

BAB 3

METODE PENELITIAN

Penelitian ini melakukan analisis sentimen dari data aplikasi KAI Access. Penelitian ini menggunakan perbandingan metode NBC dan SVM. Penelitian dimulai dengan latar belakang permasalahan yang ada, memetakan proses-proses, mencari sumber permasalahan, dan akhirnya menganalisis data mengenai KAI Access di Google Play Store untuk mengatasi permasalahan yang ada. Penelitian ini membutuhkan data aplikasi KAI access yang diperoleh melalui platform Google Play Store. Di bawah ini adalah bahan, alat, metode dan jalannya penelitian analisis sentimen tentang aplikasi KAI Access beserta langkah-langkah penelitian untuk menyelesaikan analisis sentimen menggunakan data dan merancang sistem untuk menampilkan hasil analisis yang diperoleh.

3.1 BAHAN DAN ALAT PENELITIAN

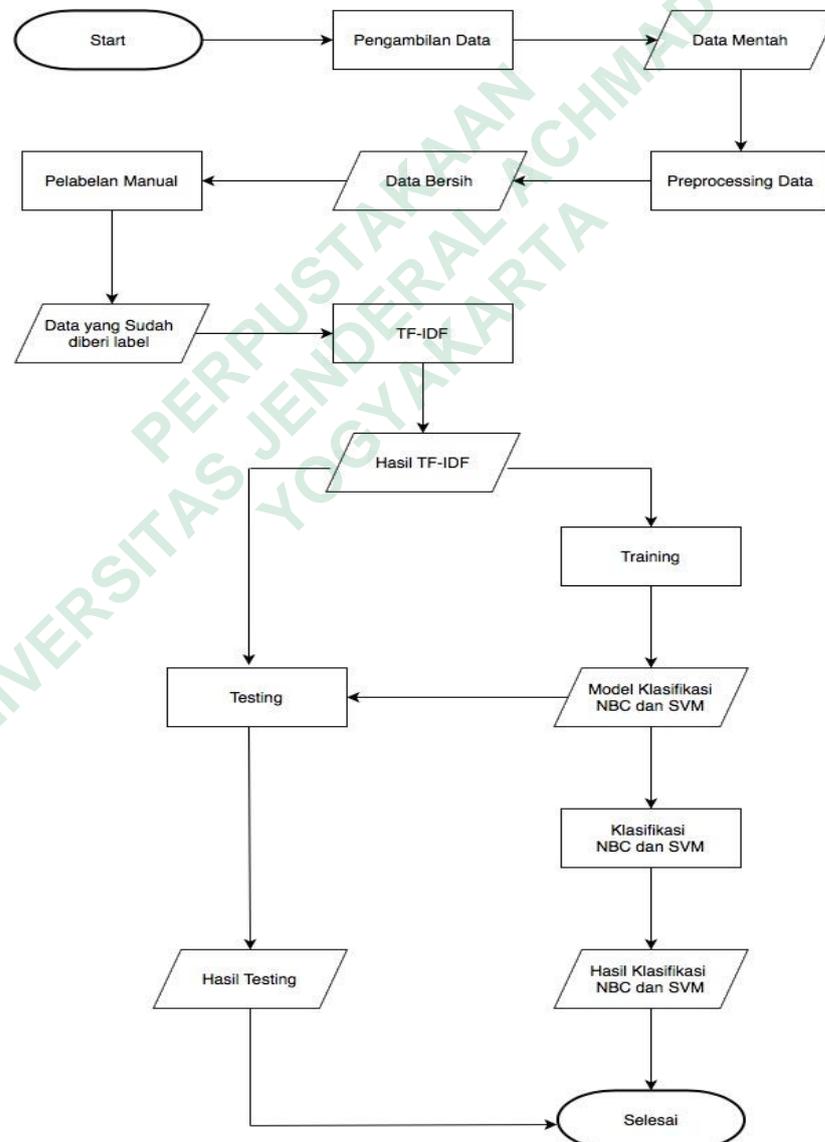
Bahan yang digunakan dalam penelitian ini berasal dari data aplikasi KAI Access di Google Play Store.

Alat yang digunakan dalam pengembangan aplikasi ini adalah :

1. Sistem Operasi : Windows 10 64-bit
2. Bahasa Pemrograman Python 3.10.5
3. Microsoft Office Excel 2019.
4. Anaconda 3 64-bit
5. Sublime Text 4
6. Google Colab
7. Framework Flask
8. Library : NLTK

3.2 JALAN PENELITIAN

Algoritma NBC menggunakan perhitungan probabilitas dan statistik untuk mengklasifikasikan data teks dengan menjumlahkan frekuensi dan kombinasi nilai dari kumpulan data (Devita et al., 2018). Sedangkan SVM merupakan salah satu metode pembelajaran terawasi yang biasa digunakan untuk klasifikasi dan regresi. Dalam pemodelan klasifikasi SVM memiliki konsep yang lebih lengkap dan jelas secara matematis, SVM juga dapat menyelesaikan masalah secara linier maupun *non linear*. Berikut ini merupakan alur penelitian serta tahap tahapan yang digunakan dalam penelitian ini ditunjukkan pada gambar 3.1.



