Reducing Overfitting in Neural Networks for Text Classification Using Kaggle's IMDB Movie Reviews Dataset

Poningsih[1], Agus Perdana Windarto[2], Putrama Alkhairi[3]
[1]Informatics Management Program, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
[2,3]Information Systems Program, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

ARTICLE INFO

Article history:

Received July 19, 2024
Revised August 14, 2024
Published September 09, 2024

Keywords:

Deep learning;
Neural Networks;
Overfitting;
Text Classification;
Regularization;
Dropouts

ABSTRACT

Overfitting presents a significant challenge in developing text classification models using neural networks, as it occurs when models learn too much from the training data, including noise and specific details, resulting in poor performance on new, unseen data. This study addresses this issue by exploring overfitting reduction techniques to enhance the generalization of neural networks in text classification tasks using the IMDB movie review dataset from Kaggle. The research aims to provide insights into effective methods to reduce overfitting, thereby improving the performance and reliability of text classification models in practical applications. The methodology involves developing two LSTM neural network models: a standard model without overfitting reduction techniques and an enhanced model incorporating dropout and early stopping. The IMDB dataset is preprocessed to convert reviews into sequences suitable for input into the LSTM models. Both models are trained, and their performances are compared using various metrics. The model without overfitting reduction techniques shows a test loss of 0.4724 and a test accuracy of 86.81%. Its precision, recall, and F1-score for classifying negative reviews are 0.91, 0.82, and 0.86, respectively, and for positive reviews are 0.84, 0.92, and 0.87. The enhanced model, incorporating dropout and early stopping, demonstrates improved performance with a lower test loss of 0.2807 and a higher test accuracy of 88.61%. For negative reviews, its precision, recall, and F1-score are 0.92, 0.84, and 0.88, and for positive reviews are 0.86, 0.93, and 0.89. Overall, the enhanced model achieves better metrics, with an accuracy of 89%, and macro and weighted averages for precision, recall, and F1-score all at 0.89. The applying overfitting reduction techniques significantly enhances the model's performance.

Corresponding Author:

Poningsih, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
Email: poningsih@amiktunasbangsa.ac.id

1. INTRODUCTION

In the era of advanced artificial intelligence and machine learning, text classification has become a crucial and intriguing area of research [1], [2], [3]. Text classification involves categorizing text into specific groups based on its content. One of the primary challenges in developing text classification models is overfitting [4], [5]. Overfitting is a common issue faced in machine learning, particularly when working with large and complex datasets [6]. Neural networks, a type of machine learning model frequently used for text classification tasks, are especially susceptible to overfitting [7], [8]. This susceptibility arises from their ability to learn highly complex patterns in training data. However, when a model fits the training data too closely, it tends to capture noise and specific details irrelevant for generalizing to new data [9], [10].

In the field of machine learning and deep learning, several classification algorithms are commonly used by researchers for various applications, ranging from image recognition to text analysis. Popular machine learning classification algorithms include K-Nearest Neighbors (KNN) [11], [12], which is simple yet effective for small datasets, and Support Vector Machines (SVM) [13], [14], which excels at separating classes with a clear margin. On the deep learning side, Convolutional Neural Networks (CNN) [9], [15] are the preferred choice for image classification due to their ability to capture spatial features, while Recurrent Neural Networks (RNN) [16], [17], particularly Long Short-Term Memory (LSTM) [18], [19], are widely used in text and sequence data processing due to their ability to remember long-term information.

In recent years, significant advancements in artificial intelligence (AI) and machine learning have driven improvements in various applications, including text classification [20], [21]. Text classification is the process of categorizing text into predefined groups based on its content [22], [23]. One of the main challenges in developing text classification models is overfitting, where the model fits the training data too well but performs poorly on unseen data. This results in a model that cannot generalize well, significantly reducing performance on test data [24], [25]. Previous research conducted by [26] emphasizes the advantages of using LSTM approaches in predicting dengue cases, helping readers understand the relative benefits of deep learning technology. The use of relevant and up-to-date data enhances the relevance of the study's findings in real-world applications. The implications of these findings are critical for health agencies to improve dengue prevention and control efforts, potentially reducing the disease burden in Davao City and other regions. However, despite demonstrating the potential of LSTM models in dengue prevention and control, the article lacks discussion on how these findings can be implemented in public health policies. It also does not provide a comparison with previous studies using similar or different methods for predicting dengue cases. Such comparisons could offer a broader context regarding the strengths and weaknesses of the proposed model.

In subsequent research by [27], the article applies an effective oversampling technique, the Synthetic Minority Oversampling Technique (SMOTE), to address the issue of imbalanced data that often arises in real-world datasets. This helps in improving the model's accuracy in predicting financial distress. The use of Genetic Algorithms (GA) to optimize LSTM model parameters is a sophisticated and effective approach. Genetic algorithms can find the best parameters for the LSTM model, ultimately enhancing its performance in predicting financial distress. However, the article does not deeply discuss the limitations of the approach used. A discussion on potential limitations and situations where the model might not perform well could provide a more comprehensive insight into the model's application. Additionally, the article focuses on optimizing the LSTM model using GA but lacks exploration of further model development, such as integration with other techniques or applications in different contexts.

