Zaini, Sari, Hasan - cek.03

by Fachri Zaini, Jessica Windi Sari, Firman Noor Hasan

Submission date: 11-Jul-2023 10:03AM (UTC+0700)
Submission ID: 2129423181
File name: Zaini,_Hasan_-_cek.03.docx (372.4K)
Word count: 6256
Character count: 33806

Jurnal Teknik Informatika (JUTIF)
Vol. x, No. y, December 2023, pp. x-y

DOI: 10.52436/jutif...
p-ISSN: 2723-3863
e-ISSN: 2723-3871

ANALYSIS OF PUBLIC SENTIMENT RELATED TO THE FAILURE OF Indonesia to Host U-20 USING MULTINOMIAL NAÏVE BAYES CLASSIFIER

Fachri Zaini [1], Jessica Windi Sari[2], Firman Noor Hasan[*3]

[1,3]Informatics Engineering Department, Faculty of Engineering, Prof. Dr. HAMKA Muhammadiyah University, Indonesia
[2]Psychology Department, Faculty of behavioral science, HELP University, Malaysia
Email: [1]fachrizaini19@gmail.com, [2]jessicatan602@gmail.com, [3]firman.noorhasan@uhamka.ac.id

(Article received: date; Revision: date; published: date)

Abstract

The case of Indonesia's failure to host the U-20 World Cup in 2023 has become a hot topic of discussion in Indonesia. The rejection of the Israel U-20 national team and security factors by FIFA are considered the main reasons for the cancellation. This raises many issues and controversies from various parties. In this study, sentiment analysis using the Naive Bayes algorithm was conducted. Researchers use the naive bayes algorithm because this algorithm has high accuracy with simple calculations. The data obtained in this study came from 250 tweets of Twitter data with a ratio of training and test data of 7:3. The results showed good data classification with 97.26% accuracy, 93.33% precision, and 100% recall. In conclusion, the classification model developed can describe public sentiment related to Indonesia's failure in the U-20 World Cup well.

Keywords: *Naive bayes, Sentiment analysis, World cup U-20*

1. INTRODUCTION

In the current era of technological advancement, social media has become one of the main platforms for people to interact and communicate quickly and easily. One of the popular social media platforms that is often used to express opinions freely is Twitter [1]. Twitter is a social media platform that has the largest number of users among several other social media with a total of 328 million users in the world [2][3].

Twitter is considered as a platform that allows users to express their thoughts and opinions more freely, easily due to its accessibility, unlimited number of followers and character limit of only 280 characters [4]. This allows users to convey their messages clearly, concisely and effectively [3]. As one of the most popular social media platforms, Twitter gives users the ability to express their objective opinions on various topics. So from this, there are many studies on sentiment analysis using Twitter data to find out a person's opinion or reaction to a phenomenon both negatively and positively.

Currently, an event that is being discussed by many people in Indonesia is the case of Indonesia's failure to host the U-20 world cup in 2023. According to PSSI Chairman Erick Thohir, the reason why FIFA canceled Indonesia as the host of the 2023 U-20 World Cup was because there were intervention factors, namely the rejection of the Israeli U-20 national team and security factors. The rejection is considered a paradox because previously Indonesia volunteered as the host and was successfully selected by FIFA [5]. Issues eventually developed in the community that blamed Ganjar Pranowo as the person who caused the failure of the U-20 world cup in Indonesia because of his statement that refused the Israeli national team to compete. Another issue was the Kanjuruhan tragedy that resulted in the death of several people [6].

The topic was discussed by many people through social media that people have. Where many also provide opinions, criticisms and suggestions on Twitter social media ranging from positive responses and negative responses. The response data on Twitter social media regarding Indonesia's failure to host the world cup can be used as a valuable source of information in understanding people's opinions and reactions to the event.

Response data or opinions written by people on Twitter social media can be classified using sentiment analysis [7]. Sentiment analysis is a technique used to analyze the viewpoints, emotions, and attitudes expressed by the public on a topic [8]. In the case of Indonesia's failure to host the world cup, it is quite difficult to determine the negative, positive, or neutral sentiment of tweets manually because it will take a lot of time and effort considering the large amount of Twitter. Therefore, a machine is needed that can automatically analyze tweets and classify the sentiment of the tweet to be negative, positive, or neutral. Text classification can

be a solution to the problem so that sentiment determination can be faster [3].

(14) classification algorithm that is often used is the *naïve bayes* algorithm. *Naïve Bayes* algorithm is an algorithm for classifying data in a very simple way in assuming attribute classification [9]. This algorithm is often used in solving problems in the machine learning process and is also known to have a high level of (20)uracy with simple calculations [10], [11]. So in this study, the naïve bayes algorithm will be used in the text clas(25)cation process because according to Fitriyyah that the naïve bayes algorithm has advantages in the speed of the process and an adequate level of accuracy when used on data with large, diverse, and complex volumes [3].

Some previous research on sentiment analysis is research by Afandi who conducted research on analyzing public sentiment regarding opinions (27)ted to the implementation of electronic systems using the *logistic regression* method. The results of this study indicate that of the 1,074 sentiments collected there are 126 sentiments that are negative, 657 sentiments that are neutral, and 291 sentiments that are positive. The Logistic Regression algorithm managed to produce an accuracy value of 79.07%. This shows that most Indonesians agree with the PSE policy implemented but there are still some people who have not accepted the policy [8](35)hen research by Hasan who conducted research to find out the number of positive and negative sentiment results from the Gra(37)ndonesia service opinion dataset and find out the results of the algorithm testing process(6)nd the accuracy value of the evaluation test using the naïve bayes meth(28) The results of this study show that there are 911 positive sentiments and 89 negative sentiments. In addition, the results of evaluation testing show that for negative sentiment the precision value is 57%, recall 67%, and f1-score 62%. As for positive sentiments, the precision value is 97%, recall 95%, and f1-score 96%. From these results it can be concluded that most customers are satisfied with Grab Indonesia's services [1].

From the problems previously described, this research will analyze the (7)entiment of the community on Twitter towar(6) the failure of the U-20 world cup in Indonesia using the naïve bayes algorithm. The result(7)f this study are expected to provide an overview of the public's response to the failure of the Indonesian U-20 soccer team in the World Cup then it is also expected to be useful in making decisions related to communication strategies and actions that can be taken by related parties in dealing with similar situations in the future.

