COMPUTER Society · ASCEE

Improving sentiment analysis on PeduliLindungi comments: a comparative study with CNN-Word2Vec and integrated negation handling

CrossMark

Herlina Jayadianti [a,1,*], Berliana Andra Arianti [b,2], Nur Heri Cahyana [b,3], Shoffan Saifullah [a,b,4], Rafał Dreżewski [b,c,5]

[a] Department of Informatics, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia
[b] Faculty of Computer Science, AGH University of Krakow, Poland
[c] Artificial Intelligence Research Group (AIRG), Informatics Department, Universitas Ahmad Dahlan, Indonesia
[1] herlina.jayadianti@upnyk.ac.id; [2] 123180065@student.upnyk.ac.id; [3] nur.hericahyana@upnyk.ac.id;
[4] saifulla@agh.edu.pl; [5] drezew@agh.edu.pl
* Corresponding Author

ARTICLE INFO

ABSTRACT

This study investigates sentiment analysis in Google Play reviews of the PeduliLindungi application, focusing on the integration of negation handling into text preprocessing and comparing the effectiveness of two prominent methods: CNN-Word2Vec CBOW and CNN-Word2Vec SkipGram. Through a meticulous methodology, negation handling is incorporated into the preprocessing phase to enhance sentiment analysis. The results demonstrate a noteworthy improvement in accuracy for both methods with the inclusion of negation handling, with CNN-Word2Vec SkipGram emerging as the superior performer, achieving an impressive 76.2% accuracy rate. Leveraging a dataset comprising 13,567 comments, this research introduces a novel approach by emphasizing the significance of negation handling in sentiment analysis. The study not only contributes valuable insights into the optimization of sentiment analysis processes but also provides practical considerations for refining methodologies, particularly in the context of mobile application reviews.

1. Introduction

The inception of 2020 brought forth an unparalleled global crisis—the COVID-19 pandemic [1], [2]. In response, governments worldwide, including Indonesia [3], initiated groundbreaking measures to combat the spread of the virus [4]. One such initiative was the development of the PeduliLindungi application, designed to serve as a pivotal tool in contact tracing and screening of COVID-19-affected individuals [5], [6]. While the application holds significant potential in augmenting public health efforts [7], its adoption and effectiveness are contingent upon user acceptance [8]. This study delves into the intricate landscape of user sentiments towards PeduliLindungi, utilizing comments extracted from the Google Play platform as a comprehensive dataset for analysis.

Recognizing the multifaceted nature of user sentiments, sentiment analysis emerges as a crucial mechanism for gauging the application's reception and identifying avenues for refinement [9]. The inherent complexity of user expressions necessitates a nuanced approach, and thus, sentiment analysis becomes instrumental in uncovering valuable insights to inform ongoing improvements and enhance overall features [10], [11].

The preprocessing stage stands as a critical precursor to the sentiment classification process, with stopwords removal being a pivotal component [12], [13]. In addressing the intricacies of negative sentence classification, this study introduces a negation handling process during preprocessing [14]–[16]. This innovative addition seeks to mitigate potential distortions in sentiment interpretation, ensuring a more accurate portrayal of user sentiments [17]–[19]. Preliminary studies incorporating negation handling demonstrate promising outcomes [20]–[23], showcasing a notable 5–7 p.p improvement in test results [4], [18].

While classical machine learning [24] methods have traditionally been employed for sentiment analysis [11], [25]–[28], their limitations in handling intricate feature extraction methods prompt a shift towards more advanced techniques [29]–[31]. This study embraces deep learning methodologies, utilizing a convolutional neural network (CNN) as the classification model and Word2Vec for word embedding. The integration of negation handling into the text preprocessing phase enriches the analytical process, promising heightened precision and reliability in sentiment analysis. As the first exploration of its kind within the context of PeduliLindungi, this research not only contributes to the ongoing discourse on sentiment analysis but also presents practical implications for refining mobile application evaluation methodologies.

This paper is structured as follows: Section 2 outlines the methods employed in this study, elucidating the intricacies of our approach. Section 3 presents the results of our analysis and engages in a comprehensive discussion of the findings. Lastly, Section 4 encapsulates the essence of our study, providing conclusive insights and implications derived from our research, thereby contributing to the broader understanding of sentiment analysis in the realm of mobile applications.

2. Method

The study follows a structured approach with distinct sections: data collection, labelling, preprocessing, word embedding, and CNN implementation. This meticulous methodology aims to extract meaningful insights from user sentiments on the PeduliLindungi application, incorporating advanced techniques to ensure robust sentiment analysis.

2.1. Data Collections

The data collection phase is paramount in ensuring the acquisition of a comprehensive and representative dataset for the subsequent sentiment analysis [32]. Employing the Google-play-scraper Python library, four essential components govern the extraction process: application id, language, country, and sort. This approach facilitates a meticulous retrieval of user comments on the PeduliLindungi application from the Google Play website.

The dataset is carefully curated to encompass a period from June 2020 to June 2021, a timeframe selected strategically to capture diverse user interactions and sentiments. Given the application's pronounced relevance during this period marked by the global COVID-19 pandemic, it ensures a robust representation of user experiences and sentiments. The temporal span ensures the inclusion of a

substantial volume of user-generated content, vital for the reliability and comprehensiveness of the analysis.

