BAB 3
METODOLOGI PENELITIAN

3.1 Metode Penelitian

Metode penelitian analisis sentimen terhadap Ulasan Aplikasi Andal by Taspen pada Google Play Store menggunakan metode klasifikasi Algoritma Multinomial Naive Bayes dilakukan dalam langkah-langkah berikut:

a. Studi Literatur : Dalam tahap pertama penelitian, dan tujuan dari tahap ini adalah untuk memperoleh pengetahuan mengenai topik yang sedang diteliti. Informasi ini dikumpulkan dari berbagai sumber, termasuk buku, artikel, jurnal, dan website. Analisis Sentimen, TF-IDF, Andal by Taspen, Google Play Store, Multinomial Naive Bayes, dan Matrix confusion adalah teori-teori yang relevan dalam penelitian ini.

b. Pengambilan Data (Data Scrapping) : Data dikumpulkan dengan menggunakan metode pengambilan data atau biasa yang dikenal scraping data melalui pustaka resmi dari yang disediakan oleh google langsung yaitu google play scraper. Google play-scraper library digunakan untuk mengumpulkan ulasan yang ada di google play store berdasarkan nama package dari aplikasi tersebut. Penelitian ini akan mengumpulkan 5071 ulasan dari aplikasi Andal oleh Taspen dari tanggal 1 Januari 2024 hingga 16 Mei 2025..

c. Perancangan Tahapan : Perancangan sistem dimulai dengan pembuatan alur, penelitian atau flowchart, pelabelan dengan berbasis lexicon based vader, proses pre-processing teks, fitur ekstrasi TF-IDF, dan pembagian data latihan dan tes.

d. Pengujian dan Evaluasi : Setelah data melalui tahap training, akan dilakukan pengujian dimana data diklasifikasikan ke bentuk sentimen. Hasil pengujian tersebut dievaluasi mengenai performa, akurasi, dan hasilnya oleh sistem.

e. Dokumentasi : Dijelaskan setiap langkah dari awal penelitian hingga desain dan model klasifikasi. Ini dilakukan secara bertahap, dari bagian pendahuluan hingga kesimpulan dan rekomendasi..

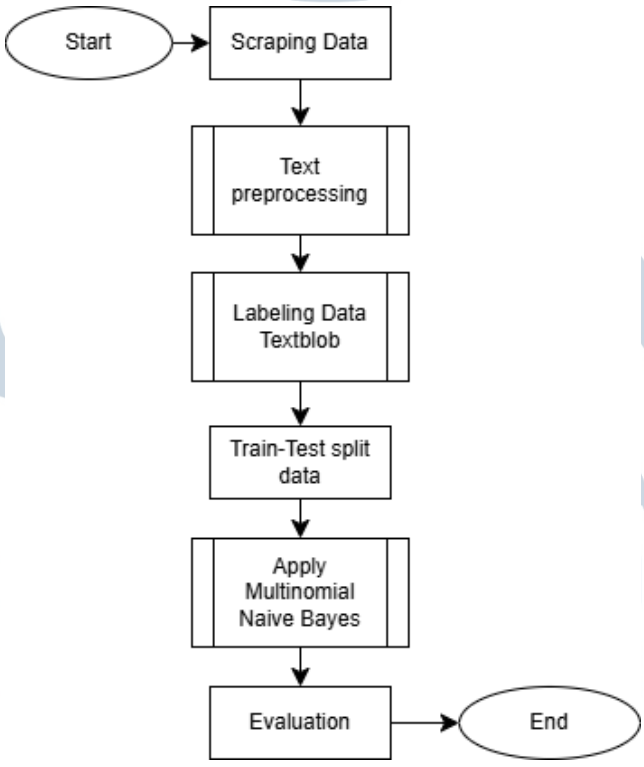
f. Konsultasi : Dalam tahap ini, merupakan berkonsultasi dan mendiskusikan penelitiannya kepada pembimbing. Tujuan dari tahap konsultasi ini adalah untuk mendapatkan saran dan pendapat tentang proses dan laporan penelitian.

3.2 Gambaran Umum Pembangunan Sistem

Diagram alur flowchart digunakan pada tahap pembangunan sistem. Diagram alur dalam penelitian ini mencakup gambaran umum sistem yang akan dibangun, Text-Preporcessing, Labelling data, dan Apply Multinomial Naive Bayes.

