปที่ 4-22

#31163 แพรวพรรณ ค่าชมภู	31163	Sep 24, 2021	Completed	฿799.00		Best Express 66850378370899
#31064 แพรวพรรณ ค่าชมภู	31064	Sep 17, 2021	Completed	฿799.00		Best Express 66850365900995
#31010 ธาร บุรณศิริ	31010	Sep 14, 2021	Completed	฿4,778.00		Best Express 66850361648774
#30607 เตชินท์ กิจพ่อค้า	30607	Aug 25, 2021	Completed	฿518.00		Kerry Express smkh0000540783r
#30288 ชัยวัฒน์ เพิ่มรนิล	30288	Aug 12, 2021	Completed	฿1,164.00		Best Express 66850342252642

รูปที่ 4-21 ข้อมูลยอดขายสินค้าก่อนการติดตั้งระบบแนะนำสินค้า

Order	Invoice Number	Date	Status	Total	Actions	Shipment Tracking
#41276 มุกดา ไพศาล ชัชวาล	41276	May 14, 2023	Completed	฿1,830.00		Best Express 66850711706319
#40450 ศรัณย์พจน์ ยะพงค์		Apr 10, 2023	Completed	฿770.00		Best Express 66850695952918
#40220 อัฐพล สมประสงค์		Mar 28, 2023	Completed	฿689.00		Best Express 66850691204975
#39977 อัฐพล สมประสงค์		Mar 15, 2023	Completed	฿1,328.00		Best Express 66850686440773
#39975 Teeraphat Traivanathum	39975	Mar 14, 2023	Completed	฿5,805.00		Best Express 66850688379637

รูปที่ 4-22 ข้อมูลยอดขายสินค้าหลังการติดตั้งระบบแนะนำสินค้า

บทที่ 5
บทสรุป อภิปรายผลและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้ได้นำเสนอการพัฒนาระบบแนะนำสินค้าแบบการรับรู้บริบทสำหรับอีคอมเมิร์ซ โดยแบ่งการพัฒนาออกเป็น 2 ส่วน คือ 1) การออกแบบและพัฒนาระบบขายสินค้าออนไลน์ 2) การออกแบบและพัฒนาระบบแนะนำ โดยสามารถสรุปผลของงานในแต่ละส่วนได้ดังนี้

5.1.1 การออกแบบและพัฒนาระบบขายสินค้าออนไลน์

การออกแบบและพัฒนาระบบขายสินค้าออนไลน์ผู้วิจัยได้ใช้ WordPress ซึ่งเป็นโปรแกรมประยุกต์สำหรับสร้างและบริหารจัดการเว็บไซต์ ผลที่ได้จากการวัดประสิทธิภาพการทำงานระบบขายสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัดที่ได้ดำเนินการออกแบบ พัฒนาและติดตั้ง โดยระบบสามารถใช้งานได้ดี ไม่มีความซับซ้อนหรือยุ่งยากแก่ผู้ใช้งาน และสะดวกในการใช้งาน ลดความผิดผลาดในการทำงานของผู้ใช้งานและผู้ดูแลระบบจากปัญหาของระบบขายสินค้าออนไลน์บริษัท วนาวัฒน์วัสดุ จำกัด ที่มีอยู่ โดยได้ทดสอบระบบการซื้อขายกับสินค้าจำนวน 4,000 รายการ

5.1.2 การออกแบบและพัฒนาระบบแนะนำ

การออกแบบและพัฒนาระบบแนะนำผู้วิจัยได้ดำเนินการวิจัยและพัฒนา Data model สำหรับระบบแนะนำสินค้า โดยใช้ข้อมูลสินค้า ได้แก่ ชื่อสินค้า รายละเอียดสินค้า และประเภทสินค้า สำหรับการจัดกลุ่มสินค้า การสร้าง Data model ได้ใช้วิธี Elbow method ในการหาค่า K ของจำนวนกลุ่มสินค้าเพื่อให้ได้ค่า K ที่ดีที่สุด และในงานวิจัยนี้ผู้วิจัยได้ใช้ K-means ในการจัดกลุ่มสินค้า โดยผลจากการทดลองการสร้าง Data model สำหรับ Context-based มีค่าจำนวนกลุ่มหรือ K เท่ากับ 9 และนำสินค้าจำนวน 4000 รายการให้ระบบดำเนินการจัดกลุ่ม ซึ่งระบบสามารถจัดกลุ่มสินค้าได้ และการวิจัยและพัฒนา Filtering model โดยผลจากการทดลองใช้วิธีการเปรียบเทียบระบบแนะนำ 2 มิติ และ 3 มิติที่พัฒนาขึ้น ผลที่ได้คือระบบแนะนำสินค้า 3 มิติ สามารถแนะนำสินค้าได้ตรงตามบริบทของผู้ใช้งานมากกว่า 2 มิติ เนื่องจากระบบแนะนำสินค้า 2 มิติจะแนะนำสินค้าที่คล้ายคลึงกับข้อมูลผู้ใช้งานเท่านั้น โดยไม่คำนึงถึงบริบทของผู้ใช้งาน แต่ระบบแนะนำ 3 มิติจะแนะนำสินค้าตามบริบทของผู้ใช้งาน ซึ่งจะมีการเลือกกลุ่มโดยใช้ Data model ที่ได้ดำเนินการวิจัย มาเป็นมิติที่ 3 ของระบบแนะนำ และสินค้าที่แนะนำจะแม่นยำกว่า 2 มิติ