A total of 13,567 comments are amassed through this scraping process, forming the basis for subsequent analysis. To facilitate robust model training and evaluation, the dataset is judiciously divided into training (10,853 comments) and testing (2,714 comments) sets, maintaining an 80:20 ratio [33]. This partitioning is essential to validate the model's performance on unseen data, contributing to the overall generalizability of the sentiment analysis model. To enhance accessibility and ease of analysis, the collected dataset is then stored in a structured .csv format. This meticulous approach to data collection establishes a robust foundation for subsequent sentiment analysis, ensuring the inclusion of a diverse range of user sentiments, experiences, and opinions related to the PeduliLindungi application during a critical period in its utilization.

2.2. Data Labeling

The data labelling stage is pivotal in assigning sentiment classes to user comments, facilitating the subsequent development of a sentiment analysis model [34], [35]. Leveraging Google Play ratings as a proxy for sentiment, each comment within the meticulously collected dataset undergoes a nuanced labelling process.

The sentiments are stratified into three distinct classes: positive, neutral, and negative. This classification is based on a rating scale where a rating of 5 is designated as positive, 3 as neutral, and 1 as negative. The rationale behind this categorization aligns with the inherent structure of Google Play ratings, providing a pragmatic approach to sentiment classification.

The distribution of sentiments within the dataset provides valuable insights into the prevailing user sentiments toward the PeduliLindungi application. Table 1 offers a comprehensive breakdown of the sentiment classes, shedding light on the distribution of positive, neutral, and negative sentiments within the 13,567 comments.

Table 1. Dataset class sentiment and splitting

Data	Category	Amount	Total
Class	Positive	5,044	
	Neutral	1,316	13,567
	Negative	7,207	
Split	Train	10,853	13,567
	Test	2.714	

This detailed classification ensures a nuanced understanding of the sentiment distribution, laying the groundwork for an in-depth sentiment analysis model. The prevalence of negative sentiments, for instance, can signal potential areas of improvement for the PeduliLindungi application, prompting developers and stakeholders to address specific pain points highlighted by users.

Furthermore, the meticulous labelling process aligns with best practices in supervised machine learning [36], [37], providing a labeled dataset that can be leveraged for model training and validation [38], [39]. The inclusion of neutral sentiments acknowledges the diversity of user opinions, ensuring that the sentiment analysis model can discern and appropriately categorize nuanced expressions that may not inherently convey positivity or negativity.

In essence, the data labelling process transforms raw comments into a structured dataset, imbued with sentiment labels, thereby setting the stage for the subsequent stages of preprocessing, feature extraction, and model development. This comprehensive sentiment labelling approach enriches the dataset, enabling a more nuanced and accurate understanding of user sentiments toward the PeduliLindungi application.

2.3. Data Preprocessing

The preprocessing of data is a pivotal stage in the sentiment analysis pipeline, involving a series of meticulous operations to refine raw text data and prepare it for subsequent analysis [40]. The following operations are conducted to ensure the dataset is transformed into a structured and standardized format [41].

- Case Folding: The conversion of all text to lowercase. This operation is imperative to ensure uniformity in the representation of words and to mitigate potential discrepancies arising from variations in letter case.

- Cleaning Text: A comprehensive cleaning process is employed to remove irrelevant characters, symbols, and any extraneous elements that do not contribute to the overall sentiment analysis. This step is crucial for eliminating noise and enhancing the quality of the text data.

- Word Normalization: Ensuring uniform representation of words by standardizing variations such as verb conjugations or plural forms. This step aids in reducing the dimensionality of the data and facilitating a more effective analysis.

- Stemming: The process of reducing words to their root form, enabling the model to recognize and categorize words with similar meanings more effectively. Stemming contributes to a more cohesive analysis by consolidating words with shared roots.

- Negation Handling: An innovative addition to the preprocessing stage, negation handling involves strategies to preserve the original sentiment of a sentence after the occurrence of negation words like "no" [42], [43]. The chosen method, "First Sentiment Word (FSW)," involves changing the polarity of the first word after the negation word [44]. This addition addresses the challenge of negation impacting sentiment categorization, ensuring a more accurate reflection of user sentiments.

- Stopword Removal: The elimination of words with minimal semantic significance, commonly known as stopwords. This step reduces noise in the dataset, focusing on words that carry more substantive meaning in the context of sentiment analysis.

- Tokenization: The segmentation of text into individual tokens or words. Tokenization is fundamental for subsequent feature extraction and model training, breaking down the text into units that the model can analyze effectively.

The incorporation of negation handling in this preprocessing stage is particularly significant. The method chosen, FSW, strategically alters the sentiment of the first word following a negation word. This approach not only enhances the accuracy of sentiment classification but also ensures that the intended meaning of negated sentences remains intact.

Negation handling is pivotal in sentiment analysis, especially in contexts where the sentiment of a statement can be drastically altered by the presence of negation words. The meticulous approach to

preprocessing outlined here sets the stage for a robust sentiment analysis model, capable of discerning nuanced sentiments and accurately categorizing user opinions on the PeduliLindungi application.

2.4. Word Embedding Word2Vec

Word embedding using Word2Vec is a transformative process designed to convert textual data into numerical vectors, imbuing the words with numerical representations that capture their semantic relationships [45]. This section delves into the intricacies of Word2Vec and its two distinct architectures: Continuous Bag of Words (CBOW) and Skip-gram.

