

ABSTRACT

Movie reviews served as a crucial source of qualitative information for evaluating various aspects of a film, ranging from storyline and acting to production quality, making sentiment analysis of movie reviews important for comprehensively understanding audience satisfaction. The varied linguistic characteristics of movie reviews required a model capable of understanding word context as well as sequential patterns simultaneously. BERT excelled in contextual word understanding, while BiLSTM and BiGRU excelled in sequential pattern modeling, with GRU being simpler in architecture than LSTM and thus potentially yielding different performance. Therefore, this study proposed a comparative evaluation of BERT-BiLSTM and BERT-BiGRU models for movie review sentiment analysis, alongside analyzing the effect of hyperparameter combinations on model performance. The dataset used was the Movie Review dataset consisting of 10,662 English-language movie reviews labeled with positive and negative sentiment. The research stages included text preprocessing, tokenization, model construction, training with hyperparameter tuning using a 5-Fold Cross Validation scheme, and evaluation using accuracy, precision, recall, and F1-score metrics. The hyperparameters analyzed included batch size, max sequence length, and hidden unit. Based on experimental results, BERT-BiLSTM achieved the best performance with an accuracy of 88.66%, F1-score of 88.70%, precision of 88.29%, and recall of 89.12%, while BERT-BiGRU achieved an accuracy of 87.16%, F1-score of 87.60%, precision of 84.62%, and recall of 90.81%. The best hyperparameter combination for BERT-BiLSTM was a batch size of 32, max sequence length of 128, and hidden unit of 128, while for BERT-BiGRU it was a batch size of 16, max sequence length of 128, and hidden unit of 128. The hyperparameter analysis showed that the effect of each hyperparameter was contextual and could not be evaluated independently. Overall, BERT-BiLSTM was selected as the best model as it produced the highest accuracy and F1-score with more balanced performance in movie review sentiment analysis.

Keywords : Sentiment Analysis, Movie Reviews, BERT, BiLSTM, BiGRU