

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Lagu

Lagu merupakan kumpulan suara yang dipadukan dan disusun dengan urutan yang sesuai dengan irama sehingga menimbulkan suara yang khas. Menurut KBBI, lagu merupakan ragam suara yang berirama (dalam bercakap, bernyanyi, membaca dan sebagainya). Pada umumnya lagu dapat mempengaruhi emosi dan psikologis bagi para pendengar karena lagu terdapat nada yang dilantunkan berdasarkan ekspresi dan ungkapan perasaan dari penyanyi tersebut, sehingga para pendengar dapat memahami dan merasakan suasana hati yang ditimbulkan dari lagu tersebut. Dengan kata lain, lagu dapat digunakan sebagai media untuk komunikasi karena merupakan bahasa universal yang menerapkan emosi [16]. Selain itu, lagu juga dapat digunakan sebagai media untuk belajar Bahasa dengan cara yang interaktif dengan menggunakan melodi, nada dan irama dalam membaca suatu frasa pada katanya, sehingga memudahkan otak untuk mencerna apa yang dipelajarinya [17].

#### 2.2 Spotify

*Spotify* merupakan salah satu platform untuk *streaming* terbesar yang diluncurkan di Swedia pada tahun 2008 oleh Daniel Ek. Dengan misi yang ditampilkan pada *website* resminya, bahwa *spotify* berperan sebagai wadah untuk membuka potensi kreativitas manusia yang dilakukan dengan memberikan kesempatan bagi jutaan seniman yang ingin membuat karya terbaik secara kreatif

agar karya tersebut dinikmati dan menjadi inspirasi untuk miliaran penggemarnya, sehingga jumlah katalog di spotify telah mencapai lebih dari 50 juta *track* lagu, termasuk 700.000 *podcast*. Selain itu, *Spotify* menawarkan layanan secara gratis kepada pelanggan hingga layanan secara premium agar pelanggan dapat mengakses fitur eksklusif musik, termasuk peningkatan kualitas suara hingga pengalaman mendengar yang sesuai permintaan, mendengarkan lagu tanpa terkoneksi ke internet (*offline*), dan bebas tayangan iklan. Hingga saat ini, *spotify* merupakan layanan *streaming* audio berskala global yang paling populer dengan pengguna yang mencapai 271 juta, termasuk 124 juta pelanggan di 79 pasar [18].

### **2.3 Algoritma K Means Clustering**

Klasterisasi atau *Clustering* adalah pengelompokan objek – objek berdasarkan kemiripan pada karakteristik dari tiap objeknya. *Clustering* bertujuan untuk menghimpun dan meletakkan benda – benda yang memiliki kemiripan ciri khas hingga serupa menjadi berhubungan satu sama lain, sedangkan yang karakteristik objeknya terdapat sedikit kemiripan hingga berbeda akan dipisahkan dan dipindahkan ke kelompok lain, dimana pada kelompok tersebut terdapat kumpulan benda – benda lain yang kemiripannya lebih mendekati. Sehingga setiap *cluster* dapat mengklasifikasikan objek berdasarkan kesamaan pada objek yang paling dekat pada objek lain dalam suatu kelompok atau *cluster* yang sama [19].

K-Means merupakan salah satu algoritma yang paling populer yang digunakan untuk *clustering* secara non hierarki. K-Means bertujuan untuk menginput dataset, kemudian membagi datanya menjadi beberapa kelompok K yang disebut *cluster*, dimana jumlah K nya ditentukan sendiri. Pengelompokan

dilakukan secara iterative dengan pengelompokan pertama dilakukan pada *cluster* yang *centroid* sebagai pusat *cluster*-nya yang paling dekat [20].

