Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 1 (2023) 1 - 6 ISSN Media Electronic: 2580-0760

Naïve Bayes and TF-IDF for Sentiment Analysis of the Covid-19 Booster Vaccine

Imelda Imelda[1], Arief Ramdhan Kurnianto[2]
[1,2] Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur
[1]imelda@budiluhur.ac.id, [2]arieframdhan8@gmail.com

Abstract

The booster vaccine polemic became a trending topic on Twitter and reaped many pros and cons. This booster vaccine began to be distributed on January 12, 2022. This booster vaccine program was implemented free of charge for the people of Indonesia to prevent the new variant of Covid-19, Omicron. The contribution of this study is to analyze the sentiment of booster vaccines to prevent covid-19 using the Naïve Bayes and TF-IDF methods. We conducted sentiment analysis to determine whether the tweet was positive, negative, or neutral. The solution used is the Naïve Bayes method and TF-IDF. The role of TF-IDF is to determine how relevant the data in the document is by utilizing word weighting. The stages of this research using CRISP-DM include Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment. The net data results show 1,557 data with a positive sentiment of 1,335, a neutral sentiment of 171 data, and a negative sentiment of 51 data. The test results with 60:40 data sharing obtained accuracy, precision, and recall values of 85.26%, 85%, and 100%. The results of this test have increased by 7.26%, 12%, and 20% from other previous studies with the same data distribution.

Keywords: naïve bayes, TF-IDF, sentiment analysis, booster vaccine

1. Introduction

Twitter is one of the largest data-producing social media on the Internet that we can use to analyze data. The flexibility of Twitter allows users to freely interact, such as writing, reading, or discussing with fellow Twitter users. The advantage, Twitter can make it easier for people to express opinions without having to meet in person [1]. We must manage the increasing amount of digital data via Twitter into helpful information. One way to collect data about the valuable tip is to analyze it. The science that discusses data analysis via Twitter is known as text mining. This Twitter analysis is unique because we can dig up people's opinions or opinions about something. This opinion is known as sentiment. Sentiment analysis is the process of understanding, extracting, and processing text data automatically contained in sentiment sentences. Sentiment analysis aims to produce a certain percentage of conditions, products, companies, institutions, or things. The analysis results include positive, negative, or neutral sentiment [2]. The results of this sentiment analysis will help more accurate and effective decision-making [3]. The booster vaccine polemic has become a trending topic on Twitter. This polemic reaps many pros and cons. Booster vaccine polemic This was widely discussed and became an issue of debate in Indonesian society. Booster vaccines will be distributed starting January 12, 2022. This booster vaccination program is carried out free of charge to help the Indonesian people prevent Omicron [4], a new variant of Covid-19. The target of this vaccination is 18 years and over. The main priority for booster vaccinations is elderly parents.

Hoax news on social media regarding the destructive consequences of administering booster vaccines has hampered efforts to distribute booster vaccines throughout Indonesia [5]. The tendency of the Indonesian people to immediately react negatively and share data when they are sick with the first and second vaccines results in high negative sentiment [6]. On the other hand, a positive response from most Indonesian people regarding this vaccination is still possible because this vaccine prevents the transmission of Covid-19 [7]. Olhang et al. [8] examine the sentiment analysis of Twitter users towards Covid-19 in Indonesia. This research uses the Naïve Bayes Classifier (NBC) method to make the data grouping consisting of positive and negative sentiment. His research uses approximately 500 tweets of data. The amount of test data is 75 tweet data. The measurement results show a recall accuracy of 32%, precision of

Accepted: 31-08-2022 | Received in revised: 26-11-2022 | Published: 01-02-2023

80%, F -Measure of 45%, and an average accuracy of 36%. Septianingrum et al. [9] also used the Naïve Bayes method. Her research is on sentiment analysis on the issue of the Covid-19 vaccine in Indonesia with classification into three classes, namely positive, negative, and neutral sentiments. This study resulted in a model accuracy rate of 78%, 80% recall, and an AUC score of 0.904.

