

KLASIFIKASI TOP 50 SPOTIFY TAHUN 2010-2019 MENGUNAKAN METODE K-MEANS CLUSTERING

TOP 50 SPOTIFY CLASSIFICATION FOR 2010-2019 USING K-MEANS CLUSTERING METHOD

Uly Laili Musyarofah¹⁾, Suci Nur Alima²⁾, Dhian Satria Yudha K.³⁾

E-mail : ¹⁾19082010007@student.upnjatim.ac.id, ²⁾19082010019@student.upnjatim.ac.id
³⁾dhian.satria@upnjatim.ac.id

^{1,2,3} Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pembangunan “Veteran” Jawa Timur

Abstrak

Penduduk seluruh dunia memiliki target pemasaran yang cukup bagus dalam promosi produk, seperti produk teknologi berlandaskan media digital musik streaming. Di era perkembangan teknologi yang semakin pesat inilah menimbulkan pengaruh pada kehidupan masyarakat, seperti munculnya media layanan streaming musik yang memudahkan masyarakat dalam mendengarkan musik. Media layanan streaming musik salah satunya aplikasi Spotify, sesuai artikel dari kompas.com Spotify mengalami pertumbuhan sebesar 130 juta pelanggan. Media platform ini memiliki fitur unik bagi anak muda. Penelitian ini akan menjelaskan klasifikasi lagu pop terpopuler pada tahun 2010-2019 menggunakan metode *K-Means* pada Spotify. Klasifikasi menggunakan dataset lagu pop terpopuler yang dianalisis menggunakan metode *K-Means* untuk mengetahui top genre serta memakai dataset yang berasal dari Website Kaggle.com yang berisikan CSV sebagai kumpulan lagu dari aplikasi Spotify. Data dari file CSV berisikan 2001 data. Dalam menemukan cluster yang sesuai, penelitian diperoleh dari hasil analisis cluster pie chart, klaster diagram batang, dan titik koordinat hubungan valensi lagu.

Kata kunci: Data Mining, Klasifikasi, K-Means, Spotify

Abstract

People all over the world have pretty good marketing targets in product promotion, such as technology products based on streaming music digital media. In this era of increasingly rapid technological development, it has an influence on people's lives, such as the emergence of music streaming media services that make it easier for people to listen to music. Music streaming service media, one of which is the Spotify application, according to an article from kompas.com, Spotify has grown by 130 million subscribers. This media platform has unique features for young people. This study will explain the classification of the most popular pop songs in 2010-2019 using the K-Means method on Spotify. The classification uses the most popular pop song dataset which is analyzed using the K-Means method to find out the top genre. The research results will be explained in detail in the results and discussion chapters.

Keywords: Data Mining, Classification, K-Means, Spotify

1. PENDAHULUAN

Pesatnya perkembangan teknologi dan globalisasi memberikan pengaruh baik dan buruk bagi masyarakat, seperti munculnya seperti munculnya media layanan streaming musik yang memudahkan masyarakat dalam mendengarkan musik. Media layanan streaming musik ini salah satunya aplikasi Spotify [1]. Aplikasi Spotify memerlukan ulasan lebih lengkap mengenai media layanan musik langsung yang memberikan dan upaya pengalaman pengguna lebih baik, serta bertujuan menumbuhkan persaingan dengan media layanan streaming musik lain melalui analisis data mining. Proses penelitian ini memanfaatkan data publik dengan melakukan klasifikasi Top 50 tahun 2010-2019 yang akan dikelompokkan [2]. Penelitian ini memakai dataset Global Top 50 tahun 2010-2019 yang akan diklasifikasikan sesuai genre, penelitian ini dilaksanakan sebab

masyarakat mendengarkan musik cenderung berdasarkan genrenya. Spotify muncul pada tahun 2008 sebagai media penyebaran penyiaran musik secara digital [3]. Layanan yang disediakan pada Spotify memiliki fitur berbayar dan gratis bagi masyarakat. Munculnya layanan musik di kehidupan masyarakat dapat memberikan kemudahan dan pilihan kepada masyarakat untuk digunakan. Kegiatan mendengarkan musik seolah-olah sudah menjadi rutinitas yang tidak boleh terlewatkan atau lupa dilakukan.

