Feature Expansion Using Word2vec for Hate Speech Detection on Indonesian Twitter with Classification Using SVM and Random Forest

Mila Putri Kartika Dewi, Erwin Budi Setiawan
School of Computing, Informatics, Telkom University, Bandung, Indonesia
Email: milakartika52@gmail.com, erwinbudisettiawan@telkomuniversity.ac.id
Correspondence Author Email: erwinbudisettiawan@telkomuniversity.ac.id

Abstract—Hate speech is one of the most common cases on Twitter. It is limited to 280 characters in uploading tweets, resulting in many word variations and possible vocabulary mismatches. Therefore, this study aims to overcome these problems and build a hate speech detection system on Indonesian Twitter. This study uses 20,571 tweet data and implements the Feature Expansion method using Word2vec to overcome vocabulary mismatches. Other methods applied are Bag of Word (BOW) and Term Frequency-Inverse Document Frequency (TF-IDF) to represent feature values in tweets. This study examines two methods in the classification process, namely Support Vector Machine (SVM) and Random Forest (RF). The final result shows that the Feature Expansion method with TF-IDF weighting in the Random Forest classification gives the best accuracy result, which is 88.37%. The Feature Expansion method with TF-IDF weighting can increase the accuracy value from several tests in detecting hate speech and overcoming vocabulary mismatches.

Keyword: Hate Speech; Feature Expansion; Word2vec; Support Vector Machine (SVM); Random Forest; Indonesian Twitter

1. INTRODUCTION

Social media is a source of information, entertainment, and communication to exchange messages online with someone without being separated by distance and time [1]. One of the social media that Indonesian people most often use is Twitter [2]. Twitter is used to send messages in the form of text and is called a tweet [3]. Based on information from the General Resources for Post and Information Technology at the Ministry of Communication and Information, Twitter users in Indonesia are 19.5 million and the fifth-ranked country with active Twitter users [4]. In uploading or commenting on tweets from other users, the Twitter application provides freedom of expression [2], so it can cause the problems such as hate speech [5]. Not only that, other problems like Twitter limits only 280 characters to upload a tweet. This restriction sometimes contains word variations, allowing for vocabulary mismatches in tweets [6]. Hate speech can occur when individuals or groups communicate with each other. The actions include provocation, incitement, or insult to various aspects, whether based on race, gender, ethnicity, disability, nationality, religion, sexual orientation, or other characteristics [7]. According to the Indonesian Ministry of Communication and Information, hate speech is the highest case on social media [5]. Classification of hate speech is a solution to deal with this problem. Many researchers have conducted research related to this topic, but the final results obtained are sometimes unsatisfactory such as low accuracy results. Low accuracy can be caused by limited training data [8]. The hate speech detection system has been developed, and is expected to reduce the problem of hate speech on twitter, so as to create healthy habits.

Studies on hate speech detection have been widely by other researchers. The study [9] used 14,509 tweet data and compared classification methods such as RF, AdaBoost, and NN. The results showed that the Random Forest method produced the best accuracy value of 72.2%. According to the authors, Random Forest results also have better recall and F1-Measure than AdaBoost and Neural Network. Other researchers compared classifications such as SVM, NB, BLR, and RFDT. Performed three test scenarios, one of which was to compare feature extraction performance using four classification algorithms. The final result shows that the Random Forest Decision Tree (RFDT) classification using the Word N-Gram feature can provide the best F-Measure value up to 93.5% [10]. The weakness of this research is that the data used is only 520 tweets. The data has gone through pre-processing, so there is a significant reduction from the initial data. Another experiment shown in research [11] used 1,000 data sets of hate tweets and applied labeling techniques based on hate speech against ethnicity, race, religion, intergroup, and neutrality. This study tested the SVM kernel, and the best accuracy results were 93% when using the RBF kernel with the TF-IDF method.

