KAMPUS AKADEMIK PUBLISHING
Jurnal Sains Student Research
Vol.4, No.1 Februari 2026
e-ISSN: 3025-9851; p-ISSN: 3025-986X, Hal 29-40
DOI: https://doi.org/10.61722/jssr.v4i1.7258

Sentiment Analysis of Coretax Tax Application Users Using IndoBERT and Web Scraping on the X (Twitter) Platform

Muhammad Al Faris Syabil
Universitas Islam Negeri Sumatera Utara
Fajar Al-Fahri
Universitas Islam Negeri Sumatera Utara
Alamat: Jl. Lap. Golf No.120, Kp. Tengah, Kec. Pancur Batu, Kabupaten Deli Serdang, Sumatera Utara 20353
Korespondensi penulis: syabilalfaris6@gmail.com

Abstract. *This study analyzes public sentiment toward the Coretax tax system based on user opinions posted on the X (Twitter) platform. The objective is to assess how the public perceives the system's stability, accessibility, and performance during periods of high usage. A quantitative text-based approach was applied using Natural Language Processing (NLP) techniques. Data were collected through web scraping of tweets containing Coretax-related keywords and processed through six preprocessing stages: case folding, cleaning, tokenizing, normalization, stopword removal, and stemming. Sentiment classification was conducted using the IndoBERT model mdhugol/indonesia-bert-sentiment-classification, which categorized tweets into positive, negative, and neutral classes. The results show that 181 tweets expressed positive sentiment, 171 negative sentiment, and 29 neutral sentiment. Negative sentiment predominantly relates to system errors and login difficulties, whereas positive sentiment commonly appears when the system functions normally. These findings demonstrate that system instability remains the primary factor influencing negative perceptions of Coretax and provide useful insights for improving the reliability of digital tax services.*
Keywords: *Coretax; IndoBERT; public perception; sentiment analysis; text preprocessing.*

INTRODUCTION

The digital transformation of Indonesia's taxation system has accelerated significantly in recent years, aiming to enhance administrative efficiency, transparency, and taxpayer accessibility. One of the key milestones of this modernization initiative is Coretax, an integrated digital tax administration platform developed by the Directorate General of Taxes (DGT). Introduced progressively since 2023, Coretax serves as the central backbone of Indonesia's updated tax ecosystem by integrating data processing, reporting mechanisms, verification procedures, and real-time service operations into a unified digital infrastructure. This transformation is designed to streamline public services, reduce administrative burdens, and strengthen taxpayer compliance through improved digital access.

Despite the intended improvements, public responses toward Coretax remain highly varied, especially across social media platforms such as X (Twitter). Many users acknowledge its enhanced accessibility and modernization, yet numerous complaints persist regarding login difficulties, system downtime, unstable performance, and interface-related issues. Social media thus functions as a dynamic public sphere where users openly express real-time opinions, making it an important resource for evaluating public perceptions of digital taxation services.

Received November 30, 2025; Revised Desember 03, 2025;Februari 01, 2026
* Muhammad Al Faris Syabil, syabilalfaris6@gmail.com

To objectively capture these perceptions, sentiment analysis has become an essential analytical method. Sentiment analysis involves classifying textual opinions into positive, negative, or neutral categories and is widely applied in assessing user satisfaction across digital platforms. By leveraging web scraping techniques, researchers can gather large volumes of user-generated posts that reflect authentic experiences and concerns regarding Coretax. This approach enables a more comprehensive understanding of sentiment patterns, user expectations, and emerging public issues related to digital tax systems.

Previous studies have investigated sentiment analysis within the domains of taxation, public policy, and digital services using various machine learning and deep learning approaches. Farhan & Setiaji, (2023) and Anindya & Kaesmetan, (2025) demonstrated the effectiveness of Naive Bayes and SVM in classifying short-text sentiment, though challenges persist due to informal linguistic patterns commonly found in Indonesian social media posts. Febriani et al., (2025) further showed that artificial neural networks yield higher accuracy when supported by robust preprocessing steps. More recent research by Putri et al., (2024), Khairani et al., (2024), and Sayarizki & Nurrahmi, (2024) emphasized that IndoBERT—an Indonesian transformer-based language model—outperforms traditional algorithms due to its ability to capture contextual and semantic nuances. In the taxation domain, studies by Fahlapi et al., (2025), Manoppo et al., (2025), and Rizkia et al., (2025) highlighted that system errors and policy changes frequently trigger spikes in negative sentiment among users.

