Islamic University Journal of Applied Sciences (IUJAS)
https://journals.iu.edu.sa/jesc

Volume VII, Issue I, 1 July 2025, Pages 227-243

Application of Machine Learning to Developing an Internet Community Model

Ozoh Patrick [1, *], Musibau Ibrahim [1], Oyinloye Olufunke [2]
[1] Faculty of Computing and Information Technology, Osun State University, Nigeria
patrick.ozoh@uniosun.edu.ng, ibrahima@uniosun.edu.ng
[2] Department of Computer Science, University of Ilesa, Nigeria

*Corresponding author: (O. Patrick), *Email Address:* patrick.ozoh@uniosun.edu.ng

Abstract

This study will assist in determining the types of elements and occasions that influence people's opinions. It aims to implement a social media sentiment Social Media Sentiment Analysis using machine learning techniques. The objectives of this study are as follows: (i) To combine natural language processing, machine learning, and financial modeling to analyze the different impacts of sentiments on social media. (ii) To implement the developed model in (i) (iii) To evaluate the performance of the developed model in (ii). The method used in this study includes the application downloads of tweets from Tweeter and inserts the data into the MongoDB database. The Natural Language Toolkit corpus called Twitter samples is available for the training dataset, the application extracts features from the training dataset. The insights would help the people for example in the National Security Department and Emergency Response Teams to understand the public emotional behaviors towards certain events and people, take appropriate actions with informed decisions, and perform situational analysis regarding public safety and security. Pymongo retrieved the text driver from the Tweets. The classifier model assigned polarity to each tweet and displayed the Tweet. Data visualization was successful for the retrieved system for Twitter user followers, friends, status counts, and charts to visualize the data. The application displayed exploratory visualization using line charts, bar charts, and tweets on a map using coordinates.

Keywords: Online social media, Sentiment analysis, Supervised learning, Natural language processing, Behavioral analysis

https://doi.org/10.63070/jesc.2025.014
Received 23 March 2025; Revised 12 May 2025; Accepted 26 May 2025.
Available online 28 May 2025.
Published by Islamic University of Madinah on behalf of *Islamic University Journal of Applied Sciences*. This is a free open access article.

1. Introduction

Due to social media's rising prominence in contemporary culture, its use has increased. They are now used by internet users for more purposes than just exchanging private information or corresponding with peers, co-workers, or relations. These subjects cover a broad range, including goods, individuals, occasions, trends, and societal problems. Customers' input is beneficial in helping businesses improve since it allows them to learn what the public thinks about their products or services [19]. The government, organizations, and prominent figures would like to learn more about how the public perceives them. From the viewpoint of the consumer, opinions, and feelings about goods are also useful as a guide when making decisions about those specific products. Numerous studies have been done to examine attitudes and feelings on social networking sites, and they cover a variety of topics, including finance and, particularly, share markets. The stock markets are extremely unpredictable, and social media reports or news frequently have a significant impact on a company's share price. Some of the variations in the stock market are hypothesized in [17].

The advancements in communication tools, made possible by the Internet, have recently brought people closer together. With the aid of the Internet, the lengthy process of contact that previously involved mail and telegraph has been transformed into an instantaneous one. People can now connect with anyone thanks to social media; one of the application software programs that appeared alongside smartphone technology. All individuals and organizations were impacted by this. Sharing a space, like a theatre, shop, or cafe, or voicing a favourable or unfavourable opinion about it, for instance, everyone and the entire community in every aspect. Social media is viewed by many as the primary setting for conversation. Through social media, people exchange news, sports, movies, personal emotions, and ideas. This has changed significant data for many different entities, such as companies marketing or selling their goods and researchers looking at thoughts and emotions.

Sentiment analysis has been made possible as a necessary instrument whom frequently share their opinions about social, economic, health, and brand-related problems. Expressions of emotion in texts are anticipated by sentiment analysis research. People's messages are evaluated for their positivity, negativity, or indifference. The advent of the internet has changed how people express their opinions. It is now carried out through blog entries, online discussion groups, product review websites, and other channels. They essentially contend that certain market disruptions can affect the equity markets. A famous news outlet's news report, speculative news, changes to market rules, or any of these things can cause a market jolt. These effects are explained by [17]. Large negative share returns are more probable than large positive share returns, according to the volatility feedback. Although the volatility feedback helps to explain some of the effects of market changes, how these news items are interpreted is of utmost significance.

The goal of this research is to combine natural language processing, machine learning, and financial modeling to examine the various effects that social media opinions have on shared values. People rely on this user-generated material in huge numbers. When looking for merchandise, a person will first look up reviews online before making a decision. A typical consumer cannot possibly investigate the volume of user-produced material. The two key strategies used in mood analysis are symbolic strategies, also known as understanding basis approaches, and machine mastering methods. The expertise-based method necessitates a sizable collection of pre-set emotions and a green expertise example for sentiment analysis. An emotional classifier that categorizes feelings is expanded using machine learning techniques using a training set. System learning is significantly simpler than the knowledge-based approach because it contains a predefined database of all feelings.

In the period of machine learning, machines are left to think and solve problems by finding the patterns in every data set on their own. Examination of secret trends and patterns helps predict and know precisely how to fit the correct algorithm with a particular learning process or resources.

