TEXT MINING IN ONLINE TRANSPORTATION USER SENTIMENT ANALYSIS ON SOCIAL MEDIA TWITTER USING THE MULTINOMIAL NAIVE BAYESIAN CLASSIFIER METHOD AND K-NEAREST NEIGHBOOR METHOD

Sartika Mandasari[1], Roslina [2], B. Herawan Hayadi [3]

[1,3] *Computer Science, Main Potential University, Indonesia*

2Department of Computer Engineering, Politeknik Negeri Medan, Indonesia

[1]sartikamandasari12@gmail.com
[2]roslinanich@gmail.com
[3]b.herawan.hayadi@gmail.com

Abstract: Text mining is the process of detecting information or something new and researching large information. Text mining can also usually perform an analysis of unstructured text. Social media users in Indonesia, which currently almost reach 200 million users, have resulted in a flood of data. This condition makes text mining a solution to extract knowledge from the flood of data [1] .
In exploring knowledge, there are various techniques or methods that can be adopted including the Multinomial Naive Bayesian Clasifier and K-Nearest Neighbor methods. Both of these methods have several phases that are able to explore the potential knowledge of a flood of supervised and unsupervised learning data. It is hoped that the combination of these two methods will help analyze public sentiment or perception towards online motorcycle taxi users in Indonesia [2].

Keywords: Text Mining, Sentiment Analisis, MNBC, K-NN

I. INTRODUCTION

Social media is one of the media that is a place for virtual gatherings. Currently, there are many social media platforms in the midst of society including Facebook, Instagram, YouTube, Linkedin, TikTok, Twitter and others. The use of social media today is not only for entertainment but can also be used commercially for product or company introduction. Some companies are currently using social media such as Twitter to see the positive and negative sentiment of moving the nation's capital [3]

Sentiment analysis or opinion mining is the process of automatically understanding, extracting, and processing textual data to obtain sentiment information contained in an opinion sentence. In general, sentiment analysis aims to determine the attitude of the speaker or author towards a topic or the overall contextual polarity of a document. One of the media that can be measured sentiment analysis is social media. Attitudes can be in the form of judgment or evaluation, the emotional side of the author at the time of writing or the effect of emotional communication that the author wants on his readers. This sentiment analysis is not only applied in the industrial world but in the health sector as during the pandemic yesterday, related to the analysis of Covid-19 sentiment on Twitter [4].

Twitter as one of the interactive social networks allows its users to criticize an issue or a service facility in real time. For example, the use of Twitter in analyzing cyberbullying sentiment in the midst of society. People who originally took a long time to convey their aspirations can now do it easily thanks to the presence of this technology. In the banking industry how can this sentiment analysis be successful in collecting data related to usury sentiment[5].

Grab Indonesia as one of the leading online motorcycle taxi companies in Indonesia has a fairly large number of customers and covers almost all regions in Indonesia. Each customer has a different level of satisfaction with the services provided by Grab Indonesia, so there are always pros and cons in the form of suggestions and complaints[6]. Processing of suggestions and complaints can now be submitted in real time via @GrabID Twitter account, so Grab Indonesia can quickly find out how customers respond to the services provided. However, considering the large number of customers, not a few suggestions and complaints addressed to @GrabID accounts are received per day through the playstore media [7].

II. METHOD

II.1 STAGES OF RESEARCH

At this stage of the study, which starts from the data collection process (data collecting) then a literature study is carried out to display the results and draw conclusions[8]. Based on this explanation, the following is a picture that explains the research methodology carried out, namely as follows:

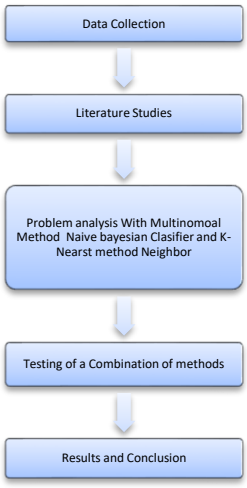

Fig. 1 Stages of research

From these results, it can be seen what the conclusions of the combination of these 2 methods are in analyzing the sentiment of the public users of the Grab Indonesia online motorcycle taxi service[9]. From the results of this combination, Grab can use this as a reference to improve customer service, and the company can get core input from users efficiently [10].

