A Comparative Analysis of BERT and VADER Models for Sentiment Analysis: Towards a Hybrid Approach

Atheek Mohamed Rafi
Computer Science
GC University
Lahore, Pakistan
mhdatheek136@gmail.com

Fathima Anizul Fathool
Computer Science
GC University
Lahore, Pakistan
anizulfathool@gmail.com

Abstract

Sentiment analysis, the process of computationally identifying and categorizing opinions in text, has advanced significantly with sophisticated deep learning models like BERT (Bidirectional Encoder Representations from Transformers) and rule-based methods such as VADER (Valence Aware Dictionary and sEntiment Reasoner). This survey paper reviews both BERT and VADER in sentiment analysis tasks, discussing their development, methodologies, strengths, and limitations. Additionally, it presents a simple hybrid approach that merges BERT's contextual comprehension with VADER's efficiency, aiming to match BERT's accuracy while operating at double the speed. By combining the best of both worlds, this hybrid model offers a promising avenue for enhancing sentiment analysis performance, making it a valuable resource for both researchers and practitioners in the field.

1 Introduction

Sentiment analysis, the process of computationally identifying and categorizing opinions expressed in text, has become increasingly vital in various domains such as customer feedback analysis, social media monitoring, and market research (Liu, 2022). With the emergence of online content and the need to extract meaningful insights from vast amounts of textual data, sentiment analysis techniques have garnered significant attention from researchers and practitioners alike.

Two prominent approaches to sentiment analysis are VADER (Valence Aware Dictionary and sEntiment Reasoner) and BERT (Bidirectional Encoder Representations from Transformers). VADER is a lexicon-based model that relies on pre-defined sentiment scores assigned to words and phrases (Hutto & Gilbert, 2014). Conversely, BERT is a deep learning model that captures contextual nuances through bidirectional processing of text (Devlin et al., 2018).

Comparing VADER and BERT is of substantial interest due to their divergent methodologies for addressing the same task of sentiment analysis. VADER offers simplicity and speed in sentiment analysis tasks, leveraging pre-defined lexicons to quickly assess sentiment polarity (Hutto & Gilbert, 2014). On the other hand, BERT excels in capturing complex linguistic patterns and contextual nuances, but at the expense of computational resources and time (Devlin et al., 2018).

The significance of comparing VADER and BERT lies in understanding the trade-offs between simplicity, speed, and accuracy in sentiment analysis. Additionally, exploring the feasibility of a hybrid approach that combines the strengths of both models presents an opportunity to potentially enhance sentiment analysis results. By leveraging the contextual understanding of BERT and the efficiency of VADER, a hybrid approach could offer improved efficiency in sentiment analysis tasks.

The objectives of this study are as follows:

- Evaluate the performance of VADER and BERT models on sentiment analysis tasks using a customer review dataset obtained from Kaggle.
- Compare the accuracy, efficiency, and computational resources required by both VADER and BERT models.

- Investigate the possibility of a hybrid approach that leverages the strengths of both models to enhance sentiment analysis results.
- Develop a classifier to determine which model performs better for a given review dataset, enabling the selection of the most suitable model based on data characteristics.
- Assess the trade-offs between accuracy and computational efficiency in using VADER, BERT, and the proposed hybrid approach.

By addressing these objectives, this study aims to provide insights into the performance and suitability of VADER, BERT, and their hybrid approach in sentiment analysis tasks, contributing to advancements in natural language processing research and applications.

2 Linguistics Challenges

Sentiment analysis, despite its growing utility, grapples with various linguistic complexities inherent in natural language. Four primary challenges—context dependence, sarcasm, slang, emoticons, and code mixing—underscore the intricacies of sentiment analysis tasks.

Context dependence presents a fundamental challenge in sentiment analysis as the sentiment expressed in a text heavily relies on its surrounding context (Birjali et al., 2021). For instance, consider the phrase *"The laptop is lighter than I expected."* Without context, this phrase might convey neutrality in VADER-like models due to the absence of both positive and negative words. However, BERT understands that laptops being lighter is perceived positively. While VADER assigns 3 points, indicating neutrality, BERT assigns 4 points, indicating positivity. This discrepancy demonstrates VADER's limitation in capturing nuanced sentiment based on context.

