

# Comparison of Random Forest and Decision Tree for Depression Detection Using Interaction Patterns

Felicia Talitha Fathin<sup>1</sup>, Warih Maharani<sup>2\*</sup>

<sup>1</sup> School of Computing, Data Science, Telkom University, Bandung, Indonesia

<sup>2</sup> School of Computing, Informatics, Telkom University, Bandung, Indonesia

Email: <sup>1</sup>feliciatalitha@student.telkomuniversity.ac.id, <sup>2\*</sup>wmaharani@telkomuniversity.ac.id

Correspondence Author Email: wmaharani@telkomuniversity.ac.id

Submitted: 09/01/2025; Accepted: 26/02/2025; Published: 01/03/2025

**Abstract**—This research focuses on evaluating the efficacy of Random Forest and Decision Tree, in detecting depression on tweets and interaction patterns on X social media. Depression as a global health problem often happens because of individuals' online behavior. This study uses data from X social media users in Indonesia who have filled out the DASS-42 questionnaire with an analysis approach that includes crawling data that includes tweets and interactions on X. The purpose of this research is to more accurately and comprehensively identify signs of depression by analyzing the interaction patterns of users on social media platforms through the integration of several many methods for feature extraction and preprocessing situations. The methods used include data preprocessing, feature combination using TF-IDF, Bag of Words, and Word2Vec and model evaluation utilizing metrics such as Precision, Recall, Accuracy, and F1-score. The findings of this research show that Random Forest performs better than Decision Tree, with a combination of TF-IDF, BoW, Word2Vec and TF-IDF, Word2Vec features obtained an accuracy of 0.60. Although Random Forest is superior, both models are difficult to identify the positive class of depression which can be seen from the relatively low F1-score and recall values. Other factors affecting model performance include lack of data relevance, low interaction rate, and limited feature extraction.

**Keywords:** Depression; X; Random Forest; Decision Tree; TF-IDF; Bag of Words; Word2Vec

## 1. INTRODUCTION

Depression, that also goes by the name major depressive disorder, is commonly found in people with anxiety disorders. The common symptoms of depression include anxiety attacks, sleep disturbances, loss of appetite, ongoing sadness (1 month), and loss of interest in favorite activities [1], [2], [3]. Depression affects 280 million people, or 3.8% of the population, with a larger rate among women. In addition, 700,000 individuals die by suicide each year, making suicide the top cause of mortality among 15-19 year olds [4] and positioned 13th among the biggest causes of mortality globally [5]. There were over 300 million individuals worldwide who experienced In 2019, the World Health Organization (WHO) reported. Indonesia accounted for 15.6 million of these individuals. However, it seems that only about 9% of patients get professional therapy based on those statistics [6].

Concurrently, social media has evolved into a digital platform for communication and sharing [7]. Depression can be experienced by anyone, but often individuals and those closest to them do not realize it. Therefore, social media becomes a means of expressing their expressions and feelings [8]. Problematic social media users (PSMU) have a greater potential to develop depression compared to social media users (SMU) [9], which means that the high intensity of media use can lead to high levels of depression experienced [10], [11]. This issue increases the possibility that social media can serve as a supportive resource that mental health experts or practitioners can use when making diagnoses and looking for indicators of mental illnesses such as depression.

Interactions on X contain significant potential for identifying users' emotions and psychological conditions, including depression. Research [12] indicates that retweets amplify the dissemination of a content, hence enhancing the probability of it being read and engaged with by numerous users. The same holds true for information pertaining to psychological illnesses, since more interactions on tweets expressing feelings or thoughts related to depression may offer insights about the user's mental health. Quote tweets enable users to append personal remarks or thoughts to a post, potentially conveying their sentiments. By examining the frequency and characteristics of interaction patterns, signs of depression can be delineated. This research establishes a foundation for employing X interactions as data for identifying depressed states, as increased engagement with content related to mental health themes correlates with a higher probability of discerning an individual's feelings or depressive conditions.

The primary issue in conducting this research is the selection of an appropriate algorithm for identifying complex patterns. This research compares two machine learning techniques, Decision Tree and Random Forest, that have distinct qualities and benefits in classification tasks. The Decision Tree algorithm is, especially with large datasets [13]. Conversely, Random Forest, as an ensemble of multiple Decision Trees, improves accuracy and mitigates overfitting. Evaluating the ability of these two methods to detect signs of depression is critical to determining the most appropriate algorithm for mental health detection applications in social media [14].

Some research has been carried out, especially on platform X, to identify depression in its users. Research conducted by Marcello H. [13] using the CART model-based Decision Tree algorithm method with TF-IDF feature extraction with a maximum tree depth of 4 achieved an accuracy of 81.25% and an F1-score of 85.71%. In this study, changes in the amount of training data, TF-IDF feature values and tree depth values for the model significantly affects performance. so that changes must be made carefully so that the model does not overfit. In research conducted by by

