**Hate Speech Detection on Indonesian Tweet Using Multinomial Naïve Bayes**

**First Author1, Second Author2, Third Author3 (10 pt)**

1,3National Chung Cheng University, Taiwan (9 pt)

2Universitas Bumigora, Indonesia (9 pt)

|  |  |  |
| --- | --- | --- |
| **Article Info** |  | **ABSTRACT** (10 PT) |
| ***Article history:***  Received mm dd, yyyy  Revised mm dd, yyyy  Accepted mm dd, yyyy |  | Detecting hate speech on social media is a significant challenge, particularly considering the increasing volume of data. Currently, the task of monitoring the spread of hate speech is officially carried out by the Ministry of Communication and Information and the National Cyber and Encryption Agency. However, it must be acknowledged that the dissemination of hate speech continues to rise over time. In order to identify content containing hate speech on social media, the Indonesian National Police has formed a team known as the "Cyber Patrol." On the other hand, combating the spread of hate speech requires active participation from all segments of society. This research aims to develop a system capable of detecting sentences containing hate speech using the Multinomial Naive Bayes algorithm and the Term Frequency-Inverse Document Frequency feature extraction technique. A total of 300 data from the Twitter platform were used as training data, while 30 data were used as testing data. The research results showed an accuracy rate of 76% with a 90%:10% data split between training and testing. These testing results illustrate that the developed system can identify tweets containing hate speech using the Multinomial Naïve Bayes method |
| ***Keywords:***  Text-Mining  Hate Speech  Multinomial Naïve Bayes  TF-IDF |
| *This is an open access article under the* [*CC BY-SA*](https://creativecommons.org/licenses/by-sa/4.0/) *license.* |
| ***Corresponding Author:***  Name of Corresponding Author,  Department of Engineering And Design,  Universitas Bumigora.  Email: corresp-author@universitasbumigora.ac.id | | |

1. **INTRODUCTION**

Since 2018, the Indonesian Ministry of Communication and Information Technology (Kominfo) has handled 3,640 pieces of content related to hate speech based on ethnicity, religion, race, intergroup relations (SARA), and politicsThe failure of Indonesia being host the U-20 World Cup has caused disappointment among some people. Many people blame politically charged narratives as the main cause of this failure. The escalation of issues on social media further increased on March 29, 2023, after FIFA officially withdrew Indonesia's status as the host of the U-20 World Cup. Since then, there has been an increase in confirmed cases of hate speech over time.Hate speech can be delivered both orally and in writing. If delivered orally, it occurs directly, face-to-face with the interlocutor. Meanwhile, when someone conveys hate speech in writing, they express their feelings and thoughts through written media, such as books or social media [1]. An example of a hate speech case in Indonesia was the case involving the artist Ahmad Dhani. He was considered to have spread hatred towards a particular group through his Twitter account. His tweet contained the following statement: Anyone who supports the blasphemer of religion is a scoundrel who deserves to be spat on their face[[1]](#footnote-1).

There have been other studies conducted previously on hate speech classification. The research focused on Sentiment Analysis of Hate Speech on Twitter users using the Naïve Bayes classifier (NBC) method with 5000 data implemented using Python. That study achieved the highest accuracy of 80% with data split of 70% training and 30% testing [2]. willianto et al [3] using data sourced from Facebook, consisting of 500 data with a data split of 80% training and 20% testing, Obtained an accuracy score of 83%. Ihsan et al [4] added abusive word classes to their research, using a data split proportion of 90%:10% and the Decision Tree Algorithm, achieving an accuracy score of 70.48%. Hakiem et al [5] using Multinomial Naïve Bayes on N-Gram with Feature Selection of Information Gain The accuracy rate achieved is 84% with a total of 250 data.

In this current research, a system has been developed to detect instances of hate speech in Indonesian tweets related to PSSI (Indonesian Football Association) and political topics. The research classifies content into three categories: non-hate speech (non-HS), contempt (penghinaan), and provocation (provokasi). The data for this study is sourced from Twitter, and guidance for labeling is provided by mrs Ema. The dataset is partitioned with a 90%:10% ratio for training and testing data, and the algorithm employed for classification, as well as feature extraction, involves Multinomial Naïve Bayes in conjunction with Term-Frequency-Inverse Document Frequency (TF-IDF).

