International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Contextual Understanding and Reprocessing in Sarcasm Detection: A Study of BERT Vs Logistic Regression

Dharmendra Bhadauria[1], Arun Bhardwaj[2], Kalyani Singh[3], S M Aqdas Hashmi[4]

[1,2,3]School of Computer Science and Application, Galgotias University, Greater Noida, UP

Abstract:
Detecting sarcasm in natural language presents distinct challenges, primarily due to its reliance on contextual clues and the use of subtle, contradictory expressions. This research examines how the BERT model and Logistic Regression (LR) perform in detecting sarcasm, emphasizing the importance of contextual understanding and various preprocessing methods. Using a social media dataset, we evaluate both models on sentiment analysis tasks after applying different preprocessing methods including tokenization, elimination of stop words, and noise reduction. We assess the models' performance using evaluation metrics like F1 scores and confusion matrices to determine their accuracy in recognizing sarcastic expressions. Our findings suggest that although BERT's bidirectional architecture provides a deeper understanding of linguistic context, it does not always outperform the simpler LR model when noise reduction preprocessing is applied. This indicates that preprocessing techniques can affect model performance in varying ways, and selecting NLP models should align with the unique needs of the sentiment analysis objective. This study provides meaningful perspectives on the role of contextual models and preprocessing methods in improving the precision of sarcasm detection.

Keywords: Sarcasm Detection, Sentiment Analysis, BERT, Logistic Regression, Tokenization, Preprocessing Techniques

1. INTRODUCTION

Identifying sarcasm in text presents a significant challenge in natural language processing (NLP) because it depends heavily on contextual cues, tone, and subtle contradictions that differentiate sarcastic remarks from literal statements. As sarcasm increasingly impacts sentiment analysis, particularly in the context of social media, its detection has become more critical. Misinterpreting sarcasm can lead to inaccurate sentiment readings, which is problematic for applications like sentiment analysis, customer feedback interpretation, and social media monitoring. Although conventional machine learning models such as Logistic Regression (LR) are valued for their straightforwardness and effectiveness in sentiment analysis and interpretability, they often fall short when trying to capture complex contextual cues, especially for sarcasm. On the other hand, Google's BERT (Bidirectional Encoder Representations from Transformers) has transformed language comprehension through its bidirectional transformer framework, allowing it to analyze word context from both preceding and succeeding directions ,thus offering a more nuanced understanding, which may be beneficial for sarcasm detection. This research compares the performance of BERT and Logistic Regression in sarcasm detection, focusing on how preprocessing techniques impact their effectiveness. The study investigates the roles of tokenization, stop word removal, and noise reduction in refining data for model input. We evaluate both models using a sarcasm-focused dataset sourced from social media, employing evaluation metrics such as F1 scores and confusion matrices to assess their performance. The findings are expected to provide valuable insights into the impact of contextual understanding and preprocessing techniques on sarcasm detection and will help guide future choices in model and preprocessing.

2. RELATED WORK

Sarcasm detection has gained significant attention in sentiment analysis, particularly as the volume of online content increases. Traditional models like Logistic Regression and Support Vector Machines have been applied to sentiment tasks but often fail to grasp the subtleties necessary for sarcasm detection. Recent studies emphasize the potential of deep learning models, particularly BERT, in capturing the rich context of language, making it more adept at sarcasm detection compared to traditional methods. Hoang et al. demonstrated that BERT's ability to understand contextual nuances far outperforms traditional models. Additionally, the importance of preprocessing techniques—such as tokenization and noise reduction—has been highlighted in improving the accuracy of models for complex tasks like sarcasm detection. Wankhade et al. found that appropriate preprocessing enhances model performance, especially in tasks involving challenging expressions such as sarcasm. This study builds on these insights, comparing the capabilities of BERT and Logistic Regression for sarcasm detection.

3. METHODOLOGY

This study employs a structured approach to compare the performance of BERT and Logistic Regression in detecting sarcasm. The methodology consists of the following stages:

1. **Data Collection:** A dataset containing labelled text samples is gathered from publicly available sources, such as Kaggle, focusing on instances of sarcastic and non-sarcastic comments.
2. **Data Preprocessing**: The collected data undergoes several preprocessing techniques to enhance model performance. This includes:
   Tokenization: The text is broken down into individual tokens (words or phrases) for analysis.
   Noise Reduction: Unnecessary elements, such as HTML tags, URLs, and special characters, are removed from the text.
   Stop Word Removal: Commonly used words (e.g., "the," "and," "is") that do not contribute to the sentiment are eliminated.
3. **Model Selection:** Two models are selected for comparison:
   Logistic Regression (LR): A traditional machine learning model used for binary classification.
   BERT: A pre-trained deep learning model that captures context and semantic meaning from text.
4. **Model Training and Evaluation:** Both models undergo training using the preprocessed dataset, and their performance is evaluated using key metrics such as F1 scores and confusion matrices to measure their accuracy in detecting sarcasm. The dataset is divided into training and testing subsets, with 80% designated for training and the remaining 20% reserved for testing. Model Comparison: The F1 scores and confusion matrices from each model are analyzed and compared to highlight their respective strengths and limitations in sarcasm detection. This analysis seeks to identify which model

performs better at detecting sarcastic expressions in text.