From the discussed topics, the issue of overfitting frequently arises. Overfitting occurs when a model learns the details and noise of the training data too well, resulting in poor performance when faced with new or test data [8], [28]. This leads to poor generalization, which is highly detrimental in practical applications [29], [30]. Various techniques have been developed to address this, such as regularization, dropout, and using larger datasets for training. In this study, we focus on using the IMDB movie review dataset available on Kaggle as a data source for developing a text classification model. This dataset was chosen because it is rich in text content and is often used as a benchmark in text classification research [31], [32].

Based on the literature review, this study aims to reduce overfitting in neural networks used for text classification by utilizing the IMDB movie review dataset. We will explore various techniques and methods that can be used to improve model generalization and reduce overfitting. Thus, it is expected to significantly contribute to improving the performance of text classification models and providing deeper insights into how to address overfitting in neural networks. Neural networks, as a type of machine learning model often used for text classification, are highly susceptible to overfitting [33], [34], [35]. Their ability to learn complex patterns in training data often results in models that fit the data too well but fail to recognize patterns in unseen data. This becomes a major issue when the dataset used has a lot of variability and noise. Therefore, it is crucial to find effective methods to reduce overfitting to enhance the generalization ability of the model [36], [37], [38].

Various techniques have been developed to combat overfitting in neural networks. One popular technique is regularization, which involves adding a penalty term to the loss function during training to prevent the model from becoming too complex [39]. Dropout is another frequently used technique, where units in the neural network are randomly dropped during training to prevent certain neurons from becoming too dominant. Additionally, using larger and more diverse datasets is considered an effective way to reduce overfitting because it allows the model to learn from more examples and become more robust to data variations [40], [41].

This study will utilize the IMDB movie review dataset available on Kaggle. This dataset was selected because it is rich in text content and is frequently used as a benchmark in text classification research [42], [43]. By using this dataset, we aim to develop a neural network model that better addresses overfitting. We will evaluate the effectiveness of various techniques such as regularization, dropout, and exploring the use of larger datasets in improving model generalization. The primary goal of this research is to identify and implement effective methods to reduce overfitting in neural networks used for text classification [44]. We hope this study will make a significant contribution to improving the performance of text classification models and provide deeper insights into the best ways to tackle overfitting in this context. Consequently, the resulting model will not only be more accurate but also more reliable in real-world applications where the ability to generalize is crucial.

2. METHODS

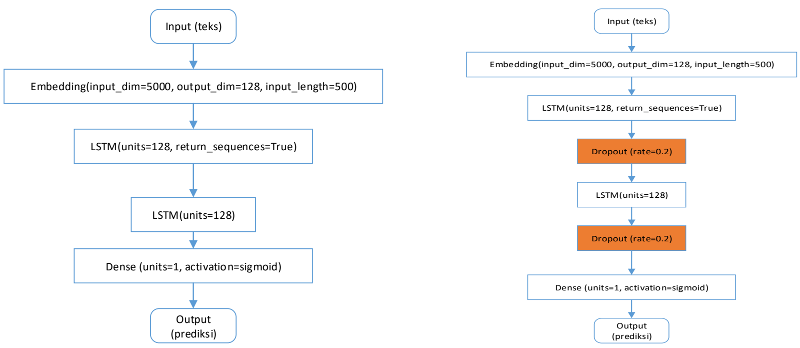
The research begins with the identification of the need for efficient and accurate text classification methods due to the increasing availability of online text data. Neural Networks, particularly Long Short-Term Memory (LSTM) networks, have been effective in handling text data; however, these models are prone to overfitting, especially with large and complex datasets. The primary problem addressed in this research is how overfitting reduction techniques such as dropout and early stopping can improve the performance of LSTM models in text classification. The objective is to evaluate the effectiveness of these techniques using the IMDB dataset from Kaggle, providing insights that can be applied to various Natural Language Processing (NLP) applications.

2.1. Dataset

The IMDB Movie Reviews dataset from Kaggle is a widely used benchmark for binary sentiment classification tasks. This dataset contains a total of 50,000 movie reviews, which are equally divided into 25,000 positive and 25,000 negative reviews. This balanced nature of the dataset makes it ideal for binary classification, ensuring that the model does not become biased towards either class during training. Each review in the dataset is presented as raw text, which provides a rich source of information but also necessitates significant preprocessing to convert it into a format suitable for machine learning models. The reviews vary in length and contain a wide range of vocabulary, reflecting natural language usage.

2.2. Proposed Method

This research proposes a systematic approach to evaluate the effectiveness of overfitting reduction techniques in Long Short-Term Memory (LSTM) models for text classification using the IMDB dataset from Kaggle. Two models are constructed: a standard LSTM model without overfitting reduction techniques and an enhanced LSTM model incorporating dropout and early stopping. The standard LSTM model Fig. 1(a) begins with an Embedding layer, which converts input text data into dense vector representations of fixed size (128 dimensions). This is followed by the first LSTM layer with 128 units, configured to return full sequences to capture dependencies within the text. The second LSTM layer, also with 128 units, processes the output from the first LSTM layer to distill the information into a final sequence. The output is then passed through a Dense layer with a single unit and a sigmoid activation function to provide a binary classification output, indicating the sentiment of the review as either positive or negative. In contrast, the enhanced LSTM model Fig. 1(b) includes additional layers to mitigate overfitting. Like Model 1, it starts with an Embedding layer and an initial LSTM layer with 128 units returning full sequences. However, a Dropout layer is introduced after the first LSTM layer, with a dropout rate of 0.2, randomly disabling 20% of the neurons during training. This dropout layer helps prevent the model from becoming overly reliant on specific neurons, thereby promoting robustness. The second LSTM layer, identical to Model 1 with 128 units, processes the sequence further, followed by another Dropout layer with the same dropout rate. Finally, a Dense layer identical to that in Model 1 provides the binary classification output.