The contributions that can be offered in this research involve the development of a web-based application capable of detecting texts falling under the category of hate speech. Additionally, the study also evaluates and compares the accuracy results when implementing the TF-IDF method for feature extraction

1. **RESEARCH METHOD (10 PT)**

Figure 1 illustrates the processes to be conducted in this research. The first step involves data collection of tweets through crawling, which will become the dataset. The gathered data will be split into two distinct sections: the training data and the testing data. The training data contain five steps, The first step is manual labeling consisting of three classes, namely " non hs (non-hate speech), penghinaan (contempt), and provokes (provocation),The second step is preprocessing to obtain clean data, and the third step is inputting the data into the database, The fourth step is creating a word corpus, and the last step is feature extraction. The testing data contains two steps,The first is preprocessing to obtain cleaned data. In the subsequent stage, the multinomial Naïve Bayes algorithm is applyed to attain classification results.

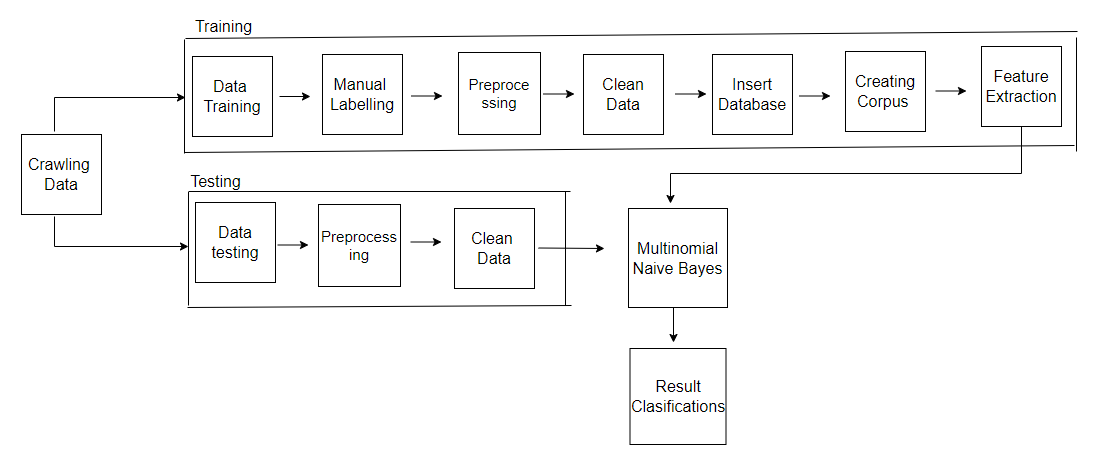


Figure 1 The program's overall process flow.

**2.1. Dataset Collection**

This research is conducted through several stages. The study starts with crawling data, which involves using a website that can scrape data from Twitter called Apify and Netlytic beside that we utilize the Twitter Application Programming Interface (API) to retrieve Twitter data. Specific keywords such as PSSI (persantuan sepakbola seluruh indonesia) and politics are used in the crawling process. The data acquired through the crawling procedure will be stored in an Excel file format.

**2.2 Labeling Data**

After collecting the data in the data crawling phase, subsequently, the data will be segregated into two categories: training data and testing data.. The training data is used to train the model or algorithm that will be used in the data analysis. Conversely, the testing data assesses the model's performance on data it has yet to encounter previously. After the data division is done, the training data will undergo a manual labeling process. The labeling in this research involves three classes: non-hs (non-hate speech), penghinaan (contempt), and provokasi (provocation). The labeling is done manually by analyzing the meaning of each tweet. In analyzing the tweet data in this research, a linguistic meaning approach is used, using the concept of conceptual meaning. This conceptual meaning approach refers to the meaning of words or sentences based on grammatical meaning without considering the context[6]

