JIKO (Jurnal Informatika dan Komputer) Accredited KEMDIKTISAINTEK, No.0173/C3/DT.05.00/2025
Vol. 9, No. 2, August 2026, pp. 100-107 p-ISSN: 2614-8897
DOI: 10.33387/jiko.v9i2.11615 e-ISSN: 2656-1948

COMPARISON OF LINEARSVC AND COMPLEMENT NAIVE BAYES ON CORETAX SENTIMENT ANALYSIS USING INDOBERT PSEUDO-LABELING AND ADASYN

Riza Febyana Shollis[1*], Ucta Pradema Sanjaya[2]

[1,2] Bachelor of Informatics Engineering, Faculty of Computer and Education, Universitas Ngudi Waluyo, Semarang Regency, Central Java 50512, Indonesia
Email: [*1] azariza321@gmail.com, [2] uctapradema@unw.ac.id

(Received: 14 March 2026, Revised: 30 May 2026, Accepted: 11 June 2026)

Abstract

Coretax is an Indonesian digital tax service platform that has attracted extensive public discussion on social media, particularly on X/Twitter. Most discussions are related to system stability and user experience when accessing the platform. This study collected sentiment data from December 1, 2024, to December 20, 2024, resulting in 10,140 tweets, which were filtered into 7,900 valid data points. The dataset underwent standard text preprocessing and was represented using TF-IDF features. Sentiment labeling was performed automatically (pseudo-labeling) using the IndoBERT model, producing an imbalanced class distribution (Negative 63.20%; Neutral 30.24%; Positive 6.56%). To address this imbalance, ADASYN was applied to the training data, and two classification models were compared: Complement Naïve Bayes and SVM (LinearSVC). Evaluation using an 80:20 train–test split showed that LinearSVC + ADASYN achieved the best performance with an accuracy of 85.25%, F1-Macro of 0.7423, MCC of 0.7120, ROC-AUC of 0.9428, and Hamming Loss of 0.1475, outperforming ComplementNB + ADASYN with an accuracy of 78.67%. Furthermore, the McNemar test confirmed that the performance difference between the two models is statistically significant. These findings indicate that LinearSVC is more effective in distinguishing sentiments related to technical complaints and procedural inquiries in the context of digital tax services.

Keywords: *Coretax, ADASYN, Naive Bayes, SVM, IndoBERT.*

**Corresponding Author: Riza Febyana Shollis*

1. INTRODUCTION

The development of digital services in the public administration sector requires information systems capable of providing accurate services, easy accessibility, and responsiveness to public needs. In the context of taxation, Coretax has become an important platform used by taxpayers to manage various tax-related activities online. The quality of services on this platform is not only determined by technical performance, but also by how users perceive their experience when interacting with the system. Understanding these perceptions has become increasingly important, considering that public opinion is now largely formed and expressed through social media platforms [1], [2].

The platform X (formerly Twitter) is a highly active digital conversation space that is often used by users to express complaints, opinions, or evaluations regarding a service. The characteristics of posts that are short, spontaneous, and rich in expression make this platform a potential data source for understanding the dynamics of public perception. Various issues related to government digital services, including tax services, often develop and spread rapidly through user interactions on this platform. This condition creates an opportunity to conduct data-driven observations of user responses toward the Coretax website [3]. Sentiment analysis has also been widely applied in evaluating public responses toward e-government services and digital public platforms [20].

The implementation of the Coretax system represents an effort to modernize tax administration in Indonesia, replacing SIDJP and several services previously available through DJP Online. However, this system still faces stability issues and operational challenges due to simultaneous connection loads when accessed by taxpayers. Problems frequently occur

during registration, profile completion, payment processing, SPT submission, e-Bupot, and e-Invoice services, which sometimes experience disruptions and even 404 errors. These issues potentially affect user satisfaction and public trust toward digital tax services.

Sentiment analysis has become a major approach in analyzing public opinion due to its ability to process large amounts of textual data and extract emotional tendencies from user-generated content. In the field of sentiment analysis research, Naïve Bayes remains one of the most widely used algorithms because of its simplicity, low computational requirements, and stable performance on high-dimensional text data [4], [19]. In addition, several studies indicate that Naïve Bayes produces competitive results for Indonesian-language text data, including informal and diverse data commonly found in social media conversations [5].