Although these studies contribute valuable insights, research focusing specifically on public sentiment toward Coretax remains limited. Most existing works analyze general taxation topics or compare machine learning models without closely examining real user experiences with Coretax. This indicates a clear research gap requiring further investigation using more advanced language models and focused data sources.

Although Coretax is designed to deliver stable, efficient, and user-friendly tax services (das sollen), real-world conditions show that users still experience frequent system errors, slow responses, login failures, and inconsistent service availability (das sein). This discrepancy illustrates a significant gap that must be addressed to evaluate whether the modernization efforts implemented by the DGT align with public expectations and operational standards. Sentiment analysis provides a systematic framework to analyze this divergence and understand users' actual experiences.

Addressing this gap, the present study aims to collect and analyze public posts related to Coretax from the X platform and classify sentiments using the IndoBERT model. The focus is to identify dominant issues discussed by users, understand overall sentiment distribution, and provide insights that may support improvements in Indonesia's digital tax service ecosystem. Through the integration of web scraping techniques and advanced transformer-based language modeling, this research contributes both methodologically and practically to the field of digital governance and sentiment analysis.

LITERATURE REVIEW

Digital Tax Systems and Coretax

The digitalization of Indonesia's tax administration has progressed rapidly, focusing on enhancing service efficiency, transparency, and data accuracy. Coretax serves as the main backbone of this digital transformation, integrating tax reporting, payment processing, and data validation within a modern unified architecture. Meilandri, (2025) explains that Coretax marks a significant milestone in Indonesia's taxation reform by improving system reliability and reducing administrative burden. However, user experiences with Coretax vary widely. Gea et al., (2024) and Hasan & Bimby, (2025) report that taxpayers frequently experience technical issues such as login failures, slow response times, and system errors, which contribute to negative sentiment. Understanding how the public perceives Coretax is therefore essential to evaluate whether Indonesia's digital tax modernization meets user expectations.

Sentiment Analysis and Social Media

Sentiment analysis is a computational method for identifying and classifying opinions within text into categories such as positive, negative, or neutral. Formally defines sentiment analysis as the systematic extraction of subjective information using natural language processing and machine learning techniques. Social media—particularly X (Twitter)—is widely used as a data source due to its real-time, expressive, and concise content. Febriani et al., (2025) demonstrate that Twitter is effective for capturing public responses to digital services and government policies. Automated data collection through web scraping also enhances data volume and representativeness, as reported by (Primatrias et al., 2025) and (Maulana & Wibowo, 2025). Therefore, sentiment analysis of social media content is an essential approach for evaluating public perceptions of digital tax systems such as Coretax.

Traditional Machine Learning Models for Sentiment Analysis

Traditional machine learning models such as Naïve Bayes and Support Vector Machine (SVM) have long been applied in sentiment classification. Farhan & Setiaji, (2023) show that Naïve Bayes performs well for short-text sentiment analysis but is sensitive to informal linguistic patterns common in Indonesian social media. Primatrias et al., (2025) confirm that Naïve Bayes remains relevant for analyzing public opinions on taxation policies. SVM is known for its strong performance in high-dimensional data classification. Fahlapi et al., (2025) demonstrate that SVM achieves stable classification results when analyzing sentiment related to Coretax user reviews. However, several studies highlight that traditional models have difficulty capturing deeper contextual meaning, sarcasm, and non-standard expressions frequently found in social media (Hasan & Bimby, 2025). These limitations provide further justification for adopting deep-learning models capable of processing more complex linguistic patterns.

Transformer Models and the Strengths of IndoBERT

Transformer architectures have significantly transformed natural language processing through their self-attention mechanism, which allows models to capture contextual relationships without relying on sequential processing. IndoBERT, a transformer-based model trained specifically on Indonesian corpora, has demonstrated superior performance across various Indonesian NLP tasks. Putri et al., (2024) show that IndoBERT improves political sentiment classification accuracy compared to traditional machine learning methods. Khairani et al., (2024) emphasize that proper preprocessing steps enhance IndoBERT's performance in handling informal social media comments. Additional studies by Sumartha, (2025), Wijaya et al., (2025), and Anindya & Kaesmetan, (2025) confirm that transformer-based models outperform traditional classifiers in processing informal and context-rich Indonesian text. Kartika et al., (2023) also demonstrate that fine-tuned IndoBERT models significantly improve performance in paraphrase identification and text classification tasks. In the context of Coretax, Rizkia et al., (2025) show that IndoBERT provides higher stability and accuracy compared to manual and lexicon-based labeling approaches.