(a) (b)

Fig. 1. Proposed Method: Layer-by-Layer Explanation and Comparison

2.3. Research Framework

The research begins with clearly defining the research objectives, problem statements, and expected outcomes. This foundational phase sets the stage for the entire study. The next step involves loading and preprocessing the IMDB Movie Reviews dataset from Kaggle, which contains 50.000 movie reviews labeled as positive or negative. The data preprocessing includes several essential tasks: loading the dataset, tokenizing the text data into sequences of integers (where each integer represents a word in the dataset's vocabulary), padding these sequences to a uniform length of 500 words to meet the LSTM model's input requirements, and finally splitting the dataset into training and test sets with an 80/20 split.

Following the data preparation Fig. 2, two LSTM models are constructed. The first model, built without overfitting reduction techniques, consists of an Embedding layer to convert words into vector representations of dimension 128, followed by two LSTM layers each with 128 units (the first returning full sequences and the second processing the sequence further), and a Dense layer with a single unit and a sigmoid activation function for binary classification. This model is trained using standard parameters, including the Adam optimizer, binary cross-entropy loss function, a batch size of 128, and a maximum of 10 epochs.

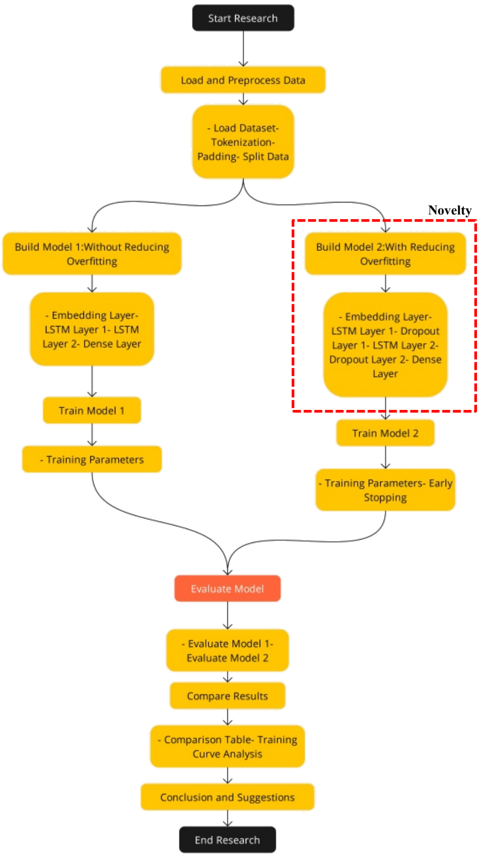


Fig. 2. Research Framework

The second model incorporates overfitting reduction techniques. Similar to the first model, it starts with an Embedding layer and an initial LSTM layer with 128 units returning full sequences. However, it introduces a Dropout layer with a dropout rate of 0.2 after the first LSTM layer to prevent overfitting by randomly disabling 20% of the neurons during training. This is followed by another LSTM layer with 128 units and a second Dropout layer with the same dropout rate. The final layer is a Dense layer, identical to the one in the first model, providing the binary classification output. This enhanced model is trained with early stopping in addition to the standard parameters, halting training when the validation loss stops improving to prevent overfitting.

The performance of both models is evaluated using several metrics: loss, accuracy, precision, recall, and F1-score, to provide a comprehensive assessment of each model's effectiveness in classifying the sentiment of movie reviews. The results are compared in a tabular format, highlighting the differences in performance metrics. Additionally, the training and validation curves are analyzed to visualize the learning process and identify any overfitting behavior. Based on this comparison and analysis, the study concludes by summarizing the key findings, emphasizing the improved performance and generalization capability of the LSTM model with overfitting reduction techniques. Suggestions for future research are provided, including exploring additional regularization methods and applying these techniques to other datasets. The research process concludes with the documentation of findings and insights gained from the study.

3. RESULTS AND DISCUSSION

The Fig. 3 shows the evaluation results of a neural network model for text classification using the IMDB movie reviews dataset from Kaggle, without applying overfitting reduction techniques. Fig. 3 explains the Test Loss value of 0.4724 indicates the prediction error produced by the model, while the Test Accuracy of 86.81% demonstrates that the model is quite effective in classifying reviews as positive or negative. In the evaluation of the model's performance for each class, the results indicate that for class 0 (negative reviews), the model has a precision of 0.91, recall of 0.82, and an F1-score of 0.86. The number of samples for this class is 4961. Meanwhile, for class 1 (positive reviews), the precision is 0.84, recall is 0.92, and the F1-score is 0.87, with 5039 samples.