Sarcasm poses another significant challenge in sentiment analysis due to its inherent irony and often contradictory expression of sentiment (Birjali et al., 2021). For example, the sentence *"Wow, what a surprise, my package arrived two days late AGAIN. Fantastic service"* appears positive due to the words, "surprise" and "fantastic." However, the intended sentiment might be negative, expressing frustration or disappointment. Both VADER and BERT assign a score of 5, indicating highly positive sentiment, thereby failing to detect the underlying sarcasm.

Slang poses a significant challenge in sentiment analysis, as many models are not adequately trained to interpret it. For example, while VADER may struggle with slang terms like "omg" in phrases such as "omg, this book," leading to a neutral sentiment score of 3, BERT is equipped to handle slang and assigns a higher score of 5, indicating positivity. This discrepancy underscores the importance of considering slang in sentiment analysis tasks and highlights the advantage of models like BERT that account for its nuances.

Code-Mixing (CM), the blending of multiple languages within a single sentence, poses a significant challenge for sentiment analysis (Birjali et al., 2021). For instance, consider a sentence that mixes English and Spanish: "This book is muy muy interesante." While a model like VADER recognizes this as neutral sentiment, it doesn't fully understand the phrase "muy muy interesante" without specialized training on Spanish. In Spanish, "muy muy interesante" translates to "very very interesting" in English. However, VADER fails to capture the intensifying effect of "muy muy" due to its lack of Spanish proficiency, resulting in a neutral sentiment score.

On the other hand, a fine-tuned BERT model, designed for sentiment analysis, has the capability to understand Spanish. Therefore, BERT assigns a score of 5, indicating highly positive sentiment. This example highlights the difficulty in ensuring accurate sentiment detection across languages, particularly in models like VADER that lack language-specific understanding.

In addition to context dependence, sarcasm, slang and code-mixing, sentiment analysis encounters challenges in interpreting **emoticons and emoji**. While emoticons and emoji can convey sentiment effectively, their interpretation can vary significantly across models and algorithms. For instance, consider the phrase *"The service was 😔."* The inclusion of the sad face emoticon suggests a negative sentiment, but accurately capturing this sentiment poses a challenge. BERT, lacking the capability to recognize emoticons, may overlook the sentiment conveyed by the emoticon, leading to incorrect sentiment analysis. Conversely, VADER, which is trained to interpret emoticons, can provide a more accurate sentiment score based on emoji analysis (Hutto & Gilbert, 2014). In this case, BERT assigned a score of 4, while VADER assigned a score of 2. Therefore, the inclusion of emoticons and emoji adds another layer of complexity to sentiment analysis, highlighting the need for models capable of effectively interpreting these visual cues in textual data.

Addressing these linguistic challenges is crucial for enhancing the accuracy and reliability of sentiment analysis systems. By leveraging advanced natural language processing techniques and contextual understanding, researchers aim to develop more robust models capable of effectively navigating the complexities of sentiment expression in textual data.

3 Related Work (Background)

In this section, we'll explore existing literature on lexicon-based approaches, deep learning approaches, and hybrid approaches for sentiment analysis.

3.1 Lexicon-Based Approaches

Lexicon-based sentiment analysis methods have been pivotal in sentiment analysis tasks, offering simplicity and effectiveness in assessing sentiment polarity in textual data. While VADER (Valence Aware Dictionary and sEntiment Reasoner) stands out as a notable model in this category, several other lexicon-based methods have been developed and utilized in sentiment analysis tasks.

