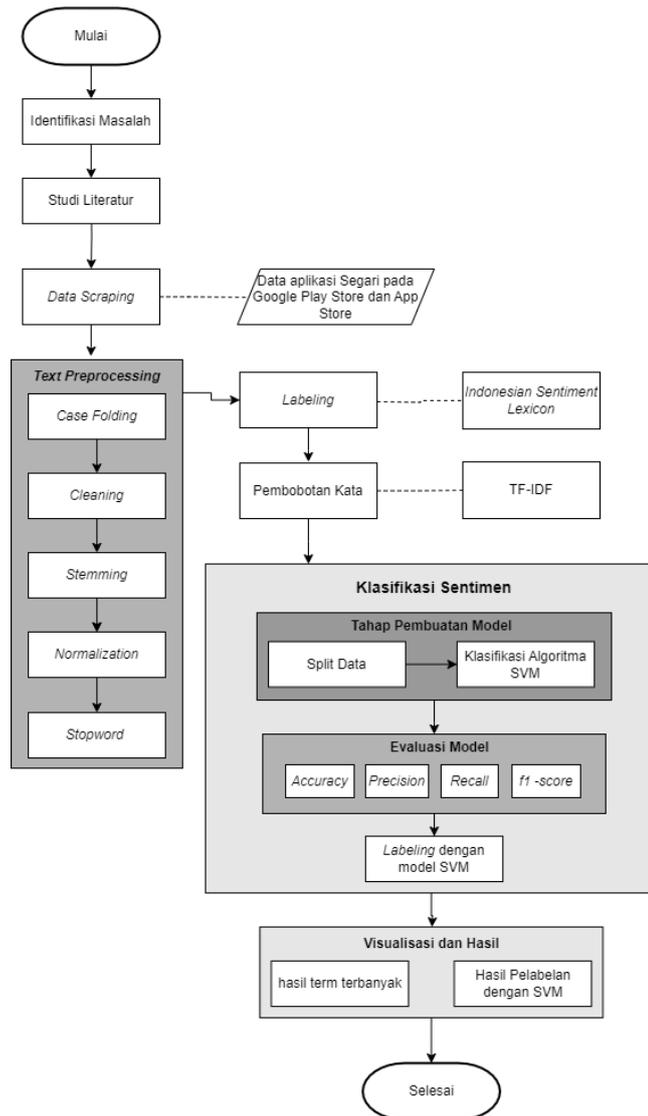


## BAB III

### METODE PENELITIAN

#### 3.1 Alur Penelitian

Penulis membuat flowchart untuk menunjukkan setiap langkah yang akan mereka ambil untuk menjalankan penelitian ini. Penelitian ini dilakukan dalam beberapa tahapan dengan algoritma *Support Vector Machine* (SVM) sebagai alat bantu untuk menyelesaikan proses analisis sentimen. Berdasarkan referensi (Idris, Mustofa, & Salihi, 2023) saya membuat alur penelitian yang akan digunakan dalam menjalankan penelitian ini.



Gambar 3. 1 Alur Penelitian

### 3.1.1 Identifikasi Masalah

Identifikasi dilakukan dengan memantau kondisi review aplikasi Segari di App Store dan Google Play Store. Pengamatan ini menunjukkan bahwa Segari menjual bahan makanan grosir kepada orang-orang di Jakarta, Tangerang, dan Bekasi dengan berbagai jenis sayuran dan buah-buahan. Penelitian ini bertujuan untuk memahami faktor keberhasilan Segari melalui analisis sentimen, yang menunjukkan hal-hal yang disukai dan tidak disukai pelanggan. Analisis sentimen dilakukan melalui algoritma *Support Vector Machine* (SVM) untuk klasifikasi teks, yang terbukti efektif dalam meningkatkan pemahaman ulasan pelanggan.

### 3.1.2 Studi Literatur

Selanjutnya, untuk melakukan penelitian literatur yang dilakukan dengan tujuan untuk membantu dalam pencarian solusi penelitian. Untuk melakukan ini, akan mencari jurnal, website, dan penelitian lainnya yang berkaitan dengan analisis emosi, pemrosesan teks, pengolahan teks, pelabelan dengan leksikon emosi Indonesia, pembobotan kata dengan TF-IDF, dan metode algoritma *Support Vector Machine*.

### 3.1.3 Data Scraping

*Data scraping* adalah langkah berikutnya. Ini adalah proses mengumpulkan review dari pengguna aplikasi Segari di Google Play Store dengan menggunakan library Python, *google-play-scraper* dan App Store dengan *app-store-scraper*. Data yang diambil diurutkan berdasarkan waktu pemeriksaan.

### 3.1.4 Text Preprocessing

Sebelum data dilabeli dan dibobotkan maka hasil dari *data scraping* harus terlebih dahulu menjalani tahap *text preprocessing*.

#### 3.1.4.1 Case Folding

Menyamakan seluruh teks ke dalam bentuk kecil (*lowercase*) sehingga tidak ada huruf besar (*uppercase*) dalam tahap case folding. Hal ini dilakukan dengan tujuan mengurangi tingkat redundansi data yang akan digunakan selama proses pengklasifikasian, sehingga proses perhitungan dapat dilakukan secara optimal. Mengubah format data menjadi *lowercase* atau *uppercase* sesuai dengan persyaratan proses pengklasifikasian adalah contohnya.

