

Perancangan Website Analisis Segmentasi Perilaku dan Minat Berlangganan *Spotify Premium* menggunakan Metode *K-Medoids* dan Regresi Polinomial

Charles ^{1*}, Edi ², Kelvin Leonardi Kosasih ³

¹⁻³ Teknik Informatika, Sekolah Tinggi Manajemen Informatika dan Komputer TIME, Indonesia

*Penulis Korespondensi: charless2432@gmail.com

Abstract. *This study aims to analyze customer behavior segmentation and subscription interest in Spotify Premium using a data mining approach. The problem addressed is the limited understanding of user behavior influencing subscription decisions, resulting in less optimal strategies for increasing premium subscribers. The methods used include the K-Medoids algorithm for customer clustering and polynomial regression to analyze relationships between behavioral variables. The dataset used is the Spotify User Behaviour Dataset, which contains various user behavior attributes. The research stages include data pre-processing, determining the optimal number of clusters using the Elbow method, clustering, and regression analysis. The results show that the optimal number of clusters is $k=9$, allowing customers to be grouped into nine segments based on similar behavior patterns, with a Silhouette Score of 0.1659 indicating weak to moderate cluster quality. Furthermore, the polynomial regression results show a Mean Squared Error (MSE) value of 1.0694 and a coefficient of determination (R^2) of -0.1109, indicating that the model has not optimally explained the data variability. However, factors such as usage intensity and content preferences still influence user's interest in subscribing. This study provides insights into customer characteristics and support more targeted strategies to increase premium subscriptions.*

Keywords: *Customer Segmentation; K-Medoids; Polynomial Regression; Silhouette Score; Spotify Premium.*

Abstrak. Penelitian ini bertujuan untuk menganalisis segmentasi perilaku serta minat pelanggan dalam berlangganan Spotify Premium menggunakan pendekatan data mining. Permasalahan yang diangkat adalah belum optimalnya pemahaman terhadap perilaku pengguna yang memengaruhi keputusan berlangganan sehingga strategi peningkatan pelanggan premium belum berjalan secara maksimal. Metode yang digunakan meliputi algoritma K-Medoids untuk pengelompokan pelanggan dan regresi polinomial untuk menganalisis hubungan antar variabel perilaku. Dataset yang digunakan berasal dari *Spotify User Behaviour Dataset* dengan berbagai atribut perilaku pengguna. Tahapan penelitian mencakup *pre-processing* data, penentuan jumlah cluster optimal menggunakan metode Elbow, proses clustering, serta analisis regresi. Hasil penelitian menunjukkan bahwa jumlah cluster optimal yang diperoleh adalah $k=9$, sehingga pelanggan dapat dikelompokkan ke dalam sembilan segmen berdasarkan pola perilaku yang serupa, dengan nilai *Silhouette Score* sebesar 0,1659 yang menunjukkan kualitas cluster berada pada kategori lemah hingga sedang. Selain itu, hasil regresi polinomial menunjukkan nilai *Mean Squared Error* (MSE) sebesar 1,0694 dan nilai koefisien determinasi (R^2) sebesar -0,1109, yang mengindikasikan bahwa model belum mampu menjelaskan variasi data secara optimal. Meskipun demikian, faktor seperti intensitas penggunaan dan preferensi konten tetap memiliki pengaruh terhadap minat pelanggan dalam berlangganan. Penelitian ini diharapkan dapat membantu memahami karakteristik pelanggan serta mendukung strategi peningkatan pelanggan premium secara lebih tepat sasaran

Kata Kunci: K-Medoids; Regresi Polinomial; Segmentasi Pelanggan; *Silhouette Score*; Spotify Premium.

1. PENDAHULUAN

Perkembangan teknologi digital telah mengubah secara signifikan cara masyarakat dalam mengakses dan mengonsumsi musik, khususnya melalui platform streaming musik (Ruddin et al., 2022 ; Nita & others, 2022). Berbagai layanan seperti Spotify, Apple Music, dan YouTube Music bersaing dalam industri ini, dengan Spotify menempati posisi dominan sebagai platform terpopuler. Hingga akhir tahun 2024, Spotify tercatat memiliki 263 juta pengguna Premium

secara global dan menguasai pangsa pasar sebesar 31,7% pada kuartal III 2023 (Mei, 2024 ; Matulessy et al., 2025). Layanan ini menawarkan dua jenis akses, yaitu versi gratis dengan iklan dan versi premium yang menyediakan pengalaman tanpa gangguan, kualitas audio yang lebih baik, serta fitur unduhan untuk pemutaran offline.

