

ABSTRAK

Dengan perkembangan *marketplace* yang sangat pesat, hal ini menyebabkan konsumen meminta untuk mendapat *service* lebih dengan maksud untuk memudahkan penarikan kesimpulan dari analisis sentimen dari sekumpulan ulasan produk tanpa melihat jumlah bintang yang diberikan pada ulasan dan menyimpulkan harga dari produk yang diinginkan. Ketika melakukan analisis ulasan produk, tidak lah cukup dengan hanya melihat jumlah banyaknya bintang yang diberikan oleh konsumen lain untuk mengerti arti atau maksud dari sebuah ulasan produk.

Algoritma *K-Nearest Neighbor* (KNN) merupakan metode yang cukup baik dalam klasifikasi dan dapat melakukan sentimen analisis dengan baik serta bekerja lebih efektif jika data analisis yang digunakan dalam jumlah yang besar, sehingga pemilihan fitur pembobotan TF-IDF dipilih untuk menangani penelitian dengan jumlah data yang cukup besar. *Term Frequency-Invers Document Frequency* (TF-IDF) merupakan numerik yang menunjukkan hubungan antara beberapa dokumen tertentu sehingga dokumen dapat diidentifikasi atau dikategorisasikan. Penambahan Algoritma *Natural Language Processing* (NLP) dikarenakan untuk menganalisa ulasan pada produk di *marketplace* terdapat banyaknya singkatan serta bahasa yang tidak baku yang akan sulit dimengerti oleh sistem, maka dari itu pendekatan NLP dibutuhkan untuk memperbaiki bahasa pada ulasan sehingga sistem analisis yang dibangun dapat bekerja maksimal. Algoritma *K-Means* sendiri dipilih karena algoritma ini mempunyai kemampuan mengelompokkan data dalam jumlah cukup besar dengan waktu komputasi yang cepat dan efisien, selain itu algoritma *K-Means* adalah algoritma terbaik dalam algoritma *Partitional Clustering* dan merupakan algoritma yang sering digunakan untuk *clustering* karena dirasa sederhana dan sangat efisien.

Dalam penelitian ini, pengujian dilakukan dengan *confusion matrix* untuk menghitung nilai akurasi, presisi, *recall* dan F1-Score. Pengujian ini akan menggunakan *split* dataset 90:10. Pengujian akan membandingkan implementasi algoritma menggunakan NLP dan tanpa menggunakan NLP. Pada model KNN menggunakan NLP proses *preprocessing* ditambahkan tahap *word normalizer*, *stemming*, *menghilangkan unused character*, *stopword removal* dan *negation handling* nilai performa yang didapat menghasilkan nilai *accuracy* tertinggi pada K5 sebesar 83,51% , nilai *precision* 80.55% , nilai *recall* 90.625% dan nilai F1-Score 87.83%. Untuk nilai performa yang didapat pada model KNN tanpa menggunakan NLP menghasilkan nilai *accuracy* tertinggi pada K1 sebesar 79,67% , nilai *precision* 76.53% , nilai *recall* 84.26% dan nilai F1-Score 83.33%.

Kata kunci : Analisis Sentimen, *K-Nearest Neighbor* (KNN), *Natural Language Processing* (NLP), *Marketplace*.