Analyzing Twitter Users' Sentiments on the Surge of Fuel Oil Prices in Indonesia using the K-Nearest Neighbor Algorithm

Rizqi Putri Nourma Budianti¹, Hanifah Citra Marvyna¹, Sritrusta Sukaridhoto²

¹Department of Information System, Universitas Nahdlatul Ulama Surabaya, Indonesia
²Department of Creative Multimedia Technology, Politeknik Elektronika Negeri Surabaya

Abstract. Sentiment analysis offers an effective solution for automating the classification of text data based on polarity, facilitating the assessment of public opinion. Among various social media platforms, Twitter stands out as a significant source of concise textual data reflecting users' viewpoints on diverse topics. Notably, the recent surge in the price of fuel oil (BBM) in Indonesia has sparked considerable discussion and expression on Twitter. In this study, our objective was to perform a comprehensive sentiment analysis of Twitter users' reactions to the rising fuel prices in Indonesia by employing the K-Nearest Neighbor (K-NN) algorithm. The research followed a structured approach encompassing data collection, text preprocessing, data labeling, feature extraction, data splitting, classification, and algorithm performance evaluation. The results revealed a dominance of negative sentiments among the 5,000 collected tweet data. The sentiments were categorized as 54.6% negative, 31.8% positive, and 13.6% neutral. This indicates a prevailing level of dissatisfaction and concern expressed by Twitter users regarding the fuel price increase. The K-NN algorithm's classification performance was most promising when evaluated in an 80:20 data ratio experiment, yielding an accuracy rate of 65%, precision of 74%, recall of 45%, and an error rate of 35%. These findings suggest that the K-NN algorithm is valuable for effectively gauging public sentiment towards the escalating fuel prices in Indonesia. This research highlights the potential of sentiment analysis and the K-NN algorithm in assessing public reactions to significant events, providing valuable insights for policymakers and stakeholders in the energy sector.

1 Introduction

Sentiment analysis is a field of study that focuses on analyzing opinions, sentiments, evaluations, and one's emotions regarding a topic such as services, products, events, even policies [1]. The aim of sentiment analysis is to extract opinions expressed in text data so that these opinions are easily understood by analyzing large sets of text data sources, such as blogs, review sites and social media [2]. The amount of text data that is collected will usually be difficult to classify manually by humans, therefore sentiment analysis can be a solution to

* Corresponding author: rizqi.putri.nb@unusa.ac.id
make it easier to classify data polarity by shortening time because it is done automatically by the system [3].

Most of the data in the form of text is obtained through social media [4]. According to data gathered by the Indonesian Internet Service Providers Association (APJII), it's projected that in 2021, Indonesia will have a total of 210 million people using the Internet. Out of these, 170 million are expected to be actively engaging on social media platforms like Facebook, Instagram, TikTok, and Twitter [5,6]. Twitter is one of the social media that presents text data in a concise manner, with a maximum of 280 characters in each uploaded content or what is known as a tweet [7]. When compared to the other three social media which focus on long text content, photos and videos, Twitter's data is simpler and has meaning in conveying an opinion [8].

Topics discussed on Twitter vary widely, ranging from hobbies to politics [9]. Recently, there has been significant attention given to the rise in fuel oil (BBM) prices in September 2022. This hike in prices has been officially confirmed via the Minister of Energy and Mineral Resources' Decree, specifically Decree Number 218.K/MG.01/MEM.M/2022. This decree outlines the new retail selling prices for various types of fuel oil and special types of fuel oil under assignment. The increase in fuel prices includes Peralite, subsidized Solar and non-subsidized Pertamax. When the decision was announced, the Indonesian people voiced a lot of opinions both pro, con, and neutral so that the topic became a trending topic with a total of 38.6 thousand tweets with the keyword "BBM Rise" as reported from the getdaytrends.com site.

Research related to the analysis of public sentiment towards the previous increase in fuel prices has been carried out. One of the studies was conducted by [10] by looking at the opinion of the Twitter community regarding Wage Subsidy Assistance (BSU) on rising fuel prices with the Naïve Bayes algorithm. The results of this study indicate that there is positive sentiment referring to the dataset with the BSU keyword of 65.2%, and negative sentiment in the dataset with the BBM keyword of 71.8% of the classification results. From the evaluation results of the system that has been built, the system can classify public opinion into positive and negative sentiments automatically. In conducting sentiment analysis an algorithm is needed to classify text [11].