5.2 ข้อจำกัดของการวิจัย

การวิจัยในครั้งนี้ มีการเก็บรวมรวบข้อมูลสินค้าจากเว็บไซต์ค้าขายวัสดุก่อสร้าง ซึ่งมีสินค้าบางหมวดหมู่ที่มีปริมาณไม่เท่ากัน ทำให้สามารถแนะนำสินค้าได้ในจำนวนจำกัด หากเก็บข้อมูลเพิ่มมากขึ้นจะทำให้สินค้าที่แนะนำมีความหลากหลาย และตรงตามความต้องการของผู้ใช้มากขึ้น

5.3 ข้อเสนอแนะ

ผู้วิจัยได้นำเสนอแนวทางในการศึกษา และพัฒนางานวิจัยนอนาคต โดยแบ่งออกเป็น 2 ส่วน ดังนี้

5.3.1 การประยุกต์ใช้งานวิจัย

1) ผลการวิจัยเกี่ยวกับระบบแนะนำสินค้าแบบการรับรู้บริบทสำหรับอีคอมเมิร์ซ สามารถนำผลที่ได้เป็นแนวทางไปประยุกต์ใช้เพื่อพัฒนาระบบแนะนำสินค้าให้มีความเหมาะสมมากยิ่งขึ้น
2) สามารถนำแนวทางในการพัฒนาระบบไปประยุกต์ใช้กับข้อมูลอื่นๆ หรือผลิตภัณฑ์อื่นๆ ตัวอย่างเช่น หนังสือ วิดีโอ เพลง เป็นต้น

5.3.2 การพัฒนางานวิจัยในอนาคต

ผู้วิจัยได้นำเสนอแนวทางการพัฒนางานวิจัยนอนาคต เพื่อต่อยอดจากงานวิจัยครั้งนี้ โดยมีรายละเอียด ดังนี้

1) ควรพิจารณาปัจจัยอื่นๆ ของธุรกิจค้าขายวัสดุก่อสร้างเพิ่มขึ้น นอกเหนือจากงานวิจัยนี้ เพื่อใช้ในการแนะนำสินค้าให้ตรงกับความต้องการของผู้ใช้มากยิ่งขึ้น
2) หากมีการเพิ่มปรีมาณของข้อมูลคำอธิบายของสินค้า (Description) อาจทำให้ระบบแนะนำสินค้ามีความแม่นยำมากยิ่งขึ้น
3) ควรนำจำนวนการสั่งซื้อสินค้าในแต่รายการของผู้ใช้ไปเป็นปัจจัยในการคำนวณสำหรับแนะนำสินค้าเพื่อความเหมาะสมมากยิ่งขึ้น

บรรณนานุกรม

[1] วรินทร์พัชร วัชรพงษ์เกษม. (2560). เอกสารประกอบการสอนรายวิชาพาณิชย์อิเล็กทรอนิกส์. มหาวิทยาลัยราชภัฏบุรีรัมย์

[2] ปาลิตา แสงศิริ, ระบบแนะนำร้านอาหารอัตโนมัติบนสมาร์ทโฟนโดยใช้ข้อมูลเชิงตำแหน่งและรายการอาหาร (วิทยานิพนธ์ปริญญามหาบัณฑิต คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์, 2561), หน้า 15.

[3] สุชาวดี ปลั่งศรี, การจัดกลุ่มขนาดผลิตภัณฑ์โดยการใช้ K-means Clustering เพื่อลดต้นทุนบรรจุภัณฑ์ (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิศวกรรมอุตสาหการและการจัดการ บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร, 2562), หน้า 6.

[4] ทัศนีย์ อุทัยสุริ, การสกัดคำสำคัญจากบทคัดย่อภาษาอังกฤษ (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาคอมพิวเตอร์บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร, 2556), หน้า 12.

[5] S. Sivapalan, A. Sadeghian, H. Rahnama, and A. M. Madni, “Recommender systems in e-commerce,” 2014 World Automation Congress (WAC), 179–184, 2014.

[6] M. Chen, “Research on recommender technology in E-commerce recommendation system,” 2010 2nd International Conference on Education Technology and Computer, 4, V4-409-V4-412, 2010.

[7] E. V. de Melo, E. A. Nogueira, and D. Guliato, “*Content-Based Filtering Enhanced by Human Visual Attention Applied to Clothing Recommendation,”* 2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI), 644–651, 2015.

[8] R. Yadav, A. Choorasiya, U. Singh, P. Khare and P. Pahade, "Recommendation System for E-Commerce Base on Client Profile", 2018 2nd International Conference on Trends in Electronics and Informatics, pp. 1316-1322, May 2018.

