JSCE (Journal of System and Computer Engineering)
Vol.7, No.3, July 2026, pp. 277-291
ISSN (online): 2723-1240
DOI : https://doi.org/10.61628/jsce.v7i3.2722

Aspect-Based Sentiment Analysis of YouTube Comments on CoreTax Using Support Vector Machine and Random Forest

Nur Vadila [1,a,*]; Josua Josen A. Limbong [2,b]; Ratna Juita [3,c]
[1,2,3] Department of Informatics Engineering, Faculty of Engineering, University of Papua, Manokwari 98314, Indonesia
[a] vadilanur27@gmail.com; [b] jja.limbong@unipa.ac.id; [c] r.juita@unipa.ac.id
** Corresponding author*

Abstract

The adoption of the CoreTax Administration System (CTAS) by the Directorate General of Taxes has generated various public responses reflected in comments on the YouTube platform. This study applies Aspect-Based Sentiment Analysis (ABSA) to analyze user sentiment towards CoreTax, compare the performance of the Support Vector Machine (SVM) and Random Forest algorithms, and evaluate the effect of hyperparameter tuning using RandomSearchCV. Data were obtained through web scraping from three YouTube videos and yielded 1,527 valid comments after preprocessing. Comments were then classified using a rule-based approach into five aspects: system, performance, ease of use, tax services, and policy. The analysis results show that the system aspect is the most discussed aspect (56.12%), while negative sentiment dominates the dataset (59.2%). The highest percentage of negative sentiment is found in the ease of use aspect (81.29%), followed by the performance aspect (76.44%). Model evaluation shows that RandomSearchCV improves SVM accuracy from 0.73 to 0.79, while Random Forest does not provide any improvement because the optimal configuration obtained is the same as the initial configuration. These findings indicate that hyperparameter tuning is more effective in improving SVM performance than Random Forest in CoreTax user comment sentiment classification.

Keywords— *CoreTax, Random Forest, Support Vector Machine, Aspect-based Sentiment Analysis, YouTube Comments*

1. Introduction

The development of social media has transformed public communication, enabling people to discuss many topics, such as digital services and government policies. YouTube is one platform that allows users to express their opinions openly thanks to its accessible comment feature. YouTube has become a crucial data source for analyzing and understanding public perception due to the views, criticism, support, and suggestions that appear in the comments section (Singh & Thomas, 2025).

CoreTax is one of the most debated topics on YouTube. The CoreTax Administration System (CTAS) was created to assist the Direktorat Jenderal Pajak (Directorate General of Taxes, DJP) at the Ministry of Finance of the Republic of Indonesia (Indryani & Setyawan, 2025). CTAS helps manage notifications, tax documents, payments, and tax collection processes (Utama & Yuliana, 2025). The implementation of this system has generated a variety of responses from the public, expressed through comments posted on the YouTube platform. These responses can range from positive, negative, or neutral, reflecting the level of public acceptance of the system (Parameswari et al., 2025). Therefore, a method is needed that can perform sentiment analysis on public opinion.

Sentiment analysis is a text mining technique that identifies and classifies opinions or emotional expressions into predefined sentiment categories, including positive, negative, and neutral (Sari & Suryono, 2024). Numeros machine learning approaches have been used to classify sentiment in text data. Two of them are the Support Vector Machine (SVM) and Random Forest algorithms. Support Vector Machine (SVM) constructs an optimal separating hyperplane that maximizes the margin between different data classes (Khaira et al., 2023). In contrast, Random Forest is an ensemble-based algorithm that integrates predictions from multiple decision trees to improve the robustness and accuracy of sentiment classification (Sidauruk et al., 2023).

Research conducted by Syafia et al. (2023) analyzing sentiment in YouTube comments about the boy group BTS, showing that SVM produced higher scores than Random Forest. However, the study only focused on general sentiment classification and did not apply the Aspect-Based Sentiment Analysis (ABSA) approach. The research conducted by Morama et al. (2022) analyzing Hotel Tentrem rewiews using Aspect-Based Sentiment Analysis (ABSA), the Random Forest Classifier algorithm only performed on room aspects and produced the same accuracy and F1-score values, namely 90%. However, this study was limited to hotel review data and only analyzed aspects related to rooms, so it did not examine user perceptions of public service systems such as CoreTax. The research conducted by Suardi et al. (2025) researching public opinion on electric vehicle policies in Indonesia, the results showed that sentiment analysis can provide insight into public perceptions of a policy. However, the study did not address the CoreTax system and did not use an aspect-based approach. Furthermore, Rizkia et al. (2025) conducted a sentiment analysis of CoreTax using IndoBERT by comparing manual, transformer-based, and lexicon-based labeling approaches. However, this study focused on the overall sentiment labeling and classification strategy without conducting aspect-based sentiment analysis. As a result, the specific aspects that elicit positive and negative user responses to the CoreTax system have not been explored in depth.

Although numerous studies on sentiment analysis on social media have been conducted, several research gaps remain that require further study. First, research on public perceptions of the implementation of the CoreTax system in Indonesia is still very limited, particularly those utilizing YouTube comment data as a source of public opinion. Second, most previous studies still use a document-level sentiment analysis approach, thus failing to identify sentiment on specific aspects discussed by users. Third, studies comparing the performance of Support Vector Machine (SVM) and Random Forest algorithms within an Aspect-Based Sentiment Analysis (ABSA) framework on YouTube comments related to digital tax services are still rare. Therefore, Aspect-based Sentiment Analysis (ABSA) is a more comprehensive and in-depth approach, as it allows for more detailed sentiment identification based on specific aspects of an object or service (Horsa & Tune, 2023). Based on this, this study aims to analyze aspect-based sentiment in YouTube comments related to CoreTax, compare the performance of the Support Vector Machine (SVM) and Random Forest algorithms, and evaluate the effect of hyperparameter tuning using RandomSearchCV on the performance of both algorithms.