**2.2 Preprocessing**

Preprocessing is the initial stage of selecting and transforming data structurely [6]. The preprocessing stage is performed to prepare the data so that the text data becomes more structured by removing noise present in the data, thus facilitating the classification process[7]. In this research, the preprocessing stage shown in Figure 2 includes several steps,begin with case folding. Case Folding is a step to convert all uppercase letters to lowercase letters. This step is performed to ensure uniformity in the text and make it easier to process, second step is data cleaning.Cleaning the data by removing extraneous elements such as emoticons, punctuation, and numerical values in order to streamline and reduce unnecessary information [8]. Third step is slang word. aims to convert non-standard words into standard words that are ready for processing. Non-standard words may include abbreviations or slang terms. Examples of non-standard words commonly seen on social media are “begajulan," which means “nakal," and “bumil," which means”ibu hamil." In this stage, these non-standard words will be converted into standard words for better processing. Fourth step is Stemming. Stemming is a word-processing process that aims to obtain the base form of a word after removing affixes according to certain rules. This process is done because words with prefixes, suffixes, or infixes can complicate the matching of related words [9]. Last is Stopword involves extracting important words and removing words that do not carry meaning in the classification process. The words in question include common words such as "yang," "di," "dan," "diri," and so on

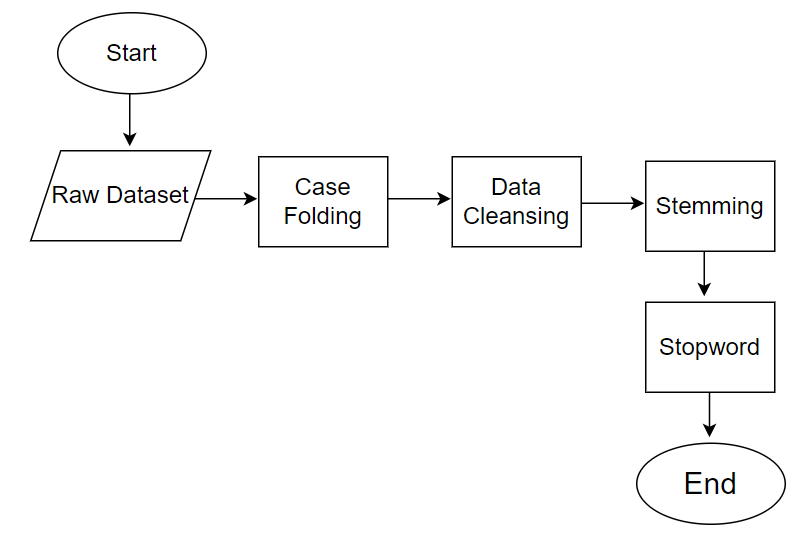


Figure 2. Preprocessing Stages

**2.3 Feature Extraction**

TF-IDF stands as one of the methods for weighting words. It combines Term Frequency (TF) and Inverse Document Frequency (IDF). Term Frequency, denoted as TF, quantifies the number of times a term appears within a document. Certain terms could occur more frequently in longer documents compared to shorter ones. On the other hand, IDF computes the occurrence of terms across the entire document collection. IDF assigns higher values to terms that appear less frequently, while more common terms receive lower values [10].TF-IDF is a method that combines two-word weighting approaches, namely by calculating word frequency and applying the inverse of the number of documents containing that word[11]. The TF-IDF calculation can be performed using equations (1), (2), and (3).

|  |  |
| --- | --- |
|  | (1) |
|  | (2) |
|  | (3) |

Where :

(𝑑,𝑡) : Frequency of the term t appearing in document d.

𝐼𝐷𝐹(𝑡) : Inverse document frequency in term t

𝑁𝑑 : Total Number of documents

𝑑(𝑡) : Number of documents containing term t.

n(d,t) : Total number of terms in a document d

TF(t) : Term frequency in term t

**2.4 Multinomial Naïve Bayes**

Multinomial Naïve Bayes Classification is a form of Bayes algorithm commonly used in text classification. Multinomial Naïve Bayes documents are considered as "bag of words," where the sequence of word occurrences in the paper is disregarded, and each word is processed using a multinomial distribution[12]. The Multinomial Naïve Bayes classification process encompasses a tripartite sequence. Initially, probabilities are computed for each class. Subsequently, probabilities for each word within every class are computed. Lastly, these calculations are juxtaposed to ascertain the anticipated class. The calculation of Multinomial Naïve Bayes can be seen in equation (4).

|  |  |
| --- | --- |
| (𝑐, 𝑑) = 𝑥 𝑃(𝑡1, 𝑐) 𝑥 … 𝑥 𝑃(𝑡𝑛, 𝑐) | (4) |

Where :

