Sentiment Analysis of Public Discourse on Education in Indonesia Using Support Vector Machine (SVM) and Natural Language Processing

B Herawan Hayadi[1,*], Ika Maulita[2]

[1] Primary School Teacher Education,Universitas Bina Bangsa, Serang, Indonesia

[2] Physics Department, Universitas Jenderal Soedirman, Indonesia

ABSTRACT

The growing influence of social media platforms like Twitter has transformed the landscape of public discourse, particularly on critical societal issues such as education. This study investigates public sentiment on education in Indonesia by analyzing tweets collected between January and June 2024, using sentiment analysis techniques powered by the Support Vector Machine (SVM) algorithm. By leveraging a dataset of 484 tweets, the analysis classified sentiments into positive, neutral, and negative categories, uncovering dominant patterns and their correlation with significant events in the education sector. The findings revealed a strong prevalence of neutral sentiments, emphasizing Twitter's role as an informational hub. Positive sentiments were linked to public approval of equity-focused reforms, such as increased funding for rural schools, while negative sentiments reflected dissatisfaction with contentious policies, particularly standardized testing reforms. The study also examined temporal trends, identifying spikes in sentiment coinciding with major policy announcements, such as the significant surge in neutral sentiments in May 2024 following the government's testing policy changes. These patterns illustrate the dynamic nature of public engagement on Twitter, shaped by real-time events and discussions. A comparative analysis with existing literature confirmed the value of social media as a barometer for public opinion, while also highlighting the unique context of Indonesian education. This research contributes to the growing field of digital society by demonstrating how sentiment analysis can provide actionable insights for policymakers and educators. It underscores the transformative potential of social media analytics in fostering inclusive and responsive governance. However, limitations such as dataset biases and classification challenges suggest avenues for future research, including multi-platform analysis and advanced natural language processing techniques. These findings serve as a foundation for leveraging sentiment analysis to enhance educational strategies and public communication in Indonesia's increasingly digitalized landscape.

Keywords Sentiment Analysis, Education, Twitter, Support Vector Machine, Public Opinion

Submitted 29 December 2024
Accepted 2 February 2025
Published 8 March 2025

Corresponding author
B Herawan Hayadi,
b.herawan.hayadi@binabangsa.ac.id

Additional Information and Declarations can be found on page 87

Introduction

The pervasive role of social media in shaping public discourse is undeniable, particularly as it intersects with the ever-critical domain of education. Platforms like Twitter, Facebook, and Instagram have evolved beyond mere communication tools into dynamic arenas where societal concerns converge and collide. This transformation underscores a profound shift in how opinions on educational policies, reforms, and practices are formed and disseminated. Unlike traditional media, which often acts as a one-way channel, social media cultivates an environment of interaction, enabling users to not only consume

How to cite this article: B. H. Hayadi and I. Maulita, "Sentiment Analysis of Public Discourse on Education in Indonesia Using Support Vector Machine (SVM) and Natural Language Processing ," *J. Digit. Soc.*, vol. 1, no. 1, pp. 69-90, 2025.

information but actively contribute to ongoing debates. Such participatory mechanisms are reshaping the contours of public engagement, emphasizing collective dialogue over individual observation.

At the heart of this phenomenon lies the democratization of opinion sharing, a feature intrinsic to social media's architecture. By providing an accessible platform for a multitude of voices—students, parents, educators, policymakers, and the public—social media disrupts hierarchies that once governed educational discourse. Empirical studies have illuminated this trend, highlighting how platforms encapsulate the diversity of societal perspectives, making visible those narratives often obscured in traditional educational policymaking [1], [2]. For instance, viral hashtags like #PendidikanIndonesia create shared spaces for nuanced discussions, allowing participants to coalesce around common concerns, whether addressing inequitable access to resources or the merits of innovative teaching methodologies.

However, this newfound openness is not without complications. The dynamic interplay of visibility and influence can, paradoxically, suppress certain voices while amplifying others. The "spiral of silence" theory, as discussed by [3], underscores how perceived majority opinions can dominate social media spaces, discouraging dissenting views from surfacing. This challenge is particularly acute in the educational sphere, where ideological divides—such as debates over standardized testing or digital learning tools—can polarize discussions. Compounding this is the pervasive issue of misinformation, as the sheer volume of content disseminated on social media often muddies the waters between fact and opinion [4]. In such an environment, the quest for constructive, evidence-based discourse becomes a formidable task.

Nevertheless, the potential of social media to galvanize grassroots movements and enact tangible change in education cannot be overlooked. Platforms have played a pivotal role in organizing advocacy campaigns, rallying support for reforms that address systemic inequities in educational systems. Research [5] highlights how grassroots digital activism, often spurred by widespread social media engagement, has been instrumental in shaping policy agendas and amplifying calls for justice in education. This dual capacity—both as a site of intellectual deliberation and a springboard for action—positions social media at the nexus of contemporary educational discourse. Its influence, while multifaceted and fraught with challenges, remains a testament to its transformative potential in navigating the complexities of public opinion and institutional reform.

The evolution of Indonesia's education system since the democratic shift of 1998 reveals a nation grappling with the complexities of reforming a foundational pillar of its society. Education, long considered a bridge to economic and social equity, has become a focal point of public discourse. Public opinion, reflecting the collective voice of diverse stakeholders, emerges not only as a mirror of societal aspirations but also as a catalyst for change. From calls to improve infrastructure in rural schools to demands for curriculum innovation, the dynamic interplay between public sentiment and policy action underscores the significance of this dialogue in shaping Indonesia's educational landscape.

A critical aspect of this transformation is the drive to create a curriculum that embodies the nation's rich cultural and religious pluralism. Historically, religious education in Indonesia has centered on unifying principles, yet contemporary discourse suggests a pivot towards embracing multiculturalism. Research [6] captures this shift, emphasizing the growing demand for an educational framework that respects and integrates the mosaic of cultural identities across

Indonesia. Public opinion, voiced through both traditional channels and emerging platforms like social media, amplifies these demands, with parents and educators advocating for a curriculum that fosters mutual respect and prepares students for participation in a globalized, pluralistic society.

The rise of social media as a conduit for educational discourse has further democratized public engagement. Platforms like Twitter and Instagram have become arenas for dialogue, enabling individuals to share grievances, insights, and aspirations about educational policies. Research by [7] highlights the pivotal role of digital platforms in empowering college students to articulate their concerns and mobilize action. Issues such as the stark disparities in education quality between urban and rural areas gain prominence through these platforms, catalyzing discussions that transcend geographic and socioeconomic barriers [8]. This digital mobilization reflects the profound impact of technology on collective advocacy, transforming passive dissatisfaction into actionable momentum.

Yet, the path to meaningful reform is fraught with systemic challenges. Despite increased investment in education and decentralization efforts, inefficiencies and corruption continue to undermine progress. As Research [9] notes, corruption erodes the effectiveness of public spending, perpetuating inequities and diminishing trust in public institutions. This distrust is both a symptom and a driver of continued advocacy for accountability within the education system. Public awareness of these issues, fueled by digital tools and grassroots movements, underscores a critical paradox: while public opinion has the power to spur reform, entrenched systemic barriers often resist transformation. Addressing this paradox requires not only robust policy frameworks but also sustained public engagement to ensure that reform efforts align with societal needs and aspirations.

The ascendancy of Twitter as a platform for public discourse in Indonesia underscores the transformative power of social media in shaping collective thought, particularly on critical topics such as education. Twitter's real-time nature enables the rapid exchange of ideas, fostering a participatory culture where users—ranging from students and parents to educators and policymakers—can converge to share insights, voice concerns, and advocate for change. In this digital agora, the discussion of educational reforms, curriculum disparities, and institutional challenges gains a breadth and immediacy that traditional media rarely achieves.

Indonesia's educational discourse on Twitter reflects a broader trend of using the platform as a site for public engagement and mobilization. Scholars such as Sauter and Bruns have observed that Twitter's integration with media consumption transforms it into a dynamic space where opinions are not only expressed but actively reorganized in response to evolving debates [10]. This is particularly evident in a nation grappling with stark regional disparities in education quality and access. Twitter's potential to amplify marginalized voices has proven vital, offering a counterbalance to top-down narratives and highlighting systemic inequities that might otherwise be overlooked.

Beyond dialogue, Twitter's interactive affordances drive awareness and advocacy. Research by Lee et al. emphasizes that platforms like Twitter are instrumental in shaping public understanding of news and policy issues, effectively broadening the scope of civic participation [11]. This dynamic resonates in Indonesia's educational sphere, where grassroots campaigns often leverage hashtags to rally support for equitable reforms. A compelling example is the 2018 East Java provincial election, during which Twitter became a pivotal tool for framing debates around education, enabling activists to mobilize collective efforts for policy changes [12]. Such instances illustrate Twitter's dual role as both a reflector of public sentiment and a catalyst for organized action.