#### 3.1.4.2 Data Cleaning

Tujuan dari tahapan ini adalah untuk menghilangkan data yang tidak memiliki nilai, seperti karakter alfanumerik, hashtag, URL, dan spasi berlebihan. Ini dilakukan untuk mengurangi kebisingan yang dapat menyebabkan proses perhitungan mengalami pengklasifikasian tidak optimal.

#### **3.1.4.3 Stemming**

Proses ini akan mengubah kata berinbunan menjadi kata dasarnya. Seperti kata membantu, meningkatkan presentasi kata dasar pada data.

#### **3.1.4.4 Normalization**

Tujuan dari langkah ini adalah untuk mengubah kata-kata yang tidak baku, singkatan, atau typo. Kata-kata ini akan diperbaiki sehingga dapat diterima.

#### **3.1.4.5 Stopword**

Pada langkah ini, kata-kata yang tidak memiliki bobot, seperti di, yang, dan dengan, akan dihapus.

### **3.1.5 Labelling**

Pada langkah berikutnya, kumpulan data yang telah dibersihkan akan diberi label. Setiap dokumen yang mengumpulkan emosi akan diidentifikasi oleh penulis. Penulis menggunakan metode *Indonesian Sentiment Lexicon (InSet Lexicon)* untuk mengidentifikasi tulisan atau komentar di setiap dokumen dengan menghitung nilai setiap kata. Nilai-nilai ini digunakan untuk mengklasifikasikan dokumen sebagai positif, netral, atau negatif. Sebuah dokumen dianggap positif jika nilai totalnya lebih dari nol, nilai negatif jika nilai totalnya kurang dari nol, dan nilai netral jika nilai totalnya sama dengan nol.

### **3.1.6 Pembobotan Kata**

Setelah data diproses sebelumnya, pembobotan kata dilakukan. Saat ini, metode pembobotan TF-IDF digunakan untuk menghitung frekuensi kemunculan setiap kata. Frekuensi kemunculan setiap kata menunjukkan seberapa penting kata tersebut. Jika sebuah istilah muncul lebih sering dalam set data, maka istilah tersebut akan memiliki nilai yang lebih besar. Karena jenis inputan data yang dibutuhkan model SVM menggunakan data bertipe vektor, tipe data hasil *preprocessing* harus diubah, proses ini diperlukan.

### 3.1.7 Model SVM

Setelah pelabelan dan pembobotan kata, langkah selanjutnya adalah penggunaan model SVM. Algoritma ini akan melakukan klasifikasi menggunakan data latihan, yang akan dibagi menggunakan fungsi *train\_test\_split* dari modul *sklearn.model\_selection*. Penulis akan menggunakan modul *sklearn.svm* yang tersedia dalam lembaga *scikit-learn* untuk menemukan hyperplane yang optimal untuk memisahkan dua kelas dalam ruang fitur.

### 3.1.8 Hasil Evaluasi Model

Selanjutnya, menggunakan data uji, hasil klasifikasi akan dievaluasi. Penulis akan memanfaatkan *confusion matrix* untuk membantu proses ini, dan hasilnya akan menunjukkan nilai akurasi, presisi, dan recall terhadap klasifikasi yang telah dilakukan. Hasil klasifikasi adalah dasar dari evaluasi ini, yang bertujuan untuk mengetahui bagaimana sistem yang telah dibuat berfungsi. berdasarkan referensi yang disebutkan sebelumnya di bab dua (Handul, Matulesy, & Kaesmetan, 2024) Hasil evaluasi *confusion matrix* terdiri dari beberapa nilai. Nilai TP (Data Positif Benar) menunjukkan data yang diprediksi positif dan benar nilai aktualnya positif; Nilai FP (Data Negatif Benar) menunjukkan data yang diprediksi negatif tetapi nilai aktualnya negatif; Nilai FN (Data Negatif Benar) menunjukkan data yang diprediksi negatif tetapi nilai aktualnya positif; dan nilai TN (Data Negatif Benar) menunjukkan data yang diprediksi negatif dan benar nilai aktualnya negatif.

### 3.1.9 Analisis Hasil Klasifikasi

Setelah tahapan evaluasi selesai, langkah berikutnya adalah melabeli data menggunakan model SVM yang telah dibangun sebelumnya. Diagram kata terbanyak akan digunakan untuk menjelaskan hasil klasifikasi sentimen ini. Visualisasi ini akan membantu menemukan fitur aplikasi Segari yang disukai dan elemen yang perlu ditingkatkan untuk meningkatkan layanan.

## 3.2 Alat Penelitian

### 3.2.1 Perangkat Keras

Perangkat keras yang akan digunakan oleh penulis pada penelitian ini merupakan laptop yang memiliki spesifikasi:

- a. Prosesor : Intel Core i3-1115G4 @ 4CPUs, 3GHz
- b. RAM : 12 GB
- c. Penyimpanan : 256 GB

### 3.2.2 Perangkat Lunak

Perangkat lunak yang akan digunakan oleh penulis pada penelitian ini adalah.

- a. Sistem Operasi : Microsoft Windows 11 64-bit
- b. Bahasa Pemrograman : Python
- c. *Text Editor* : Visual Studio Code dan Google Colab