Multitek Indonesia: Jurnal Ilmiah
Volume: 16 No. 2, Desember 2022, Hal. 75 - 89
ISSN: 1907-6223 (print)
ISSN: 2579-3497 (Online)
http://journal.umpo.ac.id/index.php/multitek

ANALYSIS SENTIMENTS IN FACEBOOK DOWN CASE USING VADER AND NAIVE BAYES CLASSIFICATION METHOD

Ilham Firman Ashari

Teknik Informatika, Institut Teknologi Sumatera
Jl Terusan Ryacudu, Lampung Selatan
Email : firman.ashari@if.itera.ac.id

Received: July 19, 2022. Accepted: January 03, 2023

Abstrak

Facebook adalah media sosial terbesar di dunia. Semua aplikasi media sosial besutan Facebook tidak bisa diakses secara bersamaan dalam waktu kurang lebih 6 jam. Hal ini tidak hanya terjadi di Indonesia, tetapi di seluruh negara di dunia, pada tanggal 5-6 Oktober waktu Indonesia. Dengan adanya kasus ini, berbagai komentar dan opini dari masyarakat di Twitter terkait kasus Facebook pun turun. Komentar positif atau negatif bermunculan di twitter. Analisis sentimen digunakan untuk mengidentifikasi komentar positif dan negatif. Pada penelitian ini, komentar positif dan negatif akan diklasifikasikan menggunakan klasifikasi Vader dan nave bayes. Data yang terkumpul sebanyak 500 data dari twitter terkait down case facebook. Dari hasil perhitungan diperoleh sentimen positif sebanyak 33,92% dan sentimen negatif dengan hasil 66,08%. Berdasarkan hasil visualisasi dengan wordcloud, kata yang paling banyak muncul adalah kata facebook down untuk sentimen positif dan negatif. Hasil yang didapatkan dari tabel Confusion Matrix dari model klasifikasi menggunakan data sharing, 80% data training dan 20% data testing, dengan metode klasifikasi menggunakan Naive Bayes dengan pembobotan kata TF-IDF, nilai akurasinya sebesar 73,69% dan untuk Count Vektorizer adalah 70,18%.

Kata Kunci: Analisis Sentimen, Facebook Down, Naïve Bayes, Vader

Abstract

Facebook is the largest social media in the world. All social media applications made by Facebook cannot be accessed simultaneously in approximately 6 hours. This happens not only in Indonesia, but in all countries in the world, on October 5-6, Indonesian time. With this case, various comments and opinios from people on Twitter related to the Facebook case were down. Positive or negative comments popping up on twitter. Sentiment analysis is used to identify positive and negative comments. In this study, positive and negative comments will be classified using Vader and nave Bayes classification. The data collected was 500 data from twitter related to the Facebook down case. From the calculation results, positive sentiment was obtained as much as 33.92% and negative sentiment with 66.08% results. Based on the results of the visualization with wordcloud, the words that appear the most are the word facebook down for positive and negative sentiments. The results obtained from the confusion matrix table from the classification model using data sharing, 80% training data and 20% testing data, with the classification method using Naive Bayes with TF-IDF word weighting, the accuracy value is 73.69% and for the Count Vectorizer is 70.18%.

Keyword: Analysis Sentiment, Facebook Down, Naïve Bayes, Vader

ISSN: 1907-6223 (Print)
ISSN:2579-3497 (Online)

INTRODUCTION

The use of technology is currently growing. This is certainly inseparable from the development of the internet. One of the technological developments that we feel the most at this time is the development of social media applications, especially during the pandemic, social media has become a close friend who is always there [1]. Social media is used to get the latest information, the information obtained can be about entertainment, news, or just sharing information with close relatives and friends [2][3]. Social media is an online media that can be used for long-distance communication between one person and several people in real-time without being limited in time [4].

From research conducted by [5], where to make a classification for social media. Social media can be categorized as collaboration, for example like Wikipedia or Medium. Then for social media content categories, for example Instagram, TikTok, and youtube.com. Social media with social networking categories, for example Facebook and Instagram. Social media with blog or microblog categories, for example Twitter. Bookmarking-based social media, for example, such as dig.com and reddit.com.

The biggest social media that are still used today to help facilitate communication are Facebook, WhatsApp, and Instagram [6][7]. According to data from statita.com quoted from research [8], where the number of Facebook users worldwide in 2018 reached 2.32 billion users. This will certainly increase with the development of time and the increase in the number of people in the world. Of the world's population of more than 7 billion quoted from research journals conducted by [9]. This shows that about 28% of the world's population uses Facebook. With this it can be said that Facebook plays an important role in helping facilitate communication for everyone in the world.

Recently, there was a case of Facebook going down, precisely from October 5 before midnight to October 6 local Indonesian time in 2021. All social media applications made by Facebook, such as Facebook, Instagram, and WhatsApp, cannot be accessed simultaneously in approximately 6 hours. This happens not only in Indonesia, but in all countries in the world. The blackout of this application from Facebook certainly makes an impact and loss not only for Facebook users but also for the Facebook company itself. Information from Facebook stated that the problem occurred due to a configuration error in the Facebook DNS [10]. The blackout happened for quite a long time, making people at that time switch to the Twitter social media application. From direct observation, that on the date Facebook was down. On social media, Twitter, this case had become a trending topic. Various comments and sentiments were conveyed by Facebook users, there were those who submitted complaints and there were those who still understood this case. this is getting culminated when Facebook can never be up again soon.

Measurement of sentiment or comments can be done using sentiment analysis. Several methods can be used to classify sentiments, including the nave Bayes method, k-means, random forest, this is like the research conducted by [11][12][13]. Research conducted by [14] related to sentiment analysis of comments on Facebook groups using nave Bayes, which uses the TF-IDF feature selection, using a ratio of 80:20. The results obtained are 75% accuracy. Another research conducted Back by [15] related to sentiment analysis from jd.id store customers using the nave Bayes classifier method, using the TF-IDF weighting feature and adding emoticon extraction. The results obtained are 98% accuracy. From research conducted by [16], related to sentiment analysis using the Naive Bayes method of film and reviews ratings. The classification accuracy level obtained with features based on stopword removal is 0.9 and the classification accuracy level is using lexicon sentiment with a result of 1 or 100%. Other research conducted by [17], where the research conducted is to conduct sentiment analysis of public comments on the relocation of the Indonesian capital. The algorithm used is nave bayes and Vader sentiment. The accuracy results obtained are approximately 76.40%. By using the sentiment Vader, a lexicon dictionary is available that contains the value of each word. The Lexicon database can be used to assess the sentiment of phrases and sentences.

