Received: Dec 22, 2025 **Revised:** Jul 6, 2025 **Accepted:** Jul 8, 2025

SENTIMENT ANALYSIS OF PUBLIC PERSPECTIVES ON PM2.5 POLLUTION IN BANGKOK AND METROPOLITAN AREAS

Chonnikarn Luangpituksa[1], Pornsiri Chatpreecha[2], Chutitanrat Uttamasiriseni[3]

[1]School of Integrated Science, Kasetsart University, Bangkok, Thailand.

[2]Department of Agro-Industrial Technology, Faculty of Agro-Industry, Kasetsart University, Bangkok, Thailand.

[3]Department of Marketing and Branding in the Digital Era, Faculty of College of Innovative Business and Accountancy, Dhurakij Pundit University, Bangkok, Thailand.

[1] chonnikarn.lu@ku.th, [2]pornsiri.cha@ku.ac.th, [3]chutitanra.utt@dpu.ac.th

Abstract

This study analyzes public opinions on PM2.5 pollution in Bangkok and its surrounding metropolitan areas using data collected from the Thai-language online discussion platform Pantip.com. A total of 390 user-generated comments were collected and processed using Natural Language Processing (NLP) techniques, including Thai word tokenization, text vectorization, and word embedding. Three machine learning algorithms—Naïve Bayes, Logistic Regression, and Support Vector Machine (SVM)—were trained and evaluated for sentiment classification into positive, negative, and neutral categories. Among the evaluated models, SVM achieved the best overall performance, with an accuracy of 88% and an F1-score of 0.86.

The sentiment distribution reported in this study was generated using the SVM model. Of the 390 collected comments, 52.1% were classified as negative, 30.0% as positive, and 17.9% as neutral. These results represent the model predictions on the collected dataset and should be interpreted within the scope of this study.

The findings suggest that sentiment analysis of social media data can provide timely insights into public perceptions of PM2.5 pollution. Such information may serve as preliminary evidence to support the improvement of environmental communication strategies and assist policymakers in developing more responsive environmental policies.

Keywords : Machine Learning, PM2.5 Pollution, Sentiment Analysis, Public Opinion, Social Media.

Introduction

The issue of fine particulate matter (PM2.5) pollution has emerged as one of the most serious environmental concerns in Thailand, particularly in recent years. The problem has intensified in various

regions across the country, especially in large urban areas such as Bangkok and its surrounding metropolitan provinces, where high population density and continuous economic activity contribute to deteriorating air quality. PM2.5 pollution poses both short-term and long-term health risks, including respiratory diseases, cardiovascular conditions, and an increased likelihood of certain types of cancer. Vulnerable populations—such as children, the elderly, and individuals with chronic illnesses—are especially at risk.

A national health impact assessment (Mueller et al., 2021). concluded that long-term exposure to PM2.5 significantly affects vulnerable populations, especially through indoor environments where pollution levels often exceed safe limits. Although the government and relevant agencies have implemented air pollution control measures and regularly disseminate air quality information to the public, institutional challenges in managing the PM2.5 crisis in Thailand were identified (Poapongsakorn et al., 2023). including critical limitations in centralized governance, overlapping agency responsibilities, and inadequate local budget allocations. The importance of enforcing stricter air quality standards has been emphasized (World Health Organization (WHO), 2021). noting that Thailand's regulatory threshold remains higher than global recommendations, which limits the effectiveness of national air quality policies.

Social media platforms not only provide venues for public expression but also facilitate value creation and influence public attitudes through interactive communication mechanisms (Mohammadiani, et al., 2022). These dynamics highlight the critical relevance of analyzing social media data to gain insight into societal concerns about environmental issues. In this context, public opinions expressed on social media platforms offer a valuable source of information, reflecting citizens' attitudes, emotions, concerns, and expectations in a natural and timely manner. Sentiment analysis, a branch of natural language processing (NLP), has emerged as a powerful tool to systematically classify and analyze these expressions. When combined with machine learning, it allows researchers and policymakers to extract meaningful insights from large volumes of user-generated content—thereby enhancing public communication and environmental policy planning.

This study applies sentiment analysis techniques to such data to systematically examine public sentiment toward the PM2.5 crisis. The findings aim to support evidence-based policymaking, enhance risk communication, and inform the development of more responsive and sustainable strategies for addressing PM2.5 pollution in alignment with public needs.

Literature Review

A Public Awareness and Behavioral Response

In recent years, the issue of PM2.5 pollution in Bangkok and its surrounding areas has prompted numerous studies exploring public awareness, health impacts, predictive modeling, and policy responses. However, research applying sentiment analysis to assess public opinions on PM2.5 pollution remains limited—particularly within the Thai context.

This section reviews relevant literature from the past five years concerning public sentiment, air quality analysis, and social media mining. Perception and response behaviors toward PM2.5 pollution among motorcycle taxi drivers in Din Daeng District, Bangkok, were examined in (Kaewphet, 2021). which found that awareness levels significantly influenced individuals' preventive actions against PM2.5 exposure. A study on public attitudes toward governmental measures to address PM2.5 pollution in Bangkok found that public understanding and engagement were directly influenced by the availability of accessible information and environmental campaigns (Tansathien, 2022).

This relationship is consistent with health behavior theories such as the Health Belief Model (HBM), which posits that awareness of risks significantly influences preventive behaviors (Champion & Skinner, 2023).

B. Forecasting, Health Impacts, and Policy Measures

Previous studies from both local and international research address the intersection of technical forecasting, health implications, and policy response. For instance, (Yar et al., 2022). outlined how artificial intelligence can be systematically integrated across the full policy cycle—from data collection and predictive modeling to policy evaluation—supporting real-time responsiveness in public decision-making. This aligns with (Polsuwan, 2021). which demonstrated that combining ARIMA with LSTM models improved the prediction of PM2.5 levels in Bangkok.

