วารสารวิศวกรรมศาสตร์และนวัตกรรม
Journal of Engineering and Innovation

บทความวิจัย

การวิเคราะห์ประเด็นอาหารไทยจากคำบรรยาย Youtube ด้วยเทคนิค LDA และ NMF

Analysis of Thai food topics from Youtube transcripts using LDA and NMF techniques

เอกชัย แซ่จึง ธนินทร์ ระเบียบโพธิ์* สรวิศ ต.ศิริวัฒนา สุภาวดี พบพิมาย

สาขาเทคโนโลยีสารสนเทศและการสื่อสาร คณะวิทยาศาสตร์และศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน 30000

Ekkachai Jueng Thanin Rabiabpho* Soravit T.Siriwattana Supawadee Phobphimai

Department of Information and Communication Technology, Faculty of Science and Liberal Arts,
Rajamangala University of Technology Isan 30000

* Corresponding author.

E-mail: thanin@sci.rmuti.ac.th; Telephone: 0 4423 3070

วันที่รับบทความ 12 มีนาคม 2568; วันที่แก้ไขบทความ ครั้งที่ 1 24 มิถุนายน 2568; วันที่ตอบรับบทความ 9 กันยายน 2568

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อวิเคราะห์มุมมองที่มีต่ออาหารไทยจากผู้ใช้งานต่างประเทศผ่านคำบรรยายในวิดีโอยูทูป โดยเปรียบเทียบประสิทธิภาพของเทคนิค Latent Dirichlet Allocation (LDA) และ Non-negative Matrix Factorization (NMF) ในการจำแนกประเด็นเกี่ยวกับอาหารไทย ข้อมูลถูกรวบรวมจากวิดีโอยูทูปจำนวน 352 รายการที่เผยแพร่ระหว่างปี พ.ศ. 2553-2567 โดยได้รับการประมวลผลข้อมูลเบื้องต้นและสร้าง N-Gram ก่อนนำไปวิเคราะห์ ผลการศึกษาพบว่าเทคนิค LDA ให้ผลลัพธ์ที่ดีที่สุดที่มีค่าคะแนนความสอดคล้องสูงที่สุดเท่ากับ 0.794 ที่จำนวน 5 หัวข้อ ได้แก่ (1) วัตถุดิบและรสชาติอาหาร (2) ประสบการณ์การท่องเที่ยวและอาหาร (3) องค์ประกอบอาหารไทย (4) อาหารริมทางและความนิยม และ (5) กระบวนการปรุงอาหารและวัตถุดิบ ในขณะที่เทคนิค NMF ให้ผลลัพธ์ที่ดีที่สุดที่มีค่าคะแนนความสอดคล้องสูงที่สุดเท่ากับ 0.956 ที่จำนวน 8 หัวข้อได้แก่ (1) วัตถุดิบและรสชาติอาหารทะเล (2) ความชื่นชอบรสชาติ (3) วัฒนธรรมอาหารริมทาง (4) ส่วนประกอบของแกงไทย (5) ประสบการณ์การทานและรสชาติอาหาร (6) การสำรวจอาหาร (7) อาหารประเภทข้าวและเนื้อสัตว์ และ (8) ประสบการณ์ร้านอาหารและการบริการ จากภาพรวมของผลลัพธ์ทั้งสองเทคนิคนั้นผู้ใช้งานต่างประเทศมองอาหารไทยใน 4 มิติหลัก ได้แก่ รสชาติอันเป็นเอกลักษณ์ วัตถุดิบเฉพาะ ประสบการณ์การรับประทาน และบริบทแวดล้อม การเปรียบเทียบทั้งสองเทคนิคพบว่า LDA มีจุดเด่นในการแสดงความเชื่อมโยงระหว่างองค์ประกอบต่างๆ ของอาหารไทยผ่านการซ้อนทับของคำสำคัญ แต่มีข้อจำกัดในการแยกแยะประเด็นย่อย ในขณะที่ NMF มีความโดดเด่นในการจำแนกประเด็นที่มีความเฉพาะเจาะจง แต่ขาดการเชื่อมโยงระหว่างหัวข้อ การเลือกใช้เทคนิคใดควรพิจารณาจากวัตถุประสงค์การใช้งาน โดยอาจผสมผสานผลลัพธ์จากทั้งสองเทคนิคเพื่อให้ได้มุมมองที่ครบถ้วนในการพัฒนากลยุทธ์ส่งเสริมอาหารไทยในต่างประเทศ

คำสำคัญ

อาหารไทย เหมืองข้อความ การวิเคราะห์หัวข้อ ยูทูป การวิเคราะห์สื่อสังคมออนไลน์

Abstract

This study aims to analyze the views of international users on Thai food through the transcripts in YouTube videos by comparing the performance of Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF) techniques in identifying Thai food-related topics. Data was collected from 352 YouTube videos published between 2010-2024, data preprocessed, and transformed into N-Grams before analysis. The results revealed that LDA achieved optimal performance with the highest coherence score of 0.794 at 5 topics: (1) ingredients and food flavors, (2) tourism and food experiences, (3) Thai food components, (4) street food and popularity, and (5) cooking processes and ingredients. Meanwhile, NMF

yielded optimal results with the highest coherence score of 0.956 at 8 topics: (1) seafood ingredients and flavors, (2) taste preferences, (3) street food culture, (4) Thai curry components, (5) dining experiences and food flavors, (6) food exploration, (7) rice and meat dishes, and (8) restaurant experiences and service. From the overall results of both techniques, international users perceive Thai food in four main dimensions: distinctive flavors, specific ingredients, dining experiences, and environmental context. Comparing both techniques, LDA excelled in showing connections between Thai food components through keyword overlap but had limitations in distinguishing subtopics. Meanwhile, NMF demonstrated superiority in identifying specific topics but lacked connections between them. The choice of technique should depend on the intended application, and combining results from both techniques may provide the most comprehensive perspective for developing strategies to promote Thai food internationally.

Keywords

Thai food; text mining; topic modeling; Youtube; online social media analysis;

1. คำนำ

อาหารไทยได้รับการยอมรับในระดับนานาชาติด้วยเอกลักษณ์ของรสชาติที่ผสมผสานทั้งเปรี้ยว หวาน เค็ม เผ็ด และการใช้สมุนไพรและเครื่องเทศที่หลากหลาย ทำให้อาหารไทยกลายเป็นหนึ่งในทรัพยากรทางวัฒนธรรมที่สำคัญของประเทศ [1] และเป็นแรงดึงดูดนักท่องเที่ยวจากทั่วโลก ตามรายงานของสำนักงานนโยบายและยุทธศาสตร์การค้า [2] ระบุว่า อุตสาหกรรมอาหารมีแนวโน้มขยายตัวต่อเนื่องใน ปี 2566 มีผลิตภัณฑ์มวลรวมภายในประเทศ (GDP) 1.14 ล้านล้านบาท ขยายตัวจากปีก่อนหน้าร้อยละ 9.1 คิดเป็นสัดส่วนร้อยละ 6.3 ของ GDP และประเทศไทยเป็นผู้ส่งออกสินค้าอาหารอันดับที่ 14 ของโลก อาหารไทยจึงเป็นส่วนสำคัญของการส่งเสริมภาพลักษณ์ของประเทศไทยในเวทีโลกในเชิงวัฒนธรรม การท่องเที่ยว และเศรษฐกิจ

ในยุคดิจิทัล แพลตฟอร์มสื่อสังคมออนไลน์อย่างยูทูป (YouTube) ได้กลายเป็นช่องทางสำคัญในการเผยแพร่และแลกเปลี่ยนข้อมูลเกี่ยวกับอาหารและวัฒนธรรมการกินของแต่ละชาติ นักท่องเที่ยว บล็อกเกอร์ และผู้สร้างเนื้อหา (Content Creator) จำนวนมากได้แบ่งปันประสบการณ์การรับประทานอาหารไทยผ่านวิดีโอ พร้อมด้วยคำบรรยายที่สะท้อนความคิดเห็น ความรู้สึก และการรับรู้ต่ออาหารไทย [3] ข้อมูลเหล่านี้จึงเป็นแหล่งข้อมูลอันมีค่าที่สามารถนำมาวิเคราะห์เพื่อทำความเข้าใจมุมมองและช่วยสร้างการรับรู้และความสนใจอาหารไทยให้กับผู้บริโภคในตลาดต่างประเทศผ่านการฑูตอาหาร (Gastrodiplomacy) [2]

แม้ว่าจะมีการศึกษาเกี่ยวกับภาพลักษณ์อาหารไทยในมุมมองของชาวต่างประเทศจำนวนหนึ่ง แต่ส่วนใหญ่มักใช้วิธีการเก็บข้อมูลแบบดั้งเดิม เช่น การสำรวจด้วยแบบสอบถามหรือการสัมภาษณ์ [4-5] ซึ่งอาจมีข้อจำกัดในการเข้าถึงกลุ่มตัวอย่างที่หลากหลายและการได้ข้อมูลที่เป็นธรรมชาติ การวิเคราะห์ข้อมูลจากสื่อสังคมออนไลน์โดยใช้เทคนิคการทำเหมืองข้อความและการเรียนรู้ของเครื่องจึงเป็นแนวทางใหม่ที่ช่วยให้สามารถค้นหาแบบแผนและประเด็นสำคัญจากข้อมูลปริมาณมากได้อย่างมีประสิทธิภาพ