The success of classification algorithms in sentiment analysis is strongly influenced by the feature representation method used. Term Frequency–Inverse Document Frequency (TF-IDF) weighting is a commonly used approach for constructing text vector representations because it assigns higher weights to relevant words while reducing the influence of common words. Recent studies demonstrate that the combination of TF-IDF and machine learning algorithms produces stable and accurate performance in various sentiment analysis tasks, including digital service reviews, public issues, and technology-related topics [6], [7].

Several previous studies have implemented sentiment analysis in different domains using machine learning approaches. Kusuma and Cahyono analyzed user sentiment toward the Shopee application on the Google Play Store using the K-Nearest Neighbor (KNN) algorithm combined with TF-IDF feature extraction [9]. Another study by Putri et al. conducted sentiment analysis on local skincare brands on Twitter using the Naïve Bayes classifier and TF-IDF representation [10]. Furthermore, Umaidah and Maulana analyzed user sentiment on Twitter toward BTS using Support Vector Machine (SVM) with TF-IDF-based features [11]. These studies demonstrate that machine learning algorithms are effective for classifying Indonesian-language sentiment data from social media platforms.

However, previous studies generally focused on e-commerce, entertainment, or consumer product domains, while sentiment analysis in the context of Indonesian digital tax services remains limited. In addition, most previous studies relied on manually labeled datasets and did not specifically address the challenge of imbalanced sentiment classes in datasets containing technical taxation terminology such as e-Bupot, NPWP, SPT, and e-Invoice. The application of IndoBERT-based pseudo-labeling for Indonesian tax-related sentiment datasets is also still rarely explored.

Based on these research gaps, this study aims to analyze public sentiment toward the Coretax platform on X/Twitter by comparing the performance of Complement Naïve Bayes and Support Vector Machine (LinearSVC) models using TF-IDF feature representation. This study also utilizes IndoBERT for pseudo-labeling and applies the ADASYN oversampling technique to address class imbalance problems in the dataset. The results of this study are expected to provide insights into public perceptions of digital tax services and contribute to the development of more responsive and reliable tax information systems.

2. RESEARCH METHOD

This study was designed to obtain an empirical description of user sentiment toward the Coretax website on the X platform through a supervised machine learning approach using the Naïve Bayes algorithm combined with TF-IDF (Term Frequency–Inverse Document Frequency) weighting. The selection of the TF-IDF and Naïve Bayes combination is based on the effectiveness and efficiency of this method in processing Indonesian-language text, including informal text such as application reviews and opinions expressed on social media [12]. In addition, Naïve Bayes is a fundamental algorithm known for its performance stability, particularly when applied to large and high-dimensional datasets, making it highly suitable for sentiment analysis on digital platforms that generate large volumes of data [13].

The research process was conducted through several main stages, starting from data collection, text preprocessing, feature extraction using TF-IDF, classification model development, and model performance evaluation. Data collection was carried out by extracting public posts and comments on the X platform using keywords relevant to Coretax services within a specific time period [14]. The collected data are unstructured and heterogeneous, therefore they must be prepared through preprocessing before further analysis can be conducted [5].

The preprocessing stage aims to improve the quality of text representation and reduce irrelevant feature dimensions. This stage includes text cleaning to remove irrelevant characters and symbols, normalization to standardize writing formats, tokenization to split text into word units, stopword removal, and stemming to convert words into their root forms [15]. Through these steps, raw data are transformed into a more homogeneous and meaningful representation, enabling the Naïve Bayes algorithm to capture word distribution patterns more accurately [16].

After preprocessing is completed, the text data are converted into numerical representations using the TF-IDF method, which assigns weights to words based on their frequency of occurrence in a document relative to the entire corpus. This representation allows the model to highlight the most informative words for determining user sentiment while minimizing the influence of common and less relevant words. This

numerical representation then becomes the input for the Naïve Bayes model, which is developed through a supervised learning approach using labeled data to classify sentiment into three categories: positive, negative, and neutral.

Model performance evaluation is conducted using accuracy, precision, recall, and F1-score metrics, in accordance with best practices in recent sentiment analysis research [17]. The use of multiple evaluation metrics allows for a comprehensive analysis of model performance. Accuracy measures the overall correctness of predictions, precision evaluates the proportion of correct predictions within each class, recall measures the model's ability to capture all instances of each class, and the F1-score provides a balance between precision and recall. This evaluation ensures that the model not only produces accurate predictions but also remains consistent when dealing with class imbalance, which commonly occurs in social media data.

