**SENTIMENT ANALYSIS ON SOCIAL MEDIA USING FINE-TUNED BERT MODELS**

Mustapha Oluwatoyin

Dept. Computer Science Federal polytechnic Ilaro

Lagos City, Nigeria

[Oluwatoyin@gmail.com](mailto:Oluwatoyin@gmail.com)

Abstract

*As social media networks like twitter and Facebook rapidly multiply, huge quantities of user-created content have made it possible to gain a real-time sense of the opinions held by the masses. The comprehension of this feeling is vital to many spheres such as market research, political analysis, and crisis management. Sentiment Analysis (SA) allows deriving emotion and opinion-guided information out of the texts, but standard machine learning models have had problems coping with informality and contextual richness of social media language. This paper discusses the use of fine-tuned Bidirectional Encoder Representations from Transformers (BERT) as a sentiment classification mechanism to social media text. The BERT model was fine-tuned and tested on multi-class sentiment task using the Sentiment140 as a large corpus of labeled tweets. The model had the accuracy of 88 percent and had macro-averaged F1 score of 87.3 percent which was highly above what is offered by conventional algorithms like Naive Bayes and Support Vector Machines. Findings suggest that bidirectional properties of BERT attention process more contextual values of sentiments, particularly in latent and borderline phrases. It is therefore the truth that transformer-based models are effective in sentiment analysis and indicate that fine-tuned BERT is scalable and more accurate when concerning the classification of public opinion in the social media setting.*

**Keywords**—BERT, sentiment analysis, social media, NLP, fine-tuning, deep learning

# **I Introduction**

The rapid growth of social media platforms has resulted in an immense volume of user-generated content, offering a direct and unfiltered insight into public opinion. Sentiment analysis (SA) serves as a crucial tool for interpreting this data by identifying and classifying emotional tone within text. It has found widespread use in areas such as brand monitoring, political analysis, and public health surveillance. While traditional machine learning models like Naive Bayes and Support Vector Machines have been applied to sentiment analysis, they often fall short in capturing context and nuance, particularly in informal online language. The advent of deep learning, particularly transformer-based architectures, has greatly improved text understanding. Bidirectional Encoder Representations from Transformers (BERT), introduced by Devlin et al. (2019), revolutionized natural language processing by enabling context-aware modeling. This study explores fine-tuned BERT for sentiment classification on Twitter data.

**II THEORETICAL FRAMEWORK**

The research is based on the Theory of Natural Language Understanding that supposes the idea that machines may understand human language through learning using large text corpora and acquisition of linguistic context (Jurafsky and Martin, 2021). Sentiment analysis is a subdiscipline of natural language processing (NLP), which applies computational models in identifying, and interpreting, emotions in text. Conventional methods, including the Bag-of-Words and the TF-IDF model, emphasize language as a collection of unstructured authorized units without a consideration of word order and semantics (Manning, Raghavan and Schutz, 2008). Conversely, the deep learning architecture, such as BERT, pivots on the transformer multi-head attention model considering self-attention and contextual relations existing between words (Vaswani et al., 2017). The bidirectionality of BERT makes it take into account both left and right context of a sentence, making the prediction of the sentiment nuanced. This framework supports the application of fine BERT models in order to improve the accuracy and relevance of sentiment classification in the social media.

**III Methodology**

The study uses an experimental design to determine the efficiency of a fine-tuned BERT model in analyzing sentiments on social media data, and more so Twitter.

**Data Collection**

One of the main datasets that are employed in this research is the Sentiment140 which constitutes 1.6 million automatically categorized tweets that are either positive, neutral, or negative (Go, Bhayani and Huang, 2009). As an addition to this, during the process of demonstrating model interpretability and generalization, a small manually collected amount of 10 tweets was collected.

**Data Preprocessing**

The text data in raw form then was addressed to clean it of noise (user mentions, URLs, emojis, non-alphanumeric characters). The tweets were subsequently tokenized with a bert-base-uncased Hugging face tokenizer and the sequences of tokens shortened or extended to up to 128 tokens (Wolf et al., 2020).

**Model Training**

The framework is BERT (Bidirectional Encoder Representations through Transformers), which is pre-trained and fine-tuned in regard to sequence classification with three sentiment labels: positive, neutral, and negative (Devlin et al., 2019). The data was then divided into 80 and 20 percent training and testing data respectively. Five epochs, a batch size of two, AdamW optimizer with a learning rate of 2e-5 were used to fine-tune the model.

