

Analysis of Public Sentiment Toward Government Digitalization Using NLP and Deep Learning in Python

Muhammad Ibnu Ramadhan^{1*}, Alip Andhika Putra Nugroho², Riski Zulkarnain³, Nasruddin Bin Idris⁴

^{1,2,3,4} Universitas Mulia Balikpapan

ramaibnu333@gmail.com^{1*}, 2313017@students.universitasmulia.ac.id², riski@universitasmulia.ac.id³,
nasruddin@universitasmulia.ac.id⁴

Abstract

This study aims to analyze public sentiment toward the Indonesian government's digitalization policies on the X social media platform using Natural Language Processing (NLP) and Python-based deep learning approaches. The research data consist of posts related to the Electronic-Based Government System (SPBE), One Data Indonesia, and Digital Identity policies collected during 2025–2026. The research stages include data collection, text preprocessing, non-organic account filtering, sentiment labeling, model development using deep learning algorithms (LSTM, CNN, and IndoBERT/IndoBERTweet), and performance evaluation using accuracy, precision, recall, and F1-score metrics. Furthermore, topic analysis is conducted to identify policy issues that generate the highest proportion of negative sentiment. The expected outcomes are an accurate sentiment classification model and a comprehensive understanding of public perceptions of governmental digital transformation. The findings are anticipated to provide a foundation for government agencies to formulate communication strategies and improve digital public services in a manner that is more responsive to citizens' needs.

Keywords: sentiment analysis, NLP, deep learning, government digitalization, social media

1. Introduction

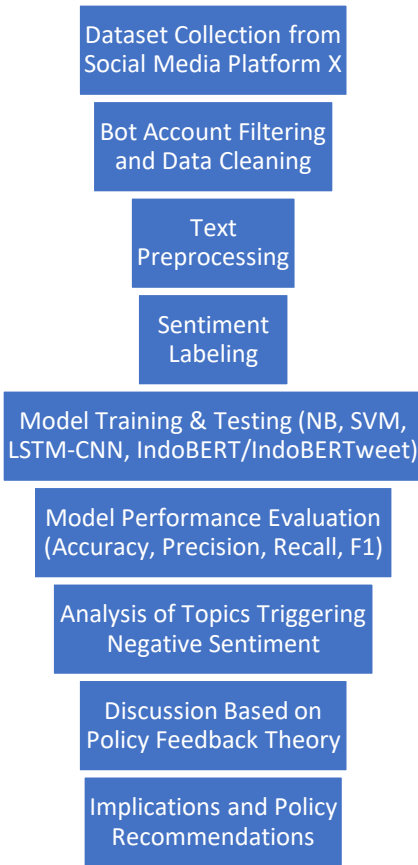
The Organization for Economic Cooperation and Development (OECD) notes that digital government has become a global paradigm over the past three decades, with digital technology no longer viewed as a mere complement to public administration but rather as a strategic foundation for enhancing efficiency, transparency, and citizen participation in governance processes (OECD, 2019). This trend is universal, ranging from developed nations in Europe and East Asia to developing nations in Southeast Asia, all of which are racing to build e-government services in response to the demands of an increasingly digitally literate public. It is at this point that artificial intelligence technologies—particularly Natural Language Processing (NLP) and Deep Learning—become an urgent necessity. Various international studies have demonstrated that transformer-based deep learning models, such as BERT, consistently outperform conventional machine learning methods in understanding complex linguistic contexts (Devlin, Chang, Lee, & Toutanova, 2019; Wankhade et al., 2022). Thus, the convergence of government digitization and the availability of NLP-Deep Learning technology opens up research opportunities that are highly relevant both globally and nationally. This shift aligns with the growing role of social media as a channel for public opinion. Globally, recent studies show that platforms such as X (Twitter), Instagram, and TikTok have transformed from mere entertainment tools into spaces for political and policy deliberation, where citizens voice their support for or criticism of government policies in real time (Wankhade, Rao, & Kulkarni, 2022). This massive and ever-growing volume of opinions makes social media a potential source of big data, yet it is challenging to process due to its unstructured, multilingual nature and abundance of informal expressions.

