Implementation Of Naive Bayes Classifier (NBC) For Sentiment Analysis On Twitter In Mobile Legends

Deden Kusnanda[1], Angga Aditya Permana[2*]

[1] Department of Engineering Informatics, Faculty of Engineering
Universitas Muhammadiyah Tangerang, Kota Tangerang, Banten, Indonesia.
[2] Department of Informatic, Faculty of Engineering and Informatic
Universitas Multimedia Nusantara, Kab Tangerang, Banten, Indonesia.
[*]Corresponding Author:
Email: angga.permana@umn.ac.id

Abstract.

On July 11, 2016, the mobile legends game was first released on the Indonesian server, and became one of the first MOBA games to enter the e-sports branch in Indonesia. The popularity of this game is growing rapidly, reaching more than 500 million downloads in the play store, and reaping a lot of controversy from mobile legends players. So that people create content and express their concerns through social media Twitter in the form of uploads and tweets. This makes the writer want to know how the public sentiment towards mobile legends online games on Twitter social media. The purpose of this research was to analyze how people's opinion about the Mobile Legends Online Game uses the Naïve Bayes Classifier method and to find out the level of accuracy that is obtained automatically by the system. As well as doing manual testing of the Confusion matrix from the results obtained. in classifying tweets. The results of community tweets taken using the data scraping method totaled 217 tweets and became a dataset, after the preprocessing process the tweets totaled 199 and became data testing, after being labeled it showed 100 positive tweets, 25 negative and 74 neutral. The level of accuracy obtained using the Naïve Bayes Classifier method is 80% with 104 positive tweets, 7 negative tweets and 88 neutral tweets. With a precision value of 83% positive, 100% negative and 76% neutral. The recall values obtained were 86% positive, 28% negative and 91% neutral. As for the F1-Score values obtained 84% positive, 44% negative and 83% neutral.

Keywords: *Mobile Legends, Analysis Sentiment, Confusion Matrix and Naïve Bayes Classifier.*

I. INTRODUCTION

Mobile legends is one of the moba games that is very popular among teenagers and children, the number of downloads of the Mobile Legends game has reached more than 500 million downloads in the Indonesian playstore. Mobile Legends was released on the Indonesian server on July 11 2016 and became one of the games included in the E-Sport branch. Cannot be separated from the many downloads of the Mobile Legends game, the Mobile Legends game has reaped a lot of controversy from many people, in terms of gameplay and Toxic words.This has made many people express their opinions through social media, because social media is a very popular digital communication platform for various internet users, as a means of communication and conveying public concerns. Specifically on the social media Twitter, Twitter is a social networking service that allows its users to post text, images and videos known as tweets, and some previous research also used Naive Bayes as an algorithm to classify opinions such as research [1]–[4], to analyze opinion sentiment in the community and retrieve data from Twitter[5]

[6]Many public opinions regarding the Mobile Legends game appear every day on popular websites that provide commenting services such as Twitter, Facebook and Instagram. The large number of opinions in the form of text and videos spread on social media, especially on Twitter, is an attraction for researchers to use this data as sentiment analysis, based on previous research entitled. Classification of poor communities using the naïve bayes method and sentiment analysis of the 2014 presidential candidates based on opinion from twitter using the naïve bayes classifier method. Sentiment analysis is a way to classify a tweet data into positive or negative data. Naive Bayes Classifier is a classification method based on an algorithm based on Bayes' theorem [7]. The tweet that is the object of this research is about the Mobile Legends Online Game. in this research python makes a model to classify or classify a "tweet" into positive, negative and neutral sentiments.

