DIJEMSS: Dinasti International Journal of Education Management and Social Science
E-ISSN: 2686-6331
P-ISSN: 2686-6358
https://dinastipub.org/DIJEMSS dinasti.info@gmail.com +62 811 7404 455
DINASTI PUBLISHER
DOI: https://doi.org/10.38035/dijemss.v6i5

Sentiment Analysis on Government Public Policies: A Systematic Literature Review

Reyhan Widyatna Harwenda[1], Michael David Angelo[2], Indra Budi[3], Aris Budi Santoso[4], Prabu Kresna Putra[5]
[1]Faculty of Computer Science University of Indonesia, Depok, Indonesia, reyhan.widyatna@ui.ac.id
[2]Faculty of Computer Science University of Indonesia, Depok, Indonesia, michael.david31@ui.ac.id
[3]Faculty of Computer Science University of Indonesia, Depok, Indonesia, indra@cs.ui.ac.id
[4]Faculty of Computer Science University of Indonesia, Depok, Indonesia, aris.budi@ui.ac.id
[5]Faculty of Computer Science University of Indonesia, Depok, Indonesia, prabu.kresna@ui.ac.id

Corresponding Author: reyhan.widyatna@ui.ac.id[1]

Abstract: In the digital era, public discourse on government policies has shifted significantly to online platforms. This presents valuable opportunities for governments to assess real-time public sentiment. However, prior studies on sentiment analysis in public policy remain fragmented, often lacking methodological consistency and domain-wide synthesis. This study conducts a Systematic Literature Review (SLR) to consolidate insights on the techniques, datasets, and trends involved in sentiment analysis applied to government development policies. The review identifies SVM, BERT, and Naive Bayes as the most frequently used and effective methods, with SVM excelling in structured data and simpler tasks, and BERT demonstrating superior performance in handling nuanced textual data. Lexicon based tools such as VADER are also used for quick sentiment classification. Social media platforms, particularly Twitter, emerge as the dominant data sources due to their high volume and real-time nature, while evaluation metrics such as precision, recall, F1-score, and confusion matrix are commonly applied to assess model performance. The findings also reveal evolving research interests from early focus on health policies to recent interest in infrastructure, environmental, and technology-related policies. Public sentiment across these areas varies, with health and environmental policies often eliciting negative responses, while technology policies show more neutral to positive sentiment. By synthesizing methods, datasets, evaluation strategies, and policy domains, this review provides a structured foundation to future research and supports policymakers in designing strategies.

Keyword: Sentiment Analysis, Systematic Literature Review, Government Policy, Machine Learning, Deep Learning, Public Opinion.

INTRODUCTION

Development policies in Indonesia often attract significant public attention and spark debate due to their wide-reaching impact. In the current era of digital democracy, public policies are no longer evaluated solely through formal political processes or structured surveys. Citizens

now actively express their opinions on various government initiatives through digital platforms such as Twitter, YouTube, and online forums. As public opinion becomes increasingly visible and accessible, understanding how people respond to policies in real-time has become crucial for creating inclusive, transparent, and responsive governance. However, despite the volume and richness of this public discourse, research efforts to capture and analyze such sentiment often lack coordination, consistency, and methodological clarity.

Sentiment analysis has emerged as a powerful tool to evaluate public responses to government programs and initiatives. Unlike traditional surveys, sentiment analysis leverages large-scale, unstructured data to detect patterns in how citizens feel about policies, programs, and their implementers. Recent developments such as aspect-based sentiment analysis (ABSA) further enhance these capabilities by enabling more granular evaluations of specific components within a policy—such as its goals, budget allocations, or execution timelines [30]. These advancements allow policymakers to go beyond surface-level approval ratings and understand nuanced public concerns.

While the application of sentiment analysis has gained traction in domains such as health, education, and transportation, much of the research remains fragmented. Many studies focus on isolated platforms, simplify sentiment into binary categories (positive/negative), or limit their analysis to specific time frames or topics [26][31]. Moreover, although there are literature reviews available, most of them emphasize algorithmic trends or model performance without critically examining how sentiment analysis methods are applied in the policy context. Some studies even decouple sentiment analysis entirely from its implications for governance and public participation [32].

These limitations highlight the need for a comprehensive synthesis of current research on sentiment analysis in the realm of public policy. A Systematic Literature Review (SLR) provides a structured and replicable approach to evaluating this body of work. Through this method, it becomes possible to identify which sentiment analysis techniques are most frequently used in policy-related studies, assess their comparative effectiveness, understand the datasets and evaluation metrics commonly employed, and trace emerging trends in public discourse across various policy domains.

This study conducts an SLR to explore how sentiment analysis is applied to assess public opinion on development policies. It synthesizes academic insights regarding the tools, methods, and policy topics examined, with a particular focus on infrastructure, environment, technology, and health. The review identifies the most used sentiment analysis techniques and compares their relevance and effectiveness across different policy domains and use cases. It also maps the types of datasets employed such as social media streams and online video comments and examines how these data sources influence analytical outcomes. In addition, this study explores how public sentiment trends have evolved over time across various policy areas. By consolidating this fragmented knowledge, the review offers practical guidance for researchers designing future studies and supports policymakers in integrating real-time public sentiment into policy design, evaluation, and communication strategies.

METHOD

This section explains the application of Systematic Literature Review (SLR) in examining research related to sentiment analysis on public opinion. SLR itself is a method for identifying, evaluating, and interpreting all available research relevant to specific research questions, topics, or phenomena of interest [17]. The process involves three key stages: planning, conducting, and documenting [17].