Prior Studies on Public Sentiment and Taxation

Several studies have examined public sentiment toward government tax policies and digital tax platforms. Fahlapi et al., (2025) reveal that technical issues within the Coretax system often trigger spikes in negative user sentiment. Similarly, Manoppo et al., (2025) report a surge in negative sentiment following the VAT increase to 12%. Primatrias et al., (2025) analyze public responses to taxation surcharge policies, while Tarigan & Idrus, (2024) utilize Multinomial Naïve Bayes to assess sentiment on general taxation topics. Sejati et al., (2024) introduce aspect-based sentiment analysis to gain a more detailed understanding of sentiment directed toward the Ministry of Finance. Collectively, these studies highlight the critical role of sentiment analysis in evaluating the performance and public acceptance of modern taxation systems such as Coretax.

Research Gap and Relevance of This Study

Although numerous studies have analyzed sentiment toward digital government services and taxation policies, several research gaps remain. Many previous works rely on traditional machine learning models and have not fully leveraged transformer-based approaches such as IndoBERT. Research focusing specifically on Coretax is also limited, with existing studies often adopting binary sentiment classification without including neutral categories, resulting in a less comprehensive sentiment representation. Furthermore, the integration of web scraping with IndoBERT-based sentiment analysis for Coretax-related discussions is rare in the current literature. This study addresses these gaps by applying a modern analytical approach that combines automated data collection and a transformer-based classification model to generate more accurate and meaningful insights into public perceptions of Indonesia's digital tax ecosystem.

METHOD

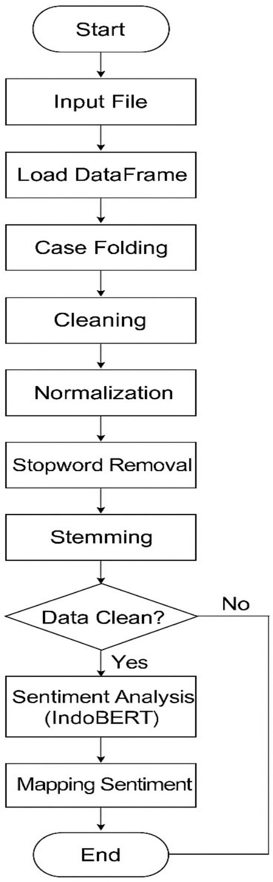

Figure 1. Flowchart of the research method

This flowchart illustrates the complete methodological pipeline, beginning from loading the dataset, performing six stages of text preprocessing, applying the IndoBERT sentiment analysis model, mapping sentiment labels, generating visualizations, and culminating in the interpretation of results.

Research Approach

This study applies a quantitative research design grounded in text analytics using Natural Language Processing (NLP). The primary purpose of the research is to classify public sentiment toward the Coretax tax application based on user-generated content extracted from the X (Twitter) platform. To achieve this objective, the study employs IndoBERT, a transformer-based language model specifically optimized for Indonesian text, ensuring linguistic accuracy and contextual relevance throughout the sentiment classification process.

Data Source and Data Collection

The dataset used in this study consists of tweets containing discussions related to "Coretax," collected through a structured web-scraping procedure. All retrieved tweets were stored in an Excel file comprising several attributes, including the original textual content and the timestamp of each post, along with preprocessing outputs and final sentiment labels. The dataset was then loaded into the working environment using the pandas library. Prior to analysis, each record was inspected to ensure its relevance to the research topic, while duplicate entries and incomplete records were removed to maintain data quality and consistency.

Text Preprocessing

Text preprocessing was conducted to prepare the dataset for sentiment analysis using IndoBERT. The preprocessing workflow consisted of several sequential steps implemented in the Google Colab environment used in this study. Case folding was applied to convert all characters into lowercase to standardize textual representation. This step was followed by the cleaning process, which involved removing URLs, user mentions, hashtags, numbers, non-alphabet symbols, and unnecessary whitespace. After cleaning, tokenization was performed by splitting each tweet into individual words. Normalization was then applied to convert informal and abbreviated expressions into their standard Indonesian forms based on a curated normalization dictionary. Common examples include transformations such as "yg" to "yang," "gak" to "tidak," and "utk" to "untuk." Stopword removal was carried out using the Indonesian stopword list from NLTK to eliminate non-informative words, and stemming was applied using the Sastrawi stemmer to reduce words to their root forms. Although preprocessing was completed through the stemming stage, the IndoBERT model ultimately received the cleaned text as input, following the configuration of the Colab implementation.

