Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1467−1476
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8310

SMOTE and BERT Approaches for Handling Class Imbalance in Sentiment Analysis of the CoreTax Application on Big Data

Meiliyani Br Ginting[1,*], Asprina Br Surbakti[2], Safarul Ilham[3], Dito Putro Utomo[4], Raheliya Br Ginting[3]

[1] Faculty of Science and Technology, Software Engineering, Institute of Technology and Business Indonesia, Medan, Indonesia
[2] Faculty of Science and Technology, Information Technology, Institute of Technology and Business Indonesia, Kabanjahe, Indonesia
[3] Faculty of Science and Technology, Informatics Engineering, Institute of Technology and Business Indonesia, Medan, Indonesia
[4] Department of Computer and Informatics Engineering, Multimedia Graphic Technology Study Program, Medan State Polytechnic, Medan, Indonesia
Email: [1,*]meiliyani.ginting@gmail.com, [2]asprinasurbakti.dosen@itbi.ac.id, [3]safarulilham@gmail.com, [4]ditoputroutomo@polmed.ac.id, [5]raheliyabrginting@gmail.com
Email Corresponding Author: meiliyani.ginting@gmail.com
Submitted: **30/08/2025**; Accepted: **30/09/2025**; Published**: 30/09/2025**

Abstract−Coretax is a tax information system developed by the Directorate General of Taxes (DJP) to support digital and integrated tax administration processes, covering everything from taxpayer registration to reporting and auditing. Although it was designed to improve efficiency, transparency, and accuracy in tax management, its implementation has sparked mixed reactions among the public due to various technical challenges and the complexity of the annual tax reporting process. This situation highlights the need for a sentiment analysis that can objectively capture public perceptions of the system's performance. In this study, Natural Language Processing (NLP) and Machine Learning techniques were applied to analyze 3,000 tweets from Twitter (X) related to Coretax. One of the main issues identified in the dataset is class imbalance, where positive sentiments significantly outnumber negative and neutral ones, leading to biased classification results. To address this issue, the Synthetic Minority Over-sampling Technique (SMOTE) was used to balance the dataset by generating synthetic samples for the minority classes. The BERT model was then employed for sentiment classification because of its strong ability to understand contextual meaning through its transformer-based architecture. Experimental results show that before applying SMOTE, the BERT model achieved an accuracy of 77%, which increased to 80% after SMOTE was implemented, along with improvements in precision, recall, and F1-score, particularly for the minority classes. These findings demonstrate that the combination of SMOTE and BERT significantly enhances the performance of sentiment analysis in understanding public responses to Coretax. This approach can serve as a valuable reference for evaluating and improving tax digitalization policies, ensuring they are more effective, inclusive, and responsive to public needs.

Keywords: SMOTE; BERT; Class Imbalance; Sentiment Analysis; Coretax Application

1. PENDAHULUAN

CoreTax is an integrated information system developed by the Directorate General of Taxes (DJP) as part of efforts to modernize tax administration in Indonesia. The system plays a key role in bringing together various tax processes such as taxpayer registration, filing of tax returns (SPT), payments, and audits into a single digital platform. By implementing CoreTax, the DJP aims to improve efficiency, transparency, and convenience in tax services, allowing taxpayers to fulfill their obligations online more quickly, accurately, and securely.[1].

However, the implementation of CoreTax in practice still faces several challenges that hinder the system from achieving its original goals. Common technical issues reported include difficulties in creating accounts, failure to receive OTP codes, slow system performance, and disruptions during the filing of tax returns (SPT). These challenges have generated a wide range of public responses, from support for the digitalization of tax services to criticism of CoreTax's effectiveness and reliability as a primary digital tax administration system[2].

Analyzing public opinion is essential to assess how well CoreTax is received by society. The diverse responses from users should be considered in the development and improvement of the application. A previous study by Agnia Suci Rizkia and colleagues in 2025 found that 30.53% of responses were negative, 47.3% were positive, and 1.79% were neutral regarding CoreTax. The testing process in that study reported an overall accuracy of 77%[3]. In addition, a study by Fathoni and colleagues in 2025 reported that 59.94% of public responses were negative and 40.06% were positive, with the analysis achieving an accuracy of 92%[4]. From previous studies, it is evident that negative public responses to the CoreTax application are higher than positive ones. This clearly presents an issue that needs to be addressed to better understand how well CoreTax is accepted by the public. One of the challenges identified in these studies is the imbalance between the number of positive and negative responses, which has resulted in suboptimal outcomes.

Based on the problems outlined above, this study will focus on performing sentiment analysis on public responses to the CoreTax application. The sentiment analysis process leverages advancements in Artificial Intelligence, as it falls under the field of Natural Language Processing, which is a key branch of AI[5]. Sentiment analysis within Natural Language Processing plays a key role in processing data and presenting it in human-readable language, making it easier to understand[6]. In general, for sentiment analysis results to be representative of all comments, the process needs to be conducted on a very large volume of data[7]. Therefore, sentiment analysis in NLP uses Machine Learning to process Big Data. Big Data–based sentiment analysis is very helpful in understanding whether responses to a particular subject are positive or negative. This approach aligns with the issues addressed in

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1467−1476
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8310

the current study. The suboptimal performance of the CoreTax application has generated a wide range of public reactions. These responses are collected into a dataset, forming what is known as Big Data. This collected data will then be analyzed through sentiment analysis to assess how CoreTax is perceived by the public. By applying sentiment analysis, it is expected that insights and recommendations can be provided to guide the future development and improvement of the CoreTax application.

