

## BAB 3 METODOLOGI PENELITIAN

### 3.1 Metode Penelitian

Berikut tahapan-tahapan dalam penelitian Analisis Sentimen Terhadap Ulasan Aplikasi Glints pada Google Playstore Menggunakan Algoritma Multinomial Naive Bayes dan didalamnya terdapat beberapa tahapan rancangan dan membangun sebuah sistem yaitu:

1. Studi Literatur : Tahap awal penelitian ini dilakukan dengan studi literatur untuk memperoleh pemahaman mengenai topik yang berkaitan dengan analisis ulasan pengguna pada aplikasi Glints. Informasi dikumpulkan dari berbagai sumber seperti buku, jurnal, artikel, dan website untuk mendukung landasan teori penelitian. Beberapa teori yang digunakan meliputi Penelitian Terdahulu, Lowongan Lapangan Pekerjaan, Glints: TapLoker, *Text Mining*, *Natural Language Processing (NLP)*, Analisis Sentimen, Naive Bayes, Multinomial Naive Bayes, TF-IDF, *Confusion Matrix*, *Text Preprocessing*, *TextBlob*. dan *Easy Data Augmentation*. Serta informasi mengenai Google Play Store dan Google Play Store Review sebagai sumber data ulasan. Melalui studi literatur ini, peneliti memperoleh gambaran yang jelas mengenai konsep dan metode yang akan digunakan dalam analisis data ulasan aplikasi Glints.
2. Pengambilan Data (Data Scrapping): Pengumpulan data pada penelitian ini dilakukan melalui metode data scraping menggunakan library *google-play-scraper*, yaitu pustaka yang digunakan untuk mengambil data aplikasi yang tersedia di Google Play Store berdasarkan nama package aplikasinya. Data yang dikumpulkan dipenelitian ini berupa ulasan pengguna aplikasi Glints. Proses pengambilan data dilakukan mulai dari 1 Mei 2025 hingga 18 November 2025 dan berhasil memperoleh sekitar 4998 ulasan. Data ulasan yang telah dikumpulkan akan digunakan untuk melakukan analisis sentimen dan tahapan pemrosesan data.
3. Pembuatan Sistem: Tahapan pembuatan sistem dalam penelitian ini dimulai dengan menyusun alur penelitian atau flowchart sebagai dasar proses kerja sistem. Data yang telah dikumpulkan tadi diproses melalui

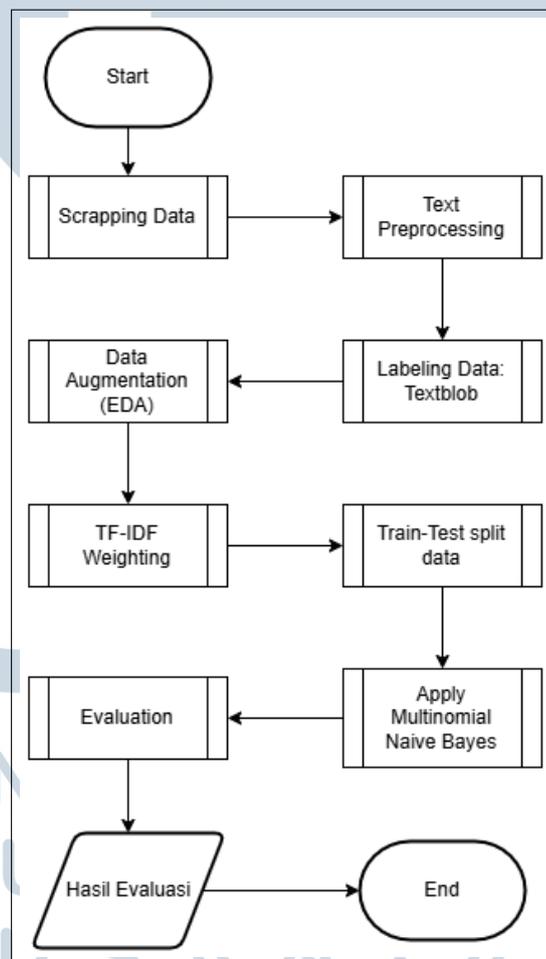
tahapan text preprocessing, lalu dilanjut dengan melakukan pelabelan sentimen menggunakan metode lexicon-based melalui TextBlob untuk mengidentifikasi kecenderungan sentimen pada ulasan pengguna aplikasi Glints, lalu didapatkan jika data set yang telah diberi pelabelannya tidak seimbang (*Imbalance dataset*) maka dilakukan tahap Augmentasi data dengan teknik *oversampling* yaitu Easy Data Augmentation (EDA) setelah melakukan tahap sebelumnya maka selanjutnya melakukan ekstraksi fitur menggunakan metode TF-IDF dan tahap akhir data dibagi menjadi data latih dan data uji sebagai persiapan untuk analisis model pada proses penelitian berikutnya.

