Naive Bayes Method In Sentiment Analysis Of Presidential Candidates For The 2024 Election Using Python

Raditia Vindua[1], Nurhayati[2]

[1,2]Faculty of Computer Science, Informatics Engineering, Pamulang University, South Tangerang, Indonesia

Article Info	ABSTRACT
Keywords: Naïve Bayes, Presidential Candidates, Python, Sentiment Analysis, The 2024 Election.	The 2024 Election in Indonesia is an interesting topic for social media users. Social media has a big impact in building public political opinions, views, sentiments and preferences. Many political figures have been nominated for President based on public opinion. There are various opinions of media social users with negative, positive and neutral sentiments. However, determining the sentiment of social media users requires quite a lot of effort and time. The large number of incoming opinions regarding presidential election candidates encourages the need for methods that help to see public opinion effectively. Python is a programming language that can be used to answer these problems. By providing a standard library that is open source and has a wide range of applications in various fields. Classification will be carried out using the Naïve Bayes Classifier to determine the level of accuracy of the classification process carried out. Sentiment analysis in this research is a process carried out to find out what the results of sentiment analysis are regarding the public's response to the presidential candidates for the upcoming 2024 election and classify them into three classes using the Naïve Bayes method using Python. The results of this research showed that Python carried out sentiment analysis with the sentiment percentage results for candidate Anies Muhaimin with a positive class of 64.91%, neutral 28.07% and negative 7.02% with a Naïve Bayes accuracy value of 75%. For candidate Prabowo Gibran, the positive class is 12.38%, neutral 6.67% and negative 80.95% with a Naïve Bayes accuracy value of 81%. Meanwhile, the candidate Ganjar Mahfud has a positive class of 40%, neutral 50.67% and negative 9.33% with a Naïve Bayes accuracy value of 60%. So that we can identify public opinion about presidential candidates for the 2024 election using the Naïve Bayes method using Python.
	Corresponding Author: Raditia Vindua Pamulang University South Tangerang Dosen02380@unpam.ac.id

INTRODUCTION

Indonesia is a democratic country. One of the main pillars in democratic system is the process of channeling people's opinion through general elections held periodically (Asshiddiqie, 2007). The 2024 Election is one of the applications of democracy in Indonesia. The democratic process can be realized through elections conducted to elect people's representatives and other public officials (Subiyanto, 2020). The 2024 Election in Indonesia are usually held

Naive Bayes Method In Sentiment Analysis Of Presidential Candidates For The 2024 Election Using Python—Raditia Vindua et.al

546 | P a g e

periodically. In 2024, Indonesia will hold simultaneous elections in electing the President. Many political figures have been nominated to be President based on public opinion. Because public opinion related to elections can be used to see a picture of public opinion towards presidential candidates.

Before the existence of technology, people expressed their opinions through newspaper where not everyone had the opportunity and ability to write and publish it. Only some people can voice their opinions through newspaper with the aim of being seen by many people, now everyone can communicate their opinions and their opinions can be seen by many people because of advances in communication technology in Indonesia. Democracy itself has run through several online media. Due to the rapid development of the times that have penetrated to the development of current communication technology, changing people's tendency to express their opinions on social media.

Social media is a media used to mingle and exchange data and communication including blogs, social networks, wikis, forums and virtual worlds (Istiani & Islamy, 2020). There are many benefits of social media, such as making new friends, knowing the game, financial aspects, travel industry data and also for political issues. The 2024 election, which is the largest democratic party in Indonesia, is an interesting topic for social media users to have opinions.

Twitter or now called X is one of the social media used to share opinions, and has features such as retweets, taking photos and videos, and sharing them to several other social networks. Twitter is already attached to human life at the present moment (Aggrawal, 2018). X is widely used for opinions because of easy access and unlimited number of followers. With a short number of characters (280 characters) so that users can convey the purpose and purpose clearly, concisely and concisely. Twitter can express objective opinions on different topics so that Twitter becomes one of the popular social media (Coletta et al., 2019).

With so many tweets circulating on X discussing the 2024 Indonesian elections, researchers are motivated to see sentiment analysis from tweets on X. Sentiment analysis is a field of study that analyzes one's opinion, one's sentiment, one's evaluation, one's attitude and one's emotions into written language. This study classifies tweets use Naïve Bayes method into three classes, namely positive, neutral and negative, because the enforcement of democracy and freedom of citizens in opinion, not only have a positive impact, but also a negative impact, such as political uproar (Jailani, 2015). Not all Twitter users are wise in choosing words in their tweets and not even a few Twitter users express expressions by writing rude and offensive words (Hidayatullah et al., 2019).

Determining the polarity of an opinion can be done manually, but as more opinions grow, it certainly takes more time and effort to classify the polarity of the opinion. Therefore, a new method was applied to classify the polarity of opinions from these vast data sources. To do that, you can use one of the functions of machine learning, namely from text mining. Text Mining aims to find valuable information hidden from both structured and unstructured information sources (Firdaus & Firdaus, 2021). However, opinion mining does not pay attention to the topic of the text, but rather focuses on the expression described from the opinion text (Indrayuni & Wahyudi, 2015).

