

Tweet Sentiment Analysis to Detect Suicide Intentions Through a Machine Learning Approach to Big Data Using the Naive Bayes Method

Noviyanti. P^{1✉}, Candra Gudianto², Listra Frigia Missianes Horhoruw³

^{1,2,3} Information Technology Study Program, Shanti Bhuana Institute, Indonesia

✉ **Corresponding Author** : Noviyanti. P (e-mail: noviyanti@shantibhuana.ac.id)

Article Information	ABSTRACT
Article History Received : December 15, 2024 Revised : January 03, 2025 Published : January 16, 2025	Suicidal ideation is a serious mental health problem and is often difficult to detect in its early stages. Social media, especially Twitter, is one of the platforms widely used by individuals to express their feelings and emotional conditions, including expressions of suicidal ideation. This study aims to develop a machine learning model that can analyze the sentiment of tweets related to suicidal ideation using big data. The data used in this study consisted of tweets that had been processed for sentiment analysis, which were then classified into three sentiment categories, namely positive, negative, and neutral. The machine learning model applied was Naive Bayes. The results of the model evaluation showed that this model had an accuracy of 72%, with precision and recall values varying depending on the sentiment category. The highest precision was recorded in the negative and neutral categories (0.91), while the highest recall was recorded in the positive category (0.97). This study provides insight into the potential use of machine learning-based sentiment analysis to detect signs of suicidal ideation through big data from social media that can help in early detection of mental health problems.
Keywords : <i>Sentiment Analysis;</i> <i>Machine Learning Approach;</i> <i>Big Data;</i> <i>Mental Health;</i> <i>Twitter;</i>	

INTRODUCTION

Suicidal ideation is a serious mental health problem and is often difficult to detect in its early stages. Every year, hundreds of thousands of individuals worldwide lose their lives due to suicide, and many of them show signs or symptoms of suicidal ideation that can be detected through their behavior (Frijanto, 2022) . One place that individuals often use to express their feelings and emotions is social media, including Twitter (Rahayu et al., 2023) .

Twitter, with billions of tweets produced each year, is a platform rich in information that reflects the feelings, emotions, and even mental states of its users (Solihin et al., 2021) . Many Twitter users share their thoughts and feelings, including expressions of sadness, anxiety, and even suicidal ideation. Given the vast amount of data available, it is important to develop an effective approach to analyzing tweets to detect potential signs of suicidal ideation.

Sentiment in text can reflect the feelings and mental state of the writer (Fadjeri et al., 2021) . Therefore, sentiment analysis becomes a useful tool to understand the attitudes or feelings of individuals written in tweets. These sentiments can be categorized into several categories, such as positive, negative, and neutral, which can provide further insight into the emotional state of the user.

However, large volumes of data cannot be processed manually because it can cause errors in data analysis (Santoso, 2020) . Therefore, machine learning offers a solution to analyze big data efficiently and effectively (Fathoni & Bhrata, 2023) . One of the things that can be done through the machine learning approach is data processing, such as overcoming missing data, relationship

analysis, and applying supervised data-based models to produce suicidal sentiment analysis models. The machine learning approach itself has been widely developed, especially in the health sector (Roihan et al., 2020) . One of the libraries that can be utilized in this study is scikit-learn (Horhoruw et al., 2024) .

Related research conducted by (Thursina Fazrian, 2023) discusses the influence of social media on the mental health of adolescents. According to the research that has been conducted, social media users among adolescents in 2015 reached more than 2,000 adolescent social media users. Based on this research, there are negative impacts of social media use among adolescents, including cases of cyber bullying, mental illness, and lack of privacy. The difference between previous research and the research to be conducted is that the previous research was conducted using a qualitative research type while the research to be conducted by the researcher uses a quantitative research type and is based on data.

Research conducted by (Arsini et al., 2023) this study aims to determine the effect of social media on the mental health of students at SMAN in Bandung City. The study was conducted using a quantitative research method by distributing questionnaires to 304 research samples. The results obtained based on the study were 53% of adolescents who experienced mental health, such as anxiety, stress, depression, and feeling lonely in the moderate category. The difference with previous research with the research to be conducted is that the research to be conducted by researchers uses a quantitative method with a big data and machine learning approach.

