

Improving Stance-based Fake News Detection using BERT Model with Synonym Replacement and Random Swap Data Augmentation Technique

Lalu M. Riza Rizky¹, Suyanto²

^{1,2}School of Computing, Telkom University, Bandung

¹zarizky@student.telkomuniversity.ac.id, ²suyanto@telkomuniversity.ac.id

Abstract

The amount of fake news on the internet remains to grow due to its low time and cost of publishing information. A fake news detection system can be implemented to combat its spread. In this research, a stance-based fake news detection model is built with a pretrained Bidirectional Encoder Representations of Transformers (BERT) model fine-tuned for stance detection between headline and body text with data augmentation. The data augmentation utilized in this research includes synonym replacement which replaces chosen words with their synonym, and random swap, which randomly replaces position between two words. The experiment is done by using the two data augmentation techniques separately, combining the two techniques where half of each augmentation is done by one technique, and mixing the two techniques. The evaluation on the test set by cross-validation shows that random swap augmentation provides the best result overall with 42.63% sensitivity, 82.14% specificity, 32.44% F1-score, with the least cost on accuracy with 71.52% accuracy.

Keywords: fake news detection, BERT, synonyms replacement, random swap, data augmentation