This study provides several contributions to the application of Aspect-Based Sentiment Analysis (ABSA) in public services. First, this study applies the ABSA approach to analyze user comments on the YouTube platform regarding the implementation of the CoreTax system, enabling sentiment identification based on specific aspects, such as the system, performance, ease of use, tax services, and policies (Saputra et al., 2025). Second, this study compares the performance of the Support Vector Machine (SVM) and Random Forest algorithms before and after hyperparameter tuning using RandomSearchCV, this providing an overview of the effect of the tuning process on the performance of each algorithm (Rizky et al., 2024). Third, this study evaluates the impact of hyperparameter tuning on sentiment classification results in both algorithms. The findings of Grthis study are expected to serve as a reference for further research in developing the application of Aspect-Based Sentiment Analysis (ABSA) in analyzing public opinion on various digital public services.

2. Method

This research was conducted through several stages designed to obtain relevant data and produce an accurate sentiment classification model. The research methodology design flow includes several main stages, starting from data collection to evaluation of the resulting classification model. A complete overview of the research methodology stages can be seen in Figure 1.

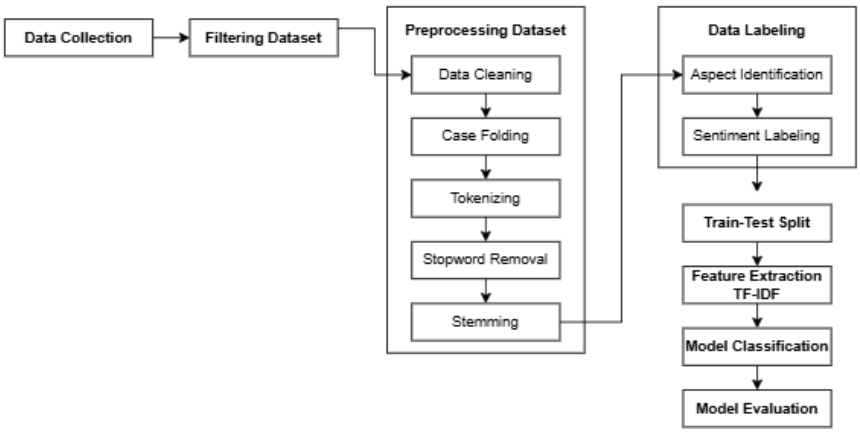

Figure 1. Research Method Diagram

2.1 Dataset Collection

The first step in this research was dataset collection, which involved collecting user comments on YouTube discussing the CoreTax system. The data collection process was conducted using web scraping techniques with the help of the YouTube Comments Scraper tool. The data came from three YouTube videos discussing the implementation of the CoreTax system. The scraping process set a maximum comment limit of 3,000. This resulted in 2,504 comments, which were then used as the initial dataset in this study.

2.2 Data Filtering

Data filtering is the stage of filtering irrelevant datasets to improve data quality before further analysis (Rivaldi & Wismarini, 2024). At this stage, data was filtered for various types of comments that did not represent user opinions. Comments that were removed included comments unrelated to the CoreTax system, comments that were spam or promotional in nature, duplicate comments that appeared more than once, and comments that did not contain meaningful opinions, such as comments containing only symbols, punctuation, or incomprehensible text. Furthermore, blank comments and comments that did not provide any information regarding the user experience with the CoreTax system were also removed from the dataset. The main goal of this stage was to ensure that the data used truly represented user views on the topic being studied. After the filtering process was completed, the total number of data used in this study was 1,527 comments.

2.3 Data Preprocessing

Data preprocessing is the initial text processing stage which aims to clean and prepare data before further analysis is carried out (Adriana et al., 2023)The steps involved include cleaning, which involves removing irrelevant elements such as URLs, mentions, hashtags, punctuation, and emojis (Rizkia et al., 2025); case folding, which is changing all characters to lowercase letters to standardize writing; tokenization, which is breaking sentences into word units (Limbong et al., 2022); stopword removal, which is removing words that do not have a significant contribution to the meaning of the text (S & Yamasari, 2024); and stemming, which is changing affixed words into basic word forms to simplify word variations in text (Prayoga et al., 2025). The preprocessing process can be seen in Table 1.

Table 1. Preprocessing Results

Preprocessing	Input	Output
Cleaning	Ciri khas aplikasi pemerintah 1. gagal verifikasi 2. tidak menerima OTP 3 kalo point 1 dan 2 berhasil, siap-siap point 3, gagal login Apps pemerintah memang fungsi utamanya hanya utk menguji kesabaran	Ciri khas aplikasi pemerintah gagal verifikasi tidak menerima OTP kalo point dan berhasil siapsiap point gagal login Apps pemerintah memang fungsi utamanya hanya utk menguji kesabaran
Case Folding	Ciri khas aplikasi pemerintah gagal verifikasi tidak menerima OTP kalo point dan berhasil siapsiap point gagal login Apps pemerintah memang fungsi utamanya hanya utk menguji kesabaran	ciri khas aplikasi pemerintah gagal verifikasi tidak menerima otp kalo point dan berhasil siapsiap point gagal login apps pemerintah memang fungsi utamanya hanya utk menguji kesabaran
Tokenizing	ciri khas aplikasi pemerintah gagal verifikasi tidak menerima otp kalo point dan berhasil siapsiap point gagal login apps pemerintah memang fungsi utamanya hanya utk menguji kesabaran	[’ciri’, ‘khas’, ‘aplikasi’, ‘pemerintah’, ‘gagal’, ‘verifikasi’, ‘tidak’, ‘menerima’, ‘otp’, ‘kalo’, ‘point’, ‘dan’, ‘berhasil’, ‘siapsiap’, ‘point’, ‘gagal’, ‘login’, ‘apps’, ‘pemerintah’, ‘memang’, ‘fungsi’, ‘utamanya’, ‘hanya’, ‘utk’, ‘menguji’, ‘kesabaran’]
Stopword Removal	[‘ini’, ‘sistem’, ‘hantu’, ‘kita’, ‘dihantuin’, ‘tiap’, ‘hari’, ‘orang’, ‘kpp’, ‘juga’, ‘masih’, ‘pada’, ‘bingung’, ‘soal’, ‘eror’, ‘coretax’, ‘yang’, ‘tiap’, ‘hari’, ‘ada’, ‘aja’, ‘masalahnya’, ‘apalagi’, ‘wajib’, ‘pajaknya’, ‘tambah’, ‘stres’, ‘lagi’]	sistem hantu dihantuin orang kpp bingung eror coretax wajib pajaknya stres
Stemming	ciri khas aplikasi pemerintah gagal verifikasi menerima otp kalo point berhasil siapsiap point gagal login apps pemerintah fungsi utamanya utk menguji kesabaran	ciri khas aplikasi perintah gagal verifikasi terima otp kalo point hasil siapsiap point gagal login apps perintah fungsi utama utk uji sabar