Hatoon A. and Mourad Y. [15], the accuracy of Random Forest was 84.7% and Decision Tree was 77.5%. This research emphasizes how crucial feature selection is for identifying depression in machine learning algorithms. Based on previous studies, this indicates that the algorithms are able to provide better results in identifying depression through analyzing interaction patterns on social media. In research [16], the prediction of anxious depression was carried out using 100 user data obtained from scraping API X. This study used 3 classifiers, Random Forest obtained an accuracy value of 81.04. This indicates that Random Forest is able to identify anxious depression based on data on social media. Research conducted by Fiza A. et al. [17] to identify depression using Support Vector Machine, Random Forest, along with Bag of Words (BoW) feature extraction resulted in an accuracy of 0.73 and 0.77 respectively. In this study Random Forest excels because it has a better F1-score, the result is 0.78 for the class that is depressed and 0.76 for the class that is not. This indicates that Random Forest is able to classify the depression class accurately. In another research [18], depression detection was carried out from the Reddit platform using several machine learning algorithms. The Random Forest model excels and produces accuracy and F1-score of 0.877. This value was obtained after augmentation using SMOTE and Word2Vec feature extraction. Meanwhile, Decision Tree obtained the maximum accuracy achieved is 0.781, with an F1-score of 0.772.

Unlike previous research that focuses primarily on the content of user posts, this study examines the relationship between depression indicators and interaction patterns, such as the number of comments, likes, quotes, and retweeted quotes. Furthermore, this research not only contrasts performance, but also integrates three feature extractions with two distinct preprocessing scenarios, providing a more comprehensive approach to depression detection. It is anticipated that this research will assist mental health professionals and social media platforms in identifying early signs of depression in users. This research has the potential to serve as a foundation for creating more effective automated monitoring systems by evaluating the effects of machine learning algorithms and understanding interaction patterns as indicators of mental health. This will help prevent depression and increase awareness of the importance of social interaction in mental health.

## 2. RESEARCH METHODOLOGY

### 2.1. Research Stages

This research applies the Random Forest and Decision Tree algorithms to identify depression through X posts and interactions. Figure 1 illustrates the research process. It starts with data collection from X social media users through crawling, followed by labeling using the DASS-42 questionnaire. The data undergoes preparation phases including case folding, cleaning, tokenization, stemming, and normalization. Following that, training and test sets of the data have been split and features are obtained applying the Word2Vec, Bag of Words (BoW), and TF-IDF techniques. The model is optimized with ADASYN when the data is unbalanced. The end result is a prediction of depression indicators from both models for performance comparison.

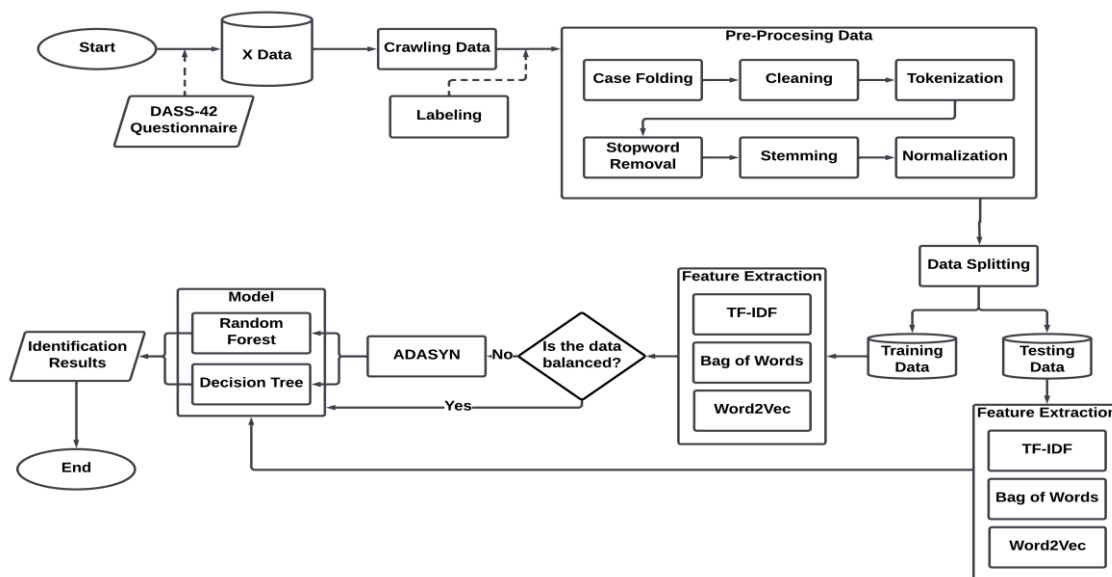


Figure 1. Research Methodology

### 2.2. Data Collection and Labeling

This research makes use of social media X user data. The DASS-42 questionnaire was utilized to collect user data, while the X API was employed to collect data. The DASS-42 questionnaire presents 42 questions to assess the level of emotional state measuring depression, anxiety and stress [19], [20]. The data is collected in CSV format and then labeling is performed. The labeling was determined by the computation of the DASS-42 questionnaire score, and

assigns a positive (1) or negative (0) classification of depression. Data was collected from March 15 to October 9, 2024, utilizing a total of 49 users from 100 respondents of the DASS-42 questionnaire, with 32 users positive for depression and 17 negative for depression. The 49 utilized accounts were chosen based on data verification outcomes, while 51 more accounts turned unusable owing to several constraints. The constraints are usernames that obtained no results after searching, along with private accounts that obstruct the data crawling operation. Therefore, 40 accounts that fulfilled the research criteria were utilized as selected data sources for the next step of analysis.