II.2 MULTINOMIAL NAIVE BAYESIAN CLASIFIER METHOD

The algorithm is an explanation of the steps for solving the problem [11]. The following is the solving algorithm of the Multinomial Naïve Bayesian Clasifier method which is as follows:

1. Prepare in advance testing data
2. Create a frequency distribution table
3. Calculating probability values
4. Calculates the probability value of a word appear
5. Calculate the final value
6. Drawing conclusions based on the final result

To estimate the prior probability of P(c) the equation is used as berkut:

$$P(c)=N\_c/N \qquad (1)$$

To estimate the conditional probability P(t|c) equation used, i.e. :

$$P(t_k|c) = \frac{T_{c_t}}{\sum t' \in v\, T_{c_{t'}}} \qquad (2)$$

To omit the value of zero (0), add-one or Laplace smoothing is used. This process adds a value of one (1) to each Tct value from the conditional probabilities calculation. So that the equation for conditional probabilities becomes:

$$P(t_k|c) = \frac{T_{c_t} + 1}{(\sum t' \in v\, T_{c_{t'}}) + B'} \qquad (3)$$

To obtain the final value of probability on the tested data document, whether the test document belongs to the class of negative, positive or negative is used equation:

$$P = \frac{N_c}{N} X \frac{T_{c_t} + 1}{(t' \in v\, T_{c_{t'}}) + B'} \qquad (4)$$

II.3 K-NEAREST NEIGHBOR METHOD

K-Nearest Neighbor (K-NN) is a simple method that is easy to implement, the data used has labels so that it facilitates the grouping process into the most suitable classes and has the advantage of being able to classify data with training data and test data and has the ease of translating the results and accuracy of predictions by accurately sorting out the nearest k values correctly [12]. K-NN will be calculated using the Simalirity formula as follows:

$$Similarity: (T,S) = \frac{\sum_{i=1}^{n} f\,(Ti,Si) * Wi}{Wi}$$

The algorithm is an explanation of the steps for solving the problem. The following is the solving algorithm of the K-Nearest Neighbor method [13], which is as follows:

1. Prepare in advance testing data (criteria and weight criteria)
2. Calculating similarity values
3. Calculate the proximity attribute value
4. Calculating the Distance value
5. Drawing conclusions

II.4 TEKS MINING

The process of extracting information carried out by a user who interacts with a set of documents using Text mining analysis tools is the process of extracting patterns (useful information and knowledge) from a number of data sources through the identification of interesting patterns [14].

1. Preprocessing

The process of extracting information carried out by a user who interacts with a set of documents using Text mining analysis tools is the process of extracting patterns (useful information and knowledge) from a number of data sources through the identification of interesting patterns[15].

2. Vector Drafting

To be understood by text mining operating systems, a vector representation of word tokens needs to be created based on the word's appearance in the document [16].

3. Information Extraction

The information extraction method used in this study is a classification that divides objects into predetermined categories (supervised method). The classification model approach used is statistical / machine learning techniques. The machine learning approach will be used in this study where the machine will operate a model that learns from examples of classified documents [17]

II.5 SENTIMENT ANALYSIS

Sentiment Analysis, is the process of analyzing, understanding, and classifying opinions, evaluations, judgments, attitudes, and emotions towards a particular entity such as products, services, organizations, individuals, events, topics, in order to obtain information [18]. Mood analysis in opinion is called Sentiment Analysis, which refers to the automatic analysis of evaluative texts by focusing on classifying texts based on their polarity. The classification of data on a particular group of sentiments (positive or negative) is carried out by building a model of the probability of the appearance of a word in a pre-grouped document according to Muthia S.[19]

II.6 PREPROCESSING

Preprocessing works to turn unstructured text data into structured data. This processing stage will reduce the texts in a signified manner, namely texts that have no effect on the document [20] Where the explanation of these stages is as follows:

a. Cleaning

Cleaning is the process of cleaning sentences from words that are not needed to reduce noise such as HTML characters, RT, emotion icons, hashtags (#), usernames (@), urls (http://situs.com), emails (nama@situs.com), symbols and punctuation marks.

