Vol. 3 No. 4 (2025)
ISSN 3007-3170(O)
ISSN :3007-3162(P)
Social Sciences & Humanity Research Review

Multi-Class Sentiment Analysis of Twitter Data Using Bert-Based Model

[1]Shahbaz Hassan Wasti, [1]Ghulam Jillani Ansari* , Dr. Jahanzeb Jahan[2]

[1]Department of Information Sciences, Division of Science and Technology, University of Education, Lahore, 54770, Pakistan,
Email: ghulamjillani@ue.edu.pk, ORCID : https://orcid.org/0000-0002-8985-1383
Email: shahbazwasti@ue.edu.pk , ORCID : https://orcid.org/0000-0001-5788-2604
[2]Department of English, Division of Arts & Social Sciences, University of Education, Lahore, Email: jahanzeb@ue.edu.pk, ORCID : https://orcid.org/0000-0001-6954-2308

ARTICLE INFO

Keywords: BERT, pre-trained model, sentiment analysis, twitter dataset

Corresponding Author:
Ghulam Jillani Ansari
Email:
ghulamjillani@ue.edu.pk
ORCID:
https://orcid.org/0000-0002-8985-1383

ABSTRACT
Modern world is a world of social media. X formerly known as Twitter is one of the leading social media platform that represents the public opinions from across the globe on every walk of life. These opinions are not just short texts but represent the sentiments of the public. The sentimental analysis of this large scale text data give insightful information to organizations, researchers, policy makers, and government institutes. Therefore, this work proposed a BERT (Bidirectional Encoder Representations from Transformers)-based sentiment analysis model. The model classifies the X (Twitter) text or tweets as positive, negative or neutral by analyzing the text data. The proposed model is evaluated on accuracy, precision, recall and F1 score. The experiment results showed that the proposed model yielded promising performance against all sentiment classes.

1 Introduction

This era arguably the era of social media platforms. These social media platform such Facebook, YouTube, Instagram, X (Twitter), Blogs and review sites have enabled public to give their opinion in all forms e.g., video, audio, text or images. X or formerly known as Twitter, is one of the leading social media application that allows its users to post their opinions in a small text format with maximum 280 characters for the standard users. The text published on X platform is referred to as tweets (Samir et al., 2021). The X application is used across every field of life, e.g., entertainment, academics, sports, politics, government, social services, journalists, news channels, and individuals. The X also authenticates the validity or identity of its user whether the user is a person or an organization, this feature makes X application more reliable source of information (Afuan et al., 2025). The restricted or short text make it more interested, while even harder for natural language processing applications (He, 2024; Marrapu et al., 2024).

Tweets generally represents opinions and sentiments of public. These sentiments can form a global opinion and can have positive and negative impact on the society (Bovet & Makse, 2019). However, the language of tweets is inherently complex due to words limit, informal language, use of images and icons, and even the ambiguous way of expressing the opinion (Akbar et al., 2020). The sentiments of these tweets can be categorized as neutral, negative or positive. It is an important task to conduct a sentiment analysis of these tweets. Several machine learning and deep learning-based solutions have been proposed for sentiment analysis (Afuan et al., 2025; Heikal et al., 2018; Jianqiang et al., 2018). The sentiment classification of tweets is a complex task, as discussed earlier it is due the language used in the tweets. But rapid growth in learning-based approaches have shown great performance. These methods require a large amount of annotated data for training and learning hidden features and semantics of the language to conduct sentiment analysis (Ain et al., 2017; Chauhan et al., 2024; Zhang et al., 2018). Such techniques requireintensive pre-processing of the text data that includes text refinement, annotation, tokenization, training of the learning models and building feature vectors (Zhang et al., 2018). These feature vectors can be used to perform sentiment analysis of the give text.

Recently, with the advancements in large language models, transformer based language models have shown improved performance in text related tasks (Bilal & Almazroi, 2022). One of the most notable transformer-based language model is called Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019). BERT is a pre-trained language model that is trained over billions of tokens to capture inherit features of the text in a corpus. Therefore, in this paper we will use BERT model to perform sentiment analysis of the tweets on three classes, e.g., positive, negative and neutral. In particular, we fine tune the BERT model on annotated tweets to perform sentiment analysis. The proposed solution is evaluated on well-known evaluation metrics, e.g., accuracy, precision, recall, and F1 score. The fine tuning also involves substantial pre-processing and data cleaning to make the data ready for the propose model.

