SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 2, August 2024, Pages 114-122
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i2.271

Sentiment Analysis on Twitter Social Media Application on Fuel Oil Price Hike Using Naïve Bayes and Decision Tree Algorithms

Muhammad Fadli[1], Agung Triayudi[2*], Endah Tri Eshti Handayani[3]

Information System Departement, Faculty of Communications and Information Technology, Universitas Nasional, Jakarta, Indonesia

Author Email: mfadli2020@student.unas.ac.id[1], agungtriayudi@civitas.unac.ac.id[2], endahtriesti@civitas.unas.ac.id[3]

Abstract. The increase in fuel prices has significant impacts on the Indonesian community's economic sector. Most people object to this policy because of its significant effects on daily life. Additionally, Indonesia's economy has not fully recovered from the Covid-19 pandemic, compounded by news of rising oil prices. With news of the increase in fuel prices, most people express sentiment regarding the rise in fuel prices on one of the social media platforms, Twitter. This study aims to differentiate the sentiment provided by the public, whether positive, negative, or neutral, using the Naïve Bayes Classifier and Decision Tree algorithms. The analysis results show that the Naïve Bayes Algorithm model, specifically Bernoulli Naïve Bayes, achieves the highest accuracy of 65.60%, with a precision of 68%, recall of 60.30%, and f1-score of 59.33%.

Keywords: Decision Tree, Fuel Oil, Naive Bayes Classifier, Price

1 Introduction

Sentiment analysis is a measurement of a person's opinion about the level of agreement on a particular topic such as a product, service, or event [1]. In conducting sentiment analysis on the increase in fuel oil prices, researchers conducted research using the Naïve Bayes and Decisson Tree algorithms. The use of this algorithm is to classify text into positive, negative, or neutral sentiment categories. Social media is a form of online media that allows individuals to easily participate, share, and create content. There are several types of social media such as online platforms and wikis, which are social media that are often used by people around the world. Examples of commonly accessed social media platforms include Twitter, Facebook, and Myspace which are places for social interaction for people in the world[2].

Twitter is one of the most widely used social networks by people in Indonesia, with an ever-increasing number of users and the ability to quickly disseminate actual information and news. Indonesia is ranked fifth globally in terms of the number of active Twitter users.[3]. News about the increase in the price of fuel oil (BBM) became the main concern of social network users in Indonesia, especially Twitter, in early September 2022. The increase in fuel prices has an impact on the economy of people in Indonesia so that many people disagree with the policy [4]

By understanding the challenges in analyzing public responses related to the increase in fuel oil prices using the Naïve Bayes Algorithm and Decision Tree becomes relevant and can provide solutions to distinguish responses that contain positive or negative properties. This also underlies the researcher to conduct sentiment analysis of public responses related to the increase in fuel oil prices with the title "Sentiment Analysis On Twitter Social Media Application On The Increase In Oil Fuel Prices (BBM) By Using The Naïve Bayes Classifier Algorithm And Decision Tree".

2 Literature Study

2.1 Data Mining

Data mining is the process of mining useful information and knowledge derived from large sets of data warehouses. It involves applying data analysis tools to detect unknown patterns and relationships in large data sets. "Data mining is a multidisciplinary area where several computational paradigms converge: decision tree construction, rule induction, artificial neural networks, example-based learning, Bayesian learning, logic

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 2, August 2024, Pages 114-122
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i2.271

programming, statistical algorithms, and so on. In addition, some of the most useful data mining tasks and methods are statistics, visualization, clustering, classification, and association rule mining. These methods reveal new, interesting, and useful knowledge based on available information [5].

2.2 Google Colaboratory

Google Collaboratory is a cloud-based development platform provided by Google. The platform allows users to write, run, and share program code in various programming languages such as Python.

2.3 Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a part of data mining that is commonly used to evaluate text data in the form of opinions. The goal is to determine the polarity of the opinion so that it can produce information that is positive, negative, or neutral [6].

2.4 Social Media

Social media is an online medium that allows individuals to participate, share and create content. There are several types of social media such as websites and wikis, which are social media platforms that are often used by people around the world. Examples of social media platforms commonly accessed by people in the world include Twitter, Facebook and Myspace [7]. Currently, social media has become a forum for people to express their opinions or opinions, whether they are positive, negative or neutral. However, negative speech is often found on social media and can harm certain parties [8].

2.5 Twitter

With the development of social media today, Twitter is one of the social media platforms that is very well known by internet users, this is due to the simplicity and ease given to users to convey opinions or opinions without significant restrictions. With the growing number of users and the rapid spread of information or news, Twitter has become one of the most popular social media platforms among the public. Currently, the use of Twitter in Indonesia ranks fifth in the world in the number of users [9].

