

BAB III

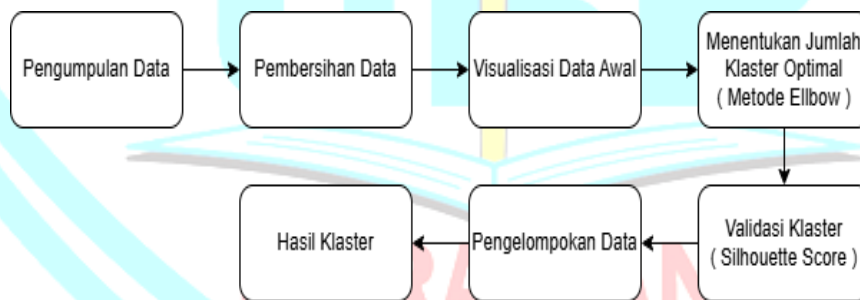
METODE PENELITIAN

3.1 Objek Penelitian

Peneliti berfokus pada analisis dan rekomendasi drama Korea berdasarkan preferensi penonton menggunakan dataset Kaggle. Dataset tersebut mencakup informasi tentang berbagai drama Korea, baik yang populer maupun yang kurang dikenal, untuk menentukan preferensi penonton terhadap drama. Penelitian mengarah kepada untuk memberikan drama Korea yang direkomendasikan dan kurang direkomendasikan.

3.2 Prosedur Penelitian

Penelitian ini menggunakan metodologi CRISP-DM (*Cross-Industry Standard Process For Data Mining*) yang terdiri dari enam tahapan berurutan. Metode ini bersifat fleksibel, dimana setiap tahapannya saling bergantung dan mengikuti hasil dari tahap sebelumnya. Tahapan-tahapan dalam CRISP-DM adalah sebagai berikut:



Gambar 3.1 Tahap Penelitian

1. Pengumpulan Data

Peneliti mengumpulkan informasi yang relevan untuk analisis. Dataset diperoleh dari Kaggle dan mencakup atribut utama seperti rating dan vote, yang menjadi indikator popularitas dan kualitas drama.

2. Pembersihan Data

Penelitian ini memahami tujuan dalam konteks bisnis dari sebuah proyek. Peneliti mengelompokkan data drama Korea guna memberikan rekomendasi film berdasarkan penilaian penonton. Tujuan dan batasan tersebut kemudian dirumuskan menjadi masalah yang akan diselesaikan melalui teknik data mining.

3. Visualisasi Awal Data

Penelitian ini menggunakan dataset yang terdiri dari drama Korea yang diperoleh dari Kaggle. Data yang digunakan untuk proses clustering mencakup atribut berupa *rating* dan *vote*.

4. Menentukan Jumlah Kluster Optimal (Metode Elbow)

Penelitian data yang telah diperoleh diperiksa untuk memastikan apakah terdapat atribut dengan nilai null (kosong) atau tidak. Hasilnya menunjukkan bahwa tidak ada data kosong dalam dataset, sehingga tidak diperlukan proses penghapusan data dengan nilai null

5. Validasi Kluster (Silhouette Score)

Untuk memperoleh hasil yang optimal, diterapkan model yang sesuai dalam proses pengklasteran. Dalam hal ini, digunakan algoritma K-Means. Adapun langkah-langkah yang dilakukan dalam penerapan algoritma tersebut mencakup beberapa tahapan.

1. Menentukan jumlah K-cluster.
2. Menentukan pusat kluster secara acak sesuai jumlah K yang ditentukan.
3. Menghitung jarak setiap data ke centroid menggunakan Euclidean Distance.

$$d(x_i, \mu_i) = (x_i - \mu_i)^2$$

Keterangan:

x_i = nilai *rating* atau jumlah *vote* dari suatu film.

μ_i = centroid kluster saat ini.

Jarak dihitung untuk semua film ke setiap kluster, lalu film dikategorikan ke kluster dengan jarak terdekat.

4. Mengelompokkan data berdasarkan jarak minimum ke centroid.
5. Memperbarui nilai centroid hingga tidak ada perubahan signifikan dalam kluster.

$$C_k = \frac{1}{n_k} \sum d_i$$

Keterangan:

C_k = Centroid baru

n_k = Jumlah film dalam kluster

d_i = Nilai rata-rata *rating* dan *vote* dalam kluster

6. Proses langkah 2-5 diulang hingga tidak ada perubahan signifikan dalam pengelompokkan

6. Pengelompokan Data

Tahap ini bertujuan untuk mengevaluasi atau menilai model yang telah ditentukan. Namun, dalam penelitian ini, tahap evaluasi tidak dilakukan karena penelitian hanya berfokus pada proses clustering untuk mengelompokkan data.

7. Hasil Klaster

Tahap ini melibatkan penerapan model yang telah dibuat dengan mengelompokkan setiap instance drama Korea ke dalam klaster masing-masing. Film-film dalam satu klaster memiliki kesamaan atribut, yaitu *rating* dan *vote*.