Naive Bayes Method In Sentiment Analysis Of Presidential Candidates For The 2024 Election Using Python–Raditia Vindua et.al

547 | P a g e

Previous research proposed a framework called the concept-level sentiment analysis model, which takes into account various natural language processing to extract opinion information from tweets. The model classifies tweets as neutral, sarcastic, positive or negative by exploiting anaphoretic adjectives and pronouns. Although the proposed model only pays attention to the essential components of tweet text, namely adjectives and pronouns, the classification will not be accurate if the set of positive and negative adjectives is a specific topic (Cambria et al., 2015).

There has been a lot of previous research on sentiment analysis. First, research conducted on the classification of twitter user sentiment in the 2019 Indonesian Presidential Election using the Naive Bayes method. Of the 240 training data, there were 134 negative sentiments and 106 positive sentiments. This study shows that the classification of tweets data using the naïve bayes classifier algorithm provides an accuracy of 73%. Negative class precision is 78% and positive class precision is 66% (Mahbubah & Zuliarso, 2019).

With the same method, a Sentiment Analysis of the 2019 Indonesian Presidential Candidate has been carried out from Twitter Social Media. In testing 3 classes of Presidential Candidate number one and Presidential Candidate number two, the accuracy results were obtained as follows: 64.6% and 58% respectively. Meanwhile, in testing 2 classes of Presidential Candidate number one and Presidential Candidate number two, the accuracy results were obtained as follows: 77.7% and 88%. The highest performance is found in the number two presidential candidate with an f-measure value of 0.88 (Fitriyyah et al., 2019).

Meanwhile, research on the analysis of Twitter user sentiment towards the 2019 Indonesian Presidential candidate using the Support Vector Machine method resulted in an accuracy of 86.82% for tweets with the keyword "Jokowi" and 86.27% for tweets with the keyword "Prabowo" (Lukmana et al., 2019). Subsequent research on the sentiment analysis of twitter users in the Presidential Candidate Debate using the Fined-Grained Method showed that sentiment tweets from both hashtags tended to be positive, more than negative and neutral sentiments (Pratama et al., 2019).

The large number of tweets about the election prompted the need for new methods to help see people's opinions effectively. Python is a programming language that can be used to answer these problems. By providing a standard library that is open source and has a wide range of applicability in various fields, Python is able to classify tweet data using the textblob library. So Python becomes one of the high-level programming languages that is easy to use (Wibowo et al., 2020).

Previous research using Python on Opinion-Based Sentiment Analysis from Twitter Social Media on "Leader Figures" resulted in sentiment with a positive class of 21.6% with a number of 108, neutral by 78% with a number of 390 and negative by 0.2% with a number of 2 (Pasek et al., 2022). However, the overarching framework, which addresses sentiment analysis using Python, has not been widely discussed in previous studies.

Seeing that cost and time efficiency are important in this case, the use of an appropriate programming language is needed to perform sentiment analysis. Naïve Bayes is one of the classifications of data mining that can be used to answer these problems. However, the

Naive Bayes Method In Sentiment Analysis Of Presidential Candidates For The 2024
Election Using Python—Raditia Vindua et.al
548 | P a g e

overarching framework, which addresses sentiment analysis using the Python programming language and Naïve Bayes classification, has not been widely discussed in previous studies. Based on the problems found, the researcher conducted a study with the title "Naïve Bayes Method in Sentiment Analysis of Presidential Candidates for the 2024 Election using Python". This research will discuss the sentiment of Twitter users conducted using the Naïve Bayes classification, which is done using the Python programming language.

METHODS

The research methodology used in this study are Data Scraping, Preprocessing, data analysis use sentiment analysis and evaluation of sentiment results use Naïve Bayes Method
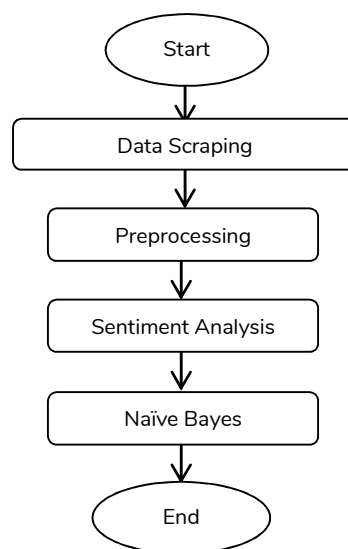

Figure 1. Research Methodology

There is a research methodology that can be seen in figure 1 which will be explained as follows:

Data Scraping

Data scraping is the process of extracting specific data from a web page or a document. For example, to scrape the names and prices of products from an e-commerce site, or the ratings and reviews of movies from a streaming platform. Data scraping can be done manually, by copying and pasting the data, or automatically, by using a script or a tool that can parse the HTML or XML code of the web page. Data scraping can be useful for collecting data for analysis, research, or comparison.