The focus of this research paper is on the evaluation of online content for a variety of websites, in terms of numbers and volumes. These websites are devoted to particular types of stock and have experience with storing customer reviews from various websites, including Amazon and a great deal more. Google Play Store/Twitter is primarily where polls are communicated through tweets, but it can take a lot of time to gather a general understanding of these unorganized records. Customers view those unstructured evaluations on a specific website, developing an opinion of the goods or services and ultimately deciding. This can be very helpful for companies and cast light on how rumours or news spreading online may favour or harm the company in the short and long term. It can also assist in determining the types of elements and occasions that influence people's opinions and confidence in the goods that a company sells. Businesses can conduct an early analysis of new goods, films, etc. thanks to sentiment analysis. Data classification is used for sentiment analysis, and done with different methods. Text is pre-processed using text extraction techniques before being classified. To illustrate these procedures, a collection of terms is created by removing symbols, commas, word stems, and stop words. This study is achieved with a categorization procedure.

This study consists of five sections. Section one contains the background and contribution to knowledge. Section two includes a review of fundamental concepts and related works. Section three showcases the presentation of the methodology and describes the techniques, technologies, and tools to be used. Section four discusses the results obtained. Section five includes the summary.

The contributions of this study are as follows:

1 To have an insight on how news or rumours circling on the internet might benefit or disadvantage the business

2 To combine natural language processing, machine learning, and financial modelling to analyse the different impacts of sentiments on social media

3 To understand factors that affect the opinion of people toward products.

2. Literature Review

[2] assess the classification performance of transformer-based pre-trained language models to correctly classify tweets dependent on heat stroke, a significant health effect as true or false. The study evaluates combining social networks with artificial intelligence for public health monitoring. The study visualizes data on classified tweets and heat stroke occurrences in animated videos. The results from the study indicate that the pre-trained language models are reliable in classifying the tweets. The animated video visualizations revealed a positive correlation. [20] Ullah et al. (2025) applied computational models to identify individuals with depression dependent on Twitter posts—feature extraction using Natural Language Processing (NLP) techniques retrieved and cleaned 1.6 million tweets. The methodology involves applying the Grey Relational Grade (GRG) to study the association between likes and shares of Twitter posts. As a result, different machine learning models were used to classify user tweets into "stressed" or "not stressed" categories. The findings suggest the importance of social media and computational models in mental health detection.

[11] presented an automated deep-learning model for identifying malicious activities. The data was collected from the social media platforms using Formspring, Instagram, and MySpace for observing cyberbullying behavior—the preprocessing processes involved removing stop words, and punctuations, and converting comments to lowercase. The study used Feature Density (FD) to compute intricate natural language datasets using the linguistically backed preprocessing model. The dataset is then input to the feature selection process that identifies important features for the predictive model. The identified features are investigated for cyberbullying behaviour detection. The Matlab software is applied for simulation. The study yielded outcomes with values of 99.12%, 94.73%, 97.45%, and 93.91% respectively for accuracy, precision, recall, and F1-Score.

The relevance of conversations and microblogs has grown due to the development of social media platforms like online reviews [8]. Because the internet has such a massive amount of opinionated data, if it is correctly examined, it may greatly aid corporations and policymakers in their decision-making. It is simple to classify attitudes even in real time.

Due to its user-friendly API and the social effect of certain powerful individuals, including politicians and celebrities, Twitter has emerged as the leading social media analytics platform. There are more than 15 billion Application Programming Interface (API) calls and over 3 billion tweets sent each day [3]. The increasing number of social media analytic applications are archived data. According to [18], a significant amount of data from users' tweets is required for opinion mining. Unstructured tweets are used. The Twitter dataset uses performance metrics. These methods are assessed. For classification approaches, the accuracy indicated SVM outperformed all others in the Twitter Sentiment assessment [7].

[15] analyses the affective content of the writings in different groups. [14] used patient remarks on Twitter about drug adverse effects to perform a sentiment analysis. [12] investigated whether the distribution of the data among groups affected the classification algorithm's success rate. [16] proposed a technique for autonomously gathering data from Twitter for opinion. In the study, N-grams and POS tags were used as features in the classifier. [1] used machine-learning techniques on movie evaluations. The paper extracted the features using TF-IDF. The support vector machine has produced superior outcomes. [5] applied sentiment analysis study. [6] proposed the sentiment analysis to classify comments using the feature vector. [9] present actions considered in the study.

3. Methodology

This section discusses the study methods used in predicting fuel consumption using machine learning. It provides details of data collection, data pre-processing, and the models used.

3.1 Data collection

Data is collected from Google Play Store and some other social media platforms like Twitter, Google Play Store, etc. (Figure1).

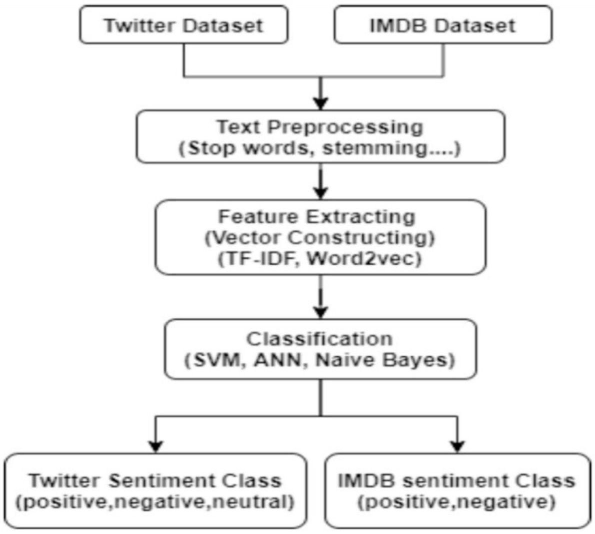

Figure 1. Flowchart of the study.

The Google Play Store/Twitter (Application Programming Interface) API was used to gather tweets. A Python program was used to pre-process the data and calculate sentiment ratings. 1680 of the tweets that were collected, and classified as neutral, 1220 as positive, and 1600 as unfavorable. The characteristics of the tweets collected by the API using Python are displayed in Table 1.

