BAB 3
METODE PENELITIAN

Penelitian ini merupakan penelitian analisis sentimen dengan menggunakan data dari Twitter. Penelitian ini menggunakan metode Naive Bayes Classification. Penelitian dimulai dengan latar belakang permasalahan yang ada, memetakan proses-proses, mencari sumber permasalahan, dan akhirnya menganalisis data mengenai Telkomsel di media sosial Twitter untuk mengatasi permasalahan yang ada. Penelitian ini membutuhkan data *tweet* yang diperoleh melalui media sosial Twitter yang berkaitan dengan Telkomsel. Dibawah ini adalah bahan, alat, metode dan jalannya penelitian analisis sentimen tentang Telkomsel beserta langkah-langkah penelitian untuk menyelesaikan analisis sentimen menggunakan data *tweet* dan merancang sistem untuk menampilkan hasil analisis yang diperoleh.

PERPUSTAKAAN UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA

3.1 BAHAN DAN ALAT PENELITIAN

Penelitian ini akan mengidentifikasi data dan informasi dari berbagai sumber. Sumber utama penelitian ini diamati dari *tweet* terkait Telkomsel di media sosial Twitter, buku tentang analisis sentimen. Selain itu, referensi lain yang terkait dengan penelitian ini diperoleh dari *literature* yang relavan seperti jurnal, penelitian sebelumnya, dan informasi internet.

Alat yang digunakan dalam penelitian ini adalah komputer dengan spesifikasi yang cukup untuk menjalankan sistem operasi dan pengembangan perangkat lunak serta koneksi Internet.

Sistem Operasi dan program aplikasi yang digunakan untuk mengembangkan aplikasi ini adalah:

1. Processor : Intel(R) Celeron(R) CPU 1005M @ 1.90GHz (2CPUs), ~1.9GHz
2. RAM : 10 GB

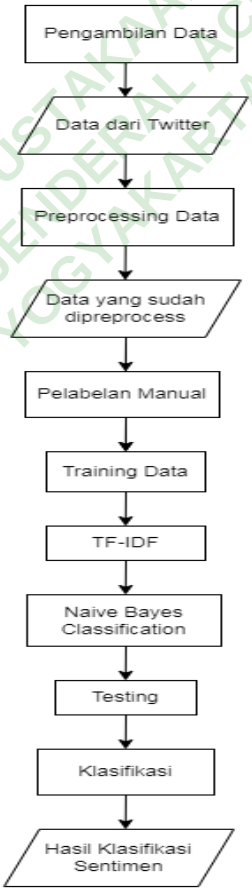
Sistem Operasi dan program-program aplikasi yang dipergunakan dalam pengembangan aplikasi ini adalah :

1. Sistem Operasi : Windows 7

2. Bahasa Pemrograman : Python 3.7.3
3. Framework : Flask, Bootstrap 4
4. Text editor : Sublime Text, Jupyter Notebook

3.2 JALAN PENELITIAN

Naive Bayes Classification merupakan salah satu metode *machine learning* dalam klasifikasi data teks yang memanfaatkan perhitungan probabilitas dan statistik dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset (Devita et al., 2018). Dalam metode ini, semua langkah akan dijalankan selangkah demi selangkah dan dieksekusi secara berurutan. Dibawah ini adalah alur penelitian dan langkah-langkah metode Naive Bayes Classification yang akan digunakan dalam penelitian ini seperti terlihat pada Gambar 3.1 dibawah ini.



