

Pengelompokan Pengguna Spotify Berdasarkan Data Tipe Campuran Menggunakan Algoritma K-Prototype

Andini Diva Luthfiyah¹, Tessy Octavia Mukhti²

^{1,2}Universitas Negeri Padang

1andinidiva2004@gmail.com

ABSTRAK

Persaingan di layanan *streaming* musik menuntut pemahaman yang lebih baik mengenai perilaku pengguna dan kecenderungan *churn*. Penelitian ini bertujuan untuk melakukan segmentasi pengguna Spotify dan menganalisis pola *churn* berdasarkan karakteristik demografis serta perilaku penggunaan layanan dengan menggunakan metode K-Prototypes pada data campuran. Jumlah kluster yang optimal ditentukan menggunakan metode Elbow, sedangkan variabel *churn* digunakan untuk mengevaluasi hasil pengelompokan. Analisis menunjukkan terbentuknya tiga kluster pengguna dengan karakteristik yang berbeda. Kluster pertama mewakili pengguna premium yang lebih muda dengan intensitas penggunaan yang relatif tinggi, kluster kedua mewakili pengguna pelajar dengan proporsi *churn* tertinggi, dan kluster ketiga mewakili pengguna gratis dengan paparan iklan yang tinggi namun proporsi *churn* terendah. Temuan ini menunjukkan bahwa metode K-Prototypes efektif dalam mengelompokkan pengguna Spotify dan memberikan informasi yang berguna untuk memahami perilaku pengguna serta kecenderungan *churn*.

Kata Kunci: K-Prototypes; Kluster; Spotify; Data Campuran; Analisis *Churn*.

ABSTRACT

Competition in *streaming* music services requires a better understanding of user behavior and *churn* tendencies. This study aims to segment Spotify users and analyze *churn* patterns based on demographic characteristics and service usage behavior using the K-Prototypes method on mixed-type data. The optimal number of Cluster was determined using the Elbow method, while the *churn* variable was used to evaluate the clustering results. The analysis shows that three user cluster were formed with distinct characteristics. The first cluster represents younger premium users with relatively high usage intensity, the second cluster represents student users with the highest *churn* proportion, and the third cluster represents free users with high ad exposure but the lowest *churn* proportion. These findings indicate that the K-Prototypes method is effective in grouping Spotify users and provides useful information for understanding user behavior and *churn* tendencies.

Keywords: K-Prototypes; Clustering; Spotify; Mixed Data; *Churn* Analysis.

PENDAHULUAN

Maraknya layanan musik digital berbasis *streaming* telah mengubah cara orang mengakses dan menikmati konten audio. Kini, pengguna dapat mendengarkan musik melalui berbagai platform digital seperti Spotify, *YouTube Music*, dan *Apple Music* tanpa terikat oleh waktu atau tempat. Perubahan ini tidak hanya memperluas akses terhadap musik, tetapi juga membuka peluang pasar yang semakin signifikan bagi industri musik digital. Dalam konteks ini, platform *streaming* dituntut untuk memiliki pemahaman yang lebih mendalam mengenai perilaku pengguna guna mempertahankan daya saing mereka di tengah pasar yang terus berkembang (Paulus et al. 2023).

Spotify adalah salah satu platform *streaming* terkemuka di dunia, platform ini menyediakan sarana komunikasi yang konsisten dengan memungkinkan pengguna menerima informasi (Nuriska et al. 2023). Spotify menawarkan beragam fitur, mulai dari layanan gratis hingga layanan premium. Layanan premium memberikan beberapa keuntungan, seperti pemutaran musik tanpa iklan, kemampuan mengunduh lagu untuk diputar secara *offline*, dan melompati lagu tanpa batas. Keragaman fitur ini menunjukkan bahwa perilaku pengguna di platform *streaming* tidak seragam,

tetapi dipengaruhi oleh berbagai karakteristik demografis dan pola penggunaan layanan. Strategi ini dirancang untuk mendorong pengguna gratis beralih ke layanan berbayar (Spotify n.d.). Untuk membantu pengambilan keputusan yang lebih tepat dalam pembuatan rencana layanan dan pemasaran, penting untuk menganalisis karakteristik pengguna.