4. ABOUT THE MODEL

This research investigates the effectiveness of two different models, BERT (Bidirectional Encoder Representations from Transformers) and Logistic Regression are analyzed for their distinct strengths in detecting sarcasm in textual data.

BERT is a deep learning-based model built on a transformer architecture that processes words in context, considering both the preceding and succeeding words within a sentence. This dual-directional analysis allows BERT to capture subtle linguistic nuances, making it highly well-suited for tasks that demand contextual comprehension, including sarcasm detection. The model undergoes a two-step training process: it first learns general language patterns from vast amounts of unlabeled data and is then fine-tuned to specialize in specific tasks. BERT has consistently demonstrated superior performance on various natural language processing benchmarks, showcasing its ability to grasp complex language features and improve understanding of human expression (Wankhade et al., 2024).

On the other hand, Logistic Regression is a simpler, well-established statistical model used for binary classification tasks. It operates by predicting the probability of an input being part of a specific category, using a logistic function applied to a linear combination of the input attributes. While not as advanced as BERT, Logistic Regression offers the advantage of being fast, computationally efficient, and highly interpretable, making it particularly useful when transparency is required. Despite its simplicity, Logistic Regression can yield competitive results in sentiment analysis, especially when paired with proper preprocessing techniques. This study compares the performance of BERT and Logistic Regression in sarcasm detection, acknowledging that sarcasm presents unique challenges that can complicate traditional sentiment analysis methods. By integrating BERT's advanced contextual capabilities with Logistic Regression's straightforward approach, this research aims to provide insights into effective methods for sarcasm detection in text.

5. PREPROCESSING TECHNIQUE

Preprocessing plays a vital role in natural language processing (NLP), improving the quality of input data and significantly influencing the effectiveness of machine learning models. In this study, we focus on three essential preprocessing techniques: tokenization, noise removal, and stop word removal. These techniques are crucial for preparing textual data for sentiment analysis tasks like sarcasm detection.

1. Tokenization

Tokenization is the process of splitting a text into smaller elements known as tokens, which can be words, phrases, or even sub words, based on the tokenization approach. The objective of tokenization is to convert raw text into a structured format that machine learning models can easily process. By fragmenting the text into digestible units, tokenization highlights significant elements, forming the groundwork for deeper analysis.

Accurate tokenization in sarcasm detection helps models understand nuanced contextual details and the connections between words. For example, BERT, a widely used transformer model, utilizes a sub word tokenization method called Word Piece. This approach allows BERT to manage uncommon or novel words by segmenting them into smaller components, thereby maintaining a manageable vocabulary size. This capability is particularly important for detecting sarcasm, where the context and word choices are crucial. Studies have shown that robust tokenization improves the performance of models in sentiment analysis, particularly when complex expressions such as sarcasm are involved (Wankhade et al., 2024).

2. Noise Removal

Noise removal involves the elimination of irrelevant or unnecessary elements from the text, such as special characters, numbers, punctuation marks, and formatting artifacts. These extraneous elements do not add value to the text's meaning and may disrupt the model's focus on the core content. The goal of noise removal is to clean the text so that the machine learning algorithms can better identify the underlying patterns relevant to sentiment analysis.

In the case of sarcasm detection, noise removal is especially important because irrelevant data can distract the model from identifying key sentiment indicators. For example, special characters or stray symbols might obscure the model's ability to detect sarcasm, which often relies on subtle cues such as tone, word choice, and sentence structure. By removing noise, we improve the clarity and relevance of the data, making it easier for algorithms like BERT and Logistic Regression to detect sarcasm accurately. Previous research has emphasized that cleaning the data by removing noise is an essential step to enhance model performance in sentiment analysis (Wankhade et al., 2024).

3. Stop Word Removal

Stop words are common words in a language, like "the," "is," "in," and "and," typically regarded as having minimal semantic importance and are usually excluded during text preprocessing. The main purpose of removing stop words is to decrease the dataset's dimensionality, allowing the model to concentrate on more significant and insightful words. By eliminating stop words, we make the dataset more manageable and help the model focus on content that plays a more significant role in sentiment analysis. This helps avoid unnecessary distractions from common words that don't carry much importance in determining the sentiment of the text. However, in the context of sarcasm detection, the role of stop words becomes more complex. Sarcasm often relies on subtle language cues and the interplay between specific words. In certain sarcastic expressions, the inclusion or omission of specific stop words can alter the tone or interpretation of the sentence. For instance, a sentence like "Oh, sure, that sounds like a great idea" may rely on the word "that" to convey sarcasm. In such cases, removing the stop word might strip away an essential element of the sarcasm, leading to a loss in the model's ability to detect the sentiment accurately. Therefore, while stop word removal is generally beneficial for most sentiment analysis tasks, it is crucial to carefully evaluate whether such removal is appropriate for sarcasm detection. In some instances, keeping specific stop words can help capture the subtle nuances of sarcastic expressions. The decision to remove or retain stop words should be made based on the specific characteristics of the dataset and the goals of the analysis. For sarcasm detection, retaining stop words that could affect the tone of the sentence is important for the model to detect sarcasm accurately.