2.6 Naive Bayes Classifier

Naive Bayes is a data mining algorithm that belongs to a grouping of data mining that is often used among other algorithms. This algorithm utilizes probability techniques to predict the likelihood of future events based on past experiences or events [10]. The Naïve Bayes algorithm utilizes training data to develop a classification model. Naïve Bayes finds significant differences between two datasets by comparing them with a larger dataset [11]. As for the stages of completing the Naïve Bayes Classifier algorithm used in classifying data, the following is the general formula for calculating the Naïve Bayes Classifier algorithm.

$$\mathbf{P(C|X)} \frac{\mathbf{P(X|C), P(C)}}{\mathbf{P(X)}}$$

Figure 1 Naive Bayes Classifier Formula

Description:
X = Sample data that has an unknown class (label).
C = Hypothesis that X is data class (label)
P(C) = Probability of hypothesis C.
P(X) = Probability of the observed sample data (probability of C).
P(X|C) = Probability based on the condition in the hypothesis.

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 2, August 2024, Pages 114-122
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i2.271

Table 1 Confusion Matrix Formula

Confusion Matrix		
	Prediction	
ACTUAL	TRUE	FALSE
TRUE	TRUE POSITIVE (TP)	FALSE POSITIFE (FP)
FALSE	FALSE NEGATIVE (FN)	TRUE NEGATICE (TP)

The following is an explanation of the table above:

a. True Positive (TP), is the number of one TRUE class that can be predicted correctly in the TRUE class.
b. True Negative (TN), is the number of one FALSE class that can be correctly predicted in the FALSE class.
c. False Positive (FP), is a condition where the predicted TRUE class is wrong for the FALSE class.
d. False Negative (FN), is a condition where the FALSE class is predicted incorrectly in the TRUE class.
e. To determine the percentage of accuracy, precision, and recall researchers use the following calculation formula:

The following is the formula to calculate :

a. Accuracy: (TP + TN) / Total of all data
b. Precision: TP / (TP+FP)
c. Recall: TP / (TP+FN)

2.7 Decision Tree Algorithm

Decision tree algorithm, also known as decision tree, is an algorithm that utilizes a set of data that has been labeled to represent a decision tree as the result. To determine its general capabilities, the decision tree is tested with unseen experimental data. This Decision Tree algorithm utilizes the principle of information entropy to form a decision tree from the research dataset [12].

Decision tree algorithms are widely used because they can clearly describe a model, knowledge or information represented in the format of a decision tree structure. The decision tree algorithm consists of a number of nodes connected by branches that run from the root node to the leaf nodes. Leaf nodes are nodes that cannot be subdivided, leaf nodes represent the predicted answer to a problem (test data). The decision tree is made upside down, where the root node is at the top, while the leaf node is at the bottom [13].

2.8 Fuel Oil

Fuel oil (BBM) is a type of fuel (fuel) produced from the refining of crude oil from the bowels of the earth. Crude oil is refined first in refineries to produce oil products, including BBM. Crude oil refining also produces various other products, such as gas, to products such as naphtha (Dewi, Y., Saryono, S., Dini, A., Maghfiroh, M., & Mauli, R. [14]. Fuel Oil (BBM) has a very important role in people's lives. In addition, fuel is one of the basic needs for urban and rural communities, both in household needs, traders, and even companies, as well as fuel has a very important role in the industrial and transportation sectors. The increase in fuel prices for people's lives raises the phenomenon of pros and cons among the public and various media. Because the impact of changes in fuel prices affects the price of production, distribution, transportation so that it also affects the price of other goods. Basic food needs are also affected by rising fuel prices, such as rice, sugar, and cooking oil [15].

3 Research Method

The initial stage is very important in conducting the research journey. Researchers use the Naïve Bayes Classifier and Decision Tree algorithms to determine the percentage of sentiment that is positive, negative, and neutral. The research stages can be seen in the figure below.

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 2, August 2024, Pages 114-122
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i2.271