Table 1. Characteristics of tweets dateset.

Dataset Attribute	Explanation of Attribute
Id	Order of tweet data frame
Text	Tweet
created at	Date and time the Tweet was posted
Retweeted	Tweet rerun status (bool)
retweet count	Number of retweets
User screen name	Username
user followers_ count	Number of followers
user location	Followers location
Hashtags	Tweet tag
sentiment score	Sentiment score
sentiment class	positive, negative, neutral

The same models were also applied to the dataset of 500 positive and 500 negative opinions, shown in Table 2.

Table 2. Dataset of opinion of users .

Dataset Attribute	Explanation of Attribute
Text	Reviews
sentiment class	positive, negative

The application accesses the Twitter messages called tweets shared by users while sharing their views, emotions, and opinions on the Twitter Social Network.

3.2 Data Pre-processing

The application pre-processes the Twitter data to remove symbols like @ mentions and # hashtags before it performs feature extraction. The words known as stop-words in the sentiment analysis context, include words such as in, on, at, it, if, is, and so on. The diagram in Figure 2 shows the overall interactions between application modules from data acquisition to classification. The application downloads the tweets posted by the users and stores in the database, then accesses the stored data for pre-processing the text and passes it to the feature extractor function using the model built from the training dataset.

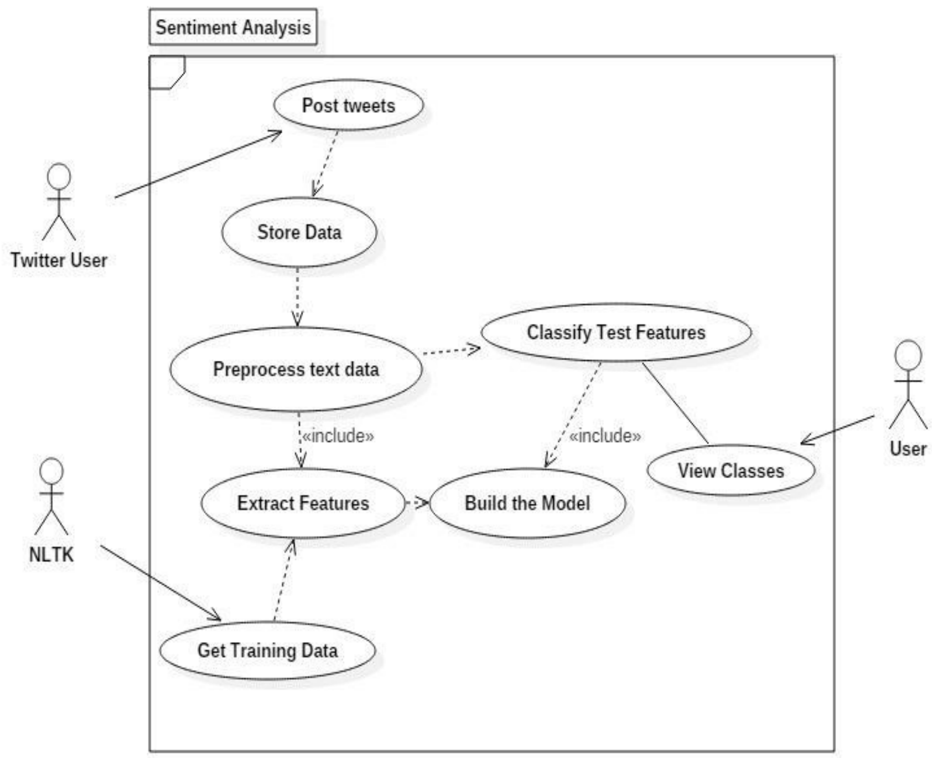

Figure 2. Use case diagram for sentiment analysis system.

3.3 Model Descriptions

This section includes the algorithms for the models and is described.

3.3.1 Naïve Bayes Algorithm

It is given as follows.

Step 1: Choose a training set.

Step 2: Train the classification algorithm using the training data.

Step 3: Use a classification algorithm to classify sample data.

Step 4: Compare the results against the known results set.

3.3.2 Bayes theorem

The Bayes theorem as given by [3] is represented in Equation 1.

$$P(c\backslash x) = \frac{P(x\backslash c)P(c)}{P(x)} \tag{1}$$

Where
$P(c\backslash x) =$ Posterior Probability function
$P(x\backslash c) =$Likelihood function
$P(c) =$Class Prior Probability function
$P(x) =$Predictor Prior Probability function

3.3.3 Maximum Entropy Algorithm

The algorithm is defined as the log-likelihood of the model p concerning the empirical distribution p, with x number of iterations derived in Equation (2) [13].

$$\log_{\vec{p}}(p) = log \prod_{x \in X} p(x)^{\overrightarrow{p(x)}} = \sum_{x \in X} p(x) \log p(x) \tag{2}$$

3.3.4 Support Vector Machines

It is represented as in Equation 3 [20].

$$\begin{aligned}
TS_i &= \max\left\{\left|\frac{\overline{x}_{ik} - \overline{x}_i}{m_k s_i}\right| \ , \quad k = 1, 2, \ldots K\right\} \\
\overline{x}_{ik} &= \sum_{j \in C_k} \overline{x}_{ij}/n_k \\
\overline{x}_i &= \sum_{j=1}^{n} x_{ij}/n \\
s_i^2 &= \frac{1}{n-K} \sum_k \sum_{j \in C_k} (x_{ij} - \overline{x}_{ik})^2 \\
m_k &= \sqrt{1/n_k + 1/n}
\end{aligned} \tag{3}$$

There are k classes, max {yk, k=1,2, 3,…K} is the maximum of all yk. Ck refers to class k which includes nk samples. xij is the expression value of I in sample j. xik is the mean expression value in k. n is the total number of samples. xi is the general mean expression value for i. si is the pooled within-class standard deviation.

