

ANALISIS SENTIMEN ISU VAKSINASI COVID-19 PADA TWITTER DENGAN METODE NAIVE BAYES

Nama Mahasiswa : Yashmine Hapsari
NIM : 11181083
Dosen Pembimbing Utama : Syamsul Mujahidin, S.Kom., M.Eng.
Dosen Pembimbing Pendamping : Nisa Rizqiya Fadhliana, S.Kom., M.T.

ABSTRAK

Wabah virus Corona, atau COVID-19 merupakan penyakit dengan tingkat penularan yang tinggi. Vaksinasi COVID-19 telah diberlakukan dan masyarakat umum yang memenuhi persyaratan wajib melakukan vaksinasi, namun status vaksin yang masih dalam tahap pengembangan merupakan salah satu faktor keraguan masyarakat untuk melakukan vaksinasi. Salah satu cara untuk mengetahui pendapat masyarakat secara umum adalah dengan melakukan analisis sentimen terhadap opini masyarakat yang tersebar di sosial media. Permasalahan yang didapatkan saat melakukan analisis sentimen yaitu pengolahan data berupa kata-kata dan pengambilan nilai dari setiap fiturnya. Pada penelitian ini klasifikasi sentimen dilakukan dengan metode *Naive Bayes* karena memiliki kelebihan yaitu dapat menangani data yang besar secara sederhana dan akurasi yang tinggi, serta memiliki keunggulan dalam performansi model dengan data yang tidak memiliki keterkaitan antar-fiturnya. *Dataset* merupakan *tweet* berjumlah 5000 data dengan kata kunci terkait vaksinasi COVID-19. Tahap pembobotan dilakukan dengan metode TF-IDF dimana dilakukan perbandingan terhadap pengaruh penggunaan token *unigram*, *bigram* dan *1-2 gram* terhadap akurasi model dan metode *cross validation* untuk menghindari terjadinya *overfitting* model. Berdasarkan percobaan yang telah dilakukan, penggunaan token *1-2 gram* memberikan pengaruh terhadap hasil evaluasi model. Pada percobaan dengan pengklasifikasi *Gaussian* dan perbandingan *train:test* yaitu 7:3, didapatkan akurasi model 67.4% untuk parameter *unigram*, 65.5% untuk parameter *bigram*, dan 70% untuk parameter *1-2 gram*, dimana model dengan gabungan token *1-2 gram* memiliki akurasi yang lebih tinggi jika dibandingkan dengan penggunaan 1 jenis token saja. Dari hasil ini dapat disimpulkan bahwa penambahan jenis tokenisasi dapat memberikan nilai tambah terhadap model untuk mengklasifikasikan data, sehingga dapat meningkatkan akurasi dalam analisis terkait sentimen publik.

Kata kunci: Analisis sentimen, COVID-19, Klasifikasi *Naive Bayes*, Twitter dan Vaksinasi.