Berikut langkah – langkah yang dikerjakan dalam algoritma *K-Means Clustering* [21]

- Menentukan jumlah k sebagai jumlah *cluster* yang dibentuk
- Tentukan titik pusat pada *cluster* K atau yang disebut *centroid* awal secara acak, dari beberapa objek yang tersedia sebanyak k clusternya
- Hitung jarak setiap objek ke masing-masing *centroid* dari masing – masing *cluster*, kemudian hitung jarak antar *centroid* nya
- Alokasikan masing – masing objek ke dalam *centroid* yang paling dekat
- Lakukan perulangan, kemudian menentukan posisi *centroid* baru dengan menggunakan persamaan
- Ulangi langkah ke 3 apabila posisi *centroid* yang baru tidak sama

Untuk menghitung *centroid cluster* ke-i, rumus yang digunakan sebagai berikut [21]

$$v = \frac{\sum_{i=1}^n x_i}{n} ; i = 1, 2, 3 \dots n$$

### **Rumus 2.1 Rumus K-Means Clustering**

Dimana :

- V, *centroid* pada *cluster*
- xi, objek ke-i
- n, banyaknya objek atau jumlah objek yang menjadi anggota *cluster*

Untuk menghitung jaraknya ke masing – masing *centroid*, dalam penelitian ini menggunakan rumus *Euclidean Distance* [21]

$$(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2 ; i = 1, 2, 3, \dots, n}$$

### **Rumus 2.2 Euclidean Distance**

Dimana:

- $x_i$ , objek x ke-i
- $y$ , data y ke-i
- $n$ , banyaknya objek

## **2.4 Kohonen Self Organizing Maps (SOM)**

*Kohonen Self Organizing Maps (SOM)* atau Jaringan Kohonen yang diperkenalkan oleh Professor Teuvo Kohonen, merupakan salah satu metode dari Jaringan Saraf atau *Neural Network* yang digunakan untuk membuat model pembelajaran tanpa pengawasan (*unsupervised learning*) karena tidak menggunakan nilai target kelas, sehingga tidak ada kelas yang ditetapkan untuk setiap data [13]. Arsitektur pada struktur Kohonen SOM terdiri dari 2 lapisan atau *layer* yaitu *layer input* untuk memasukan data yang dihubungkan ke tiap *neuron* pada *layer output* untuk menampilkan hasilnya. Tiap *neuron* di dalam *layer output* menampilkan kelas *cluster* berdasarkan data *input*-nya, jumlah pemrosesan *neuron* merupakan kelompok K, tiap data *input* dan bobot (*weights*) terkait setiap *neuron*. Arsitektur pada Kohonen SOM dijelaskan dalam bentuk topologi agar bisa memberikan visualisasi pengelompokan, maka tiap *neuron* di dalam SOM direpresentasikan dalam satu kelompok [2].

Kohonen SOM dapat menampilkan visualisasi hasil pengelompokan dalam bentuk topografi dua dimensi seperti peta, sehingga memudahkan pengamatan distribusi kelompok dari hasilnya. Penentuan fungsi seperti fungsi keterangan, laju pembelajaran, fungsi pembelajaran, jumlah kelompok  $k$  dan jumlah iterasi dengan cara percobaan secara acak, kemudian dipilih yang terbaik. Namun algoritma ini hanya cocok untuk data yang jumlah kelompoknya sudah diketahui, sehingga dalam Kohonen SOM tidak menjamin konvergensi hasil pengelompokannya, Karena SOM tidak menggunakan fungsi objektif untuk suatu kondisi yang sudah optimal pada tahap iterasi seperti *K-Means*, Maka SOM tidak akan menghentikan iterasinya sampai memenuhi jumlah iterasi yang telah ditentukan, tanpa memperdulikan hasil *Cluster*-nya yang sudah optimal atau belum [12].

Kohonen SOM bekerja dengan cara pelatihan yang berdasarkan pemenang atau *winner takes all*, dimana hanya neuron yang menjadi pemenang sebagai neuron terpilih yang bobotnya akan diperbaharui [13]. Neuron pemenang ditentukan dengan diawali penentuan bobot dan data *input* secara random yang salah satu *input*-nya dipilih, dan penentuan pemenang tersebut diperoleh dari hasil perhitungan tingkat kesamaan dengan perhitungan jarak antara data *input* dengan bobotnya, kemudian *input* dipilih berdasarkan kesamaannya dengan bobot yang ada menggunakan metode BMU (*Best Machine Unit*). Setelah neuron pemenang diperoleh, maka bobot dari data *input* diperbaharui dengan mendekati bobotnya [12].

