CYBERBULLYING SENTIMENT ANALYSIS OF INSTAGRAM COMMENTS USING NAÏVE BAYES CLASSIFIER AND K-NEAREST NEIGHBOR ALGORITHM METHODS

Fitri Anisa Nirmala*1, Muhammad Jazman2, Nesdi Evrilyan Rozanda3, Febi Nur Salisah4

1,2,3,4Information Systems, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
Email: 12050322148@student.uin-suska.ac.id, 2jazman@uin-suska.ac.id, 3nesdi.er@uin-suska.ac.id, 4febinursalisah@uin-suska.ac.id

Abstract

The high number of social media users presents major threats and risks, such as cyberbullying Cyberbullying or cyberbullying is one of the negative impacts of the rapid development of technology and social media. Sentiment Analysis is a technique for extracting text data to obtain information about positive, neutral or negative sentiment. One of Indonesian social media that often gets user sentiment through social media is Instagram. By using the Text Mining technique, the classification method will determine whether a sentiment is positive, neutral or negative. This research uses the Naïve Bayes Classifier (NBC) and K-Nearest Neighbor (KNN) methods with tf-idf weighting accompanied by the addition of an emotional icon (emoticon) conversion feature to determine the existing sentiment classes from tweets about Instagram users. The results of calculations using the first three methods using the Partitionong model, the results using the Naive Bayes method, get an accuracy value of 91.25%, a recall value of 92% and a precision value of 90% and calculations using the KNN method have an accuracy value of 67%, a recall value of 49% and a precision value of 34 %. So it can be concluded that the Naïve Bayes Classifier algorithm has the best performance.

Keywords: *Cyberbullying K-Nearest Neighbor, Naïve Bayes Classifier, Sentiment Analysis.*

1. INTRODUCTION

Progress in communication and information technology is very rapid and plays an important role in human life in fulfilling needs that must be met. Technological advances that utilize the internet have an impact on human life patterns, such as how to interact with others, exchange information, and search for information that is currently being widely discussed by the public. With the limitations of conventional media such as radio, newspapers, TV. To express an opinion. Not closing social media is an appropriate substitute or alternative for expression [1]. The use of social media is g Data, social media users are estimated to number rowing rapidly in Indonesia. According to Reportal 167 million in 2023. There are 153 million users over 18 years old, or 79.5% of the total population. Additionally, it is estimated that 78.5% of Internet users use at least one social media account. This value is predicted to continue to increase in the coming years. The existence of this technology makes it easy to interact and disseminate information to the general public. However, the benefits obtained from using information technology are often used by people for negative things, such as posting insulting words, hate speech, and spreading hoax news that can harm the party concerned [2].

The high number of internet users presents major threats and risks, such as cyberbullying. *cyberbullying* is the behavior of uploading or sending text or images that are rude and harmful using digital media or the internet [3]. According to the secretary general of APJII, Henri Kasyfi, almost half of the population of internet users in Indonesia have been victims of *cyberbullying*. This figure was obtained from a survey conducted among internet users from March to April 14 2023. The *results* obtained from the survey which had 5,900 samples stated that 49 percent admitted that they had experienced bullying on social media [4]. According to the secretary general of APJII, Henri Kasyfi, almost half of the population of internet users in Indonesia have been victims of *cyberbullying*. This figure was obtained from a survey conducted among internet users from March to April 14 2023. The *results* obtained from the survey which had 5,900 samples stated that 49 percent admitted that they had experienced bullying on social media. The basic understanding of bullying involves the intimidator and the victim being bullied both physically and verbally, which can be done directly or indirectly[5].

Cyberbullying is often carried out on social media platforms. Instagram is a social media application that is accessed by almost all groups. Through Instagram, someone can upload photos or

videos, publish them, the friendship system on Instagram uses the terms follower (*follower*) and following *(people who are followed)* [6]. There are like and comment features so that people can give appreciation in the form of likes or comments on uploaded photos. According to the We Are Social report, the number of global Instagram users will reach 1.6 billion in July 2023. The number of Instagram users in Indonesia is the 4th largest in the world, with around 103.3 million users. Meanwhile, India still occupies the top position with 332.15 million Instagram users in July 2023. The emergence of *cyberbullying* behavior is because Instagram provides comment facilities for its users on someone's video content so that other users can easily make hate speech or hate comments [7].

