

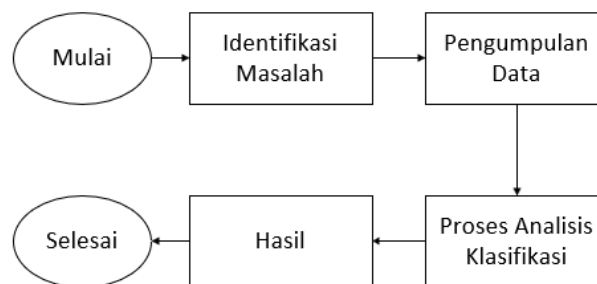
## BAB III METODE PENELITIAN

### 3.1 Objek Penelitian

Penelitian ini difokuskan pada ulasan aplikasi Wondr by BNI dengan menerapkan metode klasifikasi *Naive Bayes* untuk mengelompokkan ulasan pengguna ke dalam kategori seperti *cepat*, *biasa saja*, *lambat*, *tidak responsif*. Penelitian ini bertujuan untuk memahami persepsi serta pengalaman pengguna terhadap aplikasi tersebut dan bagaimana mesin tersebut bisa memprediksi dengan benar secara otomatis dengan kategori tersebut. Dataset yang digunakan dalam penelitian ini diperoleh dari platform *Kaggle*, yang menyediakan ulasan pengguna terkait aplikasi Wondr by BNI.

### 3.2 Kerangka Penelitian

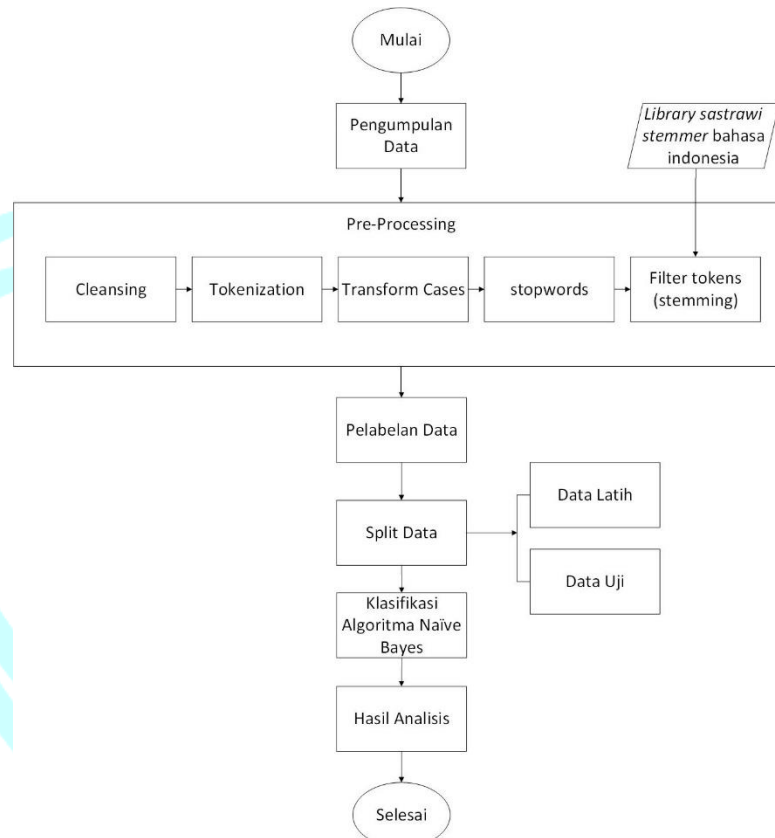
Penelitian ini dimulai dengan pengumpulan data ulasan pengguna aplikasi Wondr by BNI yang bersumber dari platform *Kaggle*, yang dikenal sebagai repositori dataset publik populer untuk keperluan analisis dan penelitian. Setelah data terkumpul, tahap selanjutnya adalah pre-processing data, meliputi cleansing, tokenization, transform cases, penghapusan stopwords, dan filter token guna mendapatkan kata dasar. Proses ini bertujuan meningkatkan kualitas data sehingga siap untuk analisis lebih lanjut. Proses ini bertujuan untuk meningkatkan kualitas dataset dan mempermudah analisis selanjutnya. Metode ini menggunakan model algoritma *Naive Bayes* untuk mengevaluasi kinerja dalam klasifikasi dan prediksi ulasan aplikasi Wondr By BNI dengan menggunakan metrix seperti Akurasi, Presisi, Recall, dan F1-score.



