

Abstract

Fake news has proliferated mass media thanks to ever advancing technology, which is becoming a major issue upsetting the society. Based on a 2019 survey by the Indonesian Telematics Society (Mastel) shows many individuals struggle to tell real from false news, particularly news with political topics. This study recommends the application of the Natural Language Inference (NLI) approach collaborated with the Indonesian Bidirection Encoder Representations from Transformers (IndoBERT) model to detect fake news in Indonesian using datasets from the TurnBackHoax and CekFakta sites. The NLI method is used to classify news as entailment (true) or contradictions (fake) by comparing news claims to trusted sources. Two approaches were compared in this study, namely a simple approach that only uses news hypotheses and an NLI approach that utilizes premises and hypotheses. With an accuracy of 86.92%, NLI outperforms the simple model with an accuracy of 80.37% in identifying fake news. The evaluation results show that the NLI-based model produces a significant increase in performance compared to the simple model, with an accuracy increase of 6.55%. The main contribution of this study is to improve the implementation of an inference-based approach that resembles a human verification method combined with IndoBERT method, which allows for more accurate detection of Indonesian-language fake news. Furthermore, this study offers suggestions for future investigation, including the creation of more diverse datasets and investigation of fine-tuning methods to raise model effectiveness in spotting more complicated fake news.

Keywords

Fake News, Natural Language Inference, Deep Learning, IndoBERT