To reduce the occurrence of *cyberbullying* cases, bullying comments on Instagram must be classified to determine whether they fall into the bullying or non-bullying category. *Naïve Bayes Classifier* (NBC) and *K-Nearest Neighbor* (KNN) are two popular classification methods[8]. This research uses the Naïve Bayes *Classifier* (NBC) and *K-Nearest Neighbor* (KNN) algorithms which will compare the accuracy of the two methods after going through the preprocessing stage. The application of the *Naïve Bayes* method is relatively easy so it is widely used in research. The *Naïve Bayes Classifier* method can also be applied to different domains. The *K-Nearest Neighbor* (KNN) method is a margin increasing method that falls into two different classes. According to previous research, the *K-Nearest Neighbor* (KNN) and Naïve Bayes methods are the classifications with the highest accuracy [9].

Research using these two methods has been carried out by several researchers. According to research by Shaikh Zainuddin Nahdlatul Wathan Anjani regarding opinions about the public's response to PPPK teachers, using the *Naïve Bayes Classifier* and *K-Nearest Neighbor* algorithms. The results of sentiment analysis on 519 data with an accuracy result of 75.53% for the *Naïve Bayes Classifier* while the *Naïve Bayes Classifier* algorithm with an *accuracy* result of 73.41%[10]. Midhat Pasha Firdaus' research in sentiment analysis of the 2024 election predictions uses the *Naïve Bayes Classifier and K-Nearest Neighbor* algorithms as methods. Using 4453 data with an accuracy value of 82.86% using NBC and 78.56% using KNN[11]. Then in Tamrizal's research regarding sentiment analysis towards BPJS Health on Twitter media using NBC and KNN with accuracy results of 80% for the NBC algorithm and 67% for the KNN algorithm[12]. This research is related research regarding *cyberbullying* Sentiment Analysis using the *Naïve Bayes Classifier* and *K-Nearest Neighbor* Algorithms with word weighting using TF-IDF.

2. RESEARCH METHODS

Explains the stages used by the author in conducting research. The research method flow can be seen in Figure 1.
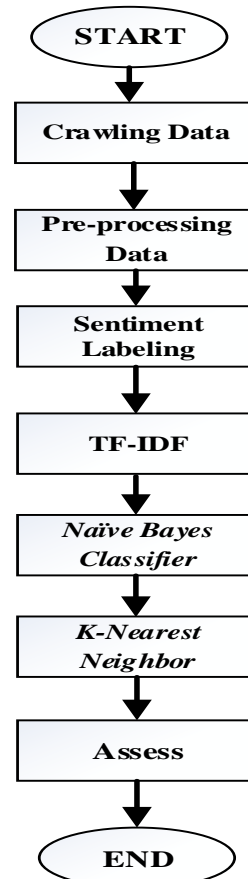

Figure 1. Research Method Flow

The research method is a flow for carrying out the stages in the research that will be carried out as shown in Figure 1.

2.1. *Crawling Data*

Crawling or data collection is collecting data for datasets [13]. *Crawling* is a technique used to collect information on the web. *Crawling* works automatically, where the information collected is based on keywords provided by the user. The tool used to crawl is called a crawler. Crawler is a program that is programmed with a certain algorithm, so that it can scan web pages, according to the web address or keywords provided by the user. When scanning, the crawler will read the existing text, hyperlinks and various tags used on the web page. Based on this information, the crawler will index the information or save the information in a file or database.

2.2. *Pre-processing Data*

Text Preprocessing is a process for improving the quality of text or selecting text data to eliminate noise [14]. *Pre-processing Data* is a data mining

technique that involves transforming raw data into an easy-to-understand format. The data pre-precossing step is necessary to resolve several types of problems including noisy data, data redundancy, missing data values, etc.

The data pre-processing steps are tokenization, case folding, stemming, filtering, and labeling.

1. Tokenization

Tokenization is the process of separating a series of characters based on space characters, and perhaps at the same time the process of removing certain characters, such as punctuation marks, is also carried out [15].

2. Case Folding

Case folding is the process of changing all the letters in a document or sentence to lowercase. The folding case is used to make searching easier. Not all data is consistent in the use of capital letters [16].