Metode *Clustering* (klaster) merupakan sebuah metode klasifikasi menggabungkan beberapa data menjadi satu bagian atau terkecil, dimana metode ini memiliki sifat dan kriteria tersendiri [4]. Dalam metode ini dapat mengkategorikan beberapa sifat dan kriteria sesuai klaster. Metode ini seringkali dipakai pada materi Data Mining, yang dilakukan dalam mengklasifikasikan beberapa data tertentu [5].

Penelitian data mining ini memakai perbandingan algoritma K-Means. Klasifikasi ini bertujuan untuk meninjau lagu pop terpopuler berdasarkan akurasi. Menurut penelitian sebelumnya, metode K-Means telah dipakai untuk proses klasifikasi pelanggan dengan analisis RFM, dimana penelitian ini memperoleh simpulan bahwa menciptakan kemampuan yang baik saat klasifikasi pelanggan [6]. Menurut klasifikasi sebelumnya menggunakan metode K-Means untuk penggolongan mahasiswa berdasarkan nilai UN dan IPK [7]. Penelitian ini menggunakan metode K-Means untuk mengclusterkan penyakit gagal ginjal (tingkat akurasinya 90%) [8].

2. METODOLOGI

2.1 Metode K-Means

Proses pelaksanaan penelitian ini memakai teknik klasifikasi dengan metode K-Means untuk klasifikasi top 50 Spotify Tahun 2010-2019 data kategorikal yang lebih stabil. Metode yang digunakan dalam penelitian ini adalah metode K-means. Tahapan dalam penelitian ini meliputi [9] :

a) Pengumpulan Data

Pengumpulan data pada penelitian ini berasal dari situs internet. Dataset berasal dari Kaggle.com, sementara studi literatur berasal dari beberapa jurnal dan artikel laporan penelitian dahulu mengenai klasifikasi menggunakan metode K-Means.

b) Implementasi/Penerapan

Implementasi/penerapan klasifikasi Global Top 50 menggunakan Google Colaboratory untuk mengetahui hasil yang disertai diagram.

c) Uji dan Analisis

Uji berasal dari hasil code dataset Global Top 50 pada Google Colaboratory, sementara analisis berasal dari analisa uji pengamatan

d) Simpulan

Simpulan berasal dari gagasan pembahasan klasifikasi Global Top 50

Pada Spotify telah memiliki beberapa fitur musik gratis dan memiliki pengklasifian penulis lagu yang berdasarkan genre musik. Untuk menjalankan dataset pada Google Colab akan diimplementasikan dengan kode python. Untuk data Global Top 50 ini, melakukan import data processing CSV file ke dalam Google Colaboratory.

