JAISI
JOURNAL OF APPLIED INFORMATON SYSTEM AND INFORMATIC

Analysis Of Twitter User Sentiment To Tiktok Shop Using Naïve Bayes And Decision Tree Algorithms

Soleh Jafar Sidiq [a] , Andi Nur Rachman[*b]

[a] *Departmen of Informatics Enginnering, Universitas Siliwangi, Tasikmalaya, Indonesia*
[b]*Departmen of Informations System, Universitas Siliwangi, Tasikmalaya, Indonesia*
Corresponding author: andy.rachman@.unsil.ac.id

***Abstract*— The growth of internet users is fantastic, before the pandemic the figure was only 175 million. While the latest data from the Asosiasi Penyelenggaraan Jasa Internet Indonesia (APJII), in 2022 internet users in Indonesia will reach around 210 million. One of the influences on the increasing number of internet users in Indonesia is the increasing number of buying and selling activities through internet media. At this time there are various kinds of e-commerce applications. One of the latest e-commerce in Indonesia is Tiktok Shop. Tiktok shop is a new feature of the Tiktok application which was established on April 17, 2021. The development of Tiktok shop cannot be separated from the people who use this feature. Many people give opinions about Tiktok Shop on one of the social media, namely Twitter. Twitter is a place to get data expressed by the public through tweets posted to the timeline. The data used are tweets in Indonesian with a dataset of 1000 tweets. The data is then processed to be analyzed for knowledge. The analysis is done with Naïve Bayes and Decision Tree methods. The accuracy results of the Naïve Bayes algorithm are 90% and the Decision tree algorithm is 93%, so the Decision Tree algorithm is better for classifying sentiment analysis of twitter users towards Tiktok Shop with a data division of 90%.**

***Keywords*— Sentiment Analysis ,Tiktok Shop, Twitter, Naïve Bayes dan Decision Tree.**

Manuscript received 15 Jun. 2023; revised 29 August. 2023; accepted 2 Nov. 2023. Date of publication Nov. 2023.
International Journal of Applied Information Systems and Informatics is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

I. INTRODUCTION

As time goes by, the development of e-commerce in Indonesia experiences very rapid growth and development every year. This is influenced by the increasing number of internet users in Indonesia who want everything that is effective and efficient. Not only that, the relatively large number of Indonesian people is a driving factor for the development of e-commerce businesses.

One of the influences on the increase in the number of internet users in Indonesia is the increasingly widespread buying and selling activities via internet media. Currently there are various kinds of e-commerce applications. One of the newest e-commerce in Indonesia is Tiktok Shop. Tiktok is a social media and music video application that was officially launched in 2016 by Zhang Yiminy from China. Tiktok shop is the newest feature developed by Tiktok which is used for buying and selling transactions directly through the Tiktok application. Tiktok shop is a unique innovation because it allows users to simultaneously use social media and carry out buying and selling transactions in one application. This is what differentiates Tiktok shop from other e-commerce platforms.

There are many public opinions about Tiktokshop on various social media. One of them is Twitter. According to (B.K, 2010), social media is a digital technology label that allows people to connect, interact, produce and share content messages. According to (Hadi, 2010), the definition of Twitter is a microlog site that provides facilities for users to send text messages with a maximum character length via SMS, instant messengers and electronic mail. Opinions expressed by the public on Twitter regarding Tiktok Shop are then processed for analysis to become knowledge. The analysis was carried out using the Naïve Bayes and Decision Tree methods which will be classified into positive, negative and neutral sentiment.

From the problem above, sentiment analysis was carried out using public opinion taken from tweets on the social media Twitter. The task of sentiment analysis is to group positive, negative and neutral texts based on the text contained. In

conducting this research the author used the Naïve Bayes and Decision Tree methods.

Naive Bayes is a simple probabilistic classification method. This method will calculate a set of probabilities by adding up the frequencies and combinations of values from a given dataset. The advantages of using the Naïve Bayes method are only requires a small amount of training data to estimate the parameters (means and variances of the variables) required for classification. Meanwhile, the Decision Tree method is a very popular and practical approach in machine learning for solving classification problems. Apart from being built relatively quickly, the results of the model built are easy to understand (Anam et al., 2021).

