

CLUSTERING KARAKTERISTIK LAGU TOP 50 SPOTIFY 2019

Rifai Nur Widyanara[†] dan Yopi Azani[‡]
(19917013@students.uii.ac.id) (19917019@students.uii.ac.id)

[†]Konsentrasi Sains Data, Magister Teknik Informatika, Fakultas Teknologi Industri
[‡]Universitas Islam Indonesia



Latar Belakang

Spotify merupakan layanan *streaming* musik digital yang memberikan akses penggunanya ke jutaan *track* atau lagu dari berbagai artis di seluruh dunia. Karena lagu-lagu yang terdapat di Spotify berasal dari berbagai negara, maka karakteristik dari setiap lagu akan berbeda antara satu dengan yang lain. Oleh karena itu kami akan melakukan pengelompokan *track* atau lagu sesuai dengan karakteristik setiap lagu yang ada di Spotify berdasarkan atribut *Popularity*, *Valence*, *Energy*, *Loudness*, *Danceability*, *Acousticness* dan *Liveness*.

Tujuan Analisis

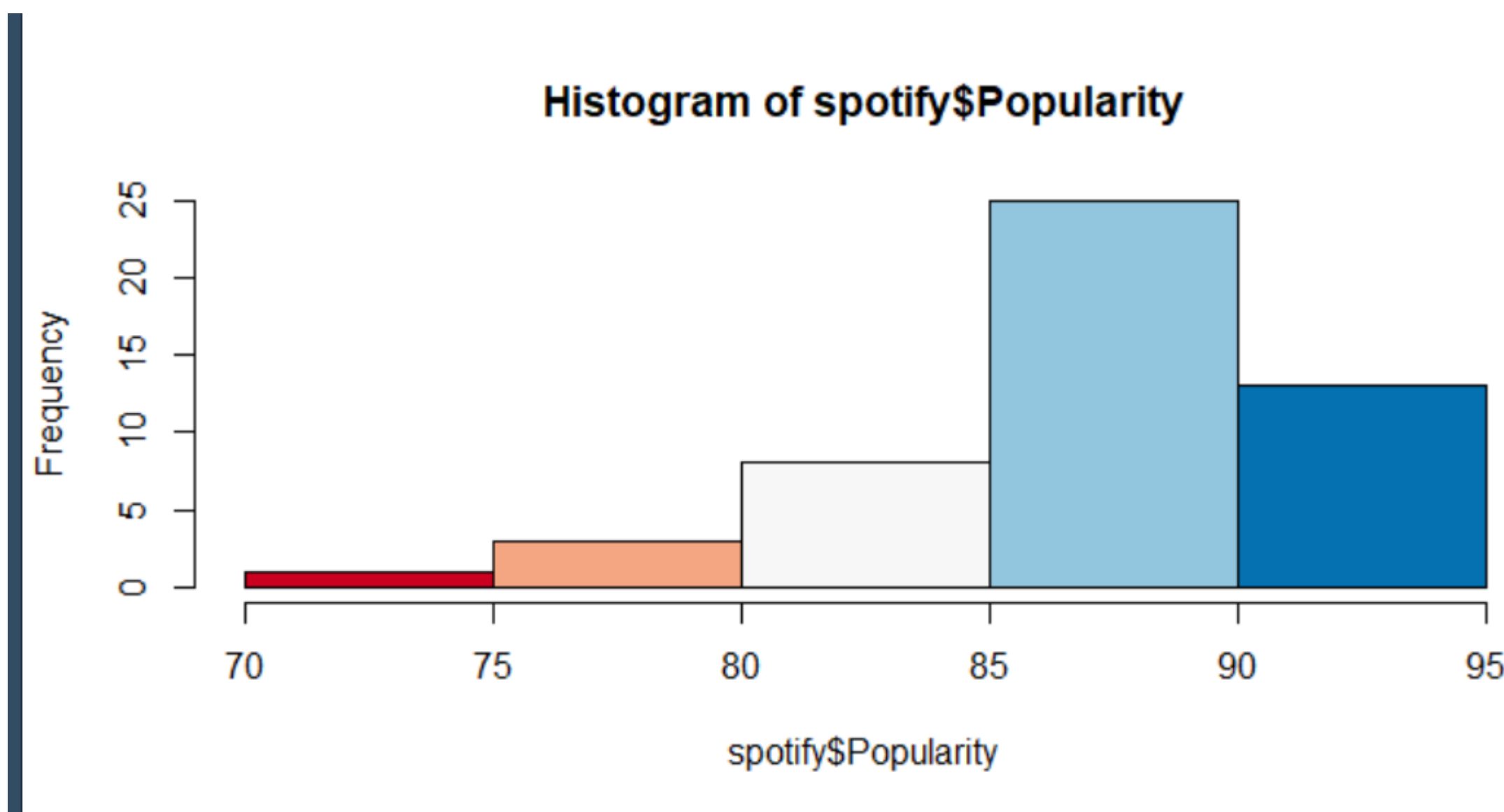
Melakukan *clustering* pada data Top 50 Spotify song 2019 sesuai dengan karakteristik dari setiap lagu tersebut. Sehingga dapat diketahui kelompok dari setiap tertentu berdasarkan karakteristiknya.