(𝑐, 𝑑) : Probability of document d belonging to class c

𝑁 : Count of class c document

: Total number of documents

𝑡 : The n word in document d

(𝑡𝑛, 𝑐): Probability of word tn given class c

The formula for calculating the probability of the n word using TF-IDF word weighting can be seen in the equation (5) [11]

|  |  |
| --- | --- |
| 𝑃 (𝑡𝑛, 𝑐) = | (5) |

Where:

𝑊𝑐 : The TF-IDF weighting value or W of term t in category c

∑𝑊′ : Total sum of W values from all terms in category c

𝐵 : Total sum of W values for unique terms (IDF value not multiplied by tf) across all documents

**2.5 Confusion Matrix**

Confusion Matrix determines how effectively a model classifies data[13].There exist four terms signifying the results of the classification procedure: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The True Negative (TN) value signifies the number of negative data instances correctly identified, whereas False Positive (FP) represents the negative data instances incorrectly identified as positive [14]. Applying the Confusion Matrix makes it possible to calculate Accuracy, Precision, and Recall values. The calculations for Accuracy, Precision, and Recall can be seen in equations (6), (7), and (8).

|  |  |
| --- | --- |
| 𝐴𝑐𝑐𝑢𝑟𝑎𝑐𝑦 = | (6) |
| 𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛 = | (7) |
| 𝑅𝑒𝑐𝑎𝑙𝑙 = | (8) |

1. **RESULTS AND ANALYSIS (10 PT)**

**3.1. Dataset**

This research utilizes a Dataset sourced from the Twitter social media platform. The total amount of successfully collected data is 300 tweets, all generated between March 29, 2023, and June 30, 2023. An instance of the collected data is shown in Table 1

Table 1 Example of Dataset

|  |  |  |  |
| --- | --- | --- | --- |
| Created At | Username | Tweets | Tweets (Bahasa) |
| 28-05-2023 | @Mira\_SasMiita | Good news: Indonesia has officially been selected by the Asian Football Confederation (AFC) as one of the 11 host countries for the qualifying round of the 2024 AFC U-23 Asian Cup. This achievement is attributed to the role of the Indonesian Football Association (PSSI) Chairman @erickthohir, who has international relations. | Good news, Indonesia resmi didapuk oleh Konfederasi Sepak Bola Asia (AFC) menjadi salah satu dari 11 negara tuan rumah babak kualifikasi Piala Asia U-23 2024. Hal ini tak terlepas dari peran Ketum PSSI @erickthohir yang memiliki relasi internasional |

**3.2. Labeling**

The next stage is data labeling. In the labeling phase, the acquired data will be classified according to the predefined categories in this study. This research applies three classes: non-hs (nonhate speech), penghinaan (contempt), and provokasi (provocation). The labeling is conducted manually by analyzing the meaning of each tweet. In the analysis of data tweets in this study, a linguistic meaning approach is utilized, applying conceptual concepts. The labeling process is shown in Table 2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Created At | Username | Tweets | Tweets (Bahasa) | Label |
| 28-05-2023 | @Mira\_SasMiita | Good news: Indonesia has officially been selected by the Asian Football Confederation (AFC) as one of the 11 host countries for the qualifying round of the 2024 AFC U-23 Asian Cup. This achievement is attributed to the role of the Indonesian Football Association (PSSI) Chairman @erickthohir, who has international relations. | Good news, Indonesia resmi didapuk oleh Konfederasi Sepak Bola Asia (AFC) menjadi salah satu dari 11 negara tuan rumah babak kualifikasi Piala Asia U-23 2024. Hal ini tak terlepas dari peran Ketum PSSI @erickthohir yang memiliki relasi internasional | Non Hs (non hate speech) |

Table 12 presents an example tweet labeled as non hs (non hate speech), as the respective tweet content does not fall within the categories of penghinaan (contempt) or provokasi (provocation). In the context of tweets categorized as defamation, they encompass content that evokes emotions and belittles the dignity of individuals or institutions. Conversely, tweets classified as provocation aim to foment hatred or enmity towards specific individuals and/or societal groups.

**3.3 Preprocessing**

After the labeling process is completed, the data will undergo the preprocessing stage. The purpose of this preprocessing is to clean the data from noise or disturbances, resulting in cleaner and more structured data. This cleaned data will be used to build a word model that will be utilized in the classification process. The preprocessing steps can be seen in Table 3.