3.3 Flowchart Sistem

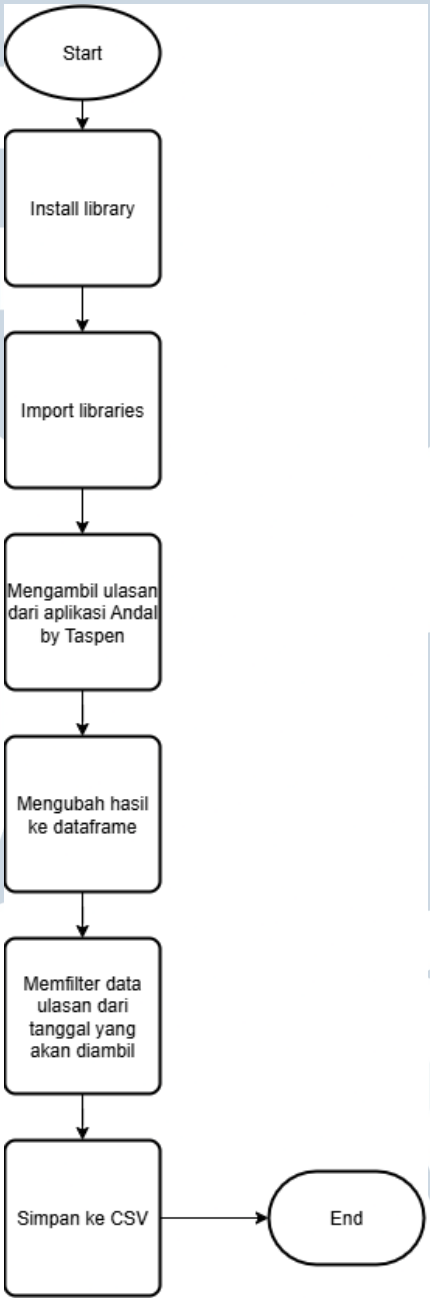
Gambar 3.1 menunjukkan flowchart yang menggambarkan sistem yang akan dibangun dalam penelitian. Flowchart penelitian dimulai dengan pengumpulan data menggunakan metode scraping data, preprocessing teks, esktrasi fitur TF-IDF, labeling data menggunakan Textblob, Split data, penerapan MultinomialNB, dan evaluasi.



Gambar 3.1. Flowchart Sistem

3.4 Scraping Data

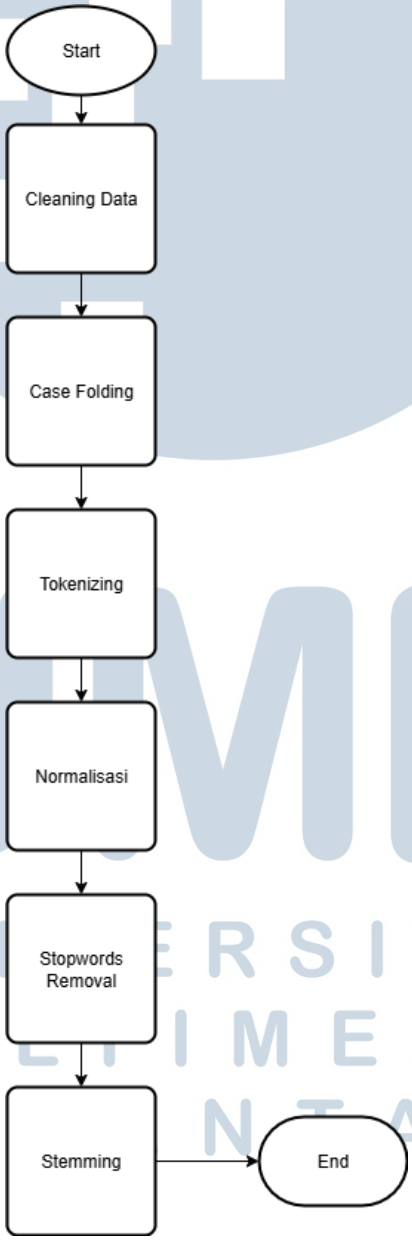
Pengambilan data atau scraping Pengumpulan data, juga disebut scraping, adalah teknik pengumpulan data yang menggunakan perangkat lunak khusus yang dapat mengunjungi halaman web untuk mengumpulkan data. Proses ini bertujuan untuk mengumpulkan data penting dan menyimpannya dalam format yang teratur, seperti database atau spreadsheet[33].



