

# Comparing TF-IDF Based SVM and Logistic Regression for Imbalanced Pertamina Corruption Tweet Sentiment Classification

Khahlil Gibran\*, Wenty Dwi Yuniarti, Khotibul Umam, Mokhamad Iklil Mustofa

Faculty Sains and Technology, Information of Technology, Universitas Islam Negeri Walisongo, Semarang, Indonesia

Email: <sup>1</sup>\*2208096096@student.walisongo.ac.id, <sup>2</sup>wenty@walisongo.ac.id, <sup>3</sup>khotibul\_umam@walisongo.ac.id, <sup>4</sup>iklil@walisongo.ac.id

Correspondence Author Email: 2208096096@student.walisongo.ac.id

Submitted: 20/04/2026; Accepted: 02/06/2026; Published: 05/06/2026

**Abstract**—The corruption case involving PT Pertamina (Persero) in early 2025 generated widespread public reactions on social media, particularly on the X (Twitter) platform. The rapid dissemination of opinions in digital environments highlights the importance of analyzing public sentiment toward socio-political issues. This study aims to examine public sentiment regarding the Pertamina corruption case using a text classification approach based on Term Frequency–Inverse Document Frequency (TF-IDF). This study contributes a controlled comparison of TF-IDF-based Support Vector Machine (SVM) and Logistic Regression on imbalanced Indonesian-language tweets related to a nationally salient corruption issue, while also emphasizing the importance of evaluating performance beyond accuracy alone through macro-F1 and minority-class recall. Two classification algorithms, Support Vector Machine (SVM) and Logistic Regression, were employed to compare their performance in predicting lexicon-derived positive and negative sentiment labels. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training data. A total of 3,058 Indonesian-language tweets collected between February 25 and March 10, 2025 underwent preprocessing and sentiment labeling using the INSET Lexicon. The results show that SVM achieved higher overall accuracy of 94.93% and a macro-F1 score of 0.80, while Logistic Regression achieved an accuracy of 90.52% and a macro-F1 score of 0.73. However, class-wise evaluation indicates that accuracy should not be interpreted independently because the dataset was dominated by negative sentiment. For the positive minority class, SVM obtained an F1-score of 0.64 and recall of 0.60, whereas Logistic Regression obtained a lower F1-score of 0.52 but a higher recall of 0.69. These findings indicate a trade-off between overall classification performance and minority-class sensitivity.

**Keywords:** Machine Learning; Text Classification; TF-IDF; SVM; Logistic Regression

## 1. INTRODUCTION

The alleged corruption case involving PT Pertamina (Persero) in early 2025 became the center of public attention and triggered a large volume of discussion in the digital space. Public attention increased significantly on the X platform, marked by the emergence of hashtags such as #korupsipertamina, which topped the national trending topics [1].

Public sentiment analysis is particularly important in this context, as public opinion on social media can influence public perception, institutional legitimacy, and encourage more responsive policy-making on corruption issues. To understand how public opinion is formed and disseminated on social media, various sentiment analysis methods have been widely used in previous studies. In the study conducted by [2], the SVM model was applied to analyze public sentiment towards the issue of lobster seed corruption in 2020.

Initial testing shows that the first framework with default parameters successfully produced an accuracy of 91.86%, *precision* of 94.05%, recall of 91.99%, and an *F1-Score* of 93.01%. All of these evaluation metric values exceeded the 90% threshold, indicating that this approach is not only effective in minimizing classification errors, but also consistent in detecting positive and negative sentiments with a high degree of reliability.

The combination of TF-IDF and Logistic Regression has also been proven to provide competitive results in previous studies [3], particularly in sentiment analysis of LinkedIn app reviews, which are not overly complex but have a large volume of data. By optimizing the coefficients to minimize classification errors, Logistic Regression is able to classify sentiment with high accuracy across various types of text datasets, including user reviews on professional platforms such as *LinkedIn*.

Thus, despite being algorithmically simple, Logistic Regression remains one of the most powerful and reliable methods of for text classification, especially in the context of sentiment analysis of user opinions on professional applications such as *LinkedIn* [3]. However, the main challenge in sentiment analysis is data class imbalance, where the amount of data with negative sentiment often dominates compared to positive or neutral sentiment. As a solution, the *Synthetic Minority Over-sampling Technique* (SMOTE) method is applied to balance the data distribution between existing classes [3].

It is important to note that the performance of classification methods such as TF-IDF + Logistic Regression and TF-IDF + SVM is highly dependent on the class distribution in the dataset. If the data is unbalanced, for example, if the number of negative tweets is much higher than positive tweets, then the model risks becoming biased towards the majority class. Based on the above, this study aims to classify lexicon-derived sentiment labels of public opinion on the X platform regarding the Pertamina corruption case by comparing the performance of SVM and Logistic Regression on data balanced using SMOTE. Although various studies have shown the effectiveness of SVM and Logistic Regression models in sentiment analysis, studies that specifically highlight the issue of national corruption with unbalanced and informal language data characteristics are still rare. Unlike previous studies that generally focus on sentiment analysis of product reviews or general social issues, this study highlights the 2025 Pertamina corruption

case as a current issue of national relevance. With this focus, this study extends prior work by comparing TF-IDF + SVM and TF-IDF + Logistic Regression on a Pertamina-related tweet dataset.

Empirical evidence regarding the performance of the Support Vector Machine (SVM) and Logistic Regression algorithms in classifying informal Indonesian-language texts with an unbalanced class distribution is discussed in this study. The application of the Synthetic Minority Over-sampling Technique (SMOTE) is an important approach in maintaining data proportionality and improving classification model performance in the context of socio-political sentiment analysis. This study provides empirical evidence on sentiment classification performance for the collected X dataset related to the Pertamina corruption case. The findings may serve as supporting evidence for understanding sentiment patterns during the observed period. Broader claims regarding transparency, public participation, or governance impact require further real-world validation.

Previous research analyzed public sentiment towards the corruption case of PT. Pertamina (Persero) on the social media platform X using *Support Vector Machine* (SVM). Data was collected through tweet harvesting with the keyword "pertamina corruption", then processed through text preprocessing stages such as tokenization, normalization, and stemming. To overcome class imbalance, the SMOTE method was used. The test results showed an increase in model accuracy from 89% to 96% after applying SMOTE, proving that this technique is effective in improving classification performance [1]. Previous research showed that Support Vector Machine can be effectively applied to sentiment analysis of public opinion data from X using TF-IDF-based text representation. However, unlike studies that compare several SVM kernels, this study specifically employs Linear SVM because TF-IDF features are typically high-dimensional and sparse. Linear SVM is therefore considered suitable for text classification tasks, especially when the objective is to build an efficient and interpretable baseline model for sentiment classification [4].

Previous research analyzed public sentiment toward insecurity phenomena on Platform X using Logistic Regression on Indonesian-language tweets. After text preprocessing and 10-fold cross-validation, the study reported that negative sentiment was dominant and that Logistic Regression achieved an average accuracy of 83.13%. This finding confirms that Logistic Regression remains effective for sentiment analysis on Platform X [5].