4. Pengujian Model : Pada tahap pengujian setelah data selesai dilatih maka sistem melakukan pengujian dengan mengklasifikasikan data ke bentuk sentimen. Hasil klasifikasi tersebut kemudian dievaluasi untuk mengetahui apakah model bekerja dengan baik. Pengujian dilakukan dengan membagi data latih dan data uji ke dalam beberapa perbandingan yaitu 60:40, 70:30, dan 80:20. Dari proses yang dihitung adalah nilai accuracy, precision, recall, dan f1-score melalui confusion matrix. Nilai-nilai tersebut digunakan untuk membandingkan label sebelum dan sesudah pemodelan sehingga dapat menghasilkan prediksi sentimen yang tepat dan stabil..
5. Dokumentasi: Penelitian ini mendokumentasikan seluruh proses secara terstruktur, dan dijelaskan secara berurutan, dimulai dari pendahuluan, perancangan metode, hingga penyusunan kesimpulan dan saran. Dokumentasi ini dibuat untuk memastikan alur penelitian jelas, sistematis, dan mudah dipahami.
6. Konsultasi: Tahap konsultasi dilakukan dengan berdiskusi secara berkala bersama dosen pembimbing untuk memastikan proses penelitian terkait analisis sentimen pada aplikasi Glints berjalan sesuai arah yang tepat. Melalui konsultasi ini, peneliti memperoleh masukan, arahan, dan evaluasi terhadap langkah-langkah penelitian maupun penulisan laporan, sehingga keseluruhan proses dapat disempurnakan dan mencapai hasil yang lebih baik.

### 3.2 Alur Pembuatan Sistem

Pada tahap pembangunan sistem, penelitian ini menggunakan diagram alur (*flowchart*) untuk menggambarkan proses kerja sistem analisis sentimen pada aplikasi Glints. Flowchart tersebut memuat alur utama sistem yaitu: text preprocessing, pelabelan data, augmentasi data, hingga penerapan algoritma Multinomial Naive Bayes. Penggunaan diagram alur (*flowchart*) ini bertujuan untuk memberikan gambaran yang jelas dan terstruktur mengenai setiap langkah yang dilakukan dalam pembuatan sistem.

### 3.3 Flowchart Pembuatan Sistem

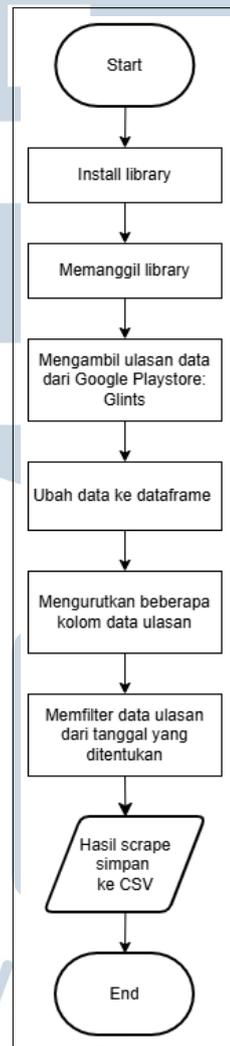


Gambar 3.1. Alur Kerja Sistem

Gambar 3.1 menampilkan flowchart yang menggambarkan alur kerja sistem analisis sentimen pada aplikasi Glints. Proses dimulai dari pengumpulan