2. METHODOLOGY

(31)

In this study data processing was carri(8) out with several processes that must be passed, which can be seen in Figure 1:

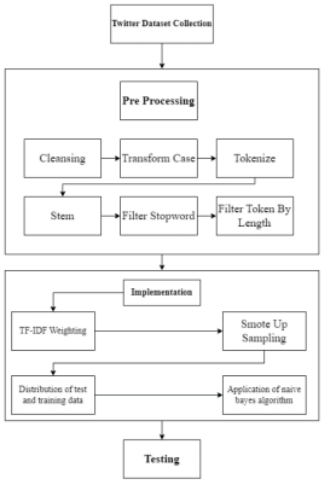


Figure 1: Research flow

Further explanation of Figure 1 is as follows:

1. Data collection
   In this study, researchers collected data from *Twitter* using keywords related to the failure of the Indonesian U-20 World Cup. The data taken are tweets posted within a certain period of time.

2. *Preprocessing* is the initial stage in data processing where raw or unstructured data is converted into data that is more structured and ready to be used f(5) further analysis [12]. In this study, six stages of *Preprocessing* were carried out in this study including : (5)
   A. *Dataset* labeling is the process of labeling or determining the class of *Twitter* responses.
   B. *Cleansing*
      The process of cleaning documents from unnecessary words or also known as the cleaning stage. The goal is to make the documents to be processed cleaner and more relevant. One way to clean documents is to remove *tweet* entities such as mentions, retweets, hashtags and URL links that do not contribute to text analysis [13].
   C. *Transform case*
      Transform cas(10)s a process in text analysis that aims to change the

letters on words in documents to lowercase or uppercase [14].

D. *Tokenize*

Tokenization is the process of converting a text document into a series of tokens or units, such as words or phrases, which are easier for computers to process [15].

E. *Stem*

The *stemming* process is done by removing affixes or prefixes in words so that it only leaves the basic word [3].

F. *Stopword filter*

Stopword filter is a process in text analysis that aims to eliminate words that do not make an important contribution to understanding the content of the document. These words are called stopwords or conjunctions [16].

G. *Filter tokens by length*

Filter tokens by length is a process in text analysis that aims to eliminate words that have a certain number of letters [17].

3. At the implementation stage, *text* classification modeling is carried out using data that can already be processed. Before modeling the data that has been *preprocessed* will go through the TF-IDF weighting stage, which is a weighting method used in text analysis to evaluate how important a word or phrase is in a document [18]. Then after that, data balancing is carried out using the smote up sampling technique. SMOTE technique is a technique to balance the amount of sample data distribution in the minority class by selecting the sample data until the amount of sample data becomes balanced with the number of samples in the majority class [19]. Then the data is divided into testing data and *training* data. Then after that, modeling is done using the *naïve bayes* algorithm.
4. Testing at this stage is testing the model performance of the model that has been generated using the *confusion matrix*. *Confusion matrix* is a calculation that compares the dataset with the Classification results according to the actual data with the total amount of data. The final result of this matrix is the level of accuracy in units of percent (%) [20].

3. RESULTS AND DISCUSSION

1. Data collection

At this stage the researcher collects *datasets* on the *Twitter* application which contains opinions, suggestions and criticisms through *tweets*. The opinion or keyword that is searched to find data is the keyword "U-20 world cup". Illustrations of *Twitter* data collection in this study are as follows.

Figure 2. Data Retrieval Process from Twitter Using Rapidminer

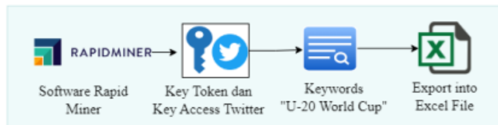


Figure 2 shows the process of pulling data from the rapidminer tool. It can be seen that the withdrawal process starts from connecting rapidminer with Twitter using the token key and Twitter access key. After connecting, just enter the keyword "U-20 world cup" in the query parameter. Later rapidminer will pull tweet data that has the word U-20 world cup. After the withdrawal process is complete, the researcher exports the data into an excel file. The data taken from March 27 to April 1, 2023. The data obtained from the withdrawal of the data is 250 Indonesian-language tweet data.

Figure 3. Partial Results of Tweet Data

Figure 3 shows the results of some of the *tweet* data that has been drawn relating to the opinions, suggestions and criticisms of the public through *Twitter* on the failure of the Indonesian U-20 soccer team in the World Cup.

No	Tweet
1	Pernyataan Sekjen @PDI_Perjuangan tentang "potensi" pemakzulan Presiden RI jika tak menolak Israel dlm Drawing Piala Dunia U20 di Bali, yg saya tafsirka
5	Tentang... https://t.co/GovStrqoLe https://t.co/Shpn2llkIE
6	Namanya juga Hasto, panik dihajar fans bola malah nyeret2 orang yang nga bersuara ??https://t.co/IWhI8rukxZ
7	YAHANA - YAHANU • MAK INI MAK ITU | "KHAS" BANGSAWAN POLITIK" @PDI_Perjuangan
10	Dalihnya selalu berubah - ubah sejak dibatalkannya Drawing Piala Dunia U20 di Bali, selalu menempatkan diri sebagai pihak YANG MAHA BENAR & pihak lain
13	Atau mungkin tabiat... https://t.co/qLOLVCqNra https://t.co/c8p12eiCll
14	RT @otnayram_me: Ganjar dengan jelas, tegas, artikulatif, dan tidak emosional kembali menegaskan betapa pentingnya sikap utk menolak kehadi...
15	@ganjarpranowo Gimana pak rasanya blunder? Terima kasih lho pak, telah membunuh mimpi anak-anak muda Indonesia untuk berlaga membela Merah Pu
16	@yunartowijaya Same energy dgn sekjen partai banteng yg minta tanggapannya ttg kasus pembatalan piala dunia u20 ?
17	
18	Waktu awal2 tampil di tv, analisanya keren bgt mas karena berimbang, skrg kok lbh nyerang pihak lain yah... #justsaying
19	RT @otnayram_me: Ganjar dengan jelas, tegas, artikulatif, dan tidak emosional kembali menegaskan betapa pentingnya sikap utk menolak kehadi...
20	Ganjar dukung adanya pengganti event piala dunia U20 di Solo
21	#KekuatanGanjarMajukanRI https://t.co/A1leqpggMx
22	RT @bamantara00: Ya, Indonesia memang terancam sanksi dari FIFA ketika gagal menyelenggarakan Piala Dunia U20.