Previous research shows a comparison of performance between the Logistic Regression algorithm and TF-IDF feature-based Support Vector Classification (SVM) in sentiment analysis of movie reviews on the IMDB platform. In this study, a dataset consisting of 2,000 review data was used, with a balanced proportion of positive and negative sentiments, 1,000 data each. The data was then processed through preprocessing stages and divided into 70% training data and 30% test data. Other studies comparing TF-IDF-based SVM and Logistic Regression also indicate that both algorithms can produce competitive results on text sentiment datasets, although their performance may vary depending on feature distribution, dataset domain, and class balance [6],[7]. In addition, other research also tested various representation methods, including TF-IDF, and found that TF-IDF remains superior in the context of Indonesian social media comments [8].

In addition, previous research has shown that class imbalance remains a major issue in sentiment analysis because the majority class can dominate the learning process and inflate accuracy values [9].

Therefore, this study addresses that gap by comparing TF-IDF-based SVM and Logistic Regression under the same preprocessing, lexicon-derived labeling, and training-only SMOTE pipeline, while emphasizing macro-F1 and positive-class recall rather than accuracy alone.

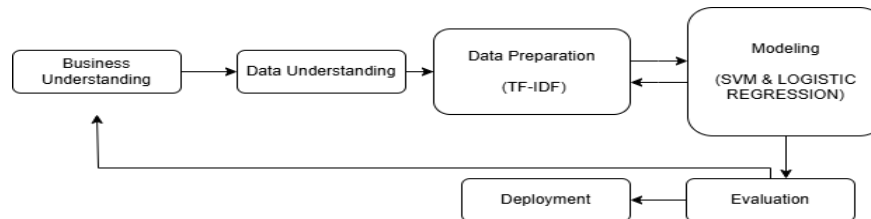
This study deliberately focuses on TF-IDF combined with Support Vector Machine and Logistic Regression for several methodological reasons. First, both algorithms provide interpretable and computationally efficient baselines, which are important for analyzing large-scale social media data. Logistic Regression, in particular, has been shown to remain a robust, transparent, and computationally efficient baseline for sentiment classification using TF-IDF features [10]. Meanwhile, SVM remains relevant in TF-IDF-based sentiment classification because previous studies have reported competitive performance of SVM/SVC when combined with TF-IDF feature representations [11], [7]. Second, previous studies have shown that TF-IDF combined with Logistic Regression and SVM remains effective for sentiment analysis, including Indonesian-language text data, where Logistic Regression and SVM achieved competitive performance in marketplace review classification [7]. Third, Because the dataset in this study contains imbalanced sentiment classes, SMOTE is applied only as a data balancing procedure to reduce potential bias toward the majority class and to ensure that the comparison between SVM and Logistic Regression is conducted under the same balanced data condition. Therefore, this study does not aim to evaluate the before-and-after effect of SMOTE, but focuses on comparing the classification performance of SVM and Logistic Regression. In this context, both algorithms are positioned as interpretable, efficient, and empirically relevant baseline models for Indonesian socio-political sentiment analysis.

Based on these previous studies, this research not only addresses the identified gap but also offers several specific contributions. First, it extends prior sentiment analysis studies by focusing on Indonesian socio-political tweets concerning the 2025 Pertamina corruption issue, a context that remains underexplored in previous comparative studies. Second, it systematically compares TF-IDF-based SVM and Logistic Regression within the same experimental setting, including identical preprocessing, INSET Lexicon-based reference labeling, and training-only SMOTE balancing. Third, this study provides methodological insight into the evaluation of imbalanced sentiment classification by showing the importance of interpreting model performance beyond accuracy alone. Accordingly, this study contributes to the development of more transparent and context-relevant sentiment classification research for Indonesian social media data

## 2. RESEARCH METHODOLOGY

### 2.1 CRISP-DM

This study uses the CRISP-DM (Cross-Industry Standard Process for Data Mining) method as a research implementation stage, using several classification models for sentiment analysis as shown in [12].



**Figure 1.** CRISP-DM Research Stages for TF-IDF-Based SVM and Logistic Regression Classification

Figure 1 shows the research stages based on the CRISP-DM framework, is a process model widely applied in data mining research. CRISP-DM is considered a de facto standard and an industry-independent process model for conducting data mining projects. This methodology consists of six iterative phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The Business Understanding phase focuses on identifying the research objectives and defining the data mining goals. The Data Understanding phase involves collecting, exploring, and assessing the quality of the data. The Data Preparation phase includes data selection, cleaning, and transformation to ensure that the data is suitable for analysis. In the Modeling phase, appropriate modeling techniques are selected and applied to build the model. The Evaluation phase is then carried out to assess whether the model results align with the research objectives. Finally, the Deployment phase involves presenting or implementing the research findings in the form of reports, recommendations, or system implementation [13].

In this study, TF-IDF is applied in the Data Preparation stage after preprocessing and sentiment labeling. The dataset is first divided into training and testing data using a stratified split. The TF-IDF vectorizer is then fitted only on the training data and applied to the test data to prevent data leakage. After the text data are transformed into TF-IDF vectors, SMOTE is applied only to the training TF-IDF vectors to balance the minority class. The prepared data are then used in the Modeling stage, where Support Vector Machine and Logistic Regression are trained as classification algorithms for positive and negative sentiment prediction

### 2.2 Business Understanding

This study aims to analyze categories public sentiment regarding the issue of corruption at Pertamina using a machine learning approach based on Term Frequency-Inverse Document Frequency. Two machine learning models, namely Support Vector Machine and Logistic Regression, were applied to build a classification model that can distinguish sentiment into positive and negative categories in tweet data. In addition, this study aims to evaluate which is the best between TF-IDF + SVM and TF-IDF-LR in the same pipeline with SMOTE. In many cases of sentiment analysis, the amount of data with negative sentiment is often far more dominant than positive sentiment, which can cause the model to be biased towards the majority class. To overcome this problem, the Synthetic Minority Over-sampling Technique (SMOTE) is used to balance the class distribution in the training data. Sentiment analysis is an approach used to identify and interpret opinions, emotions, and attitudes expressed in textual data. Sentiment analysis is a method within natural language processing that evaluates and identifies the emotional tone conveyed in text and classifies it into categories such as positive, negative, or neutral. This field has become an important area of research because it enables the extraction of valuable insights from large-scale textual data and is widely applied in natural language processing, text classification, machine learning, deep learning, and other intelligent data analysis tasks [14].

### 2.3 Data Understanding

The research process began with the collection of an Indonesian-language tweet dataset sourced from the X (Twitter) platform using a scraping process and then using tweet-harvest. The dataset was then filtered and cleaned of duplicate tweets, retweets, and irrelevant content to ensure the quality of the data to be analyzed [15]. Using the keyword "korupsi pertamina" during the period from February 25 to March 10, 2025, 3058 data points were obtained.

The dataset consisted of public tweets related to the corruption issue involving Pertamina, which became a topic of public discussion on X during the selected period. Each tweet was reviewed to understand the type of information contained in the dataset, including tweet text, posting date, and other available tweet attributes generated from the scraping process.

### 2.4 Data Preparation

#### 2.4.1 Preprocessing

After the data was collected, the preprocessing stage was carried out to prepare the raw data into a format suitable for sentiment analysis. This process includes several important stages such as text cleansing, tokenization, stopword



removal, and stemming using Indonesian NLP libraries such as Sastrawi. These stages aim to reduce noise in the data and simplify word representation [16].