The platform's collaborative potential further enhances its impact on educational discourse. Twitter serves as a bridge connecting diverse stakeholders, allowing for the exchange of ideas and resources in ways that enrich both pedagogical practices and policy dialogues. Bista highlights the capacity of social media to foster collaboration, noting that platforms like Twitter create shared spaces for students and educators to engage in meaningful discussions and mutual learning [13]. This intersection of discourse and pedagogy transforms Twitter from a mere tool for expression into a medium that actively contributes to the educational ecosystem. Yet, the inherent risks of misinformation and narrative distortion remain a significant challenge. As Miller et al. caution, the dynamics of influence and rapid information dissemination on Twitter can skew public understanding, necessitating a critical approach to the platform's content [14]. Such challenges underscore the need for vigilance as social media continues to shape Indonesia's educational future.

The complexity of Indonesia's education system lies not only in its vast and diverse landscape but also in the myriad opinions and emotions it elicits from its stakeholders. Public sentiment towards education, an often-overlooked yet deeply influential factor, plays a pivotal role in shaping perceptions of equity, quality, and accessibility. However, this sentiment is not a monolithic construct; it varies across geographic regions, socioeconomic classes, and cultural backgrounds. The lack of systematic understanding of these sentiments poses a critical gap in the policymaking process. Policymakers often grapple with conflicting narratives, where the voices of marginalized communities risk being overshadowed by dominant, urban-centric dialogues. This study seeks to address this gap by leveraging Twitter data to unearth nuanced public sentiment on education, thereby offering an empirical foundation for informed decision-making.

Recent advancements in sentiment analysis and data mining have provided robust methodologies for understanding public discourse on various topics, including education and social issues. The application of machine learning techniques, such as Support Vector Machine (SVM) and TF-IDF, has been demonstrated effectively in studies like [15] and [16], showcasing the versatility of these approaches in analyzing sentiment from digital platforms. Comparative evaluations of sentiment classification methods, as explored in [17] and [18], further emphasize the importance of selecting appropriate algorithms for context-specific analyses. In the realm of Twitter-based sentiment analysis, studies such as [19] and [20] highlight the significance of understanding sentiment trends to uncover public opinion patterns. Additionally, investigations into Twitter discourse, including [21] and [22], underline the role of social media as a dynamic medium for public engagement and analysis. In educational data mining, works such as [23] and [24] offer insights into leveraging machine learning to analyze behavioral and performance data, providing a broader perspective on data-driven decision-making in education. Together, these studies form a foundation for employing sentiment analysis and machine learning to analyze public discourse and guide actionable insights.

The objective of this study centers on harnessing data mining techniques to analyze sentiment in tweets about education in Indonesia. By employing a blend of natural language processing and machine learning methodologies, the study aims to categorize sentiments—positive, neutral, or negative—and identify prevailing themes within the discourse. Unlike traditional surveys or opinion polls, this approach captures organic, real-time expressions of public opinion, providing a dynamic snapshot of societal attitudes. Furthermore, by extracting actionable insights from the dataset, this research strives to contribute to the growing field of computational social science, where big data analytics intersects with the study of human behavior.

Understanding public sentiment is not merely an academic exercise; it has profound implications for policy and practice. Policymakers and educators operate within a context where public trust and perception significantly impact the success of reforms. A clearer picture of societal attitudes could help refine educational strategies, ensuring they resonate with the needs and aspirations of the populace. For instance, if sentiment analysis reveals widespread frustration over curriculum rigidity, this insight could drive efforts to introduce more flexible, student-centered learning models. Conversely, identifying positive sentiment towards digital learning tools might encourage greater investment in educational technology. In this way, sentiment analysis acts as a barometer, guiding stakeholders toward evidence-based interventions that align with public priorities.

The significance of this study extends beyond the immediate findings; it underscores the transformative potential of integrating data-driven insights into educational policymaking. As education systems globally face mounting challenges—from resource disparities to the pressures of globalization—understanding public sentiment becomes a cornerstone of adaptive governance. By focusing on Indonesia, this research highlights the broader applicability of data mining techniques in capturing the pulse of public opinion. It not only amplifies the voices of the many but also bridges the gap between abstract policy frameworks and the lived realities of students, educators, and families. In doing so, it lays the groundwork for a more inclusive, responsive, and equitable approach to educational reform.

Literature Review

Sentiment Analysis in Social Media

The application of sentiment analysis to social media platforms, especially Twitter, has transformed how researchers and policymakers understand public opinion on education. By employing computational techniques to parse and classify sentiments embedded in tweets, this approach offers a lens into the nuanced and dynamic landscape of societal perceptions. Education, as a pivotal domain in public discourse, benefits significantly from such insights, enabling stakeholders to discern attitudes and emotions that might otherwise remain obscured in traditional feedback mechanisms.

Xue et al. provided a compelling case for sentiment analysis by examining public discourse during the COVID-19 pandemic, showcasing how machine learning models effectively categorized tweets into emotional expressions like fear, anger, and hope [25]. Their findings underscored the capacity of sentiment analysis to capture real-time societal reactions to critical events. Similarly, in educational contexts, the same methodological frameworks can illuminate public sentiment on reforms, policies, or crises, thereby informing targeted interventions. For instance, the surge in remote learning during the pandemic heightened debates on accessibility and equity, making sentiment analysis a valuable tool for monitoring public attitudes.

Research by Aydin et al. delved deeper into education-specific applications, focusing on open and distance education (ODE) systems in Turkey. Their study highlighted how sentiment analysis could enhance understanding of student experiences, revealing that improvements in algorithmic techniques increased classification accuracy from 56% to 75% [26]. Such precision is not merely technical but profoundly practical; accurate sentiment classification enables educators to respond effectively to student concerns, fostering a more responsive and adaptive educational environment. These findings align with broader observations by Boon-Itt and Skunkan, who noted that sentiment trends during the pandemic, particularly in education, often skewed negative due to widespread disruptions and uncertainties [27].

The technical landscape of sentiment analysis is equally significant. Fütterer's comparative study on analytical tools such as VADER demonstrated that certain models outperform others when applied to education-related tweets, emphasizing the need for methodological rigor [28]. Similarly, Özyurt and Kisa adopted a dictionary-based approach to classify sentiments during the shift to distance learning, offering granular insights into public perceptions of online education [29]. However, challenges persist. Informal language and the high prevalence of neutral tweets on Twitter complicate sentiment classification, as noted by Trupthi et al. and Akbari et al., who advocated for adaptive techniques to address these complexities [30], [31]. These insights are echoed in the Indonesian context, where Irfan et al. employed advanced algorithms to analyze sentiments surrounding the 2013 curriculum, demonstrating the tangible impact of sentiment analysis on educational policy discourse [32].

Together, these studies reflect the evolving sophistication of sentiment analysis as a tool for understanding education-related discussions. They reveal a growing consensus: that sentiment analysis is not merely a technological exercise but a critical endeavor to bridge the gap between public opinion and policy action. As computational methods advance, their integration into education research holds immense promise, offering pathways to more informed, inclusive, and responsive educational systems.

Text Mining and Natural Language Processing (NLP) Techniques

The application of text mining and Natural Language Processing (NLP) techniques in sentiment analysis has revolutionized the way researchers interpret public opinion, especially within the vibrant ecosystem of social media. Twitter and similar platforms generate massive amounts of text data daily, offering a rich vein of information about societal trends and emotions. Among the arsenal of machine learning tools employed to process this textual goldmine, Support Vector Machine (SVM) stands out for its precision and adaptability. While SVM has emerged as a leading choice, other methods such as Naïve Bayes, Random Forest, and hybrid approaches have also demonstrated their utility in specific contexts. This section examines the methodologies commonly used in sentiment analysis, with a particular focus on SVM and its comparative strengths.

At the core of SVM's effectiveness is its ability to construct an optimal hyperplane that maximizes the separation between data points of different classes in a multi-dimensional space. This geometric precision enables SVM to achieve high levels of accuracy in classification tasks, particularly for binary sentiment problems, such as distinguishing between positive and negative tweets [33]. Its reliance on support vectors—key data points that define the hyperplane—ensures computational efficiency and robustness against overfitting, making it a reliable choice for large-scale sentiment analysis.