Applying the Word Embedding method is one way to represent words into vector numbers. Researchers [12] compared weighting schemes, such as Binary, TF, TF-IDF, and Word2vec. This study obtained the best accuracy of 90% when applying the SVM classification with the word2vec method. According to the author, the words similarity method is appropriate for the weighting of the Word Embedding approach. The study [13] applied the TF-IDF, N-gram Word2vec, and Doc2vec methods. Tweet data used is 14,509 and apply classification methods such as NB, SVM, LR, KNN, DT, RF, AdaBoost, and MLP. The best accuracy result in this study was 79%, obtained when implementing SVM with a combination of TF-IDF and Bigram features. From these results, the TF-IDF method is superior to Word2vec and Doc2vec. According to the author, this method cannot handle OOV words in the Twitter domain, and Word2vec also requires many training data. Another study also applies the same method [6], using a dataset of 19,401 tweets from 97 Twitter accounts and collecting on Indonews and Google

Mila Putri Kartika Dewi, Copyright © 2022, MIB, Page 979
Submitted: 19/02/2022; Accepted: 22/03/2022; Published: 25/04/2022
News data consisting of 1,111 articles. This study applies the TF-IDF method to represent values in tweets and classify tweets using the SVM, NB, and LR methods. In overcoming the vocabulary mismatch, this research applies the feature expansion method using word2vec. The news data will be as a corpus using the Word2vec method to determine semantically similar words. The best accuracy result in this study was 58.89%, obtained when applying the Logistics Regression classification method. From these results, feature expansion can increase the accuracy value.

This study aims to build a hate speech detection system and overcome vocabulary mismatches. Therefore, the researchers experimented with applying the Feature Expansion method using Word2vec as the researcher did [6]. Other methods applied are TF-IDF and Bag of Word (BOW). This method has been widely applied in previous studies and is the most common method for representing feature values. In classifying a tweet, this study uses the SVM and Random Forest methods. The SVM classification has the best accuracy based on previous research, while the Random Forest classification can process extensive data.

2. RESEARCH METHODOLOGY

2.1 System Design

This section will explain the system plan of the hate speech detection system. It consists of several steps, which include data crawling, data labeling, data preprocessing, Feature Extraction (TF-IDF, BOW), Feature Expansion (Word2vec), data classification (SVM, RF) and evaluate the system using the confusion matrix method. Figure 1 will explain the flow of the system.

![Figure 1. Hate Speech Detection System](image)

2.2 Data Crawling

In collecting data, we apply a crawling method with the Application Programming Interface (API) and use Python for building the system [5]. A total of 20,571 Indonesian-language Tweet data were collected based on specific topics or keywords, and the tweet data contains some information such as keywords, tweets and usernames.

2.3. Text Preprocessing

Text processing is applied to select and filter data that is still noisy or dirty to be cleaner and more structured [11]. It is essential to apply it before being processed to the next step, because get better/structured data quality and can improve classification accuracy [1]. This process consists of several steps, Data Cleaning is the first step taken to delete mentions, hashtags, characters/symbols such as punctuation marks, URLs, numbers (0-9), and emoticon characters [14]. Case Folding is a step to change a text that contains capital letters to lowercase [6]. Normalization is a step in checking each tweet containing abbreviated words, informal words, and slang words and then converted to actual words. Stop words removal is a step to eliminate words that are considered unimportant in the tweet text.
2.4. Bag of Word (BOW)

BOW is a set of vector numbers based on the frequency of words in the document, and its characteristic is to ignore word order and grammar but still maintain diversity [15]. This method is simpler than TF-IDF, because it does not require a special formula for processing.

2.5. Term Frequency – Inverse Document Frequency (TF-IDF)

Another method is TF-IDF, a feature weighting technique with numerical statistical methods that show relevant words/terms for several documents [16]. The concept of TF-IDF method is to calculate TF and IDF values. TF is the value of the occurrence of the word in a document based on its frequency. At the same time, IDF calculates the distribution of a word in a document [11]. The TF-IDF formula will be explained in equation (1).

\[ W_{t,d} = tf_{t,d} \times \log \left( \frac{N}{df_t} \right) \]  

Which \( W_{t,d} \) is the weight of term in document \( d \), \( tf_{t,d} \) is value of TF and \( df_t \) is the number of document containing word \( t \) [17].