Gambar 3.1 Alur Penelitian

1. Pengambilan data

Pada tahap ini dilakukan pengambilan data menggunakan metode scraping dan pengolahan data *review* KAI Access dari Google Play Store. Yang akan dilakukan menggunakan API yang sudah disediakan oleh Google Play Store, kemudian data yang telah diambil akan ditampilkan ke dalam CSV Microsoft Excel.

2. Preprocessing data

Tahap ini akan dilakukan pemrosesan data dengan cara *cleaning* data dan *normalization* yang digunakan untuk meningkatkan tingkat akurasi. Tahapan preprocessing terdiri dari 4 tahapan yaitu, data *cleaning*, data *tokenization*, *stopword removal* dan *stemming data*.

3. Pelabelan manual

Tahap ini akan dilakukan pelabelan secara manual setiap kalimat guna adanya proses analisis lebih lanjut. Label yang akan digunakan pada penelitian ini untuk menentukan hasil yaitu positif, negatif dan netral.

4. Training data

Pada tahapan ini dilakukan klasifikasi menggunakan metode NBC dan SVM yang digunakan untuk perbandingan, kemudian akan menggunakan TF-IDF untuk perbandingan setiap kata pada dokumen yang dilakukan secara otomatis.

5. Testing

Proses testing dilakukan guna mengetahui tingkat akurasi yang dihasilkan dari data yang sudah dilabeli dan di uji. Untuk mengetahui presentase dari setiap percobaan maka model akan dihitung nilainya menggunakan beberapa metode (Saputro & Sari, 2019), yaitu :

- a. Accuracy digunakan untuk melihat hasil dari keseluruhan data klasifikasi menggunakan total sampel testing yang diuji.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (5)$$

- b. Recall digunakan untuk melihat hasil dari data klasifikasi yang memiliki kategori positif dibagi dengan total sampel dalam testing dengan kategori positif.

$$\text{Recall} = \frac{TP}{TP+FP} * 100\% \quad (6)$$

- c. Precision digunakan untuk mengetahui hasil data klasifikasi kategori yang benar dibagi total sampel klasifikasi positif.

$$\text{Precision} = \frac{TP}{TP+FP} * 100\% \quad (7)$$

- d. F-Measure digunakan untuk menghitung rata-rata jumlah perhitungan recall dan precision.

$$\text{F-Measure} = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (8)$$

Keterangan :

True positive (TP) : Jumlah record positif dalam dataset yang diklasifikasikan positif.

True negative (TN) : Jumlah record negative dalam dataset yang diklasifikasikan positif.

False positive (FP) : Jumlah record negative dalam dataset yang diklasifikasikan positif.

False negative (FN) : Jumlah record positif dalam dataset yang diklasifikasikan negatif.

6. Klasifikasi

Berdasarkan data yang diuji akan menggunakan metode NBC dan SVM yang menghasilkan klasifikasi dengan kategori positif, negative dan netral. Hasil dari pengklasifikasian ini akan ditampilkan dalam bentuk dashboard.

3.3 DESAIN USER INTERFACE

User Interface adalah tampilan yang berhubungan langsung dengan pengguna sistem. Perancangan *user interface* dilakukan agar memudahkan dalam

implementasi kode program dan untuk membuat tampilan lebih dipahami oleh pengguna. Berikut merupakan deskripsi atau desain user interface untuk analisis sentimen yang mengevaluasi aplikasi KAI Access di Google Play Store dengan membandingkan metode NBC dan SVM.

3.3.1 Halaman Upload File

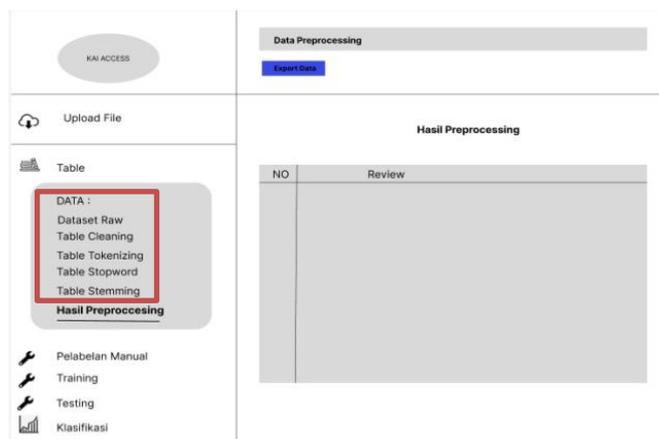
Halaman upload file KAI Access digunakan untuk melakukan upload file dan menampilkan file CSV yang sudah terupload ke dalam *directory* agar memudahkan proses preprocessing, training, testing dan klasifikasi. Implementasi halaman upload file KAI Access. Tampilan menu upload data ditunjukkan pada Gambar 3.2.