3. Stemming Words

That have been changed to lowercase need to be checked. Stemming is used to standardize words so as to reduce the list of words in the training data [17].

4. Filtering

Filtering/eliminating stopwords has many advantages, namely reducing space in the term index table by up to 40% or more . The stopword removal process is the process of removing terms that have no meaning or are irrelevant [18].

5. Labelling

Labeling comes from the word label. Label means a character or set of characters used to identify a variable or part of data or a file. There are 2 (two) labeling processes, namely labeling the token with the word exaggeration and the word negation [19].

Word weighting is the process of giving weight to each word contained in a document. In searching for ranking information based on word frequency, one of the most popular methods is the TFIDF (Term Frequency - Inversed Document Frequency) method.

In the TF-IDF method, Term Frequency focuses more on terms that frequently appear in a document, while Inverse Document Frequency focuses more on giving low weight to terms that appear in many documents.

Naive Bayes classification is a classification that is supervised learning because it has a supervisor (a human who manually classifies the data used in training) as a teacher in the learning process. Apart from that, Naive Bayes performance has a short classification time, thereby speeding up the sentiment analysis system process. In this research, testing was used using 2 classes (negative and positive) [20].

2.3. *Text Mining*

Text mining is a mining process carried out by a computer to obtain something new, previously unknown, or to rediscover implicitly implied information. The findings come from information extracted automatically from different text data sources. Text Preprocessing is part of Text Mining which is done to remove noise in sentences. Text Preprocessing aims to avoid imperfect data, interference with data and non-consumable data [21].

2.4. *Sentiment Labeling*

Sentiment labeling is giving a sentiment value to a text which can be positive, negative or neutral. This labeling can be done using various methods, including manually, namely inputting the sentiment value contained in the text yourself, then using the TF-IDF function, and with Python it can also be done using the TextBlob library.

2.5. TF-IDF *Weightingan*

TF-IDF is a method that is an integration between term frequency (TF) and inverse document frequency (IDF). Term Frequency is calculated using Equation with the ith term frequency being the frequency of appearance of the ith term in the jth document. Inverse Document Frequency (IDF) is the logarithm of the ratio of the number of all documents in the corpus to the number of documents that have the term in question as written mathematically in Equation . The value is obtained by multiplying the two which is formulated in Equation [22].

$$idf_t = \log \frac{N}{df_t} \qquad (1)$$

$$W_{d.t} = tf_{t.d} \; x \; idf_t \qquad (2)$$

2.6. *Naïve Bayes Classifier (NBC)*

Naïve Bayes is a classification method using simple probability which is rooted in Bayes' Theorem and has a high assumption of independence from each condition or event [23]. This method is part of the Bayes method used in text classification, based on a simplification model that attribute values are conditionally independent if the output value is given. The *Naïve Bayes* method is widely used in classification techniques on Twitter. This method is also used in text mining in sentiment analysis by predicting probabilities based on previous data. *Naïve Bayes* cannot detect images, but can only detect text and numbers. To calculate probability in this method, the Bayes theorem approach is used [24].

2.7. *K-Nearest Neighbor (K-NN)*

K-Nearest Neighbor (K-NN) is a simple method that is easy to implement, the data used has labels making it easier to group the process into the most appropriate classes and has the advantage of being able to classify data using training data and test data and having the ease of translating results and accuracy of predictions by accurately selecting the closest k value first [25]..

3. HASIL DAN PEMBAHASAN

3.1. *Crawling Data*

The data that has been obtained using the data crawling technique is 400 comment data. The results of data crawling can be seen in Figure 2 .

Figure 2. Data *Crawling* Results

The image above is the result of comment data obtained from Instagram social media.

3.2. *Pre-processing Data*

This stage is carried out after we have carried out the data crawling stage. Pre-processing results can be seen in table 1.

