



Sentiment Analysis of Twitter Data Related to the Ratification of the TNI Bill Using Orange Data Mining

Ilham Hibatullah^{1*}, Rizky Ichsan Nur Rahman², Hikmal Maulana³, David Utomo⁴, Sumanto⁵, Andi Diah Kuswanto⁶

^{1,2,3,4,5,6}Informatika, Bina Sarana Informatika University

ilhamhibatullh@gmail.com^{1*}, rizkyichsan29@gmail.com², hikmalmaulana212@gmail.com³, davidutomo92@gmail.com⁴, sumanto.sto@bsi.ac.id⁵, andi.ahk@bsi.ac.id⁶

Abstract

Twitter is one of the social media platforms where users can post photos, videos and talk about current issues. One of the current issues is the issue of the ratification of the TNI Bill. The method used is naïve bayes with the help of the orange data mining application. Researchers managed to group 400 tweets from Twitter based on the sentiments and emotions contained in them. The results showed that responses were Negative with a total of 166 tweets, neutral sentiment reaching 140 tweets, and 94 tweets showing positive sentiment. If the percentage of polarity analysis is calculated, the results are as large, negative (41.5%), neutral (35%), and positive (23.5%). The Naïve Bayes model used is able to classify data with fairly good accuracy, which is 82%. Although there is still an imbalance in the amount of data between positive, negative, and neutral sentiments, in general this method is quite reliable for describing public opinion on social media. In addition, this study shows that Orange Data Mining can be a practical and effective tool in analyzing texts or opinions in cyberspace.

Keywords: *twitter, orange data mining, sentiment, ekman algorithm, ruu tni*

1. Introduction

Nowadays, social media has become an important part of daily life that is difficult to separate. Among the various platforms available, Twitter is one of the most popular among the public. Twitter is a type of microblogging social media that allows users to write down and share their personal activities and views. Through this platform, users can upload photos, as well as express opinions related to various things. Trending topics on Twitter are often the subject of widespread discussion among the public. So, opinions expressed on Twitter can be used as a source of data for research. One of the issues that has been widely discussed on Twitter lately is the ratification of the Indonesian National Army Bill (RUU TNI). This issue attracted public attention because it was considered to have the potential to have a major impact on military involvement in civilian life. Some of the contents of the bill have sparked pros and cons debates, especially because of concerns that the military will once again have a major role in the civilian sphere, which has previously been limited since the reform era. Many people expressed critical opinions, support, and even concern in their tweets, thus making this issue viral and interesting to be studied more deeply[1].

In this case, sentiment analysis on the Twitter platform is important to see how the public responds, opinions, and feelings to the TNI Bill. Because data from Twitter is usually unstructured and often contains distractions such as irrelevant information, text mining techniques are needed in processing such data. Text mining itself is a process to find certain patterns or information from unorganized text data. Its main objective is to unearth trends, patterns, or knowledge hidden in the collection of texts. One of its main applications is to conduct sentiment analysis[2]. Solutions that have been carried out by several previous studies have utilized sentiment analysis using Naive Bayes algorithms. Research by showed that this method was able to classify sentiment with 73% accuracy using 240 trained data and 60 test data. Meanwhile, it got higher results, namely 93.35% accuracy with 100 training data and 150 test data. These results show that Naive Bayes is effectively used to automatically analyze public opinion on social and political issues[3].

This study uses the naïve Bayes method, Naive Bayes is one of the classification techniques that uses a basic probability approach to predict the likelihood of an event. This method is relatively easy to understand, but it is still able to provide accurate results and work quickly. In its application, Naive Bayes assumes that each feature of the analyzed data is independent or does not affect each other[4]. Based on previous research, the Naive Bayes method has been widely used to classify comments and opinions in sentiment analysis, and has proven to be effective in this task. In this study, the Naive Bayes algorithm will be used to analyze a phenomenon that is

still considered taboo, namely suicide, which is rarely discussed openly. This is similar to research from and that raises shooting cases as the object of study[5]. The Naïve Bayes algorithm has proven to be effective in classifying tweets with satisfactory results. The sentiment analysis conducted showed that this algorithm was able to achieve an accuracy level of 85%[6].