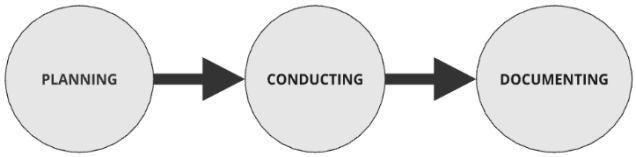

Figure 1. Systematic Literature Review Stages
(adapted from Kitchenham et al., 2009)

A. Planning

The main goal of the planning phase is to ensure that the literature review is conducted in a structured, transparent, and replicable manner. This phase consists of two key components: identifying the need for a systematic review and developing a comprehensive review protocol [17]. The need for a systematic review in this context arises from the growing application of sentiment analysis in public policy studies, which, despite its potential, remains fragmented and inconsistent across domains. Many existing studies focus on narrow topics, specific social media platforms, or isolated technical approaches, and often lack a consolidated understanding of which methods are most frequently used and how effective they are in analyzing public opinion. Moreover, there is limited comparative insight into the use of different techniques, the types of datasets employed, and how sentiment trends and variations manifest across different policy domains, such as health, infrastructure, environment, and technology. To address these gaps, this review formulates the following research questions to guide the investigation:

- What methods are commonly used in sentiment analysis and opinion mining for evaluating public sentiment toward government policies, and how do these methods compare in terms of relevance and effectiveness?
- What datasets are commonly used in sentiment analysis and opinion mining of public sentiment regarding the development policies of national infrastructure?
- What are the most common methods for evaluating sentiment and opinion mining in public policy research?
- What are the emerging trends in public policy topics analyzed using sentiment analysis, and how have these trends evolved over time?
- How does sentiment toward government policies vary, and is public sentiment predominantly positive or negative across different policy areas?

These research questions form the basis for the review protocol and guide the selection, analysis, and synthesis of relevant literature in the subsequent stages.

B. Conducting

Once the protocol is agreed upon, the review process begins. This involves several steps: identifying/searching studies, study selection, quality assessment, data extraction, and data synthesis, which may be done simultaneously, though some steps need to follow a sequential order [17]. Below is a more detailed explanation of these steps:

- Literature Search: The researcher conducts a search in selected databases, including IEEE, ScienceDirect, Scopus, ACM, Emerald Insight, and SpringerLink, using the following query format:

“TITLE-ABS-KEY (("opinion mining" OR "sentiment analysis") AND (“technique” OR “methodology”) AND (“machine learning” OR “lexicon-based”) AND (“public polic” OR “infrastructure project”))”*

- Study Selection: Each piece of literature found is filtered based on the inclusion and exclusion criteria, which can be seen in Table 1.

Table 1. Inclusion and Exclusion Criteria

Stage	Inclusion Criteria	Exclusion Criteria
Initiation	Publications from the last 5 years (2019–2024).	Publications outside the specified period.
	Focuses on sentiment analysis, opinion mining, and stance analysis.	Studies not related to sentiment analysis, opinion mining, or stance analysis.
	Uses machine learning, lexicon-based, or hybrid methods (combination of machine learning and lexicon-based).	Studies not using specific methods or methods other than machine learning, lexicon-based, or hybrid.
Stage 1 (Title and Abstract Selection)	Title and abstract describing sentiment analysis and opinion mining.	Title and abstract do not describe sentiment analysis and opinion mining.
	Studies that mention specific techniques/methods such as machine learning, lexicon, or hybrid.	Studies not applying specific/clear methods.
Stage 2 (Full-Text Selection)	Studies that describe methodologies and results of sentiment analysis and opinion mining in detail.	Studies that do not clearly describe the methodology and results.
	Studies providing experimental details of their methods.	Studies that do not provide experimental details of their methods.

- Quality Assessment: The selected studies are evaluated for quality to ensure the research findings are based on valid and reliable data [17]. Studies that meet quality standards are rated 1, while those that do not meet quality criteria are rated 0. The quality criteria for data assessment are outlined in Table 2.

Table 2. Paper Quality Checklist

Checklist	Checklist Questions
C1	Does the article clearly describe the research objectives?
C2	Does the article include a literature review, background, and context of the research?
C3	Does the article present related work to demonstrate the main contributions of the research?
C4	Does the article describe the proposed architecture or methodology used?
C5	Does the article provide research results?
C6	Does the article present conclusions relevant to the research objectives/problem?
C7	Does the article recommend future work or improvements?
C8	Is it indexed in Scopus (Q1 / Q2 / Q3 / Q4 / unindexed)?

- Data Extraction: During this phase, key data from each selected study will be extracted, including information on research objectives, methods, results, and conclusions [17].
- Data Synthesis: This process combines and analyzes information from multiple studies to draw comprehensive and reliable conclusions [17].

a. C. Documenting

The final stage of the SLR process involves documenting the results. This step includes writing and presenting the research findings in a systematic manner, ensuring that the research questions are answered, and findings are clearly communicated in the final report. The documentation phase also involves:

- Reporting Results: The key findings are summarized and presented in a coherent manner. This includes detailing the methods used, the data sets employed, the trends identified, and how these findings address the research questions.
- Providing Synthesis of Evidence: The results from different studies are compared and synthesized to form a holistic view of the subject matter. This synthesis highlights where consensus exists and where discrepancies remain in literature.
- Recommendations for Future Research: The documenting phase also includes recommendations based on the findings. These recommendations aim to guide future studies in areas where gaps were identified, particularly in terms of methodologies, datasets, or specific policy topics.

RESULTS AND DISCUSSION

This section presents the results from the conducting phase of the Systematic Literature Review (SLR), which includes the processes of identification, selection, and analysis of studies related to sentiment analysis on development policies in Indonesia. The findings from this phase will provide insights into the patterns and trends emerging from the existing literature, as well as recommendations for policymakers.

A. Findings and Paper Selection

The literature search was conducted across several predefined databases. The results from this search showed the total number of studies identified and processed through several selection stages. In the initiation stage, a total of 229 articles were found from various sources. Next, in Stage 1, which involved title and abstract selection, the number of articles passing the selection reduced to 42. Finally, in Stage 2, where full-text selection was conducted, a total of 27 articles were accepted for further analysis. The results of this selection process are summarized in Table 3.

Table 3. Paper Selection Results

Source	Initiation Stage	Stage 1	Stage 2
Scopus	12	9	5
Science Direct	3	3	2
ACM Digital Library	27	1	0
IEEE Xplore	51	15	11
SpringerLink	51	7	4
Emerald Insight	85	7	5
Total	**229**	**42**	**27**

B. Quality Test of Papers

In this phase, the quality of the papers was assessed using the quality criteria checklist, which can be seen in Table 2. The quality test was conducted on the 27 papers that had passed the full-text selection stage, as described earlier. The distribution of the papers based on their publication year and type can be seen in Figures 3 and Figure 4. After the quality assessment, the papers that met the established quality standards were identified and summarized in Table 4 and Figure 2. A more detailed breakdown of the papers that met the quality standards can be found in Table 5.