Based on the problem above, the author carried out sentiment analysis using public opinion taken from tweets on the social media Twitter. The task of sentiment analysis is to group positive, negative and neutral texts based on the text contained. In conducting this research the author used the Naïve Bayes and Decision Tree methods.

II. MATERIALS AND METHOD

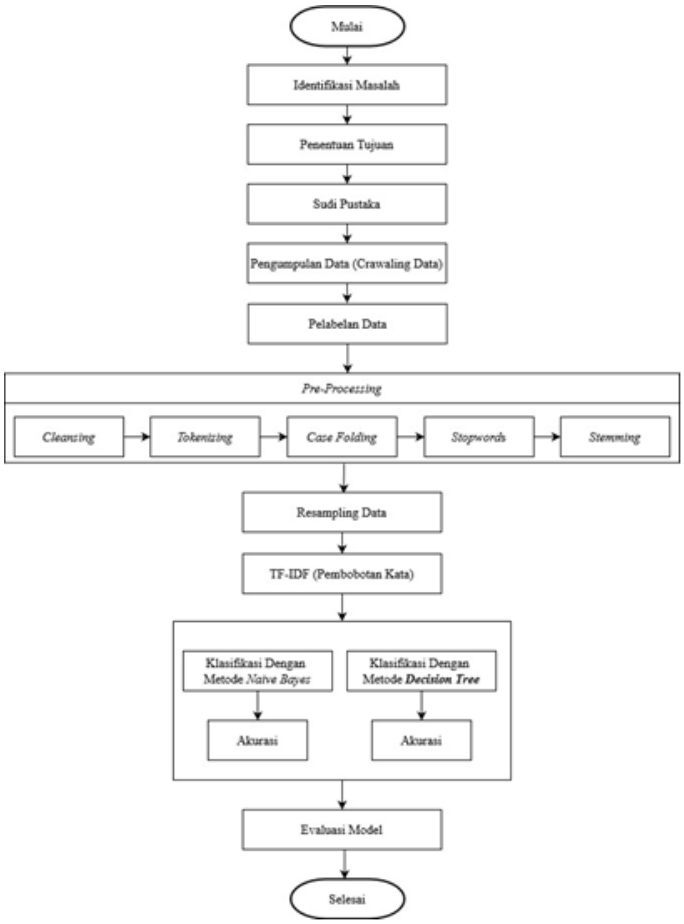

Fig. 1. Research Methodology

Aeasy way to comply with the conference paper formatting requirements is to use this document as a camera-ready template.

A. Identification of problems

Observing one of the marketplaces in Indonesia, namely Tiktok Shop. There are many opinions expressed by the public so that an analysis of the sentiment of Twitter social media users towards Tiktok Shop can be carried out.

B. Goal Setting

The goal setting stage is useful for clarifying targets in research. The aim of this research is to classify public opinion on Twitter social media about Tiktok Shop using the Naïve Bayes and Decision Tree methods

C. Literature review

Literature study is a very important stage in the research process. Literature study is defined as the stage where information is collected relating to the research being carried out. This is done by looking for materials that are reference sources for research so that the research can run well. The sources used as references are journals, theses, theses, books and other sources related to research.

D. Data collection

The data collection stage is divided into three, namely data categories and data collection from Twitter social media and data sharing (sampling technique). Determining data categories will make it easier to collect data.

E. Data Pre-Processing

At this stage, the researcher processes the data obtained from the data collection stage with the aim of eliminating several problems that can interfere with data processing:

1. Data Cleaning

   Cleaning is a technique that will make documents or text clean from unnecessary characters such as hashtags, usernames, URLs, emoticons and HTML. This cleaning is useful so that noise can be reduced so that the classification process can run well.
2. Tokenizing

   Tokenizing is a process where documents composed of words are broken down into words that break the string sequence.
3. Case Folding

   Case Folding is a process in text preprocessing where in this process the document or text which consists of a sequence of letters will be changed to lowercase as a whole so that there are no similarities in the document.
4. Stop words

   Stopwords, namely removing vocabulary that does not have a special meaning or unique words that can represent the document. Examples of such vocabulary are and, this, that, are and so on.
5. Stemming

   Stemming is a useful technique for returning words to their base words because it removes prefixes, insertions, combinations and also suffixes. For example, the word take means that in the stemming technique it will change to take.

