

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Tahun 2020, masyarakat Indonesia hidup ditengah krisis kesehatan yang diakibatkan oleh virus corona[1]. Kementrian Kesehatan telah menerapkan aturan *physical distancing*, *social distancing*, dan *lockdown* guna melawan penyebaran virus corona[1]. Akibat dari peraturan tersebut, semua warga yang berada di wilayah tersebut tidak diperbolehkan keluar rumah termasuk melakukan kegiatan jual beli secara langsung atau tatap muka[2]. Oleh karena itu, selama merebaknya virus corona masyarakat banyak yang menggunakan aplikasi Shopee untuk mendukung aktivitas jual beli[3].

Shopee merupakan aplikasi belanja online terbesar di Indonesia[4]. Menurut data SimilarWeb, Shopee berada di peringkat pertama pada kategori *marketplace* dengan total pengunjung terbesar di Indonesia dalam kuartal ke II tahun 2023 dengan nilai rata-rata 166,9 juta kunjungan perbulan. Selain itu, prestasi Shopee juga terlihat dari posisinya di Google Play Store, di mana aplikasi Shopee berada di peringkat pertama dengan jumlah total unduhan terbanyak daripada aplikasi *marketplace* lainnya.

Jumlah total unduhan Shopee di Google Play Store memang sangat banyak, ini menunjukkan bahwasanya aplikasi Shopee memang banyak di minati oleh masyarakat Indonesia. Akan tetapi, semua itu tidak dapat digunakan sebagai tolak ukur kepuasan pengguna. Dalam konteks ini, kepuasan pengguna hanya dapat diukur dari komentar dan rating yang diberikan oleh pengguna aplikasi Shopee di Google Play Store[5].

Dari latar belakang permasalahan yang sudah dijelaskan diatas, maka akan dilaksanakan sebuah penelitian yang bertujuan untuk melihat komentar masyarakat tentang aplikasi Shopee di Google Play Store cenderung positif atau negatif. Proses pelabelan data akan dilakukan menggunakan dua metode, yaitu metode pelabelan manual dan metode pelabelan otomatis menggunakan *TextBlob*. Setelah itu penelitian ini akan melakukan perbandingan terhadap metode pelabelan manual,

metode pelabelan manual + SMOTE pada implementasi algoritma klasifikasi, metode pelabelan *TextBlob*, dan metode pelabelan *TextBlob* + SMOTE pada implementasi algoritma klasifikasi, menggunakan algoritma klasifikasi *Multinomial Naïve Bayes*, *Random Forest Classifier*, *Logistic Regression*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Extra Trees Classifier*, untuk menentukan algoritma mana yang menghasilkan akurasi terbaik. Temuan dari penelitian ini dapat menjadi panduan bagi pihak manajemen Shopee untuk mengembangkan aplikasinya dan meningkatkan kualitas layanan.

### 1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, akan ada sebuah rumusan masalah sebagai berikut:

1. Bagaimana cara implementasi algoritma *Multinomial Naïve Bayes*, *Random Forest Classifier*, *Logistic Regression*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Extra Trees Classifier* dalam analisis sentimen komentar Shopee di Google Play Store?
2. Berapa hasil akurasi dari algoritma klasifikasi pada metode pelabelan manual, metode pelabelan manual + SMOTE pada implementasi algoritma klasifikasi, metode pelabelan *TextBlob*, dan metode pelabelan *TextBlob* + SMOTE pada implementasi algoritma klasifikasi, dalam analisis sentimen komentar Shopee di Google Play Store?
3. Bagaimana pengaruh pelabelan *TextBlob* dan pelabelan *TextBlob* + SMOTE dalam proses klasifikasi?
4. Bagaimana pengaruh pelabelan manual dan pelabelan manual + SMOTE dalam proses klasifikasi?

### 1.3 Batasan Masalah

Berdasarkan rumusan masalah tersebut, maka permasalahan yang akan dibahas akan dibatasi dengan batasan masalah berikut:

1. Penelitian ini akan difokuskan untuk melihat opini masyarakat tentang aplikasi Shopee di Google Play Store.
2. Penelitian ini akan memakai data komentar dan rating pengguna Shopee

pada *website* Google Play Store.

3. Penelitian ini akan membandingkan kinerja dari metode pelabelan manual, metode pelabelan manual + SMOTE pada implementasi algoritma klasifikasi, metode pelabelan *TextBlob*, dan metode pelabelan *TextBlob* + SMOTE pada implementasi algoritma klasifikasi, dalam analisis sentimen komentar Shopee di Google Play Store.
4. Penelitian ini hanya akan menggunakan dua kelas sentimen saja yaitu sentimen positif dan negatif.
5. Evaluasi performa dari algoritma *Multinomial Naïve Bayes*, *Random Forest Classifier*, *Logistic Regression*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Extra Trees Classifier* akan dilakukan dengan menggunakan *confusion matrix*.