Table 4. Results of Paper Quality Selection

Source	Number of Final Papers
Scopus	4
Science Direct	1
ACM Digital Library	0
IEEE Xplore	9
SpringerLink	4
Emerald Insight	4
Total	**22**

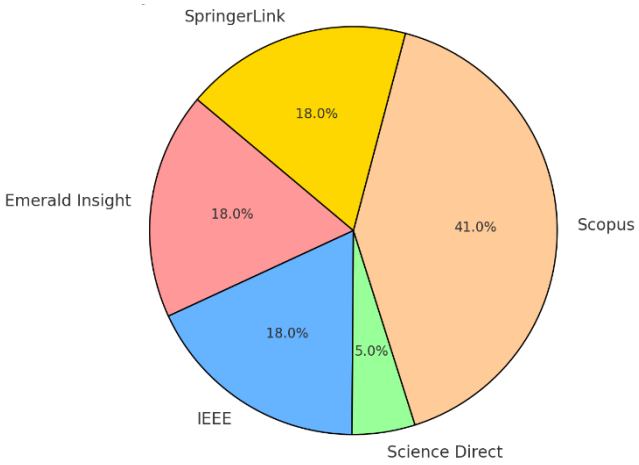

Figure 2. Journal Distribution Based on Database

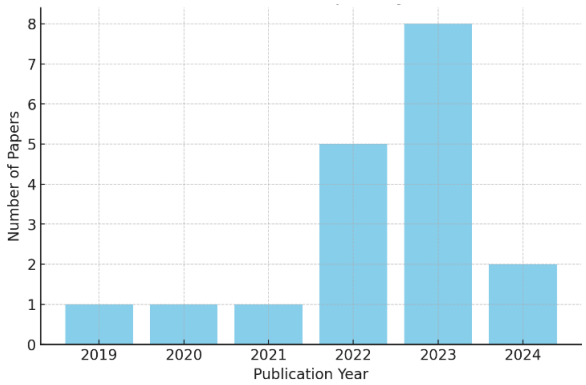

Figure 3. Distribution of Papers by Publication Year

Table 5. Mapping of Journals from Quality Test Results

No	Code	Journal Title
1	A1	A Comparative Study of Classification Algorithms for Sentiment Analysis of Covid-19 Vaccine Opinions Using Machine Learning [2]
2	A2	A Comparison of Oversampling and Undersampling Methods in Sentiment Analysis Regarding Indonesia Fuel Price Increase Using *Support Vector Machine* [8]
3	A3	A Machine Learning-Sentiment Analysis on Monkeypox Outbreak: An Extensive Dataset to Show the Polarity of Public Opinion From Twitter Tweets [7]
4	A4	An Exploratory Content and Sentiment Analysis of the Guardian

		Metaverse Articles Using Leximancer and Natural Language Processing [18]
5	A5	Analysis of Sentiment Towards Artificial Intelligent Industry Using Hybrid Natural Language Processing Technique [10]
6	A6	Analyzing online public opinion on Thailand-China high-speed train and Laos-China railway mega-projects using advanced machine learning for sentiment analysis [11]
7	A7	Comparing tweet sentiments in megacities using machine learning techniques: In the midst of COVID-19 [3]
8	A8	Comprehensive review and comparative analysis of transformer models in sentiment analysis [19]
9	A9	Decoding mood of the Twitter verse on ESG investing: opinion mining and key themes using machine learning [9]
10	A10	Induction of knowledge, attitude and practice of people toward a pandemic from Twitter: a comprehensive model based on opinion mining [20]
11	A11	Intelligent Learning based Opinion Mining Model for Governmental Decision Making [6]
12	A12	Multidimensional mining of public opinion in emergency events [15]
13	A13	Opinion Mining for National Security: Techniques, Domain Applications, Challenges, and Research Opportunities [21]
14	A14	Peruvian Presidential Elections of 2021 in Twitter/X: A Sentiment Analysis Approach [13]
15	A15	Sentiment Analysis Classification System Using Hybrid BERT Models [22]
16	A16	Sentiment Analysis in Twitter Using Lexicon Based and Polarity Multiplication [23]
17	A17	Sentiment Analysis of Air Quality Perception in Major Metro Cities of India [16]
18	A18	Sentiment Analysis of Air Quality to Identify Public Perspectives in

		Indonesia Based on Machine Learning [12]
19	A19	Sentiment Analysis of Social Media for Indonesian m-Health Peduli Lindungi Mobile-Apps (PLMA) with Lexicon-Based and *Support Vector Machine* Approach [24]
20	A20	Towards Smart City: Aspect Based Sentiment Analysis of Indonesian Public Aspiration Complaints Data Using Machine Learning [14]
21	A21	Twitter Opinion Sentiment Analysis Based on New Student Admission Zoning Issues Using the Naïve Bayes and TensorFlow Methods [5]
22	A22	Using Data Mining Technology to Analyze the Spatiotemporal Public Opinion of COVID-19 Vaccine on Social Media [4]

C. Data Extraction

Research Question 1: What methods are commonly used in sentiment analysis and opinion mining for evaluating public sentiment toward government policies, and how do these methods compare in terms of relevance and effectiveness. Based on the extraction of data from the selected literature, several methods frequently used in public sentiment analysis were identified. Support Vector Machine (SVM) was the most commonly used method, applied 8 times due to its ability to separate data with optimal margins, making it highly effective in sentiment classification. Naive Bayes was used 4 times, well-known for its simplicity and efficiency in handling simpler classification tasks. Random Forest and Logistic Regression were used 3 times each, offering advantages in ensemble classification and fast, efficient regression. Additionally, deep learning-based methods like BERT were used 3 times to handle more complex and nuanced textual data. Other methods such as Latent Dirichlet Allocation (LDA), VADER, CNN, and LSTM also appeared frequently, each with specific advantages depending on data complexity and research needs. A summary of the methods/techniques used is shown in Table 6.