The amount of Twitter user opinion tweet data on the topic of rising fuel prices in Indonesia, will make it difficult to classify sentiments if done manually [3]. Hence, this study aims to examine the sentiments of Twitter users regarding the increasing fuel prices in Indonesia by employing the K-Nearest Neighbor (K-NN) algorithm. The objective is to categorize the views of Twitter users on this price hike into three classifications: positive, negative, or neutral, utilizing the K-NN algorithm. Additionally, the study will assess the effectiveness of this algorithm in sentiment classification. This research is expected to provide knowledge about the optimal algorithm for classifying sentiment and at the same time knowing its performance based on the results of its accuracy, precision, recall, and error rate.

2 Research Methodology

2.1 Materials

Reference that related to the topic and discussion about sentiment analysis of Twitter users using K-NN been done a lot before, such as research that written previously about sentiment analysis of online learning on Twitter during a COVID-19 pandemic using Naïve Bayes shows that the topic of online learning during the pandemic classified in 2 sentiments each 30% is a positive sentiment and 69% is a negative sentiment [12].
Next, the research about sentiment analysis of public comment to Tri Indonesia in Twitter classify the data into 2 sentiments, that divided to positive and negative sentiments from 237 comments with 84.28% accuracy [13]. Related research that studied about the application Naïve Bayes algorithm for analysis sentiment of review BMKG National’s Twitter data, from this study showed that the accuracy is 69.97% and classify the opinion into 3 sentiments there are positive, negative, and neutral [14].

Then, related research that mapped public opinion regarding the impact of the Corona virus, entitled "Analysis of Twitter Document Sentiment Regarding the Impact of Corona Virus Using the Naïve Bayes Classifier Method", this study shows that the classification using the Naïve Bayes method produced stable performance which can be proven by the accuracy of the Tweet documents of 67% with an error rate of 33% [15]. The research focusing on text-based sentiment analysis on Twitter using the Naïve Bayes Classifier and Confusion Matrix involves dividing the data into five sets for training and three sets for testing. Sentiment analysis is conducted using the Naïve Bayes Classifier (NBC) method, and the confusion matrix technique is employed to calculate performance. The findings of this study show that combining NBC test results with performance testing using a confusion matrix yields an accuracy rate of 82%, a precision rate of 93%, and a recall rate of 52% [16].

The research about twitter sentiment analysis on worldwide COVID-19 outbreaks, this research was conducted because of the background of the large number of opinions and news about COVID-19. So, the results of this sentiment analysis research show that the average accuracy of the results of dataset testing on opinions about the Covid-19 pandemic is 60% [17,18].

Comparable research involves sentiment analysis using the Naïve Bayes algorithm on data collected from Twitter. This particular study centers on analyzing the sentiments on Twitter regarding the 2019 Presidential candidates of the Republic of Indonesia, employing Python as the programming language. The methodology of this research includes several phases: gathering data via the Python Library, processing text data, testing training data, and classifying the data using the Naïve Bayes approach. With results of the accuracy of each Jokowi - Ma’aruf Amin candidate is 80.90% and Prabowo - Sandiaga candidate is 80.1% [19].

Research on using Orange Data Mining for classifying student graduations with the K-Nearest Neighbor Model, Decision Tree, and Naïve Bayes reveals that the Naïve Bayes method achieves an accuracy of 89%. In comparison, the K-Nearest Neighbor (K-NN) method shows 77% accuracy, and the Decision Tree method has an accuracy of 74%. So, in conclusion the Naïve Bayes method is highly recommended for classifying graduation rate data for students of the Informatics Engineering Study Program, Madura Islamic University [20].

The study about analysis of public opinion sentiment of films on the Twitter platform using the Naïve Bayes algorithm, its results obtained are an accuracy value of 0.65%, a precision value of 0.67% and a recall value of 0.65% and a neutral percentage value of 0.83% [21].