Sentiment Analysis Using IndoBERT

Sentiment classification was conducted using the pre-trained IndoBERT model mdhugol/indonesia-bert-sentiment-classification available on HuggingFace. The model categorizes text into three sentiment classes: neutral (LABEL_0), positive (LABEL_1), and negative (LABEL_2). These default labels were mapped into human-readable sentiment categories for interpretation. The classification process utilized the pipeline function for sentiment analysis from the transformers library, with each tweet processed under a maximum input length of 512 characters to ensure compatibility with the model's tokenization and attention mechanisms.

Visualization Techniques

Visualization was employed to support the interpretation of sentiment distribution and keyword dominance within each sentiment category. A bar chart generated using matplotlib illustrated the total number of positive, negative, and neutral tweets across the dataset, enabling a clear comparison of sentiment proportions. In addition, three word clouds were produced to highlight the most frequently occurring terms in each sentiment class. Positive sentiments were visualized using green color tones, negative sentiments using red tones, and neutral sentiments using grey tones. These visualizations provided an intuitive representation of the dominant themes expressed by users and served as complementary evidence for interpreting the sentiment analysis results.

RESULTS AND DISCUSSION

This section presents the findings obtained from the dataset of 381 tweets related to the Coretax system, which were successfully processed through several stages, including text preprocessing, sentiment classification using the IndoBERT model, and visualization of sentiment distribution. The results are interpreted and related to the research objectives and prior studies.

Preprocessing Results

The preprocessing stages were carried out to ensure that all tweets were clean, standardized, and ready to be processed by the IndoBERT model. This workflow consisted of six main steps: case folding, cleaning, tokenizing, normalization, stopword removal, and stemming.

Table 1. The sample of Dataset

	Tanggal	Pengguna	Tweet
0	2025-10-10T03:14:44.000Z	part time bunda | REST.	@kring_pajak\n woy gmnsi ini coretax gangguan ...
1	2025-10-10T03:13:23.000Z	#PajakKitaUntukKita	Hai, Kak.\n\nApakah Kakak sebelumnya melakukan...
2	2025-10-10T03:11:56.000Z	#PajakKitaUntukKita	Hai, Kak.\n\nMohon maaf atas ketidaknyamananny...
3	2025-10-10T03:08:46.000Z	Danny isy	Tim coretax Jancok!!!
4	2025-10-10T03:08:17.000Z	.	coretax kapan lancarnya ampun dah

In the case folding stage, all characters were converted into lowercase to avoid capitalization inconsistencies.

Table 2. Case folding

	Tweet	case_folding
0	@kring_pajak\n woy gmnsi ini coretax gangguan ...	@kring_pajak\n woy gmnsi ini coretax gangguan ...
1	Hai, Kak.\n\nApakah Kakak sebelumnya melakukan...	hai, kak.\n\napakah kakak sebelumnya melakukan...
2	Hai, Kak.\n\nMohon maaf atas ketidaknyamananny...	hai, kak.\n\nmohon maaf atas ketidaknyamananny...
3	Tim coretax Jancok!!!	tim coretax jancok!!!
4	coretax kapan lancarnya ampun dah	coretax kapan lancarnya ampun dah

The cleaning process removed irrelevant components such as URLs, mentions, hashtags, numbers, and non-alphabetic symbols.

Table 3. Cleaning

	case_folding	cleaning

0	@kring_pajak\n woy gmnsi ini coretax gangguan ...	woy gmnsi ini coretax gangguan mulu dari kmrn ...
1	hai, kak.\n\napakah kakak sebelumnya melakukan...	hai kak apakah kakak sebelumnya melakukan akti...
2	hai, kak.\n\nmohon maaf atas ketidaknyamananny...	hai kak mohon maaf atas ketidaknyamanannya apa...
3	tim coretax jancok!!!	tim coretax jancok
4	coretax kapan lancarnya ampun dah	coretax kapan lancarnya ampun dah

Tokenizing split the text into word units, while normalization converted informal or abbreviated words into their standard forms.

