

การประยุกต์ใช้การเรียนรู้เชิงลึกและการประมาณค่าความหนาแน่นแบบเคอร์เนลเพื่อระบุพื้นที่วิกฤต: กรณีศึกษาการจัดการการท่องเที่ยวอย่างยั่งยืนในภาคเหนือ

ทบทอง ชันเจริญ^{1*}, สุวิชะ รัตตะระมย์², ทวีศักดิ์ สัมมา¹ และ ภูมิพัฒน์ อุณบ้าน¹

Deep Learning and Kernel Density Estimation for Critical Area Identification: A Sustainable Tourism Management Case Study in Northern Thailand

Tobthong Chancharoen^{1*}, Suwitchaya Rattarom², Taweesak Samma¹ and Phummipat Oonban¹

¹ Geoinformatics Program, Faculty of Computer Science and Information Technology, Rambhai Barni Rajabhat University, Chanthaburi, 22000

² Computer Science Program, Faculty of Computer Science and Information Technology, Rambhai Barni Rajabhat University, Chanthaburi, 22000

* Corresponding author: tobthong.c@rbru.ac.th

Received: 8 December 2025; Revised: 27 January 2026; Accepted: 20 April 2026

บทคัดย่อ

การท่องเที่ยวในภาคเหนือของไทยมีความสำคัญเชิงยุทธศาสตร์ต่อเศรษฐกิจภูมิภาค แต่การจัดการเพื่อความยั่งยืนยังคงเป็นความท้าทายสำคัญ งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยความไม่พึงพอใจของนักท่องเที่ยวต่างชาติต่อแหล่งท่องเที่ยวทางธรรมชาติในภาคเหนือ โดยรวบรวมบทวิจารณ์ 8,193 บทวิจารณ์ จาก TripAdvisor (พ.ศ. 2557–2566)

การศึกษานี้สร้างแนวทางใหม่ โดยการบูรณาการแบบจำลอง BERT-based Deep Learning เพื่อจำแนกความรู้สึกเชิงลบ ซึ่งเป็นเป้าหมายหลักของการวิเคราะห์ โดยแบบจำลองให้ค่าความแม่นยำ (Precision) สูงถึงร้อยละ 87 ในการระบุกลุ่มข้อมูลเชิงลบ จากนั้นประยุกต์ใช้การวิเคราะห์ N-Gram เพื่อจัดกลุ่มประเด็นปัญหา และเทคนิค Kernel Density Estimation (KDE) ด้วยแบนด์วิดท์ (h) 0.3 กม. เพื่อเปลี่ยนข้อมูลข้อความ (Textual data) ให้กลายเป็นพื้นผิวความหนาแน่นเชิงพื้นที่ (Density surface) ที่ระบุพื้นที่วิกฤตได้อย่างแม่นยำ

ผลการศึกษาพบว่าปัจจัยหลักที่ขับเคลื่อนความไม่พึงพอใจคือ ความคุ้มค่าและราคา (Value & Pricing) โดยมีหลักฐานเชิงประจักษ์จากวลี 'not worth' (83 ครั้ง) และข้อร้องเรียนเรื่องราคาค่าเข้าชมสูง ('300 บาท') ซึ่งมีความสัมพันธ์เชิงพื้นที่สอดคล้องกับปัญหาด้านสิ่งแวดล้อมและการบำรุงรักษา (เช่น 'no water' และ 'dirty') โดยพื้นที่วิกฤต (Hotspots) กระจุกตัวอย่างเข้มข้นในบริเวณอุทยานแห่งชาติดอยอินทนนท์ และอำเภอปาย จังหวัดแม่ฮ่องสอน

ผลลัพธ์จากการระบุ Hotspot นี้ทำหน้าที่เป็น ข้อมูลเชิงลึกเชิงพื้นที่ (Actionable Spatial Insights) ที่ช่วยสนับสนุนหน่วยงานภาครัฐในการตัดสินใจเชิงนโยบายอย่างตรงจุด ทั้งการทบทวนโครงสร้างค่าธรรมเนียมให้สอดคล้องกับคุณค่า และการจัดการขีดความสามารถในการรองรับนักท่องเที่ยว (Carrying Capacity) ในระดับท้องถิ่นเพื่อยกระดับความยั่งยืนของการท่องเที่ยวในพื้นที่วิกฤตได้อย่างมีประสิทธิภาพ

คำสำคัญ: การวิเคราะห์ความรู้สึก, BERT, การประมาณค่าความหนาแน่นแบบเคอร์เนล, ภาคเหนือของประเทศไทย, นโยบายการท่องเที่ยวอย่างยั่งยืน

¹ สาขาวิชาภูมิสารสนเทศ คณะวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏรำไพพรรณี จังหวัดจันทบุรี 22000

² สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏรำไพพรรณี จังหวัดจันทบุรี 22000

Abstract

Tourism in Northern Thailand is a strategic pillar of the regional economy, yet managing its sustainability remains a critical challenge. This research aims to analyze the factors driving dissatisfaction among international tourists toward natural attractions in Northern Thailand. A dataset of 8,193 reviews was collected from TripAdvisor, spanning a ten-year period from 2014 to 2023.

This study presents an academic contribution by integrating a BERT-based Deep Learning model to classify negative sentiments—the primary focus of this analysis. The model achieved a high precision of 87% in identifying negative class data. Subsequently, N-Gram Analysis was employed for issue clustering, and Kernel Density Estimation (KDE) with a bandwidth (h) of 0.3 km was applied to transform textual data into a continuous spatial density surface, enabling the precise identification of dissatisfaction hotspots.

The findings reveal that the primary driver of dissatisfaction is Value & Pricing, evidenced by the prevalence of the phrase 'not worth' (83 mentions) and specific complaints regarding high entry fees (e.g., '300 baht'). These issues are spatially correlated with environmental and maintenance concerns (e.g., 'no water' and 'dirty'). Negative hotspots are intensely concentrated in key locations, notably Doi Inthanon National Park and Pai District in Mae Hong Son Province.

The results provide actionable spatial insights to support government agencies in targeted policymaking. By identifying these crisis zones, the study facilitates informed decisions on revising entry fee structures and managing tourist carrying capacity within hotspots to enhance the long-term sustainability of Northern Thailand's natural tourism.

Keywords: Sentiment Analysis, BERT, Kernel Density Estimation (KDE), Northern Thailand, Sustainable Tourism Policy

Introduction

ภาคเหนือของประเทศไทยเป็นหนึ่งในเสาหลักของการท่องเที่ยวของประเทศ โดยมีบทบาทสำคัญในการขับเคลื่อนเศรษฐกิจในระดับภูมิภาค แหล่งท่องเที่ยวทางธรรมชาติ เช่น อุทยานแห่งชาติดอยอินทนนท์ (เชียงใหม่) และปายแคนยอน (แม่ฮ่องสอน) ดึงดูดนักท่องเที่ยวต่างชาติจำนวนมากด้วยทัศนียภาพภูเขาที่สวยงาม ภูมิอากาศที่เย็นสบาย และความหลากหลายทางชีวภาพ (Ministry of Tourism and Sports, 2016; 2020; 2023) อย่างไรก็ตาม ความสำเร็จด้านจำนวนนักท่องเที่ยวมาพร้อมกับความท้าทายที่เร่งด่วน โดยเฉพาะอย่างยิ่งปัญหาด้าน ความยั่งยืน ซึ่งรวมถึงการเสื่อมโทรมของทรัพยากรธรรมชาติ ปัญหาความแออัด (Overcrowding) ในช่วงฤดูท่องเที่ยว และปัญหาคุณภาพอากาศและมลภาวะที่ส่งผลกระทบต่อประสบการณ์ของนักท่องเที่ยว (Secretariat of Cabinet, 2011; Ministry of Tourism and Sports, 2016;)