In summary, tokenization, noise removal, and stop word removal are three key preprocessing techniques in sentiment analysis, each playing a vital role in preparing the text for machine learning models. Tokenization breaks the text down into manageable units, noise removal ensures that irrelevant elements are discarded, and stop word removal refines the dataset by eliminating non-contributory words. Together, these preprocessing methods help enhance the performance of models used for sarcasm detection by providing cleaner, more structured input data. For models like BERT and Logistic Regression, preprocessing is crucial, as sarcasm detection often depends on understanding the context, word relationships, and subtleties within the text. A well-prepared dataset ensures that these models can accurately capture patterns that indicate sarcasm, improving their predictive capabilities. The effectiveness of these preprocessing techniques significantly impacts the success of sentiment analysis, particularly in complex tasks like sarcasm detection. These preprocessing steps are essential not only for better model performance but also for enhancing the accuracy and reliability of NLP applications, such as social media monitoring, automated customer feedback analysis, and content moderation systems.

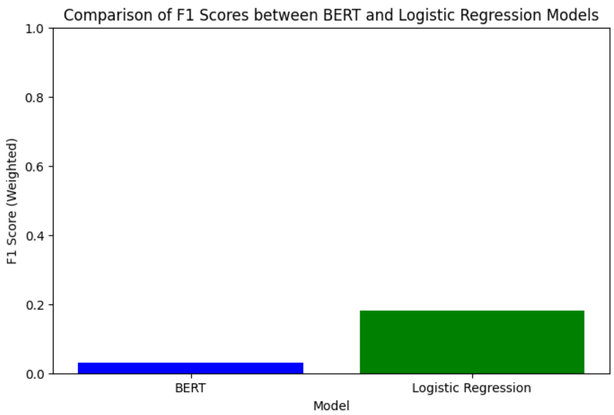
6. COMPARATIVE ANALYSIS OF BERT AND LOGISTIC REGRESSION ON TOKENIZED DATASET

This section assesses the performance of the BERT and Logistic Regression (LR) models in a sentiment analysis task using a preprocessed, tokenized dataset. The evaluation focuses on key metrics such as F1 scores and confusion matrices to determine how effectively each model classifies sentiment labels. The dataset underwent tokenization as part of the preprocessing stage, improving the feature set for both models. After preprocessing, both the Logistic Regression and BERT models were trained on the data, and their performance was evaluated based on F1 scores and confusion matrices, which offer insights into model accuracy and the distribution of predictions across different sentiment classes.

F1 Scores

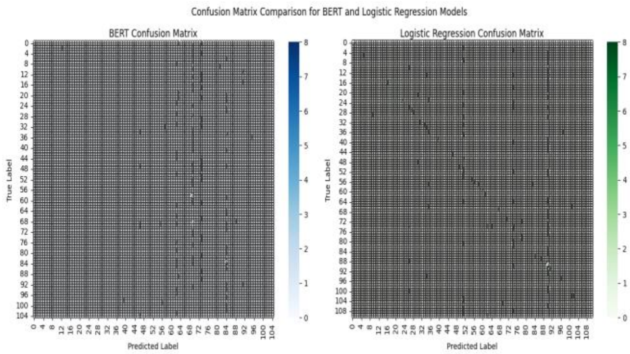
Logistic Regression: The Logistic Regression model achieved an F1 score of **0.1813**. This result indicates that the model was somewhat successful in capturing sentiment classes in the dataset, demonstrating that traditional machine learning models can still perform adequately under certain conditions.

BERT: In comparison, the BERT model attained a significantly lower F1 score of **0.0315**. This suggests that BERT faced difficulties with this particular dataset, potentially due to challenges arising from preprocessing choices or configuration settings. The relatively low F1 score highlights areas where the model could be further optimized for better sentiment classification



Confusion Matrix Analysis:

The analysis of the Logistic Regression model's confusion matrix revealed a mix of correctly and incorrectly predicted sentiment labels. This shows that while the model did not achieve perfect accuracy, it was able to correctly classify a reasonable number of cases, demonstrating some level of effectiveness. On the other hand, the confusion matrix for the BERT model showed very few instances classified correctly across the sentiment categories, indicating that the model struggled to effectively identify the appropriate sentiment. This could be attributed to the model's reliance on understanding the context of the text, which might not have been fully leveraged by the dataset's structure.



When comparing the performance of both models, it was evident that each faced distinct challenges in predicting sentiment accurately within the tokenized dataset. The F1 score analysis showed that Logistic Regression performed better than BERT in this scenario. Furthermore, the confusion matrices highlighted that while Logistic Regression was able to classify sentiment more effectively, BERT faced greater difficulty in correctly identifying the sentiment classes. These observations suggest that even though BERT is a sophisticated model with powerful capabilities, its effectiveness can be compromised depending on the quality of the dataset and preprocessing methods used. In contrast, simpler models like Logistic Regression might outperform BERT in certain situations, particularly when data processing limitations come into play. To improve model performance in future studies, it would be beneficial to refine preprocessing strategies, adjust hyperparameters, and work with larger and more representative datasets.

In conclusion, this study revealed that Logistic Regression outperformed BERT when applied to sentiment analysis of the tokenized dataset. The comparative analysis, supported by F1 scores and confusion matrix visualizations, shed light on the challenges faced by both models. This comparison offers a deeper understanding of how different models behave in NLP tasks and emphasizes the importance of careful data handling and appropriate model selection.