[9] N. M. N. Mathivanan, N. A. M. Ghani, and R. M. Janor, “Analysis of K-Means Clustering Algorithm: A Case Study Using Large Scale E-Commerce Products,” 2019 IEEE Conference on Big Data and Analytics (ICBDA), 1–4, 2019.

[10] R. Shrivastava and D. S. Sisodia, "Product Recommendations Using Textual Similarity Based Learning Models", 2019 International Conference on Computer Communication and Informatics ICCCI 2019, 2019.

[11] M. H. Mohamed, M. H. Khafagy, and M. H. Ibrahim, “Recommender systems challenges and solutions survey,” 2019 International Conference on Innovative Trends in Computer Engineering (ITCE), 149–155, 2019.

[12] J. Ramos, “Using tf-idf to determine word relevance in document queries,” In ICML, 2003.

[13] K. C. Lim, A. Selamat, M. H. Mohamed Zabil, M. H. Selamat, R. A. Alias, F. Puteh, F. Mohamed, and O. Krejcar, Quantifying Usability Prioritization Using K- Means Clustering Algorithm on Hybrid Metric Features for MAR Learning, Advancing Technology Industrialization Through Intelligent Software Methodologies, Tools and Techniques, IOS Press, 2019.

ภาคผนวก

ภาคผนวก ก.

ตัวอย่างแบบสอบถามเพื่อประเมินความพึงพอใจในการใช้งานระบบแนะนำตามบริบทสำหรับธุรกิจวัสดุก่อสร้างและอาคาร

แบบสอบถามเพื่อประเมินความพึงพอใจในการใช้งาน ระบบแนะนำสินค้าบนเว็บไซต์ขายสินค้าออนไลน์บริษัทวนาวัฒน์ วัสดุ จำกัด

คำชี้แจง แบบสอบถามความคิดเห็นนี้ เป็นการสอบถามข้อมูลความคิดเห็นของผู้ใช้ ภายหลังที่ได้ทดลองใช้ระบบแนะนำสินค้าที่พัฒนาขึ้น โดยแบ่งการประเมินออกเป็น 2 ส่วน ดังนี้

1. ประเมินผลการแนะนำของระบบแนะนำสินค้า 2 มิติ
2. ประเมินผลการแนะนำของระบบแนะนำสินค้า 3 มิติ

ขั้นตอนการดำเนินการ

- ผู้ใช้งานเข้าไปที่เว็บไซต์ขายสินค้าออนไลน์
- ผู้ใช้งานคลิกดูสินค้าที่ต้องการจะซื้อหรือสนใจ (โดยไม่ต้องซื้อ)
- จากนั้นระบบจะนำข้อมูลเหล่านั้นมาวิเคราะห์เพื่อแนะนำสินค้าที่เกี่ยวข้อง
- ผู้ใช้งานประเมินว่าสินค้าที่ระบบแนะนำนั้นตรงตามความต้องการของผู้ใช้งานหรือไม่

(เว็บไซต์อาจโหลดหน้า Home ช้าเนื่องจากวิเคราะห์ 2 ระบบพร้อมกัน)

kliangklaosutthirat@gmail.com สลับบัญชี

ไม่ใช้ร่วมกัน

ส่วนที่ 1

สถานภาพของผู้ตอบแบบสอบถาม *

- ○ พนักงาน
- ○ ผู้ใช้งานทั่วไป

ส่วนที่ 2 การประเมินผลการแนะนำสินค้าของระบบแนะนำสินค้า 2 มิติ

<u>คำถามประเมิน</u> *

ระบบแนะนำสินค้าบนเว็บไซต์ขายสินค้าออนไลน์ของบริษัทวนาวัฒน์ วัสดุ จำกัด สามารถแนะนำสินค้าได้ตรงตามความต้องการของท่านหรือไม่

**** 1 = ใกล้เคียงกับสินค้าที่คลิก 2 = ไม่ใกล้เคียงกับสินค้าที่คลิก*

	1 = ใกล้เคียง	2 = ไม่ใกล้เคียง
สินค้าแนะนำรายการที่ 1	◯	◯
สินค้าแนะนำรายการที่ 2	◯	◯
สินค้าแนะนำรายการที่ 3	◯	◯
สินค้าแนะนำรายการที่ 4	◯	◯
สินค้าแนะนำรายการที่ 5	◯	◯
สินค้าแนะนำรายการที่ 6	◯	◯
สินค้าแนะนำรายการที่ 7	◯	◯
สินค้าแนะนำรายการที่ 8	◯	◯
สินค้าแนะนำรายการที่ 9	◯	◯
สินค้าแนะนำรายการที่ 10	◯	◯

ส่วนที่ 3 การประเมินผลการแนะนำสินค้าของระบบแนะนำสินค้า 3 มิติ

<u>คำถามประเมิน</u> *

ระบบแนะนำสินค้าบนเว็บไซต์ขายสินค้าออนไลน์ของบริษัทวนาวัฒน์ วัสดุ จำกัด สามารถแนะนำสินค้าได้ตรงตามความต้องการของท่านหรือไม่

