

การวิเคราะห์บทวิจารณ์จากนักท่องเที่ยวชาวต่างชาติด้วยการประมวลผลภาษาธรรมชาติ และเครื่องมือวิเคราะห์ความรู้สึกแบบใช้พจนานุกรมเป็นฐาน

จักรินทร์ สันติรัตนภักดี^{1*} ศุภกฤษฎี นีวัฒนากุล²

^{1*}สาขาวิชาเทคโนโลยีธุรกิจดิจิทัล คณะบริหารธุรกิจ มหาวิทยาลัยวงษ์ชวลิตกุล, นครราชสีมา, ประเทศไทย

²สำนักวิทยาศาสตร์และศิลปคดีดิจิทัล มหาวิทยาลัยเทคโนโลยีสุรนารี, นครราชสีมา, ประเทศไทย

*ผู้ประพันธ์บรรณกิจ อีเมล : chakkarin_san@vu.ac.th

รับต้นฉบับ : 30 มกราคม 2568; รับประทานฉบับแก้ไข : 14 มีนาคม 2568; ตอบรับบทความ : 31 มีนาคม 2568

เผยแพร่ออนไลน์ : 26 ธันวาคม 2568

บทคัดย่อ

พฤติกรรมของนักท่องเที่ยวในปัจจุบันคือการค้นหาข้อมูลออนไลน์ประกอบการตัดสินใจท่องเที่ยว แม้บทวิจารณ์ออนไลน์จะมีข้อดีหลายประการ แต่การวิเคราะห์ข้อมูลจำนวนมากนั้นย่อมสิ้นเปลืองเวลาและทรัพยากรในการแยกข้อมูลที่สำคัญ ดังนั้นการวิเคราะห์ข้อความอย่างเป็นระบบจึงช่วยให้สามารถนำเสนอสารสนเทศได้ครอบคลุมและตรงประเด็นมากขึ้น งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์บทวิจารณ์ด้วยเครื่องมือวิเคราะห์ความรู้สึกจากข้อความแบบใช้พจนานุกรม และเพื่อสร้างรายงานเชิงวิเคราะห์บทวิจารณ์จากนักท่องเที่ยวต่างชาติต่ออุทยานแห่งชาติเขาใหญ่ โดยรวบรวมบทวิจารณ์เกี่ยวกับอุทยานแห่งชาติเขาใหญ่จากแพลตฟอร์มออนไลน์ รวมทั้งสิ้น 12,035 รายการ โดยใช้ TextBlob, Flair และ VADER จำแนกความคิดเห็นออกเป็นเชิงบวก เชิงลบ และเป็นกลาง ผลการทดสอบพบว่า VADER มีค่าความถูกต้องเฉลี่ยสูงที่สุดที่ 76% อย่างไรก็ตาม จากการวิเคราะห์ความรู้สึก ยังพบบทวิจารณ์ที่คะแนนความพึงพอใจขัดแย้งกับเนื้อหา ซึ่งแสดงถึงข้อจำกัดของการใช้ผลการวิเคราะห์ความรู้สึกเพียงอย่างเดียว ในการสะท้อนความคิดเห็น เนื่องจากความคิดเห็นมีความซับซ้อน และยากที่จะเปรียบเทียบความแตกต่างด้วยมาตรฐานเดียวกัน เพื่อแก้ไขปัญหาดังกล่าวจึงสร้างรายงานเชิงวิเคราะห์จากเปรียบเทียบความสัมพันธ์ระหว่างคำสำคัญและบทวิจารณ์ที่เกี่ยวข้องด้วยการวัดความคล้ายคลึงโคไซน์ และสรุปข้อความด้วยการประมวลผลภาษาธรรมชาติร่วมกับการนำเสนอภาพข้อมูล เพื่อสะท้อนถึงจุดเด่นและข้อจำกัดของอุทยาน ผลการวิเคราะห์ข้อมูล พบว่าแม้ว่าอุทยานแห่งชาติเขาใหญ่จะได้รับการชื่นชมในด้านความงามทางธรรมชาติ ความอุดมสมบูรณ์ของทรัพยากร และความหลากหลายทางชีวภาพ แต่ก็ยังมีข้อจำกัดบางประการที่ถูกกล่าวถึงในบทวิจารณ์ ได้แก่ ค่าธรรมเนียมที่แตกต่างกันระหว่างนักท่องเที่ยวชาวไทยและชาวต่างชาติ นอกจากนี้ ระบบขนส่งสาธารณะยังขาดความสะดวกและไม่เพียงพอ รวมถึงต้องการประสบการณ์การท่องเที่ยวที่คุ้มค่าและกิจกรรมที่เหมาะสมกับนักท่องเที่ยวทุกกลุ่ม ผลลัพธ์จากงานวิจัยยังเป็นหนึ่งในข้อมูลพื้นฐานในการตัดสินใจของนักท่องเที่ยว อีกทั้งยังเป็นประโยชน์ต่อการกำหนดแนวทาง ในการพัฒนาอุทยานแห่งชาติเขาใหญ่ให้ตอบสนองความต้องการของนักท่องเที่ยว นอกจากนี้ ยังเป็นต้นแบบในการประยุกต์ใช้กับแหล่งท่องเที่ยวอื่น ๆ เพื่อยกระดับคุณภาพการให้บริการและเสริมสร้างศักยภาพในการแข่งขันของอุตสาหกรรมการท่องเที่ยวของประเทศไทยในระยะยาว

คำสำคัญ : เครื่องมือวิเคราะห์ความรู้สึกแบบใช้พจนานุกรมเป็นฐาน การประมวลผลภาษาธรรมชาติ การวิเคราะห์ความรู้สึก การวิเคราะห์ข้อความ



Analysis of Foreign Tourist Review by Natural Language Processing and a Lexicon-Based Sentiment Analysis Tool

Chakkarin Santirattanaphakdi^{1*} Suphakit Niwattanakul²

^{1*}*Digital Business Technology Program, Faculty of Business Administration, Vongchavalitkul University,
Nakhon Ratchasima, Thailand*

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

*Corresponding Author. E-mail address: chakkarin_san@vu.ac.th

Received: 30 January 2025; Revised: 14 March 2025; Accepted: 31 March 2025

Published online: 26 December 2025

Abstract

The current behavior of tourists is to search for online information to support travel decisions. Although online reviews have many advantages, analyzing large amounts of data is time-consuming and resource-intensive in extracting important information; therefore, systematic text analysis helps present more comprehensive and targeted information. This research aims to analyze reviews using dictionary-based sentiment analysis tools and to create an analytical report of reviews from foreign tourists toward Khao Yai National Park. A total of 12,035 reviews related to Khao Yai National Park were collected from online platforms. TextBlob, Flair, and VADER were used to classify opinions as positive, negative, or neutral. The results showed that VADER had the highest average accuracy at 76%. However, sentiment analysis also found reviews in which satisfaction scores conflicted with the content, showing the limitation of using sentiment analysis alone to reflect opinions, as opinions are complex and difficult to compare using a single standard. To address this issue, an analytical report was created by analyzing the relationships between key terms and related reviews using cosine similarity and summarizing the text using natural language processing with data visualization to reflect the strengths and limitations of the park. The analysis found that although Khao Yai National Park is praised for its natural beauty, resource richness, and biodiversity, some limitations are mentioned in the reviews, including different fees for Thai and foreign tourists, inconvenient and insufficient public transportation, and the need for valuable tourism experiences and activities suitable for all tourist groups. The results of this research provide basic information for tourist decision-making and are useful for developing guidelines to improve Khao Yai National Park to better meet tourists' needs, and can also be applied as a model for other tourist destinations to enhance service quality and the competitiveness of Thailand's tourism industry in the long term.

Keywords: Lexicon-based sentiment analysis tool, Natural language processing, Sentiment analysis, Text analysis

1) บทนำ

พัฒนาการของเทคโนโลยีสารสนเทศ และการสื่อสารส่งผลให้เกิดการเปลี่ยนแปลงพฤติกรรมการใช้ชีวิตยุคปัจจุบันไปอย่างสิ้นเชิงไม่เว้นแม้แต่พฤติกรรมกาท่องเที่ยวที่ผู้ใช้มักค้นหาข้อมูลผ่านสื่อออนไลน์ต่าง ๆ ที่นำเสนอข้อมูลอย่างหลากหลาย เพื่อสนับสนุนการท่องเที่ยว โดยเฉพาะคะแนนความพึงพอใจ (rating) ที่ได้กลายเป็นเครื่องมือสำคัญในการช่วยตัดสินใจเลือกสถานที่ท่องเที่ยว [1] จากประสบการณ์ของผู้ที่เคยเดินทางไปยังสถานที่แห่งนั้นมาแล้ว ซึ่งเพิ่มความมั่นใจในการวางแผนการท่องเที่ยวให้คุ้มค่าและปลอดภัยยิ่งขึ้น

อย่างไรก็ตาม การใช้คะแนนความพึงพอใจเพียงอย่างเดียวในการประเมินคุณภาพสถานที่ท่องเที่ยวนั้นยังคงมีปัญหาหลายประการ เนื่องจากคะแนนที่เป็นตัวเลขนั้นไม่สามารถบ่งบอกถึงข้อดีหรือข้อเสียที่แตกต่างกันในหลากหลายมุมมอง จึงไม่อาจสะท้อนถึงคุณภาพที่แท้จริงได้อย่างครบถ้วน [2] อีกทั้งความคาดหวังของนักท่องเที่ยวแต่ละคนนั้นไม่มีขีดจำกัด และยากที่จะเปรียบเทียบกันได้ด้วยหลักเกณฑ์ที่เป็นมาตรฐานเดียวกัน เพื่อแก้ไขปัญหาดังกล่าว จำเป็นต้องมีการวิเคราะห์ความรู้สึกจากบทวิจารณ์ (review) ที่อยู่ในรูปแบบภาษาธรรมชาติ เพื่อจำแนกออกเป็นความคิดเห็นเชิงบวก เชิงลบ และเป็นกลาง [3] หากแต่การวิเคราะห์ความรู้สึกจากข้อความเพียงอย่างเดียวอาจมีอคติจากผู้วิจารณ์ที่มีประสบการณ์สูงต่างทั้งเชิงบวกและเชิงลบ [4] ซึ่งทำให้เกิดความเอนเอียงและไม่ได้สะท้อนภาพรวมของสถานที่ท่องเที่ยวอย่างตรงไปตรงมา ส่งผลให้ข้อมูลที่เผยแพร่ขาดความน่าเชื่อถือและไม่เป็นกลาง ตลอดจนปริมาณของบทวิจารณ์จำนวนมากในปัจจุบัน ทำให้เกิดภาวะข้อมูลท่วมท้น (information overload) จนไม่สามารถแยกข้อมูลที่สำคัญได้ เนื่องจากไม่มีเครื่องมือหรือเทคนิคในการคัดกรองข้อมูลที่สำคัญออกมา ข้อมูลจำนวนมากทั้งที่เกี่ยวข้องและไม่เกี่ยวข้อง หรือข้อมูลที่ไม่จำเป็นต่าง ๆ อาจทำให้ผู้ใช้รู้สึกสับสนหรือ ไม่สามารถตัดสินใจได้อย่างเหมาะสม อีกทั้งสิ้นเปลืองเวลาและทรัพยากรในการแยกข้อมูลที่สำคัญ

หนึ่งในสถานที่ท่องเที่ยวที่ได้รับความนิยมมากที่สุดแห่งหนึ่งของประเทศไทยคืออุทยานแห่งชาติเขาใหญ่ ซึ่งเป็นพื้นที่มรดกโลกทางธรรมชาติ (World Heritage Site) และอุทยานมรดกแห่งอาเซียน (ASEAN Heritage Park) ครอบคลุมพื้นที่ 4 จังหวัด ประกอบด้วยนครราชสีมา ปราจีนบุรี นครนายก และสระบุรี ด้วยพื้นที่เกือบ 2,206 ตารางกิโลเมตร และเปิดต้อนรับนักท่องเที่ยวตลอดทั้งปี ในปี พ.ศ. 2565 มีจำนวนนักท่องเที่ยว

ถึง 1,428,765 คน แบ่งเป็นนักท่องเที่ยวชาวไทย 1,392,078 คน และชาวต่างชาติ 36,687 คน [5] ในปี พ.ศ. 2566 จำนวนของนักท่องเที่ยวเพิ่มเป็น 1,486,532 คน [6] โดยมีจำนวนเพิ่มขึ้นทั้งนักท่องเที่ยวชาวไทยและชาวต่างชาติ และมีแนวโน้มที่จะเพิ่มขึ้นเรื่อย ๆ ในอนาคต ด้วยธรรมชาติที่อุดมสมบูรณ์และความหลากหลายทางชีวภาพ จึงทำให้นักท่องเที่ยวจำนวนมากเดินทางมาเยี่ยมชมและแบ่งปันประสบการณ์ผ่านการเขียนบทวิจารณ์สาธารณะบนแพลตฟอร์มออนไลน์จำนวนมาก ทำให้ยากต่อการนำมาประมวลผลเพื่อใช้ประโยชน์ อีกทั้งมุมมองที่แตกต่างกันของนักท่องเที่ยวชาวไทยและชาวต่างชาติที่อาจมีความคาดหวังต่อประสบการณ์ที่แตกต่างกันของแหล่งท่องเที่ยว เช่นเดียวกับมุมมองของนักท่องเที่ยวชาวไทยอาจส่งผลทั้งทางตรงและทางอ้อมต่อความพึงพอใจในแหล่งท่องเที่ยวในประเทศอย่างมีนัยสำคัญ การนำเสนอข้อมูลจากทั้งสองกลุ่มนี้ในรูปแบบเดียวกันอาจทำให้การประเมินคุณภาพของสถานที่ท่องเที่ยวไม่ครอบคลุมทุกมิติ ดังนั้นการนำเสนอข้อมูลที่หลากหลายจากบทวิจารณ์ของนักท่องเที่ยวต่างชาติจะช่วยให้ผู้ที่เกี่ยวข้องเข้าใจความต้องการในมุมมองที่ครอบคลุมและนำไปสู่การพัฒนาคุณภาพให้ตรงตามความคาดหวังของนักท่องเที่ยวได้อย่างยั่งยืน

จากที่กล่าวมา คณะผู้วิจัยจึงได้เห็นถึงความสำคัญในการพัฒนากระบวนการวิเคราะห์ข้อมูลจากบทวิจารณ์ด้วยการประมวลผลภาษาธรรมชาติ และเครื่องการวิเคราะห์ความรู้สึกจากข้อความแบบใช้พจนานุกรม เพื่อนำเสนอข้อมูลที่ชัดเจนและตรงประเด็น นอกจากนั้นยังสะท้อนภาพรวมในแต่ละมิติของแหล่งท่องเที่ยวได้อย่างสมบูรณ์และเป็นธรรม อันจะเป็นหนึ่งในข้อมูลพื้นฐานในการตัดสินใจของนักท่องเที่ยว

2) ทบทวนวรรณกรรม

2.1) การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP)

การประมวลผลภาษาธรรมชาติ เป็นการผสมผสานหลายศาสตร์เพื่อช่วยให้คอมพิวเตอร์สามารถเข้าใจและวิเคราะห์ข้อมูลในรูปแบบภาษาของมนุษย์ได้อย่างมีประสิทธิภาพ [7] ด้วยการสร้างระบบที่สามารถตีความข้อมูลเชิงภาษา ไม่ว่าจะเป็นข้อความหรือคำพูด และตอบสนองต่อมนุษย์ได้อย่างเหมาะสม

ปัจจุบันมีการประยุกต์ใช้เทคนิคการเรียนรู้เชิงลึก (deep learning) เพื่อเพิ่มประสิทธิภาพและความแม่นยำในการประมวลผลภาษาธรรมชาติ เช่น การแปลภาษาอัตโนมัติ (machine translation)

การสรุปความ (text summarization) เพื่อช่วยให้ผู้อ่านสามารถรับข้อมูลสำคัญได้อย่างรวดเร็ว และการรู้จำเสียงพูด (speech recognition) ซึ่งถูกนำไปใช้ในระบบผู้ช่วยอัจฉริยะ รวมถึงการวิเคราะห์ความรู้สึกที่มักถูกใช้ประเมินความคิดเห็นของผู้ใช้บนสื่อสังคมออนไลน์

2.2) การวิเคราะห์ข้อความ (Text Analysis)

