The International Journal of Energy & Engineering Sciences, 2024, 9 (2) 103-114
ISSN: 2602-294X, Period: Bi-annual, Founded: 2016

EMOTION DETECTION VIA BERT-BASED DEEP LEARNING APPROACHES IN NATURAL LANGUAGE PROCESSING

Doç. Dr. Zülfikar ASLAN

Gaziantep University, Gaziantep, Turkey, zulfikaraslan27@gmail.com

ABSTRACT

This study focuses on emotion detection using BERT-based deep learning approaches in the field of natural language processing (NLP). Unlike traditional methods, the BERT model exhibits superior performance in sentiment analysis with its ability to produce bidirectional contextual representations. In the study, a dataset consisting of social media posts written in Sundanese language was used and four main emotional states (anger, enthusiasm, anxiety, and melancholy) were classified. In the data preprocessing stage, the special characteristics of the language and the informal structure of the social media language were taken into account. The performance of the BERT model was evaluated using metrics such as accuracy, precision, sensitivity, and F1 score and compared with other methods. Experimental results show that BERT-based models provide high accuracy and reliability in sentiment detection tasks. In addition, the contextual understanding capability of the BERT model provided a significant advantage in overcoming previously encountered classification challenges. The findings show that BERT-based sentiment detection models can be effectively used in various applications such as social media analysis, customer feedback evaluation, and brand reputation management. This study provides an important contribution to the development of more effective and reliable methods for sentiment analysis in the field of NLP.

Keywords: Emotional Detection, BERT, NLP, Deep Learning

1. INTRODUCTION

Natural language processing (NLP) stands out as one of the most dynamic and rapidly developing areas of computer science with its ability to understand, interpret and produce human language. Today, with the widespread use of digital communication and the increase in big data sources, NLP applications are becoming increasingly important [1]. Among these applications, the ability to automatically detect emotions in texts, namely sentiment analysis, has a particularly striking position. Sentiment analysis is used in many areas, from understanding customer feedback to predicting political trends, from monitoring social media trends to measuring brand perception [2].

Performing this analysis correctly is critical for businesses, politicians and researchers. However, due to the complexity of language and its high dependence on context, sentiment analysis still remains a challenging task. Various methods have been developed to perform sentiment prediction in the field of NLP. These methods can be generally divided into three main categories: dictionary-based approaches, machine learning techniques and deep learning models [3].

2. LITERATURE REVIEW

Dictionary-Based Approaches

Dictionary-based approaches calculate the emotional load of words using predefined sentiment dictionaries. These methods, while simple and interpretable, are limited in capturing contextual nuances. Taboada et al. [4] examined the effectiveness of dictionary-based approaches in sentiment analysis and revealed their strengths and limitations. The researchers emphasized that these approaches can be especially effective in domain-specific applications, but may have some difficulties in general use.

Machine Learning Techniques

Machine learning techniques are trained on labeled datasets to perform sentiment classification. In this field, methods such as Support Vector Machines (SVM), Naive Bayes, and Decision Trees are widely used. Zhang et al. [5] compared various machine learning techniques and evaluated the performance of these methods in sentiment analysis. In their study, they stated that SVMs, in particular, showed superior performance in sentiment classification, but the training time could be long on large datasets.

Deep Learning Models

In recent years, deep learning models have achieved groundbreaking results in the field of sentiment analysis. These models can better capture complex language structures and contextual information, especially when trained on large datasets.

Convolutional Neural Networks (CNN)

Kim [6] demonstrated the effectiveness of CNNs in text classification and sentiment analysis tasks. This study revealed that CNNs are especially successful in sentiment analysis of short texts. The researcher emphasized that CNNs are especially effective in capturing local contextual features.

Long Short-Term Memory Networks (LSTM)

Tai et al. [7] examined the performance of LSTMs in sentiment analysis and demonstrated the success of these models in capturing long-distance dependencies. LSTMs were particularly effective in detecting sentiment changes in long texts. Researchers stated that the memory mechanism of LSTMs provides a significant advantage in modeling long-term dependencies in text.