Table 3. Preprocessing Steps

|  |  |
| --- | --- |
| Source Data | Case Folding |
| naik lagi... Yang teriak-teriak israel itu, pas kanjuruhan diem aja. Dia cuma butuh politik. Gak peduli nyawa.. | naik lagi... yang teriak-teriak israel itu, pas kanjuruhan diem aja. dia cuma butuh politik. gak peduli nyawa.. |
| Source Data | Data Cleansing |
| naik lagi... Yang teriak-teriak israel itu, pas kanjuruhan diem aja. Dia cuma butuh politik. Gak peduli nyawa.. | naik lagi yang teriak-teriak israel itu pas kanjuruhan diem aja dia cuma butuh politik gak peduli nyawa |
| Source Data | SlangWord |
| naik lagi... Yang teriak-teriak israel itu, pas kanjuruhan diem aja. Dia cuma butuh politik. Gak peduli nyawa.. | naik lagi yang teriak-teriak israel itu pas kanjuruhan diam saja dia hanya butuh politik tidak peduli nyawa |
| Source Data | Stemming |
| naik lagi... Yang teriak-teriak israel itu, pas kanjuruhan diem aja. Dia cuma butuh politik. Gak peduli nyawa.. | naik lagi yang teriak-teriak israel itu saat kanjuruhan diam saja dia hanya butuh politik tidak peduli nyawa |
| Souce Data | Stopword |
| naik lagi... Yang teriak-teriak israel itu, pas kanjuruhan diem aja. Dia cuma butuh politik. Gak peduli nyawa.. | teriak-teriak israel kanjuruhan diam butuh politik tidak peduli nyawa |

**3.4 Word Weighting**

The next stage is Word Weightinh, where words will be transformed into numerical representations. These representations will be utilized in the classification stage using the Multinomial Naïve Bayes method. Table 4 will serve as an example dataset for the word weighting stage up to the classification stage

Table 3. Example Of Dataset

|  |  |
| --- | --- |
| Tweet | Kelas |
| Indonesia resmi | non hs |
| Indonesia Hebat | Non hs |
| Pssi tidak prestasi | Penghinaan |
| Tamak Jabatan | Penghinaan |
| Untung rugi situasi politik | Provokasi |
| Politik tidak peduli nyawa | Provokasi |

From the example dataset, a word corpus will be created, the word corpus is obtained as follows: {indonesia, resmi, pemain, hebat, pssi, tidak, prestasi, tamak, jabatan, untung, rugi, situasi, polittik, peduli, nyawa}. The feature extraction process consists of three steps: the first is calculating the TF-IDF value for each word in the word corpus for each class, the second step is computing the total value of TF-IDF for each word in each class, and the third step is calculating the IDF value for each word in the entire datasets. Below is the outcome of computing the TF-IDF value for each term within the word corpus assigned to the non hs class, as illustrated in Table 5

Table 5 TF-IDF Calculation for Each Word in the Non hs Class

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Term | TF (Doc 1) | TF (Doc2) | DF | IDF | TF-IDF (Doc 1) | TF-IDF (Doc 2) |
| indonesia | 0,5 | 0 | 1 | 0,3010299 | 0,15051495 | 0 |
| resmi | 0,5 | 0 | 1 | 0,3010299 | 0,15051495 | 0 |
| pemain | 0 | 0,5 | 1 | 0,3010299 | 0 | 0,15051495 |
| hebat | 0 | 0,5 | 1 | 0,3010299 | 0 | 0,15051495 |

In Table 5, word explanations only cover up to the word "hebat." This is because, for the word corpus from "pssi" to "nyawa," these words do not appear in documents labeled as "non hs." Therefore, the TF-IDF values for these words are definitely 0. For the word "Indonesia" in the non hs class label, a TF (Term Frequency) value of 0.5 is obtained for the first document. This value is found by counting how many times the word "Indonesia" appears in the first document and dividing it by the length of words in the first document. The frequency of the word "Indonesia" in the first document is 1, and the word length in the first document is 2. The result of the calculation is 0,5.