2.4 Identification of Aspects

Aspect identification was conducted to group comments into specific aspect categories relevant to the CoreTax system. This process employed a rule-based approach, matching the words contained in the comments against a predetermined list of keywords. Through this process, each opinion was mapped to five aspects: system, performance, ease of use, tax services, and policy. If a comment contained more than one aspect, the aspect was determined based on the most dominant keyword occurrence. Table 2 presents aspect definitions and examples of keywords used.

Table 2. Definition of Aspects and Keywords

Aspect	Definition	Keyword
System	Comments related to CoreTax features, user accounts, login processes, application access, and overall system functionality.	Login, account, access, feature, application, website, platform, system

Performance	Comments related to system speed, stability, responsiveness, and technical issues experienced by users.	Error, bug, server, failed, down, slow, loading, response
Ease of Use	Comments related to ease of use, interface design, navigation, and user experience when operating CoreTax.	Easy, difficult, hard, complicated, menu, interface, display
Tax Services	Comments related to tax administration activities, reporting, payment processes, and services provided to taxpayers	Tax, payment, report, taxpayer, service, administration, officer
Policy	Comments related to tax regulations, government policies, and the implementation of the CoreTax System	Policy, regulation, government, DJP, project, budget

2.5 Sentiment Labeling

After successfully identifying comment aspects, the next stage is sentiment labeling, which also uses a rule-based approach. This approach is carried out by matching words in comments against a predetermined list of keywords (Abdullah, 2021). In this study, sentiment is classified into three categories: positive, negative, and neutral. Positive sentiment refers to comments expressing satisfaction, support, usefulness, or perceived ease of use. Negative sentiment refers to complaints, dissatisfaction, obstacles, errors, and criticism. Neutral sentiment refers to informative, questioning, or descriptive comments that do not explicitly state polarity (Imbiri et al., 2026).

2.6 Feature Extraction (TF-IDF)

Feature extraction is done to convert text data into a representation in numeric form (Hilmi & Indriyanti, 2025). The TF-IDF (Term Frequency-Inverse Document Frequency) method was chosen as the feature extraction method because of its ability to highlight the most relevant words in the review text. This method works by assigning weights to each word in a document based on how frequently the word appears and how important it is (Basri et al., 2024).

2.7 Train-Test Split

A train-test split is the process of dividing a dataset into training and test data, which aims to evaluate the model's ability to classify data that has not been used during the training process. The split ratio between training and test data is a crucial factor because it can affect the performance and accuracy of the resulting model (Oktafiani et al., 2023). In this study, the dataset was split 80:20, with 80% of the data used as training data and 20% as test data. Of the 1,527 comments, 1,221 were used as training data, while 306 were used as test data.

2.8 Model Classification

This study uses two machine learning algorithms, namely Support Vector Machine (SVM) and Random Forest, to classify the sentiment of user comments on the CoreTax system. SVM was selected because it is effective in handling high-dimensional data, while Random Forest can improve classification performance through an ensemble learning approach. To obtain the optimal model configuration, hyperparameter tuning was performed using RandomSearchCV. This approach randomly samples combinations of predefined hyperparameter values from the search space to identify the model configuration that achieves the best performance (Buczak et al., 2024). For the SVM algorithm, the optimized parameters include C and the linear kernel. Meanwhile, in the Random Forest algorithm, the optimization process is carried out on the parameters n_estimators (number of decision trees), max_depth (maximum tree depth), min_samples_split (minimum number of samples to split a node), and min_samples_leaf (minimum number of samples at each leaf node). Next, the best model from each algorithm is used to classify the test data and is evaluated using accuracy, precision, recall, and F1-Score metrics.

2.9 Model Evaluation

Model evaluation was conducted to compare the performance of the two algorithms and determine which model performed best in aspect-based sentiment classification. This evaluation process used four measurement metrics: accuracy, precision, recall, and F1-Score (Yutika et al., 2021). The following formula equations for each evaluation metric are presented in equations (1), (2), (3), and (4).

$$\text{Accuracy} = \frac{1}{N}\sum_{c=1}^{N}\frac{TP_c + TN_c}{TP_c + TN_c + FP_c + FN_c} \tag{1}$$

$$\text{Precision}_{macro} = \frac{1}{N}\sum_{c=1}^{N}\frac{TP_c}{TP_c + FP_c} \tag{2}$$

$$Recall_{macro} = \frac{1}{N}\sum_{c=1}^{N}\frac{TP_c}{TP_c + FN_c} \tag{3}$$

$$\text{F1}_{macro} = \frac{1}{N}\sum_{c=1}^{N}\frac{2 \times Precision_c \times Recall_c}{Precision_c + Recall_c} \tag{4}$$

3. Results and Discussion

3.1 Data Collection

Research data was collected from user comments on YouTube specifically discussing the CoreTax system. Data collection was conducted using the YouTube Comments Scraper technique, using comments from three videos relevant to the research topic as data sources. This process resulted in 2,504 comments as the initial data for the study.