However, until now, there is still not much research that specifically examines public sentiment towards the TNI Bill using Orange Data Mining. Therefore, this research specifically aims to collect tweets related to the TNI Bill, carry out the process of cleaning and pre-processing text data, and apply sentiment analysis using Orange Data Mining to classify tweets into positive, negative, and neutral sentiment categories. In addition, this study also aims to examine the distribution and proportion of sentiment that develops on social media and presents it in the form of visualization. The data is collected manually, then processed and analyzed using Orange Data Mining to see and segment sentiment trends in each tweet collected. It is hoped that the results of this research can provide a clear picture of how the public's attitude towards the policies of the TNI Bill on social media.

2. Research Methods

This study uses a descriptive quantitative approach that aims to evaluate and classify the public response to the ratification of the Indonesian National Army Bill (RUU TNI) through the analysis of tweets on social media Twitter.

The Naïve Bayes algorithm was chosen as the main classification method because it is known to have a simple structure but is able to work effectively in grouping text data into positive, negative, and neutral sentiment categories[7].

2.1. Type of Research

This research is descriptive and quantitative in nature. Descriptive research is used to provide a systematic overview of the social phenomena that occur, while the quantitative approach is used to measure and analyze numerical data obtained from text processing.

2.2 Data Sources and Collection Techniques

The research data was obtained from the Twitter platform through the manual collection method (manual crawling). The keywords used to search the data include: "TNI Bill", "revision of the TNI Law", and "ratification of the TNI Bill".

The time span for data collection is strategically determined to cover the period before and after the passage of the law, in order to represent public opinion as a whole[8].

2.3 Stages of Data Analysis

2.3.1 Pre-processing of Text

Before sentiment analysis is carried out, raw data from Twitter must go through a pre-processing process. The steps taken include:

Lowercasing,

Remove irrelevant characters such as URLs, punctuation, symbols, and numbers,

Breaking sentences into words (tokenization),

Stop word removal.

Return the word to its basic form (stemming).

This stage aims to clean up the text so that the classification results become more accurate and relevant.

2.3.2 Classification of Sentiments with Naïve Bayes

After the text cleanup process, the data is classified using the Naive Bayes algorithm. This algorithm works on the principle of probability and assumes that each feature (word) is independent of each other.

With this approach, each tweet is analyzed to determine its probability of falling into one of the sentiment categories. Naive Bayes has been shown to be effective in the classification of texts, including in the context of public opinion on social media.

The training was conducted with labeled data, while testing was conducted on unlabeled data to measure the accuracy of the predictions.

2.3.3 Model Evaluation

To test the performance of the model, the k-fold cross-validation method was used, which is a cross-validation technique that divides the dataset into several parts.

Each section is alternately used as test data, while the other part is used to train the model. The model is evaluated based on evaluation metrics such as:

Accuracy, precision, recall, F1-score.

Previous research noted that the Naïve Bayes algorithm can achieve up to 90% accuracy in the sentiment classification of tweets.

2.4 Research Stages

This stage of research was systematically prepared to achieve the purpose of sentiment analysis on the issue of the ratification of the TNI Bill. The steps taken are as follows:

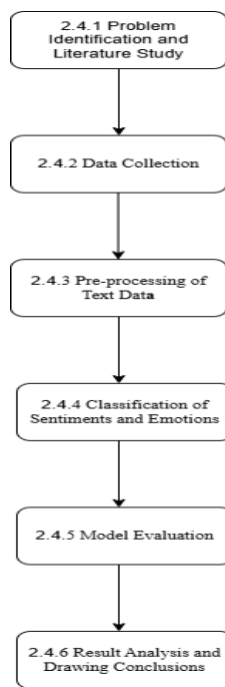


Fig. 1: Research steps

2.4.1 Problem Identification and Literature Study

The initial stage of the research was carried out by identifying the main problems, namely the pros and cons of the public to the TNI Bill. Furthermore, a literature study was conducted from previous research related to sentiment analysis, text mining methods, and Naïve Bayes algorithms.

