

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

Abstrak

Memasuki abad ke-21 seiring berkembangnya teknologi dan informasi, jumlah data yang ada di internet berkembang pesat. Ini mengakibatkan para peneliti mengakses data dan informasi untuk kebutuhan seperti penelitian akademik maupun penggunaan komersial. *Social Data* pada dunia maya dapat berisi tentang kejadian yang terjadi di kehidupan nyata, kita ambil contoh yang sekarang terjadi yaitu wabah global COVID-19 yang disebabkan oleh virus corona tersebar diseluruh dunia. Karena penyebaran virus yang cepat di berbagai tempat, World Health Organization (WHO) menyatakan keadaan darurat. Banyaknya individual termasuk berbagai media organisasi dan pemerintah menghadirkan berita terbaru dan opini terhadap virus corona. Dengan menganalisa sentimen publik terhadap virus corona memungkinkan kita untuk menyimpulkan hasil analisis pendapat masyarakat. Dataset yang digunakan adalah open dataset di website Kaggle yang diambil dari Twitter. Ada beberapa tahapan untuk menganalisis sentimen dilakukan teknik seperti *tokenization*, *stemming*, *classification*, dan lain-lain sangat berpengaruh pada akurasi. Metode *Feature Extraction* yang digunakan yaitu *Term Frequency - Inverse Document Frequency (TF-IDF)* dan *CountVectorizer*. Hasil pengujian pada model ini menghasilkan hasil akurasi terbaik 73,2% dengan *Feature Extraction TF-IDF*.

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