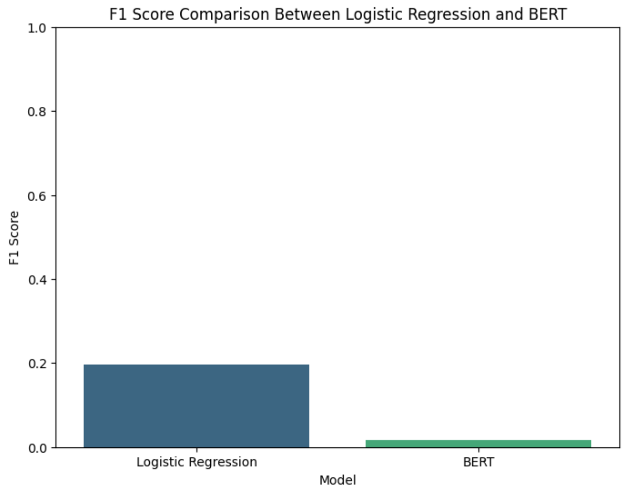
7.COMPARATIVE ANALYSIS OF BERT AND LOGISTIC REGRESSION BASED ON STOP WORD REMOVED DATASET

This section evaluates the performance of two prominent models, Logistic Regression and BERT, in performing sentiment analysis on a dataset where stop words have been eliminated. The evaluation primarily focuses on key performance metrics such as F1 scores and confusion matrices, to assess how effectively each model classifies sentiments accurately. The dataset underwent a preprocessing step where stop words were removed to improve the feature representation for both models. After this preprocessing, both the Logistic Regression and BERT models were trained and evaluated. To measure model effectiveness, F1 scores were used to balance both precision and recall, while confusion matrices were examined to understand how the models distributed their predictions across different sentiment categories.

F1 Scores:

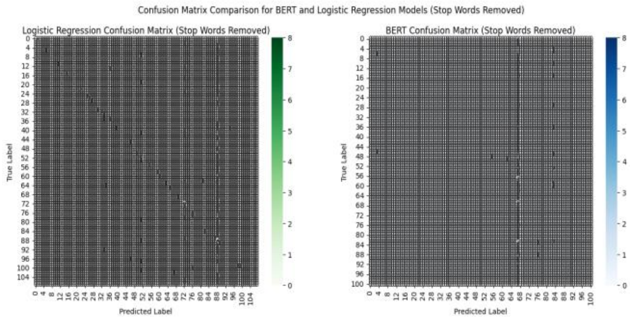
Logistic Regression: With an F1 score of **0.1955**, the Logistic Regression model demonstrated comparatively better performance in identifying sentiment categories within the dataset. This result underscores the efficiency of traditional machine learning models, particularly when tailored preprocessing techniques are applied effectively.

BERT: On the other hand, the BERT model obtained an F1 score of **0.0175**, a substantially lower score. This outcome suggests that BERT faced challenges when working with this dataset, which could be attributed to either the preprocessing decisions or the specific configuration of the model used for training.



Confusion Matrix Analysis:

The confusion matrix for Logistic Regression displayed a pattern where most cells were dark, with only a few lighter spots, indicating that the model correctly predicted some instances but mostly struggled with misclassifications. The highest frequency observed was 8, suggesting that while the model was able to correctly classify a few examples, its overall performance was scattered across the different classes. On the other hand, the confusion matrix for BERT revealed a similar pattern of sparsity, with the maximum value also being 8. This points to BERT's inability to make strong predictions in any specific class, indicating that it struggled to leverage its contextual capabilities effectively for this dataset.



The comparative analysis of Logistic Regression and BERT indicates that both models faced challenges in accurately predicting sentiment classes from the dataset with stop words removed. The F1 score results show that Logistic Regression outperformed BERT significantly in this scenario. The confusion matrices further illustrate that neither model succeeded in consistently identifying the correct classes, as reflected in the low counts across most cells. These findings suggest that while BERT is a powerful model, its performance can be context-dependent and may not always surpass simpler models, particularly when data preprocessing choices limit its effectiveness. Future research should explore more refined preprocessing techniques, hyperparameter tuning, and potentially larger datasets to enhance model performance.

In conclusion, the analysis revealed that the Logistic Regression model provided a more robust performance compared to BERT when evaluated on the dataset with stop words removed. The F1 scores and confusion matrix visualizations highlight the limitations faced by both models in sentiment classification. These insights enhance the overall comprehension of model effectiveness in natural language processing tasks and emphasize the significance of meticulous data preprocessing and thoughtful model choice.