Preprocessing

Preprocessing is a stage that aims to facilitate the process of processing data to be processed at the next stage. The preprocessing process is carried out to select data and convert it into more structured data. The preprocessing process makes a lot of unused data will be eliminated first before the dataset is processed (H, 2015). In the preprocessing system, there are three phases that are completed, namely Cleaning, Case Folding and Stemming.

Naive Bayes Method In Sentiment Analysis Of Presidential Candidates For The 2024
Election Using Python—Raditia Vindua et.al
549 | P a g e

Cleaning

In the Cleaning stage, it is useful as a cleaner of words that are expected to reduce noise such as html, connections, hashtags, usernames, and content. In addition to cleaning words, this process also removes signs such as periods (.), commas (,) and also other accent marks.

Case Folding

The Case Folding stage is one of the stages for text pre-processing, which is to convert all letters contained in a text or sentence into a uniform with lowercase to become a standard form that is consistent as a whole.

Stemming

Stemming stage, is the stage of finding the root (base) of each word resulting from the elimination of stopwords by removing affixes in front or affixes behind words. The stemming stage uses the help of a library called Sastrawi in the Python programming language.

After going through preprocessing, the data is converted into vectors and given a weighting value for each word so that it can be calculated and processed with a classification algorithm. Word weighting is the stage of giving weight to each word that appears in a document (Nurrun Muchammad Shiddieqy et al., 2016).

Sentiment Analysis

At the analysis stage, what is done is to analyze the results of calculations and program codes using sentiment analysis. Sentiment analysis is a field of science that analyzes people's opinions, sentiments, evaluations, and emotions towards certain products, services, individuals, organizations, problems, topics, events (Bhatia et al., 2018)

Sentiment analysis is carried out to determine the direction of polarity of opinion sentences, so that constituent members of positive, neutral and negative sentence groups can be found. This research was conducted using the Textblob library in Python. Textblob is one of the libraries provided by Python for processing in the field of Natural Language Processing that can provide word tagging, word extraction, word translation and sentiment analysis. Natural Language Processing (NLP) learning aims to create machines that can understand and understand the meaning of human language then give appropriate responses (Chadha et al., 2015). But currently textblobs can only be used in English, so because this study uses Indonesian data, the tweets used are translated into English first. The characteristics of sentiment analysis are as follows:

1. The categorization of sentiments that will distinguish between subjective and objective sentences
2. Level of analysis. The level of analysis is divided into 3 parts, namely message level, sentence level and aspect level.
3. Opinions that provide comparisons to something and opinions that are just opinions. This means that everyone can give an opinion by comparing one thing with another or just giving an opinion.
4. The division of opinions becomes explicit and implicit. Opinions expressed honestly, unequivocally and straightforwardly and clearly or opinions expressed unclearly.

Naive Bayes Method In Sentiment Analysis Of Presidential Candidates For The 2024 Election Using Python—Raditia Vindua et.al

550 | P a g e

Naïve Bayes

The sentiment analysis process begins with building training data, then this training data is processed with the Naïve Bayes classification algorithm to produce a classification model. This model will later be used by classifiers as a basis for conducting the classification process. In building training data, the initial data in the form of comments that have been labeled class goes through the pre-processing stage.

Furthermore, the data results will be visualized to facilitate the reading of sentiment analysis results. Analysis can be in the form of the level of accuracy of a method of this research which can later be compared with other methods so that this research can be optimal.

RESULTS AND DISCUSSION

This section contains results and discussions on Data Scraping, preprocessing, sentiment analysis and naïve bayes.

Data Scraping

In the data scrapping process, the researcher conducted a manual search using google colab which was carried out on February 6, 2024. The results of data scrapping with the keyword "Anies Muhaimin" with the number of tweets is 65 tweets that can be seen 5 examples of tweets in Table 1 below.

Table 1. Examples of data scrapping with the keyword "Anies Muhaimin"