### 2.3. Preprocessing Dataset

Data collected from social media platforms contain many errors or irrelevant text, which can make semantic analysis difficult [1]. Preprocessing is important to improve consistency and accuracy and reduce noise in sentiment analysis, as tweets often consist of incomplete sentences [21]. In preprocessing dataset, there are several steps, including case folding, cleaning, tokenization, stopword removal, stemming, and normalization.

- a. Case folding: this step is done for text cleaning by changing all the letters in each sentence in lowercase form [21].
- b. Cleaning: text cleaning is done again by removing all unnecessary characters such as symbols, punctuation marks and URLs [17], [21].
- c. Tokenization: in this section, the sentence or text is deconstructed into separate words [22].
- d. Stopword removal: in the stopword removal step, common words or conjunctions that do not provide information for analysis are eliminated, such as “yang”, “dan”, “di”, “ke”, and many others [21], [23].
- e. Normalization: normalization is used to produce a consistent representation of words. It also reduces the variability of words that have the same meaning so that they can be processed. Such as handling slang words and excessive spelling [21].
- f. Stemming: stemming is a process to remove the affixes that exist in words into the form of basic words to facilitate analysis [21].

### 2.4. Splitting Dataset

The dataset splitting technique involves separating the dataset into two categories, called training data and test data. 80% of the dataset is designated for model training, while 20% is employed for performance evaluation. This division allows the model to learn from the predominant dataset, while the residual is employed to evaluate the model's capability to forecast previously unencountered dataset, hence facilitating performance evaluation. After the separation, the training data set consists of 18,965 rows, while the test data set consists of 4,674 rows. The outcomes of the data splitting are presented in Table 1.

**Table 1.** Data Splitting Outcomes

| Data Training | Data Testing |
|---------------|--------------|
| 18,965        | 4,674        |

### 2.5. Feature Extraction

Machine learning algorithms cannot process raw text directly, therefore feature extraction is required. This algorithm converts the text into a matrix or vector and removes irrelevant data features to reduce the dimensionality of the data space [13].

- a. TF-IDF  
Term Frequency (TF) measures how often a term appears in a document, showing its importance to the subject matter. Inverse Document Frequency (IDF) measures the importance of each word in a document. The more a term appears across many texts, the lower its value. TF-IDF is a way to measure how important a word is in a text compared to other documents. [15].
- b. Bag of Words (BoW)  
The Bag of Words (BoW) model works by converting a document into a collection of words, regardless of syntax or order. Each document is represented as a fixed-length numerical vector defined by the quantity of distinct words in the corpus, with each value indicating the frequency of that word. [24].
- c. Word2Vec  
Word2Vec is an embedding technique utilizing a shallow neural network to map words into a low-dimensional vector space according to their context. This process clusters a vector space of words with the same meaning to overcome the problems of high dimensionality and context loss present in Bag of Words (BoW) and n-gram models [24].

### 2.6. Adaptive Synthetic Sampling Approach (ADASYN)

ADASYN (Adaptive Synthetic Sampling Approach) algorithm is used to perform oversampling because the dataset used is not balanced. There were more samples of depressive negative tweets (0) than depressive positive tweets (1). ADASYN helps generate synthetic samples for fewer classes so that the model can learn more balanced [25]. The training data consists of 11,714 instances for Class 0 and 6,981 instances for Class 1. After the implementation of ADASYN, the distribution of both classes is almost equal, with 11,714 instances for Class 0 and 11,409 instances for

Class 1. Table 2 shows a comparison of the distribution of the two classes both before and after the implementation of ADASYN.

**Table 2.** Class Distribution Before and After ADASYN

| Class        | Before ADASYN | After ADASYN |
|--------------|---------------|--------------|
| Negative (0) | 11,714        | 11,714       |
| Positive (1) | 6,981         | 11,409       |

## 2.7. Random Forest

Random Forest is a non-linear method that integrates many decision trees, with each tree being built through two randomized stages. First, using bootstrap samples to train the tree and second, randomly selecting a selection of variables at each internal node for partitioning. This methodology enhances identifiability, so Random Forest produces a set of classifiers with low bias and high variance. This variance is then addressed by an ensemble procedure by calculating all the averages of the identification results of the formed trees [26]. The algorithm of the Random Forest is shown in Figure 2 [14].

### Algorithm 1 Random Forest Algorithm

```

1: Given a training data set
2: Select number of trees to build ( $n\_trees$ )
3: for  $i = 1$  to  $n\_trees$  do
4:   Generate a bootstrap sample of the original data
5:   Grow a classification tree to the bootstrapped data
6:   for each split do
7:     Select  $m\_try$  variables at random from all  $p$  variables
8:     Pick the best variable/split-point among the  $m\_try$ 
9:     Split the node into two child nodes
10:  end for
11:  Use typical tree model stopping criteria to determine when a tree is complete
    (but do not prune)
12: end for
13: Output ensemble of trees
    
```

**Figure 2.** Random Forest Algorithm

Figure 2 illustrates the iterative bootstrapping of the training data, the construction of a classification tree, and the aggregation of the predictions by majority voting. The difference lies in the splitting mechanism, where Random Forest evaluates only a random selection of predictor variables to determine the optimal split, as opposed to Bagging, which considers all variables. This method produces diverse trees, reduces the correlation between them, and increases the effectiveness of the ensemble [14].