การวิเคราะห์ข้อความ คือการประมวลผลเพื่อแปลความหมายจากข้อความด้วยเทคนิคและวิธีการทางคอมพิวเตอร์ [8] เพื่อนำไปใช้ในการสกัดข้อมูล การจัดหมวดหมู่ การสรุปเนื้อหา หรือการตรวจหาความสัมพันธ์ในข้อมูล แบ่งเป็นการวิเคราะห์ข้อความเชิงโครงสร้าง (structured text analysis) จากข้อมูลที่มีรูปแบบอย่างชัดเจน เช่น ข้อความในฐานข้อมูล หรือในเอกสารที่มีการกำหนดรูปแบบการนำเสนอ และการวิเคราะห์ข้อความไร้โครงสร้าง (unstructured text analysis) เช่น บทวิจารณ์ บทความ ข่าวสาร หรือความคิดเห็นของผู้ใช้บนสื่อสังคมออนไลน์ ซึ่งนำเสนออย่างอิสระ และไม่มีมาตรฐาน เนื่องจากอยู่ในรูปแบบของภาษาธรรมชาติ

2.3) การวิเคราะห์ความรู้สึก (Sentiment Analysis)

การวิเคราะห์ความรู้สึก เป็นกระบวนการที่เกี่ยวข้องกับการตีความและประเมินความคิดเห็นหรืออารมณ์จากข้อความในรูปแบบภาษาธรรมชาติ [3] เพื่อแยกแยะและประเมินความรู้สึกที่แสดงออกในข้อความนั้น ๆ ว่าเป็นเชิงบวก (positive) เชิงลบ (negative) หรือเป็นกลาง (neutral) จากการพิจารณาจากลักษณะของคำศัพท์ที่ใช้ในข้อความ โครงสร้างประโยค และการใช้สำนวนเฉพาะที่สะท้อนถึงอารมณ์ที่มีต่อเหตุการณ์หรือประเด็นต่าง ๆ ที่กล่าวถึง

งานวิจัยในปัจจุบันที่มุ่งวิเคราะห์ความรู้สึกจากข้อความบทวิจารณ์ ข้อความจากสื่อสังคมออนไลน์ หรือบทความข่าว สามารถจำแนกได้ 3 แนวทาง ประกอบด้วย 1) การวิเคราะห์ขั้วของความรู้สึก (polarity-based sentiment analysis) โดยจำแนกความรู้สึกออกเป็นเชิงบวก เชิงลบ และเป็นกลาง ซึ่งเป็นแนวทางที่ใช้กันอย่างแพร่หลายในปัจจุบัน เนื่องจากจุดเด่นที่ความเรียบง่ายและใช้งานกับข้อมูลขนาดใหญ่ได้อย่างมีประสิทธิภาพ เนื่องจากสามารถใช้พจนานุกรมอารมณ์สำเร็จรูปในการวิเคราะห์ข้อความได้โดยไม่ต้องพัฒนาโมเดลที่ซับซ้อน นอกจากนี้ยังง่ายต่อการทำความเข้าใจและประยุกต์ใช้ได้อย่างหลากหลาย แต่ยังคงมีข้อจำกัดคือความแม่นยำ และอาจเกิดข้อผิดพลาดเมื่อต้อง

วิเคราะห์ข้อความที่ซับซ้อน เช่น การใช้ถ้อยคำเสียดสี (sarcasm) หรือการใช้คำในเชิงลบที่มีความหมายเชิงบวกในประโยค 2) การวิเคราะห์ระดับความเข้มของความรู้สึก (intensity-based sentiment analysis) ซึ่งเป็นกรวิเคราะห์ที่มุ่งเน้นการวัดระดับของความรู้สึกในระดับที่ละเอียดมากขึ้น เช่น ความรู้สึกในระดับบวกมาก บวก เป็นกลาง ลบ และลบมาก แทนที่จะจำแนกเป็นเพียงเชิงบวกหรือเชิงลบเพียงอย่างเดียว ซึ่งจะสะท้อนถึงความรู้สึกที่ซ่อนอยู่ในข้อความได้มากขึ้น เหมาะสำหรับการวิเคราะห์ข้อมูลที่ต้องการติดตามความคิดเห็นในแต่ละช่วงเวลา หรือวิเคราะห์แนวโน้มของข้อความ อย่างไรก็ตาม การใช้งานจำเป็นต้องใช้ชุดข้อมูลที่มีการระบุระดับของความรู้สึกอย่างละเอียด นอกจากนั้นความแม่นยำของโมเดลขึ้นอยู่กับบริบทของข้อความ ซึ่งอาจทำให้เกิดข้อผิดพลาดในกรณีที่มีระดับความรู้สึกไม่ครอบคลุมทุกสถานการณ์ และ 3) การวิเคราะห์ประเภทของอารมณ์ (emotion-based sentiment analysis) โดยมุ่งเน้นการจำแนกความรู้สึกตามอารมณ์พื้นฐานของมนุษย์ในระดับที่เฉพาะเจาะจง เช่น สุข (joy) โศกเศร้า (sadness) โกรธ (anger) กลัว (fear) ประหลาดใจ (surprise) และขยะแขยง (disgust) เป็นต้น เนื่องจากสามารถระบุอารมณ์ที่เฉพาะเจาะจงได้มากกว่า จึงเหมาะสำหรับการวิเคราะห์เชิงจิตวิทยา หรือการประเมินภาวะสุขภาพจิตผ่านความคิดเห็นในสื่อสังคมออนไลน์ อย่างไรก็ตาม ข้อเสียของแนวทางนี้คือความซับซ้อนในการจัดเตรียมชุดข้อมูลที่มีป้ายกำกับอารมณ์ (emotion labeled) อย่างละเอียด และบางอารมณ์อาจมีความคล้ายคลึงกันจนยากต่อการแยกแยะ นอกจากนั้นต้องสร้างโมเดลขนาดใหญ่และต้องการทรัพยากรการประมวลผลสูง จึงมีข้อจำกัดในการใช้งานกับข้อมูลขนาดใหญ่ หรืออุปกรณ์ที่มีข้อจำกัดด้านทรัพยากร

กระบวนการวิเคราะห์ขั้วของความรู้สึกเริ่มจากการเก็บรวบรวมข้อมูลจากแหล่งต่าง ๆ จากนั้นจึงเข้าสู่ขั้นตอนการเตรียมข้อมูลก่อนการประมวลผล (preprocessing) ซึ่งรวมถึงการลบอักขระพิเศษ การแปลงตัวอักษรเป็นตัวพิมพ์เล็ก การตัดคำ (tokenization) การปรับคำให้อยู่ในรูปแบบพื้นฐาน (lemmatization or stemming) และการลบคำหยุดหรือคำที่ไม่มีความหมายสำคัญ (stopword removal) ก่อนจะใช้การวิเคราะห์ความรู้สึกด้วยวิธีการต่าง ๆ ได้แก่ 1) การใช้พจนานุกรมอารมณ์ (lexicon-based approach) ที่มีการจัดอันดับความสำคัญของคำต่าง ๆ ไว้ในคลังคำศัพท์ [3] พร้อมด้วย ค่าความเข้มข้นของความรู้สึกหรืออารมณ์ (sentiment lexicon) เช่น เครื่องมือวิเคราะห์ความรู้สึกที่นิยมใช้กันมากใน

ปัจจุบัน เช่น TextBlob, Flair หรือ VADER ในการจับคู่คำศัพท์ และให้คะแนนความรู้สึก 2) การเรียนรู้ของเครื่อง (machine learning-based approach) [3] โดยอาศัยการสกัดคุณลักษณะ (feature extraction) จากการแปลงข้อความเป็นเวกเตอร์ เช่น TF-IDF ที่คำนึงถึงความสำคัญของคำในเอกสาร โดยคำที่มีความถี่สูงแต่ไม่สำคัญจะได้รับค่าน้ำหนักต่ำ รวมถึงการฝังคำ เช่น Word2Vec หรือ GloVe ที่แปลงคำให้เป็นเวกเตอร์ที่มีความหมาย ซึ่งสามารถจับความสัมพันธ์ระหว่างคำที่มีความหมายคล้ายกันได้ ตลอดจนเทคนิค one-hot encoding ซึ่งแปลงคำให้เป็นเวกเตอร์ที่มีมิติเท่ากับจำนวนคำทั้งหมดในพจนานุกรม โดยที่ทุกตำแหน่งในเวกเตอร์จะเป็นศูนย์ ยกเว้นตำแหน่งที่ตรงกับคำที่กำลังแปลงจะมีค่าเป็น 1 ก่อนจะใช้อัลกอริทึมการเรียนรู้ของเครื่องทั้งแบบมีผู้สอนและไม่มีผู้สอน เช่น ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) [9] นาอิวเบย์ (Naive Bayes) หรือการจัดกลุ่มแบบเคมีน (k-mean clustering) [10] และ 3) การเรียนรู้เชิงลึก (deep learning-based approach) [11] จากการนำโครงข่ายประสาทเทียมมาประยุกต์ใช้ เช่น โครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network: RNN) หน่วยความจำระยะสั้นแบบยาว (Long Short-Term Memory: LSTM) และโมเดลการเรียนรู้เชิงลึกด้านภาษาอย่าง transformer-based models [12] เช่น BERT หรือ DistilBERT ที่แยกแยะบริบทของคำศัพท์และความหมายแฝงในข้อความได้ดียิ่งขึ้น ทำให้สามารถระบุประเด็นที่สำคัญได้อย่างมีประสิทธิภาพ

เมื่อพิจารณาแต่ละแนวทางจะพบว่า มีข้อดีและข้อเสียแตกต่างกันไป ดังนั้นการเลือกใช้แต่ละวิธีการจึงขึ้นอยู่กับลักษณะของข้อมูล ความแม่นยำที่ต้องการ และทรัพยากรการประมวลผล โดยโมเดลการเรียนรู้ของเครื่องนั้นมีข้อดีในด้านความเร็วและความง่ายในการฝึกฝนโมเดล แต่ในขณะเดียวกันก็มีข้อจำกัดเมื่อข้อมูลที่เข้ามีความซับซ้อนหรือมีลำดับคำที่สำคัญชัดเจน ในทางกลับกันการใช้โมเดลการเรียนรู้เชิงลึกที่สามารถจัดการกับข้อมูลลำดับได้ดีนั้น [13] แต่ยังมีข้อเสียในด้านความซับซ้อนของโมเดล อีกทั้งการฝึกฝนโมเดลที่ใช้เวลานาน และต้องการทรัพยากรการคำนวณสูง จึงอาจทำให้เกิดข้อจำกัดในกรณีที่ต้องการ การประมวลผลแบบเรียลไทม์ เป็นที่มาของการเลือกใช้เครื่องมือวิเคราะห์ความรู้สึกแบบใช้พจนานุกรมบนเงื่อนไขที่ต้องการเน้นความเร็วในการประมวลผลและใช้ทรัพยากรคำนวณต่ำ แต่ยังสามารถวิเคราะห์ความรู้สึกจากข้อความได้อย่างแม่นยำ

2.4) เครื่องมือวิเคราะห์ความรู้สึกแบบใช้พจนานุกรม (Lexicon-Based Sentiment Analysis Tool)

เครื่องมือวิเคราะห์ความรู้สึกจากข้อความแบบใช้พจนานุกรมส่วนใหญ่อ้างอิงจากฐานข้อมูลที่บรรจุคำศัพท์พร้อมค่าคะแนนความรู้สึก ซึ่งมีจุดเด่นคือง่ายต่อการใช้งานโดยไม่ต้องฝึกฝนโมเดลเหมือนวิธีการเรียนรู้ของเครื่อง และสามารถวิเคราะห์ข้อมูลจากข้อความในลักษณะของคำและประโยค โดยไม่จำเป็นต้องแปลงเป็นเวกเตอร์ตัวเลขก่อน โดยเครื่องมือวิเคราะห์ความรู้สึกแบบใช้พจนานุกรมที่ได้รับความนิยมในปัจจุบัน ได้แก่

2.4.1) TextBlob [14] เป็นไลบรารีที่พัฒนาขึ้นเพื่อรองรับการวิเคราะห์ความรู้สึกจากข้อความโดยใช้คลังคำศัพท์ที่กำหนดค่าคะแนนความรู้สึกที่กำหนดไว้ล่วงหน้า ทำให้สามารถระบุได้ว่าข้อความมีแนวโน้มความรู้สึกเชิงบวก เชิงลบ หรือเป็นกลางจากการใช้ pattern library ซึ่งเป็นเครื่องมือที่ใช้สำหรับการทำเหมืองข้อความและวิเคราะห์โครงสร้างของภาษา กระบวนการวิเคราะห์ความรู้สึกมีพื้นฐานบน 2 องค์ประกอบหลัก ได้แก่ ค่าความเป็นบวกหรือลบ (polarity) ถูกกำหนดอยู่ในช่วงระหว่าง -1 ถึง +1 โดยค่าที่เข้าใกล้ +1 แสดงถึงข้อความที่มีความรู้สึกเชิงบวกสูง ในทางกลับกันค่าที่เข้าใกล้ -1 หมายถึงข้อความมีความรู้สึกเชิงลบสูง ร่วมกับค่าอัตวิสัย (subjectivity) ที่แสดงถึงความคิดเห็นส่วนตัวหรือข้อเท็จจริงที่ถูกกำหนดให้อยู่ในช่วง 0 ถึง 1 โดยค่าที่เข้าใกล้ 1 แสดงถึงข้อความที่เป็นความคิดเห็นส่วนตัวมากกว่าข้อเท็จจริง ในขณะที่ค่าที่ใกล้ 0 บ่งชี้ว่าข้อความนั้นมีลักษณะเป็นกลางและมีความเป็นข้อเท็จจริงมากกว่า

ความง่ายในการใช้งานและไม่จำเป็นต้องฝึกโมเดลเพิ่มเติมก่อนนำไปใช้วิเคราะห์ความรู้สึกจากข้อความถือเป็นจุดเด่นสำคัญของ TextBlob นอกจากนี้ยังรองรับการประมวลผลภาษาธรรมชาติในด้านอื่น ๆ เช่น การจำแนกข้อความ การวิเคราะห์ไวยากรณ์ (part-of-speech tagging) การแยกคำนาม (noun phrase extraction) และการแปลภาษา โดยเฉพาะการทำงานกับความคิดเห็นที่มีการใช้ภาษาทางการในรูปแบบภาษาอังกฤษ อย่างไรก็ตาม ยังมีข้อจำกัดในการวิเคราะห์ข้อความที่มีความหมายซับซ้อน เนื่องจากอาศัยกฎทางภาษาศาสตร์และคะแนนเชิงอารมณ์ที่ถูกกำหนดล่วงหน้าเพียงอย่างเดียว เช่นเดียวกับประสิทธิภาพอาจลดลงเมื่อใช้งานกับข้อความที่มีการใช้ภาษาเชิงประชดประชัน หรือข้อความที่มีบริบทซับซ้อนที่ต้องอาศัยความเข้าใจทางสังคมและวัฒนธรรม

2.4.2) Flair [15] เป็นหนึ่งในเครื่องมือการวิเคราะห์ความรู้สึกแบบใช้พจนานุกรมด้วยเทคนิคการฝังคำตามบริบท (contextualized word embeddings) และใช้หน่วยความจำระยะสั้นแบบยาวสองทิศทาง (bidirectional long short-term memory: BiLSTM) ที่ผ่านการฝึกฝนล่วงหน้าด้วยข้อมูลจำนวนมหาศาล ซึ่งช่วยให้สามารถจับความสัมพันธ์ระหว่างคำในประโยค และทำความเข้าใจความหมายของข้อความจากทั้งสองทิศทางได้อย่างมีประสิทธิภาพ โดยในกระบวนการนี้โมเดลจะใช้การเรียนรู้จากบริบทของข้อความในประโยคก่อนหน้าและหลังเพื่อระบุความหมายของคำเป้าหมาย ทำให้สามารถประเมินอารมณ์จากข้อความได้แม่นยำยิ่งขึ้น จึงเหมาะกับการวิเคราะห์ข้อความที่มีโครงสร้างซับซ้อนหรืองานที่ต้องการตีความหมายในระดับสูง เช่น ข้อความที่ใช้ภาษาเชิงประชดประชัน เป็นต้น

