



การสร้างแบบจำลองหัวข้อร่วมกับการจัดกลุ่มข้อความจากบทวิจารณ์ของนักท่องเที่ยวชาวต่างชาติ
ต่ออุทยานแห่งชาติเขาใหญ่ด้วยการเรียนรู้แบบไม่มีผู้สอน
Topic Modeling and Text Clustering of Foreign Tourists' Reviews
on Khao Yai National Park Using Unsupervised Learning

จักรินทร์ สันติรัตนภักดี^{1*} และ ศุภกฤษฎี นีวัฒนากุล²

Chakkarin Santirattanaphakdi^{1*} and Suphakit Niwattanakul²

¹หลักสูตรบริหารธุรกิจบัณฑิต สาขาวิชาเทคโนโลยีธุรกิจดิจิทัล คณะบริหารธุรกิจ มหาวิทยาลัยวงษ์ชวลิตกุล จังหวัดนครราชสีมา

²สำนักวิทยาศาสตร์และศิลปดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารี จังหวัดนครราชสีมา

¹Bachelor of Business Administration Program in Digital Business Technology, Faculty of Business Administration,
Vongchavalitkul University, Nakhon Ratchasima Province

²Institute of Digital Arts and Science, Suranaree University of Technology, Nakhon Ratchasima Province

*Corresponding Author: chakkarin_san@vu.ac.th

Received: March 4, 2025

Revised: March 14, 2025

Accepted: April 21, 2025

บทคัดย่อ

งานวิจัยเรื่องการสร้างแบบจำลองหัวข้อร่วมกับการจัดกลุ่มข้อความจากบทวิจารณ์ของนักท่องเที่ยวชาวต่างชาติต่ออุทยานแห่งชาติเขาใหญ่ด้วยการเรียนรู้แบบไม่มีผู้สอนนี้ รวบรวมบทวิจารณ์ของนักท่องเที่ยวชาวต่างชาติต่ออุทยานแห่งชาติเขาใหญ่จากแพลตฟอร์มออนไลน์ต่าง ๆ ตั้งแต่ 1 มกราคม 2566 ถึง 31 ธันวาคม 2567 รวมทั้งสิ้น 3,491 รายการ นำมาผ่านกระบวนการทำความสะอาดข้อมูล ก่อนจะแทนความหมายของคำ เพื่อแปลงข้อความให้อยู่ในรูปแบบเวกเตอร์เชิงตัวเลขโดยใช้เทคนิค TF-IDF เพื่อสร้างแบบจำลองหัวข้อด้วยการจัดสรรดีรีเคล พบว่าประกอบด้วย 5 หัวข้อหลัก ได้แก่ 1) สถานที่ท่องเที่ยวและสิ่งแวดล้อม 2) คุณภาพของบริการและสิ่งอำนวยความสะดวก 3) ค่าใช้จ่ายและความคุ้มค่า 4) การเดินทางและการเข้าถึง และ 5) ประสบการณ์ส่วนตัวและความสนุกสนาน สอดคล้องกับผลการจัดกลุ่มแบบเคมีนกับบทวิจารณ์เชิงลบ แบ่งออกเป็น 3 กลุ่ม ได้แก่ 1) ด้านค่าธรรมเนียมการเข้าชมและโครงสร้างการกำหนดราคา 2) ด้านการเข้าถึงและความพึงพอใจในการบริการ และ 3) ด้านสิ่งอำนวยความสะดวกและกิจกรรมการท่องเที่ยว ซึ่งส่งผลต่อความพึงพอใจของนักท่องเที่ยว การแก้ไขข้อบกพร่องเหล่านี้จะช่วยกำหนดกลยุทธ์ที่เหมาะสม เพื่อตอบสนองความต้องการของนักท่องเที่ยวที่หลากหลาย นอกจากนี้ การประยุกต์ใช้ข้อมูลขนาดใหญ่ในการพัฒนานโยบายอย่างเป็นทางการยังเป็นประโยชน์ต่อการกำหนดแนวทางการบริหารจัดการที่ตอบสนองต่อปัญหา และสร้างความยั่งยืนด้านการท่องเที่ยวไทยในอนาคต

คำสำคัญ: การเรียนรู้แบบไม่มีผู้สอน แบบจำลองหัวข้อ การจัดกลุ่มข้อความ การจัดสรรดีรีเคล การจัดกลุ่มแบบเคมีน



Abstract

This research on topic modeling combined with text clustering of international tourists' reviews on Khao Yai National Park utilizing unsupervised learning techniques collected a total of 3,491 reviews from various online platforms between January 1, 2023, and December 31, 2024. The data underwent a cleaning process before being transformed into numerical vectors using the Term Frequency-Inverse Document Frequency (TF-IDF) technique. Topic modeling was then conducted using the Latent Dirichlet Allocation (LDA) method. The analysis identified five main topics: 1) tourist attractions and the natural environment, 2) service quality and available facilities, 3) costs and value for money, 4) transportation and accessibility, and 5) personal experiences and enjoyment. These results correspond with the findings from K-means clustering of negative reviews, which were categorized into three major groups: 1) entrance fees and pricing structures, 2) accessibility and service satisfaction, and 3) facilities and tourism activities. These factors were found to significantly influence tourist satisfaction. Addressing the identified shortcomings can support the development of appropriate strategies to meet the diverse needs of tourists. Furthermore, the application of big data in systematic policy development is instrumental in shaping responsive management approaches that effectively tackle emerging issues and contribute to the long-term sustainability of Thailand's tourism industry.

Keywords: Unsupervised learning, Topic modelling, Text clustering, Latent Dirichlet Allocation (LDA), K-Means clustering

1. บทนำ

การท่องเที่ยวของไทยยังคงเติบโตอย่างต่อเนื่องเนื่องจากความต้องการท่องเที่ยวที่เพิ่มขึ้น รวมถึงการสนับสนุนจากภาครัฐที่มีมาตรการฟรีวีซ่าที่ช่วยอำนวยความสะดวกในการเดินทางเข้าประเทศ และการกลับมาของสายการบินและเที่ยวบินที่เพิ่มขึ้นส่งผลให้ประเทศไทยเป็นหนึ่งในปลายทางหลักที่ดึงดูดนักท่องเที่ยวจากทั่วโลก ในปี 2567 มีนักท่องเที่ยวต่างชาติถึง 35.54 ล้านคน และสร้างรายได้เข้าประเทศกว่า 1.67 ล้านล้านบาท (Ministry of Tourism and Sports, 2024)

หนึ่งในสถานที่ท่องเที่ยวที่ได้รับความนิยมมากที่สุดแห่งหนึ่งของประเทศไทยคืออุทยานแห่งชาติเขาใหญ่ ซึ่งเป็นพื้นที่มรดกโลกทางธรรมชาติ (World Heritage Site) และอุทยานมรดกแห่งอาเซียน (ASEAN Heritage Park) โดยในปี 2565 มีจำนวนนักท่องเที่ยวถึง 1,428,765 คน แบ่งเป็นนักท่องเที่ยวชาวไทย 1,392,078 คน และชาวต่างชาติ 36,687 คน (National Statistical Office, Ministry of Digital Economy and Society, 2022) และเพิ่มขึ้นเป็น 1,486,532 คน ในปี 2566 จากปริมาณนักท่องเที่ยวจำนวนมาก และพฤติกรรมนักท่องเที่ยวในยุคดิจิทัลจึงเกิดการแบ่งปันประสบการณ์การท่องเที่ยวผ่านแพลตฟอร์มออนไลน์ อุตสาหกรรมการท่องเที่ยวจึงต้องปรับตัวให้สอดคล้องกับการเปลี่ยนแปลงของพฤติกรรมของนักท่องเที่ยว เพื่อตอบสนองความต้องการและความคาดหวังของนักท่องเที่ยวที่มีการเปลี่ยนแปลงอย่างรวดเร็ว (Kozak & Kozak, 2018) หนึ่งในข้อมูลสำคัญที่สามารถนำมาปรับปรุงการให้บริการของแหล่งท่องเที่ยว นั่นคือบทวิจารณ์สาธารณะ (Public reviews) ที่นักท่องเที่ยวได้แบ่งปันประสบการณ์ตรงที่เกี่ยวกับความพึงพอใจจากการท่องเที่ยว จุดเด่นของสถานที่ท่องเที่ยว รวมถึงความต้องการและความคาดหวังที่ยังไม่ถูกตอบสนองบนแพลตฟอร์มออนไลน์ (Kumar et al., 2022) อย่างไรก็ตาม มุมมองของนักท่องเที่ยวชาวไทยอาจส่งผลทั้งทางตรงและทางอ้อมต่อความพึงพอใจในแหล่งท่องเที่ยวในประเทศอย่างมีนัยสำคัญ ซึ่งแตกต่างจากนักท่องเที่ยวต่างชาติที่อาจประเมินคุณภาพจากมุมมองที่เน้นประสบการณ์



ชั่วคราวหรือเฉพาะด้านมากกว่า ดังนั้นการวิเคราะห์ข้อมูลที่หลากหลายจากบทวิจารณ์ของนักท่องเที่ยวต่างชาติจะช่วยให้ผู้มีส่วนเกี่ยวข้องสามารถเข้าใจความต้องการได้อย่างรอบด้าน และนำไปสู่การพัฒนาคุณภาพให้ตรงตามความคาดหวังของนักท่องเที่ยวได้อย่างยั่งยืน (Suntornteerasut et al., 2024)