F. Resample

Resample is a technique for manipulating training data to correct the skewness of class distributions, such as random oversampling and random undersampling. This technique is used to increase accuracy to greater levels. The term original sample is used to refer to the subset that is first taken from the population, before resampling is carried out, namely the process of resampling from samples that we have taken from the population, while the term bootstrap sample (resample) is used to refer to samples that we have resampled from the original sample . (Zhang et al., 2011).

G. Word Weighting (TF-IDF)

Term Frequency Inverse Document Frequency (TF-IDF) is a weighting carried out after extracting news articles (Ariadi, & Fithriasari, 2015). The process of the TF-IDF method is to calculate weights by integrating Term Frequency (TF) and Inverse Document Frequency (IDF). The step in TF-IDF is to find the number of words we know (TF) after multiplying them by how many news articles where a word appears (IDF).

H. Classification

One of the data mining methods is classification. Classification is used to predict categories from a set of data with various attributes. Before making predictions, a learning process is first carried out. The data used in the learning process is called training data, while the data used in the prediction process is called testing data (Faizal, 2016).

I. Evaluation

Evaluation is the final process of all processes in this research. The results of the evaluation carried out are in the form of accuracy, precision and recall. So from these results you can see the classification results of the Naïve Bayes and Decision Tree algorithms.

III. Result and Discussion

A. Data collection

The data taken comes from tweets on Twitter social media. The data is identified and grouped into words that contain positive sentiment, negative sentiment or neutral. Twitter user tweets data taken about Tiktok Shop is only for the period of the month of Ramadhan, namely March 22 2023 - April 22 2023. Classification process using Naïve Bayes and Decision tree methods. Data was taken using a crawling technique totaling 1000 data. A total of 1000 came in used for the formation of a classification model which was labeled manually. The data is divided into training data and test data with a ratio of 80%:20%. The data used to form the model is resampled in the Python programming language to make better prediction decisions so that the data becomes 1932. The following is a Pie chart of the proportion of sentiment about TikTok from 1000 data tweets containing positive sentiment is 64.4%, tweets containing negative sentiment of 14.0% and neutral of 21.6%. Figure 2 explains the sentiment proportion pie chart.

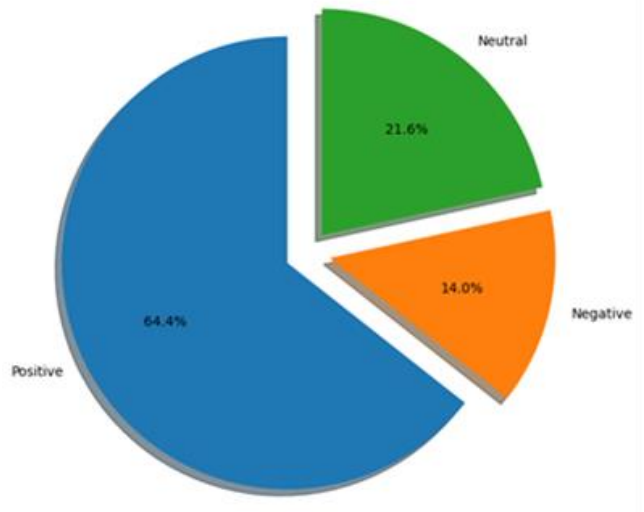

Fig. 2. Pie Chart

Table 1 explains that the documents acquired are documents labeled as text. At this stage the data is still intact or has not been cleaned so that the data to be processed is still mixed with other characters that are still attached to the data.