Fitriana and Sibaroni [10] researched sentiment analysis on KAI's Twitter posts. The research method is a Multiclass Support Vector Machine. This method is for classifying tweet data from the @KAI121 account. The ratio of training data: to test data used in this research is 90:10. The weighting used in this research is the Unigram TF-IDF because it has unique features and can produce exact values. His study results follow research from the @KAI121 account, with positive sentiments of 11%, neutral sentiments of 58%, and negative sentiments of 31%. The accuracy results show that the TF-IDF weighting with a gamma value of 0.7 obtained a value of 80.59.

Berlian et al. [11] researched sentiment analysis of public opinion on television shows on Twitter. This research also uses the Naïve Bayes method using 550 tweets of data. The data was divided into ten parts, with 55 tweet data each to training data and test data, testing the data using three categories: precision, recall, and accuracy. The test results for the three categories get a precision value of 68%, a recall of 77%, and an accuracy of 61%. In contrast, this study classifies sentiment into three parts: positive, negative, and neutral. Crawling data took Twitter data as many as 4,073 tweet data. The collection period starts from June 1, 2022, to June 15, 2022. The keyword used is booster vaccine. In Indonesian language, we call “vaksin booster”. Sentiment analysis of booster vaccines using the Naïve Bayes and TF-IDF methods via social media Twitter is the aim and contribution of this research. The booster vaccine is one of the efforts to prevent the new variant of Covid-19. This new variant is known as Omicron. The Naïve Bayes classification is made using the PHP programming language.

This research contribution discusses in more detail how the public reacted to the first phase booster vaccine plan being procured by the government. This research is essential because it is necessary to know the community's trend towards the first phase booster vaccine plan. This study is different from previous studies [1] [4] [5] [6] [7] [8] [9] [12], which only discuss vaccines in general. Meanwhile, this study examines the reaction of the Indonesian people on Twitter social media to booster vaccines.

2. Research Methods

The stages of this research using CRISP-DM include Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. Business Understanding is essential for problem analysis. Data Understanding is necessary for data collection. Data Preparation is vital for manual data elimination, preprocessing, and labeling. Modeling is crucial to modeling data, weighting words with TF-IDF, and classification using Naïve Bayes. Evaluation is vital to test the validity of the data using the confusion matrix. Deployment is crucial to ensure the application can run properly using the PHP programming language.

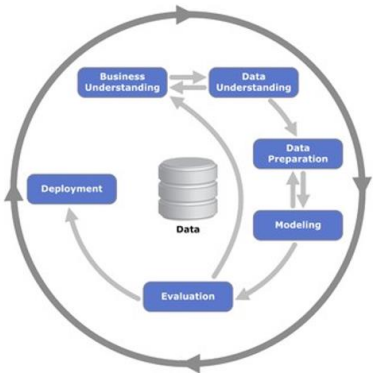

Figure 1. CRISP-DM [13]-[14]

Figure 1 shows the stages of CRISP-DM [13] - [14]. This study aims to classify positive, negative, and neutral sentiments from tweets using the Naïve Bayes method.

The first stage is Business Understanding (understanding of the business). This first stage is to gain an in-depth knowledge of the needs and definition of problems related to booster vaccines. The second stage is Data Understanding. This second stage is essential for data collection. The data used in this study comes from social media Twitter. The tweet data keyword used in this research is " booster vaccine. " Data collection uses Indonesian. The tweet data collection starts from June 1, 2022, to June 15, 2022. The data collection results obtained were 4,073 tweets. The third stage is Data Preparation. This third stage prepares the data for processing. Therefore, this stage is often called preprocessing [15]. Several preprocessing steps include: Cleansing is the process of cleaning data whose contents have nothing to do with words, such as symbol characters, emoticons, and URL links [16], Case Folding is the process of converting all text characters into lowercase [17], Tokenizing is a process of dividing sentences into parts or words. The words that this process generates are called tokenized [12], Stop-word Removal is the process of removing common words that have no meaning. Such as eliminating some verbs and adding adjectives and adverbs to the list of stop words. [18]. The stop word in this study comes from the Kaggle.com dictionary [14]. Stemming is searching for the roots of words resulting from the previous filtering process. This process is done for each word. The way stemming works is to return

DOI: https://doi.org/10.29207/resti.v7i1.4467

various forms of the stem by removing the sentences that the researcher added. The stemming process in this study uses library Literature [19].