งานวิจัยในปัจจุบันเกี่ยวกับการวิเคราะห์บทวิจารณ์การท่องเที่ยวแบ่งเป็น 4 กลุ่ม ดังนี้ 1) การวิเคราะห์เชิงปริมาณ (Quantitative analysis) เพื่อหาความสัมพันธ์และแนวโน้มต่าง ๆ ด้วยการวิเคราะห์สถิติเชิงพรรณนา (Descriptive statistics) การวิเคราะห์ความถี่ (Frequency analysis) และการวิเคราะห์ความสัมพันธ์ (Correlation analysis) เช่น ราคาทำเลที่ตั้ง หรือคุณภาพของสถานที่ท่องเที่ยว (Wonglapas & Rakha, 2024) ซึ่งสามารถประมวลผลข้อมูลขนาดใหญ่ได้อย่างรวดเร็วและให้ผลลัพธ์ที่เป็นรูปธรรม แต่ข้อจำกัดที่สำคัญคือไม่สามารถสะท้อนความรู้สึกจากนักท่องเที่ยวได้อย่างครบถ้วน 2) การวิเคราะห์เชิงคุณภาพ (Qualitative analysis) เพื่อทำความเข้าใจความรู้สึกเชิงลึกของนักท่องเที่ยว (Soltani-Nejad et al., 2022) ด้วยการวิเคราะห์เนื้อหา (Content analysis) การวิเคราะห์ธีม (Thematic analysis) และการวิเคราะห์กรณีศึกษา (Case study) ร่วมกับการสัมภาษณ์ ซึ่งมีจุดเด่นคือสามารถเข้าถึงมิติที่ซ่อนอยู่ของความคิดเห็น เช่น ความรู้สึกแรงจูงใจ หรือบริบทเฉพาะ อย่างไรก็ตาม ยังมีข้อจำกัดในด้านระยะเวลา และอคติจากผู้วิเคราะห์ที่อาจมีผลต่อการตีความข้อมูล 3) การวิเคราะห์ความรู้สึก (Sentiment analysis) ในบทวิจารณ์ว่าเป็นเชิงบวก เชิงลบ หรือเป็นกลางที่มีต่อสถานที่ท่องเที่ยวหรือบริการ (George & Ramos, 2024) จากอัลกอริทึมต่าง ๆ เช่น นาอิวเบย์ (Naïve bayes) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) หรือโมเดลการเรียนรู้เชิงลึก (Deep learning) เช่น หน่วยความจำระยะสั้นแบบยาว (Long Short-Term Memory: LSTM) โมเดลเชิงภาษาบนสถาปัตยกรรม Transformer เช่น BERT หรือ RoBERTa รวมถึงเครื่องมือประเมินความรู้สึก เช่น TextBlob, Flair หรือ VADER ข้อดีคือสามารถสรุปภาพรวมได้อย่างเป็นระบบ แต่อาจเผชิญปัญหาเมื่อพบภาษาที่มีนัยซ่อนเร้น หรือภาษาที่มีความกำกวม นอกจากนี้โมเดลขนาดใหญ่ยังต้องการทรัพยากรในการประมวลผล จึงมีข้อจำกัดในการใช้งานกับข้อมูลขนาดใหญ่ หรืออุปกรณ์ที่มีข้อจำกัดด้านทรัพยากร และ 4) การวิเคราะห์เนื้อหาเชิงลึก (Advanced text analytics) เพื่อสกัดข้อมูลเชิงลึกและค้นหาความสัมพันธ์ด้วยการประมวลผลภาษาธรรมชาติ (Natural language processing: NLP) (Nawawi et al., 2024) เพื่อแยกคำสำคัญด้วยการวิเคราะห์โครงสร้างประโยค และหาความสัมพันธ์ระหว่างเนื้อหา เช่น การวิเคราะห์เครือข่ายความสัมพันธ์ (Network analysis) การทำเหมืองข้อความ (Text mining) และการวิเคราะห์หัวข้อ (Ali et al., 2022) เพื่อค้นหาหัวข้อหลักที่พบบ่อยครั้งจากบทวิจารณ์ นอกจากนี้การจัดกลุ่มข้อความ เพื่อหาความเชื่อมโยงระหว่างเนื้อหาที่คล้ายคลึงกัน ซึ่งช่วยให้สามารถปรับปรุงและพัฒนาคุณภาพของแหล่งท่องเที่ยวและการให้บริการที่เกี่ยวข้องได้อย่างตรงจุด อย่างไรก็ตาม ข้อจำกัดของแนวทางนี้คือความซับซ้อนในการประมวลผล และความจำเป็นในการมีผู้เชี่ยวชาญเพื่อวิเคราะห์ผลลัพธ์อย่างถูกต้อง

จากที่กล่าวมาคณะผู้วิจัยจึงได้เห็นถึงความสำคัญในการวิเคราะห์เนื้อหาเชิงลึกด้วยการประมวลผลภาษาธรรมชาติ จึงเป็นที่มาของวัตถุประสงค์ในงานวิจัยนี้เพื่อสร้างแบบจำลองหัวข้อร่วมกับการจัดกลุ่มข้อความจากบทวิจารณ์ของนักท่องเที่ยวต่างชาติต่ออุทยานแห่งชาติเขาใหญ่ด้วยการเรียนรู้แบบไม่มีผู้สอน ประกอบด้วยการสร้างแบบจำลองหัวข้อเพื่อจำแนกและสกัดหัวข้อหลัก รวมถึงหัวข้อย่อยที่ซ่อนอยู่ในบทวิจารณ์ได้อย่างเป็นระบบ ซึ่งมีประโยชน์ในการวิเคราะห์เนื้อหาที่หลากหลายโดยไม่ต้องอาศัยการวิเคราะห์ทีละรายการ ส่งผลให้สามารถเข้าถึงประเด็นสำคัญที่นักท่องเที่ยวให้ความสนใจได้อย่างแม่นยำและครอบคลุม ในขณะที่การจัดกลุ่มข้อความที่มีเนื้อหาใกล้เคียงกันให้อยู่ในกลุ่มเดียวกัน ทำให้สามารถมองเห็นแนวโน้ม รูปแบบพฤติกรรม และความต้องการของนักท่องเที่ยวได้ชัดเจนยิ่งขึ้น กระบวนการดังกล่าวจึงเป็นวิธีการที่มีประสิทธิภาพในการวิเคราะห์ข้อมูลขนาดใหญ่ และผลลัพธ์สามารถเป็นปัจจัยขับเคลื่อนหลักในการวางแผนพัฒนาคุณภาพการให้บริการของแหล่งท่องเที่ยวให้สอดคล้องและตอบสนองต่อความคาดหวังของนักท่องเที่ยวได้อย่างยั่งยืน

2. วิธีดำเนินการวิจัย

งานวิจัยนี้เป็นงานวิจัยเชิงประยุกต์เพื่อสร้างแบบจำลองหัวข้อ ร่วมกับการจัดกลุ่มข้อความด้วยการเรียนรู้แบบไม่มีผู้สอน แบ่งการดำเนินงานออกเป็น 3 ขั้นตอน ประกอบด้วย การสร้างชุดข้อมูลจากบทวิจารณ์ของนักท่องเที่ยวชาวต่างชาติต่ออุทยานแห่งชาติเขาใหญ่ การวิเคราะห์ข้อมูลด้วยแบบจำลองหัวข้อ และการวิเคราะห์ข้อมูลด้วยการจัดกลุ่มข้อความ มีรายละเอียดดัง Figure 1.

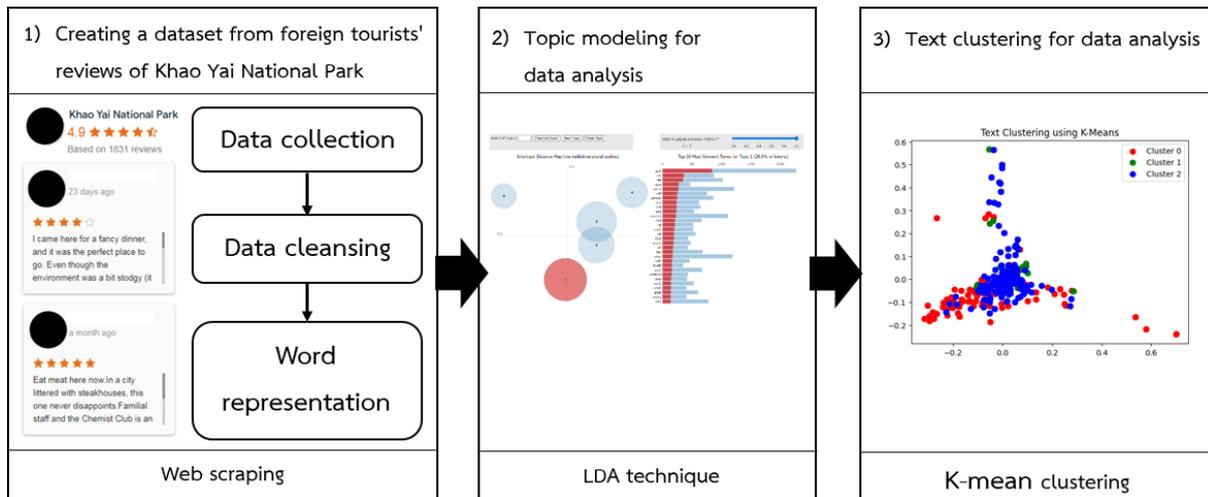


Figure 1. Conceptual framework

2.1 การสร้างชุดข้อมูลจากบทวิจารณ์ของนักท่องเที่ยวชาวต่างชาติต่ออุทยานแห่งชาติเขาใหญ่ (Creating a dataset from foreign tourists' reviews of Khao Yai National Park)

ประกอบด้วย 3 ขั้นตอน มีรายละเอียดดังนี้

2.1.1 การรวบรวมข้อมูล (Data collection)

ด้วยการสกัดและดึงข้อมูล (Web scraping) จากแพลตฟอร์มออนไลน์ด้านการท่องเที่ยวที่เปิดให้นักท่องเที่ยวเข้ามาให้คะแนนและเขียนบทวิจารณ์สาธารณะต่ออุทยานแห่งชาติเขาใหญ่ เช่น Google reviews, Agoda.com, TripAdvisor.com และ Trip.com ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2566 ถึง 31 ธันวาคม พ.ศ. 2567 เป็นระยะเวลา 24 เดือน รวมทั้งสิ้น 3,491 รายการ

2.1.2 การทำความสะอาดข้อมูล (Data cleansing)

งานวิจัยนี้ใช้ไลบรารี NLTK (Natural language toolkit) ประกอบด้วยการนำเข้าชุดคำ ไร้ความหมายจาก nltk.corpus เช่น “the”, “is”, “and” เป็นต้น คำเหล่านี้ไม่มีนัยสำคัญต่อความหมายโดยรวมของข้อความ และมักถูกตัดออกก่อนการวิเคราะห์เพื่อให้โมเดลสามารถมุ่งเน้นไปยังคำที่มีความสำคัญมากกว่า เช่นเดียวกับการนำเข้า WordNetLemmatizer จาก nltk.stem ซึ่งใช้สำหรับการแปลงคำให้อยู่ในรูปแบบรากศัพท์ (Lemma) จากการดาวน์โหลด omw-1.4 (Open multilingual wordnet) ที่ช่วยให้การแปลงคำมีความแม่นยำยิ่งขึ้น โดยพิจารณาตามบริบททางไวยากรณ์ กระบวนการนี้ช่วยลดความหลากหลายของคำที่มีรูปแบบแตกต่างกันแต่มีความหมายคล้ายคลึงกัน อย่างไรก็ตาม งานวิจัยนี้ใช้ไลบรารี spaCy เพื่อตัดคำจากข้อความภาษาอังกฤษจาก en_core_web_sm ซึ่งจะได้ผลลัพธ์ออกมาเป็นวัตถุ (Object) ที่ประกอบด้วยคำแต่ละคำและข้อมูลทางภาษาศาสตร์ที่สำคัญ เช่น ประโยค “this park is amazing!” จะถูกแยกเป็น



“this”, “park”, “is”, “amazing”, “!” จึงเป็นวิธีที่มีประสิทธิภาพมากกว่า word_tokenize จาก nltk.tokenize โดยเฉพาะเมื่อจัดการกับภาษาที่ซับซ้อนหรือคำที่มีรูปแบบพิเศษ เช่น คำย่อ เครื่องหมายวรรคตอน หรือชื่อเฉพาะ

2.1.3 การแทนความหมายของคำ (Word representation)

โดยแปลงข้อความให้กลายเป็นเวกเตอร์เชิงตัวเลขด้วยเทคนิค TF-IDF เพื่อช่วยวิเคราะห์ความสำคัญของคำในข้อความ จากการเรียกใช้ TfidfVectorizer ซึ่งอยู่ในไลบรารี scikit-learn หรือ sklearn ในการแปลงข้อความที่ผ่านการทำความสะอาดข้อมูลแล้วมาตัดคำในแต่ละข้อความออกเป็นหน่วยย่อย (Tokens) ก่อนจะคำนวณความถี่ของคำหรือ TF (Term Frequency) และความถี่ผกผันของเอกสาร หรือ IDF (Inverse Document Frequency) จากสมการที่ 1 และ 2

$$TF(t, d) = \frac{\int t, d}{\sum t' \in d \int t', d} \dots\dots\dots(\text{สมการที่ 1})$$