Table 1. Results *Pre-processing Data*

Alur Text Preprocessing	Before	After
Cleaning	@fahrianalimbong tolol!! gak ada hubungan nya keguguran dgn pake hijab syar'i yg lo bilang bayi nya kepanasan didalem gak ada hubungan nya woyyyy!! otak sama jempol lo gak singkron sih ya jadinya asal nulis komentar!	tolol gak ada hubungan nya keguguran dgn pake hijab syar'i yang lo bilang bayi nya kepanasan didalem gak ada hubungan nya woyyyy otak sama jempol lo gak singkron sih ya jadinya asal nulis komentar
Tokenization	tolol hubung gugur pakai hijab syar i bilang bayi panas dalam hubung otak jempol singkron tulis komentar	'tolol', 'hubung', 'gugur', 'pakai', 'hijab', 'syar', 'i', 'bilang', 'bayi', 'panas', 'dalam', 'hubung', 'otak', 'jempol', 'singkron', 'tulis', 'komentar'
Case Folding	@fahrianalimbong TOLOL!! Gak ada hubungan nya keguguran dgn pake hijab syar'i yang lo bilang bayi nya kepanasan didalem gak ada hubungan nya woyyyy!! Otak sama jempol lo gak	@fahrianalimbong tolol!! gak ada hubungan nya keguguran dgn pake hijab syar'i yang lo bilang bayi nya kepanasan didalem gak ada hubungan nya woyyyy!! otak sama jempol lo gak
Stemming	singkron sih ya jadinya asal nulis komentar! tolol hubungan keguguran pakai hijab syar'i bilang bayi kepanasan didalam hubungan otak jempol singkron tulis komentar	singkron sih ya jadinya asal nulis komentar! tolol hubung gugur pakai hijab syar i bilang bayi panas dalam hubung otak jempol singkron tulis komentar
Stopword Removal	tolol tidak ada hubungan nya keguguran dengan pakai hijab syar'i yang kamu bilang bayi nyakepanasan didalam tidak ada hubungan nya woi otak sama jempol kamu tidak singkron sih ya jadinya asal tulis komentar	tolol hubungan keguguran pakai hijab syar'i bilang bayi kepanasan didalam hubungan otak jempol singkron tulis komentar

From the table above, you can see the Text *Preprocessing* results in table 1. Datasets that have passed the data preprocessing stage are uniform, structured and the symbols in the text have disappeared. This is because when you enter the next stage, namely the classification stage, it is more optimal in calculations.

3.3. *Sentiment Labeling*

Sentiment labeling uses the Python programming language by creating a negative and positive dictionary in Indonesian and creating a system. In this labeling we can see the differences between words that include positive and negative things. The results of sentiment labeling can be seen in Figure 3 below.

Figure 3. *Sentiment Labeling* Results

The labeling process aims to divide the class into 2 class parts, namely negative and positive classes. Where each sentence will contain the value of each word that contains the sentiment class.

3.4. TF-IDF

To complete the word weighting on the training data and test data, the next process is to calculate the TF-IDF value. The equation formula to get the word weighting results for each word (*term*) in each document can be known as an equation. Here's the process python TF-IDF weighting as with language programming is assisted by the Scikit learn library, TfidfVectorizer in figure 4.

```
'''
Convert a collection of raw documents to a matrix of TF-IDF features
https://scikit-learn.org/stable/modules/generated/sklearn.feature_ext
'''
from sklearn.feature_extraction.text import TfidfVectorizer

tf_idf = TfidfVectorizer(ngram_range=(1,1))
tf_idf.fit(X)
```

```
▾ TfidfVectorizer
TfidfVectorizer()
```

Figure 4. Script TF-IDF

Seen in Figure 4. The process of working on TfidfVectorizer is the process of transforming text documents into vectors. The results of the TF-IDF weighting can be seen in table 2 below.