ในยุคปัจจุบัน สื่อออนไลน์ได้กลายเป็นช่องทางหลักที่นักท่องเที่ยวใช้ในการค้นคว้าข้อมูลและแบ่งปันประสบการณ์การเดินทาง (Cui & He, 2019; Lassen et al., 2014; Ministry of Tourism and Sports, 2021) แพลตฟอร์มอย่าง TripAdvisor ถือเป็นแหล่งข้อมูลขนาดใหญ่ที่รวบรวมความคิดเห็นแบบเรียลไทม์ ซึ่งรู้จักกันในชื่อ Electronic Word-of-Mouth (eWOM) (Li et al., 2017; Valdivia et al., 2017; Donthu et al., 2021) ข้อมูลเหล่านี้เป็นตัวชี้วัดที่ทรงคุณค่าและสะท้อนถึงพฤติกรรมและความพึงพอใจของนักท่องเที่ยวต่อคุณภาพของแหล่งท่องเที่ยวและบริการได้อย่างตรงไปตรงมา (Phithakkitnukoon, 2023; Ministry of Tourism and Sports, 2024) การวิเคราะห์ข้อมูลข้อความปริมาณมหาศาลนี้จึงเป็นโอกาสสำคัญในการทำความเข้าใจความต้องการและปัญหาที่แท้จริง

การจัดการและตีความข้อมูลข้อความขนาดใหญ่นี้ต้องอาศัยเทคนิคที่ซับซ้อน เช่น การวิเคราะห์ความรู้สึก (Sentiment Analysis) หรือการทำเหมืองความคิดเห็น (Opinion Mining) ซึ่งเป็นส่วนหนึ่งของการประมวลผลภาษาธรรมชาติ (Natural Language Processing – NLP) (Alaei et al., 2017; Sarker, 2021) งานวิจัยนี้เลือกใช้แบบจำลอง BERT-based Deep Learning (Pérez et al., 2024) เนื่องจากประสิทธิภาพที่เหนือกว่าในการทำความเข้าใจบริบท (Contextual Understanding) และการตีความอารมณ์ที่แฝงอยู่ในภาษาบทวิจารณ์การท่องเที่ยวซึ่งมักมีความซับซ้อนและกำกวม (Devlin et al., 2018; Yuan, 2024) เทคนิคนี้ไม่เพียงช่วยจำแนกข้อารมณได้แม่นยำ แต่ยังช่วยระบุประเด็นปัญหาเฉพาะเจาะจง (Aspect-based issues) ที่ส่งผลต่อความไม่พึงพอใจได้อีกด้วย (Hashim et al., 2020)

อย่างไรก็ตาม แม้จะมีการนำการวิเคราะห์ความรู้สึกมาใช้ในบริบทการท่องเที่ยวไทยบ้างแล้ว แต่ส่วนใหญ่ยังคงมุ่งเน้นเพียงการจำแนกความรู้สึกโดยรวมในระดับมหภาค (Kuhmanee, 2017; Leelawat, 2022) ช่องว่างงานวิจัย (Research Gap) ที่สำคัญคือ (1) การวิเคราะห์เชิงเนื้อหาแบบลงลึก (Aspect-Based) เพื่อระบุประเด็นปัญหา และ (2) ขาดการวิเคราะห์เชิงพื้นที่ (Spatial Analysis) ที่เชื่อมโยงประเด็นปัญหาเข้ากับตำแหน่งทางภูมิศาสตร์อย่างชัดเจน ทำให้ยากต่อการกำหนดมาตรการแก้ไขที่ตรงจุด

ดังนั้น งานวิจัยนี้จึงมุ่งปิดช่องว่างดังกล่าว โดยการนำเทคนิค Deep Learning ที่มีความแม่นยำสูง มาบูรณาการร่วมกับการวิเคราะห์ความหนาแน่นเชิงพื้นที่ด้วย Kernel Density Estimation (KDE) เพื่อเปลี่ยนข้อมูลความไม่พึงพอใจเชิงคุณภาพให้กลายเป็นพิกัดพื้นที่วิกฤต (Hotspots) ผลการวิจัยจะนำไปสู่ข้อเสนอแนะเชิงนโยบายที่เน้นการจัดสรรทรัพยากรและการจัดการความยั่งยืนในพื้นที่เป้าหมายได้อย่างมีประสิทธิภาพและแม่นยำยิ่งขึ้น

Objectives

1. เพื่อประเมินและจำแนกความรู้สึก (Sentiment Classification) ของนักท่องเที่ยวต่างชาติต่อแหล่งท่องเที่ยวทางธรรมชาติในภาคเหนือเชิงปริมาณ โดยเน้นประสิทธิภาพของแบบจำลองในการระบุความไม่พึงพอใจ
2. เพื่อระบุประเด็นปัญหาเฉพาะเจาะจง (Thematic Issues) และคำสำคัญเชิงบริบทที่สะท้อนถึงสาเหตุของความไม่พึงพอใจผ่านการวิเคราะห์ N-Gram ในข้อมูลระดับมหภาค
3. เพื่อวิเคราะห์การกระจายตัวและความเข้มข้นเชิงพื้นที่ (Spatial Hotspots) ของความไม่พึงพอใจโดยใช้เทคนิค Kernel Density Estimation (KDE) เพื่อเป็นข้อมูลสนับสนุนการตัดสินใจเชิงนโยบายและการวางแผนเชิงพื้นที่อย่างยั่งยืน

Methodology

วิธีการวิจัยนี้แบ่งออกเป็น 5 ขั้นตอนหลัก ได้แก่ การรวบรวมข้อมูล การประมวลผลล่วงหน้าและการเข้ารหัสเชิงพื้นที่ การวิเคราะห์ความรู้สึกเชิงปริมาณ การวิเคราะห์เชิงเนื้อหาด้วย N-Gram และ การวิเคราะห์เชิงพื้นที่ด้วย Kernel Density Estimation (KDE) โดยมีกรอบการทำงานแสดงดัง Figure 1

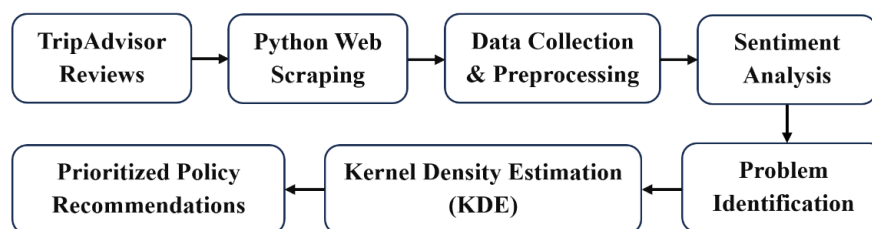


Figure 1 Research framework: Integration of deep learning and geospatial analysis