Algorithm 1. TF-IDF

```
documents = read_documents()
all_words = []
idf = []
tf = [][]

for (document in documents)
{
words = document. split()
for (word in words)
{
tf[document][word] += 1 / words.count
}

for (word in tf[document])
{
all_words[word] += 1
}
print(tf)
}

#calculate IDF
for (word in all_words)
{
idf[word] = log(documents.count / all_words[word])
}
print(idf)

for (document in documents)
{
words = document. split ()
tfidf = []
for (word in words)
{
tfidf[word] = tf[document][word] * idf[word]
}
print(tfidf)
}
```

Algorithm 1. TF-IDF Algorithm

The fourth stage is Modeling. This fourth stage is the core of data classification. The classification used in this study is Naïve Bayes. Naïve Bayes is a machine learning method using statistical and probabilistic calculations advocated by British scientist Thomas Bayes. How Naive Bayes works predicts the future based on past experiences. Class is divided into 3, namely: positive, negative, and neutral. This Naive Bayes classifier aims to measure the computational accuracy of data: positive, negative, and neutral sentiments. This classification is also assisted by using the TF-IDF method. The role of the TF-IDF is to determine how relevant the data is in the document by weighting words [20]. The goal is to count the number of comments from each sentence the researcher will use. Simply put, TF-IDF can be shown in Algorithm 1.

The fifth stage is Evaluation. This fifth stage aims to evaluate whether the results are appropriate. Evaluation is usually done by testing. Tests run on specific applications. Testing using data sampling with split data 60:40. This means that 60% of the data is used as training data, and 40% of the data is used as test data. The reason for dividing the data is 60:40 because it refers to another study [9] with quite good results, and there is a belief that using this division will increase the results. The training data becomes a data analysis model that represents the entire data. Test data is used as new data for comparison. The Confusion Matrix method takes part in this evaluation stage. The Confusion Matrix aims to get accuracy, precision, and recall values. The values obtained using the Confussion Matrix method are an accuracy of 85.26%, a precision of 0.85, and a recall of 1. Calculation and comparison of terms in the test data with each class [21] can be shown in equation (1):

$$P(w_K|v_j) = \frac{nk+1}{n+|Vocabulary|} \quad (1)$$

With wK is the word (word) k in all documents labeled as j (sentiment positive / negative / neutral), vj is all words (vocabulary) in class j (sentiment positive / negative / neutral), nk is the number of times a word appears in class j (sentiment positive / negative / neutral), n is the number of words in class j (sentiment positive / negative / neutral), and vocabulary is the total number of unique words in a document. The sixth stage is Deployment. This sixth stage is necessary for implementation. Implementation of this research using the PHP programming language. The dataset consists of 1,335 positive sentiment data, 51 negative sentiment data, and 171 neutral sentiment data.

3. Results and Discussions

Data results crawling (data collection) for this sentiment analysis obtained as many as 4073 tweet data. Table 1 shows the Data Collection. The keyword used is booster vaccine. Data collection time from June 1, 2022, until June 15, 2022.

Table 1. Data Collections

Keyword	Date	Amount
booster vaccine	June 1 – June 15, 2022	4,073 data

Table 2 shows sample data from the results of crawling with the keyword Booster Vaccines.

Table 2. Data Samples

Tweet
RT @AliBej0: Jangan ragu vaksinasi, Vaksin aman.
RT @forjakeu: Menurut para ahli vaksin Covid-19 tidak sebabkan virus baru . Penelitian terus berlanjut dan masyarakat dihimbau jangan ragu
@Beritaupdate17 Menkes Bawa Kabar Baik Soal Booster Vaksin Covid gesss
@ahmaddamarrr Jangan Lengah Rek, Ayo Vaksin Booster AstraZeneca Pfizer Dulu di Sini

Data preparation prepares the data so that it is ready to be classified. Table 3 shows the preprocessing stages: cleansing, case folding, tokenization, stopword removal, and stemming. This study uses a stopword

DOI: https://doi.org/10.29207/resti.v7i1.4467

dictionary. This study also uses literary literature in the stemming section and the Big Indonesian Dictionary as a guide.