2. *Preprocessing*

After collecting the *tweet* data, the next step is to do the *preprocessing* stage. In this study, six stages of *Preprocessing* were carried out in this study including:

A. *Dataset* labeling

The next stage is to perform the data labeling process on the *tweet* data manually by the researcher using the help of the *Microsoft Excel* application as shown in the figure below

Figure 4. Partial Result after Dataset Labeling

After the data labeling process and cleaning duplicate data, the overall result of the total data is 214 data with the highest number of *tweet* categories is negative *tweet* data as much as 121 data. Then the amount of positive category tweet data is 93 data.

B. Cleansing

After the data labeling process is carried out, the next is the process of cleaning the document from unnecessary words or also known as the cleaning stage. The cleaning process in this study uses *the rapidminer tool*. This stage will remove *tweets* such as *mentions*, *retweets*, *hashtags* and URL links that do not contribute to text analysis

Figure 5 : Cleansing Process Using rapidminer Tool

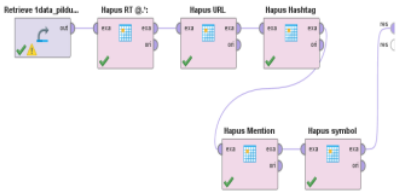


Figure 5 shows the cleansing process using the rapidminer tool. The operator used is the community sentiment data operator associated with the replace operator. The replace operator is used to replace certain values in the data with new values. As for this cleansing process, there are 5 replace operators that have different regular expressions, namely Replace remove RT @, remove URL, remove hashtag, remove mention and remove symbol. The following is a table of results from the cleansing process using the rapidminer tool, namely :

Table 1. Partial Data Cleansing Results

Before	After	Category
Namanya juga Hasto, panik dihajar fans bola malah nyeret2 orang yang nga bersuara ??	Namanya juga Hasto panik dihajar fans bola malah nyeret2 orang yang nga bersuara	Negatif
Dalihnya selalu berubah - ubah sejak dibatalkannya Drawing Piala Dunia U20 di Bali, selalu menempatkan diri sebagai pihak YANG MAHA BENAR & pihak lain pasti salah.	Dalihnya selalu berubah ubah sejak dibatalkannya Drawing Piala Dunia U20 di Bali selalu menempatkan diri sebagai pihak YANG MAHA BENAR pihak lain pasti salah	Negatif
@tempodotco laaah udh lah pildun lagi udh basi...bersyukur aj masih ad yg mo nginep...kalo pildun U20 gagal harusnya yg di salahin Ganjar n partainya...kalo orang Islam mah udh lama nolak Israel dari zaman bung Karno...ga ujug2 nolak	laaah udh lah pildun lagi udh basibersyukur aj masih ad yg mo nginepkalo pildun U20 gagal harusnya yg di salahin Ganjar n partainyakalo orang Islam mah udh lama nolak Israel dari zaman bung Karnoga ujug2 nolak	Positif
@ganjarpranowo Gimana pak rasanya blunder? Terima kasih lho pak, telah membunuh mimpi anak-anak muda Indonesia untuk berlaga membela Merah Putih di Piala Dunia U20. SEMOGA KARIR BAPAK DIPOLITIK SEGERA	Gimana pak rasanya blunder Terima kasih lho pak telah membunuh mimpi anakanak muda Indonesia untuk berlaga membela Merah Putih di	Negatif

TAMAT JUGA. SEMOGA KEGOBLOKAN BAPAK BISA JADI CONTOH AGAR TIDAK DITIRU YANG LAIN.	Piala Dunia U20 SEMOGA KARIR BAPAK DIPOLITIK SEGERA TAMAT JUGA SEMOGA KEGOBLOKAN BAPAK BISA JADI CONTOH AGAR TIDAK DITIRU YANG LAIN	
Kita Semua Harus Introspeksi Terkait Kegagalan Jadi Tuan Rumah Piala Dunia U20 https://t.co/HvovOX4bnR	Kita Semua Harus Introspeksi Terkait Kegagalan Jadi Tuan Rumah Piala Dunia U20	Negatif

C. Transform case

The next stage is to transform the case to change the letters in the words in the document to lowercase or uppercase [19]. In this study, it will be converted into lowercase letters. The following is a table of the results of the Transform case, namely :

Table 2. Partial *Transform case* results

Before	After	Category
Namanya juga Hasto panik dihajar fans bola malah nyeret2 orang yang nga bersuara	namanya juga hasto panik dihajar fans bola malah nyeret2 orang yang nga bersuara	Negatif
Dalihnya selalu berubah ubah sejak dibatalkannya Drawing Piala Dunia U20 di Bali selalu menempatkan diri sebagai pihak YANG MAHA BENAR pihak lain pasti salah	dalihnya selalu berubah ubah sejak dibatalkannya drawing piala dunia u20 di bali selalu menempatkan diri sebagai pihak yang maha benar pihak lain pasti salah	Negatif
laaah udh lah pildun lagi udh basibersyukur aj masih ad yg mo nginepkalo pildun U20 gagal harusnya yg di salahin Ganjar n partainyakalo orang Islam mah udh lama nolak Israel dari zaman bung Karnoga ujug2 nolak	laaah udh lah pildun lagi udh basibersyukur aj masih ad yg mo nginepkalo pildun u20 gagal harusnya yg di salahin ganjar n partainyakalo orang islam mah udh lama nolak israel dari zaman bung karnoga ujug2 nolak	Positif
Gimana pak rasanya blunder Terima kasih lho pak telah membunuh mimpi anakanak muda Indonesia untuk berlaga membela Merah Putih di Piala Dunia U20 SEMOGA KARIR BAPAK DIPOLITIK SEGERA TAMAT JUGA SEMOGA KEGOBLOKAN BAPAK BISA JADI CONTOH AGAR TIDAK DITIRU YANG LAIN	gimana pak rasanya blunder terima kasih lho pak telah membunuh mimpi anakanak muda indonesia untuk berlaga membela merah putih di piala dunia u20 semoga karir bapak dipolitik segera tamat juga semoga kegoblokan bapak bisa jadi contoh agar tidak ditiru yang lain	Negatif
Kita Semua Harus Introspeksi Terkait Kegagalan Jadi Tuan Rumah Piala Dunia U20	kita semua harus introspeksi terkait kegagalan jadi tuan rumah piala dunia u20	Negatif