## 2.8. Decision Tree

The Decision Tree is a model of classification that enhances the decision-making process and improves result quality through its inherent flexibility of choosing features that distinguish criteria in each of the same nodes, so it can increase the quality of decision results [13]. The Decision Tree model's calculation is presented in equations (1), (2), and (3) [27].

a. Entropy

$$Entropy = - \sum_{v \in values(A)} \left( \frac{|S_v|}{|S|} \times \log_2 \left( \frac{|S_v|}{|S|} \right) \right) \quad (1)$$

In formula (1),  $S$  represents the user interaction dataset, and  $A$  represents the attribute selected for the dataset partition. The value  $v$  represents each distinct value  $v \in values(A)$  that attribute  $A$  may assume. The subset  $S_v$  represents the elements of  $S$  for which attribute  $A$  assumes a specific value  $v$ . The quantity of elements in the subset is denoted by  $|S_v|$ , whereas the overall quantity of elements in dataset  $S$  is represented as  $|S|$ .

b. Gain (S,A)

$$Gain(S, A) = Entropy(S) - \sum_{v \in values(A)} \left( \frac{|S_v|}{|S|} \times Entropy(S_v) \right) \quad (2)$$

$Entropy(S)$  is employed to evaluate the initial degree of uncertainty in the dataset  $S$  as indicated in formula (2). The entropy value is subsequently compared with the entropy of the subsets ( $Entropy(S_v)$ ) created after partitioning the dataset according to the unique value  $v$  of attribute  $A$ . The ratio of each subset to the original dataset is represented by  $\frac{|S_v|}{|S|}$

c. Gain Ratio (S, A)

$$Gain\ Ratio(S, A) = \frac{Gain(S, A)}{SplitInformation(S, A)} \quad (3)$$

Gain Ratio is the ratio of Gain(S, A) and Split Information (SplitInformation(S, A)). Split Information in formula (3) refers to the aggregate information derived from partitioning the dataset *S* according to attribute *A*.

**2.9. Model Evaluation**

The dataset's results is assessed by a confusion matrix that gives scores for Accuracy, Precision, Recall, and F1-score [1]. Classification reports assess model results by computing True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) [3]. Classification reports assess model results by computing True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) [3]. TP denotes the ratio of correctly anticipated positive instances to the actual data. TN denotes the ratio of data expected to be negative but correct. FP indicates cases classified as positive that are, in fact, negative, while FN signifies instances classified as negative that are truly positive. These parameters are employed to evaluate the model's performance regarding accuracy, precision, recall, and F1-score [14], [28].

a. Accuracy is a calculation to measure the proportion of total correct identifications [13], [29].

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{4}$$

b. Precision is a calculation to measure the proportion of truly positive identifications [13], [29].

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

c. Recall is a calculation to measure the proportion of all positive cases that were predicted correctly [13], [29].

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

d. The F1-score is a metric that is used to determine the optimal balance between Precision and Recall [30]. It is calculated by averaging precision and recall [13].

$$F1 - Score = 2 * \frac{(Precision*Recall)}{Precision+Recall} \tag{7}$$

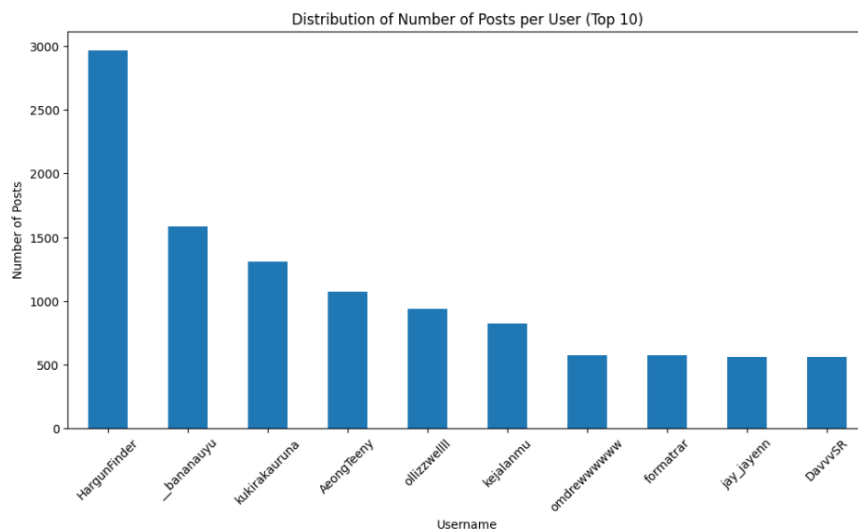
**3. RESULT AND DISCUSSION**

**3.1. Exploratory Data Analysis (EDA)**

Before moving on to additional processing and modeling steps, Exploratory Data Analysis (EDA) is carried out to comprehend the pattern and distribution of data generally. In this research, the EDA steps carried out include the distribution of the number of uploads, the distribution of interactions, the distribution of user labels, word frequency, and data visualization.