Gambar 3.2 Desain *Interface* Upload File

3.3.2 Preprocessing data

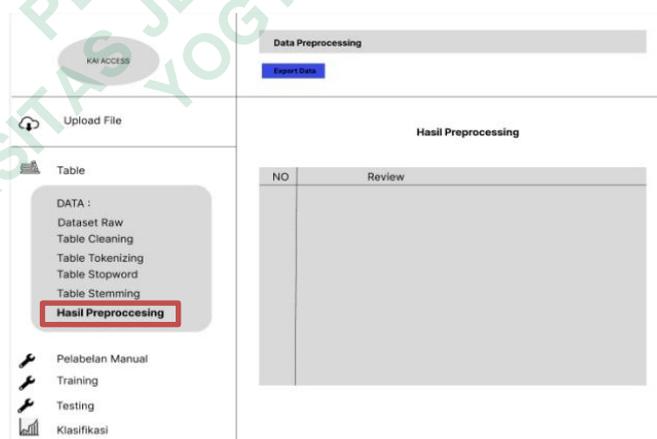
Halaman *preprocessing* pada sistem analisis sentimen merupakan proses pengolahan data *review* yang ada dengan melakukan langkah-langkah untuk menyempurnakan data *review* yang masih memiliki komponen sehingga menjadi data yang diinginkan. Tampilan menu *preprocessing* data ditunjukkan pada Gambar 3.3.



Gambar 3.3 Desain *Interface Preprocessing Data*.

3.3.3 Hasil Preprocessing data

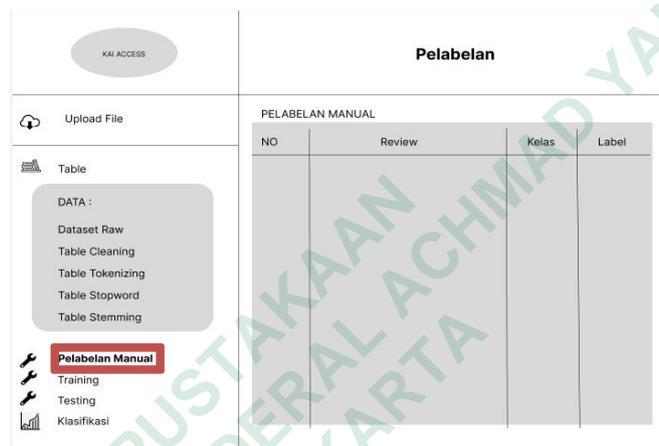
Halaman hasil *preprocessing* pada sistem analisis sentimen merupakan hasil data *review* yang sudah dilakukan proses *preprocessing*. Data *review* tersebut menjadi lebih terstruktur agar dapat digunakan untuk melakukan perhitungan di tahap selanjutnya. Tampilan menu *preprocessing* data ditunjukkan pada Gambar 3.4.



Gambar 3.4 Desain *Interface Hasil Preprocessing Data*.

3.3.4 Pelabelan Manual

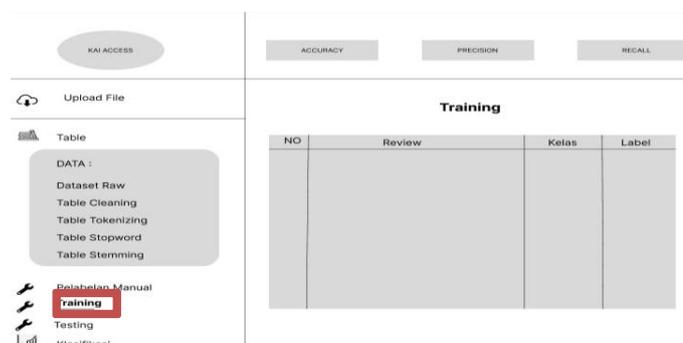
Halaman pelabelan manual pada sistem analisis sentimen merupakan proses untuk menampilkan, meninjau data yang telah diberi kelas dan label, kalimat dalam maninjau data yang telah diberi kelas dan label terhadap kalimat yang ada pada data review secara manual. Proses pelabelan manual dilakukan pada Microsoft Excel. Tampilan menu pelabelan manual ditunjukkan pada Gambar 3.5.