Empirical evidence underscores SVM's dominance in text classification. Rausanfita's study on Instagram comments revealed that SVM outperformed traditional algorithms like Naïve Bayes and Random Forest, achieving superior accuracy and recall rates [34]. Similarly, Lutfi et al. compared SVM with Maximum Entropy and Naïve Bayes, finding that SVM consistently delivered more accurate sentiment classifications, particularly when paired with optimized preprocessing techniques [35]. These findings highlight the algorithm's adaptability to diverse datasets, cementing its reputation as a cornerstone in NLP-driven sentiment analysis.

While SVM is a front-runner, other machine learning methods also play critical roles in sentiment analysis. Naïve Bayes (NB), a probabilistic classifier based on Bayes' theorem, has long been favored for its simplicity and computational efficiency. Its independence assumption, although simplistic, allows for rapid classification, particularly in balanced datasets [36]. However, studies often reveal that SVM surpasses NB in accuracy, particularly for nuanced datasets with imbalanced class distributions [37].

Random Forest and Decision Trees, both ensemble-based methods, bring added depth to sentiment classification through their hierarchical decision-making processes. By aggregating predictions from multiple decision trees, Random Forest reduces the risk of overfitting and enhances classification stability. Nonetheless, Rezki et al. noted that while these models offer competitive performance, SVM frequently edges them out in precision and scalability, particularly when applied to unstructured text data [37].

The evolution of sentiment analysis has ushered in a new era of hybrid models that integrate multiple algorithms for enhanced performance. One innovative approach involved combining SVM with Word2Vec embeddings, as demonstrated by Rezki et al., to classify sentiments related to educational policies. This fusion of SVM's classification power and Word2Vec's semantic representation achieved remarkable accuracy, illustrating the potential of hybrid techniques to address the complexities of sentiment-laden text [37]. Such models bridge the gap between statistical precision and contextual understanding, enabling richer insights from sentiment data.

However, sentiment analysis is not without its hurdles. Neutral sentiments, often more prevalent than their positive or negative counterparts, complicate classification by introducing ambiguity into the data [38]. Additionally, the informal language, slang, and abbreviations characteristic of social media platforms pose significant challenges for traditional models, which often rely on structured linguistic inputs [38]. These obstacles underscore the need for ongoing innovation in NLP techniques to ensure that sentiment analysis remains both accurate and contextually relevant in increasingly diverse digital environments.

Previous Studies on Education Sentiment in Indonesia

The study of public sentiment regarding education in Indonesia has increasingly leveraged social media as a data source, reflecting the dynamic interplay between digital discourse and societal attitudes. By employing sentiment analysis, researchers have explored diverse perspectives on educational policies, curricula, and broader socio-political issues that influence the education sector. These studies illustrate the potential of computational methods to uncover nuanced public opinions, shedding light on the evolving educational landscape.

One prominent example of sentiment analysis in Indonesia is the examination of the Merdeka Belajar Kampus Merdeka policy. Rezki et al. utilized Support Vector Machine (SVM) coupled with Word2Vec embeddings to analyze public sentiment towards this transformative educational reform, achieving high classification accuracy and revealing mixed public responses [37]. This research underscores the importance of understanding public opinion in shaping the success and reception of policy initiatives.

In a related effort, Atsqalani et al. analyzed public sentiment regarding the 2013 curriculum, a significant educational reform in Indonesia. Using the K-Nearest Neighbors (K-NN) algorithm, the study achieved an impressive accuracy rate of 96.36%, demonstrating the effectiveness of machine learning techniques in interpreting complex social media data [39]. These findings align with a broader recognition of social media as a barometer for public sentiment, offering policymakers actionable insights into public perceptions of curriculum changes.

Beyond explicit educational topics, Rausanfita's analysis of Instagram comments related to Indonesia's 2024 presidential candidates highlights an indirect connection between political discourse and educational sentiment. By utilizing SVM and other classifiers, this study identified public concerns that intersected with educational issues, such as resource allocation and national priorities [34]. Similarly, Sutoyo and Almaarif's exploration of public opinion on Indonesia's capital relocation illuminated discussions about regional educational disparities, demonstrating how significant national debates can influence public perceptions of education [40].

These studies predominantly employed machine learning techniques such as SVM, Naïve Bayes, and K-NN for sentiment classification, with SVM emerging as the preferred algorithm due to its robustness and accuracy. Lutfi et al. compared SVM with Naïve Bayes and Maximum Entropy, finding that SVM consistently outperformed other methods in text classification tasks, particularly for sentiment analysis related to education [35]. The effectiveness of SVM in distinguishing nuanced sentiments highlights its applicability in analyzing Indonesia's multifaceted educational discourse.

Despite these advancements, sentiment analysis in the Indonesian educational context faces notable challenges. The informal language, abbreviations, and emotive expressions prevalent on platforms like Twitter and Instagram complicate the classification process, often leading to a high prevalence of neutral sentiments [38]. Additionally, regional linguistic variations add another layer of complexity to sentiment analysis. To address these challenges, researchers have proposed integrating hybrid models, combining machine learning techniques with advanced natural language processing frameworks, to enhance accuracy and interpretive depth. Such innovations could significantly contribute to understanding and shaping Indonesia's educational policies and practices in an increasingly digitalized society.

Gaps in Current Research

Despite significant advancements in sentiment analysis research within Indonesia, there remains an evident void in the application of real-time sentiment analysis to education-related discussions on Twitter. While numerous studies have examined sentiment trends across various domains—ranging from tourism to political reforms—the educational sector has not been explored with the same rigor. This oversight is particularly notable given the profound societal impact of education and the growing reliance on social media platforms like

Twitter to voice public concerns and shape discourse. Addressing this gap offers an opportunity to deepen our understanding of public sentiment in a sector critical to Indonesia's development trajectory.

Existing research has largely concentrated on broader applications of sentiment analysis without prioritizing education. For example, studies have successfully mapped public sentiment toward Indonesia's tourism destinations, employing machine learning methods to highlight trends in visitor perceptions [41]. Similarly, analyses of public discourse surrounding major policy initiatives, such as the relocation of the nation's capital, have shed light on general societal attitudes [40]. However, these investigations often bypass the nuanced discussions surrounding education, despite its pivotal role in shaping Indonesia's socioeconomic fabric. This omission underscores the need for targeted research that delves into the intricacies of educational sentiment.

The absence of real-time sentiment analysis in educational contexts further compounds this research gap. While studies by Das et al. have demonstrated the utility of real-time sentiment analysis for predicting stock price fluctuations based on Twitter data, similar approaches remain underutilized in educational discourse [42]. This disparity highlights a methodological limitation, as real-time analysis enables researchers to capture the evolving dynamics of public opinion with greater precision. Existing studies on education, such as those examining the transition to online learning during the COVID-19 pandemic, often adopt retrospective or aggregated approaches, missing the immediacy and richness of real-time data [43]. This lack of temporal granularity limits the actionable insights that policymakers and educators can derive from such research.

A final critical gap lies in the platform-specific focus of current studies. Research by Pratama on educational sentiment during the pandemic has predominantly analyzed Facebook posts, thereby overlooking Twitter's distinct role as a hub for concise and dynamic public discourse [44]. This oversight is significant given that Twitter serves as a primary platform for public engagement in Indonesia, boasting one of the world's most active user bases. Unlike Facebook, where discussions may be more localized or personal, Twitter facilitates the rapid dissemination of ideas, enabling collective conversations that often influence broader narratives. Leveraging Twitter's real-time data for educational sentiment analysis thus presents an untapped opportunity to enrich understanding and inform policy decisions in ways previously unachieved.

By addressing these gaps, the present study aims to establish a comprehensive framework for analyzing real-time educational sentiment on Twitter in Indonesia. This endeavor not only bridges an existing research void but also enhances the methodological toolkit available for exploring public discourse in one of the most digitally connected societies in the world.

Method

The research method is a structured and comprehensive approach that involves carefully planned steps to ensure a thorough analysis of the subject matter. Each step is designed to contribute to the overall goal of gaining a deep understanding and generating meaningful insights. Figure 1 serves as a detailed visual representation, illustrating the sequence and nature of these steps, which are essential for conducting a well-organized and effective research study.