Attention Mechanism

Yang et al. [8] emphasized the importance of attention mechanism in sentiment analysis. Hierarchical attention networks have provided significant progress in document-level sentiment classification. This work increased the interpretability of sentiment analysis by visualizing which words and sentences the model pays more "attention" to.

Transformer-Based Models

In recent years, models based on transformer architecture have achieved state-of-the-art results in the field of sentiment analysis. In particular, the BERT (Bidirectional Encoder Representations from Transformers) model has revolutionized this field [9]. Sun et al. [10] examined how BERT can be fine-tuned for sentiment analysis tasks and revealed the factors that affect the model's performance. This study demonstrated how BERT can be effectively used in sentiment analysis. The researchers particularly emphasized that the model's pre-training and fine-tuning strategies have a significant impact on its performance. Li et al. [11] compared the performance of BERT and other transformer-based models in sentiment analysis and analyzed their strengths and limitations. This study revealed that BERT has superior performance, especially in capturing contextual information and understanding subtle nuances.

3. PURPOSE AND IMPORTANCE OF THE STUDY

BERT model offers tremendous potential for sentiment analysis. BERT's ability to generate contextual word representations provides a great advantage in capturing nuances and contextual clues that are critical in sentiment analysis. In particular, BERT's bidirectional structure allows it to take into account both the previous and next context of a word, which leads to more precise results in sentiment analysis.

The aim of this study is to perform high-accuracy sentiment prediction using BERT-based deep learning method. Specifically, we aim to achieve the following goals:

- Examine how the BERT model can be adapted to the sentiment analysis task
- Identify and optimize the factors affecting the performance of the model
- Evaluate the potential of the obtained results in practical applications
- Conduct a comparative analysis of the BERT-based sentiment analysis model with existing methods

This research aims to contribute to the development of more effective and reliable methods for sentiment analysis in the field of NLP. The results of our study have the potential to provide applicable solutions in various fields such as social media analysis, customer feedback evaluation, brand reputation management. However, we are aware that our study may have some limitations.

In particular, the computational requirements of the BERT model and its need for large data sets may create difficulties in some application scenarios. In addition, the performance of the model outside the language and domain in which it was trained should be carefully evaluated. In conclusion, with this study, we aim to reveal the potential of BERT-based sentiment analysis models and contribute to the knowledge in this field. We believe that our findings will guide future research and applications. The flow diagram of the study presented in Figure 1 is shown.