Gambar 3. 1 Kerangka Penelitian

### 3.3 Alur Proses Penelitian

Berikut ini merupakan tahapan penelitian yang ditunjukkan pada Gambar 3.2. Tahapan tersebut menggambarkan proses pelaksanaan penelitian secara sistematis dan terstruktur.



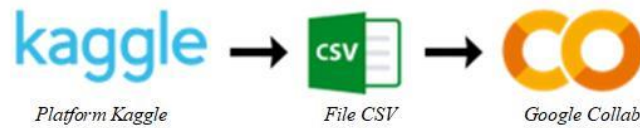
Gambar 3. 2 Alur Penelitian

#### 3.3.1 Pengumpulan Data

Pada tahap ini, penelitian menggunakan data berupa ulasan dari pengguna aplikasi Wondr by BNI. Dataset tersebut diperoleh melalui *Kaggle*, sebuah website populer yang menyediakan beragam dataset publik untuk mendukung riset berbasis data. Dataset ini berisi kumpulan teks ulasan pengguna yang mencakup pendapat, masukan, keluhan, hingga apresiasi terhadap kinerja dan layanan aplikasi yang digunakan.

Proses pengumpulan data ulasan dilakukan dengan mengunduh dataset dari *Kaggle* dalam format *CSV (Comma Separated Values)* dengan keyword aplikasi bank Wondr by BNI. Format ini dipilih karena mempermudah proses pengolahan data pada tahapan berikutnya, terutama saat menggunakan tools

seperti *Google Collab* dengan bantuan pustaka pendukung seperti *Pandas* dan *NumPy*. Berikut gambar tahapan dari pengumpulan data



Gambar 3. 3 Tahapan Pengumpulan Data

### 3.3.2 Pre-Processing

Setelah data ulasan berhasil dikumpulkan, langkah selanjutnya adalah tahapan pre-processing. Pada tahap ini, dataset akan dipersiapkan dan dibersihkan agar layak digunakan untuk proses klasifikasi menggunakan algoritma *Naive Bayes*. Tahapan ini bertujuan untuk memastikan data ulasan yang digunakan memiliki kualitas yang baik dan siap untuk dianalisis lebih lanjut. Adapun tahapan pre-processing sebagai berikut :

#### a. Cleansing

Pada tahap cleansing ini Tahap *cleansing* dilakukan untuk membersihkan teks dari elemen-elemen yang tidak relevan atau tidak diperlukan. Proses ini meliputi:

- 1) Menghapus tanda baca, angka, simbol khusus, dan karakter yang tidak memiliki kontribusi signifikan terhadap analisis.
- 2) Menghilangkan spasi berlebih maupun karakter kosong yang terdapat dalam teks.
- 3) Menghapus URL, alamat email, atau elemen lainnya yang tidak berkaitan dengan tujuan analisis teks.

Proses *cleansing* ini memastikan bahwa data ulasan yang digunakan menjadi lebih bersih dan siap untuk diproses pada tahapan selanjutnya.

#### b. Tokenization

Pada tahap selanjutnya yaitu *tokenization*, teks yang telah dibersihkan dipecah menjadi potongan-potongan kecil atau kata-kata yang disebut *token*. Tujuan tokenisasi adalah memisahkan teks menjadi unit yang lebih kecil agar lebih mudah dianalisis.

c. Transform Cases

Tahap ini bertujuan untuk menyamakan format teks dengan mengubah semua huruf menjadi huruf kecil lowercase dan bisa menjadikan huruf besar uppercase. Langkah ini dilakukan untuk memastikan bahwa tidak ada perbedaan interpretasi antara kata yang menggunakan huruf kapital maupun huruf kecil, sehingga analisis dapat dilakukan secara konsisten.

d. Stopwords

Tahap ini bertujuan untuk menghapus *stopwords*, yaitu kata-kata umum yang sering muncul

e. Filter Tokens (stemming)

Proses *filter tokens (stemming)* bertujuan untuk menyederhanakan setiap kata dalam teks menjadi bentuk dasarnya atau akar katanya. Tahapan ini bertujuan untuk mengurangi variasi kata yang memiliki arti sama tetapi berbeda dalam bentuknya, seperti "berlari", "lari", dan "lari-larian", yang semuanya diubah menjadi kata dasar "lari".

