

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kemajuan teknologi informasi yang sangat pesat telah membawa dampak signifikan terhadap berbagai aspek kehidupan manusia, termasuk dunia bisnis. E-commerce menjadi sektor yang mengalami pertumbuhan sangat pesat, dengan berbagai platform yang menyediakan layanan jual beli secara daring [1]. Tokopedia, sebagai platform e-commerce terkemuka di Indonesia, menyediakan beragam produk dan layanan serta memungkinkan pengguna untuk memberikan rating dan ulasan terhadap barang yang telah mereka beli [2].

Ulasan pelanggan menjadi hal penting dalam menentukan reputasi produk dan mempengaruhi keputusan pembelian calon pelanggan lainnya. Dengan meningkatnya jumlah ulasan yang tersedia, analisis sentimen dari ulasan-ulasan menjadi semakin penting. Analisis sentiment adalah proses otomatisasi untuk menentukan sentiment negatif atau positif dalam ulasan yang diberikan. Hal ini dapat memberikan manfaat berharga dari kepuasan pelanggan dan mendukung perusahaan untuk meningkatkan kualitas layanan dan produk.

Untuk melakukan analisis sentimen, berbagai algoritma machine learning dapat digunakan. Dua di antaranya adalah *K Nearest Neighbors(KNN)* dan *Naive Bayes*. Algoritma *Naive Bayes* dikenal karena kesederhanaan dan efisiensinya dalam mengklasifikasikan teks, sementara *K-Nearest Neighbors* dikenal karena kemampuannya dalam mengelompokkan data berdasarkan kesamaan [3]. Meskipun kedua algoritma ini memiliki kelebihan dan ciri khas tersendiri, masih sedikit penelitian yang secara komprehensif membandingkan performa keduanya dalam konteks analisis sentimen terhadap penilaian produk di platform Tokopedia.

Fokus dari pengkajian ini adalah untuk mengomparasikan kinerja *K Nearest Neighbors(KNN)* dan *Naive Bayes* guna menganalisis tingkat kepuasan pelanggan berdasarkan ulasan produk di Tokopedia. Perbandingan ini diharapkan mampu menyajikan informasi yang lebih akurat terkait dengan kelebihan dan keterbatasan dari setiap pendekatan, sehingga perusahaan akan mendapatkan keputusan yang lebih baik ketika menerapkan analisis sentimen dalam meningkatkan kepuasan pelanggan.

1.2. Rumusan Masalah

Dilatar belakangin oleh permasalahan yang telah diuraikan terdapat beberapa isu dan tantangan yang perlu diatasi, dengan demikian dalam penelitian ini disajikan rumusan masalah sebagai berikut :

1. Bagaimana proses preprocessing dan ekstraksi fitur dilakukan pada dataset ulasan produk Tokopedia untuk analisis tingkat kepuasan pelanggan?
2. Bagaimana performa algoritma *Naive Bayes* dalam mengklasifikasikan tingkat kepuasan pelanggan berdasarkan ulasan produk Tokopedia?
3. Bagaimanakah efektivitas metode algoritma *KNearest Neighbors(KNN)* dalam melakukan pengelompokkan tingkat kepuasan pelanggan berdasar pada ulasan produk Tokopedia?
4. Bagaimana perbandingan akurasi dan efektivitas antara metode *K Nearest Neighbors(KNN)* dan *Naive Bayes* dalam menganalisis tingkat kepuasan pelanggan berdasar pada ulasan produk Tokopedia?
5. Unsur-unsur apa saja yang mempengaruhi perbedaan performa kinerja antara metode *K Nearest Neighbors(KNN)* dan *Naive Bayes* untuk menganalisis sentiment dalam ulasan pelanggan?

1.3. Batasan Masalah

Berlandaskan pengkajian masalah yang telah dilaksanakan, batasan masalah didalam penelitian ini dapat diuraikan seperti berikut:

1. Dalam penelitian ini, data yang dimanfaatkan hanya mencakup ulasan produk dari platform Tokopedia sebagai sumber data utama. Ulasan yang digunakan terbatas pada periode tertentu dan kategori produk tertentu untuk memastikan konsistensi data.
2. Algoritma yang Dibandingkan Penelitian ini hanya membandingkan 2 algoritma klasifikasi, yakni *K Nearest Neighbors(KNN)* dan *Naive Bayes*, untuk analysis tingkat kepuasan pelanggan. Algoritma lain tidak diikutsertakan dalam penelitian ini.
3. Bahasa Ulasan Analisis sentimen hanya dilakukan pada ulasan yang ditulis dalam Bahasa Indonesia. Ulasan dalam bahasa lain atau yang mengandung campuran bahasa tidak termasuk dalam lingkup penelitian.

1.4. Tujuan Penelitian

Tujuan dari penelitian untuk:

1. Mengimplementasikan pre-processing dan ekstraksi fitur untuk melakukan langkah preprocessing data (seperti CaseFolding, Tokenizing, StopwordRemoval, dan Stemming) serta ekstraksi fitur memakai metode TermFrequency-Inverse Document Frequency (TF-IDF) untuk dataset ulasan produk Tokopedia.
2. Untuk Mengimplementasikan algoritma *Naive Bayes* dalam mengklasifikasikan tingkat kepuasan pelanggan berdasarkan ulasan produk Tokopedia dan mengevaluasi performanya.
3. Mengimplementasikan metode *K Nearest Neighbors(KNN)* dalam mengklasifikasikan tingkat kepuasan pelanggan berlandaskan ulasan produk Tokopedia dan mengevaluasi performanya.