

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Objek yang diteliti adalah kumpulan lagu – lagu *K-Pop* dengan data yang diperoleh dari 2 *Playlist* berbeda di *Spotify* yang dibuat oleh masing - masing *User Spotify* yaitu *Playlist* “K-Pop Hits 2021” yang dibuat oleh Filtr Éxitosp dan *Playlist* "TOP 100 KPOP Songs 2021" yang dibuat oleh David Li, sehingga data yang digunakan terdapat 237 baris dan seluruh lagu merujuk pada *Genre Pop*, meskipun terdapat campuran dari *genre* lain. Terdapat banyak musisi dan produser musik lokal yang berorientasi dengan gaya dan aliran yang berasal dari berbagai aliran baik dari aliran lokal maupun asing. Sehingga lagu yang diproduksi harus ada modifikasi berdasarkan kemiripan konten dan konteks. Maka dibuat model *clustering*, kemudian masing-masing *cluster* yang sudah terbentuk dianalisis pada distribusi terkait karakteristik lagu yang direpresentasikan oleh fitur-fitur lagu yang sebelumnya digunakan untuk pengelompokan [2].

3.2 Data Collection

Data yang dikumpulkan berasal dari *Spotify* menggunakan bantuan *python* dari library yaitu *spotipy* dengan mengakses *API* dari *Python* itu sendiri, agar data dari *Spotify* dapat diekstrak berdasarkan album atau artis yang dipilih dari lagu tersebut. Data dari *Spotify* diakses menggunakan akun *developer Spotify* agar bisa memperoleh *Client ID* dan *Secret Key* sebagai passwordnya.

Data lagu *K-Pop* masing-masing diekstrak berdasarkan 2 *playlist Spotify* yang dibuat oleh 2 pengguna berbeda, dimana masing - masing *playlist* tersebut merupakan lagu - lagu *K-Pop* terfavorit dan paling banyak didengar. Data dari *playlist* 1 berjumlah 126 baris dan data dari *playlist* 2 berjumlah 117 baris dengan jumlah kolom yang sama sebanyak 13 atribut. Masing-masing data dari *playlist* tersebut ditarik, pada mulanya setelah digabungkan menjadi 243 baris, setelah itu melakukan penghapusan duplikat pada data hingga jumlahnya menjadi 237 baris dan 17 atribut.

Variabel yang digunakan merupakan variabel – variabel yang merupakan kriteria dari sebuah audio lagu berdasarkan *audio features endpoint* yang terdapat pada *website developer Spotify*, yaitu seperti yang ditampilkan pada Tabel 3.1 sebagai berikut

Tabel 3.1 Penjelasan Variabel Atribut pada Dataset

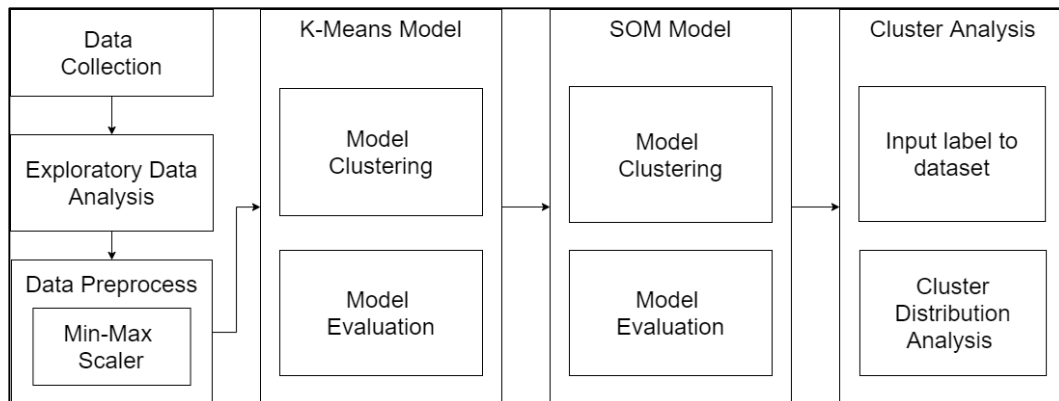
| No | Atribut | Keterangan | Tipe Data | Deskripsi |
|----|--------------|--|-----------|---|
| 1 | Key | perkiraan kunci keseluruhan track, tiap nilai menandakan nomor pada kunci tangga nadanya | Int | 0 = C, 1 = C# / D b , 2 = D dan seterusnya. kalau ga ada, nilainya -1 |
| 2 | Mode | major (1) minor (0) pada track | Int | 0 - 1 |
| 3 | Acousticness | menjelaskan seberapa akustiknya lagu | Float | 0.0 - 1.0 |
| 4 | Danceability | kecocokan lagu yang digunakan untuk menari atau tidak | Float | 0.0 - 1.0 |
| 5 | Energy | intensitas dan aktivitas yang dihasilkan dari sebuah lagu | Float | 0.0 - 1.0 |

