

BAB 3

METODE PENELITIAN

Penelitian ini melakukan analisis ulasan masyarakat pada Google Maps terhadap layanan RS ABC di DIY dengan menggunakan metode K-Means dan SVM. Berikut ini merupakan bahan, alat, metode serta tahapan penelitian guna melakukan penelitian ini.

3.1 BAHAN DAN ALAT PENELITIAN

Bahan penelitian ini menggunakan data ulasan seluruh RS ABC yang ada di DIY pada Google Maps dan mencakup seluruh ulasan yang tersedia. Data ini dikumpulkan menggunakan teknik pengambilan data (*scrapping*) artinya mengekstrak data dalam jumlah besar dari sumber daring, dalam penelitian ini mengambil dari Google Maps dan kemudian disimpan dalam bentuk file CSV. Data ulasan yang dianalisis hanya mencakup ulasan yang berisi teks (tidak hanya berisi rating) serta terdiri dari kolom nama, tanggal ulasan, rating dan teks ulasan. Data ulasan terdiri dari 2.327 data untuk RST, 285 data untuk RSV, 2.127 data untuk RSW, 2.050 data untuk RSX, 159 data untuk RSY dan 283 data untuk RSZ.

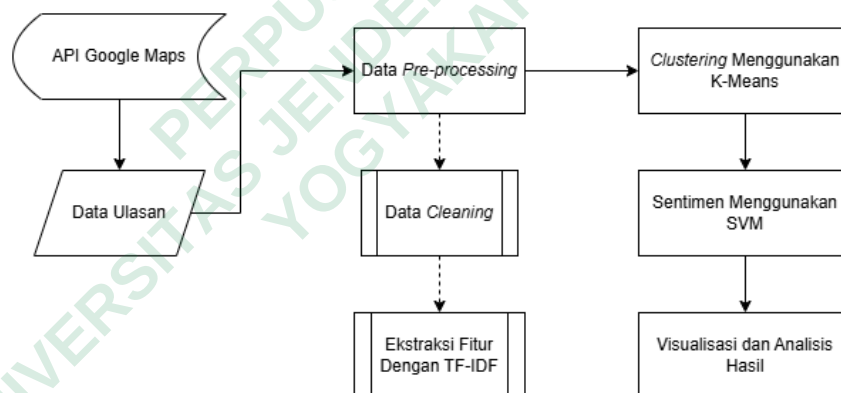
Alat yang digunakan adalah laptop dengan spesifikasi yang dapat mengeksekusi proses pengolahan data dan bisa terhubung dengan jaringan internet. Berikut ini sistem operasi, program atau aplikasi yang akan digunakan dalam proses pengolahan data dalam penelitian ini:

1. Sistem Operasi: Windows 11 64-bit.
2. Bahasa pemrograman Python 3.10.
3. Google *Colaboratory*.
4. Google *Spreadsheet*
5. *Visual Studio Code* versi 1.96.4.
6. *Application Programming Interface* (API) yaitu SerpAPI.
7. *Apify Web Scraper*.
8. *Library* Python yang digunakan:
 - a. Pandas

- b. NumPy
- c. Sastrawi
- d. Scikit-learn
- e. Matplotlib
- f. Seaborn
- g. NLTK
- h. google-search-results

3.2 JALAN PENELITIAN

Penelitian akan dilakukan melalui lima tahapan yaitu pengumpulan data penelitian berupa data ulasan masyarakat pada Google Maps RS ABC yang ada di DIY, tahapan *pre-processing* data, implementasi klastering menggunakan K-Means dan klasifikasi menggunakan algoritma SVM. Lalu dilanjutkan dengan evaluasi kedua algoritma tersebut, yang terakhir visualisasi dan analisis hasil. Adapun diagram alur penelitian dapat dilihat pada Gambar 3.1 dibawah ini.



Gambar 3.1 Alur Penelitian

Untuk penjelasan alur penelitian secara keseluruhan dijelaskan sebagai berikut:

3.2.1 Pengumpulan Data

Pengumpulan data dilakukan melalui teknik *scrapping* menggunakan bahasa pemrograman python *library google-search-results* pada *google colab* dan menggunakan API yaitu *SerpAPI* dan *Apify Web Scraper*. Data yang digunakan dalam analisis ini merupakan data ulasan masyarakat RST, RSV, RSW, RSX, RSY dan RSZ pada Google Maps. Semua data ulasan akan disimpan dalam bentuk CSV.

3.2.2 Pre-processing Data

Pre-processing pada data *mining* adalah langkah awal yang penting untuk menentukan hasil akhirnya, proses ini mengubah data supaya sesuai dengan format yang dibutuhkan pada tahap selanjutnya. Langkah ini melibatkan pengolahan dan pengaturan data guna mengidentifikasi keberadaan bentuk teks yang terstruktur atau tidak terstruktur. Dalam *pre-processing* data terdapat langkah-langkah yang dilakukan yaitu, *cleaning*, *case folding*, *normalization*, *tokenizing*, *filtering*, *stemming*, *labeling data* dan TF-IDF.