โดย $\int t, d$ คือจำนวนครั้งที่คำ t ปรากฏในเอกสาร d และตัวส่วนเป็นจำนวนรวมของคำทั้งหมดในเอกสาร d

$$IDF(t) = \log\left(\frac{N}{1 + DF(t)}\right) \dots\dots\dots(\text{สมการที่ 2})$$

โดย N คือจำนวนเอกสารทั้งหมด และ $DF(t)$ คือจำนวนเอกสารที่มีคำ t ปรากฏอยู่ ซึ่งการบวก 1 ในตัวส่วนช่วยป้องกันปัญหาการหารด้วยศูนย์

ดังนั้นค่า TF-IDF จะช่วยถ่วงน้ำหนักของคำ โดยพิจารณาทั้งความสำคัญของคำภายในเอกสารเดียวกันและความสำคัญของคำเมื่อเปรียบเทียบกับเอกสารทั้งหมดในชุดข้อมูลก่อนจะแปลงเป็นเวกเตอร์เชิงตัวเลข ดังสมการที่ 3

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \dots\dots\dots(\text{สมการที่ 3})$$

2.2 การวิเคราะห์ข้อมูลด้วยแบบจำลองหัวข้อ (Topic modeling for data analysis)

โดยนำเวกเตอร์ เชิงตัวเลขจากการแทนความหมายของคำ มากำหนดจำนวนหัวข้อที่ต้องการเป็น 3, 5, 7 และ 10 หัวข้อตามลำดับผ่านการจัดสรรดีรีเคล (Latent Dirichlet Allocation: LDA) (Blei et al., 2003) งานวิจัยนี้เลือกใช้ไลบรารี gensim ซึ่งเป็นไลบรารีที่ได้รับความนิยมอย่างแพร่หลายในงานประมวลผลภาษาธรรมชาติ โมเดลที่นำมาใช้คือ gensim.models.LdaModel โดยอาศัยอัลกอริทึมแบบ Online variational bayes ในการประมาณค่าพารามิเตอร์ที่ซ่อนอยู่ เนื่องจากมีคุณสมบัติเด่นในการรองรับข้อมูลที่มีขนาดใหญ่ โดยไม่จำเป็นต้องโหลดข้อมูลทั้งหมดเข้าสู่หน่วยความจำพร้อมกัน อีกทั้งยังทำงานแบบออนไลน์ (Online learning) เพื่ออัปเดตพารามิเตอร์ของโมเดลอย่างต่อเนื่อง และแสดงผลภาพข้อมูลผ่านไลบรารี pyLDAvis.gensim เป็นแผนภาพระยะห่างระหว่างหัวข้อ และรายการคำศัพท์ที่เกี่ยวข้องกับแต่ละหัวข้อ ซึ่งช่วยให้สามารถตีความและวิเคราะห์ความหมายของแต่ละหัวข้อได้อย่างมีประสิทธิภาพ

การตีความผลลัพธ์จากแบบจำลอง ผู้วิจัยจำเป็นต้องประเมินความเหมาะสมของจำนวนหัวข้อ โดยหัวข้อที่มากเกินไปหรือน้อยเกินไปอาจส่งผลให้หัวข้อที่ได้มีความซ้ำซ้อน หรือไม่สามารถแยกแยะได้อย่างชัดเจน ร่วมกับตรวจสอบการกระจายตัวของหัวข้อให้ห่างกันพอสมควร (Non-overlapping) เพื่อบ่งชี้ว่าหัวข้อที่ได้มีความแตกต่างกันชัดเจน และสามารถ



นำไปสู่การตีความที่มีนัยสำคัญ ตลอดจนพิจารณาค่าที่เกี่ยวข้องกับแต่ละหัวข้อด้วยค่า Relevance score จากความถี่ของคำในแต่ละหัวข้อควบคู่กับสำคัญของคำในภาพรวม ดังนั้นการเลือกค่าพารามิเตอร์ในการปรับระดับความเกี่ยวข้องจึงมีผลต่อการแสดงหัวข้อที่เหมาะสม ก่อนจะแปลความหมายแต่ละหัวข้อโดยผู้วิจัยให้สอดคล้องกับบริบทด้านการท่องเที่ยว

2.3 การวิเคราะห์ข้อมูลด้วยการจัดกลุ่มข้อความ (Text clustering for data analysis)

งานวิจัยนี้ใช้อัลกอริทึมเคมีน (K-Means) ในการจัดกลุ่มข้อมูลที่ไม่มีโครงสร้าง โดยอาศัยหลักการทางคณิตศาสตร์ในการหาความใกล้เคียงกันของข้อมูลในเชิงเวกเตอร์ ซึ่งทำงานผ่านกระบวนการที่เป็นเชิงซ้ำ (Iterative process) เพื่อหาจุดศูนย์กลางของแต่ละกลุ่มอย่างเหมาะสม เริ่มการกำหนดจำนวนกลุ่ม (K) ที่ต้องการจำแนก งานวิจัยนี้กำหนดเป็น 3, 5, 7 และ 10 กลุ่มตามลำดับ จากนั้นโมเดลจะสุ่มเลือกจุดศูนย์กลางเริ่มต้น (Initial centroids) จำนวน K จุดจากชุดข้อมูลให้โดยอัตโนมัติ เพื่อใช้เป็นตำแหน่งเริ่มต้นของจุดศูนย์กลางของแต่ละกลุ่ม ก่อนจะคำนวณระยะห่างของข้อมูลจากศูนย์กลางแต่ละกลุ่ม โดยใช้ระยะทางแบบยูคลิด (Euclidean distance) เพื่อพิจารณาว่าข้อมูลใดควรอยู่ในกลุ่มใด ดังสมการที่ 4

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots\dots\dots(\text{สมการที่ 4})$$

โดยที่ x และ y แทนค่าพิกัดของข้อมูลและศูนย์กลางในแต่ละมิติ

จากสมการที่ 4 ระยะทางแบบยูคลิดเป็นระยะทางทางเรขาคณิตในปริภูมิเวกเตอร์ n มิติ ช่วยให้สามารถวัดความใกล้เคียงระหว่างข้อมูลแต่ละรายการกับศูนย์กลางของแต่ละกลุ่ม เพื่อพิจารณาว่าข้อมูลใดควรถูกจัดให้อยู่ในกลุ่มใด เมื่อข้อมูลถูกจัดกลุ่มเสร็จสิ้นแล้ว จะคำนวณจุดศูนย์กลางใหม่ของแต่ละกลุ่มโดยการหาค่าเฉลี่ยของข้อมูลทั้งหมดในกลุ่มนั้น ซึ่งค่าที่ได้จะกลายเป็นศูนย์กลางใหม่ในแต่ละรอบของกระบวนการแบบวนซ้ำ (Iterative process) จนกว่าจุดศูนย์กลางจะไม่มีเปลี่ยนแปลง หรือเกิดการเปลี่ยนแปลงเพียงเล็กน้อย ซึ่งถือว่าเข้าสู่ภาวะคงที่ (Convergence)

3. ผลการวิจัย

งานวิจัยนี้เป็นงานวิจัยเชิงประยุกต์ เพื่อวิเคราะห์ข้อมูลจากบทวิจารณ์ของนักท่องเที่ยวชาวต่างชาติต่ออุทยานแห่งชาติเขาใหญ่ ด้วยการสร้างแบบจำลองหัวข้อร่วมกับการจัดกลุ่มข้อความด้วยการเรียนรู้แบบไม่มีผู้สอน แบ่งผลการวิจัยออกเป็น 3 ส่วน มีรายละเอียดดังนี้

3.1 ผลการสร้างชุดข้อมูลจากบทวิจารณ์ของนักท่องเที่ยวชาวต่างชาติต่ออุทยานแห่งชาติเขาใหญ่

จากการรวบรวมบทวิจารณ์สาธารณะต่ออุทยานแห่งชาติเขาใหญ่จากแพลตฟอร์มออนไลน์ด้านการท่องเที่ยว ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2566 ถึง 31 ธันวาคม พ.ศ. 2567 ประกอบด้วยค่าคะแนนความพึงพอใจ และเนื้อหาบทวิจารณ์ภาษาอังกฤษ แบ่งเป็นค่าคะแนนความพึงพอใจระดับ 1, 2, 3, 4 และ 5 เป็นจำนวน 149, 158, 462, 788 และ 1,934 ตามลำดับ รวมทั้งสิ้น 3,491 รายการ

กระบวนการทำความสะอาดข้อมูล เพื่อกำจัดข้อมูลที่ไม่จำเป็นและปรับข้อความให้อยู่ในรูปแบบที่เหมาะสมกับการประมวลผล ประกอบด้วยการแปลงข้อความเป็นตัวอักษรพิมพ์เล็กทั้งหมด การลดรูปคำให้อยู่ในรูปแบบพื้นฐาน รวมถึงการลบอักขระพิเศษและเครื่องหมายวรรคตอนที่ไม่มีความหมายของข้อความ รายละเอียดดัง Table 1. คำไร้ความหมายที่ถูกลบออก ได้แก่ “an”, “and”, “is”, “the”, “its”, “well”, “so”, “during”, “you”, “will”, “to”, “it”, “at”, “where”, “or”, “such”, “as”, “with”, “in”, “my”, “not”, “for”, “but”, “more”, “on”, “of”,



“so”, “that”, “was”, “had”, “no”, “till”, “any” และ “has” รวมถึงแปลงเป็นตัวพิมพ์เล็กทั้งหมด เช่นเดียวกับ “hands” เมื่อถูกแปลงให้อยู่ในรูปแบบพื้นฐานจะเป็น “hand” ตลอดจนการลบอักขระพิเศษและเครื่องหมายวรรคตอน

Table 1. Example of a tourist review of Khao Yai National Park after data cleansing

Rating	Plain text	Cleaned text
5	An incredibly big and scenic park. Locals pay 40 thb per head while foreigners pay 100 thb. Its well worth the visit. Wild animals roam the massive park. Go during the winter months. The weather is cool and windy. You will need a car to go through the entire park.	incredibly big scenic park locals pay 40 thb per head foreigners pay 100 thb worth visit wild animals roam massive park winter month weather cool windy need car go entire park
4	The park is not meant for hiking, tracking but more on wheels, driving or biking. At places where you can park, stop over such as viewing points or waterfalls, it is usually crowded and with small stores selling food and drinks. The chance of encounters wild animals is slim in my opinion.	park meant hiking tracking wheel driving biking place park stop viewing point waterfall usually crowded small store selling food drink chance encounter wild animal slim opinion
3	Very beautiful place beautiful waterfall and overview. I paid for entry 400 + 20 thb, the price is very high for entry, did not like that there was no water in the toilet near the first waterfall. soaped my hands and had to look for water to wash off the soap. A lot of monkeys that robbed motorcycles	beautiful place beautiful waterfall overview paid entry 400 20 thb price high entry like water toilet near first waterfall soaped hand look water wash soap lot monkey robbed motorcycle
2	Not quite worth the 2.5 hour drive from Bangkok though. Some trails are not walkable due to fallen trees. Waterfalls are the main attraction but they were just ok. There are other things to see within the Khao Yai National Park, such as elephants and cliff, however no luck seeing them, and the cliff is closed till Oct 2024.	quite worth 2.5 hour drive bangkok though trail walkable due fallen tree waterfall main attraction ok thing see within khao yai national park elephant cliff however luck seeing cliff closed oct 2024
1	400 baht for foreigner, 40 baht for thai. Where is equity? Or you think foreigners so stupid to pay 400 baht for nothing. I'll never go to any of national parks anymore as it has no sense for such money.	400 baht foreigner 40 baht thai equity think foreigner stupid pay 400 baht nothing never go national park anymore sense money