- Continuous Bag of Words (CBOW)

  - CBOW functions by predicting the target word from its surrounding context [46]. It considers the context words within a given window and attempts to predict the target word, resulting in a vector representation for each word.

  - CBOW demonstrates stability and faster training times, making it particularly suitable for larger datasets and longer texts such as news articles.

  - The model excels in performance and stability but may exhibit slightly less accuracy for frequently occurring words.

- Skip-Gram

  - In contrast, Skip-gram predicts the context words based on a given target word [12], [47]. It operates by taking a target word and predicting the words that are likely to appear in its context.

  - Skip-gram is adept at working with smaller datasets and excels in properly representing words that are rare or considered less common.

  - This model is particularly useful for capturing nuanced relationships between words and is effective in scenarios where data size might be a constraint.

The choice between CBOW and Skip-gram depends on the characteristics of the dataset and the specific requirements of the sentiment analysis task. For instance, in the context of analyzing user sentiments on the PeduliLindungi application, the choice could be influenced by the nature of comments and the diversity of language used.

The Word2Vec process involves training the model to create these vector representations [48]. The model learns to assign numerical values to words based on their contextual relationships in the training corpus. The resulting vectors capture semantic similarities, enabling the model to understand the inherent meanings and relationships between words.

Word2Vec, with its CBOW and Skip-gram architectures, serves as a powerful tool for word embedding [48], [49], enriching the dataset with numerical representations that encapsulate the semantic nuances of the language used in user comments. This step is crucial for preparing the data for the subsequent stages of sentiment classification using advanced deep learning methods [50]. The quality of word embeddings significantly influences the model's ability to discern subtle variations in sentiment and improve the overall accuracy of sentiment analysis on the PeduliLindungi application comments.

2.5. Some Common Mistakes

The Convolutional Neural Network (CNN) is a pivotal component in the sentiment analysis framework [51], bringing a powerful deep learning method to the task of classifying sentiments within user comments on the PeduliLindungi application.

- Input Layer

    - The initial layer in the CNN algorithm, where input data is accommodated. Each word's vectorized representation, obtained through Word2Vec, is stored as a numerical value. The size of this layer is determined by the length of the vector and the number of words in the dataset.

- Convolutional Layer

    - This layer is integral for feature extraction from the input text. Parameters such as the kernel and stride play crucial roles. The kernel, a matrix moving over the input data, performs a dot product with sub-regions of the input data, producing a dot product matrix. The stride value dictates the kernel's movement based on the input data.

    - The convolutional layer captures hierarchical features within the text, identifying patterns that contribute to sentiment classification.

- Activation Layer (ReLU)

    - Following the convolutional layer, the activation layer receives the values of the feature maps and applies the Rectified Linear Unit (ReLU) activation function. ReLU introduces non-linearity to the model, enhancing its ability to capture complex relationships within the data.

    - The ReLU activation function transforms negative values to zero, introducing non-linearity while ensuring computational efficiency [52].

- Pooling Layer

    - The pooling layer, or downsampling step, is crucial for reducing the dimensions of the feature maps. Max pooling, employed in this study, selects the maximum value from the pool, distilling the most salient information and expediting computation.

    - Pooling mitigates overfitting and enhances the model's ability to generalize by reducing the parameters processed.

- Fully Connected Layer

    - Neurons in this layer are fully connected to all activations in the previous layer, facilitating data processing for classification. This layer is instrumental in the multi-layer projection method.

    - The output of this layer represents the classification results, determined by the softmax activation function. Softmax calculates the probability of each target class for a given input, providing a comprehensive sentiment classification.

The CNN's architecture is adept at processing two-dimensional data, traditionally employed for image processing [53], [54] but successfully adapted for natural language processing (NLP) [55], [56]. In the context of sentiment analysis [57], CNN's efficacy lies in its weight-sharing feature, significantly reducing parameters and enhancing generalization while avoiding overfitting [58].

CNN's application to sentiment analysis in user comments on the PeduliLindungi application is characterized by its ability to discern intricate patterns and relationships within the text. The hierarchical feature extraction ensures that the model captures both local and global contextual information, contributing to a nuanced understanding of sentiment in diverse linguistic expressions. The CNN's robust architecture, as detailed in this section, positions it as a valuable asset in the sentiment analysis pipeline, ensuring accurate and reliable classification of user sentiments.

3. Results and Discussion

The analysis of sentiment classification results and model performance involves a detailed exploration of key aspects, encompassing both the outcomes of the testing scenarios and the subsequent evaluation metrics.

3.1. Sentiment Classification Result

The evaluation of sentiment classification results is pivotal in understanding the nuances within the PeduliLindungi comments dataset. The detailed analysis provides insights into the prevalence and distribution of sentiments, shedding light on the impact of preprocessing techniques and methods employed.

Table 2 illuminates the distribution of sentiments under different preprocessing scenarios and methods. Notably, negative sentiments exhibit dominance, constituting 50% to 60% of the total data. This skew towards negativity suggests a prevailing trend of dissatisfaction or critique within the user comments on the PeduliLindungi application. The distinct scenarios, considering both negation handling and the choice between CBOW and Skip-Gram methods, offer a comprehensive view of how these factors influence sentiment composition.