Berikut merupakan cara kerja algoritma Kohonen SOM [2],

1. Inisialisasi jumlah *input* dan *output* neuronnya
2. Tentukan bobot neuron dengan persamaan berikut

$$W_{ij} = \frac{Minp_i + Maxp_i}{2}$$

### Rumus 2.3 Menentukan Bobot Neuron

3. Ulangi langkah ke 3 hingga 5
4. Hitung jarak dari vektor *input* dengan bobotnya untuk tiap neuron *output* dengan persamaan berikut

$$d_j = \sqrt{\sum_{i=1}^n (W_{ij} - x_i)^2}$$

### Rumus 2.4 Jarak Antar Vektor

5. Tentukan indeks *j* sedemikian rupa hingga ke jarak minimum
6. Lakukan peningkatan pada nilai dari tiap satuan *j* antara *j* dengan persamaan berikut

$$W_{ij} \text{ baru} = W_{ij} \text{ lama} + \alpha(x_i - W_{ij} \text{ lama})$$

### Rumus 2.5 Meningkatkan Bobot I dan J

7. Modifikasi kecepatan belajar dengan (*alpha*) melalui persamaan berikut

$$\alpha(t + 1) = 0.5\alpha(t)$$

### Rumus 2.6 Modifikasi Kecepatan Belajar

## 2.5 Analisis Data Eksplorasi

Analisis Data Eksplorasi atau *Exploratory Data Analytics* merupakan metode yang seringkali digunakan untuk menganalisis berbagai macam aspek, khususnya dalam lingkungan penelitian baik ilmu pengetahuan alam maupun sosial. Analisis Data Eksplorasi seringkali digunakan untuk mencari perilaku pada pola variable untuk dipelajari, menentukan hipotesis atau dugaan sementara dengan struktur sekecil mungkin. Analisis Data Eksplorasi biasa ditampilkan dalam bentuk [22].

Analisis Data Eksplorasi juga dikenal sebagai pendekatan untuk berkomunikasi secara langsung dengan datanya. Metode ini membantu dalam menganalisis data dalam bentuk kumpulan data atau *dataset* agar dapat merangkum karakteristik – karakteristik statistik pada aspek seperti pengukuran dari pusat tendensius (*measure of central tendency*) yang mengukur nilai rata-rata (*mean*), menghitung nilai yang sering muncul (modus) dan nilai tengah (*median*), pengukuran persebaran data yang menghitung standar deviasi, varians, bentuk sebaran dan keberadaan pencilan (*outliers*) [23].

Salah satu metode pada Analisis Data Eksplorasi yaitu Eksplorasi data yaitu memahami isi dari dataset dan karakteristik dari dataset tersebut. Mulai dari ukuran data, tipe data hingga mencari keterkaitan atau korelasi antar data. Eksplorasi Data bisa dilakukan dengan cara visualisasi data dalam bentuk grafik agar dapat memperlihatkan representasi data sehingga lebih mudah dimengerti [23].

## 2.6 Data Preprocessing

*Data Preprocessing* merupakan teknik untuk mengidentifikasi data – data yang baru saja diperoleh dan dikumpulkan sebelum diolah dan diterapkan ke model. Proses ini diawali dengan mengubah data mentah menjadi bentuk yang dapat disesuaikan pada model yaitu sekumpulan algoritma yang dirancang berdasarkan teknik dan metode yang diaplikasikan. Teknik ini menjadi proses yang paling memakan waktu dan membutuhkan upaya yang signifikan, apabila data tersebut berjumlah besar dan terdapat banyak fitur. *Data Preprocessing* bertujuan untuk mengevaluasi kualitas data – data yang akan dipakai pada model tersebut [24]

Pada penelitian ini, salah satu metode *Data Preprocessing* yang digunakan adalah *min max scaler*. *Minmax* merupakan matrix yang diperoleh dengan terlebih dahulu memperoleh nilai minimal pada setiap baris dari seluruh datasetnya, kemudian mengambil nilai maksimalnya [25].