created_at	username	full_text
Tue Feb 06 07:07:07 +0000 2024	kangronee	@cakimiNOW Rakyat tidak sungkan, berani mendekati dan berani bersuara kepada Anies - Muhaimin
Tue Feb 06 06:44:24 +0000 2024	edsuryaa	@kompascom JAGA DAN KAWAL AGAR PILPRES 2024 JURDIL INSHAA ALLAH PASANGAN ANIES MUHAIMIN NO 1 MENANG SATU PUTARAN AAMIIN AAMIIN YA RABBAL'ALAMIIN
Tue Feb 06 06:35:03 +0000 2024	iyamrenzia	Monmaap Mas mau tanya, ini gak salah baca visi misi kan ya? Yang punya program dana desa jadi 5M dan pemerataan membangun 40 kota itu Anies &amp Muhaimin. Prabowo Gibran malah gencar lanjutin IKN yang mana jadi gak merata tuh, karna fokusnya cuma 1 kota dengan anggaran kira2 400T ðŸ™
Tue Feb 06 06:33:32 +0000 2024	jallainspirate	@Yurissa_Samosir @AdamVelcro @AdeDhani9 @aking_adit @AkalSehat2024 @chaidar_2000 @Kang_Toniee @NohCeukenohhh @choymarkochoy @mustafaagus007 ayoo gas 1 putaran kita goncang bsk di JIS baliho terbesar Anies Muhaimin AMIN
Tue Feb 06 06:20:25 +0000 2024	st_nurdin52033	@Yattie2023 Anies-Muhaimin. Cerdas,tegas,ber-wibawa. Perubahan yg lebih baik. Rakyat makmur adil untuk semua. Ayo tggl 14 februari 2024. Coblos No.1

Naive Bayes Method In Sentiment Analysis Of Presidential Candidates For The 2024
Election Using Python—Raditia Vindua et.al
551 | P a g e

The results of data scrapping with the keyword "Prabowo Gibran" with the number of tweets is 105 tweets, which can be seen 5 examples of tweets in Table 2 below.

Table 2. Examples of data scrapping with the keyword "Prabowo Gibran"

created_at	username	full_text
Tue Feb 06 07:09:14 +0000 2024	DiahPermad4	Pakar Hukum Tata Negara Ungkap Alasan Ini yang Bikin Status Pendaftaran Prabowo-Gibran ttp Konstitusional https://t.co/ft8Fl9sP1M https://t.co/bG9u2hmCkJ
Tue Feb 06 07:09:08 +0000 2024	pejuang_receeeh	@icannn___ @kikysaputrii Kesimpulanya indonesia maju bersama prabowo gibran âœŒï¸
Tue Feb 06 07:09:03 +0000 2024	gusput_08	benar sekali, prabowo-gibran @gibran_tweet tak layak dipilih. pelanggar HAM dan pelanggar etik, anak haram konstitusi #JokowiPerusakDemokrasi https://t.co/PwwWwq55cR
Tue Feb 06 07:08:42 +0000 2024	salshaaul_	Zulkifli Hasan ajak seluruh kader dan simpatisan PAN berjuang bersama, memenangkan Prabowo - Gibran di Pemilu 2024. dekade08
Tue Feb 06 07:08:37 +0000 2024	Rudigesper	Sementara itu, Wiranto mengungkapkan, tidak ada alasan untuk tidak memilih Prabowo-Gibran. #IndonesiaSentris #IndonesiaHijau #02Melanjutkan #AnakMudaIndonesiaEmas Prabowo Subianto

The results of data scrapping with the keyword "Ganjar Mahfud" with the number of tweets is 76 tweets that can be seen 5 examples of tweets in Table 3 below.

Table 3. Examples of data scrapping with the keyword "Ganjar Mahfud"

created_at	username	full_text
Tue Feb 06 07:11:52 +0000 2024	Ganjar_Mahfud	Ganjar Bertemu Komunitas Tukang Cukur, Dengarkan Keluhan #3sakti #GanjarMahfud #GanjarPresiden #GanjarMahfud2024 #Pilpres2024 #Pemilu2024 #SalamM3tal #M3nangTotal #SatSet #TukangCukur #Asgar #Garut #JaBar #Cukur https://t.co/a1E3ssryh5
Tue Feb 06 07:11:08 +0000 2024	ArizaSuze95777	@kumparan Masyarakat Indonesia all in ganjar mahfud ðŸ'
Tue Feb 06 07:11:07 +0000 2024	denni_sauya	Yuhuu Gaspol Hajatan Rakyat Ganjar Mahfud ðŸ¤Ÿ SABTU, 10 FEBRUARI 2024 ~ SOLO DARI NGARSOPURO KE BENTENG VASTENBURG 07.00-10.00 WIB ~ SEMARANG LAPANGAN PANCASILA SIMPANG LIMA

Naive Bayes Method In Sentiment Analysis Of Presidential Candidates For The 2024 Election Using Python—Raditia Vindua et.al

552 | P a g e

created_at	username	full_text
Tue Feb 06 07:10:47 +0000 2024	berita_jabarku	13.00-17.00 WIB @ganjarpranowo @mohmahfudmd #YakinPilihGanjar #GanjarMahfud2024 #GanjarPresidenRakyat https://t.co/GEOdPRSC4J Ganjar-Mahfud Memberi Bukti, bukan hanya janji Umay Tasya Farasya Rocky Gerung Yeyen Agak Laen King Indo Ketua KPU ISTJ INFJ Rental PS Merangkul Selen Bahlil https://t.co/3pgGqjK6SR
Tue Feb 06 07:10:35 +0000 2024	Cindyherlinn_	Meskipun menyadari bahwa pengelolaan bisnis sendiri adalah kunci keberhasilan, Pak Arsjad Rasjid menghadapi dilema antara mempertahankan pekerjaan atau fokus sepenuhnya pada bisnisnya. Ketua TPN Ganjar Mahfud #Bisnis #arsjadrasjid #inklusifkolaboratif

The results above are the results of data scrapping from Twitter using Google Colab. Based on the results of the above collection, the data is continued to the preprocessing process.