In this context, Big Data based sentiment analysis becomes crucial for understanding public perceptions and experiences regarding the implementation of CoreTax more comprehensively. Through sentiment analysis, user reviews and opinions spread across various digital platforms, such as social media, can be processed to identify public emotional patterns, both positive and negative. The Big Data approach allows for efficient collection and processing of large volumes of data, ensuring that the analysis results more accurately and objectively reflect public opinion. Consequently, sentiment analysis serves not only as a tool for evaluating system performance but also as a foundation for strategic decision-making to enhance the quality of CoreTax services and user experience[8], [9].

However, previous studies on sentiment analysis often face several key challenges, particularly related to data imbalance and limitations of the methods used. Data imbalance occurs when the number of positive and negative reviews is unequal, causing models to be biased toward the dominant class and produce less accurate predictions for the minority class. Additionally, many studies still rely on traditional algorithms such as Naive Bayes, SVM, or Logistic Regression, which have limitations in capturing the linguistic context and semantic meaning of complex review texts[10], [11], [12]. This situation often results in sentiment analysis outcomes that do not fully represent public opinion. Therefore, a more advanced and adaptive approach is needed to address data imbalance while enhancing the model's ability to understand the natural language context of users.

This study focuses on optimizing the results obtained from the sentiment analysis process. The optimization aims to ensure that the outcomes accurately reflect the reality of public comments. Sentiment analysis uses very large datasets in Big Data, which often contain imbalanced data across different classes. Such imbalance can significantly affect the results, making it necessary to implement a method to handle it. The SMOTE model provides a solution for addressing data imbalance in this context[13]. The SMOTE model aims to address data imbalance by focusing on the minority class[14]. By using the SMOTE model, it becomes possible to achieve more optimal results in the sentiment analysis process.

The sentiment analysis in this study uses the BERT model (Bidirectional Encoder Representations from Transformers) due to its ability to understand the context of text in both directions. This model was chosen to overcome the limitations of previous approaches, which were unable to accurately capture the contextual meaning in public comment data[15]. Moreover, BERT is well-suited for Big Data applications because its deep learning architecture can efficiently learn complex linguistic patterns. This capability is expected to improve both the accuracy and reliability of sentiment analysis results regarding public opinions on the CoreTax application[16].

To address the issue of class imbalance often found in public opinion data, this study uses the SMOTE (Synthetic Minority Oversampling Technique) approach. This method is chosen because it can improve the data distribution by increasing the representation of minority classes without altering their original characteristics[17]. By applying SMOTE, the sentiment analysis model can learn patterns from each sentiment class more evenly, resulting in more accurate classifications that are not biased toward the majority class. This approach is a crucial step to ensure that the analysis of public perceptions regarding the CoreTax application truly reflects real-world opinions[18].

To address this issue, this study proposes a combined approach using SMOTE and BERT. This approach is designed to improve the accuracy and reliability of sentiment analysis on public reviews of the CoreTax application. SMOTE is used as a preprocessing step to balance the class distribution in the dataset by generating synthetic samples for the minority class, preventing the model from being biased toward the majority class. Once the data distribution is balanced, the BERT model is applied to perform sentiment analysis, leveraging its deep understanding of linguistic context through the attention mechanism. The combination of these two methods is expected to enhance classification performance, reduce prediction errors for minority classes, and produce more accurate sentiment representations that objectively reflect public perceptions of the CoreTax system's effectiveness.

Based on the explanation above, the urgency of conducting this research becomes clear. This study will be highly valuable in assessing public responses to the CoreTax application in the tax reporting process. It will also facilitate sentiment analysis in a way that is faster, more accurate, and more efficient. Furthermore, the results of this research can provide insights and recommendations for the future development and improvement of the CoreTax application.

This study stands out due to its significant contributions, particularly in developing Big Data–based sentiment analysis by integrating the SMOTE data balancing technique with the BERT model. This combination is expected to improve the accuracy of sentiment classification on datasets with imbalanced distributions. Moreover, this approach can serve as a reference for future research in Natural Language Processing and Artificial Intelligence, especially in the context of large-scale social data.

The results of this study are expected to provide a more accurate understanding of public perceptions regarding the implementation of the CoreTax application. These findings can be used by the Directorate General of Taxes as a basis for evaluation and decision-making to improve the quality of digital tax services. Additionally, the sentiment analysis system developed in this research is adaptive and can be applied to analyze public opinions in other contexts,

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1467−1476
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8310

offering the potential to serve as a sustainable solution for monitoring public perceptions of various government digital services.

The problem-solving approach in this study focuses on analyzing public sentiment toward the CoreTax application. The approach combines the SMOTE and BERT models within a Big Data framework to improve the effectiveness of capturing public responses. While sentiment analysis itself is not new in technology, the integration of SMOTE and BERT offers a more effective way to assess public reactions to CoreTax. This is particularly important in a Big Data context, where data imbalances frequently occur, making SMOTE essential for enhancing BERT's performance. By combining SMOTE and BERT, this study demonstrates an approach that meets the need for an efficient and accurate sentiment analysis system capable of handling the large volume of public responses to the CoreTax application.