Meskipun jumlah pelanggan Spotify Premium terus meningkat dan melampaui kompetitornya, masih terdapat berbagai tantangan dalam mempertahankan pelanggan. Beberapa permasalahan yang muncul antara lain tingginya tingkat churn setelah masa promosi, rendahnya loyalitas pengguna terhadap fitur premium, serta kecenderungan pengguna untuk kembali ke layanan gratis. Selain itu, praktik berbagi akun dan penyalahgunaan paket keluarga juga menjadi kendala dalam menjaga keberlanjutan bisnis. Hal ini menunjukkan bahwa pemahaman terhadap perilaku dan minat pelanggan terhadap layanan premium masih belum optimal (Sangging et al., 2025 ; Prayitno & others, 2026).

Minat berlangganan pelanggan dipengaruhi oleh berbagai faktor, baik internal maupun eksternal. Faktor internal meliputi preferensi musik, intensitas penggunaan, dan persepsi nilai terhadap layanan, sedangkan faktor eksternal mencakup demografi serta persaingan antar platform streaming. Oleh karena itu, analisis perilaku pelanggan menjadi aspek penting dalam merancang strategi pemasaran yang efektif guna meningkatkan konversi pengguna serta mempertahankan loyalitas pelanggan (Salwa et al., 2024 ; Kaligis & others, 2022).

Dalam konteks tersebut, penelitian ini memanfaatkan pendekatan data mining dengan menggunakan algoritma K-Medoids untuk melakukan segmentasi pelanggan berdasarkan pola perilaku berlangganan. Metode ini efektif dalam mengelompokkan data tanpa label sehingga mampu mengidentifikasi karakteristik pelanggan secara lebih akurat (A. M. P. Hidayat & Saputra, 2026; H. Hidayat et al., 2026). Selain itu, *Regresi Polinomial* digunakan untuk menganalisis hubungan antara berbagai faktor yang memengaruhi minat berlangganan pelanggan terhadap Spotify Premium. Kombinasi kedua metode ini diharapkan dapat memberikan pemahaman yang lebih komprehensif mengenai segmentasi pasar dan faktor-faktor kunci yang memengaruhi keputusan berlangganan (H. Hidayat et al., 2026; Kaligis & others, 2022).

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk merancang sebuah website analisis yang mampu mengintegrasikan hasil segmentasi pelanggan dan analisis minat berlangganan, sehingga dapat mendukung pengambilan keputusan strategis dalam meningkatkan retensi dan konversi pengguna layanan Spotify Premium.

2. TINJAUAN PUSTAKA

Layanan Digital Streaming

Layanan digital merupakan bentuk layanan berbasis teknologi yang memungkinkan pengguna mengakses produk atau jasa melalui platform digital seperti aplikasi dan situs web tanpa interaksi langsung dengan penyedia layanan. Contohnya meliputi e-commerce, layanan keuangan digital, kesehatan digital, hiburan, serta pendidikan berbasis online (Wahyuddin et al., 2023). Salah satu bentuk layanan digital yang berkembang pesat adalah streaming, yaitu proses pengiriman data secara kontinu melalui jaringan internet sehingga dapat langsung diakses oleh pengguna tanpa perlu diunduh terlebih dahulu (Yuliani & Kurniadi, 2023).

Layanan digital streaming memungkinkan distribusi konten multimedia seperti musik dan video secara real-time (*live streaming*) maupun sesuai permintaan (*on-demand streaming*) (Yuliani & Kurniadi, 2023). Dalam industri musik, layanan ini telah mengubah pola konsumsi masyarakat dengan menyediakan akses instan terhadap jutaan lagu melalui platform seperti Spotify, Apple Music, dan YouTube Music. Keunggulan layanan ini terletak pada kemudahan akses, personalisasi berbasis algoritma, serta fleksibilitas penggunaan di berbagai perangkat. Faktor-faktor seperti kualitas layanan, biaya, kenyamanan, dan personalisasi menjadi penentu utama dalam adopsi layanan streaming oleh pengguna (Panjaitan & Andarini, 2024).

Segmentasi Perilaku

Segmentasi perilaku adalah proses pengelompokan konsumen berdasarkan pola penggunaan, kebiasaan, serta proses pengambilan keputusan terhadap suatu produk atau layanan (Nugraha et al., 2025). Berbeda dengan segmentasi demografis, pendekatan ini lebih menitikberatkan pada interaksi konsumen dengan layanan, sehingga memungkinkan penyusunan strategi pemasaran yang lebih tepat sasaran.

Dalam konteks layanan streaming musik, segmentasi perilaku dapat dianalisis melalui berbagai atribut seperti usia, jenis kelamin, lama penggunaan, perangkat yang digunakan, jenis langganan, preferensi konten, hingga kebiasaan mendengarkan musik dan podcast (Gomathy et al., 2022). Analisis atribut-atribut tersebut memberikan gambaran komprehensif mengenai karakteristik pengguna dan membantu dalam memahami kebutuhan serta preferensi pelanggan secara lebih mendalam.

