

BAB III

METODOLOGI PENELITIAN

3.1 Metode Penelitian

Metode pengumpulan data dan metode analisis data merupakan dua pendekatan yang digunakan dalam penelitian ini. Tiga metode digunakan untuk memperoleh data untuk penelitian ini yaitu observasi, studi pustaka, dan mengumpulkan data primer dari ulasan pengguna di situs web Google Play Store. Langkah-langkah pemrosesan teks, pemodelan berbasis Lexicon dan SVM, matriks konfusi, dan visualisasi semuanya dimasukkan dalam analisis data untuk penelitian ini.

3.2 Metode Pengumpulan Data

Metode pengumpulan data berikut digunakan untuk mengumpulkan data dan informasi mengenai objek yang diteliti peneliti untuk penelitian ini

3.2.1 Observasi

Untuk mengumpulkan informasi untuk penelitian awal, peneliti memasang aplikasi JakOne Mobile, melihat ulasan Google Play Store, dan menggunakan aplikasi itu sendiri. Temuan yang peneliti peroleh dari pengamatannya adalah sebagai berikut.

1. Mempelajari kendala terkait penggunaan aplikasi JakOne Mobile.
2. Mempelajari lebih lanjut kelebihan dan kekurangan aplikasi JakOne Mobile.

3.2.2 Studi Pustaka

Untuk mengumpulkan informasi dan data yang berkaitan dengan judul penelitian, dilakukan tinjauan literatur. Data dan informasi yang terkumpul kemudian dijadikan sebagai bahan pendukung dan perbandingan bagi penelitian yang sedang berlangsung.

3.2.3 *Scrapping Data*

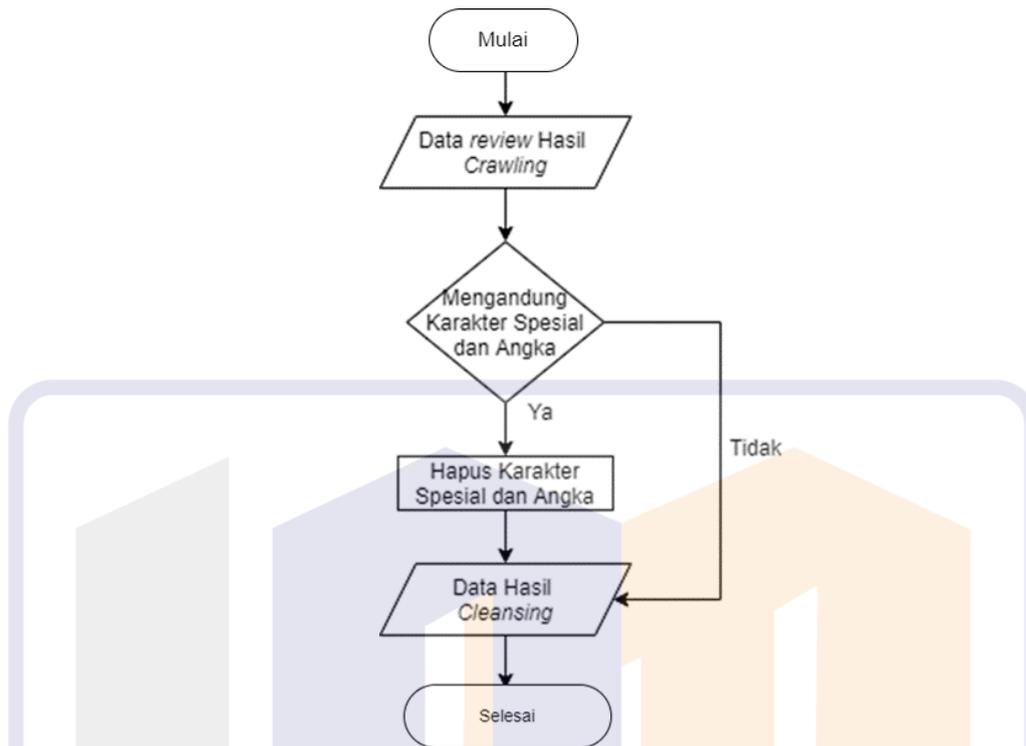
Python Scrapper digunakan untuk mengumpulkan data. Variabel metadata data ulasan mencakup nama pengguna, konten ulasan, jumlah bintang, serta tanggal dan waktu ulasan. Data dikumpulkan dari *review* di *website* Google Play. Setelah diproses, data diubah ke format *Comma Separated Value* (CSV).