Table 2. Results TF-IDF

Term	Train Data						Test Data		
	D1	D2	D3	D4	D5	D6	U1	U2	U3
tolol	0	0	0	0	0	0	0.954	0	0
hubung	0	0	0	0	0	0	1.24	0	0
gugur	0	0	0	0	0	0	0.954	0	0
pakai	0	0	0	0.477	0	0	0.477	0.477	0
hijab	0	0	0	0	0	0	0.954	0	0
syar	0	0	0	0	0	0	0.954	0	0
bilang	0	0	0	0	0	0	0.954	0	0
bayi	0	0	0	0	0	0	0.954	0	0
panas	0	0	0	0	0	0	0.954	0	0
dalam	0	0	0	0	0	0	0.954	0	0
otak	0	0	0	0	0	0	0.653	0.653	0
jempol	0	0	0	0	0	0	0.954	0	0
singkron	0	0	0	0	0	0	0.954	0	0
tulis	0	0	0	0	0	0	0.954	0	0
komentar	0	0	0	0	0	0	0.954	0	0
cium	1.412	0	0	0	0	0	0	0	0
orang	0.653	0	0	0	0	0	0	0.849	0
malu	0.849	0	0	0	0	0	0	0	0.849
kecuali	0.954	0	0	0	0	0	0	0	0
mabuk	0.954	0	0	0	0	0	0	0	0
teman	1.526	0	0	0	0	0	0	0	0
libat	0.954	0	0	0	0	0	0	0	0
bibir	1.24	0	0	0	0	0	0	0	0
coba	0.954	0	0	0	0	0	0	0	0
bayangin	0.954	0	0	0	0	0	0	0	0
sanggup	0.954	0	0	0	0	0	0	0	0
dasar	0.954	0	0	0	0	0	0	0	0
laku	0.954	0	0	0	0	0	0	0	0
joget	0	0	0	0	0	0.954	0	0	0

3.5. *Naïve Bayes Classifier*

In the *Naïve Bayes Classifier* method in this research, as previously stated in the model stage, this research divides the training data and test data into 3 parts and is then processed using the *Naïve Bayes Classifier* method. The following are the results of dividing the training data dataset: test data can be seen in table 3.

Table 3. *Naïve Bayes* Results

Data Latih : Data uji	*Accuracy*	Presisi	*Recall*	*f1-score*
70:30	91%	91.5%	91%	91%
80:20	95%	95%	95%	95%
90:10	93%	92.5%	92.5%	92.5%

The results of the confusion matrix using the *Naïve Bayes* method at a ratio of 90%:10% showed that there were 73 correct predictions for positive sentiment (*true positive*) and 7 data for correct predictions for negative sentiment (*true negative*).

3.6. *K-Nearest Neighbor* (K-NN)

The results of evaluating the dataset with the KNN algorithm with a K value by conducting experiments on the value k=3, with the partitioning validation evaluation model showing the best accuracy value with a k=3 value of 67% was obtained as shown by the confusion matrix which is formed as follows data can be seen in table 3.

Table 4. K-*Nearest Neighbor* Results

Data Latih : Data uji	*Accuracy*	Presisi	*Recall*	*f1-score*
70:30	87.5%	29.2%	33.3%	31.1%
80:20	88%	28.9%	33%	31%
90:10	90%	30%	30%	31.5%

The evaluation results using the KNN algorithm with a value of k=3 show an *accuracy* value of 67%, the confusion matrix formed from the application review data *processing* process.

3.7. The Final Result

Pada tahapan ini menjadi tahapan untuk menentukan mana hasil metode dengan akurasi terbaik. Untuk perbandingan dalam kedua metode tersebut dapat dilihat pada gambar.
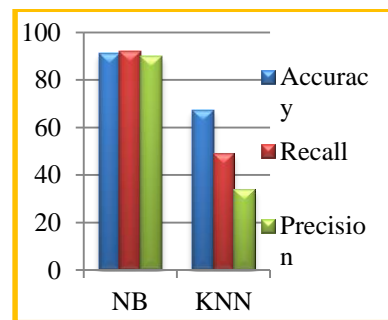
Figure 5. Comparison of Two Methods

In this research, the results of calculations using the first three methods using the Partitionong model, the results using the *Naive Bayes* method obtained an accuracy value of 91.25%, a recall value of 92% and a precision value of 90% and calculations using the KNN method had an accuracy value of 67%, a recall value of 49% and a precision value of 34%.

4. DISCUSSION

Of the three related studies that have been explained in the introduction, researchers will carry

out the same research, but with different research cases and testing with different data on the selected algorithm, with the aim of comparing which two algorithms are more accurate in determining the accuracy of *cyberbullying* levels. .

The researchers conducted sentiment analysis research to find out a person's opinion or response to *cyberbullying*, by carrying out an algorithm classification to find out the accuracy value regarding the selected topic data, to find out the opinion or response of the Indonesian people, especially on the topic of *cyberbullying* from Instagram social media comments, the researchers conducted analysis research sentiment using the *Naïve Bayes Classifier* and *K-Nearest Neighbor* algorithms.