From some of the previous descriptions, the author makes research using sentiment analysis using Vader and nave Bayes classification methods to assess comments from netizens from social media Twitter in the case of Facebook down. This of course can also be a prediction if a similar case occurs in the future. These negative comments from netizens should also be a concern for Facebook to be able to minimize this incident.

RESEARCH METHOD

This research has a workflow as shown in Figure 1. The initial stages are crawling data, then pre-processing data, labeling data, word weighting, nave Bayes classification, and evaluation with a confusion matrix.

Data Crawling

In this study, the data used is sentiment about Facebook down. The data comes from Twitter using the Search API using the Python programming language [18]. Data is an important aspect for the pre-processing process [19][20]. The data taken ranged from the Facebook down incident, around 4 -5 October 2021 with the keyword facebookdown. The crawled data is saved in CSV format to make it easy for further processing. The datasets taken for research are username, date, and tweet. The data taken with the provisions of 500 data.

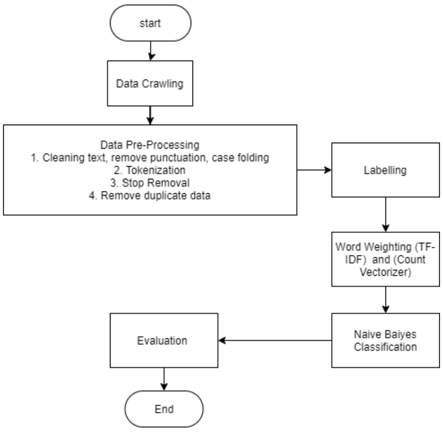

Figure 1. Sentiment Analysis Flow

Preprocessing Data

Preprocessing data is the stage of changing the original text to be used as input and implementing routines in changing and also eliminating elements that are not useful in later processing [21][22]. The dataset obtained from the crawling process is still not structured and has several components that are not needed in the process; therefore, a preprocessing process is needed before the data is processed [23]. This is done so that the data processed is good data and it is hoped that the accuracy will be better. In this sentiment analysis research, the preprocessing stages carried out are:

1. *Cleaning text, remove punctuation, case folding*

The data that has been obtained from Twitter will be cleaned of components that have nothing to do with the information in the document. Examples such as links, html, scripts, and so forth. This stage will also remove punctuation marks, symbols, URLs, numbers and change the letters contained in the document to lowercase. The results of the cleansing text can be seen in Figure 2. Case folding changes all sentences into uppercase and lowercase letters. In this study, all sentences will be converted into lowercase letters. The results of the case folding process can be seen in Figure 3.

ISSN: 1907-6223 (Print)
ISSN:2579-3497 (Online)

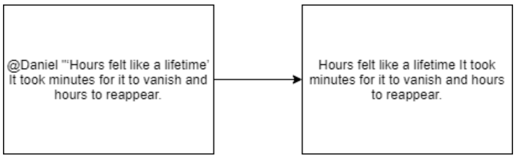

Figure 2. Cleaning Text

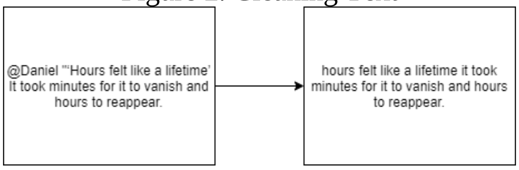

Figure 3. Case Folding

2. *Stopword removal*

Tokenization has the goal of dividing the text in a paragraph or sentence into certain parts [24]. The separator at the tokenization stage is punctuation or spaces. The results of tokenization are useful for analysis in the next process. An example of tokenization results can be seen in Figure 4.

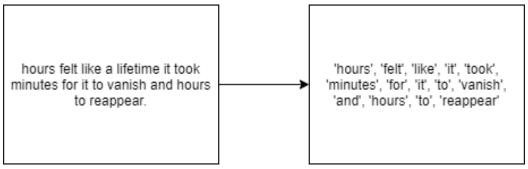

Figure 4. Tokenization

3. *Stopword removal*

Stopword removal is a process that is carried out after tokenization by removing meaningless words. One of them is conjunctions and prepositions which will be removed at this stage [25]. Examples of words are "of", "the", "an", "a", and so on.

4. *Remove duplicate data*

The last stage in this research is Remove duplicate data. Remove duplicate data is to remove or eliminate duplicate data or data that appears more than once.

A. Labelling Data

Labeling has the root word label, which means that identifying a variable requires commonly used characters. In previous research, in general, the data labeling stage will be carried out, in which sentiment will be divided into categories, namely positive, neutral, and negative [26]. But in this study, labeling data only uses positive and negative categories. Labeling is done using Vader sentiment by using a match with a lexicon dictionary [27]. In labeling, a lexicon database is used, where later each word will be compared, and the value of the words included in the positive, neutral, and negative sentiments will be calculated. In labeling for sentiment analysis, the NLTK package and the Vader Lexicon module are used, which contains a set of words with predetermined and classified values. Each score obtained for the classification of negative, positive, and neutral will be combined to produce a compound value. Compound is a value that has been normalized with a range of -1 to +1. If it is below 0, then it is declared as a negative classification. If it is equal to 0, then it is declared a neutral classification, and if it is greater than 0, it is a positive classification. The process of labeling data using Vader can be seen in Figure 5.

ISSN: 1907-6223 (Print)
ISSN:2579-3497 (Online)

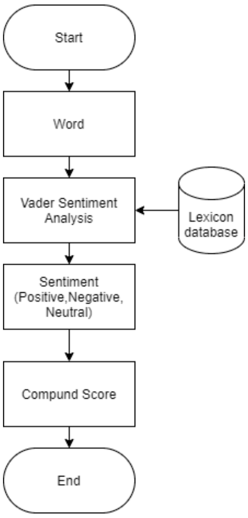

Figure 5. Vader Sentimen Process

B. Wordcloud

Wordcloud is part of the visualization mining technique, to display the most frequently occurring words or the most popular words [28]. Wordcloud in python uses the wordcloud library which needs to be installed first.