TABLE 1
DATA OBTAINED

No	Tweets
1	siap menerima pap kamu pakai mukenah yang beli di tiktokshop
2	Ini juga ada beberapa rekomendasi produk yang oke di tiktokshop. Udah aku coba. Cus beli aja di keranjang kuning, pas aku live biar makin murah atau langsung beli, bebas weh. 😄 https://t.co/PjTdfHC9wX
3	Kemarin aku beli cuma 229😭 👍 tiktokshop juga
4	@loafsajaey udah jarang beli di shopee gara gara ada tiktokshop https://t.co/5U5KMtbLWj
5	@Roseallday di tiktokshop diskon parah 😩 😩 😩 😩 😩
6	@sbtcon duh.. kalau jualan di tiktokshop mending gausah nyalain cod deh. sumpah serem banget, banyak bocil2 iseng.
7	@apesikepobgt @arionkarelll @reykalandra nanti aku cari, mau di shoppe atau tiktokshop?
8	@jaethejamal @spidwrmark katanya di tiktokshop cuma pakai jnt jadi lama
9	@princejenow 40k karna aku pertama kali beli di tiktokshop makanya bisa dapat harga segitu
10	kemarin aku jajan ini di tiktokshop murah banget loh 90rb dapat 3 aku beli barang temen-teman aku deh😱 1 nya jadi 30rb doang ayo serbu https://t.co/WOmLIk6CbX

B. PreProcessing

At this stage the data that has been acquired will be further processed using several processes. The purpose of preprocessing is to make features clear, reduce or eliminate noise, convert original data to meet needs, enlarge and reduce data according to needs.

TABLE 2
PRE-PROCESSING RESULTS

D1	Siap	terima	pap	Kamu	pakai
	mukenah		beli		tiktokshop
D2		juga	ada	Beberapa	rekomendasi
	produk		oke		tiktokshop
	Udah	aku	coba		beli
	aja		keranjang	Kuning	pas
	aku	live	biar	Makin	murah
		langsung	beli	bebas	
D3	kemarin	aku	beli	cuma	tiktokshop
	juga				
D4	udah	jarang	beli		shopee
	gara	gara	ada	tiktokshop	

D5		tiktokshop	diskon	parah	
D6	duh	kalau	jual		tiktokshop
	mending	gausah	nyalain	cod	
	sumpah iseng	seram	banget	banyak	bocil
D7	nanti	aku	cari	mau	
	shopee		tiktokshop		
D8	kata		tiktokshop	cuma	pakai
	jnt	jadi	lama		
D9	karna	aku	pertama	kali	beli
		tiktokshop	maka	bisa	dapat
	harga	segitu			
D10	kemarin	aku	jajan		
	tiktokshop	murah	banget		dapat
	aku	beli	barang	temen	teman
	aku			jadi	doang
	ayo	serbu			

C. Resample

This technique is used to make the classification model better. A total of 1000 data that had been manually labeled were then resampled so that the number of positive, negative and neutral labels was the same. A thousand data were sampled to produce 1932 data. The following is an image of the resample process in Python programming.

```
[124] from sklearn.utils import resample
      df_majority = new_data[new_data.Label=='Positive']
      df_minority1 = new_data[new_data.Label=='Negative']
      df_minority2 = new_data[new_data.Label=='Neutral']

      df_minority_upsampled1 = resample(df_minority1,
                                        replace=True,
                                        n_samples=new_data[new_data.Label =='Positive'].shape[0],
                                        random_state=123)
      df_minority_upsampled2 = resample(df_minority2,
                                        replace=True,
                                        n_samples=new_data[new_data.Label =='Positive'].shape[0],
                                        random_state=123)
      df_upsampled = pd.concat([df_majority, df_minority_upsampled1, df_minority_upsampled2])

[125] df_upsampled.info ()

      <class 'pandas.core.frame.DataFrame'>
      Int64Index: 1932 entries, 0 to 445
      Data columns (total 2 columns):
       #   Column  Non-Null Count  Dtype
      ---  ------  --------------  -----
       0   tweet   1932 non-null   object
       1   Label   1932 non-null   object
```

Fig. 3. Resample Process Using Python Programming

D. Term Frequency Inverse Document Frequency of record

Figure 8 explains the stages of word weighting or TF-IDF. The probability of their appearance in one document (D1 to D10) is calculated for words/terms.