D. Tokenize

Next is to perform a tokenization process to convert text documents into a series of tokens or units, such as words or phrases. Here are some results of the tokenization process :

Table 3. Partial Tokenization Results

Before	After	Category
namanya juga hasto panik dihajar fans bola malah nyeret2 orang yang nga bersuara	namanya, juga, hasto, panik, dihajar, fans, bola, malah, nyeret, orang, yang, nga,	Negatif

	bersuara	
dalihnya selalu berubah ubah sejak dibatalkannya drawing piala dunia u20 di bali selalu menempatkan diri sebagai pihak yang maha benar pihak lain pasti salah	dalihnya, selalu, berubah, ubah, sejak, dibatalkannya, drawing, piala, dunia, u, di, bali, selalu, menempatkan, diri, sebagai, pihak, yang, maha, benar, pihak, lain, pasti, salah	Negatif
laaah udh lah pildun lagi udh basibersyukur aj masih ad yg mo nginepkalo pildun u20 gagal harusnya yg di salahin ganjar n partainyakalo orang islam mah udh lama nolak israel dari zaman bung karnoga ujug2 nolak	laaah, udh, lah, pildun, lagi, udh, basibersyukur, aj, masih, ad, yg, mo, nginepkalo, pildun, u20, gagal, harusnya, yg, di, salahin, ganjar, n, partainya, jika, orang, islam, mah, udh, lama, nolak, israel, dari, zaman, bung, karnoga, ujug2, nolak	Positif
gimana pak rasanya blunder terima kasih lho pak telah membunuh mimpi anakanak muda indonesia untuk berlaga membela merah putih di piala dunia u20 semoga karir bapak dipolitik segera tamat juga semoga kegoblokan bapak bisa jadi contoh agar tidak ditiru yang lain	gimana, pak, rasanya, blunder, terima, kasih, lho, pak, telah, membunuh, mimpi, anakanak, muda, indonesia, untuk, berlaga, membela, merah, putih, di, piala, dunia, u20, semoga, karir, bapak, di, politik, segera, tamat, juga, semoga, kegoblokan, bapak, bisa, jadi, contoh, agar, tidak, ditiru, yang, lain	Negatif
kita semua harus introspeksi terkait kegagalan jadi tuan rumah piala dunia u20	kita, semua, harus, introspeksi, terkait, kegagalan, jadi, tuan, rumah, piala, dunia, u20	Negatif

E. Stem

Then after the tokenization process, the stemming process will be carried out, namely removing affixes or prefixes in words so that it only leaves the basic word [3]. Here are some of the results of the stem process :

Table 4. Partial Stem Results

Before	After	Category
namanya, juga, hasto, panik, dihajar, fans, bola, malah, nyeret, orang, yang, nga, bersuara	nama, juga, hasto, panik, hajar, fans, bola, malah, seret, orang, yang, tidak, suara.	Negatif
dalihnya, selalu, berubah, ubah, sejak, dibatalkannya, drawing, piala, dunia, u, di, bali, selalu, menempatkan, diri, sebagai, pihak, yang, maha, benar, pihak, lain, pasti, salah	dalih, selalu, ubah, ubah, sejak, batal, drawing, piala, dunia, u, di, bali, selalu, tempat, diri, sebagai, pihak, yang, maha, benar, pihak, lain, pasti, salah	Negatif
laaah, udh, lah, pildun, lagi, udh, basibersyukur, aj, masih, ad, yg, mo, nginepkalo, pildun, u20, gagal, harusnya, yg, di, salahin, ganjar, n, partainya, jika, orang, islam, mah, udh, lama, nolak, israel, dari, zaman, bung, karnoga, ujug2, nolak	lah, pildun, lagi, basibersyukur, aj, masih, ad, yg, mo, nginep, pildun, u20, gagal, harus, yg, di, salah, ganjar, n, partai, jika, orang, islam, mah, udah, lama, tolak, israel, dari, zaman, bung, karnoga, ujug, nolak.	Positif
gimana, pak, rasanya, blunder, terima, kasih, lho, pak, telah, membunuh, mimpi, anakanak, muda, indonesia, untuk, berlaga, membela, merah, putih, di, piala, dunia, u20, semoga, karir,	gimana, pak, rasa, blunder, terima, kasih, lho, pak, telah, bunuh, mimpi, anak, muda, indonesia, untuk, laga, bela, merah, putih, di, piala, dunia, u20, moga, karir, bapak, di, politik, segera, tamat,	Negatif

bapak, di, politik, segera, tamat, juga, semoga, kegoblokan, bapak, bisa, jadi, contoh, agar, tidak, ditiru, yang, lain	juga, moga, goblok, bapak, bisa, jadi, contoh, agar, tidak, tiru, yang, lain.	
kita, semua, harus, introspeksi, terkait, kegagalan, jadi, tuan, rumah, piala, dunia, u20	kita, semua, harus, introspeksi, terkait, gagal, jadi, tuan, rumah, piala, dunia, u20	Negatif

F. Stopword filter

After the stem process is complete, the next step is to *filter stopwords*, removing words that do not make an important contribution to understanding the content of the document. These words are called stopwords or conjunctions [17]. Here are some of the results of *the Stopword Filter* process :

Table 5. Partial *Stopword Filter* Results

Before	After	Category
nama, juga, hasto, panik, hajar, fans, bola, malah, seret, orang, yang, tidak, suara.	nama,hasto,panik, hajar, fans, bola,seret, orang, suara.	Negatif
dalih, selalu, ubah, ubah, sejak, batal, drawing, piala, dunia, u, di, bali, selalu, tempat, diri, sebagai, pihak, yang, maha, benar, pihak, lain, pasti, salah	dalih, ubah, batal, drawing, piala, dunia, bali, tempat, diri, pihak, maha, benar, pihak, salah.	Negatif
lah, pildun, lagi, basibersyukur, aj, masih, ad, yg, mo, nginep, pildun, u20, gagal, harus, yg, di, salah, ganjar, n, partai, jika, orang, islam, mah, udah, lama, tolak, israel, dari, zaman, bung, karnoga, ujug, nolak.	lah, pildun, basibersyukur, aj, masih, ad, yg, mo, nginep, pildun, u20, gagal, yg, salah, ganjar, n, partai, orang, islam, mah, udah, lama, tolak, israel, zaman, bung, karnoga, ujug, nolak.	Positif
gimana, pak, rasa, blunder, terima, kasih, lho, pak, telah, bunuh, mimpi, anak, muda, indonesia, untuk, laga, bela, merah, putih, di, piala, dunia, u20, moga, karir, bapak, di, politik, segera, tamat, juga, moga, goblok, bapak, bisa, jadi, contoh, agar, tidak, tiru, yang, lain.	gimana, pak, rasanya, blunder, terima, kasih, lho, pak, bunuh, mimpi, anak, muda, indonesia, laga, bela, merah, putih, piala, dunia, u20, moga, karir, bapak, politik, tamat, moga, goblok, bapak, contoh, tiru.	Negatif
kita, semua, harus, introspeksi, terkait, gagal, jadi, tuan, rumah, piala, dunia, u20	semua, introspeksi, terkait, gagal, tuan, rumah, piala, dunia, u20.	Negatif