b. Case Folding

Case folding is the process by which converting all uppercase or uppercase characters into lowercase or lowercase letters [21].

c. Tokenization

Tokenization is separating a series of words within a sentence, paragraph or page into a token or piece of a single word or termmed word .

d. Normalization

Substitutes a word that contains sentiment from a non-standard word to a standard word and that corresponds to its word synonyms, which has been stored in the synonym database.

e. Stopword Removal

Stopword removal is the removal of a large number of classes of conjunctions but does not affect the overall content of the document as part of pre-processing. This stage is usually done to further improve system performance so that the system can be effectively utilized in processing content that is really considered important.

f. Stemming

Stemming is the process of turning a word that has an affix into a base word. Stemming here uses a dictionary of lists of words that have the basic word by comparing the words in the document.

g. Convert Negation

At this stage, any tweets containing negated words will have their sentiment value changed. Negational words such as "no", "no", "no", "ga", "don't", "no", and "no". Convert negation is done if there is a negation word before the word that has a positive value, then the word will be changed its value to negative and vice versa.

II.7 Term Frequency (TF)

TF is the frequency with which a term appears in the document in question. The greater the number of occurrences of a term (high TF) in the document, the greater the weight or it will provide a greater conformity value [22]. In Term Frequency (TF), there are several types of formulas that can be used:

1. Binary TF (binary TF), pays attention only to a word or term whether or not it exists in the document, if there is given a value of one (1), if it is not assigned a value of zero (0).
2. Pure TF (raw TF), the TF value is given based on the number of occurrences of a term in the document. For example, if it appears five (5) times then the word will be worth five (5).
3. Logarithmic TF, this is to avoid the dominance of documents that contain few terms in the query, but have a high frequency.
4. TF normalization, using a comparison between the frequency of a term and the maximum value of the frequency set of terms or the entire term present in a document.

III. RESULTS AND DISCUSSION

III.1 SYSTEM ANALYSIS

To test the system, a training process and model testing are carried out according to research methods.

As for explaining the training process of training and testing data, 3 tweet documents are given to be used as training data and 1 tweet document as test data that has previously passed the preprocessing process [23]. Which can be seen in tables 1 and 2.

TABLE. I
STAGES OF PREPROCESSING DATA TRAINING

Document 1	Before Processing	After Processing	Manual Sentiment
1	@prozven Wahhh enjoy Kak ❤ Thank you for #Percaya on Grab	happy favors, thank you for believing in grabindonesi a service	Positive

	Indonesia service yaa - Arrange		
2	Please click the following link, https://twitter.com/messages/compose?recipient_id=2322081114 later Big Brother will be directed to DM Grab Indonesia then directly type #Bantuan yaa. Thank you - Rara	please click the grabindonesia direction link directly help thank you	Neutral
3	It was a really strange dream just now. It's really coherent the problem. Truly dystopian cyberpunk GoTo / Grab Indonesia in 2049. Gw tell me dah.	New Strange Dreams Connecting the Problem of Fear of Goto GrabIndonesia Benefits in 2049 Story	Negatif

TABLE II
SAMPLE TEST DATA

Document 1	**Before Processing**	**After Processing**	**Manual Sentiment**
1	@sakilahB Wahh joined in the fun of listening to it, sis. Thanks for sharing your experience using Grab ya kak service, continue to use Grab Indonesia ya Stay Safe service and stay always #Percaya yes - Carla	Join in the fun, thank you for the nature of Grab Indonesia's service remains safe	?

III.2 INSTALLING TERM FREQUENCY (TF) ON TRAIN AND TEST DATA

The process of calculating the term frequency will be directly sorted to sort the terms in Documents 1-10 (D1-D10) alphabetically, all the same terms must be entered in the sorting process, then in the grouping process the same term will be collected into one to calculate the term frequency in each document 1-3 [24]. The results of the sorting and grouping process can be seen in table 3.