3.3 *Preprocessing Data*

Preprocessing data adalah tahapan penting dalam analisis sentimen yang bertujuan untuk menyiapkan data teks mentah agar dapat diolah dan dianalisis dengan lebih baik. Data teks mentah seringkali tidak terstruktur dan mengandung banyak noise, seperti karakter khusus, tanda baca, kata-kata yang tidak relevan, dan lain sebagainya. *Preprocessing* data membantu membersihkan dan mentransformasi data teks menjadi format yang lebih terstruktur dan relevan untuk analisis sentimen.

3.3.1 *Cleaning*

Karakter spesial yang pengguna gunakan pada *review* JakOne Mobile di Google Play seperti angka, *hashtag*, *mention username*, URL *Website*, tag HTML, *emoticon*, koma, titik spasi berlebih akan dihapus dari data ulasan yang akan masuk ke tahap analisis. *Cleaning* ini bertujuan untuk mengurangi kesalahan acak (*noise*) pada data.



Gambar 3. 1 Gambaran tahap *cleaning* data

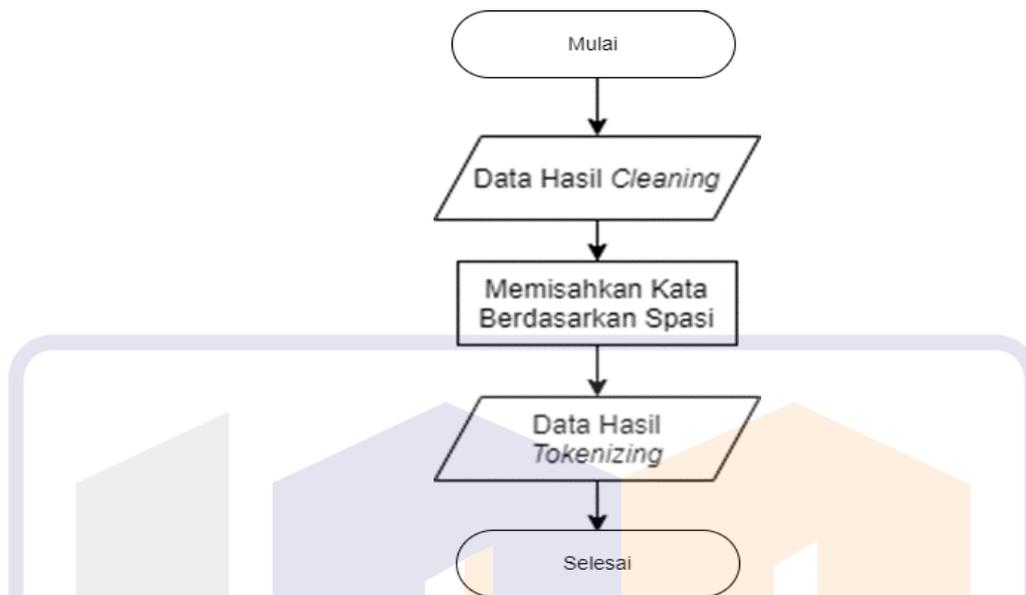
Sumber: Data yang diolah (2025)

3.3.2 Tokenisasi

Tokenisasi adalah tahap persiapan teks yang dapat digunakan untuk memecah kalimat, paragraf, atau dokumen menjadi unit-unit yang lebih kecil, seperti token atau kata-kata independen, dan menghilangkan kalimat-kalimat yang identik.

Karena Python adalah bahasa pemrograman peka huruf besar-kecil yang membedakan huruf, semua ulasan akan diubah terlebih dahulu menjadi huruf kecil sebelum dipecah menjadi kata-kata. Hal ini akan menghilangkan perbedaan antara kata seperti "Aplikasi" dan "aplikasi", yang dapat membahayakan keakuratan hasil. Huruf kecil dan huruf kapital, yang dapat menyebabkan perbedaan antara dua kata yang mempunyai arti yang sama jika tidak sama.

Casefolding adalah proses mengubah kata menjadi huruf kecil.