TABLE 3
TF-IDF PROCESS RESULTS

No	Kata	Dokumen										DF	IDF
		D1	D2	D3	D4	D5	D6	D7	D8	D9	D10		
1	Ada	0	1	0	1	0	0	0	0	0	0	2	0,698
2	Aja	0	1	0	0	0	0	0	0	0	0	1	1
3	Aku	0	1	1	0	0	0	1	0	1	3	7	0,154
4	Ayo	0	0	0	0	0	0	0	0	0	1	1	1
5	banget	0	0	0	0	0	0	0	0	0	1	1	1
6	banyak	0	0	0	0	0	1	0	0	0	0	1	1
7	Barang	0	0	0	0	0	0	0	0	0	1	1	1
8	Bebas	0	1	0	0	0	0	0	0	0	0	1	1
9	beberapa	0	1	0	0	0	0	0	0	0	0	1	1
10	Beli	1	2	1	1	0	0	0	0	1	1	7	0,1549

- .

E. Classification

Classification is the next stage after the preprocessing, resample and word weighting (TF-IDF) stages. In this processing there are 2 stages, namely classification using the Naive Bayes and Decision tree methods and data accuracy. The following is the process carried out:

1. Classification using the Naïve Bayes Method

The data to be trained and tested has been divided, then the Naïve Bayes method is implemented for the classification process. The following are the results of forming a naïve Bayes model. Table 4.11 explains that in the experiment, a random dataset was used to obtain 80% training data and 20% testing data which will be classified using the Naïve Bayes learning model, so the results obtained are in the form of a confusion matrix calculating precision, recall and f1-score.

TABEL 4
CONFUSION MATRIX NAIVE BAYES ALGORITHM

Data Prediksi	Data Aktual			*Precision*	*Recall*	*F1-Score*
	Negatif	**Netral**	**Positif**			
Negatif	112	3	2	0.88	0.96	0.91
Netral	0	123	11	0.84	0.92	0.88
Positif	16	20	100	0.88	0.74	0.80

As explained in the experiment, a random dataset was used to obtain 80% training data and 20% testing data which would be classified using the Naïve Bayes learning model, so the results obtained were Matrix confusion. The following is a description of the precision, recall and F1-score equations.

Negative Class

$$Precision = \frac{TNg}{(TNg + FP + NNg)} = \frac{112}{(112 + 16 + 0)} = 0.88$$

$$Recall = \frac{TNg}{(TNg + FN + NgN)} = \frac{112}{(112 + 2 + 3)} = 0.96$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{(Precision \times Recall)} = 2 \times \frac{0.88 \times 0.96}{(0.88 + 0.96)} = 0.91$$

Neutral Class

$$Precision = \frac{TN}{(TN + PN + NgN)} = \frac{123}{(123 + 20 + 3)} = 0.84$$

$$Recall = \frac{TN}{(TN + NP + NNg)} = \frac{123}{(123 + 11 + 0)} = 0.92$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{(Precision \times Recall)} = 2 \times \frac{0.84 \times 0.92}{(0.84 + 0.92)}$$

$$= 0.88$$

Positive Class

$$Precision = \frac{TN}{(TN + NP + FN)} = \frac{100}{(100 + 11 + 2)} = 0.88$$

$$Recall = \frac{TP}{(TP + PN + FP)} = \frac{100}{(100 + 20 + 16)} = 0.74$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{(Precision \times Recall)} = 2 \times \frac{0.88 \times 0.74}{(0.88 + 0.74)}$$

$$= 0.80$$

Classification accuracy

$$Akurasi = \frac{TP + TNg + TNet}{TP + TNg + TNet + FP + FN + FNet}$$

$$= \frac{100 + 112 + 123}{100 + 112 + 123 + 13 + 16 + 23} = 0.865$$

2. Classification using the Decision Tree Method

Decision trees are one way of processing information to predict the future by creating classification or regression models in the form of a tree structure. explains that in the experiment, a random dataset was used to obtain 80% training data and 20% testing data which will be classified using the Naïve Bayes learning model, so the results obtained are in the form of a confusion matrix calculating precision, recall and f1-score.