Table 2. Sentiment Classification Result

Method	Preprocessing	Positive	Neutral	Negative	Total Data
CNN+CBOW	Negation Handling	1,080 (39.7%)	109 (4%)	1,525 (56.1%)	2714
CNN+CBOW	Non Negation	1,202 (44.2%)	111 (4.1%)	1,401 (51.2%)	2714
CNN+Skip gram	Negation Handling	1,048 (38.6%)	25 (1%)	1,641 (60.4%)	2714
CNN+Skip gram	Non Negation	1,146 (32.2%)	32 (1.2%)	1,536 (56.5%)	2714

The prevalence of negative sentiments suggests potential issues or areas of improvement within the PeduliLindungi application. Understanding the distribution of sentiments becomes crucial for developers and policymakers to address user concerns and enhance user satisfaction. The variation in sentiment composition across different preprocessing methods and scenarios underscores the sensitivity of sentiment analysis models to these factors.

Comparing scenarios with and without negation handling provides valuable insights into the impact of this preprocessing technique. Negation handling appears to influence the categorization of sentiments, particularly in scenarios using CNN+CBOW. The nuanced differences in sentiment composition reveal that negation handling contributes to a more balanced representation of sentiments, addressing challenges posed by negations in user comments.

The choice between CBOW and Skip-Gram methods also plays a role in sentiment classification outcomes. The differences in sentiment distribution between these methods highlight the importance of selecting an appropriate word embedding technique based on the characteristics of the dataset. The

nuanced exploration of sentiment classification results sets the stage for a deeper understanding of the factors influencing the performance of sentiment analysis models on PeduliLindungi comments.

3.2. Confusion Matrix Result

The utilization of a confusion matrix in evaluating sentiment analysis models offers a granular view of the model's performance, particularly in distinguishing between positive, neutral, and negative sentiments. The detailed analysis of the confusion matrix results provides insights into the model's strengths and areas for improvement.

The confusion matrix results, as illustrated in Table 3, showcase the model's proficiency in accurately predicting positive and negative sentiments across different preprocessing scenarios and methods. True Positive (TP) values significantly outweigh False Positive (FP) and False Negative (FN) values for both positive and negative classes. This indicates that the model demonstrates robust predictive capabilities when it comes to identifying user comments expressing positive or negative sentiments regarding the PeduliLindungi application.

Table 3. Confusion Matrix Results (Positive: P(+), Neutral: Net(0), and Negative: N(-))

Method	Preprocessing	TP			FP			FN		
		P(+)	*Net(0)*	*N(-)*	*P(+)*	*Net(0)*	*N(-)*	*P(+)*	*Net(0)*	*N(-)*
CNN+CBOW	Negation Handling	764	26	1242	316	83	283	124	349	764
CNN+CBOW	Non Negation	781	26	1150	421	85	251	107	349	781
CNN+Skip gram	Negation Handling	758	5	1306	290	20	335	130	370	758
CNN+Skip gram	Non Negation	770	6	1224	376	26	312	118	369	770

However, challenges arise in the classification of neutral sentiments. The CNN model, regardless of the preprocessing scenario, struggles to predict neutral sentiments accurately. This is evident from the lower True Positive values for the neutral class compared to False Positive and False Negative values. The model tends to misclassify neutral sentiments, possibly due to the scarcity of neutral data in the training set. This points to a potential area for improvement in enhancing the model's ability to discern and accurately categorize neutral sentiments.

Comparing the results between scenarios with and without negation handling reveals nuanced differences in the model's performance. In scenarios involving negation handling, the model demonstrates improved precision in predicting positive and negative sentiments, particularly in the CNN+CBOW method. This suggests that negation handling contributes to mitigating the impact of negations in user comments, leading to more accurate sentiment classification.

Differentiating between CNN+CBOW and CNN+Skip Gram methods unveils method-specific nuances in the confusion matrix results. For instance, the CNN+CBOW method exhibits higher True Positive values for positive sentiments but struggles more with negative sentiments, as indicated by higher False Positive and False Negative values in comparison to the CNN+Skip Gram method. Understanding these method-specific nuances is crucial for optimizing model performance based on the characteristics of the dataset.

The insights derived from the confusion matrix results underscore the need for targeted enhancements, especially in improving the model's ability to classify neutral sentiments accurately. Addressing this challenge may involve augmenting the training dataset with more diverse neutral sentiments or exploring alternative preprocessing techniques to better capture the subtleties of neutral expressions in user comments.

3.3. Testing Result

The testing results, as outlined in Table 4, offer a comprehensive evaluation of the sentiment analysis models' performance across different preprocessing scenarios and methods. The detailed analysis of accuracy, precision, recall, and F1-score provides nuanced insights into the effectiveness of the models in categorizing sentiments within PeduliLindungi comments.

Table 4. Testing Result

Method	Preprocessing	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
CNN+CBOW	Negation Handling	74.9	70.7	74.9	71.5
CNN+CBOW	Non Negation	72.1	65	72.1	69
CNN+Skip gram	Negation Handling	76.2	72.3	76.2	71.3
CNN+Skip gram	Non Negation	73.7	67.2	73.7	69

Accuracy serves as a fundamental metric, reflecting the overall correctness of sentiment predictions. The comparison between scenarios with and without negation handling reveals a consistent improvement in accuracy when negation handling is employed. Specifically, both CNN+CBOW and CNN+Skip Gram methods exhibit an increase in accuracy by 3% when negation handling is integrated into the preprocessing stage. This underscores the positive impact of negation handling on the overall correctness of sentiment classifications.