Latent Dirichlet Allocation (LDA) และ Non-negative Matrix Factorization (NMF) เป็นสองเทคนิคที่ได้รับความนิยมในการจำแนกหัวข้อ (Topic Modeling) จากข้อความ โดย LDA เป็นแบบจำลองการสร้างแบบความน่าจะเป็น (Probabilistic Generative Model) ที่มองว่าเอกสารคือการผสมผสานของหัวข้อต่างๆ และแต่ละหัวข้อคือการผสมผสานของคำต่างๆ [6] ในขณะที่ NMF เป็นเทคนิคการแยกเมทริกซ์ที่มีค่าไม่เป็นลบออกเป็นเมทริกซ์ย่อยสองเมทริกซ์ ซึ่งสามารถตีความในบริบทของการจำแนกหัวข้อได้คล้ายคลึงกับ LDA [7] ทั้งสองเทคนิคมีจุดเด่นและข้อจำกัดที่แตกต่างกัน จึงมีความน่าสนใจในการเปรียบเทียบประสิทธิภาพของทั้งสองเทคนิคในการวิเคราะห์ข้อมูลประเภทเดียวกัน

การวิจัยนี้มุ่งศึกษาและวิเคราะห์มุมมองของชาวต่างประเทศที่มีต่ออาหารไทยผ่านคำบรรยายในวิดีโอยูทูป โดยใช้เทคนิคการประมวลผลภาษาธรรมชาติและการเรียนรู้ของเครื่องด้วยการเปรียบเทียบประสิทธิภาพของเทคนิค LDA และ NMF ในการจำแนกประเด็นเกี่ยวกับอาหารไทย การวิจัยนี้ไม่

เพียงแต่จะช่วยให้เข้าใจการรับรู้ของชาวต่างประเทศที่มีต่ออาหารไทยในมิติต่างๆ แต่ยังเป็นการประเมินความเหมาะสมของแต่ละเทคนิคในการวิเคราะห์ข้อมูลลักษณะนี้ ซึ่งสามารถนำไปประยุกต์ใช้กับการวิเคราะห์ข้อมูลจากสื่อสังคมออนไลน์ในบริบทอื่นๆ ต่อไป

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 การทำเหมืองข้อความและการวิเคราะห์เนื้อหา

การทำเหมืองข้อความ (Text Mining) เป็นกระบวนการสกัดข้อมูลที่มีคุณค่าจากข้อความที่ไม่มีโครงสร้าง โดยใช้เทคนิคทางการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) และการเรียนรู้ของเครื่อง (Machine Learning) เพื่อค้นหารูปแบบ แนวโน้ม และความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลเหล่านั้น [8]

ในยุคดิจิทัลที่มีการผลิตข้อมูลเนื้อหาออนไลน์จำนวนมหาศาล การทำเหมืองข้อความได้กลายเป็นเครื่องมือสำคัญในการวิเคราะห์ข้อมูลหลากหลายรูปแบบ ทั้งความคิดเห็นบนโซเชียลมีเดีย บทความข่าว บทวิจารณ์สินค้า หรือแม้แต่คำบรรยายในวิดีโอ [9] โดยเฉพาะอย่างยิ่งในการศึกษาด้านการท่องเที่ยวและอาหาร นักวิจัยได้นำเทคนิคการทำเหมืองข้อความมาประยุกต์ใช้เพื่อวิเคราะห์ความคิดเห็นและทัศนคติของนักท่องเที่ยวที่มีต่อประสบการณ์การรับประทานอาหารท้องถิ่น

กระบวนการทำเหมืองข้อความประกอบด้วยขั้นตอนหลักๆ ได้แก่ การเตรียมข้อมูล (Data Preparation) การประมวลผลเบื้องต้น (Pre-processing) การแปลงข้อความเป็นรูปแบบที่เหมาะสม (Text Transformation) และการวิเคราะห์เพื่อสกัดความรู้ (Knowledge Extraction) [10] ในขั้นตอนการประมวลผลเบื้องต้น จะมีการทำความสะอาดข้อมูล (Data Cleaning) ซึ่งรวมถึงการกำจัดคำหยุด (Stop Words) การแบ่งคำ (Tokenization) และการทำลดรูปคำ (Stemming) ให้เหลือเพียงรากศัพท์

การวิเคราะห์เนื้อหาออนไลน์ โดยเฉพาะจากคำบรรยายยูทูปเกี่ยวกับอาหารไทย เป็นแหล่งข้อมูลสำคัญที่สะท้อนมุมมองของชาวต่างประเทศได้อย่างเป็นธรรมชาติ เนื่องจากผู้ใช้งานยูทูปมักแสดงความคิดเห็นและความรู้สึกโดยไม่ถูกกำหนดกรอบของคำถาม ซึ่งต่างจากการเก็บข้อมูลด้วยแบบสอบถามหรือการสัมภาษณ์ [5] การศึกษาวิจัยของ Ibrahim Cifci และคณะ [11] ได้ใช้การวิเคราะห์ความคิดเห็นออนไลน์เพื่อทำความเข้าใจทัศนคติของนักท่องเที่ยวที่มีต่ออาหารริมทางในกรุงเทพมหานคร และพบว่าการวิเคราะห์ความคิดเห็นออนไลน์เป็นวิธีที่มีประสิทธิภาพในการค้นหาปัจจัยที่ส่งผลต่อการตัดสินใจและความพึงพอใจของนักท่องเที่ยว

2.2 ลำดับคำต่อเนื่อง (N-Gram)

ลำดับคำต่อเนื่อง (N-Gram) เป็นรูปแบบลำดับย่อยของคำหรือตัวอักษรที่ต่อเนื่องกันจำนวน N คำหรือตัวอักษร โดยถูกสกัดจากข้อความที่มีความยาวมากกว่า [8] แนวคิดนี้เป็นพื้นฐานสำคัญในการประมวลผลภาษาธรรมชาติและการทำเหมืองข้อความ เนื่องจากช่วยในการจับความสัมพันธ์ของคำที่ปรากฏร่วมกัน (Co-occurrence) ซึ่งมีความสำคัญต่อการเข้าใจความหมายและบริบทของข้อความ

การทำ N-Gram เช่น ประโยค “I like Thai food” สามารถแบ่งได้เป็นตามจำนวนคำ ได้ดังนี้

1) Unigram คือคำเดี่ยวหนึ่งคำ จะถูกสร้างเป็น [“I”, “like”, “Thai”, “food”]

2) Bigram คือคำสองคำที่อยู่ติดกัน จะถูกสร้างเป็น [“I like”, “like Thai”, “Thai food”]

3) Trigram คือคำสามคำที่อยู่ติดกัน จะถูกสร้างเป็น [“I like Thai”, “like Thai food”]

เมื่อรวม N-Gram ทุกระดับเข้าด้วยกัน ประโยค “I like Thai food” จะถูกแปลงเป็น [“I”, “like”, “Thai”, “food”, “I like”, “like Thai”, “Thai food”, “I like Thai”, “like Thai food”]

N-Gram มีประโยชน์อย่างมากในการวิเคราะห์ข้อความ เนื่องจากช่วยให้เห็นความสัมพันธ์ของคำในระดับที่สูงขึ้น ไม่ใช่เพียงแค่คำเดี่ยวๆ ซึ่งอาจไม่สามารถสื่อความหมายได้อย่างสมบูรณ์ ตัวอย่างเช่น คำว่า “hot” เพียงคำเดียวอาจไม่สามารถบ่งบอกได้ว่าหมายถึงอุณหภูมิสูงหรือรสชาติเผ็ด แต่เมื่อพิจารณาเป็น Bigram หรือ Trigram เช่น “hot weather” หรือ “hot spicy food” จะทำให้เข้าใจความหมายได้ชัดเจนขึ้น

2.3 Bag of Words

Bag of Words (BoW) เป็นโมเดลการแทนข้อความ (Text Representation Model) ที่มีความเรียบง่ายแต่มีประสิทธิภาพในการประมวลผลภาษาธรรมชาติ [12] แนวคิดพื้นฐานของ BoW คือการแปลงข้อความให้เป็นเวกเตอร์ของจำนวนครั้งที่คำแต่ละคำปรากฏในเอกสาร โดยไม่คำนึงถึงลำดับของคำหรือไวยากรณ์ภาษา เสมือนกับการใส่คำทั้งหมดลงในถุง (Bag) แล้วนับความถี่ของแต่ละคำ