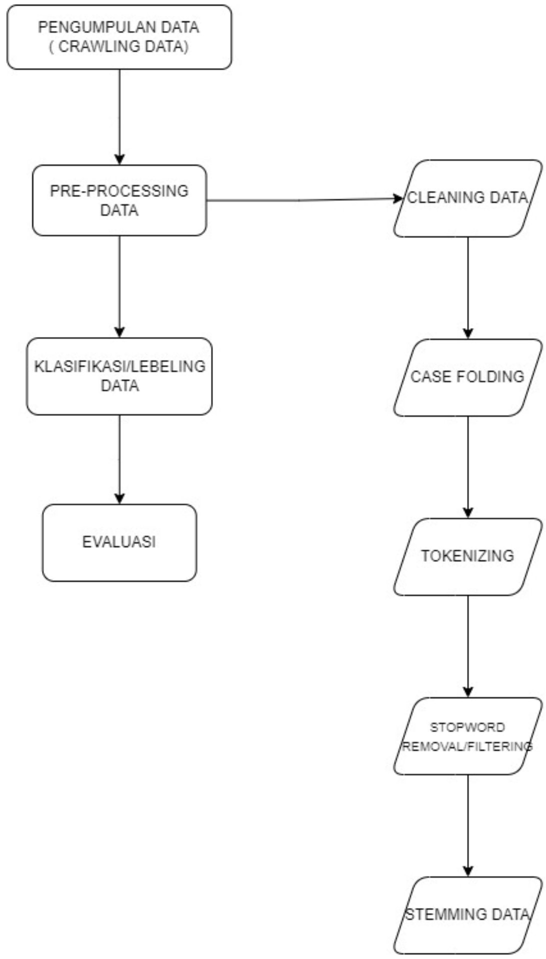

Figure 1. Research Method

In the Data Collection phase, researchers crawled data on the Twitter social media application using API tokens through Google Collaboratory as a tool for data retrieval. However, in this Data Collection phase, researchers only mined 500 data. Due to the limitations of data collection on twiiter social media applications with a maximum data collection of 800 data.

The next step is Pre-Processing, where researchers process raw data into data that is easier to understand. At this stage, researchers perform data cleaning, case folding, tokenizing, stopword removal/filtering, and data stemming. Data cleaning is the process of preparing data that will be used for analysis by deleting or modifying data that has the wrong format, incomplete, and duplicated. Furthermore, the researcher performs case folding with the aim of changing all letters contained in the data into lowercase letters.

The next process is tokenizing, which is the stage of separating text into parts in the form of tokens, either in the form of letters, numbers, words or sentences. The next stage in pre-processing is stoword removal/filtering, which is removing common words that often appear but are not meaningful. The last phase in the Pre-Processing stage is Stemming, a technique used to extract words and convert them into their basic form.

The next step is Classification, which is the process of grouping data where the data to be used has a class or to achieve the intended target.

The next phase is the Evaluation phase, which is the last step taken by researchers after obtaining data that has been classified. This stage is carried out to determine Precision, Recall, F-score, and Accuracy.

Vol 2, Issue 2, August 2024, Pages 114-122
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i2.271

4 Result and Discussion

4.1 Taking Twitter Authentification Token

The first stage in crawling data is to get a twitter authentification token first and then the token is inserted into google colab.

```
#@title Twitter Auth Token

twitter_auth_token = 'eb68cc59d0564625a49d365c32292791a11078f4'
```

Figure 2. Twitter Authentification Token

4.2 Crawling Data

At this phase, researchers continue the process of mining text or pulling data from the twitter social media application to get the data needed. At this phase, researchers include topics that will be searched by search keywords and determine the limit or amount of data.

	created_at	id_str	full_text	quote_count	reply_count	retweet_count	favorite_count	lang	user_id_str	conversation_id_str	username	tweet_url
0	Fri Dec 30 22:17:06 +0000 2022	1608950329642745856	@BennyHarmanID Wah presiden Jokowi byk bohongn...	0	0	0	0	in	1431979460249612295	1608718081341198336	Ugi7498	https://twitter.com/Ugi7498/status/16089503296...
1	Fri Dec 30 18:13:55 +0000 2022	1608889130754387968	Bupati Indramayu Serahkan Bansos Bagi Nelayan ...	0	0	0	0	in	381923504	1608889130754387968	AboutSemarangID	https://twitter.com/AboutSemarangID/status/160...
2	Fri Dec 30 11:10:29 +0000 2022	1608782570757689345	Hal yang sama pernah dikatakan Jokowi sebelum ...	0	0	0	5	in	164778726	1608782570757689345	dusrimulya	https://twitter.com/dusrimulya/status/16087825...
3	Fri Dec 30 10:15:09 +0000 2022	1608768648063365120	Tahun depan, dari sejumlah laporan dan kajian ...	0	1	1	3	in	79139422	1608763656250220544	fadlizon	https://twitter.com/fadlizon/status/1608768648...
4	Fri Dec 30 10:09:25 +0000 2022	1608767201808625664	Tetapi, kenaikan tajam harga BBM bersubsidi pa...	0	1	1	3	in	79139422	1608763656250220544	fadlizon	https://twitter.com/fadlizon/status/1608767201...
...	...	...	...	...	...	...	...	...	...	...	...	...
991	Thu Dec 15 19:24:15 +0000 2022	1603471016117735424	Selain nomor urut, yang ga berubah dari PKS ad...	0	0	0	0	in	18751414	1603471016117735424	arifunasiin	https://twitter.com/arifunasiin/status/1603471...
992	Thu Dec 15 19:23:15 +0000 2022	1603470764396941312	Selain nomor urut, yang ga berubah dari PKS ad...	0	0	0	0	in	1488435746495107076	1603470764396941312	PKS_PondokGede	https://twitter.com/PKS_PondokGede/status/1603...
993	Thu Dec 15 19:21:48 +0000 2022	1603470398599102464	Selain nomor urut, yang ga berubah dari PKS ad...	0	0	0	1	in	225322886	1603470398599102464	pksjabar	https://twitter.com/pksjabar/status/1603470398...
994	Thu Dec 15 19:18:56 +0000 2022	1603469675488497664	Selain nomor urut, yang ga berubah dari PKS ad...	0	0	0	0	in	225322886	1603469675488497664	pksjabar	https://twitter.com/pksjabar/status/1603469675...
995	Thu Dec 15 19:01:06 +0000 2022	1603465189453885440	Selain nomor urut, yang ga berubah dari PKS ad...	0	0	0	0	in	1488435746495107076	1603465189453885440	PKS_PondokGede	https://twitter.com/PKS_PondokGede/status/1603...