From the research about sentiment analysis of public opinion regarding COVID-19 on Twitter, this research uses 2 different algorithms, namely the Naïve Bayes algorithm and the KNN algorithm to classify 1098 tweet datasets taken by the keyword “COVID-19”. From this study, the classification results were obtained on 2 variables, namely positive and negative. With the classification of the Naïve Bayes algorithm, an accuracy value of 63.21% is obtained, while the KNN algorithm produces an accuracy value of 58.10%. With a total of 610 positive opinions and 488 data of negative opinions. From these results it can be concluded that public opinion towards Covid-19 is positive [22].
2.2 Data Collection Procedures

The process of data collection for this study involves data crawling using a relevant hashtag related to the topic of the increase in fuel oil prices in Indonesia, specifically the hashtag "BBM Naik." The researchers obtained this keyword by referring to Twitter trending statistical data available on the website https://getdaytrends.com/indonesia. The tweet data collected was specifically for the month of September 2022.

2.3 Data Analysis

In the Data Analysis phase of the study, the researchers aimed to analyze Twitter users' sentiments on the surge of fuel oil prices in Indonesia using the K-Nearest Neighbor (K-NN) algorithm. The process of data collection involved data crawling, which was done by utilizing a relevant hashtag related to the topic, specifically the hashtag "BBM Naik." This hashtag was chosen as it represents the discussions and opinions surrounding the increase in fuel oil prices in Indonesia.

To identify the appropriate hashtag, the researchers referred to Twitter trending statistical data from the website https://getdaytrends.com/indonesia. This website provided valuable insights into the trending topics and hashtags in Indonesia, allowing the researchers to select the most relevant hashtag for their study.

The tweet data that was collected through data crawling was specifically obtained for the month of September 2022. The researchers focused on this specific time frame to capture a substantial amount of relevant tweets discussing the surge in fuel oil prices during that period. In this study, there is several stages that performed to process the research with 9 methodologies, i.e., literature study, data crawling, text preprocessing, etc. For detail stages can be seen in Figure 1.

![Research Methodology](image)

Fig. 1. Research Methodology.
2.4 Research Methodology

Literature study is carried out by understanding the research problem, looking for basic theoretical and related methods as reference material to solve the same problem.

In collecting the data set of tweets, the process of data crawling is using related hashtag to the topic about the increase of fuel oil’s prices in Indonesia such as "BBM Naik". The keyword was obtained from referring to Twitter trending statistical data on the site https://getdaytrends.com/indonesia. The tweet data taken is tweet data for September 2022. The amount of data taken refers to previously research [23] which is 1,500 data because with this amount of data and the distribution of training and testing data 80:20 in his research resulted a high accuracy value.

At the text preprocessing stage, it is carried out to prepare documents in the form of text data that is less structured so that later it produces clean, structured data output that is ready to be processed for classification with the algorithm used. In text data generated from social media, text preprocessing will usually be carried out with several processes including data cleaning, case folding, tokenizing, filtering, stemming, removing data duplicate [10].

1. Data cleaning is performed to removing unnecessary components in the data, remove unnecessary characters, such as removing punctuation, numbers, URLs, mentions, hashtags, etc. And performing Case folding to lowercase the letters in a sentence [24,25].
2. Tokenizing is performed to split a complete sentence into several forms of word tokens [26].
3. Filtering / Stop word removal is performed to removing words, letters that are too many or too few do not affect sentiment [27].
4. Stemming is performed to reducing a word to its base word by removing prefixes and suffixes [28].
5. Removing data duplicate is stages of deleting data that is not needed because it has the same value as other data rows [29].

The process of data labelling is dividing the data into classes include positive, negative, and neutral classes [30]. With data labelling, we can see the polarity of the tweet text data. At this stage, automatic data labeling is performed using the Lexicon-based method with Indonesian language dictionary sentiment from GitHub https://github.com/fajri91/InSet. Then the determination of the label is divided into 3 classes, if the polarity score is > 0 then it is positive, if < 0 then it is negative and if = 0 then it is neutral [31].