II. METHODS

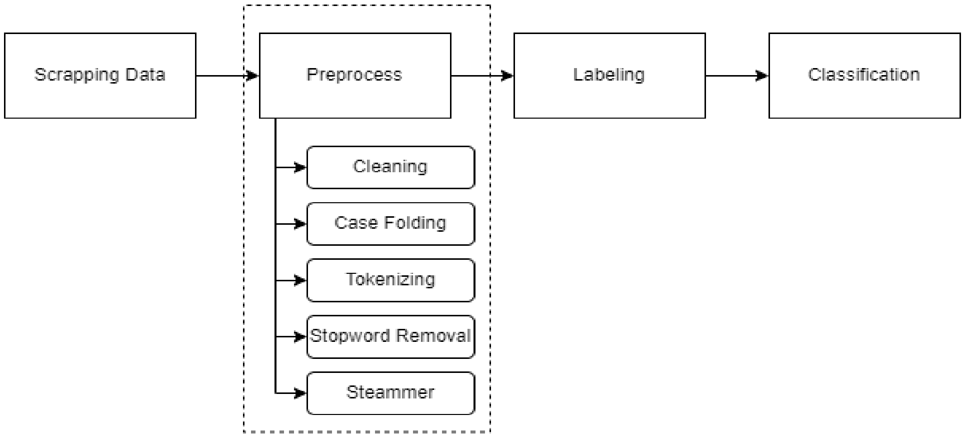

Fig 1. Research Flow

1. Scrapping Data

This is an activity carried out by taking tweets from Twitter using Python, directly to obtain the information needed for research. Scraping on Twitter was taken from 01 January 2023 - 20 July 2023 by getting 217 tweets.

2. Preprocess

It is an activity in cleaning a document that has been taken through the scraping stage such as cleaning tweets, case folding, tokenaizing, stop words, word stemmer.

a. Cleaning

Is a process carried out to clean tweets from features that are not needed.

b. Case Folding

In a tweet, there are often many differences in the use of letter forms. In this section, all capital letters are changed to lower case to make them uniform.

c. Tokenizing

Is a process carried out to divide sentences into several parts or words based on punctuation marks such as dots, periods and other marks.

d. Stopword Removal

Is a process carried out to remove words that are not needed.

e. Stemmer

Is the process of making words that have affixes turn into basic words according to Indonesian, for example, arrange becomes arrange, beginner becomes start, and play becomes play.

3. Labeling

is the process of classifying tweets into positive, negative and neutral sentiment using the textblob library.

4. Classification

This is an activity in converting tweet data into sentiment whether positive, negative or neutral using the Naive Bayes Classifier method.

III. RESULT AND DISCUSSION

Scrapping Data

Data collection in this study began with searching for previous journals related to this research, then scraping the data using Google Colab. Then the reprocessing process is carried out after that the classification is carried out.

a) datasets

The dataset is the initial data taken in the form of Indonesian text taken from the twitter.com website. data retrieved for this research using search_keyword(query: "Mobile Legend ", until : "2023-07-20" since : "2023-01-01" , lang="id", limit=1000), information request. The tweets taken are posts from several Twitter account users.

Fig 2. Scraping Data

Preprocessing Process

Data that has been taken in csv form will then be preprocessed in this process to remove words that do not affect the classification process. This process reads all tweets in Json form, so when this process is complete the results of the preprocessed data will be stored in excel form. The following is the process of preprocessing.

a) Text Preprocessing

Steps before labeling and classifying tweet data. This step includes clean tweets, case folding, tokenizing, stopwords and stemmers.