3.3.5 K-Means Clustering

Given a dataset X=[x1,…,xn], xnεRd. The dataset is partitioned into M disjoint subsets C1,…Cm. The algorithm is given as in Equation 4 [9].

$$\frac{1}{m_k}\sum_{i=1}^{m_k}\left\|x^i-\mu_c k\right\|^2 \tag{4}$$

3.3.6 Term Frequency

This technique gives the significance of a given textual context. This method involves assigning weights [10]. The weights are assigned as in Equation 5.

$$W_{i,j}=TF_{t,d}\left(\frac{N}{D_t}\right) \tag{5}$$

where TFt, Dt has the term t.

3.3.7 Predicting Users' Actions from Sentiment

1. "Share positive: Fracking topic: "Fracking saves us money; fracking creates jobs; fracking reduces greenhouse gas emissions."
2. "Spread negative vaccination topic: "Vaccination has never been proven to have saved one single life"
3. "Call for oppose action: Protect our kids and families from £10 fracking. Pls RT!".

The appendix code for the implementation of this study is displayed in Appendix A.

4. Implementation and Testing

This section explains the system development designs specified in previous sections. It highlights the algorithms that present the results for social media sentiment analysis using machine learning. The proposed system was successfully tested to denote its effectiveness and achievability. It makes possible expressions of emotion in texts as they are anticipated by sentiment analysis research. People's messages are evaluated for their positivity, negativity, or indifference.

As a purposeful goal achieved, businesses can conduct an early analysis of new goods, films, etc. The system has been fully built and is ready to be used. Figure 4 shows data exploration. In Figure 5, the data pre-processing was analyzed by loading the Google Play Store review dataset.

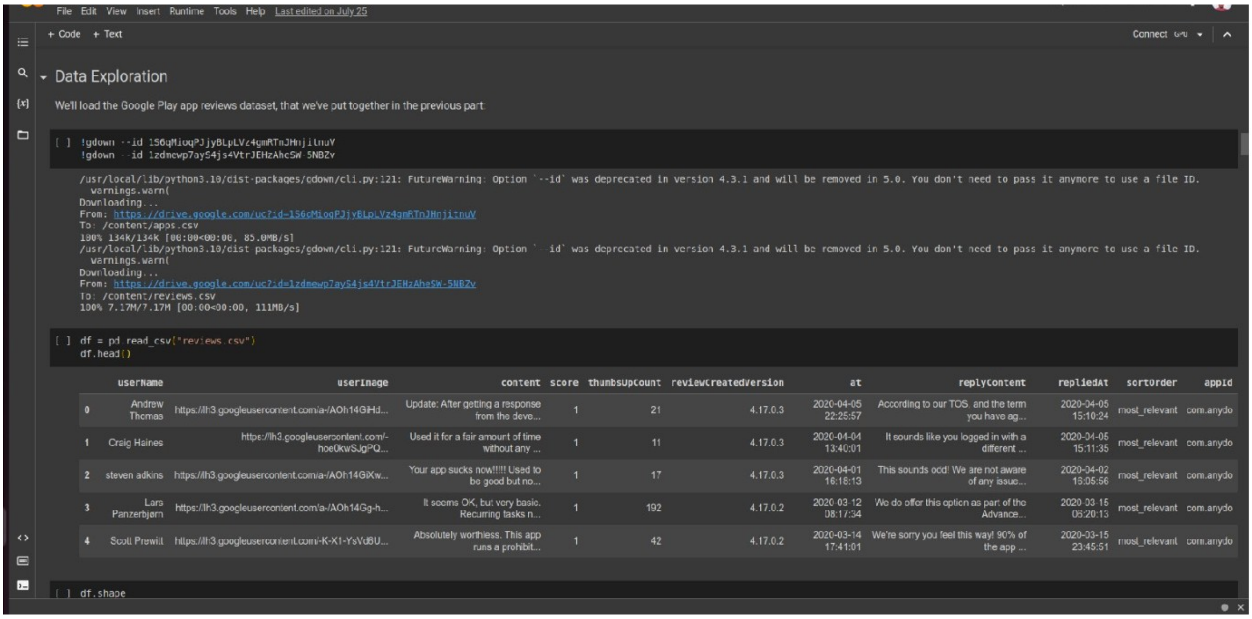

Figure 4. Data exploration.