One such method is SentiWordNet, an extension of WordNet where each synset is associated with three sentiment scores: positivity, negativity, and objectivity (Baccianella et al., 2010). SentiWordNet assigns sentiment scores to synsets based on the synset's association with positive and negative words in a large corpus. Similarly, LIWC (Linguistic Inquiry and Word Count) categorizes words into different linguistic and psychological categories and computes sentiment scores based on the frequency of occurrence of these categories in a text (Pennebaker et al., 2001).

Moreover, the General Inquirer is widely used in sentiment analysis, categorizing words into various categories, including positive and negative sentiment categories, based on semantic and syntactic rules (Stone et al., 1966). ANEW (Affective Norms for English Words) assigns emotional valence scores to words based on human ratings of word affectiveness (Bradley & Lang, 1999).

Despite the availability of these alternative lexicon-based methods, VADER remains particularly notable due to its simplicity, effectiveness, and generalizability across various contexts, especially microblog-like environments such as social media. VADER's success can be attributed to its combination of lexical features with grammatical and syntactical conventions, which enables it to capture sentiment nuances and intensity expressions effectively (Hutto & Gilbert, 2014).

Furthermore, VADER's empirical validation process, which includes constructing and validating a gold standard list of lexical features, ensures its robustness and reliability in sentiment analysis tasks. Additionally, VADER's ability to outperform individual human raters and typical benchmarks underscores its efficacy in handling the inherent challenges of sentiment analysis in social media content and other textual data sources.

3.2 Deep Learning Based Approaches

Deep learning-based approaches, particularly models like BERT (Bidirectional Encoder Representations from Transformers), have revolutionized sentiment analysis by capturing complex linguistic patterns and contextual dependencies in textual data. BERT, introduced by Devlin et al. (2018), represents a breakthrough in natural language representation models. Unlike previous models, BERT is designed to pre-train deep bidirectional representations from unlabeled text, leveraging both left and right context in all layers. This bidirectional conditioning enables BERT to capture rich contextual information and achieve state-of-the-art performance on various natural language processing tasks (Devlin et al., 2018).

Research by Putrada et al. (2023) highlights the potential of using BERT for sentiment analysis tasks, particularly focusing on Rotten Tomatoes reviews. Leveraging pre-trained transformer models like BERT and DistilBERT, the study aims to analyze sentiment in the Rotten Tomatoes dataset. Through pre-processing, tuning, and training, the authors compare the performance of BERT with traditional machine learning benchmark methods such as Support Vector Machine (SVM), Naïve Bayes (NB), and Convolutional Neural Network (CNN). Results demonstrate BERT's superiority in terms of accuracy, precision, recall, F1-scores, and area under the curve (AUC) (Putrada et al., 2023).

These findings underscore the effectiveness of deep learning-based approaches, particularly BERT, in sentiment analysis tasks, showcasing their ability to handle complex linguistic nuances and achieve superior performance compared to traditional machine learning methods.

3.3 Hybrid Based Approaches

Format Hybrid approaches that combine lexicon-based techniques with deep learning methods have emerged as promising solutions for sentiment analysis tasks. One such hybrid model, named LeBERT, is introduced by Mutinda et al. (2023), aiming to address the limitations of both lexicon-based and machine learning-based techniques in accurately representing text and classifying sentiment.

Mutinda's LeBERT model combines sentiment lexicon, N-grams, BERT, and Convolutional Neural Network (CNN) to effectively capture both semantic and sentiment-related information in textual data. By leveraging sentiment lexicon and N-grams for word vectorization and integrating BERT embeddings with CNN for sentiment classification, the model aims to achieve superior performance in sentiment analysis tasks.

The proposed LeBERT model is evaluated on various public datasets, including Amazon products' reviews, IMDb movies' reviews, and Yelp restaurants' reviews. Performance metrics such as accuracy, precision, and F-measure are used to assess the model's effectiveness. Experimental results demonstrate that the LeBERT model outperforms existing state-of-the-art models, achieving a high F-measure score of 88.73% in binary sentiment classification (Mutinda et al., 2023).

This research underscores the potential of hybrid approaches in sentiment analysis, showcasing how the integration of lexicon-based techniques with deep learning methods can lead to enhanced performance and accuracy in sentiment classification tasks.