### 3.3.3 Pelabelan Data

Setelah melewati pre-processing data selanjutnya, data ulasan pengguna yang telah terkumpul akan diberikan label sesuai dengan kategori yang telah ditetapkan. Proses pelabelan dapat dilakukan melalui dua pendekatan utama, yaitu *manual* dan *otomatis*. Proses pelabelan ini terdapat 1500 data yang akan diuji dan di labeli manual dan akan di proses klasifikasi menjadikan otomatisasi menggunakan model *Naïve Bayes* dan hasil akan menunjukkan perbandingan antara pelabelan manual dan otomatis.

**a. Pelabelan Manual:** Metode ini melibatkan peneliti untuk membaca, menganalisis, dan menentukan label ulasan secara langsung berdasarkan kata kunci atau makna konteks yang relevan.

**b. Pelabelan Otomatis:** Menggunakan pendekatan berbasis aturan (*rulebased labeling*) dengan bantuan skrip *Python* atau metode sentiment analysis untuk mempercepat proses klasifikasi. Dalam penelitian ini, ulasan pengguna dikelompokkan ke dalam beberapa kategori, seperti cepat, biasa saja, lambat, dan tidak responsif. Proses

pemberian label ini bertujuan untuk memahami sentimen atau penilaian pengguna terhadap kinerja aplikasi berdasarkan opini yang mereka sampaikan.

### **3.3.4 Split Data**

Tahap berikutnya dalam penelitian adalah pembagian data (split data), yang bertujuan untuk memisahkan dataset yang telah diproses menjadi dua bagian, yaitu data pelatihan (training data) dan data pengujian (testing data). Proses ini dengan komposisi pembagian sebesar 80% untuk data latih dan 20% untuk data uji. Data pelatihan digunakan untuk membangun dan melatih model, sedangkan data pengujian digunakan untuk mengukur kinerja model tersebut. Tujuannya adalah untuk menilai kemampuan model dalam mengklasifikasikan data baru yang belum pernah dilihat sebelumnya.

### **3.3.5 Klasifikasi Algoritma *Naïve Bayes***

Tahapan ini merupakan inti dari penelitian, yaitu melakukan klasifikasi data menggunakan algoritma *Naïve Bayes*. *Naïve Bayes* adalah metode klasifikasi berbasis probabilitas yang bekerja dengan prinsip teorema Bayes. Algoritma ini efektif untuk masalah klasifikasi teks, seperti ulasan aplikasi. Hasil dari proses klasifikasi ini akan dianalisis menggunakan confusion matrix, yang menjadi elemen penting dalam menghitung berbagai metrik evaluasi model.

### **3.3.6 Evaluasi dan Visualisasi**

Evaluasi dilakukan untuk mengukur kinerja model klasifikasi *Naïve Bayes* yang telah dibangun. Penilaian dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score guna mengetahui seberapa baik model dalam mengelompokkan ulasan pengguna ke dalam kategori yang sesuai.

Pengujian dilakukan terhadap data uji yang telah dipisahkan dari data latih dengan rasio 80:20. Salah satu alat evaluasi yang digunakan adalah confusion matrix, diagram batang, dan WordCloud yang menunjukkan jumlah prediksi benar dan salah pada tiap kategori label.

Dengan adanya proses evaluasi dan visualisasi ini, peneliti dapat menilai performa model *Naïve Bayes* secara lebih objektif melalui pendekatan numerik maupun visual. Selain itu, hasil analisis ini juga memberikan dasar pertimbangan dalam melakukan pengembangan aplikasi berdasarkan respon dan kebutuhan pengguna.