Preprocessing

This preprocessing process is done using the help of libraries in the Python programming language. In the preprocessing system, there are 3 phases that are completed, namely Cleaning, Case Folding and Stemming.

Cleaning

The cleaning process has 3 phases, namely removing duplicate data, eliminating null data and cleaning unnecessary characters. To get rid of duplicate data and null data use the pandas library. As for eliminating characters that do not need to use the regex library or it can also be called regular expression. This is the source code for the cleaning process.

```
[4]:  #cleaning duplicate data
      df=df.drop_duplicates(subset=['full_text'])
      df.duplicated().sum()

[4]:  0

[5]:  #cleaning data isnull
      df=df.dropna()
      df.isnull().sum()

[5]:  full_text    0
      username     0
      dtype: int64

[6]:  df.shape

[6]:  (57, 2)

[7]:  #cleaning character
      def cleanTxt(text):
          text = re.sub('@[^\s]+',''', text) #Removing @mentions
          text = re.sub(r'http\S+','', text) #Removing @username
          text = re.sub(r'#([^\s]+)', r'\1', text) # Removing '#' hash tag
          text = re.sub(r'RT[\s]+', '', text) # Removing RT
          text = re.sub(r'https?:\/\/\S+', '', text) # Removing hyperlink
          text = re.sub(r'\w*\d\w*', '',text).strip() # Removing number
          return text

      df['full_text'] = df['full_text'].apply(cleanTxt)
```

Figure 2. Source Code Cleaning Data

The results of cleaning for Anis Muhaimin's data from 65 data are 57 data, for Prabowo Gibran's data from 105 data, 105 data remain after going through the cleaning process, while for Ganjar Mahfud's data from 76 data, 75 data remain.

In the case folding process, it is useful to change all letters in the document to lowercase letters. This is the source code for the case folding process.

```
#Case Folding
df['full_text'] = df['full_text'].str.lower()
df
```

Figure 3. Source Code Case Folding

In the stemming process, it is useful to remove suffixes in front of or suffixes after words to find the base word. In the stemming process, it uses a sastrawi library which contained in the python programming language. In this process, the researcher installs the sastrawi library first by using the pip command in Figure 4 below.

Naive Bayes Method In Sentiment Analysis Of Presidential Candidates For The 2024 Election Using Python–Raditia Vindua et.al

554 | P a g e

Figure 4. Sastrawi library installation process

After the installation process is complete, the implementation of the stemming stage in the program code in Figure 5 below.


```python
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

#Create a function to stemming the tweets
def stemm(text):
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    return stemmer.stem(text)

#Clean tweets
df['full_text'] = df['full_text'].apply(stemm)

#Show stemming tweets
df
```

Figure 5. Source code stemming

After all the preprocessing processes are carried out on all data, the results of the preprocessing are saved into a new file which will later be continued in the sentiment analysis process by python and classified using the Naive Bayes method using the python language.

Data Analysis

Sentiment analysis is carried out to find out the direction of the polarity of opinion sentences, so that the constituent members of the positive, neutral and negative sentence groups can be found. In this study, sentiment analysis was carried out using the Textblob library in Python. However, currently textblobs are only available in English, therefore in this study using Indonesian data, it is translated into English first. There are two functions in sentiment classification that are used to classify or predict tweets, namely polarity and subjectivity as shown in the source code in Figure 6 below.

Naive Bayes Method In Sentiment Analysis Of Presidential Candidates For The 2024
Election Using Python—Raditia Vindua et.al
555 | P a g e

```
from textblob import TextBlob

#Create a function to get the subjectivity
def getSubjectivity(text):
    blob_object = TextBlob(text)
    text_indo = blob_object.translate(from_lang='id', to='en') #translate bahasa Indonesia to English
    return text_indo.sentiment.subjectivity

#Create a function to get the polarity
def getPolarity(text):
    blob_object = TextBlob(text)
    text_indo = blob_object.translate(from_lang='id', to='en') #translate bahasa Indonesia to English
    return  text_indo.sentiment.polarity

#Create two new columns 'Subjectivity' & 'Polarity'
df['Subjectivity'] = df['full_text'].apply(getSubjectivity)
df['Polarity'] = df['full_text'].apply(getPolarity)

#Show new dataframe 'Subjectivity' & 'Polarity'
df
```