Further research conducted by (Khosla, 2024) discussed the influence of social media on body image with a case study of adolescent girls. The study was conducted by directly interviewing 20 adolescent girls aged 13-16 years. Based on the interviews conducted, most respondents had negative perceptions about body image. Respondents compared their appearance with celebrities and their peers, causing feelings of dissatisfaction with themselves. The study used a descriptive method based on interviews while the research to be conducted was based on big data analysis and was objective.

Research conducted by (Aldhyani et al., 2022) discusses detecting and analyzing suicidal ideas or desires on social media using deep learning and machine learning. Based on the research conducted, it was obtained that the approach using deep learning and machine learning with the CNN - BiLSTM method can detect suicidal ideas by 95% and was also analyzed using the LIWC XGBoost method showing better performance than the CNN-BiLSTM method. So based on this research, machine learning can be used to analyze the influence of social media on mental health.

Further research conducted by (Mahur et al., 2023) discussed the detection of social media depression using machine learning. The study used Twitter social media. The research conducted aims to implement a system that can detect someone if they are depressed based on the text provided on Twitter social media. This study uses machine learning techniques to analyze tweet text and detect signs of depression. This research is an important basis for future research, which focuses on detecting suicidal ideation using a similar big data-based approach.

Other related research conducted by (Al Maeni et al., 2024) is related to sentiment analysis on the Shopee application using the Naïve Bayes method. Based on the research conducted, an accuracy of 95% was obtained, a precision value of 47.50%, and a recall value of 50.00%. Further research is related to the Naïve Bayes method which can perform balanced classification based on data on Twitter in detecting advertisements (Muhammad Firdausy & Pandu Adikara, 2022) .

The research to be conducted aims to analyze tweet sentiment on Twitter as an indicator of suicidal ideation, using a machine learning approach that can process big data and produce accurate prediction models.

RESEARCH METHODS

This research is included in the type of quantitative research and the focus of the research is on the analysis of the influence of social media on mental health using a big data and machine learning approach. The data generated from the big data approach will be processed using machine learning. The data used is text from Twitter obtained through official sources, namely Kaggle. The research flow can be seen in Figure 1.

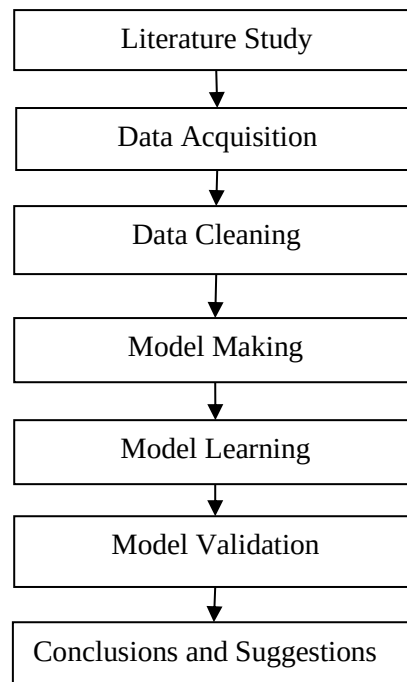


Figure 1. General Research Flow

Literature study

Conducting a literature study by looking for references related to the machine learning approach to big data in conducting sentiment analysis.

Data acquisition

Data was acquired through official sources, namely Kaggle. The data obtained was 10509 text data. The text data obtained is raw data. The data still has many punctuation marks, symbols, and numbers. In addition, the raw data does not yet have a column for sentiment.

Data cleaning

Perform data cleaning, normalization and deletion of irrelevant data, such as symbols, numbers, punctuation marks and stop words. The data cleaning process uses a machine learning approach. One of the most famous and frequently used libraries in data cleaning is scikit-learn. For example, there is the text "I need to start my life without anyone here!!!". Based on this text, the data cleansing process can be carried out by removing exclamation marks, punctuation marks, and changing all uppercase letters to lowercase, so that the text becomes "I need to start my life without anyone here"

Model building

Building a model using machine learning techniques. The model to be built uses a machine learning approach for data cleaning and sentiment analysis classification using the Naïve Bayes method. The Naïve Bayes method is a simple method that can be used to classify a text, especially sentiment classification. In the process of creating a data model that has gone through cleaning, data is used to create a sentiment column. Through a machine learning approach, the library used to generate sentiment is vaderSentiment. The process for building a model uses the naive Bayes method. The equivalent of the Naive Bayes method Formula (1).

$$P(C_k|X) = \frac{P(X|C_k) \cdot P(C_k)}{P(X)} \quad (1)$$

The algorithm of the Naive Bayes method itself ;

1. Calculate the probability of each class based on the data distribution $P(C_k)$
2. Calculate the likelihood probability of each feature x_i for each class C_k .
3. Multiply the likelihood with the prior for each class.