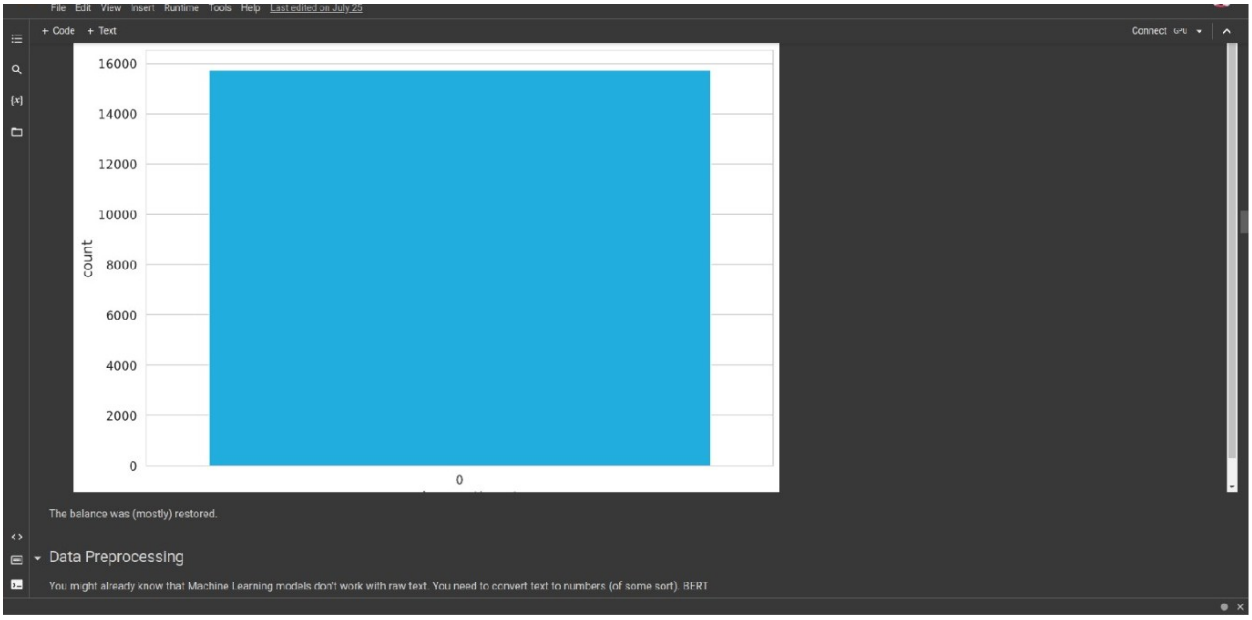

Figure 5. Data pre-processing.

The two images shown below in Figure 6 are the encoders used for modeling and analyzing social media sentiment. The second image shows its setup and configuration. Figure 7 is the screenshot that shows the predicting model on raw text; and how to use the sentiment of raw text.

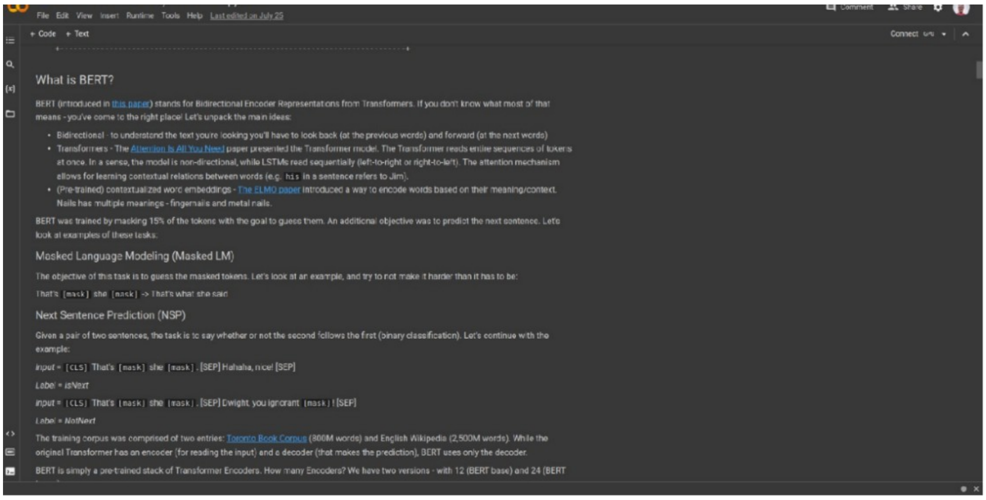

Figure 6. BERT representations

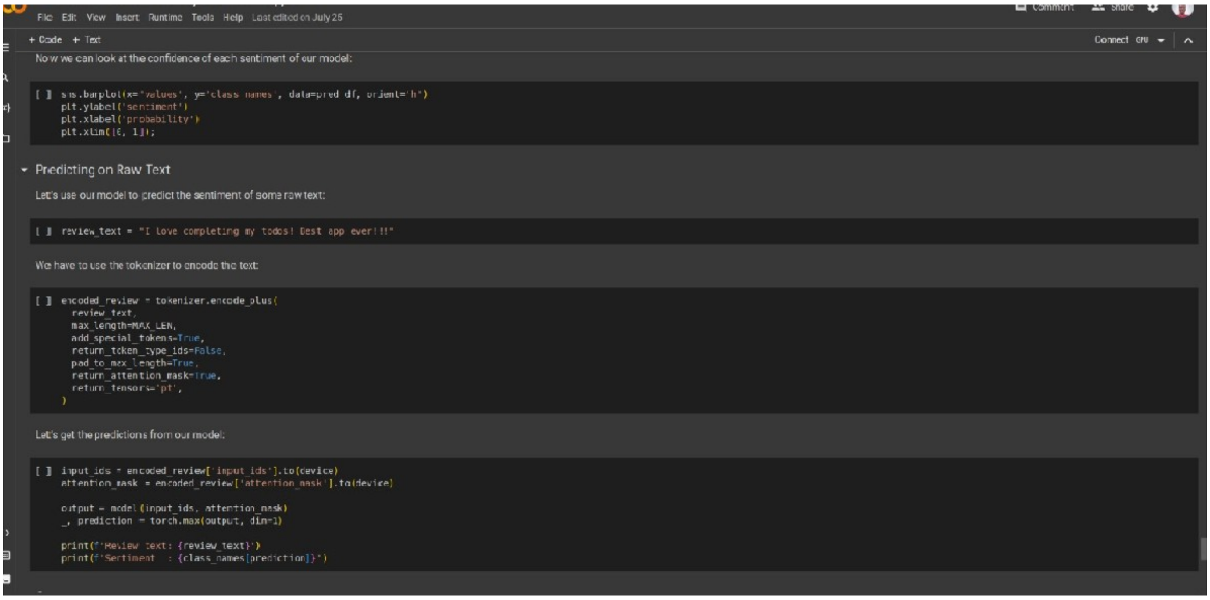

Figure 7. Predicting on raw text

The image in Figure 8 depicts the training and validation accuracy of sentimental prediction. The image in Figure 9 shows the classification report, which explains its overview and relative frequency effects.