2.4.2 Data Collection

The data was collected manually from Twitter using the keywords "TNI Bill", "revision of the TNI Law", and "ratification of the TNI Bill". The collection was carried out to cover the period before and after the ratification in order to capture the dynamics of public opinion as a whole. The amount of data collected was 400 tweets, which were then analyzed using the Orange Data Mining application for the purpose of classifying sentiments and emotions.

2.4.3 Pre-processing of Text Data

Raw data goes through text preprocessing stages such as converting letters to a lowercase format, removing punctuation, symbols and numbers, tokenization, removing stopwords, and stemming. This process aims to improve the quality of the data and the accuracy of classification.

2.4.4 Classification of Sentiments and Emotions

The cleaned data is then analyzed using a Naïve Bayes algorithm to classify tweets into positive or negative sentiment categories. In addition, emotion classification was also carried out using the Ekman algorithm through the Tweet Profiler widget in Orange Data Mining[9].

2.4.5 Model Evaluation

To evaluate the model's performance, a k-fold cross-validation method and metrics such as accuracy, precision, recall, and F1-score were used. This evaluation is important to assess the consistency and effectiveness of the model in classification.

2.4.6 Result Analysis and Drawing Conclusions

After the classification is carried out, the results are analyzed to understand public perception of the TNI Bill. From these results, conclusions and recommendations were drawn for further research and data-based decision-making.

3. Results and Discussion

The classification process at each stage of sentiment analysis on the ratification of the TNI Bill is carried out using Orange Data Mining. The design of the sentiment analysis model used can be seen in the image. This model is designed to be used after the data collection (crawling) process is complete.

The purpose of developing this model is so that in the future sentiment analysis of various issues on social media can be carried out more practically and efficiently. In this model, Ekman's algorithm is used through the Tweet Profiler widget to identify emotions from tweet data. In addition, to strengthen the sentiment classification process, the Naïve Bayes algorithm is also applied due to its good ability to group texts based on sentiment polarity automatically.

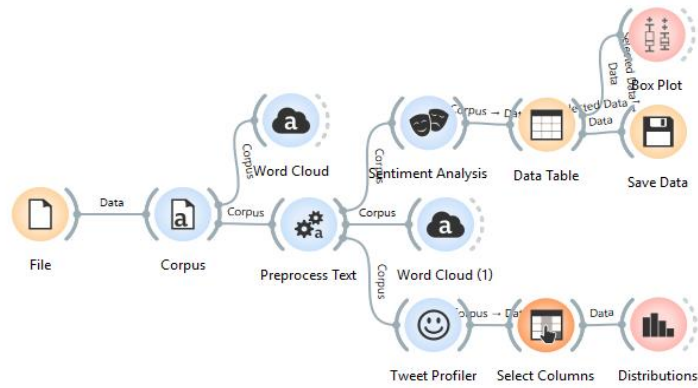


Fig. 2: Sentiment analysis model

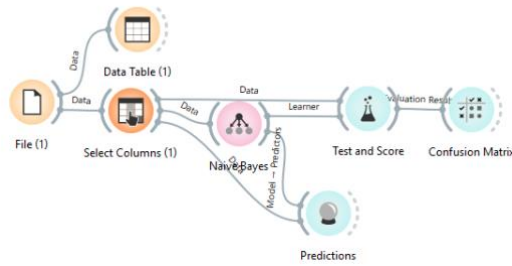


Fig. 3: Testing Flow with Naïve Bayes Algorithm

3.1 Crawling Data

The 400 data was collected manually from Twitter using the keywords "TNI Bill", "revision of the TNI Law", and "ratification of the TNI Bill". The collection was carried out to cover the period before and after the ratification in order to capture the dynamics of public opinion as a whole.