Gambar 3.5 Desain *Interface* Pelabelan Manual

3.3.5 Training Data

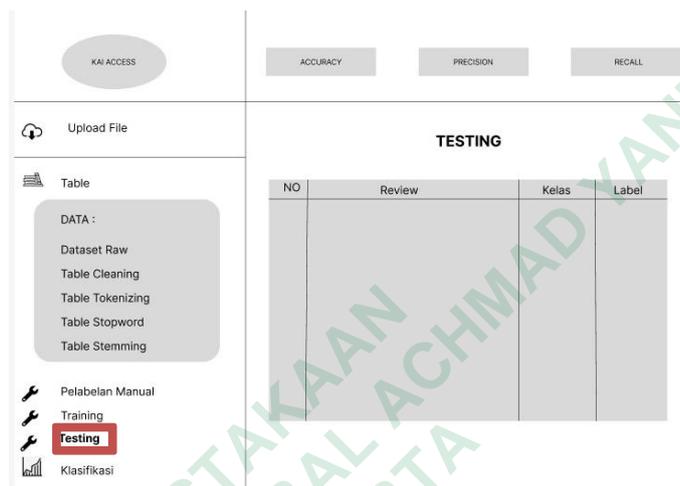
Training pada sistem analisis sentimen menggunakan ekstraksi Term Frequency-Invers Document Frequency (TF-IDF) pada data teks untuk menghasilkan model klasifikasi yang digunakan untuk menampilkan proses klasifikasi sentiment secara otomatis. Tampilan menu *Training* data ditunjukkan pada Gambar 3.6.



Gambar 3.6 Desain *Interface* Training Data.

3.3.6 Testing Data

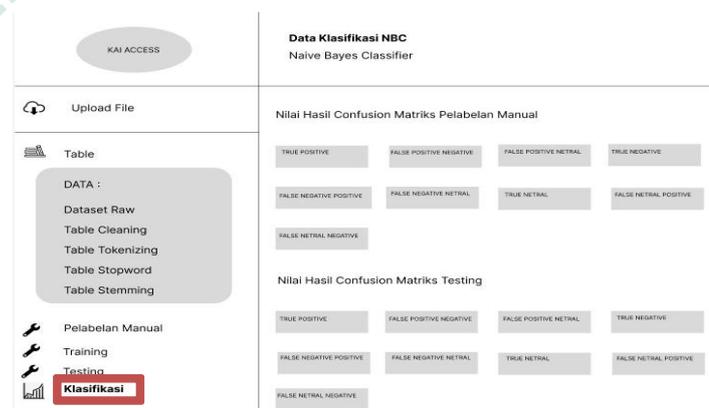
Testing pada sistem analisis sentimen dilakukan untuk mengetahui seberapa akurat model yang dibangun dalam *training* data digunakan untuk memprediksi kelas atau label dari data uji yang tersedia. Tampilan menu *testing* data ditunjukkan pada Gambar 3.7.



Gambar 3.7 Desain *Interface* Testing Data

3.3.7 Klasifikasi Naïve Bayes Classifier

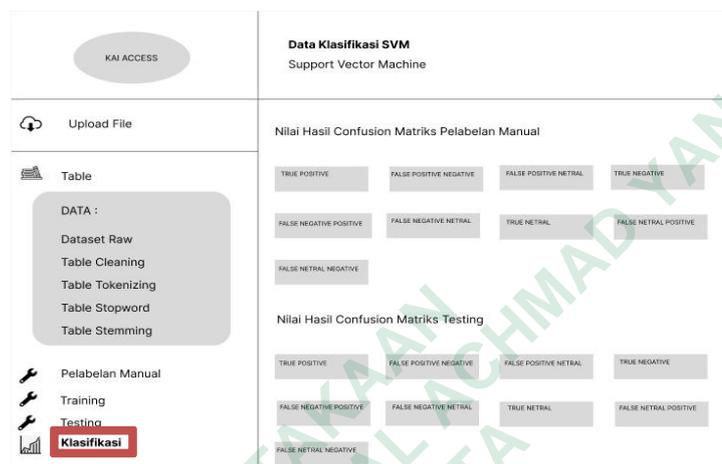
Klasifikasi NBC pada sistem analisis sentimen dilakukan untuk menampilkan data *review* yang akan diprediksi kelas, label yang sudah dibangun pada *training* data. Tampilan menu klasifikasi Naïve Bayes Classifier ditunjukkan pada Gambar 3.8.



Gambar 3.8 Desain *Interface* Klasifikasi Naïve Bayes

3.3.8 Klasifikasi Support Vector Machine

Klasifikasi SVM pada sistem analisis sentimen dilakukan untuk menampilkan data *review* yang akan diprediksi kelas, label yang sudah dibangun pada *training* data. Tampilan menu klasifikasi Support Vector Machine ditunjukkan pada Gambar 3.9.



Gambar 3.9 Desain *Interface* Support Vector Machine