Precision, recall, and F1-score provide a more nuanced understanding of model performance, especially concerning positive, neutral, and negative sentiments.

- Positive Sentiments.

    - In the CNN+CBOW method, negation handling leads to higher precision (70.7%) and recall (74.9%) for positive sentiments, resulting in an improved F1-score of 71.5%. Non-negation scenarios, while slightly lower in precision (65%), maintain a comparable F1-score (69%).

    - For the CNN+Skip Gram method, the integration of negation handling again enhances precision (72.3%) and recall (76.2%), contributing to an elevated F1-score of 71.3%. Non-negation scenarios exhibit slightly lower precision (67.2%) but maintain a competitive F1-score (69%).

- Neutral Sentiments.

    - The challenges in accurately predicting neutral sentiments are evident, with both methods and preprocessing scenarios displaying lower precision, recall, and F1-scores for neutral sentiments. This indicates a potential area for improvement, possibly through increased neutral data representation in the training set.

- Negative Sentiments

- Negation handling consistently contributes to improved precision, recall, and F1-score for negative sentiments in both CNN+CBOW and CNN+Skip Gram methods. The CNN+Skip Gram method, with negation handling, demonstrates the highest F1-score at 76.2%.

Comparing the performance of CNN+CBOW and CNN+Skip Gram methods reveals nuanced differences. The CNN+Skip Gram method consistently outperforms CNN+CBOW across all metrics and scenarios. The integration of negation handling consistently enhances the results, indicating its general applicability in improving sentiment analysis outcomes.

The testing results underscore the importance of considering both negation handling and method selection in optimizing sentiment analysis models. While negation handling contributes to improved accuracy and precision, method-specific characteristics influence the overall model performance. Addressing challenges in predicting neutral sentiments remains a crucial focus for model optimization.

3.4. Performance Enhancement with Negation Handling

The integration of negation handling significantly enhances the performance of sentiment analysis models, as observed through a thorough examination of key metrics and scenarios (shown in Fig. 1). The positive impact of negation handling is evident across both CNN+CBOW and CNN+Skip Gram methods, particularly in terms of accuracy, precision, recall, and F1-score. The accuracy improvements are pronounced, with the CNN+CBOW method showcasing a 2.8 p.p. boost (from 72.1% to 74.9%) and the CNN+Skip Gram method experiencing a 2.5 p.p. increase (from 73.7% to 76.2%) when negation handling is incorporated.
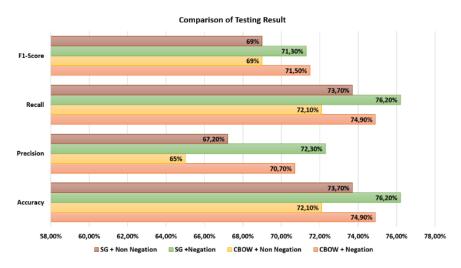

Fig. 1. Comparison of the testing results

The precision and recall optimizations further underscore the efficacy of negation handling in refining sentiment classification. For positive sentiments, both methods witness substantial improvements in precision and recall, leading to elevated F1-scores. However, challenges persist in accurately predicting neutral sentiments, emphasizing the need for additional strategies to enhance the model's performance in this specific category. In the context of negative sentiments, negation handling consistently contributes to improved precision, recall, and F1-score, with the CNN+Skip Gram method achieving the highest F1-score at 76.2%.

The comparative analysis between CNN+CBOW and CNN+Skip Gram methods highlights method-specific impacts, with the latter consistently outperforming the former across all metrics and scenarios.

These findings not only emphasize the importance of negation handling in refining sentiment predictions but also underscore the need for thoughtful method selection based on dataset characteristics. The challenges associated with neutral sentiment classification suggest avenues for further model optimization to ensure robust performance in real-world application.

4. Conclusion

This study significantly contributes to our comprehension of sentiment analysis for PeduliLindungi comments, emphasizing the pivotal roles of negation handling and method selection. The integration of negation handling proves transformative, notably enhancing key metrics, including a substantial (about 3 p.p.) increase in accuracy, about 2 p.p. increase in precision, about 3 p.p. increase in recall, and about 3 p.p. increase in F1 score. The CNN-Word2Vec Skip Gram method, coupled with negation handling, emerges as the standout performer, achieving impressive scores of 76.2% accuracy, 72.3% precision, 76.2% recall, and a notable F1 score of 71.13%.

However, this study acknowledges specific shortcomings, particularly in accuracy and system effectiveness. In light of these limitations, recommendations for system development are proposed. Enhancing accuracy could be achieved through refining the CNN architecture via hyperparameter tuning. Introducing aspect-based sentiment analysis offers a more comprehensive evaluation by identifying specific application aspects. To tackle data imbalances, employing a dataset with balanced labels is advised, and real-time retrieval of Google Play comments can further bolster the effectiveness of sentiment analysis. These recommendations are geared towards refining the system for more robust and accurate evaluations, extending its applicability beyond PeduliLindungi to a broader scope.

References

[1]   N. Naqvi and A. Saikia, "Lessons learned on building trust during a global pandemic: looking at future directions," *J. Commun. Healthc.*, pp. 1–4, Oct. 2023, doi: 10.1080/17538068.2023.2274198.