Figure 3. Crawling Data Display

4.3 Display the Frequency of Frequently Occurring Words

Before performing the pre-processing stage, researchers first display the frequency of words that often appear in documents that have been obtained through the crawling process. The barchart display of the frequency of words that often appear can be seen below this text.

.

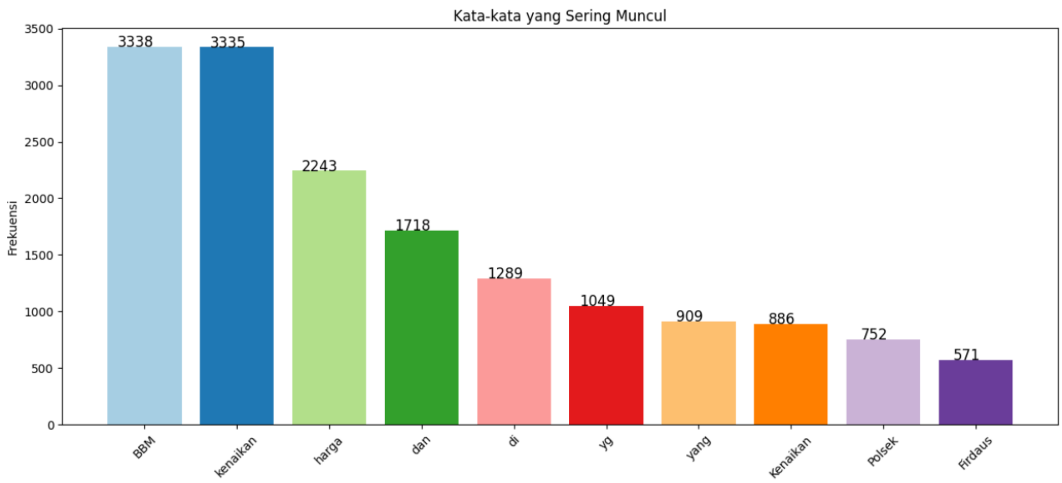

Figure 4. Word Frequency Bar Chart View

4.4 Preprocessing Data

In this process, researchers process the raw data that has been obtained through the crawling process and then clean the data by performing several stages in the preprocessing process, namely:

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 2, August 2024, Pages 114-122
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i2.271

4.4.1 Cleaning Data

This stage is carried out to prepare the raw data that has been obtained by deleting and modifying it with the correct and appropriate format.

	full_text	cleasing
0	Harga BBM di Italia Melonjak Setelah Berakhirn...	Harga BBM di Italia Melonjak Setelah Berakhirn...
1	Dampak Kenaikan Harga BBM, Faisal Rachman Bila...	Dampak Kenaikan Harga BBM Faisal Rachman Bilan...
2	#BLTBBMBantuMasyarakat #hargabbm #kenaikanharg...	BLTBBMBantuMasyarakat hargabbm kenaikanhargabb...
3	#BLTBBMBantuMasyarakat #hargabbm #kenaikanharg...	BLTBBMBantuMasyarakat hargabbm kenaikanhargabb...
4	#BLTBBMBantuMasyarakat #hargabbm #kenaikanharg...	BLTBBMBantuMasyarakat hargabbm kenaikanhargabb...
...	...	...
495	Kasus Sambo tertutup oleh kasus kenaikan BBM. ...	Kasus Sambo tertutup oleh kasus kenaikan BBM K...
496	Kadishub DKI Jakarta Syafrin Liputo mengatakan...	Kadishub DKI Jakarta Syafrin Liputo mengatakan...
497	Pada Hari Selasa Tanggal 25 Oktober 2022 Sekir...	Pada Hari Selasa Tanggal Oktober Sekira Puku...
498	Kapolres Nias AKBP Luthfi, S.I.K., bersama Kas...	Kapolres Nias AKBP Luthfi SIK bersama Kasat La...
499	Harusnya tidak perlu tunjangan BBM untuk DPRD ...	Harusnya tidak perlu tunjangan BBM untuk DPRD ...