2.6. N-Gram

N-gram is a text mining method used in this study to process text data by cutting it into a word. The types of n-grams based on their units are divided into n-grams of characters and n-grams of words [18]. In the n-gram character an input text will be cut per character based on the number of n, while in the n-gram word, the text data will be cut per word based on the number of n. This study uses the word N-gram with specifications, namely unigram, bigram, and trigram.

2.7. Word2vec

Word2vec is one of the Word Embedding methods, which converts words into a vector. It is an unsupervised learning algorithm using a neural network consisting of a hidden layer and a fully connected layer [19]. This method was developed by Mikolov in 2013 and had two architectures, namely Continuous Bag of Word (CBOW) and Skip-Gram. The CBOW architecture focuses on predicting the word target using the context of the surrounding word. In contrast, Skip-Gram focuses on predicting the target context based on word usage [20]. Figure 2 will explain the CBOW and Skip-gram architecture.

![Figure 2. CBOW and Skip-gram Architecture](image)

The input for word2vec is a text corpus, and the output is a set of vectors [21]. In this study, apply the word2vec method to determine each word's vector value and measure the closeness between words semantically in maximizing the possibility of predicting the word context or surrounding words. This method to determine the similarity of words requires calculations using the cosine similarity method.

2.8. Feature Expansion

Feature Expansion is a method to solve the problem of vocabulary mismatch, has a concept to identify features in tweets with a value of 0 (zero). That value will be replaced by related words with semantic similarities, and it is formed into a word dictionary (corpus) using the Word2vec method [6]. The data used to create a word dictionary (corpus) and as training data for the word2vec model consists of news data, tweet data, and a combination of news and tweet data. This study uses 142,544 news data, which consists of various news sources or topics, and this data is obtained from previous researchers.
The word dictionary (corpus) contains information about vocabulary and similar words. The Feature Expansion process will take the word similarity based on the order. Table 1 will show an explanation of taking similar words.

| Top Similarity | Description                                      |
|----------------|--------------------------------------------------|
| Top 1          | Taking word similarity based on the order of the top one. |
| Top 5          | Taking word similarity based on the order of the top five. |
| Top 10         | Taking word similarity based on the order of the top ten. |

Table 2 will show the results of making a word dictionary (corpus) on news data. For example, the vocabulary displayed is the word "Twitter" in the similarity Top 10.

| Vocab       | Top 1 | Top 2 | Top 3 | Top 4 | Top 5 |
|-------------|-------|-------|-------|-------|-------|
| Twitter     |       |       |       |       |       |
| mikroblogging | Top 6 | Top 7 | Top 8 | Top 9 | Top 10 |

The concept of feature expansion is explained in the following flowchart in Figure 3.

2.9. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a machine learning algorithm supervised (directed learning) and it has the concept of training on data to make predictions on class labels [4]. Other researchers widely implement support Vector Machine (SVM) because its implementation uses statistical learning to provide better results [22].

The concept of Support Vector Machine (SVM) defines the best Hyperplane (the boundary between two classes) by maximizing the distance between classes. In determining the best Hyperplane, it is necessary to calculate the margins. Based on the definition, the margin is the distance between hyperplanes from the closest point to each class (support vector) [22].

2.10. Random Forest (RF)

Random Forest is an ensemble method [23], it is a method to combine several classification which aim to find the best prediction. Random Forests are generally used to classify large amounts of data [9]. The learning concept of the Random Forest algorithm is to vote for each result from the decision tree and predict the class based on the most dominant result [13].
2.11. Confusion Matrix

Confusion Matrix is a method used to measure the performance of the classification processor used in evaluating a classification model to estimate the correct or incorrect object [4]. The output generated from this method is in the form of 2 or more classifiers. Classifier performance evaluation will calculate True positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values [4]. The four terms are formed in a Confusion Matrix table with a classifier of 2 classes, explained in Table 3.