กระบวนการสร้าง Bag of Words มีขั้นตอนหลัก ดังนี้

1) สร้างคลังศัพท์ (Vocabulary) จากคำที่ปรากฏในชุดเอกสารทั้งหมด

2) สร้างเวกเตอร์สำหรับแต่ละเอกสาร โดยนับความถี่ของแต่ละคำที่ปรากฏในเอกสารนั้น

3) สร้างเมทริกซ์คำ-เอกสาร (Term-Document Matrix) จากเวกเตอร์ของทุกเอกสาร

ตัวอย่างเช่น หากมีประโยคสองประโยคคือ “I like Thai food” และ “Thai food is spicy” คลังศัพท์จะประกอบด้วยคำทั้งหมด 6 คำ คือ [“I”, “like”, “Thai”, “food”, “is”, “spicy”] และเมทริกซ์คำ-เอกสารจะได้ดังนี้

1) เอกสาร 1 เท่ากับ [1, 1, 1, 1, 0, 0]

2) เอกสาร 2 เท่ากับ [0, 0, 1, 1, 1, 1]

ข้อดีของ Bag of Words คือความเรียบง่ายในการใช้งานและการคำนวณ ทำให้เหมาะกับการวิเคราะห์ข้อความขนาดใหญ่ อย่างไรก็ตาม ข้อจำกัดหลักคือการสูญเสียข้อมูลเกี่ยวกับลำดับคำและความสัมพันธ์ระหว่างคำ ซึ่งอาจส่งผลต่อความแม่นยำในการวิเคราะห์บางกรณี เพื่อปรับปรุงข้อจำกัดนี้ การใช้เทคนิค N-Gram ร่วมกับการใช้วิธีการถ่วงน้ำหนักคำ เช่น TF-IDF (Term Frequency-Inverse Document Frequency) [8] ซึ่งให้ความสำคัญกับคำที่ปรากฏบ่อยในเอกสารหนึ่งแต่ไม่ปรากฏบ่อยในเอกสารอื่นๆ ทำให้สามารถระบุคำที่มีความสำคัญและเป็นลักษณะเฉพาะของแต่ละเอกสารได้ดียิ่งขึ้น

2.4 เทคนิค Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) เป็นแบบจำลองการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) [6,13] ใช้ในการค้นหาโครงสร้างหัวข้อ (Topic Structure) ที่แฝงอยู่ในชุดเอกสารขนาดใหญ่ LDA เป็นแบบจำลองการสร้าง (Generative Model) ที่อยู่บนพื้นฐานของความน่าจะเป็นแบบเบย์ (Bayesian Probability) ซึ่งมีแนวคิดว่าเอกสารแต่ละฉบับประกอบด้วยหัวข้อหลายหัวข้อในสัดส่วนที่แตกต่างกัน และแต่ละหัวข้อประกอบด้วยคำหลายคำในสัดส่วนที่แตกต่างกันเช่นกัน

แนวคิดหลักของ LDA [14] คือการมองว่ากระบวนการสร้างเอกสารมีขั้นตอนดังนี้

1) เลือกการกระจายของหัวข้อสำหรับเอกสาร (Document-Topic Distribution) จากการกระจายแบบ Dirichlet Distribution

2) สำหรับแต่ละคำในเอกสารจะดำเนินการเลือกหัวข้อจากการกระจายของหัวข้อที่ได้จากขั้นตอนที่ 1 และ เลือกคำจากการกระจายของคำในหัวข้อที่เลือก

จุดเด่นของ LDA คือความสามารถในการค้นหาหัวข้อที่แฝงอยู่ในข้อมูลโดยอัตโนมัติ โดยไม่จำเป็นต้องมีการกำหนดหัวข้อล่วงหน้า ทำให้เหมาะกับการวิเคราะห์ข้อมูลที่มีขนาดใหญ่และมีความซับซ้อน อย่างไรก็ตาม การเลือกจำนวนหัวข้อที่เหมาะสมเป็นความท้าทายสำคัญในการใช้งาน LDA ซึ่งนิยมใช้เกณฑ์ความสอดคล้อง (Coherence) ในการประเมินคุณภาพของหัวข้อที่ได้จากแบบจำลอง

2.5 เทคนิค Non-negative Matrix Factorization

Non-negative Matrix Factorization (NMF) เป็นเทคนิคการลดมิติข้อมูล (Dimensionality Reduction) ที่ได้รับความนิยมในการวิเคราะห์ข้อมูลที่ถูกแปลงเป็นตัวเลขแบบไม่มีค่าติดลบ เช่น ความถี่ของคำในเอกสาร หรือความเข้มของพิกเซลในรูปภาพ การทำ Topic Modeling หรือการจัดกลุ่มหัวข้อของข้อมูลเอกสารจำนวนมาก [7,13] หลักการของ NMF ทำงานโดยแยกเมทริกซ์ข้อมูลขนาดใหญ่ออกเป็นเมทริกซ์ย่อยที่มีค่าไม่เป็นลบ ซึ่งช่วยให้สามารถค้นพบโครงสร้างหัวข้อที่ซ่อนอยู่ในชุดข้อมูลได้ ในการประยุกต์ใช้กับ Topic Modeling เมทริกซ์ข้อมูลตั้งต้นคือเมทริกซ์คำ-เอกสาร (Term-Document Matrix) ที่แสดงความถี่ของคำในแต่ละเอกสาร เทคนิค NMF จะแยกเมทริกซ์นี้ออกเป็นสองเมทริกซ์ที่มีค่าไม่เป็นลบ ได้แก่ เมทริกซ์คำ-หัวข้อ (W) ที่แสดงความสัมพันธ์ระหว่างคำและหัวข้อ และเมทริกซ์หัวข้อ-เอกสาร (H) ที่แสดงความสัมพันธ์

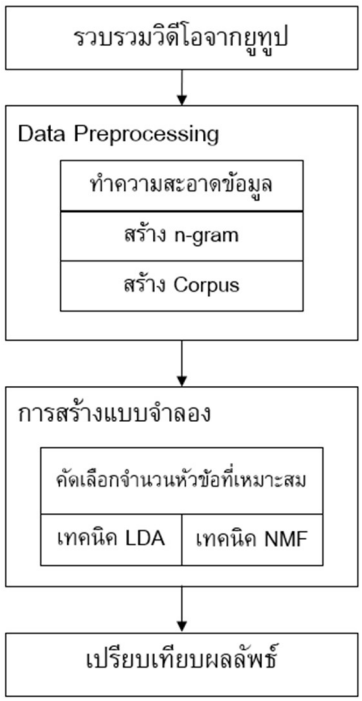
ระหว่างหัวข้อและเอกสาร โดยค่าสูงในแต่ละเมทริกซ์บ่งชี้ถึงความสำคัญที่มากขึ้น

NMF มีข้อดีหลายประการสำหรับการทำ Topic Modeling [14] เริ่มจากการแสดงผลแบบบวก (Additive Representation) ที่สอดคล้องกับลักษณะทางธรรมชาติของข้อความซึ่งเป็นการรวมกันของคำต่างๆ ประกอบกับคุณสมบัติความบางแน่น (Sparsity) ที่ทำให้แต่ละหัวข้อประกอบด้วยคำจำนวนไม่มากที่เกี่ยวข้องกันอย่างชัดเจน ส่งผลให้การตีความหัวข้อทำได้ง่าย นอกจากนี้ เมื่อมีการกำหนดเงื่อนไขอย่างเหมาะสม NMF สามารถสร้างการแสดงแบบแยกส่วน (Parts-based Representation) ของข้อมูล ทำให้เห็นโครงสร้างย่อยของหัวข้อได้ชัดเจน และด้วยข้อจำกัดที่ค่าต้องไม่เป็นลบ จึงทำให้ NMF มีความเหมาะสมกับข้อมูลที่มีค่าไม่เป็นลบอยู่แล้ว เช่น ความถี่ของคำในเอกสาร

3. วิธีการศึกษา

การศึกษาวิจัยนี้ได้วางกรอบแนวคิดให้สอดรับกับจุดมุ่งหมายของการวิจัย โดยใช้กระบวนการทางวิทยาศาสตร์ข้อมูลเป็นแนวทาง ซึ่งครอบคลุมขั้นตอนต่างๆ ได้แก่ การวิเคราะห์โจทย์ปัญหา การเก็บรวบรวมข้อมูล การจัดเตรียมข้อมูลให้พร้อมสำหรับการวิเคราะห์ การพัฒนาแบบจำลอง การทดสอบประสิทธิภาพของแบบจำลอง และการประยุกต์ใช้แบบจำลอง โดยกรอบแนวคิดของงานวิจัยนี้ได้แสดงไว้ในรูปที่ 1

การวิจัยนี้ใช้สภาพแวดล้อมการพัฒนาและเครื่องมือดังนี้ (1) Python เวอร์ชัน 3.11 (2) Microsoft Excel 2019 สำหรับการจัดเตรียมข้อมูลต้นฉบับ (3) Google Colaboratory (Colab) สำหรับการประมวลผล ไลบรารีหลักที่ใช้งานได้แก่ (1) numpy เวอร์ชัน 1.24.2 (2) pandas เวอร์ชัน 2.0.0 (3) scikit-learn เวอร์ชัน 1.3.2 (4) gensim เวอร์ชัน 4.3.3 (5) NLTK เวอร์ชัน 3.9.1