The sentiment analysis conducted in this study has the advantage of integrating sentiment analysis with Big Data. This research can serve as a reference for understanding public responses to the CoreTax application, both now and in the future. By enhancing the performance of sentiment analysis, the results obtained can be more reliable and convincing. From the research process, it is expected that this study will help address issues related to public responses to the CoreTax application. Using this system, both positive and negative public reactions can be analyzed. Being based on Big Data, it allows for processing large volumes of public responses efficiently. Additionally, the results of this study can serve as input for the development or improvement of the CoreTax application, and the approach can also be adapted to analyze public sentiment in other fields.

2. RESEARCH METHODOLOGY

This section explains the approach and steps taken in conducting the research. The research methodology is designed to provide a systematic overview of the processes for designing, collecting, processing, and analyzing data to achieve the study's objectives. The approach focuses on sentiment analysis of public opinions regarding the CoreTax application, utilizing SMOTE to address data imbalance and the BERT model for classification through Natural Language Processing (NLP). Additionally, the study adopts a Big Data Analytics paradigm to efficiently manage and analyze large volumes of data.

2.1 Research Framework

The Research Framework describes the series of processes followed in this study. These processes are illustrated in Figure 1 below:

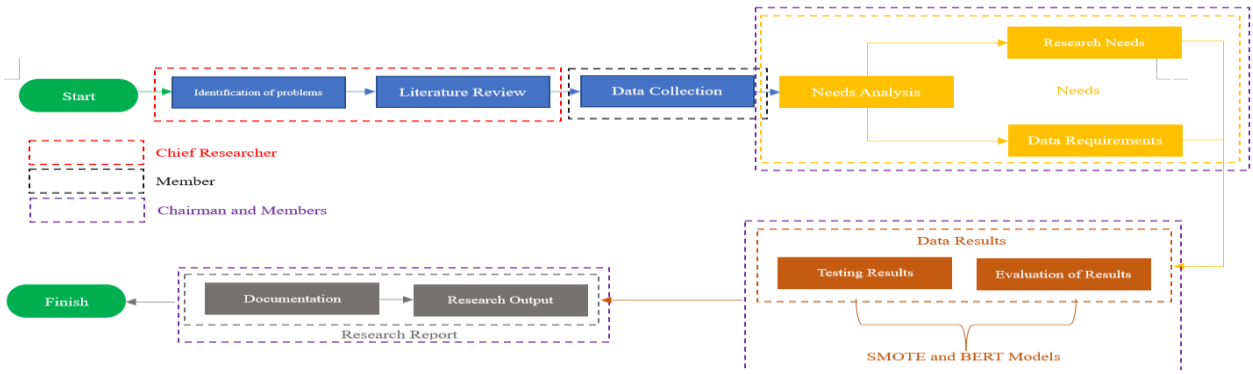

Figure 1. Research Framework

The research framework shown in Figure 1 illustrates the systematic flow of the study, starting from the collection of public opinions about the CoreTax application and continuing through to the evaluation of the results.

2.2 Natural Language Processing (NLP)

Natural Language Processing (NLP) is a field within computer science and Artificial Intelligence that relies on Machine Learning[19]. NLP enables computers to understand and process human language, whether in text, audio, or video formats[20]. In this context, NLP automatically facilitates communication between humans and computers. Today, NLP has become increasingly important due to the massive amounts of data being generated, especially with the widespread use of social media, where people freely express their opinions[21]. This is also relevant in the context of public responses to the CoreTax application and its use in tax reporting. With the large volume of feedback from users, NLP plays a crucial role by making it possible to efficiently process this data and extract meaningful insights from public responses[22].

2.3 SMOTE Model

Synthetic Minority Oversampling Technique (SMOTE) is an oversampling technique for minority classes by generating new synthetic samples[23]. The SMOTE model is generally used to resolve imbalances in datasets. Balance

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1467−1476
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8310

is crucial for data; balanced data yields more optimal results[24]. The SMOTE model process is based on the K-Nearest Neighbor method to obtain new data samples based on the percentage value of the minority class. The SMOTE model works by modifying an imbalanced dataset by allowing overfitting of the minority class data[25]. Big data-based research greatly requires the SMOTE model, as big data often presents imbalanced data sets. Therefore, in this study, the SMOTE model plays a role and aims to help balance the data to achieve optimal sentiment analysis results, as shown in Figure 2 below:

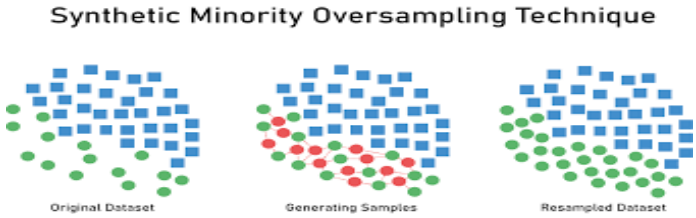

Figure 2. SMOTE Model

Figure 2 illustrates the SMOTE (Synthetic Minority Oversampling Technique) model used in this study to address data imbalance in sentiment classes. The model generates new synthetic samples for minority classes by leveraging the relationships between nearest neighbor data points, resulting in a more balanced distribution between positive and negative classes.