[2]   L. Stanca, D.-C. Dabija, and V. Câmpian, "Qualitative analysis of customer behavior in the retail industry during the COVID-19 pandemic: A word-cloud and sentiment analysis approach," *J. Retail. Consum. Serv.*, vol. 75, p. 103543, Nov. 2023, doi: 10.1016/j.jretconser.2023.103543.

[3]   S. Saifullah, Y. Fauziah, and A. S. Aribowo, "Comparison of Machine Learning for Sentiment Analysis in Detecting Anxiety Based on Social Media Data," Jan. 2021, doi: 10.26555/jifo.v15i1.a20111.

[4]   M. V. Wirastri, N. Morrison, and G. Paine, "The connection between slums and COVID-19 cases in Jakarta, Indonesia: A case study of Kapuk Urban Village," *Habitat Int.*, vol. 134, p. 102765, Apr. 2023, doi: 10.1016/j.habitatint.2023.102765.

[5]   J. Verma and A. S. Mishra, "COVID-19 infection: Disease detection and mobile technology," *PeerJ*, vol. 8, p. e10345, Nov. 2020, doi: 10.7717/peerj.10345.

[6]   R. A. Hidayat, G. Hana Zafira, N. Rahmawati Indah Nurfitriani, and A. Alfitriya Syahida, "Digital Healthcare Development for Global Citizenship: Equality of Access to Health Facilities and Services During the COVID-19 Pandemic in Indonesia," *KnE Soc. Sci.*, Feb. 2023, doi: 10.18502/kss.v8i3.12813.

[7]   Y. Mahendradhata, T. Lestari, and R. Djalante, "Strengthening government's response to COVID-19 in Indonesia: A modified Delphi study of medical and health academics," *PLoS One*, vol. 17, no. 9, p. e0275153, Sep. 2022, doi: 10.1371/journal.pone.0275153.

[8]   M. I. Wijaya *et al.*, "A Qualitative Study on Barriers to Stunting Primordial Prevention during the PentaCOME Project," *Open Access Maced. J. Med. Sci.*, vol. 11, no. E, pp. 152–161, Feb. 2023, doi: 10.3889/oamjms.2023.11289.

[9]   S. H.-W. Chuah and J. Yu, "The future of service: The power of emotion in human-robot interaction," *J. Retail. Consum. Serv.*, vol. 61, p. 102551, Jul. 2021, doi: 10.1016/j.jretconser.2021.102551.

[10] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Syst.*, vol. 226, p. 107134, Aug. 2021, doi: 10.1016/j.knosys.2021.107134.

[11] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," *Artif. Intell. Rev.*, vol. 53, no. 6, pp. 4335–4385, Aug. 2020, doi: 10.1007/s10462-019-09794-5.

[12] S. Saifullah, R. Drezewski, F. A. Dwiyanto, A. S. Aribowo, Y. Fauziah, and N. H. Cahyana, "Automated Text Annotation Using Semi-Supervised Approach with Meta Vectorizer and Machine Learning Algorithms for Hate Speech Detection," *Preprints*, p. 2023110963, 2023, doi: 10.20944/preprints202311.0963.v1.

[13] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, "Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN," *Ilk. J. Ilm.*, vol. 14, no. 3, pp. 348–354, 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.

[14] P. Mukherjee, Y. Badr, S. Doppalapudi, S. M. Srinivasan, R. S. Sangwan, and R. Sharma, "Effect of Negation in Sentences on Sentiment Analysis and Polarity Detection," *Procedia Comput. Sci.*, vol. 185, pp. 370–379, 2021, doi: 10.1016/j.procs.2021.05.038.

[15] Y. S. Mehanna and M. Mahmuddin, "The Effect of Pre-processing Techniques on the Accuracy of Sentiment Analysis Using Bag-of-Concepts Text Representation," *SN Comput. Sci.*, vol. 2, no. 4, p. 237, Jul. 2021, doi: 10.1007/s42979-021-00453-7.

[16] T. Kolajo, O. Daramola, A. Adebiyi, and A. Seth, "A framework for pre-processing of social media feeds based on integrated local knowledge base," *Inf. Process. Manag.*, vol. 57, no. 6, p. 102348, Nov. 2020, doi: 10.1016/j.ipm.2020.102348.

[17] M. T. Ari Bangsa, S. Priyanta, and Y. Suyanto, "Aspect-Based Sentiment Analysis of Online Marketplace Reviews Using Convolutional Neural Network," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 14, no. 2, p. 123, Apr. 2020, doi: 10.22146/ijccs.51646.

[18] U. Lal and P. Kamath, "Effective Negation Handling Approach for Sentiment Classification using synsets in the WordNet lexical database," in *2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, Feb. 2022, pp. 01–07, doi: 10.1109/ICEEICT53079.2022.9768641.

[19] G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng, and X. Wu, "Chinese Text Sentiment Analysis Based on Extended Sentiment Dictionary," *IEEE Access*, vol. 7, pp. 43749–43762, 2019, doi: 10.1109/ACCESS.2019.2907772.

[20] N. Punetha and G. Jain, "Optimizing Sentiment Analysis: A Cognitive Approach with Negation Handling via Mathematical Modelling," *Cognit. Comput.*, Nov. 2023, doi: 10.1007/s12559-023-10227-3.

[21] S. Jahić and J. Vičič, "Impact of Negation and AnA-Words on Overall Sentiment Value of the Text Written in the Bosnian Language," *Appl. Sci.*, vol. 13, no. 13, p. 7760, Jun. 2023, doi: 10.3390/app13137760.