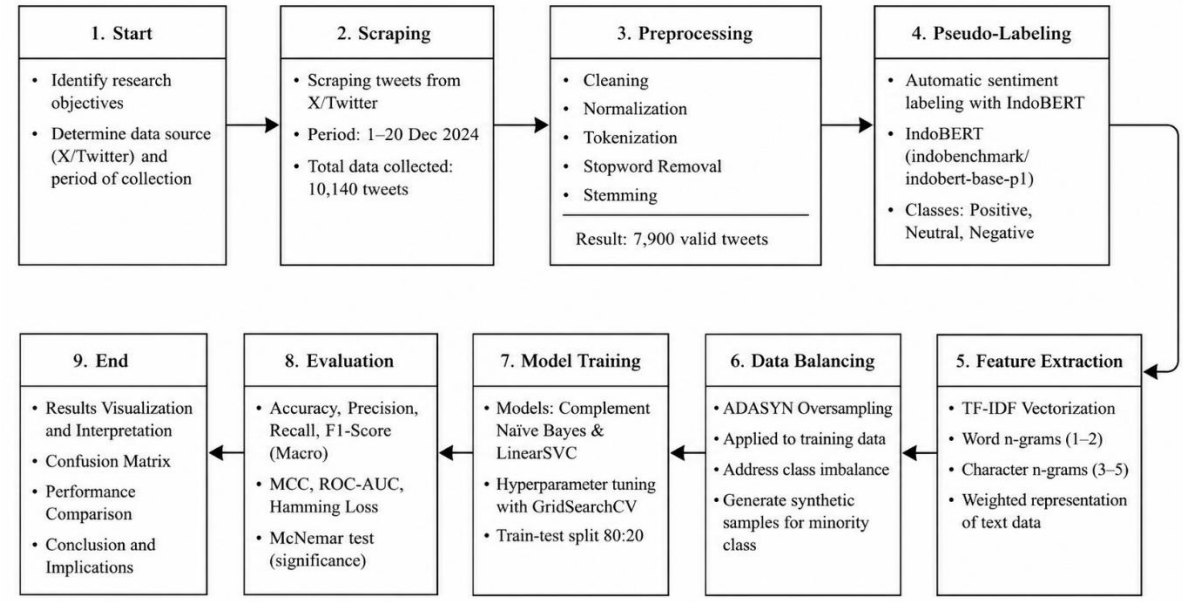

Figure 1. Sentiment Analysis Workflow

Figure 1 presents the overall research methodology flowchart, starting from data collection to model performance evaluation. This research workflow is systematically arranged to ensure that each stage contributes optimally to the research objectives. Representative data collection, comprehensive text preprocessing, accurate feature extraction, structured model development, and thorough performance evaluation form an integrated framework. Through this methodology, the study not only produces an effective sentiment classification model but also provides empirical insights that can be used to develop services and communication strategies that are more responsive to public opinion.

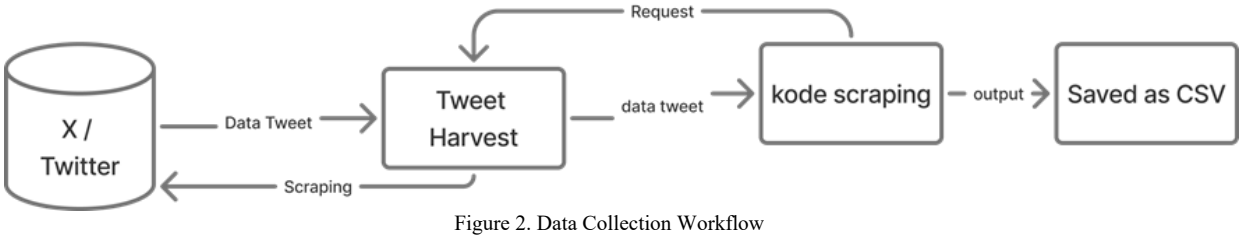

Figure 2. Data Collection Workflow

Data collection for this study was conducted by gathering tweets from the X/Twitter platform using a free tool called tweet-harvest and Python code executed in Google Colab. The data collection process began by running Python code to prepare the environment so that tweet-harvest could operate properly. After that, tweet-harvest performed data extraction according to the specified query parameters.

The keywords used were "coretax", "coretax djp", "coretax pajak", and "coretax error", with a time range from December 1, 2024, to December 20, 2025. From this process, 10,140 tweets were collected and used as the dataset for sentiment analysis.