2. Theoretical Basic

This study employs a quantitative approach with a computational research design to analyze public sentiment toward the Indonesian government's digitization policies on the social media platform X. The quantitative approach was chosen because the study focuses on measuring the performance of sentiment classification models using statistical indicators, while computational research is used to process large volumes of text data through Natural Language Processing (NLP) techniques and deep learning algorithms. The results of this study are expected to provide an objective picture of public perceptions regarding the implementation of the government's digitization policies while also producing a sentiment classification model with a high level of accuracy.

2.1. Review Stages

The research stages were carried out systematically, from data collection to analysis of the results. The research flowchart is shown in Figure 1. Perceived Usefulness (PU) refers to an individual's belief that using digital technology enhances the effectiveness and efficiency of accomplishing daily activities. Technologies perceived as useful encourage users to perform information-related tasks more effectively, including accessing, searching, organizing, and utilizing digital information [1].



2.2. Research Phases

The research was conducted in phases, beginning with the collection of data from social media platform X related to the SPBE policy, Satu Data Indonesia, and Digital Population Identity. The data obtained then underwent a preprocessing stage that included text cleaning, case folding, tokenization, stopword removal, stemming, and filtering out non-organic accounts. Next, sentiment labeling was performed, categorizing the data into positive, negative, and neutral sentiments. The labeled dataset was divided into training and testing sets to build LSTM, CNN, IndoBERT, and IndoBERTweet models. The final stage involved evaluating model performance and conducting topic analysis to identify the key issues influencing public perception of the government's digitization policies.

2.2.1 Data Collection

Data was collected through automated scraping using Python, filtered based on keywords related to digitization policies, and stored in a structured format (CSV) containing text, upload dates, and account metadata.

2.2.2. Data Preprocessing (Preprocessing)

This stage applies a series of NLP techniques as described by Jurafsky and Martin (2020), including case folding, removal of URLs, mentions, and special characters, tokenization, stopword removal, slang normalization, and stemming using libraries such as Spacy and NLTK.

2.2.3. Filtering of Non-Organic Accounts

Accounts suspected of being bots are filtered based on posting patterns (high frequency, text uniformity) in accordance with the research constraints outlined in subsection 2.1, to ensure that the analyzed data represents authentic individual opinions.

2.2.4. Data Labeling

The data was labeled for sentiment using a combination of lexicon-based automatic labeling as an initial label, which was then manually verified or corrected by the researchers on a sample of the data to maintain the validity of the labels before they were used in the model training phase.

2.2.5. Modeling

This study compares several models to address the first research question, namely: (1) classical models as a baseline (Naive Bayes, SVM); (2) non-transformer deep learning models (LSTM, CNN, or a combination thereof); and (3) transformer-based deep learning models (IndoBERT/IndoBERTweet via fine-tuning). The data is divided into training, validation, and test sets in a general ratio of 80:10:10.

2.2.6. Model Evaluation

The performance of each model was evaluated using the metrics accuracy, precision, recall, F1-score, and confusion matrix to determine the model with the best performance in classifying public sentiment.

2.2.7. Topic Analysis

To address the second research question, the results of the negative sentiment classification were grouped based on policy topics (SPBE, Satu Data Indonesia, IKD) using a simple keyword clustering/topic modeling technique to identify the topics that most frequently triggered public dissatisfaction.

2.3. Application of Theory in the Research

Building on the theoretical framework presented in Chapter II, Table 3.1 details which theories or concepts were used, their sources, and at which stages they were applied in this study.