The IDF (Inverse Document Frequency) value for the word "Indonesia" in the initial document is 0.3010299. This value is calculated by taking the logarithm of the total count of data labeled as non hs class divided by the count of occurrences of the word "Indonesia" in data labeled as a non hs class. The total data in the non hs class is 2, and the number of documents containing the word "Indonesia" with the non hs class label is 1. The result of the calculation log () is 0.3010299

Lastly, for the TF-IDF value of the word "Indonesia" in the first document, a value of 0.15051495 is obtained. This value is found by multiplying the TF value with the IDF value. The TF value for the word "Indonesia" in the first document is 0.5, and the IDF value for the word "Indonesia" is 0.3010299. The result of the calculation 0.5 multiplied by 0.3010299 is 0.15051495. The same calculation method is applied to the subsequent words. The explanation of the TF-IDF calculation results for the penghinaan (contempt) and provokasi (provocation) classes will be shown in Table 6 and Table 7

Table 6 TF-IDF Calculation for Each Word in the Penghinaan Class

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Term | TF (Doc 1) | TF (Doc2) | DF | IDF | TF-IDF (Doc 1) | TF-IDF (Doc 2) |
| tamak | 0,5 | 0 | 1 | 0,3010299 | 0,15051495 | 0 |
| jabatan | 0,5 | 0 | 1 | 0,3010299 | 0,15051495 | 0 |
| pssi | 0 | 0,3333 | 1 | 0,3010299 | 0 | 0,10033326567 |
| tidak | 0 | 0,3333 | 1 | 0,3010299 | 0 | 0,10033326567 |

Table 7 TF-IDF Calculation for Each Word in the Procokasi Class

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Term | TF (Doc 1) | TF (Doc2) | DF | IDF | TF-IDF (Doc 1) | TF-IDF (Doc 2) |
| untung | 0,25 | 0 | 1 | 0,3010299 | 0,075257475 | 0 |
| rugi | 0,25 | 0 | 1 | 0,3010299 | 0,075257475 | 0 |
| situasi | 0,25 | 0 | 1 | 0,3010299 | 0,075257475 | 0 |
| politik | 0,25 | 0,25 | 2 | 0 | 0 | 0,15051495 |

After calculating the TF-IDF values for each word list in each class, the next step is to compute the weight (w) for each word in the corpus within each class. The results of the calculations for several words within the corpus are presented in Table 8.Table 8 Calculation of the Weight for Each Word in the Provocation Class.After calculating the TF-IDF values for each word list in each class, the next step is to compute the weight (w) for each word in the corpus within each class. The results of the calculations for several words within the corpus are presented in Table 8.

Table 8 Calculation of the Weight for Each Word in the Provocation Class.

|  |  |  |  |
| --- | --- | --- | --- |
| Term | Non hs | Penghinaan | Provokasi |
| indonesia | 0,15051495 | 0 | 0 |
| resmi | 0,15051495 | 0 | 0 |
| pemian | 0,15051495 | 0 | 0 |
| hebat | 0,15051495 | 0 | 0 |
| sum | 0,6020598 | 0,60202969701 | 0,45154485 |

In Table 8, only calculations for a few words present in the corpus are provided, while the sum represents the total of calculations for all existing words. For the word "Indonesia," a weight value of 0.15051495 is obtained for the non hs (non hate speech) class. This value is obtained by adding the TF-IDF values for the word "Indonesia" in all documents labeled as non hs in the datasets.

The same method is used to calculate the weight value for each word in other classes. The next step involves the summation of the total weights for words in each class. This calculation result will be utilized in the classification phase.The final step before classification is to calculate the IDF values for all words in the corpus with all data in datasets. The IDF values for the entire word corpus in relation to the entire dataset will be presented in Table 9

Table 9 Calculation IDF Value for Each Word in Word Corpus For All Datasets.

|  |  |  |
| --- | --- | --- |
| Term | Non hs | Penghinaan |
| indonesia | 1 | 0,77815125 |
| resmi | 1 | 0,77815125 |
| pemian | 1 | 0,77815125 |
| hebat | 1 | 0,77815125 |
| tamak | 1 | 0,77815125 |
| tidak | 2 | 0,477121254 |
| indonesia | 1 | 0,77815125 |
| resmi | 1 | 0,77815125 |
| Sum | 11,070208758 | |

The IDF value for the word "Indonesia" is 0.77815125, derived from the logarithm of the total dataset documents divided by the documents containing the word "Indonesia" (log()). The value 11.070208758 represents the sum of IDF values for all words