1. Data collection

ข้อมูลบทวิจารณ์รวมทั้งพิกัดทางภูมิศาสตร์ถูกรวบรวมจากแพลตฟอร์ม TripAdvisor ซึ่งเป็นแหล่งข้อมูล Electronic Word-of-Mouth (eWOM) ที่น่าเชื่อถือ (Jeacle & Carter, 2011; Filieri et al., 2015; Valdivia et al., 2017) โดยมีขอบเขตการรวบรวมเฉพาะแหล่งท่องเที่ยวทางธรรมชาติใน 9 จังหวัดภาคเหนือของประเทศไทย ได้แก่ เชียงราย เชียงใหม่ น่าน พะเยา แพร่ แม่ฮ่องสอน ลำปาง ลำพูน และ อุตรดิตถ์ (Figure 2) ข้อมูลที่รวบรวมประกอบด้วย: ชื่อสถานที่ ประเภทของแหล่งท่องเที่ยว วันที่ หัวข้อ บทวิจารณ์ (ภาษาอังกฤษ) และ พิกัดภูมิศาสตร์ (ละติจูด/ลองจิจูด) จำนวนบทวิจารณ์ที่ใช้ในการวิเคราะห์รวมทั้งสิ้น 8,193 บทวิจารณ์ ในช่วงปี พ.ศ. 2557 ถึง 2566 รวม 10 ปี

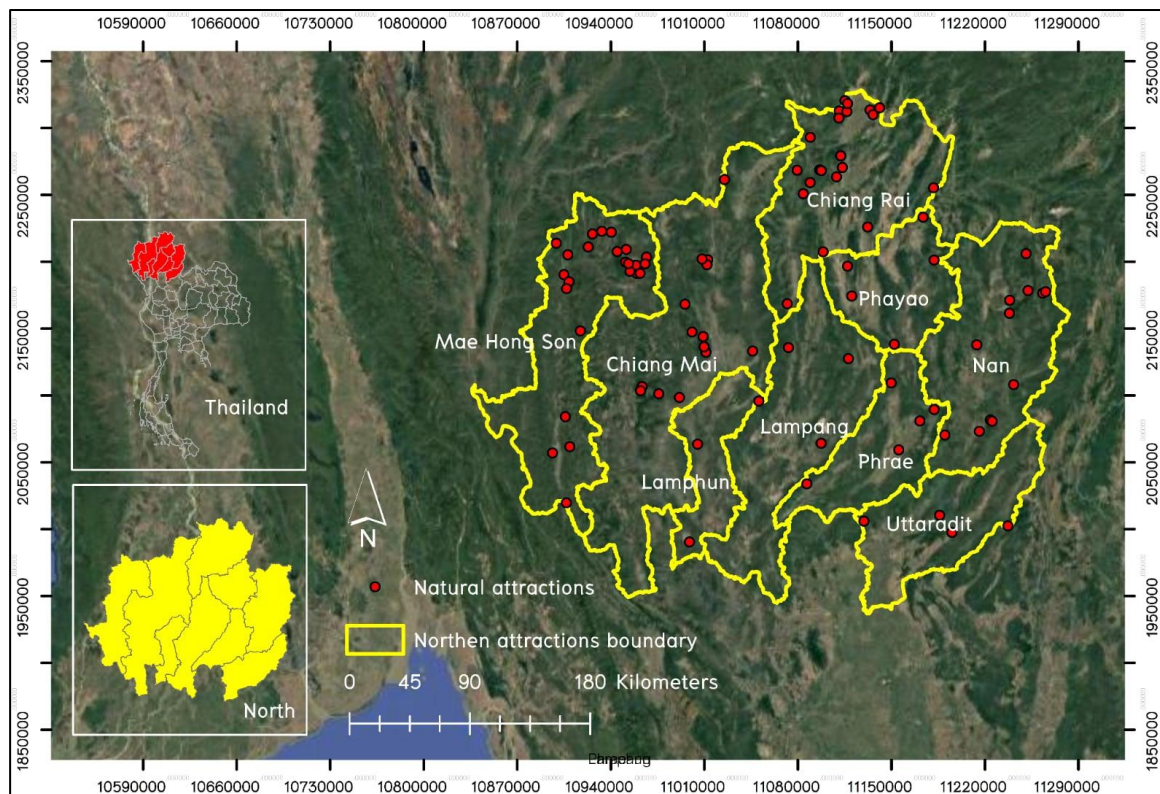


Figure 2 Study area map: natural attractions in Northern Thailand

ข้อมูล HTML ที่ไม่มีโครงสร้างจากหน้าเว็บจะถูกสกัด (Web scraping) ด้วย Python Libraries (Rossum, 1995) เช่น BeautifulSoup (Richardson, 2004) Pandas (McKinney, 2010) และ Selenium (Muthukadan, 2024) จะถูกจัดเก็บในฐานข้อมูลรูปแบบ csv

2. Preprocessing and geocoding

กรอบการทำงานเหมืองข้อความ (Text Mining) ของการศึกษานี้ดำเนินการตามลำดับขั้นตอนหลัก 3 ขั้นตอน ซึ่งเป็นการบูรณาการเทคนิค Deep Learning เข้ากับการวิเคราะห์เชิงพื้นที่ (Geospatial Analysis):

2.1 การประมวลผลข้อความ: ทำความสะอาดข้อมูลข้อความ (Text Cleaning) โดยเริ่มจากการกำจัดอักขระพิเศษ (Special Characters) (Prabha & Srikanth, 2021) การลบ Stop-words (คำที่ไม่จำเป็น เช่น 'the', 'a') (Lima et al., 2021) และการปรับรูปแบบคำให้อยู่ในรูปแบบเดียวกัน (Lemmatization) (Ramadhani et al., 2021) และ แบ่งข้อความเป็นคำ (Tokenization)

2.2 การจำแนกประเภทบทวิจารณ์: บทวิจารณ์ทั้งหมดถูกแบ่งออกเป็น บทวิจารณ์เชิงบวก (Positive Reviews) เป็นกลาง (Neutral Reviews) และ เชิงลบ (Negative Reviews) (Pérez et al., 2024) โดยมีบทวิจารณ์เชิงลบจำนวน 640 บทวิจารณ์ ที่ถูกนำไปใช้ในการวิเคราะห์เชิงลึกต่อไป

2.3 การเข้ารหัสเชิงพื้นที่ (Geocoding): พิกัดภูมิศาสตร์ (ละติจูดและลองจิจูด) ของสถานที่ท่องเที่ยวแต่ละแห่งถูกใช้เพื่อเป็นข้อมูลจุด (Point Features) เพื่อให้สามารถเชื่อมโยงข้อมูลข้อความเข้ากับตำแหน่งทางภูมิศาสตร์สำหรับการวิเคราะห์เชิงพื้นที่ (Esri, 2021)

3. Quantitative sentiment analysis

การวิเคราะห์ความรู้สึกใช้แบบจำลอง Deep Learning (BERT-based Pysentimiento) ซึ่งได้รับการฝึกฝนมาเพื่อจำแนกความรู้สึกในข้อความอย่างมีประสิทธิภาพ แบบจำลองนี้ถูกใช้เพื่อการจำแนก และการประเมินความถูกต้อง ของการจำแนกกลุ่มบทวิจารณ์ทั้งหมด (Devlin et al., 2018; Pérez et al., 2024)

ในการประเมินประสิทธิภาพใช้ชุดข้อมูลจากผู้เชี่ยวชาญ 4,509 บทวิจารณ์: สำหรับฝึกฝนแบบจำลอง 3,607 บทวิจารณ์ (80%) ทดสอบแบบจำลอง 451 บทวิจารณ์ (10%) และประเมินประสิทธิภาพของแบบจำลอง 451 บทวิจารณ์ (10%) ด้วยค่า Hyperparameter: Epochs (3, 4, 5), Batch (32) และ Learning Rate ($2e-5$ to $1e-4$)