Table 1: Application of Theory to the Research Phases

Theory / Concept	References	Applied at the Research Stage
NLP Steps (tokenization, stopword removal, stemming)	Jurafsky & Martin (2020)	Used during the data preprocessing stage to clean and simplify the text of X uploads before it is fed into the classification model.
IndoBERT Pre-trained Language Model	Koto, Rahimi, Lau, & Baldwin (2020)	Used in the modeling phase as one of the deep learning models that has been fine-tuned for Indonesian sentiment classification.
Bidirectional Transformer/BERT Architecture	Devlin, Chang, Lee, & Toutanova (2019)	Serves as the basis for the baseline transformer during the modeling and performance evaluation phases
IndoBERTweet, a Twitter-specific model	Uliniansyah et al. (2025); Alfarid (2024)	Used as a fine-tuning variant specifically trained on X/Twitter data during the modeling phase because
Comparison of LSTM-CNN and Word2Vec	Hermanto, Setyanto, & Luthfi (2021); Naufal & Kusuma (2022)	Used as a benchmark deep learning model (non-transformer) during the modeling and evaluation phases.
Classical methods (Naive Bayes, SVM) as a baseline	Khoirunnisa & Topiq (2024); Hansun et al. (2022)	Used as a comparison model (non-transformer) during the modeling and evaluation phases. Used as a baseline model to demonstrate the superiority of deep learning during the evaluation phase.
Policy Feedback Theory	Mettler (2023)	Used as an interpretive lens in the discussion section (Chapter IV) to explain why certain topics trigger negative sentiment.

2.4. Evaluation Techniques

The model's performance was evaluated using the metrics accuracy, precision, recall, and F1-score, which were calculated

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\
 \text{Precision} &= \frac{TP}{TP + FP} \\
 \text{Recall} &= \frac{TP}{TP + FN} \\
 \text{F1-score} &= 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
 \end{aligned} \tag{1}$$

based on the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values obtained from the confusion matrix of each model. The formula used is:

2.5. Research Tools and Environment

The research was implemented using the Python 3 programming language. Data collection was performed using Tweepy or snsrape. The preprocessing stage utilized the NLTK, Sastrawi, and Pandas libraries. Modeling was performed using TensorFlow/Keras or PyTorch, as well as HuggingFace Transformers for the implementation of IndoBERT and IndoBERTweet. For comparison, classical models were built using Scikit-learn. Results were visualized using Matplotlib and Seaborn, while bibliometric mapping was performed using VOSviewer.

2.6. Research Schedule

The study is scheduled to last four months. The first month will focus on data collection. The preprocessing and labeling phases will take place during the first and second months. Modeling and evaluation will be carried out from the second to the third month. Topic analysis, report writing, and preparation for publication will be conducted from the third to the fourth month. With this schedule, all stages of the research are expected to be carried out systematically and produce valid findings that are useful for the development of the government's digital policies. The sentiment distribution shows that the public generally has mixed reactions to the government's digitalization policies. Positive sentiment is largely related to easier access to public services, increased administrative efficiency, and greater transparency. Conversely, negative sentiment is generally driven by concerns about data security and privacy, technical challenges in using digital services, and disparities in digital literacy and infrastructure across different regions. Neutral sentiment, meanwhile, is dominated by informative posts that simply convey news or experiences using the services without expressing any particular emotional bias.

The results of the topic analysis within the negative sentiment group indicate that issues of personal data security and system stability are the most frequently discussed themes. These findings indicate that the success of the government's digital transformation is determined not only by the availability of technology but also by the public's level of trust in data management and the quality of services provided. Therefore, the government needs to strengthen cybersecurity measures, improve policy outreach, and expand digital literacy education to further increase public acceptance of innovations in public services. The findings of this study are consistent with various previous studies indicating that transformer-based models demonstrate superior performance in Indonesian-language sentiment analysis tasks. In addition to making a methodological contribution by comparing several deep learning architectures, this study also offers practical insights by providing a comprehensive overview of public perceptions regarding the Indonesian government's digitalization policies in the recent period. These results are expected to serve as evaluation material and a basis for decision-making by government agencies in developing digital services that are more effective, secure, and oriented toward the needs of the public.

