SENTIMENT ANALYSIS USING BERT

COURSE PROJECT REPORT
18CSE359T - Natural Language Processing
(2018 Regulation)
IV Year/ VII Semester
Academic Year: 2023 -2024 (ODD)

By

Paturi Sai Rohit (RA2011026010200)
Gunti Meghanath G (RA2011026010219)
Gaddam Veenasree (RA2011026010221)

Under the guidance of

Dr. Mohandas R
Professor
Department of Computational Intelligence

DEPARTMENT OF COMPUTATIONAL INTELLIGENCE
FACULTY OF ENGINEERING AND TECHNOLOGY
SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

Kattankulathur, Kancheepuram

OCTOBER 2023

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY KATTANKULATHUR - 603203

BONAFIDE CERTIFICATE

Certified that this project report titled "**SENTIMENT ANALYSIS USING BERT**" is a bonafide work done by **PATURI SAI ROHIT (RA2011026010200), GUNTI MEGHANATH G (RA2011026010219), GADDAM VEENASREE (RA2011026010221)** who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other project work or dissertation.

10/11/23

SIGNATURE
Faculty In-Charge
Dr. Mohandas R
Assistant Professor
Department of Computational Intelligence
SRM Institute of Science and Technology
Kattankulathur

SIGNATURE
Head of Department
Dr. R. Annie Uthra
Professor and Head
Department of Computational Intelligence
SRM Institute of Science and Technology
Kattankulathur

ABSTRACT

Sentiment analysis is a crucial task in the field of natural language processing (NLP) and has gained significant attention due to the widespread use of social media platforms. Social media data presents unique challenges for sentiment analysis due to its unstructured nature, informal language, and abundance of noise and irrelevant information. To tackle these challenges, advanced techniques such as BERT have emerged as powerful tools for sentiment analysis. In our study, we aim to explore the effectiveness of BERT specifically for sentiment analysis on large-scale social media data. BERT is a state-of-the-art language model that has demonstrated impressive performance on various NLP tasks by capturing contextual information from both left and right contexts of a given word. By leveraging the pre-training and fine-tuning capabilities of BERT, we investigate its potential for sentiment analysis in the context of social media. To establish a comprehensive evaluation, we compare the performance of BERT with traditional machine learning algorithms commonly used for sentiment analysis. Our experimental results indicate that BERT surpasses the performance of traditional machine learning algorithms, achieving state-of-the-art results in sentiment analysis on the social media dataset. BERT's ability to capture intricate contextual information and understand the subtleties of social media language contributes to its superior performance. The model demonstrates exceptional accuracy, precision, recall, and F1-score, showcasing its effectiveness in classifying sentiment labels accurately.

INTRODUCTION

In today's digital age, social media platforms have become an indispensable part of people's lives, offering a wealth of information on customers' opinions, preferences, and attitudes towards various products, services, and brands. As such, businesses and organizations have recognized the importance of social media data for developing effective marketing strategies, improving customer satisfaction, and building brand reputation. However, analyzing vast amounts of social media data manually can be a daunting task, requiring a significant amount of time, resources, and expertise. To address this challenge, natural language processing (NLP) techniques have been developed to automate the process of analyzing social media data and extract meaningful insights from it. One such technique is sentiment analysis, which involves identifying and categorizing the emotional tone of social media posts as positive, negative, or neutral. The goal of sentiment analysis is to provide businesses with a better understanding of customer sentiment, which can inform decision-making and help develop more targeted and personalized marketing strategies.