8.COMPARISON OF BERT AND LOGISTIC REGRESSION MODELS ON NOISE REDUCED DATASET

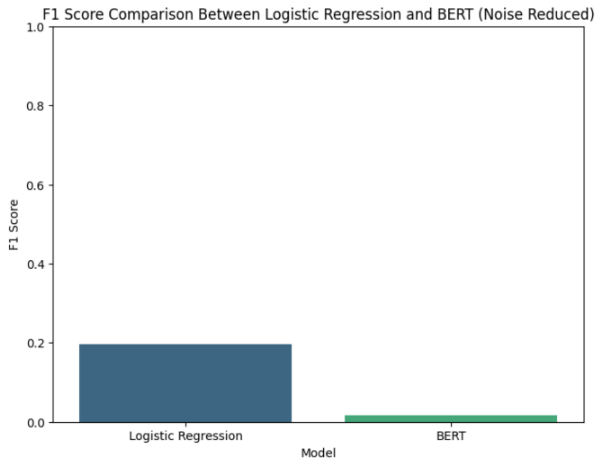
This section evaluates the performance of two prominent models—Logistic Regression and BERT—on a sentiment analysis task using a dataset with noise reduction applied. The analysis focuses on key metrics, including F1 scores and confusion matrices, to ascertain the effectiveness of each model in accurately classifying sentiments.

The dataset underwent noise reduction preprocessing to improve the feature quality for both models. Following this, The Logistic Regression model was trained together with the BERT model. The evaluation metrics included F1 scores, which offer a balance between precision and recall, and confusion matrices, which help analyze the spread of predictions across various classes

F1 Scores:

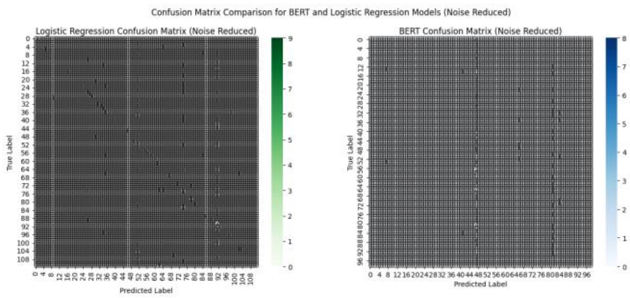
Logistic Regression: With an F1 score of **0.1839**, the model demonstrated moderate success in identifying sentiment categories, outperforming BERT in this regard.

BERT: The model recorded a significantly lower F1 score of **0.0038**, reflecting its struggle to adapt to this dataset despite noise reduction.



Confusion Matrix Analysis:

The Logistic Regression confusion matrix (left) shows a mostly dark background with scattered lighter points, indicating a few correct predictions across different classes. The maximum count in any cell is 9, suggesting that the model made infrequent but some correct predictions. The BERT confusion matrix (right) has a similar dark pattern with fewer scattered lighter spots, showing that it struggled with correct class predictions, with a maximum count of 8. This distribution points to BERT's limitations in generalizing across classes on this noise-reduced dataset.



The comparative analysis indicates that both models faced challenges in accurately predicting sentiment classes from the noise-reduced dataset. The F1 scores reveal that Logistic Regression outperformed BERT significantly in this scenario. The confusion matrices suggest that neither model excelled, with low prediction counts across classes. These findings imply that while BERT is generally a powerful model, its performance can vary based on data preprocessing and may not always surpass simpler models like Logistic Regression, particularly on preprocessed datasets. Future research might explore optimized preprocessing techniques, hyperparameter tuning, or larger datasets to improve performance.

In conclusion, the analysis revealed that the Logistic Regression model provided more robust performance than BERT when evaluated on the noise-reduced dataset. The F1 scores and confusion matrices underscore the importance of model selection and data preparation, contributing to a broader understanding of model performance in natural language processing tasks.

CONCLUSION

This research paper presents a comparative analysis of the BERT model and Logistic Regression (LR) for sentiment analysis, specifically examining how different preprocessing techniques—such as tokenization, stop word removal, and noise reduction—affect model performance. Our experiments reveal a consistent disparity in F1 scores between the two models, with Logistic Regression often outperforming BERT. Notably, Logistic Regression achieved an F1 score of 0.1955 compared to BERT's 0.0175 when stop words were removed. Similar patterns emerged with tokenization (F1: LR = 0.1813, BERT = 0.0315) and noise reduction (F1: LR = 0.1839, BERT = 0.0038), highlighting the unexpected sensitivity of BERT to these preprocessing methods in this dataset. These findings suggest that BERT's reliance on contextual embeddings, which generally enhance model robustness, may face limitations when simplified preprocessing techniques are applied. As discussed by Hao et al. in Visualizing and Understanding the Effectiveness of BERT, BERT's layered architecture is designed to generalize across tasks but can be sensitive to dataset-specific nuances. Moreover, the performance drop seen with BERT contrasts with its usual strength in handling nuanced language patterns, as identified in the literature review by Wankhade et al. on sentiment analysis challenges. Our results underscore the importance of selecting preprocessing methods that align with BERT’s strengths, as its performance here appears constrained by the lack of sophisticated data preparation. This study also underscores the efficacy of Logistic Regression in handling basic sentiment analysis tasks on this dataset, often outperforming BERT when preprocessing techniques were limited to simpler methods. While BERT generally excels in handling complex language interactions, its sensitivity to data preprocessing in this study suggests a need for more refined feature engineering or model tuning to fully leverage its capabilities. This finding aligns with literature emphasizing the impact of tokenization and preprocessing choices, as seen in Tokenization as the Initial Phase in NLP, which highlights tokenization's foundational role in model performance. In conclusion, our research illustrates that while BERT has significant potential for sophisticated language tasks, simpler approaches like Logistic Regression may prove more effective in certain contexts where preprocessing is less intricate. Future work should explore advanced preprocessing or feature extraction techniques that can better utilize BERT's architecture. This paper contributes to the discourse on preprocessing choices in NLP, offering insights that could inform enhanced methodologies in sentiment analysis and model optimization.