จุดเด่นสำคัญของ Flair คือเป็นโมเดลที่ฝึกฝนล่วงหน้าจากหลายภาษา จึงสามารถทำงานได้อย่างหลากหลายไม่จำกัดเฉพาะภาษาอังกฤษ และสามารถนำไปใช้งานได้ทันทีโดยไม่ต้องฝึกโมเดลใหม่ จึงมักถูกนำไปใช้ในการระบุองค์ประกอบในประโยค (named entity recognition) การระบุประเภทข้อความ (text classification) การตอบคำถามอัตโนมัติ (question answering) และการสรุปความ (summarization) แต่ก็มีข้อจำกัดคือการใช้ทรัพยากรและเวลาในการประมวลผลที่มากกว่า TextBlob และ VADER จึงไม่เหมาะสมกับการทำงานบนอุปกรณ์ที่มีข้อจำกัดด้านทรัพยากร

2.4.3) Valence Aware Dictionary and Sentiment Reasoner: VADER [16] เป็นไลบรารีที่ถูกรวบรวมแบบสำหรับวิเคราะห์ความรู้สึกในข้อความภาษาธรรมชาติโดยเฉพาะ จากการประยุกต์ใช้พจนานุกรมที่มีการจัดอันดับความสำคัญของคำต่าง ๆ มากกว่า 7,500 คำ ซึ่งแต่ละคำถูกกำหนดคะแนนความรู้สึก (sentiment score) ที่สะท้อนถึงความเข้มข้นของอารมณ์หรือความรู้สึกไว้ล่วงหน้า ดังตารางที่ 1

ตารางที่ 1 : คำศัพท์และตัวอย่างค่าคะแนนความรู้สึกในพจนานุกรม

word	sentiment score
happy	+2.9
sad	-2.1
excellent	+3.2
horrible	-3.1
not bad	+1.0

จากตารางที่ 1 คำศัพท์ในพจนานุกรมที่มีกำหนดตัวอย่างคะแนนความรู้สึกจะถูกนำมาใช้วิเคราะห์ข้อความร่วมกับการกำหนดฐานกฎ (rule-based) ที่กำหนดไว้ล่วงหน้าจะช่วยเพิ่มความแม่นยำในการพิจารณาลักษณะทางภาษา เพื่อตัดสินใจว่าความรู้สึกในข้อความนั้นเป็นเชิงบวก เชิงลบ หรือเป็นกลาง แทนที่จะให้ค่าความสัมพันธ์ของความรู้สึกแต่ละคำเพียงอย่างเดียวแบบเครื่องมือวิเคราะห์ความรู้สึกจากข้อความทั่วไป

หนึ่งในการพิจารณาลักษณะทางภาษาคือการกำหนดลักษณะเฉพาะที่ส่งผลต่อค่าคะแนนอารมณ์ เช่น หากใช้ตัวอักษรพิมพ์ใหญ่ทั้งหมด เช่น “AWESOME” จะกำหนดค่าความรู้สึกรุนแรงกว่า “awesome” ที่ใช้ตัวอักษรพิมพ์เล็กทั้งหมด นอกจากนี้เครื่องหมายอัศเจรีย์ (!) ยังช่วยเพิ่มระดับความเข้มของอารมณ์ เช่น “great!!” ให้ความรู้สึกเชิงบวกมากกว่าคำว่า “great” แบบปกติ ตลอดจนสามารถพิจารณาคำขยาย (modifiers) ที่มีผลต่อระดับความเข้มของความรู้สึกหรือคำวิเศษณ์ เช่น “very extremely” และ “slightly” ที่สามารถเปลี่ยนระดับความรุนแรงของคำได้ เช่นเดียวกับ “very good” จะได้รับคะแนนเชิงบวกสูงกว่าคำว่า “good” ในทางตรงกันข้าม “not good” จะถูกลดระดับความรู้สึกเชิงบวกของคำว่า “good” ทำให้ความหมายปรับลดไปเป็นกลางหรือเชิงลบมากขึ้น เช่นเดียวกับ “not bad” จะถูกลดระดับความรู้สึกเชิงลบจากคำว่า “bad” และปรับความรู้สึกให้เข้าใกล้กับคำว่า “good” ยิ่งขึ้น

แม้ว่าจะยังมีข้อจำกัดบางประการ แต่ VADER ยังสามารถตรวจจับคำประชดประชันได้ในระดับที่น่าพอใจ โดยอาศัยบริบทของข้อความ เช่น วลี “yeah, right” ที่มักใช้ในเชิงประชดประชัน ซึ่งสามารถนำมาพิจารณาเพื่อลดโอกาสในการแปลความหมายผิดพลาด ตลอดจนความสามารถในการจัดการกับการใช้ภาษาที่ไม่เป็นทางการ เช่น ข้อความที่มีการย่อคำ การใช้ตัวอักษรแทนการสื่ออารมณ์ (emoticon) และการใช้คำเสียดสีหรือแสดงความคิดเห็นในตัวเองได้ นอกจากนี้ยังมีความเร็วในการประมวลผลและไม่ต้องการทรัพยากรคำนวณมากนัก จึงเหมาะสำหรับการใช้ในงานที่ต้องการการวิเคราะห์แบบเรียลไทม์ หรือในระบบที่มีข้อจำกัดด้านทรัพยากร ตลอดจนคงทนต่อการเปลี่ยนแปลงของข้อมูลได้โดยไม่มีข้อจำกัดเรื่องมาตรฐาน จึงเหมาะสำหรับวิเคราะห์ความรู้สึกในข้อความที่ไม่เป็นทางการ จากที่กล่าวมาจึงเป็นเหตุผลหลักที่ทำให้ VADER ถูกนำมาใช้ในการประมวลผลข้อความที่เป็นภาษาธรรมชาติใน สื่อสังคมออนไลน์อย่างแพร่หลายในปัจจุบัน

2.5) บทวิจารณ์ออนไลน์ (Online Review)

บทวิจารณ์ออนไลน์ คือความคิดเห็นที่สะท้อนมุมมองหรือประสบการณ์ที่เคยได้รับผ่านแพลตฟอร์มดิจิทัล โดยบทวิจารณ์ที่เหล่านี้มักเป็นประสบการณ์ตรงจากผู้ที่เคยเดินทางท่องเที่ยว ณ สถานที่แห่งนั้นในช่วงเวลาที่ผ่านไป ซึ่งสะท้อนถึงคุณภาพของแหล่งท่องเที่ยว ที่พัก ร้านอาหาร และบริการต่าง ๆ ที่เกี่ยวข้อง เผยแพร่บนแพลตฟอร์มออนไลน์ เช่น Google reviews, Agoda.com, TripAdvisor.com, Trip.com และ สื่อสังคมออนไลน์ต่าง ๆ ซึ่งมีบทบาทในการให้ข้อมูลและกำหนดพฤติกรรมของนักท่องเที่ยว เช่น การเปรียบเทียบที่พักจากบทวิจารณ์ออนไลน์ อย่างไรก็ตาม แม้ว่าบทวิจารณ์ออนไลน์จะมีข้อดีหลายประการ แต่ก็ยังมีข้อจำกัดที่ต้องพิจารณา เช่น ความรู้สึกส่วนตัวที่มีอิทธิพลต่อการเขียนบทวิจารณ์ ดังนั้นการวิเคราะห์ข้อมูลอย่างโปร่งใสและสร้างสรรค์จะช่วยยกระดับคุณภาพของการท่องเที่ยวได้อย่างยั่งยืน

2.6) องค์ประกอบแหล่งท่องเที่ยว

การพัฒนาแหล่งท่องเที่ยวสามารถแบ่งออกได้เป็นหลายองค์ประกอบที่ช่วยส่งเสริมให้แหล่งท่องเที่ยวนั้นมีความน่าสนใจและกระตุ้นให้เกิดการเดินทางมายังสถานที่นั้น ๆ ได้ โดยอุทยานแห่งชาติเขาใหญ่เป็นแหล่งท่องเที่ยวที่มีชื่อเสียงระดับโลก มีองค์ประกอบที่สอดคล้องกับกรอบการวิเคราะห์จุดหมายปลายทางการท่องเที่ยวแบบ 6A (6A's framework for the analysis of tourism destinations) ของ Buhalis [17] ซึ่งแบ่งออกเป็น 6 ด้าน ได้แก่ 1) สิ่งดึงดูดใจ (attractions) อุทยานแห่งชาติเขาใหญ่มีสิ่งดึงดูดใจที่โดดเด่นในหลากหลายมิติ ได้แก่ สถานที่ท่องเที่ยวเชิงธรรมชาติ เช่น น้ำตกเหวสุวัต น้ำตกเหวนรก จุดชมวิวเขาเขียว และผืนป่าอันอุดมสมบูรณ์ที่เต็มไปด้วยพืชพรรณและสัตว์ป่าหลากหลายชนิด สถานที่ท่องเที่ยวที่มนุษย์สร้าง เช่น เส้นทางศึกษาธรรมชาติและอ่างเก็บน้ำสายศร สถานที่ท่องเที่ยวเชิงศิลปวัฒนธรรม เช่น ประวัติศาสตร์และวัฒนธรรมของชุมชนรอบอุทยานที่สะท้อนถึงวิถีชีวิตและความสัมพันธ์ระหว่างมนุษย์กับธรรมชาติ และสถานที่ท่องเที่ยวเชิงชุมชนสัมพันธ์ เช่น การเยี่ยมชมตลาดพื้นเมือง หรือการเข้าร่วมกิจกรรมเชิงอนุรักษ์ธรรมชาติในพื้นที่ 2) การเข้าถึงสถานที่ท่องเที่ยว (accessibility) การเดินทางมายังอุทยานแห่งชาติเขาใหญ่ผ่านถนนสายหลักที่เชื่อมต่อกันได้หลายเส้นทาง โดยเฉพาะการเดินทางด้วยรถยนต์ส่วนบุคคล หรือรถโดยสารสาธารณะ นอกจากนี้การลงทะเบียน

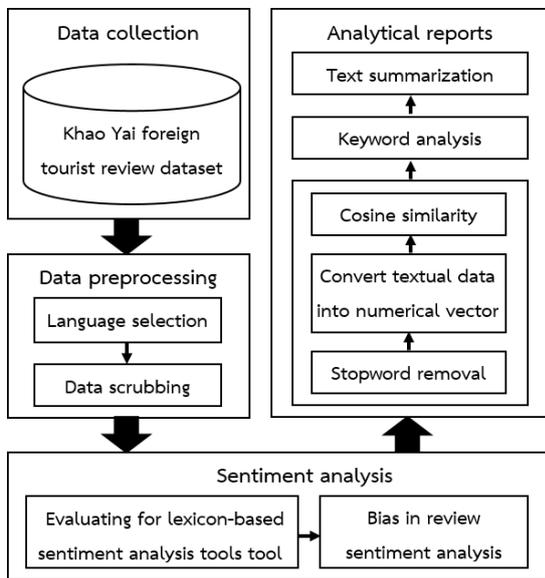
เข้าอุทยานล่วงหน้าผ่าน แอปพลิเคชัน และการจองที่พักผ่านเว็บไซต์และช่องทางออนไลน์ต่าง ๆ ก็มีส่วนช่วยให้นักท่องเที่ยวสามารถวางแผนการเดินทางได้อย่างมีประสิทธิภาพ 3) สิ่งอำนวยความสะดวก (amenities) เพื่อรองรับนักท่องเที่ยว เช่น ที่พักแบบบ้านพักและลานกางเต็นท์ ร้านอาหารภายในพื้นที่ และศูนย์บริการนักท่องเที่ยว นอกจากนี้ยังมีบริการด้านสุขอนามัย เช่น ห้องน้ำสาธารณะ และบริการสาธารณสุขประเภทต่าง ๆ 4) การให้บริการแบบเป็นชุด (available packages) ที่ตอบสนองความต้องการของนักท่องเที่ยว เช่น แพคเกจเดินป่า แพคเกจชมสัตว์ป่าในช่วงกลางคืน หรือแพคเกจเยี่ยมชมน้ำตกภายในพื้นที่ อีกทั้งยังมีการจัดกิจกรรมพิเศษในช่วงฤดูกาลต่าง ๆ เช่น การชมนกในฤดูหนาว หรือการเดินทางศึกษาธรรมชาติในช่วงฤดูฝน เป็นต้น 5) กิจกรรมการท่องเที่ยว (activities) ที่หลากหลายและตอบสนองความสนใจของนักท่องเที่ยวกลุ่มต่าง ๆ ที่เดินทางมาในรูปแบบเดี่ยว แบบคู่รัก แบบครอบครัว ตลอดจนแบบหมู่คณะ ไม่ว่าจะเป็นการเดินทางศึกษาธรรมชาติ การปั่นจักรยานในเส้นทางธรรมชาติ การส่องสัตว์ป่า และการถ่ายภาพทิวทัศน์ กิจกรรมเหล่านี้ช่วยเสริมสร้างประสบการณ์ที่น่าจดจำให้นักท่องเที่ยว และ 6) การให้บริการของแหล่งท่องเที่ยว (ancillary services) อุทยานแห่งชาติ เขาใหญ่ยังมีการจัดเตรียมบริการเสริมที่สำคัญ เช่น หน่วยบริการทางการแพทย์เบื้องต้น ธนาคาร บริการชำระเงินอิเล็กทรอนิกส์ บริการโทรศัพท์และอินเทอร์เน็ต รวมถึงจุดบริการข้อมูลที่ช่วยอำนวยความสะดวกแก่นักท่องเที่ยว จะเห็นได้ว่าองค์ประกอบทั้ง 6 ด้านนี้แสดงให้เห็นถึงศักยภาพของอุทยานแห่งชาติเขาใหญ่ในการตอบสนองความต้องการของนักท่องเที่ยวได้อย่างครบถ้วน

อย่างไรก็ดี เนื่องจากบริบทของอุทยานแห่งชาติเขาใหญ่เป็นแหล่งท่องเที่ยวทางธรรมชาติ ดังนั้นการท่องเที่ยวส่วนใหญ่จึงมุ่งที่การอนุรักษ์ธรรมชาติและสิ่งแวดล้อมเป็นหลัก งานวิจัยนี้จึงพิจารณารวมการให้บริการของแหล่งท่องเที่ยวมารวมกับองค์ประกอบด้านสิ่งอำนวยความสะดวก เนื่องจากเป็นบริการเสริมที่เกี่ยวข้องทั้งแบบเป็นทางการ และไม่เป็นทางการที่นักท่องเที่ยวควรได้รับ เช่นเดียวกับการให้บริการแบบเป็นชุดที่จะมีความทับซ้อนกับกิจกรรมการท่องเที่ยวในพื้นที่ คณะผู้วิจัยจึงรวมเข้ากับองค์ประกอบด้านกิจกรรม เนื่องจากเป็นบริบทที่มีความใกล้เคียงกัน เพื่อให้ได้ข้อมูลพื้นฐานที่ใช้เป็นแนวทางในการปรับปรุง และพัฒนาให้ตอบสนองต่อความต้องการของนักท่องเที่ยว ซึ่งจะนำไปสู่การ

พัฒนาอุตสาหกรรมการท่องเที่ยวอย่างยั่งยืนในจังหวัดนครราชสีมา และพื้นที่ใกล้เคียง

3) วิธีดำเนินการวิจัย

งานวิจัยนี้ประยุกต์ใช้การประมวลผลภาษาธรรมชาติร่วมกับเครื่องมือการวิเคราะห์ความรู้สึกจากข้อความแบบใช้พจนานุกรมเพื่อวิเคราะห์บทวิจารณ์ของนักท่องเที่ยวชาวต่างชาติต่ออุทยานแห่งชาติเขาใหญ่ แบ่งการดำเนินงานเป็น 4 ขั้นตอน ดังรูปที่ 1 กรอบการดำเนินงาน ประกอบด้วย การรวบรวมข้อมูล การประมวลผลข้อมูลเบื้องต้น การวิเคราะห์ความรู้สึก และการสร้างรายงานเชิงวิเคราะห์



รูปที่ 1 : กรอบการดำเนินงาน

3.1) การรวบรวมข้อมูล (Data Collection)

การรวบรวมบทวิจารณ์สาธารณะของนักท่องเที่ยวต่ออุทยานแห่งชาติเขาใหญ่ด้วยการสกัดและดึงข้อมูลจากเว็บไซต์ (web scraping) [18] ที่เปิดให้เขียนบทวิจารณ์สาธารณะ ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2563 ถึง 31 ธันวาคม พ.ศ. 2567 เนื่องจากข้อมูลดังกล่าวสะท้อนถึงประสบการณ์ และมุมมองของนักท่องเที่ยวในช่วงเวลาปัจจุบัน