Data Set

Dataset yang digunakan adalah data "Top 50 Most listened Songs" pada aplikasi Spotify yang diperoleh dari kaggle. Data ini terdiri dari 50 observasi (lagu) dan 13 variabel(fitur), yaitu :

- *Track Name*
- *Beats per Minutes*
- *Loudness*
- *Speechiness*
- *Artist Name*
- *Energy*
- *Liveness*
- *Popularity*
- *Genre*
- *Danceability*
- *Acousticness*

Visualisasi data Popularity:



Berdasarkan gambar Histogram di atas, dapat dilihat bahwa frekuensi terbanyak adalah pada track/lagu dengan nilai Popularity antara 85 – 90, yang menandakan bahwa lagu-lagu yang ada pada dataset ini rata-rata mempunyai nilai Popularity yang cukup tinggi dan sering diputar oleh para pendengarnya.

Algoritma

Algoritma yang digunakan adalah K-Means Clustering untuk melakukan proses kluster data dan selanjutnya menentukan label yang ada sesudah proses kluster selesai dilakukan. Menurut Santosa[1] (2007), langkah-langkah melakukan clustering dengan metode K-Means adalah sebagai berikut:

1. Tentukan jumlah kluster, k.
2. Inialisasi k pusat kluster secara random.
3. Alokasikan semua objek observasi ke kluster terdekat, yang dihitung menggunakan jarak Euclidean yang dirumuskan sebagai berikut:

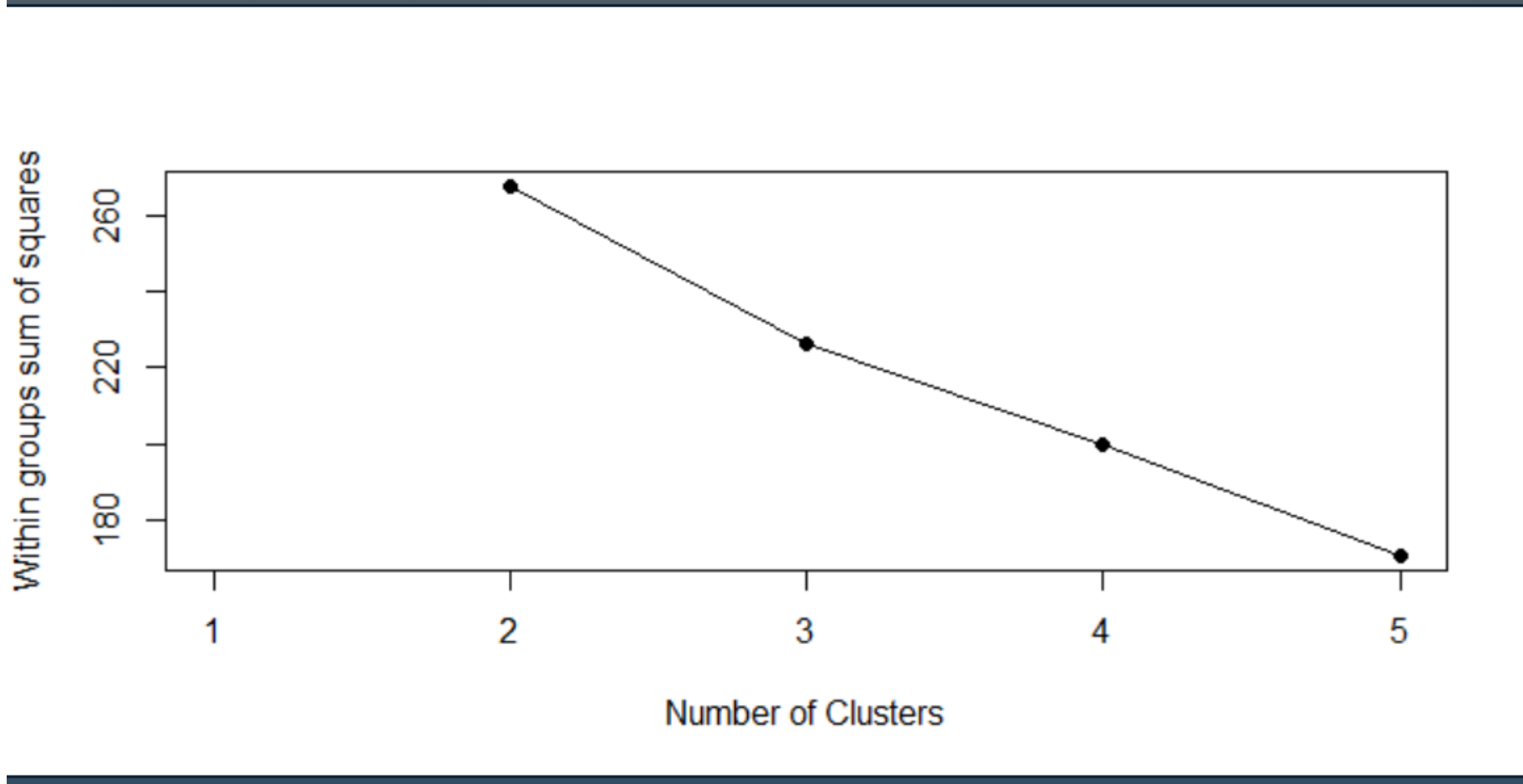
$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (1)$$