**** 1 = ใกล้เคียงกับสินค้าที่คลิก 2 = ไม่ใกล้เคียงกับสินค้าที่คลิก*

	1 = ใกล้เคียง	2 = ไม่ใกล้เคียง
สินค้าแนะนำรายการที่ 1	◯	◯
สินค้าแนะนำรายการที่ 2	◯	◯
สินค้าแนะนำรายการที่ 3	◯	◯
สินค้าแนะนำรายการที่ 4	◯	◯
สินค้าแนะนำรายการที่ 5	◯	◯
สินค้าแนะนำรายการที่ 6	◯	◯
สินค้าแนะนำรายการที่ 7	◯	◯
สินค้าแนะนำรายการที่ 8	◯	◯
สินค้าแนะนำรายการที่ 9	◯	◯
สินค้าแนะนำรายการที่ 10	◯	◯

ภาคผนวก ข.

ผลงานตีพิมพ์เผยแพร่จากวิทยานิพนธ์

2022 26th International Computer Science and Engineering Conference (ICSEC) | 978-1-6654-9198-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSEC56337.2022.10049331

2022 26th International Computer Science and Engineering Conference (ICSEC)

Contextual Data Modeling for Recommender System in Building and Construction Materials Business

Sutthirat Kliangklao
Department of Computer Engineering
Faculty of Engineering, Prince of Songkla University
Songkhla, Thailand
6310120060@email.psu.ac.th

Nikom Suvonvorn
Department of Computer Engineering
Faculty of Engineering, Prince of Songkla University
Songkhla, Thailand
nikom.s@psu.ac.th

***Abstract*— The recommendation system is one of the most important supported technologies to e-commerce that aims for recommending products or services to increase customer's satisfaction. In this paper, we propose the method for Contextual Data Modeling as an improved version of Hybrid Filtering to introduce the context-awareness in the building and construction materials business. The recommendation along with e-commerce system are built, deployed, and tested in the real situation. The evaluation score is up to 97.8 compared to baseline.**

Keywords*— *E-Commerce, Recommendation, Context-Awareness, K-means, Thai NLP

I. Introduction

The businesses have been evolved by using digital technology platform to support their activities, such as, the purchase of goods and services via the internet payment, known as e-commerce. The recommendation system is one of the most important supported technologies to e-commerce that aims for recommending products or services to increase customer's satisfaction. The system [1] relies on customer information, such as, their profiles, purchase history, likes, ratings, comments, etc. It collects and analyzes the information to obtain a list of suggesting products that meet the needs of specific customer in a circumstance or context.

Our research is focus on the recommendation system for the building and construction materials business. The sales consist of materials, supplies, and tools. Their customers are both professional builders and ordinary customers. The system will support customers by recommending products to meet their needs. The important factors affecting customer's satisfaction depend on not only obtaining the needed items, but also receiving the interesting items with good price and quality, which will be intentionally proposed for the benefit of customers by the recommending system. In the studying business, most existing online system does not have the specific recommendation method. Usually, the products are displayed by categories, and can be search by keywords. Most products are recommended based on that information.

Three main approaches are currently used for recommender system. Content-Based Filtering (CBF) [2] based on keywords and attributes of products that resemble to user profile for recommending products. Collaborative Filtering (CF) [3] relies on the preferences of similar users to recommend products. Hybrid Filtering (HF) [4] combines both techniques CBF and CF. Some unsolved problems exist, such as, cold-start problem happening when new product has not enough interaction information to decision, shilling attacks when false rating increase, synonymy when an item have different names with similar meaning, limited content or over-spec, sparsity when dispersed profile matrix leading to less accurate, and context-awareness [5].

In this paper, we propose the method of Contextual Data Modeling as an improved version of HF. The method is established based on context data modeling to improve the context-awareness for the building and construction materials business. The paper is organized as follows: section 2 describes the details of proposed method; section 3 shows results and discussions; conclusion in section 4.

II. PROPOSAED METHOD

The overall concept of Contextual Data Modeling is displayed in the Fig.1. In recent concept of recommender system, at least two dimensions of data are required for establishing the similarity matrix, such as, the profiles of users and items. The similarity measurement will determine the top-N predicted items and defined as recommended items. Our idea is to add one more dimension, the 3rd dimension for contextual information, which can be anything that suits the specific problem, for example, information from mobile services includes location, entries in the calendar, social networks, etc. In the building and construction materials business, recommending products based on only the product's profile and user's profile is far from efficiency. Customer come with specific problems or needs that require different kinds of products to be combined as a solution to meet the need, while ignoring or less considering the categories, ratings, other users needed, etc.