Moreover, (World Health Organization (WHO)., 2021). underscores the importance of aligning national policy measures with global air quality guidelines to mitigate long-term health risks, especially those involving cardiovascular and respiratory conditions, as highlighted by (Chulabhorn, 2024) Policy recommendations calling for stricter regulation of agricultural burning resonate with international approaches advocating for emission transparency and early warning systems (Saengthong, 2023).

In recent years, scholars have emphasized the role of artificial intelligence in public governance. It has been proposed that AI tools, including NLP and ML, offer transformative potential for digital governance through real-time policy adaptation (Ojo & Mellouli, 2018). Similarly, the democratizing function of social media, particularly platforms like Twitter and Facebook, in facilitating public engagement and inclusive policy discourse has been underscored (Effing et al., 2011).

C. Sentiment Analysis and Social Media Applications

Additionally, (Chompunuch et al., 2021). analyzed PM2.5-related discourse on Facebook using a Naïve Bayes classifier, while (Praphan & Komintarachat, 2021). developed a sentiment dashboard to assist decision-makers in responding to air pollution. (Ratanachai & Putthacharoen, 2021). improved sentiment classification in Thai text by integrating Word2Vec with a Support Vector Machine (SVM) model. Twitter data have been used to analyze sentiment on PM2.5, with emotions categorized into anxiety, anger, and distrust—demonstrating the potential of social media as a tool for capturing public sentiment (Wongsawat, Jindaluang, & Preechadi rek, 2022). Other researchers have applied machine learning techniques to support

emotion detection and policymaking. Lastly, Latent Dirichlet Allocation (LDA) was applied on Pantip.com data to identify key discussion topics related to air quality (Siriwan et al., 2022). Collectively, these studies reflect a growing interest in using natural language processing (NLP) and social media mining to gauge public opinion on PM2.5 pollution. Nonetheless, most do not offer integrated platforms that combine real-time data collection, automatic classification, and visual analytics.

To broaden the understanding of how sentiment analysis can inform environmental communication, recent international literature emphasizes its relevance in sustainability and climate discourse. The application of sentiment analysis in gauging public awareness of environmental sustainability has been explored, highlighting its potential for monitoring attitudes toward ecological challenges (Kularbphettong et al., 2024). A comprehensive review of sentiment analysis techniques has also been presented, comparing models such as VADER, BERT, and SVM, and evaluating their effectiveness in extracting sentiments from social media (Ligthart et al., 2021). Furthermore, a tertiary study analyzed sentiment analysis research trends, identifying common challenges such as data imbalance, and offering methodological recommendations (Hounkpevi, 2020). More recently, transformer-based models such as BERT and its multilingual variants (mBERT, XLM-RoBERTa) have demonstrated state-of-the-art performance in sentiment classification by capturing deep contextual and semantic nuances that traditional machine learning models cannot fully represent (Anderson et al., 2024). Although these deep learning architectures offer superior accuracy on large labeled datasets, their high computational demands and the limited availability of Thai-language pre-trained corpora present practical constraints—making well-validated classical models such as SVM a pragmatic choice for Thai-language sentiment analysis in resource-constrained settings.

Furthermore, in Latin America, sentiment analysis has proven valuable for monitoring Sustainable Development Goal (SDG), particularly in capturing migration-related discourse (Hounkpevi, 2020). In Indonesia, studies have reviewed sentiment analysis on social media related to air pollution, revealing that real-time public sentiment can support governments in crisis communication and policy feedback (Permana et al., 2024). At the global level, Twitter-based analyses of climate change found that public perceptions shift over time and across regions, reinforcing the value of social media in tracking dynamic environmental concerns (McDermott et al., 2024).

D. Research Gap and Contribution of the Present Study

These global insights reinforce the utility of sentiment analysis in Thailand's environmental context, where air quality concerns—especially PM2.5—have become a pressing societal issue. Although previous studies have shown the potential of using sentiment analysis techniques on platforms like Twitter and Facebook to reflect public opinion, these platforms are generally characterized by short-form expressions and rapid interactions.

This leaves a gap in understanding more elaborate and reasonable forms of discourse. Pantip.com, a Thai-language online forum, provides a unique setting where users engage in anonymous, long-form discussions that allow for the expression of detailed reasoning, emotions, and attitudes. Unlike Twitter or Facebook, Pantip facilitates extended conversations, making it possible to extract nuanced insights into how individuals perceive and discuss air quality issues. However, one limitation of Pantip is its lack of real-time data flow, which may affect its responsiveness to rapidly developing events. This gap highlights the need for research that explores Pantip as a complementary data source for environmental sentiment analysis.

The present study addresses this gap by applying machine learning–based sentiment analysis to Pantip data and visualizing public sentiment trends. Academically, this research extends the existing literature by pioneering the application of sentiment analysis on Thai-language, long-form forums—an underexplored domain in environmental communication research. It also contributes methodologically by demonstrating how machine learning models can be adapted to process complex Thai textual data, thereby providing a framework for analyzing other low-resource languages.

Furthermore, this study contributes valuable insights to a more localized understanding of public opinion and supports data-driven environmental policymaking aligned with the Sustainable Development Goals (SDGs). It also advances communication strategies for environmental risk management, thereby strengthening the integration of sentiment analysis into evidence-based governance.

Methodology

A. Research Objectives

To analyze public opinions and sentiments regarding the PM2.5 pollution issue using data collected from Pantip threads. Pantip has been identified as both a credible source of diverse and timely public sentiment (Strategy& PwC, 2015). and as a platform that has evolved into a hub for knowledge sharing (Leesa-nguansuk, 2016). To classify public comments into positive, negative, or neutral categories through sentiment analysis techniques. To propose recommendations for the management and communication of PM2.5-related information to the public.