Table 4. Tokenizing

	Cleaning	Tokenizing
0	woy gmnsi ini coretax gangguan mulu dari kmrn ...	[woy, gmnsi, ini, coretax, gangguan, mulu, dar...
1	hai kak apakah kakak sebelumnya melakukan akti...	[hai, kak, apakah, kakak, sebelumnya, melakuka...
2	hai kak mohon maaf atas ketidaknyamanannya apa...	[hai, kak, mohon, maaf, atas, ketidaknyamanann...
3	tim coretax jancok	[tim, coretax, jancok]
4	coretax kapan lancarnya ampun dah	[coretax, kapan, lancarnya, ampun, dah]

Table 5. Normalization

	Tokenizing	Normalization
0	[woy, gmnsi, ini, coretax, gangguan, mulu, dar...	[woy, gmnsi, ini, coretax, gangguan, mulu, dar...
1	[hai, kak, apakah, kakak, sebelumnya, melakuka...	[hai, kak, apakah, kakak, sebelumnya, melakuka...
2	[hai, kak, mohon, maaf, atas, ketidaknyamanann...	[hai, kak, mohon, maaf, atas, ketidaknyamanann...
3	[tim, coretax, jancok]	[tim, coretax, jancok]
4	[coretax, kapan, lancarnya, ampun, dah]	[coretax, kapan, lancarnya, ampun, dah]

Stopword removal eliminated words that did not contribute meaningful information to the analysis, and stemming reduced words to their base forms.

Table 6. Stopword removal

	Normalization	stopword_removal
0	[woy, gmnsi, ini, coretax, gangguan, mulu, dar...	[woy, gmnsi, coretax, gangguan, mulu, kmrn, lo...
1	[hai, kak, apakah, kakak, sebelumnya, melakuka...	[hai, kak, kakak, aktivasi, akun, coretax, iya...
2	[hai, kak, mohon, maaf, atas, ketidaknyamanann...	[hai, kak, mohon, maaf, ketidaknyamanannya, ke...
3	[tim, coretax, jancok]	[tim, coretax, jancok]
4	[coretax, kapan, lancarnya, ampun, dah]	[coretax, lancarnya, ampun, dah]

Table 6. Stemming

	stopword_removal	Stemming

0	[woy, gmnsi, coretax, gangguan, mulu, kmrn, lo...	[woy, gmnsi, coretax, ganggu, mulu, kmrn, logi...
1	[hai, kak, kakak, aktivasi, akun, coretax, iya...	[hai, kak, kakak, aktivasi, akun, coretax, iya...
2	[hai, kak, mohon, maaf, ketidaknyamanannya, ke...	[hai, kak, mohon, maaf, ketidaknyamanannya, ke...
3	[tim, coretax, jancok]	[tim, coretax, jancok]
4	[coretax, lancarnya, ampun, dah]	[coretax, lancar, ampun, dah]

All stages were applied to 381 tweets, resulting in a set of clean textual data used as input for the IndoBERT sentiment model.

Sentiment Analysis Results

Sentiment analysis was conducted using the IndoBERT model, which was specifically trained on Indonesian sentiment data. The model produced three sentiment categories: positive, negative, and neutral.