data dengan metode *scraping*, kemudian dilanjutkan dengan *text preprocessing*, pelabelan data dengan TextBlob, augmentasi data, ekstraksi fitur menggunakan TF-IDF, pembagian data latih dan uji, penerapan algoritma Multinomial Naive Bayes dan mengevaluasi performa model.

### 3.4 Pengumpulan Data (*Scraping*)



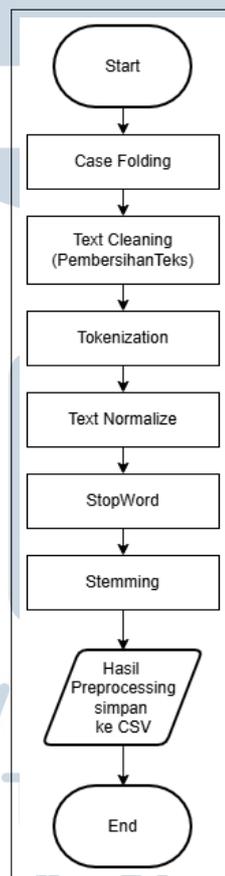
Gambar 3.2. Alur Pengumpulan Data

Gambar 3.2 adalah alur dari proses pengambilan data. Pengambilan data dalam penelitian ini dilakukan melalui teknik *web scraping*, dimana teknik ini melakukan proses otomatis untuk mengumpulkan dan menyimpan data dari halaman web menggunakan perangkat lunak atau library khusus. Metode ini

memungkinkan peneliti memperoleh informasi secara efisien dan menyimpannya dalam format terstruktur seperti CSV[8]. Pada teknik web scraping ini didapatkan sebanyak 4998 ulasan pengguna aplikasi Glints di Google Play Store dan data dikumpulkan dari tanggal 1 Mei 2025 - 18 November 2025, kemudian disaring dan disimpan dalam format CSV untuk digunakan dalam tahapan berikutnya.

### 3.5 Text Preprocessing Data

Proses ini merupakan tahap untuk membantu dalam menyeleksi dari data mentah yang telah diambil tadi akan diproses menjadi data yang lebih terstruktur dan efektif untuk di analisis. Alur yang dilakukan pada text preprocessing ini dimulai dari tahap *Case Folding*, *Cleaning Data*, *Tokenizing*, *Normalize*, *StopWord*, *Stemming* dan disimpan dari hasil text preprocessing kedalam format CSV.