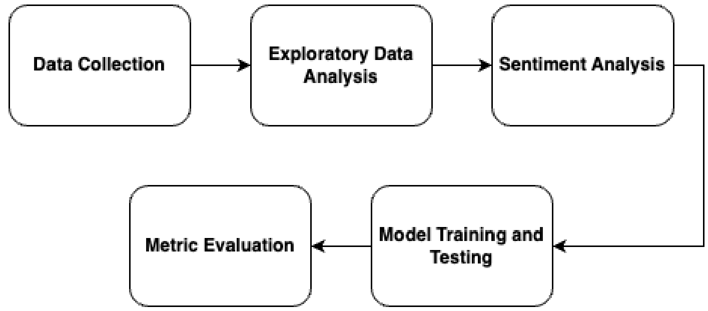

Figure 1 **Research Method Flowchart**

Data Collection

The dataset for this study was meticulously curated to provide a comprehensive representation of public sentiment on education in Indonesia. The data was collected from Twitter, a platform renowned for its dynamic and real-time discourse, using the keyword "pendidikan indonesia" over a defined period from January 1, 2024, to June 21, 2024. This specific keyword was chosen to capture discussions that directly pertain to the education system in Indonesia, ensuring relevance and thematic coherence. By focusing on a six-month timeframe, the dataset encapsulates a diverse array of conversations, including those sparked by significant events or policy announcements in the educational domain.

The dataset comprises 484 entries, each corresponding to a unique tweet. A detailed inspection revealed that the dataset includes 15 distinct columns, encompassing various attributes of the tweets. These columns range from metadata such as `conversation_id_str`, `created_at`, and `username` to engagement metrics like `favorite_count`, `retweet_count`, and `quote_count`. Additionally, textual data is represented in the `full_text` column, capturing the complete content of each tweet. The dataset also provides geographical insights via the `location` column, while links to images and external resources are stored in the `image_url` and `tweet_url` columns, respectively. This multifaceted structure ensures a rich foundation for sentiment analysis, enabling both quantitative and qualitative insights.

The preprocessing stage was crucial to prepare the dataset for analysis. The data was imported using Python's `pandas` library, employing ISO-8859-1 encoding to address potential character encoding issues. During this process, the `on_bad_lines` parameter was set to `skip` to exclude any problematic rows that could disrupt the analysis. A preliminary review of the dataset confirmed the presence of essential information while highlighting minor inconsistencies, such as missing values in columns like `image_url` and `in_reply_to_screen_name`. These omissions were expected given the optional nature of these attributes on Twitter and were addressed through targeted data cleaning steps.

By structuring the dataset to include both the textual content of tweets and associated metadata, this study ensures a holistic approach to analyzing public sentiment on education in Indonesia. The inclusion of temporal data (`created_at`) facilitates the exploration of sentiment trends over time, while engagement metrics provide additional context about the impact and reach of specific tweets. This robust dataset not only aligns with the study's objective of leveraging real-time Twitter data but also lays a solid groundwork for uncovering nuanced insights into the public discourse surrounding education in Indonesia.

Exploratory Data Analysis (EDA)

The process of EDA serves as the foundation for deriving meaningful insights from the dataset, ensuring the integrity and readiness of the data for further sentiment analysis. In this study, the dataset was scrutinized to identify missing values, clean and tokenize the textual content, and visualize key features to uncover patterns and anomalies. By systematically addressing data quality issues and preparing the textual data for analysis, this phase laid the groundwork for extracting actionable insights into public sentiment on education in Indonesia.

Initial inspection of the dataset revealed that while most critical columns, such as `full_text`, `created_at`, and engagement metrics (e.g., `favorite_count`, `retweet_count`), were complete, certain attributes like `image_url` and `in_reply_to_screen_name` contained substantial missing values. Specifically, `image_url` had 389 missing entries out of 484 rows, reflecting its optional nature on Twitter, while `location` was absent in 197 instances. These columns, being non-essential for sentiment analysis, were either excluded from the analysis or treated as auxiliary metadata. No missing values were detected in the primary text column, `full_text`, ensuring that the core data required for sentiment classification remained intact.

To prepare the text for analysis, a rigorous cleaning process was implemented. This involved removing URLs, mentions, special characters, and numerical values, leaving only alphabetic content. For instance, tweets like "Check out this amazing course on pendidikan indonesia: https://example.com" were transformed into "check out this amazing course on pendidikan indonesia." This normalization step ensured that extraneous elements did not interfere with the sentiment analysis process. Additionally, the `full_text` column was converted to lowercase, enabling consistent tokenization and reducing variability caused by case sensitivity.

Tokenization, the process of splitting text into individual words, was coupled with stopword removal to refine the dataset further. Using the Indonesian stopwords corpus from NLTK and additional custom stopwords from the WordCloud library, common but non-informative words (e.g., "dan," "yang," "di") were excluded from the analysis. For example, the sentence "pendidikan adalah kunci masa depan" was tokenized into "pendidikan, kunci, masa, depan," removing filler words that added little value to sentiment analysis. This enhanced the quality of the textual data, ensuring that the resulting tokens reflected meaningful components of the tweets.

Finally, the dataset was explored visually to uncover trends and distributions. Histograms were used to display the frequency of tweet engagements, such as `favorite_count` and `retweet_count`, revealing a skewed distribution where most tweets received minimal interaction, while a small number garnered significant attention. Additionally, a word cloud visualization highlighted the most frequently occurring terms in the cleaned tweets, emphasizing keywords like "pendidikan," "indonesia," and "masalah," which dominated the discourse. These visualizations provided an intuitive understanding of the dataset, offering both qualitative and quantitative insights into the nature of public sentiment surrounding education.

Through a meticulous EDA process, the dataset was transformed into a structured and insightful resource. The steps undertaken not only addressed inherent challenges in the data but also unlocked its potential for robust sentiment analysis, setting the stage for uncovering nuanced perspectives on education in Indonesia. This comprehensive approach ensured that the subsequent phases of analysis would be grounded in a reliable and well-prepared dataset.

Sentiment Analysis Approach

The sentiment analysis conducted in this study employed the Support Vector Machine (SVM) algorithm, a powerful supervised learning technique recognized for its effectiveness in text classification tasks. The approach encompassed three pivotal phases: training data preparation, feature extraction, and model evaluation. These steps were designed to systematically transform raw Twitter data into actionable insights, shedding light on public sentiment regarding education in Indonesia.

The preparation of training data began with cleaning the textual data to eliminate noise and ensure consistency. The raw tweet content was preprocessed by removing URLs, mentions, and special characters, followed by converting all text to lowercase. This step normalized the input, reducing variability introduced by informal and inconsistent language common in social media discourse. To facilitate sentiment labeling, TextBlob, a lexicon-based NLP tool, was employed to classify tweets into positive, neutral, or negative categories based on their polarity scores. For instance, tweets expressing optimism about education reform were tagged as "positive," while criticisms were categorized as "negative." Neutral tweets, which constituted a significant portion of the dataset, were retained to capture the full spectrum of public sentiment.

Following data preprocessing, feature extraction was conducted using the Term Frequency-Inverse Document Frequency (TF-IDF) vectorization technique. This method quantified the importance of each word in a tweet relative to the entire dataset, enabling the model to focus on terms that carried significant semantic weight. The TF-IDF representation was limited to the 5,000 most relevant features to balance computational efficiency with expressive power. For instance, frequently occurring words like "pendidikan" (education) and "kurikulum" (curriculum) were assigned higher weights, reflecting their centrality to the discourse. This transformation converted the cleaned textual data into a numerical format suitable for input into the SVM classifier.

The SVM classifier, utilizing a linear kernel, was trained on the extracted features. The algorithm's objective was to identify the optimal hyperplane that best separated the data points into their respective sentiment classes—positive, neutral, and negative. The training dataset, comprising 80% of the labeled data, provided the foundation for the model to learn the underlying patterns associated with each sentiment. The choice of a linear kernel was guided by the relatively high dimensionality of the textual features, aligning with SVM's strength in handling sparse datasets.

Model Evaluation

The model's performance was evaluated on a test set representing 20% of the dataset. Key metrics, including accuracy, precision, recall, and F1-score, were computed to assess classification effectiveness. The SVM model achieved an accuracy of 91.75%, with particularly strong performance in distinguishing neutral sentiments, which dominated the dataset. However, the model encountered challenges in classifying positive and negative sentiments, a limitation attributed to the inherent ambiguity and imbalanced distribution of sentiment classes. A confusion matrix was generated to visualize the classification outcomes, highlighting areas for potential improvement through enhanced feature engineering or the incorporation of hybrid modeling techniques.

By leveraging SVM's robustness and the structured preprocessing pipeline, this study demonstrated the feasibility of employing machine learning for sentiment analysis in educational contexts. The insights derived from this approach hold promise for informing policy decisions and fostering a deeper understanding of public sentiment on education in Indonesia.