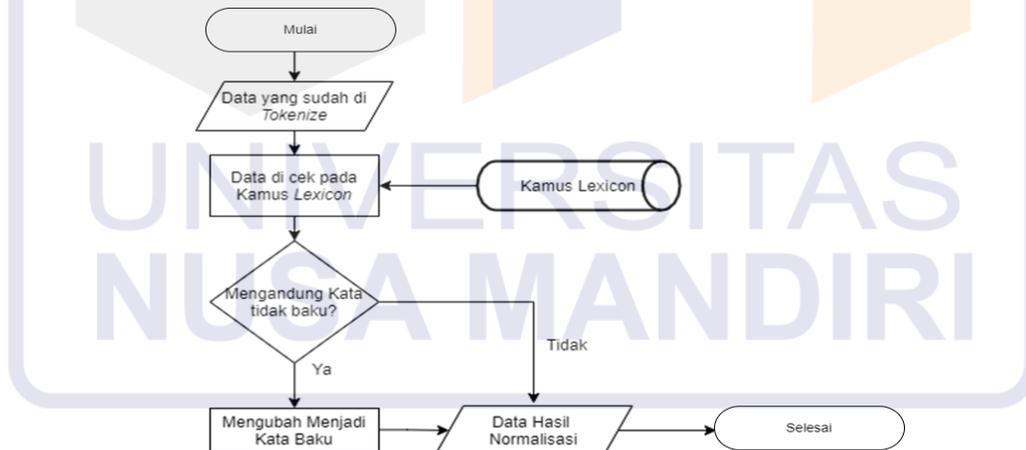


Gambar 3. 2 Alur Tokenizing

Sumber: Data yang diolah (2025)

3.3.3 Normalisasi

Normalisasi adalah tahapan mengubah kata yang tidak baku menjadi kata yang baku sesuai dengan Kamus Besar Bahasa Indonesia.

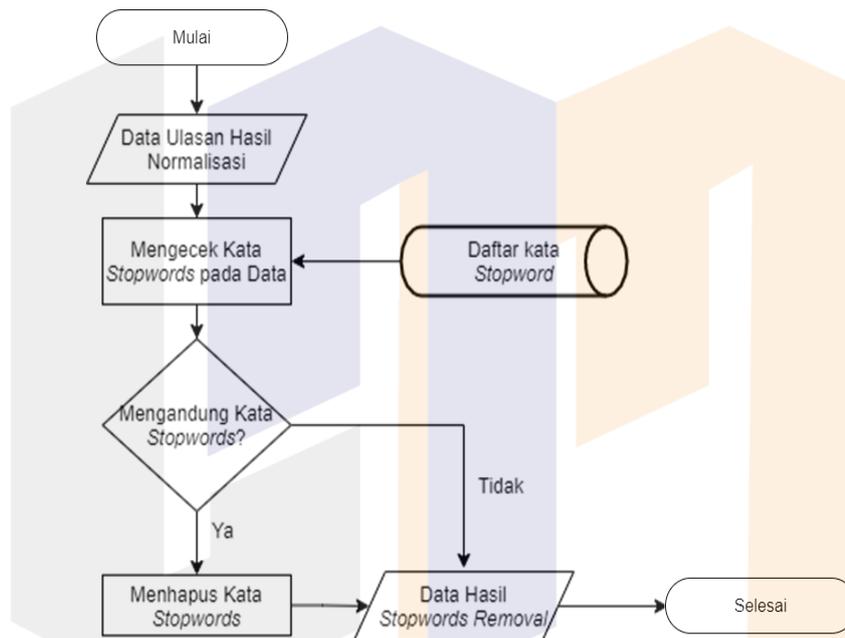


Gambar 3. 3 Tahapan Normalisasi

Sumber: Data yang diolah (2025)

3.3.4 *Stopword Removal*

Stopword adalah kata-kata seperti konjungsi, preposisi, dan sebagainya yang tidak memiliki makna apa pun dalam sebuah kalimat. Ada juga kemungkinan untuk memperoleh banyak *stopword* dalam sebuah frasa. Menghilangkan *stopword* dari kalimat atau paragraf akan memudahkan pemrosesan data dan mencegah kesalahpahaman selama analisis lebih lanjut.