The next stage was data filtering, which was carried out to remove comments irrelevant to the research context, duplicate comments, and comments that did not contain information that could be further analyzed. After the filtering process, the data used in this study was reduced to 1,527 comments.

3.2 Identification of Aspects

Aspect identification was conducted using a rule-based approach using a predetermined keyword matching technique. In this study, comment aspects were grouped into five categories: system, performance, ease of use, tax services, and policies. This aspect identification process aimed to determine which topics were most frequently discussed by users regarding the CoreTax system. The results of the aspect identification analysis can be seen in Table 3.

Table 3. Identification of Aspects

Aspect	Number of Comments	Percentage
System	857	56.12%
Performance	191	12.51%
Ease of Use	171	11.20%
Tax Services	201	13.16%
Policy	107	7.01%

The results presented in Table 3 indicate that system aspect was the most frequently discussed aspect, with 857 comments (56.12%). Next, the tax service aspect received 201 comments (13.16%), followed by performance with 191 comments (12.51%), ease of use with 171 comments (11.20%), and policies with 107 comments (7.01%). This indicates that users focus more on the technical aspects of the CoreTax system than on other aspects.

3.3 Sentiment Labeling

Sentiment labeling was performed manually by the author. The labeling process adhered to pre-established guidelines for positive, negative, and neutral sentiment categories. The sentiment labeling results can be seen in Figure 2.

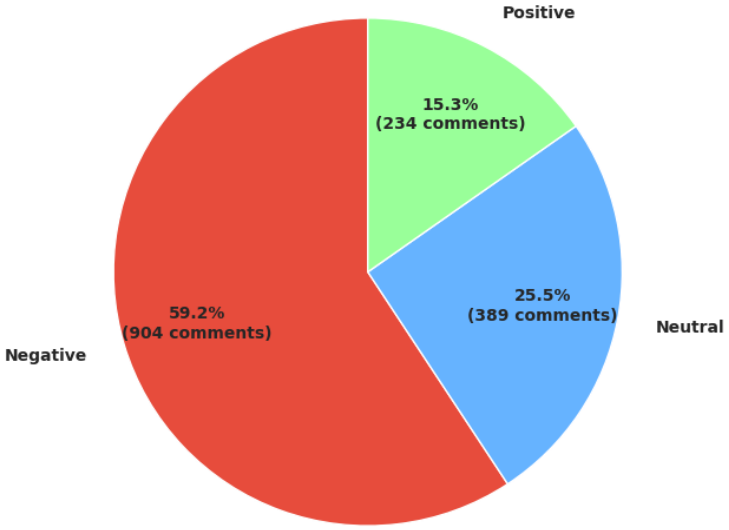

Figure 2. Sentiment Labeling

The analysis results in Figure 2 show that negative sentiment dominated user comments, accounting for 59.2%, or 905 comments. This dominance of negative sentiment reflects that most users still face various obstacles in operating the Coretax system. On the other hand, positive sentiment accounted for 15.3%, or 233 comments, generally from users who felt the system was easy to use. Neutral sentiment accounted for 25.5%, or 389 comments, found in informative comments that did not directly express opinions about the CoreTax system.

3.4 Sentiment Based on Aspect

The following analysis was conducted to determine the distribution of sentiment for each previously identified aspect. This analysis aimed to determine user perceptions of each aspect related to the CoreTax system. The complete analysis results can be seen in Table 4.

Table 4. Sentiment Based on Aspect

Aspect	Positive	Negative	Neutral
System	127	422	308
Performance	38	146	7
Ease of Use	23	139	9
Tax Services	26	127	48
Policy	19	71	17

The results of sentiment labeling based on aspects in Table 4 show that negative sentiment dominated across all aspect categories analyzed. For the system aspect, there were 422 negative comments, 308 neutral comments, and 127 positive comments. For the performance aspect, there were 146 negative comments, 7 neutral comments, and 38 positive comments. Meanwhile, for the ease of use aspect, there were 139 negative comments, 9 neutral comments, and 23 positive comments. Meanwhile, for the tax service aspect, there were 127 negative comments, 48 neutral comments, and 26 positive comments. For the policy aspect, there were 71 negative comments, 17 neutral comments, and 19 positive comments. This dominance of negative sentiment indicates that users still face various obstacles in using the CoreTax system, especially those related to system stability and ease of use.

3.5 Percentage of Negative Sentiment

In addition to analyzing the distribution of sentiment across each aspect, this study also calculated the proportion of negative sentiment to identify aspects with the highest levels of user dissatisfaction. The results are shown in Table 5.

Table 5. Percentage of Negative Sentiment

Aspect	Negative Percentage
System	49.24%
Performance	76.44%
Ease of Use	81.29%
Tax Services	63.18%
Policy	66.36%

Based on the calculation results in Table 5, the ease of use aspect had the highest percentage of negative sentiment at 81.29%, followed by the performance aspect at 76.44%, policy at 66.36%, tax services at 63.18%, and the system at 49.24%. This finding indicates that although the system aspect had the highest number of comments, the highest level of negative sentiment was found in the ease of use and performance aspects. This finding suggests that the most discussed aspects are not necessarily the ones with the highest level of dissatisfaction, thus aspect-based sentiment analysis is able to provide more detailed insights than general document-level sentiment analysis.