**3.5 Multinomial Naïve Bayes**

The subsequent phase is calculations through the Multinomial Naïve Bayes. which involves two stages.. The first stage involves calculating the probabilities for each class using the equation (1) and calculating the probabilities of each word for each class using the equation (5). Table 10 will shown the calculated values of the probability for each class

Table 10 The Probabilities of Each Class.

|  |  |  |
| --- | --- | --- |
| Non hs | Penghinaan | Provokasi |
| 0,333333 | 0,333333 | 0,333333 |

As an illustration, let's take the sentence "Indonesia gagal" and subject it to classification using the Multinomial Naive Bayes approach. After obtaining the probability values for each class, the subsequent step is determining the probability values for every word in the sentence across each existing class. To calculate the word probability, equation (5) can be used.Table 11 shows the calculated results of the word probabilities for the classes non-hs (non hate speech) class.

Table 11 Probability of words in the "non-hate speech" class.

|  |  |
| --- | --- |
| term | probabilitas |
| indonesia | P( indonesia | non hs) = = 0,0985682 |
| gagal | 𝑃( gagal | non hs) = = 0,0856731 |

The calculation for the penghinaan (contempt) and provokasi (provocation) class is performed using the same method.After obtaining the probability values for each word in each existing class, the final step is to multiply these word probability values by the probability values of each class. This step will be shown in Table 12

Table 12 Measurement Results of Multinomial Naïve Bayes

|  |  |
| --- | --- |
| Non hs | 0,333333\*0,0985682\*0,0856731= 0,0028148782702589 |
| Penghinaan | 0,333333\*0,0856733\*0,0856733= 0,0024466356643252 |
| Provokasi | 0,333333\*0,0867923\*0,0867923=0,0025109652687956 |

Referring to Table 12, the maximum value is achieved within the Nonhs category. As a result, the classification outcome for the “Indonesian gagal” is labeled as non hs (non hate speech)

**3.6 Evaluation**

The purpose of testing is to evaluate how effective the application is in detecting and identifying content that contains hate speech. This study applye a training dataset comprising 300 data and a testing dataset comprising 30 data. Based on the testing results, an accuracy of 76% was achieved when using TF-IDF, whereas the accuracy value without using TF-IDF was obtained by 50%. The detailed are shown in Table 13.

Table 13 shows the sults of the calculation accuracy,precision,recall

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm | accuracy | precision | recall |
| Multinomial Naïve Bayes + TF-IDF | 76% | 83% | 86% |
| Multinomial Naïve Bayes | 50 % | 66% | 87% |

Table 14 Result of the classification example

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tweet (Bahasa) | Tweets | Actual labeling | Result using TF -IDF | Result Without  TF-IDF |
| Politik itu tentang kedewasaan. Yang dipertengkarkan adalah ide bukan dengkul. | Politics is about maturity.It's ideas that are in dispute, not your knee. | Non hs (non hate speech) | Non hs (non-hate speech) | Provokasi  (provocation) |

Table 14 shows the classification outcomes concerning tweet data, employing TF-IDF feature extraction and a non-TF-IDF approach based on the system's testing outcomes. The system recurrently encounters errors in classifying manually labeled nonhs (non hate speech) tweet data. Upon conducting classification without using TF-IDF from a complete set of 24 manually labeled data instances assigned to the months category, the system achieves accurate sort solely for 13 data instances. However, when the type is performed utilizing the TF-IDF technique, the system attains precision in classifying 20 data instances

**3.3. Sub section 3**

xx

3.3.1. Subsub section 3

yy

3.3.2. Subsub section 2

yy

1. **CONCLUSION (10 PT)**

Provide a statement that what is expected, as stated in the "Introduction" chapter can ultimately result in "Results and Analysis" chapter, so there is compatibility. Moreover, it can also be added the prospect of the development of research results and application prospects of further studies into the next (based on result and discussion).

**ACKNOWLEDGEMENTS (10 PT)**

The Acknowledgments section is optional. Research sources can be included in this section.