| Actual Class | Prediction Class |         |         |
|--------------|-----------------|---------|---------|
| Positive (+) | TP (True Positive) | TN (True Negative) |         |
| Negative (-) | FP (False Positive) | FN (False Negative) |         |

Based on Table 3, it is possible to measure performance in determining the value of accuracy, precision, recall and F1 score [13]. Accuracy to identify the accuracy of the classification model in predicting the data correctly and will be shown in equation (2). Precision calculations to identify how accurate the results from the requested data are with the predictions provided by the model and will be shown in equation (3). Recall calculation to determine the model's success rate in finding information will be shown in equation (4). F1 – Score is the average harmonic value (Harmoni mean) of precision and recall and is shown in equation (5).

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

\[
F1 - \text{Score} = \frac{2 \times (\text{precision} \times \text{recall})}{\text{precision} + \text{recall}}
\]

3. RESULT AND DISCUSSION

3.1. Data

This research uses a collection of tweets and news data. The topic selection of these tweets is based on the trending topics that occurred on Twitter in October 2020 - June 2021, and on the topic of abusive words was determined to train the model in recognizing tweets that included hate speech. The news data consists of several sources of Indonesian news media with a total of 1,111 articles and 142,544 news data. In contrast to tweet data, news data were obtained from researchers [6] who became the reference in this study in applying the method. The news data also contains some information such as news topics, news media sources, URLs and news texts. Table 4 shows the detail crawled data with a list of keywords.

| Topic          | Keyword                        | Number of Tweet |
|----------------|--------------------------------|-----------------|
| Agama          | Agama, FPI                     | 6.651           |
| Harsh Words    | Anjing, babi, bajingan, bangsat, gila, etc. | 9.186           |
| Influencer     | Tirtha Hudi, selebgram         | 3.5.65          |
| Politic        | omnibuslaw                     | 1.169           |
| **Total**      |                                | **20.571**      |

Labeling tweet data is processed manually and involves four people. Tweet data is labeled based on two classes, tweets with hate speech and tweets with non-hate speech. In general, tweets classified as hate speech contain several elements: provocation, humiliation, discrimination against a person or group, even threats against ethnicity, religion, race, intergroup (SARA) [24], and tweets containing harsh words. Meanwhile, tweets that are not hate speech have the context of neutral words or sentences with a positive vibe. From the data labeling process, Table 5 will explain the percentage of the number of each tweet label.
3.2. Pre-processing

From several pre-processing steps that have been carried out on tweet data and news data, this section will describe the results of the process, for example in Table 6 the pre-processing results of tweet data will be shown.

Table 6. Results of P-processing Tweet Data

| Data Cleaning & Case Folding | Normalization | Filtering | Stemming |
|-----------------------------|---------------|-----------|----------|
| RT @CNNIndonesia: Kementerian Agama telah menerbitkan jadwal Imsakiyah selama bulan Ramadan 1442 Hijriah. Berikut jadwal terbaru Imsak dan Subuh di DKI Jakarta. https://bit.ly/3sgY6fK #CNNIndonesia | kementerian agama telah menerbitkan jadwal imsakiyah selama bulan ramadan hijriah berikut jadwal terbaru imsak dan subuh di daerah khusus ibukota jakarta | kementerian agama menerbitkan jadwal imsakiyah selama bulan ramadan hijriah berikut jadwal terbaru imsak dan subuh di daerah khusus ibukota jakarta | menteri agama terbit jadwal imsakiyah ramadan hijriah jadwal baru imsak subuh daerah khusus ibukota jakarta |

3.3. N-Gram

This study uses the features of unigram, bigram and trigram, and in Table 7 it will be explained the number of words that have been successfully processed from this method.

Table 7. Number of features N-gram

| N-gram | Total |
|--------|-------|
| Unigram | 17,943 |
| Bigram  | 92,682 |
| Trigram | 104,653 |

In the process, tweets with bigram and trigram features cannot process all the words at the classification step, due to the limited memory capacity of the hardware. Therefore, the solution is to set the maximum limit of features / words used, determine the features of 10,000, 20,000, and 30,000, this aims to find features with the best accuracy. However, Unigram does not apply feature restrictions, because the number of word features formed is only 17,943 and can still be processed by hardware.