นอกจากนี้ เพื่อจัดการกับปัญหาความไม่สมดุลของข้อมูล (Imbalanced Data) ระหว่างความคิดเห็นเชิงบวกและเชิงลบ งานวิจัยนี้ได้ใช้เทคนิค Class Weighting ในขั้นตอนการฝึกฝนแบบจำลอง เพื่อกำหนดน้ำหนักความสำคัญให้กับกลุ่มข้อมูลที่มีจำนวนน้อย (Minority Class) และป้องกันอคติ (Bias) ในการทำนายของแบบจำลอง (Henning et al., 2023; Wu et al., 2024) รวมถึงใช้การวัดผลด้วยเมทริกซ์ความคลุมเครือ (Confusion Matrix) นอกเหนือการพิจารณาความถูกต้องโดยรวม (Accuracy) เพียงอย่างเดียว และเพื่อยืนยันประสิทธิภาพในการจำแนกความคิดเห็นเชิงลบอย่างแม่นยำ งานวิจัยนี้จึงเน้นที่ผลการทดสอบประสิทธิภาพของแบบจำลองจากค่าความแม่นยำ (Precision) ในการจำแนกความคิดเห็นเชิงลบเป็นหลัก

Table 1 นอกจากการพิจารณาความถูกต้องโดยรวม (Accuracy) เพื่อสะท้อนประสิทธิภาพของแบบจำลองในการระบุความคิดเห็นเชิงลบอย่างแม่นยำ งานวิจัยนี้จึงคำนวณค่าดัชนีชี้วัดประสิทธิภาพตามสมการดังนี้: (Pleerux, 2023)

Table 1 The confusion matrix used to verify Pysentimiento analysis

Actual / Prediction	Positive	Neutral	Negative
Positive	<i>a</i>	<i>b</i>	<i>c</i>
Neutral	<i>d</i>	<i>e</i>	<i>f</i>
Negative	<i>g</i>	<i>h</i>	<i>i</i>

$$Precision_{negative} = \frac{i}{(c+f+i)} \quad (\text{Equation 1})$$

$$Recall_{negative} = \frac{i}{(g+h+i)} \quad (\text{Equation 2})$$

$$F1 - score_{negative} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (\text{Equation 3})$$

$$A = \frac{a+e+i}{a+b+c+d+e+f+g+h+i} \quad (\text{Equation 4})$$

4. Content analysis: N-Gram analysis

เพื่อระบุประเด็นปัญหาที่เฉพาะเจาะจง (Thematic Issues) ที่ผลักดันความไม่พึงพอใจ จึงทำการวิเคราะห์ N-Gram (Bird et al., 2009) เฉพาะในชุดข้อมูลบทวิจารณ์เชิงลบที่ได้รับการประมวลผลแล้ว การวิเคราะห์นี้มุ่งเน้นการค้นหาลีหรือคำที่มีความถี่สูง (High Frequency Keywords) ที่มีความสัมพันธ์โดยตรงกับประสบการณ์เชิงลบ จากนั้นจึงใช้การจัดกลุ่มหัวข้อโดยการตัดสินใจ (Manual Thematic Grouping) เพื่อจัดหมวดหมู่ปัญหาออกเป็นกลุ่มหลัก ได้แก่ I. ความคุ้มค่าและราคา (Value & Pricing) II. สิ่งแวดล้อมและการบำรุงรักษา (Environmental & Maintenance) และ III. การเข้าถึงและโครงสร้างพื้นฐาน (Accessibility & Infrastructure)

5. Spatial analysis: Kernel Density Estimation (KDE)

เพื่อจัดทำข้อมูลเชิงพื้นที่สำหรับการจัดการการท่องเที่ยว (Actionable Spatial Insights) การกระจายตัวของภูมิศาสตร์ของความรู้สึกเชิงลบจึงถูกวิเคราะห์โดยใช้ Kernel Density Estimation (KDE) KDE เป็นวิธีการวิเคราะห์เชิงพื้นที่แบบไม่พารามิเตอร์ที่ใช้ประมาณการฟังก์ชันความหนาแน่นความน่าจะเป็นของข้อมูลจุด โดยเปลี่ยนจุดแยกส่วน (ตำแหน่งของสถานที่ท่องเที่ยวที่มีบทวิจารณ์เชิงลบ) ให้เป็นพื้นผิวต่อเนื่องที่แสดงความเข้มข้นทางภูมิศาสตร์ของปัญหา (Hotspots) (Rogel-Salazar, 2022)

ความหนาแน่นโดยประมาณ ณ ตำแหน่ง x คำนวณได้โดยใช้ Equation 5 (Mission, 2024):

$$\hat{f}(x) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (\text{Equation 5})$$

โดยที่:	$\hat{f}(x)$	คือ	ความหนาแน่นโดยประมาณที่ตำแหน่ง x
	n	คือ	จำนวนจุดสังเกตการวิจารณ์เชิงลบ
	h	คือ	แบนด์วิดท์ (หรือพารามิเตอร์การปรับให้เรียบ) ซึ่งกำหนดรัศมีอิทธิพลรอบแต่ละจุด
	K	คือ	ฟังก์ชันเคอร์เนล (ฟังก์ชันน้ำหนัก ซึ่งมักเป็นเคอร์เนลแบบเกาส์เซียนหรือควอร์ดิก)
	x_i	คือ	ตำแหน่งของการวิจารณ์เชิงลบครั้งที่ i

การตั้งค่าและการประยุกต์ใช้:

1. ข้อมูลนำเข้า: ใช้ข้อมูลพิกัดของบวรวิหารเชิงลบทั้งหมด โดยมีจำนวนบวรวิหารเชิงลบต่อสถานที่ ถูกใช้เป็น Population Field (Weight)
2. Bandwidth (h): พิจารณาจากลักษณะภูมิภาคทางกายภาพของแหล่งท่องเที่ยวภาคเหนือซึ่งมีลักษณะเป็นจุดท่องเที่ยวเฉพาะตัว (Discreet points) แบนด์วิดท์ (h) จึงถูกกำหนดที่ 0.3 กิโลเมตร เพื่อมุ่งเน้นการตรวจจับ Localized Clustering หรือการกระจุกตัวในระดับจุลภาค วิธีการนี้ช่วยรักษาความสัมพันธ์เชิงพื้นที่ไม่ให้เกิดการกระจายตัวที่กว้างเกินไป (Over-smoothing) จนบดบังรายละเอียดความหนาแน่นในแต่ละจุดท่องเที่ยว ผลลัพธ์ที่ได้จึงมีความแม่นยำสูงในการระบุพิกัดวิกฤตเพื่อใช้ในการกำหนดเขตปฏิบัติการ (Action Zones) ในการบริหารจัดการพื้นที่
3. ซอฟต์แวร์: การวิเคราะห์ KDE ดำเนินการโดยใช้โปรแกรมประยุกต์ด้านภูมิสารสนเทศ เช่น ArcGIS หรือ QGIS ผลลัพธ์ของ KDE คือแผนที่ความหนาแน่นที่เน้นย้ำภูมิภาคที่ความไม่พอใจของนักท่องเที่ยวมีความเข้มข้นสูงสุด ซึ่งเป็นข้อมูลพื้นฐานสำหรับการกำหนดนโยบายและข้อเสนอแนะในเชิงพื้นที่