TABLE III
RESULTS OF SORTING AND GROUPING PROCESS (1-9)

No	**Term**	**D1**	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**
1	nature	0	0	0	0	0	0	0	0	0	0
2	safe	0	0	0	0	0	0	0	0	0	0
3	for	0	0	0	0	0	0	0	0	0	0
4	hear	0	0	0	0	0	0	0	0	0	0
5	grabindonesia	1	1	0	0	1	1	1	1	1	1
6	follow	0	0	0	0	0	0	0	0	0	0
7	service	0	0	0	1	1	0	0	0	0	0
8	happy	0	0	0	0	0	0	0	0	0	0
9	thank you	0	0	0	1	1	1	1	0	0	0
10	remain	0	0	0	0	0	0	0	0	0	0

III.3.CALCULATING TRAINING DATA CLASSIFICATION WITH NAIVE BAYES CLASSIFIER MULTINOMIAL ALGORITHM

After going through the process of calculating the appearance of words in each document, it will then go through the classification stage. In this process, the Multinomial Naive Bayes algorithm will be used by calculating the prior probability [25].

TABLE IV .
PRIOR PROBABILITY CALCULATION RESULTS FOR EACH CLASS

Class	***Prior Prebability***
Positive	2/62
Negative	3/62
Neutral	57/62

After calculating the prior probability, then calculate the conditional probability value and laplace smoothing each term in each class using the equation. For example, conditional and laplace smoothing calculations for the term "nature" in the negative, positive and neutral classes are:

The word "natural | negative" $\frac{0+1}{33+343} = \frac{1}{376} = \mathbf{0,0026595744680851}$

The word "natural | positive" $\frac{0+1}{5+343} = \frac{1}{348} = \mathbf{0,0028735632183908}$

The word "natural | neutral" $\frac{1+1}{597+343} = \frac{2}{940} = \mathbf{0,0021276595744681}$

III.5CALCULATING TRAINING DATA CLASSIFICATION WITH K-NEAREST NEIGHBOOR

After preprocessing, the next step is to do calculations with the sililarity formula [26].

TABLE 5
NEGATIVE CRITERIA TABLE

	Criterion	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	
New Cases	D62	0	0	0	0	0	0	0	1	0	0	???
Old Cases	D1	0	1	1	0	1	0	0	1	0	0	Negative

Calculating Proximity Values D1=Old cases and D62=New Cases

$$\text{Distance} = \frac{(a * b) + (c * d) + (e * f)}{b + d + f}$$

$$\text{Distance} = \frac{(C1 * W1) + (C2 * W2) + (C3 * W3) + (C4 * W4) + (C5 * W5) + (C6 * W6) + (C7 * W7) + (C8 * W8) + (C9 * W9) + (C10 * W10)}{W1 + W2 + W3 + W4 + W5 + W6 + W7 + W8 + W9 + W10}$$

$$\text{Distance} = \frac{(1 * 0.1) + (0.5 * 0.1) + (0.5 * 0.1) + (1 * 0.1) + (0.5 * 0.1) + (1 * 0.1) + (1 * 0.1) + (1 * 0.1) + (1 * 0.1) + (1 * 0.1)}{1}$$

= 0,8500

TABLE 6
POSITIVE CRITERIA TABLE

	Criterion	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	
New Cases	D62	0	0	0	0	0	0	0	1	0	0	???
Old Cases	D4	0	0	0	1	0	0	1	0	0	0	Positive

Calculating Proximity Values D4=Old cases and D62=New Cases

$$\text{Distance} = \frac{(a * b) + (c * d) + (e * f)}{b + d + f}$$

$$\text{Distance} = \frac{(C1 * W1) + (C2 * W2) + (C3 * W3) + (C4 * W4) + (C5 * W5) + (C6 * W6) + (C7 * W7) + (C8 * W8) + (C9 * W9) + (C10 * W10)}{W1 + W2 + W3 + W4 + W5 + W6 + W7 + W8 + W9 + W10}$$

$$\text{Distance} = \frac{(1 * 0.1) + (1 * 0.1) + (1 * 0.1) + (0.5 * 0.1) + (1 * 0.1) + (1 * 0.1) + (0.5 * 0.1) + (1 * 0.1) + (1 * 0.1) + (1 * 0.1)}{1}$$

= 0,8500

TABLE 6
NETRAL CRITERIA TABLE

	Criterion	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	
New Cases	D62	0	0	0	0	0	0	0	1	0	0	???
Old Cases	D5	0	0	0	1	0	0	1	1	0	0	Neutral