4 Implementation

The implementation strategy encompasses the utilization of two distinct models for sentiment analysis: VADER and BERT. Each model offers unique strengths and capabilities, contributing to a comprehensive sentiment analysis framework.

4.1 VADER and BERT implementation

VADER, or Valence Aware Dictionary and sEntiment Reasoner, provides a lexicon-based approach to sentiment analysis (Hutto & Gilbert, 2014). This model is particularly adept at detecting sentiment in social media text and offers simplicity in implementation. In our implementation, we modified the VADER model to align with the customer rating scale, converting its default sentiment score range from -1 to 1 to a more intuitive 1 to 5 scale, which matches the range of customer ratings.

BERT, or Bidirectional Encoder Representations from Transformers, represents a state-of-the-art deep learning model for natural language processing tasks. In our implementation, we utilize a pre-trained fine-tuned version of the BERT model in hugging face specifically designed for sentiment analysis on product reviews across six languages. This model is called *"bert-base-multilingual-uncased-sentiment"*. This model predicts the sentiment of reviews as a star rating, ranging from 1 to 5.

4.2 Data Collection

We utilized a Kaggle dataset comprising customer reviews from the Amazon Book Store category spanning from May 1996 to July 2014. The dataset encompasses a total of 3 million entries. Due to BERT BASE sentiment analysis's input length limitation of 512 characters, we reduced the dataset to approximately 1.5 million reviews to ensure compatibility with the model. Subsequently, both the BERT and VADER models are independently applied to the dataset to obtain sentiment analysis results.

4.3 Hybrid Approach Implementation

It's intriguing to note that among 1.5 million data points, both the BERT and VADER models exhibit similar performance for 7.5 million data. Considering that BERT requires nearly ten times the processing time compared to VADER for reviewing each input, there's a compelling case to develop a hybrid approach. Such a hybrid could identify reviews where both VADER and BERT are likely to perform equally well and prioritize the usage of VADER for these cases. This strategy could substantially reduce processing time while maintaining satisfactory performance.

To construct this hybrid approach, feature selection is pivotal. Here's how we perform it:

We extract various features from the reviews to enrich the accuracy of sentiment analysis. These encompass:

- Number of Words: Provides insight into the length and complexity of the review.
- Number of Characters: Helps capture the overall text length and potential verbosity.

- Emoticons: Understanding emotive expressions can significantly impact sentiment analysis.
- Punctuation Marks: Their presence and frequency can influence the tone and sentiment of the text.
- Capital Letters: Indicates emphasis or intensity of sentiment in certain contexts.
- Exclamations: Often indicative of strong emotions or sentiments in the text.
- Positive, Negative, and Neutral Words: Crucial for sentiment classification, these words help in discerning the overall sentiment of the text.
- Slang Words: Capturing informal language can be crucial as it often conveys sentiment more vividly.

By incorporating these features, we aim to create a comprehensive understanding of the text, which can then be leveraged by both VADER and BERT models to make informed sentiment predictions. This approach allows us to exploit the strengths of each model effectively while mitigating the computational burden posed by BERT's intensive processing requirements.

The count of slang words is determined using a custom list of slang terms, allowing for a more comprehensive analysis of the language used in the reviews.

BERT's strength lies in its ability to understand contextual dependencies, making it more suitable for longer texts, while VADER excels at detecting emotions conveyed through emoticons, capital letters, and exclamations (Hutto & Gilbert, 2014). Thus, features such as the number of words, characters, emoticons, capital letters, and exclamations are prioritized.

The counts of positive, negative, and neutral words are derived from the AFINN lexicon, a commonly employed resource in sentiment analysis. This lexicon consists of a broad array of English words, each assigned a numerical sentiment score, offering a straightforward means to gauge sentiment polarity (Al-Shabi, 2020).