B. Research Scope

This study focuses on the sentiment analysis of public opinions concerning the PM2.5 pollution problem in Bangkok and its surrounding metropolitan areas. The scope of the research is defined as follows:

1) Data Sources

The data comprises user-generated comments collected from Pantip.com, which ranked seventh among Thai websites in July 2024, with more than 79 million visitors (Manaskasemsak et al., 2022). With more than 15,000 topic tags, it offers rich, issue-based discussions, making it particularly valuable for understanding social concerns such as environmental health and air pollution. Only threads and comments directly related to PM2.5 pollution and air quality in the target areas were selected.

2) Geographic and Target Population

The study examines opinions from individuals residing in or connected to Bangkok and nearby provinces, including Nonthaburi, Pathum Thani, Samut Prakan, Samut Sakhon, and Nakhon Pathom. These areas are recognized as high-risk zones for PM2.5 pollution due to dense populations, extensive transportation networks, and intensive economic activities. The credibility of user-generated information on Pantip.com has been analyzed with findings indicating that users' demographic and behavioral characteristics significantly influenced how they assessed the reliability of content on the platform. Their findings support the use of Pantip as a source of well-considered public opinion, particularly in decision-oriented discussions.

3) Analytical Techniques

Sentiment analysis, a method within the Natural Language Processing (NLP) framework, is applied to classify comments into three categories: positive, negative, and neutral. Machine Learning (ML) models serve as the core analytical tools for processing and classifying textual data. The applicability of Pantip data for NLP research has been demonstrated by building a multi-turn dialogue dataset using sentence similarity techniques, highlighted the platform's linguistic diversity and its suitability for Thai-language sentiment analysis (Sae-Oueng et al., 2023).

4) Timeframe and Data Recency

The study focuses on user comments posted between 2023 and 2024 to ensure the analysis captures current public sentiments reflecting the ongoing PM2.5 situation.

C. Data Collection

This study employs a mixed-methods approach, combining qualitative and quantitative analysis through sentiment analysis—a subset of Natural Language Processing (NLP)—to examine public opinions on PM2.5 pollution. Data were collected from Pantip.com, a popular Thai online forum where users freely discuss various topics. Threads related to PM2.5, air pollution, health, Bangkok, and surrounding metropolitan areas were identified.

Using Selenium, relevant comments were scraped from Pantip.com. After screening, removing duplicates, and applying preprocessing criteria, a total of 390 Thai-language comments were retained for sentiment analysis.

D. Preprocessing

The collected textual data underwent several preprocessing steps to prepare it for sentiment analysis. Previous studies have emphasized that effective preprocessing is essential for improving sentiment classification performance on social media data (Kothandan & Murugesan, 2021). In this study, the preprocessing pipeline was carefully designed to preserve the linguistic characteristics of Thai text while ensuring consistency throughout the analysis. The specific preprocessing steps included:

- Data Cleaning: Removal of emojis, punctuation, hyperlinks, duplicate entries, and other irrelevant characters.

Thai Word Segmentation (Tokenization): Thai word segmentation was performed using the PyThaiNLP library with the newmm tokenizer before feature extraction and word cloud generation. This step segmented Thai sentences into meaningful lexical units.

- Stop Word Removal: Elimination of common Thai stop words, such as คือ (is), แล้ว (then), and มาก (very).
- Text Normalization: Standardization of text by removing special characters and unnecessary whitespace while preserving meaningful content.
- Vectorization: Conversion of text into numerical feature vectors using the Term Frequency–Inverse Document Frequency (TF–IDF) technique.
- Word Embedding: TF-IDF was used as the primary feature representation for all classification models. Word embeddings (Thai2Fit, FastText, BPEmb) were explored as optional experiments to capture contextual semantic information but were not used in the final model comparison.

The same tokenization procedure was consistently applied during both feature extraction for sentiment classification and word cloud generation to ensure consistency throughout the analysis. These preprocessing techniques are widely used for Thai-language text processing and have been shown to improve sentiment classification performance when combined with machine learning algorithms such as Support Vector Machine (SVM) (Wongsawat et al., 2022).

Figure 1 illustrates the overall sentiment analysis framework employed in this study, including data collection from Pantip.com, Thai text preprocessing, feature extraction, sentiment classification using machine learning algorithms, and result interpretation through sentiment distribution and trend analysis.

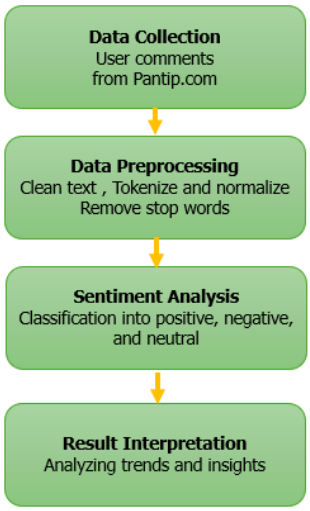

Fig. 1 Workflow of Sentiment Analysis Process

E. Model Training and Evaluation

The sentiment analysis model classifies comments into three categories: positive, negative, and neutral. The dataset was randomly divided into training (80%) and testing (20%) sets using a random state of 42 to ensure reproducibility. Three supervised machine learning algorithms—Support Vector Machine (SVM), Naïve Bayes (NB), and Logistic Regression (LR)—were trained using the same TF–IDF feature

representation. Default hyperparameter settings in the Scikit-learn library were used for all models to provide a fair comparison. Model performance was evaluated on the testing set using Accuracy, Precision, Recall, and F1-score.To strengthen the methodological contributions of this study within the context of Information Technology Management, this section presents an extended discussion on the machine learning algorithms applied, key NLP techniques, the challenges of working with Thai-language sentiment data, and the comparative performance of each classification model used in the sentiment analysis pipeline.