Calculating Proximity Values D5=Old cases and D62=New Cases.

$$\text{Distance} = \frac{(a * b) + (c * d) + (e * f)}{b + d + f}$$

$$\text{Distance} = \frac{(C1 * W1) + (C2 * W2) + (C3 * W3) + (C4 * W4) + (C5 * W5) + (C6 * W6) + (C7 * W7) + (C8 * W8) + (C9 * W9) + (C10 * W10)}{W1 + W2 + W3 + W4 + W5 + W6 + W7 + W8 + W9 + W10}$$

$$\text{Distance} = \frac{(1 * 0.1) + (1 * 0.1) + (1 * 0.1) + (0.5 * 0.1) + (1 * 0.1) + (1 * 0.1) + (0.5 * 0.1) + (1 * 0.1) + (1 * 0.1) + (1 * 0.1)}{1}$$

= 0,9000

IV. CONCLUSIONS

The Multinomial Naive Bayesian Clasifier and K-Nearest Neighbor approaches were shown to have a decent level of accuracy in identifying analytical sentiment based on the findings of this study. Other studies can use either of these techniques to improve their results and accuracy, particularly when it comes to sentiment analysis in the area of internet transportation or other sectors, particularly those based on text mining. According to the analysis's findings, 39 out of a total of 61 evaluated objects fall under the sentiment category, which has a Nearest value of 1.0.

REFERENCES

[1] N. Dwitiyanti and N. Selvia, "Analisis Sentimen Twitter Kebiasaan New Normal," *Semin. Nas. Ris. dan Teknol. dan Inov. Teknol. (SEMNAS RISTEK)*, no. 2020, pp. 832–836, 2021, [Online]. Available: http://proceeding.unindra.ac.id/index.php/semnasristek/article/viewFile/5073/912

[2] R. Alfyani and Muljono, "Comparison of Naïve Bayes and KNN algorithms to understand hepatitis," *Proc. - 2020 Int. Semin. Appl. Technol. Inf. Commun. IT Challenges Sustain. Scalability, Secur. Age Digit. Disruption, iSemantic 2020*, pp. 196–201, 2020, doi: 10.1109/iSemantic50169.2020.9234299.

[3] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.

[4] A. K. Fauziyyah, "Analisis Sentimen Pandemi Covid19 Pada Streaming Twitter Dengan Text Mining Python," *J. Ilm. SINUS*, vol. 18, no. 2, p. 31, 2020, doi: 10.30646/sinus.v18i2.491.

[5] F. A. Maulana, I. Ernawati, P. Labu, and J. Selatan, "Analisa sentimen cyberbullying di jejaring sosial twitter dengan algoritma naïve bayes," *Semin. Nas. Mhs. Ilmu Komput. dan Apl. (SENAMIKA*, pp. 529–538, 2020, [Online]. Available: https://conference.upnvj.ac.id/index.php/senamika/article/view/619

[6] M. T. Anjasmoros, I. Istiadi, and F. Marisa, "Analisis Sentimen Aplikasi Go-Jek Menggunakan Metode SVM Dan NBC (Studi Kasus: Komentar Pada Play Store)," *Conf. Innov. Appl. Sci. Technol. (CIASTECH 2020)*, no. Ciastech, pp. 489–498, 2020.

[7] F. Fahrurrozi, S. Sayyidi, and I. Ali, "Analisis Layanan Ojek Online PT. Grab Indonesia Wilayah Surabaya dalam Perspektif Bisnis Islam," *Jesya (Jurnal Ekon. Ekon. Syariah)*, vol. 3, no. 1, pp. 147–157, 2020, doi: 10.36778/jesya.v3i1.139.