Table 3. Preprocessing stage

Preprocessing	Before	After
Cleansing	RT @Anggita_lung: Capaian vaksinasi booster baru mencapai 25%. Maka dari itu, sesuai arahan Presiden @jokowi ,masyarakat diminta untuk segera…	Capaian vaksinasi booster baru mencapai Maka dari itu sesuai arahan presiden masyarakat diminta untuk segera
Case Folding	Capaian vaksinasi booster baru mencapai Maka dari itu sesuai arahan presiden masyarakat diminta untuk segera	capaian vaksinasi booster baru mencapai maka dari itu sesuai arahan presiden masyarakat diminta untuk segera
Tokenize	capaian vaksinasi booster baru mencapai maka dari itu sesuai arahan presiden masyarakat diminta untuk segera	*"capaian", "vaksinasi", "booster", "baru", "mencapai", "maka", "dari", "itu", "sesuai", "arahan", "presiden", "masyarakat", "diminta", "untuk", "segera"*
Stopword	"capaian","vaksinasi" ,"booster","baru" , "mencapai" , "maka" , "dari" , "itu" , "sesuai" , "arahan" , "presiden" , "masyarakat" , "diminta", "untuk" , "segera"	"capaian" , "vaksinasi" , "booster" , "mencapai" , "sesuai" , "arahan" , "presiden" , "masyarakat" , "segera"
Stemming	"capaian" , "vaksinasi" , "booster" , "mencapai" , "sesuai" , "arahan" , "presiden" , "masyarakat" , "segera"	"capai" , "vaksinasi" , "booster" , "capai" , "sesuai" , "arah" , "presiden" , "masyarakat" , "segera"

Table 4 shows the steps for calculating the Term Frequency-Inverse Document Frequency (TF-IDF). The term is a word that we can use to define the context of document (D). This process is used for word weighting. This process is carried out sequentially to find the Term Frequency (TF) value, Document Frequency (DF) value, Inverse Document Frequency (IDF) value, and weight value. The TF-IDF implementation in PHP is shown below.

```
<?php
$tfidf = $term_frequency *// tf
log( $total_documents_count / $documents_with_term, 2); // idf
?>
```

Table 4 shows the Term calculation, which aims to find the number of words in each document D1, D2, D3, D4, and D5. Examples:

D1 = abis vaksin booster emang ga enak gini ya badannya;
D2 = mau nanya vaksin booster tuh udah umur sih;
D3 = iyahh vaksin tersedia langsung booster;
D4 = syarat stadion vaksin booster;
D5 = booster vaksin ya buka aja begitu

Table 4. TERM calculation

TERM	TF				
	D1	D2	D3	D4	D5
abis	1	0	0	0	0
vaksin	1	1	1	1	1
booster	1	1	1	1	1
emang	1	0	0	0	0
ga	1	0	0	0	0
enak	1	0	0	0	0
gini	1	0	0	0	0
ya	1	0	0	0	1
badan	1	0	0	0	0
mau	0	1	0	0	0
nanya	0	1	0	0	0
tuh	0	1	0	0	0
udah	0	1	0	0	0
umur	0	1	0	0	0
sih	0	1	0	0	0
iyahh	0	0	1	0	0
sedia	0	0	1	0	0
langsung	0	0	1	0	0
syarat	0	0	0	1	0
stadion	0	0	0	1	0
buka	0	0	0	0	1
aja	0	0	0	0	1
begitu	0	0	0	0	1
Document Length	9	8	5	4	6

Table 5 shows the calculation results of the Normalized TF obtained from the number of words for each document. Calculation of TF Normalization is each word of the document divided by the length of the document.