978-1-6654-9198-3/22/$31.00 ©2022 IEEE

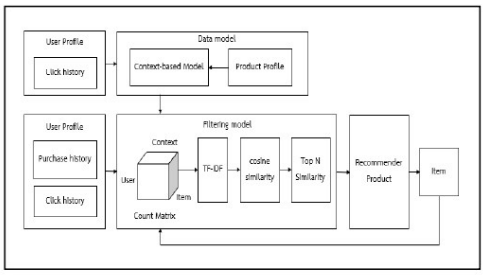


Fig. 1. Concept of Contextual Data Modeling

The product's description contains information about the possible application and characteristic of products that can be used or applied as contextual information. Products, from different categories, rating level, etc., can be grouped into the same class. For example, if customer interest in home paint color, the desirable recommended products should include not only other types of colors, but also painted plot, thinner, etc., where there are in the different categories but same application.

Our proposed concept, the product's description as contextual information must be clustered into classes using its own similarity measurement reflecting the solving problem by applying existing unsupervised learning techniques. The obtained classes as context will be assigned as the 3rd dimension data of similarity matrix. In our initial prove of proposed idea, we define the 3rd dimension as clustered products obtained by K-means clustering technique using the product's descriptions. Our method is described into two main steps: contextual data modeling and filtering model

A. Contextual Data Modeling

The objective of contextual data modeling is to obtain the classes from contextual information, which are the descriptions, and categories of products. Fig.2 describes the method compose of four steps: data pre-processing, feature extraction, clustering, and selection.

a) Data Pre-processing: the fundamental techniques of data pre-processing [6] are applied, such as, punctuation removal, tokenization for splitting paragraphs or sentences into words, stemming and lemmatization for reducing words into its root words. In this step the PyThaiNLP is used.

b) Feature extraction: in this step the Term Frequency-Inverse Document Frequency (TF-IDF) feature [7] is extracted from the description of products. TF-IDF is numerical statistic that express the important of words form each product's description. TF-IDF is calculated as:

$$TFIDF_{d,t} = FREQ_{t,d}\left(1 + \log\left(\frac{\mathrm{N}}{DFREQ_t}\right)\right) \quad (1)$$

Where $FREQ_{t,d}$ = number of word t in the product d
N = Total number of products
$DFREQ_t$ = Number of products where word t appears

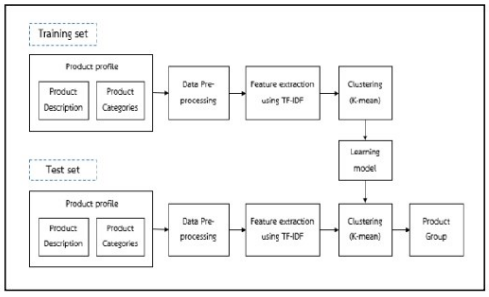


Fig. 2. Context-based model

c) Clustering: in this step, K-Means [9] is used for clustering products into clusters or classes by considering the similarity or distance of words using Euclidean distance, between each product to the centroids of clusters.

d) Cluster Selection: in this step, a cluster is selected by considering the SKU numbers from click history. Then, all products in the cluster will be counted as context profile for the later procedure of filtering model.

B. Filtering Model

In this filtering step, the count matrix is firstly established from the three types of data: user profile, item profile and context profile. Secondly, the cosine similarity is applied to calculate the similarities of count matrix between any products in context profile with the others listed in the user and item profiles. Then, the top-N ranking of similarities is provided as recommended products.

a) Count Matrix: the count matrix is built using tags (keywords) versus products. The tags are generally defined from the characteristics of products. The products of user profile are established from the click and purchase history of an individual user. The products of item profile are obtained from previous recommended items. The examples of count matrix are shown in Table I, II, and III.

TABLE I. THE EXAMPLE OF COUNT MATRIX OF USER PROFILE

User	โซฟา	จนวน	เลื่อย	เก้าอี้	ที่นั่ง
Product A	2	0	0	3	2
Product B	1	0	0	2	2
Product C	0	0	0	0	0
Product D	0	0	0	0	0

TABLE II. THE EXAMPLE OF COUNT MATRIX OF ITEM PROFILE

Item	โซฟา	ฉนวน	เลื่อย	เก้าอี้	ที่นั่ง
Product A	1	0	0	0	2
Product B	0	0	0	0	0
Product C	1	0	0	0	2
Product D	0	0	0	0	0

TABLE III. THE EXAMPLE OF COUNT MATRIX OF CONTEXT PROFILE

Context	โซฟา	ฉนวน	เลื่อย	เก้าอี้	ที่นั่ง
Product 1	3	0	0	3	2
Product 2	2	0	0	2	2
Product 3	2	0	0	0	2
Product 4	0	0	0	0	0

b) TF-IDF Extraction: The TF-IDF features are calculated from count matrix, listed in Tables I, II, and III, using the formula of Equation 1. The feature values show how important a tag can be invariant with respect to different products. The examples of TF-IDF of count matrix are shown in tables IV, V, and VI.