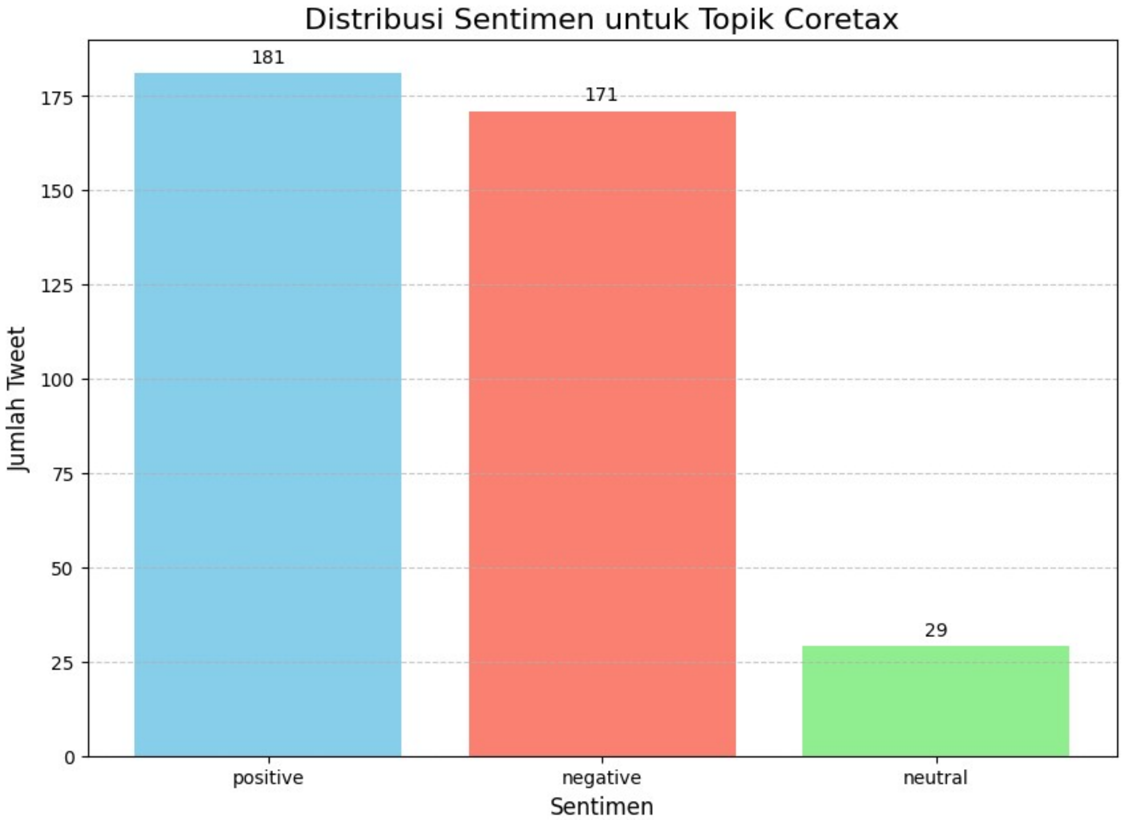

Figure 2. Coretax sentiment distribution

From a total of 381 tweets, the classification results indicated that 181 tweets expressed positive sentiment, 171 expressed negative sentiment, and 29 expressed neutral sentiment. These findings show that public perception of Coretax is relatively balanced between positive and negative sentiments, with neutral sentiment appearing in a noticeably smaller proportion.

Positive sentiment typically reflects successful user experiences when accessing Coretax or appreciation toward system improvements. Negative sentiment predominantly arises from complaints about errors, login issues, or other technical disruptions. Meanwhile, neutral sentiment mainly consists of informational tweets, user questions, and automated responses from official accounts.

Word Cloud Visualization

Word clouds were used to display the most frequently appearing words within each sentiment category.The word cloud for positive sentiment includes terms that describe successful access and satisfying user experiences.

Figure 3. Positive Sentiment Word Cloud

The negative sentiment word cloud is dominated by terms related to system problems such as error, login issues, failures, and server disruptions.

Figure 4. Negative Sentiment Word Cloud

The neutral sentiment word cloud contains vocabulary that tends to be informative or does not convey emotional expression.

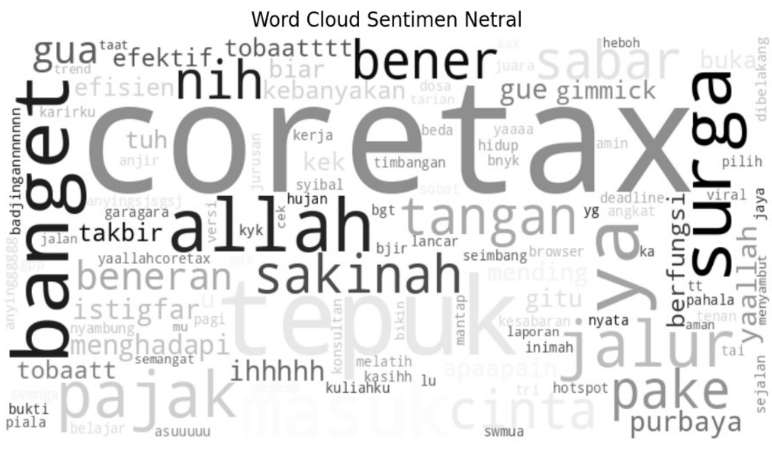

Figure 5. Neutral Sentiment Word Cloud

These three visualizations help clarify the central topics discussed by X (Twitter) users regarding Coretax.