1) Machine Learning Algorithms Applied

Three supervised machine learning algorithms were employed to classify the sentiments of public comments related to PM2.5 pollution: Support Vector Machine (SVM), Naïve Bayes (NB), and Logistic Regression (LR).Support Vector Machine (SVM) is a robust classification algorithm that maximizes the margin between classes by constructing an optimal hyperplane. It has been widely reported to achieve strong performance in high-dimensional and sparse text classification tasks (Cortes & Vapnik, 1995). Therefore, SVM was selected as the primary classification model in this study.Naïve Bayes (NB) is a probabilistic classifier based on Bayes' theorem. Due to its computational efficiency and relatively good performance on small text datasets, it was included as a baseline model for comparison (McCallum & Nigam, 1998). Logistic Regression (LR) is a linear classification algorithm commonly applied to binary and multiclass text classification problems. It estimates class probabilities and provides an additional linear baseline for evaluating classification performance (Hosmer, 2013).

To ensure a fair comparison, all three models were trained using the same TF–IDF feature representation and evaluated under the same training and testing protocol. Their performance was compared using Accuracy, Precision, Recall, and F1-score.

NLP Pipeline and Feature Engineering

The sentiment analysis pipeline involved several Natural Language Processing (NLP) techniques customized for the Thai language:

- Word Tokenization: Thai sentences were segmented using PyThaiNLP, which handles the lack of whitespace between words—a unique challenge in Thai text processing (Phatthiyaphaibun et al., 2020).
- Vectorization: The cleaned and tokenized text was converted into numerical representations using Count Vectorizer; Captures raw frequency of word occurrences, then TF-IDF; Adjusts word frequency based on document distribution, enhancing the discriminative power of words (Salton & Buckley, 1988).
- Word Embedding (Optional): Experiments were also conducted using pretrained embeddings such as Thai2Fit (Lowphanich & FastText, 2020). which help represent contextual relationships between words (Lowphanich ,2020).

These techniques ensured that the machine learning models received semantically meaningful input data for training and prediction.

3) Challenges in Thai Sentiment Analysis

Several linguistic and structural challenges were encountered during Thai text analysis:

- Lack of Explicit Word Boundaries: Unlike English, Thai does not use spaces to delimit words, requiring specialized tokenizers.
- Mixed Language Content: Public comments occasionally mixed Thai and English, requiring robust tokenization and cleaning techniques.
- Imbalanced Sentiment Distribution: The dominance of negative comments posed class imbalance issues, potentially skewing classification performance.

4) *Model Performance Comparison*

To evaluate each model's classification capability, standard performance metrics were computed, including Accuracy, Precision, Recall, and F1-score. The models were trained on manually labeled training data (as shown in Table III) and tested on cleaned Pantip comments. The manual labeling was performed independently by two trained annotators, each of whom assigned one of three sentiment labels—Negative, Positive, or Neutral—to each comment based on predefined criteria (see Table III). Inter-annotator agreement was assessed using Cohen's Kappa coefficient, yielding a score of **K** = 0.78, which indicates substantial agreement. Cases of disagreement were resolved through discussion between the two annotators, with a third annotator serving as a tiebreaker when consensus could not be reached.

The SVM model demonstrated the highest performance across all metrics, affirming its suitability for complex sentiment classification tasks involving Thai-language text. Logistic Regression offered a balance between performance and interpretability, while Naïve Bayes served as a strong baseline. These results validate the efficacy of machine learning–based sentiment analysis in capturing nuanced public emotions and underscore the potential of applying AI techniques to support real-time decision-making in environmental governance contexts.

Table I Comparative Performance Of Machine Learning Models

Algorithm	Accuracy	Precision	Recall	F1-score
Naïve bayes	0.81	0.79	0.78	0.78
Logistic regression	0.84	0.83	0.81	0.82
Support vector machine	0.88	0.87	0.86	0.86

Results and Discussion

The sentiment classification results were generated using the Support Vector Machine (SVM) model, which was identified as the highest-performing algorithm in this study. As reported in Table I, SVM outperformed Naïve Bayes and Logistic Regression across all evaluation metrics, including accuracy, precision, recall, and F1-score. These findings support the model's robustness and reliability in capturing nuanced public opinions expressed in Thai-language online forums.

Table II Sentiment Classification Results

Sentiment Type	Number of Comments	Percentage
Negative	203	52.1%
Positive	117	30.0%
Neutral	70	17.9%

The sentiment distribution presented in Table II was generated using predictions from the Support Vector Machine (SVM) model, which achieved the highest classification performance among the evaluated models (Table I). Of the 390 collected comments, 203 (52.1%) were classified as negative, 117 (30.0%) as positive, and 70 (17.9%) as neutral. Within the analyzed dataset, negative sentiment was the most prevalent category, accounting for 52.1% of the comments. These comments frequently expressed concerns, dissatisfaction, and criticism regarding PM2.5 pollution and the perceived effectiveness of governmental responses. The reported sentiment distribution should be interpreted as the prediction output of the SVM model for the collected dataset rather than a direct representation of the opinions of the entire population.

Positive comments comprised 117 entries, representing 30.0% of the total. This typically conveyed encouragement, shared protective measures (such as wearing masks), or praised specific actions taken by relevant agencies. Neutral comments totaled 70 entries or 17.9%, and were generally characterized by questions, information sharing, or emotionally neutral observations.

These results indicate a predominantly negative public sentiment toward the PM2.5 crisis, highlighting widespread concern and a strong demand for more effective and transparent government intervention. The findings emphasize the value of sentiment analysis in understanding public perceptions and informing environmental policy responses.