2.4 BERT Model

BERT (Bidirectional Encoder Representations from Transformers) is an Artificial Intelligence model based on Deep Learning. The BERT model was developed to improve the performance of NLP[26]. The BERT model is based on the ability to understand words in sentences by analyzing the relationships between words in two directions. The BERT model combines Encoder Representations and Transformers[27]. By utilizing this, the BERT model can understand the full meaning of words, making it very popular for use in data analysis. In general, the BERT model process is divided into five stages[28]:

a. Tokenization: Text is broken down into small units (tokens), including words or subwords, so that it can be understood by the model.
b. Embedding: Each token is converted into a numeric vector that represents its meaning in a dimensional space.
c. Neural Network Processing: These vectors are processed in a neural network, where BERT analyzes the relationships between words bidirectionally, understanding the meaning of the entire text, not just from one direction.
d. Masked Language Model (MLM) Technique: BERT randomly masks some words in a sentence and tries to guess them based on the context of the surrounding words. This technique allows BERT to understand the meaning of words in various contexts.
e. Next Sentence Prediction (NSP) Technique: BERT is also trained to determine whether two sentences are related or not.

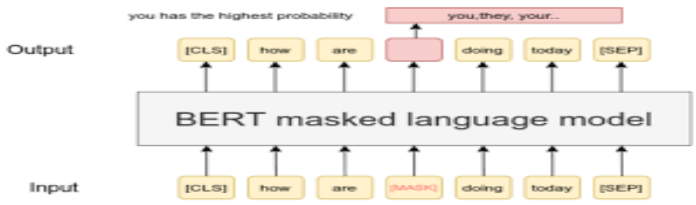

Figure 3. BERT Model

Figure 3 shows the architecture of the BERT (Bidirectional Encoder Representations from Transformers) model used in this study for sentiment analysis of public opinions on the CoreTax application. BERT understands the context of words in both directions left to right and right to left allowing it to capture the meaning of sentences more deeply compared to conventional one-directional models.

2.5 Data Analysis Techniques

This study employs a quantitative approach based on sentiment analysis, utilizing the BERT model alongside the SMOTE data balancing technique. The methodology is designed to identify public perceptions of the CoreTax application through user reviews collected from various digital sources, while also addressing class imbalance in the dataset that could affect the accuracy of the analysis.

2.5.1 Data Sources and Structure

The data used in this study comes from user reviews on the social media platform Twitter (X), where people actively share their experiences with the CoreTax application. Data collection was conducted from January 2025 to April 2025,

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1467−1476
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8310

resulting in a total of 3,000 text review entries. Each entry includes a user ID, posting date, review text, and sentiment label (positive, negative, or neutral). The dataset was stored in CSV format and processed using Python for further analysis.

2.5.2 Text Data Preprocessing

Preprocessing is carried out to ensure that the text data used for model training is clean and ready for analysis. The steps applied include:

a. Case Folding: converting all text to lowercase to prevent duplicate words caused by capitalization.
b. Text Cleaning: removing non-alphabetic characters, links, emojis, and irrelevant punctuation.
c. Tokenization: splitting sentences into words (tokens) for easier processing by the model.
d. Stopword Removal: removing common words with little semantic value, such as “yang,” “dan,” and “dengan.”
e. Stemming: reducing words to their root form using an Indonesian stemming algorithm.
f. Labeling: assigning sentiment labels to each review.

These steps aim to improve the quality of text representation before training the BERT model.

2.5.3 Class Imbalance Analysis

Initial data exploration revealed a significant class imbalance, with 40% positive reviews, 17% negative, and 43% neutral. This imbalance can cause the model to be biased toward majority classes, making it necessary to apply SMOTE to generate synthetic samples for the minority class. This approach is expected to enhance the model’s ability to recognize sentiment patterns more fairly across all classes.

2.5.4 SMOTE Implementation

Data balancing was performed using the SMOTE algorithm with k-neighbors set to 5 and an oversampling ratio of 100% for the minority class. These parameters were chosen based on initial trials, which showed that this ratio provided optimal balance without causing overfitting. After applying SMOTE, each class—positive, negative, and neutral contained 1,300 samples.

2.5.5 BERT Model and Fine-Tuning Process

The model used in this study is IndoBERT-base, a pre-trained BERT version adapted for the Indonesian language. It was fine-tuned using the CoreTax dataset after preprocessing and balancing. Training was configured with a learning rate of 2e-5, batch size of 16, and 4 epochs. The AdamW optimizer was used, along with an early stopping strategy to prevent overfitting. Fine-tuning allows the model to adjust its transformer weights to better understand the linguistic context of CoreTax reviews.

2.5.6 Model Evaluation and Validation Scheme

Model performance was evaluated using Accuracy, Precision, Recall, and F1-Score, providing a comprehensive view of classification performance for each sentiment class. Macro Average was also calculated to assess performance evenly across classes without bias toward class distribution. Validation was conducted using an 80:20 train-test split, with 80% of the data used for training and 20% for testing. This scheme ensures that the model is evaluated on previously unseen data, allowing an objective assessment of its generalization ability.