Tabel 6. Public Sentiment Analysis Methods

Method	**Frequency**
SVM (*Support Vector Machine*)	8
Naive Bayes	4
Random Forest	3
Logistic Regression	3
BERT	3
Latent Dirichlet Allocation (LDA)	3
VADER	3
Convolutional Neural Network (CNN)	2
Long Short-Term Memory (LSTM)	2

RoBERTa	2
TextBlob	2
DistilBERT	2
Gradient Boosting Classifier	1
Extra Trees	1
Multilayer Perceptron (MLP)	1
K-Nearest Neighbor (KNN)	1
GPT-3.5	1

In comparing the sentiment analysis methods identified across the reviewed studies, several patterns emerge regarding their typical use cases and levels of complexity. Traditional machine learning methods (SVM, Naive Bayes, Random Forest, Logistic Regression) are still commonly used for sentiment analysis tasks based on structured datasets. On the other hand, deep learning methods (CNN, LSTM, BERT, RoBERTa) are employed for large and complex datasets, particularly in fields like social media and politics, where more complex context needs to be analyzed. Lastly, lexicon-based approaches such as VADER and TextBlob are often used for quick sentiment analysis that relies on predefined lexicons of words.

Table 7. Explanation of Sentiment Analysis Algorithms

Algorithm	Context of Use
Support Vector Machine (SVM)	Commonly used in sentiment analysis to classify data into different sentiment categories (positive, negative, neutral).
Naive Bayes	Used for text classification due to its efficiency in analyzing labeled data like public opinions.
Random Forest (RF)	An ensemble technique that improves classification accuracy, often used in public policy analysis.
Logistic Regression (LR)	Used to predict the probability of text being positive or negative in sentiment analysis.
Convolutional Neural Networks (CNN) and LSTM	CNN detects patterns in data, while LSTM handles long-term text sequence data.
BERT and RoBERTa	Transformer models used for NLP tasks with high accuracy in sentiment classification.
Lexicon-Based Approach	Lexicon-based approaches like VADER and TextBlob are used to determine sentiment polarity.

Latent Dirichlet Allocation (LDA)	Used for topic modeling to extract main themes from textual data.
K-Nearest Neighbor (KNN)	A neighbor-based algorithm for sentiment classification.
GPT-3.5	Used for deep learning-based sentiment classification, handling complex tasks.

To determine which sentiment analysis methods are most relevant and effective in evaluating public sentiment toward government policies, this review highlights six techniques that stand out based on their frequency of use and demonstrated performance in the selected studies. These methods are Support Vector Machine (SVM), Naive Bayes, Random Forest, Latent Dirichlet Allocation (LDA), BERT, and RoBERTa were consistently applied across a variety of policy contexts, including health, education, infrastructure, and social development. The selection of these six methods is supported by their ability to handle both structured and unstructured data, making them versatile tools for sentiment and opinion mining in complex government-related discourse. Among them, SVM emerged as the most widely used technique, praised for its accuracy and robustness in classifying sentiment across different datasets. BERT, a transformer-based model, follows closely due to its strength in capturing contextual meaning in long and nuanced text, which is especially useful in analyzing public reactions to multifaceted policies. Naive Bayes and Random Forest are also frequently used, particularly for simpler or more structured sentiment analysis tasks where computational efficiency is a priority. Meanwhile, LDA is commonly employed for topic modeling to uncover underlying themes in public discussions, and RoBERTa demonstrates high accuracy in deeper contextual analysis. The prevalence and performance of these six methods are summarized in Table 8 and visually illustrated in Figure 6, highlighting their central role in sentiment analysis studies focused on public policy.

Table 8. Frequency of Algorithm Occurrence

Algorithm	Number of Papers	Percentage
SVM	7	39%
BERT	4	22%
Naïve bayes	2	11%
Random Forest	2	11%
Extra Trees	1	6%
Latent Dirichlet Allocation (LDA)	1	6%
Others	1	6%

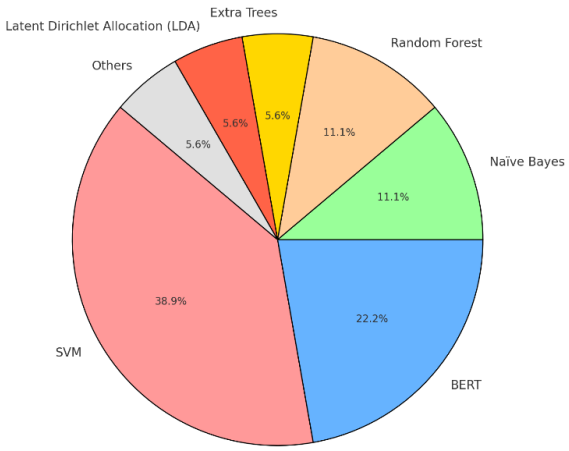

Figure 4. Comparison of Best Algorithms

Based on the data extraction results, the most relevant and effective methods for sentiment analysis and opinion mining of public sentiment toward government policies can be seen in Table 8 and Figure 4. Detailed explanations of these methods are as follows.

1. *Support Vector Machine* (SVM)

SVM is highly suitable for sentiment analysis of government policies due to its ability to classify opinions with clear margins between sentiment categories (positive, negative, neutral). In various studies, SVM has been used to analyze public opinion on policies such as fuel price hikes and large infrastructure projects. SVM also often provides accurate results, particularly when used with techniques for handling data imbalance, such as SMOTE. It is a good choice for datasets that are not too large and have imbalanced sentiment classes. An example of SVM usage can be found in studies regarding fuel price policy in Indonesia, where SVM demonstrated high effectiveness with accuracy in classifying sentiment [21].

2. Transformer Models (BERT, RoBERTa)

Transformer models like BERT and RoBERTa are highly relevant when the nuances of public opinion on policies need to be understood in depth. These models allow for more contextual sentiment analysis, which is important for capturing complex sentiments related to policies with broad social impacts. Transformer models could understand the context of long and complex texts, resulting in more accurate sentiment analysis compared to traditional methods. They are particularly effective for handling large datasets such as social media, which is often used to discuss government policies. A paper that used RoBERTa to analyze public opinion on ESG investments shows that this model performs exceptionally well in understanding public opinions and sentiments in complex and lengthy discussions [9]. For quick and efficient sentiment analysis related to public policies, such as on social media with smaller datasets, SVM and Naive Bayes are highly relevant and effective choices. For more complex and contextual sentiment analysis, especially when dealing with large datasets, such as public opinions from various social media sources regarding government policies, BERT, RoBERTa, and Random Forest yield better results, although they require more computational resources.