Results

1. Overview and statistics of the data

ในการรวบรวมบวรวิหารทั้งหมด 8,193 บวรวิหาร จากแหล่งท่องเที่ยวทางธรรมชาติใน 9 จังหวัดภาคเหนือ ผลการวิเคราะห์ความรู้สึกใน 3 ประเภท คือ ความคิดเห็นเชิงบวก เชิงลบ และเป็นกลาง โดยพบการกระจายตัวของความรู้สึก แสดงดัง Table 2 ผลการวิเคราะห์ยืนยันว่า บวรวิหารเชิงลบมีสัดส่วนร้อยละ 7.81 ของบวรวิหารทั้งหมด ซึ่งเป็นชุดข้อมูลหลักที่ใช้ในการวิเคราะห์เชิงลึกเพื่อระบุประเด็นปัญหา

Table 2 The results of sentiment analysis of natural attractions in Northern Thailand

Sentiment Category	Number of Reviews	%
Positive	6,150	75.06
Neutral	1,403	17.12
Negative	640	7.81
Total	8,193	100

โดยผลการทดสอบประสิทธิภาพของแบบจำลองจากจำนวนบวรวิหารทั้งหมด 451 บวรวิหาร เมทริกซ์ความคลุมเครือแสดงค่าความถูกต้องโดยรวม (Accuracy) ร้อยละ 76 และค่าความแม่นยำ (Precision) ของบวรวิหารเชิงลบร้อยละ 87 แสดงดัง Table 3

Table 3 Performance metrics of Pysentimiento model for sentiment classification

Actual / Prediction	Positive	Neutral	Negative	Recall	F1-score	Total
Positive	235	27	3	0.89	0.86	265
Neutral	29	33	8	0.47	0.42	70
Negative	15	26	75	0.66	0.74	116
Precision	0.84	0.38	0.87			
Accuracy					0.76	451

2. Thematic results from N-Gram analysis

การวิเคราะห์ N-Gram เฉพาะในชุดข้อมูลบทวิจารณ์เชิงลบจำนวน 640 บทวิจารณ์ ได้ระบุค่าและวลีที่มีความถี่สูงสุด โดยสามารถจัดกลุ่มปัญหาหลักออกได้เป็น 3 หมวดหมู่ (Table 4) โดยกลุ่มปัญหาสิ่งแวดล้อมและการบำรุงรักษา มีความถี่ของคำที่รุนแรงและแสดงอารมณ์เชิงลบสูงที่สุดในชุดข้อมูล

Table 4 Thematic results of N-Gram analysis driving negative sentiment

Sentiment category	Representative keywords & N-Grams	Description of the main issues
I. Value & Pricing	not worth (83), 300 baht (28), entrance fee (35), per person (37), money (69)	Highest Driver of Dissatisfaction. The most frequent N-gram, "not worth" (83), serves as a conclusive judgment that the high cost (specifically the 300 baht entry fee) does not align with the perceived value of the experience.
II. Environmental & Maintenance	water (205), dirty (57), dry season (21), no water (11), smell bad	High Frequency of Physical Issues. The high frequency of descriptive terms like water and dirty indicates immediate, visible issues with cleanliness and poor maintenance. Seasonal factors (no water / dry season) heavily compound the negative experience.
III. Accessibility & Infrastructure	chiang mai (69), waste time (33), not recommend (31), too many people	Management and Travel-Related Issues. Long travel times to destinations (e.g., from chiang mai) were frequently evaluated as a waste of time, leading to explicit negative recommendations (not recommend), signaling issues with site management and capacity.

3. Spatial results from Kernel Density Estimation (KDE)

การประยุกต์ใช้ Kernel Density Estimation (KDE) กับพิกัดภูมิศาสตร์ของบทวิจารณ์เชิงลบทั้งหมด โดยใช้ Bandwidth (h) ที่ 0.3 กิโลเมตร แสดงให้เห็นการกระจุกตัวของความไม่พึงพอใจในภูมิภาค ดังแสดงในแผนที่ความหนาแน่นความรู้สึกเชิงลบ (Figure 3)

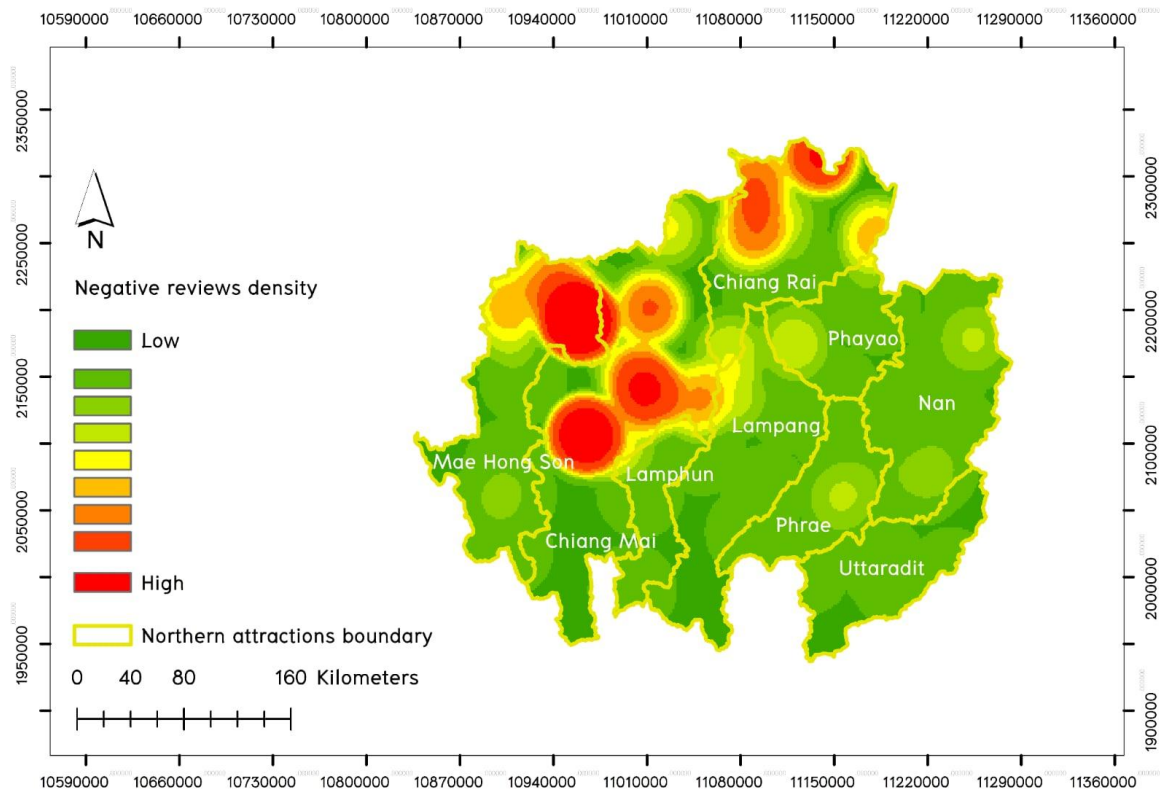


Figure 3 Kernel density map of negative reviews in Northern Thailand

ผลการวิเคราะห์ KDE ระบุการกระจุกตัวของความรู้สึกเชิงลบ (Hotspots) อย่างมีนัยสำคัญ ซึ่งสะท้อนถึงพื้นที่ที่มีปัญหาการจัดการเชิงพื้นที่อย่างเข้มข้น ดังนี้:

- Hotspot A (ศูนย์กลางวิกฤต): พื้นที่อุทยานแห่งชาติดอยอินทนนท์ และดอยสุเทพ-ปุย จังหวัดเชียงใหม่ แสดงค่าความหนาแน่นสูงสุด ซึ่งสัมพันธ์กับจำนวนนักท่องเที่ยวที่หนาแน่นและความคาดหวังต่อการบริการในระดับสูง
- Hotspot B (พื้นที่วิกฤตต่อเนื่อง): อำเภอปาย จังหวัดแม่ฮ่องสอน พบการกระจุกตัวของปัญหาในแหล่งท่องเที่ยวทางธรรมชาติหลายจุดที่อยู่ใกล้เคียงกัน สะท้อนถึงปัญหาในระดับ Cluster ของแหล่งท่องเที่ยว
- Hotspot C (พื้นที่เฝ้าระวัง): บริเวณสามเหลี่ยมทองคำ จังหวัดเชียงราย แม้จะมีค่าความหนาแน่นรองลงมา แต่ยังคงเป็นจุดยุทธศาสตร์ที่ต้องการการปรับปรุงคุณภาพบริการเฉพาะจุด

Conclusion and Discussion

1. Interpretation of quantitative and thematic results

ผลการวิเคราะห์ความรู้สึกที่มีสัดส่วนบทวิจารณ์เชิงลบถึงร้อยละ 7.81 ซึ่งให้เห็นว่าแม้ภาคเหนือจะยังคงเป็นที่นิยม แต่ปัญหาที่สะสมอยู่เริ่มส่งผลกระทบต่อความพึงพอใจโดยรวมของนักท่องเที่ยวต่างชาติ ซึ่งสอดคล้องกับงานวิจัยก่อนหน้านี้ระบุว่า การจัดการคุณภาพประสบการณ์ท่องเที่ยวในประเทศไทยเป็นสิ่งท้าทาย (Thongsawang, 2021)

การวิเคราะห์ N-Gram ยืนยันว่า ประเด็นด้านความคุ้มค่าและราคา (I. Value & Pricing) คือปัจจัยหลักที่ก่อให้เกิดความไม่พึงพอใจสูงสุด โดยมีหลักฐานเชิงประจักษ์จากวลี "not worth" (83) และการระบุราคาที่สูง (เช่น 300 baht) ซึ่งเน้นย้ำว่าความคาดหวังด้านคุณภาพไม่สอดคล้องกับค่าใช้จ่าย ที่ถูกเรียกเก็บ ปัญหานี้ถูกเชื่อมโยงอย่างมีนัยสำคัญกับ ประเด็นสิ่งแวดล้อมและการบำรุงรักษา (II. Environmental & Maintenance) โดยเฉพาะปัญหาคุณภาพน้ำ (water, dirty) และการลดลงของปริมาณน้ำตกในช่วง 'dry season' ซึ่งส่งผลกระทบต่อประสบการณ์ทางกายภาพและตอกย้ำความรู้สึกว่าไม่คุ้มค่า (Chumwatana & Chuaychoo, 2016) ในทางตรงกันข้าม ปัญหาด้านการเข้าถึงและโครงสร้างพื้นฐาน (III. Accessibility & Infrastructure) ซึ่งเป็นปัญหาเชิงโครงสร้าง ก็ยังคงมีความสำคัญ โดยเฉพาะในด้านความแออัด ซึ่งหมายถึงการจัดการด้านความสามารถในการรองรับ (Carrying capacity) ยังไม่มีประสิทธิภาพเพียงพอ (Tokarchuk et al., 2021)

2. Spatial integration and policy implications

ผลการวิเคราะห์ Kernel Density Estimation (KDE) ถือเป็นส่วนเสริมที่สำคัญที่สุดของงานวิจัยนี้ โดยแปลงปัญหาเชิงเนื้อหาให้เป็นข้อมูลเชิงพื้นที่ Hotspot A (ดอยอินทนนท์/ดอยสุเทพ-ปุย, เชียงใหม่) และ Hotspot B (อำเภอปาย, แม่ฮ่องสอน) ไม่เพียงแต่เป็นจุดที่มีจำนวนนักท่องเที่ยวมากเท่านั้น แต่ยังเป็นบริเวณที่มีความเข้มข้นสูงสุดของปัญหาเชิงลบ งานวิจัยนี้ต่อยอดองค์ความรู้โดยการเปลี่ยนข้อมูลข้อความ (Textual Data) ให้กลายเป็นพื้นผิวความหนาแน่นเชิงพื้นที่ (Density Surface) การบูรณาการข้อมูลเชิงเนื้อหาและเชิงพื้นที่ที่มีนัยสำคัญต่อการจัดการความยั่งยืนดังนี้:

- การจัดสรรทรัพยากรแบบเจาะจงพื้นที่: ข้อมูล Hotspot ที่ชัดเจนใน Figure 3 ช่วยให้หน่วยงานภาครัฐ (เช่น กรมอุทยานแห่งชาติ, องค์การบริหารส่วนจังหวัด) สามารถมุ่งเน้นงบประมาณและทรัพยากรในการแก้ไขปัญหา กลุ่มสิ่งแวดล้อม และการบำรุงรักษา ไปยังพื้นที่ Hotspot A และ B เป็นอันดับแรก การดำเนินการที่แม่นยำนี้จะช่วยเพิ่มผลกระทบของนโยบาย (Policy Impact) ได้สูงสุด

- การจัดการความแออัด: การกระจุกตัวของปัญหาใน Hotspot ชี้ให้เห็นถึงความจำเป็นในการใช้ระบบจองล่วงหน้า หรือ การกำหนดจำนวนนักท่องเที่ยวสูงสุด (Visitor quotas) ในช่วงที่มีจำนวนนักท่องเที่ยวสูง (Peak season) ในบริเวณเหล่านี้ โดยมีวัตถุประสงค์เพื่อลดความแออัด (ตามปัญหาในกลุ่มการเข้าถึงและโครงสร้างพื้นฐาน) โดยเฉพาะในจุดที่มีความอ่อนไหวทางธรรมชาติสูง

- การพัฒนาโครงสร้างพื้นฐานรองรับ: รัฐบาลควรพิจารณาการลงทุนในโครงสร้างพื้นฐานที่เกี่ยวข้องกับการจัดการของเสียและสิ่งอำนวยความสะดวกที่ยั่งยืนในพื้นที่ Hotspot (เช่น ระบบจัดการน้ำเสีย, ห้องน้ำมาตรฐาน) เพื่อยกระดับประสบการณ์โดยรวมอย่างเร่งด่วน

การเลือกใช้ Kernel Density Estimation (KDE) มีวัตถุประสงค์เพื่อวิเคราะห์ความเข้มข้นของปัญหา (Intensity of Dissatisfaction) ในเชิงพื้นที่ การใช้ KDE ยังช่วยให้ผู้วางแผนนโยบายสามารถเห็นภาพรวมของพื้นที่อิทธิพล (Area of Influence) ที่ปัญหาหนึ่ง ๆ ส่งผลกระทบต่อพื้นที่ข้างเคียง โดยจะเป็นประโยชน์ในการกำหนดเขตวิกฤต (Action Zones) เพื่อบริหารจัดการสิ่งแวดล้อมในระดับพื้นที่ มากกว่าการมุ่งเน้นเพียงจุดพิภคเดียว (Isolated point) นอกจากนี้ การกำหนดแบนด์วิธ (h) ที่ 0.3 กม. มีวัตถุประสงค์เพื่อจำกัดวงวิเคราะห์ให้อยู่ในระดับ Localized Clustering สำหรับระบุจุดบกพร่องภายในแหล่งท่องเที่ยวที่มีความละเอียดสูง เช่น บริเวณศูนย์บริการนักท่องเที่ยวหรือจุดชมวิวหลัก และจำกัดค่าความหนาแน่นไม่ให้กระจายตัวกว้างเกินไปในภูมิประเทศภูเขา