3.2 Preprocessing

The data obtained is then processed using Preprocess Text in Orange data mining whose stages include *i) Transformation, ii) Tokenization, iii) Normalization, and iv) Filtering*. This approach is in line with previous research that stated that text cleanup is a critical step in sentiment analysis. Furthermore, the widget used, namely Word Cloud, functions to provide word visualization before and after the preprocessing process. Figure 4. Showing pre-processing flows using orange data mining.

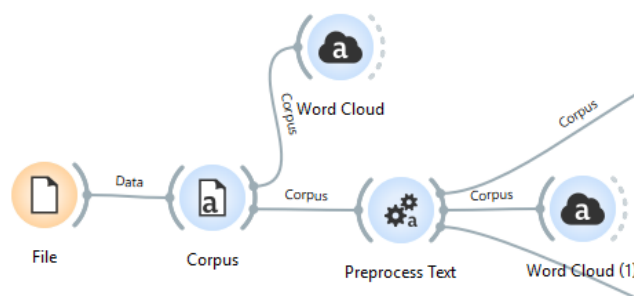


Fig. 4: Preprocessing pipeline

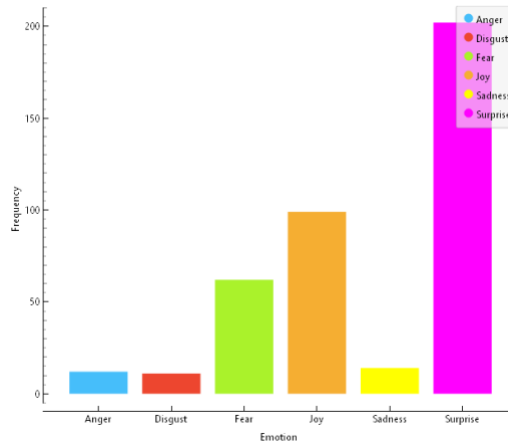


Fig. 8: Results of emotional analysis

Based on the results of the calculations shown in figure 8, the number of emotions detected in the tweet is as follows: anger = 12, insult = 11, fear = 62, happiness = 99, sadness = 14, and surprise = 202. From this data, it can be concluded that the majority of twitter users were surprised by the passage of this TNI Bill.

3.4 Polarity Classification

The classification of sentiment polarity was carried out using *the Multilingual sentiment technique*. Orange data mining provides this tool through a *sentiment analysis widget*. This stage of analysis involves several widgets, including *file widgets, corpus, preprocess text, sentiment analysis, data tables, select columns, box plots, and save data*.

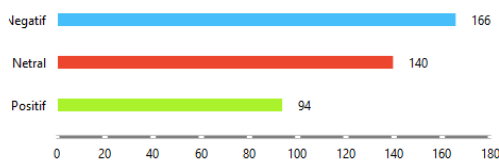


Fig. 9: Sentiment analysis results

Based on Figure 9, sentiment analysis of 400 tweets discussing the ratification of the TNI Bill shows that the majority of responses are negative with a total of 166, neutral sentiment of 140, and positive sentiment of 94. If the percentage of polarity analysis is calculated, the results are as large, negative sentiment = 41.5%, neutral = 35%, and positive = 23.5%