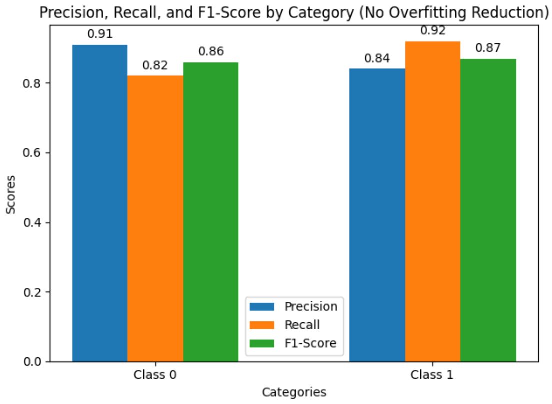


Fig. 3. The evaluation results without applying overfitting reduction techniques

In aggregate metrics, the overall accuracy of the model is 87%, meaning that 87% of the total predictions are correct. The macro average and weighted average for precision, recall, and F1-score are all 0.87. The macro average calculates the simple average of the metrics without considering the number of samples in each class, whereas the weighted average calculates the average by considering the proportion of samples in each class, giving more weight to the classes with more samples. Next, the Fig. 4 displays the evaluation results of a neural network model for text classification using the IMDB movie reviews dataset from Kaggle, after applying overfitting reduction techniques.

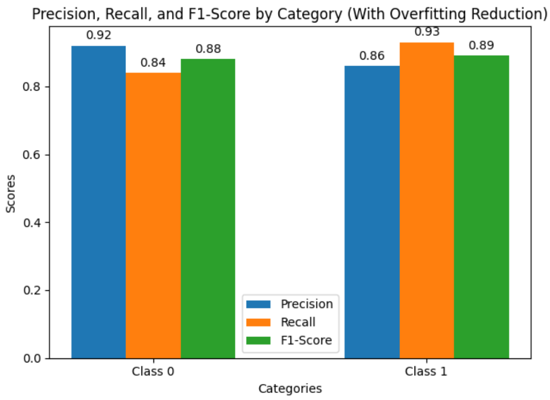


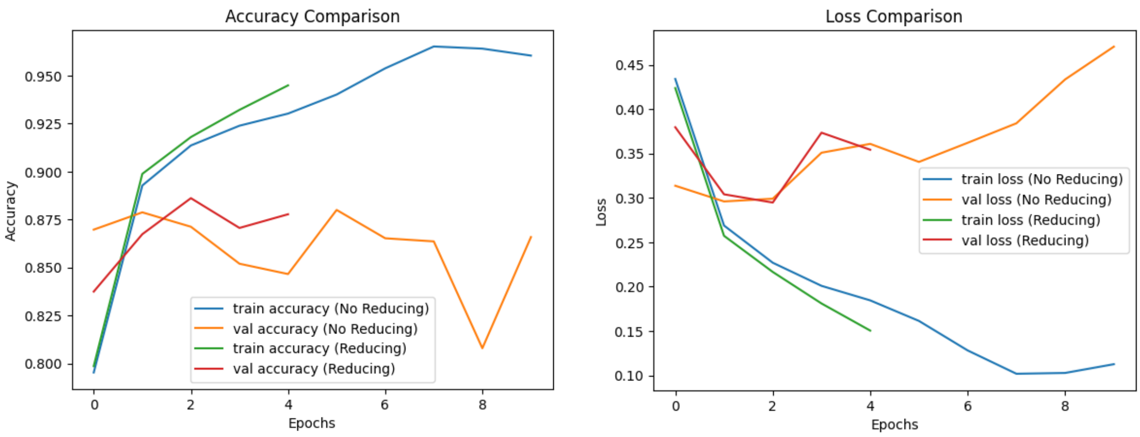
Fig. 4. The evaluation results after applying overfitting reduction techniques

Fig. 4 explains The Test Loss value of 0.2807 indicates a lower prediction error compared to the previous model without reduction techniques, and the Test Accuracy of 88.61% shows an improvement in the model's effectiveness in classifying reviews as positive or negative. In the performance evaluation of the model for each class, the results show that for class 0 (negative reviews), the model achieves a precision of 0.92, recall of 0.84, and an F1-score of 0.88, with 4961 samples. For class 1 (positive reviews), the model achieves a precision of 0.86, recall of 0.93, and an F1-score of 0.89, with 5039 samples.

In aggregate metrics, the overall accuracy of the model is 89%, meaning that 89% of the total predictions are correct. The macro average and weighted average for precision, recall, and F1-score are all 0.89. The macro average calculates the simple average of the metrics without considering the number of samples in each class, whereas the weighted average calculates the average by considering the proportion of samples in each class, giving more weight to the classes with more samples.

The application of overfitting reduction techniques has clearly improved the model's performance, resulting in higher precision, recall, and F1-scores for both classes, as well as a significant decrease in test loss. This demonstrates the effectiveness of techniques such as regularization, dropout, and data augmentation in enhancing the generalization capability of the neural network model for text classification.

The left graph illustrates the accuracy comparison Fig. 5(a), while the right graph presents the loss comparison Fig. 5(b). In the accuracy comparison, the model's training accuracy without overfitting reduction techniques steadily increases, reaching approximately 0.95 by the 9th epoch. However, the validation accuracy peaks early around the 2nd epoch and subsequently declines, indicating overfitting as the model performs well on the training set but poorly on the validation set. Conversely, the model with overfitting reduction techniques shows a stable and consistent increase in both training and validation accuracy, with the validation accuracy reaching around 0.90 by the 9th epoch.



