Journal of Information Systems and Informatics
Vol. 7, No. 3, September 2025 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v7i3.1279 Published By DRPM-UBD

Sentiment Analysis on Coretax Data Using SVM and Random Forest with SMOTE and Tomek-Link

Hery Oktafiandi[1], Winarnie[2], M. Fajar Ramadhan[3], Febriyanti Panjaitan[4]

[1]Information System Department, Satu University, Palembang, Indonesia
[2,3,4]Computer Science Department, Satu University, Palembang, Indonesia
Email: [1]hery_oktafiandi@univ.satu.ac.id, [2]winarnie@univ.satu.ac.id, [3]m.ramadhan@univ.satu.ac.id, [4]pebrianti.panjaitan@univ.satu.ac.id

Abstract

This study is motivated by the increasing adoption of digital tax platforms in Indonesia, particularly Coretax, which enables online tax reporting and payment. Understanding user sentiment is crucial for evaluating system effectiveness and identifying areas for improvement. However, sentiment data is often imbalanced, making it challenging to detect the sentiments of the minority class. This research evaluates the performance of Support Vector Machine (SVM) and Random Forest (RF) in classifying sentiment from Coretax related reviews collected between March and September 2025 from Twitter, YouTube, and the DJP application. Lexicon-based labeling and preprocessing were applied, followed by class balancing using Tomek-Link, SMOTE, and SMOTE-Tomek techniques. On the original data, SVM achieved an accuracy of 98.56%, while Random Forest reached 98.43%, both performing strongly on the majority class. However, minority class detection was improved through SMOTE and SMOTE-Tomek, albeit with a slight decrease in overall accuracy due to the risk of overfitting. The novelty of this study lies in its focus on Coretax 2025 data and a comparative analysis of multiple resampling techniques, providing practical insights into improving sentiment analysis performance on imbalanced digital tax data.

Keywords: Coretax, Sentiment Analysis, SVM, Random Forest, Resampling Techniques

1. INTRODUCTION

Coretax is a digital tax reporting platform introduced to enhance the efficiency, transparency, and accessibility of tax administration in Indonesia [1], [2]. By enabling taxpayers to report and pay taxes online, Coretax aims to eliminate time-consuming manual procedures and improve tax compliance. However, in its implementation, the platform still encounters several obstacles, particularly in terms of technical stability and user adoption [3]. Common issues include system errors, connectivity problems, and limited access to infrastructure, which are often exacerbated by the varying levels of digital literacy across the population [4], [5]. These challenges have led to a wide range of user experiences, reflected in reviews shared through digital platforms. Sentiment analysis of such reviews provides

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

valuable insights into public perception; however, a significant challenge in this process is data imbalance, where certain sentiment classes, such as favorable reviews, are significantly underrepresented compared to others. This imbalance often leads to biased classification models that perform well on majority classes but fail to detect patterns in minority classes, resulting in reduced overall reliability[6], [7].

In response to this issue, various data resampling techniques have been developed to mitigate the impact of class imbalance in sentiment classification. Among the most commonly used methods are SMOTE (Synthetic Minority Over-sampling Technique), Tomek-Link, and a hybrid approach that combines both. SMOTE balances the dataset by generating synthetic samples for underrepresented classes; however, it can introduce noise and increase the risk of overfitting. On the other hand, Tomek-Link removes borderline or ambiguous samples near class boundaries, which helps reduce class overlap and enhance model generalization—although it may also eliminate valid data points. The SMOTE-Tomek hybrid method aims to combine the advantages of both techniques by simultaneously oversampling minority classes and refining the decision boundary, thereby achieving a better balance between precision and recall. This study implements all three resampling approaches—SMOTE, Tomek-Link, and SMOTE-Tomek—to evaluate their effectiveness in improving the performance of sentiment analysis models applied to Coretax user review data.

This research evaluates the performance of two machine learning algorithms: the Support Vector Machine (SVM) and the Random Forest (RF). SVM is known for its ability to handle high-dimensional data, while Random Forest offers advantages in ensemble learning and generalization across complex datasets [8], [9], [10]. Hyperparameter optimization using GridSearch is employed to fine-tune each model, ensuring optimal performance on the imbalanced dataset [11], [12]. By conducting a comparative analysis of these models and resampling techniques, this study contributes to improving sentiment classification accuracy for real-world applications in the public sector. The results are expected to inform digital tax administration policy and support the evaluation of a better user experience in future Coretax development.

To guide this investigation, the following research questions are proposed: (1) How do SVM and Random Forest perform in classifying sentiment on imbalanced Coretax review data? (2) What is the impact of applying Tomek-Link, SMOTE, and SMOTE-Tomek resampling methods on classification performance, particularly for the minority class? In response, the study sets the following objectives: (1) to evaluate and compare the performance of SVM and Random Forest on sentiment classification tasks using the Coretax dataset; (2) to assess the effectiveness of different resampling techniques in handling class imbalance; and

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

(3) to provide actionable insights for enhancing public service platforms through data-driven policy and sentiment analysis.