Dimana:
 $D(i, j)$ = Jarak antara observasi ke-i ke pusat kluster
 X_{ki} = Observasi ke-i pada atribut ke-k.
 X_{kj} = Titik pusat ke-j pada atribut ke-k.
 $i = 1, 2, \dots, n$ $j = 1, 2, \dots, n$ $k = 1, 2, \dots, n$

4. Hitung kembali pusat kluster dengan keanggotaan kluster yang sekarang.
5. Alokasikan kembali setiap objek memakai pusat kluster yang baru. Jika pusat kluster tidak berubah lagi, proses clustering selesai.

Hasil dan Pembahasan

Salah satu cara yang sering digunakan untuk menentukan jumlah kluster pada K-Means Clustering adalah Elbow Method. Elbow Method menggunakan nilai total WSS (Within the Sum of Squares) sebagai penentu nilai k yang optimal.



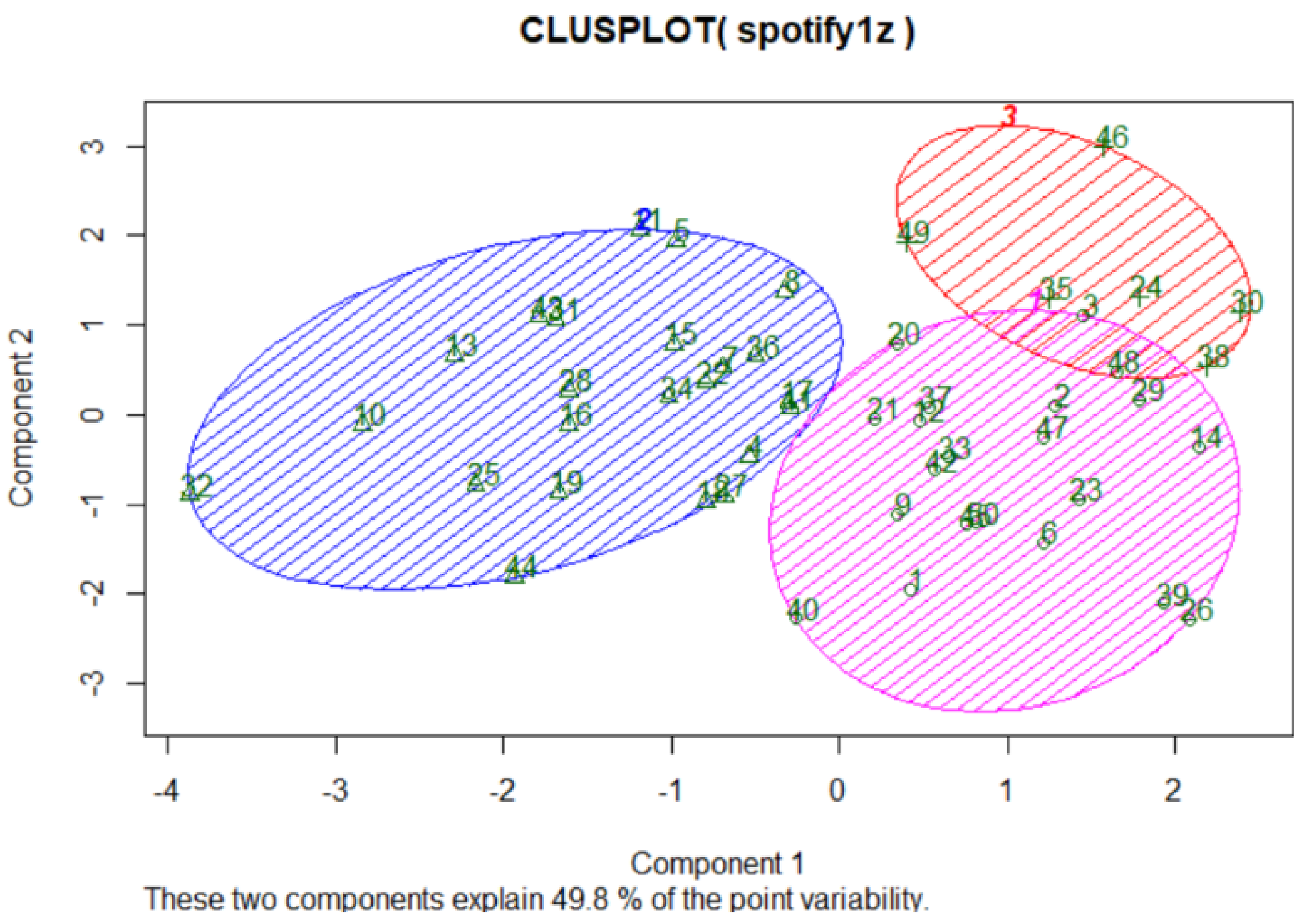
Dengan melihat plot-nya, garis mengalami patahan yang membentuk "Elbow" atau siku pada saat k=3, sehingga dipilih 3 sebagai jumlah kluster untuk analisis ini.

Cluster	Energy	Danceability	Loudness	Liveness	Valence	Acousticness	Popularity
<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0.581	0.307	0.529	-0.458	0.608	-0.279	-0.398
2	-0.759	-0.243	-0.652	-0.204	-0.693	0.328	0.295
3	0.874	-0.144	0.645	2.38	0.525	-0.280	0.260