ผลการศึกษาสอดคล้องกับกรณีอุทยานแห่งชาติหวู่หยีซาน ประเทศจีน (Fu & Zhou, 2024) ที่แสดงให้เห็นว่าแม้ความพึงพอใจในทรัพยากรธรรมชาติจะอยู่ในระดับสูง แต่ความรู้สึกเชิงลบมักเกิดจากจุดบกพร่องด้านสิ่งอำนวยความสะดวก ระบบการจอง และการจัดการที่ไม่คุ้มค่าเมื่อเทียบกับราคาและเวลา ซึ่งข้อมูลเหล่านี้ทำหน้าที่เป็นสัญญาณเตือนให้ผู้บริหารจัดการสามารถปรับปรุงโครงสร้างพื้นฐานได้ตรงจุด

นอกจากนี้ ปัญหาความเสื่อมโทรมตามฤดูกาล (เช่น น้ำตกแห้งในหน้าแล้ง) ใน Hotspot A และ B ยังสอดคล้องกับข้อค้นพบในมณฑลฝูเจี้ยน (Lu et al., 2025) ที่ระบุว่าความแปรผันตามฤดูกาลส่งผลกระทบต่อความพึงพอใจอย่างรุนแรง อย่างไรก็ตาม ในขณะที่ปัญหาในต่างประเทศมักเกิดจากพฤติกรรมนักท่องเที่ยวเป็นหลัก แต่กรณีภาคเหนือของไทยกลับพบว่าอุปสรรคสำคัญคือ การบำรุงรักษาเชิงกายภาพและความคุ้มค่าในการเข้าชม ซึ่งชี้ให้เห็นว่าการจัดการการท่องเที่ยวเชิงนิเวศของไทยจำเป็นต้องยกระดับมาตรฐานการจัดการความคาดหวังควบคู่ไปกับการฟื้นฟูระบบนิเวศ

3. Limitations and conclusion

งานวิจัยนี้มีข้อจำกัดบางประการที่ควรพิจารณา ประการแรกคือแหล่งข้อมูลที่รวบรวมจาก TripAdvisor เพียงแหล่งเดียว ซึ่งสะท้อนมุมมองของนักท่องเที่ยวต่างชาติที่สื่อสารด้วยภาษาอังกฤษเป็นหลัก แต่อาจไม่ครอบคลุมพฤติกรรมของนักท่องเที่ยวกลุ่มอื่น ประการต่อมา การใช้พิกัดที่ตั้งคงที่ (Point Location) เป็นตัวแทนจุดเกิดปัญหาอาจยังไม่ครอบคลุมประสบการณ์เชิงพื้นที่ตลอดเส้นทางท่องเที่ยว (Spatial Journey) และสุดท้ายคือปัจจัยด้านความผันผวนตามฤดูกาล เช่น ปัญหาหมอกควันหรือปริมาณน้ำที่จำกัดในฤดูแล้ง ซึ่งส่งผลกระทบต่อความเข้มข้นของปัญหาในช่วงเวลาที่ต่างกัน อย่างไรก็ตาม ผลการศึกษาที่ได้ยังคงเป็นแนวทางสำคัญที่ให้ความแม่นยำสูงในการชี้เป้าพื้นที่วิกฤต (Crisis Mapping) ได้อย่างเป็นรูปธรรม

งานวิจัยนี้แสดงให้เห็นถึงประสิทธิภาพของการบูรณาการเทคนิค Deep Learning (BERT-based) ร่วมกับการวิเคราะห์ N-Gram เพื่อถอดรหัสความไม่พึงพอใจเชิงลึก และการใช้ Kernel Density Estimation (KDE) เพื่อระบุตำแหน่งเชิงยุทธศาสตร์ ผลการศึกษพบว่าประเด็นด้านความคุ้มค่าและราคามีความสัมพันธ์อย่างมีนัยสำคัญกับปัญหาสิ่งแวดล้อมและการบำรุงรักษา โดยมีความเข้มข้นสูงสุดในพื้นที่ Hotspot หลักของเชียงใหม่ (ดอยอินทนนท์ และดอยสุเทพ-ปุย) และแม่ฮ่องสอน (ปาย)

งานวิจัยนี้ได้ส่งมอบ แนวทางใหม่ (Policy Contribution) โดยการเปลี่ยนข้อมูลข้อความที่กระจัดกระจายให้กลายเป็นแผนที่นำทางสำหรับการตัดสินใจ ช่วยให้หน่วยงานที่เกี่ยวข้องสามารถจัดสรรงบประมาณและทรัพยากรเพื่อจัดการความยั่งยืนในพื้นที่ Hotspot ได้อย่างแม่นยำ ตรงจุด และสอดคล้องกับความคาดหวังของนักท่องเที่ยวในระดับสากล

Acknowledgements

ผู้เขียนขอขอบคุณเว็บไซต์ TripAdvisor สำหรับข้อมูลอันมีค่าที่ใช้ในการวิจัยนี้ ขอขอบคุณเป็นพิเศษแก่ผู้ทรงคุณวุฒิของวารสารวิชาการเพื่อการพัฒนาอุตสาหกรรมเชิงพื้นที่ สำหรับความคิดเห็นและข้อเสนอแนะเชิงสร้างสรรค์ ซึ่งช่วยพัฒนาคุณภาพของต้นฉบับบทความนี้

References

- Alaei, A., Becken, S. & Stantic, B. (2017). Sentiment Analysis in Tourism: Capitalising on Big Data. Sage Journals. *Journal of Travel Research*, 8. DOI: 10.1177/0047287517747753.
- Bird, S., Klein, E. and Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media, Inc.
- Chumwatana, T. & Chuaychoo, I. (2016). Automatic filtering non-English complaint emails in tourism industry using N-gram extraction and classification techniques. 4th International Symposium on Computational and Business Intelligence (ISCBI), September 5–7, 2016, Olten, Switzerland.
- Cui, Y. & He, Q. (2019). Travel Behavior Analysis with Social Media Data and Smartphone GPS Data. *Final Report*. Transportation Informatics Tier I University Transportation Center, University of Buffalo, NY, USA.
- Devlin, J., Chang, M. W., Lee, K. & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv*. DOI: 10.48550/arXiv.1810.04805.
- Donthu, N., Kumar, S., Pandey, N., Pandey, N. & Mishra, A. (2021). Mapping the Electronic Word-of-Mouth (eWOM) research: A Systematic Review and Bibliometric Analysis. *Journal of Business Research*, 135, 758–773. DOI: 10.1016/j.jbusres.2021.07.015.
- Esri. (2025, Dec 12). *What is geocoding? 2021*. Form <https://desktop.arcgis.com/en/arcmap/latest/manage-data/geocoding/what-is-geocoding.htm>
- Filieri, R., Alguezaui, S. & McLeay, F. (2015). Why do Travelers Trust TripAdvisor? Antecedents of Trust Towards Consumer-generated Media and its Influence on Recommendation Adoption and Word of Mouth. *Tourism Management*, 51. DOI: 10.1016/j.tourman.2015.05.007.
- Fu, W. & Zhou, B. (2024). Theme Exploration and Sentiment Analysis of Online Reviews of Wuyishan National Park. *Land*, 13, 629. DOI: 10.3390/land13050629.
- Hashim, R., Omar, B. & Al-Samarraie, H. (2020). The Application of Sentiment Analysis in Tourism Research: A Brief Review. *Journal of Business Tourism and Applied Sciences (JBTS)*, 8(1), 51–60.
- Henning, S., Beluch, W., Fraser, A. & Friedrich, A. (2023). *A Survey of Methods for Addressing Class Imbalance in Deep-Learning Based Natural Language Processing*. Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, Dubrovnik, Croatia, 523–540.
- Jeacle, I. & Carter, C. (2011). In TripAdvisor We Trust: Rankings, Calculative Regimes and Abstract Systems. *Accounting Organizations and Society*, 36(4–5), 293–309. DOI: 10.1016/j.aos.2011.04.002.