C. Word Weighting

Word weighting or also included in the feature selection process based on the weight contained in the document. The weight value becomes a measure in seeing the distribution of a word in determining the class or category in the document [29]. In this study, there are 2-word weights used, namely TF-IDF and Count Vectorizer.

1. Weighting with TF-IDF

The next stage is the TF-IDF weighting which is used to evaluate how important a word is in a document. Term Frequency or (TF) the more occurrences of words contained in the document, the higher the weight value will be. Meanwhile, the Inverse Document Frequency (IDF) process is the opposite of the TF process [30]. In IDF, the greater the value of the frequency of the words that appear, it will result in the value of the resulting weight will be small [30]. In IDF, the higher the frequency of occurrence of the term, the smaller the weight value of the term itself.

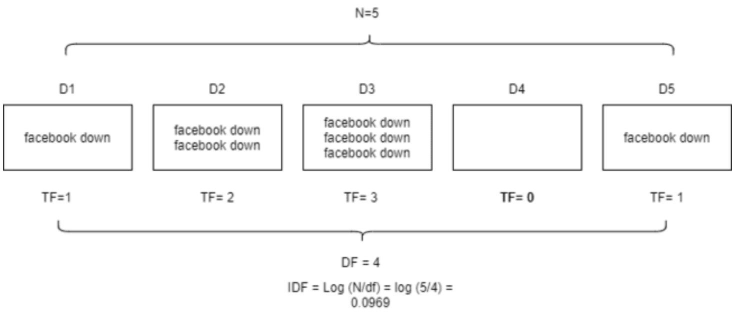

Figure 6. Vader Sentiment Process

Information:

D1…D5 = Document

TF = The number of words from each document

N = The Number of documents that will be compared

DF = The number of documents that have the word is looking for

ISSN: 1907-6223 (Print)
ISSN:2579-3497 (Online)

With the weighting formula as follows (1) and (2) [31].

$$W_{t,d} = tf_{t,d} * idf_{t,d} \quad (1)$$

$$= tf_{t,d} * \log\left(\frac{N}{dft}\right) \quad (2)$$

Information:

$W_{t,d}$ =TF-IDF weight

$tf_{t,d}$ =The number of word frequencies

$idf_{t,d}$ =The number of document frequency inverses per word

dft = The number of document frequencies per word

N = The number of documents

The Scikit-Learn (SK-Learn) library is one of the libraries in python that is used to assist the TF-IDF weighting process.

2. Count Vectorizer Weighting

The count vectorize weighting is used to calculate the frequency that data appears contained in the document. With this weighting feature, it will change the features contained in the text into a vector representation [32]. For example:

a. Ilham makan nasi padang
b. Doni makan bubur ayam

From these sentences or what we can call the corpus, a dictionary can be compiled consisting of:

Ilham, makan, nasi, doni, padang, ayam

Where the vector representation of the existing data is 6 vectors (for each word). Where each vector element shows the number of word features. Example:

Ilham makan nasi padang: [1, 1, 1, 0, 1, 0]

D. Naïve Bayes

Naive Bayes is a method using probability and statistics with the assumption of independence based on Bayes' theorem [33][34]. Equation (3) is the equation for the nave Bayes classification.

$$P(H|S) = \frac{P(S|H)P(H)}{P(S)} \quad (3)$$

Information:

H : *Class of unknown data*

S : Hypothesis data is class specific

$P\,(H|S)$: The probability hypothesis H from conditions of S

$P\,(H)$: Hypothesis from probability H

$P(S|H)$: The probability hypothesis S from conditions of H

$P\,(X)$: Hypothesis probability X

The data that will be processed is divided into training and testing data.

a. Training Process

In this process, training data is carried out with known labels. The purpose of this process is to build a model that will be used in the testing process

b. Testing Process

This stage is carried out to test the accuracy of the model built in the training process. The data used is called a test set and the model will predict the label that has been given based on the training process that has been carried out.

E. Confusion Matrix

After performing the Naïve Bayes classification stage, the next step is to calculate the level of accuracy using the Confusion Matrix [35]. The confusion matrix can be seen in Figure 7.

ISSN: 1907-6223 (Print)
ISSN:2579-3497 (Online)

Predicted Values		Actual Values: Positive (1)	Actual Values: Negative (0)
	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figure 7. Confusion Matrix

Confusion Matrix can be used to calculate the value of accuracy, furthermore recall, and F-1 score [35]:

a. Accuracy: Indicates a value related to the percentage of the number of data that is correctly predicted to the overall data. The formula can be seen in the equation (4).

$$\frac{Tp+Tn}{Tp+Fp+Fn+Tn} \quad (4)$$

b. Precision: shows a value related to the proportion of data predicted by the model with a positive label from the overall data that indeed the data has a positive label. The formula can be seen in the equation (5).

$$\frac{Tp}{Tp+Fp} \quad (5)$$

c. Recall: shows a value related to the proportion of data predicted by the positive label model of all data that is indeed positively label. The formula can be seen in the equation (6).

$$\frac{Tp}{Tp+Fn} \quad (6)$$

d. F-1 Score: Shows values related to precision and recall by taking the average value. The formula is in equation (7).

$$(2 * Recall * Precision) / (Recall + Precision) \quad (7)$$

Information:

TP = Positive data predicted by the engine correctly (positive)
TN = Negative data predicted by the engine correctly (negative)
FP = Negative data predicted by the engine incorrectly (positive)
FN = Positive data that the engine predicts incorrectly (negative)

RESULT AND DISCUSSION

This sentiment analysis research on data retrieval processing using twitter crawling using the tweepy and pandas libraries, obtained 500 data. The data that has been obtained is then stored in .csv format. The next step after getting the data is to delete null data. Where this data has no value that can be processed and is not needed. The results of the process can be seen in Figure 8.