Seiring dengan terus meningkatnya jumlah pengguna layanan *streaming* musik, persaingan di antara berbagai platform pun semakin ketat. Setiap penyedia layanan berupaya menawarkan kualitas audio, keragaman konten, dan fitur tambahan yang mampu menarik serta mempertahankan pengguna. Dalam konteks ini, salah satu tantangan utama yang harus diatasi adalah *churn*, fenomena di mana pengguna membatalkan langganan mereka atau beralih ke platform lain (Royana et al. 2025). Di pasar global yang kompetitif ini, Spotify harus secara strategis mengatasi tantangan-tantangan ini dalam strategi pemasarannya, dengan memberikan nilai tambah yang membedakannya dari para pesaingnya (Salwa et al. 2024). *Churn* merupakan isu strategis karena dapat memengaruhi pertumbuhan pengguna dan keberlanjutan platform digital. Studi-studi sebelumnya menunjukkan bahwa prediksi *churn* merupakan isu penting dalam layanan *streaming* musik karena erat kaitannya dengan retensi pelanggan dan keberlanjutan pendapatan (Gaddam et al. 2022). Dalam praktiknya, keputusan pengguna untuk berhenti menggunakan layanan tidak dipengaruhi oleh satu faktor saja, melainkan oleh kombinasi berbagai karakteristik pengguna dan pola keterlibatan mereka dengan layanan tersebut.

Salah satu pendekatan yang dapat digunakan untuk memahami karakteristik pengguna adalah analisis kluster. Teknik analisis multivariat yang disebut analisis kluster bertujuan untuk mengelompokkan item berdasarkan atributnya. Analisis kluster bekerja dengan mengelompokkan objek penelitian sehingga objek yang memiliki kesamaan paling dekat satu sama lain ditempatkan dalam kluster yang sama. Kluster yang terbentuk dalam satu kumpulan memiliki karakteristik yang relatif serupa (homogen), sementara kluster-kluster tersebut berbeda satu sama lain (heterogen). Pengelompokan didasarkan pada variabel yang diamati (Rais, Annas, and Refaldy 2024). Melalui pengelompokan, pengguna dapat dikelompokkan ke dalam segmen yang relatif homogen, sehingga memungkinkan analisis pola perilaku dan tren *churn* yang lebih terarah. Dalam data pengguna Spotify, karakteristik yang diamati tidak hanya berupa variabel numerik, seperti usia, waktu mendengarkan, jumlah lagu per hari, tingkat melewati lagu, dan jumlah iklan per minggu, tetapi juga mencakup variabel kategorik, seperti jenis kelamin, negara, jenis langganan, jenis perangkat, dan mode *offline*. Hal ini menunjukkan bahwa data pengguna Spotify bersifat campuran, sehingga metode pengelompokan yang digunakan harus mampu mengakomodasi variabel numerik dan kategorik secara bersamaan.

Salah satu metode untuk menangani data bertipe campuran adalah algoritma K-Prototypes. Algoritma K-Prototypes merupakan gabungan dari metode pengelompokan K-Means dan K-Modes. Algoritma ini tidak terlalu rumit, mampu menangani kumpulan data besar, dan mempertahankan efisiensi algoritma K-Means dalam menangani kumpulan data besar sekaligus menghilangkan keterbatasan K-Means, yang hanya dapat digunakan untuk data numerik yang menjadi keunggulan utama dalam menggunakan algoritma K-Prototypes (Purba, Mustafid, and Kartikasari 2025). Dengan menggunakan metode ini, pengelompokan pengguna diharapkan menghasilkan kluster yang lebih akurat mencerminkan karakteristik pengguna Spotify.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk melakukan segmentasi pengguna Spotify berdasarkan karakteristik demografis dan perilaku penggunaan layanan dengan menggunakan metode K-Prototypes. Selain itu, penelitian ini bertujuan untuk mengidentifikasi karakteristik masing-masing kelompok yang dihasilkan serta menganalisis tren *churn* di dalam setiap kelompok pengguna. Hasil penelitian ini diharapkan dapat memberikan informasi yang bermanfaat sebagai landasan untuk mengembangkan strategi pemasaran yang lebih efektif, personalisasi layanan, dan upaya retensi pengguna.

METODE PENELITIAN

Sumber Data dan Variabel Penelitian

Dalam penelitian ini digunakan data sekunder dari situs website [Kaggle](#). Data tersebut terdiri dari data pengguna Spotify yang mencakup karakteristik demografis, pola penggunaan layanan, dan status *churn* pengguna. Variabel-variabel dalam penelitian ini dikategorikan menjadi variabel numerik, variabel kategorik, dan variabel target. Dalam analisis K-Prototypes, hanya variabel numerik dan kategorik yang digunakan dalam proses pembentukan kluster, sedangkan variabel target digunakan dalam evaluasi dan interpretasi hasil kluster. Tabel 1 mencantumkan variabel-variabel yang digunakan dalam penelitian ini.