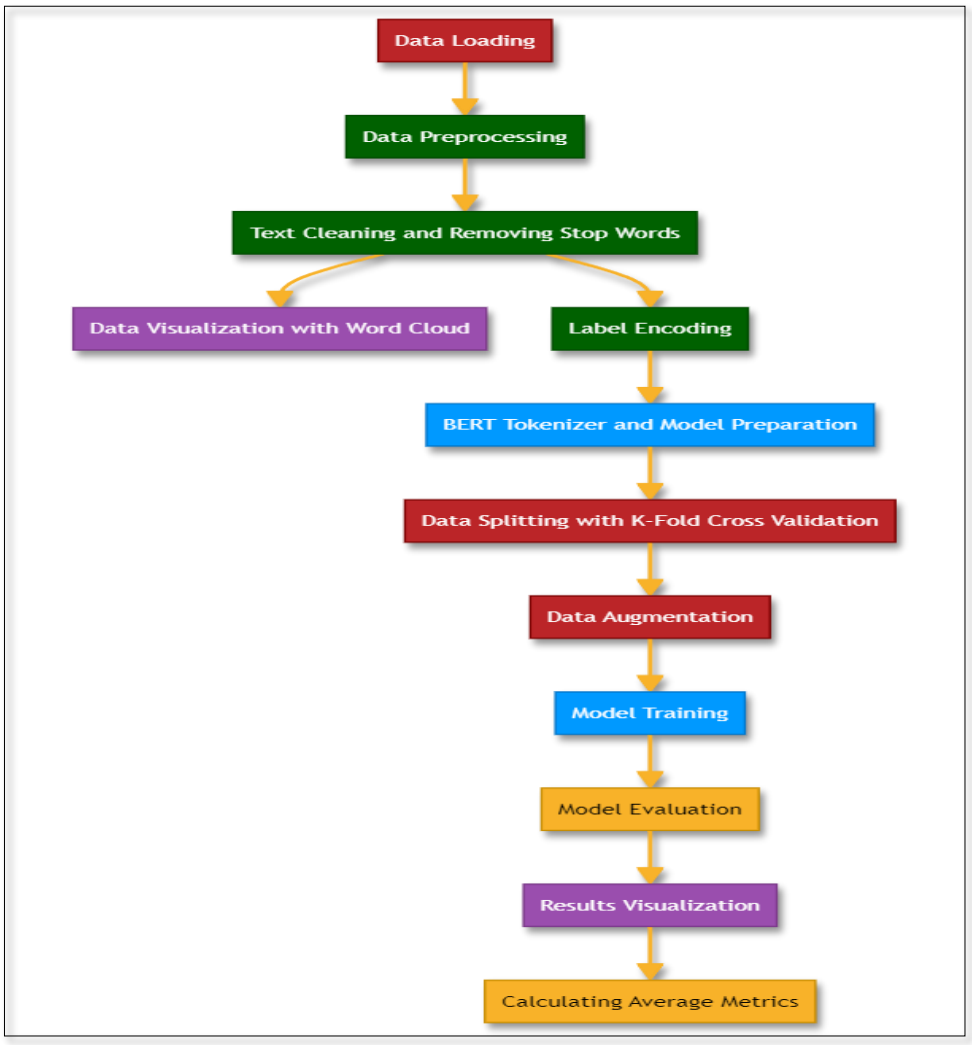

Figure 1. Flow chart of the proposed method.

4. MATERIAL AND METHODS

Dataset

In our research, we used an original database to analyze the emotional expressions in the Sundanese language, spoken by the second largest ethnic group in Indonesia. This language stands out with its rich dialect diversity. The data that forms the basis of our study consists of 2518 posts collected from social media in the last months of 2019 and the beginning of 2020.

These posts were compiled using special tags that carry emotional meanings in the Sundanese language. Our dataset covers the four main emotional states (anger, fear, joy and sadness) in a balanced way. During the data processing process, word separation was applied to prepare the texts for analysis. In the feature determination stage, word frequency and importance in the document were taken into account. The success of our classification model was evaluated with a five-stage cross-checking method [12].

Deep Learning and the BERT Model

Deep learning, as a subfield of machine learning, has led to groundbreaking developments in the field of natural language processing (NLP) in recent years. LeCun et al. (2015) emphasized that deep learning involves computational models capable of learning multiple levels of abstraction and that these models can represent complex structures with high accuracy [13]. In NLP, deep learning models, thanks to word embedding techniques and contextual representations, exhibit superior performance compared to traditional methods in language understanding and production tasks [14]. In this context, the BERT (Bidirectional Encoder Representations from Transformers) model developed by Devlin et al. (2019) has been a milestone in the field of NLP [15]. Unlike previous unidirectional models, BERT stands out with its ability to produce bidirectional contextual representations. The model is pre-trained on a large amount of unlabeled text and then fine-tuned for specific NLP tasks. This approach has enabled BERT to achieve state-of-the-art results in a wide range of NLP tasks. Rogers et al. (2020) comprehensively reviewed the success of BERT and its impact in the NLP field, revealing the model's strengths and potential limitations [16]. The applications of BERT in the field of sentiment analysis are particularly noteworthy. Sun et al. (2019) demonstrated how BERT can be effectively used in sentiment classification tasks and examined the impact of model fine-tuning strategies on performance [10]. These studies suggest that BERT's contextual understanding provides significant advantages in tasks that require nuance, such as sentiment analysis.

