

ABSTRAK

Sentimen analisis produk dalam e-commerce telah menjadi perhatian utama dalam industri tersebut. Analisis sentimen memberikan wawasan tentang preferensi dan perilaku konsumen dalam pembelian. Tujuan proyek ini adalah mengembangkan model machine learning dengan algoritma k-NN untuk menganalisis sentimen pada ulasan produk Lazada Indonesia. Metode meliputi pengumpulan data, pemrosesan data, analisis, pemodelan, dan evaluasi. Data ulasan diambil dari Kaggle.com dan diolah dengan Natural Language Processing (NLP) untuk mengidentifikasi sentimen positif, negatif, atau netral. Model k-NN dibangun dan dievaluasi. Hasil menunjukkan bahwa pendekatan NLP dan k-NN dapat diterapkan untuk menganalisis sentimen ulasan produk e-commerce. Namun, data yang digunakan cenderung tidak seimbang. Model awal memiliki akurasi 87%, tetapi penanganan data tidak seimbang dengan Random UnderSampler terjadi penurunan akurasi menjadi 57% namun kemampuan mengenali kelas minoritas meningkat. Teknik Random OverSampler mencapai akurasi 96%, dan Synthetic Minority Over-sampling Technique (SMOTE) 94%. Dalam pengembangan lebih lanjut, perlu untuk memikirkan dampak data tidak seimbang pada prediksi model dan mengatasinya. Dengan penerapan teknik penanganan data yang lebih canggih, diharapkan model dapat memberikan prediksi yang lebih akurat pada ulasan produk e-commerce.

Kata Kunci: sentimen analisis, e-commerce, NLP, machine learning, k-NN, Imbalanced data