Abstract

Twitter is one of the social media that is still widely used today. Messages on Twitter are limited to only 280 characters, restricting users' ability to exchange information and forcing users to use various strategies, such as slang and abbreviations, for condensed details. This makes it difficult for users to understand the context or topic of the information, so a solution is needed for this problem. Many studies have classified topics on Twitter but have provided low accuracy results due to limitations and have yet to apply the hybrid deep learning method to topic classification. Therefore, this study involves an expansion feature to overcome these limitations by looking for similar words using gloves and applying a hybrid deep learning method to the classification model to outperform a single model. The models used in this study are Bidirectional Long Short-Term Memory (Bi-LSTM), Convolutional Neural Network (CNN), BILSTM-CNN hybrid, and CNN-BILSTM hybrid. The results showed that the Bi-SLTM model with feature expansion produced an accuracy of 93.82%, followed by CNN with feature expansion produced an accuracy of 93.75%. The best model for topic classification in this study with the best ratio (80:20), the best N-gram variation (unigram + bigram + trigram), and the glove expansion feature with top 5 similarity, namely the CNN-BILSTM hybrid with an accuracy of 94.34% resulted in an accuracy increase of 1.65% for Bi-LSTM and 1.47% for CNN.

Keywords: topic classification, Glove, Bi-LSTM, CNN, hybrid deep learning