Figure 6. Source code sentiment analysis

Once all tweets have been classified, the polarity of all tweets will be displayed in an average form, and the labels 'Positive', 'Neutral' or 'Negative' correspond to the mean of the polarity of all tweets. Source code the determination of positive, neutral and negative classes can be seen in Figure 7

```
#Create function to compute negative (-1), neutral (0) and positive (+1) analysis
def getAnalysis(score):
    if score > 0:
        return 'Positif'
    elif score == 0:
        return 'Netral'
    else:
        return 'Negatif'
df['Analysis'] = df['Polarity'].apply(getAnalysis)

#Show dataframe
df
```

Figure 7. The result of sentiment analysis

Subjectivity is used to see the value of a tweet that is an opinion or factual. The higher the subjectivity, the more the tweet can be said to be an opinion. Then the results of the sentiment analysis of the above tweets are calculated using the source code to determine the number of positive tweets, neutral tweets and negative tweets.

Naive Bayes Method In Sentiment Analysis Of Presidential Candidates For The 2024
Election Using Python–Raditia Vindua et.al
556 | P a g e

```python
#count the number of negative (-1), neutral (0) and positive (+1) values
tweet_positif = df[df.Analysis == 'Positif']
tweet_netral = df[df.Analysis == 'Netral']
tweet_negatif = df[df.Analysis == 'Negatif']

#Show the number
print("Hasil Sentimen")
print("positif: ", len(tweet_positif))
print("netral: ", len(tweet_netral))
print("negatif: ", len(tweet_negatif))
```

Figure 8. Count the number of negative, neutral and positive values

After all tweets were analyzed using python, the data was then classificated using the Naive Bayes classification.

Naive Bayes

Classification is carried out by applying the Naïve Bayes classification method so that it can be used to draw conclusions from the results of the classification accuracy.

```python
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import GaussianNB

vectorizer = TfidfVectorizer (max_features=2500)
model_g = GaussianNB()
```

```python
v_data = vectorizer.fit_transform(df['full_text']).toarray()

print (v_data)
```

```
[[0. 0. 0. ... 0. 0. 0. ]***
```

```python
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(v_data, df['Analysis'], test_size=0.2, random_state=0)
model_g.fit(X_train,y_train)
```

```
▸ GaussianNB
```

```python
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score

y_preds = model_g.predict(X_test)

print(confusion_matrix(y_test,y_preds))
print(classification_report(y_test,y_preds))
print('nilai akurasinya adalah ',accuracy_score(y_test, y_preds))
```

Figure 9. Source code Naive Bayes classification

From the results above, Anies Muhaimin got an accuracy score of 75%, for Prabowo Gibran got an accuracy score of 81%, while for Ganjar Mahfud got an accuracy score of 60%.

Naive Bayes Method In Sentiment Analysis Of Presidential Candidates For The 2024 Election Using Python–Raditia Vindua et.al

557 | P a g e

Then the results of the classification data will be visualized to make it easier to read the results of sentiment analysis using the source code below.

```python
import matplotlib.pyplot as plt

labels = ["Positif", "Netral", "Negatif"]
quantity = [len(tweet_positif), len(tweet_netral), len(tweet_negatif)]

colors = ['yellowgreen', 'lightskyblue', 'lightcoral']

plt.title('Hasil Sentimen Anies Muhaimin tahun 2024')
plt.pie(quantity, labels=labels, colors=colors,
        autopct='%1.2f%%', shadow=True)

plt.axis('equal')
plt.show()
```

Figure 10. Source code Pie Chart Visualization

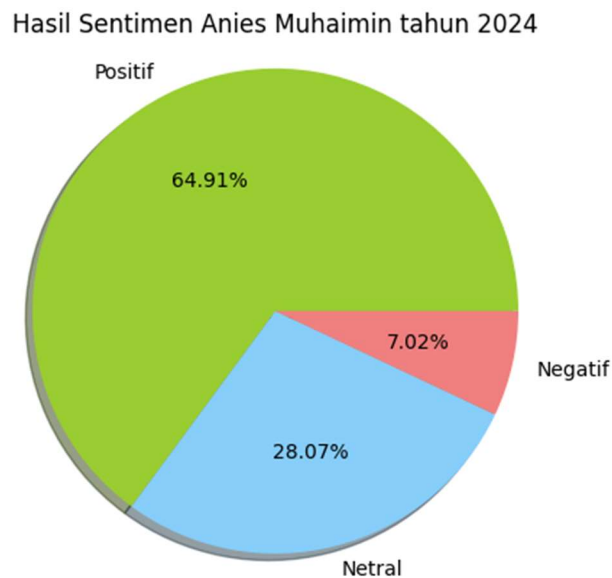This is the result of the Pie Chart visualization for data from Anies Muhaimin, Prabowo Gibran and Ganjar Mahfud.