3. RESULT AND DISCUSSION

The next stage of this research involved addressing the problem through sentiment analysis using SMOTE and BERT. The testing process began by collecting data from Twitter (X), followed by experimentation conducted on Google Colaboratory. Through the data crawling process on social media, a collection of tweets related to the research topic was obtained. The crawled dataset included key information such as the posting date, tweet content, number of likes, and number of retweets. This information is summarized in:

Table 1. Data Crawling Results

No	Tanggal	Tweet Raw	Likes	Retweets
1	2025-01-05 08:21:34	Coretax makin ribet aja, login error terus 😡	15	3
2	2025-01-06 12:10:45	Seneng banget, laporan SPT tahunan lancar via Coretax! 👍	42	10
3	2025-01-06 18:32:12	Baru coba aplikasi Coretax, lumayan membantu sih walaupun agak lemot.	20	2
4	2025-01-07 09:15:02	Coretax error pas mau upload dokumen, bikin kesel 😫	11	1
5	2025-01-07 14:27:55	Aplikasi Coretax sangat membantu untuk UMKM. Terima kasih! 🙏	55	7

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1467−1476
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8310

No	Tanggal	Tweet Raw	Likes	Retweets
6	2025-01-08 21:05:11	Susah banget akses Coretax malam ini, down terus. 😔	9	0
7	2025-01-09 11:41:23	Coretax oke juga, fitur pelaporan lebih rapi daripada sebelumnya.	27	5
8	2025-01-10 10:03:19	Lagi testing Coretax, sejauh ini biasa aja.	7	1
9	2025-01-10 16:48:47	Akhirnya berhasil lapor via Coretax tanpa error. Terima kasih DJP! 🙌	39	12
10	2025-01-11 20:29:05	Coretax bikin pusing, selalu error pas submit 😡	13	2

Table 1 serves as the initial foundation for the analysis stage, as it contains raw data that will later undergo several preprocessing steps, including cleaning, text preprocessing, and sentiment labeling. From this table, early patterns of public discussions related to the research topic can be identified, while also preparing the dataset for further analytical processes. After the crawling stage, the collected data still contained various irrelevant elements such as excessive punctuation, emojis, links, and hashtags. Therefore, a data cleaning process was carried out to ensure the quality and consistency of the dataset to be used in the next stage of analysis.

Table 2. Data Cleaning Results

No	Tweet Raw	Tweet Clean
1	Coretax makin ribet aja, login error terus 😡	coretax ribet login error
2	Seneng banget, laporan SPT tahunan lancar via Coretax! 👍	senang lapor spt tahun lancar coretax
3	Baru coba aplikasi Coretax, lumayan membantu sih walaupun agak lemot.	coba aplikasi coretax lumayan bantu lemot
4	Coretax error pas mau upload dokumen, bikin kesel 😤	coretax error upload dokumen kesel
5	Aplikasi Coretax sangat membantu untuk UMKM. Terima kasih! 🙏	aplikasi coretax bantu umkm terima kasih
6	Susah banget akses Coretax malam ini, down terus. 😔	susah akses coretax malam down terus
7	Coretax oke juga, fitur pelaporan lebih rapi daripada sebelumnya.	coretax oke fitur pelapor rapi sebelumnya
8	Lagi testing Coretax, sejauh ini biasa aja.	lagi testing coretax sejauh ini biasa aja
9	Akhirnya berhasil lapor via Coretax tanpa error. Terima kasih DJP! 🙌	berhasil lapor coretax tanpa error terima
10	Coretax bikin pusing, selalu error pas submit 😡	coretax pusing error submit

The results of the cleaning process are shown in Table 2. In this table, it can be seen that the tweet texts have been normalized, leaving only the core words that represent users' opinions. This cleaning step makes the data more structured and ready for the preprocessing and sentiment labeling stages. After the cleaning process, the refined data is automatically assigned sentiment labels based on keyword analysis or the machine learning model applied. The labeling process aims to classify each text into positive, negative, or neutral categories, allowing researchers to identify public perceptions related to the research topic.

Table 3. Sentiment Labeling Results

No	Tweet Clean	Sentiment
1	coretax ribet login error	Negative
2	senang lapor spt tahun lancar coretax	Positive
3	coba aplikasi coretax lumayan bantu lemot	Neutral
4	coretax error upload dokumen kesel	Negative
5	aplikasi coretax bantu umkm terima kasih	Positive
6	susah akses coretax malam down terus	Negative
7	coretax oke fitur pelapor rapi sebelumnya	Positive
8	lagi testing coretax sejauh ini biasa aja	Neutral
9	berhasil lapor coretax tanpa error terima	Positive
10	coretax pusing error submit	Negative

After completing the preprocessing stage as shown in Table 3 above, the next step is to conduct testing based on the data obtained. The initial phase of this research involves presenting the results of data distribution to illustrate the initial class distribution.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1467−1476
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8310

```
positif    1200
negatif     500
netral     1300
Name: label, dtype: int64
```

Figure 4. Initial Class Distribution

In Figure 4, the data distribution across sentiment classes can be seen, with 1,200 positive samples, 500 negative samples, and 1,300 neutral samples. This imbalance reveals a significant class disparity, where the negative class has far fewer data points compared to the others. Such an imbalance can cause the learning model to become biased toward the majority classes, as it is more frequently exposed to patterns from them during training. As a result, the model's ability to recognize or predict minority class samples decreases, leading to lower accuracy and fairness in classification. This imbalance naturally arises from public opinion patterns about the CoreTax application, where users tend to express more positive or neutral sentiments than explicit complaints (negative sentiments). Therefore, analyzing this distribution is crucial not only to describe the dataset's condition but also to justify the methodological decision to apply the SMOTE technique in the next phase — aimed at reducing bias caused by majority class dominance. The results of the data distribution after applying SMOTE are shown below:

```
Sebelum SMOTE: [1200  500 1300]
Sesudah SMOTE: [1300 1300 1300]
```