	tweet
0	RT @danyork: Good piece explaining #BGP and wh...
1	RT @Anubha26358265: It is as if the earth has ...
2	RT @RaunakJr: Whatsapp,insta , Facebook down\n...
3	RT @RubraRuby: Hey, We're giving away $200 to ...
4	RT @omarsuleiman504: Hope everyone is having a...
...	...
495	RT @DxbAwan: All People coming on Twitter to c...
496	RT @LaLaBemo: Sad Day 🤐 💔 #ไอจีล่ม #เฟซล่ม #เฟสบ...
497	RT @ChrisGaskill: Is #facebookdown again?\n\nT...
498	RT @PrUkkera: Twitter reaction after #whatsapp...
499	#Facebookdown for Me Right Now #Houston #Weather

500 rows × 1 columns

Figure 8. Pre-Processing Eliminate null data

ISSN: 1907-6223 (Print)
ISSN:2579-3497 (Online)

The next step is data preprocessing which consists of cleaning text, remove punctuation, case folding, tokenization, stop removal, and remove duplicate data. Below are the results of the data preprocessing stage. Where there are 2 main columns, namely tweets and cleaning tweets. Where tweet is the original data from twitter and cleaning tweets is data that has been preprocessed data. The results can be seen in Figure 9. The next step after cleaning text, removing punctuation, and case folding is the tokenization process. The tokenization process is carried out to sort out words from each sentence. Where the results of the tokenization process can be seen in Figure 10.

.

	tweet	cleaning_tweets
0	RT @danyork: Good piece explaining #BGP and wh...	good piece explaining bgp and what happened f...
1	RT @Anubha26358265: It is as if the earth has ...	the earth has trembled facebook and instagram...
2	RT @RaunakJr: Whatsapp,insta , Facebook down\n...	whatsapp insta facebook down twitter its busi...
3	RT @RubraRuby: Hey, We're giving away $200 to ...	hey giving away help people out there who nee...
4	RT @omarsuleiman504: Hope everyone is having a...	hope everyone having productive day instagram...
...	...	...
495	RT @DxbAwan: All People coming on Twitter to c...	all people coming twitter check what happened...
496	RT @LaLaBemo: Sad Day 😢 💔 #ไอจีล่ม #เฟซล่ม #เฟสบ...	sad day facebook facebookdown
497	RT @ChrisGaskill: Is #facebookdown again?\n\nT...	facebookdown again twitter about become more ...
498	RT @PrUkkera: Twitter reaction after #whatsapp...	twitter reaction after whatsappdown facebookd...
499	#Facebookdown for Me Right Now #Houston #Weather	facebookdown for right now houston weather

500 rows × 2 columns

Figure 9. The results of the pre-processing cleaning text, remove punctuation, case folding

	tweet	cleaning_tweets	Tokenization
0	RT @danyork: Good piece explaining #BGP and wh...	good piece explaining bgp and what happened f...	[, good, piece, explaining, bgp, and, what, ha...
1	RT @Anubha26358265: It is as if the earth has ...	the earth has trembled facebook and instagram...	[, the, earth, has, trembled, facebook, and, i...
2	RT @RaunakJr: Whatsapp,insta , Facebook down\n...	whatsapp insta facebook down twitter its busi...	[, whatsapp, insta, facebook, down, twitter, i...
3	RT @RubraRuby: Hey, We're giving away $200 to ...	hey giving away help people out there who nee...	[, hey, giving, away, help, people, out, there...
4	RT @omarsuleiman504: Hope everyone is having a...	hope everyone having productive day instagram...	[, hope, everyone, having, productive, day, in...
...	...	...	...
495	RT @DxbAwan: All People coming on Twitter to c...	all people coming twitter check what happened...	[, all, people, coming, twitter, check, what, ...
496	RT @LaLaBemo: Sad Day 😢 💔 #ไอจีล่ม #เฟซล่ม #เฟสบ...	sad day facebook facebookdown	[, sad, day, facebook, facebookdown,]
497	RT @ChrisGaskill: Is #facebookdown again?\n\nT...	facebookdown again twitter about become more ...	[, facebookdown, again, twitter, about, become...
498	RT @PrUkkera: Twitter reaction after #whatsapp...	twitter reaction after whatsappdown facebookd...	[, twitter, reaction, after, whatsappdown, fac...
499	#Facebookdown for Me Right Now #Houston #Weather	facebookdown for right now houston weather	[, facebookdown, for, right, now, houston, wea...

500 rows × 3 columns

Figure 10. Result of Pre-Processing Stop Removal

After that the step taken is stop removal. Where at this stage removes words that are not important from the tweet. The results of the stopword removal process can be seen in Figure 11.

	tweet	cleaning_tweets	Tokenization	Stop_removal
0	RT @danyork: Good piece explaining #BGP and wh...	good piece explaining bgp and what happened f...	[, good, piece, explaining, bgp, and, what, ha...	[, good, piece, explaining, bgp, happened, fac...
1	RT @Anubha26358265: It is as if the earth has ...	the earth has trembled facebook and instagram...	[, the, earth, has, trembled, facebook, and, i...	[, earth, trembled, facebook, instagram, faceb...
2	RT @RaunakJr: Whatsapp,insta , Facebook down\n...	whatsapp insta facebook down twitter its busi...	[, whatsapp, insta, facebook, down, twitter, i...	[, whatsapp, insta, facebook, twitter, busines...
3	RT @RubraRuby: Hey, We're giving away $200 to ...	hey giving away help people out there who nee...	[, hey, giving, away, help, people, out, there...	[, hey, giving, away, help, people, needs, als...
4	RT @omarsuleiman504: Hope everyone is having a...	hope everyone having productive day instagram...	[, hope, everyone, having, productive, day, in...	[, hope, everyone, productive, day, instagramd...
...	...	...	...	...
495	RT @DxbAwan: All People coming on Twitter to c...	all people coming twitter check what happened...	[, all, people, coming, twitter, check, what, ...	[, people, coming, twitter, check, happened, w...
496	RT @LaLaBemo: Sad Day 😢 💔 #ไอจีล่ม #เฟซล่ม #เฟสบ...	sad day facebook facebookdown	[, sad, day, facebook, facebookdown,]	[, sad, day, facebook, facebookdown,]
497	RT @ChrisGaskill: Is #facebookdown again?\n\nT...	facebookdown again twitter about become more ...	[, facebookdown, again, twitter, about, become...	[, facebookdown, twitter, become, dumpster, fi...
498	RT @PrUkkera: Twitter reaction after #whatsapp...	twitter reaction after whatsappdown facebookd...	[, twitter, reaction, after, whatsappdown, fac...	[, twitter, reaction, whatsappdown, facebookdo...
499	#Facebookdown for Me Right Now #Houston #Weather	facebookdown for right now houston weather	[, facebookdown, for, right, now, houston, wea...	[, facebookdown, right, houston, weather]