- a. **Cleaning.** Proses pembersihan data dilakukan dengan menghapus karakter yang tidak diperlukan seperti tanda baca, angka, tautan, hashtag, mention, emoji, dan lainnya. Hal ini bertujuan untuk mengurangi *noise* serta elemen-elemen yang tidak relevan dalam data teks.
- b. **Case Folding.** Proses mengubah semua huruf yang terdapat pada teks menjadi huruf kecil.
- c. **Data Normalization.** Mengubah kata-kata *slang* (gaul) menjadi kata baku dengan menggunakan *slangwords*.
- d. **Tokenizing.** Berfungsi untuk membagi kalimat menjadi kata-kata penyusunnya, yang disebut sebagai *term* atau *token*. Pemisahan kata ini berdasarkan spasi.
- e. **Filtering.** Proses menghapus kata-kata yang dianggap tidak memiliki makna atau tidak penting.
- f. **Stemming.** Proses menghapus imbuhan dari kata untuk memperoleh kata dasar.
- g. **Labeling Data.** Proses memberikan label pada data berdasarkan kategori sentimennya. Labeling ini bertujuan untuk menghasilkan data yang siap digunakan untuk proses training model. Proses pelabelan dibagi menjadi dua label yaitu positif dan negatif untuk mengurangi penilaian secara subjektif (Hidayat & Sugiyono, 2023). Pelabelan data menerapkan *skema average labeling*, dimana rating 4 dan 5 dikategorikan sebagai positif, sedangkan rating 1, 2, dan 3 dikategorikan sebagai negatif (Kharisma et al., 2023).
- h. **TF-IDF.** Representasi vektor TF-IDF berguna untuk mengubah kata menjadi numerik yang digunakan untuk pelatihan model *machine learning*.

3.2.3 *Clustering* Menggunakan K-Means

Data yang sudah dilakukan *pre-processing* diklasterkan dengan K-Means ke dalam beberapa kelompok berdasarkan karakteristik topik yang sejenis sehingga dapat diketahui aspek apa saja yang diperbincangkan. Penelitian ini menerapkan metode *elbow* untuk menentukan jumlah *cluster* yang optimal. Metode ini menganalisis grafik *Within-Cluster Sum of Squares* (WCSS), dimana nilai WCSS akan menurun seiring bertambahnya jumlah klaster (k), namun penurunan tersebut tidak selalu signifikan. Titik optimal ditandai dengan bentuk siku (*elbow*) pada grafik, yang menunjukkan bahwa penambahan jumlah klaster setelah titik tersebut tidak memberikan pengurangan WCSS yang signifikan. Setelah data dikelompokkan ke dalam klaster, visualisasi *wordcloud* berfungsi untuk mengidentifikasi dan memahami topik utama di setiap klaster.

3.2.4 Sentimen Menggunakan SVM

Setelah topik pembahasan pada setiap *cluster* telah teridentifikasi, selanjutnya yaitu mengklasifikasikan sentimen dari setiap *cluster* tersebut supaya mengetahui opininya apakah bersifat positif atau negatif. Penerapan SVM mencakup proses melatih model menggunakan dataset yang tersedia serta menguji kinerjanya menggunakan data yang baru. Data yang sudah diproses pada tahap *pre-processing* kemudian dipisahkan menjadi dua bagian, yaitu data pelatihan (*data training*) dan data pengujian (*data testing*). Dengan pendekatan ini, opini masyarakat mengenai aspek pelayanan rumah sakit di setiap *cluster* dapat dianalisis secara lebih mendalam.

Proses klasifikasi dilakukan menggunakan algoritma SVM, yaitu metode yang bertujuan menemukan *hyperplane* optimal untuk memisahkan dua kelas dalam *input space* dengan menerapkan prinsip SRM (Azzahra & Mailoa, 2025). Untuk evaluasi model klasifikasi, dilakukan dengan menghitung *confusion matrix* untuk menganalisis performa model dengan membandingkan prediksi dan label sebenarnya. Selain itu, juga dilakukan evaluasi kemampuan model dalam mengklasifikasikan data melalui metrik perhitungan seperti *F1-score*, *precision*, *accuracy* dan *recall*.

3.2.5 Visualisasi dan Analisis Hasil

Tahap akhir dalam penelitian ini adalah visualisasi dan analisis hasil untuk mempermudah pemahaman. Visualisasi disajikan dalam bentuk dashboard yang memuat jenis diagram seperti *pie chart*, *bar chart*, dan lain-lain yang dapat digunakan untuk menampilkan distribusi sentimen serta topik yang dominan dalam data ulasan. Visualisasi ini membantu menyajikan informasi secara lebih jelas dan informatif. Hasil visualisasi ini kemudian dianalisis guna mengidentifikasi pola atau tren yang signifikan, sehingga dapat memberikan wawasan mendalam terkait opini masyarakat terhadap layanan RS ABC di DIY, sekaligus mendapatkan gambaran aspek layanan yang perlu diperbaiki berdasarkan ulasan yang tersedia.

PERPUSTAKAAN
UNIVERSITAS JENDERAL ACHMAD YOHANES
YOGYAKARTA