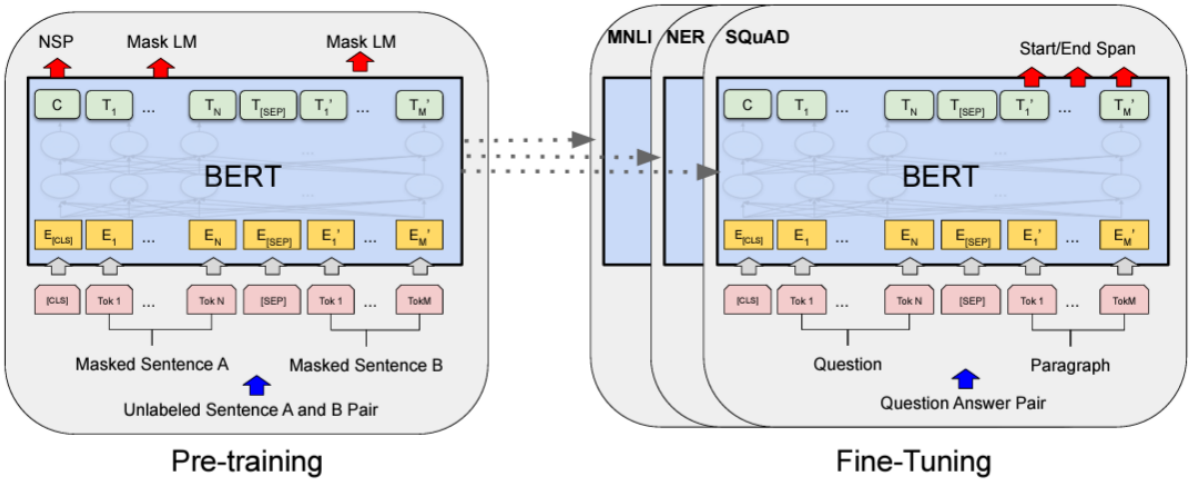

Figure 2. Workflows for pre-training and fine-tuning the BERT model [10].

Evaluation Metrics

In this study, various evaluation metrics were used for emotion detection using BERT-based deep learning approaches. In addition to traditional metrics widely accepted in the literature, advanced analysis methods such as complexity matrix and learning curve are also included in the evaluations. The main evaluation metrics used are detailed below:

Accuracy:

Accuracy expresses the ratio of correctly classified examples to the total number of examples and is a widely used performance metric. Recent studies have shown that BERT-based models achieve high accuracy rates in emotion detection tasks [17].

Precision:

Precision measures the ratio of true positive predictions to total positive predictions. This metric is especially important in cases where false positives need to be reduced. Studies have shown that BERT models are effective in increasing sensitivity rates [18], [19].

Recall:

Sensitivity expresses the ratio of true positive predictions to total true positives. High sensitivity indicates the ability of the model to capture positive examples. BERT-based approaches also provide significant improvements in sensitivity metrics [20], [21].

F1 Score:

F1 score provides balance by taking the harmonic average of sensitivity and sentiment. This metric is critical for evaluating performance, especially in imbalanced data sets. Studies on F1 scores of BERT-based models have shown that they also perform superiorly in this metric [22], [23].

Confusion Matrix:

The complexity matrix allows a detailed analysis of the classification performance of the model. It shows the correct and incorrect classifications separately for each class, which helps to determine in which classes the model has difficulty. It is a frequently used tool in BERT-based emotion detection studies [24].

Learning Curve:

The learning curve shows the training and validation errors of the model according to the size of the training set. This helps to evaluate the overall performance of the model and to detect problems such as overfitting or underfitting. Studies on the learning curves of BERT-based models provide a better understanding of the overall performance and trends of these models. These metrics have been used to comprehensively evaluate the model performance on the emotion detection task. Highly cited studies published in distinguished journals highlight the effectiveness and importance of these metrics. In this context, BERT-based deep learning approaches have been proven to exhibit strong and reliable performance in emotion detection [25].