4. Determine the maximum probability class.

Model learning

In the training process of this model, the Naïve Bayes method plays an important role in classifying data that has been added with a sentiment column. The classification process using the Naïve Bayes method is based on the simple bayes theorem referring to raw data that has been equipped with sentiment classes, namely positive sentiment classes, negative sentiment classes, and neutral sentiment classes. The model training process will be divided into 2 data, namely training data and test data.

Model validation

Model validation is done by looking at Accuracy, Precision, Recall, and F1-Score. Accuracy is the accuracy of the method that can classify correctly. Precision is an validation value that shows the actual prediction with the prediction made by the model. Recall shows the intensity in the classification process, whether it is more inclined towards positive, negative, or neutral sentiment classification results. Precision F1-score aims to see the class balance between precision and recall.

Conclusion and Suggestions

Making conclusions from research results and providing suggestions for further development in the future.

RESULTS AND DISCUSSION

The data used in this study were 10509 text data obtained from official Kaggle sources based on Twitter data on suicidal ideation. The data previously contained raw data consisting of text, numbers, punctuation, and symbols. So the step after obtaining the data is to preprocess the data through a machine learning approach. The data link can be seen at the following address <https://www.kaggle.com/datasets/arshkandroo/behavioural-tweets>.

Data Cleaning

Sentiment analysis in this study uses the Google Collaboratory tool. Google Collaboratory uses the Python programming language. There are several stages in performing data cleaning, namely having to prepare a library or library that must be available to perform data preprocessing through a machine learning approach.

```
!pip install scikit-learn
!pip install wordcloud
!pip install matplotlib
!pip install seaborn
!pip install pandas
!pip install numpy

import pandas as pd
import re
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from sklearn.feature_extraction.text import TfidfVectorizer
import matplotlib.pyplot as plt
from wordcloud import WordCloud
```

```
nltk.download('stopwords')
nltk.download('punkt')
```

	tweet
0	alone time precious love alone want spend time...
1	wanna vent dont feel safe venting cause right ...
2	shit brings back cause know love care even fee...
3	tweet purely afraid go lonely nigga dont like ...
4	im gonna take shower im sure help calm know ca...

Figure 2 is the raw data head that has been successfully uploaded. Raw data is original data that has not been preprocessed at all. The following is the code to preprocess text data on the raw data that has been uploaded.

```
def clean_text(text):
    """Function to clean text"""
    text = re.sub(r'http\S+|www.\S+', "", text) # remove URL
    text = re.sub(r'@\w+', "", text) # delete mention (@username)
    text = re.sub(r'#', "", text) # remove hashtag
    text = re.sub(r'[^a-zA-Z\s]', "", text) # remove characters other than letters and spaces
    text = text.lower().strip() # Convert to lower case and remove spaces at the beginning/end
    return text
```

[illegible]

Tweet Sentiment Analysis to Detect Suicide.....

Model Building

The model creation process is carried out using the naive bayes method. The raw data obtained is purely text without any sentiment, so in the process of creating this model, it is necessary to create a sentiment column in clean data. The library used in the process of creating a sentiment column is !pip install vaderSentiment. The process of creating a sentiment column refers to clean data using the following code

```
analyzer = SentimentIntensityAnalyzer()

# Function to get sentiment
def get_sentiment(text):
    score = analyzer.polarity_scores(text)
    if score['compound'] >= 0.05:
        return 'positive'
    elif score['compound'] <= -0.05:
        return 'negative'
    else:
        return 'neutral'
```

The code above is sentiment analysis using the vaderSentiment library using the SentimentIntensityAnalyzer command. The raw data obtained has not been subjected to sentiment analysis at all so it is necessary to carry out sentiment analysis on the raw data. Determining the score of a text as positive, negative or neutral is based on a threshold value. If the threshold value is greater than or equal to 0.05 then it is detected as a positive sentence, the threshold value is less than or equal to -0.05 then it is detected as a negative sentence, and if the threshold value is between -0.05 and 0, 05 is detected as a neutral sentence.

```

                                clean_tweet sentiment
0  alone time precious love alone want spend time...  positive
1  wanna vent dont feel safe venting cause right ...  positive
2  shit brings back cause know love care even fee...  neutral
3  tweet purely afraid go lonely nigga dont like ...  negative
4  im gonna take shower im sure help calm know ca...  positive
```

Figure 4. Head Data Clean Tweet and Sentiment

Figure 4 is the head data that has successfully created a sentiment column based on clean tweet data. The data is then used for model training using naive bayes.