3.4. Built Corpus Data

The formation of a word dictionary (corpus) uses news data, tweet data and a combination of news and tweet data. At this step, the Word2vec method will be used to find the similarity of the vocabulary. This study has tested each data, by taking based on the level of similarity (Top1, Top 5 and Top 10) which has been described in section 2, then in Table 8 will describe the number of vocabularies formed from each data tested.

Table 8. Number of Vocabulary

| Data | Total |
|------|-------|
| News | 231,553 |
| Tweet | 17,943 |
| News + Tweet | 239,786 |

3.5. Test Result

This study tested several scenarios by applying the Bag of Word (BOW), TF-IDF, and N-gram methods to represent values and weights in tweet data. Then, the Feature Expansion method will be applied to deal with vocabulary mismatches and classify the data using the Support Vector Machine (SVM) and Random Forest methods. The classification stage was repeated five times in determining the accuracy value and will be taken from the average value.

At the classification stage, this study applies a ratio of 90:10, where 90% is for training data and 10% for test data. Determination of this ratio because the resulting accuracy value is better than other ratios. This study
3.5.1. Determining Baseline Value

The first scenario will determine the baseline value, which is the fundamental value before weighting. At this step, the Bag of Word (BOW) method will be used. The baseline is determined based on the most optimal value of the two classifications (SVM and RF). Table 9 will show the results of the first scenario using SVM and Random Forest classification.

| Feature | SVM | Random Forest |
|---------|-----|---------------|
|         | Unigram | Bigram | Trigram | Unigram | Bigram | Trigram |
| 10,000  | 86,55    | 81,06  | 75,52   | 86,51    | 86,74  | 76,41   |
| 20,000  | -        | 81,84  | 76,67   | -        | 82,4   | 76,7    |
| 30,000  | -        | 81,75  | 76,26   | -        | 82,66  | 76,07   |
| All     | 86,60    | -      | 86,94   | -        | -      | -       |

Based on Table 9, the best accuracy at Support Vector Machine (SVM) baseline is 86,60% and at Random Forest (RF) baseline is 86,49%. In this scenario, the comparison of values for the two baselines (SVM and RF) is not too different. However, the random forest classification still gives the best accuracy results.

3.5.2. Effect of TF-IDF on Baseline

The second scenario is testing using the TF-IDF method. This method will be applied to each baseline to see the effect of the weighting method. Table 10 will show the results of the second scenario using SVM and Random Forest classification.

| Feature                          | Accuracy (%) |
|---------------------------------|--------------|
| SVM (Baseline) + TF-IDF         | 87,18 (+0,673)|
| Random Forest (Baseline) + TF-IDF | 87,96 (+1,174)|

Based on Table 10, the TF-IDF (weighting) method can increase the accuracy value of the two classification methods. At SVM baseline is increased accuracy up to 0,673%, while Random Forest baseline up to 1,174%.

3.5.3. Feature Expansion

The third scenario applies the Feature Expansion method using a word dictionary (corpus) from news data, tweet data, and a combination of news and tweet data. The similarity of the corpus applied is Top 1, Top 5, and Top 10. This section will implement the Feature Expansion method on the baseline to see its effect if it is applied to the Baseline (SVM and Random Forest). Table 11 will show the results of the third scenario on the SVM and Random Forest Baseline.

| Similarity | SVM | Random Forest |
|------------|-----|---------------|
|            | News + Tweet | News + Tweet |
| Top 1      | 87,41 (+0,931) | 87,07 (+0,201) |
|            | 86,93 (+0,382) | 87,11 (+0,123) |
| Top 5      | 87,07 (+0,101) | 87,07 (+0,342) |
|            | 86,72 (+0,135) | 87,10 (+0,342) |
| Top 10     | 87,07 (+0,202) | 87,43 (+0,570) |
|            | 86,76 (-2,222) | 87,43 (+0,570) |

Based on Table 11, the Feature Expansion method can increase the accuracy of Baseline SVM. Accuracy results of 87,41% are obtained from news data in Top 1 and can increase accuracy up to 0,93%. The Feature Expansion method can increase the accuracy of the Baseline Random Forest up to 0,57%. Accuracy results of 87,43% were obtained from the combination of news and tweet data included in the Top 10. Figure 4 will show...
replacement statistics for each word feature based on the most optimal results of the two classification methods in news data and combinations data (news + tweet) from the feature expansion process.