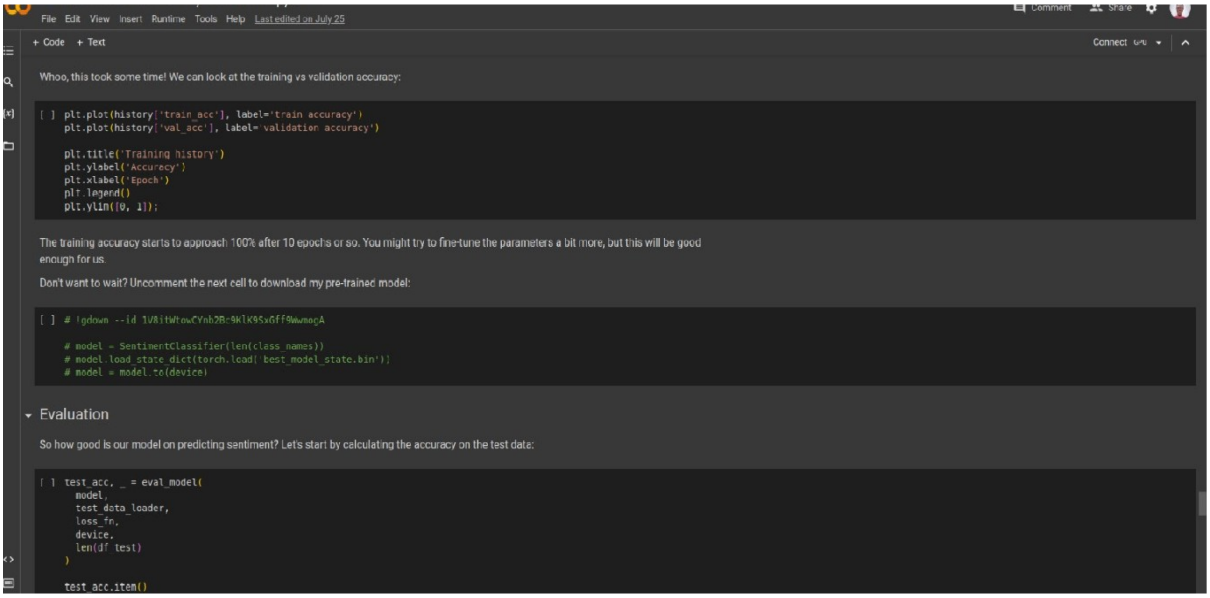

Figure 8. Training vs validation accuracy

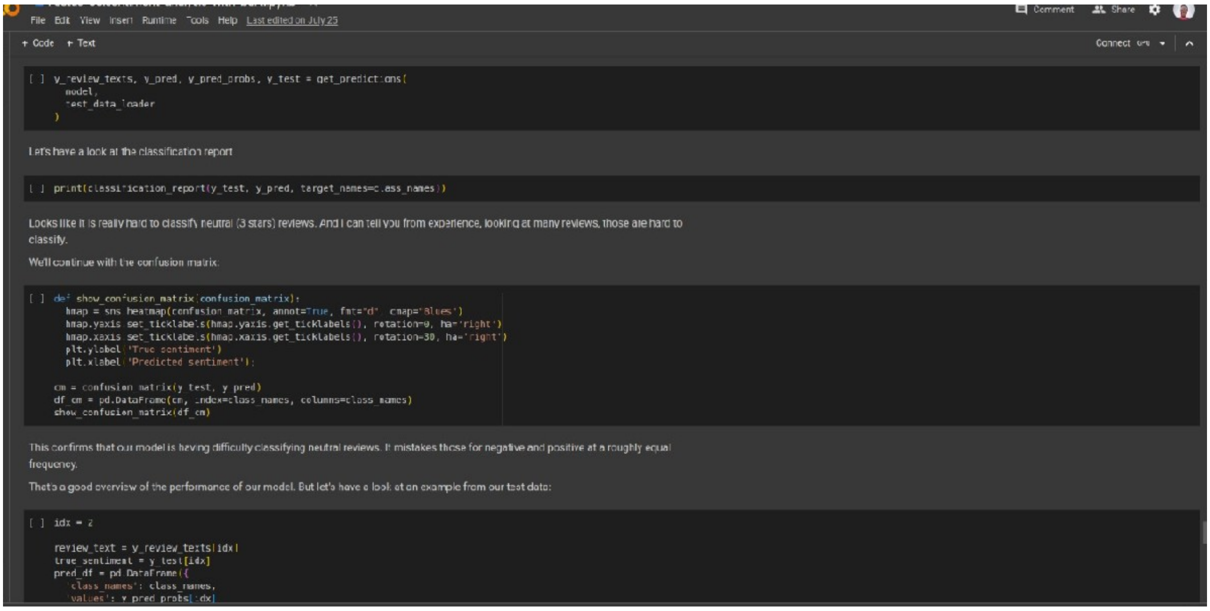

Figure 9. Classification report

The evaluation of models is shown in Figure 10.