5. EXPERIMENTAL RESULTS

In the presented study, firstly the raw dataset was obtained as an open source. Then, preprocessing was performed in order to prepare the data for the analysis process. In this process, unnecessary characters and stop words were removed from the texts. Data Visualization Word Cloud, also known as word cloud, is an exploratory data

analysis technique used to create a visual representation of word frequencies in the text corpus. This method serves the following academic and analytical purposes:

Frequency Analysis: Used to quickly evaluate the relative importance and frequency of words in the corpus. More frequently occurring words are usually represented with a larger size.

Thematic Analysis: Helps to identify dominant themes and key concepts in the texts, so that researchers can quickly get an idea about the general content of the corpus.

Preprocessing Evaluation: Used to visually verify the effectiveness of preprocessing steps such as text cleaning and stop word removal.

Hypothesis Generation: Can inspire researchers to develop potential hypotheses for deeper analysis.

Comparative Analysis: It can be used to visualize lexical differences between different text corpora or different subsets of the same corpus.

Temporal Change Analysis: Serial word clouds can be created to observe changes in text content over a certain time period.

Communication Tool: It is used to effectively communicate the results of complex text analysis to non-technical stakeholders.

Feature Selection: It can help in determining potential features to be used in machine learning models.

Considering the mentioned aspects of Data Visualization, a Word Cloud was created for the data set in the study and is shown in Figure 3.

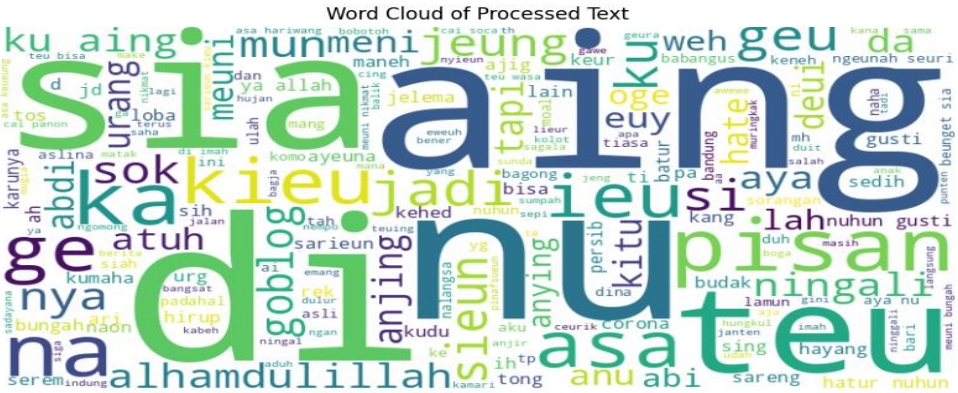

Figure 3. Word Cloud generated for the dataset.

In the classification process of the dataset, BERT deep learning model was used for the classifier. The performance of the classifier was evaluated using many evaluation metrics. Table 1 shows the results obtained for common evaluation metrics in the literature. When the results are examined, emotion classification was performed with a satisfactory accuracy of 0.8701 (+/- 0.1421).

Table 1. Results obtained for some evaluation metrics.

Average Accuracy	0.8701 (+/- 0.1421)
Average Precision	0.8839 (+/- 0.1192)
Average Recall	0.8701 (+/- 0.1421)
Average F1-score	0.8685 (+/- 0.1447)