Figure 5. Data Distribution After SMOTE

Figure 5 shows the change in class distribution before and after applying SMOTE (Synthetic Minority Over-sampling Technique). Initially, the classes were imbalanced with counts of 1200, 500, and 1300, but after oversampling, all classes were balanced at 1300 each. This change not only increases the number of samples in minority classes but also highlights the methodological impact of SMOTE on the learning model. The initial imbalance made the model favor the majority classes, leading to biased predictions and lower accuracy for minority classes. After applying SMOTE, synthetic data generated by interpolating features between minority samples provides a better representation of all classes. This helps the model generalize more evenly and reduces overfitting to dominant classes. Therefore, the improved model performance after SMOTE is not just due to having more data but also because the training distribution now better reflects the entire data population.

```
== HASIL TANPA SMOTE ==
              precision    recall  f1-score   support

    negative       0.71      0.55      0.62        80
     neutral       0.68      0.75      0.71       160
    positive       0.85      0.88      0.86       240

    accuracy                           0.77       480
   macro avg       0.75      0.73      0.73       480
weighted avg       0.77      0.77      0.77       480
```

Figure 6. Test Results Without SMOTE

Figure 6 presents the results of sentiment analysis testing using the BERT model without applying SMOTE, showing an overall accuracy of 77%. While this may seem reasonable, the results reveal a clear imbalance in performance across classes. The positive class achieves high precision and recall (0.85 and 0.88), whereas the negative class performs much worse, especially in recall (0.55). This difference indicates that the model is better at recognizing majority classes than minority classes. The imbalance arises from the uneven number of training samples, where classes with fewer examples do not provide enough variation for the model to learn representative patterns. As a result, BERT exhibits classification bias, often predicting positive or neutral sentiments, while many negative samples are misclassified. This shows that even though BERT has strong contextual understanding, the quality of its classification heavily depends on balanced training data. Therefore, applying SMOTE in the next stage is essential to reduce this bias and improve the model's ability to recognize minority classes more proportionally.

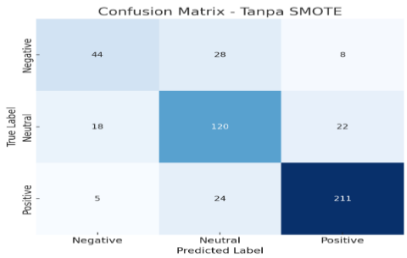

Figure 7. Confusion Matrix Without SMOTE

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1467−1476
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8310

Figure 7 shows the confusion matrix for the BERT model evaluation before applying SMOTE, highlighting the model's tendency to predict majority classes more accurately than minority classes. The model performs well in identifying positive and neutral classes but struggles significantly with the negative class, as seen from the low number of correct predictions in the "Negative" row. This not only reflects the imbalance in the data distribution but also indicates that the semantic representation of the minority class is not strong enough for the model to learn effectively. In other words, because the model "sees" majority class data more often during training, BERT's internal weights become more focused on patterns that frequently occur in these classes. As a result, the model's ability to generalize to negative expressions is limited, leading to many negative samples being misclassified as neutral or positive. This phenomenon demonstrates that class imbalance is not just a matter of data quantity; it also affects the formation of linguistic representations within the model, ultimately reducing the overall performance of the sentiment analysis system.

== HASIL DENGAN SMOTE ==

	precision	recall	f1-score	support
negative	0.74	0.65	0.69	80
neutral	0.70	0.77	0.73	160
positive	0.87	0.89	0.88	240
accuracy			0.80	480
macro avg	0.77	0.77	0.77	480
weighted avg	0.80	0.80	0.80	480

Figure 8. Evaluation Results with SMOTE

Figure 8 illustrates the evaluation results of the BERT model after the application of SMOTE, where the model's accuracy increased to 80% compared to the previous 77%. This improvement was not merely the result of adding more samples to the minority class, but rather due to the model gaining a more balanced linguistic representation across all sentiment categories. With the dataset distribution now evened out, the model was no longer overly exposed to patterns from the majority class, allowing the learning process to become more proportional. As a result, the classification layer in BERT became better at distinguishing the unique semantic features of each sentiment class, particularly in the negative class, which had previously shown low recall. The increase in the F1-score from 0.62 to 0.69 for the negative class indicates that SMOTE not only expanded the data volume but also improved the model's ability to recognize rare linguistic patterns that were underrepresented in the original dataset. Therefore, the main impact of SMOTE on the model is not just quantitative but also qualitative it enhances BERT's contextual learning, making it more representative of the full spectrum of sentiment data.

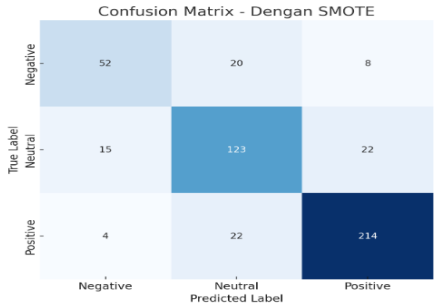

Figure 9. Confusion Matrix with SMOTE

Figure 9 presents the confusion matrix of the BERT model's performance after applying SMOTE, showing a noticeable improvement in recognizing minority classes compared to the results before SMOTE was used. The increase in true positives, especially within the negative and neutral classes, occurred because SMOTE's synthetic oversampling successfully enriched the linguistic feature variations that were previously underrepresented in the training data. By generating more synthetic samples for these minority classes, BERT was able to learn contextual relationships more evenly across all classes, leading to a more balanced distribution of attention weights within the transformer mechanism. As a result, the number of misclassifications—both false negatives and false positives—was reduced, since the model no longer relied too heavily on the dominant language patterns of the majority class. In other words, the improvement in model accuracy after SMOTE was not merely due to the addition of more data, but rather because the model gained a deeper and more comprehensive understanding of the semantic nuances across different sentiment expressions. These findings highlight that balancing data during the training phase plays a crucial role in enhancing BERT's generalization ability for large-scale sentiment analysis in Big Data contexts.

