

Perbandingan Metode Random Forest dan Decision Tree untuk Klasifikasi Emosi Berdasarkan Unggahan di Media Sosial

Muhammad Abiyyu Tsaqif¹, Warih Maharani²

Fakultas Informatika, Universitas Telkom, Bandung

abiyyutsaif@student.telkomuniversity.ac.id,

wmaharani@telkomuniversity.ac.id

Abstrak

Klasifikasi emosi dalam media sosial menjadi tugas penting dengan aplikasi dalam kesehatan mental, pemantauan sentimen publik, dan analisis umpan balik pelanggan. Penelitian ini membandingkan algoritma *Random Forest* dan *Decision Tree* untuk mengklasifikasikan emosi seperti bahagia, sedih, marah, dan takut dari postingan media sosial. Data dikumpulkan melalui *crawling tweet* dan pelabelan manual. Tahap prapemrosesan mencakup tokenisasi, stemming, dan penghapusan *stopword*, dengan ekstraksi fitur menggunakan TF-IDF dan *Bag of Words*. Skenario eksperimen menguji rasio pembagian data, *resampling* untuk keseimbangan kelas, dan penyetelan parameter. Parameter *Decision Tree* yang diuji meliputi *criterion* (*gini*, *entropy*), *max depth* (*none*, *fixed value*), *min samples split* (2, 5), dan *min samples leaf* (1, 2). Sementara itu, parameter *Random Forest* yang disesuaikan mencakup *n_estimators* (100–400), *max depth* (*none*, *fixed value*), *min samples split* (2, 5, 10), dan *min samples leaf* (1, 2). Hasil menunjukkan *Random Forest* mencapai akurasi maksimum 76,17%, melampaui *Decision Tree* yang mencapai 72,62%. Kombinasi TF-IDF dan *Bag of Words* memberikan akurasi tertinggi untuk kedua model. Studi ini menekankan pentingnya *preprocessing*, keseimbangan data, dan optimasi parameter dalam klasifikasi emosi. Temuan ini memberikan wawasan dalam pengembangan analisis sentimen dan pemrosesan bahasa alami, yang dapat diterapkan dalam pemantauan opini publik, peningkatan pengalaman pelanggan, dan manajemen krisis.

Kata Kunci : *Bag of Words*; *Decision Tree*; Klasifikasi Emosi; *Random Forest*; TF-IDF