อย่างไรก็ตาม งานวิจัยนี้ให้ความสำคัญต่อข้อกำหนดทางกฎหมายและจริยธรรมในการดึงข้อมูลจากแพลตฟอร์มออนไลน์โดยยึดมั่นในหลักการปฏิบัติตามนโยบายและข้อกำหนดที่ระบุโดยเว็บไซต์ที่ใช้เป็นแหล่งข้อมูล นอกจากนี้ยังตระหนักถึงความสำคัญของการปกป้องข้อมูลส่วนบุคคลของผู้ใช้ ซึ่งรวมถึงการ

หลีกเลี่ยงข้อมูลที่สามารถระบุตัวบุคคลได้โดยตรงที่อาจสร้างความเสี่ยงต่อความเป็นส่วนตัวของผู้วิจารณ์ รวมถึงภาพถ่ายหรือสื่ออื่น ๆ ที่มีลักษณะของผู้ใช้มาใช้เป็นส่วนหนึ่งของการวิเคราะห์เพื่อให้มั่นใจว่างานวิจัยดำเนินการภายใต้กรอบของความถูกต้องตามกฎหมายและจริยธรรมที่ไม่กระทบต่อความเป็นส่วนตัวหรือสิทธิของผู้ใช้งานแพลตฟอร์มออนไลน์ในทางใดทางหนึ่ง

3.2) การประมวลผลข้อมูลเบื้องต้น (Data Preprocessing)

หลังจากที่ได้ข้อมูลดิบจากการรวบรวมข้อมูลแล้ว ขั้นตอนต่อมาคือการประมวลผลข้อมูลเบื้องต้น เพื่อแปลงข้อมูลให้อยู่ในรูปแบบที่พร้อมสำหรับการวิเคราะห์ มีกระบวนการดังนี้

3.2.1) การคัดเลือกภาษา (Language Selection) เนื่องจากบทวิจารณ์สาธารณะมักประกอบด้วยหลายภาษา งานวิจัยนี้มุ่งวิเคราะห์บทวิจารณ์จากนักท่องเที่ยวชาวต่างชาติ จึงคัดกรองเฉพาะบทวิจารณ์ที่เขียนด้วยภาษาอังกฤษที่เกี่ยวข้องกับการท่องเที่ยวในอุทยานเท่านั้น เพื่อให้ได้ข้อมูลที่มีความเฉพาะเจาะจงและสอดคล้องกับวัตถุประสงค์ของการศึกษา

3.2.2) การทำความสะอาดข้อมูล (Data Scrubbing) ก่อนการนำไปใช้งาน โดยตรวจสอบข้อมูลที่สูญหาย (missing values) ซึ่งอาจมีผลกระทบต่อความถูกต้องของการวิเคราะห์ นอกจากนี้ยังมีการตรวจสอบและลบข้อมูลที่ไม่เกี่ยวข้อง ซึ่งอาจรวมถึงข้อมูลที่ไม่มีความหมายหรือเป็นข้อมูลรบกวน เช่น สัญลักษณ์ หรืออักขระพิเศษที่ไม่เกี่ยวข้อง ตลอดจนการปรับเปลี่ยนรูปแบบของข้อมูลให้ข้อมูลมีความสม่ำเสมอและพร้อมสำหรับการประมวลผล

อย่างไรก็ดี เนื่องจากบทวิจารณ์อยู่ในรูปแบบของภาษาอังกฤษ ซึ่งมีการเว้นวรรคเพื่อระบุขอบเขตของคำอย่างชัดเจน จึงไม่จำเป็นต้องผ่านกระบวนการตัดคำ ซึ่งจะช่วยให้การประมวลผลเป็นไปอย่างรวดเร็วและมีประสิทธิภาพมากขึ้น

3.3) การวิเคราะห์ความรู้สึก (Sentiment Analysis)

การวิเคราะห์ความรู้สึกจากบทวิจารณ์ในรูปแบบภาษาอังกฤษจำนวน 12,035 รายการที่ผ่านการประมวลผลข้อมูลเบื้องต้น รายละเอียดดังตารางที่ 2

ตารางที่ 2 : จำนวนบทวิจารณ์จากนักท่องเที่ยวชาวต่างชาติ

Rating	Quantity (items)
1	546
2	560
3	1,644

ตารางที่ 2 : จำนวนบทวิจารณ์จากนักท่องเที่ยวยาวต่างชาติ (ต่อ)

Rating	Quantity (items)
4	2,814
5	6,471
รวม	12,035

จากตารางที่ 2 บทวิจารณ์ที่ผ่านการประมวลผลข้อมูลเบื้องต้นแล้วจะนำมาการวิเคราะห์ความรู้สึกจากข้อความร่วมกับค่าคะแนนความพึงพอใจ 5 ระดับ เพื่อประเมินผลด้วยค่าความถูกต้องร่วมกับการวิเคราะห์ความเอนเอียงในบทวิจารณ์ มีรายละเอียดดังนี้

3.3.1) การประเมินผลเครื่องมือสำหรับวิเคราะห์ความรู้สึกจากข้อความแบบใช้พจนานุกรม (*Evaluating for Lexicon-Based Sentiment Analysis Tools*) ที่สามารถวิเคราะห์ข้อความในรูปแบบข้อความธรรมดาได้เลย โดยไม่จำเป็นต้องแปลงเป็นเวกเตอร์เชิงตัวเลขก่อน

งานวิจัยนี้ประเมินผลด้วยค่าความถูกต้อง (accuracy) เพื่อวัดประสิทธิภาพของเครื่องมือสำหรับวิเคราะห์ความรู้สึกแบบใช้พจนานุกรมที่ได้แก่ TextBlob, Flair และ VADER ในการวิเคราะห์และจำแนกความรู้สึกจากข้อความ ดังสมการที่ 1

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of test instances}} \quad (1)$$

จากสมการที่ 1 ค่าความถูกต้องจะคำนวณจากสัดส่วนของการจำแนกข้อมูลที่ทดสอบว่าเป็นเชิงบวก เชิงลบ หรือเป็นกลาง เมื่อเทียบกับผลจริงที่ได้จากการให้คะแนนความพึงพอใจ โดยกำหนดให้ค่าคะแนนความพึงพอใจที่เท่ากับ 5 และ 4 ถือว่าเป็นความรู้สึกเชิงบวก หากเท่ากับ 3 ถือว่าเป็นกลาง และค่าคะแนนความพึงพอใจที่เท่ากับ 2 และ 1 ถือว่าเป็นความรู้สึกเชิงลบ คิดเป็นจำนวนผลลัพธ์จริงของความรู้สึกเชิงบวกเท่ากับ 9,285 รายการ เชิงลบ เท่ากับ 1,106 รายการ และเป็นกลางเท่ากับ 1,644 รายการ ดังนั้นค่าความถูกต้องที่ได้จะแสดงให้เห็นถึงความสามารถในการจำแนกความรู้สึก และเป็นแนวทางในการนำไปประยุกต์ใช้งานในสถานการณ์จริงต่อไปในอนาคต

3.3.2) การวิเคราะห์ความเอนเอียงในบทวิจารณ์ (*Bias in Review Sentiment Analysis*) เนื่องจากในการวิเคราะห์ข้อความหนึ่งในแนวคิดที่ได้รับความนิยมอย่างมากคือการประเมินความรู้สึกจากบทวิจารณ์ออนไลน์ อย่างไรก็ตาม ปัญหาหนึ่งที่มีเกิดขึ้นคือความเอนเอียง (bias) ในบทวิจารณ์ ซึ่งทำให้ผลลัพธ์ของการวิเคราะห์มีความคลาดเคลื่อนจากความเป็นจริง

ความเอนเอียงในบทวิจารณ์อาจเกิดจากปัจจัยหลายประการ เช่น แรงจูงใจในการให้คะแนน โดยบางแพลตฟอร์มให้รางวัลหรือมีสิทธิพิเศษสำหรับบทวิจารณ์เชิงบวก ซึ่งอาจทำให้คะแนนความพึงพอใจมีแนวโน้มสูงกว่าความเป็นจริง ตลอดจนความเข้าใจผิดเกี่ยวกับระบบการให้คะแนน อารมณ์ชั่วขณะ หรือการใช้ภาษาที่มีลักษณะประชดประชัน งานวิจัยนี้จึงมุ่งวิเคราะห์ความเอนเอียงจากบทวิจารณ์ที่มีคะแนนความพึงพอใจเท่ากับ 5 แต่มีความรู้สึกจากข้อความเชิงลบ และในทางกลับกัน บทวิจารณ์ที่มีคะแนนความพึงพอใจเท่ากับ 1 แต่มีความรู้สึกจากข้อความเชิงบวกจากการใช้เครื่องมือวิเคราะห์ความรู้สึกจากข้อความแบบใช้พจนานุกรม โดยแต่ละค่าจะได้รับคะแนนความเข้มข้น (valence score) [19] ที่จะช่วยให้ระบุทิศทางของอารมณ์ในข้อความ หากมีคะแนนมากกว่า 0.05 จะถูกพิจารณาว่าเป็นเชิงบวก และหากคะแนนน้อยกว่า -0.05 จะถูกพิจารณาว่าเป็นเชิงลบ นอกจากนี้หากคะแนนอยู่ระหว่าง -0.05 ถึง 0.05 จะถูกพิจารณาว่าเป็นกลาง และคะแนนรวม (compound score) ที่พิจารณารูปร่างทั้งหมดของข้อความที่ในบทวิจารณ์ เนื่องจากในบทวิจารณ์เดียวกันอาจมีการผสมผสานของความคิดเห็นทั้งบวกและลบ คะแนนรวมจึงช่วยให้สามารถประเมินความรู้สึกโดยรวมของข้อความได้อย่างแม่นยำจากการรวมคะแนนของค่าต่าง ๆ ที่ปรากฏในบทวิจารณ์

คะแนนรวมที่มีเกณฑ์ในการตัดสินใจที่ชัดเจนสามารถนำไปหาความเอนเอียงระหว่างคะแนนความพึงพอใจเปรียบเทียบกับผลการวิเคราะห์ความรู้สึก เช่น หากคะแนนความพึงพอใจสูง แต่แสดงความรู้สึกเชิงลบในข้อความ หมายความว่านักท่องเที่ยวอาจมีความคาดหวังที่สูง หรืออาจมีข้อร้องเรียนบางประการที่ไม่สามารถสะท้อนในคะแนนได้ ซึ่งแสดงให้เห็นถึงปัญหาที่ต้องได้รับการแก้ไข ในทางตรงกันข้าม คะแนนความพึงพอใจต่ำ แต่กลับแสดงความรู้สึกเชิงบวกในข้อความ อาจบ่งชี้ถึงความไม่ชัดเจนในระบบการประเมิน หรือการให้คะแนนที่อาจตีความไม่ตรงกัน จึงต้องตรวจสอบวิธีการให้คะแนนหรือนโยบายการรวบรวมข้อมูลเพิ่มเติม

3.4) การสร้างรายงานเชิงวิเคราะห์ (Analytical Reports)

รายงานเชิงวิเคราะห์เพื่อมุ่งเน้นการนำเสนอข้อมูลเชิงลึกในรูปแบบที่เข้าใจง่ายจากการเชื่อมโยงข้อมูลเพื่อวิเคราะห์ในหลากหลายมิติ ร่วมกับการสื่อสารผลลัพธ์ผ่านการสร้างภาพข้อมูล (data visualization) ประกอบด้วย

3.4.1) การลบคำหยุด (Stop Removal) ที่พบได้บ่อยครั้งในข้อความแต่ไม่มีนัยสำคัญ จึงสามารถตัดออกได้เลยโดยไม่กระทบต่อความหมายหลักของข้อความ เช่น “a”, “the”, “is” หรือ “and” เป็นต้น ซึ่งช่วยลดขนาดของข้อมูลและทำให้สามารถมุ่งเน้นการวิเคราะห์ไปยังคำที่มีความหมายสำคัญได้มากขึ้น จากนั้นจะลบเครื่องหมายต่าง ๆ ที่ไม่เกี่ยวข้อง ก่อนจะแปลงข้อความทั้งหมดเป็นตัวอักษรพิมพ์เล็กในรูปแบบพื้นฐาน เช่น “watching” เปลี่ยนเป็น “watch” ขั้นตอนนี้ช่วยลดความหลากหลายของคำที่มีรูปแบบแตกต่างกันแต่มีความหมายเดียวกัน ให้ข้อมูลมีความสม่ำเสมอและพร้อมสำหรับการประมวลผล

3.4.2) การแปลงข้อความที่เป็นเวกเตอร์เชิงตัวเลขจากการพิจารณาความสำคัญของคำในแต่ละข้อความผ่านการคำนวณ TF-IDF (Term Frequency-Inverse Document Frequency) [20] โดย term Frequency (TF) คือการวัดความถี่ของคำที่ปรากฏในเอกสาร และ inverse document frequency (IDF) คือการวัดความสำคัญของคำในเอกสาร โดยคำที่ปรากฏบ่อยครั้งในเอกสารจะมีค่า IDF ต่ำ ตรงกันข้ามกับคำที่ปรากฏในเอกสารน้อยครั้งจะมีค่า IDF สูง ดังนั้นการคูณค่า TF และ IDF จะสามารถกำหนดได้ว่าแต่ละคำในข้อความมีความสำคัญมากน้อยเพียงใดต่อข้อความนั้น ๆ ซึ่งจะช่วยในการวิเคราะห์ข้อความที่มีความยาวและความแตกต่างกันเป็นไปได้อย่างมีประสิทธิภาพ

กระบวนการแปลงข้อความที่เป็นเวกเตอร์เชิงตัวเลขด้วยการคำนวณค่า TF-IDF ช่วยวิเคราะห์ความสำคัญของคำในข้อความ โดยความถี่ของคำหรือ TF และความถี่ผกผันของเอกสารหรือ IDF คำนวณได้จากสมการที่ 2 และ 3

$$TF(t, d) = \frac{\int t, d}{\sum t' \in d \int t', d} \quad (2)$$

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

$$IDF(t) = \log \left(\frac{N}{1 + DF(t)} \right) \quad (3)$$

โดย N คือจำนวนเอกสารทั้งหมด และ $DF(t)$ คือจำนวนเอกสารที่มีคำ t ปรากฏอยู่ ซึ่งการบวก 1 ในตัวส่วนช่วยป้องกันปัญหาการหารด้วย 0

ดังนั้น TF-IDF เป็นค่าที่ช่วยถ่วงน้ำหนักของคำ โดยพิจารณาทั้งความสำคัญของคำภายในเอกสารเดียวกันและความสำคัญของคำเมื่อเปรียบเทียบกับเอกสารทั้งหมดในชุดข้อมูล ดังสมการที่ 4 ซึ่งจะช่วยระบุค่าสำคัญในข้อความได้อย่างมีประสิทธิภาพ

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \quad (4)$$

3.4.3) การวัดความคล้ายคลึงโคไซน์ (Cosine Similarity) [20] สำหรับวัดความคล้ายคลึงของเวกเตอร์ มักใช้ในงานที่เกี่ยวข้องกับข้อความและข้อมูลที่มีมิติสูง เนื่องจากไม่ถูกกระทบจากขนาดของเวกเตอร์ และคำนวณได้รวดเร็ว งานวิจัยนี้วัดความคล้ายคลึงระหว่างบทวิจารณ์แต่ละรายการกับคำสำคัญที่อยู่ในรูปแบบเวกเตอร์เชิงตัวเลข ซึ่งใช้ในการเปรียบเทียบว่าเนื้อหาของข้อความมีความคล้ายคลึงกันมากน้อยเพียงใดระหว่างบทวิจารณ์และคำสำคัญจากบทวิจารณ์การท่องเที่ยว จากสมการที่ 5

$$\text{Cosine similarity} = \frac{A \cdot B}{\|A\| \|B\|} \quad (5)$$

โดย $A \cdot B$ คือผลคูณเชิงสเกลาร์ (dot product) ระหว่างเวกเตอร์ A และ B และ $\|A\| \|B\|$ คือขนาดของเวกเตอร์ A และ B ตามลำดับ

จากสมการที่ 5 ค่าความคล้ายคลึงโคไซน์ จะมีค่าระหว่าง -1 ถึง 1 โดยค่าเข้าใกล้กับ 1 ยิ่งมากแสดงว่าเวกเตอร์ทั้งสองมีทิศทางเดียวกัน หรือมีความคล้ายคลึงกันสูง ในทางกลับกัน หากค่าเข้าใกล้กับ 0 แสดงว่าเวกเตอร์ทั้งสองไม่มีความคล้ายคลึงกัน และหากค่าเข้าใกล้กับ -1 แสดงว่าเวกเตอร์ทั้งสองมีทิศทางตรงกันข้าม