REFERENCES

1. Islam, Md Saiful & Zhang, Long. (2024). A Review on BERT: Language Understanding for Different Types of NLP Task. 10.20944/preprints202401.1857.v1.
2. Hao, Y., Dong, L., Wei, F., & Xu, K. (2019). Visualizing and Understanding the Effectiveness of BERT. ArXiv. https://arxiv.org/abs/1908.05620
3. Mickel Hoang, Oskar Alija Bihorac, and Jacobo Rouces. 2019. Aspect-Based Sentiment Analysis using BERT. In Proceedings of the 22nd Nordic Conference on Computational Linguistics, pages 187–196, Turku, Finland. Linköping University Electronic Press.
4. Wankhade, M., Rao, A.C.S. & Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev 55, 5731–5780 (2022). https://doi.org/10.1007/s10462-022-10144-1
5. Sentiment analysis using product review data
a. Xing Fang and Justin Zhan
6. Koroteev, Mikhail. (2021). BERT: A Review of Applications in Natural Language Processing and Understanding. 10.48550/arXiv.2103.11943.
7. Chakkarwar, Vrishali & Tamane, Sharvari & Thombre, Ankita. (2023). A Review on BERT and Its Implementation in Various NLP Tasks. 10.2991/978-94-6463-136-4_12.
8. Yu, Shanshan & Jindian, Su & Luo, Da. (2019). Improving BERT-Based Text Classification With Auxiliary Sentence and Domain Knowledge. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2953990.
9. Peters, Matthew & Neumann, Mark & Iyyer, Mohit & Gardner, Matt & Clark, Christopher & Lee, Kenton & Zettlemoyer, Luke. (2018). Deep contextualized word representations. 10.48550/arXiv.1802.05365.
10. Gu, Jiatao & Hassan, Hany & Devlin, Jacob & Li, Victor. (2018). Universal Neural Machine Translation for Extremely Low Resource Languages. 10.48550/arXiv.1802.05368.
11. Liu, Pengfei & Qiu, Xipeng & Huang, Xuanjing. (2016). Recurrent Neural Network for Text Classification with Multi-Task Learning. 10.48550/arXiv.1605.05101.
12. Jain, Praphula & Pamula, R. & Admasu Yekun, Ephrem. (2022). A multi-label ensemble predicting model to service recommendation from social media contents. The Journal of Supercomputing. 78. 1-18. 10.1007/s11227-021-04087-7.
13. Alhumoud, Sarah & Al Wazrah, Asma. (2022). Arabic sentiment analysis using recurrent neural networks: a review. Artificial Intelligence Review. 55. 10.1007/s10462-021-09989-9.
14. Mrityunjay, Singh & Jakhar, Amit & Pandey, Shivam. (2021). Sentiment analysis on the impact of coronavirus in social life using the BERT model. Social Network Analysis and Mining. 11. 10.1007/s13278-021-00737-z.
15. Acheampong, Francisca & Nunoo-Mensah, Henry & Chen, Wenyu. (2021). Transformer models for text-based emotion detection: a review of BERT-based approaches. Artificial Intelligence Review. 54. 1-41. 10.1007/s10462-021-09958-2.