The remaining sections of this paper are: the Section 2 after introduction describes the related work on sentiment analysis for Twitter-based datasets. The methodology of the proposed solution is presented in the Section 3, while Section 4 discussed the results, and, finally the Section 5 concludes the paper.

2 Related Work

Over the recent years, the machine and deep learning methods have been used to develop solution of different fields. Similarly, these approaches have been also widely used for social media content analysis.. Therefore, in this section we will briefly cover the basics and latest solutions developed for sentiment analysis.

One of the fundamental development in this field was use of artificial neural network to generate vector representation of words referred to as word embedding (Fauzi, 2019). The emeddings were generated by training the model on the corpus of billions of tokens or words. The Word2Vec has been one of the most used word embedding in natural language processing applications. Several Word2Vec-based methods have been proposed over the years for sentiment analysis (Al-Saqqa & Awajan, 2019). Alayba et al., (2018) used Word2Vec skip-gram and bag of words model to perform sentiment analysis on Arabic text. The Naïve bayes and Support Vector Machine (SVM) classification are used in the proposed model to classify the tweets in positive, negative or neutral. The model achieved good performance. Ashi et al., (2018) also used Word2Vec embedding and linear SVM to perform sentiment analysis of airline tweets dataset. Deho et al., (2018) proposed a random forest-based classifier using Word2Vec embedding to classify the tweets. Following the word embedding work, recent advance in natural language processing has focused on language transformer based model of word vectors. The BERT is one of the prominent solution in this regard (Afuan et al., 2025). Zhang et al., (2020) have used transfer learning approaches to analyze sentiments about HPV vaccines. In their approach they used GPT and BERT as pre-trained model. The results showed that BERT outperformed all the other counterparts. Pathak et al., (2021) proposed aspect-based sentiment analysis by fine tuning BERT model. The model showed competitive performance on different languages. Afuan et al., (2025) also developed a BERT-based sentiment analysis model to analyze tweets in Indonesian language.

3 Methodology

Figure 1 provides the graphical view of the proposed methodology. A pre-trained transformer language model BERT has the ability to learn the language insights even from unlabeled data. Moreover, it can be further fine-tuned with labeled dataset for a specific use case like sentiment analysis of twitter data.