Data Mining

Data mining adalah proses penggalian informasi dan pola dari kumpulan data berukuran besar untuk menghasilkan pengetahuan yang bermanfaat. Proses ini mencakup pengumpulan, ekstraksi, analisis, dan interpretasi data, serta merupakan bagian dari *Knowledge Discovery in Database* (KDD) (Amna et al., 2023).

Data mining memiliki beberapa fungsi utama, yaitu deskripsi, estimasi, prediksi, klasifikasi, clustering, dan asosiasi (Amna et al., 2023). Dalam penelitian ini, teknik yang digunakan adalah clustering untuk segmentasi pelanggan serta regresi untuk menganalisis hubungan antar variabel.

Artificial Intelligence (AI)

Artificial Intelligence (AI) merupakan kemampuan sistem komputer untuk meniru kecerdasan manusia dalam memahami informasi, menganalisis data, dan mengambil keputusan. AI bertujuan untuk mengembangkan sistem yang mampu menyelesaikan tugas kompleks secara otomatis. Beberapa tujuan utama AI meliputi pengolahan bahasa alami, pembelajaran mesin, pengenalan pola, serta penalaran dan pengambilan keputusan, yang semuanya berperan dalam pengolahan data secara cerdas (Muttaqin et al., 2023).

Machine Learning

Machine Learning (ML) adalah cabang dari AI yang berfokus pada pengembangan algoritma yang mampu belajar dari data dan membuat prediksi tanpa pemrograman eksplisit. Kinerja model akan meningkat seiring bertambahnya data yang digunakan. ML diklasifikasikan menjadi *supervised learning*, *unsupervised learning*, *semi-supervised learning*, dan *reinforcement learning* (Damayanti et al., 2024). Dalam penelitian ini, pendekatan yang digunakan adalah *unsupervised learning* melalui teknik clustering untuk menemukan pola tersembunyi dalam data (Permana et al., 2022).

Clustering

Clustering merupakan metode dalam *unsupervised learning* yang bertujuan mengelompokkan data berdasarkan tingkat kemiripan karakteristik. Data dalam satu cluster memiliki kesamaan tinggi, sedangkan antar cluster memiliki perbedaan yang signifikan (Permana et al., 2022 ; Simorangkir et al., 2024).

Tujuan clustering meliputi identifikasi pola tersembunyi, penyederhanaan data, serta mendukung pengambilan keputusan. Metode clustering terdiri dari beberapa jenis seperti *partitioning*, *hierarchical*, *density-based*, dan *grid-based* (Saputro & Sucihermayanti, 2021a).

Dalam penelitian ini, clustering digunakan untuk mengelompokkan pelanggan Spotify berdasarkan perilaku berlangganan sehingga diperoleh segmentasi yang lebih homogen .

Algoritma K-Medoids

K-Medoids adalah algoritma clustering yang mengelompokkan data ke dalam K cluster berdasarkan kedekatan jarak antar data, dengan medoid sebagai pusat cluster (Saputro & Sucihermayanti, 2021b). Prosesnya meliputi pemilihan medoid awal, perhitungan jarak menggunakan *Manhattan distance*, pembentukan cluster, serta iterasi hingga diperoleh hasil

optimal (Faran & Aldisa, 2024a). Algoritma ini memiliki keunggulan dalam ketahanan terhadap outlier dibandingkan metode seperti K-Means, sehingga lebih stabil dalam mengolah data dengan noise (Faran & Aldisa, 2024b).

Regresi Polinomial

Regresi Polinomial merupakan metode regresi yang digunakan untuk memodelkan hubungan non-linear antara variabel independen dan dependen dalam bentuk persamaan polinomial . Metode ini mampu merepresentasikan hubungan yang lebih kompleks dibandingkan regresi linear. Model regresi dinyatakan dalam bentuk persamaan dengan variabel berpangkat hingga orde tertentu, di mana koefisien regresi dihitung untuk meminimalkan kesalahan prediksi dan meningkatkan akurasi model (Laksono et al., 2023).

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan data mining untuk menganalisis segmentasi perilaku pengguna serta minat berlangganan Spotify Premium. Tahapan penelitian meliputi pengumpulan data, pre-processing data, analisis clustering menggunakan algoritma K-Medoids, serta pemodelan menggunakan regresi polinomial.

Data yang digunakan merupakan data sekunder yang diperoleh dari *platform Kaggle*, yaitu *Spotify User Behaviour Dataset* yang berisi berbagai atribut perilaku pengguna seperti usia, perangkat, jenis langganan, preferensi konten, serta intensitas penggunaan.