Since the dataset was collected from platform X, the tweets may contain informal language, abbreviations, non-standard spelling, and slang expressions. These characteristics are common in social media text and may affect sentiment classification because informal words are not always represented in standard dictionaries or sentiment lexicons. In this study, informal textual forms were handled in a limited manner through text cleansing, tokenization, stopword removal, and stemming. However, this study did not apply a dedicated slang normalization dictionary. Therefore, slang terms that were not recognized by the INSET Lexicon or were not converted into standard forms may still affect the accuracy of sentiment labeling and classification. Recent studies on Indonesian sentiment analysis show that slang normalization can improve classification performance because informal words, abbreviations, and non-standard expressions on Twitter may reduce the quality of sentiment analysis if they are not normalized properly [17].

**Table 1.** Pre-processing Result

Full text	Clean text	Tokenization	Stop Removal	STEMMED
@DoankWarto Sejak era orla dipertamina banyak..	sejak era orla di pertamina banyak korupsi..	, 'sejak', 'era', 'orla', 'di', 'pertamina'...	'sejak', 'era', 'orla', 'pertamina', 'kor...'.	'sejak', 'era', 'orla', 'pertamina', 'kor...'.

Table 1 shows the results of the pre-processing stages, which include clean\_text, tokenization, stopword removal, and stemming. The clean\_text stage serves to clean the text of irrelevant elements such as URLs, numbers, symbols, emojis, and excessive punctuation so that the data becomes more structured and ready for processing. Next, the tokenization process is carried out to break sentences into word units so that analysis can take place at a more granular level. After that, stopword removal is applied to eliminate common words that do not contribute significantly to the main meaning of the text, thereby reducing noise and improving the quality of the features produced. Finally, stemming is performed to return each word to its base form in order to unify variations of words with the same meaning. This entire process ensures that the text data is in optimal condition for use in the feature extraction and classification model stages.

### 2.4.2 Labeling

After the pre-processing stage, this study uses a lexicon-based approach, which is a method that relies on an opinion dictionary to generate lexicon-based sentiment labels. In this context, the INSET Lexicon (Indonesian Sentiment Lexicon) is used, which is an Indonesian sentiment dictionary developed specifically for sentiment analysis on social media and microblogs[18].

INSET Lexicon was selected because it is an Indonesian sentiment lexicon specifically developed for microblog and social media contexts. It contains more than 10,000 sentiment words, consisting of 3,609 positive words and 6,609 negative words, with polarity weights ranging from -5 to +5, each tweet that has undergone preprocessing will be analyzed based on the presence of words found in the INSET Lexicon. The sentiment score was calculated as a lexicon-derived polarity score based on sentiment-bearing words found in each tweet.

INSET is widely used in Indonesian lexicon-based sentiment analysis [19]. While other research use InSet Lexicon as training data in supervised machine learning and describe it as an effective alternative for large-scale data labeling [20].

**Table 2.** Labeling Result

Text	Label	Score
sejak era orla pertamina korupsi rezim bongka...	Negative	-26.0
maaf ga bikin uang masyarakat turun harga per...	Negative	-4.0

Table 2 presents an example of the results of the sentiment labeling process, which consists of three main components, namely the original text (*Text*), sentiment category (*Label*), and sentiment score (*Score*). The *Text* column contains excerpts from the collected data, which in this example reflect public opinion on the issue of corruption at Pertamina. The *Label* column shows the automatic sentiment labels generated by the INSET Lexicon-based labeling procedure with the category ‘Negative’ indicating that, according to the lexicon-based scoring scheme, the text contains more or stronger negative polarity terms. The *Score* column provides a quantitative value that represents the lexicon-derived polarity strength, not a definitive measurement of human-perceived sentiment intensity. Within the lexicon-based scoring scheme, the lower the negative score, the stronger the negative polarity tendency. Thus, this table not only describes sentiment categories but also shows the weight or degree of sentiment strength in each text.

This methodological framing is also consistent with self-supervised sentiment classification research. Other research explain that when annotated data are unavailable, a lexicon-based method can be used to generate reference labels before training a supervised machine learning classifier [21]. The generated labels are treated as lexicon-derived reference labels, not as human ground-truth labels. Therefore, model performance should be interpreted as agreement

with INSET-based labels. This study also acknowledges that lexicon-based labeling may not fully capture sarcasm, irony, implicit sentiment, or target-dependent meaning [22].

### 2.4.3 Feature Extraction Using TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) is a statistical weighting method used to represent text numerically by combining the frequency of a term in a document with its inverse frequency across the corpus. A term receives a higher weight when it appears frequently in a document but rarely appears in other documents [23]. The Term Frequency (TF) measures how often a term appears in a document and is formulated as:

$$TF = \frac{t}{d} \tag{1}$$

In this equation,  $t$  indicates the frequency of a word's appearance in a document, while  $d$  describes the total number of words in the document. The Inverse Document Frequency (IDF) formula is presented in Equation (2).

$$idf = \log\left(\frac{N}{df(t)}\right) \tag{2}$$

In this case,  $N$  represents the total number of available documents, while  $df(t)$  indicates the number of documents containing word  $t$ . The formulation for calculating TF-IDF can be seen in Equation (3).

$$TFidf = TF \cdot idf \tag{4}$$

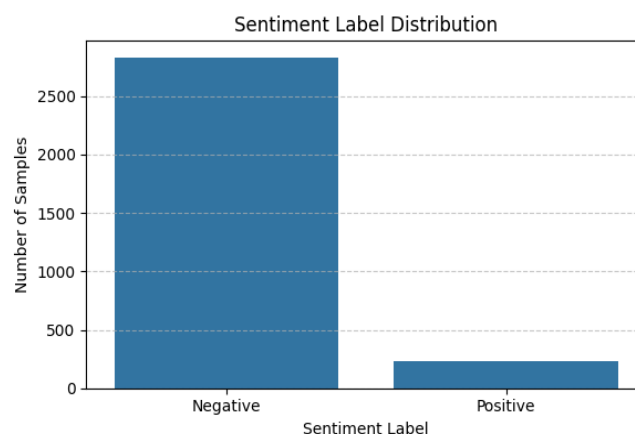
In this research TF-IDF was applied after Labeling process with splitting the dataset into training and testing data using a stratified split then transforming the text into TF-IDF vectors, where the vectorizer was fitted on the training data and applied to the test data after that applying SMOTE only to the training TF-IDF vectors to balance the minority class and prevent data leakage

### 2.4.4 SMOTE

SMOTE is an oversampling method used to address class imbalance by generating synthetic samples for the minority class based on the characteristics of nearby minority samples [24], [25]. The following is the formula for the SMOTE technique:

$$x_{syn} = x_i + (x_{knn} - x_i) \times \delta \tag{5}$$

Where  $x_{syn}$  represents the synthetic data point,  $x_i$  represents the minority sample to be replicated,  $x_{knn}$  represents one of its nearest neighbors, and  $\delta$  is a random value between 0 and 1. The SMOTE (Synthetic Minority Over-sampling Technique) technique was used to address data disparity issues in the sentiment classification process. SMOTE serves to balance the proportion of data between classes by adding synthetic samples to the minority class. In this study, SMOTE was applied only after the train-test split and only to the training TF-IDF vectors to prevent data leakage [26], [27]. The test data remained unchanged and non-synthetic. Therefore, SMOTE outputs are interpreted as synthetic numerical vectors in the TF-IDF feature space, not as linguistically meaningful tweets.