3. RESULT AND DISCUSSION

3.1 Data Preprocessing

1. The initial dataset consisted of 10,140 records obtained from the data collection process. The available information included message text, date, user ID, number of replies, number of retweets, and other related metadata.
2. The data then underwent ID filtering and full-text column filtering to ensure that the sentiment data used in the study were written in the Indonesian language.

Table 1. Total Data

Label	Amount of Data
Initial scraping results	10,140
Filter lang:ID	9,065
Filter full text column	7,900

3. The labeling process was performed using IndoBERT through a pretrained model from Hugging Face, namely "mdhugol/indonesian-roberta-base-sentiment-classifier", to generate pseudo-labels.
4. Data Splitting and Baseline Model Setup: The dataset was divided into 80% training data and 20% testing data to objectively evaluate model performance.
5. The sentiment distribution results from Linear SVC, IndoBERT, and Naïve Bayes were visualized using bar charts. Additionally, word clouds were generated for each sentiment class (Positive, Negative, and Neutral) to observe the most frequently used words and identify keyword patterns that contribute to each sentiment category.

Stage	Text
Original Text	@DitjenPajakRI min mau tanya. Saya mau buat npwp. Tapi di https://t.co/L4C0vukcRh saat mau daftar pengguna baru NIK sudah terdaftar. Dan saat aktivasi akun wajib pajak selalu ada error saat memasukan email dan nomor telepon. Lalu saat di submit juga error (Terjadi kesalahan saat permintaan)??
Cleaned	min mau tanya saya mau buat npwp tapi di saat mau daftar pengguna baru nik sudah terdaftar dan saat aktivasi akun wajib pajak selalu ada error saat memasukan email dan nomor telepon lalu saat di submit juga error terjadi kesalahan saat permintaan
Normalized	admin mau tanya saya mau buat npwp tapi di saat mau daftar pengguna baru nik sudah terdaftar dan saat aktivasi akun wajib pajak selalu ada error saat memasukkan email dan nomor telepon lalu saat submit juga error terjadi kesalahan saat permintaan
Stemmed & Stopword	tanya buat npwp daftar pengguna baru nik daftar aktivasi akun pajak error masuk email nomor telepon submit error kesalahan permintaan

3.2 Labeling Using IndoBERT and Handling Imbalanced Data

This study used automatic labeling with the IndoBERT model (*mdhugol/indonesian-roberta-base-sentiment-classifier*). This model was selected due to its strong capability in processing Indonesian-language data. Given the large volume of sentiment data, labeling could be performed efficiently using this model.

The ADASYN technique was applied only to the training data to prevent data leakage.

Table 2. IndoBERT Labeling Results

Label	Total	Percentage
Negative	4,993	63.20%
Neutral	2,389	30.24%
Positive	518	6.56%

The dominance of the Negative class (63.2%) and the scarcity of the Positive class (6.5%) indicate that the dataset is highly imbalanced. Therefore, a data balancing technique (ADASYN) was applied during the model training stage.

Before and After Applying ADASYN:

- Before ADASYN:
  The training data consisted of 6,320 records with an uneven class distribution:
  - Neutral: 3,994
  - Negative: 1,911
  - Positive: 415
- After ADASYN:
  The training dataset increased to 11,797 records with a more balanced distribution:
  - Neutral: 3,994
  - Positive: 3,984
  - Negative: 3,819

The implementation of ADASYN proved effective in balancing the proportion of the minority class (Positive), thereby preventing the model from being biased toward the majority class.

The dataset was divided using a stratified train–test split with a ratio of 80:20. There were 6,320 training samples before ADASYN and 1,580 testing samples. After applying ADASYN, the number of synthetic samples increased to 11,797 training samples, with a balanced class distribution (Neutral: 3,994; Positive: 3,984; Negative: 3,819), enabling the model to learn patterns from the minority class (Positive) more effectively.

3.3 Naïve Bayes & Svm (Linear Svc) Classification Results

1. Complement Naïve Bayes Model

This model uses TF-IDF word n-grams (1,2) combined with ADASYN oversampling. The performance results of this model are:

- Accuracy: 0.7867 (78.67%)
- F1-Macro: 0.6833
- MCC: 0.6412
- ROC-AUC Macro (OvR): 0.9303
- Hamming Loss: 0.1475

The detailed performance for each class is presented below:

Table 3. Naïve Bayes Classification Results

Class	Precision	Recall	F1-score
Negative	0.730	0.822	0.774
Neutral	0.976	0.780	0.867
Positive	0.291	0.689	0.409

The Neutral class has very high precision (0.976), meaning that when the model predicts "Neutral," the prediction is mostly correct. However, the recall for the Neutral class is only 0.780, indicating that many Neutral instances are incorrectly classified into other classes.