รูปที่ 1 กรอบแนวคิดงานวิจัย

3.1 การรวบรวมข้อมูล

การรวบรวมข้อมูลสำหรับงานวิจัยนี้มุ่งเน้นการศึกษามุมมองของชาวต่างประเทศที่มีต่ออาหารไทย โดยผู้วิจัยได้ดำเนินการสืบค้นและคัดเลือกวิดีโอจากเว็บไซต์ยูทูปด้วยคำค้น "Thai food" ซึ่งมีเกณฑ์การคัดเลือกที่สำคัญ ด้านผู้สร้างเนื้อหา (Content Creator) ได้แก่ วิดีโอจะต้องสร้างโดยชาวต่างประเทศ หรือมีชาวต่างประเทศเป็นผู้พูดหลัก หรือเป็นวิดีโอที่สร้างโดยคนไทยหรือร้านอาหารไทยที่ตั้งอยู่ในต่างประเทศ เกณฑ์ด้านเนื้อหา (Content Focus) ได้แก่ วิดีโอที่มีเนื้อหาหลักเกี่ยวกับการรีวิวอาหารไทย หรือ การปรุงอาหารไทยโดยชาวต่างประเทศ หรือประสบการณ์การรับประทานอาหารไทยในประเทศไทยหรือต่างประเทศ เกณฑ์ด้านภาษา (Language Criteria) ได้แก่ วิดีโอที่มีการพูดเป็นภาษาอังกฤษเป็นหลัก เกณฑ์ด้านความยาวของวิดีโอ (Duration Criteria) ได้แก่ ความยาวขั้นต่ำ 2 นาที เพื่อให้มีเนื้อหาเพียงพอต่อการวิเคราะห์ ความยาวสูงสุดไม่เกิน 60 นาที เพื่อหลีกเลี่ยงเนื้อหาที่ไม่เกี่ยวข้องมากเกินไป

การคัดเลือกวิดีโอได้พิจารณาเฉพาะวิดีโอที่สามารถทำสำเนาคำบรรยายจากบริการของยูทูปได้ จากการรวบรวมทั้งหมดได้วิดีโอที่ตรงตามเกณฑ์จำนวน 352 วิดีโอ ระหว่างปี

พ.ศ. 2553 ถึง 2567 ข้อมูลทั้งหมดถูกจัดเก็บอย่างเป็นระบบในรูปแบบตารางโดยใช้โปรแกรม Microsoft Excel ประกอบด้วยข้อมูล 5 ส่วนสำคัญ ได้แก่ แหล่งที่อยู่ของวิดีโอ (URL) ข้อความคำบรรยายภาษาอังกฤษ (Script) วันที่โพสต์วิดีโอ (Postdate) จำนวนยอดวิว (Views) และชื่อช่องวิดีโอ (Channel)

3.2 การเตรียมข้อมูลก่อนการประมวลผล

ข้อมูลที่ได้รวบรวมจำนวน 352 รายการ จะได้รับการตรวจสอบและทำความสะอาดข้อมูลให้เหมาะสมก่อนนำไปประมวลผลสร้างแบบจำลอง ตามขั้นตอนดังนี้

3.2.1 การจัดการตัวอักขระพิเศษและข้อมูลรบกวน (Noise Data)

ขั้นตอนนี้ดำเนินการ (1) ลบคำที่ไม่เกี่ยวข้อง เช่น audio-related tags: [Music] [Applause] [Song] (2) ตัดเครื่องหมายวรรคตอน เช่น - : ; "" รวมทั้งตัวเลขและสัญลักษณ์พิเศษ (3) กำจัดตัวอักษรซ้ำที่ไม่มีความหมาย เช่น aa, bb, cc

3.2.2 การปรับข้อความให้รูปแบบเป็นมาตรฐาน (Text Normalization)

ขั้นตอนนี้ดำเนินการ (1) แปลงทุกตัวอักษรเป็นตัวพิมพ์เล็ก (2) แบ่งข้อความเป็นคำ (Tokenization) โดยใช้ชุดคำสั่งไลบรารี NLTK [15]

3.2.3 การกรองคำเชื่อม (Stop Words)

ขั้นตอนนี้ดำเนินการกำจัด English stop words มาตรฐาน เช่น 'the', 'and', 'to'

ข้อมูลคำบรรยายแต่ละชุดจะถูกนำไปสร้าง N-Gram เพื่อสร้างกลุ่มคำใหม่จากการรวมคำอยู่ติดกันเพื่อให้คงความหมายของคำหรือประโยคนั้น โดยสร้างทั้ง unigrams (คำเดี่ยว) bigrams (คำต่อเนื่อง 2 คำ) และ trigrams (คำต่อเนื่อง 3 คำ) และแต่ละกลุ่มคำจะถูกคั่นด้วยเครื่องหมายคอมม่า และบันทึกเป็นไฟล์ CSV ต้นฉบับพร้อมใช้งาน

การสร้างคลังคำและการแปลงข้อมูลผู้วิจัยใช้ไลบรารี gensim [16] ในการสร้างพจนานุกรมคำ (Dictionary) เพื่อกำหนดรหัสให้กับคำแต่ละคำ และแปลงข้อความให้อยู่ในรูปแบบ document-term matrix โดยใช้ฟังก์ชัน doc2bow ซึ่งจะเก็บค่าความถี่ของคำในแต่ละเอกสาร เพื่อให้การวิเคราะห์มีความแม่นยำมากขึ้น ผู้วิจัยได้ทำการปรับค่าความถี่ของคำให้อยู่ในช่วง 0 ถึง 1 โดยการหารค่าความถี่ของแต่ละคำด้วยค่าความถี่สูงสุดที่พบในเอกสารนั้นๆ (Max normalization) เพื่อช่วยลดผลกระทบจากความแตกต่างของความยาวเอกสาร

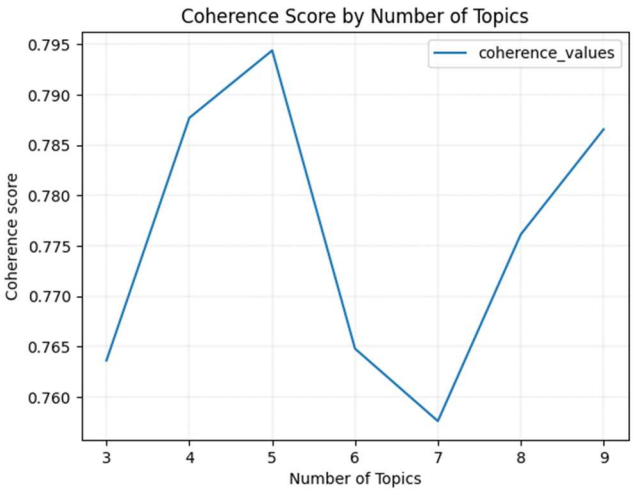
การประมวลผลข้อความจากวิดีโอทั้ง 352 รายการส่งผลให้ได้คลังข้อมูล (corpus) ประกอบด้วยคำทั้งหมด 358,738 คำ หลังจากผ่านกระบวนการทำ N-gram และการทำความสะอาดข้อมูล ข้อมูลถูกแปลงเป็น normalized corpus ที่มีค่าความถี่อยู่ในช่วง 0 ถึง 1 เพื่อเตรียมความพร้อมสำหรับการวิเคราะห์ด้วยเทคนิค LDA และ NMF ในขั้นตอนถัดไป

3.3 การสร้างแบบจำลอง

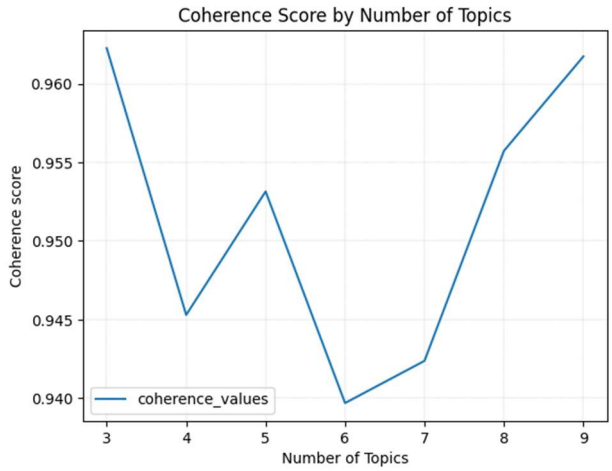
การวิเคราะห์เพื่อกำหนดจำนวนหัวข้อที่เหมาะสมสำหรับเทคนิค LDA และ NMF ดำเนินการโดยพิจารณาจากสองปัจจัยหลัก ได้แก่ ค่าคะแนนความสอดคล้อง (Coherence score) และความเหมาะสมในการแปลความหมายของหัวข้อ