การแทนความหมายของคำ เริ่มจากนำข้อความที่ผ่านการทำความสะอาดข้อมูลแล้ว (Cleaned text) มาตัดคำในแต่ละข้อความออกเป็นหน่วยย่อย (Tokens) ก่อนจะแปลงเป็นข้อมูลเชิงตัวเลขในรูปแบบเวกเตอร์ (Vector) ด้วยการคำนวณค่า TF-IDF ช่วยวิเคราะห์ความสำคัญของคำในข้อความ ตัวอย่างประโยค “the park is good and awesome” และ “the park is disappointed” แทนด้วย doc1 และ doc2 ตามลำดับ รายละเอียดดัง Table 2. เวกเตอร์คุณลักษณะ (Feature vector) จากการคำนวณ TF-IDF จะแสดงถึงความสำคัญของคำในเอกสารแต่ละฉบับที่ต่างกัน เช่น “good”, “awesome” และ “disappointed” ที่ปรากฏในเอกสาร doc1 และ doc2 จึงสามารถใช้ในการแยกแยะคำสำคัญได้ดีขึ้น ในทางกลับกัน “the” และ “is” มีค่า TF-IDF ต่ำหรืออาจเท่ากับ 0 แสดงถึงเป็นคำที่ปรากฏบ่อย แต่ไม่ค่อยมีความสำคัญในเอกสาร และถือเป็นคำไร้ความหมายที่อาจถูกกำจัดตั้งแต่กระบวนการทำความสะอาดข้อความ เช่นเดียวกับ “and”

Table 2. Example of TF-IDF calculation for two documents

Word	TF		df	IDF(log ₂)	TF * IDF	
	doc1	doc2			doc1	doc2
the	1/7	1/7	2	log(2/2) = 0	0	0
park	1/7	1/7	2	log(2/2) = 0	0	0
is	1/7	1/7	2	log(2/2) = 0	0	0
good	1/7	0	1	log(2/1) = 0.3010	0.3010 × 1/7 ≈ 0.043	0
and	1/7	0	1	log(2/1) = 0.3010	0.043	0
awesome	1/7	0	1	log(2/1) = 0.3010	0.043	0
disappointed	0	1/7	1	log(2/1) = 0.3010	0	0.043

3.2 ผลการวิเคราะห์ข้อมูลด้วยแบบจำลองหัวข้อ

จากการนำเวกเตอร์เชิงตัวเลขที่ผ่านการแทนความหมายของคำมาใช้ในการวิเคราะห์ด้วยการจัดสรรดีรีเคล เพื่อระบุหัวข้อที่ซ่อนอยู่ในบทความ ภายใต้แนวคิดที่ว่าแต่ละบทความเป็นการรวมกันของประเด็นย่อยหลายประเด็น และแต่ละประเด็นย่อยถูกประกอบขึ้นจากชุดของคำศัพท์ที่มีความสัมพันธ์กัน การวิเคราะห์ข้อความจึงต้องคำนึงถึงโครงสร้างเชิงเนื้อหา และการกระจายตัวของคำภายในแต่ละหัวข้อ ดัง Figure 2.

จาก Figure 2. ผลการวิเคราะห์หัวข้อด้วยการจัดสรรดีรีเคล และนำเสนอผลลัพธ์ในรูปแบบภาพข้อมูลเชิงโต้ตอบผ่าน pyLDAvis ที่ช่วยให้สามารถตีความและวิเคราะห์โครงสร้างของหัวข้อได้อย่างชัดเจนขึ้น ประกอบด้วย

3.2.1 แผนที่ระยะห่างระหว่างหัวข้อ (Intertopic distance map)

แสดงการกระจายตัวของหัวข้อในรูปแบบ Multidimensional Scaling (MDS) โดยแต่ละวงกลมมีหมายเลขกำกับภายในแทนหัวข้อที่ได้จากการวิเคราะห์ และขนาดของวงกลมสะท้อนถึงสัดส่วนของข้อความที่เกี่ยวข้องกับหัวข้อนั้นในชุดข้อมูลทั้งหมด

3.2.2 ความเกี่ยวข้องของคำศัพท์กับหัวข้อที่เลือก (Relevance metric: λ) เป็นพารามิเตอร์ที่ใช้ควบคุม

ความสำคัญของคำศัพท์แต่ละคำในแต่ละหัวข้อ โดยค่าความเกี่ยวข้องจะอยู่ระหว่าง 0 ถึง 1 คำนวณได้จากสมการที่ 5

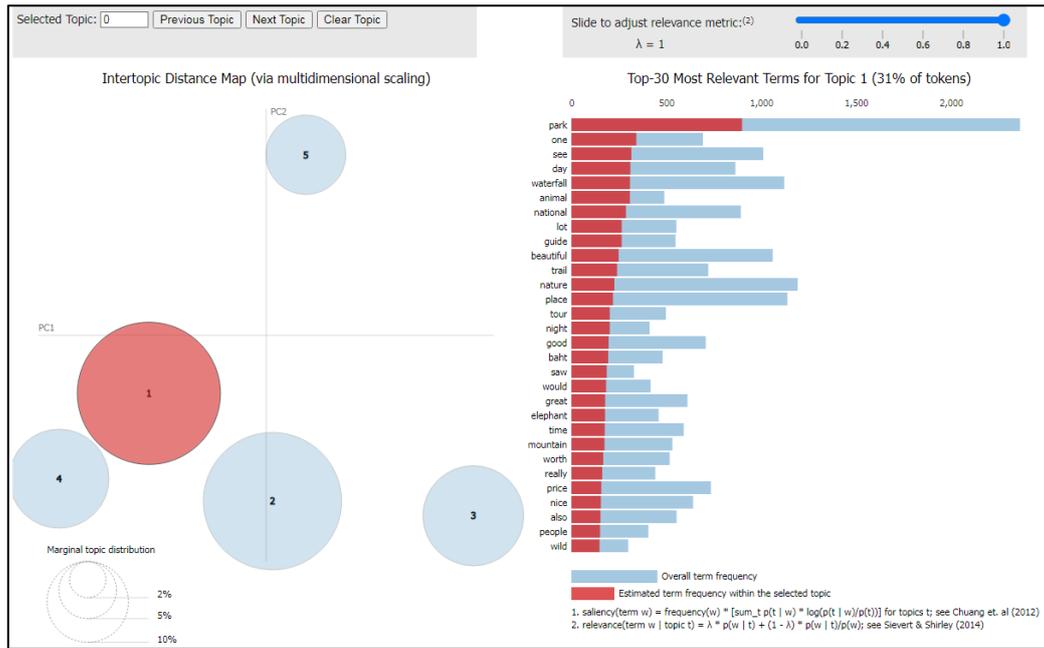


Figure 2. Topic Modeling Analysis Results with topic = 5

$$relevance(w|t) = \lambda P(w|t) + (1 - \lambda) \frac{P(w|t)}{P(w)} \dots\dots\dots(\text{สมการที่ 5})$$

โดย $P(w|t)$ คือ ความน่าจะเป็นของคำ w ที่จะปรากฏในหัวข้อ t
 $P(w)$ คือ ความน่าจะเป็นของคำ w ทั้งหมดในเอกสาร
 λ คือ พารามิเตอร์ที่ควบคุมความสมดุลระหว่างสององค์ประกอบข้างต้น

จากสมการที่ 5 หากกำหนดค่าพารามิเตอร์ $\lambda = 1$ คำที่ปรากฏจะเป็นคำที่พบได้บ่อยที่สุดในหัวข้อที่เลือก ซึ่งหมายความว่าคำเหล่านี้เป็นคำที่ใช้ซ้ำบ่อยภายในหัวข้อนั้นโดยเฉพาะ โดยไม่มีการพิจารณาว่าคำเหล่านั้นกระจายอยู่ในเอกสารทั้งหมดมากน้อยเพียงใด วิธีนี้เหมาะสำหรับการระบุคำหลักที่เป็นเอกลักษณ์ของหัวข้อนั้น อย่างไรก็ตาม คำที่ใช้ทั่วไปในหลายหัวข้ออาจถูกนำมาแสดงด้วย ในทางตรงกันข้าม หากกำหนดค่า $\lambda = 0$ คำที่ถูกแสดงจะเป็นคำที่มีความจำเพาะกับหัวข้อที่เลือก โดยเปรียบเทียบกับ การกระจายของคำเหล่านั้นในเอกสารทั้งหมด ดังนั้นคำที่อาจมีความถี่ต่ำแต่มีบทบาทสำคัญในการแยกแยะหัวข้อนั้นออกจากหัวข้อนั้นจะถูกเน้นให้เห็นเด่นชัด วิธีนี้เหมาะสำหรับค้นหาคำที่ช่วยจำแนกหัวข้อต่าง ๆ ออกจากกันได้อย่างชัดเจน และเมื่อค่า λ อยู่ระหว่าง 0 และ 1 ผลลัพธ์ที่ได้จะแสดงทั้งคำที่มีความถี่สูงในหัวข้อที่เลือกและคำที่มีนัยสำคัญในการจำแนกหัวข้อ วิธีนี้ช่วยให้สามารถมองเห็นภาพรวมของหัวข้อนั้นได้อย่างสมดุล ทั้งในแง่ของคำที่ใช้บ่อยและคำที่มีบทบาทสำคัญต่อโครงสร้างของโมเดล

3.2.3 คำที่มีความเกี่ยวข้องกับหัวข้อ (Most relevant terms)

ซึ่งแสดงคำศัพท์ที่มีอิทธิพลต่อหัวข้อที่เลือกประกอบด้วยความถี่ของคำในเอกสารทั้งหมด (Overall term frequency) แทนด้วยสีฟ้า และความถี่ของคำในหัวข้อที่เลือก (Estimated term frequency within the selected topic) แทนด้วยสีแดง



จากผลการวิเคราะห์หัวข้อด้วยการจัดสรรดีรีเคล พบว่า การกระจายตัวของหัวข้อและคำที่เกี่ยวข้องในหัวข้อที่ 1 เป็นหัวข้อหลักของชุดข้อมูล เนื่องจากมีขนาดใหญ่ที่สุดและประกอบด้วยคำที่เกี่ยวข้องกับการท่องเที่ยวและธรรมชาติ ในขณะที่หัวข้ออื่น ๆ มีขนาดเล็กกว่า สะท้อนให้เห็นถึงเนื้อหาที่เฉพาะเจาะจงมากขึ้น นอกจากนี้ระยะห่างระหว่างวงกลมแสดงถึงความแตกต่างของเนื้อหา หากหัวข้อได้อยู่ห่างจากกลุ่มของหัวข้ออื่นมากหมายถึงหัวข้อนั้นมีเนื้อหาที่แตกต่างกันอย่างชัดเจน ในทางกลับกันวงกลมที่อยู่ใกล้กันสะท้อนถึงหัวข้อที่มีเนื้อหาคล้ายคลึงกันหรือมีคำที่ซ้อนทับกันสูง อันจะเห็นได้จากหัวข้อที่ 5 ซึ่งแยกตัวออกจากกลุ่มของหัวข้ออื่นอย่างชัดเจน อาจหมายถึงเนื้อหาที่แตกต่างจากหัวข้ออื่น ๆ อย่างมีนัยสำคัญ รวมถึงอาจเป็นเรื่องที่ไม่เกี่ยวข้องกับการท่องเที่ยวหรือมีคำศัพท์เฉพาะที่แตกต่างจากหัวข้อหลักดัง Table 3.

Table 3. Topic analysis results using LDA: 5 key topics emphasized in the reviews

Topics	Keywords
1. Tourist attractions and environment	park, nature, beautiful, trail, waterfall, mountain, scenario, guide
2. Service quality and amenities	staff, service, friendly, helpful, hotel, restaurant, clean, food
3. Cost and value for money	price, expensive, worth, money, cheap, ticket, entrance, cost
4. Transportation and accessibility	transport, bus, taxi, road, drive, parking, far, location
5. Personal experience and enjoyment	fun, experience, amazing, enjoy, exciting, great, love, adventure

จากตารางที่ 3 สามารถแบ่งผลการวิเคราะห์หัวข้อจากบทวิจารณ์ของนักท่องเที่ยวชาวต่างชาติต่ออุทยานแห่งชาติเขาใหญ่เป็นกลุ่มใหญ่ ๆ ที่สะท้อนถึงสิ่งที่นักท่องเที่ยวให้ความสำคัญ ได้แก่ 1) สถานที่ท่องเที่ยวและสิ่งแวดล้อม (Tourist attractions and environment) จากบทวิจารณ์ในกลุ่มนี้ที่กล่าวถึงความงามของธรรมชาติ น้ำตก เส้นทางเดินป่า สัตว์ป่า และภูเขา รวมถึงการแนะนำสถานที่ และเส้นทางที่น่าสนใจในการเยี่ยมชม 2) คุณภาพของบริการและสิ่งอำนวยความสะดวก (Service quality and amenities) หัวข้อนี้เกี่ยวข้องกับคุณภาพบริการของโรงแรม ร้านอาหาร และผู้ให้บริการ โดยที่นักท่องเที่ยวให้ความสำคัญเน้นเรื่องความสะดวกของที่พัก ความเป็นกันเองของเจ้าหน้าที่ และคุณภาพอาหาร 3) ค่าใช้จ่ายและความคุ้มค่า (Cost and value for money) ที่สะท้อนถึงความพึงพอใจเกี่ยวกับค่าธรรมเนียม โดยเฉพาะการเปรียบเทียบระหว่างชาวไทยและชาวต่างชาติ ทั้งความรู้สึกว่าเหมาะสมหรือสูงเกินไปสำหรับประสบการณ์ที่ได้รับจากการท่องเที่ยวในสถานที่อื่น ๆ 4) การเดินทางและการเข้าถึง (Transportation and accessibility) ตลอดจนปัญหาและอุปสรรคในการเดินทาง เช่น จำนวนยานพาหนะในช่วงเทศกาลหรือวันหยุดยาว การใช้ความเร็วเกินกฎหมายกำหนดในเส้นทางสาธารณะ หรือค่าบริการขนส่งสาธารณะที่ไม่ครอบคลุม ยากต่อการเข้าถึงและมีราคาแพง และ 5) ประสบการณ์ส่วนตัวและความสนุกสนาน (Personal experience and enjoyment) โดยใช้คำที่มีความหมายเชิงบวกจากความประทับใจในการท่องเที่ยว ซึ่งสอดคล้องกับกิจกรรมที่หลากหลายและตอบสนองความต้องการของนักท่องเที่ยวได้ทุกรูปแบบ

การวิจัยนี้แบ่งการทดสอบโดยกำหนดค่าพารามิเตอร์ความเกี่ยวข้อง (λ) ที่แตกต่างกัน เพื่อแสดงถึงผลกระทบของการเลือกจำนวนหัวข้อในการสรุปข้อมูล โดยแบ่งออกเป็น 3, 5, 7 และ 10 หัวข้อตามลำดับ มีรายละเอียดดังนี้ เมื่อแบ่งออกเป็น 3 หัวข้อ กำหนด $\lambda = 1.0$ ข้อมูลที่ได้รับจะเป็นหัวข้อที่กว้างและครอบคลุม เช่น สถานที่ การบริการและค่าใช้จ่ายจากคำที่พบได้บ่อย เช่น “hotel”, “food” และ “transport” ซึ่งยังไม่สามารถแยกแยะหัวข้อย่อยได้อย่างชัดเจน เมื่อใช้ $\lambda = 0.6$ เริ่มเห็นคำที่เฉพาะเจาะจงมากขึ้น เช่น “mountain”, “adventure” และ “experience” สำหรับหัวข้อที่เกี่ยวกับสถานที่ แต่ยังไม่สามารถแยกแยะหัวข้อย่อยได้ดีมากนัก แต่เมื่อใช้ $\lambda = 0.3$ คำที่เลือกมีความสำคัญเชิงจำแนกที่ชัดเจนมากขึ้น เช่น “parking”, “taxi” และ “expensive” แต่ยังคงมีข้อจำกัดในการแยกหัวข้อย่อยได้อย่างสมบูรณ์ เช่นเดียวกับการแบ่งเป็น 5 หัวข้อ โดยกำหนด $\lambda = 1.0$ พบว่าการแบ่งหัวข้อยังคงค่อนข้างคลุมเครือ และคำที่โดดเด่นยังเป็น



คำที่ทั่วไป แต่เมื่อใช้ $\lambda = 0.6$ หัวข้อเริ่มมีความเฉพาะเจาะจงมากขึ้น ประกอบด้วย สถานที่ท่องเที่ยว เช่น “park”, “waterfall” และ “trail” การบริการ เช่น “staff”, “hotel” และ “friendly” คำใช้จ่าย เช่น “expensive”, “ticket” และ “worth” การเดินทาง เช่น “bus”, “taxi” และ “drive” และประสบการณ์โดยรวม เช่น “fun”, “amazing” และ “enjoy” เป็นต้น โดยคำที่เกี่ยวข้องกับหัวข้อจะมีความเฉพาะเจาะจงมากขึ้น ซึ่งสามารถทำให้การแยกแยะหัวข้อย่อยทำได้ดีขึ้น และเมื่อใช้ $\lambda = 0.3$ การจำแนกหัวข้อชัดเจนขึ้น เช่น หัวข้อค่าใช้จ่ายสามารถแยกออกจากการเดินทางได้ดีขึ้น ดังนั้นการใช้ 5 หัวข้อจึงเหมาะสมสำหรับการอธิบายข้อมูลให้ครบถ้วน โดยการลดค่าของ λ ช่วยแยกแยะคำที่เฉพาะเจาะจงได้ดีขึ้น

นอกจากนั้นเมื่อแบ่งเป็น 7 หัวข้อ กำหนด $\lambda = 1.0$ พบว่าหัวข้อที่แยกออกมาหลัก ๆ ยังคงมีความคล้ายคลึงกับ 5 หัวข้อเดิม แต่เมื่อใช้ $\lambda = 0.6$ จะสามารถแยกคำที่เฉพาะเจาะจงได้มากขึ้น ประกอบด้วยอาหาร โรงแรม และการบริการ ซึ่งแยกออกจากกันได้อย่างชัดเจน แต่เมื่อใช้ $\lambda = 0.3$ หัวข้อที่ได้มีรายละเอียดมากขึ้น ประกอบด้วยอาหารและร้านอาหาร เช่น “food”, “taste” และ “delicious” ตลอดจนรีวิวเชิงบวก เช่น “wonderful”, “love” และ “unforgettable” หรือรีวิวเชิงลบ เช่น “dirty”, “bad” และ “terrible” ซึ่งมีความเฉพาะเจาะจงมากขึ้น แต่บางหัวข้อเริ่มเล็กลงไปและอาจไม่มีความหมายที่ชัดเจน เช่น “spicy” สำหรับอาหาร หรือ “overprice” สำหรับค่าใช้จ่าย ตลอดจนการแบ่งเป็น 10 หัวข้อ เมื่อใช้ $\lambda = 1.0$ พบว่าหัวข้อที่ได้มีความซ้ำซ้อนกันมาก จึงไม่สามารถแยกแยะข้อมูลได้อย่างชัดเจน หรือเมื่อใช้ $\lambda = 0.6$ พบว่าการแยกรีวิวเชิงบวก และรีวิวเชิงลบออกจากกันทำได้ดีขึ้น แต่บางหัวข้ออาจแตกย่อยมากเกินไป เช่น บ้านพักอุทยานกับลานกางเต็นท์ ซึ่งมีบริบทที่แตกต่างกันในด้านความสะดวกสบาย แต่เมื่อใช้ $\lambda = 0.3$ คำที่ปรากฏมีความเฉพาะทางมาก เช่น “vegan” และ “rush hour” ซึ่งบางครั้งยากต่อการตีความข้อมูล ส่วนเมื่อใช้ $\lambda = 0.0$ หัวข้อเริ่มแตกเป็นกลุ่มย่อยที่อาจไม่จำเป็น และบางหัวข้อเริ่มซ้ำซ้อนกันเกินไปและยากต่อการทำความเข้าใจ

3.3 ผลการวิเคราะห์ข้อมูลด้วยการจัดกลุ่มข้อความแบบเคมีน

จากการนำเวกเตอร์เชิงตัวเลขที่ผ่านการแทนความหมายของคำมาใช้ในการคำนวณระยะห่างระหว่างข้อมูล โดยกำหนดจำนวนกลุ่มที่ต้องการ และสุ่มเลือกจุดเซ็นทรอยด์จำนวน K จุด เพื่อเป็นจุดตั้งต้นของแต่ละกลุ่ม ก่อนที่จะจัดกลุ่มข้อมูลจากการคำนวณระยะห่างจากข้อมูลนั้นไปยังจุดเซ็นทรอยด์ของแต่ละกลุ่ม และจัดกลุ่มข้อมูลให้ใกล้เคียงกับจุดเซ็นทรอยด์ที่มีระยะห่างน้อยที่สุด แล้วจะทำการปรับค่า เพื่อคำนวณค่าเฉลี่ยของจุดเซ็นทรอยด์ใหม่สำหรับแต่ละกลุ่มซ้ำไปเรื่อย ๆ จนกว่าจุดเซ็นทรอยด์ไม่มีการเปลี่ยนแปลง

งานวิจัยนี้กำหนดจำนวนกลุ่ม (K) เริ่มต้นเท่ากับ 3 เมื่อนำผลการจัดกลุ่มมาผ่านกระบวนการลดมิติข้อมูลด้วย PCA (Principal Component Analysis) โดยการหาทิศทางที่ข้อมูลมีความแปรปรวนสูงสุด และใช้ทิศทางเหล่านั้นเป็นแกนใหม่เพื่อแทนข้อมูลเดิม ซึ่งช่วยให้สามารถแสดงผลข้อมูลใน 2 มิติได้โดยที่ยังคงลักษณะสำคัญของข้อมูลไว้ ก่อนจะนำมาแสดงผลด้วยแผนภาพการกระจาย (Scatter plot) และวิเคราะห์คำสำคัญในแต่ละกลุ่ม รายละเอียด Figure 3.

