

ABSTRAK

Perkembangan teknologi informasi yang pesat telah memberikan dampak signifikan dalam penyebaran informasi, terutama melalui media sosial. Namun, kemudahan ini juga membawa tantangan berupa maraknya penyebaran berita palsu atau hoax yang dapat memengaruhi persepsi publik dan mengganggu stabilitas sosial. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi teks berbahasa Indonesia guna mengidentifikasi berita hoax dengan menggunakan metode Word2Vec sebagai teknik representasi kata dan algoritma K-Nearest Neighbor (KNN) sebagai metode klasifikasinya. Word2Vec digunakan untuk mengubah teks menjadi vektor numerik yang merepresentasikan makna semantik kata, sehingga dapat digunakan dalam proses klasifikasi. Sementara itu, algoritma KNN digunakan untuk mengklasifikasikan teks ke dalam kategori hoax atau non-hoax berdasarkan kemiripan fitur. Dataset yang digunakan terdiri dari 60 berita berbahasa Indonesia yang diambil dari situs turnbackhoax.id dan detik.com. Hasil evaluasi menunjukkan bahwa kombinasi Word2Vec dan KNN mampu memberikan performa klasifikasi yang cukup baik, dengan tingkat akurasi mencapai 76.92%. Penelitian ini menunjukkan bahwa pendekatan Word2Vec dan KNN efektif untuk tugas identifikasi berita hoax, serta berpotensi dikembangkan lebih lanjut dalam sistem pendeteksi berita palsu berbasis aplikasi. Temuan ini diharapkan dapat berkontribusi terhadap upaya peningkatan literasi digital dan pemberdayaan masyarakat dalam menghadapi informasi yang menyesatkan di era digital.

Kata Kunci: Berita Hoax, Word2Vec, K-Nearest Neighbor, Klasifikasi Teks.

ABSTRACT

The rapid development of information technology has had a significant impact on the spread of information, especially through social media. However, this convenience also brings challenges in the form of spreading fake news or hoaxes that can affect public perception and disrupt social stability. This study aims to develop an Indonesian text classification system to identify hoax news by using Word2Vec method as word representation technique and K-Nearest Neighbor (KNN) algorithm as its classification method. Word2Vec is used to convert text into numerical vectors representing the semantic meaning of words, so it can be used in the classification process. Meanwhile, the KNN algorithm is used to classify text into hoax or non-hoax categories based on feature similarity. The dataset used consists of 60 Indonesian-language news stories taken from turnbackhoax.id and detik.com . The evaluation results show that the combination of Word2Vec and KNN can provide a fairly good classification performance, with an accuracy rate of 76.92%. This study shows that Word2Vec and KNN approaches are effective for hoax news identification tasks, as well as potentially being further developed in application-based fake news detection systems. This finding is expected to contribute to efforts to improve digital literacy and community empowerment in dealing with misleading information in the digital era.

Keywords: Hoax News, Word2Vec, K-Nearest Neighbor, Text Classification.