Figure 5. Cleaning Data Display

4.4.2 Case Folding & Tokenizing

This process aims to convert all letters in the dataset into lowercase letters in order to harmonize and this process is also done to break the text into small units called tokens.

	full_text	cleasing	Tokenization and Case Folding
0	Harga BBM di Italia Melonjak Setelah Berakhirn...	Harga BBM di Italia Melonjak Setelah Berakhirn...	[harga, bbm, di, italia, melonjak, setelah, be...
1	Dampak Kenaikan Harga BBM, Faisal Rachman Bila...	Dampak Kenaikan Harga BBM Faisal Rachman Bilan...	[dampak, kenaikan, harga, bbm, faisal, rachman...
2	#BLTBBMBantuMasyarakat #hargabbm #kenaikanharg...	BLTBBMBantuMasyarakat hargabbm kenaikanhargabb...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...
3	#BLTBBMBantuMasyarakat #hargabbm #kenaikanharg...	BLTBBMBantuMasyarakat hargabbm kenaikanhargabb...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...
4	#BLTBBMBantuMasyarakat #hargabbm #kenaikanharg...	BLTBBMBantuMasyarakat hargabbm kenaikanhargabb...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...
...	...	...	...
495	Kasus Sambo tertutup oleh kasus kenaikan BBM. ...	Kasus Sambo tertutup oleh kasus kenaikan BBM K...	[kasus, sambo, tertutup, oleh, kasus, kenaikan...
496	Kadishub DKI Jakarta Syafrin Liputo mengatakan...	Kadishub DKI Jakarta Syafrin Liputo mengatakan...	[kadishub, dki, jakarta, syafrin, liputo, meng...
497	Pada Hari Selasa Tanggal 25 Oktober 2022 Sekir...	Pada Hari Selasa Tanggal Oktober Sekira Puku...	[pada, hari, selasa, tanggal, oktober, sekira,...
498	Kapolres Nias AKBP Luthfi, S.I.K., bersama Kas...	Kapolres Nias AKBP Luthfi SIK bersama Kasat La...	[kapolres, nias, akbp, luthfi, sik, bersama, k...
499	Harusnya tidak perlu tunjangan BBM untuk DPRD ...	Harusnya tidak perlu tunjangan BBM untuk DPRD ...	[harusnya, tidak, perlu, tunjangan, bbm, untuk...

Figure 6. Case Folding & Tokenizing Display

4.4.2 Filtering / Stopword Removal

This process is practiced with the aim of eliminating frequently occurring but less important words, such as "and", "at", "from". This process improves the quality of text analysis by focusing on meaningful and informative words.

	full_text	cleasing	Tokenization and Case Folding	Filtering/stopword removal
0	Harga BBM di Italia Melonjak Setelah Berakhirn...	Harga BBM di Italia Melonjak Setelah Berakhirn...	[harga, bbm, di, italia, melonjak, setelah, be...	[harga, bbm, italia, melonjak, diskon, cukai, ...
1	Dampak Kenaikan Harga BBM, Faisal Rachman Bila...	Dampak Kenaikan Harga BBM Faisal Rachman Bilan...	[dampak, kenaikan, harga, bbm, faisal, rachman...	[dampak, kenaikan, harga, bbm, faisal, rachman...
2	#BLTBBMBantuMasyarakat #hargabbm #kenaikanharg...	BLTBBMBantuMasyarakat hargabbm kenaikanhargabb...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...
3	#BLTBBMBantuMasyarakat #hargabbm #kenaikanharg...	BLTBBMBantuMasyarakat hargabbm kenaikanhargabb...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...
4	#BLTBBMBantuMasyarakat #hargabbm #kenaikanharg...	BLTBBMBantuMasyarakat hargabbm kenaikanhargabb...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...
...	...	...	...	...
495	Kasus Sambo tertutup oleh kasus kenaikan BBM. ...	Kasus Sambo tertutup oleh kasus kenaikan BBM K...	[kasus, sambo, tertutup, oleh, kasus, kenaikan...	[sambo, tertutup, kenaikan, bbm, knaikan, bbm,...
496	Kadishub DKI Jakarta Syafrin Liputo mengatakan...	Kadishub DKI Jakarta Syafrin Liputo mengatakan...	[kadishub, dki, jakarta, syafrin, liputo, meng...	[kadishub, dki, jakarta, syafrin, liputo, tari...
497	Pada Hari Selasa Tanggal 25 Oktober 2022 Sekir...	Pada Hari Selasa Tanggal Oktober Sekira Puku...	[pada, hari, selasa, tanggal, oktober, sekira,...	[selasa, tanggal, oktober, sekira, wib, person...
498	Kapolres Nias AKBP Luthfi, S.I.K., bersama Kas...	Kapolres Nias AKBP Luthfi SIK bersama Kasat La...	[kapolres, nias, akbp, luthfi, sik, bersama, k...	[kapolres, nias, akbp, luthfi, sik, kasat, lan...
499	Harusnya tidak perlu tunjangan BBM untuk DPRD ...	Harusnya tidak perlu tunjangan BBM untuk DPRD ...	[harusnya, tidak, perlu, tunjangan, bbm, untuk...	[tunjangan, bbm, dprd, tolak, kenaikan, tunjan...