Tahap awal penelitian adalah pre-processing data yang meliputi pembersihan data, transformasi data kategorikal menjadi numerik, serta standarisasi data. Selanjutnya dilakukan penentuan jumlah cluster optimal menggunakan metode Elbow, kemudian dilakukan proses clustering menggunakan algoritma K-Medoids.

Untuk mempermudah visualisasi hasil clustering, digunakan *Principal Component Analysis* (PCA) untuk mereduksi dimensi data menjadi dua dimensi utama.

Setelah proses clustering, dilakukan analisis hubungan antar variabel menggunakan regresi polinomial, dengan variabel dependen berupa *premium subscription willingness* dan variabel independen berupa atribut perilaku pengguna seperti usia, preferensi konten, perangkat, serta kebiasaan penggunaan Spotify.

Evaluasi model dilakukan menggunakan *Silhouette Score* untuk *clustering* serta *Mean Squared Error* (MSE) dan *R-Squared* (R^2) untuk regresi.

Silhouette Score dihitung menggunakan persamaan:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \dots \dots \dots (i)$$

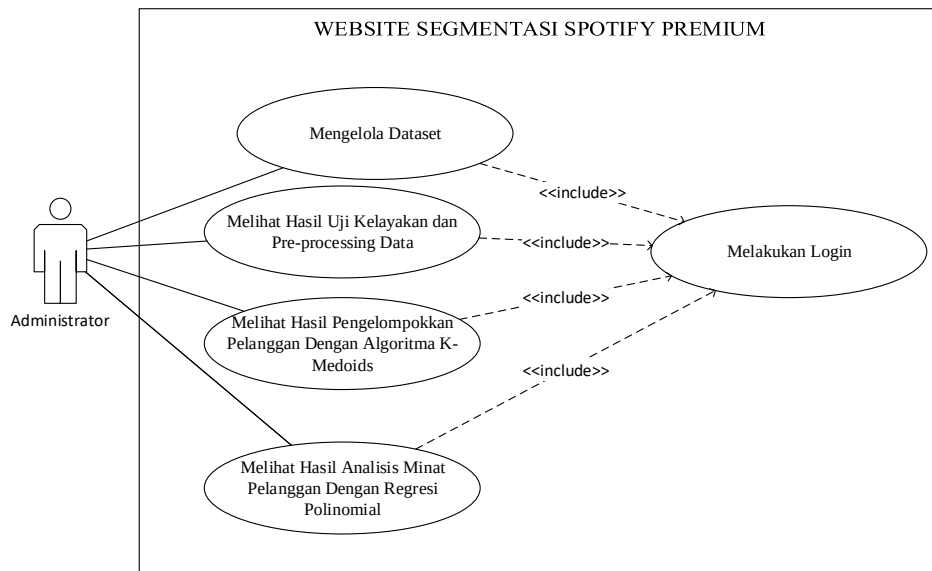
di mana $a(i)$ adalah rata-rata jarak data terhadap cluster sendiri dan $b(i)$ adalah jarak minimum ke cluster lain.

Mean Squared Error (MSE) dihitung dengan:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \dots \dots \dots (ii)$$

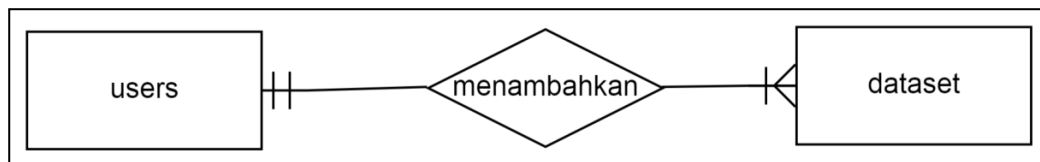
Sedangkan koefisien determinasi (R^2) dihitung dengan:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \dots \dots \dots (iii)$$



Gambar 1. Use Case Diagram Sistem Usulan.

Selain itu, perancangan basis data dilakukan menggunakan *Entity Relationship Diagram* (ERD) untuk memastikan struktur data yang digunakan mampu mendukung kebutuhan sistem secara menyeluruh. ERD menggambarkan hubungan antar entitas seperti pengguna, permintaan layanan, dan jenis layanan yang dikelola dalam sistem.



Gambar 2. Entity Relationship Diagram (ERD).

Sistem yang dibangun menggunakan arsitektur berbasis web dengan konsep client-server. Pengguna mengakses sistem melalui browser, kemudian data diproses di sisi server menggunakan *framework Flask* dan bahasa pemrograman *Python*. Proses analisis dilakukan melalui tahapan *pre-processing*, clustering menggunakan algoritma K-Medoids, serta pemodelan menggunakan regresi polinomial, kemudian hasilnya ditampilkan dalam bentuk

visualisasi interaktif pada antarmuka pengguna.