Berikut merupakan persamaan matrix pada *min-max scaler* [25]

$$M = \begin{pmatrix} X_{1,1} & \cdots & X_{1,n} \\ \vdots & \ddots & \vdots \\ X_{m,1} & \cdots & X_{m,n} \end{pmatrix}$$

### Rumus 2.7 Min-max Scaler

Dimana :

- $X_{1,1}$ , merupakan nilai pada baris dan kolom ke 1
- $X_{1,n}$ , merupakan nilai pada baris ke 1 dan kolom ke n
- $X_{m,1}$ , merupakan nilai pada baris ke m dan kolom ke 1
- $X_{m,n}$ , merupakan nilai pada baris ke m dan kolom ke n



Dengan perhitungan nilai Max  $\Delta = \{\Delta_1, \dots, \Delta_n\}$  dimana  $\Delta$  merupakan satuan nilai maksimum dalam satu baris data, dan nilai Min  $\nabla = \{\nabla_1, \dots, \nabla_n\}$  dimana  $\Delta$  merupakan satuan nilai minimum [25].

## 2.7 Optimisasi

Optimisasi merupakan proses untuk meningkatkan performa model dengan mengoptimalkan parameter – parameternya dan juga menggunakan cara mengekstraksi pada fitur – fiturnya. Proses Optimisasi yang digunakan dalam penelitian ini, antara lain Metode *Elbow*, Metode *Silhouette* dan *Davies-Bouldin Index* [26].

### 2.7.1 Elbow Method

Metode Elbow yang merupakan metode tertua untuk menentukan jumlah *cluster* yang tepat dari dataset yaitu metode optimisasi dengan menjalankan klusterisasi k-means dari dataset dengan rentang interval nilai dari data tertentu dari k nya, kemudian menghitung rata - rata kesalahan (*Sum of Squared Errors*) dan hasilnya ditampilkan melalui visualisasi dalam bentuk grafik garis. Tujuannya untuk mencari nilai kesalahan SSE terkecilnya dari semua nilai yang terdapat pada K nya [27].

Semakin besar jumlah cluster K, maka nilai SSE semakin kecil. Untuk menghitung SSE (*Sum of Squared Errors*) dari masing – masing cluster, berikut rumus persamaannya [28].

$$SSE = \sum_{K=1}^k \sum_{X_i \in S_K} \|X_i - C_k\|_2^2$$

### Rumus 2.8 Elbow Method

Dimana :

- K, merupakan nilai k, mulai dari yang diinisiasi hingga mengalami penurunan
- $X_i$ , merupakan data ke i
- $C_k$ , merupakan *Cluster* pada K nya

Algoritma perhitungan Metode *Elbow* dalam menentukan nilai K pada K-Means

- Inisiasi awal nilai K
- Tambahkan nilai K
- Hitung hasil SSE (*Sum of Squared Errors*) dari tiap nilai K
- Melihat hasil SSE (*Sum of Squared Errors*) dari nilai K yang turun secara drastis
- Tetapkan nilai K yang berbentuk siku

#### 2.7.2 Silhouette Method

Metode silhouette merupakan teknik untuk memperoleh nilai yang tiap *data poinnya* dihitung berdasarkan relasi antar *data point* untuk masing – masing *cluster*, kemudian dibandingkan dengan *cluster* lainnya [29]

Berikut merupakan persamaan dari *elbow method*

$$j = \frac{b - a}{(a_j, b_j)}$$

### Rumus 2.9 Silhouette Method

- $a_j$ , merupakan rata – rata dari jarak data point untuk *data point* yang lain dalam *cluster* yang sama
- $b_j$ , merupakan rata – rata minimum dari jarak data point untuk *data point* yang lain pada *cluster* yang berbeda dan diminimalkan dalam *cluster*-nya
- dimana tiap nilai dari *silhouette* berada di antara -1 sampai +1, yang merupakan nilai *silhouette* yang lebih besar menggambarkan bahwa  $j$  terdefinisi dengan baik pada *cluster*-nya sendiri

### 2.7.3 Davies Bouldin Index

Metode Indeks Davies-Bouldin atau DBI merupakan teknik untuk validasi *cluster* yang diperoleh dari hasil *clustering* dengan menggunakan fungsi rasio jumlah antar *cluster scatter* sampai pemisahan *cluster*, kemudian memaksimalkan jarak antar *cluster* dan meminimalkan jarak data di dalam *cluster*. Sehingga diperoleh nilai DBI yang semakin kecil dan mendekati 0, maka merepresentasikan nilai pada kualitas *cluster* yang semakin optimal [15].