Figure 7. Filtering / Stopword Removal Display

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 2, August 2024, Pages 114-122
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i2.271

4.4.3 Stemming

This stage involves converting the words in the text to their basic form or base word. This stage is practiced with the aim of simplifying the words in the text so that the various variations of the same word are considered as one unit.

	full_text	cleasing	Tokenization and Case Folding	Filtering/stopword removal	stemming_data
0	Harga BBM di Italia Melonjak Setelah Berakhirn...	Harga BBM di Italia Melonjak Setelah Berakhirn...	[harga, bbm, di, italia, melonjak, setelah, be...	[harga, bbm, italia, melonjak, diskon, cukai, ...	harga bbm italia lonjak diskon cukai diskon ha...
1	Dampak Kenaikan Harga BBM, Faisal Rachman Bila...	Dampak Kenaikan Harga BBM Faisal Rachman Bilan...	[dampak, kenaikan, harga, bbm, faisal, rachman...	[dampak, kenaikan, harga, bbm, faisal, rachman...	dampak naik harga bbm faisal rachman bilang se...
2	#BLTBBMBantuMasyarakat #hargabbm #kenaikanharg...	BLTBBMBantuMasyarakat hargabbm kenaikanhargabb...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...	bltbbmbantumasyarakat hargabbm kenaikanhargabb...
3	#BLTBBMBantuMasyarakat #hargabbm #kenaikanharg...	BLTBBMBantuMasyarakat hargabbm kenaikanhargabb...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...	bltbbmbantumasyarakat hargabbm kenaikanhargabb...
4	#BLTBBMBantuMasyarakat #hargabbm #kenaikanharg...	BLTBBMBantuMasyarakat hargabbm kenaikanhargabb...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...	bltbbmbantumasyarakat hargabbm kenaikanhargabb...
...	...	...	...	...	...
495	Kasus Sambo tertutup oleh kasus kenaikan BBM. ...	Kasus Sambo tertutup oleh kasus kenaikan BBM K...	[kasus, sambo, tertutup, oleh, kasus, kenaikan...	[sambo, tertutup, kenaikan, bbm, knaikan, bbm,...	sambo tutup naik bbm knaikan bbm tutup kanjuru...
496	Kadishub DKI Jakarta Syafrin Liputo mengatakan...	Kadishub DKI Jakarta Syafrin Liputo mengatakan...	[kadishub, dki, jakarta, syafrin, liputo, meng...	[kadishub, dki, jakarta, syafrin, liputo, tari...	kadishub dki jakarta syafrin liputo tarif angk...
497	Pada Hari Selasa Tanggal 25 Oktober 2022 Sekir...	Pada Hari Selasa Tanggal Oktober Sekira Puku...	[pada, hari, selasa, tanggal, oktober, sekira,...	[selasa, tanggal, oktober, sekira, wib, person...	selasa tanggal oktober sekira wib personil pol...
498	Kapolres Nias AKBP Luthfi, S.I.K., bersama Kas...	Kapolres Nias AKBP Luthfi SIK bersama Kasat La...	[kapolres, nias, akbp, luthfi, sik, bersama, k...	[kapolres, nias, akbp, luthfi, sik, kasat, lan...	kapolres nias akbp luthfi sik kasat lantas pol...

Figure 8. Stemming Display

4.4.4 Drop Duplicates

a stage of the process of removing entries or rows that are exactly the same in a data set. This process is done to ensure that each entry in the data is unique, and eliminate unnecessary duplicates.