Table 5. Calculation of Normalized TF

Normalization TF				
D1	D2	D3	D4	D5
0.111	0	0	0	0
0.111	0.125	0.2	0.25	0.167
0.111	0.125	0.2	0.25	0.167
0.111	0	0	0	0
0.111	0	0	0	0
0.111	0	0	0	0
0.111	0	0	0	0
0.111	0	0	0	0.167
0.111	0	0	0	0
0	0.125	0	0	0
0	0.125	0	0	0
0	0.125	0	0	0
0	0.125	0	0	0
0	0.125	0	0	0
0	0.125	0	0	0
0	0	0.2	0	0
0	0	0.2	0	0
0	0	0.2	0	0
0	0	0	0.25	0
0	0	0	0.25	0
0	0	0	0	0.167
0	0	0	0	0.167
0	0	0	0	0.167

DOI: https://doi.org/10.29207/resti.v7i1.4467

Table 6 shows the DF calculation process by counting the number of documents that contain terms; for example, the term "abis" only appears in 1 document, namely in D1, and not in documents D2, D3, D4, or D5. An example of the term "vaksin" in each document shows us that the DF is 5. The IDF calculation process is summed up by calculating the LOG (entire documents divided by DF), as in the example LOG (full documents totaling five divided by DF).

Table 6. IDF calculations

DF	IDF
1	0.699
5	0
5	0
1	0.699
1	0.699
1	0.699
1	0.699
2	0.398
1	0.699
1	0.699
1	0.699
1	0.699
1	0.699
1	0.699
1	0.699
1	0.699
1	0.699
1	0.699
1	0.699
1	0.699
1	0.699
1	0.699
1	0.699

Table 7 shows the last step, namely the calculation of TF-IDF which is the multiplication of the Normalized TF value and the IDF value for each term in each document, namely D1, D2, D3, D4, and D5. Table 8 shows the modeling of the Naïve Bayes method. This modeling uses several samples of existing tweets.

Table 7. Calculation of TF-IDF

TF-IDF				
D1	D2	D3	D4	D5
0.078	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0.078	0	0	0	0
0.078	0	0	0	0
0.078	0	0	0	0
0.078	0	0	0	0
0.044	0	0	0	0.066
0.078	0	0	0	0
0	0.087	0	0	0
0	0.087	0	0	0
0	0.087	0	0	0
0	0.087	0	0	0
0	0.087	0	0	0
0	0.087	0	0	0
0	0	0.14	0	0
0	0	0.14	0	0
0	0	0.14	0	0
0	0	0	0.175	0
0	0	0	0.175	0
0	0	0	0	0.117
0	0	0	0	0.117
0	0	0	0	0.117

Table 8. Sample Tweets for Naïve Bayes

Tweets	Label
mau nanya vaksin booster tuh udah umur sih	Positive
booster vaksin ya buka aja begitu	Positive
iyahh vaksin sedia langsung booster	Positive
syarat stadion vaksin booster	Positive
abis vaksin booster emang ga enak gini ya badan	Negative

This study uses 5 sample tweets. Then the probability of a positive document is P($positive$) $= \frac{4}{5}$= 0.8 and the probability of a negative document is P ($negative$) $= \frac{1}{5}$= 0.2. Based on the formula written in equation (1), the probability calculation is carried out to determine the negative or positive class as follows:

P(mau|positive) $= \frac{1}{23} + \frac{1}{32} = \frac{2}{55}$ = 0.03636

P(begitu|positive) $= \frac{1}{23} + \frac{1}{32} = \frac{2}{55}$ = 0.03636

P(sedia|positive) $= \frac{1}{23} + \frac{1}{32} = \frac{2}{55}$ = 0.03636

P(syarat|positive) $= \frac{1}{23} + \frac{1}{32} = \frac{1}{55}$= 0.03636

P(ga|negative) $= \frac{1}{9} + \frac{1}{32} = \frac{2}{41}$= 0.04878

Table 9 shows an example for determining the class of a *tweet.*

Table 9. Determining Sentiment

Tweets	Label
sehabis vaksin booster badan ko pegel	-

Researchers calculate the value of Naive Bayes to get positive and negative classes based on tweets shown in Table 9: (positive). P(sehabis|positive). p(vaksin|positive). p(booster|positive). p(badan|positive). p(ko|positive). p(pegel|positive)