Feature extraction is a method used to identify an object in the data based on the histogram of the object [32]. One method in the feature extraction process is TF-IDF. TF-IDF, which
stands for term frequency-inverse document frequency, is an algorithm for weighting the term index resulting from the text preprocessing process [33].

\[ W = TF \times IDF \]  

(1)

Description:
- **W**: IDF weight of each searched word
- **TF**: Total data
- **IDF**: The number of times the word appears in the data

\[ IDF = \log \frac{N}{n_t} \]  

(2)

Description:
- **IDF**: IDF value of each searched word
- **N**: Total data
- **n_t**: The number of times the word appears in the data

Prior to the classification process, clean data will be used as input, and divided into training data and testing data. Training data is used to train the specified algorithm. While data testing is used as data that will be used to test the results of the training that has been carried out. The division is with a ratio of 80:20, 70:30, and 60:40 for each training data and testing data [34].

The purpose of this stage is to classify a document which is divided into certain categories based on the words in the document. Then the group of words will be assessed and will produce output classification results which have certain meanings such as positive, negative, and neutral [21]. K-NN is an algorithm used as a classifier that uses training data with the closest distance to the object [35]. The K-NN classification theorem is as follows:

\[ d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \]  

(3)

Description:
- **D**: distance between x and y
- **X**: cluster datacenter
- **Y**: Target data
- **i**: Each data
- **n**: Total data
- **x_i**: Data cluster at-i
- **y_i**: Data in each at-i

The algorithm performance evaluation stage is carried out to validate the correctness of the classification results in the previous stage [36]. At this stage using the cross-validation method which will be measured by the confusion matrix to know the results of accuracy, precision, recall, and error rate.
Table 1. Cross validation with confusion matrix.

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>Predicted Class</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Neutral</td>
<td>Negative</td>
</tr>
<tr>
<td>Positive</td>
<td>TP</td>
<td>FNt</td>
<td>FN</td>
</tr>
<tr>
<td>Neutral</td>
<td>FP</td>
<td>TNg</td>
<td>FN</td>
</tr>
<tr>
<td>Negative</td>
<td>FP</td>
<td>FNt</td>
<td>TN</td>
</tr>
</tbody>
</table>

And the formula for accuracy, precision, recall, and error rate is as follows:

\[
\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{4}
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{5}
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{6}
\]

\[
\text{Error Rate} = \frac{FP + FN}{TP + FN + FP + TN} \tag{7}
\]

3 Result and discussion

In this chapter, a detailed presentation and explanation of the research work process will be carried out from beginning to end.

3.1 Data crawling

Data collection in this study was carried out by collecting tweet set data from Twitter with the help of the Python library, namely Snscrape. Snscrape is commonly used for data mining on social media, including Twitter with certain keyword filters [37]. The keyword used in this study was "BBM naik" and the number of tweets collected was 5,000 data.

3.2 Data cleansing

At this stage, data cleansing is carried out by removing attributes that are not needed for the classification process. The results of the cleaning data are found in Table 2.

Table 2. Data cleaning result.

<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Police #Patroli # fuelgoesup 'n' 'n As the fuel prices increased as of August 3, 2022, the Aceh Regional Police along with all its units have started patrolling and providing security at each location SPBU 'n' 'n' <a href="https://t.co/lC1hYXn2wr">https://t.co/lC1hYXn2wr</a> lewat @serambinews</td>
<td>As fuel prices rose last August, the Aceh Regional Police along with their entire team began patrolling and guarding every gas station</td>
</tr>
</tbody>
</table>

3.3 Tokenizing

Tokenizing is done to convert a whole sentence into multiple word tokens. The results of the cleaning data are found in Table 3.
With the increase in fuel prices since last August, the Aceh Regional Police along with their entire ranks have started conducting patrols and guarding every gas station.

3.4 Filtering / stop word removal

At the filtering stage, word deletion is carried out that has no effect on sentiment. The filtering results are found in Table 4.

<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>[along, rising, price, fuel, per, August, Yesterday, regional police, aceh, along with, whole, ranks, start, do, patrol, And, guard, in, every, gas station, past]</td>
<td>[along, rising, price, fuel, August, Yesterday, regional police, aceh, along with, ranks, patrol, guard, gas station]</td>
</tr>
</tbody>
</table>

3.5 Stemming

Stemming is the stage of reducing a word into its base word by removing suffix and prefix affixes. The filtering results are found in Table 5.