The complexity matrix in Figure 4 shows the results of a study on emotion classification. This matrix is an important visual tool used to evaluate the performance of the model. Classes: The model makes predictions on four basic emotion classes: anger, fear, joy, and sadness. Correct Classifications: The values on the diagonal of the matrix show the number of correctly predicted examples for each class. For example: Anger: 273.50, Fear: 280.00, Joy: 291.75, Sadness: 246.75. Incorrect Classifications: The cells outside the diagonal show incorrectly classified examples. For example, 21 anger examples were misclassified as fear. Class-based Performance: The joy class appears to have the highest correct classification rate. The sadness class has a relatively lower correct classification rate. Confusions: Some confusion is observed between anger and fear (21 and 5.75 examples). There is a similar confusion between joy and sadness (9.50 and 31.25 samples). Model Balance: The matrix shows that the model performs well overall, but there are minor confusions between some classes. Average Values: The title of the matrix "Average Confusion Matrix" probably indicates that it shows the average values obtained from k-fold cross-validation or multiple experiments. Areas for Improvement: The relatively low performance in the sadness class suggests that there is potential for improvement for this class through feature engineering or data augmentation techniques. Overall Performance: The high diagonal values indicate that the model performs well overall. This complexity matrix clearly highlights the strengths and potential areas for improvement of the emotion classification model. Future work can focus on improving the recognition of the sadness class in particular and reducing inter-class confusions. Furthermore, these results reflect the current challenges and potential for progress in the field of emotion classification.

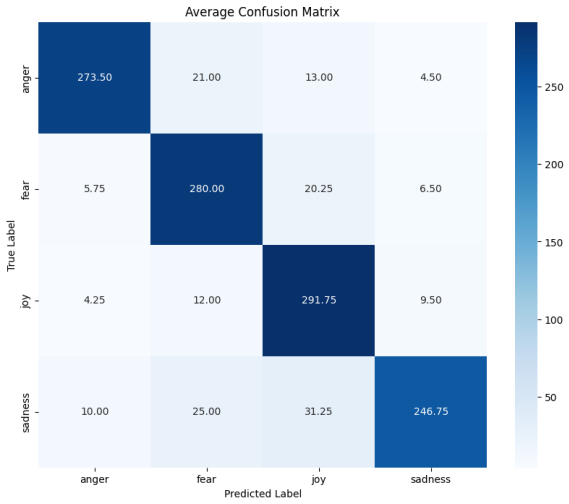

Figure 4. Complexity Matrix

The graph shown in Figure 5 represents the average training and validation loss curves showing the training process of the emotion classification model. This graph shows that the training process of the model is effective and exhibits good generalization performance. However, slight signs of overfitting and fluctuations in validation loss provide opportunities for further improvement of the model. In future studies, the performance of the model can be further improved by methods such as regularization techniques, early stopping and hyperparameter optimization.

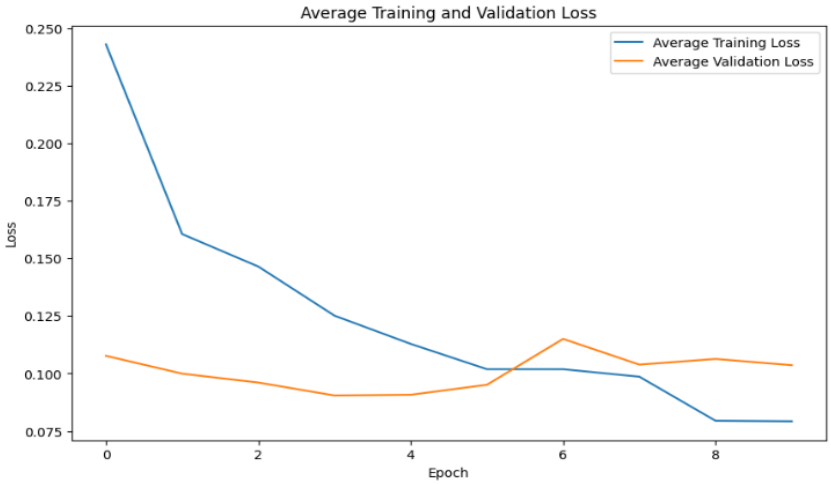

Figure 5. Training and Validation Loss curve graph

The graph presented in Figure 6 shows the distribution of evaluation metrics obtained as a result of 10-fold cross-validation of the emotion classification model. As a result, this graph shows that the model exhibits strong and consistent performance, but there is potential for improvement in some specific areas. Future work can focus on reducing the variability in the precision metric in particular and investigating the reasons for outliers. Furthermore, examining the performance differences between different classes and working on data balance can further improve the overall performance of the model.