3. Naive Bayes

Naive Bayes is often used in sentiment analysis related to government policies due to its simplicity and efficiency. This algorithm works well for fast text classification in relatively large datasets, such as social media reviews about public policies. Despite its simplicity, Naive Bayes provides satisfactory performance with adequate accuracy in public sentiment analysis.

It is often used as a baseline in many studies, including those related to government policies. An example of Naive Bayes usage can be seen in papers analyzing sentiment regarding air quality policies in India and Indonesia, where Naive Bayes was used successfully to distinguish positive and negative opinions [16].

4. Random Forest (RF)

Random Forest (RF) is a robust algorithm and highly relevant for sentiment analysis and opinion mining, where collective decision-making from several "decision trees" provides more stable and accurate predictions, especially when the dataset analyzed involves public opinion on policies. RF yields excellent results in terms of accuracy and precision, particularly when managing more complex datasets with diverse text features. An example of RF application can be found in studies using RF to analyze public opinion related to the COVID-19 vaccine, where RF proved to be highly effective in delivering more accurate predictions compared to other algorithms [15].

5. Latent Dirichlet Allocation (LDA)

LDA is highly relevant for topic modeling in public opinion toward policies. This algorithm is used to discover the main themes emerging in public discussions on specific policies, such as COVID-19 vaccination or public transportation initiatives. LDA is effective in identifying dominant topics from large datasets related to policies, helping policymakers understand the main issues being discussed by the public. A paper using LDA to analyze public opinion on the Thailand-China high-speed rail project found that LDA efficiently helped identify the key issues discussed by the public [13].

Research Question 2: What datasets are commonly used in sentiment analysis and opinion mining of public sentiment regarding national infrastructure development policies. Based on the data extraction results, several datasets have been used in research related to sentiment analysis and opinion mining of public sentiment toward national infrastructure development policies. Some of the most frequently used datasets include data from social media platforms, public review platforms, and data specific to particular infrastructure projects. The data sets most used are summarized in Table 9 and Figure 5.

Table 9. Percentage of Datasets Used in Papers

Dataset	Frequency	Percentage
Twitter	11	50%
Youtube	3	14%
Weibo	3	14%
Tiktok	1	5%
Website Resmi	1	5%
Kaggle	1	5%
The Guardian	1	5%
Baidu Tieba	1	5%

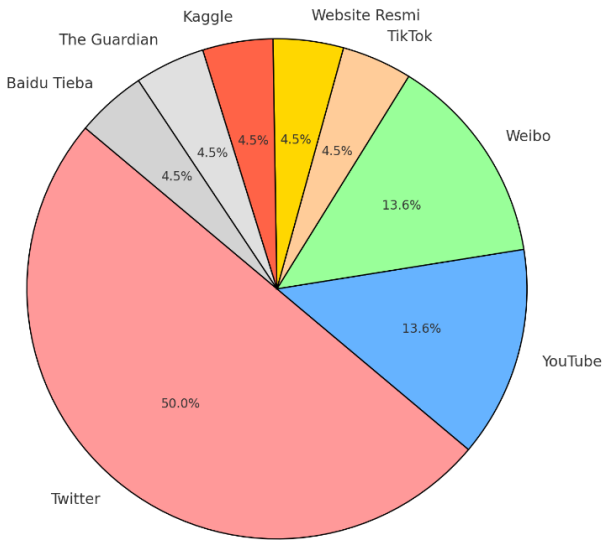

Figure 5. Dataset Usage in Sentiment Analysis Studies

The reason Twitter is the most frequently used dataset is due to its status as a primary data source in sentiment analysis research. With billions of tweets posted daily, Twitter provides real-time data from the public on various policies, opinions, and current events, including major infrastructure projects. Additionally, Twitter offers a large volume of data that encompasses not only positive or negative opinions but also discussions on various aspects of infrastructure policies, including both criticisms and support.

Research Question 3: What are the most common methods for evaluating sentiment and opinion mining. Based on the data from the studies presented, the most used evaluation method for sentiment analysis and opinion mining is the confusion matrix. The confusion matrix is the most widely used evaluation tool in binary or multi-class classification tasks, including sentiment analysis. This matrix provides information about True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) from the classification results. It is commonly used because it offers a comprehensive way to evaluate the performance of sentiment analysis models, including their ability to correctly classify sentiment and handle class imbalance. In addition to the confusion matrix, other evaluation metrics such as precision, recall, and F1-score are often derived from it. These metrics provide more nuanced insights into model performance, especially in the case of imbalanced datasets where the number of positive and negative samples may differ significantly. Precision focuses on the accuracy of positive predictions, while recall measures how many actual positive instances were identified by the model. The F1-score provides a balance between precision and recall, making it a useful metric when both false positives and false negatives carry significant consequences. In conclusion, while the confusion matrix remains the most used method for evaluating sentiment analysis and opinion mining, integrating other performance metrics such as precision, recall, F1-score, and cross-validation offers a more comprehensive assessment of model performance, especially when dealing with complex and imbalanced datasets. These evaluation methods are essential for refining sentiment analysis models and ensuring their effectiveness in real-world applications, such as policy analysis and public opinion monitoring.

Research Question 4: What are the emerging trends in public policy topics analyzed using sentiment analysis, and how have these trends evolved over time. Based on the literature reviewed, several emerging trends have been identified in the application of sentiment analysis to public policy. Initially, a large portion of studies focused on health policies, particularly regarding COVID-19 measures, where public sentiment was often negative due to government

restrictions and vaccination campaigns [2][4]. As time progressed, attention shifted toward infrastructure and economic policies, with public sentiment toward large-scale projects, such as high-speed rail initiatives and smart city development, becoming key areas of analysis [6][14]. Additionally, environmental policies and public opinions on air quality have gained prominence, reflecting growing concerns about sustainability and environmental impact [16]. The rise of AI-related policies has also attracted attention, with public sentiment regarding technological advancements, particularly artificial intelligence, revealing a mix of optimism and caution regarding its societal implications [18][19]. These emerging trends highlight a growing interest in analyzing public opinion across a broader range of policy areas, especially as social media platforms continue to provide real-time, diverse insights into public sentiment.