Model Learning

Model learning using the naive bayes method of the scikit-learn library. Naive bayes is one of the machine learning approach methods that is widely used to perform classification. Naive bayes is based on the bayes theorem by looking at the probability of an event based on existing data. Therefore, if using the naive bayes method, you must already have a class classification in this case consisting of 3 sentiment classes, namely neutral, negative, and positive.

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import classification_report

vectorizer = TfidfVectorizer(max_features=1000)
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)

model = MultinomialNB()
model.fit(X_train_tfidf, y_train)
```

The above code consists of the scikit-learn library used in the naive bayes method and the code to convert text into numeric features with TF-IDF and initialize and train the Naive Bayes model. To learn a model, the data must be divided into training data and testing data. Model learning has been successfully carried out by achieving an accuracy of 72% in the classification process.

Model Validation

Based on the process that has been done, starting from cleaning, building, and learning of data, Precision for negative: 0.91, meaning that of all predictions that the model classifies as "negative", 91% of them are correct. Precision for neutral: 0.91, meaning that of all predictions that the model classifies as "neutral", 91% of them are correct. Precision for positive: 0.65, meaning that of all predictions that the model classifies as "positive", 65% of them are correct. Recall for negative: 0.44, meaning that the model only managed to predict 44% of all data that was actually labeled "negative". Recall for neutral: 0.46, meaning that the model only managed to predict 46% of all data that was actually labeled "neutral". Recall for positive: 0.97, meaning that the model successfully predicted 97% of all data that was actually labeled "positive". F1-score for negative: 0.59, is the average value between precision (0.91) and recall (0.44). F1-score for neutral: 0.61, which is the average value between precision (0.91) and recall (0.46). F1-score for positive: 0.78, which is the average value between precision (0.65) and recall (0.97).

	precision	recall	f1-score	support
negative	0.91	0.44	0.59	604
neutral	0.91	0.46	0.61	425
positive	0.65	0.97	0.78	1073
accuracy			0.72	2102
macro avg	0.83	0.62	0.66	2102
weighted avg	0.78	0.72	0.69	2102

Figure 5. Precision, Recall, and F1-Score Results

CONCLUSION

Based on the research that has been done, an accuracy of 72% was obtained, which shows that the model is generally able to predict tweet sentiment with a fairly good level of success. However, this accuracy can still be improved, especially to identify tweets with negative and neutral sentiment. Based on research that has been carried out, an accuracy of 72% is obtained, which shows that the model is generally able to predict tweet sentiment with a fairly good success rate. However, this accuracy can still be improved, especially for identifying tweets that have negative and neutral sentiment. Apart from that, the accuracy obtained at 72% can be influenced by several things, including raw data that is not sentiment from the start. The processed raw data only consists of 1 column, namely the tweets column, so the sentiment column is created based on the vaderSentiments library. Apart from that, the amount of data that is less than optimal can be a determinant of the level of accuracy of a model. However, the model obtained is able to provide fairly good accuracy so that in the future it can still be improved by considering several of the causes above.

The model shows excellent precision for the negative and neutral categories, with values reaching 0.91, which means the model tends to be accurate in classifying tweets as negative or neutral. However, the precision for the positive category is lower (0.65), indicating that there are some misclassifications between positive and other categories. Meanwhile, recall is high in the positive category (0.97), meaning the model successfully identifies most tweets that show positive signs well. However, recall for the negative and neutral categories is lower (0.44 and 0.46), indicating that the model often fails to identify tweets with negative and neutral sentiments. The