500 rows × 4 columns

Figure 11. Result of Pre-Processing Stop Removal

After the stopword removal process is done, the next step is the process to eliminate duplicate data. The results can be seen in Figure 12.

	tweet	cleaning_tweets	Tokenization	Stop_removal	tweet_preprocessing
0	RT @danyork: Good piece explaining #BGP and wh...	good piece explaining bgp and what happened f...	[, good, piece, explaining, bgp, and, what, ha...	good piece explaining bgp happened facebookdo...	good piece explaining bgp happened facebookdo...
1	RT @Anubha26358265: It is as if the earth has ...	the earth has trembled facebook and instagram...	[, the, earth, has, trembled, facebook, and, i...	earth trembled facebook instagram facebookdow...	earth trembled facebook instagram facebookdow...
3	RT @RubraRuby: Hey, We're giving away $200 to ...	hey giving away help people out there who nee...	[, hey, giving, away, help, people, out, there...	hey giving away help people needs also grow p...	hey giving away help people needs also grow p...
4	RT @omarsuleiman504: Hope everyone is having a...	hope everyone having productive day instagram...	[, hope, everyone, having, productive, day, in...	hope everyone productive day instagramdown fa...	hope everyone productive day instagramdown fa...
5	RT @Salman100_DH: .Facebook and Instagram user...	facebook and instagram users coming twitter r...	[, facebook, and, instagram, users, coming, tw...	facebook instagram users coming twitter right...	facebook instagram users coming twitter right...
...	...	...	...	...	...
493	RT @anil_loniya: Whatsapp, Facebook and Instag...	whatsapp facebook and instagram are down now ...	[, whatsapp, facebook, and, instagram, are, do...	whatsapp facebook instagram twitter people li...	whatsapp facebook instagram twitter people li...
495	RT @DxbAwan: All People coming on Twitter to c...	all people coming twitter check what happened...	[, all, people, coming, twitter, check, what, ...	people coming twitter check happened whatsapp...	people coming twitter check happened whatsapp...
496	RT @LaLaBemo: Sad Day🥺💔 #ไอจีล่ม #เฟซล่ม #เฟสบ...	sad day facebook facebookdown	[, sad, day, facebook, facebookdown,]	sad day facebook facebookdown	sad day facebook facebookdown
497	RT @ChrisGaskill: Is #facebookdown again?\n\nT...	facebookdown again twitter about become more ...	[, facebookdown, again, twitter, about, become...	facebookdown twitter become dumpster fire usual	facebookdown twitter become dumpster fire usual
499	#Facebookdown for Me Right Now #Houston #Weather	facebookdown for right now houston weather	[, facebookdown, for, right, now, houston, wea...	facebookdown right houston weather	facebookdown right houston weather

283 rows × 5 columns

Figure 12. Remove Duplicate Data

The data will enter the data labeling stage after passing the preprocessing stage. The data labeling function is used to find out whether sentiment falls into the positive or negative category. Below is the code for labeling data using Vader, it can be seen in Figure 13. The results of labeling can be seen in Figure 14.

```
import nltk
nltk.download('vader_lexicon')
nltk.download('punkt')

from nltk.sentiment.vader import SentimentIntensityAnalyzer
vid = SentimentIntensityAnalyzer()
df['sentiment'] = df['tweet_preprocessing'].apply(lambda x: vid.polarity_scores(x))
def convert(x):
    if x < 0:
        return "negative"
    elif x > 0:
        return "positive"
    else:
        return "neutral"
df['label'] = df['sentiment'].apply(lambda x:convert(x['compound']))
```

Figure 13. Labelling Data Using Vader Lexicon

	tweet	cleaning_tweets	Tokenization	Stop_removal	tweet_preprocessing	sentiment	label
0	RT @danyork: Good piece explaining #BGP and wh...	good piece explaining bgp and what happened f...	[, good, piece, explaining, bgp, and, what, ha...	good piece explaining bgp happened facebookdo...	good piece explaining bgp happened facebookdo...	{'neg': 0.0, 'neu': 0.707, 'pos': 0.293, 'comp...	positive
1	RT @Anubha26358265: It is as if the earth has ...	the earth has trembled facebook and instagram...	[, the, earth, has, trembled, facebook, and, i...	earth trembled facebook instagram facebookdow...	earth trembled facebook instagram facebookdow...	{'neg': 0.231, 'neu': 0.769, 'pos': 0.0, 'comp...	negative
3	RT @RubraRuby: Hey, We're giving away $200 to ...	hey giving away help people out there who nee...	[, hey, giving, away, help, people, out, there...	hey giving away help people needs also grow p...	hey giving away help people needs also grow p...	{'neg': 0.0, 'neu': 0.662, 'pos': 0.338, 'comp...	positive
4	RT @omarsuleiman504: Hope everyone is having a...	hope everyone having productive day instagram...	[, hope, everyone, having, productive, day, in...	hope everyone productive day instagramdown fa...	hope everyone productive day instagramdown fa...	{'neg': 0.0, 'neu': 0.674, 'pos': 0.326, 'comp...	positive
5	RT @Salman100_DH: .Facebook and Instagram user...	facebook and instagram users coming twitter r...	[, facebook, and, instagram, users, coming, tw...	facebook instagram users coming twitter right...	facebook instagram users coming twitter right...	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...	neutral
...	...	...	...	...	...	...	...
493	RT @anil_loniya: Whatsapp, Facebook and Instag...	whatsapp facebook and instagram are down now ...	[, whatsapp, facebook, and, instagram, are, do...	whatsapp facebook instagram twitter people li...	whatsapp facebook instagram twitter people li...	{'neg': 0.0, 'neu': 0.737, 'pos': 0.263, 'comp...	positive
495	RT @DxbAwan: All People coming on Twitter to c...	all people coming twitter check what happened...	[, all, people, coming, twitter, check, what, ...	people coming twitter check happened whatsapp...	people coming twitter check happened whatsapp...	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...	neutral
496	RT @LaLaBemo: Sad Day 😢 💔 #ไอจีล่ม #เฟซล่ม #เฟสบ...	sad day facebook facebookdown	[, sad, day, facebook, facebookdown,]	sad day facebook facebookdown	sad day facebook facebookdown	{'neg': 0.508, 'neu': 0.492, 'pos': 0.0, 'comp...	negative
497	RT @ChrisGaskill: Is #facebookdown again?\n\nT...	facebookdown again twitter about become more ...	[, facebookdown, again, twitter, about, become...	facebookdown twitter become dumpster fire usual	facebookdown twitter become dumpster fire usual	{'neg': 0.5, 'neu': 0.5, 'pos': 0.0, 'compound...	negative
499	#Facebookdown for Me Right Now #Houston #Weather	facebookdown for right now houston weather	[, facebookdown, for, right, now, houston, wea...	facebookdown right houston weather	facebookdown right houston weather	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...	neutral