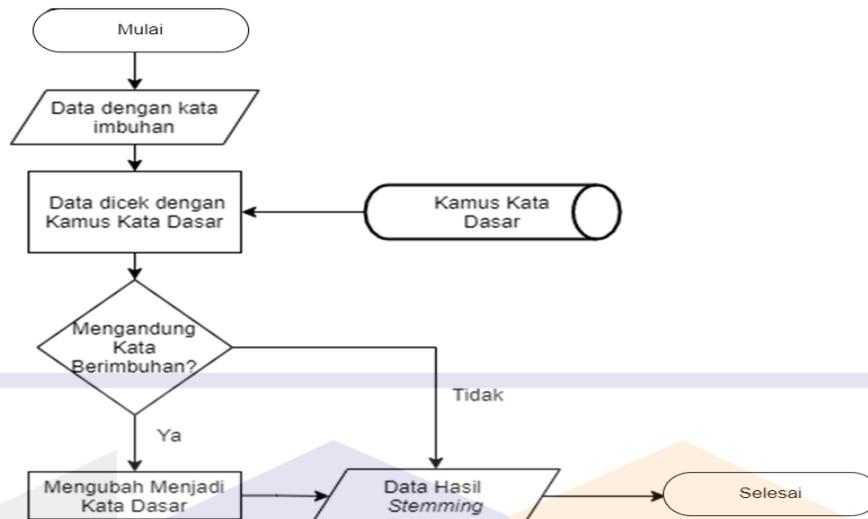


Gambar 3. 4 Alur Stopword Removal

Sumber: Data yang diolah (2025)

3.3.5 *Stemming*

Teknik *preprocessing* teks yang disebut *stemming* digunakan untuk mengidentifikasi kata dasar suatu istilah dengan menghilangkan semua imbuhan dari kata turunan, termasuk prefiks, infiks, sufiks, dan gabungan imbuhan. *Stemming* adalah proses mengubah bentuk kata menjadi kata dasar yang sesuai dengan struktur Bahasa Indonesia yang baik dan benar.



Gambar 3. 5 Alur Stemming Data

Sumber: Data yang diolah (2025)

3.4 Pelabelan Data Metode *Lexicon Based*

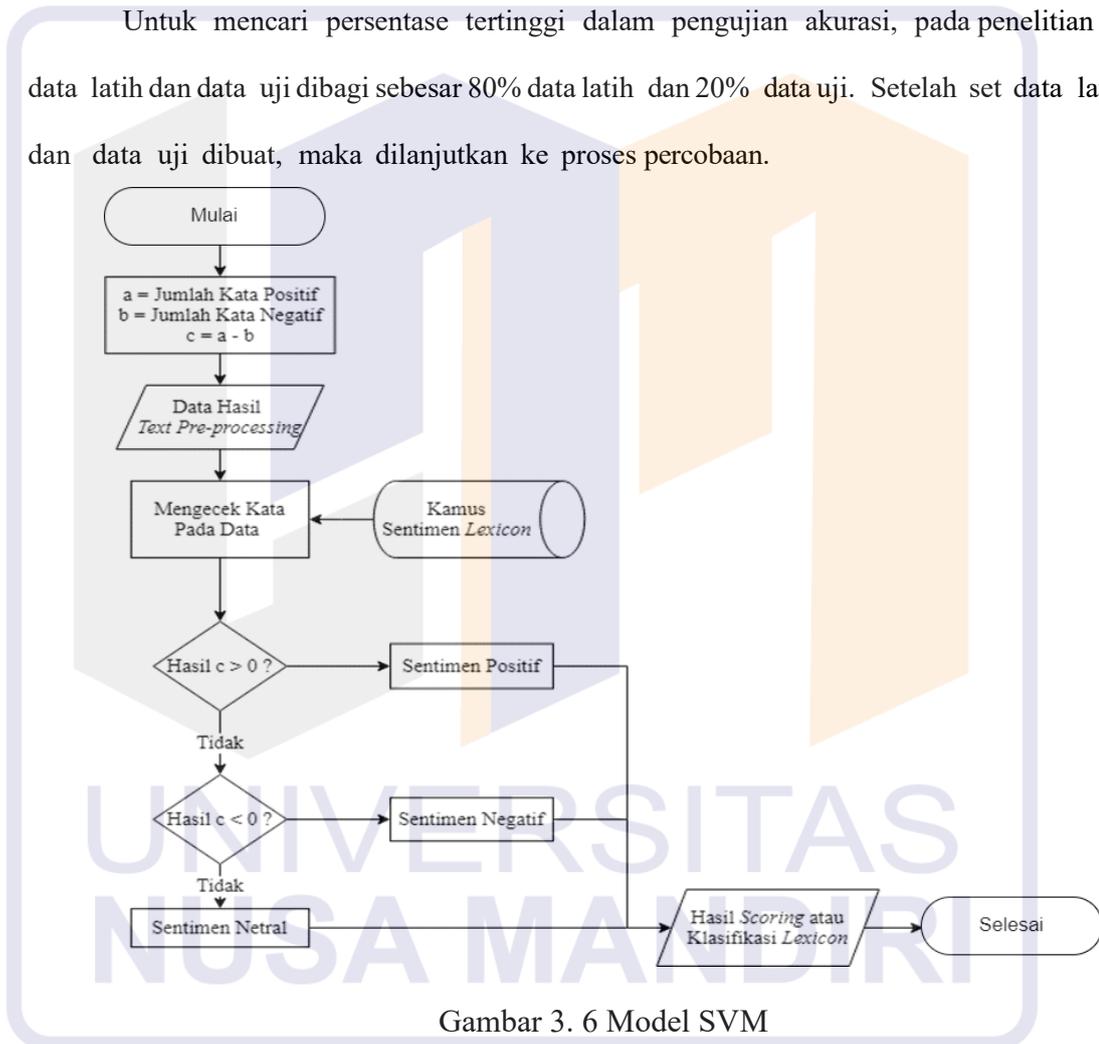
Setiap tag kelas data komentar harus memiliki data ulasan yang digunakan untuk melatih *classifier* metode pembelajaran mesin yang terhubung. Terlepas dari apakah ulasannya positif atau negatif, setiap informasi yang telah melewati langkah *preprocessing* kata sekarang akan diberi tag atau label kelas. Label kelas teks dapat ditentukan dengan berbagai cara, seperti dengan menetapkan label kelas secara manual berdasarkan penilaian sendiri, mengevaluasi berdasarkan kuantitas nilai atau peringkat yang diberikan, atau menggunakan Metode *Lexicon Based* untuk kategorisasi teks.