| No | Atribut | Keterangan | Tipe Data | Deskripsi |
|----|----------|---|-----------|-----------------------|
| 6 | Liveness | probabilitas lagu tersebut direkam dengan penonton langsung pada saat acara <i>live</i> | Float | 0.0 - 1.0 |
| 7 | Loudness | menggambarkan kerasnya suara pada lagu | Float | Decibels (dB) |
| 8 | Valence | menjelaskan <i>mood</i> dari sebuah lagu | Float | 0.0 - 1.0 |
| 9 | Tempo | kecepatan ketukan pada lagu tersebut | Float | Beat Per Minute (BPM) |

Semua variabel fitur yang diperoleh dari library *spotify* diatas merupakan variabel sebagai fitur, karena metode penelitian ini menggunakan algoritma yang unsupervised karena belum ada labelnya buat diprediksi, jadinya belum diketahui seperti apa targetnya.

3.3 Alur Penelitian

Berikut merupakan *flowchart* untuk menggambarkan prosedur- prosedur yang dilakukan untuk mencapai tujuan dalam penelitian ini, *flowchart* pada Gambar 3.1 dibuat berdasarkan penalaran yang diperoleh dari pemahaman jurnal referensi penelitian terkait menurut Puteri & Irawan [2] tentang “*Analysis of Song Popularity in Business Digital Music Streaming for Increasing Quality Using Kohonen SOM Algorithm*” dan Widiarina & Wahono [15] tentang “*Algoritma Cluster Dinamik untuk Optimasi Cluster pada Algoritma K-Means dalam pemetaan Nasabah Potensial*” sebagai basis jurnalnya.



Gambar 3.1 Flowchart penelitian

3.3.1 Data collection

Penelitian ini menggunakan *dataset* yang sebelumnya diperoleh dari Spotify dengan mengekstrak data pada *Web API* melalui *Client ID* dan *Secret Key* dari akun developer, kemudian data di ekstrak dengan mengeksplor *Web API Reference*-nya. Untuk mengekstrak lagu, keduanya dilakukan berdasarkan *playlist* dari 2 pengguna *spotify* yang berbeda diperoleh dengan cara mengakses *user playlist* dan *tracks*-nya. Kemudian masing – masing dari data tersebut diperoleh menggunakan metode *request* untuk mengambil data dari *API Spotify* tersebut. Kemudian diubah menjadi tabel dan dicetak dalam format CSV.

Setelah data dibaca menggunakan metode dari *library* *pandas* pada *python*, maka dilakukan proses inspeksi data untuk memeriksa kebersihan data yaitu memastikan bahwa tidak ada data kosong yang disebut *null values* atau *missing values* dan data yang redundan atau duplikat yaitu data yang berulang dengan nilai yang sama.