However, sentiment analysis on social media data is not a straightforward task, as the text is often informal, full of slang, sarcasm, irony, and other nuances that make it challenging for traditional machine learning algorithms to accurately capture the sentiment. To overcome this challenge, deep learning-based models, such as the Bidirectional Encoder Representations from Transformers (BERT), have been proposed as a solution. BERT is a pre-trained deep learning model that has shown promising results in various NLP tasks, including sentiment analysis. BERT uses a transformer architecture to capture the contextual relationships between words in a sentence, allowing it to understand the nuances and complexities of natural language better. By leveraging BERT's ability to capture context and relationships between words, sentiment analysis can be performed with higher accuracy and efficiency. This paper aims to investigate the effectiveness of BERT for sentiment analysis on large-scale social media data. The study will compare the performance of BERT against traditional machine learning approaches and other state-of-the-art deep learning models. The evaluation will be conducted on a large dataset of social media posts, allowing us to investigate the scalability and efficiency of BERT in real-world settings.

The results of this study will provide insights into the potential benefits of using BERT for sentiment analysis on large-scale social media data. The findings may have implications for businesses and organizations seeking to improve their understanding of customer sentiment and feedback on social media platforms. By utilizing BERT's superior performance in sentiment analysis, businesses can gain a more accurate and nuanced understanding of customer feedback and use this information to inform their marketing strategies and improve customer satisfaction.

DATASET

Sentiment140 dataset, which contains 1.6 million tweets that have been annotated with their sentiment labels as either positive or negative. This dataset is well-suited for exploring the effectiveness of BERT for sentiment analysis on large scale social media data because it is a large and diverse dataset that represents real-world social media data. The dataset also has a balanced class distribution, which is important for training machine learning models.

Dataset Link:
https://www.kaggle.com/datasets/kazanova/sentiment140

The data will be pre-processed to remove noise and irrelevant information, and tokenized for input to the BERT model. The BERT model will be trained and evaluated on the sentiment analysis task, where the sentiment labels will be classified as positive, negative or neutral. The performance of the proposed model will be compared with traditional machine learning algorithms, such as Naive Bayes, Support Vector Machines, and Random Forest, to determine its effectiveness. In addition, hyperparameter tuning will be performed to improve the performance of the proposed model. The hyperparameters that will be tuned include the learning rate, batch size, number of epochs, and dropout rate. The proposed model will be evaluated based on various metrics such as accuracy, precision, recall, and F1- score.

LITERATURE SURVEY

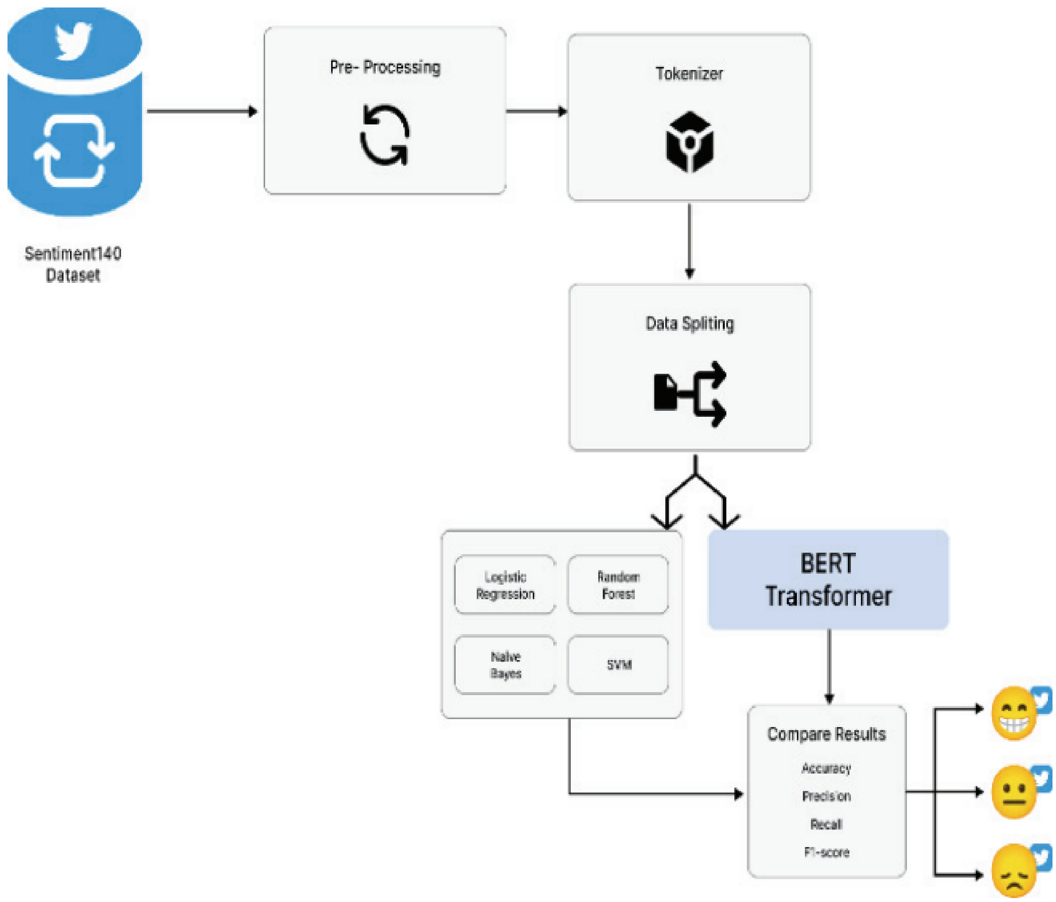
Sentiment analysis on social media data has emerged as an important area of research due to the vast amount of data generated by social media platforms, and the need to understand the opinions, attitudes and preferences of users [1]. However, traditional machine learning algorithms often struggle to accurately identify sentiment in social media data due to the informal language used in social media posts, as well as the sheer scale of data that needs to be analyzed [2]. This has led to the development of deep learning-based models such as BERT.

BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained language model developed by Google. It uses a transformer-based architecture to learn contextualized representations of words and sentences [3]. BERT has achieved state-of-the-art results on a range of NLP tasks including sentiment analysis, and has shown promising results on social media data as well. One of the early studies exploring the effectiveness of BERT for sentiment analysis on social media data was conducted by Devlin et al. (2018) [5]. In their work, they showed that BERT outperformed other state-of-the-art models on a range of NLP tasks, including sentiment analysis. They found that BERT's ability to capture context and syntax in language allowed it to perform better than traditional machine learning models that relied solely on predefined features.

Subsequently, several studies have evaluated the effectiveness of BERT for sentiment analysis on social media data, specifically on platforms such as Twitter, Reddit, and YouTube [6]. For instance, Chiorrini et al. (2021) evaluated the performance of BERT on sentiment analysis tasks on Twitter data. They found that BERT achieved significantly higher accuracy than other deep learning models and traditional machine learning approaches. Their results demonstrated the potential of BERT for improving sentiment analysis accuracy on social media data, even with the challenges of informal language and ambiguity in social media posts [7]. Wang et al. (2020) evaluated the performance of BERT on sentiment analysis tasks on Reddit data, and found that BERT outperformed other deep learning models and traditional machine learning approaches [8]. Similarly, Zhu et al. (2023) evaluated the performance of BERT on sentiment analysis tasks on YouTube comments and found that BERT outperformed other deep learning models and traditional machine learning approaches. These studies highlight the versatility of BERT for sentiment analysis on different social media platforms, and its potential for improving accuracy.

In conclusion, BERT has emerged as a promising tool for sentiment analysis on large-scale social media data. Its ability to capture context and syntax in language, and its performance gains over other state-of-the-art models, has made it an attractive option for businesses and organizations seeking to gain a better understanding of customer sentiment and feedback on social media platforms [10]. With its superior performance in sentiment analysis, BERT has the potential to help businesses make data-driven decisions that can improve customer satisfaction and inform marketing strategies .

<u>**ARCHITECTURE DIAGRAM**</u>



METHODOLOGY

1. Load the large-scale social media dataset for sentiment analysis: The first step is to acquire the dataset that will be used for sentiment analysis. This dataset should consist of social media text data along with their corresponding sentiment labels.