**Figure 2.** Data Distribution Before SMOTE

Figure 2 shows the initial distribution of sentiment labels in the dataset before the balancing process. It can be seen that the number of data with Negative labels is very dominant, reaching more than 2.700 data, while Positive labels are only in the range of 250 data. Figure 2 shows the initial distribution of sentiment labels before the balancing process. The Negative class is highly dominant, while the Positive class contains far fewer samples. This imbalance indicates that the dataset has a significant class imbalance problem. Such a condition may cause classification models to become biased toward the majority class and fail to recognize minority-class patterns effectively

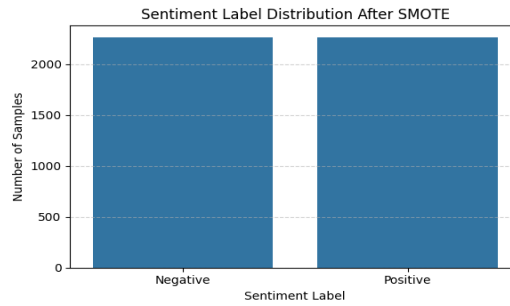


Figure 3. Data Distribution After SMOTE

Figure 3 shows the label distribution after applying the Synthetic Minority Over-sampling Technique (SMOTE). This method works by synthetic numerical feature vectors in the TF-IDF feature space. The visualization results show that both classes have the same amount of data, approximately 2,200 data points each. Figure 3 shows the class distribution after applying SMOTE to the training data. The visualization indicates that SMOTE produced a more balanced class distribution between the Negative and Positive classes in the training feature space.

## 2.5 Modeling

### 2.5.1 Classification Using Support Vector Machine

Support Vector Machine (SVM) is a supervised learning algorithm that aims to determine the optimal hyperplane separating two classes with maximum margin. The hyperplane acts as a decision boundary between classes. SVM works by maximizing the distance between the closest training samples (support vectors) and the decision boundary.

Although originally developed for binary classification, SVM has been extensively applied in text classification problems[28]. In this study, Linear SVM was employed to classify tweets into positive and negative sentiment categories using the balanced TF-IDF feature vectors obtained from the previous stage. Linear SVM was selected because it is effective for high-dimensional sparse text features and provides a strong baseline for text classification tasks. The trained model was then used to predict sentiment labels in the test set for performance comparison. The hyperplane function is defined as:

$$(\mathbf{w} \cdot \mathbf{x}) + \mathbf{b} = 0 \quad (6)$$

The decision boundaries for class separation are:

$$(\mathbf{w} \cdot \mathbf{x}) + \mathbf{b} = +1$$

$$(\mathbf{w} \cdot \mathbf{x}) + \mathbf{b} = -1 \quad (7)$$

Previous research [29] applying SVM for sentiment analysis related to internet service providers in Indonesia demonstrates its effectiveness in handling complex and unstructured social media data.

### 2.5.2 Classification Using Logistic Regression

Logistic Regression is a probabilistic classification method used to model the relationship between independent variables and a binary dependent variable. The dependent variable is encoded as 1 (positive) and 0 (negative) [30]. In sentiment analysis. In this study, Logistic Regression was trained using the same balanced TF-IDF training features as the SVM model, so that both classifiers were compared under identical preprocessing and data balancing conditions. Logistic Regression was included as a comparative baseline because of its interpretability, computational efficiency, and widespread use in binary text classification. Logistic Regression estimates the probability that a document belongs to a particular class. The logistic regression model is expressed as:

$$\text{logit}(S) = \mathbf{b}_0 + \mathbf{b}_1\mathbf{M}_1 + \mathbf{b}_2\mathbf{M}_2 + \mathbf{b}_3\mathbf{M}_3 + \dots + \mathbf{b}_k\mathbf{M}_k \quad (8)$$

The logit function, denoted as  $\text{logit}(S)$ , represents the logarithm of the odds that an observation belongs to a particular class. In this context,  $S$  is the probability that an observation falls into the target class, such as positive sentiment, while  $1 - S$  represents the probability that it does not.

The term  $\mathbf{b}_0$  refers to the constant or intercept in the logistic regression model. It represents the baseline value of the log-odds when all predictor attributes are equal to zero. The coefficients  $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_k$  represent the logistic regression coefficients, which indicate the direction and magnitude of each predictor's influence on the dependent variable. The variables  $\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_k$  represent the predictor attributes or features used in the model. In text classification, these may include values such as the TF-IDF weight of particular words or terms. The symbol  $k$  denotes the total number of predictor attributes included in the model

A previous study [24] demonstrated that Logistic Regression combined with SMOTE effectively addressed data imbalance in sentiment analysis related to the 2024 Indonesian elections. The results indicate that Logistic Regression remains competitive when paired with appropriate feature representation and balancing techniques.

Based on these findings, Logistic Regression is employed in this study as a comparative baseline to SVM, providing a comprehensive evaluation of both models in the context of public sentiment classification. In the modeling stage, the sentiment-labeled dataset was transformed into TF-IDF feature vectors after being split into stratified training and testing sets. The TF-IDF vectorizer was fitted on the training data and then applied to the test data. To address class imbalance while avoiding data leakage, SMOTE was applied only to the training TF-IDF vectors. The balanced training features were then used to train two classification models, namely Linear Support Vector Machine (SVM) and Logistic Regression, under the same experimental setting

## 2.6 Evaluation

In this study, model evaluation is conducted to assess the extent to which the Support Vector Machine (SVM) and Logistic Regression algorithms are capable of accurately performing text classification on social media X data after undergoing feature representation using the Term Frequency–Inverse Document Frequency (TF-IDF) method. This evaluation stage aims to measure the performance of both models based on the testing dataset, which has not been previously seen during the training process.

Model performance is evaluated using several key metrics commonly applied in text classification tasks, namely accuracy, precision, recall, and F1-score, which are derived from the confusion matrix. In this context, \*True Positive\* (TP) represents the number of text data instances that are correctly classified into a particular sentiment category in accordance with their actual labels. Meanwhile, True Negative (TN) indicates the number of instances that are correctly predicted as not belonging to a given class. False Positive (FP) occurs when a data instance is incorrectly classified into a certain class, whereas False Negative (FN) represents instances that should belong to a specific class but are not correctly identified by the model.

These four components form the basis for calculating the evaluation metrics and provide a comprehensive overview of the model’s performance in terms of correctness (accuracy), precision (precision), completeness (recall), and balance between precision and recall (F1-score). This evaluation enables a systematic comparison between SVM and Logistic Regression in performing text classification on social media X data.

### a. Accuracy

Accuracy is used to measure the proportion of correct predictions compared to all predictions made by the model [31]. The calculation of the accuracy value in this study was carried out using Equation (9)

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

### b. Precision

Precision describes the extent to which the model is able to provide accurate predictions for each class without producing many false positives[32].

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

### c. Recall

Recall is computed as the ratio of positive samples that were properly categorized as positive to the total number of positive samples[33].

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

### d. F1-Score

The F1-score is the harmonic mean between precision and recall, which provides an overview of the balance of model performance in each class[34]. The F1-score calculation in this study refers to Equation (11).