2. METHODS

This study [13],[14], [15] follows a methodological approach consisting of several main stages, ranging from data collection to model evaluation. This approach aligns with those used in previous studies on sentiment analysis. Figure 1 illustrates the overall research process, facilitating a clearer understanding of the steps undertaken.

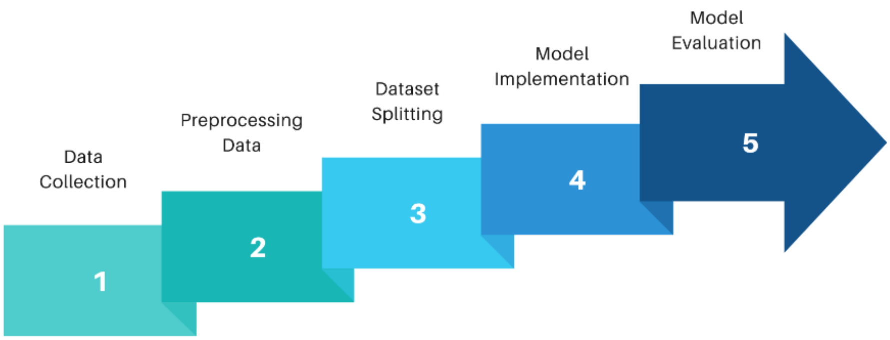


Figure 1. Methodological Approach

2.1. Data Collection

Data for this study were collected from three main sources: Twitter, YouTube, and the official Directorate General of Taxes (DJP) application available on Google Play Store, during the period from March to September 2025. Data collection was conducted using keywords relevant to user experiences with Coretax, such as "Coretax error," "Coretax tax report," and "DJP Online." This strategy allowed the retrieval of reviews reflecting various aspects of user experience, including technical complaints, ease of use, and satisfaction with the service. After the initial data collection and cleaning process, a total of 12,026 reviews were obtained for further analysis (see Table 1).

Tabel 1. Distribution of Coretax Reviews by Source

Data Source	Number of Reviews
Youtube	2913
Twitter	177
DJP Application	7936
Total	12.026

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

2.2. Preprocessing Data

The review data underwent preprocessing, including cleaning, tokenisation, case folding, stopword removal, stemming, and labelling. After cleaning, the dataset was reduced from 12,026 to 11,972 reviews, as 54 empty or invalid entries were removed (Figures 2 and Figure 3).

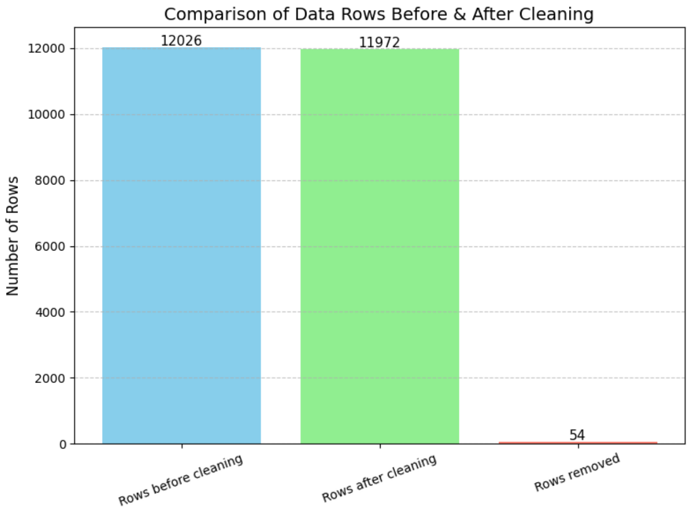


Figure 2. Comparison of Data Quantity Before and After Cleaning

	tweet
0	sudah september coretax masih sering begini ng...
1	min coretax ga bs diakses kah kenapa terimaka...
2	ndakikndakik fokus moneter dan fiskal tapi gak...
3	terima kasih atas partisipasi kawanpajak di ku...
4	halo min ini coretax lagi trouble kah mau pos...
...	...
11967	ok nyaman
11968	mantap aplikasi
11969	jos banget
11970	oke
11971	oc

11972 rows × 1 columns

Figure 3. Data View After Cleaning

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

The next process combined tokenisation, case folding, and stopword removal to produce more structured and consistent text. The data were then labelled into neutral, negative, and positive sentiments using a lexicon-based approach [16]. The labelling results revealed class imbalance, with neutral reviews dominating and positive reviews being the least frequent. Additional cleaning was performed for unreadable data during labelling. Out of 11,972 reviews, 160 entries failed to be labelled because they were empty, too short, or contained words not found in the lexicon dictionary. This resulted in a final dataset of 11,812 reviews ready for analysis (Figures 4 and Figure 5).