ผู้วิจัยได้ใช้ไฟล์ข้อมูลที่เตรียมความพร้อมจากขั้นตอนก่อนหน้าและทดลองแบ่งหัวข้อตามหลักการของ Zhang และคณะ [3] ที่เสนอให้เริ่มทดลองกับจำนวนหัวข้อที่มีจำนวนน้อยและพิจารณาค่าคะแนนความสอดคล้อง [17] ที่เหมาะสมร่วมด้วย เพื่อหลีกเลี่ยงการสร้างหัวข้อที่มีความหมายกว้างเกินไปจากกรณีที่จำนวนหัวข้อน้อย) และป้องกันความซ้ำซ้อน การตีความยาก (Topic mixing) จากกรณีที่จำนวนหัวข้อมากเกินไป

สำหรับเทคนิค LDA ผู้วิจัยได้ทดลองแบ่งจำนวนหัวข้อตั้งแต่ 3 ถึง 9 หัวข้อ และวัดประสิทธิภาพด้วยค่า Coherence score (c_v) ดังแสดงในรูปที่ 2 ผลการวิเคราะห์พบว่าการแบ่งเป็น 5 หัวข้อให้ค่าคะแนนความสอดคล้องสูงที่สุดเท่ากับ 0.794 และมีความเหมาะสมในการแปลความหมาย กล่าวคือแต่ละหัวข้อมีความแตกต่างกันอย่างชัดเจนและสามารถสื่อความหมายได้ครอบคลุมประเด็นสำคัญเกี่ยวกับอาหารไทย ด้วยเหตุนี้ ผู้วิจัยจึงกำหนดให้แบบจำลอง LDA สุดท้ายมีจำนวน 5 หัวข้อ โดยใช้พารามิเตอร์ที่สำคัญ ได้แก่ การกรองคำที่มีความถี่ต่ำและสูงเกินไป (no_below=100, no_above=0.5) กำหนดจำนวนรอบการเรียนรู้ (passes=10, iterations=100) และกำหนดค่าเริ่มต้นสุ่ม (random_state=42) เพื่อความสม่ำเสมอของผลลัพธ์



รูปที่ 2 เปรียบเทียบค่าคะแนนความสอดคล้องด้วยเทคนิค LDA



รูปที่ 3 เปรียบเทียบค่าคะแนนความสอดคล้องด้วยเทคนิค NMF

ในส่วนของเทคนิค NMF ดังรูปที่ 3 แม้ว่าการแบ่งเป็น 3 หัวข้อจะให้ค่าคะแนนความสอดคล้องเท่ากับ 0.962 มีค่าสูงที่สุด แต่จำนวนหัวข้อที่น้อยเกินไปส่งผลให้แต่ละหัวข้อมีความหมายกว้างและไม่เฉพาะเจาะจง ในขณะที่การแบ่งเป็น 9 หัวข้อซึ่งมีค่าคะแนนความสอดคล้องเท่ากับ 0.961 ซึ่งสูงเป็นลำดับที่สอง ก็ไม่เหมาะสมเนื่องจากเกิดความซ้ำซ้อนของคำสำคัญระหว่างหัวข้อ ทำให้การตีความหมายมีความคลุมเครือ ดังนั้นผู้วิจัยจึงเลือกใช้จำนวน 8 หัวข้อที่มีค่าคะแนนความสอดคล้องเท่ากับ 0.956 สำหรับการวิเคราะห์ด้วยเทคนิค NMF เพื่อให้ได้การจำแนกหัวข้อที่มีความสมดุลระหว่างความละเอียดของการแบ่งกลุ่มและความชัดเจนในการแปลความหมาย โดยใช้พารามิเตอร์ที่สำคัญ ได้แก่ random_state=42 และประมวลผลกับเมทริกซ์ TF-IDF ที่สร้างจากคลังข้อมูลเดียวกันกับที่ใช้ในแบบจำลอง LDA โดยกำหนดค่า max_df=0.95 และ min_df=2 เพื่อกรองคำที่มีความถี่สูงหรือต่ำเกินไป

4. ผลการศึกษา

การวิเคราะห์ประเด็นอาหารไทยจากคำบรรยาย Youtube ด้วยเทคนิค LDA และ NMF ได้ผลการศึกษาดังนี้

4.1 สำรวจข้อมูล (Exploratory Data Analysis : EDA)

การสำรวจข้อมูลจากวิดีโอยูทูปตามเกณฑ์การคัดเลือก ได้วิดีโอจำนวน 352 รายการ ซึ่งช่องยูทูป 5 อันดับแรกที่มีการรวบรวมข้อมูลมากที่สุด ประกอบด้วย ช่อง Mark Wiens (18 วิดีโอ) ช่อง ร้านเด็ด อเมริกา - Landed America (16 วิดีโอ) ช่อง MOSSALA101 (7 วิดีโอ) ช่อง Mickey Stotch (7 วิดีโอ) และ ช่อง DancingBacons (6 วิดีโอ) จำนวนวิดีโอที่รวบรวมในแต่ละปีมีดังนี้

ตารางที่ 1 การจำแนกจำนวนวิดีโอตามปี

ปี พ.ศ.	จำนวนวิดีโอ
2567	192
2566	90
2565	39
2564	10
2563	3
2562	4
2561	6
2560	3
2559	3
2557	1
2553	1

การแสดงให้เห็นถึงมุมมองของชาวต่างประเทศที่มีต่ออาหารไทย โดยคำที่ปรากฏบ่อยที่สุดในคำบรรยายได้แสดงในแผนภาพเมฆกลุ่มคำ (Word Cloud) ไว้ในรูปที่ 4

รูปที่ 4 คำที่ปรากฏบ่อยที่สุดในคำบรรยาย

ผลการวิเคราะห์พบว่าคำที่มีความถี่สูงสุด 5 อันดับแรกเป็นคำที่เกี่ยวข้องกับการบริโภคและคุณลักษณะทั่วไปของอาหาร ได้แก่ “food” (3,277 ครั้ง) “eat” (2,497 ครั้ง) “try” (1,767 ครั้ง) “delicious” (1,752 ครั้ง) และ “rice” (1,623 ครั้ง) ซึ่งสะท้อนให้เห็นว่าผู้บริโภคให้ความสำคัญกับประสบการณ์การรับประทานและรสชาติอาหารเป็นหลักในส่วนของวัตถุดิบพื้นฐานที่เป็นเอกลักษณ์ของอาหารไทย พบคำที่มีความถี่สูง ได้แก่ “rice” (1,623 ครั้ง) “sauce” (1,558 ครั้ง) “chicken” (1,365 ครั้ง) “spicy” (1,337 ครั้ง) และ “fish” (1,255 ครั้ง) ซึ่งแสดงให้เห็นถึงองค์ประกอบสำคัญในอาหารไทยที่เป็นที่รู้จักในวงกว้าง

นอกจากนี้ ข้อมูลยังแสดงให้เห็นถึงประเภทของอาหารไทยที่ได้รับความนิยม โดยมีคำว่า “curry” (1,057 ครั้ง) “fried” (991 ครั้ง) “sweet” (989 ครั้ง) และ “noodles” (972 ครั้ง) ปรากฏในความถี่สูง สะท้อนให้เห็นถึงรูปแบบการปรุงอาหารและรสชาติหลักที่เป็นเอกลักษณ์ของอาหารไทย

ผลการวิเคราะห์นี้แสดงให้เห็นว่าการรับรู้เกี่ยวกับอาหารไทยในกลุ่มผู้บริโภคมุ่งเน้นที่รสชาติอันโดดเด่น วัตถุดิบพื้นฐาน และประสบการณ์การรับประทานที่หลากหลาย

4.2 แบบจำลองด้วยเทคนิค LDA

ผลการวิเคราะห์ด้วยเทคนิค LDA ได้จำนวนหัวข้อเท่ากับ 5 ซึ่งให้ค่าคะแนนความสอดคล้องสูงที่สุด สามารถจำแนกประเด็นอาหารไทยออกเป็น 5 หัวข้อหลักดังแสดงในตารางที่ 2

ตารางที่ 2 การจำแนกหัวข้อและคำสำคัญจากแบบจำลอง LDA

หัวข้อ	คำสำคัญ	การตีความ
1	market, bangkok, meat, sticky, noodles, mango, sticky rice, flavor, fresh, sugar, coconut, egg, shrimp, chili, crispy, street, night, soup, seafood, sour	วัตถุดิบและรสชาติอาหาร
2	water, feel, hot, tell, sticky, need, fun, mango, together, bangkok, sticky rice, noodles, probably, city, coconut, cool, better, walk, help, pretty	ประสบการณ์การท่องเที่ยวและอาหาร
3	curry, dish, flavor, chili, meat, soup, noodles, shrimp, dishes, crispy, egg, basil, green, fresh, coconut, garlic, famous, milk, sour, need	องค์ประกอบอาหารไทย
4	street, street food, bangkok, market, crispy, grilled, noodles, amazing, seafood, fresh, egg, popular, restaurants, coconut, night, tell, famous, style, great, curry	อาหารริมทางและความนิยม
5	cook, curry, ingredients, green, stuff, shrimp, pretty, price, chili, making, need, fish sauce, night, morning, fresh, though, help, coconut, vegetables, together	กระบวนการปรุงอาหารและวัตถุดิบ