Gambar 3.3. Alur Text Preprocessing

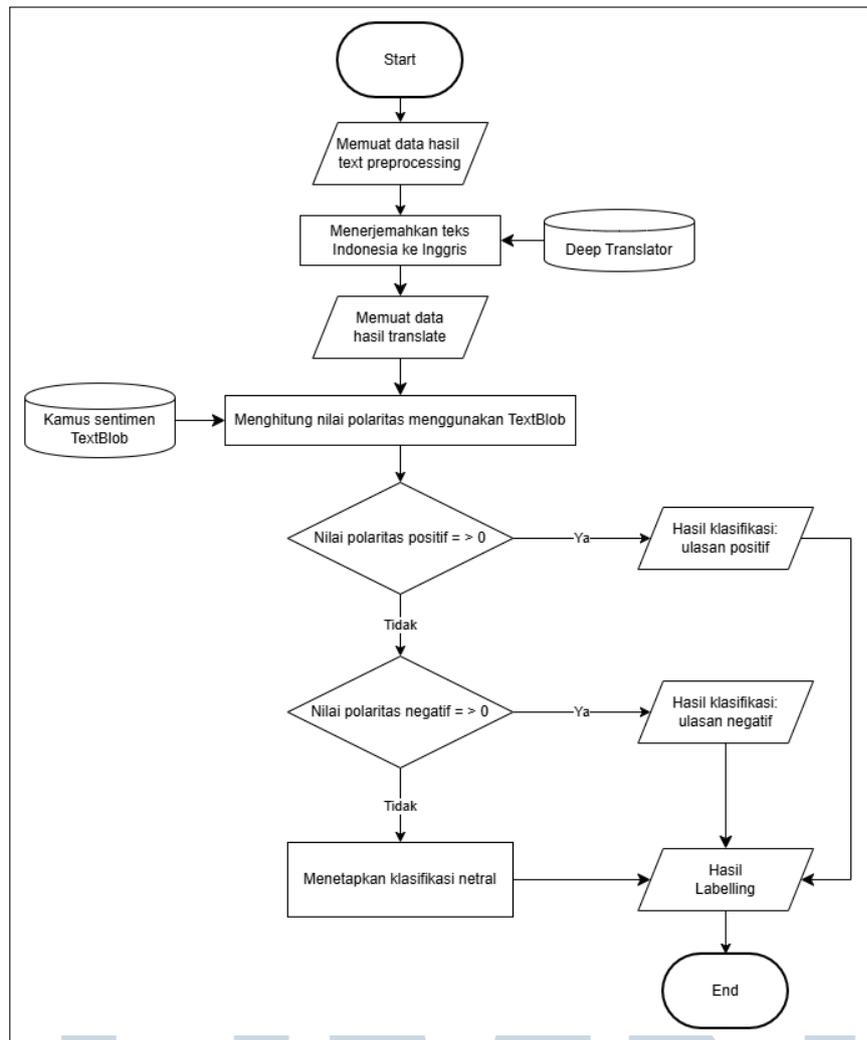
Gambar 3.3 ini merupakan gambar visualisasi alur flowchart tahapannya. Berikut dibawah ini penjelasan singkat pengerjaan pada setiap prosesnya;

1. Case Folding : Tahapan ini ditunjukkan untuk mengubah teks yang tadi nya terdapat huruf kapital menjadi huruf kecil.
2. Text Cleaning: Tahapan ini untuk membersihkan dan menyeleksi data teks berupa emoji, simbol, single-char, link, angka, white-space dan tanda baca yang tidak memiliki sentimen.
3. Tokenization: Proses memecah kalimat ulasan yang panjang akan dipotong menjadi kata-kata terpisah sehingga setiap kata dapat dianalisis secara individual. Proses ini mengubah teks yang awalnya berbentuk kalimat lengkap menjadi daftar kata-kata tunggal.
4. Normalize: Proses untuk mengoreksi kesalahan kata-kata atau ejaan dan memperbaiki daftar kata berbentuk file dengan format CSV.
5. Stopword: Proses menghapus kata-kata umum yang sering muncul dalam teks namun tidak memiliki arti penting.
6. Stemming: Proses ini menggunakan library Sastrawi khusus dirancang untuk bahasa Indonesia. Proses ini mengubah kata berimbuhan menjadi kata dasar dengan menghilangkan kata depan dan belakangnya..

### 3.6 Pelabelan Data (*TextBlob*)

Proses pelabelan data dilakukan secara otomatis menggunakan pendekatan lexicon-based melalui library TextBlob, cara kerjanya dengan menghitung dari "kamus sentimen" untuk mendapatkan sentimen di setiap katanya. Proses pelabelan dimulai setelah data ulasan melalui tahap text preprocessing.