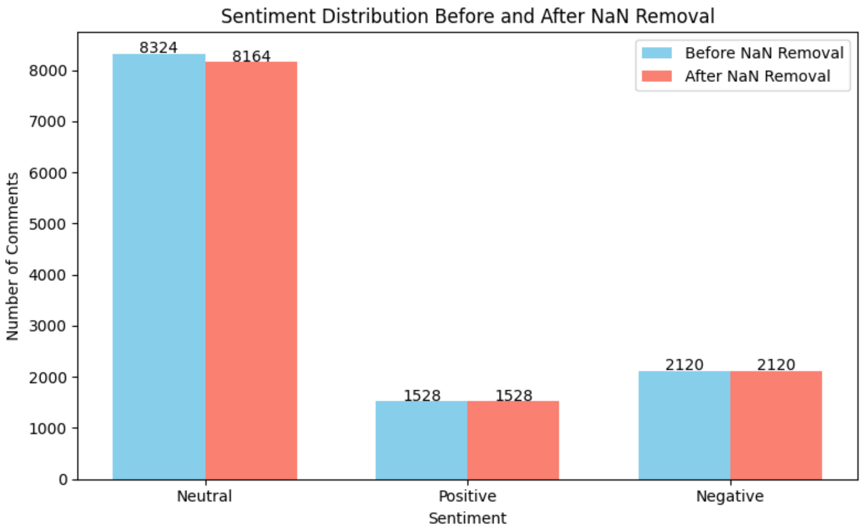


Figure 4. Sentiment Distribution Before and After Cleaning

	tweet	preprocessing	stemming_ulasan	label
0	sudah september coretax masih sering begini ng...	['september', 'coretax', 'ngeklik', 'gak', 'mu...	september coretax ngeklik gak muncul pilih paj...	positif
1	min coretax ga bs diakses kah kenapa terimaka...	['min', 'coretax', 'ga', 'bs', 'akses', 'kah',...	min coretax ga bs akses kah terimakasih	netral
2	ndakikndakik fokus moneter dan fiskal tapi gak...	['ndakikndakik', 'fokus', 'moneter', 'fiskal',...	ndakikndakik fokus moneter fiskal gak peduli b...	negatif
3	terima kasih atas partisipasi kawanpajak di ku...	['terima', 'kasih', 'partisipasi', 'kawanpajak...	terima kasih partisipasi kawanpajak kuis kemar...	positif
4	halo min ini coretax lagi trouble kah mau pos...	['halo', 'min', 'coretax', 'trouble', 'kah', '...	halo min coretax trouble kah posting spt gabisa	netral
...	...	...	...	...
11807	telkomsel roli	['telkomsel', 'rol']	telkomsel rol	netral
11808	ok nyaman	['ok', 'nyaman']	ok nyaman	positif
11809	mantap aplikasi	['mantap', 'aplikasi']	mantap aplikasi	positif
11810	jos banget	['jos']	jos	netral
11811	oc	['oc']	oc	netral

11812 rows × 4 columns

Figure 5. Data After Preprocessing and Labeling

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

2.3. Dataset Splitting

Next, the process of dataset splitting and handling data imbalance applied to the training data is described. The Coretax review dataset was divided into 80% for training and 20% for testing, allowing the model to learn from the majority of the data while retaining a representative test set for evaluation. Table 2 shows the number of data points per sentiment label before and after the split .

Table 2. Initial Data and Dataset Split

Sentiment Label	Total Data	Training Data	Test Data
Negative	2.120	1686	434
Neutral	8164	6547	1617
Positive	1528	1216	312

After splitting, the training data underwent imbalance handling using several sampling techniques, namely TOMEK, SMOTE, and the SMOTE-Tomek combination. Table 3 summarizes the label distribution in the training data after applying these techniques.

Table 3. Data Imbalance in the Training Set

Data Label	Negatif	Netral	Positif
Negative	2.120	1686	434
Neutral	8164	6547	1617
Positive	1528	1216	312

2.4. Model Implementation

In this study, SVM and Random Forest models were applied to classify sentiment in Coretax data, which includes user reviews regarding their experiences with the digital tax reporting system. The SVM model was chosen for its ability to handle high-dimensional data and optimally separate classes using a hyperplane. Meanwhile, Random Forest was used for its strength in combining predictions from multiple decision trees, enhancing classification stability and accuracy while reducing the risk of overfitting. Both models were evaluated to assess their capability in distinguishing negative, neutral, and positive sentiments, providing deeper insight into users' perceptions of the Coretax system. A literature review on the use of SVM and Random Forest methods is presented in Table 4.

Table 4. Literature Review on SVM and Random Forest Methods

Researchers and Year	Method	Challenges	Results
Ida Ayu Mirah Cahya	1) Random Forest 2) Neural Network 3) KNN	There is a dataset imbalance, with negative data	The SVM method demonstrated better accuracy performance,

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

at all. (2023) [17]	4) SVM	outnumbering positive data.	with an improvement of 0.48% after applying sampling, increasing from 80.528% to 81.017%.
Dhiaka shabrina Assyifa at all (2024) [10]	Random Forest	Testing data imbalance handling methods using Tomek Links, SMOTE, and SMOTE-Tomek with hyperparameter optimization.	Random Forest with SMOTE-Tomek, after hyperparameter optimization, improved by 1%, increasing from 84% to 85%.
Raniya Rakarahayu at all. (2024)[18]	SVM dan Naive Bayes	Considering and exploring classification methods from other machine learning approaches.	The accuracy of SVM is 82%, while Naive Bayes achieves 80%, indicating that SVM performs better than Naive Bayes.