Policy Area	References
Health Policies	[2] [4] [16] [12]
Infrastructure Policies	[6] [14]
Environmental Policies	[16] [12]
Technology Policies	[10] [19] [23]
Others	[18] [7] [8] [13] [5]

The shift in focus from health policies to infrastructure, environmental policies, and technology-related policies can be attributed to several factors. First, the COVID-19 pandemic prompted an immediate need for sentiment analysis in health policy to gauge public reaction to government interventions. As the pandemic subsided, attention shifted naturally to long-term infrastructure development and economic recovery, especially in response to global financial disruptions. Infrastructure policies, particularly those involving large-scale projects such as high-speed rail and smart cities, are seen as vital for revitalizing economies. Meanwhile, growing concerns about climate change have made environmental sustainability a key area of public interest, prompting more research into public sentiment regarding environmental policies. Additionally, the rapid advancements in artificial intelligence (AI) have sparked debates about its societal impacts, prompting increasing interest in sentiment analysis regarding AI-related policies and their implications for privacy, employment, and governance. The increasing use of social media platforms has also played a significant role in these trends, as these platforms offer real-time, diverse public opinions that are invaluable for policy evaluation and adjustments.

Research Question 5: How does sentiment toward government policies vary and is public sentiment predominantly positive or negative across different policy areas. The sentiment toward government policies varies significantly depending on the policy area. In the context of health policies, such as COVID-19 vaccine distribution, the sentiment is generally negative, with public dissatisfaction over the government's handling of the pandemic, especially in terms of communication and implementation [2][4]. For economic and infrastructure policies, such as fuel price increases or large-scale development projects, sentiment tends to be more mixed. While some public groups may show support for long-term economic stability, there is also a significant portion of the population expressing negative sentiment due to the immediate financial burden these policies impose on them [8][21]. Environmental policies, particularly those addressing air quality and pollution, tend to receive predominantly negative sentiment, reflecting public frustration over government inaction or perceived inefficiency in addressing environmental concerns [16][12].

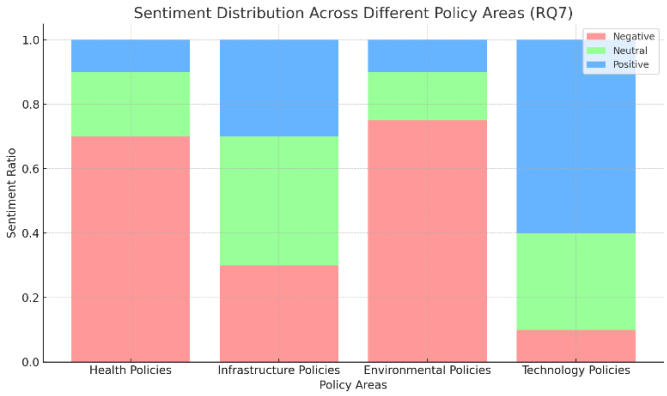

Figure 6. Sentiment Distribution Across Different Policy

In contrast, policies related to technology and AI governance tend to elicit more neutral to positive sentiments, as public opinion leans toward optimism regarding technological advancements, though concerns about privacy and job displacement remain [10][9]. Technology policies related to artificial intelligence and machine learning are increasingly viewed as a pathway to economic growth and social development, which is why they often receive positive sentiment from communities eager for innovation. However, there are some negative reactions as well, especially in relation to the ethical implications of AI and its potential effects on the workforce.

Overall, while negative sentiment is common in response to policies that directly affect daily life or financial well-being, there is also significant nuanced sentiment that includes mixed or neutral reactions, especially for more complex or long-term policies. This highlights the complexity of public sentiment, which cannot always be categorized simply as positive or negative but requires a more detailed analysis to capture subtle shifts in public opinion over time.

D. Synthesis Result

From the analysis of several papers, many of them utilize Natural Language Processing (NLP) techniques and sentiment analysis to examine public opinion on various topics. For instance, the algorithms Leximancer and VADER were used to analyze topics and sentiment related to the metaverse [18]. Other papers employ similar methods, such as lexicon-based analysis or machine learning approaches (e.g., Random Forest, Support Vector Machine). Compared to machine learning methods, lexicon-based approaches like VADER are often used for processing large text datasets in sentiment analysis. However, machine learning approaches like Random Forest and SVM tend to perform better in classifying more complex text data.

Although methods like Lexicon-based or VADER are effective for basic sentiment analysis, some papers mention that these methods tend to be less accurate when dealing with more complex or ambiguous contexts. In contrast, deep learning models like BERT or RoBERTa excel at understanding deeper contexts but require significant computational resources and more intensive training. Additionally, papers analyzing public opinion on government policies or the metaverse often criticize issues like data bias and dependency on major tech platforms. The use of non-transparent algorithms and the lack of oversight over harmful content, such as misinformation or digital harassment, are concerns highlighted in these studies.

The primary differences between these studies lie in the use of different techniques and tools based on data complexity. Papers focusing on basic sentiment analysis often use methods like VADER or Leximancer, which are simpler and easier to implement. However, studies dealing with more complex data or requiring deeper classification, such as deep learning models

(BERT, GPT-3.5), show superior results, particularly in handling ambiguity and nuanced sentiment.

Based on the analysis of various papers, it is clear that NLP methods play a significant role in sentiment analysis and public opinion mining, particularly for emerging topics like government policies or new technologies like the metaverse. These studies suggest using more advanced methods like BERT to improve analysis accuracy, especially when dealing with unstructured data. Additionally, many papers emphasize the need for hybrid methods, such as combining topic modeling with sentiment analysis, to provide a more comprehensive view of the data. Tools like Leximancer or Monkeylearn also make it easier to automate large-scale data analysis.

Over time, the focus of sentiment analysis research has shifted from health policies to infrastructure, environmental policies, and technology-related policies. Initially, health policies, particularly those related to COVID-19 and vaccination, dominated the research landscape, with negative sentiment being most prevalent due to public dissatisfaction with government handling of the pandemic [2][4]. Over time, attention shifted to infrastructure policies, especially large-scale projects like high-speed rail and smart city development, which gained significant attention for their impact on economic stability and urban growth [6][14]. Alongside this, environmental policies related to air quality and pollution became a growing focus as climate change concerns began to shape public opinion [16]. Moreover, the rise of artificial intelligence (AI) and related technological policies began attracting attention, with sentiment showing a mix of optimism and caution regarding the societal implications of technological advancements [18][19].