Visualization Techniques

Visualizations play a critical role in interpreting and presenting the results of sentiment analysis, offering intuitive and accessible insights into complex datasets. These visual tools not only clarified patterns within the dataset but also enhanced the narrative by grounding abstract findings in tangible visual evidence. The word cloud generated from the cleaned tweet data in Figure 2, excluding stopwords, offers a visual representation of the most frequently occurring terms within the discourse surrounding education in Indonesia. Dominant words such as "pendidikan" (education), "indonesia", and "ukt" (Uang Kuliah Tunggal or single tuition fee) highlight the central themes of public discussions. The prominence of "pendidikan" and "indonesia" reflects the national focus on education as a critical societal issue, while "ukt" underscores ongoing debates about the affordability of higher education, a topic that resonates deeply with students and parents alike. Other significant terms, such as "mahasiswa" (students), "kampus" (campus), and "kuliah" (lectures), emphasize the concerns of higher education stakeholders, particularly students navigating rising costs and policy changes.

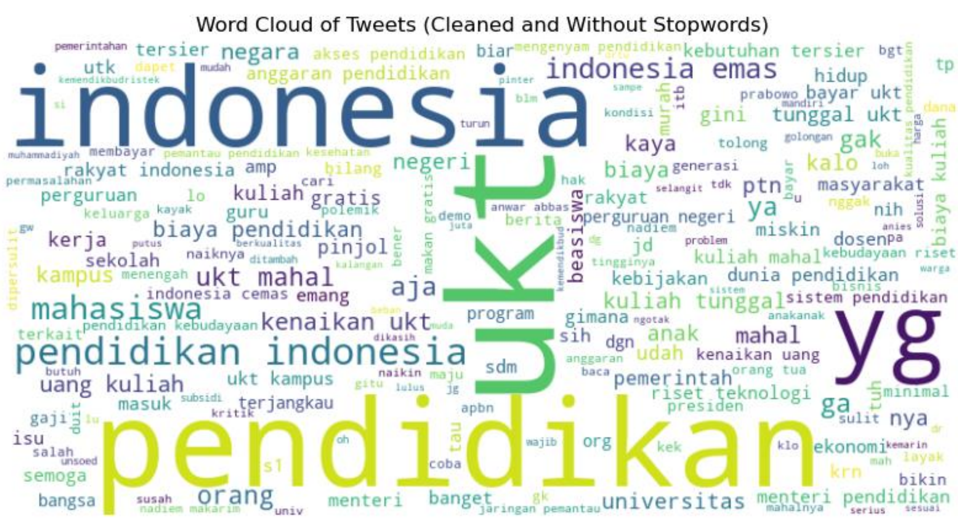

Figure 2 **Word Cloud of Tweets**

Notably, words like "biaya" (cost), "tinggi" (high), and "kenaikan" (increase) suggest a sentiment of financial strain and frustration, with discussions centered on tuition hikes and economic barriers to accessing education. The presence of "gratis" (free) and "program" indicates conversations around government initiatives aimed at reducing education costs, such as scholarships and subsidies. This contrast between financial burden and policy support highlights the public's mixed sentiments toward education reforms. Furthermore, terms like "kebijakan" (policy), "menteri" (minister), and "sdm" (human resources) reveal an engagement with broader governance issues, reflecting the public's awareness of education as a driver for national development and its reliance on effective policymaking.

The recurring appearance of "kampus" (campus), "universitas" (university), and "negeri" (state institutions) underscores the focus on public universities and the perceived gap in quality and affordability between public and private education systems. Additionally, informal terms like "yg" (short for "yang" or "which") and "ga" (informal "no") reflect the conversational tone of tweets, characteristic of social media discourse. The word cloud, therefore, not only highlights key topics but also captures the intersection of financial concerns, policy expectations, and societal pressures that shape public opinion on education in Indonesia. This visualization serves as a qualitative complement to the quantitative analysis, offering policymakers a clearer understanding of prevailing themes in the public discourse.

Result and Discussion

Sentiment Distribution

The analysis of public sentiment surrounding education in Indonesia revealed a striking distribution of emotional responses, as captured through Twitter data. By categorizing tweets into three sentiment classes—positive, neutral, and negative—patterns emerged that provided a lens into how the public perceives and engages with educational issues. The sentiment distribution, visualized through a bar chart (Figure 3), offered both a quantitative overview and deeper insights into the underlying dynamics of public discourse.

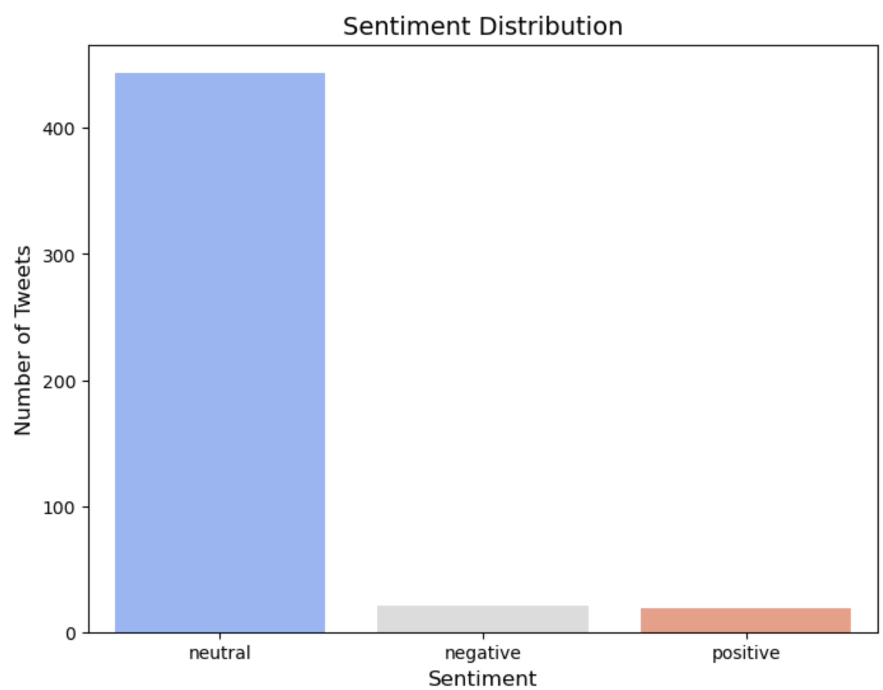

Figure 3 Sentiment Distribution

Neutral sentiments dominated the dataset, accounting for the overwhelming majority of tweets analyzed. These tweets often conveyed information, observations, or questions without an overt emotional tone, reflecting a preference for neutral communication in public discourse about education. For instance, tweets discussing curriculum updates or examination schedules typically fell into this category, as they served an informational purpose rather than expressing approval or dissatisfaction. This trend underscores the role of social media as a platform for disseminating facts and updates, particularly in domains like education, where clarity and neutrality are often prioritized.

The positive sentiments, while comparatively fewer in number, highlighted public approval or optimism about certain educational developments. Tweets classified as positive frequently praised policy reforms, the achievements of students, or government initiatives aimed at improving access to education. For example, discussions around increased funding for rural schools or the implementation of digital learning tools often elicited supportive reactions. Although limited in volume, these positive expressions are critical indicators of public alignment with specific strategies, offering policymakers valuable feedback on the initiatives that resonate most effectively with the population.

Negative sentiments, though also a minority, were significant for their depth and specificity. These tweets often critiqued perceived inefficiencies, such as disparities in educational quality between urban and rural areas or frustrations with rigid standardized testing. Such expressions of dissatisfaction serve as a barometer for identifying areas in need of urgent attention. While the proportion of negative tweets was small compared to neutral sentiments, their qualitative content provided rich insights into the systemic challenges faced by educators, students, and parents alike. These voices of critique are particularly valuable in shaping targeted reforms.

The distribution of sentiments—predominantly neutral, followed by positive and negative—reflects a complex tapestry of public engagement with education in Indonesia. Neutral tweets, while numerically dominant, may mask latent opinions, emphasizing the need for nuanced analysis to uncover underlying attitudes. Meanwhile, the positive and negative sentiments, though fewer, highlight the extremes of public opinion, offering critical feedback loops for educational stakeholders. The visualization of this distribution, as shown in the bar chart, underscores the diverse ways in which the public interacts with educational discourse on Twitter. These findings provide a foundation for informed policy-making and underscore the value of sentiment analysis in capturing the multifaceted nature of public opinion.

Sentiment Trends Over Time

The temporal analysis of sentiment trends revealed significant fluctuations over the six-month period, reflecting public responses to key developments in Indonesia's education sector. By plotting the frequency of positive, neutral, and negative sentiments across months (Figure 4), clear patterns emerged that were often aligned with notable policy announcements or events. This time-based approach provided a dynamic perspective on how public opinion evolved, emphasizing the interplay between sentiment and real-world events.