Hyperparameter tuning was conducted using GridSearch to identify the optimal configuration that supports the best performance on imbalanced data. The hyperparameters used for both SVM and Random Forest are presented in Table 5. These configurations have also been utilized in previous studies aimed at improving classification accuracy, particularly when applying data balancing techniques.

Table 5. Hyperparameters for Random Forest and SVM

Model	**Hyperparameters**
Random Forest	1) 'n_estimators': [100, 200, 300] 2) max_depth: [None, 10, 20, 30] 3) min_samples_split: [2, 5, 10] 4) min_samples_leaf: [1, 2, 4] 5) max_features: ['auto', 'sqrt', 'log2']
SVM	1) 'C': [0.1, 1, 10] 2) kernel: ['linear', 'rbf'] 3) gamma: ['scale', 'auto']

The optimized models were then evaluated using the test dataset to assess their ability to classify negative, neutral, and positive sentiments, thereby providing deeper insights into user perceptions of the Coretax system.

2.5. Model Evaluation

After the training process, the performance of SVM and Random Forest models was evaluated using several metrics to measure classification capability comprehensively:

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

1) Accuracy, accuracy measures the proportion of correct predictions out of the total test data. This metric provides a general overview of how well the model classifies data, but does not always reflect performance for each class, especially in imbalanced datasets.
2) Precision, precision assesses the correctness of positive predictions, that is, how many of all predictions labelled as positive are actually positive. This metric is important to understand how often the model makes correct positive predictions, thereby reducing the risk of false positives.
3) Recall, recall measures the model's ability to detect all actual positive instances. In other words, it shows how well the model identifies all instances of the positive class. This metric is relevant for minimising false negatives, i.e., positive instances missed by the model.
4) F1-Score, the F1-Score is the harmonic mean of precision and recall, providing a balanced view of both metrics. It is particularly useful when class distribution is imbalanced, as it accounts for both positive and negative errors.

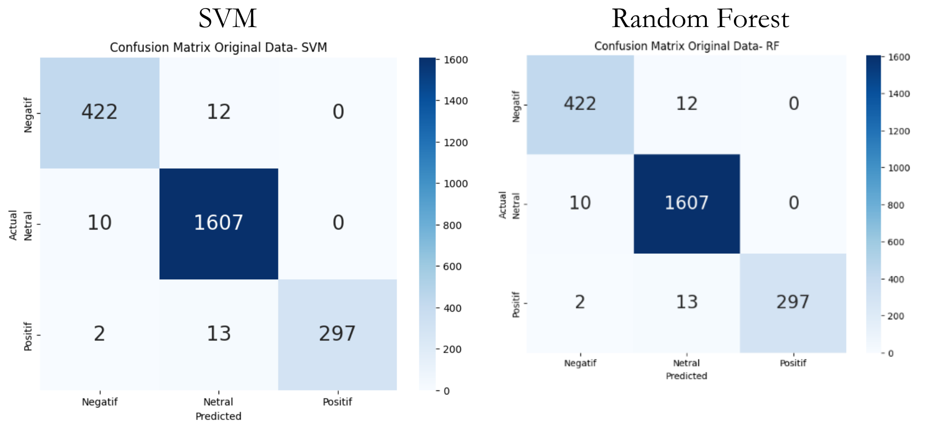
3. RESULTS AND DISCUSSION

3.1. Performance Evaluation

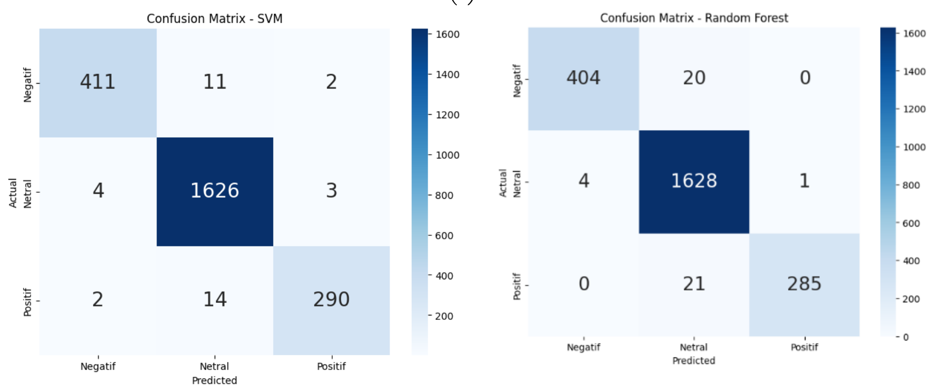
Based on performance metric evaluation, the classification results of SVM and Random Forest models on the Coretax dataset reveal interesting patterns regarding the models' ability to recognise each sentiment class. The analysis was conducted on the original data as well as after applying data balancing techniques, including Tomek-Link, SMOTE, and the SMOTE-Tomek combination, to understand the impact of resampling on accuracy and class-wise prediction distribution. Figure 6 shows the model performance evaluation. On the original data, both models (SVM and Random Forest) demonstrated excellent performance, particularly for the negative and neutral classes, with correct predictions nearly matching the total samples in each class. Prediction errors were minimal; for example, in SVM, 13 positive samples were misclassified as neutral.