**REFERENCES (10 PT)**

The main references are international journals and proceedings. All references should be to the most pertinent, up-to-date sources **and the minimum of references are 15**. References are written in IEEE style. Please use a consistent format for references – see examples below (9 pt):

[1] Y. Suryani, R. Istianingrum, and S. U. Hanik, “Linguistik Forensik Ujaran Kebencian terhadap Artis Aurel Hermansyah di Media Sosial Instagram,” *BELAJAR Bhs. J. Ilm. Progr. Stud. Pendidik. Bhs. dan Sastra Indones.*, vol. 6, no. 1, pp. 107–118, 2021, doi: 10.32528/bb.v6i1.4167.

[2] I. Riadi and A. Fadlil, “Analisis Sentimen HateSpeech pada Pengguna Layanan Twitter dengan Metode Naïve Bayes Classifier ( NBC ),” vol. 10, no. 2, pp. 0–9, 2023, doi: 10.30865/jurikom.v10i2.5984.

[3] Willianto, I. A. Musdar, Junaedy, and H. Angriani, “Implementasi Teori Naive Bayes Dalam Klasifikasi Ujaran Kebencian,” *J. Inform. Univ. Pamulang*, vol. 6, no. 4, pp. 666–671, 2021.

[4] F. Ihsan, I. Iskandar, N. S. Harahap, and S. Agustian, “Decision tree algorithm for multi-label hate speech and abusive language detection in Indonesian Twitter,” *J. Teknol. dan Sist. Komput.*, vol. 9, no. 4, pp. 199–204, 2021, doi: 10.14710/jtsiskom.2021.13907.

[5] M. Hakiem, M. A. Fauzi, and Indriati, “Klasifikasi Ujaran Kebencian pada Twitter Menggunakan Metode Naïve Bayes Berbasis N-Gram Dengan Seleksi Fitur Information Gain,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2443–2451, 2019, [Online]. Available: http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4682

[6] N. L. Ratniasih and N. W. Ninik Jayanti, “Sentiment Analysis of Stakeholder Satisfaction Measurement,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 16, no. 2, p. 217, 2022, doi: 10.22146/ijccs.72245.

[7] Yuyun, Nurul Hidayah, and Supriadi Sahibu, “Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 820–826, 2021, doi: 10.29207/resti.v5i4.3146.

[8] Febiana Anistya and Erwin Budi Setiawan, “Hate Speech Detection on Twitter in Indonesia with Feature Expansion Using GloVe,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1044–1051, 2021, doi: 10.29207/resti.v5i6.3521.

[9] R. Rasenda, H. Lubis, and R. Ridwan, “Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter,” *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 369, 2020, doi: 10.30865/mib.v4i2.2051.

[10] M. Johari and A. Laksito, “The Hybrid Recommender System of the Indonesian Online Market Products using IMDb weight rating and TF-IDF,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 5, pp. 977–983, 2021, doi: 10.29207/resti.v5i5.3486.

[11] F. A. Ramadhan, S. H. Sitorus, and T. Rismawan, “Penerapan Metode Multinomial Naïve Bayes untuk Klasifikasi Judul Berita Clickbait dengan Term Frequency - Inverse Document Frequency Application of Multinomial Naïve Bayes News Clickbait with Term Frequency - Inverse Document Frequency,” vol. 11, no. 1, pp. 70–76, 2023, doi: 10.26418/justin.v11i1.57452.

[12] L. M. Siniwi, A. Prahutama, and A. R. Hakim, “QUERY EXPANSION RANKING PADA ANALISIS SENTIMEN MENGGUNAKAN KLASIFIKASI MULTINOMIAL NAÏVE BAYES (Studi Kasus : Ulasan Aplikasi Shopee pada Hari Belanja Online Nasional 2020),” *J. Gaussian*, vol. 10, no. 3, pp. 377–387, 2021, doi: 10.14710/j.gauss.v10i3.32795.

[13] A. F. Fandisyah, N. Iriawan, and W. S. Winahju, “Deteksi Kapal di Laut Indonesia Menggunakan YOLOv3,” *J. Sains dan Seni ITS*, vol. 10, no. 1, 2021, doi: 10.12962/j23373520.v10i1.59312.

[14] Karsito and Susanti Santi, “Klasifikasi Kelayakan Peserta Pengajuan Kredit Rumah Dengan Algoritma Naive Bayes Di Perumahan Azzura Residence,” *J. Teknol. Pelita Bangsa*, vol. 9, pp. 43–48, 2019.

1. https://twitter.com/AHMADDHANIPRAST/status/838660282222178304 [↑](#footnote-ref-1)