TABLE 5
CONFUSION MATRIX DECISION TREE ALGORITHM

Data Aktual	Data Prediksi			*Precision*	*Recall*	*F1-Score*
	Negatif	**Netral**	**Positif**			
Negatif	114	1	2	0.85	0.97	0.91
Netral	0	128	6	0.91	0.96	0.93
Positif	20	11	105	0.93	0.77	0.84

As explained in the experiment, a random dataset was used to obtain 80% training data and 20% testing data which will be classified using the Decision Tree learning model, so the results obtained are in the form of a confusion matrix calculation. The following is a description of the precision, recall and F1-score equations.

Negative Class

$$Precision = \frac{TNg}{(TNg + FP + NNg)} = \frac{114}{(114 + 20 + 0)} = 0.85$$

$$Recall = \frac{TNg}{(TNg + FN + NgN)} = \frac{114}{(114 + 2 + 1)} = 0.97$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{(Precision \times Recall)} = 2 \times \frac{0.85 \times 0.97}{(0.85 + 0.97)} = 0.93$$

Neutral Class

$$Precision = \frac{TN}{(TN + PN + NgN)} = \frac{128}{(128 + 11 + 1)} = 0.91$$

$$Recall = \frac{TN}{(TN + NP + NNg)} = \frac{128}{(128 + 6 + 0)} = 0.96$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{(Precision \times Recall)} = 2 \times \frac{0.91 \times 0.96}{(0.91 + 0.96)}$$

$$= 0.93$$

Positive Class

$$Precision = \frac{TN}{(TN + NP + FN)} = \frac{105}{(105 + 6 + 2)} = 0.93$$

$$Recall = \frac{TP}{(TP + PN + FP)} = \frac{105}{(105 + 11 + 20)} = 0.77$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{(Precision \times Recall)} = 2 \times \frac{0.93 \times 0.77}{(0.93 + 0.77)}$$

$$= 0.84$$

Classification accuracy

$$Akurasi = \frac{TP + TN + TNet}{TP + TN + TNet + FP + FN + FNet}$$

$$= \frac{105 + 114 + 128}{105 + 114 + 128 + 8 + 20 + 12} = 0.8966$$

F. Evaluation

Data that has gone through the preprocessing stage is then divided randomly into training data and testing data. The total amount of data is 1000, then it is resampled, that is, the unbalanced data is balanced first so that the total data becomes 1932. The data is divided using a ratio of 80% for training data and 20% for testing data, namely 1545: 387. The following table 4.13 shows the model evaluation from the formation of the model that has been created.

The next step is to build a classification model using the Naive Bayes and Decision tree methods on training data. The model obtained is then tested on testing data to see the accuracy of the model classification..

TABEL 6
EVALUATION

Metode	Evaluasi
Naïve Bayes	0.865
Decision tree	0.896

G. Visualization of Results with Word Cloud

Visualization is done using a word cloud. The function of this word cloud is to find words that appear frequently and have an impact on working on the classification model. By using a word cloud, you can find out the frequency of words that frequently appear in the three categories positive, negative, and neutral.

```
from wordcloud import WordCloud
text = df_upsampled.loc[df_upsampled['Label'] == 'Positive']
wordcloud = WordCloud(max_font_size=50,
max_words=200, background_color="white").generate(str(text))
plt.figure()
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()
```

Fig. 4. Word cloud sentiment positive

```
from wordcloud import WordCloud
text = df_upsampled.loc[df_upsampled['Label'] == 'Negative']
wordcloud = WordCloud(max_font_size=50,
max_words=200, background_color="white").generate(str(text))
plt.figure()
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()
```

Fig. 5. Word cloud sentiment negative

```
from wordcloud import WordCloud
text = df_upsampled.loc[df_upsampled['Label'] == 'Neutral']
wordcloud = WordCloud(max_font_size=50,
max_words=200, background_color="white").generate(str(text))
plt.figure()
plt.imshow(wordcloud, interpolation="bilinear")
plt.axis("off")
plt.show()
```