Based on Figure 4, the SVM classification gets the highest accuracy of 87.41%, obtained using the news data corpus in the Top 1 retrieval. However, the percentage of feature replacement is relatively lower than the others, which is 15.65%. So that the best accuracy value in the SVM classification tends to increase with a lower percentage of replacement. While the Random Forest classification gets the highest accuracy of 87.43%, obtained when using corpus from a combination of news and tweet data in the Top 10. Based on the graph above, the percentage of feature replacement is relatively higher than the others, which is 77.91%. The feature replacement in Random Forest affects increasing the accuracy value and tends to increase with the feature replacement that has the highest percentage.

The fourth scenario is to implement Feature Expansion with the TF-IDF method on the baseline. This scenario has the same concept as the third scenario. However, the difference is that it only adds the TF-IDF method. This section also uses a word dictionary (corpus) of news, tweet, combination data (news and tweets), and Top similarity implementation.

Table 12. Baseline (SVM and Random Forest) + TF-IDF + Feature Expansion Performance

| Similarity | SVM | Random Forest |
|------------|-----|---------------|
|            | News | Tweet | News + Tweet | News | Tweet | News + Tweet |
| Top 1      | 87.28 | 87.11 | 87.55 | 87.38 | 87.89 | 87.76 |
|            | (+0.786) | (+0.595) | (+1.100) | (+0.503) | (+1.095) | (+0.950) |
| Top 5      | 87.43 | 86.89 | 86.71 | 87 | 87.92 | 88.1 |
|            | (+0.964) | (+0.337) | (+0.123) | (+0.067) | (+1.129) | (+1.341) |
| Top 10     | 87.32 | 87.25 | 87.39 | 87.16 | 87.56 | 88.37 |
|            | (+0.830) | (+0.752) | (+0.909) | (+0.257) | (+0.715) | (+1.643) |

Based on Table 12, the Feature Expansion + TF-IDF method increases the accuracy of the SVM baseline. The best classification is obtained based on the use of corpus from news + tweet data in the Top 1, which is 87.55% and has an increase of 1.1%. The Feature Expansion + TF-IDF method also improves the accuracy of the Random Forest baseline. The best classification is obtained based on the corpus of combined data (news + tweet) in the Top 10, which is 88.37% and has an increase of 1.64%. Figure 5 will explain the statistics of the replacement of each word feature based on the most optimal results in both classification methods. This graph will identify the percentage of feature replacement based on the best accuracy results from using corpus data.

Figure 5. Word feature replacement graphic on Baseline + TF-IDF + Feature Expansion
Based on Figure 5, the SVM classification gets the highest accuracy of 87.55%, obtained using the corpus of combined data (news + tweets) in the Top 1 retrieval. However, the percentage of feature replacement is relatively lower than the others, which is 39.80%. So that the best accuracy value in the SVM classification tends to increase with a lower percentage of replacement. While the Random Forest classification gets the highest accuracy of 88.37%, obtained when using corpus from a combination of news and tweet data in the Top 1. Based on the graph above, the percentage of feature replacement is relatively higher than the others, which is 77.91%. The replacement feature in Random Forest affects the increase in the accuracy value and tends to increase with the feature replacement that has the highest percentage.