Tabel 1. Variabel Penelitian

Variabel	Deskripsi	Skala
X_1	Usia Pengguna (tahun)	Numerik
X_2	Rata-rata waktu mendengarkan harian (menit)	Numerik
X_3	Jumlah lagu yang diputar per hari	Numerik
X_4	Persentase lagu yang dilewati	Numerik
X_5	Jumlah iklan yang didengar per minggu	Numerik
X_6	Jenis Kelamin (Laki-laki, Perempuan, Lainnya)	Kategorik
X_7	Negara asal pengguna	Kategorik
X_8	Jenis Langganan (Gratis, Premium, Keluarga, dan Pelajar)	Kategorik
X_9	Jenis Perangkat (Seluler, <i>Desktop</i> , dan Web)	Kategorik
X_{10}	Mode <i>Offline</i> (Ya/Tidak)	Kategorik
	<i>Churn</i>	Kategorik

Berdasarkan Tabel 1, variabel numerik dan kategorik digunakan secara bersamaan dalam proses pengelompokan menggunakan metode K-Prototypes, sedangkan variabel *churn* tidak dimasukkan dalam pembentukan kelompok. Variabel *churn* digunakan pada tahap analisis lanjutan untuk mengevaluasi kecenderungan pembatalan langganan di setiap kelompok yang terbentuk. Pendekatan ini memungkinkan interpretasi yang lebih informatif dan relevan terhadap kelompok-kelompok tersebut dalam konteks analisis perilaku dan retensi pengguna Spotify.

Tahap Analisis Data

Analisis data dilakukan menggunakan perangkat lunak *RStudio* dengan metode K-Prototypes. Proses analisis meliputi pembersihan data, standarisasi variabel numerik, penentuan jumlah kluster optimal menggunakan metode Elbow, serta pengelompokan menggunakan algoritma K-Prototypes. Variabel *churn* digunakan pada tahap evaluasi untuk menganalisis tren *churn* di setiap kluster yang terbentuk. Pembersihan data dilakukan sebelum analisis dengan mengubah variabel kategorik menjadi faktor agar dapat diproses oleh algoritma K-Prototypes. Langkah-langkah selanjutnya adalah sebagai berikut:

1. Standarisasi Data

Standarisasi dilakukan ketika atribut numerik memiliki satuan yang berbeda. Standarisasi dilakukan dengan mengubah data asli menjadi Z-score. Z-score adalah ukuran seberapa jauh data menyimpang dari rata-ratanya (μ), yang diukur berdasarkan simpangan baku (σ) (Karo Karo and Hendriyana 2022).

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

2. Kluster Optimal

Analisis kluster adalah teknik statistik multivariat yang dirancang untuk mengelompokkan objek yang diamati ke dalam beberapa kluster berdasarkan karakteristiknya (Hasmira, Wahidah Alwi, and Khalilah Nurfadilah 2023). Ada dua metode dalam analisis kluster: hierarki dan non-hierarki. Metode hierarki menghasilkan dekomposisi hierarki dari

kumpulan data berdasarkan kemiripan karakteristik objek. Sebaliknya, metode non-hierarki digunakan untuk mengelompokkan objek, di mana jumlah kluster yang akan dibentuk dapat ditentukan terlebih dahulu (Syafiyah et al. 2022).

Dengan menghitung persentase hasil perbandingan antara jumlah kluster yang akan membentuk tren pada waktu tertentu, Metode elbow dapat digunakan untuk mengidentifikasi jumlah kluster yang ideal (Rais et al. 2024). Selain itu, perbandingan jumlah kluster tambahan disajikan melalui perhitungan persentase yang dihasilkan. Grafik dapat digunakan sebagai sumber informasi untuk menampilkan hasil persentase yang bervariasi untuk setiap nilai kluster. Nilai kluster pertama adalah yang terbaik jika membentuk sudut pada grafik atau menunjukkan penurunan nilai terbesar jika dibandingkan dengan nilai kluster kedua. Tentukan SSE (*Sum of Square Error*) untuk setiap nilai kluster guna mendapatkan perbandingan ini. Hal ini karena nilai SSE menurun seiring dengan meningkatnya jumlah kluster (k). Rumus SSE k-prototype:

$$SSE = \sum_{k=1}^K \sum_{i=1}^n \|x_i - u_k\|^2 \quad (2)$$

Di mana K adalah jumlah kluster, x_i adalah nilai atribut ke- i , dan u_k adalah nilai atribut titik pusat kluster ke- k .