Untuk perhitungan dalam validasi *cluster* tersebut, berikut metrik yang digunakan [2]

- Jumlah akar di dalam *cluster* atau *Sum of Square Within Cluster* (SSW) sebagai metrik kohesi. dengan persamaan sebagai berikut

$$SSW = \frac{1}{m_i} \sum_{n=1}^m d_n(x_j, c_i)$$

#### **Rumus 2.10 Sum of Square Within Cluster**

- Jumlah akar di dalam *cluster* atau *Sum of Square Between Cluster* (SSB) sebagai metrik pemisahan untuk mengukur jarak antar pusat *cluster* atau *centroid*. Dengan persamaan sebagai berikut,

$$SSB_{i,j} = d(c_i, c_j)$$

#### **Rumus 2.11 Sum of Square Between Cluster**

- Rasio perbandingan antara *cluster* i dan j, nilai yang diperoleh dari hasil perhitungan menggunakan metrik SSW dan SSB. Dimana *cluster* terbaik diperoleh dari kohesi sekecil mungkin atau sebesar mungkin, dengan persamaan sebagai berikut,

$$R_{ij} = SSW_i + \frac{SSW_j}{SSB_{i,j}}$$

#### **Rumus 2.12 Rasio Perbandingan**

dengan karakteristik sebagai berikut

- $R_{ij} \geq 0$
  - $R_{ij} = R_{ji}$
  - Jika  $SSW_j \geq SSW_r$  dan  $SSB_{i,j} = SSB_{i,r}$  maka  $R_{i,j} = R_{i,r}$
  - Jika  $SSW_j = SSW_r$  dan  $SSB_{i,j} = SSB_{i,r}$  maka  $R_{i,j} > R_{i,r}$
- Setelah diperoleh rasio perbandingan antara *cluster*, maka dicari nilai DBI dimana nilai DBI diperoleh dari nilai rata-rata perhitungan menggunakan komponen metrik diatas dengan nilai SSW terkecil yang memberikan hasil *clustering* terbaik. Dengan persamaan sebagai berikut,

$$DBI = \frac{1}{k} \sum_{i=1}^k \frac{k}{i} \max_{i=j} (R_{i,j})$$

### **Rumus 2.13 Davies-Bouldin Index**

## **2.8 Tools**

*Library* Python yang digunakan dalam penelitian ini yaitu *Scikit-Learn*, yang merupakan modul python yang terintegrasi secara ultra modern untuk menangani masalah yang berskala menengah. *Library* ini menekankan pada kemudahan dalam penggunaan dan pengoperasiannya pada kinerja, dokumentasi dan konsistensi dari *API*-nya. *Library* ini pada pengembangannya berada di bawah lisensi dari BSD [30]. Selain itu, *Library* lain yang digunakan yaitu *Spotipy* (singkatan dari *Spotify-Python*), *Spotipy* merupakan modul Python yang digunakan untuk mengekstrak data dari *API Spotify* dengan cara memperoleh *Client ID* dan

*Secret Key* nya terlebih dahulu, sehingga dengan akses dari ID dan kunci passwordnya ini dapat terhubung ke *database* pada *Spotify* nya. Hal ini harus didahului dengan membuat akun pengembang (*developer account*) pada *Spotify* nya terlebih dahulu [31].

## 2.9 Penelitian terdahulu

**Tabel 2.1 Penelitian terdahulu**

Judul	Penulis	Jurnal	Metode
<i>Analysis of Song Popularity in Business Digital Music Streaming for Increasing Quality Using Kohonen SOM Algorithm</i>	Chyntia Kumalasari Puteri dan M. Isa Irawan	<i>IPTEK Journal of Proceedings Series</i> (2019), Vol. 0 No. 5	<i>Clustering with Kohonen SOM dan Multiple Linear Regression</i>
<i>Indonesian's Traditional Music Clustering Based on Audio Features</i>	Aisha Gemala Jondya dan Bambang Heru Iswanto	<i>Procedia Computer Science Journal</i> (2017), Vol. 116, P. 174-181	<i>Principal Component Analysis, Agglomerative Hierarchical Clustering dan X-means Clustering</i>
<i>What Makes a Song Trend? Cluster Analysis of Musical Attributes for Spotify Top Trending Songs</i>	Zayd Al-Beitawi, Mohammad Saleban, and Sonya Zhang	<i>Journal of Marketing Development and Competitiveness</i> (2020), Vol. 14 No. 3	<i>K-Means Clustering dan Agglomerative Clustering</i>
<i>Understanding User-Curated Playlists on Spotify</i>	Martin Pichl, Eva Zangerle, dan Günther Specht	<i>International Journal of Multimedia Data Engineering and Management</i> (2017), Vol. 8 No. 4	<i>Principal Component Analysis dan Singular Value Decomposition</i>