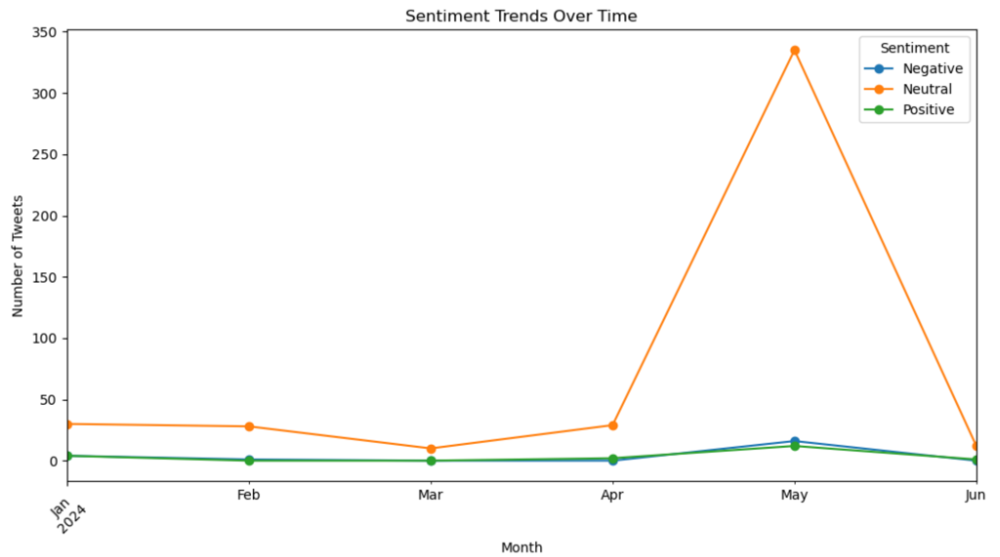

Figure 4 **Sentiment Trends Over Time**

Neutral sentiments maintained a steady presence throughout the study period, acting as a foundation for public discourse. In May 2024, a notable increase occurred, with neutral tweets exceeding 300 instances. This surge aligned with the government's announcement of reforms to the "Merdeka Belajar" (Emancipated Learning) initiative, a policy shift that attracted significant public attention and prompted widespread discussion. During this month, neutral tweets predominantly featured debates and clarifications about the policy, highlighting Twitter's role as a platform for disseminating and seeking clarity on educational reforms. The general consistency of neutral sentiments, aside from these peaks, underscores their function as the backbone of educational discussions on social media.

Positive sentiments remained relatively low but showed slight increases during certain months. In March 2024, the Indonesian government announced a significant increase in the education budget, allocating approximately IDR 660.8 trillion (20% of the state budget) to enhance educational infrastructure and quality, with a focus on rural areas. This substantial investment aimed to improve teacher competencies, distribute educational resources more equitably, and upgrade facilities in underdeveloped regions. The policy was part of a broader strategy to prepare Indonesia's human resources for future challenges, including technological disruptions and global competitiveness. The government's commitment to education was further reflected in the 2024 National Indicative Budget, which saw an almost 20% increase, amounting to IDR 660 trillion, up from IDR 552 trillion in 2023. This budget allocation was directed towards improving infrastructure and the quality of schools, increasing scholarships, raising the quality of higher education institutions, and strengthening research and innovation. These initiatives underscored the government's dedication to building superior, innovative, and competitive human resources through substantial investment in the education sector.

Negative sentiments, while representing a minority, indicated critical moments of dissatisfaction. In May 2024, the Indonesian government's proposal to increase university tuition fees sparked significant public outcry and student protests. The policy, which aimed to raise tuition fees for new students, led to concerns about the affordability of higher education and its potential impact on access for students from lower-income backgrounds. In response to the widespread criticism, the Ministry of Education and Culture, under Minister Nadiem Makarim, announced the cancellation of the tuition hike on May 27, 2024, following a meeting with President Joko Widodo. Minister Makarim stated, "The Ministry of Education and Culture has made a decision to cancel the UKT increase this year." This decision was made after considering public concerns over the affordability of higher education and discussions with university rectors. Additionally, in early 2024, the Bandung Institute of Technology (ITB) faced criticism for encouraging students to pay tuition using loan schemes managed by third-party financial institutions. This initiative led to student protests, with demonstrators arguing that such schemes could burden students with debt and limit access to education. The Ministry of Education, Culture, Research, and Technology intervened, urging ITB to seek alternative funding solutions that would not impose financial hardships on students. These events highlight the challenges and sensitivities surrounding education funding policies in Indonesia. The government's responsiveness to public sentiment, as demonstrated by the reversal of the tuition fee increase, underscores the importance of considering stakeholder perspectives in policy formulation. The incidents also reflect broader concerns about equitable access to education and the financial pressures faced by students in the country.

The public's reaction to these policies, as evidenced by the surge in negative sentiments on social media during this period, illustrates the critical role of public opinion in shaping education policy in Indonesia. It also emphasizes the need for policymakers to engage in transparent and inclusive dialogues with stakeholders to ensure that education reforms are both effective and socially equitable. In summary, the analysis of sentiment trends over time reveals that neutral sentiments dominated the discourse, with significant fluctuations corresponding to key policy announcements. Positive sentiments, though less prevalent, increased in response to policies aimed at enhancing educational equity. Negative sentiments surfaced during debates over policy reforms, reflecting public concerns about the implications of such changes. These findings underscore the dynamic nature of public sentiment in response to educational policies in Indonesia and highlight the importance of social media platforms like Twitter in capturing and reflecting these sentiments.

The sentiment trends over time illustrated a dynamic and multifaceted public engagement with education-related topics on Twitter. The interplay between surges in sentiment and policy announcements underscores the platform's utility as a real-time feedback mechanism for stakeholders. While neutral sentiments predominated, the temporal shifts in positive and negative sentiments revealed moments of public optimism and concern, respectively. This temporal analysis not only captured the pulse of public opinion but also underscored the importance of correlating sentiment dynamics with real-world events to derive actionable insights. These findings highlight the potential for using time-based sentiment trends as a tool for policymakers to monitor and respond to public sentiment effectively.

Comparison with Previous Studies

The findings of this study align with and, in some cases, diverge from existing literature on sentiment analysis and public discourse in education. Similar to the work by Atsqalani et al., which analyzed sentiments surrounding the 2013 curriculum, this research found that neutral sentiments dominated public discussions, reflecting the informational nature of much of the content shared on Twitter [39]. However, while Atsqalani's work focused primarily on curriculum debates, this study expanded its scope to include broader educational themes, capturing diverse sentiments about policies, equity, and institutional challenges.

In contrast to Pratama's research on educational sentiment during the pandemic, which highlighted a more polarized distribution of positive and negative sentiments due to the contentious nature of online learning, this study observed a relatively lower prevalence of polarized opinions [44]. The disparity can be attributed to the topics under discussion; whereas pandemic-related discourse often provoked strong emotional responses, the themes analyzed here, such as funding reforms and testing policies, elicited more neutral reactions. This divergence underscores the contextual nuances of public sentiment, shaped by the nature and immediacy of the issues at hand.

Interestingly, the temporal trends in sentiment identified in this research resonate with findings from Das et al., who demonstrated that real-time analysis is particularly effective in detecting sentiment spikes tied to major events or announcements [42]. The pronounced spike in neutral sentiments during May 2024, following the announcement of standardized testing reforms, parallels Das's observations, reinforcing the argument that public sentiment on Twitter can serve as an early indicator of societal engagement with policy changes. This study thus extends the applicability of real-time sentiment analysis to the Indonesian education sector, providing a framework for tracking public reactions over time.

Implications for Policy and Public Discourse

The implications of these findings for policymakers and educators in Indonesia are both profound and multifaceted. The predominance of neutral sentiments indicates that Twitter serves as a critical space for disseminating information and clarifying educational policies. Policymakers can leverage this platform to enhance transparency and communication, ensuring that key stakeholders, including students, parents, and educators, are well-informed about policy objectives and implementation strategies. This is particularly crucial during periods of significant reform, as evidenced by the neutral sentiment spike in May 2024.

The insights into positive sentiments provide a roadmap for amplifying public approval of educational initiatives. Policies such as increased funding for rural schools, which garnered optimistic reactions, demonstrate that targeted interventions addressing systemic inequities resonate strongly with the public. Policymakers can build on these findings by prioritizing equity-focused strategies, ensuring that such efforts are prominently communicated to foster greater trust and support among stakeholders.