3.4.4) การวิเคราะห์คำสำคัญ (Keyword Analysis) จากบทวิจารณ์ของนักท่องเที่ยวต่างชาติต่ออุทยานแห่งชาติเขาใหญ่ ด้วยการประมวลผลภาษาธรรมชาติ จากการสร้างคำสำคัญจากการวิเคราะห์คำอัตโนมัติ [21] เพื่อสร้างคำสำคัญ โดยเน้นไปที่องค์ประกอบของแหล่งท่องเที่ยว 4 องค์ประกอบ ประกอบด้วย สิ่งดึงดูดใจ การเข้าถึงสถานที่ท่องเที่ยว สิ่งอำนวยความสะดวก และกิจกรรมการท่องเที่ยว เช่น คำว่า “staff and service” สำหรับองค์ประกอบด้านสิ่งอำนวยความสะดวก หรือคำว่า “hiking trail” สำหรับองค์ประกอบเกี่ยวกับกิจกรรมในอุทยาน เพื่อนำมาเปรียบเทียบความคล้ายคลึงกันระหว่างบทวิจารณ์ และคำสำคัญที่ถูกแปลงเป็นเวกเตอร์เชิงตัวเลขด้วย ค่าความคล้ายคลึงโคไซน์ โดยบทวิจารณ์ที่มีค่าความคล้ายคลึงสูงกว่าค่าที่กำหนดไว้จะถูกเลือกมาวิเคราะห์คะแนนความรู้สึก เพื่อประเมินความคิดเห็นของนักท่องเที่ยว เช่น ข้อความระบุว่า “the staff provided detailed and polite information” จะได้รับคะแนนเชิงบวก ในขณะที่ข้อความ เช่น “the hiking trail lacks sufficient signage” จะได้รับคะแนนเชิงลบ ก่อนจะนำมารายงานผลตามองค์ประกอบของแหล่งท่องเที่ยวร่วมกับการนำเสนอภาพข้อมูล

เพื่อให้เห็นภาพรวมของข้อมูลในเชิงปริมาณและเชิงคุณภาพที่สะท้อนถึงความต้องการของนักท่องเที่ยว

3.4.5) การสรุปข้อความ (Text Summarization) เป็นเทคนิคในการสรุปเนื้อหาของข้อความให้สั้นลง โดยยังคงรักษาสาระสำคัญของเนื้อหาไว้ให้ได้มากที่สุด [21] เทคนิคนี้เป็นส่วนหนึ่งของการประมวลผลภาษาธรรมชาติที่มีบทบาทสำคัญในการช่วยให้สามารถสื่อสารข้อมูลที่สำคัญไปยังผู้รับได้อย่างรวดเร็ว

ปัจจุบันการสรุปข้อความแบ่งเป็น 2 ประเภทหลัก ได้แก่ extractive summarization โดยการเลือกประโยคหรือวลีที่สำคัญจากต้นฉบับมาใช้โดยตรง จึงมีข้อดีในด้านความแม่นยำ เนื่องจากสามารถสรุปโดยยังคงความถูกต้องและตรงกับเนื้อหาดั้งเดิม แต่ยังมีข้อจำกัดในเรื่องของการกระชับข้อมูล รวมถึงอาจทำให้ข้อความบางส่วนที่มีความสำคัญถูกละเลยไปได้ และ abstractive summarization วิธีนี้ใช้โมเดลปัญญาประดิษฐ์ในการทำความเข้าใจบริบทของข้อความต้นฉบับและเขียนสรุปขึ้นมาใหม่ในรูปแบบที่กระชับและเป็นธรรมชาติมากขึ้น ซึ่งช่วยให้การสรุปมีความเป็นธรรมชาติ และง่ายต่อการทำความเข้าใจ เนื่องจากโมเดลสามารถเข้าใจบริบทของข้อความต้นฉบับและเลือกคำที่เหมาะสมเพื่อสร้างข้อความที่มีความหมายใหม่ได้อย่างไรก็ตาม การสร้างข้อความใหม่อาจทำให้เกิดการเข้าใจผิดหรือการสูญเสียข้อมูลสำคัญไป นอกจากนี้ยังต้องการทรัพยากรในการคำนวณที่สูงและมีความซับซ้อนในการฝึกโมเดล

จะเห็นได้ว่าทั้งสองวิธีต่างมีจุดเด่นและข้อจำกัดที่ต่างกัน การเลือกใช้จึงควรพิจารณาตามลักษณะของข้อมูลและวัตถุประสงค์การใช้งาน งานวิจัยนี้พิจารณาเลือกประโยคหรือวลีที่สำคัญจากต้นฉบับมาใช้โดยตรงจากการจับคู่ความสัมพันธ์ระหว่างเวกเตอร์ก่อนจะปรับแต่งกระบวนการสรุปข้อความสำหรับใช้ในเนื้อหาที่เฉพาะเจาะจงสำหรับการท่องเที่ยวและนำเสนอในรูปแบบความเรียงประกอบภาพข้อมูลต่อไป

4) ผลการวิจัยและอภิปรายผล

งานวิจัยนี้พัฒนาขึ้นการเขียนโปรแกรมภาษาไพธอน (python) บน Google Colaboratory แบ่งผลการวิจัยและอภิปรายผลออกเป็น 2 ส่วนตามวัตถุประสงค์ ได้แก่ เพื่อวิเคราะห์บทวิจารณ์ด้วยเครื่องมือวิเคราะห์ความรู้สึกจากข้อความแบบใช้พจนานุกรม และเพื่อสร้างรายงานเชิงวิเคราะห์บทวิจารณ์จากนักท่องเที่ยวชาวต่างชาติต่ออุทยานแห่งชาติเขาใหญ่ มีรายละเอียดดังนี้

4.1) ผลการวิเคราะห์บทวิจารณ์ด้วยเครื่องมือวิเคราะห์ความรู้สึกแบบใช้พจนานุกรม

การวิเคราะห์ความรู้สึกจากข้อความโดยอาศัยพจนานุกรมที่มีการสร้างคำศัพท์และกำหนดคะแนนความรู้สึกไว้ล่วงหน้าแบ่งผลการวิจัยและอภิปรายผลเป็น 5 ขั้นตอน ได้แก่

4.1.1) ผลการรวบรวมข้อมูลบทวิจารณ์สาธารณะ ด้วยการสกัดและดึงข้อมูลจากเว็บไซต์ที่เปิดให้นักท่องเที่ยวเขียนบทวิจารณ์สาธารณะต่ออุทยานแห่งชาติเขาใหญ่ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2563 ถึง 31 ธันวาคม พ.ศ. 2567 รวมระยะเวลาทั้งสิ้น 4 ปี ซึ่งถือว่าสะท้อนภาพรวมของแหล่งท่องเที่ยวในปัจจุบัน และเหมาะสมต่อการนำมาวิเคราะห์เพื่อเป็นแนวทางพัฒนาศักยภาพของแหล่งท่องเที่ยว

งานวิจัยนี้รวบรวมข้อมูลด้วยการสกัดและดึงข้อมูลจากเว็บไซต์อัตโนมัติ เริ่มจากการส่งคำขอไปยังเว็บไซต์เป้าหมาย โดยใช้โปรโตคอล HTTP เพื่อดึงข้อมูลหน้าเว็บออกมาในรูปแบบของโครงสร้าง HTML หลังจากนั้น จะทำการวิเคราะห์และสกัดข้อมูลที่ต้องการผ่านโครงสร้างของ HTML เช่น แท็ก <div>, หรือ <table> ซึ่งเป็นส่วนที่เก็บข้อมูลหลักของเว็บไซต์ เพื่อเก็บรวบรวมข้อมูล ประกอบด้วยวันที่ ค่าคะแนนความพึงพอใจ และบทวิจารณ์สาธารณะที่สามารถนำมาใช้ในการวิเคราะห์จุดเด่นและข้อเสนอแนะของอุทยานแห่งชาติเขาใหญ่ได้

4.1.2) ผลการประมวลผลข้อมูลเบื้องต้น งานวิจัยนี้มุ่งเน้นการวิเคราะห์ข้อความจากนักท่องเที่ยวชาวต่างชาติ โดยคัดกรองเฉพาะบทวิจารณ์ที่เขียนด้วยภาษาอังกฤษ ประกอบด้วย บทวิจารณ์และคะแนนความพึงพอใจ จำนวน 12,035 รายการ ผ่านการทำความสะอาดข้อมูล โดยมีวัตถุประสงค์เพื่อปรับปรุงคุณภาพของข้อมูลให้มีความถูกต้อง สมบูรณ์ และพร้อมสำหรับการนำไปใช้งาน เริ่มต้นจากการระบุและจัดการคำที่หายไป ซึ่งอาจเกิดจากการบันทึกข้อมูลที่ไม่สมบูรณ์หรือข้อผิดพลาดในการรวบรวมข้อมูล โดยหากพบรายการใดที่มีค่าคะแนนความพึงพอใจหายไป จะพิจารณาลบแถวที่มีค่าหายไป เนื่องจากไม่ส่งผลกระทบต่อความสมบูรณ์ของข้อมูล นอกจากนี้ ยังมีการจัดรูปแบบข้อมูลให้เป็นมาตรฐาน เช่น การแปลงค่าข้อความให้เป็นตัวอักษรพิมพ์เล็กทั้งหมด และการลบอักขระพิเศษที่ไม่จำเป็น เพื่อให้ข้อมูลมีคุณภาพและความน่าเชื่อถือก่อนนำไปใช้

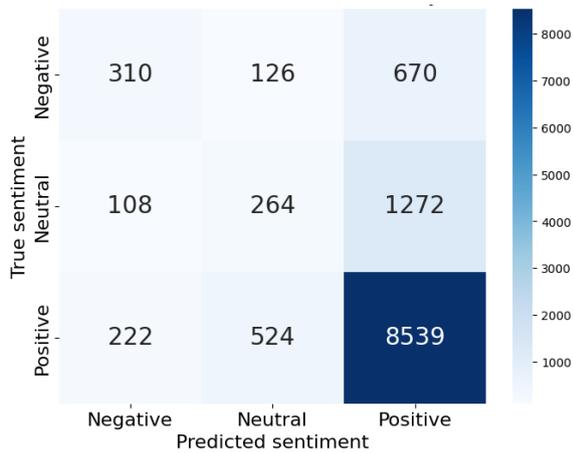
4.1.3) ผลการประเมินผลเครื่องมือสำหรับวิเคราะห์ความรู้สึกจากข้อความแบบใช้พจนานุกรมด้วยค่าความถูกต้องที่คำนวณจากสัดส่วนของการจำแนกข้อมูลว่าเป็นเชิงบวก เชิงลบ

หรือเป็นกลาง เมื่อเทียบกับผลจริงที่ได้จากการให้คะแนนความพึงพอใจ ได้แก่ TextBlob, Flair และ VADER จากการวัดซ้ำจนครบ 3 รอบ แล้วนำผลการประเมินมาหาค่าเฉลี่ย ดังตารางที่ 3

ตารางที่ 3 : ค่าความถูกต้องเฉลี่ยจากการวิเคราะห์ความรู้สึกจากข้อความ

sentiment analysis tool	accuracy
TextBlob	0.7198
Flair	0.7324
VADER	0.7575

จากตารางที่ 3 ค่าความถูกต้องเฉลี่ยจากการวิเคราะห์ความรู้สึกจากข้อความระหว่าง TextBlob, Flair และ VADER มีค่าความถูกต้องเฉลี่ยใกล้เคียงกันคิดเป็นร้อยละ 72, 73 และ 76 ตามลำดับโดย VADER มีค่าความถูกต้องเฉลี่ยสูงที่สุด



รูปที่ 2 : เมทริกซ์ความสับสนการจำแนกความรู้สึกจากข้อความด้วย VADER

เมื่อพิจารณาจากเมทริกซ์ความสับสน (confusion matrix) ซึ่งใช้ประเมินประสิทธิภาพของการจำแนกความรู้สึกเป็นเชิงบวก เชิงลบ และเป็นกลางของ VADER ดังรูปที่ 2 พบว่า การจำแนกความรู้สึกจากข้อความได้อย่างถูกต้องประมาณ 3 ใน 4 ของทั้งหมด เมื่อพิจารณาในรายละเอียด พบว่า สามารถจำแนกข้อความเชิงบวกได้ดีมาก โดยมีจำนวนตัวอย่างที่ทำนายถูกต้องถึง 8,539 ตัวอย่าง แต่ยังมีข้อผิดพลาดในการจำแนกข้อความที่เป็นกลางและข้อความเชิงลบอยู่ในระดับที่ค่อนข้างสูง โดยจำแนกความรู้สึกผิดว่าเป็นเชิงบวก ทั้งที่เป็นเชิงลบหรือเป็นกลางรวมกันถึง 746 ตัวอย่าง แสดงให้เห็นถึงแนวโน้มที่ให้น้ำหนักกับข้อความเชิงบวกมากเกินไป (bias toward positive) เช่นเดียวกับการจำแนกข้อความที่เป็นกลาง ซึ่งทำนายผิดไปเป็นเชิงบวกจำนวน

1,272 ตัวอย่าง ถือเป็นอัตราส่วนที่สูงมากเมื่อเทียบกับข้อความที่เป็นกลางทั้งหมด นอกจากนี้ ยังมีข้อผิดพลาดในการจำแนกข้อความเชิงลบบางส่วนที่ถูกจำแนกผิดเป็นเชิงบวกหรือเป็นกลางรวมกันถึง 796 ตัวอย่าง

อย่างไรก็ดี นอกเหนือจากค่าความถูกต้องที่เป็นตัวชี้วัดประสิทธิภาพในการทำงานแล้ว เครื่องมือสำหรับวิเคราะห์ความรู้สึกจากข้อความแบบใช้พจนานุกรมแต่ละแบบ ยังมีจุดเด่นและข้อจำกัดในการทำงานแตกต่างกันไป โดย TextBlob เหมาะสำหรับการใช้งานทั่วไปที่ต้องการความง่ายในการใช้งานส่วน Flair เหมาะสำหรับงานที่ต้องการเข้าใจบริบทของข้อความ แต่อาจต้องแลกกับการใช้ทรัพยากรการประมวลผลมากขึ้น ในทางกลับกัน VADER นั้นเหมาะสำหรับงานที่ต้องการความเร็วในการวิเคราะห์ข้อมูลและรองรับข้อความภาษาธรรมชาติจากสื่อสังคมออนไลน์ ดังนั้นการเลือกใช้งานควรพิจารณาจากลักษณะของข้อมูล และข้อจำกัดด้านทรัพยากรในการประมวลผลเป็นปัจจัยร่วมในการพิจารณาอีกด้วย

งานวิจัยนี้พิจารณาเลือก VADER ที่เป็นเครื่องมือสำหรับวิเคราะห์ความรู้สึกจากข้อความแบบใช้พจนานุกรมที่ออกแบบมาให้ใช้งานกับข้อความที่ไม่มีโครงสร้างได้ทันที โดยไม่ต้องผ่านกระบวนการตัดคำ การกำจัดคำหยุด และสามารถวิเคราะห์ข้อมูลจากข้อความทั้งหมดได้ทันที โดยไม่จำเป็นต้องแปลงเป็นเวกเตอร์ตัวเลขก่อน นอกจากนี้ยังมีค่าความถูกต้องเฉลี่ยสูงที่สุดแล้วยังเหมาะสำหรับงานที่ต้องการความเร็วในการวิเคราะห์ข้อความภาษาธรรมชาติบนอุปกรณ์ที่มีข้อจำกัดด้านทรัพยากรอย่างสมาร์ทโฟนได้อย่างมีประสิทธิภาพ ซึ่งเป็นปัจจัยสำคัญที่ควรพิจารณาสำหรับการใช้งานในสถานการณ์จริง

4.1.4) ผลการวิเคราะห์ความเอนเอียงในบทวิจารณ์ เพื่อประเมินว่าการวิเคราะห์ความรู้สึกจากบทวิจารณ์นั้นสามารถสะท้อนความคิดเห็นของนักท่องเที่ยวดังกล่าวได้อย่างถูกต้องหรือไม่ โดยเฉพาะในกรณีที่คะแนนความพึงพอใจ และเนื้อหาของบทวิจารณ์มีความขัดแย้งกัน โดยนำบทวิจารณ์จำนวน 12,035 รายการที่ผ่านการวิเคราะห์ความรู้สึกจาก VADER มาพิจารณาบทวิจารณ์ที่มีความเอนเอียง เพื่อทำความเข้าใจทิศทางและแนวโน้มของบทวิจารณ์จากนักท่องเที่ยวยุคต่างชาตินี้ในแง่บวก แง่ลบ หรือเป็นกลาง โดยพิจารณาบทวิจารณ์ที่มีความเอนเอียงในทางบวกหรือลบอย่างชัดเจน พบว่า บทวิจารณ์ที่คะแนนเท่ากับ 5 แต่มีความรู้สึกเป็นลบ มีจำนวนเท่ากับ 2,912 รายการ ในทางกลับกัน บทวิจารณ์ที่คะแนนเท่ากับ 1 แต่มีความรู้สึกเป็นบวก มีจำนวน