Dari output hasil KMeans, diperoleh bahwa untuk jumlah kluster, k=3, terbentuk kluster 1 Sebanyak 21 track/song, kluster 2 sebanyak 23 track/song, dan kluster 3 sebanyak 6 track/song. Untuk merepresentasikan krakteristik tiap kluster, dapat menggunakan acuan Nilai means tiap kelompok yang terbentuk. Untuk profilisasi tiap kelompok yang terbentuk adalah :

1. Kluster 1 merupakan kumpulan track/song dengan nilai Valence dan Danceability paling tinggi dibanding kluster yang lain
2. Kluster 2 merupakan kumpulan track/song dengan nilai Acousticness dan Popularity yang paling tinggi dibanding kluster yang lain.
3. Kluster 3 merupakan kumpulan track/song dengan nilai Energy, Loudness, dan Liveness yang paling tinggi dibanding kluster yang lain.

Visualisasi Clustering Karakteristik Lagu yang masuk dalam Top 50 Spotify 2019



Kesimpulan

Berdasarkan hasil dari analisis ini dapat diambil kesimpulan bahwa untuk data Top 50 Spotify tahun 2019 sebagian besar lagu yang masuk dalam Top 50 Spotify 2019 adalah lagu dengan karakteristik *Valence*, *Danceability*, *Acousticness*, dan *Popularity* yang tinggi

Daftar Pustaka

[1] Johan Oscar Ong. "IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN STRATEGI MARKETING PRESIDENT UNIVERSITY". In: *Jurnal Ilmiah Teknik Industri* (2013).