

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Perkembangan teknologi yang dibantu dengan berkembangnya akses jaringan internet memiliki pengaruh yang cukup signifikan bagi penggunaan di Indonesia. Sebuah survei yang dilakukan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) menunjukkan bahwa jumlah pengguna internet di Indonesia akan mencapai 215,63 juta pada tahun 2022-2023. Ini adalah peningkatan 2,67% dari 210,03 juta pengguna pada periode sebelumnya. Jumlah pengguna internet ini setara dengan 78,19% dari 275,77 juta orang Indonesia.[1] Tokopedia telah menjadi salah satu platform terpenting bagi pelanggan dan pengusaha dalam berkomunikasi dan berinteraksi. Dalam hal ini, Tokopedia menjadi salah satu platform yang populer digunakan sebagai platform jual beli online. Dalam lingkungan bisnis ini, pengusaha perlu mengumpulkan data opini pelanggan mereka dari Tokopedia untuk memahami preferensi, kebutuhan, dan keinginan mereka.

Tokopedia adalah toko online paling populer di Indonesia, dengan 1,2 miliar pengunjung, 863,1 juta menggunakan web mobile dan 329,8 juta menggunakan desktop.[2] Namun, mengumpulkan data opini pelanggan dari Tokopedia secara manual dapat memakan waktu dan tidak efisien. Oleh karena itu, teknik Sentiment Analysis dapat digunakan untuk membantu pengusaha mengumpulkan dan menganalisis opini pelanggan secara otomatis. Analisis sentimen, juga dikenal sebagai opinion mining, adalah proses memahami, mengekstrak, dan mengolah data teks secara otomatis untuk mendapatkan informasi tentang sentimen yang terkandung dalam suatu kalimat opini. Karena besarnya pengaruh dan manfaatnya, penelitian dan aplikasi berbasis analisis sentimen telah berkembang pesat.[3]

Algoritma K-Nearest Neighbors adalah alat penentu klasifikasi yang bergantung pada label kategori pada dokumen pelatihan, seperti dokumen tes, dan contoh dasar yang tidak membangun dengan representasi deklaratif kategori yang

jelas.[4] Metode KNN adalah metode belajar yang malas, Algoritma ini bekerja dengan mencari k titik data terdekat yang akan diklasifikasikan atau diregresi. Dalam klasifikasi, titik data ini akan diklasifikasikan berdasarkan mayoritas kelas dari k titik data terdekat.

Penelitian ini tidak hanya menggunakan metode KNN, tetapi juga menggunakan metode Decision Tree. Untuk pengenalan pola statistic, salah satu algoritma yang paling sering digunakan untuk pengambilan keputusan adalah algoritma Decision Tree. Pohon keputusan juga berfungsi sebagai representasi model prediktif untuk sebuah keputusan dengan struktur hirarki atau pohon; mereka mencari solusi untuk masalah dengan mengubah kriteria menjadi node yang terhubung untuk membentuk struktur seperti pohon. Setiap pohon memiliki cabang. Cabang-cabang ini menunjukkan tugas yang harus dilakukan untuk tumbuh ke cabang berikutnya hingga semuanya berakhir pada daun.[5]

Penelitian ini menggunakan metode KNN dan Decision Tree untuk mengklasifikasikan sentiment. Kedua metode tersebut dipilih karena memiliki kelebihan dan kekurangan masing-masing. KNN merupakan metode klasifikasi yang sederhana dan mudah diimplementasikan, namun memerlukan waktu yang lama untuk melakukan klasifikasi pada dataset yang besar. Sedangkan Decision Tree merupakan metode klasifikasi yang mudah dipahami dan dapat menghasilkan model yang mudah diinterpretasikan, namun rentan terhadap *overfitting*. Selain itu, akan mengolah data dari repository Kaggle dan menggunakan bahasa pemrograman Python. Selain itu, akan dilakukan evaluasi performa pada kedua teknik analisis sentimen.

Selain itu, dalam penelitian ini juga diterapkan teknik SMOTE (Synthetic Minority Over-sampling Technique) untuk mengatasi masalah ketidakseimbangan data. SMOTE adalah metode oversampling yang bekerja dengan cara membuat contoh sintetik baru dari data minoritas, sehingga memperbesar jumlah data minoritas tanpa melakukan duplikasi. Teknik ini penting karena data yang tidak seimbang dapat mempengaruhi kinerja model klasifikasi, terutama dalam kasus analisis sentimen di mana opini positif dan negatif bisa sangat tidak seimbang.

Dengan menggunakan SMOTE, diharapkan performa model KNN dan Decision Tree dapat ditingkatkan dalam mengklasifikasikan sentimen secara lebih akurat dan handal.

## **1.2 Rumusan Masalah**

Rumusan permasalahan berdasarkan identifikasi masalah pada penelitian ini adalah “Bagaimana implementasi Algoritma Synthetic Minority Over-sampling Technique (SMOTE)?”.

## **1.3 Batasan Masalah**

Batasan masalah yang ditentukan oleh peneliti adalah sebagai berikut :

1. Penelitian ini berfokus pada implementasi Algoritma Synthetic Minority Over-sampling Technique (SMOTE).
2. Data yang digunakan dalam penelitian ini adalah data teks.
3. Algoritma klasifikasi yang digunakan dalam penelitian ini menggunakan Decision Tree dan KNN.
4. Implementasi Algoritma SMOTE dan Algoritma klasifikasi menggunakan Bahasa pemrograman python.
5. Penulisan kode python menggunakan Google Colab.
6. Sumber dataset yang digunakan menggunakan dataset sekunder dari Kaggle.

## **1.4 Tujuan Penelitian**

Tujuan dari penelitian ini adalah:

- a. Mengetahui implementasi Algoritma Synthetic Minority Over-sampling Technique (SMOTE).
- b. Mengetahui pengaruh sebelum dan sesudah implementasi algoritma SMOTE pada algoritma klasifikasi.
- c. Mengetahui hasil tingkat akurasi, presisi, recall dan f1-score sebelum dan sesudah implementasi SMOTE data dataset.

## **1.5 Manfaat Penelitian**

1. Segi Pembaca:

- a. Memberikan kontribusi pada pengembangan metode implementasi Algoritma Synthetic Minority Over-sampling Technique (SMOTE).
  - b. Memperkaya pengetahuan dan pemahaman tentang metode klasifikasi Decision Tree dan KNN.
2. Manfaat bagi Peneliti Selanjutnya:
- a. Mendorong pengembangan metode Algoritma SMOTE yang lebih canggih dan efektif.
  - b. Menjadi sumber data dan referensi untuk penelitian selanjutnya dalam bidang yang sama dengan penelitian ini.

## **1.6 Metodologi Penelitian**

### **1.6.1 Metode Pengumpulan Data**

Dataset yang digunakan pada penelitian ini merupakan data sekunder yang bersumber dari *repository* Kaggle.

### **1.6.2 Metode Klasifikasi**

Algoritma klasifikasi yang digunakan pada penelitian ini dimaksudkan untuk menguji *dataset* menggunakan algoritma klasifikasi KNN dan Decision tree.

### **1.6.3 Metode Mengatasi Ketidak seimbangan Kelas**

Metode yang akan digunakan untuk mengatasi ketidak seimbangan data yaitu Teknik resampling. Teknik resampling ini terdapat 2 metode yaitu *oversampling* dan *undersampling*. Pada penelitian ini penulis menggunakan teknik *resampling oversampling* untuk mengatasi ketidak seimbangan kelas pada dataset yang diimplementasikan terhadap kelas minoritas. Algoritma *oversampling* yaitu Synthetic Minority Over-sampling Technique (SMOTE).

### **1.6.4 Metode Evaluasi**

Pada tahap ini, perbandingan kinerja kedua algoritma klasifikasi dilakukan tanpa teknik resampling dan dengan penambahan SMOTE.

Indikator evaluasi yang digunakan dalam penelitian ini adalah confusion matrix.

### **1.7 Sistematika Penulisan**

Sistematika penulisan ini dibuat untuk mempermudah dalam penyusunan skripsi ini maka perlu ditentukan sistematika penulisan yang baik dan benar. Sistematika penulisannya di bagi dalam beberapa bab, yaitu:

#### **BAB I PENDAHULUAN**

Bab I pendahuluan mendeskripsikan mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan, manfaat, metodologi penelitian, dan sistematika penulisan.

#### **BAB II TINJAUAN PUSTAKA**

Bab II Tinjauan pustaka ini, membahas tinjauan literatur yang berkaitan dengan topik penelitian. Tinjauan ini diambil dari buku, jurnal, seminar, atau prosedur yang sesuai dengan topik penelitian.

#### **BAB III METODE PENELITIAN**

Bab III Mengemukakan metode penelitian yang akan dilakukan dalam alur penelitian untuk Implementasi Algoritma SMOTE.

#### **BAB IV HASIL DAN PEMBAHASAN**

Bab IV Memaparkan dari hasil-hasil tahapan penelitian, mulai dari analisis, hasil testing, dan implementasinya.

#### **BAB V PENUTUP**

Bab V Berisi kesimpulan dan saran dari seluruh penelitian yang telah dilakukan.