G. Filter tokens by length

The last step in the *preprocessing* process is *Filter tokens by length* which is to eliminate words that have a certain number of letters [17]. Here are some of the results of the *Filter tokens by length* process with letter parameters of 4-15 letters per word :

Table 6. Partial results of *filtering tokens by length*

Before	After	Category
nama, juga, hasto, panik, hajar, fans, bola, malah, seret, orang, yang, tidak, suara.	nama,hasto,panik, hajar, fans, bola,seret, orang, suara.	Negatif
dalih, selalu, ubah, ubah, sejak, batal, drawing, piala, dunia, u, di, bali, selalu,	dalih, ubah, batal, drawing, piala, dunia, bali, tempat, diri, pihak, maha, benar, pihak, salah.	Negatif

tempat, diri, sebagai, pihak, yang, maha, benar, pihak, lain, pasti, salah		
lah, pildun, lagi, basibersyukur, aj, masih, ad, yg, mo, nginep, pildun, u20, gagal, harus, yg, di, salah, ganjar, n, partai, jika, orang, islam, mah, udah, lama, tolak, israel, dari, zaman, bung, karnoga, ujug, nolak.	pildun, basibersyukur, masih, nginep, pildun, gagal, salah, ganjar, partai, orang, islam, udah, lama, tolak, israel, zaman, bung, karnoga, ujug, nolak.	Positif
gimana, pak, rasa, blunder, terima, kasih, lho, pak, telah, bunuh, mimpi, anak, muda, indonesia, untuk, laga, bela, merah, putih, di, piala, dunia, u20, moga, karir, bapak, di, politik, segera, tamat, juga, moga, goblok, bapak, bisa, jadi, contoh, agar, tidak, tiru, yang, lain.	gimana, rasanya, blunder, terima, kasih, bunuh, mimpi, anak, muda, indonesia, laga, bela, merah, putih, piala, dunia, moga, karir, bapak, politik, tamat, moga, goblok, bapak, contoh, tiru.	Negatif
kita, semua, harus, introspeksi, terkait, gagal, jadi, tuan, rumah, piala, dunia, u20	semua, introspeksi, terkait, gagal, tuan, rumah, piala, dunia	Negatif

As for helping the data preprocessing process starting from the *transform case stage, tokenize, stem, Stopword Filter to filter tokens by length*, researchers will use *the rapidminer tool*. The following are the operators used, namely :

Figure 6. Preprocessing Process Using rapidminer Tool

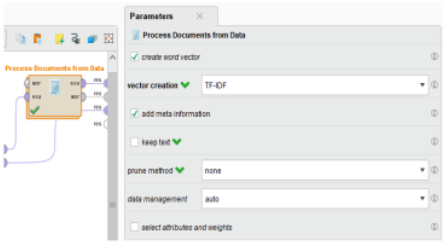
Figure 6 shows the process of advanced preprocessing starting from the transform case stage, tokenize, stem, Stopword Filter to filter tokens by length. As for the Stem and Stopword Filter process, researchers first looked for a stem dictionary containing words based on KBBI and an Indonesian stopword dictionary containing conjunctions. Researchers managed to find both dictionaries via the internet, namely : https://www.kaggle.com/oswinrh/indonesian-stoplist (Indonesian Stopword Dictionary) and https://github.com/sastrawi/sastrawi/tree/master/data (Indonesian Stem Dictionary). After that, the two dictionaries are included in the stem dictionary parameters and stopwords filter. The result of this process is data that is ready to be modeled.

3. Implementation

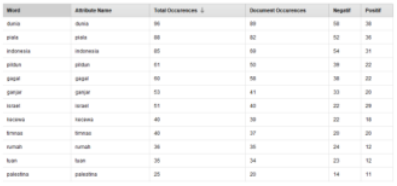
A. TF-IDF Weighting

In the first step of the implementation process, TF-IDF weighting is first carried out to determine the weight of words in the document. One way to do TF-IDF weighting is to use the Process *Documents from Data* operator on the RapidMiner platform. As for the operator, the researcher includes the stages of *the transform case, tokenize, stem, Stopword Filter to filter tokens by length*. The following is the process in TF-IDF weighting using the *rapidminer tool*:

Figure 7. TF-IDF Weighting Process Using *rapidminer Tool*



The results of the TF-IDF weighting process which shows the number of times the word appears in the *dataset* are as follows :



Word	Attribute Name	Total Occurrences ↓	Document Occurrences	Negatif	Positif
dunia	dunia	96	69	58	38
piala	piala	88	62	52	36
indonesia	indonesia	85	59	54	31
pildun	pildun	61	50	38	22
gagal	gagal	60	58	38	22
ganjar	ganjar	53	41	33	20
israel	israel	51	40	22	29
kecewa	kecewa	40	39	22	18
timnas	timnas	40	37	20	20
rumah	rumah	36	36	24	12
tuan	tuan	35	34	23	12
palestina	palestina	25	20	14	11

Figure 8. Preprocessing Process Using rapidminer Tool

Based on Figure 8. Shows that the 7 words that appear most often are the word world by 96 words, trophy by 88 words, indonesia by 85 words, pildun by 61 words, failure by 60 words and ganjar by 53 words.