Fig. 6. Word cloud sentiment Neutral

IV. Conclusion

Based on the accuracy results of the Naïve Bayes algorithm, which is 86.5% and the Decision Tree algorithm, which is 89.6%, the Decision Tree algorithm is better for classifying Twitter user sentiment analysis of Tiktok Shop with a data division of 80%:20%. Sentiment analysis from the classification model uses a Decision Tree with 387 test data, so there are 114 negative data that correctly predict negative or True Negative. There were 20 negative data that were predicted incorrectly or False Negative, with 0 data that were predicted to be neutral and 20 data that were predicted to be positive. Furthermore, there were 128 neutral data that were correctly predicted to be neutral or True Neutral and there were 12 neutral data that were incorrectly predicted or False Neutral with 1 data that was predicted to be negative and 11 data that were predicted to be positive.

References

[1] Anam, M. K., Pikir, B. N., & Firdaus, M. B. (2021). Penerapan Na ̈Ive Bayes Classifier, K-Nearest Neighbor (Knn) Dan Decision Tree Untuk Menganalisis Sentimen Pada Interaksi Netizen Danpemeritah. *Matrik : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(1), 139–150. Https://Doi.Org/10.30812/Matrik.V21i1.1092

[2] Devita, R. N., Herwanto, H. W., & Wibawa, A. P. (2018). Perbandingan Kinerja Metode Naive Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *5*(4), 427. Https://Doi.Org/10.25126/Jtiik.201854773.

[3] Dragut, E., Fang, F., Sistla, P., Yu, C., & Meng, W. (2009). Stop Word And Related Problems In Web Interface Integration. *Proceedings Of The Vldb Endowment*, *2*(1), 349–360. Https://Doi.Org/10.14778/1687627.1687667.

[4] Efendi, Z. (2018). *Analisis Opini Masyarakat Terhadap Jasa Transportasi Online Menggunakan Text Mining Classification Pada Twitter*. Uin Suska Riau.

[5] Efendi, Z., & Mustakim. (2017). *Text Mining Classification Sebagai Rekomendasi Dosen Pembimbing Tugas Akhir Program Studi Sistem Informasi*. (Hal. 138–143). Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (Sntiki)

[6] Ernamia, E. M. A., & Herliana, A. (2022). Analisis Sentimen Kuliah Daring Dengan Algoritma Naïve Bayes, K-Nn Dan Decision Tree. *Jurnal Responsif : Riset Sains Dan Informatika*, *4*(1), 70–80. Https://Doi.Org/10.51977/Jti.V4i1.614.

[7] Faizal, M. R. (2016). *Seni Belajar Pemrograman Supervised Learning Dengan R*. Indc Indonesia.

[8] Hozairi, H., Anwari, A., & Alim, S. (2021). Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes. *Network Engineering Research Operation*, *6*(2), 133. Https://Doi.Org/10.21107/Nero.V6i2.237.

[9] Jati, B. D. (2020). *Naskah Publikasi Analisis Sentimen Produk Iphone Menggunakan Naive Bayes Classifier*.

[10] Jogiyanto, H. . (2015). *Analisa Dan Desain Sistem Informasi: Pendekatan Terstruktur Teori Dan Praktik Aplikasi Bisnis*. Andi.

[1] Jtik, J., Teknologi, J., Putra, T. W., & Triayudi, A. (2022). *Analisis Sentimen Pembelajaran Daring Menggunakan Metode Naïve Bayes , Knn , Dan Decision Tree*. *6*(1).

[12] Liu, B. (2010). Sentiment Analysis And Subjectivity. *Handbook Of Natural Language Processing, Second Edition*, 627–666.

[13] Mujahidin, S., Prasetio, B., & Utomo, M. C. C. (2022). Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga Bbm Pada Komentar Youtube Dengan Metode Gaussian Naïve Bayes. *Voteteknika (Vocational Teknik Elektronika Dan Informatika)*, *10*(3), 17. Https://Doi.Org/10.24036/Voteteknika.V10i3.118299

[14] Nafalski, A., & Wibawa, A. P. (2016). Machine

Translation With Javanese Speech Levels' Classification. *Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska*.