Additionally, the TF-IDF weighting scheme can be employed to assess the significance of terms within the document (Birjali et al., 2021). By applying TF-IDF, 2,000 additional features are generated from the review text, enhancing the sentiment analysis process.

In the hybrid approach, a random forest classifier with 100 n-estimators is employed to predict the three classes: class 0 (indicating VADER outperformed), class 1 (indicating BERT outperformed), and class 2 (indicating both models equally performed) for each review based on its characteristics. This model undergoes training on 80% of the selected 1.5 million entries, specifically those where BERT and VADER scores diverged from human ratings. The remaining 20% (comprising 298,129 entries) are reserved for testing.

Reviews displaying distinctive attributes such as emoticons, capital letters, and exclamations may be better suited for VADER, while others could benefit from BERT's contextual analysis. Furthermore, reviews that demonstrate equal effectiveness in both BERT and VADER can leverage VADER to enhance efficiency.

The classifier generates a calculation report showcasing metrics such as accuracy, precision, recall, and F1-score.

Table 1. Binary Classifier Calculation Report.

	Precision	**Recall**	**F1score**	**support**
Class 0	0.99	0.41	0.58	48226
Class 1	0.73	0.78	0.75	102268
Class 2	0.77	0.88	0.82	147635
Accuracy			0.77	298129

With an accuracy of 77%, the classifier effectively predicts the most suitable model for sentiment analysis. This trained model is then utilized for the hybrid approach, where reviews are processed using either VADER or BERT based on the classifier's predictions.

5 Evaluation

Evaluating the performance of sentiment analysis models and approaches is important to understanding their effectiveness. In this section, we conduct a comparative analysis of VADER, BERT, and our hybrid approach, focusing on accuracy, efficiency, and computational resource utilization. The evaluation is conducted on a dataset comprising 641,171 Kindle reviews obtained from Kaggle, ensuring a comprehensive assessment on unseen data.

5.1 Error Analysis

The first aspect of our evaluation involves comparing the predicted sentiment scores generated by VADER, BERT, and the hybrid approach against human-rated sentiment scores. Table 2 provides a summary of the Mean Absolute Error (MAE) and Mean Squared Error (MSE) values for each model in predicting sentiment scores.

Table 2. Error Analysis Table.

Model	MAE	MSE
Hybrid	0.5026	0.7972
BERT	0.4812	0.7295
VADER	0.6605	1.1541

The MAE and MSE values serve as indicators of the average magnitude and spread of errors between predicted and actual sentiment scores. Lower values of MAE and MSE suggest better alignment between predicted and human-rated sentiment scores. Upon examination of Table 2, it becomes apparent that the hybrid approach exhibits slightly higher MAE and MSE values compared to BERT, yet lower than VADER. Consequently, the hybrid approach manages to maintain nearly the same level of error as BERT while outperforming VADER in sentiment score prediction.

5.2 Classification Metrics

we'll delve into the classification metrics and overall accuracy of each approach.

Table 3. Hybrid Approach Calculation Report.

Ratings	Precision	Recall	F1score	support
1	0.43	0.43	0.43	16471
2	0.30	0.46	0.36	22074
3	0.40	0.36	0.38	60227
4	0.36	0.34	0.35	152161
5	0.75	0.76	0.76	390238
Accuracy			0.61	641171

Table 4. BERT Approach Calculation Report.

Ratings	Precision	Recall	F1score	support
1	0.42	0.48	0.45	16471
2	0.31	0.55	0.40	22074
3	0.41	0.51	0.45	60227
4	0.40	0.57	0.47	152161
5	0.86	0.64	0.74	390238
Accuracy			0.61	641171

Table 5. VADER Approach Calculation Report.

Ratings	Precision	Recall	F1score	support
1	0.21	0.15	0.18	16471
2	0.15	0.22	0.18	22074
3	0.20	0.15	0.17	60227
4	0.27	0.25	0.26	152161
5	0.69	0.72	0.71	390238
Accuracy			0.53	641171

In evaluating the performance of the sentiment analysis models, several metrics are considered, including accuracy, precision, recall, and F1-score, as presented in Tables 3, 4, and 5. These metrics provide insights into the models' ability to classify reviews accurately across different sentiment categories.