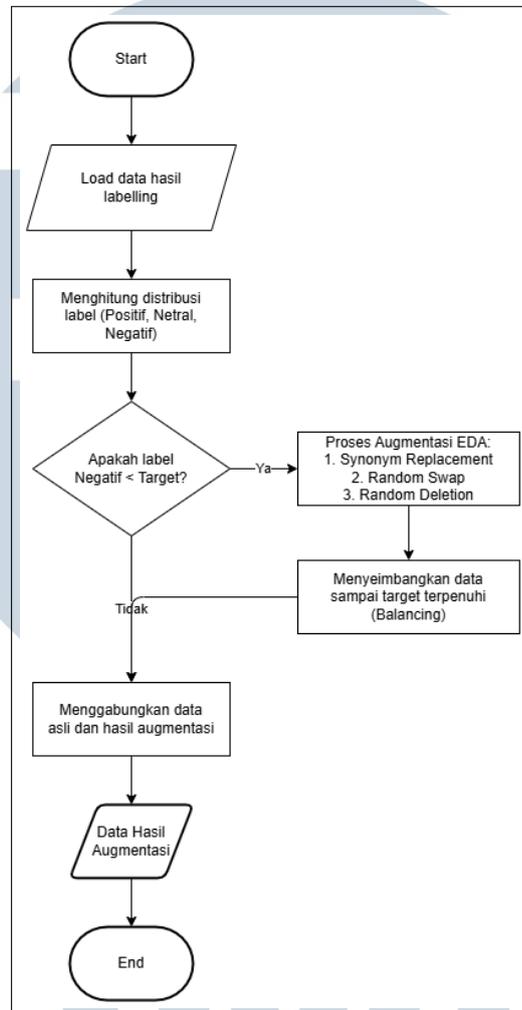
U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A



Gambar 3.4. Pelabelan: TextBlob

Gambar 3.4 merupakan tahap dari pelabelan sentimen ulasan aplikasi Glints diproses menggunakan TextBlob melalui alur yang dimulai dari membaca data yang telah melewati tahap text preprocessing, setelahnya didapatkan diseluruh ulasan dalam bahasa Indonesia dan diterjemahkan terlebih dahulu menggunakan pustaka DeepTranslator, oleh karena itu TextBlob hanya dapat menganalisis teks berbahasa Inggris. Proses penerjemahan telah selesai maka data akan diolah oleh TextBlob untuk menghitung nilai polaritas setiap ulasan. Nilai polaritas tersebut digunakan untuk menentukan kategori sentiment lalu sistem dapat memberikan label secara otomatis ke dalam tiga kelas yaitu positif, netral, dan negatif.

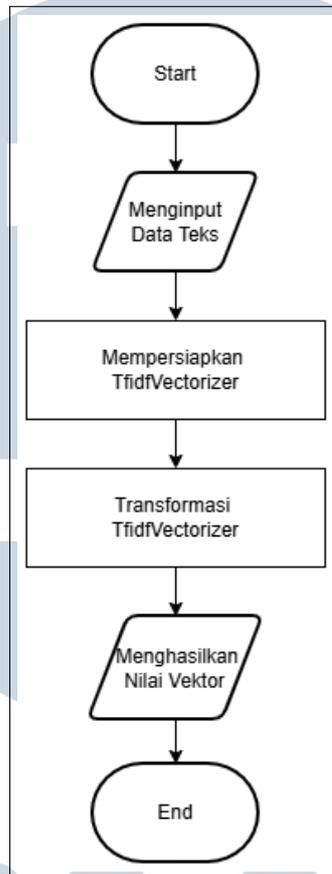
### 3.7 Augmentasi Data (*Easy Data Augmentation*)



Gambar 3.5. Augmentasi Data

Gambar 3.5 merupakan alur dari penggunaan augmentasi data menggunakan teknik *oversampling* yaitu Easy Data Augmentation (EDA), tujuannya adalah untuk meningkatkan jumlah data minoritas karena didapatkan setelah melakukan pelabelan data tersebut imbalance, polaritas negatif cenderung lebih sedikit dibanding positif dan netral. Terdapat 3 proses didalam augmentasi EDA yaitu dengan melakukan *Synonym Replacement*, *Random Swap*, dan *Random Deletion*. Proses ini membuat data negatif akan ditambah hingga batas target dan setelah nya akan digabungkan antara data pelabelan dengan data yang telah diaugmentasi. Cara ini bertujuan semata-mata untuk meningkatkan hasil kinerja dari model klasifikasi dengan menghitung metrik evaluasinya.