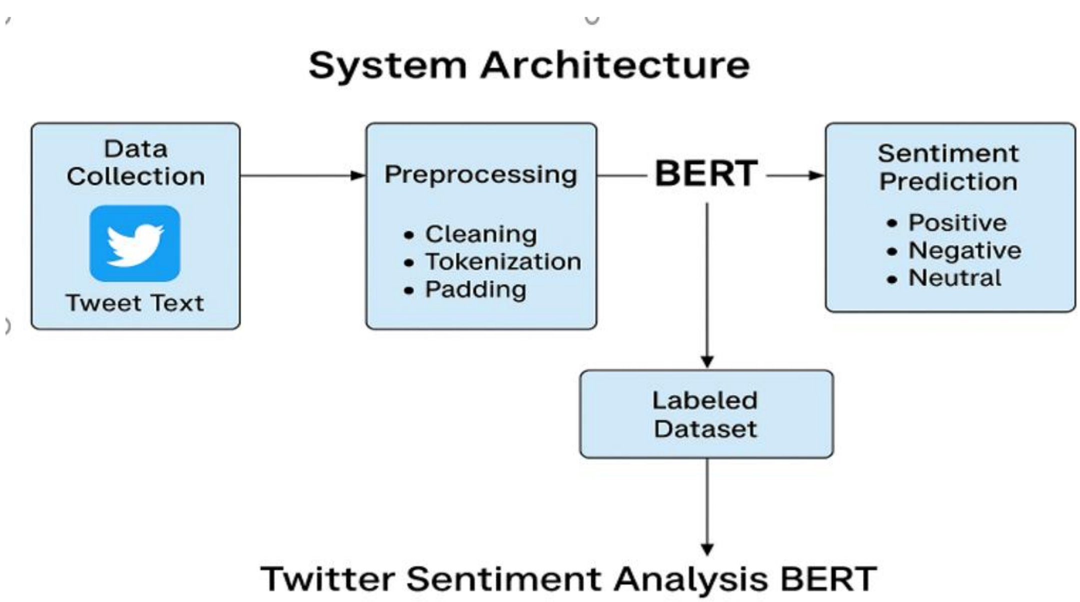


Figure 1 Proposed Methodology

3.1 Dataset

We obtained the dataset from Kaggle, public dataset repository. The dataset comprises on sentiment corpus of over 4500 tweets in English language. The tweets in the dataset are annotated into three sentiment classes: positive, negative, and neutral. The dataset used in this study consists of a publicly available Twitter sentiment corpus containing over 4,500 English-language tweets annotated into three sentiment classes: positive, negative, and neutral. Since, the dataset is downloaded from a public repository and it is designed to perform sentiment analysis of social media opinions. This paper will use the dataset in 80%, 10% and 10% ratio for training, validation and testing purposes using BERT pre-trained model. **Table 1**provides the summary of the dataset used in training BERT model.

Table 1 Training dataset for sentiment analysis

Sr. No.	Sentiment	Number of Tweets
1	Negative	1001
2	Neutral	1430
3	Positive	1103

3.2 Pre-processing and data cleaning

It is an important step before feeding the data to a machine or deep learning pipeline. We used Python language and NLTK library to clean the data. The data cleaning process involved following steps:

1. Removing of unnecessary characters, symbols, punctuations and stop words.
2. Uncasing of capitalized words.
3. Lemmatization and tokenization of the words using NLTK stemmer.

Table 2 provides the output of the pre-processing step.

Table 2 Dataset before and after pre-processing

No.	Original text	Refined text
1	Huh, another ScarePoint coding Sunday	huh another scarepoint coding sunday
2	resorted to eating Mickey Ds ALONE.	restore eating mickey ds alone
3	Watching Body of Lies...good film	watch body lies good film
4	Now I have a sunburn	now i have sunburn
5	I want to see David cook!!	i want to see david cook

3.3 Model training and tuning

The BERT model was trained on annotated dataset with while adjusting model hyperparameters like batch size, learning rate, optimizer, activation function and epochs mentioned in the Table 3.

Table 3 BERT Hyperparameters

No	Hyperparameter	Value Used
1.	Model	BERT-Base-Uncased
2.	Batch Size	16
3.	Learning Rate	2.00E-05
4.	Optimizer	AdamW
5.	Weight Decay	0.01
6.	Activation Function	GELU
7.	Number of Epochs	3
8.	Maximum Sequence Length	128
9.	Loss Function	Cross-Entropy Loss
10.	Warm-up Steps	10% of total steps
11.	Dropout Rate	0.1

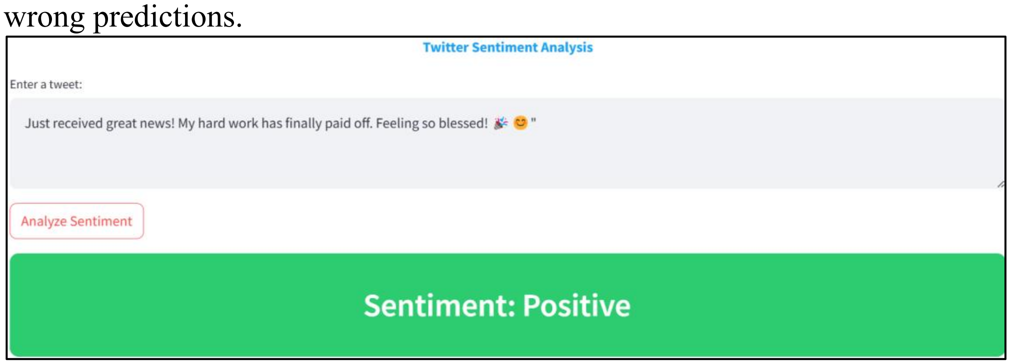
4 RESULTS AND DISCUSSIONS

4.1 Experiment detail

Google Colaboratory with NVIDA Tesla T4 GPU, and 16 GB VRAM was used to execute the model training and testing process. The model was developed using Python, relatd libraries for transformers to fine tune the BERT Model. A light-weight web application is also developed which takes input from the user and based on the training classifies the user input into positive, negative or neutral.