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (11)$$

## 2.7 Deployment

The CRISP-DM deployment stage is the implementation and utilization of the model in the real world or production environment. Within the CRISP-DM framework, deployment is treated as a potential future application rather than an implemented system in this study. The evaluated pipeline could be adapted in future work for integration into dashboards or monitoring tools, subject to additional validation on continuously collected data.

# 3. RESULT AND DISCUSSION

## 3.1 Result

### 3.1.1 Data Collection and Preprocessing

This study began with the collection of Indonesian-language tweets from the X platform using the keyword “korupsi pertamina” during the period from February 25, 2025 to March 10, 2025. A total of 3,058 tweets were obtained and

used as the dataset for this research. After the data were collected, an initial filtering process was conducted to remove duplicate tweets, retweets, and irrelevant content in order to ensure that the dataset was relevant to the research topic and suitable for further analysis.

Following data collection, the dataset underwent a preprocessing stage to transform raw social media text into a cleaner and more structured format. The preprocessing steps included text cleaning, tokenization, stopword removal, and stemming using Indonesian natural language processing tools such as Sastrawi. In the text cleaning stage, irrelevant elements such as URLs, mentions, numbers, symbols, emojis, and excessive punctuation were removed. The tokenization process then split each tweet into individual word units to enable more detailed analysis. After that, stopword removal was applied to eliminate common words that do not contribute significantly to sentiment meaning. Finally, stemming was performed to convert inflected words into their root forms, thereby reducing lexical variation and improving consistency in text representation. The overall results of these preprocessing stages are presented in Table 1, which shows the transformation from `full_text` into `clean_text`, tokenization, `stop_removal`, and stemmed forms. Overall, this stage produced a more structured and analysis-ready dataset for the subsequent sentiment labeling and classification processes.

### 3.1.2 Sentiment Labeling

After preprocessing, each tweet was assigned a sentiment label using a lexicon-based approach with the INSET Lexicon. In this stage, the sentiment of each tweet was determined by identifying sentiment-bearing words contained in the lexicon and calculating a sentiment score based on their polarity. Based on the resulting score, each tweet was classified into either the positive or negative sentiment category.

However, the labels generated through this process are not treated as absolute human ground-truth labels. Instead, they are treated as lexicon-derived reference labels. This distinction is important because the collected social media dataset does not contain manually annotated sentiment labels. Therefore, INSET Lexicon is used as an automatic labeling mechanism to enable supervised sentiment classification on an unlabeled Indonesian social media dataset.

The labeling results are illustrated in Table 2, which includes three main components: Text, Label, and Score. The Text column contains the processed tweet content, the Label column shows the assigned sentiment category, and the Score column represents the intensity of the sentiment polarity. A lower negative score indicates a stronger negative sentiment tendency. This labeling stage produced a sentiment-labeled dataset that served as the basis for the classification models. In addition, the labeling results showed that the negative class was substantially more dominant than the positive class, indicating the presence of class imbalance in the dataset.

### 3.1.3 Data Balancing Using SMOTE

Based on the sentiment labeling results, the dataset exhibited a clear class imbalance, where the number of negative tweets was much larger than the number of positive tweets. Such an imbalance can affect the learning process of classification algorithms, as models tend to favor the majority class and may fail to adequately capture patterns from the minority class. To address this issue, this study applied the Synthetic Minority Over-sampling Technique (SMOTE) as a data balancing method.

The distribution of sentiment classes before balancing is shown in Figure 2, where the negative class clearly dominates the dataset. After applying SMOTE to the training data, the class distribution became more balanced, as illustrated in Figure 3. This more proportional class distribution provided a better training condition for the classification models, especially in learning the characteristics of the minority class. In this study, SMOTE was used as part of the experimental pipeline so that the comparison between Support Vector Machine (SVM) and Logistic Regression was conducted under the same balanced training condition.

SMOTE was applied only to the training data after TF-IDF feature extraction. Therefore, the synthetic samples generated by SMOTE were not interpreted as new tweets or linguistically meaningful textual data. Instead, they were treated as synthetic numerical feature vectors in the TF-IDF feature space. This distinction is important because TF-IDF represents text as weighted numerical values, not as natural language sentences.

The use of SMOTE in this study was therefore limited to balancing the class distribution in the training feature space. The generated synthetic vectors were used to help the classification models learn minority-class patterns more effectively under imbalanced data conditions. However, this study acknowledges that SMOTE does not preserve full linguistic or semantic meaning when applied to TF-IDF features. Thus, the results should be interpreted as the performance of machine learning models trained on balanced numerical feature representations, not as evidence that synthetic samples represent valid natural-language tweets.

### 3.1.4 Modeling and Evaluation

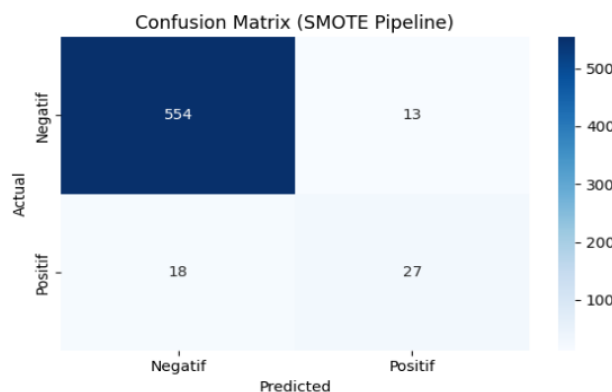
In the modeling stage, the sentiment-labeled text data were transformed into numerical features using Term Frequency–Inverse Document Frequency (TF-IDF). These features were then used as input for two classification algorithms: Support Vector Machine (SVM) and Logistic Regression. Model performance was evaluated on the test set using accuracy, precision, recall, and F1-score.

Based on the testing results, the SVM model achieved an accuracy of 0.95. However, this value should be interpreted cautiously because the test set remained highly imbalanced, with 567 negative samples and only 45 positive samples. A majority-class prediction strategy would already produce a high baseline accuracy of approximately 92.65%. Therefore, accuracy alone is insufficient to evaluate the model’s performance. Class-specific recall, precision, F1-score, macro average, and confusion matrix analysis are needed to assess whether the model can recognize both negative and positive sentiment categories.

**Table 3.** SVM Evaluation Matrix

	Precision	Recall	F1-Score	Support
Negative	0.97	0.98	0.97	567
Positive	0.68	0.60	0.64	45
Accuracy			0.95	612
Macroavg	0.82	0.79	0.80	612
Weighted avg	0.95	0.95	0.95	612

Table 3 shows that Linear SVM achieved an accuracy of 0.95 and a macro-F1 score of 0.80. However, this accuracy should be interpreted carefully because the test set was dominated by negative sentiment. For the negative class, SVM achieved strong performance with precision of 0.97, recall of 0.98, and F1-score of 0.97. In contrast, the positive class obtained lower performance, with precision of 0.68, recall of 0.60, and F1-score of 0.64. The positive-class recall indicates that SVM missed 40% of positive tweets. Therefore, although SVM produced strong overall performance, it remained less sensitive to minority positive sentiment. The macro-F1 score of 0.80 provides a more balanced view of model performance than accuracy because it gives equal weight to both classes. Meanwhile, the weighted average F1-score of 0.95 is heavily influenced by the dominant negative class. Therefore, the SVM model can be considered strong in overall classification and majority-class recognition, but still limited in detecting positive sentiment.