Using Tomek-Link reduced the number of minority samples that were very close to other classes, causing a slight decrease in correct predictions for some classes, such as negative and positive in SVM. Still, the model became clearer in distinguishing class boundaries. With SMOTE, all classes were balanced by adding synthetic samples. This led to increased errors in minority classes (positive and neutral), as some synthetic data could confuse the model; for instance, in SVM, 162 neutral samples were misclassified. The SMOTE-Tomek combination slightly reduced errors compared to SMOTE alone. For example, in Random Forest for the positive class, correct predictions only dropped from 279 to 277, while class distribution became more balanced and class overlap was reduced.

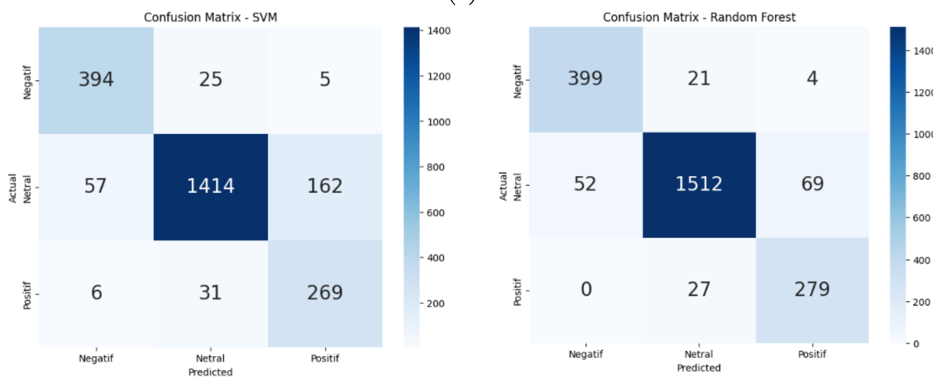
p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882



(a)

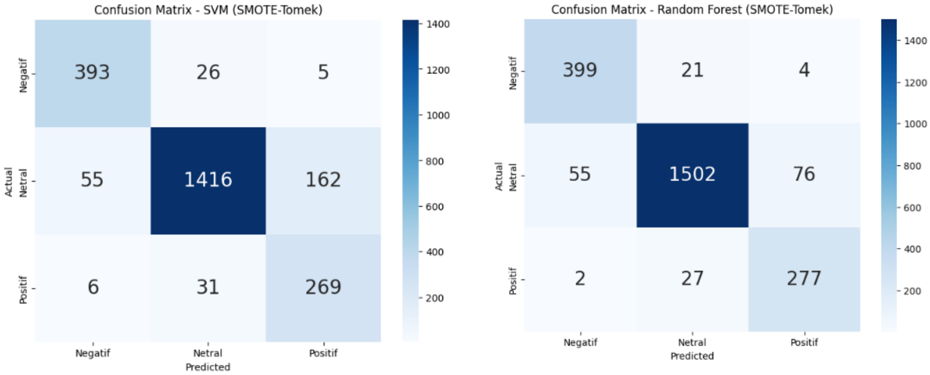


(b)



(c)

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882



(d)

Figure 6. Confusion matrices of SVM and Random Forest classification results on the Coretax dataset before and after applying resampling techniques: (a) Original, (b) Tomek-Link, (c) SMOTE, (d) SMOTE-Tomek.

The data in Table 6 shows the performance of both models on the original data (Original Data). SVM achieved an accuracy of 98.56%, precision of 98.74%, recall of 97.42%, and F1-score of 98.06%, while Random Forest recorded an accuracy of 98.43%, precision of 98.57%, recall of 97.27%, and F1-score of 97.90%. These results indicate that both models can classify the majority class, particularly the "neutral" class with the largest number of samples, very well. However, there were a small number of errors in minority classes, such as "positive" and "negative," as seen in each model's confusion matrix, where some samples were misclassified. The application of SMOTE increased the number of samples in minority classes through synthetic data augmentation. As a result, SVM accuracy decreased to 87.89%, while Random Forest reached 92.67%. Although overall accuracy declined, recall for minority classes improved, indicating that the models were better able to capture minority samples, even though precision slightly decreased due to potential overfitting on synthetic data. The combination of SMOTE and Tomek-Link (SMOTE-Tomek) produced performance nearly equivalent to SMOTE alone, with minimal differences in metrics, suggesting that this approach of oversampling and minor sample cleaning is relatively stable in maintaining a balance between precision and recall.