The hybrid approach achieves parity with the BERT model, boasting an accuracy of 61%. Moreover, it demonstrates competitive precision, recall, and F1-score metrics across all sentiment categories, indicating a well-balanced performance in review classification.

Conversely, BERT achieves an accuracy of 61% as well. However, it excels in precision, recall, and F1-score, particularly for sentiment categories 4 and 5, showcasing its proficiency in accurately categorizing positive sentiment reviews.

In contrast, VADER lags in accuracy, attaining 53%. It also exhibits lower precision, recall, and F1-score compared to both the hybrid and BERT approaches. This discrepancy suggests that while VADER may suffice in certain contexts, its lexicon-based approach may not capture sentiment nuances as effectively as deep learning models like BERT.

Overall, the evaluation underscores the hybrid approach's effectiveness in optimizing sentiment analysis outcomes. By leveraging the strengths of both VADER and BERT models, it achieves comparable performance to BERT while retaining the advantages of both methodologies.

5.3 Accuracy and Off-by-1 Accuracy

The accuracy (off-by-1) metric provides a more lenient evaluation criterion, allowing for predictions that are one level off from the human-assigned sentiment rating. This metric is particularly insightful as it considers cases where the model's prediction is in close alignment with the human judgment.

Table 6. Accuracy (Off-by-1)

Model	*Exact Accuracy*	*Accuracy Off-by-1*
Hybrid	0.6051	0.9238
BERT	0.6059	0.9413
VADER	0.5255	0.8657

Analysing the accuracy (off-by-1) results across the different sentiment analysis approaches, we observe the following:

Hybrid Sentiment: Achieves an accuracy (off-by-1) of 92.38%, indicating that the hybrid sentiment approach accurately predicts sentiment ratings within one level of discrepancy from the human-assigned ratings in most cases.

BERT Sentiment: Demonstrates a slightly higher accuracy (off-by-1) of 94.13% compared to the final sentiment approach.

VADER Sentiment: Exhibits a notably lower accuracy (off-by-1) of 86.57%, indicating that VADER's predictions deviate more frequently by one level from the human-assigned ratings. Despite its effectiveness in certain scenarios, VADER's performance lags behind BERT and the Hybrid sentiment approach in terms of accurately predicting sentiment with a one-level discrepancy.

Overall, the accuracy (off-by-1) metric underscores the effectiveness of the Hybrid approach and BERT in aligning closely with human-assigned sentiment ratings. Conversely, VADER demonstrates a higher rate of off-by-one errors, suggesting a lower level of accuracy in predicting sentiment compared to the other approaches.

5.4 Processing Time

The processing time analysis reveals significant differences in the computational efficiency of the sentiment analysis approaches. These assessments were performed on a system featuring an AMD Ryzen 7 4800H CPU and 16 GB of RAM. Additionally, the results are presented as average seconds, obtained by running each approach ten times.

In the hybrid approach, both BERT and VADER are leveraged to analyze sentiment in the dataset. Specifically, BERT handles 202,923 entries while VADER handles 438,248 entries in in dataset used for evaluation. This division of labor allows for a more efficient processing of the dataset compared to using BERT alone. By distributing the workload between the two models, the overall time required for sentiment analysis is reduced.

Table 7. Processing Time for 10,000 reviews

Model	***Processing Time (seconds)***
Hybrid	604.88
BERT	1271.13
VADER	99.80

Hybrid Model: This approach, which combines the strengths of both VADER and BERT models, demonstrates moderate processing time, taking approximately 604.88 seconds to analyze 10000 reviews. Despite incorporating more complex processing logic, the hybrid model's efficiency remains commendable.

BERT Model: Utilizing the powerful contextual analysis capabilities of BERT, this model exhibits a longer processing time compared to the hybrid approach. It takes around 1271.13 seconds to analyze the same dataset of 10000 reviews. The increased processing time can be attributed to BERT's deep neural network architecture, and the computational resources required for fine-tuning.