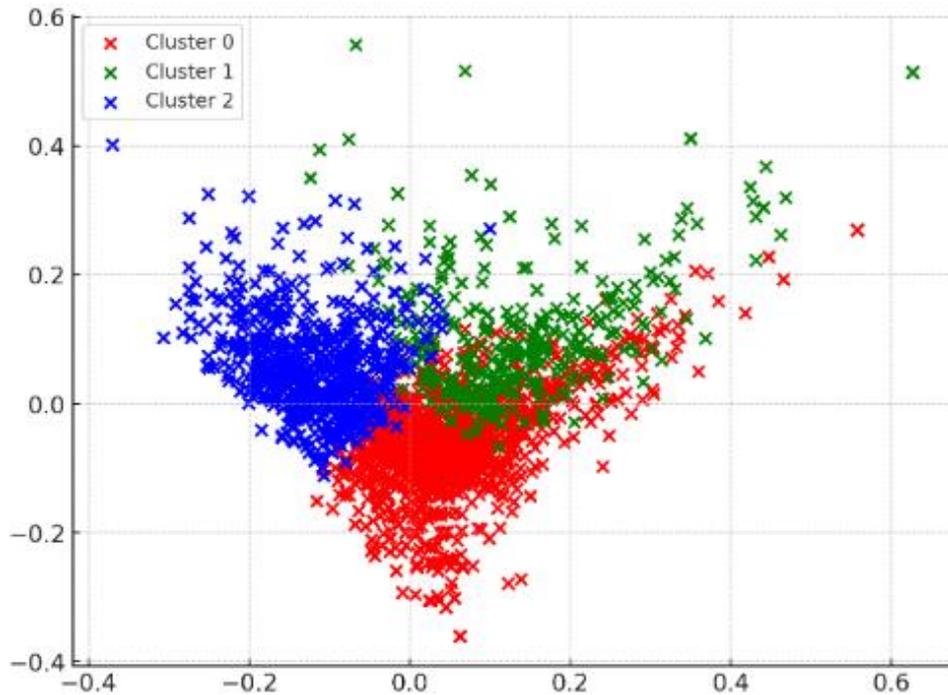


Figure 3. Clustering Results Using K-Means with K = 3

จาก Figure 3. ผลการวิเคราะห์ห้บทวิจารณ์สามารถจำแนกได้เป็น 3 กลุ่มหลักในประเด็นที่แตกต่างกันตามบริบท ได้แก่ 1) ประสบการณ์เกี่ยวกับสัตว์ป่าและบริการนำเที่ยว (Wildlife experience and guided tours) 2) ความสวยงามของธรรมชาติและแหล่งท่องเที่ยวภายในอุทยาน (Natural beauty and tourist attractions) และ 3) ค่าธรรมเนียมและความคุ้มค่า (Fees and value for tourism) ดัง Table 4.

Table 4. Key terms from K-Means clustering with K = 3

Clusters	Keywords
1. Wildlife experience and guided tours	park, guide, saw, day, animals, elephants, tour, great, wildlife, monkeys
2. Natural beauty and tourist attractions	park, mountain, atmosphere, reservoir, visit, beautiful, waterfalls, nature, view, day
3. Fees and value for tourism	place, park, beautiful, 400, good, thai, visit, nature, baht, nice

จาก Table 4. ผลลัพธ์จากการจัดกลุ่มข้อความแบบเคมีน ประกอบด้วย Cluster 0 (สีแดง) นักท่องเที่ยวในกลุ่มนี้ มุ่งเน้นไปที่ประสบการณ์เกี่ยวกับสัตว์ป่า โดยเฉพาะการพบเห็นช้าง ลิง และสัตว์อื่น ๆ นอกจากนี้ยังมีกรกล่าวถึงบริการนำเที่ยวโดยผู้นำทางหรือไกด์ ซึ่งช่วยอำนวยความสะดวกและให้ความรู้เกี่ยวกับระบบนิเวศของอุทยาน ซึ่งสะท้อนถึงความ ต้องการของนักท่องเที่ยวที่มองหาโอกาสในการชมสัตว์ป่าในสภาพแวดล้อมธรรมชาติ Cluster 1 (สีเขียว) บทวิจารณ์ในกลุ่มนี้ กล่าวถึงลักษณะทางธรรมชาติของอุทยานแห่งชาติเขาใหญ่ เช่น น้ำตก ป่าไม้ และความสวยงามของระบบนิเวศ นักท่องเที่ยว ส่วนใหญ่แสดงความชื่นชมต่อทัศนียภาพของอุทยาน และให้ความเห็นว่าควรค่าแก่การเดินทางมาเยี่ยมชม และ Cluster 2

(สีน้ำเงิน) บทวิจารณ์ในกลุ่มนี้มุ่งเน้นไปที่ค่าธรรมเนียมเข้าอุทยาน ซึ่งบางบทวิจารณ์แสดงความคิดเห็นว่ามีอัตราที่สูงเกินไปเมื่อเทียบกับค่าธรรมเนียมสำหรับชาวไทย นอกจากนี้ยังมีความคิดเห็นเกี่ยวกับความคุ้มค่าของการเข้าชมอุทยาน ทั้งในด้านสิ่งอำนวยความสะดวกและประสบการณ์ที่ได้รับสะท้อนถึงข้อกังวลเกี่ยวกับค่าธรรมเนียมเข้าอุทยาน โดยบทวิจารณ์จำนวนไม่น้อยแสดงความคิดเห็นว่าค่าธรรมเนียมมีอัตราสูงกว่าเมื่อเทียบกับนักท่องเที่ยวชาวไทย ปัจจุบันนี้เป็นประเด็นสำคัญที่ควรพิจารณาเพื่อหาสมดุลระหว่างการสร้างรายได้จากการท่องเที่ยวและการเข้าถึงของนักท่องเที่ยวต่างชาติ

เนื่องจากบทวิจารณ์ที่มีคะแนนสูงมักจะไม่มีระบุข้อบกพร่องหรือปัญหามากนัก จึงอาจไม่เหมาะสมสำหรับการวิเคราะห์เชิงลึก งานวิจัยนี้จึงได้คัดเลือกเฉพาะบทวิจารณ์ที่มีคะแนนความพึงพอใจเท่ากับ 3, 2 และ 1 จำนวน 462, 158 และ 149 ตามลำดับ รวมทั้งสิ้น 769 รายการเพื่อทำการจัดกลุ่มข้อความอีกครั้ง ผลการวิเคราะห์บทวิจารณ์เชิงลบจากนักท่องเที่ยวต่างชาติ พบว่า จำแนกเป็น 3 กลุ่ม ดัง Figure 4.

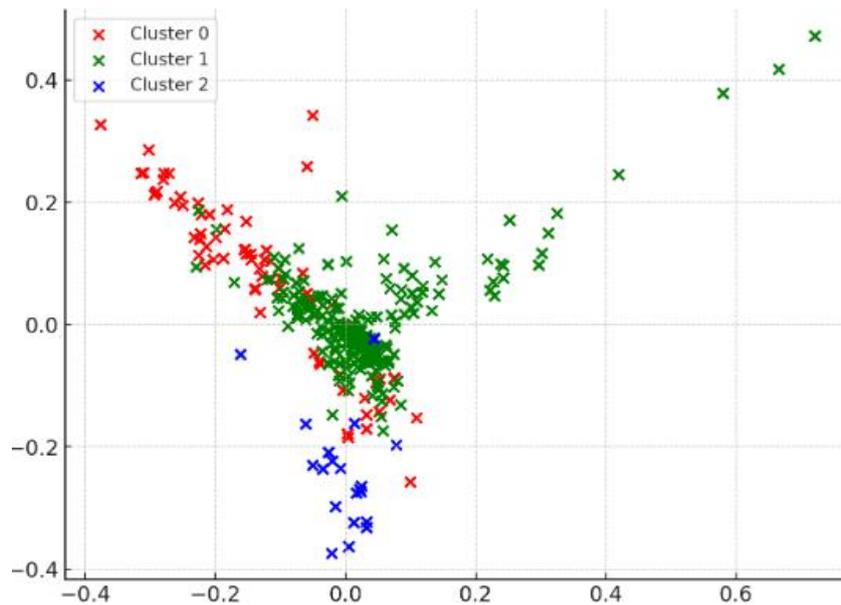


Figure 4. K-means clustering results for negative reviews with K = 3

จาก Figure 4. เมื่อพิจารณาการกระจายตัวของข้อมูลของแต่ละกลุ่มในภาพรวม พบว่าไม่มีจุดข้อมูลใดที่แยกตัวออกห่างจากกลุ่มอย่างชัดเจน (Outlier) สะท้อนว่าข้อมูลที่ใช้จัดกลุ่มเป็นไปในทิศทางเดียวกัน และทุกกลุ่มมีบทบาทสำคัญต่อการวิเคราะห์ความคิดเห็นของนักท่องเที่ยวอย่างครอบคลุม เมื่อพิจารณาในรายละเอียด พบว่าการวิเคราะห์และจัดกลุ่มบทวิจารณ์เชิงลบออกเป็น 3 กลุ่ม ประกอบด้วย Cluster 0 (สีแดง) ด้านค่าธรรมเนียมการเข้าชมและโครงสร้างการกำหนดราคาถือเป็นหนึ่งในประเด็นหลักที่นักท่องเที่ยวให้ความสนใจ คำสำคัญที่ปรากฏในกลุ่มนี้ได้แก่ “entrance”, “fee”, “thai”, “pay”, “price” และ “foreign” เป็นต้น ซึ่งอาจสะท้อนถึงข้อสังเกตเกี่ยวกับค่าธรรมเนียมการเข้าชมและอัตราค่าบริการ อาจมาจากความต้องการข้อมูลที่ชัดเจนยิ่งขึ้นเกี่ยวกับโครงสร้างราคา ตลอดจนแนวทางการเรียกเก็บค่าธรรมเนียมที่ควรมีความสอดคล้องกับคุณภาพของบริการและสิ่งอำนวยความสะดวก เช่นเดียวกับ Cluster 1 (สีเขียว) ด้านการเข้าถึงและความพึงพอใจในการบริการ โดยเฉพาะการเดินทางและการใช้บริการภายในอุทยานเป็นอีกหนึ่งปัจจัยที่นักท่องเที่ยวให้ความสำคัญ คำสำคัญที่เกี่ยวข้องกับกลุ่มนี้ได้แก่ “hard”, “new”, “disappointed”, “help” และ “watch” เป็นต้น ซึ่งอาจสะท้อนถึงข้อเสนอแนะเกี่ยวกับความสะดวกสบายในการเดินทาง เช่น ความต้องการระบบขนส่งสาธารณะที่ง่ายต่อการเข้าถึงและเชื่อมโยงอย่างเป็นระบบ การเพิ่มจำนวนป้ายบอกทางและพัฒนาให้ชัดเจนยิ่งขึ้น เส้นทางที่เหมาะสมกับการเดินเท้า รวมถึง



การให้บริการข้อมูลและความช่วยเหลือจากเจ้าหน้าที่ที่รองรับความต้องการของนักท่องเที่ยวได้อย่างมีประสิทธิภาพ และ Cluster 2 (สีน้ำเงิน) ด้านสิ่งอำนวยความสะดวกและกิจกรรมการท่องเที่ยว นอกเหนือจากเส้นทางธรรมชาติและทัศนียภาพที่งดงาม สิ่งอำนวยความสะดวกและตัวเลือกในการทำกิจกรรมภายในอุทยานก็เป็นอีกหนึ่งปัจจัยที่มีผลต่อความพึงพอใจของนักท่องเที่ยว คำสำคัญที่เกี่ยวข้องกับกลุ่มนี้ได้แก่ “activities”, “choice”, “time”, “rent”, “signs” และ “camp” รวมถึง “watch” ที่เกี่ยวข้องกับกิจกรรมสองชีวิตสัตว์ป่า ซึ่งสะท้อนถึงข้อเสนอแนะเกี่ยวกับความหลากหลายของกิจกรรมและสิ่งอำนวยความสะดวก เช่น เพิ่มตัวเลือกในการท่องเที่ยว มีป้ายบอกทางที่ครอบคลุมพื้นที่สำคัญ และการพัฒนาพื้นที่ให้บริการ เป็นต้น