**Figure 4.** Confusion Matrix Result of the SVM Algorithm

Figure 4 presents the confusion matrix results of the SVM model using the TF-IDF and SMOTE pipeline. The model was evaluated on the test data using lexicon-derived reference labels generated from the InSet Lexicon. Based on the confusion matrix, the SVM model shows strong performance in identifying negative sentiment, with 554 true negatives and only 13 false positives. This indicates that most tweets labeled as negative by the InSet Lexicon were also predicted as negative by the SVM model.

However, the model still shows limitations in detecting positive sentiment. The SVM model correctly classified 27 positive tweets, while 18 positive tweets were misclassified as negative. This result indicates that the model is more effective in recognizing negative sentiment patterns than positive ones.

From a linguistic perspective, these false negative errors may occur because positive tweets in socio-political discourse often contain negative lexical items. For example, a tweet may express support for a certain political actor or policy while simultaneously criticizing an opposing group. In this case, negative words may dominate the TF-IDF representation, causing the SVM model to classify the tweet as negative even though the overall sentiment label from the InSet Lexicon is positive.

In addition, socio-political tweets in Indonesian frequently contain sarcasm, irony, informal expressions, abbreviations, hashtags, and context-dependent political terms. These linguistic characteristics are difficult for TF-IDF-based models to capture because TF-IDF mainly represents word frequency and importance, rather than pragmatic meaning or sentence-level context. Therefore, the misclassification produced by the SVM model is not only related to class imbalance, but also to the linguistic complexity of Indonesian socio-political tweets.

Overall, the SVM model performs well in minimizing false positive errors in the negative class. Nevertheless, its lower sensitivity toward positive sentiment suggests that further improvements are needed, particularly by incorporating contextual or semantic features that can better capture implicit sentiment, sarcasm, and target-dependent

expressions. As a comparison, a model was trained using the Logistic Regression (LR) algorithm with a similar preprocessing and oversampling pipeline, namely using the SMOTE method to handle class imbalance.

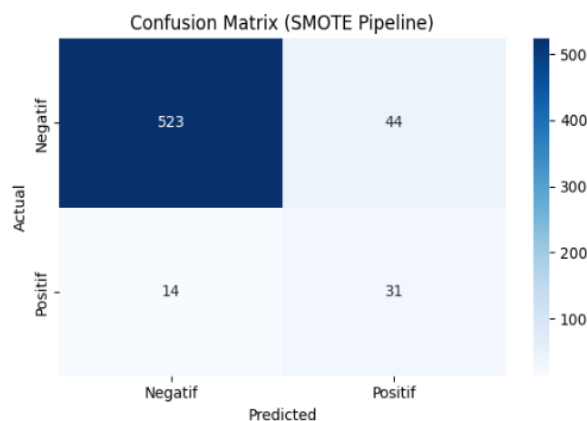
**Table 4.** Logistic Regression Evaluation Matrix

	Precision	Recall	F1-Score	Support
Negative	0.97	0.92	0.95	567
Positive	0.41	0.69	0.52	45
Accuracy			0.91	612
Macro avg	0.69	0.81	0.73	612
Weighted avg	0.93	0.91	0.92	612

Table 4 shows that Logistic Regression achieved lower overall accuracy than SVM, namely 0.91. This accuracy is lower than the SVM accuracy and should also be interpreted cautiously because the test set is dominated by the negative class. Unlike SVM, Logistic Regression showed higher sensitivity toward the positive class, as reflected by its positive-class recall of 0.69. This means that Logistic Regression correctly identified 69% of positive tweets, which is higher than the SVM positive recall of 0.60.

However, this improvement in positive recall came with a trade-off. Logistic Regression produced a low positive-class precision of 0.41, meaning that many tweets predicted as positive were actually negative. This is also reflected in the confusion matrix, where 44 negative tweets were misclassified as positive. Therefore, Logistic Regression was more aggressive in detecting positive sentiment, but less selective in ensuring that positive predictions were correct.

In contrast, SVM produced fewer false positives but missed more positive tweets. This indicates a different decision behavior between the two models. Logistic Regression tends to increase minority-class sensitivity, while SVM tends to maintain a more conservative decision boundary that produces better overall balance as shown by its higher macro-F1 score



**Figure 5.** Confusion Matrix Results of the Logistic Regression Algorithm

Figure 5 shows the confusion matrix results of the Logistic Regression model using the same TF-IDF and SMOTE pipeline. Similar to the SVM model, Logistic Regression was evaluated using lexicon-derived reference labels generated from the InSet Lexicon. The model correctly classified 523 negative tweets as true negatives, while 44 negative tweets were incorrectly classified as positive. This indicates that Logistic Regression is more likely than SVM to predict tweets as positive, even when the reference label is negative. For the positive class, Logistic Regression achieved better performance than SVM. The model correctly classified 31 positive tweets, while 14 positive tweets were misclassified as negative. This shows that Logistic Regression has higher sensitivity toward the positive class, as reflected in its better positive recall.

The false positive errors in Logistic Regression may be caused by the presence of positive words that are used in a negative or sarcastic context. In socio-political tweets, users often use seemingly positive expressions to criticize political figures, policies, or opposing groups. As a result, the model may interpret these positive lexical cues as indicators of positive sentiment, even though the intended meaning of the tweet is negative.

Meanwhile, false negative errors may occur when tweets labeled as positive by the InSet Lexicon contain negative expressions directed at an opposing political group rather than at the main target of support. This reflects the problem of target-dependent sentiment, where the sentiment orientation depends on which actor, issue, or group is being evaluated. Since Logistic Regression with TF-IDF features does not fully capture this contextual relationship, some positive tweets are still classified as negative.

These findings suggest that linguistic misclassification in the Logistic Regression model is influenced by sarcasm, negation, informal language, political slang, and context-dependent expressions. Therefore, the errors shown

in the confusion matrix should not be interpreted only as numerical classification errors, but also as evidence of the linguistic challenges involved in classifying Indonesian socio-political tweets.

Overall, Logistic Regression demonstrates better sensitivity in detecting the positive class than SVM, although this comes with a higher number of false positives. This trade-off shows that Logistic Regression is more responsive to positive sentiment patterns, but it is also more vulnerable to misinterpreting positive lexical cues that appear in negative or sarcastic contexts.

### 3.2 Discussion

Based on the evaluation results, SVM and Logistic Regression showed different classification behaviors in handling Indonesian social media text represented using TF-IDF features. The SVM model achieved higher overall accuracy and macro-F1 than Logistic Regression. However, the high accuracy should be interpreted cautiously because the test set was dominated by the negative class. A majority-class baseline would already produce high accuracy, so accuracy alone is not sufficient to determine the best model.

The SVM model achieved an accuracy of 0.95 and a macro-F1 score of 0.80, which indicates better overall balance than Logistic Regression. However, its positive-class recall was only 0.60, meaning that 40% of positive tweets were misclassified as negative. This shows that SVM was effective in recognizing dominant negative sentiment patterns, but less sensitive to minority positive sentiment. In other words, SVM provided more stable overall performance but still showed weakness in detecting positive opinions.