The Positive class shows a significant issue with precision (0.291). Although the recall for Positive is relatively high (0.689), the low precision indicates that the model frequently predicts Positive when the actual sentiment is not Positive.

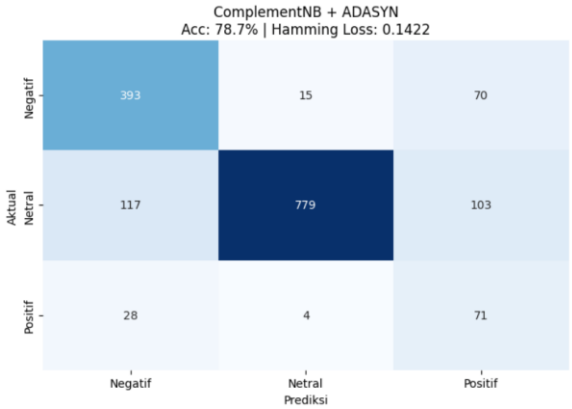

Figure 3. Naïve Bayes Confusion Matrix

Based on the confusion matrix results, the Complement Naïve Bayes model with ADASYN demonstrates fairly good performance with an accuracy of 78.67%, correctly predicting 1,243 out of 1,580 test samples. Most correct predictions occur in the majority classes, namely Neutral (779 data) and Negative (393 data). However, misclassifications still occur due to word ambiguity and similarities in sentence patterns between technical reports and complaint-related messages.

The main challenge lies in the Positive class. Although the recall rate is relatively high (0.689), the low precision (0.291) indicates that the model often classifies data as Positive when it is not actually Positive (false positives), making Positive predictions less reliable.

2. SVM Model (Linear SVC)

This model uses TF-IDF word n-grams (1,1) or (1,2) combined with TF-IDF character n-grams (4,5), ADASYN oversampling, SVM (Linear SVC) with class_weight='balanced', and parameter optimization using GridSearchCV (5-fold cross-validation).

The performance results of this model are:

- Accuracy: 0.8525 (85.25%)
- F1-Macro: 0.7423
- MCC: 0.7120
- ROC-AUC Macro (OvR): 0.9428
- Hamming Loss: 0.0983

The detailed results for each class are shown below:

Table 4. SVM (Linear SVC) Classification Results

Class	Precision	Recall	F1-score
Negative	0.800	0.822	0.811
Neutral	0.923	0.900	0.911
Positive	0.478	0.534	0.505

The Neutral class becomes significantly more stable, with recall increasing from 0.780 (Naïve Bayes) to 0.900 (SVC). This indicates that the model is much better at correctly identifying Neutral sentiments.

The Positive class also experiences an improvement in precision (0.478). The SVM (Linear SVC) model is more selective when assigning the Positive label, which reduces the number of false positives. Additionally, character n-grams help the model handle typos and slang expressions commonly found in social media text. The results obtained in this study are consistent with previous findings that TF-IDF-based sentiment classification methods are effective for social media text analysis [21]

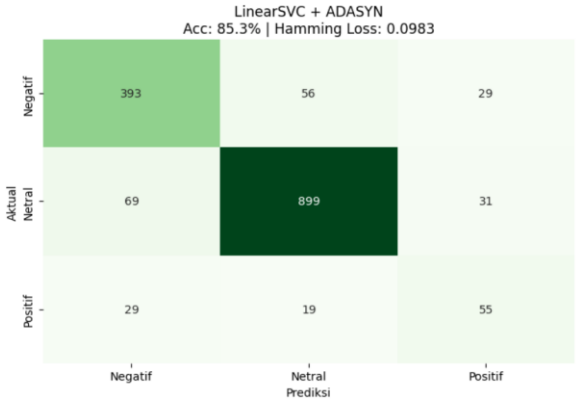

Figure 4. Confusion Matrix of SVM (Linear SVC)

The confusion matrix results for the SVM (Linear SVC) model with ADASYN show a significant improvement in performance compared to the previous model, achieving an accuracy of 85.25% with 1,347 correct predictions out of 1,580 test samples.

The most notable improvement occurs in the Neutral class, which is correctly predicted for 899 data points, indicating that the model is much more effective at distinguishing informational or procedural tweets from complaint-related and appreciation-related tweets.