B. Data Visualization

After the previous stages are completed, the next step is to connect the *Process Documents from Data* operator to the *WordList to Data* operator. This operator is responsible for calculating the weight value and frequency of occurrence of each word in the *dataset* that has gone through the stages of *transform case, tokenize, stem, stopword filter, and filter token by length*. The following is the *WordList to Data* process using the *rapidminer tool* :

Figure 9. Data Visualization Process with *WordCloud*

Figure 9 shows the data visualization process with WordCloud. The operator used is the community sentiment data operator which is connected to the *replace* operator previously described in cleansing. Then the operator is connected to the nominal to text operator to convert the input data into text form. Then it is connected again with the *Process Documents from Data operator which collects transform case, tokenize, stem, Filter Stopword to filter tokens by length operators*. Then it is connected to the *WordList to Data* operator which is used to convert the word list into a data format that can be used for further analysis. This operator needs to be connected again with the *sort* operator and *Filter Example Range* to be able to visualize the result data from the *wordcloud*.

After this process, the next step is to visualize wordcloud on the results that have been obtained to be able to more easily understand the information generated, namely:

Figure 10. Data Visualization Results with *WordCloud*

Figure 10. Shows the results of data visualization using *wordcloud*. These results show that the greater the total words in the document, the greater the words displayed. As for this study, it can be seen that 8 words that often appear in sentiment data regarding the failure of the U-20 world cup in Indonesia are the words world, cup, indonesia, pildun, failed, ganjar, israel and disappointed.

The words that appear can be concluded that the public sentiment towards the failure of the 2023 U-20 world cup is negative. People feel disappointed regarding the failure of the event. Then there are the words "Ganjar" and "Israel" which refer to Ganjar Pranowo, the Governor of Central Java, who the public considers as the person who thwarted the event because of his statement that refused the Israeli national team to compete.

C. Classification using Naïve Bayes algorithm

At the implementation stage, *text* classification modeling is carried out using the *naïve bayes* algorithm. However, before the modeling process is carried out, the data balancing process will first be carried out using the SMOTE technique due to data imbalance between positive and negative sentiment data. After that the data is divided into test data and

training data using split data with a ratio of 3: 7. The ratio comparison used in this study is 3: 7 because based on research [21] states that the greater the percentage difference or ratio between training data and testing data, the higher the accuracy obtained. So in this study, a ratio comparison of 3: 7 ratio between test data and training data to determine the accuracy that will be generated. So if in the division of the 3: 7 ratio, modeling is found that has a good *performance model*, it can be concluded that the accuracy value at a ratio of 1: 9 between test data and training data will produce an even better *performance model*. The following is the modeling process using the rapidminer tool and can be seen in Figure 11.

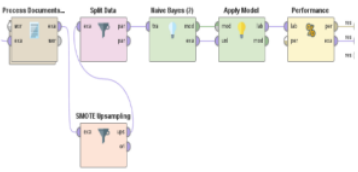


Figure 11. Modeling Process Using *rapidminer Tool*

Figure 11 shows the process of the modeling stage in this study using the *naïve bayes* algorithm. The operators used include the Process Documents from Data operator which collects *transform case operators, tokenize, stem, Stopword Filter to filter tokens by length*. Then connected with the *SMOTE Upsampling* operator which is used to balance negative and positive data. Then for the modeling process, it is necessary to divide the data between testing data and training data using the split data operator. Then the operator is connected to the *naïve bayes* operator as a classification algorithm. Then it is connected to the *apply model* operator to apply the model to the testing data to generate sentiment predictions. And finally connected with the performance operator to provide a metric for evaluating the performance of the model in this study, namely *accuracy, precision and recall*.

As for using the SMOTE Upsampling operator, it makes the data balanced where the total data after balancing the data is 242 data with the number of classes between positive and negative sentiments as much as 121 data.

D. Testing

At this stage, model performance testing is carried out from the model that has been generated using the confusion matrix. This test is carried out to measure the accuracy of the model in classifying data. As in Figure 11, it can be seen that there is a split data operator which is used to divide the data into test data and training data. Then the operator is reconnected with the naïve bayes operator and apply model for the modeling process. After that, it is finally connected with performance operator to determine the performance of the model using the confusion matrix. The results of this process are as follows:

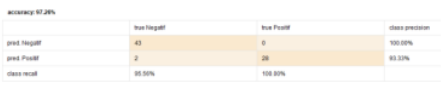


Figure 12. Accuracy Results

Figure 12. shows the results of the accuracy of the modeling results. The results of the classification of public sentiment about Indonesia's U-20 world cup failure using the *naïve bayes classifier* obtained an accuracy of 97.26% which shows that the model built has a good ability to classify public sentiment. This shows that the Naive Bayes *classifier* model is able to recognize well whether a text contains positive or negative sentiments related to Indonesia's failure in the tournament.

Figure 13. Precision result

Figure 13 shows the results of the precision on the modeling results. The results of the classification of public sentiment about the failure of the Indonesian U-20 world cup using the naïve bayes classifier obtained a precision of 93.33% which shows that of all the positive classification results, 93.33% of them are truly positive.

Figure 14. Recall Results

Figure 14 shows the results of recall on the modeling results. The results of the classification of public sentiment about the failure of the Indonesian U-20 world cup using the naïve bayes classifier obtained a Recall of 100% which shows that the model is able to recognize negative sentiment well with the model's ability to recognize negative sentiment of 100%.

4. CONCLUSION

Based on the results of data collection, 214 tweet data were obtained regarding public responses to the failure of the Indonesian U-20 football team in the World Cup from March 27 to April 1, 2023. From the results of data analysis obtained information that there are more negative sentiments than positive sentiments which indicate that public responses regarding the failure of the Indonesian U-20 soccer team in the World Cup tend to be negative seen from the large amount of negative sentiment data in the data. As for the research conducted, it was also obtained information that some people were disappointed with Indonesia's failure to host the U-20 World Cup which the community believed was caused by Ganjar Pranowo's response who refused the State of Israel to compete in the match. This can be seen from the 8 words that often appear in sentiment data regarding the failure of the U-20 world cup in Indonesia, namely the words world, cup, indonesia, pildun, failed, ganjar, israel and disappointed. Then the results of data classification using naïve bayes show good results, namely accuracy of 97.26%, precission of 93.33% and Recall of 100% so it can be concluded that the classification model obtained can classify well the public sentiment related to Indonesia's failure in the U-20 World Cup. The results of this study are expected to be useful in making decisions regarding communication strategies and actions that can be taken by related parties in dealing with similar situations in the future.

Suggestions for future research include using other techniques or methods besides Naive Bayes classifier to compare performance and accuracy between different models. In addition, future research can expand the range of sentiment analysis on other topics by using a larger dataset.