<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>[accompanied, increased, price, fuel, August, yesterday, regional police, Aceh, as well as, lined up, patrolled, guarded, gas stations]</td>
<td>[accompanied, increased, price, fuel, August, yesterday, regional police, Aceh, as well as, lined up, patrolled, guarded, gas stations]</td>
</tr>
</tbody>
</table>

3.6 Deletion of duplicate data

Then the deletion of duplicate data in the data is also carried out at this stage, the deletion uses the help of the Python library, namely Pandas. So that the tweet data that originally amounted to 5,000 tweets now has 4,857 tweets left.

3.7 Data labelling

After going through several stages above, the output of these stages is clean data. This clean data will then be automatically labeled with the Lexicon-based feature into 3 different types of labels, namely, positive, negative, or neutral. At this stage, data labeling is carried out automatically with the help of the Lexicon-based method. How it works by using the help of an Indonesian dictionary or lexicon that contains words that contain positive and negative sentiments. The total score (compound) of a sentence is added and will later produce a polarity score. Words that are not in the dictionary, will be given a score of 0. For more details regarding the score of each word in the InSet Lexicon dictionary can be seen in...
Furthermore, the determination of the label based on its polarity value can be initialized as follows, if the polarity value is > 0 then it is positive, if it is < 0 then it is negative, and if it is = 0 then it is neutral [38]. The results of the labelling data are found in Table 6.

Table 6. Labelling Result.

<table>
<thead>
<tr>
<th>Labelling</th>
<th>Polarity Score</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>[accompanied, increased, price, fuel, August, yesterday, regional police, Aceh, as well as, lined up, patrolled, guarded, gas stations]</td>
<td>7</td>
<td>positive</td>
</tr>
</tbody>
</table>

### 3.8 Feature extraction

The weighting of the words in this study data was carried out using the TF-IDF. Weighting is done to change the shape of the text data to numeric. Word weighting calculations are done automatically using the Python library Scikit-learn with the TfidfVectorizer module. For the TF-IDF source code can be seen at Fig 2.

![TF-IDF Results](Fig. 2).

### 3.9 Data splitting

The data used in this study was then divided into training data and testing data. Data sharing is done automatically, using the Scikit-learn library. With a comparison ratio of training and testing data of 80:20, 70:30, and 60:40. The results of data splitting can be seen in Table 7.

Table 7. Data Splitting Result.

<table>
<thead>
<tr>
<th>Data Ratio</th>
<th>Training Data</th>
<th>Testing Data</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>80:20</td>
<td>3.885</td>
<td>972</td>
<td>4.857</td>
</tr>
<tr>
<td>70:30</td>
<td>3.399</td>
<td>1.458</td>
<td>4.857</td>
</tr>
<tr>
<td>60:40</td>
<td>2.914</td>
<td>1.943</td>
<td>4.857</td>
</tr>
</tbody>
</table>

### 3.10 Classification using K-NN

At the classification stage, a type of experimental research method will be carried out with the division of experiments to carry out the process of classifying existing data. In addition, using the the K-NN algorithm as a classification method.

### 3.11 Evaluate algorithm performance with Confusion Matrix

After the data classification process is complete, the researcher will then conduct an evaluation. The algorithm performance evaluation stage is carried out to validate the correctness of the classification
results in the previous stage. This stage is carried out by calculating the evaluation with the cross-validation method with a confusion matrix of each experiment to find out the results of accuracy, precision, recall, and error rate.

3.12 Discussion

It can be seen from the classification results of K-NN algorithm, that the best classification results were obtained by using the K-NN algorithm with a data ratio of 80:20 which resulted in an accuracy value of 22%. And from the classification results, it resulted in segmentation of each class sentiment including negative at 54.6%, positive at 31.7%, and neutral at 13.7%. Segmentation of classification results can also be seen on the pie chart Fig. 3.

Fig. 2. TF-IDF Results.

Fig. 3. Pie chart of classification results.