3. K-Prototype

Algoritma K-Prototypes adalah teknik pengelompokan yang didasarkan pada pemisahan. Teknik ini dikembangkan untuk menangani pengelompokan data dengan karakteristik campuran, khususnya fitur numerik dan kategorik, berdasarkan algoritma K-Means. Perbedaan utama antara K-Prototypes dan metode pengelompokan lainnya terletak pada cara mengukur kemiripan atau ketidaksamaan antara suatu objek dan pusat kluster (prototype). Untuk atribut numerik, perhitungan dilakukan menggunakan jarak Euclidean kuadrat, sedangkan untuk atribut kategorik, digunakan ukuran ketidaksesuaian kategori, mirip dengan metode K-Modes. Dengan demikian, K-Prototypes memungkinkan proses pengelompokan dilakukan secara bersamaan pada dua jenis data yang berbeda (Fadilah and Wijayanto 2023). Metrik jarak dalam algoritma K-Prototypes dapat dinyatakan sebagai berikut:

$$d_{ij} = \sum_{k=1}^p (x_{ik} - x_{jk})^2 + \gamma \sum_{k=p+1}^{p+m} \delta (x_{ik} - x_{jk}) \quad (3)$$

Di mana d_{ij} adalah jarak antara objek i dan j , $\sum_{k=1}^p (x_{ik} - x_{jk})^2$ merupakan ukuran jarak untuk data numerik, γ adalah kompensator jarak, dan $\sum_{k=p+1}^{p+m} \delta (x_{ik} - x_{jk})$ merupakan ukuran jarak untuk data kategorik. Langkah-langkah algoritma K-Prototypes (Refaldy, Annas, and Rais 2023) dimulai dengan:

- a. Menentukan jumlah kelompok (k) yang akan dibentuk,
- b. Tetapkan prototype awal sebagai pusat awal masing-masing kelompok,
- c. Hitung jarak setiap pengamatan dalam kumpulan data dari semua prototype awal, lalu masukkan setiap objek ke dalam kelompok dengan jarak terpendek,
- d. Setelah proses alokasi awal selesai, prototype baru dihitung ulang berdasarkan anggota kluster yang telah terbentuk,
- e. Selanjutnya, semua objek dialokasikan ulang berdasarkan jaraknya dari prototype baru. Proses ini diulangi hingga tidak ada lagi perubahan dalam keanggotaan kluster.

4. Hasil Klusterisasi

Hasil pengelompokan dari analisis ini menggambarkan karakteristik masing-masing kelompok dengan menghitung rata-rata setiap variabel di dalam kluster tersebut. Interpretasikan perbedaan pola di antara kluster tersebut.

HASIL DAN PEMBAHASAN

Standarisasi Data

Analisis awal menunjukkan bahwa variabel numerik dalam penelitian ini memiliki satuan pengukuran yang berbeda-beda; oleh karena itu, dilakukan proses standarisasi dengan menggunakan metode Z-score. Standarisasi diterapkan pada variabel Usia, Waktu Mendengarkan, Jumlah Lagu per Hari, Tingkat Melewati Lagu, dan Iklan per Minggu untuk memastikan bahwa setiap variabel berada pada skala yang setara dan tidak mendominasi perhitungan jarak selama proses pengelompokan. Hasil standarisasi data dapat dilihat pada Tabel 2.

Tabel 2. Data Spotify Setelah Standarisasi

No	X_1	X_2	X_3	X_4	X_5
1	1.28	-1.52	-0.95	-0.58	1.77
2	-0.37	-0.16	0.42	0.23	-0.51
⋮	⋮	⋮	⋮	⋮	⋮
7999	-0.29	1.08	1.54	-0.17	-0.51
8000	0.58	0.67	0.63	0.92	-0.07