ผลการวิเคราะห์แสดงให้เห็นว่า LDA สามารถจำแนกหัวข้อที่มีความหมายเชิงบริบท โดยหัวข้อที่ 1 และ 3 เกี่ยวข้องกับองค์ประกอบของอาหารไทย ในขณะที่หัวข้อที่ 2 เน้นประสบการณ์และบรรยากาศในการรับประทาน หัวข้อที่ 4 เน้นสถานที่และวัฒนธรรมอาหารริมทาง และหัวข้อที่ 5 เน้นกระบวนการปรุงและส่วนประกอบของอาหาร

สังเกตได้ว่ามีความซ้อนทับของคำสำคัญระหว่างหัวข้อ เช่น "coconut" ปรากฏในทุกหัวข้อ และ "curry" ปรากฏในหลายหัวข้อ สะท้อนลักษณะของ LDA ที่อนุญาตให้คำหนึ่งๆ สามารถปรากฏในหลายหัวข้อได้ การกระจายตัวนี้แสดงให้เห็น

ถึงความเชื่อมโยงระหว่างวัตถุดิบหลัก รสชาติ และบริบทการรับประทานในมุมมองของชาวต่างประเทศ

4.3 แบบจำลองด้วยเทคนิค NMF

ผลการวิเคราะห์ด้วยเทคนิค NMF ได้จำนวนหัวข้อเท่ากับ 8 เป็นจำนวนที่เหมาะสมที่สุดสำหรับการตีความ สามารถจำแนกประเด็นอาหารไทยได้ดังแสดงในตารางที่ 3

ตารางที่ 3 การจำแนกหัวข้อและคำสำคัญจากแบบจำลอง NMF

หัวข้อ	คำสำคัญ	การตีความ
1	fish, sauce, dish, flavor, chili, garlic, pork, fresh, shrimp, seafood	วัตถุดิบและรสชาติอาหารทะเล
2	music, spicy, egg, pretty, rice, love, know, best, restaurants, flavor	ความชื่นชอบรสชาติ
3	street, bangkok, food, market, fried, crispy, night, sweet, coconut, amazing	วัฒนธรรมอาหารริมทาง
4	curry, green, coconut, ingredients, rice, red, milk, vegetables, making, dishes	ส่วนประกอบของแกงไทย
5	delicious, eat, spicy, try, taste, rice, pork, hot, soup, looks	ประสบการณ์การทานและรสชาติอาหาร
6	think, know, look, eat, try, place, water, looks, market, tell	การสำรวจอาหาร
7	chicken, rice, fried, pork, spicy, noodles, sticky, mango, sweet, crispy	อาหารประเภทข้าวและเนื้อสัตว์
8	food, restaurant, eat, famous, restaurants, cook, price, help, best, popular	ประสบการณ์ร้านอาหารและการบริการ

ผลการวิเคราะห์แสดงให้เห็นว่า NMF สามารถแยกหัวข้อที่มีความเฉพาะเจาะจงมากขึ้น โดยเฉพาะด้านประเภทอาหารและวัตถุดิบ เช่น แยกอาหารทะเล (หัวข้อ 1) แกงไทย (หัวข้อ 4) และอาหารประเภทข้าวและเนื้อสัตว์ (หัวข้อ 7) ออกจากกัน นอกจากนี้ ยังแยกแยะประสบการณ์การรับประทานออกเป็นหลายมิติ ได้แก่ ความชื่นชอบประทับใจด้านรสชาติ (หัวข้อ 2) ประสบการณ์การทานและรสชาติอาหาร (หัวข้อ 5) การค้นพบและสำรวจอาหาร (หัวข้อ 6) และประสบการณ์ร้านอาหารและการบริการ (หัวข้อ 8)

การวิเคราะห์น้ำหนักของคำในแต่ละหัวข้อพบว่า NMF มีแนวโน้มที่จะกำหนดคำสำคัญให้อยู่ในหัวข้อที่เฉพาะเจาะจงมากกว่า LDA ซึ่งสะท้อนลักษณะของ NMF ที่พยายามแยกองค์ประกอบออกจากกันอย่างชัดเจน

4.4 เปรียบเทียบผลลัพธ์จากแบบจำลอง

การเปรียบเทียบผลลัพธ์ระหว่างแบบจำลอง LDA และ NMF แสดงให้เห็นถึงความแตกต่างสำคัญในการจำแนกประเด็นเกี่ยวกับอาหารไทยในมุมมองของชาวต่างประเทศ ตารางที่ 4 สรุปความแตกต่างหลักระหว่างผลลัพธ์ของทั้งสองเทคนิค

ตารางที่ 4 การเปรียบเทียบผลลัพธ์การวิเคราะห์หัวข้อระหว่าง LDA และ NMF

คุณลักษณะ	LDA	NMF
จำนวนหัวข้อที่เหมาะสม	5 หัวข้อ	8 หัวข้อ
ลักษณะการแบ่งหัวข้อ	ครอบคลุมประเด็นกว้าง เช่น "องค์ประกอบของอาหารไทย" "อาหารริมทางและความนิยม"	เฉพาะเจาะจง เช่น "องค์ประกอบของแกงไทย" "อาหารหลักประเภทข้าวและโปรตีน"
การซ้อนทับของคำสำคัญ	คำสำคัญปรากฏร่วมกันในหลายหัวข้อ เช่น "coconut" พบในทุกหัวข้อ	คำสำคัญมักปรากฏเฉพาะในหัวข้อที่เกี่ยวข้องโดยตรง
ค่าน้ำหนักคำ	กระจายตัวค่อนข้างสม่ำเสมอระหว่างหัวข้อ	เกิดการกระจุกตัวในหัวข้อที่เฉพาะเจาะจง
การจัดกลุ่มประเด็น	จัดกลุ่มตามบริบทการใช้งาน เช่น "วัตถุดิบและรสชาติอาหาร"	จัดกลุ่มตามคุณลักษณะ เช่น แยกระหว่าง "วัตถุดิบและรสชาติอาหารทะเล" และ "องค์ประกอบของแกงไทย"
การตีความหัวข้อ	สะดวกต่อการมองภาพรวมและความเชื่อมโยง	เหมาะสำหรับการวิเคราะห์เชิงลึกในแต่ละประเด็น

ผลการเปรียบเทียบในตารางที่ 4 แสดงให้เห็นถึงความแตกต่างที่เป็นผลมาจากหลักการพื้นฐานของแต่ละเทคนิค โดย LDA เป็นแบบจำลองความน่าจะเป็นที่อนุญาตให้หนึ่งเอกสารมีหลายหัวข้อและหนึ่งคำสามารถปรากฏในหลายหัวข้อได้ ในขณะที่ NMF เป็นเทคนิคการแยกเมทริกซ์ที่มุ่งแยกองค์ประกอบออกจากกันอย่างชัดเจน

5. อภิปรายผลการศึกษา

การศึกษามุมมองของชาวต่างประเทศที่มีต่ออาหารไทยและเปรียบเทียบประสิทธิภาพของเทคนิค LDA และ NMF ในการจำแนกประเด็น ผลการวิจัยมีประเด็นสำคัญที่นำมาอภิปรายดังนี้

5.1 มุมมองของชาวต่างประเทศต่ออาหารไทย

ผลการวิเคราะห์ความถี่ของคำและการจัดกลุ่มหัวข้อทั้งจากเทคนิค LDA และ NMF สะท้อนให้เห็นว่าชาวต่างประเทศมองอาหารไทยในมิติที่หลากหลาย โดยให้ความสำคัญกับองค์ประกอบหลักสี่ประการ ได้แก่ รสชาติอันเป็นเอกลักษณ์ วัตถุดิบเฉพาะ ประสบการณ์การรับประทาน และบริบทแวดล้อม

ประการแรก รสชาติอันเป็นเอกลักษณ์ของอาหารไทยเป็นจุดเด่นที่ได้รับการกล่าวถึงอย่างมาก โดยคำว่า "spicy" (เผ็ด) "sweet" (หวาน) และ "sour" (เปรี้ยว) ปรากฏในความถี่สูงและอยู่ในหัวข้อสำคัญของทั้งสองเทคนิค สะท้อนให้เห็นว่าชาวต่างประเทศรับรู้และจดจำรสชาติแบบเฉพาะตัวของอาหารไทยที่มีการผสมผสานรสชาติอย่างลงตัว สอดคล้องกับงานวิจัยของ Promsivapallop และ Kannaovakun [18] ที่พบว่ารสชาติเป็นปัจจัยหลักที่ทำให้นักท่องเที่ยวเลือกรับประทานอาหารไทย