TABLE IV. THE EXAMPLE OF TF-IDF OF USER PROFILE

User	โซฟา	ฉนวน	เลื่อย	เก้าอี้	ที่นั่ง
Product A	0.38	0	0	0.56	0.38
Product B	0.26	0	0	0.52	0.52
Product C	0	0	0	0	0
Product D	0	0	0	0	0

TABLE V. THE EXAMPLE OF TF-IDF OF ITEM PROFILE

Item	โซฟา	ฉนวน	เลื่อย	เก้าอี้	ที่นั่ง
Product A	0.43	0	0	0	0.87
Product B	0	0	0	0	0
Product C	0.43	0	0	0	0.87
Product D	0	0	0	0	0

TABLE VI. THE EXAMPLE OF TF-IDF OF CONTEXT PROFILE

Context	โซฟา	ฉนวน	เลื่อย	เก้าอี้	ที่นั่ง
Product 1	0.42	0	0	0.49	0.28
Product 2	0.27	0	0	0.43	0.37
Product 3	0.56	0	0	0	0.56
Product 4	0	0	0	0	0

a) Cosine Similarity: in this step the cosine similarity is applied to determine the similarity between products of context profile and products from user profile and item profile. Cosine similarity is calculated as:

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

The cosine similarity calculation is performed twice using the TF-IDF values. The first time, A is the vector of TF-IDF of user profile and B is the vector of the TF-IDF of context profile. Then, in the second time, A is the vector of TF-IDF of item profile and B is the vector of TF-IDF of context profile.

TABLE VII. THE EXAMPLE OF SIMILARITY

	User				Item			
Context	Product A	Product B	Product C	Product D	Product A	Product B	Product C	Product D
Product 1	0.99	0.93	0	0	0.62	0	0.62	0
Product 2	0.99	099	0	0	0.72	0	0.72	0
Product 3	0.69	0.71	0	0	0.95	0	0.95	0
Product 4	0	0	0	0	0	0	0	0

Table VII shows the examples of similarities of user, item, and context. Note that the similarity results of product 1, product 2, and product 3 are similar with product A and product B of user profile, and with product A and product C of item profile.

The top-N ranking of similarities are considered as the recommended products.

III. RESULT

This research is funded by Higher Education for Industry Consortium (HIFI) program. The proposed e-commerce platform is developed, evaluated, and deployed during the two years program by researcher. The recommender system as backend service of the platform is also developed, evaluated, deployed for real application, as shown at wehome.co.th. The parameters of system are calibrated and evaluated as follows.

A. Dataset

The dataset using for experiment received from the e-commerce platform includes 4,000 products of building and construction materials.

B. Experiment

a) Optimal value of k-clusters: in k-means clustering, the number of k cluster is needed to be determined before performing the analysis. The elbow method is applied to runs k-means on the dataset for values k from 2 to 20, then for each value of k computes an average score for all clusters. The distortion score is computed, the sum of square distances from each point to its assigned center.

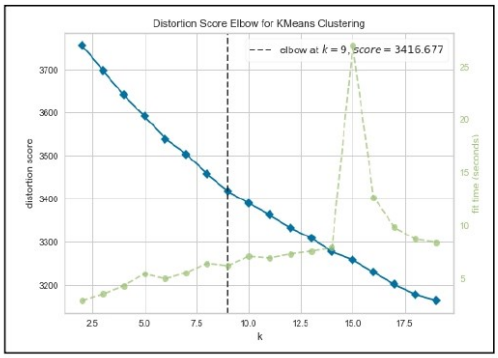


Fig. 3. Optimal value of k-clusters

Fig. 3 visualizes the distortion values versus k-cluster, we found that k = 9 is an optimal value with a distortion score of 3416.677.

b) K-means Results: Table VIII shows the example of products and its categories manually defined by sale department.

TABLE VIII. EXAMPLE OF PRODUCTS AND CATEGORIES

ID	Product	Categories
101	โซฟา 3 ที่นั่ง แบบปรับเอนนอน	เฟอร์นิเจอร์
102	สว่านไร้สาย	เครื่องมือช่าง/อุปกรณ์ก่อสร้าง
103	ฉนวนกันความร้อน	วัสดุก่อสร้าง
104	เสื่อยางเตือน	เครื่องมือช่าง/อุปกรณ์ก่อสร้าง
105	หม้อหุงข้าว	เครื่องใช้ไฟฟ้า

Table IX shows example of clusters obtained after the data pre-processing, feature extraction, and clusters using K-Mean.

TABLE IX. EXAMPLE OF CLUSTERS USING K-MEANS

ID	Product	Clusters
101	โซฟา 3 ที่นั่ง แบบปรับเอนนอน	3
102	สว่านไร้สาย	6
103	ฉนวนกันความร้อน	3
104	เสื่อยางเตือน	6
105	หม้อหุงข้าว	3

From table VIII and IX, we can notice that products in the different categories can be grouped in the same cluster, for example, โซฟา 3 ที่นั่ง แบบปรับเอนนอน, ฉนวนกันความร้อน, หม้อหุงข้าว are in different categories but the same 3-cluster.

c) Evaluation: the performance evaluation for recommendation system is executed by the comparison between two dimensions and three dimensions data modeling. The evaluation are follows:

- 300 products tested
- 20 users, experts form store, randomly select 5 products, 10 products are recommended for each selection
- 20 users score the recommended products, 1 (bad) to 10 (good) scores.
- blind evaluation for both two dimensions and three dimensions data modeling

Fig. 4 shows the results that the overall score for three dimensions, 97.8, is much better than the two dimensions, 80.5, in all cases. The lowest score is 60.00 for two dimensions, and the highest is 100.00 for three dimensions. Because of three dimensions has cluster selection in contextual data modeling before sending data to context profile, which reduces the recommended data and more accurate, but two dimensions do not have contextual data modeling, resulting in many suggested products and less accurate than three dimensions.