Statistik Deskriptif

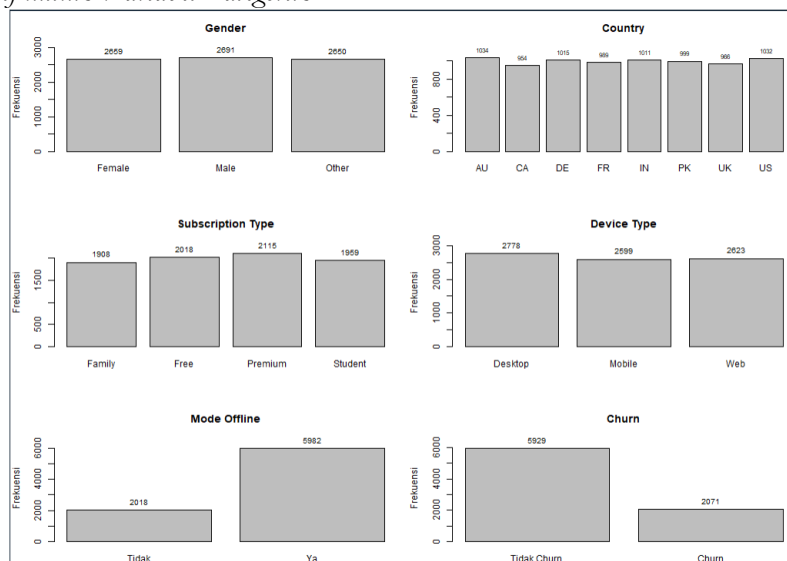
Statistik Deskriptif untuk Variabel Numerik

Tabel 3. Statistik Deskriptif Numerik

Variabel	Mean	Min	Max	Std.Dev
X_1	37.66	16.00	59.00	12.74
X_2	154.10	10.00	299.00	154.07
X_3	50.13	1.00	99.00	28.45
X_4	0.30	0.00	0.60	0.30
X_5	6.94	0.00	49.00	13.62

Statistik deskriptif untuk variabel numerik disajikan pada Tabel 3. Berdasarkan tabel tersebut, variabel Usia memiliki rata-rata 37,66 dengan rentang 16,00–59,00. Variabel Waktu Mendengarkan memiliki rata-rata 154,10 menit per hari dengan rentang 10,00–299,00. Variabel Lagu per Hari memiliki rata-rata 50,13 dengan rentang 1,00–99,00. Variabel Tingkat Lewati memiliki rata-rata 0,30 dengan rentang 0,00–0,60, sedangkan Iklan per Minggu memiliki rata-rata 6,94 dengan rentang 0,00–49,00. Hasil ini menunjukkan bahwa data pengguna Spotify menunjukkan keragaman yang cukup besar, terutama pada variabel waktu mendengarkan dan paparan iklan.

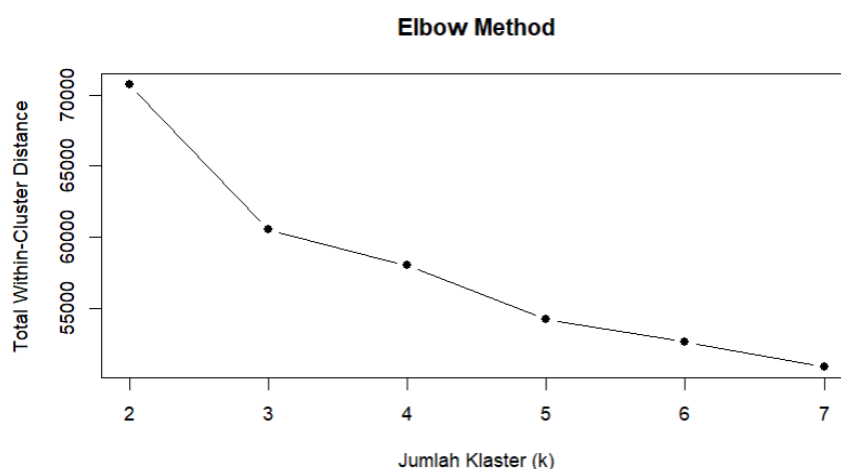
Statistik Deskriptif untuk Variabel Kategorik



Gambar 1. Distribusi Kategori Pengguna Spotify

Distribusi variabel kategorik ditunjukkan pada Gambar 1. Berdasarkan gambar tersebut, komposisi jenis kelamin relatif seimbang antara Perempuan (2.659), Laki-laki (2.691), dan Lainnya (2.650). Variabel Negara juga menunjukkan distribusi yang cukup merata di delapan kategori, dengan frekuensi tertinggi di AU (1.034) dan AS (1.032), sedangkan frekuensi terendah ada di CA (954). Untuk variabel Tipe Langganan, kategori Premium memiliki frekuensi tertinggi (2.115), diikuti oleh Gratis (2.018), Mahasiswa (1.959), dan Keluarga (1.908). Untuk variabel Tipe Perangkat, mayoritas pengguna menggunakan *Desktop* (2.778), diikuti oleh Web (2.623) dan *Mobile* (2.599). Sementara itu, untuk variabel Mode *Offline*, mayoritas pengguna termasuk dalam kategori Ya (5.982), sedangkan kategori Tidak berjumlah 2.018. Hasil ini menunjukkan bahwa data penelitian menunjukkan variasi kategori yang memadai dan mendukung penggunaan metode K-Prototypes untuk menganalisis data numerik dan kategorik campuran.