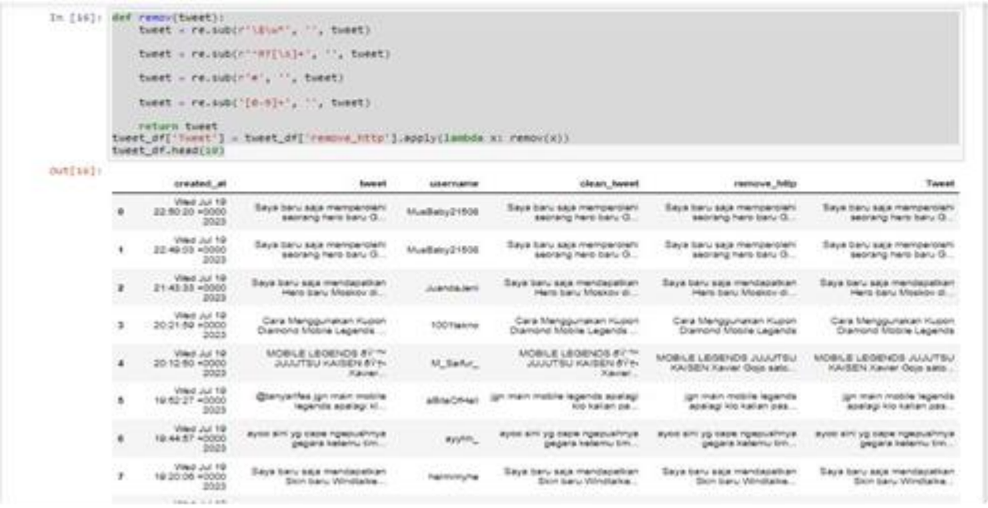

Fig 3. Clean Tweet

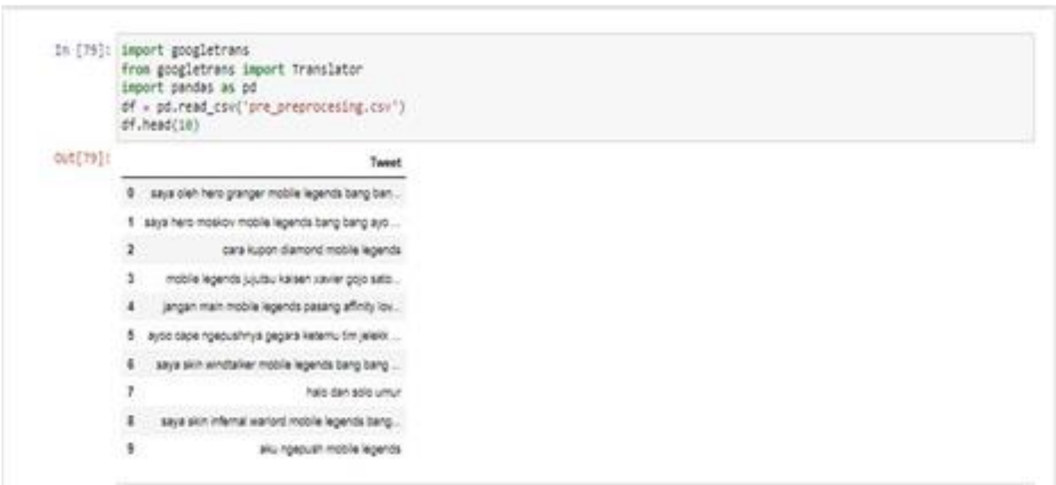

Fig 4. Preprocessing Results

Labeling

Labeling is the process of classifying tweets into positive, negative and neutral sentiments. As a comparison that will be carried out on the classification using the Naive Bayes Classifier method.

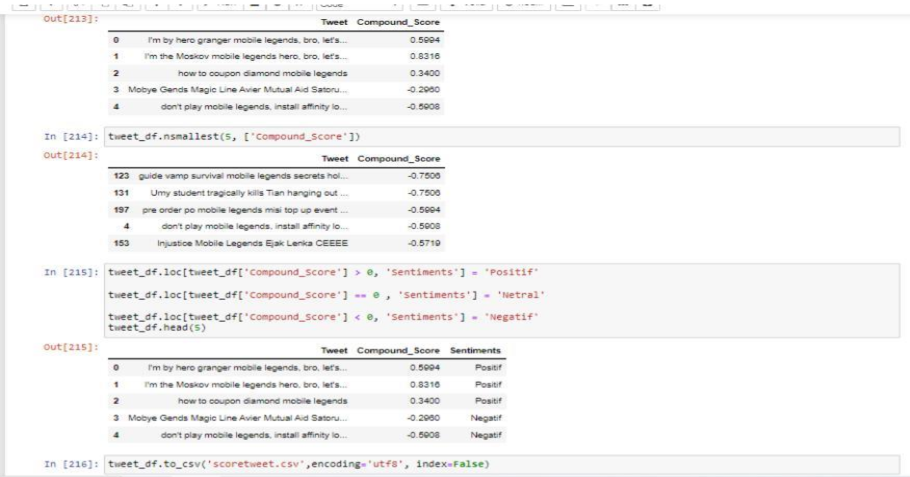

Fig 5. Tweet Score

Fig 6. Classification

Classification Naïve Bayes Classifier

Sentiment is determined by calculating the probability of document scraping with reference to sentiment classification, this is done automatically using the Naïve Bayes Classifier algorithm.