3. HASIL DAN PEMBAHASAN

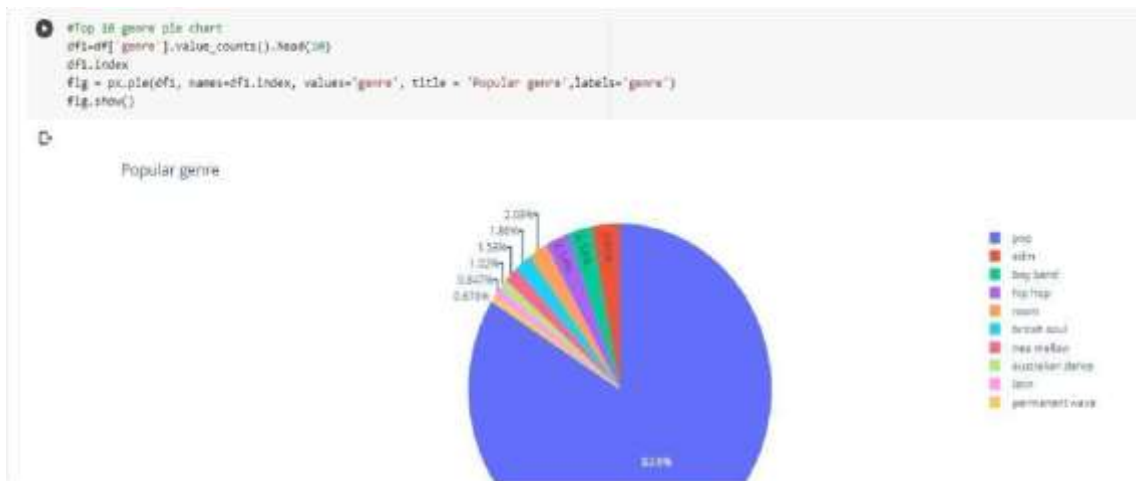
Pada penelitian ini dataset memakai dataset yang berasal dari Website Kaggle.com yang berisikan CSV sebagai kumpulan lagu dari aplikasi Spotify. Data dari file CSV berisikan 2001 data. Data dari Website Kaggle.com memiliki beberapa sifat dan kriteria, meliputi artis (penyanyi), lagu, durasi lagu, explicit (kecocokan lagu), popularitas, *danceability* (ritme lagu), tingkat energi lagu, *key* (kunci lagu), *loudness* (kekerasan lagu), mode, *speechiness* (tingkat speech lagu), *acousticness* (tingkat akustik lagu), *instrumentalness* (tingkat instrumen lagu), *liveness* (tingkat hidup lagu), valensi lagu (pembawaan lagu yg mempengaruhi suasana pendengar), tempo lagu, genre lagu [10].

```
#Reorganize genre
for i in yearless_df['top genre']:
    if 'pop' in i:
        yearless_df['top genre'] = yearless_df['top genre'].replace(i, 'pop')
    elif 'hip hop' in i:
        yearless_df['top genre'] = yearless_df['top genre'].replace(i, 'hip hop')
    elif 'edm' in i:
        yearless_df['top genre'] = yearless_df['top genre'].replace(i, 'edm')
    elif 'r&b' in i:
        yearless_df['top genre'] = yearless_df['top genre'].replace(i, 'pop')
    elif 'latin' in i:
        yearless_df['top genre'] = yearless_df['top genre'].replace(i, 'latin')
    elif 'room' in i:
        yearless_df['top genre'] = yearless_df['top genre'].replace(i, 'room')
    elif 'electro' in i:
        yearless_df['top genre'] = yearless_df['top genre'].replace(i, 'edm')
    elif 'house' in i:
        yearless_df['top genre'] = yearless_df['top genre'].replace(i, 'house')

for i in df10['top genre']:
    if 'pop' in i:
        df10['top genre'] = df10['top genre'].replace(i, 'pop')
    elif 'hip hop' in i:
        df10['top genre'] = df10['top genre'].replace(i, 'hip hop')
    elif 'edm' in i:
```

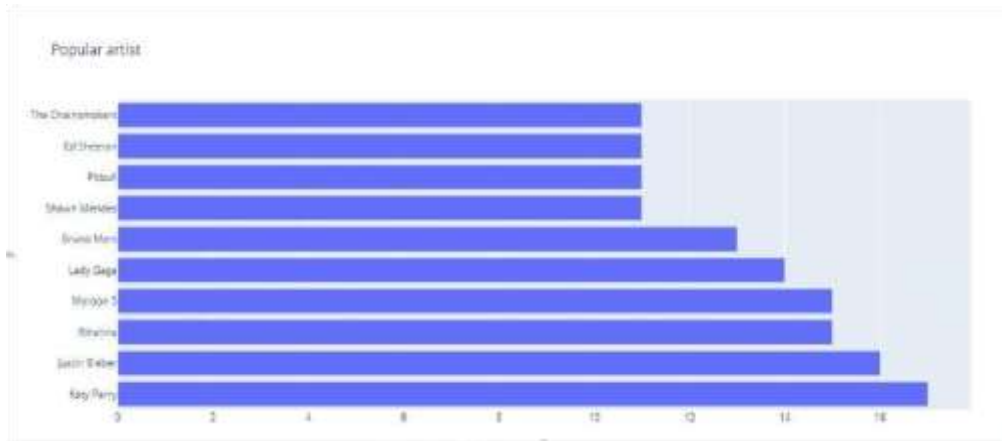
Gambar 2. Klasifikasi Pengaturan Genre Lagu

Selanjutnya, kriteria jenis data diintegrasikan berdasarkan kolom file CSV ke dalam satu file. Penelitian ini menggunakan Metode K-Means dengan tujuan klasifikasi data ke dalam beberapa kriteria/karakteristik sesuai dengan Global Top 50. Gambar 2, menunjukkan sebuah pengaturan top genre lagu pada Spotify, seperti genre pop. Genre lagu ini akan dilakukan analisis kluster berdasarkan sifat dan kriteria lagu sesuai dataset Spotify. Pada proses ini diwujudkan untuk meminimalisir variasi antar data yang ada di dalam suatu cluster dan mengembangkan variasi data yang ada di cluster lainnya. Setelah mengetahui kriteria jenis klasifikasi fitur Global Top 50 akan menampilkan Global Top 10 genre lagu pada Diagram Pie (Chart).