f1-score for positive is the highest (0.78), indicating a good balance between precision and recall in this category. However, the f1-score for negative and neutral (0.59 and 0.61) is lower, reflecting a trade-off between precision and recall that needs to be improved.

This study shows that machine learning, especially sentiment classification approaches, have potential in detecting tweets related to suicidal ideation based on model evaluations that have been carried out. Although the model provides fairly good accuracy, there are challenges in improving the model's ability to detect positive, negative and neutral sentiment more accurately. Furthermore, this research can be developed by improving data preprocessing, using more complex algorithms, or optimizing model parameters to improve performance, especially in detecting tweets with negative sentiments related to suicidal ideation. The results of this research can be used as a basis for developing an early detection system for mental health problems through analysis of social media such as Twitter with an approach based on big data and machine learning.

REFERENCES

- Al Maeni, N., Suarna, N., & Prihartono, W. (2024). Shopee User Sentiment Analysis Based on Tweet Data From Twitter Using the Naive Bayes Classifier Method. *JATI (Journal of Informatics Engineering Students)* , 8 (2), 1834–1840. <https://doi.org/10.36040/jati.v8i2.8298>
- Aldhyani, THH, Alsubari, SN, Alshebami, AS, Alkahtani, H., & Ahmed, ZAT (2022). Detecting and Analyzing Suicidal Ideation on Social Media Using Deep Learning and Machine Learning Models. *International Journal of Environmental Research and Public Health* , 19 (19). <https://doi.org/10.3390/ijerph191912635>
- Arsini, Y., Azzahra, H., Tarigan, KS, & Azhari, I. (2023). The Influence of Social Media on Adolescent Mental Health. *MUDABBIR Journal Research and Education Studies* , 3 (2), 50–54. <https://doi.org/10.56832/mudabbir.v3i2.370>
- Fadjeri, A., Hidayat, K., & Handayani, DR (2021). Emotion Detection in Text Using the Naïve Bayes Algorithm. *Journal of Information and Computer Technology Research (JURISTIK)* , 1 (2), 1–4. <https://doi.org/10.53863/juristik.v1i02.365>
- Fathoni, MA, & Bhrata, AA (2023). Smart Manufacturing Management System Utilizing Big Data and Machine Learning Algorithms for MSME Production. *Journal Science Innovation and Technology (SINTECH)* , 3 (1), 36–47. <https://doi.org/10.47701/sintech.v3i1.2523>
- Frijanto, A. (2022). *Depression and Suicide* . Ministry of Health. https://yankes.kemkes.go.id/view_artikel/1450/depresi-dan-bunuh-diri
- Horhoruw, LFM, Pratama, A., Asri, Y., & P, N. (2024). *Machine Learning Exploration with Scikit-Learn : Machine Learning Learning Strategy* (Vol. 16, Issue 1). UWAIS Inspirasi Indonesia.
- Khosla, S. (2024). *Exploring the Influence of Social Media on Body Image: A Study on Adolescent Girls* . 12 (1). <https://doi.org/10.25215/1201.104>
- Mahur, L. S., Gupta, S., & Gupta, P. (2023). Review on Depression Detection on Social Media Using Machine Learning. *Lecture Notes in Networks and Systems* , 726 LNNS (June), 79–88. https://doi.org/10.1007/978-981-99-3716-5_8
- Muhammad Firdausy, T., & Pandu Adikara, P. (2022). *Advertisement Detection on Twit using the Naive Bayes Method* . 6 (1), 204–209.
- Rahayu, K., Fitria, V., Septhya, D., Rahmaddeni, R., & Efrizoni, L. (2023). Text Classification to Detect Depression and Anxiety in Twitter Users Based on Machine Learning. *MALCOM: Indonesian Journal of Machine Learning and Computer Science* , 3 (2), 108–114. <https://doi.org/10.57152/malcom.v3i2.780>
- Roihan, A., Sunarya, P.A., & Rafika, A.S. (2020). Utilization of Machine Learning in Various Fields: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)* , 5 (1),

75–82. <https://doi.org/10.31294/ijcit.v5i1.7951>

Santoso, JT (2020). Big Data Analysis. In *Yayasan Prima Agus Teknik Publisher* .

Solihin, F., Awaliyah, S., Muid, A., & Shofa, A. (2021). Utilization of Twitter as a Media for Disseminating Information by the Communication and Informatics Service. *Journal of Social Science Education (JPIPS)* , 1 (13), 52–58.

Thursina Fazrian. (2023). The Influence of Social Media on Students' Mental Health at a Senior High School in Bandung City. *West Science Journal of Psychology and Counseling* , 1 (01), 19–30.