Untuk menentukan apakah ulasan itu positif atau negatif untuk penelitian ini, peneliti akan menggunakan kamus *Lexicon*. Untuk menentukan label kategori data ulasan, peneliti pertama-tama menentukan sentimen yang ada dalam komentar. Kemudian, mereka menggunakan skor sentimen untuk mendapatkan label untuk setiap komentar. Setiap ulasan sekarang dapat digunakan untuk melatih *classifier* dan berfungsi sebagai referensi untuk menentukan akurasi, presisi, *recall*, dan *f-measure* dari temuan analisis sentimen setelah diberi label sentimen dengan benar menggunakan kamus *Lexicon*.

3.5 Model Algoritma *Support Vector Machine*

Tahapan berikutnya setelah mendapatkan hasil *Lexicon Based*, dapat dilakukan klasifikasi dengan metode *Support Vector Machine*. SVM adalah metode klasifikasi yang dilakukan dengan cara mencari term atau kondisi serupa dengan parameter tertentu agar dapat menentukan kategori akhirnya.

Untuk mencari persentase tertinggi dalam pengujian akurasi, pada penelitian ini data latih dan data uji dibagi sebesar 80% data latih dan 20% data uji. Setelah set data latih dan data uji dibuat, maka dilanjutkan ke proses percobaan.



Gambar 3. 6 Model SVM

Sumber: Data yang diolah (2025)

3.6 Confusion Matrix

Keempat jenis hasil kinerja *machine learning* yaitu akurasi, presisi, *recall*, dan *F-Measure* yang diperoleh dari analisis sentimen menggunakan metode *Support Vector Machine* pada ulasan JakoOne Mobile ditentukan dengan melakukan *Confusion*

Matrix setelah prosedur pemodelan *Laxicon* dan *support vector machine* selesai. Untuk menemukan hasil akurasi maksimum, peneliti menghitung temuan analisis sentimen SVM tergantung pada setiap variabel.[13]

3.7 Visualisasi

Untuk menentukan istilah mana yang paling sering ditemukan dalam data ulasan dan seberapa sering istilah tersebut muncul dalam ulasan, kata-kata yang sering muncul dalam ulasan sekarang akan divisualisasikan menjadi *Word Cloud*. Diagram batang yang terdiri dari beberapa karakteristik emosi adalah cara lain visualisasi digunakan untuk menunjukkan kuantitas emosi dalam data ulasan.