4. CONCLUSION

This study focuses on sentiment analysis of public responses toward the CoreTax application developed by the Directorate General of Taxes (DJP) as part of its effort to modernize tax administration in Indonesia. Although

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1467−1476
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8310

CoreTax was designed to simplify the process of tax filing and payment, its implementation has faced several challenges, such as difficulties in account creation, issues with OTP verification, and slow system performance. These problems have sparked various reactions on social media, ranging from support to strong criticism. The analysis revealed a class imbalance between positive and negative sentiments, which can affect the accuracy of sentiment classification. To address this issue, the study applied the SMOTE (Synthetic Minority Oversampling Technique) method to balance the data distribution and the BERT (Bidirectional Encoder Representations from Transformers) model to enhance sentiment classification accuracy using Natural Language Processing (NLP) techniques on Big Data. The results showed that applying SMOTE improved BERT's accuracy from 77% to 80%, indicating better recognition of minority-class patterns and reduced bias toward the majority class. However, the 3% improvement should be interpreted cautiously, as it may not be statistically significant without further validation through methods such as cross-validation or additional performance metrics like precision, recall, and F1-score. Another limitation lies in the use of Twitter as the sole data source, which tends to feature informal language and context-specific expressions that may not fully represent the broader public perception. Therefore, future research should expand the dataset to include multiple social media platforms, experiment with alternative balancing techniques such as ADASYN, and compare different NLP models to achieve more generalizable and accurate sentiment analysis outcomes.

REFERENCES

[1] Nataherwin and A. E. Defin, "Pelatihan Penggunaan Coretax System untuk Pelaporan Perpajakan di PT Koilima Putra Mandiri," *J. Pustaka Mitra*, vol. 5, no. 5, pp. 265–269, 2025, doi: https://doi.org/10.55382/jurnalpustakamitra.v5i5.1155.

[2] D. F. Nurhaeni, D. Masitoh, H. Shofurani, N. K. Livtanta, and Ridwan, "Analisis Efektifitas dan Efisiensi Sistem CORETAX: Mengukur Kepercayaan Publik di Tengah Transisi Sistem Perpajakan 2025," *J. Sos. Polit.*, vol. 6, no. 1, pp. 21–37, 2025, doi: 10.54144/jsp.v6i1.103.

[3] A. S. Rizkia, Wufron, and F. F. Roji, "Sentiment Analysis of Coretax: A Comparison of Manual, Transformers- Based, and Lexicon-Based Data Labeling on IndoBERT Performance," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 3, pp. 1037–1048, 2025, doi: https://doi.org/10.57152/malcom.v5i3.2151 1037.

[4] Fathoni, A. F. Ansori, I. N. Ramadhani, C. R. Anissa, and S. A. Putri, "Analisis Sentimen Masyarakat Indonesia di Twitter Terhadap Sistem Perpajakan 'Coretax' Menggunakan Metode Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 4, pp. 6749–6753, 2025, doi: https://10.36040/jati.v9i4.14214.

[5] S. Pais, J. Cordeiro, and M. L. Jamil, "NLP-based platform as a service: a brief review," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00603-5.

[6] M. Chiny, M. Chihab, O. Bencharef, and Y. Chihab, "Netflix Recommendation System based on TF-IDF and Cosine Similarity Algorithms," in *Proceedings ofthe 2nd International Conference on Big Data, Modelling and Machine Learning*, Science and Technology Publications, Lda, 2022, pp. 15–20. doi: 10.5220/0010727500003101.

[7] S. Jaradat, R. Nayak, A. Paz, and M. Elhenawy, "Ensemble Learning with Pre-Trained Transformers for Crash Severity Classification: A Deep NLP Approach," *Algorithms*, vol. 17, no. 7, 2024, doi: 10.3390/a17070284.

[8] L. Nurina, S. H. Hairuddin, A. A. Bakri, and A. Pilua, "Tinjauan Bibliometrik Terhadap Pemanfaatan Big Data, Analisis Sentimen, dan Kriptokurensi dalam Analisis Pajak," *Sanskara Akunt. dan Keuang.*, vol. 2, no. 01, pp. 66–76, 2023, doi: 10.58812/sak.v2i01.257.

[9] Putri Angraini Aziz, S. B. Nur Ilahi, Sumiarni Moka, and A. M. Sajiah, "Penerapan Hadoop untuk Analisis Sentimen Berbasis Big Data pada Ulasan Aplikasi Transportasi Online," *SATESI J. Sains Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 51–60, 2025, doi: 10.54259/satesi.v5i1.4051.

[10] B. Ramadhani and R. R. Suryono, "Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse," *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 714, 2024, doi: 10.30865/mib.v8i2.7458.

[11] E. R. Lidinillah, T. Rohana, and A. R. Juwita, "Analisis sentimen twitter terhadap steam menggunakan algoritma logistic regression dan support vector machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 154–164, 2023, doi: 10.37373/tekno.v10i2.440.