REFERENCES

[1] F. N. Hasan and M. Dwijayanti, "Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinominal Naïve Bayes Classifier," *J. Linguist. Komputasional*, vol. 4, no. 2, pp. 52–58, 2021, doi: https://doi.org/10.26418/jlk.v4i2.61.

[2] A. L. Fairuz, R. D. Ramadhani, and N. A. F. Tanjung, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter," *J. Dinda Data Sci. Inf. Technol. Data Anal.*, vol. 1, no. 1, pp. 42–51, 2021, doi: 10.20895/dinda.v1i1.180.

[3] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 3, p. 279, 2019, doi: 10.26418/jp.v5i3.34368.

[4] E. Nofianto, Fitriyah, and Supratiwi, "Media Sosial sebagai Sarana Pendidikan Politik oleh Pejabat Publik (Studi pada Akun Media Sosial Nur Hidayat Sardini)," *J. Ilm.*, vol. 23, no. 1, pp. 855–864, 2023, doi: 10.33087/jiubj.v23i1.3060.

[5] A. Rismawan, "Dikritik Gara-gara Batal Piala Dunia U-20 di Indonesia, Ganjar Pranowo : Ini Bukan Kiamat." https://bandung.viva.co.id/bola/17997-dikritik-gara-gara-batal-piala-dunia-u-20-di-indonesia-ganjar-pranowo-ini-bukan-kiamat.

[6] B. Baskoro, "Paradoks Indonesia Saat Batal Jadi Tuan Rumah Piala Dunia U-20." https://sport.detik.com/sepakbola/liga-indonesia/d-6651105/paradoks-indonesia-saat-batal-jadi-tuan-rumah-piala-dunia-u-20-2023.

[7] Alfandi Safira and F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier," *Zo. J. Sist. Inf.*, vol. 5, no. 1, pp. 59–70, 2023, doi: 10.31849/zn.v5i1.12856.

[8] I. R. Afandi, I. F. Hanif, F. N. Hasan, E. Sinduningrum, Z. Halim, and N. Pratiwi, "Analisis Sentimen Opini Masyarakat Terkait Penyelenggaraan Sistem Elektronik Menggunakan Metode Logistic Regression," *J. Linguist. Komputasional*, vol. 5, no. 2, pp. 77–84, 2022, [Online]. Available: https://inacl.id/journal/index.php/jlk/article/view/103

[9] M. Siddik, H. Hendri, R. N. Putri, Y. Desnelita, and G. Gustientiedina, "Klasifikasi Kepuasan Mahasiswa Terhadap Pelayanan Perguruan Tinggi Menggunakan Algoritma Naïve Bayes," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 3, no. 2, pp. 162–166, 2020, doi: 10.31539/intecoms.v3i2.1654.

[10] R. Sari and R. Y. Hayuningtyas, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website," *Indones. J. Softw. Eng.*, vol. 5, no. 2, pp. 51–60, 2019, doi: 10.31294/ijse.v5i2.6957.

[11] E. Indrayuni, "Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes," *J. Khatulistiwa Inform.*, vol. 7, no. 1, pp. 29–36, 2019, doi: 10.31294/jki.v7i1.5740.

[12] L. Hermawan and M. Bellaniar Ismiati, "Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval," *J. Transform.*, vol. 17, no. 2, p. 188, 2020, doi: 10.26623/transformatika.v17i2.1705.

[13] M. Syarifuddinn, "Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn," *INTI Nusa Mandiri*, vol. 15, no. 1, pp. 23–28, 2020, doi: 10.33480/inti.v15i1.1347.

[14] A. Muhammad Iqbal, G. Dudih, and S. Falentino, "Analisis Sentiment Masyarakat terhadap Kasus Covid-19 pada Media Sosial Youtube dengan Metode Naive bayes," *J. Sains Komput. Inform. (J-SAKTI*, vol. 5, no. 2, pp. 807–814, 2021.

[15] M. A. Rofiqi, A. C. Fauzan, A. P. Agustin, and A. A. Saputra, "Implementasi Term-Frequency Inverse Document Frequency (TF-IDF) Untuk Mencari Relevansi Dokumen Berdasarkan Query," *Ilk. J. Comput. Sci. Appl. Informatics*, vol. 1, no. 2, pp. 58–64, 2019, doi: 10.28926/ilkomnika.v1i2.18.

[16] A. H. Anshor and A. Safuwan, "Analisis Sentimen Opini Warganet Twitter Terhadap Tes Screening Genose Pendeteksi Virus Covid-19 Menggunakan Metode Naïve Bayes Berbasis Particle Swarm Optimization," *J. Inform. Teknol. dan Sains*, vol. 5, no. 1, pp. 170–178, 2023, doi: 10.51401/jinteks.v5i1.2229.

[17] I. R. Afandi, F. N. Hasan, A. A. Rizki, N. Pratiwi, and Z. Halim, "Analisis Sentimen Opini Masyarakat Terkait Pelayanan Jasa Ekspedisi Anteraja Dengan Metode Naive Bayes," *J. Linguist. Komputasional*, vol. 5, no. 2, pp. 63–70, 2022, [Online]. Available: https://t.co/2HAdwg1drL

[18] 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.

[19] and U. P. A. N. Kasanah, M. Muladi, "Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN," *Masa Berlaku Mulai*, vol. 1, no. 3, pp. 196–201, 2017.

[20] R. L. Hasanah, M. Hasan, W. E. Pangesti, F. F. Wati, and W. Gata, "Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode Knn (K-Nearest Neighbor)," *J. Techno Nusa Mandiri*, vol. 16, no. 1, pp. 1–6, 2019, doi: 10.33480/techno.v16i1.25.

[21] A. Nugroho, "Analisa Splitting Criteria Pada Decision Tree dan Random Forest untuk Klasifikasi Evaluasi Kendaraan," *JSITIK J. Sist. Inf. dan Teknol. Inf. Komput.*, vol. 1, no. 1, pp. 41–49, 2022, doi: 10.53624/jsitik.v1i1.154.