Table 6. Summary of Evaluation Metric Scores

Model	Sampling Technique	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	Original	98.56	98.74	97.42	98.06
	Tomek-Link	98.47	98.45	97.09	97.75
	SMOTE	87.89	81.36	89.14	84.36
	SMOT-Tomek	87.93	81.46	89.10	84.40
	Original	98.43	98.57	97.27	97.90

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Model	Sampling Technique	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	Tomek-Link	98.05	98.73	96.03	97.33
	SMOTE	92.67	88.21	92.62	90.23
	SMOT-Tomek	92.17	87.33	92.20	89.53

Several previous studies have found that the application of SMOTE-Tomek often results in better performance balance compared to using SMOTE or undersampling alone. This improvement is evident in the increased accuracy and the model's enhanced ability to handle minority class data, particularly in cases of imbalanced sentiment classification. A more detailed comparison of the studies is presented in Table 7.

Table 7. Comparison of Related Studies

Researchers and Year	Method	Challenges	Results
Ida Ayu Mirah Cahya at all. (2023) [17]	5) Random Forest 6) Neural Network 7) KNN 8) SVM	There is a dataset imbalance, with negative data outnumbering positive data.	The SVM method demonstrated better accuracy performance, with an improvement of 0.48% after applying sampling, increasing from 80.528% to 81.017%.
Dhiaka shabrina Assyifa at all (2024) [10]	Random Forest	Testing data imbalance handling methods using Tomek Links, SMOTE, and SMOTE-Tomek with hyperparameter optimization.	Random Forest with SMOTE-Tomek, after hyperparameter optimization, improved by 1%, increasing from 84% to 85%.
Raniya Rakarahayu at all. (2024)[18]	SVM dan Naive Bayes	Considering and exploring classification methods from other machine learning approaches.	The accuracy of SVM is 82%, while Naive Bayes achieves 80%, indicating that SVM performs better than Naive Bayes.
This Study (2025)	SVM and Random Forest with Tomek-Link, SMOTE, SMOTE-Tomek	Handling severe sentiment imbalance in Coretax reviews	SMOTE-Tomek balanced performance best. SVM: 87.93% accuracy; RF: 92.17% after resampling.

This study has several limitations. The dataset was restricted to user reviews collected from Twitter, YouTube, and the official DJP application, which may not fully represent the complete user experience with Coretax. Furthermore, the

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

application of resampling techniques, such as SMOTE, may introduce overfitting, particularly in minority classes, which can potentially affect model precision. The analysis was also limited to two machine learning algorithms, SVM and Random Forest, without exploring other potentially effective models such as Neural Networks or XGBoost. Additionally, the study focuses on the period from March to September 2025, meaning the findings may vary if applied to different time frames or if there are significant changes in the Coretax system. Despite these limitations, the results offer valuable insights for improving Coretax services. A balanced sentiment classification model can help platform managers detect minority complaints (whether positive or negative) that are often overlooked. Techniques like SMOTE-Tomek improve the model's sensitivity to these sentiments without significantly compromising accuracy for majority classes, making them practical for integration. These findings support the development of automated sentiment monitoring tools, alert systems for sudden surges in negative sentiment, and data-driven recommendations for service policy adjustments. Moreover, the methodology presented can be extended to other digital tax platforms to create more responsive and user-centered services.

3.2. Discussion

The findings of this study highlight the significant role of resampling techniques in addressing the class imbalance problem commonly encountered in sentiment analysis tasks, particularly in the context of Coretax user reviews. The results underscore how different resampling approaches—Tomek-Link, SMOTE, and the SMOTE-Tomek combination—affect the performance of two widely-used machine learning algorithms, SVM and Random Forest, in classifying sentiment.

Both SVM and Random Forest models demonstrated strong overall performance, especially when applied to the original dataset. With an accuracy of 98.56% for SVM and 98.43% for Random Forest, it is evident that both models effectively classified the neutral class, which was the dominant sentiment in the dataset. However, both models showed minor misclassification errors, particularly with the minority classes—positive and negative reviews. These misclassifications were mostly observed in the neutral sentiment, where positive reviews were often misclassified as neutral, suggesting some difficulty in distinguishing between these closely related sentiment categories.

The introduction of resampling techniques, especially SMOTE, resulted in an increase in recall for the minority classes (positive and negative sentiments), as shown by the improvement in recall metrics for SVM and Random Forest models. However, this came at the cost of a decrease in overall accuracy, indicating a trade-off between improving minority class recognition and maintaining accuracy for the majority class. This pattern is typical in imbalanced datasets, where the model may

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

overfit to the synthetic samples generated by SMOTE, leading to a slight decline in precision.

When applying the Tomek-Link technique, the model showed a slight reduction in misclassification errors, particularly by removing borderline samples that were close to other classes. While this improved the model's ability to distinguish between classes, it also led to some loss of valid minority samples, particularly for the positive and negative reviews. The application of SMOTE, which synthesizes new samples for the minority classes, resulted in noticeable performance changes. While it improved recall for the positive and negative classes, it also introduced synthetic samples that sometimes confused the model, as evidenced by the misclassification of neutral reviews. The addition of synthetic samples may lead to overfitting, especially for the SVM model, which struggled to generalize well on the augmented data, resulting in lower accuracy.