Jumlah Kluster Optimal



Gambar 2. Grafik Metode Elbow

Berdasarkan Gambar 2, terlihat penurunan nilai SSE ketika jumlah kluster adalah k=3, yang menunjukkan jumlah kluster yang optimal. Untuk k>3, tidak ada peningkatan lebih lanjut dalam kualitas kluster; oleh karena itu, k=3 dipilih sebagai jumlah kluster yang optimal. Pemilihan tiga kluster dianggap mampu merepresentasikan struktur data pengguna Spotify.

Hasil Klasterisasi

Tabel 4. Karakteristik Kluster Berdasarkan Variabel Numerik

Kluster	X_1	X_2	X_3	X_4	X_5	Jumlah Pengguna	Proporsi <i>Churn</i>
1	29.39	164.27	48.71	0.304	0.33	3174	0.26
2	46.62	143.09	52.14	0.296	0.18	2994	0.27
3	37.34	154.32	49.29	0.300	29.46	1832	0.25

Tabel 5. Karakteristik Kluster Berdasarkan Variabel Dominan

Kluster	X_6	X_7	X_8	X_9	X_{10}	Jumlah Pengguna	Proporsi <i>Churn</i>
1	29.39	164.27	48.71	0.304	0.33	3174	0.26
2	46.62	143.09	52.14	0.296	0.18	2994	0.27
3	37.34	154.32	49.29	0.300	29.46	1832	0.25

Tabel 4 dan 5 menunjukkan bahwa proses pengelompokan berhasil membagi pengguna Spotify menjadi tiga kelompok berdasarkan karakteristik masing-masing. Kluster 1 memiliki usia rata-rata 29,39 tahun, waktu mendengarkan 164,27 menit, 48,71 lagu per hari, tingkat melewati lagu sebesar 0,304, dan 0,33 iklan per minggu. Kluster ini didominasi oleh pengguna laki-laki dari DE, dengan langganan Premium, menggunakan perangkat web, dan mayoritas menggunakan

mode *offline*. Tingkat *churn* untuk Klaster ini adalah 0,26. Dengan karakteristik tersebut, Klaster 1 dapat diartikan sebagai kelompok pengguna premium yang relatif lebih muda, aktif, dan cukup stabil.

Klaster 2 memiliki usia rata-rata 46,62 tahun, durasi mendengarkan 143,09 menit, 52,14 lagu per hari, tingkat melewati lagu sebesar 0,296, dan 0,18 iklan per minggu. Klaster ini didominasi oleh pengguna perempuan dari PK, dengan jenis langganan pelajar, yang menggunakan perangkat seluler, dan mayoritas menggunakan mode *offline*. Tingkat *churn* di klaster ini adalah 0,27, yang merupakan nilai tertinggi di antara ketiga klaster. Dengan demikian, Klaster 2 mewakili kelompok pengguna pelajar yang lebih tua dengan kecenderungan *churn* yang sedikit lebih tinggi dibandingkan dengan klaster lainnya.

Klaster 3 memiliki usia rata-rata 37,34 tahun, waktu mendengarkan 154,32 menit, 49,29 lagu per hari, tingkat melewati lagu sebesar 0,300, dan 29,46 iklan per minggu. Klaster ini didominasi oleh pengguna perempuan dari UK, yang berlangganan paket Gratis, menggunakan perangkat seluler, dan tidak menggunakan mode *offline*. Tingkat *churn* untuk klaster ini adalah 0,25, yang merupakan yang terendah di antara ketiga klaster. Hasil ini menunjukkan bahwa Klaster 3 mewakili kelompok pengguna layanan gratis dengan paparan iklan yang tinggi, namun tidak menunjukkan tingkat *churn* tertinggi.