Comparison With Previous Studies

The results obtained in this study align with findings from previous research on digital government services and social media–based sentiment analysis. Prior studies indicate that technical disruptions in government digital platforms often lead to increases in negative sentiment. These patterns are consistent with the findings of Fahlapi et al., (2025) and Manoppo et al., (2025), which reported that system instability is a dominant factor influencing negative public perception.

Furthermore, the effectiveness of IndoBERT in handling informal text is in line with the work of Putri et al., (2024) and Khairani et al., (2024), who confirmed that transformer-based models outperform traditional algorithms such as Naive Bayes and SVM in analyzing Indonesian social media text.

Discussion and Implications

The analysis indicates that the stability and reliability of the Coretax system are the main determinants shaping public perception. When the system functions smoothly, positive sentiment increases significantly; conversely, when technical issues occur, negative sentiment rises rapidly. These findings have direct implications for the development of digital taxation policies. Enhancing server capacity, optimizing login performance, and providing rapid responses to technical disruptions can significantly improve public perception and trust in digital tax systems. Additionally, this study opens opportunities for future research, such as aspect-based sentiment analysis, topic modeling, or comparisons with other transformer-based models like IndoBERTweet, which is specifically optimized for social media text.

CONCLUSION

This study successfully analyzed public sentiment toward the Coretax system using an NLP-based approach with IndoBERT. The results show that system reliability is the main factor influencing public perception: technical disruptions frequently lead to negative sentiment, while stable performance generates positive responses. The findings also confirm the effectiveness of IndoBERT in processing informal Indonesian social-media text, aligning with previous studies on transformer-based sentiment analysis.

The implications highlight the need for improving system stability, enhancing user support, and monitoring real-time feedback to strengthen public trust in digital tax services. Future research may apply aspect-based sentiment analysis, topic modeling, or compare multiple transformer models to provide deeper insights into user concerns.

REFERENCES

Anindya, F. S., & Kaesmetan, Y. R. (2025). IMPLEMENTASI METODE BERT DAN SVM PADA ANALISIS SENTIMEN GAME GENSHIN IMPACT. *Jurnal Manajamen Informatika Jayakarta*, *5*(1), 52–60. https://journal.stmikjayakarta.ac.id/index.php/JMIJayakarta/article/view/1781

Fahlapi, R., Ramadani, L. A., Julianti, D. A., Aldilah, R., & Amalia, D. (2025). ANALISIS SENTIMEN ULASAN PENGGUNA CORETAX DI APLIKASI TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE. *Kohesi: Jurnal Multidisiplin Saintek*, *8*(6). https://ejournal.cahayailmubangsa.institute/index.php/kohesi/article/view/2166/1878/6231

Farhan, N. M., & Setiaji, B. (2023). Komparasi Metode Naive Bayes dan SVM pada Sentimen Twitter Mengenai Persoalan Perpu Cipta Kerja: Comparison of Naive Bayes and SVM Methods on Twitter Sentiment Regarding the Government Regulations on Job Creation Issue. *The Indonesian Journal of Computer Science*, *12*(5). http://ijcs.net/ijcs/index.php/ijcs/article/view/3375

Febriani, S., Wijayanti, Y., & Siswanto, I. (2025). Twitter Sentiment Analysis on Digital Payment in Indonesia Using Artificial Neural Network. *Journal of Applied Informatics* *and Computing*, *9*(2), 526–533. https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/8988/2689/29015

Gea, J. S., Budiati, H., Lase, K. J. D., & Berutu, S. S. (2024). Analisis Sentimen Masyarakat Terhadap Direktorat Jenderal Pajak. *Infact: International Journal of Computers*, *8*(01), 28–34. https://journal.ukrim.ac.id/index.php/JIF/article/view/466/377/1812

Hasan, M. A., & Bimby, N. (2025). Analisis Sentimen Publik Terhadap Kenaikan Pajak PPN di Indonesia Tahun 2024 Menggunakan Algoritma Machine Learning. *JURNAL FASILKOM*, *15*(1), 179–184. https://ejurnal.umri.ac.id/index.php/JIK/article/view/8556

Kartika, B. V., Alfredo, M. J., & Kusuma, G. P. (2023). Fine-Tuned IndoBERT Based Model and Data Augmentation for Indonesian Language Paraphrase Identification. *Revue d'Intelligence Artificielle*, *37*(3). https://www.iieta.org/journals/ria/paper/10.18280/ria.370322