VADER Model: Leveraging a lexicon-based approach, VADER demonstrates the fastest processing time among the three approaches, taking only 99.80 seconds to analyze 10000 reviews. The efficiency of VADER stems from its simplicity and reliance on pre-defined sentiment lexicons, requiring fewer computational resources compared to deep learning models like BERT.

In evaluating processing time, it's essential to consider the trade-offs between efficiency and accuracy. While VADER offers rapid analysis, it may sacrifice nuanced understanding in complex linguistic contexts. On the other hand, BERT's deeper analysis capabilities come at the cost of increased processing time. The hybrid approach strikes a balance between efficiency and accuracy by leveraging the strengths of both models. Depending on specific application requirements, stakeholders can prioritize either efficiency or accuracy in their sentiment analysis tasks.

6 Conclusion

Performance The evaluation of sentiment analysis approaches, encompassing VADER, BERT, and a hybrid model, has provided valuable insights into their performance, efficiency, and computational requirements.
Across the board, each approach exhibits distinct strengths and weaknesses:

- VADER, with its lexicon-based approach, offers rapid processing and simplicity but may lack nuanced understanding in complex linguistic contexts.
- BERT, leveraging deep contextual analysis, demonstrates superior accuracy but

requires more computational resources and longer processing time.
- The hybrid approach, combining VADER's efficiency with BERT's accuracy, strikes a balance between speed and precision. It offers a versatile solution that can adapt to various application requirements.

Moreover, the evaluation highlights the importance of considering trade-offs between accuracy, efficiency, and computational resources in selecting a sentiment analysis approach. Depending on the specific needs of a project, stakeholders can prioritize either speed or accuracy to optimize performance.

However, while our study represents a significant step forward in sentiment analysis techniques, several challenges remain unaddressed. The inherent complexity of sarcasm detection, for example, poses a persistent challenge that warrants further exploration. Additionally, refining models to better understand linguistic nuances and cultural context is crucial for improving accuracy across diverse datasets and languages.

Another limitation lies in the constraint imposed by BERT base, which can only handle text inputs of up to 512 characters. As a result, our research is confined within this limitation, and there is a need to extend our analysis to handle longer texts effectively.

Moving forward, further research and experimentation could focus on refining the hybrid approach to enhance its efficiency and accuracy. Additionally, exploring alternative sentiment analysis methodologies and models may offer new insights into optimizing sentiment analysis for diverse applications.

In conclusion, the evaluation underscores the significance of tailored sentiment analysis solutions that align with the objectives and constraints of specific projects, ultimately facilitating more informed decision-making and insights generation.

References

Al-Shabi, M. (2020). Evaluating the performance of the most important Lexicons used to Sentiment analysis and opinions Mining. IJCSNS, 20(1), 1.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010, May). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec* (Vol. 10, No. 2010, pp. 2200-2204).

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134.

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (Vol. 30, No. 1, pp. 25-36). Technical report C-1, the center for research in psychophysiology, University of Florida.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805.*

Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).

Liu, B. (2022). *Sentiment analysis and opinion mining.* Springer Nature.

Mutinda, J., Mwangi, W., & Okeyo, G. (2023, January 21). Sentiment Analysis of Text Reviews Using Lexicon-Enhanced Bert Embedding (LeBERT) Model with Convolutional Neural Network. *Applied Sciences*, *13*(3), 1445.

Pennebaker, James & Francis, Martha & Booth, Roger. (1999). Linguistic inquiry and word count (LIWC).

Putrada, A. G., Alamsyah, N., & Fauzan, M. N. (2023, August). BERT for Sentiment Analysis on Rotten Tomatoes Reviews. In *2023 International Conference on Data Science and Its Applications (ICoDSA)* (pp. 111-116). IEEE.

Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). *The general inquirer: A computer approach to content analysis.* M.I.T. Press.