อย่างไรก็ดี เมื่อทดสอบจัดกลุ่มข้อความออกเป็น 5, 7 และ 10 กลุ่ม เพื่อเปรียบเทียบแนวโน้มของปัญหาและศึกษาผลกระทบของจำนวนกลุ่มต่อความชัดเจนของผลลัพธ์ พบว่า เมื่อเพิ่มจำนวนเป็น 5 กลุ่มจะมีการแยกย่อยออกไปโดยจำแนกเพิ่มเติมเป็นด้านการบริการของเจ้าหน้าที่และข้อมูลสำหรับนักท่องเที่ยว และด้านกิจกรรมและคุณภาพของแหล่งท่องเที่ยว นอกเหนือจากสามกลุ่มเดิม ส่งผลให้สามารถมองเห็นปัญหาที่เกี่ยวข้องกับการให้บริการและความคาดหวังของนักท่องเที่ยวได้ชัดเจนขึ้น โดยเฉพาะเรื่องความช่วยเหลือจากเจ้าหน้าที่และคุณภาพของกิจกรรมที่น่าเสนอ อย่างไรก็ตาม บางกลุ่มอาจยังมีประเด็นที่ทับซ้อนกัน ทำให้ยังไม่สามารถแยกย่อยปัญหาได้อย่างสมบูรณ์ เช่นเดียวกับการจัดเป็น 7 กลุ่มช่วยให้สามารถจำแนกปัญหาได้อย่างเฉพาะเจาะจงมากขึ้น โดยแยกปัญหาเกี่ยวกับเส้นทางเดินป่าและสิ่งแวดล้อมภายในอุทยาน ความเท่าเทียมด้านค่าธรรมเนียม และคุณภาพของอาหารและร้านค้าภายในอุทยานออกจากกลุ่มเดิมได้ ซึ่งการเพิ่มจำนวนกลุ่มทำให้สามารถมองเห็นข้อบกพร่องที่เป็นปัญหาเชิงลึก เช่น การบำรุงรักษาสิ่งอำนวยความสะดวกและป้ายบอกทางที่อาจยังไม่มี ความชัดเจนเพียงพอ นอกจากนี้ นักท่องเที่ยวบางกลุ่มยังสะท้อนถึงความไม่คุ้มค่าของค่าธรรมเนียมเมื่อเทียบกับคุณภาพของสถานที่และบริการที่ได้รับ การจัดกลุ่มในลักษณะนี้ช่วยให้เห็นภาพรวมของความต้องการที่มีความซับซ้อนมากขึ้น แต่บางกลุ่มอาจมีขนาดเล็กเกินไปจนทำให้ขาดความสำคัญในภาพรวมของปัญหา และเมื่อจัดเป็น 10 กลุ่ม พบว่าปัญหา มีการกระจายตัวอย่างละเอียด และเฉพาะเจาะจงมากขึ้น เช่น ด้านป้ายบอกทางและระบบนำทาง ด้านความปลอดภัยของนักท่องเที่ยวจากสัตว์มีพิษหรือสัตว์ที่ลงมาหากินในพื้นที่ยามค่ำคืน และด้านระบบขนส่งสาธารณะในการเข้าถึงอุทยาน ตลอดจนยังสามารถจำแนกปัญหาด้านค่าธรรมเนียมออกเป็นหลายมิติ เช่น ค่าธรรมเนียมที่ไม่สมเหตุผล และค่าบริการเพิ่มเติมที่ไม่ได้รับแจ้งล่วงหน้า อย่างไรก็ตาม จำนวนกลุ่มที่มากอาจทำให้การสรุปผลทำได้ยากขึ้น เนื่องจากต้องพิจารณาประเด็นที่มีความซับซ้อนมากขึ้น และบางกลุ่มอาจมีเนื้อหาที่คล้ายคลึงกันมากเกินไป

4. การอภิปรายผลการวิจัย

งานวิจัยนี้จึงมุ่งเน้นการวิเคราะห์บทวิจารณ์ของนักท่องเที่ยวต่างชาติต่ออุทยานแห่งชาติเขาใหญ่ผ่านการเรียนรู้แบบไม่มีผู้สอน โดยใช้การสร้างแบบจำลองหัวข้อและการจัดกลุ่มข้อความ มีรายละเอียดดังนี้

4.1 ผลการสร้างแบบจำลองหัวข้อ

พบว่า นักท่องเที่ยวให้ความสำคัญกับประเด็นหลัก 5 ด้าน ได้แก่ 1) สถานที่ท่องเที่ยวและสิ่งแวดล้อม 2) คุณภาพของบริการและสิ่งอำนวยความสะดวก 3) ค่าใช้จ่ายและความคุ้มค่า 4) การเดินทางและการเข้าถึง และ 5) ประสบการณ์ส่วนตัวและความสนุกสนาน ซึ่งสามารถเชื่อมโยงกับกรอบการวิเคราะห์จุดหมายปลายทางการท่องเที่ยวแบบ 6A (Six As Framework for the Analysis of Tourism Destinations) ของ Buhalis (2000) ประกอบด้วยด้านสิ่งดึงดูดใจ (Attractions) ของอุทยานแห่งชาติเขาใหญ่ เช่น ทิวทัศน์ น้ำตก จุดชมวิว และเส้นทางศึกษาธรรมชาติที่มีระบบนิเวศอุดมสมบูรณ์ ต่อมาคือการเข้าถึง (Accessibility) นักท่องเที่ยวส่วนใหญ่เห็นว่าการเดินทางมายังอุทยานทำได้สะดวกโดยใช้รถยนต์ส่วนตัว อย่างไรก็ตาม มีข้อจำกัดในเรื่องของระบบขนส่งสาธารณะที่อาจทำให้นักท่องเที่ยวบางกลุ่มเข้าถึงพื้นที่ได้ยากในส่วน of สิ่งอำนวยความสะดวก (Amenities) แม้ว่าจะมีบริการพื้นฐาน เช่น ที่พัก ลานกางเต็นท์ ร้านอาหาร และ



ศูนย์บริการนักท่องเที่ยว แต่นักท่องเที่ยวบางส่วนให้ข้อเสนอแนะเกี่ยวกับคุณภาพของห้องน้ำสาธารณะ ความสะอาดของพื้นที่ และความพร้อมของระบบสาธารณูปโภค เช่น ไฟฟ้าและน้ำประปา นอกจากนี้ ยังมีข้อเรียกร้องให้พัฒนาสิ่งอำนวยความสะดวกเพิ่มเติมเพื่อรองรับกลุ่มนักท่องเที่ยวที่มีความต้องการเฉพาะ เช่น กลุ่มครอบครัวหรือผู้สูงอายุ นอกจากนี้การให้บริการแบบเป็นชุด (Available packages) และกิจกรรมการท่องเที่ยว (Activities) ที่หลากหลายตอบสนองความต้องการของกลุ่มนักท่องเที่ยวที่แตกต่างกัน ไม่ว่าจะเป็นการเดินป่า ปั่นจักรยาน หรือส่องสัตว์ โดยนักท่องเที่ยวส่วนใหญ่ให้ความเห็นเชิงบวกเกี่ยวกับโอกาสในการสัมผัสธรรมชาติอย่างใกล้ชิด และการให้บริการของแหล่งท่องเที่ยว (Ancillary services) นักท่องเที่ยวแสดงความพึงพอใจต่อบริการข้อมูลจากศูนย์บริการนักท่องเที่ยว และระบบจองที่พักออนไลน์ อย่างไรก็ตาม ยังมีข้อเสนอแนะให้เพิ่มการชำระเงินแบบอิเล็กทรอนิกส์เพื่อรองรับพฤติกรรมของนักท่องเที่ยว

4.2 ผลการจัดกลุ่มข้อความ

พบว่า นักท่องเที่ยวให้ความสนใจในประเด็นที่แตกต่างกันสามารถจำแนกได้เป็น 3 กลุ่มหลัก ได้แก่ 1) ประสบการณ์เกี่ยวกับสัตว์ป่าและบริการนำเที่ยว 2) ความสวยงามของธรรมชาติและแหล่งท่องเที่ยวภายในอุทยาน และ 3) ประเด็นเกี่ยวกับค่าธรรมเนียมและความคุ้มค่า เมื่อพิจารณาในรายละเอียด เพื่อการศึกษาปัญหาหรือข้อบกพร่องที่นักท่องเที่ยวชาวต่างชาติพบเจอในอุทยานแห่งชาติเขาใหญ่ พบว่า ความคิดเห็นเชิงลบสามารถจำแนกออกเป็น 3 กลุ่มหลัก ได้แก่ 1) ด้านค่าธรรมเนียมการเข้าชมและโครงสร้างการกำหนดราคาที่มีราคาสูงเมื่อเทียบกับสิ่งอำนวยความสะดวกและบริการที่ได้รับ นอกจากนี้ การกำหนดราคาที่แตกต่างกันระหว่างชาวไทยและชาวต่างชาติ ยังถูกมองว่าอาจส่งผลกระทบต่อ การรับรู้ความคุ้มค่า 2) ด้านการเข้าถึงและความพึงพอใจในการบริการ จากข้อจำกัดด้านระบบขนส่งสาธารณะที่ไม่ครอบคลุม ส่งผลให้ต้องใช้บริการรถบริการท้องถิ่นหรือทัวร์นำเที่ยว ซึ่งอาจมีค่าใช้จ่ายสูงและมีมาตรฐานแตกต่างกัน และ 3) ด้านสิ่งอำนวยความสะดวกและกิจกรรมการท่องเที่ยว จากข้อจำกัดในด้านความหลากหลาย โดยเฉพาะอย่างยิ่งในกลุ่มนักท่องเที่ยวที่เดินทางเป็นครอบครัวหรือผู้สูงอายุ สอดคล้องกับทฤษฎีคุณค่าโดยรวม (Overall value theory) ของ Zeithaml (1988) ที่อธิบายว่าความคุ้มค่าที่รับรู้ (Perceived value) เกิดจากการเปรียบเทียบระหว่างสิ่งที่ได้รับ (Benefits) กับสิ่งที่ต้องเสียไป (Sacrifices) ในการประเมินประสบการณ์การท่องเที่ยวจึงต้องให้ความสำคัญทั้งปัจจัยที่เป็นรูปธรรม เช่น ค่าใช้จ่ายและคุณภาพของบริการ และปัจจัยที่เป็นนามธรรม เช่น ความสะดวกในการเดินทางและความประทับใจต่อธรรมชาติ โดยองค์ประกอบของความคุ้มค่าในการท่องเที่ยว โดยเฉพาะเมื่อเปรียบเทียบระหว่างคุณค่าทางการเงิน (Economic value) ที่สะท้อนถึงการที่นักท่องเที่ยวชาวต่างชาติเปรียบเทียบค่าธรรมเนียมเข้าชมกับประสบการณ์ที่ได้รับ ซึ่งอาจเกิดความรู้สึกไม่คุ้มค่าเมื่อเทียบกับประสบการณ์ที่ได้รับกับคุณค่าทางประสบการณ์ (Experiential value) เกิดจากความประทับใจที่ได้รับจากธรรมชาติและสัตว์ป่า เช่น การเดินป่าหรือการชมสัตว์ป่าหายาก ซึ่งเป็นจุดเด่นของอุทยานแห่งชาติเขาใหญ่

