

Analisis Sentimen Tweet COVID-19 Menggunakan Metode K-Nearest Neighbors

Muhammad Hafizh Mahendra¹, Danang Triantoro Murdiansyah², Kemas Muslim Lhaksamana³

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

¹madfrog@students.telkomuniversity.ac.id, ²danangtri@telkomuniversity.ac.id,

³kemasmuslim@telkomuniversity.ac.id

Abstract

Entering the 21st century as technology and information develop, the amount of data on the internet is growing rapidly. This has resulted in researchers accessing data and information for purposes such as academic research and commercial use. Social Data in cyberspace can contain events that occur in real life, we take an example that is now happening, namely the global COVID-19 outbreak caused by the corona virus spread throughout the world. Due to the rapid spread of the virus in various places, the World Health Organization (WHO) declared a state of emergency. Many individuals including various media organizations and governments present the latest news and opinions on the corona virus. By analyzing public sentiment towards the corona virus, it is possible to conclude the results of the analysis of public opinion. The dataset used is an open dataset on the Kaggle website taken from Twitter. There are several stages to analyzing sentiment, techniques such as tokenization, stemming, classification, and others are very influential on the accuracy. Feature Extraction methods used are Term Frequency - Inverse Document Frequency (TF-IDF) and CountVectorizer. The test results on this model produce the best accuracy of 73.2% with the TF-IDF Feature Extraction.

Keywords: *Sentiment Analysis, KNN, COVID-19, TF-IDF, CountVectorizer*