Figure 11. Pie Chart Anies Muhaimin

By entering the keyword "Anies Muhaimin", the public's opinion about Anies Muhaimin was 64.91% Positive, 28.07% Neutral and 7.02% Negative.
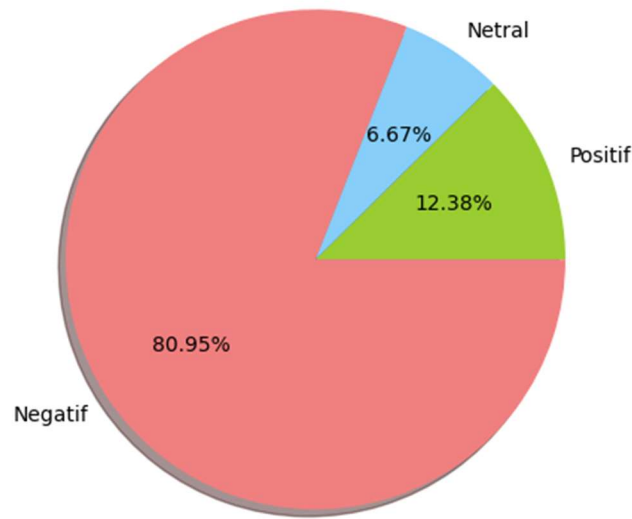
Hasil Sentimen Prabowo Gibran tahun 2024

Figure 11. Pie Chart Prabowo Gibran

By entering the keyword "Prabowo Gibran", the public's opinion about Prabowo Gibran is 12.38% Positive, 6.67% Neutral and 80.95% Negative.

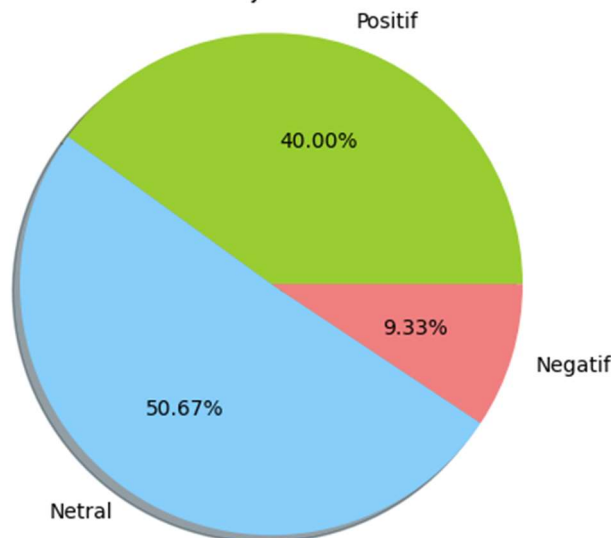Hasil Sentimen Ganjar Mahfud tahun 2024

Figure 11. Pie Chart Ganjar Mahfud

Meanwhile, with the keyword "Ganjar Mahfud", the public's opinion about Ganjar Mahfud is 40% Positive, 50.67% Neutral and 9.33% Negative. From the above results, it can be concluded that python and Naïve Bayes classification can classify tweets from twitter so that we can identify public opinion about presidential candidates for the 2024 election.

Naive Bayes Method In Sentiment Analysis Of Presidential Candidates For The 2024
Election Using Python—Raditia Vindua et.al
559 | P a g e

However, the decision to choose a presidential candidate which is better can be returned to the opinion of each community.

CONCLUSION

Based on the results of the research conducted, it can be concluded that the Python programming language can be applied to analyze opinion sentiment from Twitter about the 2024 Election Presidential Candidate using the naïve bayes classification method using the Python programming language. Where before conducting sentiment analysis, a preprocessing process must be carried out to select data and convert it into more structured data. The results of this research showed that Python carried out sentiment analysis with the sentiment percentage results for candidate Anies Muhaimin with a positive class of 64.91%, neutral 28.07% and negative 7.02% with a Naïve Bayes accuracy value of 75%. For candidate Prabowo Gibran, the positive class is 12.38%, neutral 6.67% and negative 80.95% with a Naïve Bayes accuracy value of 81%. Meanwhile, the candidate Ganjar Mahfud has a positive class of 40%, neutral 50.67% and negative 9.33% with a Naïve Bayes accuracy value of 60%. Naïve Bayes also managed to make calculations with an accuracy score of 75% for Anies Muhaimin, for Prabowo Gibran got an accuracy score of 81%, while for Ganjar Mahfud got an accuracy score of 60%. And Python can visualize the data into various kinds of graphs so that the data is easier to understand. Therefore, it can be concluded that python and Naïve Bayes classification can classify tweets from twitter so that we can identify public opinion about presidential candidates for the 2024 election.

REFERENCE

Aggrawal, N. (2018). Detection of Offensive Tweets: A Comparative Study. *Computer Reviews Journal*, *Vol 1 No 1*, 75–89.