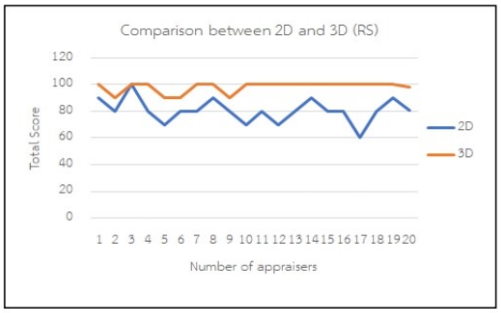


Fig. 4. Comparison between two dimensions and three dimensions data modeling

IV. CONCLUSION

We propose the Contextual Data Modeling as contextual data modeling in recommender system applied to the building and construction materials business. The TF-IDF features are extracted from product's description and used by K-means to group relevant products into clusters. In addition to user profiles and product profiles, the clusters are defined as contextual information, the third dimension, for calculating the similarities for recommending products. The e-commerce platform along with the backend for recommender system are

built, deployed, and evaluated in the real situation. The experiments are executed using 4,000 data records from real situation dataset and evaluated with 20 expert users. The evaluation score of three dimension is up to 97.8 compared to two dimensions.

REFERENCES

[1] S. Sivapalan, A. Sadeghian, H. Rahnama, and A. M. Madni, "Recommender systems in e-commerce," 2014 World Automation Congress (WAC), 179–184, 2014.

[2] M. Chen, "Research on recommender technology in E-commerce recommendation system," 2010 2nd International Conference on Education Technology and Computer, 4, V4-409-V4-412, 2010.

[3] E. V. de Melo, E. A. Nogueira, and D. Guliato, "*Content-Based Filtering Enhanced by Human Visual Attention Applied to Clothing Recommendation,"* 2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI), 644–651, 2015.

[4] T. Badriyah, E. T. Wijayanto, I. Syarif and P. Kristalina, "A hybrid recommendation system for E-commerce based on product description and user profile", 2017 Seventh International Conference on Innovative Computing Technology (INTECH), 95-100, 2017.

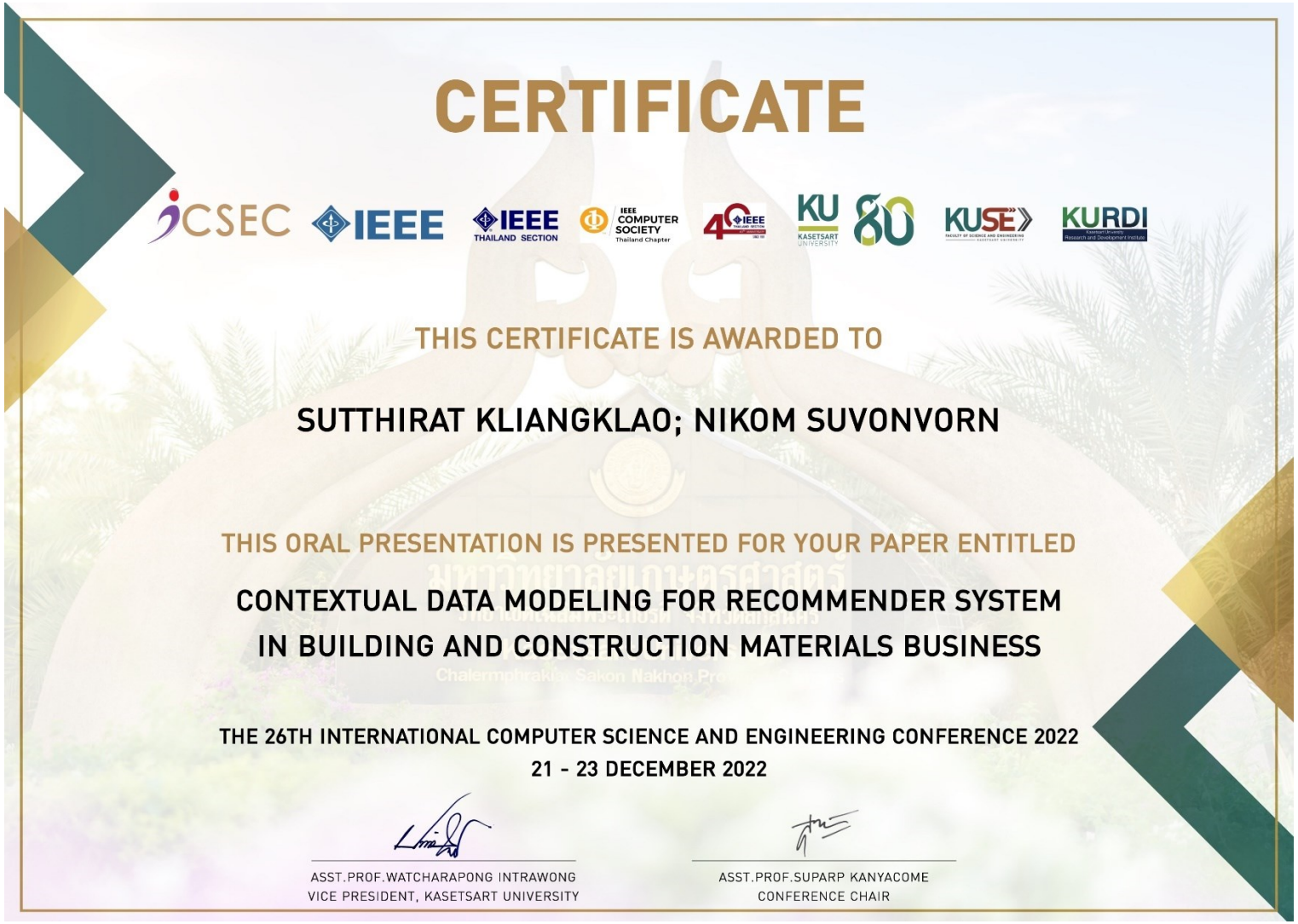
[5] M. H. Mohamed, M. H. Khafagy, and M. H. Ibrahim, "Recommender systems challenges and solutions survey," 2019 International Conference on Innovative Trends in Computer Engineering (ITCE), 149–155, 2019.