ประการที่สอง วัตถุดิบหลักและส่วนประกอบของอาหารไทยได้รับความสนใจอย่างมาก โดยเฉพาะ "rice" (ข้าว) "chicken" (ไก่) "pork" (หมู) "fish" (ปลา) และ "seafood" (อาหารทะเล) ซึ่งแสดงให้เห็นว่าชาวต่างประเทศให้ความสำคัญกับวัตถุดิบพื้นฐานที่ใช้ในการปรุงอาหารไทย และสามารถระบุถึงเอกลักษณ์ของวัตถุดิบที่นิยมใช้ในครัวไทย ผลการวิจัยนี้สอดคล้องกับการศึกษาของ เณรัญชรา กิจวิกรานต์ [4] ที่พบว่าวัตถุดิบประจำถิ่นเป็นองค์ประกอบสำคัญในการสร้างเอกลักษณ์ให้กับอาหารไทยในสายตาชาวต่างประเทศ

ประการที่สาม ประสบการณ์การรับประทานอาหารไทยเป็นมิติที่ได้รับการกล่าวถึงอย่างมีนัยสำคัญ คำว่า "eat" (ทาน) "try" (ลอง) และ "delicious" (อร่อย) มีความถี่สูงและปรากฏในหลายหัวข้อ แสดงให้เห็นว่าชาวต่างประเทศมองการรับประทานอาหารไทยเป็นประสบการณ์ที่น่าค้นหาและทดลอง ทั้งนี้ การที่ NMF สามารถแยกประสบการณ์การรับประทานออกเป็นหลายมิติ เช่น "ความชื่นชอบรสชาติ" "การค้นพบและสำรวจอาหาร" และ "ประสบการณ์ร้านอาหาร" แสดงให้เห็นถึงความลึกซึ้งของการรับรู้ในมิตินี้

ประการที่สี่ บริบทแวดล้อมของการรับประทานอาหารไทย เช่น "street" (ถนน) "market" (ตลาด) "bangkok" (กรุงเทพฯ) และ "restaurant" (ร้านอาหาร) มีการกล่าวถึงอย่างมีนัยสำคัญ สะท้อนให้เห็นว่าชาวต่างประเทศมองอาหารไทยไม่ได้แยกขาดจากบริบทแวดล้อมทางวัฒนธรรมและสถานที่ โดยเฉพาะวัฒนธรรมอาหารริมทางซึ่งเป็นหัวข้อที่ถูกจัดกลุ่มชัดเจนในทั้งสองเทคนิค ซึ่งสอดคล้องกับงานวิจัยของ Chavarria และ Phakdee-auksorn [19] ที่พบว่าอาหารริมทางเป็นสิ่งดึงดูดนักท่องเที่ยวที่สำคัญของประเทศไทย

5.2 ประสิทธิภาพของเทคนิค LDA และ NMF

ผลการเปรียบเทียบประสิทธิภาพของเทคนิค LDA และ NMF ในการจำแนกประเด็นอาหารไทยจากคำบรรยายยูทูป แสดงให้เห็นข้อดีและข้อจำกัดที่แตกต่างกันของแต่ละเทคนิค

LDA มีจุดเด่นในการแสดงความเชื่อมโยงระหว่างหัวข้อต่างๆ ผ่านการซ้อนทับของคำสำคัญ ทำให้เห็นภาพรวมความสัมพันธ์ระหว่างองค์ประกอบต่างๆ ของอาหารไทย เช่น การที่คำว่า "coconut" ปรากฏในทุกหัวข้อสะท้อนให้เห็นถึงความสำคัญของวัตถุดิบนี้ที่แทรกซึมอยู่ในทุกมิติของอาหารไทย อย่างไรก็ตาม LDA มีข้อจำกัดในการจัดหัวข้อที่ขาดความเฉพาะเจาะจง ทำให้บางหัวข้อมีความหมายกว้างและครอบคลุมประเด็นที่หลากหลายเกินไป ซึ่งอาจทำให้การตีความหมายเชิงลึกทำได้ยาก

ในทางตรงข้าม NMF แสดงจุดเด่นในการแยกแยะประเด็นย่อยที่มีความเฉพาะเจาะจง ทำให้สามารถเจาะลึกลงไปในแต่ละมิติได้อย่างชัดเจน เช่น การแยก "วัตถุดิบและรสชาติอาหาร

ทะเล" ออกจาก "องค์ประกอบของแกงไทย" และ "อาหารประเภทข้าวและเนื้อสัตว์" อย่างชัดเจน ซึ่งเป็นประโยชน์ต่อการวิเคราะห์เชิงลึกในแต่ละประเด็น อย่างไรก็ตาม ข้อจำกัดของ NMF คือการขาดการเชื่อมโยงระหว่างหัวข้อ ทำให้อาจมองไม่เห็นความสัมพันธ์ระหว่างองค์ประกอบต่างๆ ของอาหารไทยในภาพรวม

ผลการวิจัยนี้สอดคล้องกับงานวิจัยของ Chen และคณะ [14] ที่พบว่า NMF มักให้ผลลัพธ์หัวข้อที่มีความเฉพาะเจาะจงและแยกแยะชัดเจนมากกว่า LDA ในขณะที่ LDA มีแนวโน้มที่จะสร้างหัวข้อที่มีความหมายกว้างและมีการซ้อนทับระหว่างหัวข้อมากกว่า

ประเด็นสำคัญที่พบจากการเปรียบเทียบทั้งสองเทคนิคคือ จำนวนหัวข้อที่เหมาะสมในการวิเคราะห์มีความแตกต่างกัน โดย LDA ให้ผลลัพธ์ที่ดีที่สุดที่ 5 หัวข้อ ในขณะที่ NMF ต้องการ 8 หัวข้อเพื่อให้ได้ผลลัพธ์ที่เหมาะสม สะท้อนให้เห็นถึงธรรมชาติของทั้งสองเทคนิคที่แตกต่างกัน โดย LDA มีความยืดหยุ่นในการจัดกลุ่มคำและประเด็นเข้าด้วยกัน ในขณะที่ NMF ต้องการจำนวนหัวข้อมากกว่าเพื่อแยกแยะประเด็นย่อยได้อย่างชัดเจน

5.3 การประยุกต์ใช้ผลการวิจัย

ผลการวิจัยนี้มีนัยสำคัญต่อการประยุกต์ใช้ในหลายด้าน โดยเฉพาะในด้านการส่งเสริมการท่องเที่ยวเชิงอาหาร การพัฒนาผลิตภัณฑ์อาหารไทยสำหรับตลาดต่างประเทศ และการปรับปรุงการนำเสนอร้านอาหารไทยในต่างประเทศ

ผลลัพธ์จาก NMF ที่แยกแยะประเด็นย่อยอย่างชัดเจน และผลลัพธ์จาก LDA ที่แสดงความเชื่อมโยงระหว่างองค์ประกอบต่างๆ สามารถนำไปประยุกต์ใช้ในการออกแบบประสบการณ์ร้านอาหารแบบองค์รวมที่เชื่อมโยงรสชาติ วัตถุดิบ และบรรยากาศเข้าด้วยกัน เพื่อสร้างประสบการณ์ที่สอดคล้องกับการรับรู้ของชาวต่างประเทศ ซึ่งสอดคล้องกับงานวิจัยของ Prapasawasdi และคณะ [1] ที่พบว่าการสร้างประสบการณ์แบบองค์รวมในด้านความคาดหวัง ทัศนคติ คุณค่าที่รับรู้และบรรทัดฐานทางสังคมมีผลต่อการรับรู้ของนักท่องเที่ยวในการรับประทานอาหารท้องถิ่น

อย่างไรก็ตาม การศึกษานี้มีข้อจำกัดที่สำคัญหลายประการที่ต้องพิจารณาได้แก่ (1) การใช้ข้อมูลเฉพาะภาษาอังกฤษ เฉพาะแพลตฟอร์มยูทูปเพียงแหล่งเดียว อาจไม่สะท้อนความหลากหลายของมุมมองจากชาวต่างประเทศ (2) ขอบเขตช่วงเวลาเก็บข้อมูลระหว่างปี พ.ศ.2553-2567 อาจไม่สามารถจับการเปลี่ยนแปลงของการรับรู้อันเนื่องมาจากผลกระทบจากเหตุการณ์สำคัญต่างๆ เช่น วิกฤตการณ์โควิด-19 หรือนโยบายส่งเสริมอาหารไทยของรัฐบาล ที่อาจส่งผลต่อพฤติกรรมการบริโภคสื่อและการรับรู้อาหารต่างประเทศ (3) การมีอคติทางวัฒนธรรม (Cultural bias) เป็นข้อจำกัดที่สำคัญซึ่งสอดคล้องกับการศึกษาของ Antón และคณะ [20] เนื่องจากผู้ใช้แพลตฟอร์มยูทูปในแต่ละภูมิภาคมีพฤติกรรมการใช้สื่อสังคมออนไลน์ ความแตกต่างทางวัฒนธรรมส่งผลต่อการประเมินประสบการณ์อาหารทำให้ผลการศึกษาอาจมีความเอนเอียงไปทางมุมมองของกลุ่มชาติพันธุ์หรือวัฒนธรรมใดวัฒนธรรมหนึ่ง