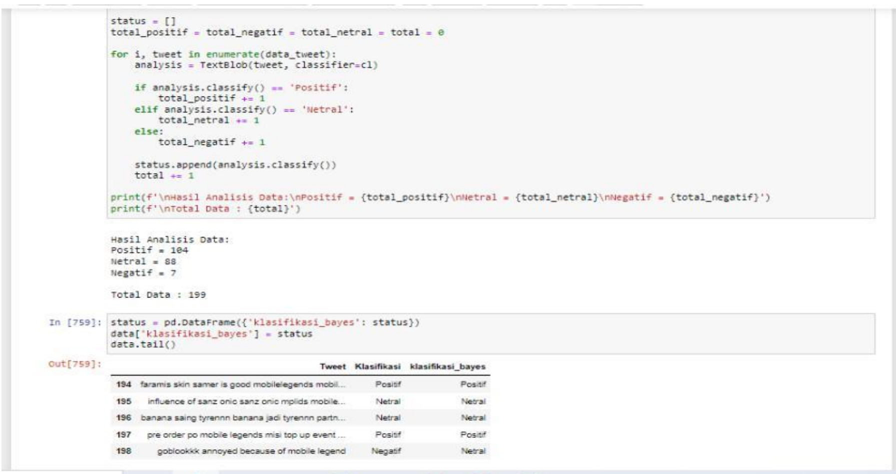

Fig 7. Classification Naïve Bayes Classifier

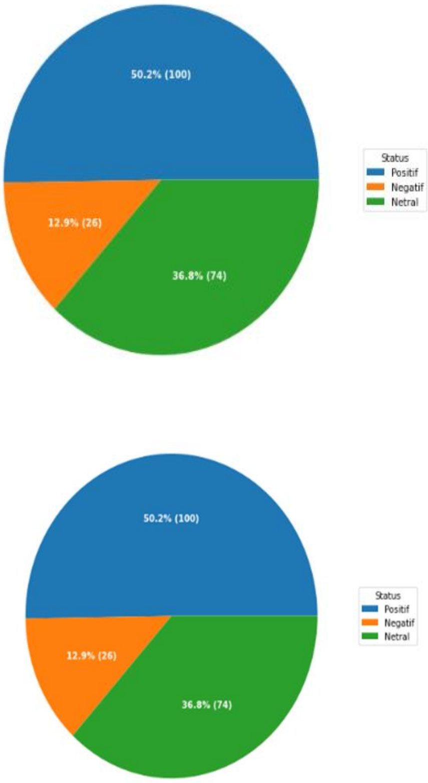

Fig 8. Diagram Classification

```
In [754]: import random

set_positif = []
set_negatif = []
set_netral = []

for n in dataset:
    if(n[1] == 'Positif'):
     set_positif.append(n)
    elif(n[1] == 'Negatif'):
     set_negatif.append(n)
    else:
     set_netral.append(n)

set_positif = random.sample(set_positif, k=int(len(set_positif)/2))
set_negatif = random.sample(set_negatif, k=int(len(set_negatif)/2))
set_netral = random.sample(set_netral, k=int(len(set_netral)/2))

train = set_positif + set_negatif + set_netral

train_set = []

for n in train:
     train_set.append(n)

In [755]: from textblob.classifiers import NaiveBayesClassifier
cl = NaiveBayesClassifier(train_set)
print('Akurasi Test:', cl.accuracy(dataset))

Akurasi Test: 0.8040201005025126
```

Fig 9. Acuraccy

Testing the Confusion Matrix

Data on sentiment classification 199 data. Positive 100 data, Negative 25 data, Neutral 74 data. Data on the classification of the Naive Bayes Classifier Positive 104 data true positive 87, false negative 11, false neutral 6. Negative 7 data true negative 7, positive flase 0, false neutral 0. Neutral 88 data true neutral 67, false positive 14, false negative 7 .

Tp, Tnet, Tneg = data true (data match)

Fp, Fnet, Fneg = data false (data does not match)

TF = true + false data result

AB = accuracy result

Table 1. Data true and False

	Positif	Negatif	Netral
Positif	87	11	6
Negatif	0	7	0
Netral	14	7	67

Precision

1 Precision positif confusion matrix :

$$\frac{87}{87+11+6} = \frac{87}{104} = 0{,}83$$

2 Precision negatif confusion matrix :

$$\frac{7}{7+0+0} = \frac{7}{7} = 1$$

3 Precision netral confusion matrix :

$$\frac{67}{67+14+7} = \frac{67}{88} = 0{,}76$$

Recall

a) Recall positif confusion matrix :

$$\frac{87}{87+0+14} = \frac{87}{101} = 0{,}86$$

b) Recall negatif confusion matrix :

$$\frac{7}{7+7+11} = \frac{7}{25} = 0{,}28$$

c) Recall netral confusion matrix :

$$\frac{67}{67+6+0} = \frac{67}{73} = 0{,}91$$

F1-Score

a) F1-Score positif confusion matrix :

$$\frac{2\ x\ 0{,}83\ x\ 0{,}86}{0{,}83+0{,}86} = \frac{1{,}4276}{1{,}69} = 0{,}84$$

b) F1-Score negatif confusion matrix :

$$\frac{2\ x\ 1\ x\ 0{,}28}{1+0{,}28} = \frac{0{,}56}{1{,}28} = 0{,}437$$

c) F1-Score netral confusion matrix :

$$\frac{2\ x\ 0{,}76\ x\ 0{,}91}{0{,}76+0{,}91} = \frac{1{,}3832}{1{,}67} = 0{,}828$$