#### 1.4 Tujuan Penelitian

Tujuan yang akan dicapai oleh peneliti dalam penelitiannya adalah sebagai berikut:

1. Penelitian bertujuan untuk mengevaluasi kinerja dari algoritma klasifikasi *Multinomial Naïve Bayes*, *Random Forest Classifier*, *Logistic Regression*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Extra Trees Classifier*.
2. Tujuan dari penelitian ini adalah untuk melihat opini masyarakat tentang aplikasi Shopee di Google Play Store.

#### 1.5 Manfaat Penelitian

Manfaat yang diinginkan dari penelitian ini adalah sebagai berikut:

1. Manfaat Teoritis  
 Penelitian ini akan menghasilkan sebuah pengetahuan yang lebih luas tentang bagaimana algoritma-algoritma yang berbeda berkerja dalam kasus yang sama. Hal ini tentu saja dapat meningkatkan pengetahuan dan pemahaman tentang algoritma-algoritma tersebut.
2. Manfaat Praktis bagi Objek Penelitian  
 Hasil penelitian dapat menghasilkan sebuah informasi yang berharga terhadap pengembang aplikasi Shopee untuk memperbaiki kualitas produk

berdasarkan komentar pengguna.

### 3. Manfaat bagi Peneliti Selanjutnya

Hasil dari penelitian ini seperti dataset dan metodologi yang dipakai, dapat dipakai sebagai sumber referensi bagi peneliti berikutnya yang ingin melakukan penelitian yang sama atau ingin memperbesar penelitian ini ke area yang lebih luas.

## 1.6 Sistematika Penulisan

Penyusunan laporan skripsi akan menggunakan sistematika penulisan yang akan terbagi menjadi beberapa bab sebagai berikut:

### BAB I PENDAHULUAN

Berisi Latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

### BAB II TINJAUAN PUSTAKA

Bab II Tinjauan pustaka ini mencakup :

- A. Study literatur, berisi tentang penelitian sebelumnya yang ada kaitannya dengan penelitian ini.
- B. Dasar teori, berisi penjelasan tentang konsep dasar algoritma *Machine Learning* yang dipakai dalam penelitian. Ini mencakup bagaimana algoritma-algoritma ini bekerja, kelebihan dan kelemahannya. Kemudian dasar teori juga berisi tentang analisis sentimen yang menjelaskan tentang definisi, tujuan dan metode yang digunakan dalam klasifikasi sentimen. Ini mencakup teknik-teknik analisis teks seperti data *cleaning*, *case folding*, *stopword removal*, *tokenizing*, *stemming*, TF-IDF, SMOTE, *TextBlob*, dan klasifikasi sentimen. Dasar teori juga berisi tentang Google Play Store yang digunakan sebagai latar belakang penelitian. Dan yang terakhir dasar teori berisi tentang *confusion matrix* yang dipakai untuk mengevaluasi kinerja algoritma. Ini mencakup *accuracy*, *presisi*, *recall*, dan *F1-Score*.



### **BAB III METODE PENELITIAN**

Bab III akan membahas alur penelitian yang diterapkan dalam penelitian ini. Alur penelitian ini mencakup pengumpulan data, pelabelan, *preprocessing*, *exploratory data analysis*, visualisasi *word clouds*, ekstraksi fitur, SMOTE, klasifikasi dan evaluasi hasil klasifikasi.

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

Bab IV akan membahas hasil dari penelitian yang sudah dilakukan. Ada beberapa aspek yang akan dijelaskan di antaranya adalah ada berapa banyak komentar yang diklasifikasikan sebagai sentimen positif dan negatif. Selain itu, penulis juga akan membahas hasil kinerja dan pengaruh dari metode pelabelan manual, metode pelabelan manual + SMOTE pada implementasi algoritma klasifikasi, metode pelabelan *TextBlob*, dan metode pelabelan *TextBlob* + SMOTE pada implementasi algoritma klasifikasi, menggunakan algoritma klasifikasi *Multinomial Naïve Bayes*, *Random Forest Classifier*, *Logistic Regression*, *Support Vector Machine*, *K-Nearest Neighbors*, dan *Extra Trees Classifier*, disini juga akan dijabarkan hasil evaluasi menggunakan *confusion matrix* yang mencakup akurasi, *precision*, *recall* dan *f1-score*. Selanjutnya, penulis akan membahas tentang kata kata yang sering muncul dalam visualisasi *word clouds*.

### **BAB V PENUTUP**

Bab V akan mencakup tentang kesimpulan dan saran. Kesimpulan dalam penelitian ini digunakan untuk menyajikan ringkasan utama atau pokok dari penelitian yang telah dilaksanakan, dan saran digunakan untuk dapat membantu peneliti dimasa depan dalam membuat penelitian yang lebih relevan.