### 3.8 Pembobotan TF-IDF



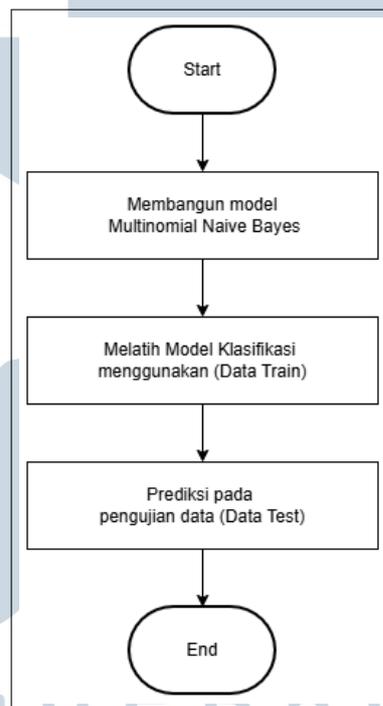
Gambar 3.6. Pembobotan TF-IDF

Gambar 3.6 adalah proses pembobotan dengan metode TF-IDF yang digunakan untuk mengubah teks ulasan aplikasi Glints menjadi representasi numerik agar dapat diproses oleh algoritma Multinomial Naive Bayes. Proses ini dilakukan menggunakan *TfidfVectorizer* untuk mengonversi kumpulan ulasan menjadi matriks vektor berdasarkan bobot setiap kata. Bobot tersebut dihitung dari frekuensi kemunculan kata dalam sebuah ulasan serta tingkat kelangkaannya di seluruh dokumen. TF-IDF dapat menunjukkan seberapa penting suatu kata dalam konteks keseluruhan data. Tahapan ini memastikan bahwa model klasifikasi mampu memahami struktur teks lebih efektif.

### 3.9 Train-Test Split Data

Setelah proses pembobotan TF-IDF dilakukan langkah selanjutnya adalah membagi data menggunakan metode *train-test split*. Tahap nya adalah memisahkan data aplikasi Glints menjadi dua bagian yaitu data pelatihan dan data pengujian. Data pelatihan (*data train*) digunakan untuk melatih model klasifikasi algoritma Multinomial Naive Bayes sedangkan data pengujian (*data test*) digunakan untuk menguji kualitas dari performa model klasifikasi. Tahap ini bertujuan untuk menilai model mampu mempelajari pola teks secara optimal dan dievaluasi dari hasil pemodelannya.

### 3.10 Input Multinomial Naive Bayes Classification



Gambar 3.7. Multinomial Naive Bayes Classification

Setelah proses pembagian data, Gambar 3.7 menjelaskan dari tahap berikutnya yaitu menerapkan algoritma Multinomial Naive Bayes untuk melakukan klasifikasi teks pada data aplikasi Glints. Proses ini dimulai dengan membangun model klasifikasi lalu melatih model tersebut menggunakan seluruh data pelatihan (*data train*) yang sebelumnya telah melalui tahap pembobotan. Model yang telah dilatih selanjutnya digunakan untuk menghasilkan prediksi pada pengujian data

(*data test*). Dengan langkah ini sistem dapat menilai kemampuan model dalam mengklasifikasikan teks secara akurat.

### 3.11 Evaluation Multinomial Naive Bayes

Tahap ini untuk mengevaluasi kinerja model klasifikasi pada data ulasan Glints menggunakan algoritma Multinomial Naive Bayes. Evaluasi dilakukan melalui confusion matrix yang memuat tiga label sentimen yaitu positif, negatif, dan netral. Data yang diperoleh dari confusion matrix tersebut kemudian digunakan untuk menghitung metrik penilaian yaitu accuracy, precision, recall, dan F1-score. Perhitungan ini bertujuan untuk mengetahui kualitas dari performa algoritma Multinomial Naive Bayes sebagai model klasifikasi.

