

ABSTRAK

Ulasan pelanggan pada platform digital seperti *Google Maps* merupakan aset penting bagi evaluasi kualitas layanan restoran, termasuk Rumah Makan Bebek Sinjay. Namun, volume data tekstual yang masif dan struktur bahasa Indonesia yang kompleks menuntut adanya pendekatan komputasional untuk mengekstraksi informasi sentimen secara otomatis. Penelitian ini bertujuan untuk mengimplementasikan dan mengoptimasi model analisis sentimen menggunakan algoritma *K-Nearest Neighbor* (K-NN) dengan memanfaatkan arsitektur *Orange Data Mining*. Kebaruan (*novelty*) pada penelitian ini terletak pada integrasi modul bawaan dengan injeksi *script* kustom Python pada tahap awal pra-pemrosesan. *Script* ini bertugas mengeksekusi *stemming* khusus tata bahasa Indonesia serta pelabelan *Lexicon-Based* menggunakan *Custom Dictionary* domain kuliner untuk menstrukturisasi 1.587 ulasan mentah menjadi kelas biner (Positif dan Negatif). Ekstraksi fitur dilakukan menggunakan metode TF-IDF yang dioptimasi dengan normalisasi vektor *L2 (Euclidean)*. Evaluasi model menggunakan *10-fold Cross Validation* pada variasi nilai k (3, 5, 7, 9) dengan pengaturan *Distance-weighted K-NN*. Hasil eksperimen mengungkap adanya fenomena paradoks akurasi akibat ketimpangan kelas (*class imbalance*). Meskipun skenario $k=9$ menghasilkan akurasi tertinggi (74,1%), pengujian menetapkan $k=5$ sebagai batas keputusan optimal (*optimal decision boundary*) dengan tingkat Akurasi sebesar 72,3%. Pemilihan ini didasarkan pada kemampuan model $k=5$ dalam menahan laju bias *over-smoothing*, sehingga mampu mengklasifikasikan 875 ulasan Positif (*True Positive*) dan 273 ulasan Negatif (*True Negative*) secara proporsional tanpa mengorbankan nilai *Recall* pada kelas minoritas. Pendekatan hibrida ini terbukti efektif dalam memproses teks bahasa Indonesia dan memitigasi bias pada dataset ulasan yang tidak seimbang.

Kata Kunci : Analisis Sentimen, Bebek Sinjay, K-Nearest Neighbor, Lexicon-Based, Python Script, TF-IDF.

ABSTRACT

Customer reviews on digital platforms such as Google Maps are a crucial asset for evaluating restaurant service quality, including Rumah Makan Bebek Sinjay. However, the massive volume of textual data and the complex structure of the Indonesian language necessitate a computational approach to extract sentiment information automatically. This study aims to implement and optimize a sentiment analysis model using the K-Nearest Neighbor (K-NN) algorithm by utilizing the Orange Data Mining architecture. The novelty of this research lies in the integration of built-in modules with the injection of a custom Python script in the initial preprocessing stage. This script is tasked with executing stemming specific to Indonesian grammar and Lexicon-Based labeling using a culinary domain Custom Dictionary to structure 1,587 raw reviews into binary classes (Positive and Negative). Feature extraction is performed using the TF-IDF method optimized with L2 (Euclidean) vector normalization. The model evaluation uses 10-fold Cross-Validation on variations of k values (3, 5, 7, 9) with a Distance-weighted K-NN configuration. The experimental results reveal an accuracy paradox phenomenon due to class imbalance. Although the k=9 scenario produces the highest accuracy (74.1%), the evaluation establishes k=5 as the optimal decision boundary with an Accuracy rate of 72.3%. This selection is based on the k=5 model's ability to withstand the rate of over-smoothing bias, enabling it to proportionally classify 875 Positive reviews (True Positive) and 273 Negative reviews (True Negative) without sacrificing the Recall value in the minority class. This hybrid approach is proven effective in processing Indonesian text and mitigating bias in imbalanced review datasets.

Keywords: Sentiment Analysis, Bebek Sinjay, K-Nearest Neighbor, Lexicon-Based, Python Script, TF-IDF.