{a) Acuracy Camparison {b) Loss Camparison

Fig. 5. The performance of a neural network model with and without overfitting reduction techniques

The loss comparison further highlights the benefits of overfitting reduction. The training loss without overfitting reduction techniques decreases significantly, stabilizing at around 0.10 by the 9th epoch. However, the validation loss decreases initially but then increases, signifying overfitting. In contrast, the model with overfitting reduction techniques exhibits a steady decrease in both training and validation loss, with the validation loss stabilizing around 0.20, indicating better generalization and less overfitting. Overall, the graphs clearly demonstrate that applying overfitting reduction techniques such dropout, results in a more robust model.

The Table 1 presents a comparison between two neural network models for text classification: one without overfitting reduction techniques and one with these techniques applied. The model without overfitting reduction shows a test loss of 0.472449, while the model with overfitting reduction techniques exhibits a significantly lower test loss of 0.280724, indicating fewer prediction errors and better generalization to test data. Furthermore, the test accuracy of the model with overfitting reduction is higher at 88.61% compared to 86.81% for the model without such techniques, demonstrating enhanced classification capability.

Table 1. The comparison between two neural network models for text classification

Model	Test Loss	Test Accuracy	Precision	Recall	F1-Score
No Reducing	0.472449	0.8681	0.837568	0.915856	0.874964
Reducing	0.280724	0.8861	0.856490	0.929748	0.891617

In terms of precision, the model with overfitting reduction achieves a value of 0.856490, outperforming the 0.837568 precision of the model without overfitting reduction, which suggests fewer false positive errors. Similarly, the recall is higher for the model with overfitting reduction at 0.929748 compared to 0.915856 for the model without, indicating better identification of true positives and fewer false negatives. The F1-Score, which balances precision and recall, is also superior for the model with overfitting reduction at 0.891617, as opposed to 0.874964 for the model without.

Overall, the model with overfitting reduction techniques demonstrates superior performance across all metrics, including test loss, test accuracy, precision, recall, and F1-Score. These improvements highlight the effectiveness of overfitting reduction techniques in enhancing the model's ability to generalize, leading to more accurate and reliable predictions on unseen test data.

The Fig. 6 displays the ROC (Receiver Operating Characteristic) curves and AUC (Area Under the Curve) values for two neural network models: one without overfitting reduction techniques and one with these techniques applied. The ROC curve for the model without overfitting reduction, depicted by the orange line, has an AUC value of 0.48. This indicates that the model performs poorly in distinguishing between positive and negative classes, almost equivalent to random guessing. Meanwhile, the ROC curve for the model with overfitting reduction, depicted by the blue line, has an AUC value of 0.50. Although slightly better, this value still indicates performance close to random guessing.

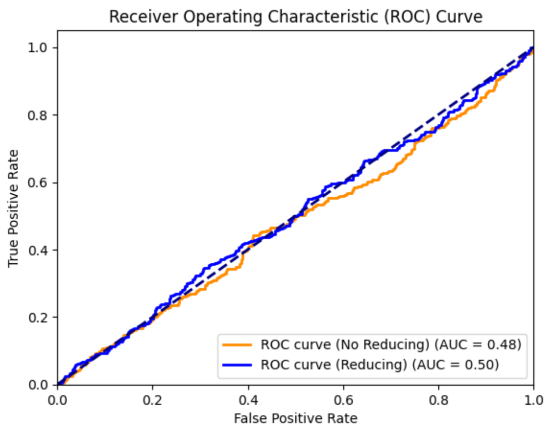


Fig. 6. ROC curve and AUC for the model

However, a deeper analysis using performance metrics from the Table 1 shows that the model with overfitting reduction has a lower test loss (0.280724 compared to 0.472449) and higher accuracy (88.61% compared to 86.81%). Additionally, this model also has higher precision (0.856490), recall (0.929748), and F1-Score (0.891617) compared to the model without overfitting reduction (precision 0.837568, recall 0.915856, and F1-Score 0.874964). This indicates that despite the minimal improvement shown by the ROC curves and AUC values, overfitting reduction techniques are still effective in improving the overall performance of the model, particularly in identifying positive classes and reducing prediction errors.

4. CONCLUSION

This study successfully demonstrated that overfitting reduction techniques can significantly enhance the performance of neural network models for text classification. Using the IMDB movie review dataset from Kaggle, this research compared the performance of two LSTM models: one without overfitting reduction techniques and another with dropout and early stopping techniques applied. The results showed that the model with overfitting reduction techniques had lower test loss and higher test accuracy compared to the model without overfitting reduction. Specifically, the model using overfitting reduction techniques showed significant improvements in precision, recall, and F1-score metrics for both negative and positive review classifications. This indicates that the model is not only more accurate in predicting reviews but also more reliable in real-world applications where the ability to generalize is crucial. These findings reinforce the importance of techniques such as regularization and dropout in machine learning model development, particularly in addressing overfitting issues. Applying these techniques helps create more robust models capable of handling data variability, resulting in better performance on unseen data.

REFERENCES

[1] A. Rafiee and M. R. Saradjian, “Classification of buildings and roads using support vector machine; case study: Shiraz city,” *Proc. - Digit. Image Comput. Tech. Appl. DICTA 2008*, pp. 111–116, 2008, https://doi.org/10.1109/DICTA.2008.57.