Conversely, the analysis of negative sentiments highlights critical areas for improvement. Public dissatisfaction with standardized testing reforms, as reflected in the May 2024 sentiment spike, underscores the need for policymakers to engage in more inclusive dialogues with affected communities. By addressing these concerns through participatory decision-making processes, educators and policymakers can mitigate resistance and build consensus around contentious issues. These findings emphasize the importance of listening to critical voices on social media, which often highlight overlooked challenges and offer valuable perspectives for refining policy interventions.

At a broader level, this research underscores the transformative potential of sentiment analysis as a tool for guiding educational reforms. By providing a real-time understanding of public opinion, sentiment analysis enables policymakers to anticipate and address societal reactions, ensuring that reforms are not only effective but also widely accepted. Furthermore, these insights contribute to a growing body of literature that positions social media as a powerful feedback mechanism, bridging the gap between top-down policy frameworks and grassroots public discourse. This dynamic interplay between sentiment and action represents a promising avenue for fostering more inclusive and responsive education systems in Indonesia.

Conclusion

This study revealed insightful trends in public sentiment on education in Indonesia, derived from analyzing Twitter data spanning January to June 2024. Neutral sentiments dominated the discourse, highlighting the informational nature of many tweets and the use of Twitter as a platform for disseminating updates and clarifying policy changes. Positive sentiments were associated with initiatives like increased rural school funding, reflecting public approval of equity-focused reforms, while negative sentiments were tied to contentious topics such as standardized testing reforms, indicating areas of dissatisfaction. Temporal trends, including sentiment spikes correlating with policy announcements, emphasized the dynamic and event-driven nature of public discourse. These findings provide a nuanced understanding of how Indonesians engage with educational topics online, offering valuable feedback loops for policymakers and educators.

This research contributes significantly to the burgeoning field of digital society, particularly in understanding the intersection of social media and public discourse on societal issues. By focusing on Twitter, a platform with unparalleled immediacy and reach, the study underscores the role of social media as a real-time barometer for public opinion. Education, a critical pillar of societal development, emerges as a domain where digital conversations mirror broader societal aspirations and concerns. The methodology and findings of this study demonstrate how sentiment analysis can bridge the gap between grassroots opinion and top-down policy-making, enriching the discourse around digital engagement in a rapidly evolving social landscape. This alignment with the principles of a digitally connected society highlights the potential of leveraging technology to foster inclusivity and responsiveness in governance.

While the study provides robust insights, certain limitations must be acknowledged. The dataset, derived exclusively from Twitter, may not comprehensively represent public opinion, particularly from demographics less active on the platform. The reliance on English and Indonesian stopwords for text preprocessing might have excluded certain nuanced expressions, potentially impacting sentiment classification. Additionally, the dominance of neutral sentiments, while reflecting informational tweets, could mask latent opinions or ambivalence that were not fully captured by the classification model. These factors underscore the inherent challenges in analyzing social media data, where informal language, abbreviations, and cultural context can complicate interpretation.

Building on this study, future research could explore sentiment dynamics across multiple social media platforms, such as Instagram or Facebook, to capture a more diverse spectrum of public opinion. Extending the analysis to longer timeframes could provide deeper insights into how public sentiment evolves in response to sustained educational reforms or societal changes. Moreover, incorporating advanced natural language processing techniques, such as deep learning models, could enhance the accuracy of sentiment classification, especially for informal and ambiguous language. Finally, exploring thematic sentiment analysis focused on specific issues, such as digital literacy or teacher training, could provide more granular insights into public discourse on education in Indonesia. These avenues hold promise for further advancing the understanding of social media's role in shaping public opinion within the digital society.

Declarations

Author Contributions

Conceptualization: B.H.H.; Methodology: I.M.; Software: B.H.H.; Validation: I.M.; Formal Analysis: I.M.; Investigation: B.H.H.; Resources: I.M.; Data Curation: I.M.; Writing Original Draft Preparation: I.M.; Writing Review and Editing: B.H.H.; Visualization: B.H.H.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. C. McGregor, "Social Media as Public Opinion: How Journalists Use Social Media to Represent Public Opinion," *Journalism*, vol. 20, no. 8, pp. 1070–1086, 2019, doi: 10.1177/1464884919845458.

[2] S. Rafique, "Empowering ESL Students: Leveraging Social Media Blogs for Effective Opinion Writing Skills Development," *JPR*, vol. 9, no. 4, pp. 103–114, 2023, doi: 10.61506/02.00132.

[3] H. Chen, "Spiral of Silence on Social Media and the Moderating Role of Disagreement and Publicness in the Network: Analyzing Expressive and Withdrawal Behaviors," *New Media Soc.*, vol. 20, no. 10, pp. 3917–3936, 2018, doi: 10.1177/1461444818763384.

[4] R. S. Daud, “The Role of Political Communication in Shaping Public Opinion: A Comparative Analysis of Traditional and Digital Media,” *J. Public Represent. Soc. Provis.*, vol. 1, no. 2, pp. 63–69, 2021, doi: 10.55885/jprsp.v1i2.241.

[5] M. Xiao, “Research on the Orientation of Public Opinion by New Media: Taking the Jiang Ge Case as an Example,” *Commun. Humanit. Res.*, vol. 13, no. 1, pp. 250–255, 2023, doi: 10.54254/2753-7064/13/20230346.

[6] S. Mappiasse, “Similar Trends, Different Agendas: Religious Instruction in Public Schools in Indonesia, Malaysia and Singapore,” *J. Islam. Educ. Policy*, vol. 1, no. 1, 2017, doi: 10.30984/j.v1i1.403.

[7] I. U. Sumaryanti and J. Yuniar, “The Implication of Social Media Toward College Students’ Online Behavior in Bandung,” *Mediat. J. Komun.*, vol. 15, no. 1, pp. 55–64, 2022, doi: 10.29313/mediator.v15i1.8845.

[8] F. Mansir, “The Role of the Principal for Educational Reform in Indonesia,” *-Turats*, vol. 16, no. 1, pp. 13–21, 2022, doi: 10.24260/at-turats.v16i1.2205.

[9] D. Suryadarma, “How Corruption Diminishes the Effectiveness of Public Spending on Education in Indonesia,” *Bull. Indones. Econ. Stud.*, vol. 48, no. 1, pp. 85–100, 2012, doi: 10.1080/00074918.2012.654485.

[10] T. Sauter and A. Bruns, “Tweeting the TV Event, Creating ‘Public Sphericules’: AD HOC Engagement With SBS’s Go Back to Where You Came From — Season 2,” *Media Int. Aust.*, vol. 152, no. 1, pp. 5–15, 2014, doi: 10.1177/1329878x1415200103.

[11] N. Y. Lee, Y. Kim, and J. Kim, “Tweeting Public Affairs or Personal Affairs? Journalists’ Tweets, Interactivity, and Ideology,” *Journalism*, vol. 17, no. 7, pp. 845–864, 2016, doi: 10.1177/1464884915585954.

[12] A. W. Prakasita, “The Use of Social Media (Twitter) in the Winning of East Java Province Election in 2018,” *Indones. J. Polit. Stud. Ijps*, vol. 2, no. 2, pp. 110–130, 2022, doi: 10.15642/ijps.2022.2.2.110-130.

[13] K. Bista, “Is Twitter a Pedagogical Tool in Higher Education? Perspectives of Education Graduate Students,” *J. Scholarsh. Teach. Learn.*, pp. 83–102, 2015, doi: 10.14434/josotl.v15i2.12825.

[14] E. Miller, J. L. Jolly, J. N. Latz, and K. Listman, “Influencers and Major Themes in a Gifted Education Community of Practice on Twitter,” *J. Adv. Acad.*, vol. 33, no. 3, pp. 469–504, 2022, doi: 10.1177/1932202x221099590.

[15] M. Liebenlito, N. Inayah, E. Choerunnisa, T. E. Sutanto, and S. Inna, “Active learning on Indonesian Twitter sentiment analysis using uncertainty sampling,” *J. Appl. Data Sci.*, vol. 5, no. 1, Art. no. 1, Jan. 2024, doi: 10.47738/jads.v5i1.144.

[16] J. R. Batmetan and T. Hariguna, “Sentiment Unleashed: Electric Vehicle Incentives Under the Lens of Support Vector Machine and TF-IDF Analysis,” *J. Appl. Data Sci.*, vol. 5, no. 1, Art. no. 1, Jan. 2024, doi: 10.47738/jads.v5i1.162.