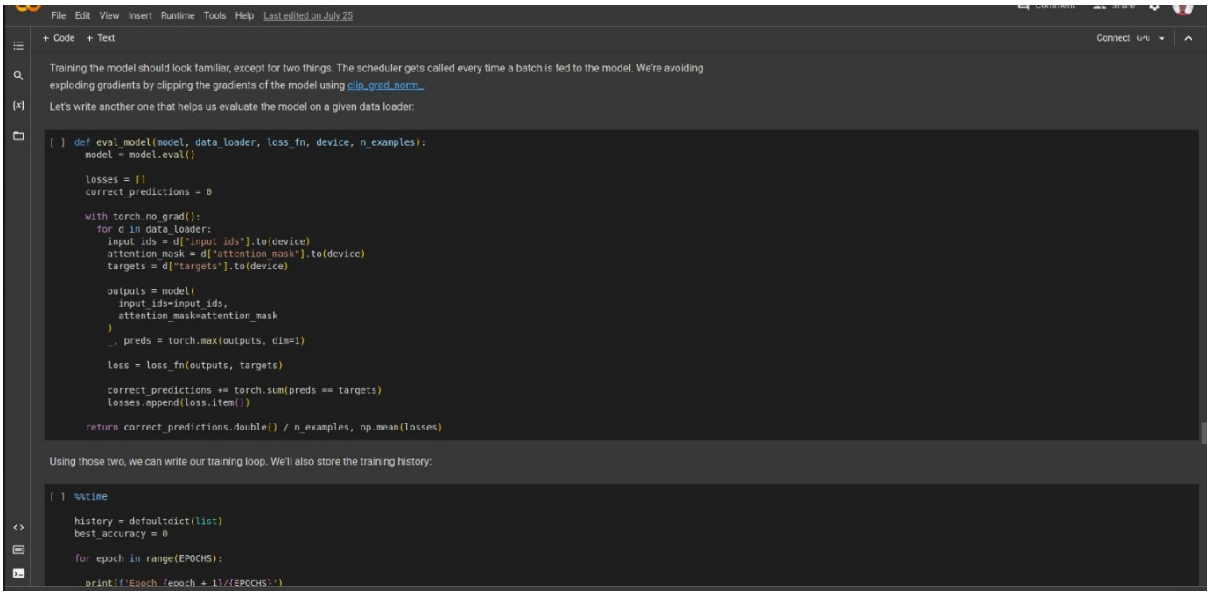

Figure 10. Model evaluation

5 Summary, Conclusion, and Future Work

5.1 Summary

The tests performed during the development of the implementation of social media sentiment analysis using machine learning to verify and validate the module functionalities to provide necessary data; for database connection testing, pre-processed text data for data preprocessing testing, polarity for the tweet text for sentiment results testing, data visualization testing, and response time testing. Database connection was successful using the MongoDB connection driver Pymongo, for sentiment analysis, the system retrieved the text data from the tweets. Data pre-processing was successful in that the repeating characters were removed, replaced #hashtags and @mention symbols, and removed URLs. Sentiment results were successful in that using the classifier model was able to assign polarity to each tweet and display the tweet along with its polarity. However, unigrams did not perform well on some words. Finally, data visualization was successful in that the system retrieved numeric data for Twitter user followers, friends, and status counts and created the charts to visualize the data.

5.2 Conclusion

The Social Media applications and algorithms has successfully been developed with the basic functionalities as in System Requirements Specifications. The application was able to connect and download tweets from the Twitter App using Twitter Streaming API and saved the data in the MongoDB database. It also performed data pre-processing to remove, replace and convert some tweet/message characters such as repetitions, replaces all URLs with the word URL, @mention with the word AT_USER, and #hashtag with hashtag word. In addition, it converted the tweet texts into lowercase letters. Using the framework, Jinja2 template engine, and HTML, the application displayed the sentiment analysis results on the web application. It also displayed exploratory visualization using charts and tweets on a map using coordinates.

5.3 Future Work

Future research can aid in the better enhancement of Social Media Sentiment Analysis using machine learning and should be able to create a new-age project by creating a search engine so that users can search for a particular word or text from the database. Otherwise, creating a bag of words using unigrams that some positive emotional words negated classified as positive such as not good or not happy. With this, future enhancements should consider the bigrams for improved classification accuracy.

References

[1] Y . Acar, G. Demet, and M. Kaynar, “Data and discourse: an assessment of Taksim urban design competition in terms of populism and participation”, Journal of Urban Design, no. 28, vol. 6, pp 682-698, 2023.

[2] S. Anno, Y. Kimura, and S. Sugita, “Using transformer-based models and social media posts for heat stroke detection”, Scientific Reports, vol. 15, no. 1, pp. 742, 2025.

[3] M. G. Albino, “Opinion-Mining Technique on Generative Artificial Intelligence Topic Using Data Classification Algorithms”.

[4] J. P. Bharadiya, “A review of Bayesian machine learning principles, methods, and applications”, International Journal of Innovative Science and Research Technology, vol. 8, no. 5, pp. 2033-2038, 2023.

[5] S. H. Chowdhury, M. Pathan, K.E. Arefin, D. Mohd, and Noorul Huq, “Prevalence and Effect of Computer Vision Syndrome during COVID-19: among Bangladeshi People”, SAS J Med, vol. 4, pp. 204-211, 2024.

[6] K. Chhutlani, V. Takrani, A. Motwani, T. Harchandani, and S. Sahu, “Sentiment Analysis of OYO Hotel Reviews Using NLP”, In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1-6, IEEE, 2023.

[7] R. K. Godara, A. Mengi, A. Sharma, and S. Sharma, “Detecting Depressive Symptoms on Social Media: A Comprehensive Review of Methodologies and Strategies for Suicide Prevention”, In International Conference on Cognitive Computing and Cyber Physical Systems, pp. 87-100, Singapore: Springer Nature Singapore, 2023.

[8] A. K. Jha, and N. K. Verma, “Social media sustainability communication: an analysis of firm behaviour and stakeholder responses”, Information Systems Frontiers, vol. 25, no. 2, pp. 723-742, 2023.