3.5 Testing of the Naïve Bayes Algorithm

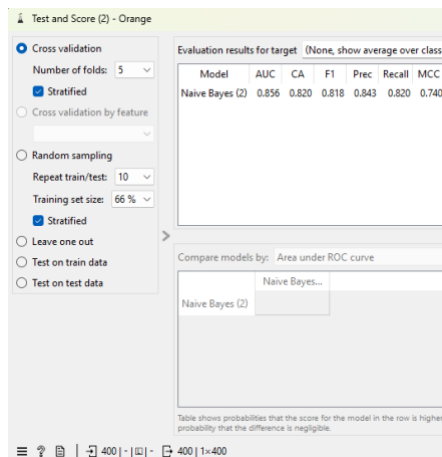


Fig. 10: Results of testing the naïve Bayes model

Based on Figure 10, the Naive Bayes model shows quite good performance, shown by an Area Under Curve (AUC) value of 0.856 and an accuracy (CA) of 82.0%. The F1-Score obtained of 0.818 indicates a balance between precision and recall values. A precision of 0.843

indicates that most of the positive predictions the model produces are correct, while a recall of 0.820 reflects the model's ability to detect positive cases. On the other hand, the Matthews Correlation Coefficient (MCC) value of 0.740 indicates a strong correlation between the prediction results and the actual data.

Overall, these results indicate that the Naive Bayes algorithm is able to classify the data quite well. A high AUC value indicates the model's ability to distinguish between classes, while a sufficiently high F1 and MCC values indicate the consistency of the model's performance in the classification process.

To measure the performance of the model in more detail, a confusion matrix was used which showed the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) created by the model. Figure 11 shows the results of the confusion matrix from the naïve Bayes model.

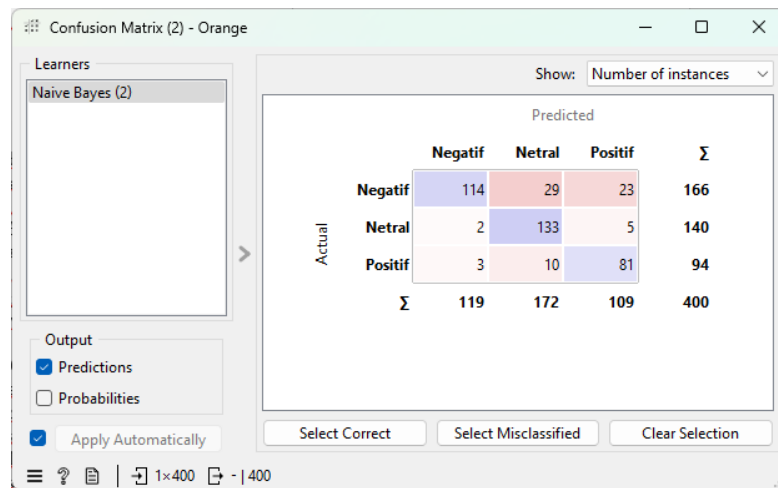


Fig. 11: Confusion Matrix

Based on the confusion matrix in Figure 11, it can be seen that the model is able to identify most of the data accurately. Of the total 166 data labeled Negative, 114 were correctly predicted, while 29 were classified as Neutral and 23 as Positive. For the Neutral class, out of 140 actual data, 133 of them were correctly classified, while 2 were misclassified as Negative and 5 as Positive. In the Positive class, out of 94 data, the model managed to correctly predict 81, while 10 were classified as Neutral and 3 as Negative.

In general, the model shows quite good classification performance, especially in the Neutral class, which obtains the highest number of correct classifications. However, there are still mistakes in distinguishing certain classes, especially in Negative data which is often classified as Neutral. This indicates that despite the high overall accuracy, the model still has the potential to be improved in terms of separation between similar classes.

4. Conclusion

This research proves that sentiment analysis on the issue of the ratification of the TNI Bill can be done well using the Naïve Bayes algorithm through the Orange Data Mining application. Of the 400 tweets collected manually, most of the public showed a negative response to this bill, namely 166 tweets (41.5%), neutral as many as 140 tweets (35%), and 94 tweets (23.5%) showing positive opinions. In addition, the emotional analysis of the tweets showed that many people felt surprised, followed by excitement and fear. This shows that the ratification of the TNI Bill caused a considerable emotional reaction from the public.