เท่ากับ 448 รายการ เมื่อเรียงลำดับตามค่าคะแนนความรู้สึกจากมากไปน้อย รายละเอียดดังตารางที่ 4 และ 5

ตารางที่ 4 : ตัวอย่างบทวิจารณ์ที่มีค่าคะแนนความพึงพอใจเท่ากับ 5 แต่ค่าความรู้สึกเป็นลบ (negative)

review	sentiment score
<p>it is worth the effort to get there. a treat to see wildlife in its natural unspoiled environment.</p> <p>used to see wildlife at home, but this topped it to see snakes, scorpions, millions of bats, deer, giant squirrels, gibbons, macaques, and even the only croc in the park near the waterfall trail, all in one day.</p> <p>only negative is the late morning opening and late afternoon closing of the park. probably why we missed the elephants.</p>	-0.7263
<p>a must visit if you're in khao yai being city dweller, 1st time exploring and experiencing the nature and didn't even bother about being sucked dry by leeches.</p> <p>encountered a couple of wildlife and the view on the top is to die for.</p>	-0.6194
...	...
<p>our best experience in Thailand. mostly visited by Thai tourists, and not really by foreigners it seems.</p> <p>it's a real wild jungle, with many animals to see crocodiles, monkeys, spiders, snakes. for a real jungle experience, go visit it with a ranger and stay there overnight in a tent.</p> <p>but closely follow the rangers or guides advice if you don't want to be attacked by leeches.</p>	-0.1343

จากตารางที่ 4 บทวิจารณ์จากนักท่องเที่ยวต่างชาติที่มีค่าคะแนนความพึงพอใจเท่ากับ 5 แต่ค่าความรู้สึกเป็นลบในหลายแง่มุม ซึ่งสามารถจำแนกได้เป็นประเด็นหลัก ๆ ได้แก่ ค่าธรรมเนียมที่ไม่เท่าเทียม ข้อจำกัดในการเข้าถึง ปัญหาด้านการบริการ และความคาดหวังที่ไม่ตรงกับความเป็นจริง โดยเฉพาะการเปรียบเทียบค่าใช้จ่ายระหว่างนักท่องเที่ยวชาวไทยและชาวต่างชาติที่แตกต่างกันอย่างชัดเจน ประเด็นต่อมาคือข้อจำกัดในด้านการเข้าถึงและกิจกรรมต่าง ๆ ที่ไม่ตอบสนองต่อความ

ต้องการของผู้เยี่ยมชมอย่างเต็มที่ อีกทั้งการจำกัดเวลาเปิดและปิดอุทยานทำให้พลาดโอกาสในการชมสัตว์ป่าในช่วงเวลาที่แตกต่างกันตามแต่ละฤดูกาล นอกจากนี้ ยังมีการกล่าวถึงการบริการที่ไม่เป็นไปตามความคาดหวัง บางบทวิจารณ์ได้ระบุว่าผู้นำเที่ยวขาดความพร้อมในการให้บริการ เช่น ไม่แจ้งข้อมูลหรือคำแนะนำที่ชัดเจนในเรื่องความปลอดภัย หรือคำแนะนำต่อการปรับตัวตามสภาพอากาศและสถานการณ์ต่าง ๆ ที่เหมาะสมจึงทำให้รู้สึกที่ไม่คุ้มค่ากับราคา อีกทั้งพลาดการเห็นสัตว์ที่มีชื่อเสียงอย่างช้าง หรือการที่สภาพเส้นทางเดินไม่เป็นไปตามคำแนะนำ อาจนำไปสู่ความผิดหวังนักท่องเที่ยว แม้สถานที่นั้นจะมีความงดงาม จึงแสดงความรู้สึกเชิงลบแม้ว่าจะได้รับประสบการณ์ที่ดีในบางแง่มุมจากการท่องเที่ยว

ตารางที่ 5 : ตัวอย่างบทวิจารณ์ที่มีค่าคะแนนความพึงพอใจเท่ากับ 1 แต่ค่าความรู้สึกเป็นบวก (positive)

review	sentiment score
<p>we were excited to visit but were put off right away by the \$400 baht charge per adult for foreigners. for Thai people, it's only \$40 baht, so visitors are charged 10 times the price.</p> <p>they also charge extra based on the vehicle you're driving. crazy. besides the price, the main road going through the park is more like a speedway for people crossing from one side to the other.</p>	0.9536
<p>it's kind of like a cheap toll road for Thais. if you've ever visited a national park in other countries, it's a different experience altogether.</p> <p>we weren't expecting to see animals but were hopeful.</p> <p>we saw a few monkeys on the road and one deer. that's it. it's a scenic park to drive through, but literally, that's all we did, drive through.</p> <p>i wouldn't recommend coming here.</p> <p>save your \$400 baht and have a nice meal out instead.</p>	0.9536



ตารางที่ 5 : ตัวอย่างบทวิจารณ์ที่มีค่าคะแนนความพึงพอใจเท่ากับ 1 แต่ค่าความรู้สึกเป็นบวก (positive) (ต่อ)

Review	sentiment score
did not bother, fee too high for foreigners, 400 baht. for locals it's 30 baht, why the drastic difference. what a turn off for foreigners. we have visited great national parks all over the world which have reasonable fees or are free.	0.8568
...	...
i do not recommend the night tour offered from the park, since the guide spoke no English (not even the name of the animals), and showed no interest in looking for animals. just saw deer.	0.1002

เช่นเดียวกับ บทวิจารณ์ที่มีค่าคะแนนความพึงพอใจเท่ากับ 1 แต่ค่าความรู้สึกเป็นบวกจากตารางที่ 5 พบว่า หลายบทวิจารณ์ แสดงความผิดหวังจากปัญหาด้านค่าธรรมเนียมที่ไม่เท่าเทียม การบริการที่ไม่ได้มาตรฐาน และประสบการณ์จริงที่ไม่ตรงกับความคาดหวัง แม้จะพึงพอใจในแง่ของความงดงามของธรรมชาติ และสัตว์ป่า แต่ปัญหาด้านการจัดการ และค่าใช้จ่ายที่สูงเกินไป กลับสร้างความผิดหวังและส่งผลกระทบต่อคะแนนความพึงพอใจโดยรวมที่ลดลง

4.2) ผลการสร้างรายงานเชิงวิเคราะห์

รายงานเชิงวิเคราะห์จากบทวิจารณ์ของนักท่องเที่ยวชาวต่างชาติต่ออุทยานแห่งชาติเขาใหญ่ที่ระบุจุดเด่น รวมถึงข้อจำกัดของการท่องเที่ยว อันจะเป็นแนวทางในการพัฒนาและยกระดับแหล่งท่องเที่ยวให้สอดคล้องกับมาตรฐานสากล แบ่งผลการวิจัยและอภิปรายผลออกเป็น 5 ขั้นตอน ประกอบด้วย

4.2.1) ผลการลบคำหยุด เพื่อกำจัดข้อมูลที่ไม่จำเป็น และปรับข้อความให้อยู่ในรูปแบบที่เหมาะสมกับการประมวลผล ประกอบด้วยการแปลงข้อความเป็นตัวอักษรพิมพ์เล็กทั้งหมด การแปลงคำให้อยู่ในรูปแบบพื้นฐาน รวมถึงการลบอักขระพิเศษ และเครื่องหมายวรรคตอนที่ไม่มีความหมายของข้อความ ดังตารางที่ 6 คำหยุดที่ถูกลบออก ประกอบด้วย “a”, “an”, “and”, “are”, “as”, “at”, “be”, “but”, “by”, “can”, “do”,

“for”, “from”, “has”, “have”, “if”, “in”, “is”, “it”, “its”, “not”, “of”, “on”, “or”, “so”, “that”, “the”, “this”, “to”, “was”, “were”, “will”, “with”, “you” และ “your”

ตารางที่ 6 : ตัวอย่างบทวิจารณ์ที่ผ่านกระบวนการประมวลผลข้อมูลเบื้องต้น

No.	Review	Rating
1.	nice park many trail waterfall hike trail without guide map available visitor center people complain pay foreigner see difference salary foreigner earn much common thai people support higher entrance fee conserve park	5
2.	incredibly big scenic park local pay 40 thb per head foreigner pay 100 thb well worth visit wild animal roam massive park go winter month weather cool windy need car go entire park	5
3.	expect much see animal may find elephant solid waste go jungle trekking without get tired national park route not long almost no climbing take 2 hr 3.3 km hundred year old tree make noise place frighten animal throw away garbage feed animal bring water snack	4
...
12035.	park lovely want go camping forget despite numerous campsite mark map one genuine one like supermarket car park	1

จากตารางที่ 6 ข้อความที่ผ่านการลบคำหยุด และผ่านการลบอักขระ เช่น ., “, ”, !, ? และ , จากนั้นแปลงคำให้อยู่ในรูปแบบพื้นฐาน ได้แก่ “trails” เป็น “trail”, “waterfalls” เป็น “waterfall”, “hiking” เป็น “hike”, “guides” เป็น “guide”, “maps” เป็น “map”, “visitors” เป็น “visitor”, “complaining” เป็น “complain”, “paying” เป็น “pay”, “foreigners” เป็น “foreigner”, “earning” เป็น “earn”, “supports” เป็น “support”, “conserving” เป็น “conserve”, “pays” เป็น “pay”, “heads” เป็น “head”, “roaming” เป็น “roam”, “going” เป็น “go”, “expecting” เป็น “expect”, “seeing” เป็น “see”, “animals” เป็น “animal”, “finding” เป็น “find”, “takes” เป็น “take”, “years” เป็น “year”, “trees” เป็น “tree”, “making” เป็น “make”, “frightening” เป็น

“frighten”, “throwing” เป็น “throw”, “feeding” เป็น “feed”, “bringing” เป็น “bring”, “wanting” เป็น “want”, “camping” เป็น “camp”, “forgetting” เป็น “forget”, “marking” เป็น “mark” และ “likes” เป็น “like”

4.2.2) ผลการแปลงข้อความเป็นเวกเตอร์เชิงตัวเลขด้วย TF-IDF ซึ่งช่วยกำหนดค่าความสำคัญของแต่ละคำในเอกสารได้ โดยคำที่ปรากฏบ่อยในข้อความเดียวกันจะได้รับค่าน้ำหนักที่ต่ำลง ในขณะที่คำที่มีความสำคัญต่อเนื้อหาจะได้รับค่าน้ำหนักที่สูงขึ้น โดยเรียกใช้ TfidfVectorizer เพื่อคำนวณค่าความสำคัญของคำในประโยคตัวอย่าง 3 ประโยค ประกอบด้วย “the park is good”, “the park is bad” และ “awesome is the awesome” แทนด้วย doc1, doc2 และ doc3 ตามลำดับ รายละเอียดดังตารางที่ 7 ถึง 9

ตารางที่ 7 : การหาความถี่ของคำแต่ละคำในเอกสาร

TF (term-frequency)						
	the	park	is	good	bad	awesome
doc1	1/4 =0.25	1/4 =0.25	1/4 =0.25	1/4 =0.25	0	0
doc2	1/4 =0.25	1/4 =0.25	1/4 =0.25	0	1/4 =0.25	0
doc3	1/4 =0.25	0	1/4 =0.25	0	0	2/4 =0.50

ตารางที่ 8 : การคำนวณค่าความสำคัญของคำ

IDF (inverse document-frequency)						
	the	park	is	good	bad	awesome
value	$\log(3/3)$ =0	$\log(3/2)$ =0.1761	$\log(3/3)$ =0	$\log(3/1)$ =0.4771	$\log(3/1)$ =0.4771	$\log(3/1)$ =0.4771

ตารางที่ 9 : ผลการแปลงข้อความเป็นเวกเตอร์คุณลักษณะ

feature vector						
	the	park	is	good	bad	awesome
doc1	0	0.0440	0	0.1193	0	0
doc2	0	0.0440	0	0	0.1193	0
doc3	0	0	0	0	0	0.2386

จากตารางที่ 7 การคำนวณค่าความถี่ของคำจากจำนวนครั้งที่คำปรากฏในเอกสาร ทหารด้วยจำนวนคำทั้งหมดในเอกสารนั้น คำที่ได้แสดงความถี่ของคำแต่ละคำในเอกสารหนึ่ง ๆ เพื่อคำนวณค่าความสำคัญของคำ ดังตารางที่ 8 คำที่มีค่าสูงกว่า

แสดงว่าคำปรากฏในเอกสารน้อยกว่า และมีความสำคัญมากกว่า และตารางที่ 9 เวกเตอร์คุณลักษณะ (feature vector) ที่ได้จากการคำนวณ TF-IDF จะแสดงถึงความสำคัญของคำในเอกสารแต่ละฉบับที่มีความแตกต่างกัน เช่น “good”, “bad” และ “awesome” ที่ปรากฏในเอกสาร doc1, doc2 และ doc3 จึงสามารถใช้ในการแยกแยะคำสำคัญได้ดีขึ้น ในทางกลับกัน “the” และ “is” มีค่า TF-IDF ต่ำหรืออาจเท่ากับ 0 แสดงถึงเป็นคำที่ปรากฏบ่อย แต่ไม่ค่อยมีความสำคัญในเอกสาร

4.2.3) ผลการวัดความคล้ายคลึงโคไซน์ เป็นวิธีการวัดความคล้ายคลึงระหว่างเวกเตอร์ของสองประโยคหรือเอกสาร โดยคำนวณจากผลคูณจุดของเวกเตอร์ทั้งสองและย่อด้วยความยาวของเวกเตอร์แต่ละอัน

เมื่อข้อความถูกแปลงเป็นเวกเตอร์เชิงตัวเลขด้วย TF-IDF จะนำมาคำนวณผลคูณจุด (dot product) ระหว่างเวกเตอร์แต่ละคู่ประโยค โดยนำค่าของ TF-IDF จากเอกสารแต่ละคำมาคูณกันแล้วหาผลรวม หากเอกสารมีค่าเหมือนกันมาก ค่าจะสูงขึ้น ดังตารางที่ 10 จากนั้นจะคำนวณความยาว (norm) ของเวกเตอร์ เพื่อแสดงขนาดของเวกเตอร์ หากค่า TF-IDF ยิ่งสูงแสดงว่าเอกสารมีค่าเฉพาะมากขึ้น ดังตารางที่ 11 ก่อนจะคำนวณค่าความคล้ายคลึงโคไซน์สำหรับแต่ละคู่ของประโยค

ตารางที่ 10 : การคำนวณผลคูณจุดระหว่างเวกเตอร์แต่ละคู่ประโยค

pair	calculation	dot product
doc1 * doc2	$(0 \times 0) + (0.044 \times 0.044) + (0 \times 0) + (0.119 \times 0) + (0 \times 0.119) + (0 \times 0)$	0.001936
doc1 * doc3	$(0 \times 0) + (0.044 \times 0) + (0 \times 0) + (0.119 \times 0) + (0 \times 0) + (0 \times 0.238)$	0
doc2 * doc3	$(0 \times 0) + (0.044 \times 0) + (0 \times 0) + (0 \times 0) + (0.119 \times 0) + (0 \times 0.238)$	0

ตารางที่ 11 : การคำนวณความยาวของเวกเตอร์

document	calculation	norm
doc1	$\sqrt{(0^2 + 0.044^2 + 0^2 + 0.119^2 + 0^2 + 0^2)}$	0.1268
doc2	$\sqrt{(0^2 + 0.044^2 + 0^2 + 0^2 + 0.119^2 + 0^2)}$	0.1268
doc3	$\sqrt{(0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0.238^2)}$	0.238

ตารางที่ 12 : การคำนวณค่าความคล้ายคลึงโคไซน์

pair	calculation	cosine similarity
doc1 , doc2	$\frac{0.001936}{0.1268 \times 0.1268} = \frac{0.001936}{0.016097}$	0.1204
doc1 , doc3	$\frac{0}{0.1268 \times 0.2380} = \frac{0}{0.0302}$	0
doc2 , doc3	$\frac{0}{0.1268 \times 0.2380} = \frac{0}{0.0302}$	0