Akurasi

Akurasi confusion matrix :

$$\frac{87+7+67}{87+11+6+0+7+0+14+7+67} = \frac{161}{199} = 0{,}80$$

IV. CONCLUSION

The Naïve Bayes method can analyze sentiment well. The trials were carried out using real-time testing data, and were classified as positive, negative or neutral sentiments. The Naïve Bayes Classifier method is also very effective in classifying sentiments up to an accuracy rate of 80%.

V. ACKNOWLEDGMENTS

The authors would like to thank Universitas Multimedia Nusantara for the support and facilities given for this research project.

REFERENCES

[1] B. Gunawan, H. Sasty, P. #2, E. Esyudha, and P. #3, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," vol. 4, no. 2, pp. 17–29, 2018.

[2] Simamora, R. N. H., & Elviani, S. (2022). Carbon emission disclosure in Indonesia: Viewed from the aspect of board of directors, managerial ownership, and audit committee. ***Journal of Contemporary Accounting***, 1-9.

[3] D. Garbian Nugroho, Y. Herry Chrisnanto, A. Wahana Jurusan Informatika, and F. Matematika dan Ilmu Pengetahuan Alam Universitas Jenderal Achmad Yani Jalan Terusan Jenderal Sudirman, *Analisis Sentimen Pada Jasa Ojek Online Menggunakan Metode Naïve Bayes*.

[4] Tarigan, N. M. R., Syahputra, R. A., & Yudha, T. K. (2022). The Analysis of Quality of Work Life and Work Achievement in Department of Agriculture Simalungun Regency. SIASAT, 7(1), 55-70.

[5] A. Syakuro, "Pada Media Sosial Menggunakan Metode Naïve Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG) Halaman Judul Skripsi Oleh : Abdan Syakuro," *Anal. sentimen Masy. terhadap e-commerce pada media Sos. menggunakan Metod. naive bayes Classif. dengan Sel. fitur Inf. gain*, pp. 1–89, 2017.

[6] Tarigan, N. M. R., & Wasesa, S. (2020). The Influence of Organizational Culture on Increasing Employee Motivation in Sumut Bank of Sharia Unit, North Sumatera. ***Britain International of Humanities and Social Sciences (BIoHS) Journal,*** 2(2), 440-449.

[7] A. A. Permana And W. A. Noviyanto, "Comparison of the Accuracy of the Lexicon-Based and Naive Bayes Classifier Methods To Public Opinions About Removing Masks on Social Media Twitter," ***J. Theor. Appl. Inf. Technol.***, vol. 101, no. 3, pp. 1174–1183, 2023.

[8] Sri, R., Mahdi, F., Julkarnain, J., Kurnia, H. N. T., & Habibie, A. (2022). Intellectual capital and islamic corporate social responsibility on the financial performance of sharia commercial banks in Indonesia. In E3S Web of Conferences (Vol. 339, p. 05003). EDP Sciences.

[9] C. Destitus, W. Wella, and S. Suryasari, "Support Vector Machine VS Information Gain: Analisis Sentimen Cyberbullying di Twitter Indonesia," ***Ultim. InfoSys*** **J. Ilmu Sist. Inf.,** vol. 11, no. 2, pp. 107–111, 2020, doi: 10.31937/si.v11i2.1740.

[10] Yusrita, Y., Fahmi, N. A., Yudha, T. K., & Nasution, I. (2020). Capabilities, Commitments and Effect on the Competitiveness of Small and Medium Enterprises (SME) in Medan. ***Budapest International Research and Critics Institute-Journal (BIRCI-Journal)***, 3(3), 2442-2450.

[11] A. Budiman, J. C. Young, and A. Suryadibrata, "Implementasi Algoritma Naïve Bayes untuk Klasifikasi Konten Twitter dengan Indikasi Depresi," ***J. Inform. J. Pengemb. IT,*** vol. 6, no. 2, pp. 133–138, 2021.

[12] N. Istiani and A. Islamy, "Fikih Media Sosial Di Indonesia," ***Asy Syar'Iyyah J. Ilmu Syari'Ah Dan Perbank. Islam,*** vol. 5, no. 2, pp. 202–225, 2020, doi: 10.32923/asy.v5i2.1586.