283 rows × 7 columns

Figure 14. Result of Labeling Data Process Using Vader Lexicon

The next step is to remove neutral sentiment, so that positive sentiment gets a percentage of 33.92 % and negative sentiment gets 66.08 % which means that the community where the datasets are 283 data. The code to delete data with neutral sentiment can be seen in Figure 15 and the plot results for positive or negative sentiment can be seen in Figure 16 .

```
#menghapus data dengan sentimen netral
df = df.drop(df[df.label == "neutral"].index)
print('Dataset size:', df.shape)
print('columns are:', df.columns)
```

Figure 15. Code To Remove Neutral Classification

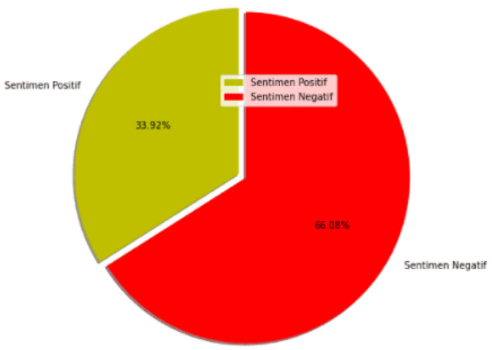

Figure 16. Sentiment percentage distribution based on positive and negative sentiment categories

This study uses wordcloud in conducting a classification based on known categories. Figures 17 is the results from positive sentiments and figure 18 is the result from negative sentiments. The same result was obtained for wordcloud positive and negative sentiments, that's the word facebookdown. These two words can be said to be words that often appear or are popular because they have the largest form than the other words.

ISSN: 1907-6223 (Print)
ISSN:2579-3497 (Online)

Figure 17. Wordcloud Positive Sentiment

Figure 18. Wordcloud Negative Sentiment

Confusion matrix is one way that can be used to measure performance in solving table classification problems. The table is a process of classifying test data whose actual values are known. In this study, the weighting will be compared using the Count Vectorizer with TF-IDF. The weighting process with the count vectorizer can be seen in Figure 19.

```
from sklearn.feature_extraction.text import CountVectorizer
count_vectorizer = CountVectorizer(ngram_range=(1,2))    # Unigram and Bigram
final_vectorized_data = count_vectorizer.fit_transform(X.values.astype(str))
print(final_vectorized_data.toarray())
print("Size of matrix " , final_vectorized_data.shape)
```

```
[[0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 ...
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]]
Shape of matrix  (150, 1970)
```

Figure 19. Count Vectorizer Process and Result

By using TF-IDF weighting, the results can be seen as shown in Figure 20.

```
from sklearn.feature_extraction.text import TfidfVectorizer

tf_idf_vectorizer = TfidfVectorizer(use_idf=True,ngram_range=(1,3))
tf_final_vectorized_data = tf_idf_vectorizer.fit_transform(X.values.astype(str))
tf_final_vectorized_data.shape
print(tf_final_vectorized_data)

```

```
  (0, 1330)	0.21315251176282463
  (0, 1734)	0.24122681345125177
  (0, 316)	0.24122681345125177
  (0, 1017)	0.24122681345125177
  (0, 2947)	0.24122681345125177
  (0, 1672)	0.24122681345125177
  (0, 4204)	0.20411460530497624
  (0, 1326)	0.1596178898948866
  (0, 1733)	0.24122681345125177
  (0, 315)	0.24122681345125177
  (0, 1016)	0.24122681345125177
  (0, 2946)	0.24122681345125177
  (0, 1671)	0.24122681345125177
  (0, 1976)	0.13154358820645948
  (0, 4192)	0.13892809547027357
  (0, 1216)	0.05897257110794269
  (0, 1732)	0.21315251176282463
  (0, 314)	0.24122681345125177
  (0, 1015)	0.24122681345125177
  (0, 2945)	0.24122681345125177
  (0, 1666)	0.21315251176282463
  (1, 3400)	0.25438961978893915
  (1, 1299)	0.25438961978893915
  (1, 1943)	0.25438961978893915
  (1, 1113)	0.2247834129573665
  :	:
  (280, 1216)	0.09824881673751272
```

Figure 20. TF-IDF Process and Result

In Figure 21, the results of the confusion matrix using the count vectorizer produce True Positive: 30, False Positive: 7, True Negative: 10, and False Negative: 10. While the confusion matrix using TF-IDF produces True Positive:40, False Positive:15, True Negative:2, and False Negative:0, the result can be seen in figure 22.

