

Topic Classification Using the Long Short-Term Memory (LSTM) Method with FastText Feature Expansion on Twitter

Bella Adriani Putri
School of Computing
Telkom University
Bandung, Indonesia
bellaadrp@student.telkomuniversity.ac.id

Erwin Budi Setiawan
School of Computing
Telkom University
Bandung, Indonesia
erwinbudisetiawan@telkomuniversity.ac.id

Abstract—Twitter serves as a social network where people can connect by sending short messages, known as tweets. The limited character count often leads to variations in wording, making it challenging to understand tweets without proper classification. In this research, feature expansion is employed to solve these issues. The main goal of this research is to evaluate the impact of FastText feature expansion in the context of classifying Indonesian-language tweets based on their topic classification. This research utilized FastText to decrease vocabulary mismatches in topic classification by calculating the similarity between words in the corpus. As a result, the anticipated outcome of this research is that FastText could increase the performance of topic classification. The stage of feature extraction involves word weighting using the TF-IDF must be carried out before testing the feature expansion. This research also utilized Long Short-Term Memory (LSTM) as a classification method because it works better using memory cells than recurrent neural networks in general. The best result from this research achieves an accuracy of 95.30% using the LSTM method and feature expansion with FastText based on top similarity 5 by constructing the News corpus. These experiments prove that the LSTM method using FastText could increase the model's accuracy value of topic classification on Twitter.

Keywords—*tweet, topic classification, feature expansion, FastText, Long Short-Term Memory (LSTM)*