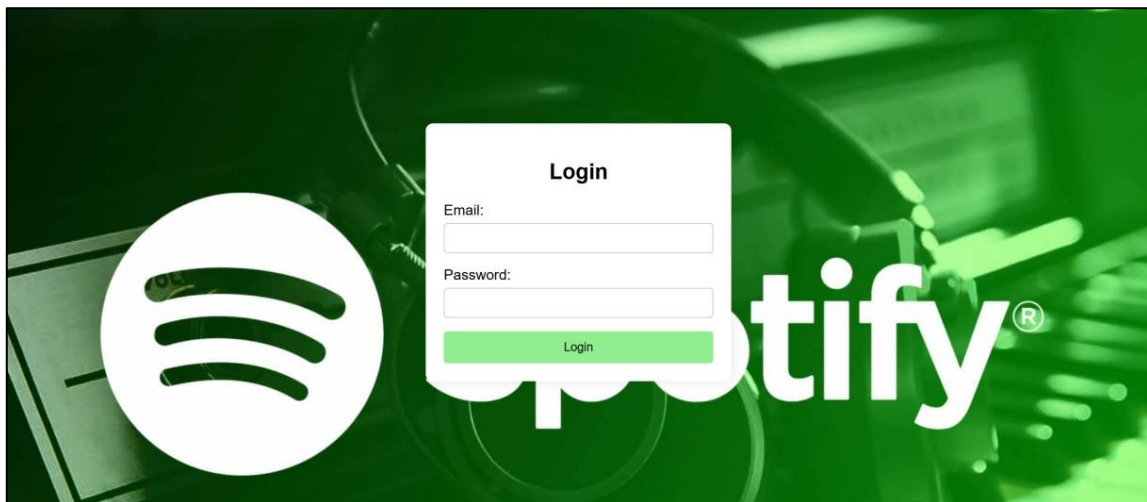
Setelah tahap perancangan selesai, sistem diimplementasikan dalam bentuk aplikasi berbasis web yang terintegrasi dengan basis data. Sistem ini memungkinkan administrator untuk mengelola dataset, melakukan proses analisis, serta melihat hasil segmentasi pelanggan dan analisis minat berlangganan secara *real-time*.

Pengujian sistem dilakukan menggunakan metode *blackbox testing* untuk memastikan bahwa setiap fungsi dalam sistem berjalan sesuai dengan kebutuhan yang telah ditentukan. Pengujian dilakukan dengan memberikan berbagai skenario input dan mengamati output yang dihasilkan oleh sistem tanpa melihat struktur kode program.

4. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini adalah sebuah website analisis segmentasi perilaku dan minat berlangganan Spotify Premium menggunakan metode K-Medoids dan regresi polinomial.

Implementasi sistem dimulai dari halaman login yang berfungsi sebagai pintu masuk bagi pengguna untuk mengakses sistem. Setelah berhasil melakukan autentikasi, pengguna akan diarahkan ke halaman utama yang menampilkan ringkasan informasi layanan.



Gambar 3. Tampilan Halaman Login.

Setelah berhasil melakukan login, pengguna dapat mengakses halaman utama untuk melakukan pengelolaan data layanan. Halaman ini menampilkan data layanan dalam bentuk tabel yang terstruktur sehingga memudahkan proses pengelolaan dan pencarian data.

Dataset											
Tambah Data Import CSV Kosongkan Data 10 entries per page Search:											
ID	Age	Gender	Usage Period	Listening Device	Subscription Plan	Premium Willingness	Preferred Premium Plan	Listening Content	Favorite Genre	Music Time Slot	
1	20-35	Female	More than 2 years	Smart speakers or voice assistants	Free (ad-supported)	Yes	Family Plan-Rs 179/month	Podcast	Melody	Night	
2	12-20	Male	More than 2 years	Computer or laptop	Free (ad-supported)	Yes	Individual Plan- Rs 119/ month	Podcast	Rap	Afternoon	
3	35-60	Others	6 months to 1 year	Smart speakers or voice assistants	Free (ad-supported)	Yes	Student Plan-Rs 59/month	Podcast	Pop	Night	
4	20-35	Female	1 year to 2 years	Smartphone, Smart speakers or voice assistants	Free (ad-supported)	No	None	Music	Melody	Night	
5	20-35	Female	1 year to 2 years	Smartphone	Free (ad-supported)	No	None	Music	Melody	Night	
6	20-35	Male	More than 2 years	Smartphone	Free (ad-supported)	No	None	Music	Pop	Night	
7	20-35	Female	1 year to 2 years	Smartphone	Free (ad-supported)	No	None	Music	Melody	Morning	
8	20-35	Female	Less than 6 months	Smartphone	Free (ad-supported)	No	None	Music	Melody	Night	
9	20-35	Female	Less than 6 months	Smartphone	Free (ad-supported)	Yes	Individual Plan- Rs 119/ month	Music	Melody	Afternoon	
10	20-35	Female	More than 2 years	Smartphone	Free (ad-supported)	No	None	Music	Melody	Night	
Showing 1 to 10 of 520 entries											

Gambar 4. Tampilan Kelola Dataset.