[22] Y. Mehmood and V. Balakrishnan, "An enhanced lexicon-based approach for sentiment analysis: a case study on illegal immigration," *Online Inf. Rev.*, vol. 44, no. 5, pp. 1097–1117, Jun. 2020, doi: 10.1108/OIR-10-2018-0295.

[23] M. Neumann and R. Linzmayer, "Capturing Student Feedback and Emotions in Large Computing Courses: A Sentiment Analysis Approach," in *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, Mar. 2021, pp. 541–547, doi: 10.1145/3408877.3432403.

[24] R. I. Kurnia, "Classification of User Comment Using Word2vec and SVM Classifier," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 1, pp. 643–648, Feb. 2020, doi: 10.30534/ijatcse/2020/90912020.

[25] M. E. Basiri, M. Abdar, M. A. Cifci, S. Nemati, and U. R. Acharya, "A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques," *Knowledge-Based Syst.*, vol. 198, p. 105949, Jun. 2020, doi: 10.1016/j.knosys.2020.105949.

[26] P. K. Jain, R. Pamula, and G. Srivastava, "A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews," *Comput. Sci. Rev.*, vol. 41, p. 100413, Aug. 2021, doi: 10.1016/j.cosrev.2021.100413.

[27] K. Paramesha, H. L. Gururaj, A. Nayyar, and K. C. Ravishankar, "Sentiment analysis on cross-domain textual data using classical and deep learning approaches," *Multimed. Tools Appl.*, vol. 82, no. 20, pp. 30759–30782, Aug. 2023, doi: 10.1007/s11042-023-14427-9.

[28] E. M. Mercha and H. Benbrahim, "Machine learning and deep learning for sentiment analysis across languages: A survey," *Neurocomputing*, vol. 531, pp. 195–216, Apr. 2023, doi: 10.1016/j.neucom.2023.02.015.

[29] T. Cui, N. Du, X. Yang, and S. Ding, "Multi-period portfolio optimization using a deep reinforcement learning hyper-heuristic approach," *Technol. Forecast. Soc. Change*, vol. 198, p. 122944, Jan. 2024, doi: 10.1016/j.techfore.2023.122944.

[30] N. S. Alharbi, H. Jahanshahi, Q. Yao, S. Bekiros, and I. Moroz, "Enhanced Classification of Heartbeat Electrocardiogram Signals Using a Long Short-Term Memory–Convolutional Neural Network Ensemble: Paving the Way for Preventive Healthcare," *Mathematics*, vol. 11, no. 18, p. 3942, Sep. 2023, doi: 10.3390/math11183942.

[31] S. N. Kigo, E. O. Omondi, and B. O. Omolo, "Assessing predictive performance of supervised machine learning algorithms for a diamond pricing model," *Sci. Rep.*, vol. 13, no. 1, p. 17315, Oct. 2023, doi: 10.1038/s41598-023-44326-w.

[32] A. Gandhi, K. Adhvaryu, S. Poria, E. Cambria, and A. Hussain, "Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions," *Inf. Fusion*, vol. 91, pp. 424–444, Mar. 2023, doi: 10.1016/j.inffus.2022.09.025.

[33] M. S. Başarslan and F. Kayaalp, "Sentiment analysis using a deep ensemble learning model," *Multimed. Tools Appl.*, Oct. 2023, doi: 10.1007/s11042-023-17278-6.

[34] M. Bordoloi and S. K. Biswas, "Sentiment analysis: A survey on design framework, applications and future scopes," *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 12505–12560, Nov. 2023, doi: 10.1007/s10462-023-10442-2.

[35] S. M. Sherif *et al.*, "Lexicon annotation in sentiment analysis for dialectal Arabic: Systematic review of current trends and future directions," *Inf. Process. Manag.*, vol. 60, no. 5, p. 103449, Sep. 2023, doi: 10.1016/j.ipm.2023.103449.

[36] G. Fatouros, J. Soldatos, K. Kouroumali, G. Makridis, and D. Kyriazis, "Transforming sentiment analysis in the financial domain with ChatGPT," *Mach. Learn. with Appl.*, vol. 14, p. 100508, Dec. 2023, doi: 10.1016/j.mlwa.2023.100508.

[37] M. Y. Khan, A. Qayoom, M. S. Nizami, M. S. Siddiqui, S. Wasi, and S. M. K.-R. Raazi, "Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep Learning Techniques," *Complexity*, vol. 2021, pp. 1–18, Sep. 2021, doi: 10.1155/2021/2553199.

[38] M. Giatsoglou, M. G. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, and K. C. Chatzisavvas, "Sentiment analysis leveraging emotions and word embeddings," *Expert Syst. Appl.*, vol. 69, pp. 214–224, Mar. 2017, doi: 10.1016/j.eswa.2016.10.043.

[39] K. L. Tan, C. P. Lee, K. M. Lim, and K. S. M. Anbananthen, "Sentiment Analysis With Ensemble Hybrid Deep Learning Model," *IEEE Access*, vol. 10, pp. 103694–103704, 2022, doi: 10.1109/ACCESS.2022.3210182.