3.6 Model Evaluation

Model evaluation was conducted to measure the algorithm's performance in classifying sentiments in user comments about the CoreTax system using accuracy, precision, recall, and F1-Score metrics. Prior to model evaluation, a hyperparameter tuning process using RandomSearchCV was performed to obtain the optimal model configuration. The hyperparameter tuning results showed that the best configuration for the SVM algorithm used a linear kernel with a C value of 100. Meanwhile, the Random Forest algorithm produced the best configuration with

n_estimators=100, max_depth=None, min_samples_split=2, and min_samples_leaf=1. The evaluation results before and after the hyperparameter tuning process can be seen in Table 6.

Table 6. Comparison of Model Evaluation Results Before and After RandomSearchCV

Model	Configuration	Accuracy	Precision	Recall	F1-Score
Support Vector Machine	Default	0.73	0.80	0.63	0.68
Support Vector Machine	RandomSearchCV	0.79	0.79	0.76	0.77
Random Forest	Default	0.80	0.83	0.73	0.77
Random Forest	RandomSearchCV	0.80	0.83	0.73	0.77

Based on the test results presented in Table 6, the application of hyperparameter tuning using RandomSearchCV had different effects on the performance of the two algorithms. For the SVM algorithm, the tuning process successfully improved model performance. Accuracy increased from 0.73 to 0.79, recall increased from 0.63 to 0.76, and the F1-Score increased from 0.68 to 0.77. Although the precision value decreased, from 0.80 to 0.79, the improvements in the other evaluation metrics indicate that the overall performance of the SVM model improved after the optimization process.

In contrast, in the Random Forest algorithm, the hyperparameter tuning process did not result in changes to model performance. RandomSearchCV results showed that the best configuration obtained was the same as the initial configuration used, thus providing no increase in accuracy, precision, recall, or F1-Score. This indicates that the initial Random Forest configuration was able to provide optimal performance. Therefore, in the CoreTax user comment dataset used in this study, hyperparameter tuning had a more significant impact on the performance of the SVM algorithm compared to Random Forest.

To provide a more in-depth analysis of the performance of each model, a detailed evaluation was also conducted for each sentiment class using precision, recall, and F1-Score metrics. The comparison results for the precision metric are shown in Figure 3, the recall metric in Figure 4, and the F1-Score metric in Figure 5.

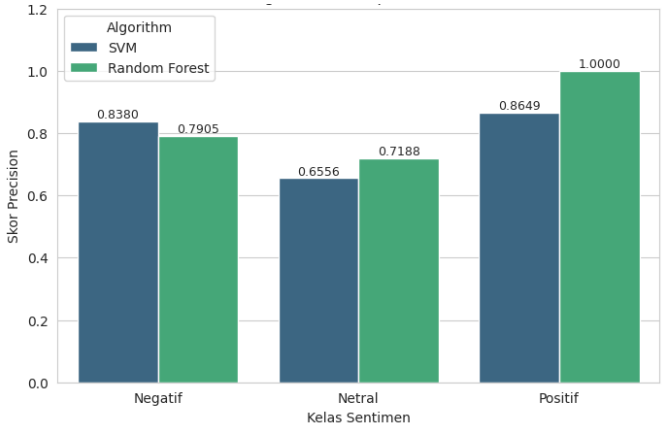

Figure 3. Precision Graphics

Based on the test results in Figure 3, the precision values of the two algorithms show differences in each sentiment class. In the negative sentiment class, the SVM algorithm obtained a precision value of 0.8380, higher than Random Forest, which reached 0.7905. In the neutral class, Random Forest showed slightly better performance with a precision value of 0.7188, while SVM obtained a value of 0.6556. Meanwhile, in the positive sentiment class, Random Forest achieved a perfect precision value of 1.0000, outperforming SVM, which obtained a value of 0.8649. These results indicate that Random Forest is more accurate in identifying comments that

fall into the positive sentiment class, while SVM has an advantage in classifying comments with negative sentiment.

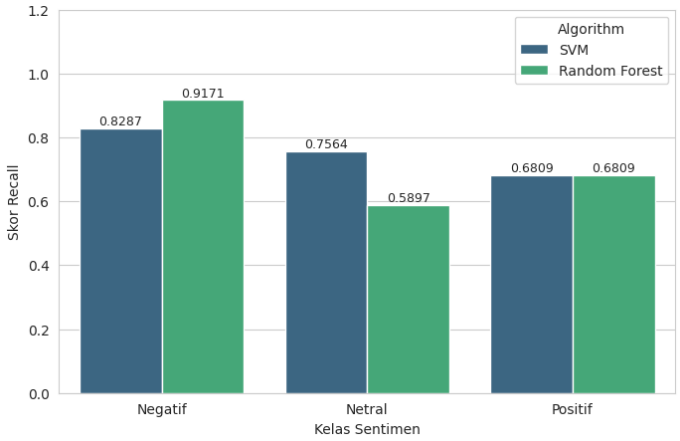

Figure 4. Recall Graphics

Based on the test results in Figure 4, the recall values of the two algorithms show variations in each sentiment class. In the negative sentiment class, the Random Forest algorithm obtained a value of 0.9171, higher than the SVM algorithm which reached 0.8287. This indicates that Random Forest is able to recognize most of the data that truly belongs to the negative sentiment class. Conversely, in the neutral sentiment class, SVM showed better performance with a recall value of 0.7564, while Random Forest obtained a value of 0.5897. Meanwhile, in the positive sentiment class, both algorithms produced the same recall value, namely 0.6809. These results indicate that Random Forest has a better ability to detect data in the negative sentiment class, while SVM is superior in recognizing data in the neutral sentiment class. In the positive sentiment class, both algorithms have equal ability to identify data that belongs to that class.