Selanjutnya, sistem menyediakan fitur untuk menambahkan data layanan baru melalui form input yang telah disediakan.


[Dataset](#)
[Hasil Uji Kelayakan dan Pre-processing](#)
[Hasil Pengelompokan Pelanggan](#)
[Hasil Analisis Minat Pelanggan](#)
Keluar

Tambah Dataset

Kembali

Age *

Pilih

Gender *

Pilih

Spotify Usage Period *

Pilih

Listening Device *

☐ Computer or laptop
 ☐ Smartphone
 ☐ Wearable devices
 ☐ Smart speakers or voice assistants

Subscription Plan *

Pilih

Willing to Subscribe Premium *

Pilih

Preferred Premium Plan

None

Preferred Listening Content *

Pilih

Favorite Music Genre *

Pilih

Music Time Slot *

Pilih

Influential Mood *

☐ Relaxation and stress relief
 ☐ Uplifting and motivational
 ☐ Sadness or melancholy
 ☐ Social gatherings or parties

Music Listening Frequency *

☐ Leisure Time
 ☐ Workout Session
 ☐ Study Hours
 ☐ While Traveling
 ☐ Office Hours
 ☐ Night Time
 ☐ When Cooking
 ☐ Random
 ☐ Before Bed

Music Explore Method *

Pilih

Music Recommendation Rating *

Pilih

Podcast Listening Frequency *

Pilih

Favorite Podcast Genre *

Pilih

Podcast Format

Pilih

Host Preference *

Pilih

Podcast Duration *

Pilih

Podcast Variety Satisfaction *

Pilih

Simpan

Gambar 5. Tampilan Tambah Dataset.

Selain itu, sistem juga menyediakan fitur untuk mengubah data yang telah ada sehingga memungkinkan pengguna untuk memperbarui informasi sesuai dengan kondisi terbaru.

The screenshot shows a web form titled "Ubah Dataset" with a navigation bar at the top containing links: Dataset, Hasil Uji Kelayakan dan Pre-processing, Hasil Pengelompokkan Petanggan, Hasil Analisis Minat Petanggan, and Keluar. A red "Kembali" (Back) button is in the top left. The form fields include:

- Age * (dropdown: 20-35)
- Gender * (dropdown: Female)
- Spotify Usage Period * (dropdown: More than 2 years)
- Listening Device * (checkboxes: Computer or laptop, Smartphone, Wearable devices, Smart speakers or voice assistants - checked)
- Subscription Plan * (dropdown: Free)
- Willing to Subscribe Premium * (dropdown: Yes)
- Preferred Premium Plan * (dropdown: Free (ad-supported))
- Preferred Listening Content * (dropdown: Podcast)
- Favorite Music Genre * (dropdown: Melody)
- Music Time Slot * (dropdown: Night)
- Influential Mood * (checkboxes: Relaxation and stress relief, Uplifting and motivational, Sadness or melancholy - checked, Social gatherings or parties)
- Music Listening Frequency * (checkboxes: Leisure Time - checked, Workout Session, Study Hours, While Traveling, Office Hours, Night Time, When Cooking, Random, Before Bed)
- Music Explore Method * (dropdown: Playlists)
- Music Recommendation Rating * (dropdown: 3)
- Podcast Listening Frequency * (dropdown: Never)
- Favorite Podcast Genre * (dropdown: Comedy)
- Podcast Format (dropdown: Interview)
- Host Preference * (dropdown: Both)
- Podcast Duration * (dropdown: Both)
- Podcast Variety Satisfaction * (dropdown: Ok)

A blue "Simpan" (Save) button is at the bottom right of the form.

Gambar 6. Tampilan Ubah Dataset.

Sistem juga dilengkapi dengan fitur untuk menampilkan hasil pengolahan data serta monitoring layanan yang sedang berjalan.

Selain implementasi antarmuka, hasil utama dari penelitian ini juga mencakup analisis data yang dilakukan menggunakan metode K-Medoids dan regresi polinomial. Hasil perhitungan total biaya (*inertia*) untuk setiap nilai k disajikan pada Tabel berikut:

Tabel 1. Hasil Total Biaya *Inertia* Dari Metode *Elbow*.