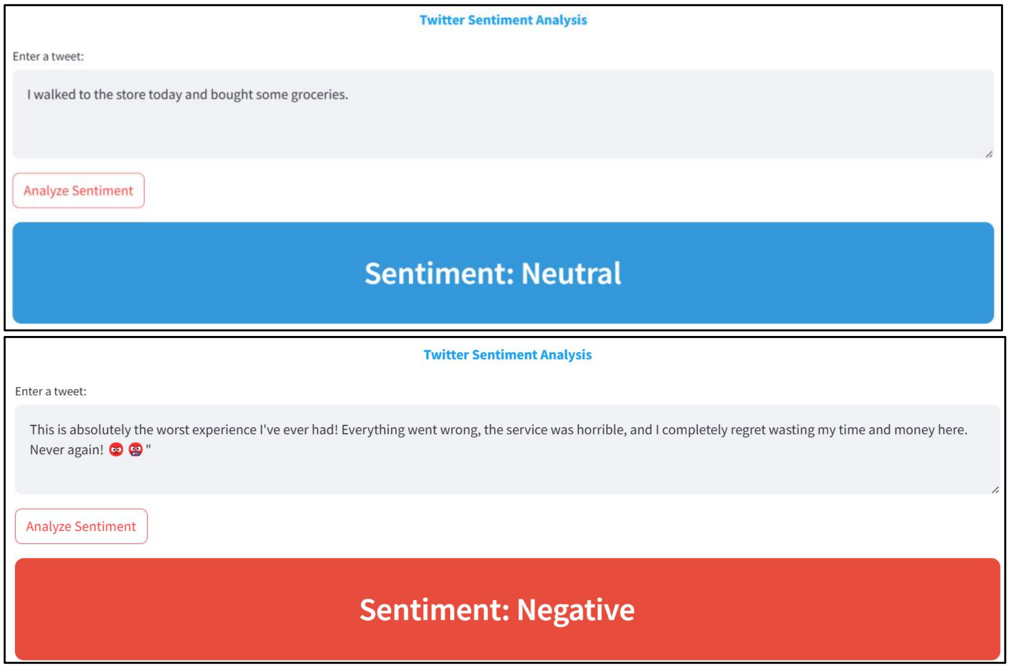
4.2 Results

The proposed model was tested on two grounds qualitative and quantitative. **Figure 2** shows the qualitative results via user input from the web application. The application successfully classifies the user input according to the provided text, i.e., positive, negative and neutral. There were chances where we got false positive results also. It was observed that the model was trained on a smaller dataset therefore, it has generated few wrong predictions.



(a)

(b)



(c)

Figure 2 Qualitative result using web application

Table 4 Evaluation results

Sentiment	Precision	Recall	F1-score
positive	0.84	0.76	0.79
negative	0.72	0.81	0.79
neutral	0.74	0.72	0.73

Table 4 lists the results on evaluation metrics like precision, recall, and F1-score. However, the overall accuracy of the model is 82%. Although some modern or state of the art methods have achieved higher accuracy even higher than 85%. Since we have used a smaller annotated dataset and for deep learning models a higher dataset gives better result. But the other metrics especially reflect an overall balanced and reliable performance of the model.

The overall, results assured that a fine-tuned pre-trained model BERT can effectively use for sentiment analysis. However, it is also observed that a larger dataset is required for better learning of parameters. Therefore, as a future work we will built a larger dataset to further enhance this model.

5 Conclusions

This work used a pre-trained transformer-based language model BERT for multi-class sentiment analysis. The model was fine-tuned with annotated X (Twitter) dataset. The proposed model shows a good performance and achieved a reasonable accuracy of 82%. While other metrics also performed in an acceptable range. The results of both the qualitative and quantitative evaluation are very encouraging. However, an important finding of this study is the size of the dataset used. It is observed that model like BERT requires a large amount of data to improve and tune its parameters. Therefore, in future we would like to train this dataset on a larger dataset to get better results.