[17] Henderi and Q. Siddique, “Comparative Analysis of Sentiment Classification Techniques on Flipkart Product Reviews: A Study Using Logistic Regression, SVC, Random Forest, and Gradient Boosting,” *J. Digit. Mark. Digit. Curr.*, vol. 1, no. 1, Art. no. 1, May 2024, doi: 10.47738/jdmdc.v1i1.4.

[18] B. Berlilana, A. M. Wahid, D. Fortuna, A. N. A. Saputra, and G. Bagaskoro, “Exploring the Impact of Discount Strategies on Consumer Ratings: An Analytical Study of Amazon Product Reviews,” *J. Appl. Data Sci.*, vol. 5, no. 1, Art. no. 1, Jan. 2024, doi: 10.47738/jads.v5i1.163.

[19] T. Wahyuningsih and S. C. Chen, “Analyzing Sentiment Trends and Patterns in Bitcoin-Related Tweets Using TF-IDF Vectorization and K-Means Clustering,” *J. Curr. Res. Blockchain*, vol. 1, no. 1, Art. no. 1, Jun. 2024, doi: 10.47738/jcrb.v1i1.11.

[20] Hery and A. E. Widjaja, “Predictive Modeling of Blockchain Stability Using Machine Learning to Enhance Network Resilience,” *J. Curr. Res. Blockchain*, vol. 1, no. 2, Art. no. 2, Sep. 2024, doi: 10.47738/jcrb.v1i2.15.

[21] S. Yadav and A. R. Hananto, “Comprehensive Analysis of Twitter Conversations Provides Insights into Dynamic Metaverse Discourse Trends,” *Int. J. Res. Metaverese*, vol. 1, no. 1, Art. no. 1, Jun. 2024, doi: 10.47738/ijrm.v1i1.2.

[22] D. Sugianto and A. R. Hananto, “Geospatial Analysis of Virtual Property Prices Distributions and Clustering,” *Int. J. Res. Metaverese*, vol. 1, no. 2, Art. no. 2, Sep. 2024, doi: 10.47738/ijrm.v1i2.10.
[23] M. Irfan, “Comparing Random Forest and Logistic Regression for Predicting Student Completion in Online University Courses Using Behavioral Data,” *Artif. Intell. Learn.*, vol. 1, no. 1, Art. no. 1, Nov. 2024.
[24] Y. Durachman and A. W. B. A. Rahman, “Clustering Student Behavioral Patterns: A Data Mining Approach Using K-Means for Analyzing Study Hours, Attendance, and Tutoring Sessions in Educational Achievement,” *Artif. Intell. Learn.*, vol. 1, no. 1, Art. no. 1, Nov. 2024.
[25] J. Xue, J. Chen, C. Chen, C. Zheng, S. Li, and T. Zhu, “Public Discourse and Sentiment During the COVID 19 Pandemic: Using Latent Dirichlet Allocation for Topic Modeling on Twitter,” *Plos One*, vol. 15, no. 9, p. e0239441, 2020, doi: 10.1371/journal.pone.0239441.
[26] Z. E. Aydin, Z. K. Öztürk, and Z. Í. E. Cicek, “Turkish Sentiment Analysis for Open and Distance Education Systems,” *Turk. Online J. Distance Educ.*, pp. 124–138, 2021, doi: 10.17718/tojde.961825.
[27] S. Boon-itt and Y. Skunkan, “Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study,” *Jmir Public Health Surveill.*, vol. 6, no. 4, p. e21978, 2020, doi: 10.2196/21978.
[28] T. Fütterer, “ChatGPT in Education: Global Reactions to AI Innovations,” *Sci. Rep.*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-42227-6.
[29] Ö. Özyurt and N. Kisa, “Sentiment Analysis of Tweets Relating to Distance Education During the Covid-19 Pandemic,” *J. Comput. Educ. Res.*, vol. 9, no. 18, pp. 853–868, 2021, doi: 10.18009/jcer.950790.
[30] M. Trupthi, S. Pabboju, and G. Narsimha, “Possibilistic Fuzzy C-Means Topic Modelling for Twitter Sentiment Analysis,” *Int. J. Intell. Eng. Syst.*, vol. 11, no. 3, pp. 100–108, 2018, doi: 10.22266/ijies2018.0630.11.
[31] W. A. Akbari, T. Tukino, B. Huda, and M. Muslih, “Sentiment Analysis of Twitter User Opinions Related to Metaverse Technology Using Lexicon Based Method,” *Sinkron*, vol. 8, no. 1, pp. 195–201, 2023, doi: 10.33395/sinkron.v8i1.11992.
[32] M. Irfan, M. A. Fauzi, T. Tibyani, and N. D. Mentari, “Twitter Sentiment Analysis on 2013 Curriculum Using Ensemble Features and K-Nearest Neighbor,” *Int. J. Electr. Comput. Eng. Ijece*, vol. 8, no. 6, p. 5409, 2018, doi: 10.11591/ijece.v8i6.pp5409-5414.
[33] R. Umar, “Comparing the Performance of Data Mining Algorithms in Predicting Sentiments on Twitter,” *J. Resti Rekayasa Sist. Dan Teknol. Inf.*, vol. 7, no. 4, pp. 817–823, 2023, doi: 10.29207/resti.v7i4.4931.
[34] A. Rausanfita, “Sentiment Analysis of Comments on Instagram Posts of Indonesia’s 2024 Presidential Candidates Using the Support Vector Machine Method,” *Tech-E*, vol. 7, no. 1, pp. 21–31, 2023, doi: 10.31253/te.v7i1.2289.
[35] A. A. Lutfi, A. E. Permanasari, and S. Fauziati, “Sentiment Analysis in the Sales Review of Indonesian Marketplace by Utilizing Support Vector Machine,” *J. Inf. Syst. Eng. Bus. Intell.*, vol. 4, no. 1, p. 57, 2018, doi: 10.20473/jisebi.4.1.57-64.
[36] O. S. Simbolon, “Sentiment Analysis of Mypertamina Application Using Support Vector Machine and Naïve Bayes Algorithms,” *J. Sist. Inf. Dan Ilmu Komput. Primajusikom Prima*, vol. 7, no. 1, pp. 177–189, 2023, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v7i1.4078.
[37] N. Rezki, S. A. Thamrin, and S. Siswanto, “Sentiment Analysis of Merdeka Belajar Kampus Merdeka Policy Using Support Vector Machine With Word2vec,” *Barekeng J. Ilmu Mat. Dan Terap.*, vol. 17, no. 1, pp. 0481–0486, 2023, doi: 10.30598/barekengvol17iss1pp0481-0486.
[38] U. I. Larasati, M. A. Muslim, R. Arifudin, and A. Alamsyah, “Improve the Accuracy of Support Vector Machine Using Chi Square Statistic and Term Frequency Inverse Document Frequency on Movie Review Sentiment Analysis,” *Sci. J. Inform.*, vol. 6, no. 1, pp. 138–149, 2019, doi: 10.15294/sji.v6i1.14244.
[39] H. Atsqalani, N. Hayatin, and C. S. K. Aditya, “Sentiment Analysis From Indonesian

Twitter Data Using Support Vector Machine and Query Expansion Ranking," *J. Online Inform.*, vol. 7, no. 1, pp. 116–122, 2022, doi: 10.15575/join.v7i1.669.
[40] E. Sutoyo and A. Almaarif, "Twitter Sentiment Analysis of the Relocation of Indonesia's Capital City," *Bull. Electr. Eng. Inform.*, vol. 9, no. 4, pp. 1620–1630, 2020, doi: 10.11591/eei.v9i4.2352.
[41] Christian Lilik Henri Setiawan Nivak, "Sentiment Analysis of Ten Tourism Destinations in Indonesia," *J. Indones. Tour. Dev. Stud.*, vol. 9, no. 3, pp. 162–168, 2021, doi: 10.21776/ub.jitode.2021.009.03.03.
[42] S. Das, R. K. Behera, M. Kumar, and S. K. Rath, "Real-Time Sentiment Analysis of Twitter Streaming Data for Stock Prediction," *Procedia Comput. Sci.*, vol. 132, pp. 956–964, 2018, doi: 10.1016/j.procs.2018.05.111.
[43] M. Mujahid *et al.*, "Sentiment Analysis and Topic Modeling on Tweets About Online Education During COVID-19," *Appl. Sci.*, vol. 11, no. 18, p. 8438, 2021, doi: 10.3390/app11188438.
[44] A. R. Pratama, "Sentiment Analysis of Facebook Posts Through Special Reactions: The Case of Learning From Home in Indonesia Amid COVID-19," *J. Ilm. Tek. Elektro Komput. Dan Inform.*, vol. 8, no. 1, p. 83, 2022, doi: 10.26555/jiteki.v8i1.23615.