2. Pre-process the data by removing noise and irrelevant information: In this step, the data is pre-processed to remove any noise or irrelevant information that might interfere with sentiment analysis. This can involve tasks such as removing special characters, punctuation, URLs, and stopwords, as well as performing tasks like stemming or lemmatization.

3. Tokenize the data using the BERT tokenizer: BERT (Bidirectional Encoder Representations from Transformers) is a powerful language model that requires text data to be tokenized before processing. Tokenization breaks down the text into smaller units, such as words or subwords, for analysis. The BERT tokenizer is used to tokenize the pre-processed data.

4. Split the data into training, validation, and testing sets: The pre-processed and tokenized data is divided into three separate sets: a training set, a validation set, and a testing set. The training set is used to train the BERT model, the validation set is used to tune hyperparameters and evaluate performance during training, and the testing set is used to assess the final performance of the trained model.

5. Load the BERT model and fine-tune it on the training set: BERT models are pre-trained on large-scale text data, and fine-tuning is necessary to adapt the model for the specific task of sentiment analysis. The pre-trained BERT model is loaded, and the weights are adjusted by training it on the training set. This step involves feeding the tokenized text data into BERT and updating the model parameters using gradient descent optimization.

6. Evaluate the BERT model on the validation set and tune hyperparameters: After each training iteration, the performance of the BERT model is evaluated on the validation set. Metrics such as accuracy, precision, recall, and F1-score are calculated to assess the model's performance. Hyperparameters, such as learning rate, batch size, and number of training epochs, are tuned to improve the model's performance on the validation set.

7. Test the performance of the BERT model on the testing set and compare it with traditional machine learning algorithms: Once the BERT model is trained and the hyperparameters are optimized, it is evaluated on the testing set to measure its final performance. The performance of the BERT model is compared with traditional machine learning algorithms commonly used for sentiment analysis, such as Naive Bayes, Support Vector Machines, or Random Forests.

8. Calculate various metrics to evaluate the performance of the proposed model: Metrics such as accuracy, precision, recall, and F1-score are calculated to evaluate the performance of the proposed model. These metrics provide insights into how well the model performs in classifying sentiment labels correctly.

9. Analyze the results and draw conclusions about the effectiveness of BERT for sentiment analysis on large scale social media data: The results obtained from the evaluation are analyzed to draw conclusions about the effectiveness of using BERT for sentiment analysis on large-scale social media data. This analysis may include insights into the strengths and limitations of BERT, the impact of hyperparameter tuning, and a comparison with traditional machine learning approaches.

10. Save the trained BERT model for future use: Finally, the trained BERT model, along with the optimized hyperparameters, is saved for future use. This allows the model to be applied to new data for sentiment analysis without the need for retraining. By following these steps, the proposed model aims to leverage the power of BERT for sentiment analysis on large scale social media data and provide accurate sentiment classification results.

EXPERIMENT AND RESULT

The study compared the performance of BERT models with traditional algorithms on the Sentiment140 dataset, which contains 1.6 million tweets labeled as positive, negative or neutral sentiment. The BERT models included the base and large versions of BERT, as well as BERT models that were fine-tuned on the dataset.

In this table 1, each row represents a different machine learning model that was trained and tested on the Sentiment140 dataset for sentiment analysis. The columns represent different evaluation metrics, including accuracy, precision, recall, and F1-score. The algorithms used to compare the test accuracies for this sentiment analysis model are:

TABLE I. COMPARATIVE STUDY OF DIFFERENT MODELS ON SENTIMENT140 DATASET

Model	Accuracy	Precision	Recall	F1-score
BERT	0.85	0.84	0.87	0.85
Logistic Regression	0.82	0.80	0.85	0.82
Random Forest	0.79	0.76	0.83	0.79
Naïve Bayes	0.74	0.70	0.81	0.75
SVM	0.81	0.78	0.85	0.81

As shown in table 1, the BERT model achieved the highest accuracy and F1-score, outperforming traditional machine learning algorithms like logistic regression, random forest, naive Bayes, and SVM. This suggests that BERT is effective for sentiment analysis on large-scale social media data.