Although the detection capability for the Negative class remains relatively stable, challenges still exist for the Positive class, which has relatively low recall (only 55 correctly predicted data points) due to similarities in word patterns with the Neutral class and the limited amount of training data.

The Neutral class is often misclassified as Positive or Negative because many neutral tweets contain technical complaint-related keywords such as “error,” “login,” or “failed” without explicitly expressing emotional polarity. In several cases, neutral tweets are written in the form of procedural questions or status reports, making their linguistic patterns highly similar to complaint-related tweets. This overlap in vocabulary distribution causes ambiguity during classification, particularly in short social media texts.

Overall, LinearSVC demonstrates superior performance, as it is capable of producing more stable and reliable classification results.

3.4 MODEL COMPARISON RESULTS

Table 4. Classification Results from Several Journals

Topic + Model	Data Size	Accuracy	Precision	F1-Macro	MCC	ROC-AUC	Hamming Loss
E-Commerce with KNN	895	0.8268	Negative: 0.79 Positive: 0.86	0.828	–	–	–
3 Local Skincare Brands with Naïve Bayes	4,500 (1,500/brand)	79% / 78% / 79%	85%, 88%, 83%	57%, 55%, 59%	–	–	–
K-POP BTS Group	1,500	90%	95%	95%	–	–	–
Coretax with CNB + ADASYN	7,900	0.7867	Negative: 0.730 Neutral: 0.976 Positive: 0.291	0.6832	0.6412	0.9303	0.1475
Coretax with SVM (Linear SVC) + ADASYN	7,900	0.8525	Negative: 0.800 Neutral: 0.923 Positive: 0.478	0.7423	0.7120	0.9428	0.0983

When compared with related studies from other journals, the proposed Coretax SVM (LinearSVC) + ADASYN method demonstrates competitive performance with an accuracy of 0.8525 on a relatively larger dataset (7,900 samples) and a three-class sentiment setting (negative/neutral/positive).

Some previous studies reported higher headline scores such as 90% accuracy on the K-POP BTS dataset (1,500 samples) or an F1-macro of 0.828 in the KNN e-commerce study (895 samples). However, these results were obtained using different datasets, label distributions, and evaluation settings, so direct comparisons should be interpreted cautiously.

Notably, compared with the skincare sentiment study using Naïve Bayes, where the reported F1 values ranged around 55–59%, the Coretax SVM model demonstrates more balanced multi-class performance, supported by a high ROC-AUC (0.9428) and a lower Hamming Loss (0.0983). This indicates strong class separation and a lower overall prediction error rate in this study.

3.5 ROC CURVE

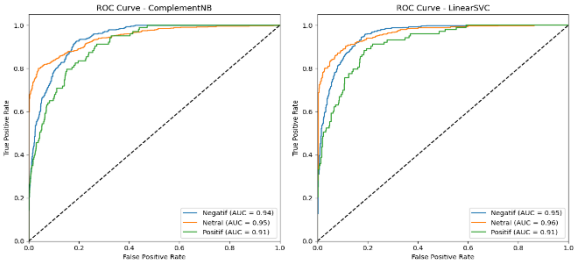

Figure 5. ROC Curve of Naïve Bayes & SVM (Linear SVC)

Based on the visualization of the Receiver Operating Characteristic (ROC) curves, both models demonstrate excellent performance in distinguishing each sentiment class using the One-vs-Rest (OvR) approach.

In the ComplementNB model, classification performance is consistent with Area Under the Curve (AUC) values of:

- 0.94 for the Negative class
- 0.95 for the Neutral class
- 0.91 for the Positive class

Meanwhile, the LinearSVC model shows slightly superior performance with AUC values of:

- 0.95 for the Negative class
- 0.96 for the Neutral class
- 0.91 for the Positive class

Overall, the ROC curves for both models bend sharply toward the upper-left corner, away from the diagonal baseline. This indicates that both ComplementNB and LinearSVC achieve high sensitivity and specificity, effectively separating the Negative, Neutral, and Positive sentiment classes.

3.6 MCNEMAR TEST

The McNemar test was conducted to validate the statistical significance of the performance difference between the Naïve Bayes and LinearSVC models. The test results show a Chi-square value of 46.531 with a p-value of 0.000 ($p < 0.05$), indicating that the null hypothesis (H_0) is rejected. This confirms that the difference in accuracy between the two models is statistically significant.