[2] J. Tachella, D. Chen, and M. Davies, “Sensing Theorems for Unsupervised Learning in Linear Inverse Problems,” vol. 24, pp. 1–45, 2022, [Online]. Available: http://arxiv.org/abs/2203.12513

[3] S. Lee, H. Kim, Q. X. Lieu, and J. Lee, “Knowledge-Based Systems CNN-based image recognition for topology optimization,” *Knowledge-Based Syst.*, vol. 198, p. 105887, 2020, https://doi.org/10.1016/j.knosys.2020.105887.

[4] W. Baccouch, S. Oueslati, B. Solaiman, and S. Labidi, “ScienceDirect ScienceDirect performance for for automatic automatic A comparative comparative study study of of CNN CNN and and U-Net U-Net performance segmentation of of medical medical images : images : application application to to cardiac cardia MRI ,” *Procedia Comput. Sci.*, vol. 219, no. 2022, pp. 1089–1096, 2023, https://doi.org/10.1016/j.procs.2023.01.388.

[5] R. A. Saputra and K. M. Lhaksmana, “Prediction Retweet Using User-Based and Content-Based with Artificial Neural Network-Harmony Search,” *Building of Informatics, Technology and Science (BITS),* vol. 5, no. 2, pp. 407–413, 2023, https://doi.org10.47065/bits.v5i2.4079.

[6] P. Alkhairi, E. R. Batubara, R. Rosnelly, W. Wanayaumini, and H. S. Tambunan, “Effect Effect of Gradient Descent With Momentum Backpropagation Training Function in Detecting Alphabet Letters,” *Sink. J. dan Penelit. Tek. Inform.*, vol. 8, no. 1, pp. 574–583, 2023, https://garuda.kemdikbud.go.id/documents/detail/3307794.

[7] M. S. B. Kusuma, H., Kardhana, Y. Suryadi, F. I. W. Rohmat, “Hourly Discharge Prediction Using Long Short-Term Memory Recurrent Neural Network (Lstm-Rnn) In The Upper Citarum River,” *Int. J. GEOMATE*, vol. 23, no. 98, pp. 147–154, 2022, https://doi.org/10.21660/2022.98.3462.

[8] D. Malhotra, “Hybrid Deep Learning Model for COVID-19 Prediction Using Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (LSTM) Network,” *Int. J. Comput. Theory Eng.*, vol. 15, no. 3, pp. 125–129, 2023, https://doi.org/10.7763/IJCTE.2023.V15.1341.

[9] P. Alkhairi and A. P. Windarto, “Classification Analysis of Back propagation-Optimized CNN Performance in Image Processing,” *J. Syst. Eng. Inf. Technol.*, vol. 2, no. 1, pp. 8–15, 2023, https://doi.org/10.29207/joseit.v2i1.5015.

[10] M. A. Lubis, D. G. S. Saragih, I. D. Anastasia, A. P. Windarto, and P. Alkhairi, “Application of the ANN Algorithm to Predict Access to Drinkable Water in North Sumatra Regency/City,” *Int. J. Informatics Data Sci.*, vol. 1, no. 1, pp. 18–25, 2023, https://journals.adaresearch.or.id/ijids/article/view/16.

[11] K. V. Rani, “Lung Lesion Classification Scheme Using Optimization Techniques and Hybrid (KNN-SVM) Classifier,” *IETE J. Res.*, vol. 68, no. 2, pp. 1485–1499, 2022, https://doi.org/10.1080/03772063.2019.1654935.

[12] Z. A. Sejuti and M. S. Islam, “A hybrid CNN–KNN approach for identification of COVID-19 with 5-fold cross validation,” *Sensors Int.*, vol. 4, p. 100229, 2023, https://doi.org/10.1016/j.sintl.2023.100229.

[13] A. A. A. Ali, “Intelligent handwritten recognition using hybrid CNN architectures based-SVM classifier with dropout,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3294–3300, 2022, https://doi.org/10.1016/j.jksuci.2021.01.012.

[14] S. Vijayarani, S. Dhayanand, A. Professor, and M. Phil, “Kidney Disease Prediction Using Svm and Ann Algorithms,” *Int. J. Comput. Bus. Res. ISSN (Online*, vol. 6, no. 2, pp. 2229–6166, 2015, https://www.researchmanuscripts.com/March2015/2.pdf.

[15] A. P. Windarto, T. Herawan, and P. Alkhairi, “Prediction of Kidney Disease Progression Using K-Means Algorithm Approach on Histopathology Data,” I*nternational Journal of Computing and Business Research (IJCBR),* vol. 6, no. 2, pp. 1-12, 2023, https://doi.org/10.1007/978-3-031-48465-0_66.

[16] L. Zhao, “CNN, RNN, or ViT? An Evaluation of Different Deep Learning Architectures for Spatio-Temporal Representation of Sentinel Time Series,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 16, pp. 44–56, 2023, https://doi.org/10.1109/JSTARS.2022.3219816.