Khairani, U., Mutiawani, V., & Ahmadian, H. (2024). Pengaruh tahapan preprocessing terhadap model Indobert dan Indobertweet untuk mendeteksi emosi pada komentar akun berita Instagram. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *11*(4), 887–894. https://jtiik.ub.ac.id/index.php/jtiik/article/view/8315

Manoppo, M. R., Kolang, I. C., Fiat, D. N. N., Mawara, R. M. C., Sumarno, A. D. P., Yusupa, A., & Tarigan, V. (2025). ANALISIS SENTIMEN PUBLIK DI MEDIA SOSIAL TERHADAP KENAIKAN PPN 12% DI INDONESIA MENGGUNAKAN INDOBERT. *Jurnal Kecerdasan Buatan Dan Teknologi Informasi*, *4*(2), 152–163. https://ojs.ninetyjournal.com/index.php/JKBTI/article/view/322/108

Maulana, I. A., & Wibowo, A. P. W. (2025). Analysis of public sentiment text clustering on tax increases using Orange data mining on Twitter. *Brilliance: Research of Artificial Intelligence*, *5*(1), 93–99. https://jurnal.itscience.org/index.php/brilliance/article/view/5787/4403

Meilandri, D. (2025). *Transformation of Indonesia's Tax System through Coretax: A Qualitative Study in the Digital Era. Sustainability Accounting Journal, 2 (1), 51-56.* https://e-journal.upr.ac.id/index.php/SAJ/article/view/19497/6980/44827

Primatrias, R., Huda, B., & Hilabi, S. (2025). Analisis Sentimen Masyarakat Terhadap Opsen Pajak pada Media Sosial X Menggunakan Metode Naïve Bayes. *IKRAM: Jurnal Ilmu Komputer Al Muslim*, *4*(2), 54–61. https://journal.almuslim.ac.id/index.php/ikram/id/article/view/96

Putri, D. I., Alfian, A. N., Putra, M. Y., & Mulyo, P. D. (2024). IndoBERT Model Analysis: Twitter Sentiments on Indonesia's 2024 Presidential Election. *Journal of Applied Informatics and Computing*, *8*(1), 7–12. https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/7440

Rizkia, A. S., Wufron, W., & Roji, F. F. (2025). Analisis Sentimen Coretax: Perbandingan Pelabelan Data Manual, Transformers-Based, dan Lexicon-Based pada Performa IndoBERT: Sentiment Analysis of Coretax: A Comparison of Manual, Transformers-Based, and Lexicon-Based Data Labeling on IndoBERT Performance. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *5*(3), 1037–1048. https://www.journal.irpi.or.id/index.php/malcom/article/view/2151/1034

Sayarizki, P., & Nurrahmi, H. (2024). Implementation of indobert for sentiment analysis of indonesian presidential candidates. *Indonesian Journal on Computing (Indo-JC)*, *9*(2), 61–72. https://socjs.telkomuniversity.ac.id/ojs/index.php/indojc/article/view/934

Sejati, P. T., Alzami, F., Marjuni, A., Indrayani, H., & Puspitarini, I. D. (2024). Aspect-Based Sentiment Analysis for Enhanced Understanding of'Kemenkeu'Tweets. *Journal of Applied Informatics and Computing*, *8*(2), 487–498. https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/8558/2513/27124

Sumartha, D. K. (2025). IMPLEMENTASI INDOBERT UNTUK ANALISIS SENTIMEN

OPINI PUBLIK TERHADAP KEBIJAKAN KENAIKAN UKT DI ERA PEMERINTAHAN. *Jurnal Informatika Dan Teknik Elektro Terapan*, *13*(3S1). https://journal.eng.unila.ac.id/index.php/jitet/article/view/7880

Tarigan, D. D., & Idrus, S. A. I. (2024). Journal of Informatics and Data Science (J-IDS) Sentiment Analysis of Twitter Users Regarding Taxation Topics in. *Journal of Informatics and Data Science (J-IDS)*, *03*(01). https://doi.org/10.24114/j-ids.xxxxx https://jurnal.unimed.ac.id/2012/index.php/jids/article/view/52465

Wijaya, W., Seputra, K. A., & Dewi, N. P. N. P. (2025). FINE TUNNING MODEL INDOBERT UNTUK ANALISIS SENTIMEN BERITA PARIWISATA INDONESIA. *Jurnal Pendidikan Teknologi Dan Kejuruan*, *22*(2), 195–204. https://ejournal.undiksha.ac.id/index.php/JPTK/article/view/104056