Zaini, Sari, Hasan - cek.03

ORIGINALITY REPORT

11%	6%	8%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Dede Kurniadi, Fitri Nuraeni, Nia Abania, Leni Fitriani, Asri Mulyani, Yoga Handoko Agustin. "Scholarship Recipients Prediction Model using k-Nearest Neighbor Algorithm and Synthetic Minority Over-sampling Technique", 2022 12th International Conference on System Engineering and Technology (ICSET), 2022 Publication	1%
2	Aghus Sofwan, Imam Santoso, Himawan Pradipta, M. Arfan, Ajub Ajulian Zahra M. "Normal and Murmur Heart Sound Classification Using Linear Predictive Coding and k-Nearest Neighbor Methods", 2019 3rd International Conference on Informatics and Computational Sciences (ICICoS), 2019 Publication	1%
3	www.researchgate.net Internet Source	1%
4	time.com Internet Source	1%

5	Febbi Sena Lestari, Harliana Harliana, Muhamat Maariful Huda, Tito Prabowo. "Sentiment Analysis of iPusnas Application Reviews on Google Play Using Support Vector Machine", Proceedings of the International Seminar on Business, Education and Science, 2022 Publication	1 %
6	Sulthan Rafif, Pramana Yoga Saputra, Moch Zawaruddin Abdullah. "Classification of Trends in Lecturer Research Fields Using Naive Bayes Method", 2021 International Conference on Electrical and Information Technology (IEIT), 2021 Publication	1 %
7	Submitted to Universitas Lancang Kuning Student Paper	1 %
8	seaninstitute.org Internet Source	1 %
9	republik-online.com Internet Source	<1 %
10	ejournal.nusamandiri.ac.id Internet Source	<1 %
11	journal.uinjkt.ac.id Internet Source	<1 %

12	Asrul Sani, Samuel, Djaka Suryadi, Firman Noor Hasan, Ade Davy Wiranata, Siti Aisyah. "Predicting the Success of Garment Sales on Transaction Data using the Classification Method with the Naïve Bayes Algorithm", 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), 2023 Publication	<1 %
13	Nur Iksan, Djoko Adi Widodo, Budi Sunarko, Erika Devi Udayanti, Etika Kartikadharma. "Sentiment Analysis of Public Reaction to COVID19 in Twitter Media using Naïve Bayes Classifier", 2021 IEEE International Conference on Health, Instrumentation & Measurement, and Natural Sciences (InHeNce), 2021 Publication	<1 %
14	repository.unika.ac.id Internet Source	<1 %
15	eprints.intimal.edu.my Internet Source	<1 %
16	jurnal.una.ac.id Internet Source	<1 %
17	Manaarul Hidayat, Rahmat Hidayat, Dwi Otik Kurniawati. "Comparison of The Use of Bigrams and Stopword Removal for	<1 %

Classification Using Naive Bayes (Case Study on Sentiment Analysis of By.U Internet Users)", 2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM), 2021
Publication

---

18 Nurfadillah Tanjung, Deci Irmayani, Volvo Sihombing. "Implementation of C5.0 Algorithm for Prediction of Student Learning Graduation in Computer System Architecture Subjects", Sinkron, 2022 <1%
Publication

---

19 Praise Setiawan Saragih, Deden Witarsyah, Faqih Hamami, Jose Manuel Machado. "Sentiment Analysis of Social Media Twitter with Case of Large Scale Social Restriction in Jakarta using Support Vector Machine Algorithm", 2021 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS), 2021 <1%
Publication

---

20 prosiding-icostec.respati.ac.id <1%
Internet Source

---

21 Anggit Dwi Hartanto, Ema Utami, Sumarni Adi, Harish Setyo Hudnanto. "Job Seeker Profile Classification of Twitter Data Using the <1%

Naïve Bayes Classifier Algorithm Based on the DISC Method", 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2019
Publication

---

22 Irsyad Fakhrur Roji, Sari Widya Sihwi, Afrizal Doewes. "Development of Regional Public Election Analytics Application In Indonesia Using Data from Twitter", 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2018
Publication

<1 %

---

23 Mussalimun, Elvien Hastatomo Khasby, Gempita Icky Dzikrillah, Muljono. "Comparison of K- N earest Neighbor (K -NN) and Naïve Bayes Algorithm for Sentiment Analysis on Google Play Store Textual Reviews", 2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE), 2021
Publication

<1 %

---

24 Steven Johan, Friska Natalia, Ferry Vincenttius Ferdinand, Sud Sudirman. "Image-Based Skin Cancer Early Detection using CNN Algorithm", 2021 6th International Conference on New Media Studies (CONMEDIA), 2021
Publication

<1 %

25	jurnal.unimed.ac.id Internet Source	<1%
26	ojs.stmikpringsewu.ac.id Internet Source	<1%
27	www.ncbi.nlm.nih.gov Internet Source	<1%
28	Abdurrahim Abdurrahim, Lailis Syafa'ah, Merinda Lestandy. "Sentiment analysis of Covid-19 vaccine tweets utilizing Naïve Bayes", AIP Publishing, 2022 Publication	<1%
29	Dede Kurniadi, Shopi Nurhidayanti, Indri Tri Julianto, Teguh Wahyono, Yosep Septiana, Hetty Rohayani. "Classification of Television Programs Based on Public Opinion in Social Media Using Random Forest and Decision Tree", 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), 2023 Publication	<1%
30	Fajar Sidik, Ibnu Suhada, Azhar Haikal Anwar, Firman Noor Hasan. "Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier", Jurnal Linguistik Komputasional (JLK), 2022 Publication	<1%

31	Hiya Nalatissifa, Hilman Ferdinandus Pardede. "Customer Decision Prediction Using Deep Neural Network on Telco Customer Churn Data", Jurnal Elektronika dan Telekomunikasi, 2021 Publication	<1 %
32	Submitted to Nanyang Technological University, Singapore Student Paper	<1 %
33	Nidaul Hasanati, Qurrotul Aini, Arndini Nuri. "Implementation of Support Vector Machine with Lexicon Based for Sentiment Analysis on Twitter", 2022 10th International Conference on Cyber and IT Service Management (CITSM), 2022 Publication	<1 %
34	Vivine Nurcahyawati, Zuriani Mustaffa. "Feature Selection based on Particle Swarm Optimization Algorithm for Sentiment Analysis Classification", 2021 International Conference on Intelligent Technology, System and Service for Internet of Everything (ITSS-IoE), 2021 Publication	<1 %
35	Yohanes Alvinika, Wahyu Prasetyo, Paulus Mudjihartono. "Chapter 2 Analysis of Twitter User's Sentiment Against COVID-19 Vaccination Using the Lexicon Based	<1 %

Method", Springer Science and Business Media LLC, 2022
Publication

36	jutif.if.unsoed.ac.id Internet Source	<1%
37	repository.uin-suska.ac.id Internet Source	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		