The SMOTE-Tomek hybrid method, which combines both oversampling and cleaning techniques, demonstrated balanced performance. It managed to improve recall for minority classes while minimizing the overfitting risk compared to SMOTE alone. For instance, in the Random Forest model, while accuracy slightly dropped from 92.67% to 92.17%, the recall for the minority class remained robust, and class overlap was effectively reduced. These results suggest that SMOTE-Tomek is a promising method for handling imbalanced datasets without sacrificing too much accuracy, making it an ideal candidate for improving sentiment analysis models in imbalanced sentiment scenarios like those found in Coretax reviews.

The comparison of this study's results with related studies further reinforces the efficacy of resampling techniques in imbalanced sentiment classification. Previous work has shown that SMOTE-Tomek outperforms other resampling methods, such as SMOTE alone or undersampling, in terms of achieving better performance balance. For example, Dhiaka Shabrina Assyifa et al. (2024) reported an improvement in Random Forest performance by 1% after applying SMOTE-Tomek with hyperparameter optimization [10]. Similarly, studies by Ida Ayu Mirah Cahya (2023) and Raniya Rakarahayu (2024) highlighted SVM's improved accuracy after applying sampling techniques, reinforcing the reliability of resampling methods like SMOTE-Tomek in handling dataset imbalance in sentiment analysis.

The results of this study have significant implications for the future development of the Coretax platform and similar digital services. By leveraging the insights gained from this analysis, Coretax can better understand the user sentiment distribution, especially in cases where minority class sentiments (whether positive or negative) might otherwise be overlooked. The use of SMOTE-Tomek as a resampling strategy can be effectively integrated into sentiment analysis models to detect these minority sentiments more accurately without compromising the

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

detection of neutral sentiments, which dominate most user feedback. Moreover, these findings can help in the creation of automated sentiment monitoring tools, which could alert platform administrators to sudden surges in negative or positive sentiment. This could facilitate a more proactive response to emerging issues or positive feedback, helping to improve user engagement and satisfaction with the platform. Additionally, the insights from sentiment analysis can inform service policy adjustments, allowing Coretax to address user concerns more effectively and enhance its digital tax administration experience.

While the study offers valuable insights, it also has several limitations. The dataset was limited to reviews from three primary sources: Twitter, YouTube, and the official Directorate General of Taxes (DJP) application. This restriction may result in an incomplete representation of the overall user experience with Coretax, especially if significant user feedback is found outside these channels. Additionally, the application of resampling techniques, such as SMOTE, might lead to overfitting on synthetic data, which could slightly affect the model's performance when generalized to real-world data. Another limitation is the exclusive focus on SVM and Random Forest models, leaving other advanced models like Neural Networks or XGBoost unexamined. Future research should explore these models and assess their performance in improving sentiment classification for Coretax reviews, as well as consider a broader set of data sources to better capture the full spectrum of user experiences. Moreover, the temporal limitation of the dataset—spanning only from March to September 2025—raises the possibility that changes in Coretax's features, user interface, or policies during or after this period may affect sentiment distributions. Therefore, applying the model to data from subsequent periods would be essential to evaluate the robustness and adaptability of the sentiment classification system over time.

4. CONCLUSION

Based on the model performance evaluation, both SVM and Random Forest demonstrated strong capability in classifying the majority sentiment class (neutral), achieving over 98% accuracy on the original, imbalanced dataset. The application of resampling techniques, particularly SMOTE and SMOTE-Tomek, enhanced the models' ability to recognize minority classes (positive and negative), though SMOTE alone slightly reduced overall accuracy due to the risk of overfitting on synthetic data. In contrast, the SMOTE-Tomek combination offered more balanced performance by improving recall for minority classes while maintaining precision and reducing class overlap. Random Forest showed slightly greater stability than SVM after resampling, though both performed competitively. These findings suggest that careful selection of resampling techniques can significantly improve model robustness in real-world sentiment analysis tasks. For platform administrators, such balanced sentiment models can support more effective

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

monitoring of user feedback, enabling timely responses to negative sentiment trends and guiding data-driven service improvements. Future research should consider exploring more advanced classification models, such as Neural Networks or XGBoost, and expanding datasets to include more diverse sources and timeframes for broader applicability and deeper insights.

REFERENCES

[1] M. R. Panjaitan and Y. Yuna, "Pengaruh Coretax terhadap Transparansi dan Akuntabilitas Sistem Perpajakan," *Jurnal Riset Akuntansi*, vol. 2, no. 4, pp. 51–60, 2024.

[2] G. Dimetheo, A. Salsabila, and N. C. A. Izaak, "Implementasi Core Tax Administration System sebagai Upaya Mendorong Kepatuhan Pajak di Indonesia," in *Prosiding Seminar Nasional Ekonomi dan Perpajakan*, 2023, pp. 10–25.