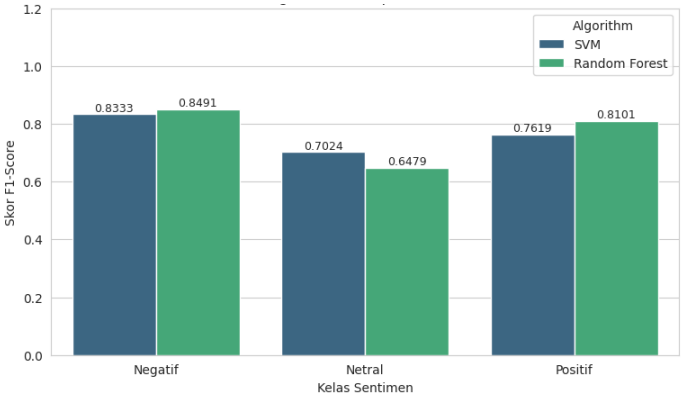

Figure 5. F1-Score Graphics

Based on the test results in Figure 5, the F1-Score values of the two algorithms show differences in each sentiment class. In the negative sentiment class, Random Forest obtained a value of 0.8491, slightly higher than SVM which reached 0.8333. This indicates that both algorithms have almost the same balance in classifying negative sentiment. In the neutral class, SVM showed better performance with a value of 0.7024, while Random Forest obtained a value of 0.6479. Meanwhile, in the positive sentiment class, Random Forest again showed better performance with a value of 0.8101, outperforming SVM which obtained a value of 0.7619. Overall, these results indicate that Random Forest has better performance in classifying negative and positive sentiment because it is able to maintain a balance between precision and recall, while SVM excels in the neutral sentiment class.

To obtain a more detailed picture of the model's ability to classify each sentiment class, further analysis was performed using a confusion matrix. The confusion matrix was utilized to evaluate the classification performance by displaying the number of correct and incorrect predictions for each sentiment class. The complete analysis results can be seen in Figures 6 and 7.

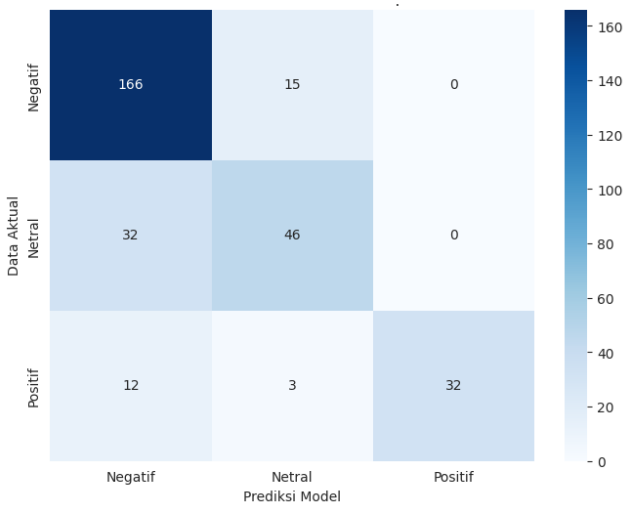

Figure 6. Confusion Matrix Random Forest

Based on the test results in Figure 6, the Random Forest confusion matrix shows that out of 181 negative comments, 166 were correctly classified, while 15 were incorrectly classified as neutral. In the neutral class, out of 78 comments, 46 were correctly classified, and 32 were incorrectly classified as negative. Meanwhile, out of 47 positive comments, 32 were correctly classified, while 12 were incorrectly classified as negative and 3 were neutral. These results indicate that the Random Forest algorithm is capable of classifying negative and positive sentiments well. However, misclassification errors still occur, especially in the neutral and negative classes, indicating similar characteristics between the two classes. Nevertheless, overall, Random Forest provides good classification performance, indicated by the high number of predictions that match their classes.

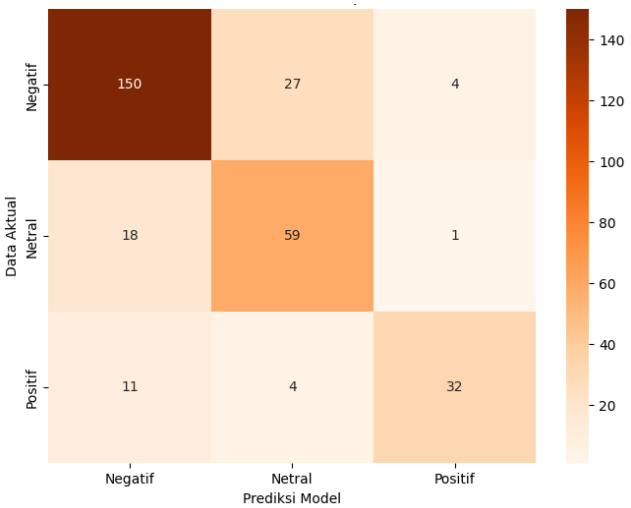

Figure 7. Confusion Matrix SVM

Based on the test results in Figure 7, the SVM confusion matrix shows that out of 181 negative comments, 150 were correctly classified, while 27 were incorrectly classified as neutral and 4 as

positive. In the neutral class, out of 78 comments, 59 were correctly classified, while 18 were incorrectly classified as negative and 1 as positive. Meanwhile, out of 47 positive comments, 32 were correctly classified, while 11 were incorrectly classified as negative and 4 as neutral. These results indicate that the SVM algorithm is capable of classifying negative sentiment well, although misclassifications still occur, especially between the negative and neutral classes. Furthermore, some positive comments are still predicted as either negative or neutral. This indicates that the similarity in language characteristics between classes makes it difficult for the SVM to accurately distinguish sentiments.