Asshiddiqie, J. (2007). *Pokok-Pokok Hukum Tata Negara Indonesia Pasca Reformasi* (xviii). PT Bhuana Ilmu Populer. https://simpus.mkri.id/opac/detail-opac?id=6455

Bhatia, S., Sharma, M., & Bhatia, K. K. (2018). Sentiment Analysis and Mining of Opinions. *Studies in Big Data*, *30*(May), 503–523. https://doi.org/10.1007/978-3-319-60435-0_20

Cambria, E., Poria, S., Bisio, F., Bajpai, R., & Chaturvedi, I. (2015). The CLSA Model: A Novel Framework for Concept-Level Sentiment Analysis. *Springer International Publishing Switzerland*, 3–22. https://www.researchgate.net/publication/273635373_The_CLSA_Model_A_Novel_Framework_for_Concept-Level_Sentiment_Analysis

Chadha, N., Gangwar, R. C., & Bedi, R. (2015). Current Challenges and Application of Speech Recognition Process using Natural Language Processing: A Survey. *International Journal of Computer Applications*, *131*(11), 28–31. https://doi.org/10.5120/ijca2015907471

Coletta, L. F. S., Ponti, M., Hruschka, E. R., Acharya, A., & Ghosh, J. (2019). Combining clustering and active learning for the detection and learning of new image classes. *Neurocomputing*, *358*(September), 150–165.

Naive Bayes Method In Sentiment Analysis Of Presidential Candidates For The 2024 Election Using Python—Raditia Vindua et.al

560 | P a g e

https://doi.org/10.1016/j.neucom.2019.04.070

Firdaus, A., & Firdaus, W. I. (2021). Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan). *Jurnal JUPITER*, *13*(1), 66.

Fitriyyah, S. N. J., Safriadi, N., & Pratama, E. E. (2019). Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *5*(3), 279. https://doi.org/10.26418/jp.v5i3.34368

H, A. T. J. (2015). Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining. *Informatika UPGRIS*, *1*, 1–9.

Hidayatullah, A. F., Aulia, A., Yusuf, F., Juwairi, K. P., Abida, R., & Nayoan, N. (2019). Identifikasi konten kasar pada tweet bahasa Indonesia. *Jurnal Linguistik Komputasional*, *2*(1), 1–5. http://inacl.id/journal/index.php/jlk/article/view/15

Indrayuni, E., & Wahyudi, M. (2015). Penerapan Charachter N-Gram Untuk Sentiment Review Hotel Menggunakan Algoritma Naive Bayes. *Konfrensi Nasional Ilmu Pengetahuan Dan Teknologi (KNIT)*, 88–93.

Istiani, N., & Islamy, A. (2020). Fikih Media Sosial Di Indonesia. *Asy Syar'iyyah: Jurnal Ilmu Syari'Ah Dan Perbankan Islam*, *5*(2), 202–225. https://doi.org/10.32923/asy.v5i2.1586

Jailani. (2015). Sistem Demokrasi di Indonesia Ditinjau dari Sudut Hukum Ketatanegaraan. *Inovatif: Jurnal Ilmu Hukum, Vol. 8, No.*, *8*(1), 134–147.

Lukmana, D. T., Subanti, S., & Susanti, Y. (2019). Analisis Sentimen Terhadap Calon Presiden 2019 Dengan Support Vector Machine Di Twitter. *Seminar Nasional Penelitian Pendidikan Matematika (SNP2M) 2019 UMT*, *2002*, 154–160.

Mahbubah, L. D., & Zuliarso, E. (2019). Analisa Sentimen Twitter Pada Pilpres 2019 Menggunakan Algoritma Naive Bayes. *Sintak*, 194–195. https://www.unisbank.ac.id/ojs/index.php/sintak/article/view/7585

Nurrun Muchammad Shiddieqy, H., Paulus Insap, S., & Wing Wahyu, W. (2016). Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter. *Seminar Nasional Teknologi Informasi Dan Komunikasi*, *2016*(March), 57–64.

Pasek, P., Mahawardana, O., Arya, G., Agus, I. P., & Pratama, E. (2022). Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap " Figure Pemimpin " Menggunakan Python. *JITTER-Jurnal Ilmiah Teknologi Dan Komputer*, *3*(1).

Pratama, S. F., Andrean, R., & Nugroho, A. (2019). Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analysis. *JOINTECS (Journal of Information Technology and Computer Science)*, *4*(2), 39. https://doi.org/10.31328/jointecs.v4i2.1004

Subiyanto, A. E. (2020). Pemilihan Umum Serentak yang Berintegritas sebagai Pembaruan Demokrasi Indonesia. *Jurnal Konstitusi*, *17*(2), 355. https://doi.org/10.31078/jk1726

Wibowo, W., Ulama, B. S. S., & Azies, H. Al. (2020). *Modul-Prokom-DSB-versi-18-agustus-BW-compressed_2.pdf* (p. 135).

Naive Bayes Method In Sentiment Analysis Of Presidential Candidates For The 2024 Election Using Python–Raditia Vindua et.al

561 | P a g e