Pembahasan

Hasil penelitian menunjukkan bahwa metode K-Prototypes mampu mengelompokkan pengguna Spotify ke dalam tiga klaster dengan karakteristik yang berbeda. Perbedaan utama antar klaster terlihat pada jenis langganan, jenis perangkat, mode *offline*, usia, dan jumlah iklan yang didengarkan per minggu. Temuan ini menegaskan bahwa segmentasi pengguna berdasarkan data campuran dapat mengungkap pola penggunaan layanan yang lebih spesifik, karena baik variabel numerik maupun kategorik berkontribusi pada pembentukan klaster. Dalam hal ini, studi ini sejalan dengan konsep umum analisis klaster, yang mengelompokkan objek dengan karakteristik serupa ke dalam klaster yang relatif homogen. Temuan ini juga mendukung studi sebelumnya yang menunjukkan bahwa K-Prototypes cocok untuk data campuran karena menggabungkan keunggulan K-Means dan K-Modes dalam menangani atribut numerik dan kategorik secara bersamaan (Purba et al. 2025). Dalam studi ini, keunggulan tersebut tercermin dalam pembentukan klaster yang tidak hanya berbeda dalam hal perilaku numerik, seperti usia dan intensitas mendengarkan, tetapi juga dalam hal karakteristik kategorik, seperti jenis langganan, jenis perangkat, dan mode *offline*. Dengan demikian, penggunaan K-Prototypes memberikan segmentasi pengguna Spotify yang lebih komprehensif dibandingkan pendekatan yang hanya berfokus pada satu jenis variabel.

Dari sudut pandang praktis, Klaster 1 mewakili pengguna premium yang relatif lebih muda, aktif, dan cukup stabil; Klaster 2 mewakili pengguna pelajar dengan tingkat *churn* tertinggi; sedangkan Klaster 3 mewakili pengguna gratis yang sering terpapar iklan namun memiliki tingkat *churn* terendah. Hasil ini menunjukkan bahwa *churn* tidak ditentukan oleh satu faktor saja. Temuan ini sejalan dengan penelitian terbaru yang menunjukkan bahwa *churn* pelanggan di platform *streaming* musik merupakan masalah yang kompleks dan tidak dapat dijelaskan sepenuhnya hanya dengan informasi pengguna statis (Cheng and He 2025). Meskipun Klaster 3 mengalami paparan iklan tertinggi, proporsi *churn*-nya lebih rendah daripada Klaster 2. Hal ini menunjukkan bahwa hubungan antara fitur layanan dan *churn* lebih kompleks dan bergantung pada kombinasi karakteristik pengguna dan pola penggunaan layanan, yang sejalan dengan studi sebelumnya yang menyatakan bahwa *churn* dalam layanan *streaming* dipengaruhi oleh berbagai faktor, bukan hanya oleh satu variabel saja (Royana et al. 2025).

PENUTUP

Penelitian ini menggunakan metode K-Prototypes untuk mengelompokkan pengguna Spotify berdasarkan data campuran yang terdiri dari variabel numerik dan kategorik. Hasil analisis menunjukkan bahwa jumlah kluster optimal yang terbentuk adalah tiga. Setiap kluster memiliki karakteristik yang berbeda, baik dari segi demografi maupun pola penggunaan layanan. Kluster 1 didominasi oleh pengguna Premium yang relatif lebih aktif dan stabil, Kluster 2 didominasi oleh pengguna Mahasiswa dengan tingkat *churn* tertinggi, sedangkan Kluster 3 didominasi oleh pengguna Gratis dengan tingkat *churn* terendah. Hasil ini menunjukkan bahwa metode K-Prototypes mampu memberikan segmentasi pengguna yang jelas dalam data campuran dan membantu dalam memahami tren *churn* di dalam setiap kelompok pengguna. Secara umum, hasil penelitian ini dapat menjadi dasar untuk mendukung strategi retensi pengguna, personalisasi layanan, dan penentuan target promosi yang lebih selaras dengan karakteristik masing-masing segmen pengguna Spotify. Penelitian lebih lanjut dapat membandingkan hasil K-Prototypes dengan metode pengelompokan data campuran lainnya untuk mencapai segmentasi pengguna yang lebih komprehensif.