Nilai k	Total Biaya (<i>Inertia</i>)
2	1.019,41
3	998,53
4	976,84
5	945,22
6	936,92
7	929,59
8	915,39
9	905,74
10	902,03

Berdasarkan hasil pengujian menggunakan metode Elbow, diperoleh jumlah cluster optimal sebanyak $k=9$. Hasil clustering menunjukkan bahwa pelanggan dapat dikelompokkan ke dalam sembilan segmen berdasarkan pola perilaku penggunaan Spotify. Setiap cluster merepresentasikan karakteristik pengguna yang berbeda, seperti kelompok pengguna dengan intensitas penggunaan tinggi, pengguna dengan preferensi konten tertentu, serta pengguna dengan tingkat interaksi rendah. Segmentasi ini memberikan gambaran yang lebih spesifik mengenai perilaku pelanggan. Hasil segmentasi ini sejalan dengan konsep segmentasi perilaku yang menyatakan bahwa konsumen dapat dikelompokkan berdasarkan pola penggunaan dan preferensi layanan.

Evaluasi clustering menggunakan Silhouette Score menghasilkan nilai sebesar 0,1659 yang menunjukkan bahwa kualitas cluster berada pada kategori lemah hingga sedang. Hal ini mengindikasikan bahwa masih terdapat overlap antar cluster.

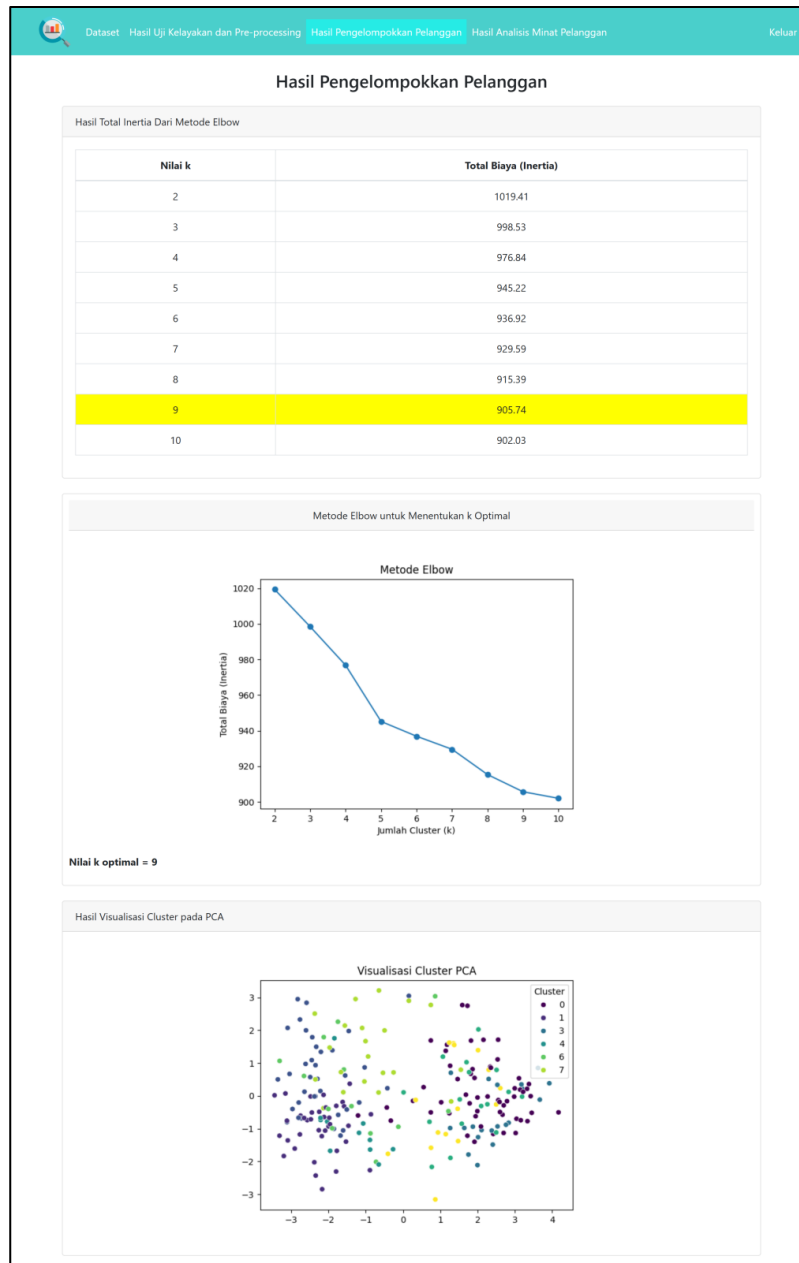
Selanjutnya, hasil analisis regresi polinomial menunjukkan nilai Mean Squared Error (MSE) sebesar 1,0694 dan nilai koefisien determinasi (R^2) sebesar -0,1109, yang menunjukkan bahwa model belum mampu menjelaskan variasi data secara optimal. Kondisi ini kemungkinan disebabkan oleh kompleksitas hubungan antar variabel yang tidak sepenuhnya dapat direpresentasikan oleh model polinomial, serta adanya variabel lain di luar dataset yang turut memengaruhi minat pelanggan. Meskipun demikian, beberapa atribut seperti intensitas penggunaan dan preferensi konten tetap menunjukkan pengaruh terhadap minat berlangganan pengguna. Temuan bahwa intensitas penggunaan berpengaruh terhadap minat berlangganan juga didukung oleh penelitian sebelumnya yang menyatakan bahwa frekuensi penggunaan layanan digital berhubungan dengan loyalitas pengguna. Fitur ini bertujuan untuk memberikan informasi yang lebih terstruktur kepada pengguna dan pengelola dalam memahami kondisi data serta perkembangan layanan secara keseluruhan.