[15] Ng, A. (2017). Machine Learning Yearning. *Url: Http://Www. Mlyearning. Org/(96)*.

[16] Nuansa, E. P. (2017). *Analsis Sentimen Pengguna Twitter Terhadap Pemilihan Gubernur Dki Jakarta Dengan Metode Naïve Bayesian Classification Dan Support Vector Machine*.

[17] Of, A., Social, T., Sentiment, M., The, O. N., Reaction, P. S., The, T. O., Of, D., Creation, J. O. B., Using, L. A. W., Classification, T. H. E., & Naive, M. (2021). *Analisis Sentimen Media Sosial Twitter Terhadap Reaksi Masyarakat Naive Bayes Analysis Of Twitter Social Media Sentiment On The Public ' S Reaction To The Drafts Of Job Creation Law Using The Classification Method Naive Bayes*. *8*(5), 9007–9016.

[18] Prasetyo, E. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Cv. Andi.

[19] Ramadhan, M. A., & Wahyudin, M. I. (2022). Analisis Sentimen Mengenai Keberhasilan Indonesia Di Ajang Thomas Cup 2020 (Studi Kasus Media Sosial Twitter) Menggunakan Metode Naïve Bayes Dan Decision Tree. *Jurnal Jtik (Jurnal Teknologi Informasi Dan Komunikasi)*, *6*(4), 505–511. Https://Doi.Org/10.35870/Jtik.V6i4.560.

[20] Reynaldhi, M. A. R., & Sibaroni, Y. (2021). Analisis Sentimen Review Film Pada Twitter Menggunakan Metode Klasifikasi Hybrid Naïve Bayes Dan Decision Tree. *E-Proceeding Of Engineering*, *8*(5), 10127–10137.

[21] Rozaq, A., Yunitasari, Y., Sussolaikah, K., Sari, E. R. N., & Syahputra, R. I. (2022). Analisis Sentimen Terhadap Implementasi Program Merdeka Belajar Kampus Merdeka Menggunakan Naïve Bayes, K-Nearest Neighboars Dan Decision Tree. *Jurnal Media Informatika Budidarma*, *6*(2), 746. Https://Doi.Org/10.30865/Mib.V6i2.3554.

[22] Salmon Pattiiha, F. (2022). Perbandingan Metode K-Nn, Naïve Bayes, Decision Tree Untuk Analisis Sentimen Tweet Twitter Terkait Opini Terhadap Pt Pal Indonesia. *Jurnal Riset Komputer)*, *9*(2), 2407–389. Https://Doi.Org/10.30865/Jurikom.V9i2.4016.

[23] Schneider, K.-M. (2005). Techniques For Improving The Performance Of Naive Bayes For Text Classification. *International Conference On Intelligent Text Processing And Computational Linguistics*, 682–693.

[24] Sutopo. (2018). *Penentuan Jumlah Sampel Dalam Penelitian*. Http://Ejurnal.Stiedharmaputra-Smg.Ac.Id/Index.Php/Jema/Article/Download/156/128.

[25] Tri Romadloni, N., Santoso, I., & Budilaksono, S. (2019). Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl Commuter Line. *Jurnal Ikra-Ith Informatika*, *3*(2), 1–9.

[26] Tuffery, S. (2011). *Data Mining And Statistic For Decision Making*. John Wiley And Sons, Ltd.

[27] Weiss, S. M. (2010). *Text Mining: Predictive Methods For Analyzing Unstructured Information*. Springer.

[28] Yulian, E. (2018). *Text Mining Dengan K-Means Clustering Pada Tema Lgbt Dalam Arsip Tweet Masyarakat Kota Bandung*. *04(01)*, *27*. Matematika "Mantik"

[29] Zhang, D., Liu, W., Gong, X., & Jin, H. (2011). A Novel Improved Smote Resampling Algorithm Based On Fractal. *Journal Of Computational Information Systems*, *7*(6), 2204–2211.