0.8, x 0.03636 x 0.03636 x 0.03636 x 0.03636 x 0.03636 =$5.085\ x\ 10^{-8}$

(negative). P(mau|negative). p(vaksin|negative). p(booster|negative). p(sudah|negative). p(sedia|negative). p(ya|negative)

0.2, x 0.04878 x 0.0487 x 0.0487 x 0.0487 x 0.0487 =$5.487\ x\ 10^{-8}$

Table 10 shows the results of the multiplication value having a high negative value $5.487x\ 10^{-8}$and a low positive value $5.085\ x\ 10^{-8}$, so we can conclude that the sentence has a negative sentiment.

Table 10. Prediction Results of the Naïve Bayes Method

Tweets	Label
sehabis vaksin booster badan ko pegel	**Negative**

Table 11 shows the final stages of the test results from all the *tweets* obtained. The results show that the accuracy value is 85.26%, the precision of 0.85 or 85%, and the recall of 1 or 100%. This data comes from

DOI: https://doi.org/10.29207/resti.v7i1.4467

crawling results *Tweets* containing the keyword "vaksin *booster*" have 1,557 *tweets*. The test results show that this research is better than this [9] , which both used a 60:40 division. The increase of this research is an accuracy of 7.26%, a precision of 12%, and a recall of 20%.

Table 11. Test Results for the Naïve Bayes Method

Train Data	Test Data	Split Data	accuracy	precision	recall
933	624	60:40	85.26%	0.85	1

4. Conclusion

Based on the tests and analysis carried out for the sentiment analysis of the booster vaccine prevention of Covid-19 using Naïve Bayes and TF-IDF through the social network Twitter with a total of 1,557 data and using a ratio of 60:40, the data is manually divided into two parts, namely training data and test data. The dataset consists of positive sentiment consisting of 1,335 data, negative sentiment composed of 51 data, and neutral sentiment consisting of 171 data. Each of these data has a different value, 933 data for training data and 634 data for test data. The test was carried out using the Naïve Bayes method with a ratio of 60:40 to obtain an accuracy score of 85.26%, 85% precision, and 100% recall. The test results show that this research is better than before, using the same 60:40 division with an increase in accuracy of 7.26%, precision of 12%, and recall of 20%. This analysis is still limited to quantitative analysis using Naïve Bayes and TF-IDF. In further research, it can also be elaborated with critical analysis in the qualitative form to make it more comprehensive.

References

[1] S. Suhardiman and F. Purwaningtias, "Analisis Sentimen Masyarakat Terhadap Virus Corona Berdasarkan Opini Masyarakat Menggunakan Metode Naïve Bayes Classifier," *J. Pengemb. Sist. Inf. dan Inform.*, vol. 2, no. 4, pp. 220–232, 2020.

[2] M. W. A. Putra, S. Susanti, E. Erlin, and H. Herwin, "Analisis Sentimen Dompet Elektronik Pada Media Sosial Twitter Menggunakan Naïve Bayes Classifier," *IT J. Res. Dev.*, vol. 5, no. 1, pp. 72–86, 2020.

[3] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," *J. Tekno Kompak*, vol. 15, no. 1, pp. 131–145, 2021.

[4] N. P. G. Naraswati, D. C. Rosmilda, D. Desinta, F. Khairi, R. Damaiyanti, and R. Nooraeni, "Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification," *Sist. J. Sist. Inf.*, vol. 10, no. 1, pp. 228–238, 2021.

[5] A. A. Permana, M. F. Fahrezi, D. Y. Priyanggodo, D. A. Kristiyanti, and M. Sihotang, "Sentimen Analisis Opini Masyarakat Pada Media Sosial Twitter Terhadap Vaksin Berbayar Menggunakan Metode Naïve Bayes Classifier (NBC)," *JTS J. Tek.*, vol. 10, no. 2, pp. 84–92, 2021.