Logistic Regression, on the other hand, achieved lower accuracy of 0.91 and lower macro-F1 of 0.73, but produced higher positive-class recall of 0.69. This indicates that Logistic Regression was more sensitive in detecting positive tweets than SVM. However, this sensitivity came at the cost of lower positive-class precision of 0.41, meaning that many negative tweets were incorrectly predicted as positive. Therefore, Logistic Regression showed a recall-oriented behavior for the minority class, while SVM showed a more precision-oriented and conservative behavior.

From a computational and feature-distribution perspective, both SVM and Logistic Regression are linear models that operate on high-dimensional sparse TF-IDF vectors. SVM attempts to find a maximum-margin hyperplane, which can produce a more conservative separation between classes when the dominant negative class has clearer and more frequent lexical patterns. This may explain why SVM produced fewer false positives and achieved higher macro-F1. In contrast, Logistic Regression estimates class probabilities through a linear combination of TF-IDF features. After SMOTE balancing, Logistic Regression became more sensitive to minority-class regions in the feature space, which improved positive recall but also increased false positive predictions.

The confusion matrix results further show that model errors were not merely numerical errors but were also related to linguistic characteristics of social media text. Tweets about corruption often contain negative lexical cues, even when the intended sentiment may not be fully negative. As a result, positive tweets containing corruption-related terms may be misclassified as negative. In addition, slang, informal expressions, implicit sentiment, and sarcasm may reduce the ability of TF-IDF-based models to capture the intended meaning of a tweet. Therefore, the results suggest that future work should include linguistic error analysis, slang normalization, sarcasm-aware modeling, and comparison with contextual embedding-based models

## 4. CONCLUSION

Based on the results and discussion, both Support Vector Machine and Logistic Regression were able to perform sentiment classification on Indonesian-language tweets related to the Pertamina corruption issue using the TF-IDF and SMOTE-based pipeline. However, the two models showed different performance characteristics. SVM achieved higher overall accuracy and macro-F1, indicating stronger overall balance across evaluation metrics. Nevertheless, the high accuracy should be interpreted cautiously because the test set was dominated by negative sentiment. A majority-class baseline would already produce high accuracy, so class-specific metrics provide a more meaningful interpretation of model performance. The SVM model achieved stronger performance in recognizing the dominant negative class and produced fewer false positive errors. However, its positive-class recall of 0.60 indicates that 40% of positive tweets were still misclassified as negative. Logistic Regression, although having lower overall accuracy and macro-F1, achieved higher positive-class recall of 0.69. This indicates that Logistic Regression was more sensitive in detecting positive sentiment, but this came with lower precision and more false positive predictions. Therefore, the main finding of this study is not simply that SVM outperformed Logistic Regression, but that both models produced different trade-offs. SVM was more stable for overall classification, while Logistic Regression was more sensitive to minority-class positive sentiment. This finding is important because sentiment classification on social media data should not rely only on accuracy, especially when the dataset is imbalanced. The contribution of this study lies not only in comparing two traditional classifiers, but also in providing an empirical benchmark for imbalanced Indonesian socio-political sentiment classification on Platform X under the same preprocessing, lexicon-derived labeling, and training-only SMOTE setting. In addition, this study shows that model selection in imbalanced sentiment classification should not rely solely on accuracy, but should also consider macro-F1 and minority-class recall to obtain a more balanced interpretation of performance. This study also has several limitations. First, the sentiment labels were generated using a lexicon-based approach and should be interpreted as lexicon-derived reference labels rather than



human ground-truth labels. Second, SMOTE was applied only in the TF-IDF numerical feature space and does not generate linguistically meaningful tweets. Third, slang, sarcasm, implicit sentiment, and contextual meaning were not fully handled by the current pipeline. Future research should include manual label validation, inter-annotator agreement, comparison between lexicon-derived labels and human-annotated labels, slang normalization, sarcasm detection, linguistic error analysis of misclassified tweets, class weighting, threshold tuning, and contextual embedding or transformer-based models. These future directions are added to address the limitations related to labeling validity, minority-class detection, slang, sarcasm, and contextual meaning.

## REFERENCES

- [1] R. D. Pebrianti, "Analisis Sentimen Masyarakat Platform X Terhadap Korupsi PT. Pertamina (PERSERO) Menggunakan SVM," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 2, Apr 2025, doi: 10.23960/jitet.v13i2.6399.
- [2] B. Pamungkas, M. E. Purbaya, dan D. J. A. K., "Analisis Sentimen Twitter Menggunakan Metode Support Vector Machine (SVM) pada Kasus Benih Lobster 2020," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 3, no. 2, 2021, doi: 10.20895/inista.v3i2.243.
- [3] N. S. Wardana, F. P. Aditiawan, dan A. P. Sari, "Logistic Regression Classification with TF-IDF and FastText for Sentiment Analysis of LinkedIn Reviews," *VISA J. Vis. Ideas*, vol. 4, no. 3, Agu 2024, doi: 10.47467/visa.v4i3.2835.
- [4] M. Rahardi, A. Aminuddin, F. F. Abdulloh, dan R. A. Nugroho, "Sentiment Analysis of Covid-19 Vaccination using Support Vector Machine in Indonesia," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 6, 2022, doi: 10.14569/IJACSA.2022.0130665.
- [5] E. Safitri, W. A. Syukrilla, dan I. N. L. Fitriana, "Logistic Regression for Sentiment Analysis of Insecurity Phenomena on Platform X," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 18, no. 1, Jul 2025, doi: 10.36456/jstat.vol18.no1.a10545.
- [6] D. S. Ramdan, R. D. Apnena, dan C. A. Sugianto, "Film Review Sentiment Analysis: Comparison of Logistic Regression and Support Vector Classification Performance Based on TF-IDF," *J. Appl. Intell. Syst.*, vol. 8, no. 3, Nov 2023, doi: 10.33633/jais.v8i3.9090.
- [7] C. A. Hutagalung dan V. B. Lestari, "Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes," *J-KOMA J. Ilmu Komput. dan Apl.*, vol. 8, no. 1, hal. 33–42, Jun 2025, doi: 10.21009/j-koma.v8i1.05.
- [8] H. Suroyo dan E. J. Pratama, "Comparison of Text Representation Methods for Sentiment Analysis Using Support Vector Machine," *J. Adv. Inf. Ind. Technol.*, vol. 7, no. 1, hal. 21–30, Mei 2025, doi: 10.52435/jaiit.v7i1.610.
- [9] Y. Irawan, R. Wahyuni, R. Ordila, dan Herianto, "Comparative Analysis of Machine Learning Algorithms with SMOTE and Boosting Techniques in Accuracy Improvement," *Indones. J. Comput. Sci.*, vol. 13, no. 5, Okt 2024, doi: 10.33022/ijcs.v13i5.4368.
- [10] N. A. Smary, W. Ahmed, K. Amin, P. Pławiak, dan M. Hammad, "Enhancing Machine Learning-Based Sentiment Analysis Through Feature Extraction Techniques," *PLoS One*, vol. 19, no. 2, hal. e0294968, Feb 2024, doi: 10.1371/journal.pone.0294968.
- [11] D. M. Ulya, J. Juhari, R. E. Yuliana, dan M. Jamhuri, "Reliable and Efficient Sentiment Analysis on IMDb with Logistic Regression," *CAUCHY J. Mat. Murni dan Apl.*, vol. 10, no. 2, hal. 821–834, Agu 2025, doi: 10.18860/cauchy.v10i2.33809.
- [12] L. A. Fitriana, S. Linawati, N. Herlinawati, R. Sa'adah, dan S. Seimahuria, "Analisis Sentimen Pengguna Twitter terhadap Brand Indosat Menggunakan Metode Naive Bayes Classifier," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 3, hal. 4291–4297, Jun 2024, doi: 10.36040/jati.v8i3.9866.
- [13] C. Schröer, F. Kruse, dan J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Comput. Sci.*, vol. 181, hal. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [14] J. R. Jim, M. A. R. Talukder, P. Malakar, M. M. Kabir, K. Nur, dan M. F. Mridha, "Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review," *Nat. Lang. Process. J.*, vol. 6, no. 12, hal. 100059, Mar 2024, doi: 10.1016/j.nlp.2024.100059.
- [15] R. I. Syah, H. Hoiriyah, dan M. Walid, "Analisis Sentimen Pengguna Media Sosial Terhadap Aplikasi M-Health Peduli Lindungi Dengan Metode Lexicon Based Dan Naïve Bayes," *Indones. J. Bus. Intell.*, vol. 6, no. 1, Jun 2023, doi: 10.21927/ijubi.v6i1.3275.
- [16] A. Addiga dan S. Bagui, "Sentiment Analysis on Twitter Data Using Term Frequency-Inverse Document Frequency," *J. Comput. Commun.*, vol. 10, no. 08, hal. 117–128, 2022, doi: 10.4236/jcc.2022.108008.
- [17] A. Bustamin, A. A. Prayogi, D. Siswanto, M. Rafrin, dan A. Nurdin, "Text Normalization for Indonesian Slang Words in Sentiment Analysis Development," *ICIC Express Lett. Part B Appl.*, vol. 16, no. 2, 2025, doi: 10.24507/iceiclb.16.02.121.
- [18] F. Koto dan G. Y. Rahmangtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," in *2017 International Conference on Asian Language Processing (IALP)*, IEEE, Des 2017, hal. 391–394. doi: 10.1109/IALP.2017.8300625.
- [19] A. Rufaida, A. Permanasari, dan N. Setiawan, "Lexicon-Based Sentiment Analysis Using Inset Dictionary: A Systematic Literature Review," in *Proceedings of the 5th International Conference on Applied Engineering, ICAE 2022, 5 October 2022, Batam, Indonesia*, EAI, 2023. doi: 10.4108/eai.5-10-2022.2327474.
- [20] N. A. Daulay, Rifqi Ramadhan, dan Lya Hulliyatus Suadaa, "Sentiment Classification of Community towards COVID-19 Issues on Twitter (Case Study: Indonesia, March-May 2020)," *Proc. Int. Conf. Data Sci. Off. Stat.*, vol. 2023, no. 1, hal. 201–217, Des 2023, doi: 10.34123/icdsos.v2023i1.360.
- [21] S. Sazzed dan S. Jayarathna, "SSentiA: A Self-supervised Sentiment Analyzer for classification from unlabeled data," *Mach. Learn. with Appl.*, vol. 4, hal. 100026, Jun 2021, doi: 10.1016/j.mlwa.2021.100026.
- [22] Y. Y. Tan, C.-O. Chow, J. Kanesan, J. H. Chuah, dan Y. Lim, "Sentiment Analysis and Sarcasm Detection using Deep Multi-Task Learning," *Wirel. Pers. Commun.*, vol. 129, no. 3, hal. 2213–2237, Apr 2023, doi: 10.1007/s11277-023-10235-4.
- [23] Y. Wang, "Research on the TF-IDF algorithm combined with semantics for automatic extraction of keywords from network news texts," *J. Intell. Syst.*, vol. 33, no. 1, hal. 81–88, Jul 2024, doi: 10.1515/jisys-2023-0300.
- [24] N. Sulistianingsih dan I. N. Switrayana, "Enhancing Sentiment Analysis for the 2024 Indonesia Election Using SMOTE-



- Tomek Links and Binary Logistic Regression,” *Int. J. Educ. Manag. Eng.*, vol. 14, no. 3, hal. 22–32, Jun 2024, doi: 10.5815/ijeme.2024.03.03.
- [25] M. P. Pulungan, A. Purnomo, dan A. Kurniasih, “Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Kepribadian MBTI Menggunakan Naive Bayes Classifier,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 5, hal. 1033–1042, Okt 2024, doi: 10.25126/jtiik.2024117989.
- [26] A. Demircioğlu, “Applying oversampling before cross-validation will lead to high bias in radiomics,” *Sci. Rep.*, vol. 14, no. 1, hal. 11563, Mei 2024, doi: 10.1038/s41598-024-62585-z.
- [27] H. Barus, I. N. Fajri, dan Y. Pristyanto, “Sentiment Classification Analysis of Tokopedia Reviews Using TF-IDF, SMOTE, and Traditional Machine Learning Models,” *J. Appl. Informatics Comput.*, vol. 9, no. 5, hal. 2552–2561, Okt 2025, doi: 10.30871/jaic.v9i5.10524.
- [28] F. Abdusyukur, “Penerapan Algoritma Support Vector Machine (SVM) untuk Klasifikasi Pencemaran Nama Baik di Media Sosial Twitter,” *Komputa J. Ilm. Komput. dan Inform.*, vol. 12, no. 1, hal. 73–82, Mei 2023, doi: 10.34010/komputa.v12i1.9418.
- [29] N. Fachrurrozy, A. A. Amalia, dan S. Y. K. Dhian, “Analysis Sentiment Of Users Internet Service Providers In Indonesia On Social Media X Using Support Vector Machine,” *Data Sci. J. Comput. Appl. Informatics*, vol. 8, no. 2, hal. 88–95, Jul 2024, doi: 10.32734/jocai.v8.i2-16317.
- [30] K. Bhargava dan R. Katarya, “An improved lexicon using logistic regression for sentiment analysis,” in *2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN)*, IEEE, Okt 2017, hal. 332–337. doi: 10.1109/IC3TSN.2017.8284501.
- [31] D. Liang, X. Jin, Y. Yuan, dan R. Zou, “Performance Analysis of Machine Learning Methods,” *J. Phys. Conf. Ser.*, vol. 2428, no. 1, hal. 012039, Feb 2023, doi: 10.1088/1742-6596/2428/1/012039.
- [32] I. Imantoko, A. Hermawan, dan D. Avianto, “Comparative analysis of support vector machine and k-nearest neighbors with a pyramidal histogram of the gradient for sign language detection,” *Matrix J. Manaj. Teknol. dan Inform.*, vol. 11, no. 2, hal. 107–118, Jul 2021, doi: 10.31940/matrix.v11i2.2433.
- [33] O. Rainio, J. Teuvo, dan R. Klén, “Evaluation metrics and statistical tests for machine learning,” *Sci. Rep.*, vol. 14, no. 1, hal. 6086, Mar 2024, doi: 10.1038/s41598-024-56706-x.
- [34] N. R. Ramadhan dan E. R. Pramudya, “Prediksi Periode Fosil Trilobita Menggunakan XGBoost dengan Seleksi Fitur Geologi – Geospasial dan Hyperparameter Tuning,” vol. 7, no. 4, hal. 2181–2192, 2026, doi: 10.47065/bits.v7i4.8862.