Table III Sample Training Data For Sentiment Analysis Model

No.	Original Thai Sentence	English Translation	Actual Label	Predicted Label	Explanation
1	ฝุ่น PM2.5 เยอะมาก ออกไปไม่ได้เลย	PM2.5 pollution is very severe. I cannot go outside.	Negative	Negative	Correct prediction
2	วันนี้อากาศดีขึ้นมาก	The air quality has improved today.	Positive	Positive	Correct prediction
3	ค่า PM2.5 วันนี้อยู่ในระดับปานกลาง	PM2.5 is at a moderate level today.	Neutral	Neutral	Correct prediction
4	ฝุ่นก็เยอะ แต่ยังต้องออกไปทำงาน	PM2.5 is high, but I still have to go to work.	Negative	Neutral	Misclassified because of mixed sentiment

As shown in Table III, the sample training data set consists of user-generated comments related to the PM2.5 pollution situation. These comments are categorized based on sentiment into three types: Negative – comments that express dissatisfaction, disappointment, or concern, such as “the air is terrible” or “tired of the dust”; Positive – comments that convey constructive or appreciative sentiments, such as “thank you to the officers” or “the air is better”; and Neutral – comments that do not express clear emotional tones, such as “just okay” or “nothing has changed.” This classification illustrates how the model distinguishes emotional nuances in public discourse regarding air quality issues. To ensure the quality and consistency of the labeled dataset, the annotation was carried out by two independent annotators who applied the above criteria to all 390 comments. Inter-annotator agreement was measured using Cohen’s Kappa (**K** = 0.78), indicating substantial agreement. Discrepancies between the two annotators were resolved through discussion, with a third annotator consulted in cases where consensus could not be reached. The final agreed-upon labels were used as ground truth for model training and evaluation.

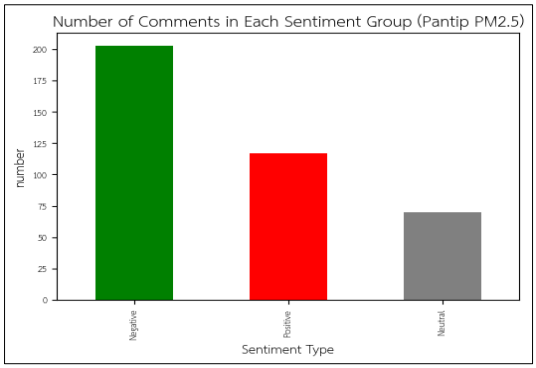

Fig. 2 Bar Chart of Sentiment Distribution

Figure 2 presents a bar chart illustrating the distribution of public sentiments derived from user comments on Pantip.com regarding the PM2.5 pollution situation in Bangkok and the surrounding metropolitan areas. The largest group of comments was classified as negative, totaling 203 comments or 52.1% of all responses. This reflects widespread concern, dissatisfaction, and criticism from the public regarding the PM2.5 issue, particularly in relation to government response, health impacts, and delays in effective problem-solving.

The positive sentiment group included 117 comments or 30.0% of the total. Although this proportion is lower than negative sentiments, these comments reflect constructive attitudes such as encouraging public officials, expressing appreciation for certain mitigation efforts, and recognizing improvements in specific periods. The neutral sentiment group contained 70 comments, accounting for 17.9%. These comments generally lacked strong emotional cues and often involved inquiries, information exchanges, or objective observations without clear positive or negative tones.

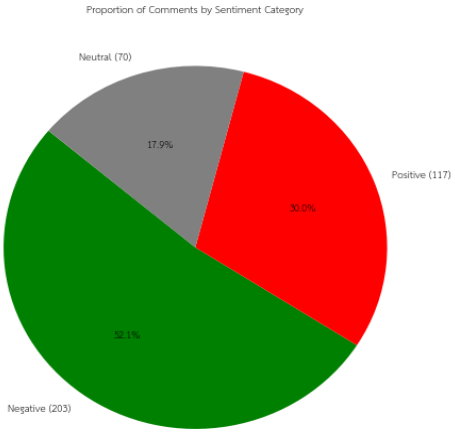

Fig. 3 Proportion of Public Sentiment Categories

Figure 3 shows the proportion of public sentiment categories using a pie chart. Overall, the analysis indicates that a substantial portion of the public holds negative sentiments toward the PM2.5 situation. This may signify heightened awareness, mistrust, or dissatisfaction with current efforts. Therefore, applying sentiment analysis to social media data can serve as a powerful tool for capturing public attitudes, supporting risk communication, informing policy planning, and developing responsive measures that align more closely with societal needs.

Negative sentiment accounted for the majority of comments (52.1%), clearly dominating public opinion. Positive sentiment comprised 30.0% of the comments, reflecting appreciation for mitigation efforts and protective behavior. Neutral sentiment made up the remaining 17.9%, which typically consisted of factual statements or questions. The visual representation emphasizes the public's prevailing dissatisfaction and concern over PM2.5 pollution, reinforcing the need for more transparent communication and responsive policymaking.

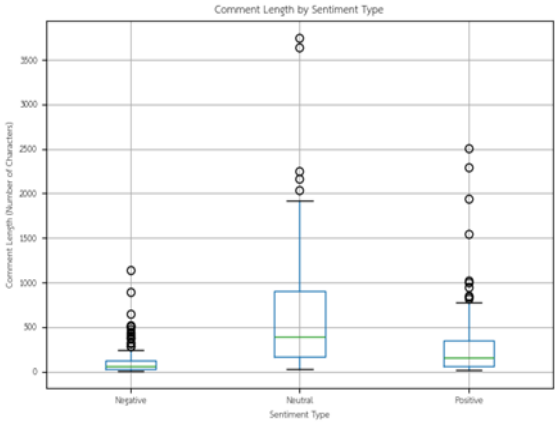

Fig. 4 Boxplot of Comment Length by Sentiment Category

Figure 4 illustrates the distribution of comment lengths for each sentiment category using a boxplot. The analysis reveals that neutral comments exhibit the widest variation and longest average lengths, often involving detailed factual descriptions or inquiries. Negative comments tend to be more concise but

emotionally expressive, while positive comments are generally shorter and contain expressions of encouragement or gratitude. These patterns suggest that public sentiment influences the verbosity and structure of online discourse, offering deeper insights into emotional intensity and public engagement.