[9] A. Likas, N. Vlassis, and J. J. Verbeek, “The global k-means clustering algorithm”, Pattern recognition, vol. 36, no. 2, pp. 451-461, 2023.

[10] T. Mahmud, R, Karim, R. Chakma, T. Chowdhury, M. S. Hossain, and K. Andersson, “A benchmark dataset for cricket sentiment analysis in bangla social media text”, Procedia Computer Science, vol. 238, pp. 377-384, 2024.

[11] M. K. Mali, R. R. Pawar, S. A. Shinde, S. D. Kale, S. V. Mulik A. A. Jagtap, and P. U. Rajput, “Automatic detection of cyberbullying behaviour on social media using Stacked Bi-Gru attention with BERT model”. Expert Systems with Applications, vol. 262, pp. 125641, 2025.

[12] M. Z. Naeem, F. Rustam, A. Mehmood, I. Ashraf, and G. S. Choi, “Classification of movie reviews using term frequency-inverse document frequency and optimized machine learning algorithms”, PeerJ Computer Science, vol. 8, e914, 2022.

[13] K. Nizam and N. Ahsan, “THE IMPACT OF E-DETERMINANTS ON CUSTOMER LOYALTY: A SOCIAL SUSTAINABILITY PARADIGM”, Journal of Research in Social Development and Sustainability, vol. 2, no. 2, pp. 87-115, 2023.

[14] D. Patel, S. Saxena, T. Verma, and P. G. Student, “Sentiment analysis using maximum entropy algorithm in big data”, International Journal of Innovative Research in Science, Engineering and Technology, vol. 5, no. 5, 2016.

[15] K. Qian, A. Belyi, F. Wu, S. Khorshidi, A. Nikfarjam, R. Khot, and Y. Li, “Open Domain Knowledge Extraction for Knowledge Graphs”, arXiv preprint arXiv:2312.09424, 2023.

[16] R. S. Rana, K. Jain, and M. Hema, “Information retrieval on disaster tweets using NLP”, In AIP Conference Proceedings, vol. 3075, no. 1, AIP Publishing, 2024.

[17] B. Rosenthal and Airoldi, “Consumer morality formation on social media platforms: the case of guns in Brazil”, Journal of micromarketing, vol. 44, no. 1, pp. 178-198, 2024.

[18] C. Song, S. Chen, X. Cai, and H. Chen, “Sentiment Analysis of Spanish Political Party Tweets Using Pre-Trained Language Models”, arXiv e-prints, arXiv-2411, 2024.

[19] S. Suryawanshi, B. R. Chakravarthi, P. Verma, M. Arcan, J. P. McCrae, and P. Buitelaar, “ A dataset for troll classification of TamilMemes”, In Proceedings of the WILDRE5–5th workshop on indian language data: resources and evaluation, pp. 7-13, 2020.

[20] W. Ullah, P. Oliveira-Silva, M. Nawaz, R. M. Zulqarnain, I. Siddique, and M. Sallah, “Identification of depressing tweets using natural language processing and machine learning: Application of grey relational grades”. Journal of Radiation Research and Applied Sciences, vol. 18, no. 1, 2025.

[21] S. Vyavahare, S. Teraiya, and S. Kumar, “FDM manufactured auxetic structures: an investigation of mechanical properties using machine learning techniques”. International Journal of Solids and Structures, vol. 265, pp. 112-126, 2023.

[22] L. Wang, “Support vector machines: theory and applications”, vol. 177, Springer Science & Business Media, 2005.

APPENDIX A Implementation code

```
!nvidia-smi
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 525.105.17   Driver Version: 525.105.17   CUDA Version: 12.0     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  Tesla T4            Off  | 00000000:00:04.0 Off |                    0 |
| N/A   55C    P8    10W /  70W |      0MiB / 15360MiB |      0%      Default |
|                               |                      |                  N/A |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
|        ID   ID                                                   Usage      |
|=============================================================================|
|  No running processes found                                                 |
!pip install -q -U watermark
!pip install -qq transformers
%reload_ext watermark
%watermark -v -p numpy,pandas,torch,transformers
#@title Setup & Config
import transformers
from transformers import BertModel, BertTokenizer, AdamW, get_linear_schedule_with_warmup
import torch
import numpy as np
import pandas as pd
import seaborn as sns
from pylab import rcParams
import matplotlib.pyplot as plt
from matplotlib import rc
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix, classification_report
from collections import defaultdict
from textwrap import wrap
from torch import nn, optim
from torch.utils.data import Dataset, DataLoader
import torch.nn.functional as F
%matplotlib inline
%config InlineBackend.figure_format='retina'
sns.set(style='whitegrid', palette='muted', font_scale=1.2)
HAPPY_COLORS_PALETTE = ["#01BEFE", "#FFDD00", "#FF7D00", "#FF006D", "#ADFF02", "#8F00FF"]
sns.set_palette(sns.color_palette(HAPPY_COLORS_PALETTE))
rcParams['figure.figsize'] = 12, 8
```

```
RANDOM_SEED = 42
np.random.seed(RANDOM_SEED)
torch.manual_seed(RANDOM_SEED)
device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
device
!gdown --id 1S6qMioqPJjyBLpLVz4gmRTnJHnjitnuV
!gdown --id 1zdmewp7ayS4js4VtrJEHzAheSW-5NBZv
df = pd.read_csv("reviews.csv")
df.head() sns.countplot(df.score)
plt.xlabel('review score');
def to_sentiment(rating):
  rating = int(rating)
  if rating <= 2:
    return 0
  elif rating == 3:
    return 1
  else:
    return 2
df['sentiment'] = df.score.apply(to_sentiment)
```