Gambar 3.2. Flowchart scraping data

3.5 Text-preprocessing

Preprocessing adalah proses awal mengolah data mentah agar dapat digunakan untuk model atau analisis lanjutan. Text mining menggunakan pengolahan data awal untuk mendapatkan informasi berharga dari data teks yang tidak teratur.Gambar 3.3, menampilkan berbagai langkah yang diperlukan untuk melakukan proses pengolahan teks. Tahapan-tahapan ini termasuk, cleaning data, casefolding, tokenizing, normalisasi, stopwords, dan stemming



Gambar 3.3. Flowchart Text-Preprocessing

1. Cleaning Data

   Dalam proses ini, username, hashtag, simbol, emotikon, tautan, karakter tunggal, angka, ruang kosong, dan tanda baca yang tidak mengandung informasi emosional dikeluarkan dari data teks.

2. Case Folding

   Tahapan proses yang mengubah dari huruf besar menjadi huruf kecil.

3. Tokenizing

   Dalam tahap ini, proses memisahkan teks menjadi beberapa kata.

4. Normalisasi

   Pada tahap ini dilakukan dengan membuat sebuah file csv yang isinya dari kata yang salah dalam penulisan serta dengan perbaikannya.

5. Stopwords Removal

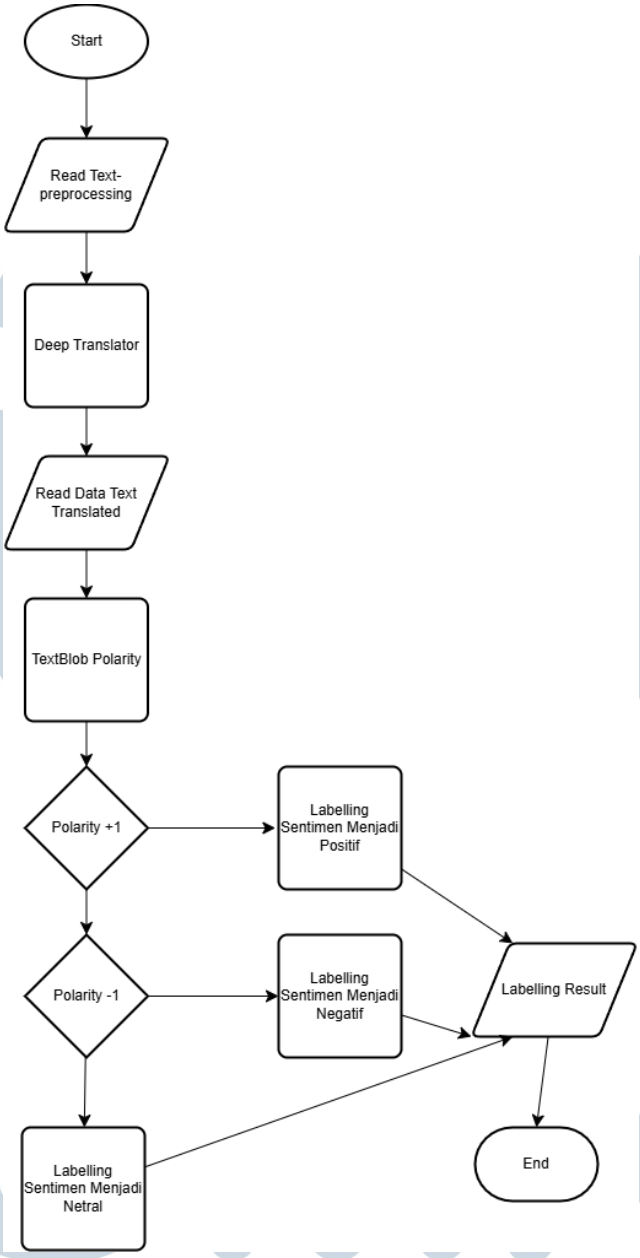
   Dalam tahap stopwords, kata-kata yang tidak mempunyai makna khusus, seperti penghubung, dihapuskan.

6. Stemming

   Dalam proses stemming, pustaka sastrawi digunakan untuk menghapus atau mengambil kata-kata yang memiliki afiks hingga bentuk dasarnya dapat dihasilkan.