[6] N. M. N. Mathivanan, N. A. M. Ghani, and R. M. Janor, "Improving Classification Accuracy Using Clustering Technique," Bulletin of Electrical Engineering and Informatics, vol. 7, no. 3, pp. 465–470, 2018.

[7] J. Ramos, "Using tf-idf to determine word relevance in document queries," In ICML, 2003.

[8] N. M. N. Mathivanan, N. A. M. Ghani, and R. M. Janor, "Analysis of K-Means Clustering Algorithm: A Case Study Using Large Scale E-Commerce Products," 2019 IEEE Conference on Big Data and Analytics (ICBDA), 1–4, 2019.

[9] K. C. Lim, A. Selamat, M. H. Mohamed Zabil, M. H. Selamat, R. A. Alias, F. Puteh, F. Mohamed, and O. Krejcar, Quantifying Usability Prioritization Using K- Means Clustering Algorithm on Hybrid Metric Features for MAR Learning, Advancing Technology Industrialization Through Intelligent Software Methodologies, Tools and Techniques, IOS Press, 2019.


CERTIFICATE
ICSEC
IEEE
IEEE THAILAND SECTION
IEEE COMPUTER SOCIETY Thailand Chapter
KU KASETSART UNIVERSITY
KUSE
KURDI
THIS CERTIFICATE IS AWARDED TO
SUTTHIRAT KLIANGKLAO; NIKOM SUVONVORN
THIS ORAL PRESENTATION IS PRESENTED FOR YOUR PAPER ENTITLED
CONTEXTUAL DATA MODELING FOR RECOMMENDER SYSTEM
IN BUILDING AND CONSTRUCTION MATERIALS BUSINESS
THE 26TH INTERNATIONAL COMPUTER SCIENCE AND ENGINEERING CONFERENCE 2022
21 - 23 DECEMBER 2022
ASST.PROF.WATCHARAPONG INTRAWONG
VICE PRESIDENT, KASETSART UNIVERSITY
ASST.PROF.SUPARP KANYACOME
CONFERENCE CHAIR

ประวัติผู้เขียน

ชื่อ สกุล นายสุทธิรัตน์ เกลี้ยงเกลา

รหัสประจำตัวนักศึกษา 6310120060

วุฒิการศึกษา

วุฒิ	ชื่อสถาบัน	ปีที่สำเร็จการศึกษา
วิทยาศาสตรบัณฑิต (เทคโนโลยีสารสนเทศ)	มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย	2562

ทุนการศึกษา

ทุนโครงการการอุดมศึกษาเพื่ออุตสาหกรรม (Higher Education for Industry: Hi-FI)

ตำแหน่งและสถานที่ทำงาน

พ.ศ. 2563 - ปัจจุบัน

โปรแกรมเมอร์ บริษัท วนาวัฒน์วัสดุ จำกัด สงขลา

การตีพิมพ์เผยแพร่ผลงาน

S. Kliangklao and N. Suvonvorn, “Contextual Data Modeling for Recommender System in Building and Construction Materials Business,” in Proceedings of ICSEC 2022 International Computer Science and Engineering Conference, Sakon Nakhon, Thailand, 2022, pp. 172-176.