

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Perkembangan teknologi yang pesat telah mengubah cara bisnis dan layanan berkomunikasi dengan pelanggan mereka. Web telah menjadi bagian yang tak terpisahkan dari kehidupan sehari-hari, termasuk dalam penggunaan sistem tiket online untuk transportasi umum. Salah satunya adalah aplikasi KAI Access yang diluncurkan oleh PT. KAI untuk memudahkan konsumen dalam memesan tiket. Namun, banyak sekali review negatif terhadap aplikasi tersebut di google play store. Sehingga pihak pengembang aplikasi perlu memperhatikan hal tersebut.

Penelitian ini menggunakan algoritma Naïve Bayes dan SVM untuk analisis sentimen terhadap ulasan produk. Algoritma Naïve Bayes memanfaatkan teorema Bayes untuk menghitung probabilitas sentimen tertentu berdasarkan fitur-fitur yang ada pada teks ulasan. Naïve Bayes juga dapat disebut idiot's Bayes, simple Bayes, dan independence Bayes karena mudah diaplikasikan tanpa memerlukan proses iterasi yang rumit. Algoritma Support Vector Machine adalah Machine Learning yang mengaplikasikan fungsi hyperplane pada data sehingga terbentuk daerah-daerah pada setiap kelas. Hyperplane adalah suatu fungsi yang digunakan sebagai pemisah antar kelas yang ada. Dalam memprediksi suatu kelas dari data, SVM akan melabelinya berdasarkan daerah kelas mana yang merupakan tempat dari data tersebut.

Perbandingan antara metode *Naïve bayes* dan *SVM* pada Comparison of Naïve Bayes and SVM in Sentiment Analysis of Product Reviews on Marketplaces menghasilkan nilai akurasi tertinggi sebesar 85.53% pada fold 2 dan nilai akurasi terendah sebesar 77.16% pada fold 3. Sedangkan algoritma SVM memperoleh nilai akurasi tertinggi sebesar 88.58% pada fold 2 dan nilai akurasi terendah sebesar 82.99% pada fold 7. Dengan ini dinyatakan bahwa SVM dapat bekerja lebih baik untuk analisis sentimen ulasan produk perawatan kecantikan pada marketplace

Shopee karena mendapatkan nilai rata-rata akurasi yang lebih tinggi yaitu 86,14% dibandingkan dengan algoritma Naïve Bayes[1].

Pada penelitian yang dilakukan oleh Agus Heri Yunial dengan judul Analisa Perbandingan Algoritma Klasifikasi *Support Vector Machine*, *Decision Tree* Dan *Naïve Bayes* didapatkan nilai akurasi rata-rata menggunakan algoritma decision tree adalah sebesar 87.43 % dan merupakan nilai akurasi tertinggi dari kedua algoritma lainnya. Algoritma *Support Vector Machine* adalah sebesar 87.16 %, dan nilai akurasi rata-rata menggunakan algoritma *Naive Bayes* adalah sebesar 84.92 %[2].

Pada penelitian lain yang membahas tentang Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Dengan Menggunakan Algoritma Naïve Bayes menunjukkan Hasil evaluasi algoritma Naïve Bayes dengan tiga skenario berbeda dalam klasifikasi sentimen ulasan aplikasi E-Commerce menghasilkan hasil terbaik pada skenario 1 pada aplikasi Shopee (80% data training dan 20% data testing) yang menghasilkan accuracy 92%, precision 92,13%, recall 98,8% dan f1-score 95,35%. Pada aplikasi Tokopedia menghasilkan hasil terbaik pada skenario 3 (60% data training dan 40% data testing) menghasilkan accuracy 83,5%, precision 82,58%, recall 91,6% dan f1-score 91,6%. Pada aplikasi Lazada menghasilkan hasil terbaik pada skenario pada 3 (60% data training dan 40% data testing) menghasilkan accuracy 79,5%, precision 79,4%, recall 100% dan f1-score 88,52%[3].

Pada penelitian ini, dilakukan analisis sentimen terhadap dataset ulasan KAI ACCESS yang diambil dari hasil scrape Google Playstore dengan menggunakan metode Naive Bayes dan SVM. Kumpulan ulasan yang tidak sesuai dengan tujuan penelitian dihapus, sehingga dataset menjadi bersih. Perasaan positif dan negatif.

diberi label menggunakan lexicon Vader, dan menggunakan labelling manual sebagai perbandingan untuk memproses data. Langkah-langkah preprocessing seperti case folding, pembersihan (data cleaning), normalisasi kata, penghilangan stopword, stemming, dan tokenizing juga dilakukan. Setelah itu, data dipisahkan dengan menggunakan teknik K-Fold Cross Validation, TF-IDF. Evaluasi model

menggunakan confusion matrix untuk menghasilkan laporan klasifikasi berupa akurasi, precision, recall, dan F1 score.

Penelitian ini dilakukan perbandingan antara dua algoritma klasifikasi teks Naïve Bayes dan Support Vector Machine dari segi akurasi dan kecepatan proses analisis sentimen terhadap ulasan pada aplikasi KAI ACCESS. Komparasi dari kedua algoritma tersebut diharapkan dapat mengetahui algoritma mana yang memiliki akurasi terbaik dan waktu proses yang cepat.

### **1.2 Rumusan Masalah**

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana hasil analisis sentimen ulasan KAI ACCESS dapat digunakan untuk meningkatkan pengalaman pengguna dan kualitas aplikasi KAI ACCESS?
2. Bagaimana perbandingan antara algoritma *Naive Bayes* dan *SVM* dalam analisis sentimen aplikasi KAI ACCESS?
3. Bagaimana hasil perbandingan antara kinerja labeling menggunakan Lexicon dan Labeling secara manual pada aplikasi KAI ACCESS?

### **1.3 Batasan Masalah**

Permasalahan dalam penelitian ini akan dibatasi dengan beberapa hal yang tercantum pada point-point berikut:

1. Penelitian hanya fokus penelitian terhadap ulasan aplikasi Kai Access pada PlayStore
2. Penelitian ini hanya menggunakan Data yang di scraping sendiri pada aplikasi Kai Access
3. Jumlah data set yang di fokuskan sebanyak 3000 data.
4. Algoritma yang di fokuskan dalam penelitian ini algoritma Naive Bayes, Support vector machine.
5. Data pada penelitian ini diambil pada tanggal 15 mei 2024.

6. Pada proses pelabelan hanya menggunakan labelling lexicon dan labelling manual.

#### **1.4 Tujuan Penelitian**

Tujuan yang akan dicapai pada penelitian ini adalah:

1. Menganalisis kinerja algoritma *Naive Bayes* dan *SVM* pada analisis sentimen aplikasi KAI ACCESS
2. Membandingkan akurasi algoritma *Naive Bayes* dan *SVM* untuk mengetahui algoritma klasifikasi teks yang paling tepat pada analisis sentimen pada aplikasi KAI ACCESS
3. Membandingkan akurasi ketika melakukan labeling menggunakan Lexicon dan labeling secara manual untuk mengetahui pengaruh dari kedua proses labeling tersebut
4. Mengetahui hasil analisis sentimen ini dapat digunakan untuk meningkatkan pengalaman pengguna.

#### **1.5 Manfaat Penelitian**

Manfaat penelitian yang dilakukan adalah:

1. secara praktis, hasil dari penelitian ini dapat membantu dalam mengevaluasi tingkat kepuasan pengguna aplikasi terhadap aplikasi KAI ACCESS
2. secara teoritis, hasil penelitian ini dapat menjadi referensi bagi peneliti selanjutnya mengenai analisis sentimen ulasan aplikasi KAI ACCESS dengan menggunakan algoritma *Naive Bayes* dan *svm*.

#### **1.6 Sistematika Penulisan**

Untuk mempermudah dalam membaca dan mengikuti apa yang dipaparkan dalam penelitian ini, maka penulis telah melakukan penyusunan sistematika penulisan sebagai berikut:

## BAB I PENDAHULUAN

Pada bab ini berisi latar belakang dilakukannya penelitian, perumusan masalah, batasan masalah, tujuan masalah, manfaat penelitian, dan sistematika penulisan yang digunakan dalam menyusun penelitian ini

## BAB II TINJAUAN PUSTAKA

Pada bab ini membahas tentang kajian pustaka dan dasar teori-teori yang memiliki kaitan dengan topik penelitian. kurang lebih pada bab ini akan membahas mengenai teori *Naive Bayes*, *SVM*, analisis sentimen aplikasi KAI ACCESS.

## BAB III METODE PENELITIAN

Pada bab ini memaparkan *flowchart* alur penelitian yang dimulai dari pengumpulan data menggunakan *scraping data*, pengolahan data, implementasi algoritma *Naive Bayes* dan *SVM* untuk data ulasan aplikasi KAI ACCESS, sampai tahapan evaluasi.

## BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang hasil implementasi algoritma *Naive Bayes* dan *SVM* pada analisis sentimen ulasan aplikasi KAI ACCESS dan interpretasi hasil analisis sentimen kedua algoritma berupa penyajian data maupun gambar yang dapat menunjang visualisasi data.

## BAB V PENUTUP

Pada bab ini penulis akan menjelaskan kesimpulan berupa hasil komparasi dari algoritma *Naive Bayes* dan *SVM* pada analisis sentimen ulasan aplikasi KAI ACCESS. penulis juga menyajikan saran untuk memperbaiki kekurangan yang ada pada penelitian ini.