16. Kumar, K. & Uma, V.. (2021). Intelligent sentinet-based lexicon for context-aware sentiment analysis: optimized neural network for sentiment classification on social media. The Journal of Supercomputing. 77. 10.1007/s11227-021-03709-4.
17. Jain, Praphula & Admasu Yekun, Ephrem & Pamula, R. & Srivastava, Gautam. (2021). Consumer recommendation prediction in online reviews using Cuckoo optimized machine learning models. Computers & Electrical Engineering. 95. 107397. 10.1016/j.compeleceng.2021.107397.
18. Jain, Praphula & Saravanan, Vijayalakshmi & Pamula, R.. (2021). A Hybrid CNN-LSTM: A Deep Learning Approach for Consumer Sentiment Analysis Using Qualitative User-Generated Contents. ACM Transactions on Asian and Low-Resource Language Information Processing. 20. 1-15. 10.1145/3457206.
19. Joshi, Mandar & Chen, Danqi & Liu, Yinhan & Weld, Daniel & Zettlemoyer, Luke & Levy, Omer. (2020). SpanBERT: Improving Pre-training by Representing and Predicting Spans. Transactions of the Association for Computational Linguistics. 8. 64-77. 10.1162/tacl_a_00300.
20. Jin, Di & Jin, Zhijing & Zhou, Joey & Szolovits, Peter. (2020). Is BERT Really Robust? A Strong Baseline for Natural Language Attack on Text Classification and Entailment. Proceedings of the AAAI Conference on Artificial Intelligence. 34. 8018-8025. 10.1609/aaai.v34i05.6311.
21. Yu, Shanshan & Jindian, Su & Luo, Da. (2019). Improving BERT-Based Text Classification With Auxiliary Sentence and Domain Knowledge. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2953990.
22. Chi, Sun & Qiu, Xipeng & Xu, Yige & Huang, Xuanjing. (2019). How to Fine-Tune BERT for Text Classification?. 10.1007/978-3-030-32381-3_16.
23. Lee, Jinhyuk & Yoon, Wonjin & Kim, Sungdong & Kim, Donghyeon & Kim, Sunkyu & So, Chan & Kang, Jaewoo. (2019). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics (Oxford, England). 36. 10.1093/bioinformatics/btz682.
24. Yoon, Wonjin & So, Chan & Lee, Jinhyuk & Kang, Jaewoo. (2019). CollaboNet: collaboration of deep neural networks for biomedical named entity recognition. BMC Bioinformatics. 20. 249. 10.1186/s12859-019-2813-6.
25. Zhao, Wei & Peyrard, Maxime & Liu, Fei & Gao, Yang & Meyer, Christian & Eger, Steffen. (2019). MoverScore: Text Generation Evaluating with Contextualized Embeddings and Earth Mover Distance. 563-578. 10.18653/v1/D19-1053.
26. Beltagy, Iz & Lo, Kyle & Cohan, Arman. (2019). SciBERT: A Pretrained Language Model for Scientific Text. 3606-3611. 10.18653/v1/D19-1371.
27. Wada, Takashi & Iwata, Tomoharu & Matsumoto, Yuji. (2019). Unsupervised Multilingual Word Embedding with Limited Resources using Neural Language Models. 3113-3124. 10.18653/v1/P19-1300.
28. Stanchev, Peter & Wang, Weiyue & Ney, Hermann. (2019). EED: Extended Edit Distance Measure for Machine Translation. 514-520. 10.18653/v1/W19-5359.
29. Lo, Chi-Kiu. (2019). YiSi - a Unified Semantic MT Quality Evaluation and Estimation Metric for Languages with Different Levels of Available Resources. 507-513. 10.18653/v1/W19-5358.
30. Cohan, Arman & Ammar, Waleed & Zuylen, Madeleine & Cady, Field. (2019). Structural Scaffolds for Citation Intent Classification in Scientific Publications. 3586-3596. 10.18653/v1/N19-1361.
31. Jurgens, David & Kumar, Srijan & Hoover, Raine & McFarland, Dan & Jurafsky, Dan. (2018). Measuring the Evolution of a Scientific Field through Citation Frames. Transactions of the

Association for Computational Linguistics. 6. 391-406. 10.1162/tacl_a_00028.
32. Nye, Benjamin & Li, Junyi & Patel, Roma & Yang, Yinfei & Marshall, Iain & Nenkova, Ani & Wallace, Byron. (2018). A Corpus with Multi-Level Annotations of Patients, Interventions and Outcomes to Support Language Processing for Medical Literature. Proceedings of the conference. Association for Computational Linguistics. Meeting. 2018. 197-207.
33. Liang, Bin & Li, Hongcheng & Su, Miaoqiang & Bian, Pan & Li, Xirong & Shi, Wenchang. (2018). Deep Text Classification Can be Fooled. 10.24963/ijcai.2018/585.
34. Zhang, Lei & Liu, Bing. (2016). Sentiment Analysis and Opinion Mining. 1-10. 10.1007/978-1-4899-7502-7_907-1.
35. Zhou, Shusen & Chen, Qingcai & Wang, Xiaolong. (2013). Active deep learning method for semi-supervised sentiment classification. Neurocomputing. 120. 536-546. 10.1016/j.neucom.2013.04.017.
36. Zhang, Yaowen & Xiang, Xiaojun & Yin, Cunyan & Shang, Lin. (2013). Parallel Sentiment Polarity Classification Method with Substring Feature Reduction. 7867. 121-132. 10.1007/978-3-642-40319-4_11.
37. Mukherjee, Arjun & Liu, Bing & Glance, Natalie. (2012). Spotting Fake Reviewer Groups in Consumer Reviews. WWW'12 - Proceedings of the 21st Annual Conference on World Wide Web. 10.1145/2187836.2187863.
38. Lin, Yuming & Zhang, Jingwei & Wang, Xiaoling & Zhou, Aoying. (2012). An information theoretic approach to sentiment polarity classification. ACM International Conference Proceeding Series. 10.1145/2184305.2184313.
39. Sarvabhotla, Kiran & Pingali, Prasad & Varma, Vasudeva. (2011). Sentiment classification: A lexical similarity based approach for extracting subjectivity in documents. Inf. Retr.. 14. 337-353. 10.1007/s10791-010-9161-5.
40. Jiang, Long & Yu, Mo & Zhou, Ming & Liu, Xiaohua & Zhao, Tiejun. (2011). Target-dependent Twitter Sentiment Classification.. ACL. 151-160.
41. Liu, Bing. (2010). Sentiment analysis and subjectivity.
42. Go, Alec & Bhayani, Richa & Huang, Lei. (2009). Twitter sentiment classification using distant supervision. Processing. 150.
43. Pang, Bo & Lee, Lillian. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval. 2. 1-135. 10.1561/1500000011.
44. Zhu, Xin & Wu, Han & Zhang, Lanfang. (2022). Automatic Short Answer Grading via BERT-based Deep Neural Networks. IEEE Transactions on Learning Technologies. 15. 1-1. 10.1109/TLT.2022.3175537.
45. Kannan, Mrinalini & Vijayalakshmi, P. & Nagarajan, T.. (2022). SBSim: A Sentence-BERT Similarity-Based Evaluation Metric for Indian Language Neural Machine Translation Systems. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 30. 1-1. 10.1109/TASLP.2022.3161160.
46. Zhou, Yuxiang & Liao, Lejian & Gao, Yang & Wang, Rui & Huang, Heyan. (2021). TopicBERT: A Topic-Enhanced Neural Language Model Fine-Tuned for Sentiment Classification. IEEE Transactions on Neural Networks and Learning Systems. PP. 1-14. 10.1109/TNNLS.2021.3094987.
47. Pan, Qian & Rong, Huan & Qian, Yurong & Tian, Yuan & Al-Nabhan, Najla. (2021). T-BERTSum: Topic-Aware Text Summarization Based on BERT. IEEE Transactions on Computational Social Systems. PP. 1-12. 10.1109/TCSS.2021.3088506.