[3] M. A. Al Maliki, "Studi Literatur: Analisis Penerapan Aplikasi CoreTax dalam Sistem Perpajakan," *EKOMA: Jurnal Ekonomi, Manajemen, Akuntansi*, vol. 4, no. 3, pp. 5132–5140, 2025.

[4] N. Cindy and C. Chelsya, "Persepsi Mahasiswa Terhadap Penerapan Core Tax Administration System (CTAS) di Indonesia," *Economics and Digital Business Review*, vol. 5, no. 2, pp. 1029–1040, 2024.

[5] N. K. J. S. Mayoni, "Pengaruh Persepsi Kegunaan, Persepsi Kemudahan, dan Persepsi Risiko terhadap Niat Untuk Menggunakan Coretax dalam Sistem Administrasi Pajak di Kabupaten Badung," 2024, *Politeknik Negeri Bali*.

[6] M. A. Pryono, S. H. Wijoyo, and F. A. Bachtiar, "Analisis Sentimen Terhadap Program Merdeka Belajar Kampus Merdeka Pada Sosial Media Twitter Menggunakan K-Means Clustering, Support Vector Machine (SVM) dan Syntethic Minority Oversampling Technique (SMOTE)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 9, 2024.

[7] N. A. Riskiwibowo, "Analisis Emosi Terhadap Komentar Video Youtube 'Penyebab Kegagalan Adopsi Sistem Pendidikan Finlandia di Indonesia' Menggunakan Metode Random Forest," in *Proceedings of the National Conference on Electrical Engineering, Informatics, Industrial Technology, and Creative Media*, 2024, pp. 812–827.

[8] M. R. Saputra and P. Parjito, "Analisis Sentimen Twitter Terhadap Konflik di Papua Menggunakan Perbandingan Naive Bayes dan SVM," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, pp. 1197–1208, 2025.

[9] S. P. Tanzil, "Analisis Sentimen Pengguna Instagram terhadap Timnas Indonesia U-23 pada Piala AFC menggunakan Algoritma K-Nearest Neighbor (K-NN) dengan SMOTE," *Telekontran: Jurnal Ilmiah*

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Telekomunikasi, Kendali dan Elektronika Terapan, vol. 12, no. 1, pp. 68–80, 2024.

[10] D. Shabrina Assyifa and A. Luthfiarta, "SMOTE-Tomek Re-sampling Based on Random Forest Method to Overcome Unbalanced Data for Multi-class Classification," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 9, no. 2, pp. 151–160, 2024, doi: 10.25139/inform.v9i2.8410.

[11] H. Oktafiandi, W. Winarnie, and S. M. R. Olajuwon, "Perbandingan Algoritma untuk Analisis Sentimen Terhadap Google Play Store Menggunakan Machine Learning," *Jurnal Ekonomi Dan Teknik Informatika*, vol. 11, no. 2, pp. 16–21, 2023.

[12] N. A. Haqimi and T. A. Roshinta, "Analisis Spam Komentar Instagram menggunakan Support Vector Machine dengan Variasi Hyperparameter," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 9, no. 3, pp. 242–253, 2024.

[13] F. Panjaitan, W. Ce, H. Oktafiandi, G. Kanugrahan, Y. Ramdhani, and V. H. C. Putra, "Evaluation of Machine Learning Models for Sentiment Analysis in the South Sumatra Governor Election Using Data Balancing Techniques," *Journal of Information Systems and Informatics*, vol. 7, no. 1, pp. 461–478, Mar. 2025, doi: 10.51519/journalisi.v7i1.1019.

[14] B. M. Iqbal, K. M. Lhaksmana, and E. B. Setiawan, "2024 Presidential Election Sentiment Analysis in News Media Using Support Vector Machine," *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 2, pp. 397–404, 2023.

[15] U. R. H. Baba, "Analisa Sentimen Menjelang Komparasi Analisis Sentimen Masyarakat Terhadap Isu Penundaan Pemilu 2024 Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine," *Innovative: Journal Of Social Science Research*, vol. 4, no. 3, pp. 11972–11990, 2024.

[16] R. H. Muhammadi, T. G. Laksana, and A. B. Arifa, "Combination of support vector machine and lexicon-based algorithm in twitter sentiment analysis," *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, vol. 8, no. 1, pp. 59–71, 2022.

[17] Ida Ayu Mirah Cahya Dewi, I Komang Dharmendra, and Ni Wayan Setiasih, "Analisis Sentimen Review Aplikasi Satu Sehat Mobile Menggunakan Model Sampling Tomek Links," *Jurnal Teknologi Informasi dan Komputer*, vol. 9, no. 5, pp. 497–504, 2023, doi: 10.36002/jutik.v9i5.2644.

[18] R. Rakarahayu Putri and N. Cahyono, "Analisis Sentimen Komentar Masyarakat Terhadap Pelayanan Publik Pemerintah Dki Jakarta Dengan Algoritma Super Vector Machine Dan Naive Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 2363–2371, 2024, doi: 10.36040/jati.v8i2.9472.