PERPUSTAKAAN UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA

Gambar 3.1 Alur Penelitian

1. Tahap Pengambilan data, pada tahap ini dilakukan pengambilan data di media sosial Twitter untuk mendapatkan *tweet* mengenai Telkomsel. Proses ini dilakukan menggunakan Anaconda Prompt dan dieksekusi pada Jupyter Notebook, data *tweet* yang di dapat berformat file *Comma Separated Values* (CSV) sehingga dapat dibuka di Microsoft Excel.
2. Tahap *preprocessing* data, pada tahap ini dilakukan proses pengolahan data teks untuk data yang belum terstruktur, sehingga perlu diperbaiki dengan melakukan tahapan-tahapan berikut ini:
    a. *Tokenization* merupakan proses untuk pemecahan atau pemisahan kalimat dalam suatu teks yang disebut token.
    b. *Case folding* merupakan proses untuk mengubah setiap huruf pada *tweet* menjadi *lowercase* atau huruf kecil.
    c. *Stopword removal* merupakan penghapusan kata yang memiliki informasi rendah pada teks.
    d. *Number removal* merupakan proses untuk membersihkan data teks dari karakter angka.
    e. *Punctuation removal* merupakan proses penghapusan karakter pada data teks.
    f. *Whitespaces removal* merupakan proses menghilangkan spasi dari awal dan di akhir kalimat.
    g. *Normalization* merupakan standarisasi kata dengan salah eja atau penggunaan bahasa yang tidak biasa.
    h. *Stemming* merupakan penghilangan infleksi pada kata menjadi bentuk dasarnya.
3. Tahap Pelabelan manual, pada tahap ini dilakukan proses pelabelan manual untuk memberikan sentimen positif, netral atau negatif tehadap kata yang ada pada dokumen agar dapat dianalisis.
4. Tahap *training* data, pada tahap ini dilakukan proses *training data* yang di awali dengan fitur ekstraksi dengan menggunakan *term frequency-invers document frequency* (TF-IDF) pada data teks. Proses *training data* digunakan untuk melakukan klasifikasi sentimen secara otomatis.

PERPUSTAKAAN UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA

5. Tahap *testing*, pada tahap ini dilakukan proses *testing* untuk mengukur dan menentukan seberapa akurat model yang telah dibangun pada tahap *training*¸ hasil *testing* yang diprediksi kelas dan label akan dihitung dengan cara menghitung *confusion matrix* untuk mengetahui *persentase* dari setiap pengujian.
6. Tahap Klasifikasi, pada tahap ini merupakan proses dimana hasil data *tweet* yang sudah diprediksi kelas dan labelnya secara otomatis akan ditampilkan dalam bentuk dashboard menggunakan *framework* Flask bahasa pemrograman Python.

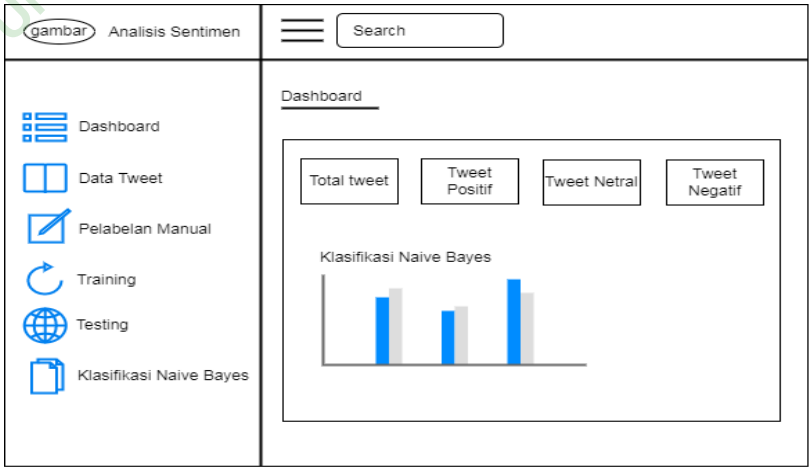
3.3 DESAIN INTERFACE

Interface merupakan gambaran yang berhubungan langsung dengan pengguna sistem. Perancangan antar muka dibuat agar sistem yang dihasilkan lebih menarik dan mudah dipahami pada saat perancangan. Berikut ini adalah deskripsi atau desain-desain *interface* pada Analisis Sentimen Kepuasan Pelanggan Perusahaan Telekomunikasi Seluler Telkomsel Di Twitter:

PERPUSTAKAAN UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA

3.3.1 Dashboard Analisis Sentimen

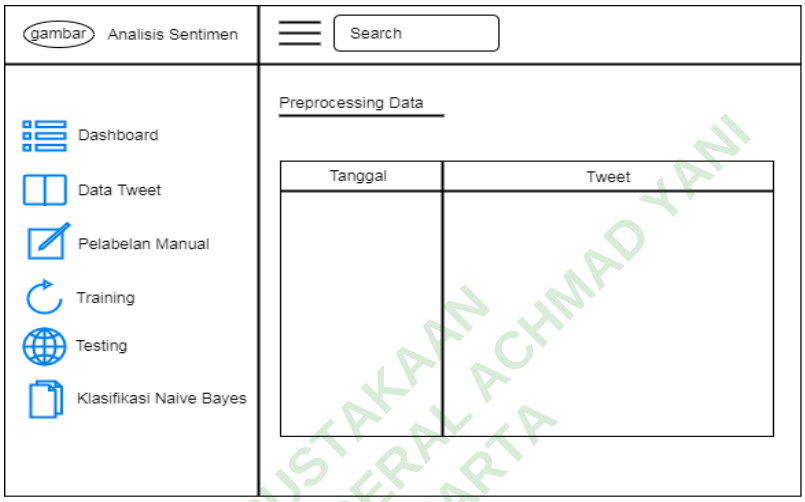
Pada sistem analisis sentimen dapat mengetahui data *tweet* yang telah didapatkan kelas dan labelnya berdasarkan prediksi yang sudah dibuat pada model *training* dan ditampilkan pada menu dashboard. Tampilan menu dashboard ditunjukkan pada Gambar 3.2.



Gambar 3.2 Desain *Interface* Halaman Dasboard

3.3.2 *Preprocessing*

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

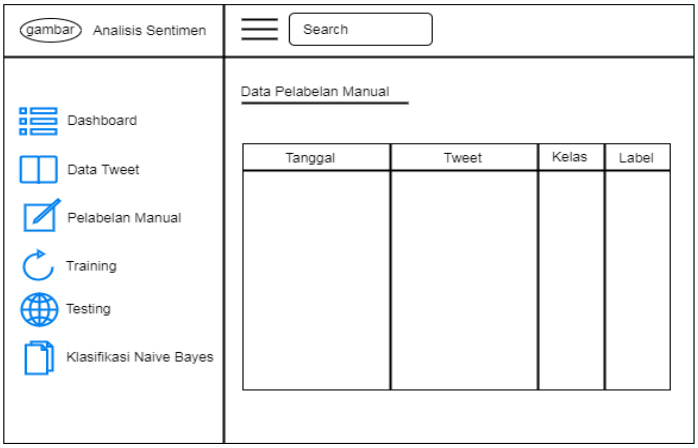


Gambar 3.3 Desain *Interface* Halaman *Preprocessing* Data

PERPUSTAKAAN UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA

3.3.3 Pelabelan Manual

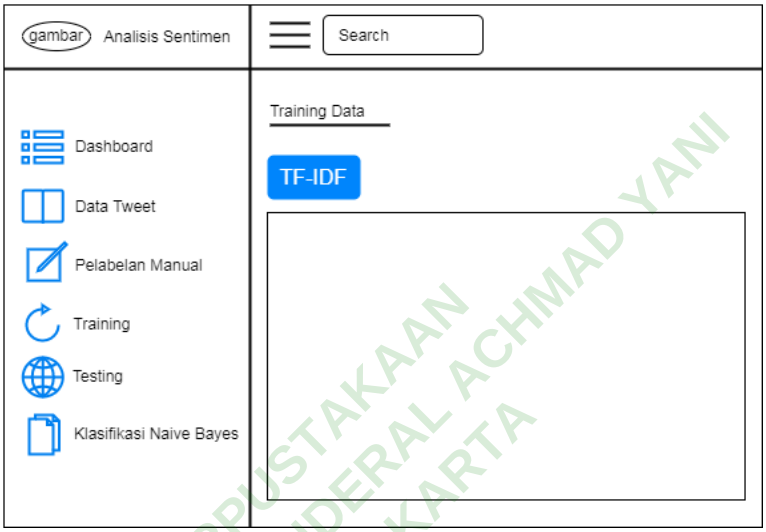
Pelabelan manual pada sistem analisis sentimen merupakan proses untuk menampilkan data *tweet* yang sudah diberi kelas dan label terhadap kalimat yang ada pada data *tweet* secara manual. Proses pelabelan manual dilakukan pada Microsoft Excel. Tampilan menu pelabelan manual ditunjukkan pada Gambar 3.4.