ตั้งตารางที่ 12 ค่าความคล้ายคลึงระหว่าง doc1 กับ doc2 เท่ากับ 0.1204 แสดงถึงความคล้ายคลึงกันเล็กน้อย เนื่องจากมีคำที่ใช้เหมือนกันหลายคำ ในทางตรงกันข้าม ความคล้ายคลึงระหว่าง doc1 กับ doc3 และ doc2 กับ doc3 เท่ากับ 0 เนื่องจากไม่มีคำใดเลยที่ใช้ร่วมกัน

4.2.4) ผลการวิเคราะห์คำสำคัญ งานวิจัยนี้สร้างคำสำคัญพิจารณาจากการวิเคราะห์คำอัตวิสัย (subjectivity analysis) ของบทวิจารณ์ด้วย TextBlob ที่เป็นหนึ่งในเครื่องมือสำหรับวิเคราะห์ความรู้สึกจากข้อความแบบใช้พจนานุกรมที่มีการระบุคำอัตวิสัยไว้ล่วงหน้า บนแนวคิดของข้อความที่เป็นอัตวิสัย (subjective) มักจะเป็นข้อความที่สะท้อนถึงความคิดเห็น ความรู้สึก หรือการตีความส่วนบุคคล ในทางตรงกันข้าม ข้อความที่เป็นภาววิสัย (objective) จะเป็นข้อความที่เกี่ยวข้องกับข้อมูลที่เป็นข้อเท็จจริง หรือเป็นคำอธิบายที่ไม่เกี่ยวข้องกับความรู้สึกหรือความคิดเห็นส่วนตัว ดังสมการที่ 6

$$Subjectivity = \frac{\sum (subjectivity\ of\ each\ word)}{number\ of\ word\ with\ subjectivity\ value} \quad (6)$$

จากสมการที่ 6 คำอัตวิสัยจะอยู่ในช่วงตั้งแต่ 0 ถึง 1 โดยค่า 0 หมายถึงเป็นข้อความที่มีความเป็นกลางสูง ในทางกลับกัน ค่า 1 หมายถึงข้อความที่เต็มไปด้วยความคิดเห็นหรือความรู้สึกส่วนบุคคล และค่ากลางระหว่าง 0 และ 1 แสดงถึงระดับของความเป็นส่วนตัวที่มีความหลากหลายจากข้อความที่เป็นกลางไปสู่ข้อความที่มีความเป็นส่วนตัวสูง ตัวอย่างข้อความคือ “nice park with many trails and waterfall”

ตารางที่ 13 : ตารางตัวอย่างคำอัตวิสัยของคำแต่ละคำในคลังคำศัพท์

subjectivity scores of each word							
word	nice	park	with	many	trails	and	waterfall
subjectivity	1.0	0	0	0.5	0	0	0

จากตารางที่ 13 คำอัตวิสัยที่บ่งบอกถึงค่าที่เป็นความคิดเห็นส่วนบุคคลที่แสดงออกผ่านคำหรือกลุ่มคำในข้อความ โดยคำว่า “nice” มีคำอัตวิสัยเท่ากับ 1.0 สะท้อนให้เห็นว่าคำนี้เป็นความคิดเห็นส่วนบุคคลต่อมุมมองเชิงบวกอย่างเด่นชัด เช่นเดียวกับคำว่า “many” มีคำอัตวิสัยเท่ากับ 0.5 ซึ่งบ่งบอกถึงค่าที่เป็นกลางระหว่างข้อเท็จจริงและความคิดเห็นส่วนบุคคล ต่อลักษณะการประเมินที่เป็นเชิงปริมาณ แต่ในขณะเดียวกันก็สะท้อนถึงการยอมรับหรือความพึงพอใจในความหลากหลายหรือจำนวนที่มากในทางตรงกันข้าม คำว่า “park”, “trails” และ “waterfall” ที่มีคำอัตวิสัยเท่ากับ 0.0 ที่บ่งชี้ถึงความเป็นข้อเท็จจริง และไม่มีความคิดเห็นส่วนตัวแฝงอยู่ ดังนั้นคำเหล่านี้ทำหน้าที่บ่งบอกลักษณะเฉพาะของสถานที่หรือกิจกรรม โดยไม่เพิ่มความรู้สึกเข้าไปในข้อความ เมื่อนำมาคำนวณคำอัตวิสัยตารางที่ 14 จะเห็นได้ว่าค่าที่มีผลต่อคำอัตวิสัย คือ “nice” และ “many” เท่านั้น ในขณะที่คำอื่น ๆ เป็นข้อเท็จจริง ทำให้คำอัตวิสัย โดยรวมอยู่ที่ 0.75 ดังตารางที่ 15

ตารางที่ 14 : ตัวอย่างการคำนวณคำอัตวิสัย

subjectivity calculation
$Subjectivity = \frac{(1.00 + 0.50)}{2} = \frac{1.50}{2} = 0.75$

ตารางที่ 15 : ผลลัพธ์จากการคำนวณคำอัตวิสัย และความระดับความเข้มข้น

subjectivity	nice	park	with	many	trails	and	waterfall
0.75	1.0	0	0	0.5	0	0	0

จากตารางที่ 15 งานวิจัยนี้คัดเลือกคำสำคัญจากคำที่มีคำอัตวิสัยเท่ากับ 0.0 เท่านั้น จากการพิจารณาความสอดคล้องกับองค์ประกอบของแหล่งท่องเที่ยว 4 องค์ประกอบ ได้แก่ สิ่งดึงดูดใจ การเข้าถึงสถานที่ท่องเที่ยว สิ่งอำนวยความสะดวก และกิจกรรมการท่องเที่ยว องค์ประกอบละ 5 คำ เช่น “waterfall” ซึ่งเป็นหนึ่งในคำสำคัญด้านสิ่งดึงดูดใจ ก่อนจะถูกแปลงให้อยู่ในรูปแบบเวกเตอร์คำสำคัญ (keyword vector) ด้วย TF-IDF ที่สามารถนำไปเปรียบเทียบกับเวกเตอร์บทวิจารณ์ (review vector) ด้วยคำนวณค่าความคล้ายคลึงโคไซน์ ผลลัพธ์ที่ได้จะเป็นค่าตัวเลขที่บ่งบอกระดับความใกล้เคียงระหว่างคำสำคัญกับแต่ละบทวิจารณ์ ก่อนจะจัดเรียงผลลัพธ์จากค่าที่สูงสุดไปหาต่ำสุดสำหรับนำมาคัดเลือกเฉพาะบทวิจารณ์ที่มีความเกี่ยวข้องสูง โดยใช้เกณฑ์ที่กำหนด (threshold) เท่ากับ 0.5 ในการกรองข้อมูล กระบวนการนี้

ช่วยให้มั่นใจว่าข้อความที่นำมาใช้ในการสรุปจะมีเนื้อหาที่สอดคล้องกับคำค้นหาในขั้นตอนต่อไป

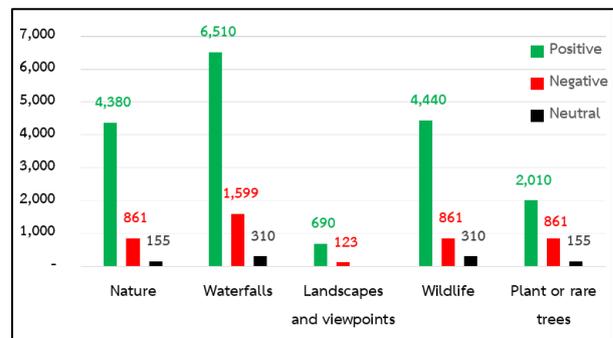
4.2.5) การสรุปข้อความที่ผ่านการคัดเลือกจะถูกปรับแต่งให้อยู่ในรูปแบบที่กระชับและเข้าใจง่ายขึ้น กระบวนการนี้ประกอบด้วยหลายขั้นตอนผสมผสานกัน ได้แก่ การกรองคำซ้ำ เช่นเดียวกับการเลือกคำสำคัญและการตัดคำที่ไม่จำเป็น ร่วมกับการลบคำที่ไม่มีนัยสำคัญต่อเนื้อหา จะสามารถช่วยให้ข้อความมีความกระชับขึ้นโดยไม่ทำให้สาระสำคัญเสียไป ดังตารางที่ 16 ตัวอย่างคำสำคัญคือ “waterfall”

ตารางที่ 16 : ตัวอย่างบทวิจารณ์ที่เกี่ยวข้องกับคำสำคัญ และผลการสรุปข้อความ

relative review	text summarization
My wife and i visited the national park on a small group tour to see the waterfall and to have an elephant ride, we found the park quite beautiful and a nice walk down to the waterfall, the steps are quite steep and a lot of them getting down to the viewing point for the waterfall.	Visited the National park on a small group tour to see waterfall have an elephant ride, we found quite beautiful nice walk down waterfall, steps are steep lot of them getting viewing point for waterfall.
We were at the waterfall, it's beautiful there, you can't miss it!	We were at it's there, you can't miss it! Haew Su Wat beautiful Khao Yai waterfall, adjacent Narok waterfall and it is also most famous Khao Park.
Haew su wat beautiful khao yai waterfall, adjacent to haew narok waterfall. And it is also the most famous waterfall of khao yai national park.	During rainy season, there will be water, but specialty this in summer because when water ebbs, cave or caves behind. Which can up mountain explore cave, area around covered with various trees, covering sunlight, making atmosphere inside only cool, shady comfortable.
During the rainy season, there will be a lot of water. But the specialty of this waterfall is in summer. Because when the water ebbs, it will see a small cave or caves. Behind the waterfall which can walk up the mountain to explore this small cave, the area around the waterfall is covered with various trees.	
Covering the sunlight, making the atmosphere inside only cool, shady and comfortable.	
...	
National park with a beautiful waterfall Great to see	

จากตารางที่ 16 การกรองคำซ้ำจะช่วยให้ข้อความมีความกระชับขึ้น โดยคงเนื้อหาสำคัญไว้ เช่น คำว่า “waterfall” ปรากฏหลายครั้งในบทวิจารณ์จะคงไว้เพียงครั้งเดียว เช่นเดียวกับการเลือกคำสำคัญและการตัดคำที่ไม่จำเป็น เช่น “a”, “an”, “the” หรือ “is” เป็นต้น ซึ่งเป็นคำบุพบทหรือ คำสรรพนามที่ไม่ส่งผลต่อความหมายหลักของข้อความ ก่อนจะจัดระเบียบข้อความใหม่ให้มีโครงสร้างที่เหมาะสม ง่ายต่อการอ่าน และสื่อความหมายได้อย่างชัดเจน โดยใช้การรวมประโยคที่มีเนื้อหาใกล้เคียงกัน หรือการปรับเปลี่ยนลำดับของข้อความเพื่อให้มีความต่อเนื่อง ก่อนจะนำไปใช้สรุปบทวิจารณ์อัตโนมัติจากข้อความจำนวนมาก ซึ่งจะช่วยให้สามารถสรุปข้อมูลที่ง่ายต่อการทำความเข้าใจของผู้ใช้ได้อย่างมีประสิทธิภาพมากขึ้น

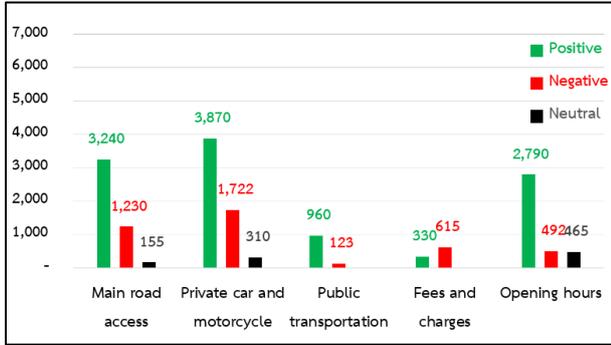
การนำเสนอข้อมูลในรูปแบบกราฟจากการนำค่าความรู้สึกจากแต่ละบทวิจารณ์ที่มีความเกี่ยวข้องมานำเสนอผลลัพธ์ในรูปแบบภาพข้อมูล จำแนกแต่ละประเด็นออกเป็นความรู้สึกเชิงบวก เชิงลบ และเป็นกลาง พร้อมกับอภิปรายผลอีกครั้งใน 4 องค์ประกอบด้านการท่องเที่ยวโดยผู้วิจัย ประกอบด้วยด้านสิ่งดึงดูดใจ ด้านการเข้าถึง ด้านสิ่งอำนวยความสะดวก และด้านกิจกรรม ดังรูปที่ 4 ถึง 7



รูปที่ 4 : การวิเคราะห์คำสำคัญจากองค์ประกอบด้านสิ่งดึงดูดใจ

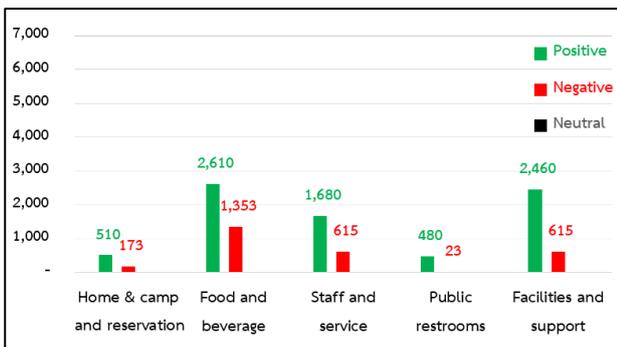
จากภาพที่ 4 อุทยานแห่งชาติเขาใหญ่ได้รับการชื่นชมอย่างสูงจากนักท่องเที่ยวในด้านสิ่งดึงดูดใจจากธรรมชาติที่งดงามและความหลากหลายทางชีวภาพ รวมถึงความอุดมสมบูรณ์ของพรรณไม้หายาก และสัตว์ป่านานาชนิด อีกทั้งมีน้ำตกที่สวยงามเหมาะกับการไปเยี่ยมชม อย่างไรก็ตาม บางช่วงเวลาอาจขาดแคลนน้ำส่งผลให้ไม่สามารถสัมผัสกับความงดงามของน้ำตกได้ตามที่คาดหวัง นอกจากนี้บางจุดในอุทยานที่ต้องเดินขึ้นบันไดที่ชันและจุดที่ไม่มีราวจับ ซึ่งไม่เหมาะสมกับผู้สูงอายุ หรือผู้พิการ ส่วนจุดชมวิวดังกล่าว ซึ่งมียุคที่คนทั้งหลายได้รับการ

ยกย่องว่าเป็นสถานที่ที่ควรเยี่ยมชม แต่ยังมีบางส่วนที่ไม่พบเห็น สัตว์ป่าบางชนิดตามที่คาดหวัง แม้จะมีการนำเที่ยวโดยผู้เชี่ยวชาญ แต่ยังมีบางข้อเสนอแนะที่ต้องการให้การเยี่ยมชมเป็นไปตามความคาดหวังและสะดวกสบายยิ่งขึ้น



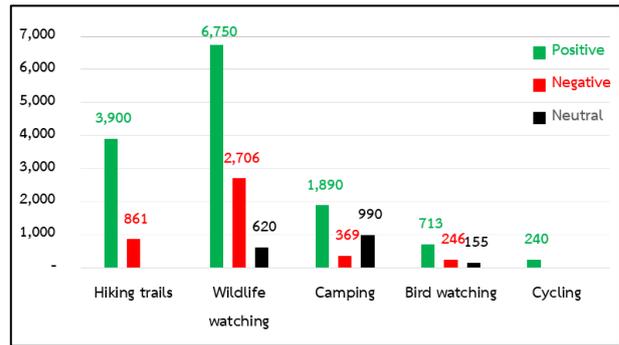
รูปที่ 5 : การวิเคราะห์คำสำคัญจากองค์ประกอบด้านการเข้าถึง

จากรูปที่ 5 การเข้าถึงอุทยานแห่งชาติเขาใหญ่ผ่านถนนสายหลักมีความสะดวกสบาย โดยเฉพาะเส้นทางที่เชื่อมต่อกับจุดท่องเที่ยวสำคัญต่าง ๆ เหมาะสำหรับการเดินทางด้วยรถยนต์หรือจักรยานยนต์ เนื่องจากสามารถแวะตามจุดท่องเที่ยวได้ง่าย แม้จะยังมีข้อจำกัดด้านความแออัดในช่วงฤดูการท่องเที่ยว อีกทั้งการใช้ความเร็วสูงในเขตอุทยานนั้นเสี่ยงต่อความปลอดภัยทั้งของผู้ร่วมใช้ทางและสัตว์ป่า รวมถึงการใช้รถใช้ถนนนอกเวลาทำการยังอาจรบกวนผู้พักค้างคืนและสัตว์ป่า ขณะที่บริการขนส่งสาธารณะยังมีข้อจำกัด ทำให้ต้องพึ่งพารถรับจ้างท้องถิ่นในราคาที่ไม่เป็นมาตรฐาน อย่างไรก็ตาม เวลาทำการยังเป็นข้อจำกัดในการชมสัตว์ป่าในช่วงเช้ามีดหรือช่วงค่ำที่แตกต่างกันไปในแต่ละฤดูกาล ขณะเดียวกันค่าธรรมเนียมที่แตกต่างระหว่างชาวไทยและชาวต่างชาติยังเป็นอีกประเด็นที่ได้รับการวิจารณ์ถึงความเหมาะสม แม้จะเป็นชาวต่างชาติที่มีบัตรอนุญาตทำงานในประเทศไทยก็ตาม



รูปที่ 6 : การวิเคราะห์คำสำคัญจากองค์ประกอบด้านสิ่งอำนวยความสะดวก

จากรูปที่ 6 สิ่งอำนวยความสะดวกภายในอุทยานมีความเหมาะสมกับความต้องการของนักท่องเที่ยว โดยเฉพาะที่พักท่ามกลางธรรมชาติ ห้องพักรีสอร์ทและมีสิ่งอำนวยความสะดวกพื้นฐานครบครัน และการบริการอาหารมีความหลากหลายในราคาที่เหมาะสม อย่างไรก็ตาม ยังมีข้อเสนอแนะที่สะท้อนถึงระยะทางที่ค่อนข้างไกลระหว่างที่พักกับร้านอาหาร หรือมีการใช้เสียงดังในยามวิกาล อาจรบกวนผู้เข้าพักในอุทยาน โดยการบริการของเจ้าหน้าที่ได้รับคำชมในด้านการดูแลรักษาพื้นที่และสัตว์ป่า แต่อาจมีบ้างที่พบกับอุปสรรคทางภาษา นอกจากนี้ยังมีข้อวิจารณ์ถึงการให้ข้อมูลที่ไมชัดเจนเกี่ยวกับแผนที่และเส้นทางในอุทยาน หรือขาดข้อมูลที่จำเป็นประโยชน์เกี่ยวกับอันตรายจากสัตว์มีพิษต่าง ๆ ที่อาจพบระหว่างการเดินทาง อย่างไรก็ตาม ทัศนียภาพที่สวยงามของน้ำพุร้อน ไฟฟ้า น้ำประปา สัญญาณอินเทอร์เน็ต หรือจุดชาร์จโทรศัพท์ก็ได้รับการชื่นชมจากนักท่องเที่ยวเช่นกัน



รูปที่ 7 : การวิเคราะห์คำสำคัญจากองค์ประกอบด้านกิจกรรม

จากรูปที่ 7 กิจกรรมต่าง ๆ ในอุทยานแห่งชาติเขาใหญ่สามารถตอบสนองความต้องการจากนักท่องเที่ยวส่วนใหญ่ได้อย่างน่าพึงพอใจ ด้วยความหลากหลายและเหมาะสมกับสมกับนักท่องเที่ยวในหลากหลายบริษัท โดยการเดินป่าเป็นกิจกรรมที่ได้รับความนิยมอย่างมากจากนักท่องเที่ยวชาวต่างชาติ ซึ่งสามารถสัมผัสธรรมชาติและความสะดวกสบายทางชีวภาพได้อย่างใกล้ชิด เช่นเดียวกับการเดินทางจักรยานที่ชัดเจน แต่บางเส้นทางยังขาดป้ายบอกทางที่ชัดเจน เช่นเดียวกับการส่องสัตว์เป็นอีกกิจกรรมที่น่าสนใจ โดยสามารถพบเห็นสัตว์ป่าร่วมกับการชมธรรมชาติ แต่การมีนักท่องเที่ยวจำนวนมากในช่วงวันหยุดยาวอาจส่งผลต่อความสะดวกในการใช้บริการ รวมถึงลานกางเต็นท์ในพื้นที่ที่ขาดความเป็นระเบียบ และเสียงรบกวนโดยสัตว์ป่าที่ออกหาอาหารใน

ยามค่าคืน นอกจากนั้น การคูณในอุทยานที่ได้รับความนิยมอย่างสูงในปัจจุบัน เนื่องจากความหลากหลายของสายพันธุ์นกที่สามารถพบเห็นได้ในพื้นที่ ซึ่งเป็นจุดดึงดูดนักท่องเที่ยวจากทั่วโลก ดังนั้นหากมีเทคโนโลยีที่มีความรู้จะช่วยให้นักท่องเที่ยวได้รับประสบการณ์ที่ดีขึ้น

5) สรุปผล

พฤติกรรมนักท่องเที่ยวในยุคดิจิทัลมักค้นหาข้อมูลผ่านสื่อออนไลน์ต่าง ๆ เพื่อสนับสนุนการท่องเที่ยว โดยหนึ่งในข้อมูลสำคัญคือคะแนนความพึงพอใจและบทวิจารณ์สาธารณะจากแพลตฟอร์มดิจิทัล แม้ว่าบทวิจารณ์ออนไลน์จะมีข้อดีหลายประการ แต่ปริมาณของบทวิจารณ์จำนวนมากทั้งที่เกี่ยวข้องและไม่เกี่ยวข้อง อาจสิ้นเปลืองเวลาและทรัพยากรในการแยกข้อมูลที่สำคัญ ดังนั้นการวิเคราะห์ข้อความอย่างเป็นระบบจึงเป็นวิธีที่ช่วยให้สามารถเข้าใจมุมมองของนักท่องเที่ยวได้อย่างรอบด้าน

งานวิจัยนี้มุ่งเน้นการวิเคราะห์บทวิจารณ์จากนักท่องเที่ยวชาวต่างชาติที่มีต่ออุทยานแห่งชาติเขาใหญ่ด้วยเครื่องมือวิเคราะห์ความรู้สึกแบบใช้พจนานุกรมเป็นฐาน โดยรวบรวมข้อมูลจากบทวิจารณ์สาธารณะเกี่ยวกับอุทยานแห่งชาติเขาใหญ่จากแพลตฟอร์มออนไลน์ระหว่างวันที่ 1 มกราคม พ.ศ. 2563 ถึง 31 ธันวาคม พ.ศ. 2567 รวมทั้งสิ้น 12,035 รายการ เพื่อจำแนกความคิดเห็นของนักท่องเที่ยวออกเป็นเชิงบวก เชิงลบ และเป็นกลางจาก TextBlob Flair และ VADER

ผลการทดสอบ พบว่า VADER มีค่าความถูกต้องเฉลี่ยสูงที่สุดเท่ากับร้อยละ 76 ในการจำแนกความรู้สึกของบทวิจารณ์ อย่างไรก็ตาม ยังพบแนวโน้มของอคติบางประการในกระบวนการวิเคราะห์ เช่น การให้ค่าน้ำหนักกับข้อความเชิงบวกมากเกินไป ส่งผลให้การจำแนกความรู้สึกในบางกรณีไม่ตรงกับความรู้สึกจากบทวิจารณ์ นอกจากนี้ยังพบตัวอย่างของบทวิจารณ์ที่คะแนนความพึงพอใจขัดแย้งกับเนื้อหา เช่น บทวิจารณ์ที่ให้คะแนนสูงแต่มีเนื้อหาวิพากษ์วิจารณ์เกี่ยวกับค่าธรรมเนียมและการให้บริการหรือบทวิจารณ์ที่ให้คะแนนต่ำแต่มีเนื้อหาชื่นชมธรรมชาติของอุทยาน ซึ่งแสดงให้เห็นถึงข้อจำกัดของการใช้เพียงผลการวิเคราะห์ความรู้สึกจากข้อความในการสะท้อนความคิดเห็นของนักท่องเที่ยว ซึ่งอาจไม่เพียงพอ เนื่องจากความคิดเห็นของนักท่องเที่ยวมีความซับซ้อนและหลากหลาย อีกทั้งความคาดหวังของนักท่องเที่ยวแต่ละคนนั้นไม่มีขีดจำกัด และยากที่จะเปรียบเทียบความแตกต่างกันในหลากหลายมุมมองด้วยหลักเกณฑ์ที่เป็นมาตรฐานเดียวกัน

เพื่อลดข้อจำกัดของการใช้ผลการวิเคราะห์ความรู้สึกจากข้อความเพียงอย่างเดียวในการสะท้อนความคิดเห็นของนักท่องเที่ยว และเพิ่มการนำผลลัพธ์ไปใช้ประโยชน์อย่างมีประสิทธิภาพ งานวิจัยนี้วิเคราะห์ข้อมูลเชิงคุณภาพจากบทวิจารณ์ โดยการแปลงข้อความเป็นเวกเตอร์ด้วยวิธี TF-IDF และเปรียบเทียบความสัมพันธ์ระหว่างคำสำคัญและบทวิจารณ์ที่เกี่ยวข้องด้วยการวัดความคล้ายคลึงโคไซน์ ก่อนจะสรุปข้อความด้วยการประมวลผลภาษาธรรมชาติร่วมกับการนำเสนอภาพข้อมูลตามองค์ประกอบด้านการท่องเที่ยว ผลการวิเคราะห์พบว่าแม้ว่าอุทยานแห่งชาติเขาใหญ่จะได้รับการชื่นชมในด้านความงามทางธรรมชาติ แต่ก็ยังมีข้อจำกัดบางประการที่ถูกกล่าวถึงในบทวิจารณ์ของนักท่องเที่ยวชาวต่างชาติ ปัญหาหลักที่พบ ได้แก่ ค่าธรรมเนียมเข้าชมที่แตกต่างกันระหว่างนักท่องเที่ยวชาวไทยและชาวต่างชาติ นอกจากนี้ ระบบขนส่งสาธารณะยังขาดความสะดวกและไม่เพียงพอ ทำให้การเดินทางภายในอุทยานต้องพึ่งพารถส่วนตัวหรือรถเช่าเป็นหลัก ตลอดจนเวลาทำการของอุทยานและข้อจำกัดในการเข้าถึงบางพื้นที่อาจไม่สอดคล้องกับการชมสัตว์ป่าในช่วงเวลาที่แตกต่างกัน

ผลลัพธ์จากงานวิจัยนี้แสดงให้เห็นว่าการประมวลผลภาษาธรรมชาติร่วมกับการวิเคราะห์ความรู้สึกจากข้อความสามารถช่วยให้เข้าใจความคิดเห็นของนักท่องเที่ยวได้ดียิ่งขึ้น ทั้งในแง่ของข้อดีและข้อจำกัดของแหล่งท่องเที่ยว ซึ่งจะนำไปสู่การตัดสินใจของนักท่องเที่ยว อีกทั้งยังสามารถนำมากำหนดแนวทางการพัฒนาอุทยานแห่งชาติเขาใหญ่ให้สามารถปรับปรุงคุณภาพการให้บริการและตอบสนองความต้องการของนักท่องเที่ยว นอกจากนี้ ยังเป็นต้นแบบในการประยุกต์ใช้กับแหล่งท่องเที่ยวอื่น ๆ เพื่อยกระดับคุณภาพการให้บริการและเสริมสร้างศักยภาพในการแข่งขันของอุตสาหกรรมการท่องเที่ยวของประเทศไทยในระยะยาว

6) ข้อเสนอแนะ

การวิเคราะห์ความรู้สึกจากข้อความในครั้งต่อไปควรใช้การเรียนรู้เชิงลึกอย่างโมเดล Transformer เพื่อเพิ่มประสิทธิภาพในการประมวลผลข้อมูล นอกจากนี้ ควรมีการประเมินผลจากผู้ใช้งานเพื่อพิจารณาว่าผลลัพธ์จากการวิเคราะห์ข้อความมีความถูกต้อง น่าเชื่อถือ และตรงตามความต้องการของกลุ่มเป้าหมาย ทั้งนี้การประเมินจะช่วยให้สามารถปรับปรุงและเพิ่มประสิทธิภาพของผลลัพธ์ก่อนนำไปใช้งานต่อไป



REFERENCES

- [1] Z. Zhu, Y. Zhao, and J. Wang, "The impact of destination online review content characteristics on travel intention: experiments based on psychological distance perspectives," *Aslib J. Inf. Manage.*, vol. 76, no. 1, pp. 42–64, 2024.
- [2] A. Sihombing, L.-W. Liu, and P. Pahrudin, "The impact of online marketing on tourists' visit intention: Mediating roles of trust," *J. Tourism, Heritage Serv. Marketing*, vol. 10, no. 2, pp. 15–23, 2024.
- [3] L. Lei and D. Liu, *Conducting Sentiment Analysis*. Cambridge, U.K.: Cambridge University Press, 2021.
- [4] R. Rachmiani, N. Kintan Oktadinna, and T. Rachmat Fauzan, "The impact of online reviews and ratings on consumer purchasing decisions on e-commerce platforms," *Int. J. Manage. Sci. Inf. Technol.*, vol. 4, no. 2, pp. 504–515, 2024.
- [5] National Statistical Office of Thailand, "Number of tourists in national parks, fiscal year 2022," 2022. [Online]. Available: <https://catalog.dnp.go.th/dataset/91d66c6a-5b88-41ee-8597-11be8aec5aa6/resource/525f31b3-d572-41fa-a822-04f5f3c12ab7/download/tourism65.xlsx>
- [6] National Statistical Office of Thailand, "Number of tourists in national parks, fiscal year 2023," 2023. [Online]. Available: <https://catalog.dnp.go.th/dataset/91d66c6a-5b88-41ee-8597-11be8aec5aa6/resource/70ae333d-19fc-4cdb-a60c-dc43bbf74ad6/download/tourism66.xlsx>
- [7] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2008.
- [8] T. Snake, *Natural Language Processing Practical Using Transformers with Python*. 2022.
- [9] L. T. Dela Cruz *et al.*, "An enhancement of support vector machine in the context of sentiment analysis applied in scraped data from Tripadvisor hotel reviews," *Int. J. Comput. Sci. Res.*, vol. 8, pp. 3235–3251, 2024.
- [10] S. Jardim and C. Mora, "Customer reviews sentiment-based analysis and clustering for market-oriented tourism services and products development or positioning," *Procedia Comput. Sci.*, vol. 196, pp. 199–206, 2022.
- [11] K. Puh and M. Bagić Babac, "Predicting sentiment and rating of tourist reviews using machine learning," *J. Hosp. Tour. Insights*, vol. 6, no. 3, pp. 1188–1204, 2023.
- [12] I. Perikos and A. Diamantopoulos, "Explainable aspect-based sentiment analysis using transformer models," *Big Data Cogn. Comput.*, vol. 8, no. 11, 2024, Art. no. 141.
- [13] K. W. Trisna, J. Huang, H. Lei, and E. M. Dharma, "From context-independent embedding to transformer: Exploring sentiment classification in online reviews with deep learning approaches," *J. Theor. Appl. Inf. Technol.*, vol. 102, no. 19, pp. 6980–7003, 2024.
- [14] P. Suanpang, P. Jamjuntr, and P. Kaewyong, "Sentiment analysis with a TextBlob package Implications for tourism," *J. Manage. Inf. Decis. Sci.*, vol. 24, no. S6, pp. 1–9, 2021.
- [15] I. Khumtaveeporn and K. Wattanasuwan, "AI sentiment analysis for destination branding: A case study of Buriram, Thailand," *J. Bus. Adm.*, vol. 46, no. 180, pp. 50–74, 2023.
- [16] K. Barik and S. Misra, "Analysis of customer reviews with an improved VADER lexicon classifier," *J. Big Data*, vol. 11, 2024, Art. no. 10.
- [17] D. Buhalis, "Marketing the competitive destination of the future," *Tourism Manage.*, vol. 21, no. 1, pp. 97–116, 2000.
- [18] R. Mitchell, *Web Scraping with Python*, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2024.
- [19] M. Arief and N. A. Samsudin, "Hybrid approach with VADER and multinomial logistic regression for multiclass sentiment analysis in online customer review," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 12, pp. 311–320, 2023.
- [20] D. Tay, *Data Analytics for Discourse Analysis with Python: The Case of Therapy Talk*. New York, NY, USA: Routledge, 2024.
- [21] B. Liu, *Sentiment Analysis and Opinion Mining*. Cham, Switzerland: Springer, 2022.