Further analysis of prediction disagreement cases shows that LinearSVC correctly predicted 166 data points that were misclassified by Naïve Bayes, which is substantially higher than the reverse situation where Naïve Bayes correctly predicted only 62 data points that LinearSVC misclassified.

The dominance of LinearSVC in correcting Naïve Bayes prediction errors confirms its superiority, therefore LinearSVC is identified as the best-performing model in this study.

3.7 WORD CLOUD

Figure 6. Sentiment WordCloud

The word cloud visualization generated from the LinearSVC model predictions successfully maps dominant words that represent the unique characteristics of each sentiment class.

- In the positive sentiment class, keywords such as “fast,” “good,” and “thank you” appear frequently, indicating user appreciation and expectations for responsive service.
- The neutral sentiment class contains more technical and interrogative terms, such as “code,” “login,” “NPWP,” and “please,” reflecting service inquiries or status reports without strong emotional expression.
- Meanwhile, the negative sentiment class is dominated by system performance

complaints, with words such as "error," "slow," "difficult," and "failed."

Overall, this visualization confirms that the model consistently captures the context of the taxation domain and accurately distinguishes user intentions and sentiment patterns.

3.8 OVERVIEW

Based on the analysis of the experimental results, the LinearSVC algorithm combined with the ADASYN oversampling technique proves to be the most effective model for classifying Coretax sentiment. The model achieves the best performance with 85.25% accuracy, 0.9428 ROC-AUC, and statistical validity confirmed by the McNemar test.

This advantage is mainly driven by the ability of LinearSVC to handle high-dimensional text data and maximize the separation margin in datasets with extreme class imbalance. In contrast, the Complement Naïve Bayes model still struggles to distinguish between technical complaints (Negative) and procedural inquiries (Neutral).

Furthermore, the successful application of ADASYN in balancing the training data distribution not only improved evaluation metrics but also enabled the model to better learn patterns from minority sentiment classes, resulting in more reliable sentiment classification performance.

4. CONCLUSION

This study processed textual data by collecting 10,140 sentiment data points from the X platform, which were then filtered to retain only Indonesian-language content, resulting in 9,065 data points. After further filtering of the full-text column, the final dataset consisted of 7,900 data points.

Various preprocessing steps such as symbol cleaning, URL removal, mention removal, normalization, stemming, and stopword removal proved useful in preparing the data so that it could be optimally utilized in the classification process.

Automatic labeling using the IndoBERT model produced an imbalanced sentiment distribution, with the Negative class dominating at 63.20% (4,993 data points), followed by the Neutral class at 30.24% (2,389 data points) and the Positive class at 6.56% (518 data points). This indicates a class imbalance problem that could potentially bias the model toward the majority class.

To address this issue, the ADASYN oversampling technique was applied to the training data. The results showed that ADASYN was effective in balancing the minority class distribution, particularly the Positive class. The number of training samples increased from 6,320 data points to 11,797 data points after applying ADASYN, with a more balanced class distribution. This allowed the model to better learn sentiment patterns from the minority class.

Based on the evaluation results, the LinearSVC + ADASYN model emerged as the best-performing model compared with Complement Naïve Bayes + ADASYN. The LinearSVC model achieved 85.25% accuracy, F1-Macro of 0.7423, MCC of 0.7120, ROC-AUC of 0.9428, and Hamming Loss of 0.0983. Meanwhile, the ComplementNB model showed lower performance with 78.67% accuracy, F1-Macro of 0.6833, and Hamming Loss of 0.1475.

Future research is recommended to perform manual labeling on a portion of the dataset to validate the pseudo-labels generated by IndoBERT, ensuring better label quality and more accurate model evaluation.

In addition, further approaches should be developed to improve classification performance for the Positive class, since this class contains the smallest number of samples and often exhibits word patterns similar to the Neutral class. Several strategies may be explored, such as experimenting with alternative data balancing techniques like SMOTE as a substitute for ADASYN.

Finally, the dataset should be expanded by collecting data over a wider time range and using more varied data retrieval queries, which would increase data diversity and make the sentiment analysis results more representative and comprehensive.

5. REFERENCE

[1] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014, doi: 10.1016/j.asej.2014.04.011.

[2] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Information*, vol. 10, no. 4, Art. no. 150, Apr. 2019, doi: 10.3390/info10040150.

[3] H. Imaduddin, F. Y. A'la, and Y. S. Nugroho, "Sentiment analysis in Indonesian healthcare applications using IndoBERT approach," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 8, pp. 113–119, 2023, doi: 10.14569/IJACSA.2023.0140813.

