

Komparasi Logistic Regression dan Naive Bayes dalam Mengklasifikasi Emosi di Media Sosial X Indonesia

Gerald Shabran Rasyad¹, Warih Maharani²
Fakultas Informatika, Universitas Telkom, Bandung
geraldsrasyad@student.telkomuniversity.ac.id,
wmaharani@telkomuniversity.ac.id

Abstrak

Emosi merupakan bagian integral dari interaksi manusia dan pengambilan keputusan, sering kali diungkapkan di platform media sosial seperti X, yang menyediakan data berharga untuk analisis sentimen. Namun, menganalisis teks dari X menimbulkan tantangan karena bahasa informal, bahasa gaul, dan fitur tekstual yang unik. Penelitian ini membandingkan Regresi Logistik dan Naive Bayes dalam mengklasifikasikan emosi dari tweet Indonesia, mengatasi kesenjangan pada penelitian sebelumnya dengan mengeksplorasi metode ekstraksi fitur, rasio pemisahan data, dan penyetelan hyperparameter. Data yang dikumpulkan dari 100 mahasiswa Telkom University menghasilkan 8.978 tweet yang diberi label empat emosi: Senang, Sedih, Marah, dan Takut. Setelah dilakukan preprocessing, diterapkan metode ekstraksi fitur TF-IDF dan Bag of Words (BoW). Model dilatih dan diuji pada pemisahan data 10%, 20%, dan 30%, dan performa dievaluasi menggunakan akurasi, presisi, perolehan, dan skor F1. Penyetelan hyperparameter dilakukan untuk Regresi Logistik menggunakan GridSearch. Hasil menunjukkan Regresi Logistik mengungguli Naive Bayes, mencapai akurasi 73,49% dibandingkan dengan 70,27%, dengan BoW memberikan hasil yang lebih unggul dibandingkan TF-IDF. Pembagian data sebesar 20% memberikan keseimbangan terbaik untuk pelatihan dan pengujian. Penelitian ini menunjukkan efektivitas Regresi Logistik dan menyoroti pentingnya ekstraksi fitur yang disesuaikan dan optimalisasi parameter untuk klasifikasi emosi dalam kumpulan data teks informal, khususnya untuk tweet berbahasa Indonesia.

Kata Kunci : Logistic Regression, Naive Bayes, emotion classification, TF-IDF, BoW