3.3.2 Analisis Data Eksplorasi

Analisis Data Eksplorasi dilakukan untuk mempelajari datanya lebih lanjut setelah melakukan inspeksi atau pemeriksaan apabila data tidak bersih karena terdapat data kosong dan data yang redundan, yaitu dengan pembersihan data yang dikenal dengan *data cleansing* untuk menangani data kosong atau *handling missing values* dengan menghapus data kosongnya atau melakukan imputasi yang menggantikan isi nilai kosong dengan nilai tertentu, dan menangani data duplikat atau *handling duplicate values* dengan menghapus data duplikat.

Pada dasarnya, Analisis Data Eksplorasi merupakan metode untuk memahami isi data dan kondisi datanya dengan menemukan pola, memeriksa asumsi, menemukan anomali dengan bantuan statistik ringkasan dan representasi grafis yaitu melakukan visualisasi data agar dapat memperlihatkan representasi data dalam bentuk gambar sehingga lebih mudah dimengerti. Seperti mengamati distribusi pada kolom data numerik menggunakan histogram atau mengamati nilai pada kolom data kategorik menggunakan diagram batang atau *bar chart*.

3.3.3 Data Preprocess

Setelah data di ekstrak dengan *request* pada *API Spotify* hingga di cetak dalam format CSV, dataset tersebut harus diidentifikasi terlebih dahulu. Namun pada penelitian ini metode *preprocessing* data yang digunakan adalah *min-max scaler*, yaitu metode untuk memperoleh nilai per baris dari seluruh dataset hingga yang dihasilkan berbentuk *matrix*. Nilai yang diperoleh merupakan nilai minimum terlebih dahulu, kemudian nilai maksimumnya.

3.3.4 Pemodelan Data

Pada tahap ini, merupakan tahap dimana model *Clustering K-Means* diaplikasikan dimana hasil *output* dari model ini adalah *centroid*, yaitu titik pusat dari *cluster* dan tiap lagu akan masuk ke *cluster* dengan *centroid* terdekat pada lagu tersebut. Dalam proses penerapan model ini dilakukan pengelompokan model terlebih dahulu menggunakan lagu – lagu *K-Pop*. Kemudian *cluster* dari hasil *clustering* ditampilkan dengan cara visualisasi untuk menampilkan *scatter plot* untuk melihat *centroid* terdekat pada masing – masing *cluster* dari masing – masing data. Hal yang sama dilakukan pada penerapan Algoritma SOM yang melibatkan melalui *neuron* sebagai tempat data *input* diproses, kemudian *neuron* terbaik dipilih sebagai pemenang berdasarkan kesamaannya dengan bobot yang ada menggunakan metode *BMU*.

Setelah model dijalankan dalam proses pelatihan, validasi diterapkan untuk meningkatkan performa model tersebut. Metode Validasi yang digunakan yaitu *Elbow Method* yang digunakan untuk mencari kesalahan SSE terkecilnya dari seluruh nilai dalam suatu *cluster*, kemudian *Silhouette Method* untuk memperoleh nilai yang tiap *data poinnya* dihitung berdasarkan relasinya pada jarak antar *data point* terdekat pada *cluster* yang sama dan jarak antar *data point* terjauh pada *cluster* yang berbeda. Kemudian *cluster* dari hasil *clustering* dievaluasi menggunakan metode Indeks Davies-Bouldin (DBI) untuk mencari nilai terkecilnya yang mendekati 0 sebagai representasi kualitas *cluster* terbaik.

3.3.5 Analisis *Cluster*

Pada tahap ini, model *cluster* yang memberikan kinerja terbaik diukur dari nilai *Silhouette* dan *DBI* diterapkan pada dataset dengan memasukan label dari hasil *model cluster* yang sebelumnya dihasilkan pada pemodelan data. Setelah itu dilakukan analisis distribusi melalui histogram pada tiap *cluster* lagu *K-POP* berdasarkan karakteristik yang direpresentasikan oleh fitur-fitur yang tadinya digunakan pada proses pengelompokan, untuk mengidentifikasi karakteristik lagu K-Pop dari masing-masing *cluster*.