	full_text	cleasing	Tokenization and Case Folding	Filtering/stopword removal	stemming_data
0	Harga BBM di Italia Melonjak Setelah Berakhirn...	Harga BBM di Italia Melonjak Setelah Berakhirn...	[harga, bbm, di, italia, melonjak, setelah, be...	[harga, bbm, italia, melonjak, diskon, cukai, ...	harga bbm italia lonjak diskon cukai diskon ha...
1	Dampak Kenaikan Harga BBM, Faisal Rachman Bila...	Dampak Kenaikan Harga BBM Faisal Rachman Bilan...	[dampak, kenaikan, harga, bbm, faisal, rachman...	[dampak, kenaikan, harga, bbm, faisal, rachman...	dampak naik harga bbm faisal rachman bilang se...
2	#BLTBBMBantuMasyarakat #hargabbm #kenaikanharg...	BLTBBMBantuMasyarakat hargabbm kenaikanhargabb...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...	bltbbmbantumasyarakat hargabbm kenaikanhargabb...
3	#BLTBBMBantuMasyarakat #hargabbm #kenaikanharg...	BLTBBMBantuMasyarakat hargabbm kenaikanhargabb...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...	bltbbmbantumasyarakat hargabbm kenaikanhargabb...
4	#BLTBBMBantuMasyarakat #hargabbm #kenaikanharg...	BLTBBMBantuMasyarakat hargabbm kenaikanhargabb...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...	[bltbbmbantumasyarakat, hargabbm, kenaikanharg...	bltbbmbantumasyarakat hargabbm kenaikanhargabb...
...	...	...	...	...	...
495	Kasus Sambo tertutup oleh kasus kenaikan BBM. ...	Kasus Sambo tertutup oleh kasus kenaikan BBM K...	[kasus, sambo, tertutup, oleh, kasus, kenaikan...	[sambo, tertutup, kenaikan, bbm, knaikan, bbm,...	sambo tutup naik bbm knaikan bbm tutup kanjuru...
496	Kadishub DKI Jakarta Syafrin Liputo mengatakan...	Kadishub DKI Jakarta Syafrin Liputo mengatakan...	[kadishub, dki, jakarta, syafrin, liputo, meng...	[kadishub, dki, jakarta, syafrin, liputo, tari...	kadishub dki jakarta syafrin liputo tarif angk...
497	Pada Hari Selasa Tanggal 25 Oktober 2022 Sekir...	Pada Hari Selasa Tanggal Oktober Sekira Puku...	[pada, hari, selasa, tanggal, oktober, sekira,...	[selasa, tanggal, oktober, sekira, wib, person...	selasa tanggal oktober sekira wib personil pol...
498	Kapolres Nias AKBP Luthfi, S.I.K., bersama Kas...	Kapolres Nias AKBP Luthfi SIK bersama Kasat La...	[kapolres, nias, akbp, luthfi, sik, bersama, k...	[kapolres, nias, akbp, luthfi, sik, kasat, lan...	kapolres nias akbp luthfi sik kasat lantas pol...

Figure 9. Drop Duplicates Display

4.5 Labeling

The labeling process is done to identify or categorize the data. This process involves labeling or classifying each data entry. After performing the lebeling process, data entries can be given sentiment labels in the form of positive, negative, or neutral.

	stemming_data	sentiment_polarity	sentiment
0	Italian fuel prices jump, excise tax discounts...	0.000000	netral
1	impact of rising fuel prices, Faisal Rachman s...	0.000000	netral
2	bltbbmhelps the community price of fuel oil in...	0.000000	netral
3	target subsidy bltbbm subsidybbm price incre...	0.000000	netral
4	subsidies are right on target for BLT fuel pri...	0.285714	positif

Figure 10. Lebeling Data Display

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 2, August 2024, Pages 114-122
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i2.271

Datasets that have been classified or labeled as positive, negative, or neutral in every tweet posted by users on the twitter social media application.

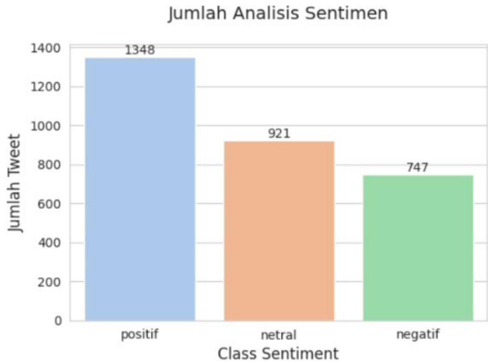

Figure 11. Barchart View Sentiment Analysis

4.6 Naive Bayes Algorithm Result

The accuracy results of each naïve bayes classification model include Gaussian Naïve Bayes with an accuracy of 58.60%, Multinomial Naïve Bayes with an accuracy of 65.40%, and Bernoulli Naïve Bayes with an accuracy of 65.60%.

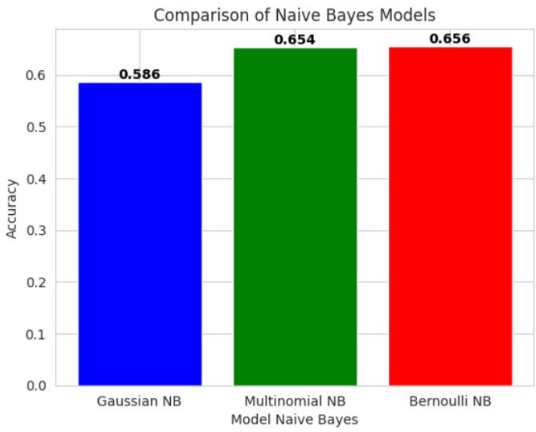

Figure 3 Barchart View Naive Bayes Algorithm Result

4.7 Decision Tree Algorithm Result

The accuracy results of the Decision Tree algorithm from the dataset that has been analyzed with an accuracy of 61.92%, precision of 62.66%, recall of 64%, and f1-score of 63%.

```
Accuracy - Decision Tree: 0.6324503311258278

Classification Report - Decision Tree:
              precision    recall  f1-score   support

     negatif       0.60      0.61      0.61       150
      netral       0.57      0.70      0.63       179
     positif       0.71      0.60      0.65       275

    accuracy                           0.63       604
   macro avg       0.63      0.64      0.63       604
weighted avg       0.64      0.63      0.63       604
```