[12] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM: Comparative Evaluation of SVM Kernels for Sentiment Classification in Fuel Price Increase Analysis," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 153–160, 2023, [Online]. Available: https://journal.irpi.or.id/index.php/malcom/article/view/897%0Ahttps://journal.irpi.or.id/index.php/malcom/article/download/897/421

[13] D. Dablain, B. Krawczyk, and N. V. Chawla, "DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 34, no. 9, pp. 6390–6404, 2023, doi: 10.1109/TNNLS.2021.3136503.

[14] I. S. Ramadhan and A. Salam, "Teknik Random Undersampling untuk Mengatasi Ketidakseimbangan Kelas pada CT Scan Kista Ginjal," *Techno.Com*, vol. 23, no. 1, pp. 20–28, 2024, doi: 10.62411/tc.v23i1.9738.

[15] Y. Wu, Z. Jin, C. Shi, P. Liang, and T. Zhan, "Research on the application of deep learning-based BERT model in sentiment analysis," *Appl. Comput. Eng.*, vol. 71, no. 1, pp. 14–20, 2024, doi: 10.54254/2755-2721/71/2024ma.

[16] Y. Wen, Y. Liang, and X. Zhu, "Sentiment analysis of hotel online reviews using the BERT model and ERNIE model—Data from China," *PLoS One*, vol. 18, no. 3 March, pp. 1–14, 2023, doi: 10.1371/journal.pone.0275382.

[17] Vidya Chandradev, I Made Agus Dwi Suarjaya, and I Putu Agung Bayupati, "Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT," *J. Buana Inform.*, vol. 14, no. 02, pp. 107–116, 2023, doi: 10.24002/jbi.v14i02.7244.

[18] A. Ripa'i, F. Santoso, and F. Lazim, "Deteksi Berita Hoax dengan Perbandingan Website Menggunakan Pendekatan Deep Learning Algoritma BERT," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1749–1758, 2024, doi: 10.33379/gtech.v8i3.4541.

[19] N. Nurwanda, N. Suarna, and W. Prihartono, "Penerapan Nlp (Natural Language Processing) Dalam Analisis Sentimen Pengguna Telegram Di Playstore," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1841–1846, 2024, doi:

Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1467−1476
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.8310

10.36040/jati.v8i2.8469.
[20] M. R. A. Prasetya and A. M. Priyatno, "Dice Similarity and TF-IDF for New Student Admissions Chatbot," *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 1, no. 1, pp. 13–18, 2022, doi: 10.31004/riggs.v1i1.5.
[21] S. K. Rongali, "Natural Language Processing in Artificial Intelligence," *World J. Adv. Res. Rev.*, vol. 25, no. 1, pp. 1931–1935, 2025, doi: 10.1201/9780367808495.
[22] F. M. Sinaga, W. S. Lestari, S. Winardi, and K. H. Rambe, "ENHANCING SENTIMENT ANALYSIS ACCURACY WITH BERT AND SILHOUETTE METHOD OPTIMIZATION," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 11, no. 1, pp. 76–86, 2025, doi: 10.33480/jitk.v11i1.6392.Transformers.
[23] K. Pramayasa, I. M. D. Maysanjaya, and I. G. A. A. D. Indradewi, "Analisis Sentimen Program Mbkm Pada Media Sosial Twitter Menggunakan KNN Dan SMOTE," *SINTECH (Science Inf. Technol. J.*, vol. 6, no. 2, pp. 89–98, 2023, doi: 10.31598/sintechjournal.v6i2.1372.
[24] Candra, K. W. Chandra, and H. Irsyad, "Efektifitas SMOTE dalam Mengatasi Imbalanced Class Algoritma K-Nearest Neighbors pada Analisis Sentimen terhadap Starlink," *J. Ilmu Komput. dan Inform.*, vol. 4, no. 1, pp. 31–42, 2024, doi: 10.54082/jiki.132.
[25] B. Kurniawan, A. Suwarisman, I. Afriyanti, A. Wahyudi, and D. D. Saputra, "Analisis Sentimen Complain dan Bukan Complain pada Twitter Telkomsel dengan SMOTE dan Naïve Bayes," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 1, pp. 106–113, 2023, doi: 10.35870/jtik.v7i1.691.
[26] Z. A. Sriyanti, D. S. Y. Kartika, and A. R. E. Najaf, "Implementasi Model Bert Pada Analisis Sentimen Pengguna Twitter Terhadap Aksi Boikot Produk Israel," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 2335–2342, 2024, doi: 10.23960/jitet.v12i3.4743.
[27] Ardiansyah, Adika Sri Widagdo, Krisna Nuresa Qodri, F. E. N. Saputro, and Nisrina Akbar Rizky P, "Analisis sentimen terhadap pelayanan Kesehatan berdasarkan ulasan Google Maps menggunakan BERT," *J. Fasilkom*, vol. 13, no. 02, pp. 326–333, 2023, doi: 10.37859/jf.v13i02.5170.
[28] P. Wulff, L. Mientus, A. Nowak, and A. Borowski, "Utilizing a Pretrained Language Model (BERT) to Classify Preservice Physics Teachers' Written Reflections," *Int. J. Artif. Intell. Educ.*, vol. 33, no. 3, pp. 439–466, 2023, doi: 10.1007/s40593-022-00290-6.