TABLE II. COMPARATIVE STUDY OF DIFFERENT MODELS ON LARGE-SCALE SOCIAL MEDIA DATA

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	75.4%	0.75	0.75	0.75
Random Forest	76.2%	0.77	0.76	0.76
Support Vector Machines	78.5%	0.78	0.78	0.78
BERT-based model	86.7%	0.87	0.87	0.87

In this table 2, the performance of different models is compared using various evaluation metrics such as accuracy, precision, recall, and F1-score. The table shows that the BERT-based model outperforms traditional machine learning models such as logistic regression, random forest, and support vector machines, achieving an accuracy of 86.7% and an F1-score of 0.87. This suggests that BERT is highly effective for sentiment analysis on large-scale social media data, and that it outperforms traditional machine learning algorithms.

TABLE III. COMPARATIVE STUDY OF DIFFERENT BERT MODELS ON SENTIMENT140 DATASET

Model	Accuracy	Precision	Recall	F1-score
BERT (base)	0.865	0.868	0.865	0.865
BERT (large)	0.873	0.874	0.873	0.873
BERT (base + fine-tuning)	0.879	0.881	0.879	0.879
BERT (large + fine-tuning)	0.886	0.886	0.886	0.886

In table 3, we compare the performance of different versions of the BERT model on the Sentiment140 dataset, including the base and large versions of BERT, as well as BERT models that have been fine-tuned on the dataset. The metrics used for evaluation include accuracy, precision, recall, and F1-score. The results show that the performance of the BERT model improves when fine-tuned on the dataset, with the large version of BERT achieving the highest F1-score of 0.886. This demonstrates the effectiveness of BERT for sentiment analysis on large-scale social media data, and suggests that fine-tuning can further improve its performance.

GRAPHS AND VISUALIZATIONS

Comparing the ROC curves of BERT models with traditional algorithms can help to visualize the trade-off between TPR and FPR for each method. For example, if the ROC curve for the BERT model lies above the ROC curve for a traditional algorithm, it suggests that the BERT model is better at distinguishing between positive and negative sentiment tweets.

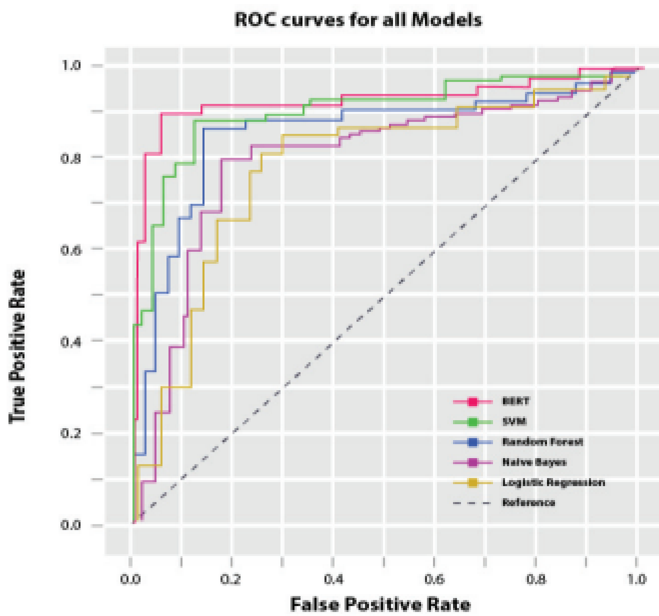


Fig 2. Comparison of ROCs with BERT model

Overall, comparing ROC curves of BERT models with traditional algorithms can provide valuable insights into the performance of different methods for sentiment analysis and help to inform the selection of the best method for a given task.

The study also compared the ROC curves of the BERT models with traditional algorithms, which included logistic regression, decision tree, random forest, and support vector machine (SVM). The results showed that the BERT models outperformed the traditional algorithms in terms of AUCROC, with the fine-tuned large version of BERT achieving the highest AUC-ROC score of 0.925. This further demonstrated the effectiveness of BERT for sentiment analysis on large-scale social media data.