[6] F. Fathonah and A. Herliana, "Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naïve Bayes," *J. Sains dan Inform.*, vol. 7, no. 2, pp. 155–164, 2021, doi: 10.34128/jsi.v7i2.331.

[7] N. M. A. J. Astari, D. G. H. Divayana, and G. Indrawan, "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier," *J. Sist. dan Inform.*, vol. 15, no. 1, pp. 22–29, 2020, doi: 10.30864/jsi.v15i1.332.

[8] M. M. M. Olhang, S. Achmadi, and F. X. Ariwibisono, "Analisis Sentimen Pengguna Twitter Terhadap Covid-19 Di Indonesia Menggunakan Metode Naïve Bayes Classifier (NBC)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 4, no. 2, pp. 214–221, 2020.

[9] F. Septianingrum, J. H. Jaman, and U. Enri, "Analisis Sentimen Pada Isu Vaksin Covid-19 di Indonesia dengan Metode Naive Bayes Classifier," *J. Media Inform. Budidarma*, vol. 5, no. 4, pp. 1431–1437, 2021, doi: 10.30865/mib.v5i4.3260.

[10] D. N. Fitriana and Y. Sibaroni, "Sentiment Analysis on KAI Twitter Post Using Multiclass Support Vector Machine (SVM)," *Resti J.*, vol. 4, no. 5, pp. 846–853, 2020.

[11] T. F. Berlian, A. Herdiani, and W. Astuti, "Analisis Sentimen Opini Masyarakat Terhadap Acara Televisi pada Twitter dengan Retweet Analysis dan Naïve Bayes Classifier," in *e-Proceeding of Engineering*, 2019, vol. 6, no. 2, pp. 8660–8667.

[12] W. Yulita, E. D. Nugroho, and M. H. Algifari, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid - 19 Menggunakan Algoritma Naïve Bayes Classifier," *JDMSI*, vol. 2, no. 2, pp. 1–9, 2021.

[13] Y. Suhanda, I. Kurniati, and S. Norma, "Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik," *J. Teknol. Inform. dan Komput. MH Thamrin*, vol. 6, no. 2, pp. 12–20, 2020.

[14] T. Wurijanto, H. B. Setiawan, and A. B. Tjandrarini, "Penerapan Model CRISP-DM pada Prediksi Nasabah Kredit yang Berisiko Menggunakan Algoritma Support Vector Machine," vol. 10, 2022.

[15] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 4, no. 2, pp. 113–118, 2018.

[16] A. Hendra, "Analisis Sentimen Review Halodoc Menggunakan Naive Bayes Classifier," *JISKa*, vol. 6, no. 2, pp. 78–89, 2021.

[17] D. D. 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.

[18] A. P. Maharani and A. Triayudi, "Sentiment Analysis of Indonesian Digital Payment Customer Satisfaction Towards GOPAY, DANA, and ShopeePay Using Naïve Bayes and K-Nearest Neighbour Methods," *J. Media Inform. Budidarma*, vol. 6, no. 1, pp. 672–680, 2022, doi: 10.30865/mib.v6i1.3545.

[19] M. T. A. Bangsa, S. Priyanta, and Y. Suyanto, "Aspect-Based Sentiment Analysis of Online Marketplace Reviews Using Convolutional Neural Network," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 14, no. 2, pp. 123–134, 2020, doi: 10.22146/ijccs.51646.

[20] L. Mayasari and D. Indarti, "Klasifikasi Topik Tweet Mengenai Covid Menggunakan Metode Multinomial Naive Bayes Dengan Pembobotan TF-IDF," *J. Ilm. Inform. Komput.*, vol. 27, no. 1, pp. 43–53, 2022.

[21] L. A. Andika, P. A. N. Azizah, and R. Respatiwulan, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," *Indones. J. Appl. Stat.*, vol. 2, no. 1, pp. 34–41, 2019.

DOI: https://doi.org/10.29207/resti.v7i1.4467