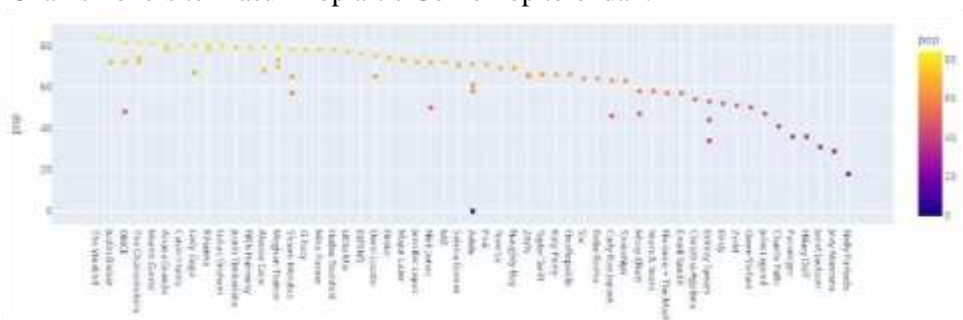
Gambar 3. Global Top 10 Genre Pie Chart

Pada gambar 3, Diagram Pie diperoleh penjelasan bahwa terdapat Top 10 populer genre lagu. Posisi pertama dimiliki oleh genre pop dengan presentase sejumlah 83,9%. Posisi kedua dimiliki oleh genre EDM dengan presentase sejumlah 3,05%. Posisi ketiga dimiliki oleh genre Boyband dan Hip Hop dengan presentase yang sama yaitu 2,54%. Sementara, Diagram Pie dengan presentase terendah diperoleh *permanent wave*, yaitu album musik dari Kanada dengan presentase 0,678%.



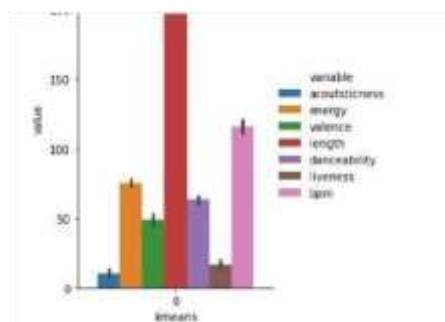
Gambar 4. Diagram Popular artist

Sesuai diagram batang diatas, diketahui bahwa index populer Top artis Genre Pop posisi atas diduduki oleh penyanyi Katy Perry dan Justin Bieber. Sementara penyanyi Rihanna dan Maroon 5 memiliki posisi kedudukan yang sama, serta posisi kedudukan paling rendah setara diduduki oleh Pitbull, Ed Sheeran, dan The Chainsmokers. Sehingga, Pitbull, Ed Sheeran, dan The Chainsmokers termasuk Top artis Genre Pop terendah.



Gambar 5. Hubungan Populer Artis Tahun 2016

Diketahui bahwa semakin tinggi tingkat Popular artist, maka titik akurat berwarna kuning. Jika semakin rendah tingkat Popular artist, maka titik akurat berwarna biru tua. Dari gambar 5, titik akurat berwarna kuning diduduki oleh artis Hailee S., Mike Posner, G.Eazy. Artis The Weekend menduduki posisi tertinggi sebagai Top Popular artist di tahun 2016. Sebaliknya, artis tidak terpopuler di tahun 2016 diduduki oleh Adele.



Gambar 6. Klaster Diagram Batang

Dari gambar 6 dapat diketahui, bahwa hasil klaster kriteria tertinggi genre Pop ditentukan berdasarkan valence lagu. Valence lagu pada Spotify berarti semakin tinggi nilai valensi atau titik akuratnya, maka lagu tersebut memiliki pesan bahagia (nada mayor). Valensi lagu dapat

