

ABSTRAK

Interaksi pada masa kini tidak mengharuskan bertatap muka, sosial media menjadi salah satu wadah untuk berinteraksi secara tidak langsung. X sebagai salah satu media sosial dengan pengguna sebanyak 27,5 juta, menjadi tempat interaksi sosial berbentuk teks dengan banyaknya data yang dapat dikumpulkan dan digunakan dalam penelitian ini untuk menemukan perbandingan performa antara metode *Logistic Regression* dan *Naive Bayes* dalam melakukan prediksi kepribadian berdasarkan data dari media sosial X.

Dalam penelitian ini, percobaan dilakukan dengan tiga skenario: tanpa data interaksi, dengan data interaksi, dan augmentasi data. *Logistic Regression* lebih unggul dalam menangkap pola dalam data teks dengan akurasi tertinggi sebesar 52,63%, tetapi *Naive Bayes* lebih stabil ketika dihadapkan pada data yang tidak seimbang dengan akurasi tertinggi sebesar 51,58%. Penambahan data interaksi tidak memberikan peningkatan signifikan karena distribusi yang kurang bervariasi. Namun, augmentasi data dapat sedikit meningkatkan akurasi, terutama untuk *Naive Bayes*. Faktor utama yang memengaruhi kinerja model adalah metode ekstraksi fitur, kualitas dan variasi data, serta ketidakseimbangan data. Sehingga berdasarkan penelitian ini, pemilihan fitur yang tepat dan strategi pengolahan data yang optimal berperan penting dalam meningkatkan akurasi prediksi kepribadian.

Kata Kunci : X, Interaksi, *Logistic Regression*, *Naive Bayes*.