**3.1.1. Distribution of Number of Posts**

The distribution of the number of posts is used to get an overview of user behavior and identify users who are active on X. In this step, 10 users with the highest number of uploads are displayed. The user known as "Hargumfinder" has the most posts in Figure 3, with over 3,000, considerably more than any other user. In contrast, among this top 10 group, users like "Jay\_Jayem" and "DavvSR" have comparatively fewer postings.



**Figure 3.** Distribution of Number Post per User



**Table 5.** TF-IDF and BoW Outcomes

| Word    | TF-IDF   | BoW |
|---------|----------|-----|
| alami   | 0,218218 | 1   |
| banget  | 0,218218 | 1   |
| bingung | 0,218218 | 1   |
| buat    | 0,218218 | 1   |
| tiba    | 0,436436 | 2   |

### 3.3.2. Word2Vec

Word2Vec converts words into numerical vectors, where words that have the same meaning have adjacent vectors. This model utilizes context to produce a more meaningful word representation. The outcomes of the Word2Vec implementation can be seen in Table 6.

**Table 6.** Word2Vec Outcomes

| Dimension | 0         | 1         | 2         | ... | 149       |
|-----------|-----------|-----------|-----------|-----|-----------|
| Value     | -0.000677 | -0.001203 | -0.001203 | ... | -0.000652 |

### 3.4. Model Evaluation

In this research, the hyperparameter tuning was performed using both of models with 10 iterations, where the parameters used in this research include `n_estimators`, `max_depth`, `min_samples_split`, `min_samples_leaf` and the values shown in Table 7. The tuning results show that the optimal configuration for the Random Forest model is `n_estimators = 500`, `max_depth = 30`, `min_samples_split = 10`, and `min_samples_leaf = 2`. Meanwhile, for the Decision Tree model, the best configuration is obtained with `max_depth = 30` and `min_samples_split = 5`.

**Table 7.** Model Parameters

| Paramater                      | Tested Values   |   |
|--------------------------------|---|---|
|                                | Random Forest   | Decision Tree                                 |
| <code>n_estimators</code>      | 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 | None  |
| <code>max_depth</code>         | None, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100         | None, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 |
| <code>min_samples_split</code> | 2, 5, 10, 15, 20, 25, 30                              | 2, 5, 10, 15, 20, 25, 30                      |
| <code>min_samples_leaf</code>  | 1, 2, 4, 8, 16, 32                                    | 1, 2, 4, 8, 16, 32                            |

This research uses 2 preprocessing scenarios with 3 feature combinations as shown in Table 4, Table 5 and Table 6 to measure the performance of Random Forest and Decision Tree, which can be seen in Table 8 and 9. In Scenario 1, both Random Forest and Decision Tree models using TF-IDF and BoW features showed a more balanced performance in predicting Class 0 and Class 1 by achieving 0,55 Accuracy. Both can properly regulate the prediction distribution, which is crucial for recognizing how user posts and interactions signify indicators of depression. Compared to other features such as TF-IDF, BoW and Word2Vec, these two models show superiority in class 0, despite a slightly lower recall rate for class 1. The other features are suboptimal due to a pronounced imbalance, resulting in a significant decrease in performance for class 0, despite an increase in recall for class 1. Table 8 presents the model evaluation outcomes for Scenario 1.

**Table 8.** Model Evaluation Outcomes for Scenario 1

| Model         | Feature Extraction    | Precision |      | Recall |      | F1-Score |      | Accuracy |
|---------------|-----------------------|-----------|------|--------|------|----------|------|----------|
|               |                       | 0         | 1    | 0      | 1    | 0        | 1    |          |
| Random Forest | TF-IDF, BoW, Word2Vec | 0.33      | 0.62 | 0.15   | 0.82 | 0.21     | 0.70 | 0.57     |
|               | TF-IDF, BoW           | 0.40      | 0.64 | 0.39   | 0.65 | 0.39     | 0.64 | 0.55     |
|               | TF-IDF, Word2Vec      | 0.37      | 0.62 | 0.18   | 0.82 | 0.24     | 0.71 | 0.58     |
| Decision Tree | TF-IDF, BoW, Word2Vec | 0.36      | 0.62 | 0.12   | 0.87 | 0.18     | 0.73 | 0.59     |
|               | TF-IDF, BoW           | 0.40      | 0.64 | 0.39   | 0.65 | 0.39     | 0.64 | 0.55     |
|               | TF-IDF, Word2Vec      | 0.38      | 0.64 | 0.82   | 0.20 | 0.52     | 0.39 | 0.43     |

In Scenario 2, the combinations of TF-IDF, BoW, Word2Vec and TF-IDF, Word2Vec with the Random Forest model exhibited strong performance in identifying Class 1, with the highest Recall demonstrating the model's ability in detecting positive cases of depression. Both feature combinations achieved an overall accuracy of 0.60, showing a balanced outcome in detecting Class 1 while preserving general model performance. Nevertheless, despite the increased overall accuracy, performance on Class 0 significantly declined, resulting in a prediction disparity. These variables are highly effective in optimizing identification for Class 1, however the accuracy for Class 0 is significantly lower, suggesting that the model shows greater sensitivity in identifying positive depressions than negative ones.

Alternative methods, including TF-IDF and BoW within the Random Forest model, as well as all feature combinations in the Decision Tree, failed to achieve the recall and F1-score for Class 1, hence giving lower results in this context. The results of the evaluation of the model for Scenario 2 are presented in Table 9.