In this case, sentiment analysis is very helpful to find out the responses or opinions of the Indonesian people regarding how to make positive and negative comments so that bullying or cyberbullying does not occur. So, in determining whether comments are positive or negative, the author takes data from Instagram social media comments. In determining the results, the author uses two algorithms to compare which algorithm is more accurate in identifying *Cyberbullying*. The results can be proven by the percentage of sentiment classes in Figures 5 and 6. From these two algorithms, there is an algorithm that is very effective in finding out how big the percentage accuracy of *cyberbullying* comments is.

5. CONCLUSION

Based on the results of the tests that have been carried out, several conclusions can be drawn in analyzing cyberbullying sentiment in Instagram comments using the Naïve Bayes Classifier and K-Nearest Neighbor methods. What is concluded from this research is as follows:
1. TF-IDF helps the *Naïve Bayes Classifier and K-Nearest Neighbor* algorithm in weighting words in text analysis.
2. The *Naïve Bayes Classifier and K-Nearest Neighbor* algorithms can be used to analyze *cyberbullying* sentiment in the Instagram comments column. The classification results are in the form of positive classes and negative classes which are divided into positive sentiment and negative sentiment.
3. Testing was carried out with two algorithms with the same data, the results of the first method were *Naïve Bayes Classifier* with an *accuracy* value of 91.25%n with an average value for each sentiment class of 90% *precision* and 92% *recall* and calculations using the KNN method with an *accuracy* value of 67 %, *recall* value 49% and *precision* value 34%.
4. From these two methods, the author found the method with the best accuracy in cyberbullying sentiment analysis, namely the *Naïve Bayes Classifier* method with an *accuracy* value of 91.25%n with an average value for each

sentiment class of 90% *precision* and 92% *recall*.
5. Text *Preprocessing* has a big influence on classification to increase *accuracy* in sentiment analysis.

REFERENCES

[1] W. P. Onno, *Text mining analisis medsos kekuatan brand & intelejen di internet*, 1st ed. yogyakarta: andi (anggota IKAPI), 2019

[2] Rusdiaman, D., *dan* Rosiyadi, D. Analisa Sentimen Terhadap Tokoh Publik Menggunakan Metode *Naïve Bayes* Classifier Dan Support Vector Machine. *CESS (Journal of Computer Engineering System and Science)*, *4*(2), 230– 235.2019

[3] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *J. Ilm. Edutic Pendidik. dan Inform.*, vol. 7, no. 1, pp. 1–11, 2020.

[4] Sukesti, Restu. "Bianglala mimpi: antologi esai bengkel bahasa dan sastra Indonesia siswa SLTA Kabupaten Sleman."*J. Ilm. Pendidik Bahasa dan Sastra Indonesia.*, vol. 5, no. 2, pp 1-8, 2019.

[5] Arofah, Zahrotul, Muhammad Roisul Basyar, and Anggraeny Puspaningtyas. "Strategi Penanggulangan Bullying pada Sekolah Menengah Pertama dalam Perspektif Collaborative Governance: Studi pada SMP Islam Tikung." *Mutiara: Jurnal Penelitian dan Karya Ilmiah.*, vol. 1, no 6, pp 227-235, 2023.

[6] Saingo, Yakobus Adi. "Pemanfaatan Media Sosial Sebagai Sarana Pembentukan Karakter Kristiani Tingkat Sekolah Menengah Pertama." *Jurnal Shanan.*, vol. 6, no.1, pp 89-110, 2022.

[7] Zulkarnain, Iskandar, et al. "Penerapan Digital Marketing Sebagai Strategi Komunikasi Pemasaran Usaha Kecil Dan Menengah Warga Depok Jaya." *DEDIKASI .*, vol. 1, no. 1, pp. 78–86, 2020.

[8] Anam, M. Khairul, et al. "Penerapan Na ̈ive Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen dan Pemeritah." *MATRIK: J. Manaje, T. Informatika dan Rekayasa Komputer.*, vol. 21, no. 1, pp. 139-150, 2021.

[9] Ilmawan, Lutfi Budi, and Muhammad Aliyazid Mude. "Perbandingan metode klasifikasi Support Vector Machine dan Naïve Bayes untuk analisis sentimen pada ulasan tekstual di Google Play Store." *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 154-161, 2020.