[4] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, "Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN," *ILKOM Jurnal Ilmiah*, vol. 14, no. 3, pp. 348–354, Dec. 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.

[5] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, Hong Kong, China, Jun. 2008, pp. 1322–1328, doi: 10.1109/IJCNN.2008.4633969.

[6] M. Lestandy, "Sentiment analysis vaccine COVID-19: TF-IDF, RNN, Naïve Bayes," *Jurnal RESTI*

(Rekayasa Sistem dan Teknologi Informasi), vol. 5, no. 4, pp. 802–808, 2021, doi: 10.29207/resti.v5i4.3308.

[7] S. A. Helmayanti and F. Hamami, "Penerapan algoritma TF-IDF dan Naïve Bayes untuk analisis sentimen berbasis aspek ulasan aplikasi Flip," *Jurnal Ilmiah Manajemen Informatika dan Komputer*, vol. 4, no. 3, pp. 182–191, 2023, doi: 10.35870/jimik.v4i3.415.

[8] N. Z. B. Jannah and K. Kusnawi, "Comparison of Naïve Bayes and SVM in sentiment analysis of product reviews on marketplaces," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 1, pp. 145–154, 2024.

[9] A. S. Muliana, "Analysis of public sentiment on election results using Naïve Bayes and TF-IDF," *STMSI: Jurnal Sistem dan Teknologi Informasi*, vol. 7, no. 2, pp. 101–110, 2024.

[10] V. B. Lestari and C. A. Hutagalung, "Evaluation of TF-IDF extraction techniques in sentiment analysis of Indonesian marketplaces," *J-KOMA: Jurnal Ilmu Komputer dan Aplikasi*, vol. 8, no. 1, pp. 25–34, 2025.

[11] L. D. Cahya, "Sentiment analysis on artificial intelligence developments using Naïve Bayes classifier," *Jurnal Ilmu dan Pengetahuan Teknologi*, vol. 6, no. 2, pp. 77–86, 2024.

[12] S. Khoerunnisa, A. Nugraha, and R. Pratama, "Penerapan algoritma Naïve Bayes dengan teknik TF-IDF dan cross validation untuk analisis sentimen terhadap Starlink," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 1, pp. 112–121, 2025.

[13] I. H. Kusuma and N. Cahyono, "Analisis sentimen masyarakat terhadap penggunaan e-commerce menggunakan algoritma K-Nearest Neighbor," *Jurnal Informatika dan Sistem Informasi*, vol. 8, no. 3, pp. 236–243, 2023.

[14] K. S. Putri, R. Setiawan, and A. Pambudi, "Analisis sentimen terhadap brand skincare lokal menggunakan Naïve Bayes classifier," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 4, pp. 701–710, 2023.

[15] T. Safitri, Y. Umaidah, and I. Maulana, "Analisis sentimen pengguna Twitter terhadap BTS menggunakan algoritma Support Vector Machine," *Journal of Applied Informatics and Computing*, vol. 7, no. 1, pp. 48–55, 2023.

[16] N. K. A. Juliana and M. A. Raharja, "Analisis sentimen pada ulasan aplikasi myIM3 menggunakan Multinomial Naïve Bayes dengan TF-IDF," *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, vol. 2, no. 3, pp. 521–530, 2024.

[17] P. W. A. Wibawa and C. R. A. Pramartha, "Analisis sentimen pada teks berbahasa Bali menggunakan metode Multinomial Naïve Bayes dengan TF-IDF dan BoW," *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, vol. 2, no. 1, pp. 89–98, 2023.

[18] A. Ardiansyah and Kurniawan, "Optimization of Naïve Bayes classifier using TF-IDF for sentiment analysis," *Journal of Software and Algorithmic Intelligence*, vol. 7, no. 3, pp. 155–164, 2024.

[19] T. Hidayat, R. Maulana, and D. Wulandari, "Analisis sentimen opini masyarakat terhadap Pilkada 2024 menggunakan Naïve Bayes," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, pp. 45–54, 2025.

[20] S. Rahmah, A. F. Maulana, and D. Utami, "Sentiment analysis of public opinions on e-government services using Naïve Bayes," *International Journal of Computing and Digital Systems*, vol. 13, no. 2, pp. 215–226, 2024.

[21] H. N. Dewi and R. Pratama, "Improved sentiment classification on social media using TF-IDF and Naïve Bayes," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 6, no. 1, pp. 33–42, 2023.