

Deteksi Disinformasi pada Pemilihan Presiden Indonesia 2024 dengan IndoBERT

Andhika Bayu Yudhistira Arda Putra¹, Dr. Yuliant Sibaroni, S.Si., M.T.², Dr. Aditya Firman Ihsan, S.Si., M.Si.³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹andhikabayu@students.telkomuniversity.ac.id, ²yuliant@telkomuniversity.ac.id,

³adityaihsan@telkomuniversity.ac.id

Abstrak

Media sosial tidak hanya digunakan untuk komunikasi sosial, tetapi juga untuk penyebaran berita dan informasi yang komprehensif dan efektif. Twitter adalah salah satu media sosial terbesar yang juga digunakan untuk menyebarkan berita dan informasi. Informasi yang dipublikasikan di Twitter tidak selalu dapat diverifikasi. Hal ini dapat menyebabkan disinformasi tersebar di Twitter. Penyebaran disinformasi di media sosial telah menjadi masalah yang berkembang, terutama pada saat pemilihan presiden. Tujuan dari penelitian ini adalah menggunakan model IndoBERT untuk mengidentifikasi dan meminimalisir penyebaran disinformasi di Twitter terkait pemilihan presiden Indonesia 2024. Penelitian ini dilakukan dalam beberapa tahap yaitu pengumpulan dataset, preprocessing, pelabelan data, penyematan kata dengan Word2Vec, klasifikasi dengan IndoBERT, validasi dan evaluasi dengan K-Fold Cross Validation. Hasil penelitian menunjukkan bahwa penggunaan IndoBERT yang dikombinasikan dengan NLTK Tokenizer dan BERT AutoTokenizer memberikan hasil yang cukup baik dalam meminimalisir penyebaran disinformasi di media sosial. Hasil akurasi yang dicapai adalah 85% ketika menggunakan IndoBERT dengan BERT AutoTokenizer dan 87% ketika menggunakan IndoBERT dengan NLTK Tokenizer dan BERT AutoTokenizer. Secara keseluruhan, penelitian ini menunjukkan efektivitas penggunaan model NLP yang canggih seperti IndoBERT dalam mendeteksi dan meminimalisir penyebaran disinformasi di media sosial.

Kata kunci: indoBERT, Twitter, disinformasi, media sosial, Pemilihan Presiden 2024.

Abstract

Social media is not only used for social communication, but also for the comprehensive and effective dissemination of news and information. Twitter is one of the largest social media also used to spread news and information. Information published on Twitter may not always be verifiable. This can lead to disinformation being spread on Twitter. The spread of disinformation on social media has become a growing problem, especially around the time of the presidential election. The purpose of this study is to use the IndoBERT model to identify and minimize the spread of disinformation on Twitter related to the 2024 Indonesian presidential election. This study was conducted in several phases including dataset collection, preprocessing, data labeling, word embedding with Word2Vec, classification with IndoBERT, validation and evaluation with K-Fold Cross Validation. The results show that using IndoBERT in combination with NLTK Tokenizer and BERT AutoTokenizer yields promising results in minimizing the spread of disinformation on social media. Accuracy results achieved were 85% when using IndoBERT with BERT AutoTokenizer and 87% when using IndoBERT with NLTK Tokenizer and BERT AutoTokenizer. Overall, this study demonstrates the effectiveness of using advanced NLP models like IndoBERT in detecting and minimizing the spread of disinformation on social media.

Keywords: indoBERT, Twitter, disinformation, social media, 2024 Indonesian presidential election.