3.6 Labelling Data Textblob

Pelabelan data dilakukan secara otomatis menggunakan pendekatan lexicon-based, yaitu dengan memanfaatkan library TextBlob. Metode ini menggunakan kamus sentimen untuk menentukan nilai polaritas dari setiap ulasan. Proses ini dimulai dengan membaca data teks yang telah melalui tahapan text preprocessing, kemudian menerjemahkannya ke dalam bahasa Inggris menggunakan layanan DeepL Translator. Hal ini dilakukan karena TextBlob hanya dapat membaca dan memproses teks berbahasa Inggris. Setelah proses penerjemahan selesai, TextBlob menghitung nilai polaritas masing-masing ulasan. Berdasarkan nilai polaritas tersebut, sistem secara otomatis memberikan label sentimen ke dalam tiga kategori: positif, netral, dan negatif.



Gambar 3.4. Flowchart TextBlob

Gambar 3.4 Text-Preprocessing diatas merupakan sebuah alur dari pelabelan menggunakan Textblob dimulai dari membaca data yang sudah diproses text-preprocessing kemudian melakukan terjemahan bahasa indonesia ke dalam bahasa inggris menggunakan layanan DeepL Translator. Dikarenakan library textblob menggunakan kamus yang berisi dengan daftar kata bahasa inggris. Maka, textblob hanya membaca data teks dalam bahasa inggris. Dalam proses penerjemahan library yang dipakai dari api deepl translator. Setelah data diterjemahkan dengan sukses

langkah selanjutnya adalah menghitung polaritas dengan menggunakan library textblob. TextBlob membaca teks berbahasa inggris, menghitung polaritas kata atau kalimat menggunakan kamus sentimen internalnya, lalu mengklasifikasikan teks menjadi positif, netral, atau negatif secara otomatis.

3.7 Pembobotan TF-IDF

Tahap pembobotan atau penilian TF-IDF memberikan nilai angka pada data teks kalimat. Tujuannya adalah agar algoritma multinomial naive Bayes bisa dapat mengerti teks yang diberikan.

Gambar 3.5. Flowchart Pembobotan TF-IDF

Pada gambar 3.5 fungsi dari TfidfVectorizer mengubah kumpulan teks

kalimat/ulasan menjadi angka dalam bentuk matriks vektor, untuk menghitung bobot setiap kata berdasarkan frekuensi kemunculan kata dan seberapa jarang kata itu muncul di seluruh dokumen. Dengan kata lain TfidfVectorizer mengukur seberapa penting suatu kata dalam suatu dokumen dibandingkan dengan seluruh kumpulan dokumen.

3.8 Train-Test Split Data

Setelah melakukan tahap pembobotan TF-IDF, kemudian menggunakan train-test split data untuk melakukan tahap pembagian data. Data yang tersedia saat ini akan dibagi menjadi dua kategori yaitu, data pelatihan dan data pengujian. Data pelatihan atau train data berfungsi untuk membuat model dengan menggunakan algoritma dari naive bayes yaitu multinomial naive bayes, sedangkan data pengujian atau bisa disebut train data berperan dalam menilai performa model tersebut.

3.9 Apply Multinomial Naive Bayes

Setelah membagi data menjadi dua bagian, langkah berikutnya adalah mengklasifikasikan data menggunakan algoritma dari naive bayes, multinomial naive bayes. Dimulai dengan membuat model klasifikasinya terlebih dahulu dan kemudian menggunakan data pelatihan atau data train untuk melakukan pembobotan. Prediksi data pengujian akan dibuat dengan menggunakan model yang dibuat dengan algoritma ini. Pada Gambar 3.6 menunjukkan apply multinomial naive bayes .

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

Gambar 3.6. Flowchart Apply MultinomialNB

3.10 Evaluation

Pada tahap ini, penelitian menggunakan algoritman Multinomial Naive Bayes untuk melakukan evaluasi yang akan mengukur kinerja dari pemodelan yang telah berhasil dibuat. Penilaian dilakukan dengan memanfaatkan confusion matrix yang memiliki tiga jenis label perbandingan, yaitu positif, negatif, dan netral. Kemudian, data dari confusion matrix itu akan dipakai untuk menghitung akurasi, presisi, recall, dan skor F1 untuk mengevaluasi kinerja model.