**Table 9.** Model Evaluation Result for Scenario 2

| Model         | Feature Extraction    | Precision |      | Recall |      | F1-Score |      | Accuracy |
|---------------|-----------------------|-----------|------|--------|------|----------|------|----------|
|               |                       | 0         | 1    | 0      | 1    | 0        | 1    |          |
| Random Forest | TF-IDF, BoW, Word2Vec | 0.31      | 0.62 | 0.06   | 0.92 | 0.10     | 0.74 | 0.60     |
|               | TF-IDF, BoW           | 0.39      | 0.63 | 0.37   | 0.65 | 0.38     | 0.64 | 0.55     |
|               | TF-IDF, Word2Vec      | 0.41      | 0.63 | 0.13   | 0.89 | 0.20     | 0.74 | 0.60     |
| Decision Tree | TF-IDF, BoW, Word2Vec | 0.37      | 0.62 | 0.34   | 0.65 | 0.36     | 0.64 | 0.54     |
|               | TF-IDF, BoW           | 0.38      | 0.63 | 0.36   | 0.65 | 0.37     | 0.64 | 0.54     |
|               | TF-IDF, Word2Vec      | 0.37      | 0.61 | 0.76   | 0.23 | 0.50     | 0.33 | 0.43     |

It can be seen that Scenario 2 is better in Accuracy, but Scenario 1 tends to be more stable in handling the class distribution than Scenario 2. This is consistent with research [31], [32], which shows that proper preprocessing greatly affects model performance. This research indicated that the implementation of stopword removal and stemming did not significantly affect the accuracy of the model, although a slight improvement was observed. The results of the experiment show that the Random Forest model outperforms the Decision Tree., aligning with the conclusions of other previous research, Random Forest is generally more effective in handling data with high complexity and more feature variables [15], [19], [33].

Model evaluation results show that Random Forest and Decision Tree evaluation results tend to show low accuracy and F1-score, especially for Class 0. Based on the results of the evaluation, there are several factors that affect model performance, such as data quality, lack of interaction, class imbalance, and the effect of model parameters. The first factor is data quality and tweet relevance, where most tweets do not represent deep emotional expressions. Many tweets in the dataset were about giveaway, sharing information irrelevant to the user's mental state, or reposting other users' posts. This type of data makes it difficult for the model to recognize specific patterns related to depression, as the relevant information is very limited. Table 10 shows some irrelevant post based on the dataset.

**Table 10.** Irrelevant Post Content from Dataset

| Username  | Label | Post   |
|-----------|-------|--|
| dkss012   | 1     | @samsungID Jawaban Fitur #GalaxyAI di #GalaxyS24FE yang ada di event #FunExperienceElRumi adalah Photo Assist.   |
| strawiy   | 1     | @gqgrowth @2 done kaak makasih banyak ya kak atas give awaynya semoga apapun yang kakak jalani lancar terus dimudahkan segalagalanya. semoga rezeki kakak juga dilancarkan! semoga kakak |
| kejalanmu | 1     | @FIFAWorldCup POWER by @weareoneEXO #WorldCup #WorldCupEXOPower #EXO   |

Another factor is class imbalance. In Table 2, the dataset used has more Class 1 with 11,714 rows, while Class 0 has only 6,981 rows. This imbalance makes the classification results tend to show bias towards the majority class, resulting in low recall for Class 0. The amount of data for the positive depression class was effectively increased to 11,409 after the use of ADASYN, and is now close to the amount of data for the negative class. However, the evaluation findings indicate that this problem has not been fully corrected. Features such as TF-IDF, BoW, and Word2Vec are limited in their ability to capture the nuances of depression-related online behavior. TF-IDF gives greater weight to words that appear infrequently in the text, but emotionally relevant words do not always receive high weight. BoW calculates the frequency of word occurrence directly without considering meaning or context, so important information related to order is often lost. Meanwhile, Word2Vec works to capture semantic relationships between words, but is less capable when implemented on text with a small number of words, despite normalization. In addition, minimal user interaction patterns make it difficult for the model to distinguish between Class 0 and Class 1, as both classes tend to have few or inconspicuous interactions such as the interaction distribution shown in Table 3.

The hyperparameter selection in Table 7 shows that Random Forest and Decision Tree achieve optimal performance with certain configurations. In Random Forest, using  $n\_estimators=200$ ,  $max\_depth=30$ , and  $min\_samples\_split=5$  provides a balance between complexity and generalization. The maximum depth helps the model capture complex patterns without overfitting, while the minimum limit on node splits prevents the formation of too small nodes. Meanwhile, Decision Tree uses  $max\_depth=30$  and  $min\_samples\_split=5$ , which keeps the model simple while preventing the formation of less informative nodes. This tuning hyperparameter had a significant impact on model performance. Random Forest achieved an accuracy of 0.57 with a high recall for the positive class of 0.82, indicating its ability to detect more cases of depression, although the precision for the negative class was still low. In contrast, Decision Tree had a slightly higher accuracy of 0.59 but showed a low recall for the negative class of 0.12, reflecting its sensitivity to overfitting.