[17] S. Y. Xiong, “A Proposed Hybrid CNN-RNN Architecture for Student Performance Prediction,” *Int. J. Intell. Syst. Appl. Eng.*, vol. 10, no. 3, pp. 347–355, 2022, https://researchbank.swinburne.edu.au/items/6fb71acd-1dd9-4f3f-97fa-9b743a8fa480/1/.

[18] B. Ruhani, “Hydrogen production via renewable-based energy system: Thermoeconomic assessment and Long Short-Term Memory (LSTM) optimization approach,” *Int. J. Hydrogen Energy*, vol. 52, pp. 505–519, 2024, https://doi.org/10.1016/j.ijhydene.2023.03.456.

[19] E. Rahimi, “The Efficiency of Long Short-Term Memory (LSTM) in Phenology-Based Crop Classification,” *Korean J. Remote Sens.*, vol. 40, no. 1, pp. 57–69, 2024, https://doi.org/10.7780/kjrs.2024.40.1.6.

[20] Z. Li, “Transient Stability Assessment of Power System Based on Multi-granularity Neighborhood Rough Set and Improved Bi-directional Long-short-term Memory Network,” *J. Electr. Eng.*, vol. 18, no. 3, pp. 232–241, 2023, https://doi.org/10.11985/2023.03.025.

[21] D. N. Kirupanithi, “Cybersecurity in Blockchain by Secure Identity Management using Long Short Term Memory (LSTM),” *SSRG Int. J. Electr. Electron. Eng.*, vol. 10, no. 4, pp. 152–163, 2023, https://doi.org/10.14445/23488379/IJEEE-V10I4P115.

[22] A. P. Windarto, T. Herawan, and P. Alkhairi, “Early Detection of Breast Cancer Based on Patient Symptom Data Using Naive Bayes Algorithm on Genomic Data,” in *Artificial Intelligence, Data Science and Applications*, pp. 478-484, 2023, https://doi.org/10.1007/978-3-031-48465-0_64.

[23] S. Defit, A. P. Windarto, and P. Alkhairi, “Comparative Analysis of Classification Methods in Sentiment Analysis: The Impact of Feature Selection and Ensemble Techniques Optimization,” *Telematika*, vol. 17, no. 1, pp. 52–67,

2024, https://ejournal.amikompurwokerto.ac.id/index.php/telematika/article/view/2824.

[24] P. Alkhairi, W. Wanayumini, and B. H. Hayadi, "Analysis of the adaptive learning rate and momentum effects on prediction problems in increasing the training time of the backpropagation algorithm," *AIP Conf. Proc.*, vol. 3048, no. 1, p. 20049, 2024, https://doi.org/10.1063/5.0203374.

[25] A. Kumar, "Bitcoin Price Prediction Using Sentiment Analysis and Long Short-Term Memory (LSTM)," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 7, pp. 480–485, 2023, https://ijisae.org/index.php/IJISAE/article/view/2988.

[26] K. D. B. Ligue, "Deep Learning Approach to Forecasting Dengue Cases in Davao City Using Long Short-term Memory (LSTM)," *Philipp. J. Sci.*, vol. 151, no. 3, pp. 797–812, 2022, https://philjournalsci.dost.gov.ph/images/pdf/pjs_pdf/vol151no3/deep_learning_approach_to_forecasting_dengue_cases_in_Davao_City_.pdf.

[27] J. A. Adisa, "The Effect of Imbalanced Data and Parameter Selection via Genetic Algorithm Long Short-Term Memory (LSTM) for Financial Distress Prediction," *IAENG Int. J. Appl. Math.*, vol. 53, no. 3, 2023, https://www.iaeng.org/IJAM/issues_v53/issue_3/IJAM_53_3_04.pdf.

[28] S. K. Selvakumar, "Divination of stock market exploration using long short-term memory (LSTM)," *Multidiscip. Sci. J.*, vol. 6, no. 1, 2023, https://doi.org/10.31893/multiscience.2024001.

[29] Z. Xuan, "DRN-LSTM: A Deep Residual Network Based On Long Short-term Memory Network For Students Behaviour Recognition In Education," *J. Appl. Sci. Eng.*, vol. 26, no. 2, pp. 245–252, 2023, https://doi.org/10.6180/jase.202302_26(2).0010.

[30] I. Brahma, "Using Long Short-Term Memory (LSTM) networks with the toy model concept for compressible pulsatile flow metering," *Meas. J. Int. Meas. Confed.*, vol. 223, 2023, https://doi.org/10.1016/j.measurement.2023.113782

[31] Y. Yi, "Digital twin-long short-term memory (LSTM) neural network based real-time temperature prediction and degradation model analysis for lithium-ion battery," *J. Energy Storage*, vol. 64, 2023, https://doi.org/10.1016/j.est.2023.107203.

[32] Y. Liu, "Long Short-Term Memory (LSTM) Based Model for Flood Forecasting in Xiangjiang River," *KSCE J. Civ. Eng.*, vol. 27, no. 11, pp. 5030–5040, 2023, https://doi.org/10.1007/s12205-023-2469-7.

[33] N. W. Shen, "Univariate and Multivariate Long Short Term Memory (LSTM) Model to Predict Covid-19 Cases in Malaysia Using Integrated Meteorological Data," *Malaysian J. Fundam. Appl. Sci.*, vol. 19, no. 4, pp. 653–667, 2023, https://doi.org/10.11113/mjfas.v19n4.2814.