[40] R. Kaushal and R. Chadha, "A Survey of Various Sentiment Analysis Techniques of Whatsapp," in *2023 2nd International Conference for Innovation in Technology (INOCON)*, Mar. 2023, pp. 1–6, doi: 10.1109/INOCON57975.2023.10101105.

[41] H. Jayadianti, B. Santosa, J. Cahyaning, S. Saifullah, and R. Drezewski, "Essay auto-scoring using N-Gram and Jaro Winkler based Indonesian Typos," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 22, no. 2, pp. 325–338, Mar. 2023, doi: 10.30812/matrik.v22i2.2473.

[42] C. P. Chai, "Comparison of text preprocessing methods," *Nat. Lang. Eng.*, vol. 29, no. 3, pp. 509–553, May 2023, doi: 10.1017/S1351324922000213.

[43] K. S. Eljil, F. Nait-Abdesselam, E. Hamouda, and M. Hamdi, "Enhancing Sentiment Analysis on Social Media with Novel Preprocessing Techniques," *J. Adv. Inf. Technol.*, vol. 14, no. 6, pp. 1206–1213, 2023, doi: 10.12720/jait.14.6.1206-1213.

[44] A. M. Ningtyas and G. B. Herwanto, "The Influence of Negation Handling on Sentiment Analysis in Bahasa Indonesia," *Proc. 2018 5th Int. Conf. Data Softw. Eng. ICoDSE 2018*, pp. 1–6, 2018, doi: 10.1109/ICODSE.2018.8705802.

[45] S. Saifullah, R. Dreżewski, F. A. Dwiyanto, A. S. Aribowo, and Y. Fauziah, "Sentiment Analysis Using Machine Learning Approach Based on Feature Extraction for Anxiety Detection," in *Computational Science – ICCS 2023: 23rd International Conference, Prague, Czech Republic, July 3–5, 2023, Proceedings, Part II*, Berlin, Heidelberg: Springer-Verlag, 2023, pp. 365–372, doi: 10.1007/978-3-031-36021-3_38.

[46] H. Xia, "Continuous-bag-of-words and Skip-gram for word vector training and text classification," *J. Phys. Conf. Ser.*, vol. 2634, no. 1, p. 012052, Nov. 2023, doi: 10.1088/1742-6596/2634/1/012052.

[47] F. Xiao, S. Yu, and Y. Li, "Efficient Large-Capacity Caching in Cloud Storage Using Skip-Gram-Based File Correlation Analysis," *IEEE Access*, vol. 11, pp. 111265–111273, 2023, doi: 10.1109/ACCESS.2023.3322725.

[48] S. J. Johnson, M. R. Murty, and I. Navakanth, "A detailed review on word embedding techniques with emphasis on word2vec," *Multimed. Tools Appl.*, Oct. 2023, doi: 10.1007/s11042-023-17007-z.

[49] F. Incitti, F. Urli, and L. Snidaro, "Beyond word embeddings: A survey," *Inf. Fusion*, vol. 89, pp. 418–436, Jan. 2023, doi: 10.1016/j.inffus.2022.08.024.

[50] B. Jang, I. Kim, and J. W. Kim, "Word2vec convolutional neural networks for classification of news articles and tweets," *PLoS One*, vol. 14, no. 8, p. e0220976, Aug. 2019, doi: 10.1371/journal.pone.0220976.

[51] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.

[52] S. Saifullah, B. Yuwono, H. C. Rustamaji, B. Saputra, F. A. Dwiyanto, and R. Drezewski, "Detection of Chest X-ray Abnormalities Using CNN Based on Hyperparameters Optimization," *Eng. Proc.*, vol. 52, pp. 1–7, 2023. [Online]. Available at: https://sciforum.net/paper/view/16260.

[53] S. Saifullah *et al.*, "Nondestructive chicken egg fertility detection using CNN-transfer learning algorithms," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 9, no. 3, pp. 854–871, 2023. [Online]. Available at: https://arxiv.org/abs/2309.16257.

[54] S. Saifullah and R. Dreżewski, "Enhanced Medical Image Segmentation using CNN based on Histogram Equalization," *2023 2nd Int. Conf. Appl. Artif. Intell. Comput.*, pp. 121–126, May 2023, doi: 10.1109/ICAAIC56838.2023.10141065.

[55] P. Y. Mahajan and D. P. Rana, "Text mining approach for the prediction of disease status from discharge summaries using CCBE and NEROA-CNN," *Expert Syst. Appl.*, vol. 227, p. 120310, Oct. 2023, doi: 10.1016/j.eswa.2023.120310.

[56] M. Umer *et al.*, "Impact of convolutional neural network and FastText embedding on text classification," *Multimed. Tools Appl.*, vol. 82, no. 4, pp. 5569–5585, Feb. 2023, doi: 10.1007/s11042-022-13459-x.

[57] H. I. Abdalla, A. A. Amer, and S. D. Ravana, "BoW-based neural networks vs. cutting-edge models for single-label text classification," *Neural Comput. Appl.*, vol. 35, no. 27, pp. 20103–20116, Sep. 2023, doi: 10.1007/s00521-023-08754-z.

[58] S. Balasubramaniam, Y. Velmurugan, D. Jaganathan, and S. Dhanasekaran, "A Modified LeNet CNN for Breast Cancer Diagnosis in Ultrasound Images," *Diagnostics*, vol. 13, no. 17, p. 2746, Aug. 2023, doi: 10.3390/diagnostics13172746.