Fig. 5 Word Cloud – Positive Sentiment

Figure 5 illustrates a word cloud generated from comments classified as positive sentiment. Frequently occurring keywords include endure, good, air, mask, wear, filter, thank you, and reassured. These terms reflect the public's adaptive behaviors and coping strategies, such as wearing particulate-filtering masks and expressing appreciation for governmental or organizational efforts in implementing dust mitigation measures.



Fig. 6 Word Cloud - Neutral Sentiment

Fig.6 Word Cloud – Neutral Sentiment

Figure 6 presents the word cloud derived from comments labeled as neutral. Commonly observed words include PM2.5, dust, explain, ask, reason, time, news, and system. These comments often represent inquiries, factual descriptions, or opinion exchanges without overt emotional expression. The neutral sentiment group plays a critical role in capturing public discourse that is informative or knowledge-seeking in nature.

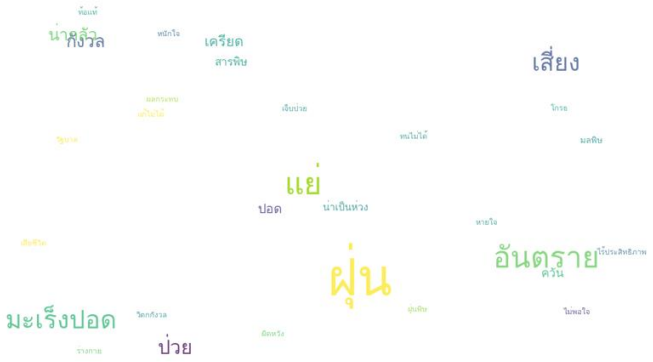

Fig. 7 Word Cloud - Negative Sentiment

Fig. 7 Word Cloud – Negative Sentiment

Figure 7 shows the word cloud for comments categorized as negative sentiment. Prominent terms include terrible, stressed, none, unsolvable, discouraged, exhausted, sick, and ineffective. These words strongly reflect public frustration and dissatisfaction with the PM2.5 situation, especially regarding health impacts and the perceived inefficiency of governmental responses.

The results of the sentiment analysis reveal that negative sentiment was the most dominant among public comments, reflecting widespread dissatisfaction, concern, and high expectations toward government action in addressing PM2.5 pollution. Many comments expressed frustration related to health impacts, delayed responses, and perceived inefficiencies in public sector measures.

These findings are consistent with previous research that conducted sentiment analysis on PM2.5-related tweets and found that negative emotions such as anxiety, anger, and distrust were predominant during pollution events (Wongsawat et al., 2022). Similarly, previous research reported that negative sentiment was more prevalent in Facebook discussions than either positive or neutral opinions, often associated with skepticism toward governmental efforts and perceived inaction (Chompunuch et al., 2021).

Although positive sentiment was less prevalent in the dataset, it reflected public resilience and adaptability. Comments in this category included messages of gratitude, support for responsible agencies, and the adoption of protective behaviors such as wearing masks. These remarks demonstrate optimism and hope for improvement despite ongoing challenges.

Meanwhile, neutral sentiment—though emotionally unmarked—offered valuable insights into public engagement. Many neutral comments consisted of information requests, experience sharing, or factual observations. These interactions reflect the public's desire for understanding and clarity, and they may serve as an important basis for improving risk communication strategies and policy engagement frameworks in the future.

Discussion

Thailand has formulated strategic plans for managing PM2.5 in alignment with the World Health Organization (WHO) guidelines, comparable to other countries in the region (Sukcharoen et al., 2020).

However, public sentiment continues to reflect dissatisfaction with governmental measures. This discontent may indicate a gap between public expectations and the actual outcomes of policy implementation.

Research has shown that sentiment analysis is effective in capturing public support for environmental sustainability, especially when government communication is clear and policy directions are transparent (Kularbphettong et al., 2024). In such contexts, positive sentiment and support for state initiatives tend to increase. On the other hand, findings from this study show that negative sentiments remain dominant in online discourse, especially on platforms such as Pantip.com. This reflects ongoing concern, dissatisfaction, and unmet expectations regarding governmental responses to PM2.5 pollution.

The influence of social media on individual decision-making and public perceptions has been widely recognized. Social media applications significantly impact students' choices regarding higher education enrollment. This finding supports the notion that social media platforms, including online forums such as Pantip.com, can shape public attitudes and behaviors toward environmental issues like air pollution (Ali et al., 2021). In line with this perspective, the predominance of negative sentiment regarding PM2.5 pollution observed in this study may reflect broader societal concerns. This underscores the critical importance of integrating social media sentiment analysis into environmental policy development frameworks to better align governmental responses with public expectations.

Furthermore, there is an urgent need for more effective communication strategies, trust-building, and enhanced public engagement in air quality management. In a modern digital governance framework, digital-era governance should guide information technologies, especially AI tools such as natural language processing (NLP) and machine learning (ML) to monitor public discourse in real time and adapt policies more responsively to societal needs (Ojo & Mellouli, 2018). Similarly, digital platforms that enhance citizen participation have the potential to promote more inclusive and transparent governance. In this regard, government agencies should adopt citizen-centered systems to facilitate participatory governance by creating opportunities for citizens to take part in environmental policy decision-making (Kantaros et al., 2021). Moreover, incorporating AI and machine learning in policy development can enhance the responsiveness and effectiveness of governance, effectively bridging the gap between policy formulation and public needs (Yar et al., 2022).