The Naïve Bayes model used succeeded in classifying the data with fairly high accuracy, which was 82%. Although there is still an imbalance between the number of positive, negative, and neutral tweets, the results are still quite good. This can also be seen from other metrics such as F1-score and AUC which show that the model's performance is quite stable. Overall, this method has proven to be effective and practical for finding out people's opinions on social media on important issues. Orange Data Mining has also been proven to help the analysis process become easier and more efficient. This research can be an example or basis for analyzing other issues that are widely discussed on the internet.

References

- [1] A. Liken Anggoro, L. V. P. Ken, and M. G. Setiawan, "Analisis media text clustering pada twitter akan kasus selebriti menggunakan orange data mining," *remik*, vol. 7, no. 1, pp. 189–195, Jan. 2023, doi: 10.33395/remik.v7i1.12001.
- [2] Ade Tiara Susilawati, Nur Anjeni Lestari, and Puput Alpria Nina, "Analisis sentimen publik pada twitter terhadap boikot produk israel menggunakan metode Naïve Bayes," *Nian Tana Sikka: Jurnal ilmiah Mahasiswa*, vol. 2, no. 1, pp. 26–35, Dec. 2023, doi: 10.59603/niantanasikka.v2i1.240.
- [3] Y. Mayasari, R. Nasution, and N. Sumatera Utara, "Post-Election Sentiment Analysis 2024 via Twitter (X) Using the Naïve Bayes Classifier Algorithm," *Journal of Dinda Data Science, Information Technology, and Data Analytics*, vol. 4, no. 2, pp. 123–134, 2024, [Online]. Available: <http://journal.itelkom-pwt.ac.id/index.php/dinda>

- [4] Woro Isti Rahayu, "PERBANDINGAN ALGORITMA K-MEANS DAN NAÏVE BAYES UNTUK MEMPREDIKSI PRIORITAS PEMBAYARAN TAGIHAN RUMAH SAKIT BERDASARKAN TINGKAT KEPENTINGAN PADA PT. PERTAMINA (PERSERO)," *Jurnal Teknik Informatika*, Vol. 13, No. 2, April 2021, vol. 13, no. PERBANDINGAN ALGORITMA K-MEANS DAN NAÏVE BAYES UNTUK MEMPREDIKSI PRIORITAS PEMBAYARAN TAGIHAN RUMAH SAKIT BERDASARKAN TINGKAT KEPENTINGAN PADA PT. PERTAMINA (PERSERO), May 2021.
- [5] Fenilinas Adi Artanto, "Analisis Sentimen Opini Publik terhadap Fenomena Bunuh Diri Mahasiswa di Twitter Menggunakan Algoritma Naive Bayes," *SATESI: Jurnal Sains Teknologi dan Sistem Informasi*, vol. 4, no. 1, pp. 70–77, Apr. 2024, doi: 10.54259/satesi.v4i1.2908.
- [6] R. Harun, R. Ishak, and S. S. Panna, "Analisis Sentimen Opini Publik Pengguna Twitter Terhadap Kenaikan Harga BBM Menggunakan Algoritma Naive Bayes," *Copyright @BALOK*, vol. 2, no. 1, 2023.
- [7] J. Homepage, N. C. Agustina, D. Herlina Citra, W. Purnama, C. Nisa, and A. Rozi Kurnia, "MALCOM: Indonesian Journal of Machine Learning and Computer Science The Implementation of Naive Bayes Algorithm for Sentiment Analysis of Shopee Reviews on Google Play Store Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store," vol. 2, pp. 47–54, 2022.
- [8] A. F. Setyaningsih, D. Septiyani, and S. R. Widiyarsi, "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Masyarakat pada Twitter mengenai Kepopuleran Produk Skincare di Indonesia," *Jurnal Teknologi Informatika dan Komputer*, vol. 9, no. 1, pp. 224–235, May 2023, doi: 10.37012/jtik.v9i1.1409.
- [9] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," vol. 15, no. 1.