Judul	Penulis	Jurnal	Metode
Sebuah Pola Cluster Geospasial Eksplorasi Kejahatan Narkoba di DKI Jakarta	Raymond Sunardi Oetama, Tan Thing Heng, dan David Tjahjana	Ultima InfoSys : Jurnal Ilmu Sistem Informasi (2020), Vol. 11 No. 1	<i>K-Means Clustering</i> dan Visualisasi <i>Geospasial</i>
K-Means Clustering Video Trending di Youtube Amerika Serikat	Kevin Widjaja, dan Raymond Sunardi Oetama	Ultima InfoSys : Jurnal Ilmu Sistem Informasi (2020), Vol. 11 No. 2	<i>K-Means Clustering</i> , <i>Dunn Index</i> (Validasi) dan <i>Silhouette Coefficient</i> (Validasi)

Sesuai dengan Tabel 2.1 mengenai informasi jurnal, bahwa referensi penelitian mulai dari jurnal [2] mengenai “*Analysis of Song Popularity in Business Digital Music Streaming for Increasing Quality Using Kohonen SOM Algorithm*” dengan permasalahannya yaitu Perdagangan musik sangat membutuhkan layanan *streaming* music untuk prospek jangka panjang, penelitian ini membagi 4 cluster (61 lagu pada *cluster* 1, 27 lagu pada *cluster* 2, 59 lagu pada *cluster* 3, 53 lagu pada *cluster* 4) dengan *cluster* ke 2 yang dipilih karena merupakan *cluster* dengan jumlah *stream* terbanyak atau yang sering diputar oleh pengguna *spotify*. Dalam hal ini yang diadopsi yaitu metode pengelompokan menggunakan algoritma *clustering* pada data lagu dari *Spotify* dengan fitur audio sebagai atribut atau variabelnya, sehingga menghasilkan *cluster* lagu berdasarkan kemiripan karakteristiknya dari fitur audio tersebut.

Penelitian [9] mengenai “*Indonesian's Traditional Music Clustering Based on Audio Features*” yang permasalahannya yaitu terdapat tantangan dalam menganalisis kesamaan musik yaitu membedakan antar gaya musik yang tidak

dapat dibandingkan secara langsung, dengan menggunakan X- Means dan AHC menghasilkan 4 *cluster* dari 101 lagu (25 lagu pada *cluster* 0, 27 lagu pada *cluster* 1, 15 lagu pada *cluster* 2, 34 lagu pada *cluster* 3), dimana persebaran lagu dari kedua metode cenderung mirip, dimana hasil dari X- Means dapat dilihat dari komposisinya. Sebaliknya X-Means tidak mengklasifikasikan lagu berdasarkan *cluster* yang sama, sedangkan AHC menempatkan kedua lagu provinsi pada *cluster* yang sama. Dalam hal ini yang diadopsi yaitu metode pengelompokan menggunakan algoritma *clustering* pada data lagu yang merupakan lagu-lagu daerah dari 18 provinsi di Indonesia dengan ciri khas asalnya masing-masing pada karakteristik lagu provinsi tersebut.

Penelitian [5] tentang “*What Makes a Song Trend? Cluster Analysis of Musical Attributes for Spotify Top Trending Songs*” yang membahas bagaimana memahami atribut tertentu yang membuat lagu tersebut menjadi *trending*, dapat membantu layanan menciptakan pengalaman pelanggan yang lebih baik serta upaya pemasaran yang lebih efektif, dengan menggunakan top 100 lagu *trending* dari 2017 dan 2018 menggunakan variable lagu yaitu, *danceability*, *energy*, *loudness*, *speechiness*, *acousticness*, *instrumentalness*, *liveness*, *valence*, dan *tempo*. Hasil penelitian menunjukkan bahwa struktur musik dengan *danceability* yang tinggi dan *instrumentalness* yang rendah meningkatkan popularitas sebuah lagu dan membawa mereka ke puncak kesuksesan. Dalam hal ini yang diadopsi yaitu metode pengelompokan data lagu menggunakan algoritma *clustering* dengan k yang optimal menjadi beberapa *cluster*, kemudian memahami ciri-ciri dari masing-masing *cluster* dan menganalisis polanya berdasarkan variabel tertentu yang



berkontribusi secara krusial sehingga lagu tersebut menjadi tren. Kemudian menyelidiki jika tiap *cluster* cocok dengan *genre* lagu tertentu yang berpotensi memberikan informasi terkait variabel yang membentuk *genre* tertentu.