ISSN: 1907-6223 (Print)
ISSN:2579-3497 (Online)

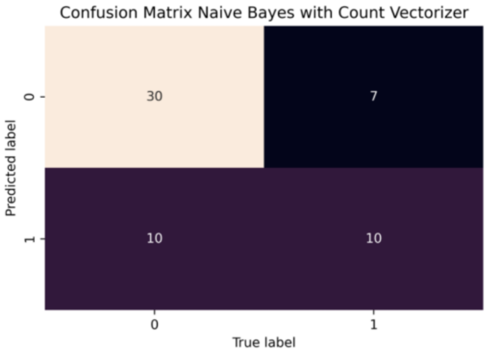

Figure 21. Confusion Matrix using Count Vectorizer

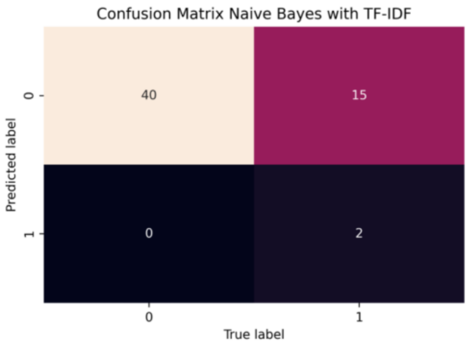

Figure 22. Confusion Matrix using TF-IDF

The last stage is calculation to get the accuracy value. It can be seen in the image below, the Count Vectorizer gets an accuracy 70.18% and an accuracy value using TF-IDF which is 73.69%, this can be seen from Figure 23 and Figure 24.

Count Vectorizer

	precision	recall	f1-score	support
0	0.81	0.75	0.78	40
1	0.50	0.59	0.54	17
accuracy			0.70	57
macro avg	0.66	0.67	0.66	57
weighted avg	0.72	0.70	0.71	57

Score Accuracy with count vectorizer: 0.7017543859649122

Figure 23. *Count Vectorizer Accuracy Value*

TF-IDF

	precision	recall	f1-score	support
0	0.73	1.00	0.84	40
1	1.00	0.12	0.21	17
accuracy			0.74	57
macro avg	0.86	0.56	0.53	57
weighted avg	0.81	0.74	0.65	57

Score Accuracy with TF-IDF : 0.7368421052631579

Figure 24. TF-IDF *Accuracy* Value

CONCLUSIONS

Based on the research that has been done, which has collected as much as 500 data from Twitter related to the Facebook down case. Positive sentiment was obtained as much as 33.92% and negative sentiment with a yield of 66.08 %, which had previously been omitted for neutral sentiment. From the results of the visualization with wordcloud, the words that appear the most are the word facebook down for positive and negative sentiments. Form the research, the results of confusion matrix table from the classification model using data sharing, 80% training data and 20% testing data, with the

ISSN: 1907-6223 (Print)
ISSN:2579-3497 (Online)

classification method using Naive Bayes with TF-IDF word weighting, the accuracy value is 73.69 % and for the Count Vectorizer is 70.18 %. From this study the accuracy results are still not good, this is because the data used for the process still needs to be filtered and the cleaning process is carried out properly. Suggestions for future research can be developed further with other classification methods and use better filtering methods.

REFERENCES

[1] M. A. Harahap and S. Adeni, "Tren penggunaan media sosial selama pandemi di indonesia," *J. Prof. FIS UNIVED*, vol. 7, no. 2, pp. 13–23, 2020.

[2] I. F. Ashari, "Implementation of Cyber-Physical-Social System Based on Service Oriented Architecture in Smart Tourism," *J. Appl. Informatics Comput.*, vol. 4, no. 1, pp. 66–73, 2020, doi: 10.30871/jaic.v4i1.2077.

[3] I. F. Ashari, M. Idris, and M. A. Nasrulah, "Analysis of Combination of Parking System with Face Recognition and QR Code using Histogram of Oriented Gradient Method," *IT J. Res. Dev.*, vol. 7, no. 1, pp. 94–110, 2022, doi: 10.25299/itjrd.2022.9958.

[4] Y. Fadhillah, "Sistem Pakar Menggunakan Forward Chaining dan Certainty Factor untuk Diagnosa Kerusakan Smartphone," *J. Educ. Dev.*, vol. 8, no. 2, pp. 337–342, 2020, [Online]. Available: http://journal.ipts.ac.id/index.php/ED/article/view/1966/1028.

[5] A. Setiadi, "Pemanfaatan media sosial untuk efektifitas komunikasi," *J. Ilm. Matrik*, vol. 16, no. 1, 2014.

[6] A. Saputra, "Survei Penggunaan Media Sosial Di Kalangan Mahasiswa Kota Padang Menggunakan Teori Uses and Gratifications," *Baca J. Dokumentasi Dan Inf.*, vol. 40, no. 2, p. 207, 2019, doi: 10.14203/j.baca.v40i2.476.

[7] I. F. Ashari, M. F. Zuhdi, M. T. Gagaman, and S. T. Denira, "Kolepa Mobile Application Development Based on Android Using SCRUM Method (Case Study : Kolepa Minigolf and Coffe Shop)," *J. Appl. Informatics Comput.*, vol. 6, no. 1, pp. 104–112, 2022.

[8] W. Presthus and D. M. Vatne, "A Survey on Facebook Users and Information Privacy," *Procedia Comput. Sci.*, vol. 164, pp. 39–47, 2019, doi: 10.1016/j.procs.2019.12.152.

[9] A. Clark *et al.*, "Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study," *Lancet Glob. Heal.*, vol. 8, no. 8, pp. e1003–e1017, 2020, doi: 10.1016/S2214-109X(20)30264-3.

[10] I. F. Ashari, M. Alfarizi, M. N. K, and M. A. H, "Vulnerability Analysis and Proven On The neonime. co Website Using OWASP ZAP 4 and XSpear," *J. Teknol. Komput. dan Sist. Inf.*, vol. 5, no. 2, pp. 75–81, 2022.

[11] M. Cindo, D. P. Rini, and Ermatita, "Literartur Review : Metode Klasifikasi Pada Sentimen Analisis," *Semin. Nas. Teknol. Komput. Sains*, pp. 66–70, 2019.

[12] I. F. Ashari, R. Banjarnahor, and D. R. Farida, "Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies," vol. 6, no. 1, pp. 7–15, 2022.