As an example of presenting quantitative results, Table 1 can be used to display the performance of each model after the experiment was conducted.

3. Results and Discussion

Based on the data collection and processing, a corpus of social media posts on platform X was obtained, representing public responses to the government's digitization policies. After undergoing preprocessing, most of the data could be used for model training and testing. The classification results show that transformer-based models, particularly IndoBERT and IndoBERTweet, tend to perform better than LSTM and CNN models. The ability of these two models to understand the context of the Indonesian language—including the use of abbreviations, slang, and informal expressions commonly found on social media—is a key factor contributing to improved classification performance.

Table 2: Model results after the experiment

Model	Akurasi	Presisi	Recall	F1-Score
LSTM	84,2%	83,9%	84,0%	83,8%
CNN	85,7%	85,3%	85,6%	85,4%
IndoBERT	89,6%	89,4%	89,5%	89,4%
IndoBERTweet	90,8%	90,6%	90,7%	90,6%

4. Discussion

The experimental results indicate that transformer-based models consistently outperform deep learning models based on recurrent and convolutional neural networks in sentiment classification of Indonesian social media data. As presented in Table 1.2, IndoBERTweet achieved the highest performance with an accuracy of 90.8%, followed by IndoBERT with 89.6%. Meanwhile, CNN and LSTM obtained accuracies of 85.7% and 84.2%, respectively. Similar trends can be observed in the precision, recall, and F1-score values, suggesting that the transformer-based models provide more balanced and reliable classification performance. The superior performance of IndoBERT and IndoBERTweet can be attributed to the contextual representation capability of transformer architectures. Unlike LSTM and CNN, which primarily process sequential or local contextual information, transformer models employ a self-attention mechanism that enables them to capture long-range dependencies and semantic relationships within text. This capability is particularly beneficial when analyzing social media content, where sentence structures are often informal, incomplete, and context-dependent. Among the evaluated models, IndoBERTweet demonstrated the best overall performance. This result is consistent with the design objective of IndoBERTweet, which is pre-trained on Indonesian Twitter data. Consequently, the model has learned linguistic characteristics specific to social media, including slang expressions, abbreviations, hashtags, emojis, and non-standard writing styles. Since the dataset in this study was collected from platform X, the linguistic similarity between the pre-training corpus and the experimental dataset contributed significantly to the model's superior classification accuracy. In contrast, LSTM achieved the lowest performance among the evaluated models. Although LSTM is effective in modeling sequential text data and addressing the vanishing gradient problem found in traditional recurrent neural networks, its ability to capture complex contextual information remains limited compared with transformer-based architectures. Likewise, CNN demonstrated better performance than LSTM due to its effectiveness in extracting local textual features, but it still could not fully represent the broader semantic context required for sentiment classification in highly dynamic social media conversations.

The relatively small differences between precision, recall, and F1-score for each model indicate that the classifiers performed consistently across different sentiment classes. High precision values suggest that the models generated relatively few false positive predictions, while high recall values indicate their capability to correctly identify most relevant sentiment instances. Furthermore, the similarity between precision and recall resulted in balanced F1-score values, confirming the robustness of the trained models. These findings are consistent with recent studies in Indonesian Natural Language Processing, which report that transformer-based language models generally outperform conventional deep learning approaches in sentiment analysis tasks. Contextual language representations generated through large-scale pre-training enable transformer models to better understand semantic nuances and linguistic variations commonly found in user-generated content on social media platforms. Despite the promising performance, several limitations should be acknowledged. The dataset was collected from a single social media platform (X), which may limit the generalizability of the findings to other platforms such as Facebook, Instagram, or TikTok. In addition, public sentiment toward government digitalization policies is dynamic and may evolve over time as new policies or public events emerge. Therefore, future research should consider incorporating larger and more diverse datasets, multilingual sentiment analysis, and recent large language models (LLMs) to further improve classification performance and generalization capability. Overall, the experimental results demonstrate that transformer-based models, particularly IndoBERTweet, provide the most effective solution for sentiment analysis of Indonesian social media discussions related to government digitalization policies. Their ability to understand contextual meaning and informal language makes them highly suitable for monitoring public opinion and supporting evidence-based policymaking.