mempengaruhi pendengar, pendengar menjadi lebih senang dan bahagia. Maka dari itu, klaster diagram batang Genre Pop dari Top Popular artis menunjukkan memiliki tingkat valensi yang tinggi. Posisi kedua diperoleh kriteria Beats Per Minute (BPM), bahwa Genre Pop lagu Popular artis dipengaruhi oleh ketukan/tempo lagu untuk memastikan cepat atau lambatnya lagu tersebut.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Dilaksanakannya penelitian ini untuk mengetahui hasil klasifikasi Top 50 Spotify Tahun 2010-2019 menggunakan metode K-Means Clustering. Perkembangan teknologi yang semakin pesat inilah yang menimbulkan pengaruh pada kehidupan masyarakat, seperti munculnya media layanan streaming musik yang memudahkan masyarakat dalam mendengarkan musik. Klasifikasi menggunakan dataset lagu pop terpopuler yang dianalisis menggunakan metode K-Means untuk mengetahui top genre. Dataset berasal dari Website Kaggle.com. Dalam penerapannya, penelitian ini dipengaruhi oleh *valence* (valensi lagu), genre lagu, tempo, Beats Per Minute/ketukan (BPM), dan sebagainya. Dalam menemukan cluster yang sesuai, penelitian diperoleh dari hasil analisis cluster pie chart, klaster diagram batang, dan titik koordinat hubungan valensi lagu. Penelitian dengan metode K-Means memudahkan pengguna dalam mengidentifikasi lagu artis terpopuler dan tidak terpopuler pada tahun tertentu.

4.2 Saran

Setelah melaksanakan penelitian memakai metode K-Means, tentunya penelitian ini masih memiliki beberapa kekurangan. Maka dari itu, dibutuhkan saran yang membangun untuk penelitian kedepannya. Saran yang diperlukan oleh penelitian, yakni :

- a) Hasil proses pelaksanaan penelitian ini mencapai klasifikasi Top genre Spotify dengan metode K-Means, sehingga pengguna dengan mudah memutar musik genre top terpopuler
- b) Eksplorasi data dan variabel lebih lanjut untuk mencapai variabel secara maksimal

5. DAFTAR RUJUKAN

- [1] S. Y. M. Netti and I. Irwansyah, "Spotify: Aplikasi Music Streaming untuk Generasi Milenial", Jurnal Komunikasi, vol. 10, no. 1, pp. 1–16, Juli, 2018.
- [2] Husna, A. "Segmentasi Pelanggan Menggunakan Model RFM dan Teori Rough Set Untuk Memahami Karakteristik Pelanggan (Studi Kasus: PT. ABBOTT INDONESIA, Malang)", Institut Teknologi Sepuluh Nopember, Jurusan Sistem Informasi, 2015.
- [3] Josi, Putra, G., Arindawati, A., "Nurkinan, 'Motif Penggunaan Aplikasi Musik Spotify pada Generasi-Z di SMA XYZ Bekasi", WARTA: Ikatan Sarjana Komunikasi Indonesia", vol.3, no.2, Desember, 2020
- [4] Harsemadi, I. G., & Sudarma, I. M, "Penggolongan Musik Terhadap Suasana Hati Menggunakan Metode K-Means", E-Proceedings KNS&I STIKOM Bali, pp. 49-54, Agustus, 2017
- [5] Novandya, A., "Penerapan Algoritma Klasifikasi Data Mining C4. Pada Dataset Cuaca Wilayah Bekasi", Konferensi Nasional Ilmu Sosial dan Teknologi, vol.1, no.1, pp. 368~372 Maret, 2017
- [6] Suprawoto, T., "Klasifikasi data mahasiswa menggunakan metode K-means untuk menunjang pemilihan strategi pemasaran", JIKO (Jurnal Informatika dan Komputer), vo.1, no.1, 2016.
- [7] Adiana, B. E., Soesanti, I., & Permanasari, A. E., 'Analisis segmentasi pelanggan menggunakan kombinasi RFM model dan teknik clustering', Jurnal Terapan Teknologi Informasi, vol.2, no.1, pp. 23-32, 2018
- [8] Sri, I. A., "Klasifikasi Penyakit Ginjal dengan Metode K-Means. JTT (Jurnal Teknologi Terpadu)", vol. 5, no.1, pp. 49-53, 2017.

- [9] Metisen, B. M., & Sari, H. L., “Analisis clustering menggunakan metode K-Means dalam pengelompokan penjualan produk pada Swalayan Fadhila”, Jurnal media infotama, 1 vol.1, no.2, 2015
- [10] Amirah, M. M. A., Widodo, A. W., & Dewi, C., “Pengelompokan Lagu Berdasarkan Emosi Menggunakan Algoritma Fuzzy C-Means”, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2548, 964X, 2017