- Kuhamanee, T., Talmongkol, N., Chaisuriyakul, K., San-Um, W., Pongpisuttinun, N. & Pongyupinpanich, S. (2017). *Sentiment analysis of foreign tourists to Bangkok using data mining through online social network*. IEEE 15th International Conference on Industrial Informatics (INDIN), Emden, Germany.
- Lassen, N. B., Madsen, R. & Vatrappu, R. (2014). *Predicting iPhone Sales from iPhone Tweets*. IEEE 18th International Enterprise Distributed Object Computing Conference, Ulm, Germany.
- Leelawat, N., Jariyapongpaiboon, S., Promjun, A., Boonyarak, S., Saengtabtim, K., Laosunthara, A., Yudha, A. K. & Tang, J. (2022). Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning. *Heliyon*, 8(10), e10894. DOI: 10.1016/j.heliyon.2022.e10894.
- Li, X., Pan, B., Law, R. & Huang, X. (2017). Forecasting Tourism Demand with Composite Search Index. *Tourism Management*, 59, 57–66. DOI: 10.1016/j.tourman.2016.07.005.
- Lima, T. O., Colaço, M. Jr., Prado, K. H. J., Jesus, I. D. Jr. & Oliveira, F. R. (2021). *A Big Data Experiment to Evaluate the Effectiveness of Traditional Machine Learning Techniques Against LSTM Neural Networks in the Hotels Clients Opinion Mining*. IEEE International Conference on Big Data (Big Data), December 15–18, 2021, Orlando, FL, USA.
- Lu, Z., Chen, S., Qiu, C., Lin, Y., Lu, Y. & Xu, Y. (2025). Exploring the impact of landscape environments on tourists' emotional fluctuations in Fujian's Coastal National Parks using machine learning. *PLoS One*, 20(8), e0329118. DOI: 10.1371/journal.pone.0329118.
- McKinney, W., (2010). *Data Structures for Statistical Computing in Python*. Proceedings of the 9th Python in Science Conference (SCIPY 2010), June 28, 2010, Tacoma Convention Center | Tacoma, WA, USA.
- Ministry of Tourism and Sports. (2025, Dec 12). *National Tourism Development Plan No. 2 (2017–2021) 2016*. Form https://secretary.mots.go.th/strategy/more_news.php?cid=9. (in Thai)
- _____. (2025, Dec 12). *Tourism Economic Review. Tourism Economic Report, Year 1, Issue 3/2020 2020*. Form https://www.mots.go.th/more_news_new.php?cid=605. (in Thai)
- _____. (2025, Dec 12). *Development Strategy. National Tourism Development Plan No. 3 (2023–2027) 2023*. Form <https://anyflip.com/zzfck/kiel>. (in Thai)
- _____. (2025, Dec 12). *Survey Project for Analysis of International Tourist Behavior 2024. Final Report*. Form <https://intelligencecenter.tat.or.th/articles/35286>. (in Thai)
- _____. (2025, Dec 12). *Tourism Economic Review. Tourism Economic Report, Year 2, Issue 1/2021*. Form <https://www.mots.go.th/news/category/708>. (in Thai)
- Mission, R. S. (2024). Spatial Analysis Leveraging Geographic Information System and Kernel Density Estimation for Crime Prediction. *Journal of Innovative Technology Convergence*, 6(1), 95–104. DOI: 10.69478/JITC2024v6n2a10.
- Muthukadan, B., (2025, Dec 12). *Selenium Python Bindings 2024*. Form <https://selenium-python.readthedocs.io/>

- Pérez, J. M., Rajngewerc, M., Giudici, J. C., Furman, D. A., Luque, F., Alemany, L. A. & Martinez, M. V. (2024). pysentimiento: A Python Toolkit for Opinion Mining and Social NLP tasks. *arXiv*. DOI: 10.48550/arXiv.2106.09462.
- Phithakkitnukoon, S. (2023). *Analysis of Tourist Behavior Using Mobile Phone GPS Data*. In: Urban Informatics Using Mobile Network Data. Springer Nature, 183–214.
- Pleerux, N., Boonpook, W., Chaiyo, A. and Nardkulpat, A., (2023). *Trip Advisor Restaurant Customer Reviews Analysis in Bangkok, Thailand*. International Conference on Cyber Management and Engineering (CyMaEn), Jan 26–27, 2023, Bangkok, Thailand.
- Prabha, M. I. & Srikanth, G. U. (2021). *Survey of Sentiment Analysis Using Deep Learning Techniques*. 1st International Conference on Innovations in Information and Communication Technology (ICIICT), April 25–26, 2019, Chennai, India.
- Ramadhani, A., Sutoyo, E. & Widartha, V. P. (2021). *LSTM-based Deep Learning Architecture of Tourist Review in Tripadvisor*. Sixth International Conference on Informatics and Computing (ICIC), November 3–4, 2021, Jakarta, Indonesia.
- Richardson, L. (2025, Dec 12). *Beautiful Soup 2004*. Form <https://www.crummy.com/software/BeautifulSoup/>.
- Rogel–Salazar, J. (2025, Dec 12). *Density–Based Clustering 2022*. Form <https://domino.ai/data-science-dictionary/density-based-clustering>.
- Rossum, G., van. (1995). *Python tutorial*. Department of Computer Science, College of Western Idaho, USA.
- Sarker, I. H. (2021). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. Springer Nature Link, *Computer Science*, 2, 420. DOI: 10.1007/s42979–021–00815–1.
- Secretariat of Cabinet, the. (2025, Dec 12). National Tourism Development Plan 2012–2016 2011. Form <https://www.ratchakitcha.soc.go.th/DATA/PDF/2554/E/067/14.PDF>. (in Thai)
- Thongsawang, S. (2021). GUIDELINES FOR PROMOTING THAILAND TOURISM FOR CHINESE TOURISTS IN THE CHONBURI PROVINCE. *Journal of Social Science and Buddhistic Anthropology*, 6(8). 307–324. (in Thai)
- Tokarchuk, O., Gabriele, R. & Maurer, O. (2021). Estimating tourism social carrying capacity. *Annals of Tourism Research*, 86, 102971. DOI: 10.1016/j.annals.2020.102971.
- Valdivia, A., Luzón, M. V. & Herrera, F. (2017). *Sentiment Analysis on TripAdvisor: Are There Inconsistencies in User Reviews?* In: Martínez de Pisón, F., Urraca, R., Quintián, H., Corchado, E. (eds). Hybrid Artificial Intelligence Systems. HAIS, Lecture Notes in Computer Science, 10334, 15–25.
- Wu et al. (2024). Stratified random sampling for neural network test input selection. *Information and Software Technology*, 165, 107331. DOI: 10.1016/j.infsof.2023.107331
- Yuan, X. (2024). Evaluation of Rural Tourism Development Level using BERT-enhanced Deep Learning Model and BP Algorithm. *Scientific Reports*, 14, 25748. DOI: 10.1038/s41598–024–77444–0.