Sentiment toward government policies varies significantly depending on the policy area. For health policies, such as COVID-19 vaccine distribution, public sentiment is generally negative, reflecting dissatisfaction with the government's handling of the pandemic, particularly in terms of communication and implementation [2][4]. For economic and infrastructure policies, such as fuel price increases or large-scale development projects, sentiment tends to be mixed. Some segments of the public express support for long-term economic stability, while others show negative sentiment due to the immediate financial burden these policies impose on them [8][21]. Environmental policies, especially those related to air quality and pollution, tend to receive predominantly negative sentiment, reflecting frustration with government inaction or perceived inefficiency in addressing environmental issues [16][12]. In contrast, policies related to technology and AI governance tend to elicit more neutral to positive sentiments, with public opinion generally leaning toward optimism about technological advancements, although concerns about privacy and job displacement remain [10][9].

CONCLUSION

This systematic literature review (SLR) provides a comprehensive synthesis of sentiment analysis and opinion mining methods applied to the evaluation of government policies. The findings reveal that Support Vector Machine (SVM) is the most frequently used and effective method, applied in 39% of the studies. SVM's strength lies in its ability to separate sentiment classes with optimal margins, making it highly suitable for structured datasets with clearly defined positive, negative, or neutral opinions. Naive Bayes and Random Forest also demonstrate strong performance, especially in scenarios where simplicity and computational efficiency are prioritized. Logistic Regression continues to be used due to its interpretability and solid performance with structured input. Meanwhile, deep learning approaches such as BERT (22%), RoBERTa, CNN, and LSTM are gaining traction, particularly for complex sentiment analysis tasks involving longer, unstructured, and more nuanced text data. BERT has shown strong contextual understanding, making it highly effective in capturing subtle sentiment shifts.

Regarding datasets, Twitter emerges as the dominant source, used in 50% of the studies, due to its real-time nature, accessibility, and high volume of public policy discourse. Other platforms like YouTube and Weibo are also used, particularly in regional studies or when analyzing sentiment toward infrastructure-related policies. This prevalence highlights the utility of social media as a primary data source for gauging public opinion on government actions. As such, method selection should align with the data source and complexity: traditional machine learning techniques are well-suited for smaller, structured tasks, while transformer-based models like BERT are ideal for analyzing large, diverse social media datasets with more complex linguistic features. By consolidating commonly used techniques, evaluation strategies (e.g., precision, recall, F1-score, confusion matrix), and key policy domains (health, infrastructure, environment, and technology), this review serves as a guide for both researchers and practitioners in developing data-driven, citizen-responsive strategies for public policy evaluation.

While this review provides significant insights, several areas offer opportunities for future research. First, most studies rely on single-platform data; future work could integrate sentiment data from multiple platforms (e.g., combining Twitter with YouTube or TikTok) for a more holistic view of public opinion across demographics. Second, multi-lingual sentiment analysis should be prioritized, particularly in diverse or multi-ethnic regions, to ensure inclusive policy feedback from non-English-speaking communities. Third, building real-time sentiment analysis systems would enhance governments' ability to respond rapidly to public concerns. Fourth, combining topic modelling techniques such as Latent Dirichlet Allocation (LDA) with sentiment classification could improve understanding of both the content and tone of public discussions. Finally, with the growing use of AI in analyzing social behaviour, future studies must critically address ethical concerns such as bias, misclassification, and the potential misuse of sentiment data in political or administrative decisions. Addressing these gaps will further enhance the value of sentiment analysis as a strategic tool for participatory and accountable governance.

REFERENCE

B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering – A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, Jan. 2009, doi: 10.1016/j.infsof.2008.09.009.

D. S. Zainulabdeen, M. Çevik, and M. M. Abdulrazzaq, "A Comparative Study of Classification Algorithms for Sentiment Analysis of Covid-19 Vaccine Opinions Using Machine Learning," in *2024 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, Istanbul, Turkiye: IEEE, May 2024, pp. 1–6. doi: 10.1109/HORA61326.2024.10550643.

Z. Yao, J. Yang, J. Liu, M. Keith, and C. Guan, "Comparing tweet sentiments in megacities using machine learning techniques: In the midst of COVID-19," *Cities*, vol. 116, p. 103273, Sep. 2021, doi: 10.1016/j.cities.2021.103273.

T. Li, Z. Zeng, J. Sun, and S. Sun, "Using data mining technology to analyse the spatiotemporal public opinion of COVID-19 vaccine on social media," *Electron. Libr.*, vol. 40, no. 4, pp. 435–452, Jan. 2022, doi: 10.1108/EL-03-2022-0062.

A. A. Dwisanny and S. Supatmi, "Twitter Opinion Sentiment Analysis Based on New Student Admission Zoning Issues Using the Naïve Bayes and TensorFlow Methods," *ICSPIS 2023 - Proceedings of the 9th International Conference on Signal Processing and Intelligent Systems*. 2023. doi: 10.1109/ICSPIS59665.2023.10402745.

A. Sharma and H. Shekhar, "Intelligent Learning based Opinion Mining Model for Governmental Decision Making," *Procedia Computer Science*, vol. 173. pp. 216–224, 2020. doi: 10.1016/j.procs.2020.06.026.

S. Bengesi, T. Oladunni, R. Olusegun, and H. Audu, "A Machine Learning-Sentiment Analysis on Monkeypox Outbreak: An Extensive Dataset to Show the Polarity of Public Opinion From Twitter Tweets," *IEEE Access*, vol. 11, pp. 11811–11826, 2023, doi: 10.1109/ACCESS.2023.3242290.

Z. A. Faridzi, D. Pramesti, and R. Y. Fa'rifah, "A Comparison of Oversampling and Undersampling Methods in Sentiment Analysis Regarding Indonesia Fuel Price Increase Using Support Vector Machine," in *2023 International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS)*, Bali, Indonesia: IEEE, Aug. 2023, pp. 1–6. doi: 10.1109/ICADEIS58666.2023.10270851.

R. Jaiswal, S. Gupta, and A. K. Tiwari, "Decoding mood of the Twitterverse on ESG investing: opinion mining and key themes using machine learning," *Manag. Res. Rev.*, vol. 47, no. 8, pp. 1221–1252, Jan. 2024, doi: 10.1108/MRR-07-2023-0526.