[13] I. F. Ashari, M. C. Untoro, and E. M. Sutowo, "Hyperparameter Tuning Feature Selection with Genetic Algorithm and Gaussian Naïve Bayes for Diabetes Disease Prediction," no. November, 2022.

[14] R. P. Sidiq, B. A. Dermawan, and Y. Umaidah, "Sentimen Analisis Komentar Toxic pada Grup Facebook Game Online Menggunakan Klasifikasi Naïve Bayes," *J. Inform. Univ. Pamulang*, vol. 5, no. 3, p. 356, 2020, doi: 10.32493/informatika.v5i3.6571.

[15] A. Sari, F. V., & Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd. Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 2, no. 2, pp. 681–686, 2019, [Online]. Available: https://jurnal.umk.ac.id/index.php/simet/article/view/3487.

[16] N. S. Fathullah, Y. A. Sari, and P. P. Adikara, "Analisis Sentimen Terhadap Rating dan Ulasan Film dengan menggunakan," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 2, pp. 590–593, 2020.

[17] J. Watori, R. Aryanti, A. Junaidi, and A. Yani, "Penggunaan Algoritma Klasifikasi Terhadap Analisa Sentimen Pemindahan Ibukota Dengan Pelabelan Otomatis," *J. Inform.*, vol. 7, no. 1, pp. 85–90, 2020, doi: 10.31311/ji.v7i1.7528.

[18] I. F. Ashari and R. R. Muharram, "Pengembangan Antarmuka Pengguna Kolepa Mobile App Menggunakan metode Design Thinking dan System Usability Scale," *J. Sist. Inf.*, vol. 9, no. 2, pp. 168–176, 2022, doi: 10.30656/jsii.v9i2.4993.

[19] I. F. Ashari, M. D. Satria, and M. Idris, "Parking System Optimization Based on IoT using Face and Vehicle Plat Recognition via Amazon Web Service and ESP-32 CAM (Case Study : Institut Teknologi Sumatera)," vol. 11, no. 2, pp. 137–153, 2022.

[20] I. F. Ashari and V. Adhelia, "Expert System and IoT for Diagnose of Feline Panleukopenia Virus Using Certainty Factor," *Matrik J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 451–462, 2022, doi: 10.30812/matrik.v21i2.1517.

[21] H. Najjichah, A. Syukur, and H. Subagyo, "Pengaruh Text Preprocessing Dan Kombinasinya Pada Peringkas Dokumen Otomatis Teks Berbahasa Indonesia," *J. Teknol. Inf.*, vol. XV, no. 1, pp. 1–11, 2019.

[22] I. F. Ashari, "The Evaluation of Audio Steganography to Embed Image Files Using Encryption and Snappy Compression," *Indones. J. Comput. Sci.*, vol. 11, no. 2, pp. 318–336, 2022.

[23] I. F. Ashari, "The Evaluation of Image Messages in MP3 Audio Steganography Using Modified Low-Bit Encoding," *Telematika*, vol. 15, 2021.

[24] A. N. Ulfah and M. K. Anam, "Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM)," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 1, pp. 1–10, 2020, doi: 10.35957/jatisi.v7i1.196.

[25] S. Mujilahwati, "Pre-Processing Text Mining Pada Data Twitter," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2016, no. Sentika, pp. 2089–9815, 2016.

[26] A. R. Prananda and I. Thalib, "Sentiment Analysis for Customer Review: Case Study of GO-JEK Expansion," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 6, no. 1, p. 1, 2020, doi: 10.20473/jisebi.6.1.1-8.

[27] S. Elbagir and J. Yang, "Twitter sentiment analysis using natural language toolkit and Vader sentiment," *Lect. Notes Eng. Comput. Sci.*, vol. 2239, pp. 12–16, 2019.

[28] M. G. Pradana, "Penggunaan fitur wordcloud dan document term matrix dalam text mining," *J. Ilm. Inform.*, vol. 8, no. 1, pp. 38–43, 2020.

[29] A. Deolika, K. Kusrini, and E. T. Luthfi, "Analisis Pembobotan Kata Pada Klasifikasi Text Mining," *J. Teknol. Inf.*, vol. 3, no. 2, p. 179, 2019, doi: 10.36294/jurti.v3i2.1077.

[30] W. A. Prabowo and F. Azizah, "Sentiment Analysis for Detecting Cyberbullying Using TF-IDF and SVM," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 6, pp. 11–12, 2020, doi: 10.29207/resti.v4i6.2753.

ISSN: 1907-6223 (Print)
ISSN:2579-3497 (Online)

[31] N. M. Saadah, R. W. Atmagi, D. Rahayu, and A. Z. Arifin, "Information Retrieval of Text Document With Weighting TF-IDF and LCS," *J. Comput. Sci. Inf.*, pp. 34–37, 2013.

[32] V. A. Kozhevnikov and E. Pankratova, "RESEARCH OF THE TEXT DATA VECTORIZATION AND CLASSIFICATION ALGORITHMS OF MACHINE LEARNING," *Int. Sci. J. Theor. Appl. Sci.*, vol. 85, no. 05, 2020, doi: 10.15863/TAS.

[33] C. D. Manning, P. Raghavan, and H. Schutze, "Text classification and Naive Bayes," *Introd. to Inf. Retr.*, no. c, pp. 234–265, 2012, doi: 10.1017/cbo9780511809071.014.

[34] M. C. Untoro, M. Praseptiawan, I. F. Ashari, and A. Afriansyah, "Evaluation of Decision Tree, K-NN, Naive Bayes and SVM with MWMOTE on UCI Dataset," *J. Phys. Conf. Ser.*, vol. 1477, no. 3, 2020, doi: 10.1088/1742-6596/1477/3/032005.

[35] J. Novakovic, A. Veljovi, S. Iiic, Z. Papic, and M. Tomovic, "Evaluation of Classification Models in Machine Learning," *Theory Appl. Math. Comput. Sci.*, vol. 7, no. 1, pp. 39–46, 2017, [Online]. Available: https://uav.ro/applications/se/journal/index.php/TAMCS/article/view/158.