48. Nogueira, Rodrigo & Cho, Kyunghyun. (2019). Passage Re-ranking with BERT.
49. Reimers, Nils & Gurevych, Iryna. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. 3973-3983. 10.18653/v1/D19-1410.
50. V. Chakkarwar and S. Tamane: "Social Media Analytics during Pandemic for Covid19 using Topic Modeling," 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC), 2020, pp. 279-282, doi: https://doi.org/10.1109/ICSIDEMPC4902s0.2020.9299617.
51. Dat Quoc Nguyen, Thanh Vu and Anh Tuan Nguyen: "BERTweet: A pre-trained language model for English Tweets"
52. Jacob Devlin, Ming-Wei Cheng, Kenton Lee, Kristina Tautonova: "BERT: Pre-training of deep bidirectional transformers for language understanding" in Proc. Annu. Conf. North Ameri. Chapter Assoc. Comp. Linguistics, Hum. Lang. Technol., 2019, p-41714186.
53. Victor Sanh, Lysandre Debut, Julien Chaumond, Thomas Wolf: "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter"
54. K. Mrinalini, P. Vijayalakshmi, T. Nagaranjan: "SBSim: A Sentence-BERT Similarity-Based Evaluation Metric for Indian Language Neural Machine Translation Systems", IEEE/ACM Transactions on Audio, Speech and Language Processing, Vol. 30, 2022.
55. Baevski, Alexei & Edunov, Sergey & Liu, Yinhan & Zettlemoyer, Luke & Auli, Michael. (2019). Cloze-driven Pretraining of Self-attention Networks. 5363-5372. 10.18653/v1/D19-1539.
56. Liu, Xiaodong & He, Pengcheng & Chen, Weizhu & Gao, Jianfeng. (2019). Multi-Task Deep Neural Networks for Natural Language Understanding. 4487-4496. 10.18653/v1/P19-1441.
57. Tenney, Ian & Das, Dipanjan & Pavlick, Ellie. (2019). BERT Rediscovers the Classical NLP Pipeline. 4593-4601. 10.18653/v1/P19-1452.
58. Peters, Matthew & Ruder, Sebastian & Smith, Noah. (2019). To Tune or Not to Tune? Adapting Pretrained Representations to Diverse Tasks. 7-14. 10.18653/v1/W19-4302.
59. Liu, Nelson & Gardner, Matt & Belinkov, Yonatan & Peters, Matthew & Smith, Noah. (2019). Linguistic Knowledge and Transferability of Contextual Representations. 1073-1094. 10.18653/v1/N19-1112.
60. Wang, Alex & Singh, Amanpreet & Michael, Julian & Hill, Felix & Levy, Omer & Bowman, Samuel. (2018). GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. 353-355. 10.18653/v1/W18-5446.
61. Howard, Jeremy & Ruder, Sebastian. (2018). Universal Language Model Fine-tuning for Text Classification. 328-339. 10.18653/v1/P18-1031.
62. Kuncoro, Adhiguna & Dyer, Chris & Hale, John & Yogatama, Dani & Clark, Stephen & Blunsom, Phil. (2018). LSTMs Can Learn Syntax-Sensitive Dependencies Well, But Modeling Structure Makes Them Better. 1426-1436. 10.18653/v1/P18-1132.
63. Williams, Adina & Nangia, Nikita & Bowman, Samuel. (2018). A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference. 1112-1122. 10.18653/v1/N18-1101.
64. Peters, Matthew & Neumann, Mark & Iyyer, Mohit & Gardner, Matt & Clark, Christopher & Lee, Kenton & Zettlemoyer, Luke. (2018). Deep Contextualized Word Representations. 2227-2237. 10.18653/v1/N18-1202.
65. Li, Hao & Xu, Zheng & Taylor, Gavin & Goldstein, Tom. (2017). Visualizing the Loss Landscape of Neural Nets. 10.48550/arXiv.1712.09913.

66. McCann, Bryan & Bradbury, James & Xiong, Caiming & Socher, Richard. (2017). Learned in Translation: Contextualized Word Vectors. 10.48550/arXiv.1708.00107.
67. Vaswani, Ashish & Shazeer, Noam & Parmar, Niki & Uszkoreit, Jakob & Jones, Llion & Gomez, Aidan & Kaiser, Lukasz & Polosukhin, Illia. (2017). Attention Is All You Need. 10.48550/arXiv.1706.03762.