3.7 Wordcloud Visualization

As a complement to the analysis of sentiment classification results, word visualization was carried out using Wordcloud to identify the most dominant and frequently appearing words in user comments, as presented in Figures 8, 9 and 10.

Figure 8. Positive Wordcloud

Figure 9. Neutral Wordcloud

Figure 10. Negative Wordcloud

To strengthen the interpretation of the WordCloud visualization results, a frequency analysis of the most frequently occurring words in each sentiment category was performed. This analysis aims to provide a more measurable picture of the dominant terms used by users to express their opinions regarding the CoreTax system. The results of the word frequency analysis can be seen in Tables 7, 8, and 9.

Table 7. Positive Sentiment Frequency

Positive Sentiment	Frequency
kerja	40
bagus	39
solusi	38
baik	37
lancar	35

Table 8. Neutral Sentiment Frequency

Neutral Sentiment	Frequency
coretax	53
pajak	53
email	47
npwp	31
perintah	24

Table 9. Negative Sentiment Frequency

Negative Sentiment	Frequency
bayar	123
sulit	110
susah	95
login	92
sistem	86

Based on Tables 7, 8, and 9, the results of the word frequency analysis reinforce the aspect-based findings, which indicate that ease of use and performance are the primary sources of negative sentiment. The dominance of the words "sulit," "susah," and "login" indicates that users are still experiencing difficulties using and accessing the CoreTax system. Furthermore, the high frequency of the word "bayar" indicates that some complaints also relate to the tax administration process carried out through the CoreTax system.

4. Conclusions

This study successfully applied Aspect-Based Sentiment Analysis (ABSA) to analyze user sentiment towards the CoreTax system based on comments on the YouTube platform. The analysis results showed that the system aspect was the most discussed aspect by users with 857 comments (56.12%). Meanwhile, negative sentiment dominated all aspects with a percentage of 59.2%, reflecting the various obstacles that users still face in operating the CoreTax system. Although the system aspect was the most frequently discussed aspect, the ease of use aspect had the highest level of negative sentiment at 81.29%, followed by the performance aspect at 76.44%.

These findings indicate that ease of use and performance are aspects that need to be a top priority in the development and improvement of CoreTax.

The comparison of model performance before and after hyperparameter tuning shows that RandomSearchCV has different impacts on both algorithms. In Support Vector Machine (SVM), the tuning process improves model performance by increasing accuracy from 0.73 to 0.79, recall from 0.63 to 0.76, and F1-score from 0.68 to 0.77, although precision decreases slightly from 0.80 to 0.79. In contrast, in Random Forest, the tuning process does not result in an increase in performance because the initial configuration has provided optimal results, namely accuracy 0.80, precision 0.83, recall 0.73, and F1-score 0.77. RandomSearchCV has a greater impact in improving the performance of the SVM model compared to the Random Forest model on the dataset used in this study.

This study demonstrates that Aspect-Based Sentiment Analysis (ABSA) provides more detailed insights into user concerns than document-level sentiment analysis. However, the study has several limitations, including the use of comments from only three YouTube videos, a rule-based approach for aspect identification, and a comparison involving only the Support Vector Machine (SVM) and Random Forest algorithms. Future studies are encouraged to incorporate data from multiple social media platforms, adopt more adaptive aspect identification methods, and explore transformer-based models to further improve sentiment analysis performance.

Based on these findings, the Direktorat Jenderal Pajak (Directorate General of Taxes, DJP) is encouraged to improve CoreTax by enhancing system usability, optimizing performance, ensuring the stability of key features, strengthening user support services, and providing clearer policy information to improve user experience and foster a more positive perception of the system.

References

Abdullah, R. (2021). *Aspect Based Sentiment Analysis for Explicit and Implicit Aspects in Restaurant Review using Grammatical Rules , Hybrid Approach , and SentiCircle*. *14*(5), 294–305. https://doi.org/10.22266/ijies2021.1031.27

Adriana, N. M. T. O., Suarjaya, I. M. A. D., & Githa, D. P. (2023). *ANALISIS SENTIMEN PUBLIK TERHADAP AKSI DEMONSTRASI DI INDONESIA MENGGUNAKAN SUPPORT VECTOR MACHINE DAN RANDOM FOREST*. *3*(2), 257–267.

Basri, H., Junianto, M. B. S., & Kusyadi, I. (2024). *Enhancing Usability Testing Through Sentiment Analysis : A Comparative Study Using SVM , Naive Bayes , Decision Trees and Random Forest*. *7*(4), 1603–1610. https://doi.org/10.32493/jtsi.v7i4.45117

Buczak, P., Groll, A., Pauly, M., Rehof, J., & Horn, D. (2024). Using sequential statistical tests for efficient hyperparameter tuning. *AStA Advances in Statistical Analysis*, *108*(2), 441–460. https://doi.org/10.1007/s10182-024-00495-1

Hilmi, F. H., & Indriyanti, A. D. (2025). *Sentiment Analysis of 2024 Election Fraud Using SVM and Naïve Bayes Algorithms*. *5*(4), 353–364.

Horsa, O. G., & Tune, K. K. (2023). *Aspect-Based Sentiment Analysis for Afaan Oromoo Movie Reviews Using Machine Learning Techniques*. *2023*. https://doi.org/10.1155/2023/3462691

Imbiri, M. A., Baisa, L. Y., & Limbong, J. A. (2026). Sentiment Classification and Influential Actor Detection on Twitter (Case Study : The Raja Ampat Mining Conflict). *Indonesian Journal of Data and Science*, *7*(1), 1–21. https://doi.org/https://doi.org/10.56705/ijodas.v7i1.376

Indryani, A. P., & Setyawan, N. D. (2025). *ANALYSIS OF UNESA STUDENTS' RECEPTIONS OF THE IMPLEMENTATION OF THE CORE TAX ADMINISTRATION SYSTEM (CTAS) IN INDONESIA IN 2025*. *1*(3), 144–156.