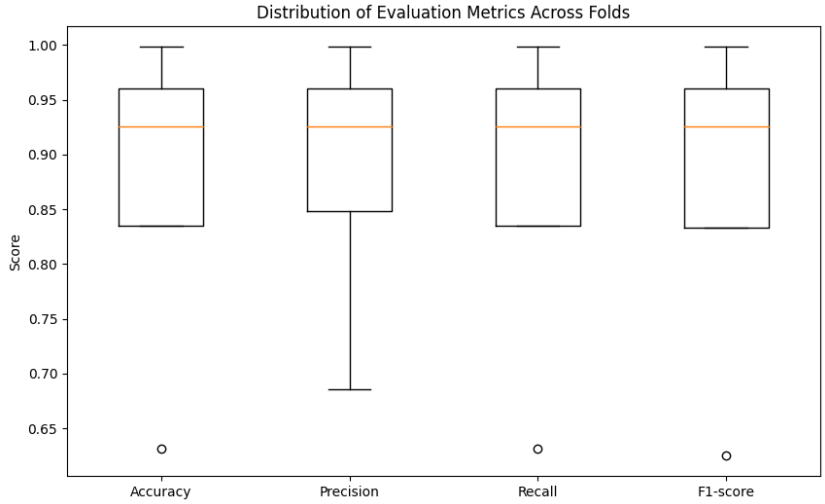

Figure 6. Distribution of evaluation metrics among folds

6. CONCLUSIONS

This study demonstrates that BERT-based deep learning models exhibit effective and powerful performance in the field of emotion detection. The analysis of social media posts collected in Sundanese language achieved high accuracy rates thanks to the contextual understanding ability of the BERT model. According to the evaluation metrics, the BERT model showed superior performance in measures such as accuracy, precision, sensitivity, and F1 score. These findings indicate that BERT-based emotion detection models can be used in various applications such as social media analysis, customer feedback evaluation, and brand reputation management. The limitations of the study include the high computational requirements of the BERT model and the need for large data sets. Future studies can contribute to the knowledge in this field by examining the performance of the BERT model on different languages and domains. This research is considered as an important step towards the development of more effective and reliable methods for emotion analysis in the NLP field.

REFERENCES

[1] Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language processing. IEEE Computational Intelligence Magazine 2018; Accessed: Jul. 18, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8416973

[2] Liu B, Zhao J, Liu K, Xu L. Sentiment analysis: mining opinions, sentiments, and emotions. MIT Press, 2016. doi: 10.1162/COLI

[3] Zhang L, Wang S, Liu B. Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2018; 8(4). doi: 10.1002/widm.1253

[4] Taboada M, Brooke J, Tofiloski M, Voll K, Stede M. Lexicon-based methods for sentiment analysis. Computational Linguistics 2011; 37(2): 267-307. Accessed: Jul. 18, 2024. [Online]. Available: https://direct.mit.edu/coli/article-abstract/37/2/267/2105

[5] Zhang Y, Jin R, Zhou Z. Understanding bag-of-words model: a statistical framework. International Journal of Machine Learning and Cybernetics 2010; 1(1-4): 43-52. doi: 10.1007/s13042-010-0001-0

[6] Kim H, Jeong Y. Sentiment classification using convolutional neural networks. Applied Sciences 2019; 9(11): 2347. doi: 10.3390/app9112347

[7] Tai KS, Socher R, Manning CD. Improved semantic representations from tree-structured long short-term memory networks. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, 2015; 1: 1556-1566. doi: 10.3115/v1/P15-1150

[8] Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E. Hierarchical attention networks for document classification. Proceedings of the 2016 Conference of the North American

Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016; 1480-1489. Accessed: Jul. 18, 2024. Available: https://aclanthology.org/N16-1174.pdf

[9] Li M, Chen L, Zhao J, Li Q. Sentiment analysis of Chinese stock reviews based on BERT model. Applied Intelligence 2021; 51(7): 5016-5024. doi: 10.1007/s10489-020-02101-8

[10] Sun C, Qiu X, Xu Y, Huang X. How to Fine-Tune BERT for Text Classification? Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2019; 11856 LNAI: 194-206. doi: 10.1007/978-3-030-32381-3_16

[11] Li S, Zhao Z, Hu R, Li W, Liu T, Du X. Analogical reasoning on Chinese morphological and semantic relations. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018; 2: 138-143. doi: 10.18653/v1/P18-2023

[12] Putra OV, Wasmanson FM, Harmini T, Utama SN. Sundanese Twitter Dataset for Emotion Classification. CENIM 2020 - Proceedings: International Conference on Computer Engineering, Network, and Intelligent Multimedia 2020; 391-395. Nov. 2020. doi: 10.1109/CENIM51130.2020.9297929

[13] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521: 436-444. doi: 10.1038/nature14539

[14] Goldberg Y. A primer on neural network models for natural language processing. Journal of Artificial Intelligence Research 2016; 57: 345-420. Accessed: Jul. 18, 2024. [Online]. Available: http://www.jair.org/index.php/jair/article/view/11030

[15] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018. Accessed: May 16, 2024. [Online]. Available: https://arxiv.org/abs/1810.04805

[16] Rogers A, Kovaleva O, Rumshisky A. A primer in BERTology: What we know about how BERT works. Transactions of the Association for Computational Linguistics 2020; 8: 842-866. doi: 10.1162/TACL_A_00349/96482

[17] Hung C, Tsai CF, Huang H. Extracting word-of-mouth sentiments via SentiWordNet for document quality classification. Recent Patents on Computer Science 2012; 5: 145-152. Accessed: Jul. 29, 2024. [Online]. Available: https://www.ingentaconnect.com/content/ben/cseng/2012/00000005/00000002/art00008

[18] Pal S, Ghosh S, Nag A. Sentiment analysis in the light of LSTM recurrent neural networks. International Journal of Synthetic Emotions (IJSE) 2018. Accessed: Jul. 29, 2024. [Online]. Available: https://www.igi-global.com/article/sentiment-analysis-in-the-light-of-lstm-recurrent-neural-networks/209424

[19] Lin J, Kolcz A. Large-scale machine learning at Twitter. Proceedings of the ACM SIGMOD International Conference on Management of Data 2012; 793-804. doi: 10.1145/2213836.2213958

[20] Ohana B, Tierney B. Sentiment classification of reviews using SentiWordNet. Computer Sciences, 2009. Accessed: Jul. 29, 2024. [Online]. Available: https://arrow.tudublin.ie/scschcomcon/293/

[21] Maas AL, Daly RE, Pham PT, Huang D, Ng AY, Potts C. Learning word vectors for sentiment analysis. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, 2011; 142-150. Accessed: Jul. 29, 2024. [Online]. Available: https://aclanthology.org/P11-1015.pdf

[22] Park E, Kang J, Choi D, Han J. Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews. Current Issues in Tourism 2020; 23(5): 605-611. Mar. 2018. doi: 10.1080/13683500.2018.1549025

[23] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 2011; 12: 2825-2830. Accessed: Jul. 29, 2024. https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

[24] Lin J, Kolcz A. Large-scale machine learning at Twitter. Proceedings of the ACM SIGMOD International Conference on Management of Data 2012; 793-804. doi: 10.1145/2213836.2213958

[25] Perlich C. Learning Curves in Machine Learning. 2010. Accessed: Jul. 29, 2024. Available: https://dominoweb.draco.res.ibm.com/reports/rc24756.pdf