4. CONCLUSION

The hate speech detection system that has been built in this study uses the Feature Expansion (Word2vec) and Feature Extraction (TF-IDF and BOW) methods. In the Feature Expansion section, a corpus has been created using a set of data by taking the similarity of words in Top 1, Top5, and Top 10. The classification methods used in this research are Support Vector Machine (SVM) and Random Forest (RF). Based on the research results, applying the Feature Expansion method with the weighting method (TF-IDF) can increase the accuracy of both classifications and improve performance better than other test scenarios. The Support Vector Machine (SVM) classification method gets an accuracy of 87.55% or an increase of 1.10% from the baseline value. In comparison, the Random Forest method gets an accuracy of 88.37% or an increase of 1.64% from the baseline value. The highest accuracy results were obtained from the corpus of combined data (news+tweet), taking on the similarities in the Top 1 and Top 10. The researcher observed that taking similarity in the Top 5 for all data in the corpus could increase accuracy. However, the optimal and stable results remain on the similarity of Top 1 and Top 10 in the combined data. Researchers also observed the accuracy results in both classification methods (SVM and RF). Most of the highest accuracy of the test scenarios was obtained through the random forest classification. However, sometimes the SVM method gets better accuracy than random forest, which happens when the feature expansion method is implemented.

ACKNOELEDGMENT

Praise and gratitude the authors pray to Allah SWT who has provided smoothness in this research and in the preparation of this journal. As well as thanks to parents who always provide support and Supervising Professors who provide a lot of advice and scientific guidance to complete this research properly.

REFERENCES

[1] D. A. N. Taradhita and I. K. G. D. Putra, “Hate speech classification in Indonesian language tweets by using convolutional neural network,” J. ICT Res. Appl., vol. 14, no. 3, pp. 225–239, 2021, doi: 10.5614/itib.ict.res.appl.2021.14.3.2.
[2] K. M. Hana, Adiwijaya, S. Al Faraby, and A. Bramantoro, “Multi-label Classification of Indonesian Hate Speech on Twitter Using Support Vector Machines,” 2020 Int. Conf. Data Sci. Its Appl. ICODSA 2020, no. January 2021, 2020, doi: 10.1109/ICODSA50139.2020.9212992.
[3] A. P. Sitor, H. Murfi, S. Nurrohmah, and A. Akbar, “Sensing Trending Topics in Twitter for Greater Jakarta Area,” Int. J. Electr. Comput. Eng., vol. 7, no. 1, pp. 330–336, 2017, doi: 10.11591/ijece.v7i1.p330-336.
[4] P. P. A. Arsy Denica Pravina, Imam Cholissodin, “Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM),” J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 3, no. 3, pp. 2789–2797, 2019, [Online]. Available: http://j-iptik.ub.ac.id/index.php/jiptik/article/view/4793.
[5] T. T. A. Putri, S. Sriadhi, R. D. Sari, R. Rahmadani, and H. D. Hutahaean, “A comparison of classification algorithms for hate speech detection,” IOP Conf. Ser. Mater. Sci. Eng., vol. 830, no. 3, 2020, doi: 10.1088/1757-899X/830/3/032006.
[6] E. B. Setiawan, D. H. Widyantoro, and K. Surendro, “Feature expansion using word embedding for tweet topic classification,” Proceeding 2016 10th Int. Conf. Telecommun. Syst. Appl. TSSA 2016 Spec. Issue Radar Technol., 2017, doi: 10.1109/TSSA.2016.7871085.
[7] D. Winia, “Analysis of the use of the hate speech on social media in the case of presidential election in 2019,” J. Appl. Stud. Lang., vol. 3, no. 2, pp. 158–167, 2019, doi: 10.31940/jasl.v3i2.1541.
[8] F. A. Wenando, “Detection of Hate Speech in Indonesian Language on Twitter Using Machine Learning Algorithm,” PROCEEDING Cel/SciTech-UMRI 2019, vol. 4, pp. 6–8, 2019.
[9] K. Nugroho et al., “Improving random forest method to detect hatespeech and offensive word,” 2019 Int. Conf. Inf. Commun. Technol. ICOCIAT 2019, no. July, pp. 514–518, 2019, doi: 10.1109/ICOCIAT6746704.2019.8938451.
[10] I. Alfina, R. Mulia, M. I. Fanany, and Y. Ekana, “Hate speech detection in the Indonesian language: A dataset and preliminary study,” 2017 Int. Conf. Adv. Comput. Sci. Inf. Syst. ICACIS 2017, vol. 2018-Janua, no. October, pp. 233–237, 2018, doi: 10.1109/ICACIS.2017.8355039.
[11] Oryza Habibie Rahman, Gunawan Abdillah, and Agus Komarudin, “Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine,” J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 5, no. 1, pp. 17–23, 2021, doi: 10.29207/resti.v5i1.2700.
[12] U. G. Student, “Machine Learning Based Sentiment Classification,” vol. 29, no. 3, pp. 1062–1071, 2020.
13. S. Abro, S. Shaikh, Z. Ali, S. Khan, G. Mujtaba, and Z. H. Khand, “Automatic hate speech detection using machine learning: A comparative study,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 8, pp. 484–491, 2020, doi: 10.14569/IJACSA.2020.0110861.