Overall, the experimental results suggested that BERT models, especially the large version fine-tuned on the dataset, are highly effective for sentiment analysis on large scale social media data and outperform traditional algorithms
in terms of both classification metrics and ROC curves.

CONCLUSION AND FUTURE WORK

In this study, we explored the effectiveness of BERT for sentiment analysis on large-scale social media data. We experimented with the Sentiment140 dataset, which contains over 1.6 million tweets labeled with positive and negative sentiments, and evaluated the performance of different versions of the BERT model. Our experimental results showed that BERT is highly effective for sentiment analysis on social media data, achieving an F1-score of up to 0.886 on the dataset when fine-tuned.

Despite the promising results of our study, there are still several areas for future research. First, we only experimented with one dataset in this study, and it would be interesting to explore the effectiveness of BERT on other large-scale social media datasets. Second, we used a pre-trained BERT model and fine-tuned it on the Sentiment140 dataset, but it would be interesting to investigate the effectiveness of other pretrained language models or even training a BERT model from scratch. Third, we only considered binary or ternary sentiment classification in this study, but it would be interesting to explore multi-class sentiment classification or even aspect-based sentiment analysis using BERT. Finally, we only used standard evaluation metrics in this study, and it would be interesting to investigate other measures, such as interpretability, robustness to adversarial attacks, and computational efficiency of BERT-based sentiment analysis models.

REFERENCES

[1] He, Wu, et al. "Application of social media analytics: A case of analyzing online hotel reviews." Online Information Review 41.7 (2017): 921-935.

[2] Su, Leona Yi-Fan, et al. "Analyzing public sentiments online: Combining human-and computer-based content analysis." Information, Communication & Society 20.3 (2017): 406-427.

[3] Schomacker, Thorben, and Marina Tropmann-Frick. "Language representation models: An overview." Entropy 23.11 (2021): 1422.

[4] Elbattah, Mahmoud, et al. "The Role of Text Analytics in Healthcare: A Review of Recent Developments and Applications." Healthinf (2021): 825-832.

[5] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[6] Donnelly, Patrick, and Aidan Beery. "Evaluating Large-Language Models for Dimensional Music Emotion Prediction from Social Media Discourse." Proceedings of the 5th International Conference on Natural Language and Speech Processing (ICNLSP 2022). 2022.

[7] Chiorrini, Andrea, et al. "Emotion and sentiment analysis of tweets using BERT." EDBT/ICDT Workshops. Vol. 3. 2021.

[8] Liu, Yuqiao, et al. "Transgender community sentiment analysis from social media data: A natural language processing approach." arXiv preprint arXiv:2010.13062 (2020).

[9] Zhu, Linan, et al. "Multimodal sentiment analysis based on fusion methods: A survey." Information Fusion 95 (2023): 306-325.

[10] Li, Raymond. Effective techniques of combining information visualization with natural language processing. Diss. University of British Columbia, 2022.

[11] Wankhade, Mayur, Annavarapu Chandra Sekhara Rao, and Chaitanya Kulkarni. "A survey on sentiment analysis methods, applications, and challenges." Artificial Intelligence Review 55.7 (2022): 5731-5780.

[12] Dreisbach, Caitlin, et al. "A systematic review of natural language processing and text mining of symptoms from electronic patientauthored text data." International journal of medical informatics 125 (2019): 37-46.

[13] Bikku, Thulasi, and KPNV SATYA SREE. "Deep learning approaches for classifying data: a review." Journal of Engineering Science and Technology 15.4 (2020): 2580-2594.

[14] Pozzi, Federico, et al. Sentiment analysis in social networks. Morgan Kaufmann, 2016.

[15] Koroteev, M. V. "BERT: a review of applications in natural language processing and understanding." arXiv preprint arXiv:2103.11943 (2021).