Penelitian [4] mengenai “*Understanding User-Curated Playlists on Spotify*” yang membahas bagaimana membangun sistem rekomendasi yang menyarankan musik yang sesuai dengan konteks pengguna berdasarkan seleranya, melalui pemahaman mendalam tentang karakteristik dari *playlist* yang dibuat secara mandiri oleh pengguna tersebut yang lebih dipersonalisasi dengan menganalisis *playlist* yang dibuat secara mandiri oleh pengguna *Spotify* tersebut. Hasil penelitian menunjukkan bahwa temuan tentang karakteristik *playlists* dapat dieksploitasi dengan sistem rekomendasi berbasis *Singular Value Decomposition* dan pendekatan pengelompokan untuk menemukan daftar kelompok *playlist* mudah diintegrasikan pada sistem rekomendasi menggunakan teknik *pre* atau *post-filtering*. Dalam hal ini yang diadopsi yaitu bagaimana penerapan sistem rekomendasi pada *Spotify* dengan algoritma yang berdasarkan pemahaman mendalam sesuai dengan variabel pada lagu yang merepresentasikan karakteristik lagunya yang berada pada *playlist Spotify* yang dibuat oleh penggunanya secara mandiri.

Penelitian [32] mengenai “Sebuah Pola Cluster Geospasial Eksplorasi Kejahatan Narkoba di DKI Jakarta” yang membahas Tingkat kejahatan yang muncul di Indonesia yang semakin meningkat setiap 1 Menit 33 detik, khususnya di DKI Jakarta yang merupakan provinsi dengan kejahatan tertinggi sebanyak

34.655 dengan data yang diperoleh dari Biro Pusat Statistik, Databoks dan Jakarta *Open Data*. Dengan menggunakan visualisasi *geospatial* untuk membagi wilayah daerah di Jakarta berdasarkan derajat tingkatan kejahatan dengan jenis kejahatannya, Maka hasil penelitian menunjukkan *clustering* pada jenis-jenis kejahatan yang terjadi di masing-masing wilayah Jakarta dengan tingkat kejahatan yang beragam. Dalam hal ini yang diadopsi yaitu Pada penggunaan metode *clustering* untuk mengelompokan data kasus kejahatan yang terjadi per wilayah di DKI Jakarta dan Jenis Kejahatannya, yang penerapannya dibantu dengan teknik Visualisasi *Geospatial* karena kasus yang dihadapi berupa informasi geografis.

Dan penelitian [33] mengenai “K-Means Clustering Video Trending di Youtube Amerika Serikat” yang mengenai para seniman yang bekerja di eniman yang bekerja di seniman yang disebut sebagai yang disebut sebagai *Youtuber* yang bekerja di *Youtube* sebagai platform tempat mereka berkarya dan mengunggah hasil karyanya berupa video untuk menghasilkan uang hingga menjadi sarana bisnis, dengan membuat video tersebut mencapai posisi *trending*. Video yang *Trending* sendiri ditentukan berdasarkan karakteristik yaitu *views*, *likes* dan *comment* sebagai variabel yang digunakan untuk model *clustering* untuk menentukan ciri-ciri video yang *trending* berdasarkan kategori yang dibuat oleh model *cluster* tersebut. Hasil penelitian menampilkan 3 *cluster* sebagai kategori kelompok yang terdiri dari banyaknya video dengan rata-rata *views*, *likes* dan *dislikes*-nya. Dalam hal ini mengadopsi penerapan model *clustering* untuk mengelompokan data berdasarkan kemiripan karakteristiknya menjadi beberapa *cluster* dan dikategorikan sesuai isi *cluster*-nya.