## 4. CONCLUSION

The results indicate that Random Forest and Decision Tree have specific advantages in identifying depression through post patterns and user interactions. In Scenario 1, basic features such as TF-IDF and BoW provide more balanced predictions for both classes, showing nearly identical performance between the two models. This combination is better suited for applications that require a proportional distribution of predictions, but the recall for Class 1 is slightly lower compared to other features. On the other hand, in Scenario 2, the TF-IDF, BoW, and Word2Vec features with Random Forest showed higher accuracy in identifying Class 1, indicating an improved ability to detect positive cases of depression. However, their performance in class 0 dropped significantly, resulting in a prediction disparity. The decision tree showed a similar pattern, although it was more susceptible to overfitting on certain data. This experiment shows that the selection of features and parameters significantly influences model outcomes. Advanced features such as Word2Vec provide a more nuanced identification of interaction patterns, but often lead to imbalance. Meanwhile, simple features yield better predictive balance. This research highlights the need to focus on the quality of user interaction data to enhance the model's prediction accuracy. Future research may explore alternative methods, such as deep learning, to more effectively mitigate class imbalance.

## REFERENCES

- [1] S. Gupta, L. Goel, A. Singh, A. Prasad, and M. A. Ullah, “Retracted: Psychological Analysis for Depression Detection from Social Networking Sites,” *Comput. Intell. Neurosci.*, vol. 2023, no. 1, Jan. 2023, doi: 10.1155/2023/9796187.
- [2] T. Dhaker, A. Kumar, and D. A. G., “Detecting Depression on Social Media: A Comprehensive Review of Data Analysis, Deep Learning, NLP, and Machine Learning Approaches,” *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, pp. 103–117, Sep. 2023, doi: 10.32628/CSEIT2390510.
- [3] J. G. Choi, K. O. Inhwon, and S. Han, “Depression Level Classification Using Machine Learning Classifiers Based on Actigraphy Data,” *IEEE Access*, vol. 9, pp. 116622–116646, 2021, doi: 10.1109/ACCESS.2021.3105393.
- [4] World Health Organization, “Depressive disorder (depression).” Accessed: Dec. 19, 2024. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/depression>
- [5] P. Jain, K. R. Srinivas, and A. Vichare, “Depression and Suicide Analysis Using Machine Learning and NLP,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jan. 2022. doi: 10.1088/1742-6596/2161/1/012034.
- [6] Nabila Hisanah Yusri, “Depresi, Kesehatan Mental yang Tak Boleh Disepelekan.” Accessed: Mar. 23, 2024. [Online]. Available: [https://www.its.ac.id/news/2023/05/22/\\_trashed-5/](https://www.its.ac.id/news/2023/05/22/_trashed-5/)
- [7] M. Liu, K. E. Kamper-Demarco, J. Zhang, J. Xiao, D. Dong, and P. Xue, “Time Spent on Social Media and Risk of Depression in Adolescents: A Dose–Response Meta-Analysis,” May 01, 2022, *MDPI*. doi: 10.3390/ijerph19095164.
- [8] J. Angskun, S. Tipprasert, and T. Angskun, “Big data analytics on social networks for real-time depression detection,” *J. Big Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40537-022-00622-2.
- [9] O. Ahmed, E. I. Walsh, A. Dawel, K. Alateeq, D. A. Espinoza Oyarce, and N. Cherbuin, “Social media use, mental health and sleep: A systematic review with meta-analyses,” Dec. 15, 2024, *Elsevier B.V.* doi: 10.1016/j.jad.2024.08.193.
- [10] A. Abidah and A. Aziz, “Hubungan Antara Intensitas Penggunaan Media Sosial dan Tingkat Depresi pada Mahasiswa,” *Acta Psychol.*, vol. 2, no. 2, pp. 92–107, 2020, [Online]. Available: <http://journal.uny.ac.id/index.php/acta-psychologia>
- [11] L. Azem *et al.*, “Social Media Use and Depression in Adolescents: A Scoping Review,” Jun. 01, 2023, *MDPI*. doi: 10.3390/bs13060475.
- [12] Y. Yan, F. Toriumi, and T. Sugawara, “Understanding how retweets influence the behaviors of social networking service users via agent-based simulation,” *Comput. Soc. Networks*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40649-021-00099-8.
- [13] M. R. Hidayatullah and Warih Maharani, “Depression Detection on Twitter Social Media Using Decision Tree,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 4, pp. 677–683, Aug. 2022, doi: 10.29207/resti.v6i4.4275.
- [14] A. Renaldi and W. Maharani, “Depression Detection of User in Media Social Twitter Using Random Forest,” *J. Inf. Syst. Res.*, vol. 3, no. 4, pp. 410–416, Jul. 2022, doi: 10.47065/josh.v3i4.1837.
- [15] H. Alsagri and M. Ykhlef, “Quantifying Feature Importance for Detecting Depression using Random Forest,” 2020. [Online]. Available: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- [16] A. Kumar, A. Sharma, and A. Arora, “Anxious Depression Prediction in Real-time Social Data,” 2019. doi: 10.48550/arXiv.1903.10222.
- [17] F. Azam, M. Agro, M. Sami, M. H. Abro, and A. Dewani, “Identifying Depression among Twitter Users using Sentiment Analysis,” in *2021 International Conference on Artificial Intelligence, ICAI 2021*, Institute of Electrical and Electronics Engineers Inc., Apr. 2021, pp. 44–49. doi: 10.1109/ICAI52203.2021.9445271.
- [18] Sampath Kayalvizhi and Thenmozhi Durairaj, “Data set creation and empirical analysis for detecting signs of depression from social media postings,” in *International Conference on Computational Intelligence in Data Science*, Springer International Publishing, Feb. 2022, pp. 136–151. doi: 10.1007/978-3-031-16364-7.
- [19] S. A. Sulak and N. Koklu, “Analysis of Depression, Anxiety, Stress Scale (DASS-42) With Methods of Data Mining,” *Eur. J. Educ.*, vol. 59, no. 4, Dec. 2024, doi: 10.1111/ejed.12778.
- [20] S. R. Marsidi, “Identification Of Stress, Anxiety, And Depression Levels Of Students In Preparation For The Exit Exam Competency Test,” *J. Vocat. Heal. Stud.*, vol. 5, no. 2, p. 87, Nov. 2021, doi: 10.20473/jvhs.V5.I2.2021.87-93.
- [21] F. Resyanto, Y. Sibaroni, and A. Romadhony, “Choosing The Most Optimum Text Preprocessing Method for Sentiment Analysis ( Case:iPhone Tweets),” in *2019 Fourth International Conference on Informatics and Computing (ICIC)*, Semarang, Indonesia, 2019, p. 1. doi: 10.1109/ICIC47613.2019.8985943.
- [22] M. T. Hossain, M. A. R. Talukder, and N. Jahan, “Depression prognosis using natural language processing and machine learning from social media status,” *Int. J. Electr. Comput. Eng.*, vol. 12, no. 3, pp. 2847–2855, Jun. 2022, doi: 10.11591/ijece.v12i3.pp2847-2855.