[8] I. A. Safra and E. Zuliarso, "Analisa Sentimen Persepsi Masyarakat Terhadap Pemindahan Ibukota Baru di Kalimantan Timur pada Media Sosial Twitter," *Proceeding SENDIU 2020*, pp. 978–979, 2020, [Online]. Available: https://unisbank.ac.id/ojs/index.php/sendi_u/article/view/7984/2947

[9] M. A. Muzakki, "Klasifikasi dan Analisa Sentimen Kuesioner Fasilitas dan Layanan untuk Universitas Qomaruddin Gresik," *J. Comput. Sci. Vis. …*, vol. 5, pp. 68–76, 2020, [Online]. Available: https://www.journal.unusida.ac.id/index.php/jik/article/view/291

[10] M. N. Randhika, J. C. Young, A. Suryadibrata, and H. Mandala, "Implementasi Algoritma Complement dan Multinomial Naïve Bayes Classifier Pada Klasifikasi Kategori Berita Media Online," *Ultim. J. Tek. Inform.*, vol. 13, no. 1, pp. 19–25, 2021, doi: 10.31937/ti.v13i1.1921.

[11] K. L. Liu and T. T. Wong, "Naïve Bayesian classifiers with multinomial models for rRNA taxonomic assignment.," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 10, no. 5, pp. 1334–1339, 2013, doi: 10.1109/tcbb.2013.114.

[12] P. Kasemsumran and E. Boonchieng, "EEG- Based Motor Imagery Classification Using Novel String Grammar Fuzzy K-Nearest Neighbor Techniques with One Prototype in Each of Classes," *2020 Int. Conf. Artif. Intell.*

Inf. Commun. ICAIIC 2020, pp. 742–745, 2020, doi: 10.1109/ICAIIC48513.2020.9065236.
[13] T. Wiratama Putra, A. Triayudi, and A. Andrianingsih, “Analisis Sentimen Pembelajaran Daring Menggunakan Metode NaÃ¯ve Bayes, KNN, dan Decision Tree,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 1, pp. 20–26, 2022, doi: 10.35870/jtik.v6i1.368.
[14] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, “Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia,” *Edutic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.
[15] A. Noviriandini, H. Hermanto, and Y. Yudhistira, “Klasifikasi Support Vector Machine Berbasis Particle Swarm Optimization Untuk Analisa Sentimen Pengguna Aplikasi Pedulilindungi,” *JIKA (Jurnal Inform.*, vol. 6, no. 1, p. 50, 2022, doi: 10.31000/jika.v6i1.5681.
[16] A. Hidayat, “Aplikasi Teks Mining Untuk Mendeteksi Spam Pada Email Berbasis Naive Bayes,” vol. 2, no. 8, pp. 1–10, 2022.
[17] H. P. Doloksaribu and Y. T. Samuel, “Komparasi Algoritma Data Mining Untuk Analisis Sentimen Aplikasi Pedulilindungi,” *J. Teknol. Inf. J. Keilmuan dan Apl. Bid. Tek. Inform.*, vol. 16, no. 1, pp. 1–11, 2022.
[18] I. Fahrur Rozi, A. Taufika Firdausi, and K. Islamiyah, “Analisis Sentimen Pada Twitter Mengenai Pasca Bencana Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram,” *J. Inform. Polinema*, vol. 6, no. 2, pp. 33–39, 2020, doi: 10.33795/jip.v6i2.316.
[19] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
[20] L. Pandwinata, “Klasifikasi Text Mining Untuk Analisa Keluhan Masyarakat Terhadap Pembelian Produk Kosmetik Online Dengan Menggunakan Metode Naïve Bayes,” vol. 1, no. 2, pp. 36–42, 2022.
[21] A. Hamano, A. Yamada, T. Takada, and M. Takeuchi, “田口敦清 1) , 齊藤結花 2) , 河田聡 1),2),” no. 1, pp. 1–2.
[22] P. Djodi, “Implementasi Algoritma Text Mining TF-IDF Untuk Fitur Autoresponder,” vol. 10, no. 1, pp. 31–38, 2022.
[23] R. T. Aldisa and P. Maulana, “Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM,” *Technol. Sci.*, vol. 4, no. 1, pp. 106–109, 2022, doi: 10.47065/bits.v4i1.1581.
[24] J. A. Septian, T. M. Fachrudin, and A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor,” *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.
[25] C. F. Hasri and D. Alita, “Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, pp. 145–160, 2022, [Online]. Available: http://jim.teknokrat.ac.id/index.php/informatika
[26] A. Deviyanto and M. D. R. Wahyudi, “Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 1, p. 1, 2018, doi: 10.14421/jiska.2018.31-01.