6 REFERENCES

Afuan, L., Hidayat, N., Hamdani, H., Ismanto, H., Purnama, B. C., & Ramdhani, D. I. (2025). Optimizing BERT Models with Fine-Tuning for Indonesian Twitter Sentiment Analysis. *J Wirel Mob Netw Ubiquitous Comput Dependable Appl*, *16*(2), 248–267.

Ain, Q. T., Ali, M., Riaz, A., Noureen, A., Kamran, M., Hayat, B., & Rehman, A. (2017). Sentiment analysis using deep learning techniques: a review. *International Journal of Advanced Computer Science and Applications*, *8*(6).

Akbar, M. R., Slamet, I., & Handajani, S. S. (2020). Sentiment analysis using tweets data from Twitter of Indonesian's Capital City changes using classification method support vector machine. *AIP Conference Proceedings*, *2296*(1), 20041.

Al-Saqqa, S., & Awajan, A. (2019). The use of word2vec model in sentiment analysis: A survey. *Proceedings of the 2019 International Conference on Artificial Intelligence, Robotics and Control*, 39–43.

Alayba, A. M., Palade, V., England, M., & Iqbal, R. (2018). Improving sentiment analysis in Arabic using word representation. *2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR)*, 13–18.

Ashi, M. M., Siddiqui, M. A., & Nadeem, F. (2018). Pre-trained word embeddings for Arabic aspect-based sentiment analysis of airline tweets. *International Conference on Advanced Intelligent Systems and Informatics*, 241–251.

Bilal, M., & Almazroi, A. A. (2022). Effectiveness of fine-tuned BERT model in classification of helpful and unhelpful online customer reviews. *Electronic Commerce Research*, 1.

Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, *10*(1), 7.

Chauhan, P., Sharma, N., & Sikka, G. (2024). On the importance of pre-processing in small-scale analyses of twitter: a case study of the 2019 Indian general election. *Multimedia Tools and Applications*, *83*(7), 19219–19258.

Deho, B. O., Agangiba, A. W., Aryeh, L. F., & Ansah, A. J. (2018). Sentiment analysis with word embedding. *2018 IEEE 7th International Conference on Adaptive Science & Technology (ICAST)*, 1–4.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186.

Fauzi, M. A. (2019). Word2Vec model for sentiment analysis of product reviews in Indonesian language. *International Journal of Electrical and Computer Engineering*, *9*(1), 525.

He, L. (2024). Enhanced twitter sentiment analysis with dual joint classifier integrating RoBERTa and BERT architectures. *Frontiers in Physics*, *12*, 1477714.

Heikal, M., Torki, M., & El-Makky, N. (2018). Sentiment analysis of Arabic tweets using deep learning. *Procedia Computer Science*, *142*, 114–122.

Jianqiang, Z., Xiaolin, G., & Xuejun, Z. (2018). Deep convolution neural networks for twitter sentiment analysis. *IEEE Access*, *6*, 23253–23260.

Marrapu, S., Senn, W., & Prybutok, V. (2024). Sentiment analysis of twitter discourse on omicron vaccination in the USA Using VADER and BERT. *Journal of Data Science and Intelligent Systems*.
Pathak, A., Kumar, S., Roy, P. P., & Kim, B.-G. (2021). Aspect-based sentiment analysis in Hindi language by ensembling pre-trained mBERT models. *Electronics*, *10*(21), 2641.
Samir, A., Elkaffas, S. M., & Madbouly, M. M. (2021). Twitter Sentiment Analysis Using BERT. *2021 31st International Conference on Computer Theory and Applications (ICCTA)*, 182–186. https://doi.org/10.1109/ICCTA54562.2021.9916614
Zhang, L., Fan, H., Peng, C., Rao, G., & Cong, Q. (2020). Sentiment analysis methods for HPV vaccines related tweets based on transfer learning. *Healthcare*, *8*(3), 307.
Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1253.