6. สรุปผลการศึกษา

ผลการศึกษาพบว่าชาวต่างประเทศมองอาหารไทยในหลากหลายมิติ โดยให้ความสำคัญกับ 4 องค์ประกอบหลัก ได้แก่ รสชาติอันเป็นเอกลักษณ์ วัตถุดิบเฉพาะ ประสบการณ์การรับประทาน และบริบทแวดล้อม รสชาติที่หลากหลายของอาหารไทยโดยเฉพาะความเผ็ด ความหวาน และความเปรี้ยว เป็นจุดเด่นที่ได้รับการกล่าวถึงมากที่สุด ตามด้วยวัตถุดิบพื้นฐานอย่างข้าว ไก่ หมู ปลา และอาหารทะเล ซึ่งสะท้อนการรับรู้ของชาวต่างประเทศต่อเอกลักษณ์ของครัวไทย

การวิเคราะห์ด้วยเทคนิค LDA และ NMF แสดงให้เห็นความแตกต่างในการจำแนกประเด็นอาหารไทย โดย LDA ให้ผลลัพธ์ที่ดีที่สุดที่ 5 หัวข้อ มีจุดเด่นในการแสดงความเชื่อมโยงระหว่างหัวข้อผ่านการซ้อนทับของคำสำคัญ ทำให้เห็นภาพรวมความสัมพันธ์ระหว่างองค์ประกอบต่างๆ ของอาหารไทย แต่มีข้อจำกัดในการจัดหัวข้อที่ขาดความเฉพาะเจาะจง ในขณะที่ NMF ให้ผลลัพธ์ที่เหมาะสมที่ 8 หัวข้อ มีความโดดเด่นในการแยกประเด็นย่อยที่มีความเฉพาะเจาะจง ทำให้เจาะลึกในแต่ละมิติได้ชัดเจนกว่า แต่ขาดการเชื่อมโยงระหว่างหัวข้อในภาพรวม

การเลือกใช้ผลลัพธ์จากเทคนิคใดควรพิจารณาจากวัตถุประสงค์การใช้งาน หากต้องการความเข้าใจเชิงองค์รวมเพื่อการวางแผนกลยุทธ์การท่องเที่ยวเชิงอาหารใน

ระดับประเทศ ผลจาก LDA อาจเหมาะสมกว่า ในขณะที่หากต้องการข้อมูลละเอียดเพื่อการพัฒนาผลิตภัณฑ์อาหารเฉพาะประเภทหรือการปรับปรุงเมนูในร้านอาหารไทยต่างประเทศ ผลจาก NMF จะให้ข้อมูลที่มีประโยชน์มากกว่า

ทั้งนี้ การผสมผสานผลลัพธ์จากทั้งสองเทคนิคอาจให้มุมมองที่สมบูรณ์ที่สุด โดยใช้ LDA เพื่อเข้าใจภาพรวมความเชื่อมโยงระหว่างองค์ประกอบของอาหารไทย ควบคู่กับการใช้ NMF เพื่อวิเคราะห์เชิงลึกในแต่ละองค์ประกอบ

วิธีการวิจัยที่นำเสนอเป็นแนวทางที่สามารถปรับใช้กับข้อมูลจากแหล่งต่างๆ ได้ เช่น รีวิวสินค้าและบริการ รีวิวโรงแรมออนไลน์ หรือข้อมูลจากแพลตฟอร์มสื่อสังคมออนไลน์อื่นๆ ผลการศึกษาและแนวทางการวิจัยต่อเนื่อง สามารถนำไปศึกษาและประยุกต์ใช้ในหลายสาขาอุตสาหกรรม การใช้เทคนิค LDA และ NMF ในการวิเคราะห์เนื้อหาขนาดใหญ่ เช่น การวิเคราะห์ความคิดเห็นของผู้บริโภคจากโซเชียลมีเดียและรีวิวออนไลน์ เพื่อเข้าใจพฤติกรรมและความต้องการของกลุ่มเป้าหมาย การวิเคราะห์รีวิวโรงแรม สถานที่ท่องเที่ยว และร้านอาหาร เพื่อพัฒนาแพ็กเกจท่องเที่ยวและปรับปรุงคุณภาพการบริการ

กิตติกรรมประกาศ

งานศึกษาครั้งนี้ได้รับการสนับสนุนเครื่องมืออุปกรณ์ สถานที่และทรัพยากรต่างๆ จากสาขาเทคโนโลยีสารสนเทศและการสื่อสาร คณะวิทยาศาสตร์และศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน

เอกสารอ้างอิง

[1] Prapasawasdi U, Wuttisittikulkij L, Borompichaichartkul C, Changkaew L, Saadi M. Cultural tourism behaviors: enhancing the influence of tourists' perceptions on local Thai Food and Culture. *The open psychology journals*. 2018 Oct 30;11(1).

[2] สำนักงานนโยบายและยุทธศาสตร์การค้า, กระทรวงพาณิชย์. แนวทางการขยายตลาดสินค้าอาหารไทย. กรุงเทพฯ: สำนักงาน; 2567.

[3] Zhang S, Lopez D, Lam L, Yenumula A, Vu T. A comparative analysis of text mining methodologies for online consumer reviews. *Journal of Applied Business and Economics*. 2023;25(7): doi:10.33423/jabe.v25i7.6724.

[4] เณรัญชรา กิจวิกรานต์. ภาพลักษณ์อาหารไทย การรับรู้คุณภาพอาหารไทยและแนวโน้มพฤติกรรม การท่องเที่ยวเชิงอาหารของนักท่องเที่ยวชาวต่างประเทศ. *วารสารวิชาการการท่องเที่ยวไทยนานาชาติ*. 2558;10(1):12–28.

[5] ธัมมามนตร์ คุณรัตนากรณ์, สมจิตร ล้วนจำเริญ, ราณี อิสิชัยกุล, วัชรีภรณ์ ไชยมงคล. อิทธิพลของส่วนประสมการตลาด และภาพลักษณ์ของผลิตภัณฑ์การท่องเที่ยวเชิงวัฒนธรรมที่มีต่อความภักดีในการท่องเที่ยวเพื่อเรียนรู้วัฒนธรรมไทยของนักท่องเที่ยวชาวต่างประเทศ. *วารสารดุษฎีบัณฑิตทางสังคมศาสตร์.* 2557;4(3):100-117.

[6] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *Journal of Machine Learning Research*. 2003; 3:993-1022.

[7] Seung D, Lee L. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*. 2001;13: 3.

[8] Aggarwal CC, Zhai C. An Introduction to Text Mining. In *Mining Text Data* 2012 Jan 7 (pp. 1-10). Boston, MA: Springer US.

[9] Yadav A, Vishwakarma DK. Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*. 2020;53(6):4335-4385.

[10] Feldman R, Sanger J. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge university press; 2007.

[11] Cifci I, Atsız O, Gupta V. The street food experiences of the local-guided tour in the meal-sharing economy: the case of Bangkok. *British Food Journal*. 2021; 123(12):4030-4048.

[12] Qader WA, Ameen MM, Ahmed BI. An overview of bag of words; importance, implementation, applications, and challenges. In *2019 international engineering conference (IEC).* 2019 Jun 23 (pp. 200-204). IEEE.

[13] Scikit-learn: Machine Learning in Python [Internet]. Version 1.3.2. [cited 2024 Dec 11]. Available from: https://scikit-learn.org/stable/modules/decomposition.html

[14] Chen Y, Zhang H, Liu R, Ye Z, Lin J. Experimental explorations on short text topic mining between LDA and NMF based Schemes. *Knowledge-Based Systems*. 2019 Jan 1;16.

[15] NLTK: Natural Language Toolkit [Internet]. [cited 2024 Dec 11]. Available from: https://www.nltk.org/.

[16] Gensim: Topic Modeling for Human. [cited 2024 Dec 11]. Available from: https://radimrehurek.com/genism.

[17] George S, Srividhya V. Implicit aspect based sentiment analysis for restaurant review using LDA topic modeling and ensemble approach. *International Journal of Advanced Technology and Engineering Exploration*. 2023;10(102): 554 – 568.

[18] Promsivapallop P, Kannaovakun P. Destination food image dimensions and their effects on food preference and consumption. *Journal of destination marketing & management*. 2019; 11: 89-100.

[19] Chavarria LC, Phakdee-Auksorn P. Understanding international tourists' attitudes towards street food in Phuket, Thailand. *Tourism Management Perspectives*. 2017; 21:66-73.

[20] Antón C, Camarero C, Laguna M, Buhalis D. Impacts of authenticity, degree of adaptation and cultural contrast on travellers' memorable gastronomy experiences. *Journal of Hospitality Marketing & Management*. 2019;28(7):743–764.