14. J. Patihullah and E. Winarko, “Hate Speech Detection for Indonesia Tweets Using Word Embedding And Gated Recurrent Unit,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*., vol. 13, no. 1, p. 43, 2019, doi: 10.22146/ijccs.40125.

15. W. Trisari, H. Putri, R. Hendrowati, and L. Belakang, “Penggalian Teks Dengan Model Bag of Words Terhadap,” vol. 2, no. 1, pp. 129–138, 2020.

16. S. Qaiser and R. Ali, “Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents,” *Int. J. Comput. Appl.*, vol. 181, no. 1, pp. 25–29, 2018, doi: 10.5120/ijca2018917395.

17. M. A. Lestari, P. P. Adikara, and S. Adinugroho, “Rekomendasi Lagu berdasarkan Lirik dan Genre Lagu menggunakan Metode Word Embedding (Word2Vec),” vol. 3, no. 8, pp. 2548–964, 2019. [Online]. Available: http://j-ptiik.ub.ac.id.

18. Z. Pratama, E. Utami, and M. R. Arief, “Analisa Perbandingan Jenis N-GRAM Dalam Penentuan Similarity Pada Deteksi Plagiat,” *Creat. Inf. Technol. J.*, vol. 4, no. 4, p. 254, 2019, doi: 10.24076/citec.2017v4i4.118.

19. A. Nurdin, B. Anggo Seno Aji, A. Bustamining, and Z. Abidin, “Perbandingan Kinerja Word Embedding Word2Vec, Glove, Dan Fasttext Pada Klasifikasi Teks,” *J. Tekno Kompak*, vol. 14, no. 2, p. 74, 2020, doi: 10.33365/jtk.v14i2.732.

20. T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *1st Int. Conf. Learn. Represent. ICLR 2013 - Work. Track Proc.*, no. January 2013, 2013.

21. irwan budiman, M. R. Faisal, and D. T. Nugrahadi, “Studi Ekstraksi Fitur Berbasis Vektor Word2Vec pada Pembentukan Fitur Berdimensi Rendah,” *J. Komputasi*, vol. 8, no. 1, pp. 62–69, 2020, doi: 10.23960/komputasi.v8i1.2517.

22. R. A. Rizal, I. S. Girsang, and S. A. Prasetyo, “Klasifikasi Wajah Menggunakan Support Vector Machine (SVM),” *REMIK (Riset dan E-Jurnal Manaj. Inform. Komputer)*, vol. 3, no. 2, p. 1, 2019, doi: 10.33395/remik.v3i2.10080.

23. A. Primajaya and B. N. Sari, “Random Forest Algorithm for Prediction of Precipitation,” *Indones. J. Artif. Intel1. Data Min.*, vol. 1, no. 1, p. 27, 2018, doi: 10.24014/ijaidm.v1i1.4903.

24. K. Antariksa, Y. S. Purnomo WP, and E. Ernawati, “Klasifikasi Ujaran Kebencian pada Cuitan dalam Bahasa Indonesia,” *J. Buana Inform.*, vol. 10, no. 2, p. 164, 2019, doi: 10.24002/jbi.v10i2.2451.