In addition, sentiment analysis is not simply a technical tool but rather serves as a “bridge” connecting society and government, fostering co-creation, interaction, and the exchange of knowledge (Verma, 2023). This idea is further supported by the development of large language models (LLMs) for analyzing public opinions from platforms such as X (Twitter). By capturing the “voices of the people” on a community level, such an approach enables government agencies to tailor policies to meet the specific contextual needs of each community (Seki, 2023).

Future research should extend beyond Pantip.com to include digital platforms such as Facebook, X (Twitter), and YouTube to capture a more diverse and representative set of public views. Integrating deep

learning techniques and context-aware models could further enhance classification accuracy—especially for mixed-emotion or ambiguous content. Additional exploration of cross-platform comparisons, multilingual datasets, and multimodal sentiment analysis (e.g., combining text with images) could significantly expand the potential for public sentiment to inform inclusive and responsive environmental policymaking.

Conclusion

The study revealed that among the three algorithms tested, Support Vector Machine (SVM) outperformed both Naïve Bayes and Logistic Regression across all evaluation metrics: accuracy, precision, recall, and F1-score. These findings confirm the suitability of SVM for sentiment classification in Thai, particularly in its ability to consistently interpret complex and emotionally nuanced public opinions. This suggests that SVM could serve as a foundation for developing real-time public sentiment monitoring systems to support feedback-driven policy. Eventually, from a policy standpoint, government agencies should leverage social media data to monitor and evaluate environmental impact closely. In parallel, communication strategies should be crafted to ensure transparency, accessibility, and emotional resonance with the public.

This research contributes to the expanding body of knowledge on social media analytics applied to environmental governance. It offers a practical framework for incorporating public sentiment into data-driven policymaking and risk communication, particularly within the context of urban air quality management and policy evaluation—bridging the gap between "policy" and "people" through the power of artificial intelligence and digital participation. In addition to its practical applications, this study demonstrates the role of artificial intelligence in the environmental policy domain. By implementing Natural Language Processing (NLP) tailored for Thai language and using Support Vector Machines (SVM) for sentiment classification, this research presents a scalable model for extracting public opinion from online platforms. The technical contributions not only enhance responsiveness but also align with Sustainable Development Goals (SDG 11 and SDG 13), reinforcing goals around urban resilience and climate action through citizen-informed governance.

References

Ali, A. M., Shamsudin, M. F., Ali, A. M., Ishak, M. F., & Esa, S. A. (2021). The Use of Social Media Application as a Factor Influencing the Students' Decision-Making to Enroll at Private Higher Education Institutions using SmartPLS. *Journal of Information Technology Management*, 13(3), 187–19.

Anderson, T., Sarkar, S., & Kelley, R. (2024). Analyzing public sentiment on sustainability: A comprehensive review and application of sentiment analysis techniques.” *Natural Language Processing Journal,* 3(1), 12–27.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information, *Transactions of the Association of Computational Linguistics*, 5, 135–146.

Champion, S., & Skinner, C. (2023). The health belief model," in *Health Behavior and Health Education: Theory, Research, and Practice*, 4th ed., San Francisco, CA: Jossey-Bass, 45–65.

Chompunuch, S., Rattana, P., & Limsirorat, K. (2021). PM2.5 public perception via Facebook: Sentiment classification using Naïve Bayes, *J. Info. Tech. & Innovation*, 10(3), 45–54.

Chulabhorn Research Institute, Health impacts from PM2.5 exposure, Bangkok, Thailand, 2024. [Online]. Available: https://www.cri.or.th.

Cortes, C., & Vapnik, V. (1995). Support-vector networks, Machine Learning, 20(3), 273–297.

Effing, J., Van Hillegersberg, & Huibers, T. (2011). Social media and political participation: Are Facebook, Twitter and YouTube democratizing our political systems?, *Electronic Participa tion (ePart 2011), Lecture Notes in Computer Science*, 6847, Springer, Berlin, Heidelberg, 25–35, doi: 10.1007/978-3-642-23333-3_3.

Homchuangsub, K., & Tongngam, S. (2023). The credibility of the information on the website Pantip.com that impacts the decision to purchase products: Case study—Beauty web board, *Grad. Sch. Appl. Stat., Nat. Inst. Dev. Admin*, Bangkok, Thailand.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression, 3rd ed. Hoboken, NJ, USA: Wiley.

Hounkpevi, A. (2020). Sentiment analysis to inform about the SDGs: Immigration in Colombia and Costa Rica, *Cepei, Bogotá, Colombia, Technical Brief*. [Online]. Available: https://cepei.org.

Kaewphet, S. (2021). Perception and response behaviors towards particulate matter 2.5 microns (PM2.5): A case study of motorcycle taxi drivers in Din Daeng district, Bangkok, *M.S. thesis*, Faculty of Political Science, Chulalongkorn University, Bangkok, Thailand.

Kantaros, N., Santivong, A., Rungsang, A., & Chumphonwong, T. (2021). Participatory governance: From the concept to the public sector development in Thailand 4.0, *Journal of Suvarnabhumi Institute of Technology* (in Thai), 7(2), 130–144. [Online]. Available: https://so04.tci-thaijo.org/index.php/svittj/article/view/203232.

Kothandan, J., & Murugesan, P. (2021). ML based social media data emotion analyzer and sentiment classifier with enriched preprocessor, *Journal of Information Technology Management*, 13, Special Issue, 6–20.

Kularbphettong, K., Roonrakwit, P., & Boonseng, C. (2024). Sentiment analysis of the awareness of vironmental sustainability. *Edelweiss Applied Science and Technology*, 8(3), 145–155.