Gambar 3.4 Desain *Interface* Halaman Pelabelan Manual

3.3.4 *Training*

Training pada sistem analisis sentimen menggunakan ekstraksi *Term Frequency-Invers Document Frequency* (TF-IDF) pada data teks untuk menghasilkan model klasifikasi yang dapat digunakan untuk menampilkan proses klasifikasi sentimen secara otomatis. Tampilan menu *training* data ditunjukkan pada Gambar 3.5.

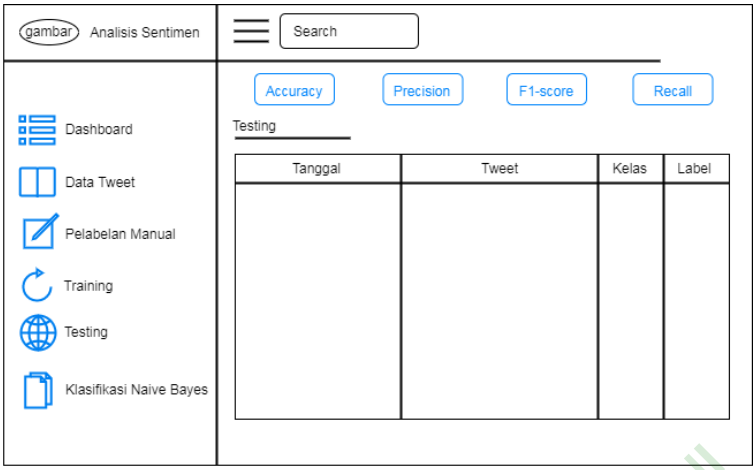


Gambar 3.5 Desain *Interface* Halaman *Training*

PERPUSTAKAAN UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA

3.3.5 *Testing*

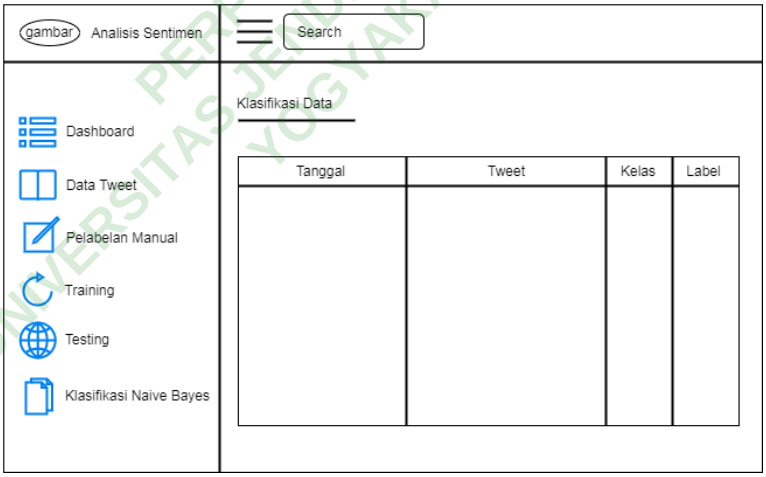
Testing pada sistem analisis sentimen dilakukan untuk mengetahui seberapa akurat model yang dibangun dalam *training* data digunakan untuk memprediksi kelas atau label dari data uji yang tersedia. Tampilan menu *training* data ditunjukkan pada Gambar 3.6.



Gambar 3.6 Desain *Interface* Halaman *Testing*

3.3.6 Klasifikasi Naive Bayes

Klasifikasi naive bayes pada sistem analisis sentimen dilakukan untuk menampilkan data *tweet* yang akan diprediksi kelas, label yang sudah dibangun pada *training* data. Tampilan menu klasifikasi Naive Bayes ditunjukkan pada Gambar 3.7.



Gambar 3.7 Desain *Interface* Klasifikasi Naive Bayes

PERPUSTAKAAN UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA