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

Arvia Dwi Cahyani[1], Dr. Yuliant Sibaroni, S.Si., M.T[2]., Sri Suryani Prasetiyowati P, S.Si., M.Si[3],
[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
arviadc@students.telkomuniversity.ac.id[1], yuliant@telkomuniversity.ac.id[2],
srisuryani@telkomuniversity.ac.id[3]

_____

Abstract
It is undeniable that nowadays the spread of news is very fast on social media. With the ease of obtaining news on social media, it has resulted in many users spreading the news without knowing the authenticity of the news. On Twitter social media users have an important role in spreading the news. At the beginning of 2020 Covid cases began to enter Indonesia and many people spread news about Covid and the truth was not known, the news was increasingly spread on Twitter because it was spread by several irresponsible people. This research builds a system that can detect hoax news on social media. The stages in this research start from crawling data, data preprocessing, word embedding, data splitting, then the modeling process uses the Convolutional Neural Network (CNN) and Support Vector Machine (SVM) methods. In addition, the dataset used in this study regarding the Covid-19 case in 2020 was 3897 news, with the number of factual news as high as 2021 and fake news as many as 1876 data. The system built uses data divided into 90:10 (90% train data and 10% test data). After several processes, it was successfully improved using TF-IDF (Term Frequency Inverse Document Frequency) feature extraction and N-Gram weighting of the two methods to produce fairly good accuracy. The research results indicate that using a combination of Unigram + Bigram + Trigram N-Gram in CNN yields an accuracy of 75.8%, while in SVM modeling, the accuracy reaches 77.9%.

Keywords: news, hoax, covid-19, social media, twitter, cnn, svm