Figure 13. Result View Of Descision Tree Algorithm

SaNa: Journal of Blockchain, NFTs and Metaverse Technology
Vol 2, Issue 2, August 2024, Pages 114-122
ISSN: 3030-9832 (Media Online)
DOI: https://doi.org/10.58905/sana.v2i2.271

5 Conclusion

Naive Bayes is a commonly applied data mining algorithm for clustering data mining. This algorithm utilizes the principle of probability to measure the likelihood of future events based on past experiences or events. While the Decission Tree Algorithm or decision tree is an algorithm that takes a collection of data that has a label and represents a decision tree as its output. The accuracy results that can be obtained from the Naïve Bayes Algorithm Model are, Gaussian Naïve Bayes with an accuracy of 58.60%, precision of 60.33%, recall of 57.66%, and f1-score of 58.33%. Multinomial Naïve Bayes with accuracy 65.40%, precision 66.66%, recall 60.66%, and f1-score 62%. Bernoulli Naïve Bayes with accuracy 65.60%, precision 68%, recall 60.30%, and f1-score 59.33%. While the accuracy results obtained using the Decision Tree Algorithm are 61.92%, precision 62.66%, recall 64%, and f1-score 63%. The conclusion obtained is that the most accurate accuracy results obtained are using the Bernoulli Naïve Bayes Algorithm Model.

References

[1] Gusti, I., Indrawan, A., Ayu, D., Cahya Dewi, I., Putu, I. A., & Wisdantini, A. (2023). Analisis Sentimen Terhadap Presidensi G20 2022 pada Media Sosial Twitter Menggunakan Metode Naïve Bayes. *Media Online)*, *4*(1), 553–561.

[2] Muzaki, A., & Witanti, A. (2021). SENTIMENT ANALYSIS OF THE COMMUNITY IN THE TWITTER TO THE 2020 ELECTION IN PANDEMIC COVID-19 BY METHOD NAIVE BAYES CLASSIFIER. Jurnal Teknik Informatika (Jutif), 2(2), 101–107.

[3] Putu Wibina Karsa Gumi, I., Syafrianto, A., Ilmu Komputer, F., Studi Informatika STMIK RAHMA Yogyakarta, P. el, Sisingamangaraja Jl Karangkajen No, J., Mergangsan, K., & Yogyakarta Author Emails, K. (2022). Perbandingan Algoritma Naïve Bayes dan Decision Tree Pada Sentimen Analisis. In IJCSR: The Indonesian Journal of Computer Science Research (Vol. 1).

[4] Arfah Wahlil Pratama, M., fuad, M., & Komputer, S. (n.d.). Penentuan Status Penerima Bantuan Indonesia Pintar Pada Smkn 9 Bulukumba Dengan Metode Naive Bayes.

[6] Križanić, S. (2020). Educational data mining using cluster analysis and decision tree technique: A case study. International Journal of Engineering Business Management, 12.

[7] Aulia Sari, C., Sukmawati, A., Putri Aprilli, R., Sarah Kayaningtias, P., & Yudistira, N. (2022). PERBANDINGAN METODE NAÏVE BAYES, SUPPORT VECTOR MACHINE DAN DECISION TREE DALAM KLASIFIKASI KONSUMSI OBAT. 3, 33–41.

[8] Musyaffa, I., & Kamayani, M. (2022). Dipresentasikan pada Tanggal 3 Desember. Universitas Muhammadiyah Prof. Dr. Hamka Jl. Tanah Merdeka, 7(6), 7261226.

[9] Khatib Sulaiman, J., Danirmala, T., Sulistyo Nugroho, Y., & Muhammdiyah Surakarta, U. (n.d.). Analisis Sentimen Terhadap Topik Kenaikan Harga Bahan Bakar Minyak (BBM) pada Media Sosial Twitter. Indonesian Journal of Computer Science.

[9] Rejeki, F., & Ayumi, V. (2023). Analisa Sentimen Mengenai Kenaikan Harga Bbm Menggunakan Metode Naïve Bayes Dan Support Vector Machine. JSAI : Journal Scientific and Applied Informatics, 6(1).

[10] J., Akbar, F., Wira Saputra, H., Karel Maulaya, A., & Fikri Hidayat, M. (2022). MALCOM: Indonesian Journal of Machine Learning and Computer Science Implementation of Decision Tree Algorithm C4.5 and Support Vector Regression for Stroke Disease Prediction Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke. 2, 61–67.