[34] B. Sabzipour, "Comparing a long short-term memory (LSTM) neural network with a physically-based hydrological model for streamflow forecasting over a Canadian catchment," *J. Hydrol.*, vol. 627, 2023, https://doi.org/10.1016/j.jhydrol.2023.130380.

[35] B. Xu, "A novel long short-term memory artificial neural network (LSTM)-based soft-sensor to monitor and forecast wastewater treatment performance," *J. Water Process Eng.*, vol. 54, 2023, https://doi.org/10.1016/j.jwpe.2023.104041.

[36] Z. Gao, "A novel multivariate time series prediction of crucial water quality parameters with Long Short-Term Memory (LSTM) networks," *J. Contam. Hydrol.*, vol. 259, 2023, https://doi.org/10.1016/j.jconhyd.2023.104262.

[37] I. Janbain, "Use of long short-term memory network (LSTM) in the reconstruction of missing water level data in the River Seine," *Hydrol. Sci. J.*, vol. 68, no. 10, pp. 1372–1390, 2023, https://doi.org/10.1080/02626667.2023.2221791.

[38] C. H. Ho, "PM<inf>2.5</inf> Forecast in Korea using the Long Short-Term Memory (LSTM) Model," *Asia-Pacific J. Atmos. Sci.*, vol. 59, no. 5, pp. 563–576, 2023, https://doi.org/10.1007/s13143-022-00293-2.

[39] R. Huang, "Well performance prediction based on Long Short-Term Memory (LSTM) neural network," *J. Pet. Sci. Eng.*, vol. 208, 2022, https://doi.org/10.1016/j.petrol.2021.109686.

[40] S. R. Kothuri, "MALO-LSTM: Multimodal Sentiment Analysis Using Modified Ant Lion Optimization with Long Short Term Memory Network," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 5, pp. 326–335, 2022, https://doi.org/10.22266/ijies2022.1031.29.

[41] K. E. ArunKumar, D. V. Kalaga, C. M. S. Kumar, M. Kawaji, T. M. Brenza, "Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends," *Alexandria Eng. J.*, vol. 61, no. 10, pp. 7585–7603, 2022, https://doi.org/10.1016/j.aej.2022.01.011.

[42] M. Aparna, "A Hybrid Siamese-LSTM (Long Short-Term Memory) for Classification of Alzheimer's Disease," *Int. J. Softw. Innov.*, vol. 10, no. 1, 2022, https://doi.org/10.4018/IJSI.309720.

[43] J. Liu, "Prediction of nucleosome dynamic interval based on long-short-term memory network (LSTM)," *J. Bioinform. Comput. Biol.*, vol. 20, no. 3, 2022, https://doi.org/10.1142/S0219720022500093.

[44] D. Hopp, "Economic Nowcasting with Long Short-Term Memory Artificial Neural Networks (LSTM)," *J. Off. Stat.*, vol. 38, no. 3, pp. 847–873, 2022, https://doi.org/10.2478/jos-2022-0037.

BIOGRAPHY OF AUTHORS

Poningsih was born in Pematangsiantar on June 4, 1981. She completed her Doctoral (Ph.D.) education at the University of North Sumatra in March 2024. The author has been teaching at STIKOM Tunas Bangsa since 2005, lecturing in the Informatics Management Study Program. Her research focuses on the field of Artificial Intelligence (data mining and decision support systems). Email: poningsih@amiktunasbangsa.ac.id.

Agus Perdana Windarto the authors were born in Pematangsiantar on August 30 1986. They completed their Masters in Computer Science and Doctoral (Ph.D.) degrees at Putra Indonesia University 'YPTK' Padang in 2014 and 2024. The author has been an active lecturer at STIKOM Tunas Bangsa since 2012, teaching in the Information Systems program. Their research focuses on artificial intelligence. Additionally, the author has served as a reviewer for various nationally accredited journals (SINTA 2 - SINTA 6) and manages a community called “Pemburu Jurnal” at STIKOM Tunas Bangsa. They have won multiple research grant proposals from DIKTI (twice in 2018-2019), DIKTI Community Service Grant (once in 2019), PKM-P Grant (as a student advisor in 2018), and PKM-AI Grant (as a student advisor in 2019). The author is also part of the Relawan Jurnal Indonesia (RJI) community in North Sumatra, the Data Science Indonesia Researchers Association (PDSI), the Forum of Higher Education Communities (FKPT), and is a co-founder of the Yayasan Adwitiya Basurata Inovasi (Yayasan Abivasi) foundation with fellow professors. He can be contacted at email: agus.perdana@amiktunasbangsa.ac.id.

Putrama Alkhairi The writer Putrama Alkhairi was born in Sinaksak on April 17 1997. They completed a master's degree in Computer Science in 2023 at the Main Potential University (UPU) Medan. The author has been an active lecturer at STIKOM Tunas Bangsa since 2021, teaching in the Information Systems program. Their research focuses on artificial intelligence (decision support systems, data mining, artificial neural networks, deep learning, and genetic algorithms). Apart from that, the author has also been active as a member of a community called "Journal Hunters" at STIKOM Tunas Bangsa as a student who took part in joint research with lecturers and has won several awards as the best article writer at the national seminar held by SENARIS 2018. Can be contacted via email : putrama@amiktunasbangsa.ac.id.