

Deteksi Hoax Mengenai Berita Covid Pada Twitter Menggunakan Convolutional Neural Network (CNN) dan Support Vector Machine (SVM)

Arvia Dwi Cahyani¹, Dr. Yuliant Sibaroni, S.Si., M.T.², Sri Suryani Prasetyowati P, S.Si., M.Si.³,

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

arviadc@students.telkomuniversity.ac.id¹, yuliant@telkomuniversity.ac.id²,

srisuryani@telkomuniversity.ac.id³

Abstrak

Tidak dapat dipungkiri lagi bahwa saat ini penyebaran berita sangat cepat terjadi pada media sosial. Dengan kemudahan dalam memperoleh berita di media sosial, mengakibatkan banyaknya pengguna yang menyebarkan berita tanpa mengetahui keaslian berita tersebut. Pada media sosial twitter pengguna memiliki peran yang penting dalam penyebaran berita. Pada permulaan tahun 2020 kasus Covid mulai memasuki Indonesia dan banyak orang yang menyebarkan berita mengenai Covid dan tidak diketahui kebenarannya, berita kian tersebar di twitter karena disebarluaskan oleh beberapa orang yang tidak bertanggung jawab. Penelitian ini membangun sebuah sistem yang dapat mendeteksi berita *hoax* pada media sosial. Tahapan pada penelitian ini dimulai dari *crawling data*, *preprocessing data*, *word embedding*, *splitting data*, kemudian proses *modelling* menggunakan metode *Convolutional Neural Network (CNN)* dan *Support Vector Machine (SVM)*. Selain itu, dataset yang digunakan dalam penelitian ini mengenai kasus Covid-19 pada tahun 2020 sebanyak 3897 berita, dengan jumlah berita fakta sebanyak 2021 dan berita hoax sebanyak 1876 data. Sistem yang dibangun menggunakan data yang terbagi menjadi 90:10 (90% data *train* dan 10% data *test*). Setelah dilakukan dengan beberapa proses, hingga berhasil ditingkatkan menggunakan ekstraksi fitur *TF-IDF (Term Frequency Inverse Document Frequency)* dan pembobotan N-Gram kedua metode dapat menghasilkan akurasi yang cukup baik. Hasil penelitian menunjukkan bahwa dengan menggunakan kombinasi N-Gram Unigram + Bigram + Trigram pada CNN mempunyai nilai akurasi sebesar 75.8%, dan pada pemodelan SVM sebesar 77.9 %.

Kata kunci : berita, *hoax*, covid-19, media sosial, twitter, cnn, svm