5. สรุปผลการวิจัยและข้อเสนอแนะ

การดำเนินการวิจัยครั้งนี้ คณะผู้วิจัยสามารถสรุปผลการวิจัยและข้อเสนอแนะได้ ดังนี้

5.1 สรุปผลการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างแบบจำลองหัวข้อร่วมกับการจัดกลุ่มข้อความจากบทวิจารณ์ของนักท่องเที่ยวชาวต่างชาติต่ออุทยานแห่งชาติเขาใหญ่ด้วยการเรียนรู้แบบไม่มีผู้สอน รวบรวมบทวิจารณ์เกี่ยวกับอุทยานแห่งชาติเขาใหญ่จากแพลตฟอร์มออนไลน์ออนไลน์ ตั้งแต่วันที่ 1 มกราคม 2566 ถึง 31 ธันวาคม 2567 รวมทั้งสิ้น 3,491 รายการ ผ่านกระบวนการทำความสะอาดข้อมูล ก่อนจะแทนความหมายของคำ เพื่อแปลงข้อความให้อยู่ในรูปแบบเวกเตอร์เชิงตัวเลขด้วยเทคนิค TF-IDF จากนั้นสร้างแบบจำลองหัวข้อด้วยการจัดสรรทีรีเคล ประกอบด้วย 5 หัวข้อหลัก ได้แก่ 1) สถานที่ท่องเที่ยวและสิ่งแวดล้อมเป็นหัวข้อที่ได้รับการกล่าวถึงในเชิงบวกมากที่สุด ซึ่งเป็นปัจจัยสำคัญที่ดึงดูดนักท่องเที่ยว



2) คุณภาพของบริการและสิ่งอำนวยความสะดวก จากข้อเสนอแนะเกี่ยวกับการพัฒนาโครงสร้างพื้นฐานให้มีมาตรฐานที่ดีขึ้น เพื่อรองรับจำนวนนักท่องเที่ยวที่เพิ่มขึ้น 3) ค่าใช้จ่ายและความคุ้มค่าเป็นประเด็นที่มีความคิดเห็นในหลากหลายมุมมอง บางกลุ่มอาจเห็นว่าค่าธรรมเนียมอาจสูงเกินไปเมื่อเปรียบเทียบกับสิ่งอำนวยความสะดวกและบริการที่มีอยู่ 4) การเดินทาง และการเข้าถึงที่ระบบขนส่งสาธารณะมีข้อจำกัด ซึ่งอาจเป็นอุปสรรคสำหรับนักท่องเที่ยวที่ไม่มีพาหนะส่วนตัว และ 5) ประสบการณ์ส่วนตัวและความสนุกสนานจากข้อจำกัดในการเลือกทำกิจกรรมในบางช่วงเวลา เมื่อพิจารณาในรายละเอียด พบว่า การปรับค่าพารามิเตอร์ ส่งผลโดยตรงต่อความเฉพาะเจาะจงของคำภายในแต่ละหัวข้อ โดยการกำหนดค่าที่ต่ำช่วยให้ โมเดลสามารถระบุคำที่มีลักษณะเฉพาะของแต่ละกลุ่มหัวข้อได้ชัดเจนยิ่งขึ้น ในทางกลับกัน จำนวนหัวข้อที่มากเกินไป อาจทำให้เกิดความซ้ำซ้อนระหว่างหัวข้อหรือทำให้ข้อมูลกระจายตัวมากเกินไปจนยากต่อการตีความ ดังนั้นการเลือกจำนวนหัวข้อ และกำหนดค่าพารามิเตอร์อย่างเหมาะสมจึงมีความสำคัญอย่างยิ่งในการเพิ่มประสิทธิภาพของการวิเคราะห์เนื้อหา

นอกจากนั้น ผลการจัดกลุ่มข้อความแบบเคมีนออกเป็น 3 กลุ่มหลัก ได้แก่ 1) ด้านค่าธรรมเนียมการเข้าชมและ โครงสร้างการกำหนดราคา 2) ด้านการเข้าถึงและความพึงพอใจในการบริการ และ 3) ด้านสิ่งอำนวยความสะดวกและ กิจกรรมการท่องเที่ยว โดยทุกกลุ่มมีการกระจุกตัวและทิศทางกระจายที่ต่างกันอย่างชัดเจน ซึ่งบ่งชี้ถึงลักษณะของเนื้อหา ภายในกลุ่มที่มีความเฉพาะเจาะจงในตัวเอง เมื่อพิจารณารูปแบบการกระจายของข้อมูลแต่ละกลุ่ม พบว่า ด้านการเข้าถึงและความพึงพอใจในการบริการมีการกระจายตัวกว้างมากกว่ากลุ่มอื่น และแผ่กระจายออกไปในหลายทิศทาง ซึ่งอาจสะท้อนถึง ความคิดเห็นที่หลากหลายภายในกลุ่มที่ครอบคลุมหลายด้าน ในขณะที่ด้านค่าธรรมเนียมการเข้าชมและโครงสร้างการกำหนด ราคา และด้านสิ่งอำนวยความสะดวกและกิจกรรมการท่องเที่ยวนั้นมีลักษณะการกระจุกตัวหนาแน่นกว่า สื่อให้เห็นถึงความ คล้ายคลึงกันของข้อมูลภายในที่อาจมุ่งเน้นไปยังประเด็นใดประเด็นหนึ่งเป็นหลัก นอกจากนี้ยังพบว่า ความคิดเห็นเชิงลบใน ด้านค่าธรรมเนียมการเข้าชมและโครงสร้างการกำหนดราคามีแนวโน้มอยู่ห่างจากกลุ่มอื่นอย่างเด่นชัด ขณะที่ด้านการเข้าถึง และความพึงพอใจในการบริการ และด้านสิ่งอำนวยความสะดวกและกิจกรรมการท่องเที่ยว ยังมีบางจุดที่แทรกอยู่ใกล้กัน ซึ่งอาจเป็นตัวสะท้อนถึงความสัมพันธ์หรือความทับซ้อนของเนื้อหาในบางประเด็น แม้จะอยู่คนละกลุ่มก็ตาม

จะเห็นได้ว่าผลการวิจัยนี้สามารถสกัดข้อมูลเชิงลึกที่เป็นประโยชน์ต่อการวางแผนและพัฒนาการจัดการแหล่งท่องเที่ยว โดยเฉพาะในบริบทของการท่องเที่ยวเชิงธรรมชาติในประเทศไทย ช่วยให้สามารถมองเห็นแนวโน้มของปัญหาที่เกิดขึ้นในมุมมองของนักท่องเที่ยวชาวต่างชาติได้อย่างเป็นระบบ ซึ่งสามารถนำไปใช้เป็นข้อมูลพื้นฐานในการพัฒนาการ ให้บริการให้สอดคล้องกับความคาดหวังของนักท่องเที่ยว และช่วยเพิ่มระดับความพึงพอใจจากการท่องเที่ยวในอนาคต

5.2 ข้อเสนอแนะการวิจัย

การเลือกแบบจำลองสำหรับการวิเคราะห์ข้อความถือเป็นปัจจัยสำคัญที่ส่งผลต่อประสิทธิภาพของผลลัพธ์โดยตรง แม้การจัดสรรทรัพยากรจะเป็นวิธีที่ใช้กันอย่างแพร่หลาย แต่ยังมีข้อจำกัดในด้านการจับความสัมพันธ์เชิงบริบทของคำภายใน ประโยค ในทางกลับกัน โมเดลเชิงภาษาที่พัฒนาบนสถาปัตยกรรม Transformer เช่น BERT (Bidirectional Encoder Representations from Transformers) หรือ RoBERTa (Robustly Optimized BERT Pretraining Approach) สามารถ เรียนรู้บริบทของคำภายในประโยคได้อย่างมีประสิทธิภาพ โดยอาศัยกลไก Self-attention ที่สามารถพิจารณาความสัมพันธ์ ระหว่างคำทั้งก่อนหน้าและถัดไปภายในข้อความเดียวกัน นอกจากนี้ ยังสามารถปรับแต่ง (Fine-tuning) ให้สามารถวิเคราะห์ ข้อความเฉพาะทาง ได้อย่างเหมาะสม อย่างไรก็ตาม การเลือกใช้ครุพิจารณาทั้งในด้านวัตถุประสงค์ในการวิเคราะห์และ ข้อจำกัดด้านทรัพยากร เพื่อให้ได้ผลลัพธ์ที่เหมาะสมที่สุดกับลักษณะของข้อมูลและเป้าหมายของงานวิจัย



6. เอกสารอ้างอิง

- Ali, T., Omar, B., & Soulimane, K. (2022). Analyzing tourism reviews using an LDA topic-based sentiment analysis approach. *MethodsX*, 9, 101894.
- Blei, DM., Ng, AY., & Jordan, MI. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(2003), 993–1022.
- Buhalis, D. (2000). Marketing the competitive destination of the future. *Tourism Management*, 21(1), 97–116
- George, OA., & Ramos, CMQ. (2024). Sentiment analysis applied to tourism: Exploring tourist-generated content in the case of a wellness tourism destination. *International Journal of Spa and Wellness*, 7(2), 139–161.
- Kozak, M., & Kozak, N. (2018). *Tourist behavior* (1st ed.). Springer International Publishing.
- Kumar, N., Sousa, BB., & Sharma, S. (2022). *Tourist behavior: Past, present, and future*. CRC Press/Apple Academic Press.
- Ministry of Tourism and Sports. (2024). Preliminary statistics on international tourists entering Thailand, Ministry of Tourism and Sports. Retrieved from <https://www.mots.go.th/news/category/759>. (in Thai)
- National Statistical Office, Ministry of Digital Economy and Society. (2022). Number of tourists in national parks, fiscal year 2022. Retrieved from <https://directory.gdcatalog.go.th/Dataset/Content/83aedce9-4cc1-4127-9d44-1b2f7c423add>. Accessed ??? (in Thai)
- Nawawi, I., Ilmawan, KF., Maarif, MR., & Syafrudin, M. (2024). Exploring tourist experience through online reviews using aspect-based sentiment analysis with zero-shot learning for hospitality service enhancement. *Information*, 15(8), 499.
- Soltani-Nejad, N., Rastegar, R., Shahriari-Mehr, G., & Taheri-Azad, F. (2024). Conceptualizing tourist journey: Qualitative analysis of tourist experiences on TripAdvisor. *Journal of Quality Assurance in Hospitality & Tourism*, 25(2), 343-364.
- Suntornteerasut, P., Darawong, C., & Wongvedvanij, R. (2024). The impact of travel experience on tourist satisfaction and intention to revisit towards wellness tourism in Thailand. *Journal of Business, Innovation and Sustainability*, 19(4), 160-179. (in Thai)
- Wonglapas, S. & Rakha, C. (2024). Structural equation modeling analysis of service quality affecting tourist loyalty in the Mekong River lifestyle-based tourism development area for sustainable grassroots economy development. *Southern Technology Journal*, 17(1), 14–26. (in Thai)
- Zeithaml, VA. (1988). Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence. *Journal of Marketing*, 52(3), 2–22.