[10] Hidayat, Fajar Nurrachmat, and Sugiyono Sugiyono. "Analisis Sentimen Masyarakat Terhadap Perekrutan Pppk Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine." *Jurnal Sains dan Teknologi.,* vol. 5, no. 2, pp. 665-672, 2023.

[11] Zain, Haekal Hilmi, Rolly Maulana Awannga, and Woro Isti Rahayu. "Perbandingan Model Svm, Knn Dan Naïve Bayes Untuk Analisis Sentiment Pada Data Twitter: Studi Kasus Calon Presiden 2024." JIM: *Jurnal Ilmiah Mahasiswa Pendidikan Sejarah.,* vol.8, no. 3, pp. 2083-2093, 2023.

[12] Puspita, Rani, and Agus Widodo. "Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS.*" Jurnal Informatika* Universitas Pamulang., vol.5, no. 4, pp. 646-654, 2021.

[13] A. R. Isnain et al., "Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm," *Universitas. Teknokr. Indonesia.,* vol. 2, no. 1, pp. 31– 37, 2021.

[14] A. P. Natasuwarna, "Seleksi Fitur Support Vector Machine pada Analisis Sentimen Keberlanjutan Pembelajaran Daring," *Techno.Com,* vol. 19, no. 4, pp. 437–448, 2020, doi: 10.33633/tc.v19i4.4044

[15] Singgalen, Yerik Afrianto. "Analisis Sentimen Konsumen terhadap Food, Services, and Value di Restoran dan Rumah Makan Populer Kota Makassar Berdasarkan Rekomendasi Tripadvisor Menggunakan Metode CRISP-DM dan SERVQUAL." *Building of Informatics, Technology and Science (BITS).,* vol. 4, no. 4. Pp. 1899-1914. , 2023.

[16] Alita, D., & Isnain, A. R. Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. Jurnal Komputasi., vol. 8, no. 2, pp. 50-58, 2020.

[17] Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. Text preprocessing for text mining in organizational research: Review and recommendations. Organizational Research Methods., vol. 25, no. 1, pp. 114-146, 2022.

[18] Rianto, et al. "Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation." *Journal of Big Data* ., vol. 8, pp. 1-16, 2021.

\[19] Siregar, Amril Mutoi, and Maulana Abdur Rofik. "Mengukur Tingkat Kepuasan Mahasiswa Dalam Pembelajaran Dengan Naïve Bayes." *Prosiding Konferensi Nasional Penelitian Dan Pengabdian*

Universitas Buana Perjuangan Karawang., vol. 1, no. 1, pp. 1-19. 2021.

[20] Juniarsih, Sri, Eva Faja Ripanti, and Enda Esyudha Pratama. "Implementasi Naive Bayes Classifier pada Opinion Mining Berdasarkan Tweets Masyarakat Terkait Kinerja Presiden dalam Aspek Ekonomi." *JUSTIN (Jurnal Sistem dan Teknologi Informasi).,* vol. 8, no. 3, pp. 239-249, 2022.

[21] Sari, Fransiska Vina, and Arief Wibowo. "Analisis sentimen pelanggan toko online Jd. Id menggunakan metode Naïve Bayes Classifier berbasis konversi ikon emosi." Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer 10.2 (2019): 681-686.

[22] Sumayah, Siti, Falentino Sembiring, and Wisuda Jatmiko. "Analysis of sentiment of Indonesian community on metaverse using support vector machine algorithm." *Jurnal Teknik Informatika (JUTIF).,* vol. 4,no, 1, pp. 143-150, 2023

[23] Watratan, Alvina Felicia, and Dikwan Moeis. "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia." *Journal of Applied Computer Science and Technology.,* vol. 1, no. 1, pp. 7-14, 2020

[24] Prabowo, W. A., dan Wiguna, C. Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM. *Jurnal Media Informatika Budidarma.,* vol. 5, no. 1, pp. 149-156, 2021.

[25] Singgalen, Yerik Afrianto. "Analisis Sentimen Konsumen terhadap Food, Services, and Value di Restoran dan Rumah Makan Populer Kota Makassar Berdasarkan Rekomendasi Tripadvisor Menggunakan Metode CRISP-DM dan SERVQUAL." *Building of Informatics, Technology and Science (BITS).,* vol. 4, no. 4. Pp. 1899-1914. , 2023.