REFERENSI

- Cheng, Haiyan, and Jie He. 2025. "Advanced Customer *Churn* Prediction for a Music *Streaming* Digital Marketing Service Using Attention Graph-Based Deep Learning Approach." *Scientific Reports* 15(1). doi: 10.1038/s41598-025-28357-z.
- Fadilah, Zahra Rizky, and Arie Wahyu Wijayanto. 2023. "Perbandingan Metode Klusterisasi Data Bertipe Campuran : One-Hot- Encoding , Gower Distance , Dan K-Prototype Berdasarkan Akurasi (Studi Kasus : Chronic Kidney Disease Dataset)." 7(1).
- Gaddam, Lahari, Sree Lakshmi, Hiranmayee Kadali, and Shahrooz Abghari. 2022. "Comparison of Machine Learning algorithms on Predicting *Churn* Within Music *Streaming* Service." (June).
- Hasmira, Wahidah Alwi, and Khalilah Nurfadilah. 2023. "Penentuan Kluster Hirarki Optimum Dalam Pengelompokan Kabupaten/Kota Di Provinsi Sulawesi Selatan Dengan Indikator Kemiskinan." *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)* 11(1):93–102. doi: 10.24252/msa.v11i1.33910.
- Karo Karo, Ichwanul Muslim, and Hendriyana. 2022. "KLASIFIKASI PENDERITA DIABETES MENGGUNAKAN ALGORITMA MACHINE LEARNING DAN Z-SCORE." *Jurnal Teknologi Terpadu* 8(2):94–99.
- Nuriska, Dwi, Bambang Irawan, Agus Bahtiar, Arif Rinaldi Dikananda, Teknik Informatika, Kota Cirebon, Suasana Hati, Yudhistira Arie Wijaya, and Cangehgar Cyber. 2023. "KLAUSTERISASI DATA LAGU TERPOPULER SPOTIFY 2023 BERDASARKAN SUASANA HATI MENGGUNAKAN ALGORITMA K-MEANS." 7(6):3843–50. doi: 10.36040/jati.v7i6.8232.
- Paulus, Yohanes, Hanny Wadhi, Fitri Ciptosari, and Reynaldo Angga Siagian. 2023. "Strategi Pemasaran Digital Musik Daerah Flores Melalui Platform Musik Digital Spotify." 06(3):309–16. doi: 10.32493/jpkpk.v6i3.27791.
- Purba, Ariani Fitri, Mustafid Mustafid, and Puspita Kartikasari. 2025. "PENERAPAN ALGORITMA K-PROTOTYPE UNTUK PENGELOMPOKAN DESA DI KABUPATEN BEKASI BERDASARKAN INFRASTRUKTUR DIGITAL." *Jurnal Gaussian* 13(2):479–89. doi: 10.14710/j.gauss.13.2.479-489.
- Rais, Zulkifli, Suwardi Annas, and Muhammad Refaldy. 2024. "Algoritma K-Prototype Dalam Pengelompokan Kabupaten/Kota Di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat Tahun 2020." *VARLANSI: Journal of Statistics and Its Application on Teaching and Research* 6(3):144–51. doi: 10.35580/variansium20.
- Refaldy, Muhammad, Suwardi Annas, and Zulkifli Rais. 2023. "K-Prototype Algorithm in Grouping Regency/City in South Sulawesi Province Based on 2020 People's Welfare." *ARRUS Journal of Mathematics and Applied Science* 3(1):11–19. doi:

10.35877/mathscience1763.

- Royana, Akge Ninov, Yisti Vita Via, Chrystia Aji Putra, Jalan Raya, Rungkut Madya, and Apple Music. 2025. "Evaluasi Kinerja LightGBM Dan CatBoost Untuk Prediksi *Churn* Berdasarkan Dataset Pelanggan Layanan *Streaming* Musik." 9(4):6977–82.
- Salwa, Azahra, Indri Sri Asdini, Sylvia Amanda, Aurora Elise Putriku, Susanna Saragih, Alamat Jl, William Iskandar, Kenangan Baru, Kec Percut, Sei Tuan, Kabupaten Deli Serdang, and Sumatera Utara. 2024. "Strategi Pemasaran Aplikasi Spotify Dalam Global Spotify ' s Global Marketing Strategy." 2(4).
- Spotify. n.d. "Apa Itu Spotify." *Spotify*. Retrieved December 17, 2025 (<https://support.spotify.com/id-id/article/what-is-spotify/>).
- Syafiyah, Umi, Iqbal Asrafi, Brian Wicaksono, Dita Putri Puspitasari, and Mayland Sirait. 2022. "Analisis Perbandingan Metode Klaster Data Indikator Ketenagakerjaan Di Jabar Tahun 2020." *Seminar Nasional Official Statistics* 2020:803–12.