Khaira, U., Aryani, R., & Hardian, R. W. (2023). *Perbandingan Algoritma Naïve Bayes dan Support Vector Machine (SVM) dalam Analisis Sentimen Kebijakan Kemdikbudristek Tentang Kuota Internet Selama Pandemi Covid-19*. *18*(2), 183–191.

Limbong, J. J. A., Sembiring, I., Hartomo, K. D., Kristen, U., Wacana, S., & Korespondensi, P. (2022). *ANALISIS KLASIFIKASI SENTIMEN ULASAN PADA E-COMMERCE SHOPEE*

BERBASIS WORD CLOUD DENGAN METODE NAIVE BAYES DAN K-NEAREST ANALYSIS OF REVIEW SENTIMENT CLASSIFICATION ON E-COMMERCE SHOPEE WORD CLOUD BASED WITH NAÏVE BAYES AND K-NEAREST NEIGHBOR METHODS. *9*(2), 347–356. https://doi.org/10.25126/jtiik.202294960

Morama, H. C., Ratnawati, D. E., & Arwani, I. (2022). *Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier*. *6*(4), 1702–1708.

Oktafiani, R., Hermawan, A., & Avianto, D. (2023). *Pengaruh Komposisi Split Data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning*. *9*(April), 19–28. https://doi.org/10.34128/jsi.v9i1.622

Parameswari, S. D., Lubis, M., Suakanto, S., Ramadhan, Y. Z., Amanah, N., & Dila, R. A. (2025). *Studi Perbandingan Naï ve Bayes dan Support Vector Machine (SVM) dalam Analisis Sentimen Pengguna Metaverse*. *4*(3), 1059–1065.

Prayoga, M. B., Cahyono, N., Subektiningsih, & Kamarudin. (2025). *PENERAPAN GRID SEARCH UNTUK OPTIMASI MODEL MACHINE LEARNING DALAM KLASIFIKASI SENTIMEN KOMENTAR YOUTUBE*. *9*(3), 3817–3824.

Rivaldi, R. C., & Wismarini, T. D. (2024). *Analisis Sentimen Pada Ulasan Produk Dengan Metode Natural Language Processing (NLP) (Studi Kasus Zalika Store 88 Shopee)*. *17*(1), 120–128.

Rizkia, A. S., Wufron, & Roji, F. F. (2025). *Sentiment Analysis of Coretax : A Comparison of Manual , Transformers- Based , and Lexicon-Based Data Labeling on IndoBERT Performance*. *5*(July), 1037–1048.

Rizky, M. H., Faisal, M. R., Budiman, I., & Kartini, D. (2024). Effect of Hyperparameter Tuning Using Random Search on Tree-Based Classification Algorithm for Software Defect Prediction. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *18*(1), 95–106. https://doi.org/: https://doi.org/10.22146/ijccs.90437

S, R. A., & Yamasari, Y. (2024). *Eksplorasi Fitur Seleksi pada SVM dan Random Forest dalam Analisis Sentimen Aplikasi GoPay*. *06*, 55–65.

Saputra, E. I., Anam, M. K., Yenni, H., Zamsuri, A., Studi, P., Informatika, T., Ilmu, F., Universitas, K., & Kuning, L. (2025). *OPTIMALISASI ALGORITMA SUPPORT VECTOR MACHINE PADA ASPECT-BASED SENTIMENT ANALYSIS MENGGUNAKAN GRIDSEARCHCV*. *7*(1), 271–278. https://doi.org/https://doi.org/10.31849/zn.v7i1.17800

Sari, P. K., & Suryono, R. R. (2024). *KOMPARASI ALGORITMA SUPPORT VECTOR MACHINE DAN RANDOM FOREST UNTUK ANALISIS SENTIMEN METAVERS*. *7*(1), 31–39.

Sidauruk, N. B., Riza, N., & Fatonah, R. N. S. (2023). *PENGGUNAAN METODE SVM DAN RANDOM FOREST UNTUK ANALISIS SENTIMEN ULASAN PENGGUNA TERHADAP KAI ACCES DI GOOGLE PLAYSTORE*. *7*(3), 1901–1906.

Singh, R. K., & Thomas, A. (2025). *A SYSTEMATIC LITERATURE REVIEW OF YOUTUBE COMMENTS SENTIMENT ANALYSIS : CHALLENGES AND EMERGING TRENDS*. *9292*(December), 947–961. https://doi.org/10.21917/ijdsml.2025.0184

Suardi, A., Chairat, N., Rizal, R., & Himawan, H. (2025). *Comparative Sentiment Analysis of YouTube Comments on Indonesia ' s Electric Vehicle Incentive Policy Using TF-IDF and IndoBERTweet Models*. *6*(6), 5988–6002.

Syafia, A. N., Hidayattullah, M. F., & Suteddy, W. (2023). *Studi Komparasi Algoritma SVM Dan Random Forest Pada Analisis Sentimen Komentar Youtube BTS*. *8*(3), 207–212.

Utama, K. C., & Yuliana, L. (2025). *Implementasi Pembaruan Sistem Inti Administrasi Perpajakan (Coretax) terhadap Efisiensi Kinerja Pegawai di Direktorat Jenderal Pajak Universitas Terbuka , Indonesia Universitas Paramadina , Indonesia Undang-Undang*.

Yutika, C. H., Adiwijaya, & Faraby, S. Al. (2021). *Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes*. *5*(April), 422–430. https://doi.org/10.30865/mib.v5i2.2845