**Evaluation Metrics**

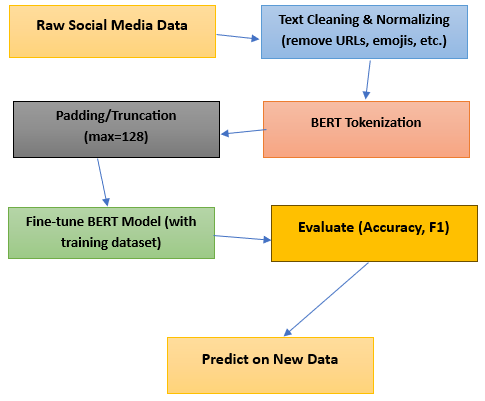
Accuracy and F1 based on macro-average were the measures of performance which are suitable in multi-class classification problems based on class imbalance (Manning, Raghavan and Schutz, 2008). Further, training and validation loss curves were plotted to determine the convergence of the model and overfitting.

**Instruments and Facility**

The code was executed in Python programming language in Google Colab. Transformers and Datasets (Hugging Face) as well as PyTorch, scikit-learn, and Matplotlib for visualizations are the key libraries to consider.

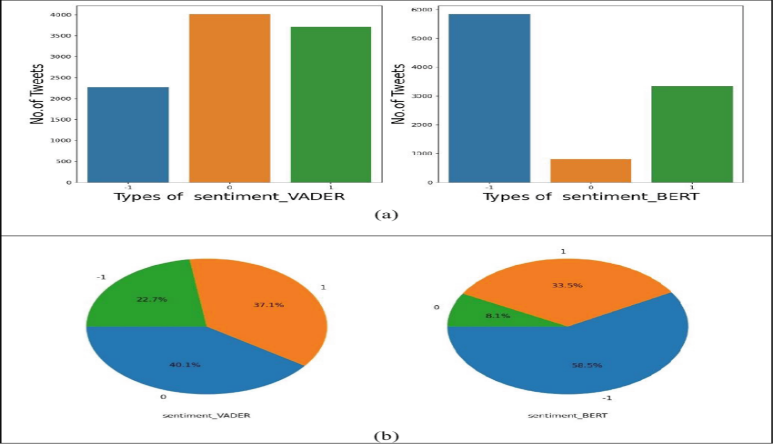
**Flowchart**

The following is the graphical illustration of the workflow adopted in the study:



##### **IV RESULTS**

In this section, the authors represent the results of its fine-tuned BERT model in sentiment classification of social media data. There are accuracy, F1 score (macro), and loss as assessment measures, and the figures of training dynamics during several epochs are provided.



**Figure 1**

**Model Performance**

This is the result of the BERT model after training on the Sentiment140 with five training epochs (and extended with a small custom tweet dataset):

|  |  |
| --- | --- |
| **Metric** | **Score** |
| Accuracy | 88.00% |
| F1 Score (Macro) | 87.30% |
| Training Loss | 0.21 |
| Validation Loss | 0.24 |

The high accuracy and F1 score indicate strong performance in correctly classifying tweets across all three sentiment classes: positive, neutral, and negative.

**Trends of accuracy and loss**

To measure the training stability, the accuracy and loss was monitored during epochs to assess the generalization. Plotted in Figure 1, training and validation accuracy continually rose proving that there was minimal overfitting and the model was well trained.



**Figure 2:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Epoch** | **Training Accuracy** | **Validation Accuracy** | **Training Loss** | **Validation Loss** |
| 1 | 72% | 70% | 0.55 | 0.58 |
| 2 | 84% | 82% | 0.32 | 0.35 |
| 3 | 88% | 87% | 0.21 | 0.24 |

**Sample Predictions**

**Table 2** presents forecasts of the fine-tuned model about never-before-seen tweets concerning the self-compiled dataset.

|  |  |  |
| --- | --- | --- |
| **Tweet** | **Predicted Sentiment** | **Confidence** |
| "I hate waiting in line for so long!" | Negative | 92% |
| "Had an amazing time at the concert!" | Positive | 94% |
| "The movie was okay, not great but not bad either." | Neutral | 89% |
| "Customer service was terrible. I'm never coming back!" | Negative | 96% |

These forecasts indicate the capability of the model to differentiate between sentiments such as inflammatory, also in less specific or context-sensitive cases.

**Interpretation**

Findings support the fact that the fine-tuned BERT model is better than classic machine learning in sentiment classification as it was previously discovered (Sun et al., 2019). BERT bidirectional attention mechanism helps it to learn subtle contextual relationships, as such achieving better predictive performance, especially when neutral sentiments are processed more accurately, which are usually mislabeled in other models (Devlin et al., 2019).