H. Chang, "Analysis of Sentiment Towards Artificial Intelligent Industry Using Hybrid Natural Language Processing Technique," in *2023 IEEE 8th International Conference On Software Engineering and Computer Systems (ICSECS)*, Penang, Malaysia: IEEE, Aug. 2023, pp. 363–370. doi: 10.1109/ICSECS58457.2023.10256397.

M. Nokkaew *et al.*, "Analyzing online public opinion on Thailand-China high-speed train and Laos-China railway mega-projects using advanced machine learning for sentiment analysis," *Social Network Analysis and Mining*, vol. 14, no. 1. 2024. doi: 10.1007/s13278-023-01168-8.

N. Agustina, A. Hadhiwibowo, I. K. A. S. Saputra, W. Cahyati, Z. Pupi Indriarti, and B. Rustandi, "Sentiment Analysis of Air Quality to Identify Public Perspectives in Indonesia Based on Machine Learning," in *2023 17th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, Lombok, Indonesia: IEEE, Oct. 2023, pp. 1–6. doi: 10.1109/TSSA59948.2023.10366899.

V. A. Ayma Quirita, J. David Cárdenas, W. Aliaga, V. H. Ayma Quirita, A. Palacios, and R. B. Sierra, "Peruvian Presidential Debates in the Elections of 2021 in Twitter/X: A Sentiment Analysis Approach," *IEEE Access*, vol. 12, pp. 138386–138398, 2024, doi: 10.1109/ACCESS.2024.3465503.

M. A. Khadija, I. S. D. Jayanti, and F. U. Nimah, "Towards Smart City: Aspect Based Sentiment Analysis of Indonesian Public Aspiration Complaints Data Using Machine Learning," *Proceedings - International Conference on Informatics and Computational Sciences*. pp. 215–220, 2024. doi: 10.1109/ICICoS62600.2024.10636859.

Q. Zhou and M. Jing, "Multidimensional mining of public opinion in emergency events," *Electron. Libr.*, vol. 38, no. 3, pp. 545–560, Jul. 2020, doi: 10.1108/EL-12-2019-0276.

K. Wadhwa, D. Mehra, H. Gosain, and A. K. Haritash, "Sentiment Analysis of Air Quality Perception in Major Metro Cities of India," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Delhi, India: IEEE, Jul. 2023, pp. 1–5. doi: 10.1109/ICCCNT56998.2023.10307448.

B. Kitchenham, "Procedures for Performing Systematic Reviews".

S. Tunca, B. Sezen, and V. Wilk, "An exploratory content and sentiment analysis of the guardian metaverse articles using leximancer and natural language processing," *J. Big Data*, vol. 10, no. 1, May 2023, doi: 10.1186/s40537-023-00773-w.

H. Bashiri and H. Naderi, "Comprehensive review and comparative analysis of transformer models in sentiment analysis," *Knowl. Inf. Syst.*, Sep. 2024, doi: 10.1007/s10115-024-02214-3.

P. Reisinezhad and M. Fakhrahmad, "Induction of knowledge, attitude and practice of people toward a pandemic from Twitter: a comprehensive model based on opinion mining," *Kybernetes*, vol. 52, no. 7, pp. 2507–2537, Jan. 2023, doi: 10.1108/K-05-2022-0758.

N. A. M. Razali *et al.*, "Opinion mining for national security: techniques, domain applications, challenges and research opportunities," *J. Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00536-5.

A. S. Talaat, "Sentiment analysis classification system using hybrid BERT models," *J. Big Data*, vol. 10, no. 1, Jun. 2023, doi: 10.1186/s40537-023-00781-w.

Kusrini and M. Mashuri, "Sentiment Analysis In Twitter Using Lexicon Based and Polarity Multiplication," in *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT)*, Yogyakarta, Indonesia: IEEE, Mar. 2019, pp. 365–368. doi: 10.1109/ICAIIT.2019.8834477.

Kartini, A. K. Darmawan, R. I. Syah, and M. Makruf, "Sentiment Analysis of Social Media for Indonesian m-Health PeduliLindungi Mobile-Apps(PLMA) with Lexicon-Based and Support Vector Machine Approach," in *2023 IEEE 9th Information Technology International Seminar (ITIS)*, Batu Malang, Indonesia: IEEE, Oct. 2023, pp. 1–7. doi: 10.1109/ITIS59651.2023.10419994.

Kamal, M. R., Affandy, Pujiono, & Hasibuan, Z. A. (2024). Systematic Literature Review on Aspect-Based Sentiment Analysis: Insights from Community Reports. Proceedings - 2024 International of Seminar on Application for Technology of Information and Communication: Smart And Emerging Technology for a Better Life, ISemantic 2024, 392–397. https://doi.org/10.1109/iSemantic63362.2024.10762054.

Verma, S. (2022). Sentiment analysis of public services for smart society: Literature review and future research directions. Government Information Quarterly, 39(3), 101708. https://doi.org/10.1016/j.giq.2022.101708.

Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. In Artificial Intelligence Review (Vol. 54, Issue 7). Springer Netherlands. https://doi.org/10.1007/s10462-021-09973-3

Renteria, C., & Gil-Garcia, J. R. (2017). A systematic literature review of the relationships between policy analysis and information technologies: Understanding and integrating multiple conceptualizations. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10429 LNCS(July), 112–124. https://doi.org/10.1007/978-3-319-64322-9_10.

Hua, Y. C., Denny, P., Wicker, J., & Taskova, K. (2024). A systematic review of aspect-based sentiment analysis: domains, methods, and trends. In Artificial Intelligence Review (Vol. 57, Issue 11). Springer Netherlands. https://doi.org/10.1007/s10462-024-10906-z.

Zhao, Y., Li, Y., Liu, Y., Li, Q. (2022). Aspect Based Fine-Grained Sentiment Analysis for Public Policy Opinion Mining. In: Chen, J., Hashimoto, T., Tang, X., Wu, J. (eds) Knowledge and Systems Sciences. KSS 2022. Communications in Computer and Information Science, vol 1592.

Salah, A. A., Almalki, A., & Zaki, N. (2019). A survey on sentiment analysis of social media in smart cities. Journal of Big Data, 6, Article 13. https://doi.org/10.1186/s40537-019-0183-2.

Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis – A review of research topics, venues, and top cited papers. Computer Science Review, 27, 16–32. https://doi.org/10.1016/j.cosrev.2017.10.002.