- [23] S. Fransiska, Rianto Rianto, and A. Irham Gufroni, “Sentiment Analysis Provider by.U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method,” *Sci. J. Informatics*, vol. 7, no. 2, pp. 2407–7658, 2020, doi: <https://doi.org/10.15294/sji.v7i2.25596>.
- [24] H. D. Abubakar and M. Umar, “Sentiment Classification: Review of Text Vectorization Methods: Bag of Words, Tf-Idf, Word2vec and Doc2vec,” *SLU J. Sci. Technol.*, vol. 4, no. 1&2, pp. 27–33, Aug. 2022, doi: 10.56471/slujst.v4i.266.
- [25] Chao Lu, Shaofu Lin, Xiliang Liu, and Hui Shi, “Telecom Fraud Identification Based on ADASYN and Random Forest,” in *2020 5th International Conference on Computer and Communication Systems (ICCCS)*, Shanghai, China: Institute of Electrical and Electronics Engineers (IEEE), 2020, pp. 447–452. doi: 10.1109/ICCCS49078.2020.9118521.
- [26] M. Aria, C. Cuccurullo, and A. Gnasso, “A comparison among interpretative proposals for Random Forests,” *Mach. Learn. with Appl.*, vol. 6, p. 100094, Dec. 2021, doi: 10.1016/j.mlwa.2021.100094.
- [27] Abdussalam Sulaiman Olainiyi, Saheed Yakub Kayode, Hambali Moshood Abiola, Salau-Ibrahim Taofeekat Tosin, and Akinbowale Nathaniel Babatunde, “Student’s Performance Analysis Using Decision Tree Algorithms,” *Ann. Comput. Sci. Inf. Syst.*, 2017.
- [28] P. E. Sumolang and W. Maharani, “Depression Detection on Twitter Using Bidirectional Long Short Term Memory,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 2, pp. 369–376, Sep. 2022, doi: 10.47065/bits.v4i2.1850.
- [29] D. Kurniasari, R. Nurul Hidayah, and R. Khoirun Nisa, “Classification Models For Academic Performance: A Comparative Study Of Naïve Bayes And Random Forest Algorithms In Analyzing University Of Lampung Student Grades,” *J. Tek. Inform.*, vol. 5, no. 5, pp. 1267–1276, 2024, doi: 10.52436/1.jutif.2024.5.5.2066.
- [30] D. Valero-Carreras, J. Alcaraz, and M. Landete, “Comparing two SVM models through different metrics based on the confusion matrix,” *Comput. Oper. Res.*, vol. 152, Apr. 2023, doi: 10.1016/j.cor.2022.106131.
- [31] A. W. Pradana and M. Hayaty, “The Effect of Stemming and Removal of Stopwords on the Accuracy of Sentiment Analysis on Indonesian-language Texts,” *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, pp. 375–380, Oct. 2019, doi: 10.22219/kinetik.v4i4.912.
- [32] M. A. Palomino and F. Aider, “Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis,” *Appl. Sci.*, vol. 12, no. 17, Sep. 2022, doi: 10.3390/app12178765.
- [33] K. Rahayu, V. Fitria, D. Septhya, R. Rahmaddeni, and L. Efrizoni, “Klasifikasi Teks untuk Mendeteksi Depresi dan Kecemasan pada Pengguna Twitter Berbasis Machine Learning,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 108–114, Sep. 2023, doi: 10.57152/malcom.v3i2.780.