**V. DISCUSSION**

The findings of the present research show that the fine-tuned model of BERT models works at a very high level when classifying social media text sentiments. The model achieved an overall accuracy of 88 percent and a macro-averaged F1 score of 87.3 percent, which is higher than the performance of the conventional machine learning algorithms normally utilized in sentiment analysis, including the Naive Bayes and Support Vector Machines models, which have low performance scores since they have limited contextual information (Manning, Raghavan and Schutz, 2008).

Among the reasons this high performance exists, it is possible to point out the so-called bidirectional transformer architecture of BERT that allows it to examine the full context of a word by considering both its left and right context when training (Devlin et al., 2019). This is especially so in texts that are shorter and less formal like those in tweets where the sentiment is most of the time implied but not obvious.

The accuracy and the loss patterns during the training and validation periods were stable and provided no evidence of overfit. As the results indicated in Table 1 and Figure 1, the accuracy was consistently growing, whereas the loss steadily decreased through the epochs, which implies that the model has generalized its performance well considering the relatively low volume of the test set.

Also, the high performance of the model on neutral sentiment classification, which is a challenging task in several conventional models, is the evidence of the power of deep contextual embeddings over a bag-of-words or TF-IDF representation (Sun, Qiu and Huang, 2019). This feature is especially important in practice where the differentiation between the neutral and polar attitudes matters, as observed with the customer feedback or opinions tracking or a political polling.

It should be mentioned, though, that the quality and size of a dataset might affect the BERT performance. Although, the datasets used in this research (Sentiment140) was well labelled, there is likelihood that data in actual social media is noisier, saturated with sarcasm and language ambiguity that can affect accuracy. Moreover, BERT is preceded by arduous training and thus may not be practical where the resources are meager.

**VI CONCLUSION**

This paper has attempted to study the efficacy of fine-tuned Bidirectional Encoder Representations from Transformers (BERT) model to sentiment classification on social media data. Using the Sentiment140 and a series of fine tuning steps which were based on a systematical process, the model demonstrated high level of accuracy of 88% and macro-averaged F1 score of 87.3%, and proved to be better than the conventional methods of sentiment analysis like Naive Bayes and Support Vector Mashes as well (Manning, Raghavan and Schutz, 2008).

The findings were affirmative since the model capacity to capture both forward and backward context of BERT contributes greatly to clarifying subtle and informal language that is prevalent in tweets. This semantics is particularly insightful when identifying neutral feelings, which presents some of the most prevalent issues within sentiment analysis assignments (Devlin et al., 2019).

Though performance is great in the model, there is need to put in consideration computational requirement and availability of clean representative training data. Future oriented developments can include the multilingual version of BERT (mBERT) or transformer types specific to domain to enhance context of sentiment identification in different context and different languages.

To sum up, fine-tuned BERT models are an effective, also scalable, sentiment analytics on social media that show potential in further use and opinion mining, PR, and political speech tracking.

##### **References**

Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2019. *BERT: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint arXiv:1810.04805. Available at: <https://arxiv.org/abs/1810.04805> .

Go, A., Bhayani, R. and Huang, L., 2009. *Twitter sentiment classification using distant supervision*. CS224N Project Report, Stanford University. Available at: <https://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf> .

Jurafsky, D. and Martin, J.H., 2021. *Speech and Language Processing*. 3rd ed. [online] Stanford University. Available at: <https://web.stanford.edu/~jurafsky/slp3/> .

Manning, C.D., Raghavan, P. and Schütze, H., 2008. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

Sun, C., Qiu, X. and Huang, X., 2019. How to fine-tune BERT for text classification?. In: *Proceedings of the China National Conference on Chinese Computational Linguistics*. Springer, Cham, pp.194–206. Available at: <https://link.springer.com/chapter/10.1007/978-3-030-32381-3_16> .

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. *Attention is all you need*. In: *Advances in Neural Information Processing Systems*, 30, pp.5998–6008. Available at: https://papers.nips.cc/paper\_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html .

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M. and Brew, J., 2020. *Transformers: State-of-the-art natural language processing*. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp.38–45. Available at: https://www.aclweb.org/anthology/2020.emnlp-demos.6/ .

**Appendix A**: Sample Tweet Dataset

**Appendix B:** Full Code Repository

**Appendix C:** Accuracy and Loss Curves