Leesa-nguansuk, S. (2016). Pantip to be knowledge source: Site to bring in experts in various fields, *Bangkok Post*. [Online]. Available:https://www.bangkokpost.com/business/general/1108297/ pantip-to-be-knowledge-source.

Ligthart, J., Catal, R., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: A tertiary study, *Artificial Intelligence Review*, 54, 4997–5053.

Lowphanich, P. (2020). Thai2Fit: Transfer learning approach for Thai NLP, in *Proc. 1st Workshop on NLP for South and Southeast Asian Languages*.

Manaskasemsak, B., Puttitanun, S., Tantisuwankul, J., & Rungsawang, A. (2022). Personalized thread commendation on Thai Internet forum, in *Proc. 2022 Assoc. for Computing Machinery*. 267–272. [Online]. Available: https://doi.org/10.1145/35075483507589.

McCallum, A., & Nigam, K. (1998). "A comparison of event models for naive Bayes text classification," in AAAI-98 Workshop on Learning for Text Categorization.

McDermott, J. D., Zhang, F., & Sakamoto, M. (2024). Evolution of international sentiment towards climate change on Twitter, *Commun. Earth Environ*, 5, article 163.

Mohammadiani, R. P., Abbasimehrb, H., & Malikc, Z. (2022). Social media value creation practices and interactivity of electronic word of mouth systems, *Journal of Information Technology Management*, 14(3), 73–91.

Mueller, W., Vardoulakis, S., & Steinle, S. et al. (2021). A health impact assessment of long-term exposure to particulate air pollution in Thailand, *Environmental Research Letters*, 16(5), 055018.

Ojo, T., & Mellouli, J. (2018). Deploying governance through digital transformation: A systematic literature review, *Government Information Quarterly,* 35(4).

Permana, D., Setiawan, R., & Rahmatullah, A. F. (2024). Sentiment analysis of air pollution on social media: Systematic literature review, *Sistemasi*, 13(1), 65–76.

Phatthiyaphaibun, S., Inrut, T., & Khamhaeng, S. (2020). Thai natural language processing toolkit and text classification*, International journal of advanced computer science and applications*, 11(9), 592–598.

Poapongsakorn, N., Pantakua, K., & Ratchakom, S. (2023). Challenges in tackling PM2.5 crisis, *Bangkok Post.* [Online]. Available:https://www.bangkokpost.com/opinion/opinion/2548171

Polsuwan, T. (2021). Forecasting PM2.5 using hybrid ARIMA-LSTM model in Bangkok area, *M.S. thesis*, Dept. Comp. Sci., Dhurakij Pundit Univ., Bangkok, Thailand. [Online]. Available: https://libdoc. dpu.ac.th

Praphan, T., & Komintarachat, T. (2021). Sentiment dashboard for PM2.5 policy decision-making, *Proceedings of the 18th National Conference on Computing and Information Technology*, 2021, 121–127.

Ratanachai, A., & Putthacharoen, S. (2021). Improving sentiment classification for Thai texts using Word2Vec and SVM, *IEEE Access*,9, 112345–112355.

Saengthong, T. (2023). Policy recommendations for mitigating PM2.5 impact in Bangkok, *J. Humanities Soc.Sci. Mahasarakham university,* 41(2), 221–237.

Sae-Oueng, A., Kerdthaisong, K., Phasook, P., Chuangkrud, P., & Kongyoung, S. (2023). Pantip multi-turn datasets generating from Thai large social platform forum using sentence similarity techniques, in *Proc. Int. Conf. Artif. Intell. Data Sci. (AiDAS)*, Thailand.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval, *Information Processing & Management*, 24(5), 513–523.A.

Seki, K. (2023). Analyzing public opinion on Twitter using large language models for community-driven policy design, in *Proc. Int. Conf. AI for Social Good (AI4SG)*, Tokyo, Japan, 112–119.

Siriwan, P., Rattanawan, K., & Chokchai Charoen, T. (2022). Topic modeling of Pantip data related to PM2.5issues*, J. Digital Society and Data Science*, 6(2), 93–104.

Strategy& PwC. (2015). The rise of content democratization in Southeast Asia, *Strategy& Southeast Asia*. [Online]. Available: https://www.strategyand.pwc.com/gx/en/insights/archive/content-emocratization/content-democratization.pdf.

Sukcharoen, W., Tangaromsuk, P., Sontiatcha ra, M., Waithayakul, K., Savedkairop, C., Poopong pet, J., Kengkoom, R., Bhubhanil, S., & Lapmanee, S. (2020). The Study on Thailand's Particulate Matter 2.5 (PM2.5) Management in Accordance with World Health Organization (WHO) Guidelines. *Vajira Medical Journal: Journal of Urban Medicine,* 64(5), 345–356.

Tansathien, K. (2022). Public attitudes toward government measures to resolve PM2.5 issues in Bangkok, *Rajabhat Buriram Research Journal*, 15(1), 50–61.

Verma, R. (2023). Sentiment Analysis as a Bridge for Civic Engagement: Connecting Society and Policy through Digital Dialogue, *Int. J. Public Adm. Digit. Age*, 5(2), 45–60.

Wongsawat, P., Jindaluang, R., & Preechadi rek, S. (2022). Sentiment analysis of PM2.5-related tweets in Thailand, *In Proceedings of the International Conference on Data Mining and Applications*, 100–107.

World Health Organization (WHO). (2021). WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide, Geneva, Switzerland. [Online]. Available: https://www.who.int/publications/i/item/9789240034228

Yar, S., Virtanen, L., & Ritala, P. (2022). Artificial intelligence in public sector service delivery: *A systematic literature review and research agenda government information quarterly,* 39(1), 101650. https://doi.org/10. 1016/j.giq. 2021. 101650