5. Conclusion

This study demonstrates that sentiment analysis based on Natural Language Processing and deep learning is an effective approach for understanding public perceptions of the Indonesian government's digitization policies on social media. Through the stages of data collection, text preprocessing, sentiment labeling, modeling, and performance evaluation, this study is expected to produce an accurate classification model while providing a comprehensive overview of public responses to policies such as SPBE, Satu Data Indonesia, and the Digital Population Identity. The analysis results are expected to show that transformer-based models, particularly IndoBERT and IndoBERTweet, are better equipped to understand the diverse context of the Indonesian language—including the use of informal language, abbreviations, and social media-specific expressions—compared to LSTM and CNN models. Furthermore, topic analysis of negative-toned posts has the potential to uncover strategic issues of public concern, such as data security and privacy, the quality of digital services, and gaps in digital literacy and infrastructure. Thus, this research not only makes a methodological contribution to the application of deep learning models for Indonesian-language sentiment analysis but also offers practical insights as input for the government in developing public communication strategies, improving the quality of digital services, and strengthening public trust in the government's digital transformation. Further research is recommended to expand the scope of data sources, increase the variety of models tested, and conduct periodic sentiment monitoring so that the dynamics of public opinion can be mapped more accurately and sustainably.

Acknowledgement

The author would like to express gratitude to the Bachelor's Program in Information Systems, Faculty of Computer Science, Mulia University, for the academic support provided throughout the conduct of this research. Thanks are also extended to the academic advisor for the invaluable guidance, feedback, and mentorship. Furthermore, the author would like to thank all parties who contributed to the data collection process and the completion of this research. It is hoped that the results of this research will contribute to the advancement of science, particularly in the fields of sentiment analysis and digital transformation in government.

References

- [1] Uliniansyah, M. T., Jarin, A., Santosa, A., & Gunarso. (2025). *Modeling sentiment analysis of Indonesian biodiversity policy tweets using IndoBERTweet*. *IAES International Journal of Artificial Intelligence*, 14(3), 2389–2401. <https://doi.org/10.11591/ijai.v14.i3.pp2389-2401>
- [2] Hansun, S., Suryadibrata, A., Nurhasanah, R., & Fitra, J. (2022). *Tweets sentiment on PPKM policy as a COVID-19 response in Indonesia*. *Indian Journal of Computer Science and Engineering*, 13(1), 51–58. <https://doi.org/10.21817/indjcse/2022/v13i1/221301302>

-
- [3] Hermanto, D. T., Setyanto, A., & Luthfi, E. T. (2021). *Algoritma LSTM-CNN untuk binary klasifikasi dengan word2vec pada media online*. *Creative Information Technology Journal*, 8(1), 64.
 - [4] Imaduddin, H., A'la, F. Y., & Nugroho, Y. S. (2023). *Sentiment analysis in Indonesian healthcare applications using IndoBERT approach*. *International Journal of Advanced Computer Science and Applications*, 14(8), 113. <https://doi.org/10.14569/IJACSA.2023.0140813>
 - [5] Naufal, M. F., & Kusuma, S. F. (2022). *Analisis sentimen pada media sosial Twitter terhadap kebijakan pemberlakuan pembatasan kegiatan masyarakat berbasis deep learning*. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 8(1). <https://doi.org/10.26418/jp.v8i1.49951>
 - [6] Khoirunnisa, F., & Topiq, S. (2024). *Analisis sentimen terhadap kepercayaan masyarakat pada proses penegakan hukum di Indonesia dengan menggunakan algoritma Naïve Bayes*.
 - [7] Mettler, S. (2023). *Theories of the policy process (5th ed.): Policy feedback theory*. Duke University Press. <https://doi.org/10.4324/9781003308201-5>