The following are the results of the evaluation on the Experiment using the K-NN algorithms can be seen in Table 8.

<table>
<thead>
<tr>
<th>Percobaan</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-NN 80:20</td>
<td>22%</td>
<td>70%</td>
<td>14%</td>
<td>78%</td>
</tr>
<tr>
<td>Negatif</td>
<td>14%</td>
<td>89%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netral</td>
<td>70%</td>
<td>7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positif</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K-NN 70:30</td>
<td>21%</td>
<td>75%</td>
<td>10%</td>
<td>79%</td>
</tr>
<tr>
<td>Negatif</td>
<td>15%</td>
<td>93%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netral</td>
<td>77%</td>
<td>4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positif</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K-NN 60:40</td>
<td>21%</td>
<td>74%</td>
<td>11%</td>
<td>79%</td>
</tr>
<tr>
<td>Negatif</td>
<td>15%</td>
<td>94%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netral</td>
<td>90%</td>
<td>4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positif</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 8 and Figure 4 in the study indicate that among various experiments using the K-Nearest Neighbor (K-NN) algorithms, the one with an 80:20 data ratio showcased the most efficient performance. This specific experiment yielded a 22% accuracy rate, coupled with commendable precision, recall values, and a minimal error rate. The detailed performance evaluation results, as depicted in Table 8 and the bar chart in Figure 4, reveal that this particular K-NN algorithm experiment, with the 80:20 data ratio, not only achieved an accuracy of 22% but also produced precision results of 70% for negative, 14% for neutral, and 70% for positive sentiments. The recall results were 14% for negative, 89% for neutral, and 7% for positive sentiments, with an overall error rate of 78%.

4 Conclusion

The research analyzing Twitter user sentiments about the rising fuel prices in Indonesia utilized two algorithms for classification and evaluation: the Naïve Bayes algorithm and the K-Nearest Neighbor (K-NN) algorithm. For this analysis, a dataset comprising 5,000 data entries was employed.

When applying the Naïve Bayes algorithm, the sentiment distribution was as follows: negative sentiment accounted for 54.6%, positive sentiment for 31.8%, and neutral sentiment for 13.8%. However, the results from the data classification process directly using the Naïve Bayes algorithm with a data ratio of 80:20 revealed a relatively low accuracy value of 22%, a precision result of 51.3%, a recall result of 36.7%, and an error rate of 78%.

Conversely, the K-Nearest Neighbor algorithm was employed to examine Twitter user sentiments regarding the rise in fuel prices in Indonesia. However, specific details about the performance metrics of the K-NN algorithm, such as accuracy, precision, recall, and error rate, were not provided in this context. It would be beneficial to include a comprehensive evaluation of the K-NN algorithm's performance to complement the analysis. In conclusion, the sentiment classification analysis showed that Twitter users' responses to the increase in fuel prices in Indonesia were predominantly negative. The relatively low accuracy and precision values from the Naïve Bayes algorithm indicate that the classification results might require further optimization and fine-tuning to achieve more accurate sentiment predictions. It is important to explore other machine learning algorithms and fine-tune the model parameters to improve the performance of the sentiment analysis system. Additionally, providing a more detailed evaluation of the K-NN algorithm would enhance the overall analysis and strengthen the research findings.
References

3. Y. Harfian, Klasifikasi Sentimen Aplikasi Dompet Digital DANA Pada Komentar Di Instagram Menggunakan Metode Naive Bayes Classifier, UIN Sultan Syarif Kasir Riau, 2021
5. S. Kemp, Datareportal (2021)
21. Y. Nurtikasari, Syariful Alam, and Teguh Iman Hermanto, INSOLOGI J. Sains Dan Teknol. 1, 411 (2022)
22. M. Syarifuddin, INTI Nusa Mandiri 15, 23 (2020)
25. N. Hikmah, J. Antart. 08, (2018)
35. T. Wiratama Putra, A. Triayudi, and A. Andrianingsih, J. JTIK (Jurnal Teknol. Inf. Dan Komunikasi) 6, 20 (2022)
37. R. A. Firmansah, R. E. Indrajir, and E. Dazki, SATIN - Sains Dan Teknol. Inf. 8, 01 (2022)