Pada tahap awal, sistem menampilkan hasil uji kelayakan dan proses pre-processing data. Halaman ini berfungsi untuk memastikan bahwa data yang digunakan dalam sistem telah memenuhi standar kualitas sebelum diproses lebih lanjut. Informasi yang ditampilkan mencakup kondisi data, seperti kelengkapan data serta hasil transformasi yang dilakukan selama tahap pre-processing. Dengan adanya fitur ini, administrator dapat mengetahui apakah data yang digunakan sudah layak untuk dianalisis.

Dataset Hasil Uji Kelayakan dan Pre-processing Hasil Pengelompokan Pelanggan Hasil Analisis Minat Pelanggan Keluar					
Hasil Uji Kelayakan & Pre-processing					
Missing Value Setiap Column					
Kolom	Jumlah Missing				
age	0				
gender	0				
spotify_usage_period	0				
spotify_listening_device	0				
spotify_subscription_plan	0				
premium_sub_willingness	0				
preferred_premium_plan	208				
preferred_listening_content	0				
fav_music_genre	0				
music_time_slot	0				
music_influential_mood	0				
music_lis_frequency	0				
music_expl_method	0				
music_recc_rating	0				
pod_lis_frequency	0				
fav_pod_genre	148				
preferred_pod_format	140				
pod_host_preference	141				
preferred_pod_duration	129				
pod_variety_satisfaction	0				
Pemeriksaan Tipe Data					
Kolom	Tipe Data				
age	object				
gender	object				
spotify_usage_period	object				
spotify_listening_device	object				
spotify_subscription_plan	object				
premium_sub_willingness	object				
preferred_premium_plan	object				
preferred_listening_content	object				
fav_music_genre	object				
music_time_slot	object				
music_influential_mood	object				
music_lis_frequency	object				
music_expl_method	object				
music_recc_rating	int64				
pod_lis_frequency	object				
fav_pod_genre	object				
preferred_pod_format	object				
pod_host_preference	object				
preferred_pod_duration	object				
pod_variety_satisfaction	object				
Hasil Standarisasi Data					
age	gender	spotify_usage_period	spotify_listening_device	spotify_subscription_plan	premium_sub_wi
0.16418781318183337	-0.4735048871614975	1.3844373104863459	-0.2325712764925474	-0.709475654760603	0.75261780900631
0.16418781318183337	-0.4735048871614975	0.5417363388859614	0.36623292964918375	-0.709475654760603	0.75261780900631
0.16418781318183337	-0.4735048871614975	0.5417363388859614	0.36623292964918375	-0.709475654760603	0.75261780900631
0.16418781318183337	-0.4735048871614975	1.3844373104863459	0.9650371357909149	1.4094916341243982	0.75261780900631
0.16418781318183337	1.3552036425656653	1.3844373104863459	0.6656350327200493	1.4094916341243982	0.75261780900631
-1.5871486607577224	1.3552036425656653	-0.3009646327144231	1.8632434450035116	1.4094916341243982	0.75261780900631
0.16418781318183337	-0.4735048871614975	1.3844373104863459	0.6656350327200493	-0.709475654760603	0.75261780900631
-1.5871486607577224	-0.4735048871614975	1.3844373104863459	0.36623292964918375	-0.709475654760603	0.75261780900631
1.9155244871213892	1.3552036425656653	-1.1436656043148075	1.8632434450035116	-0.709475654760603	0.75261780900631
0.16418781318183337	1.3552036425656653	-1.1436656043148075	0.9650371357909149	1.4094916341243982	0.75261780900631

Gambar 7. Tampilan Lihat Hasil Uji Kelayakan dan Pre-processing Dataset.

Selanjutnya, sistem menampilkan hasil pengelompokan data yang telah diproses sebelumnya. Halaman ini menyajikan hasil segmentasi dalam bentuk visualisasi sehingga memudahkan pengguna dalam memahami pola data yang terbentuk. Informasi yang ditampilkan pada halaman ini membantu pengelola dalam mengidentifikasi karakteristik kelompok data serta memahami distribusi data secara lebih komprehensif.

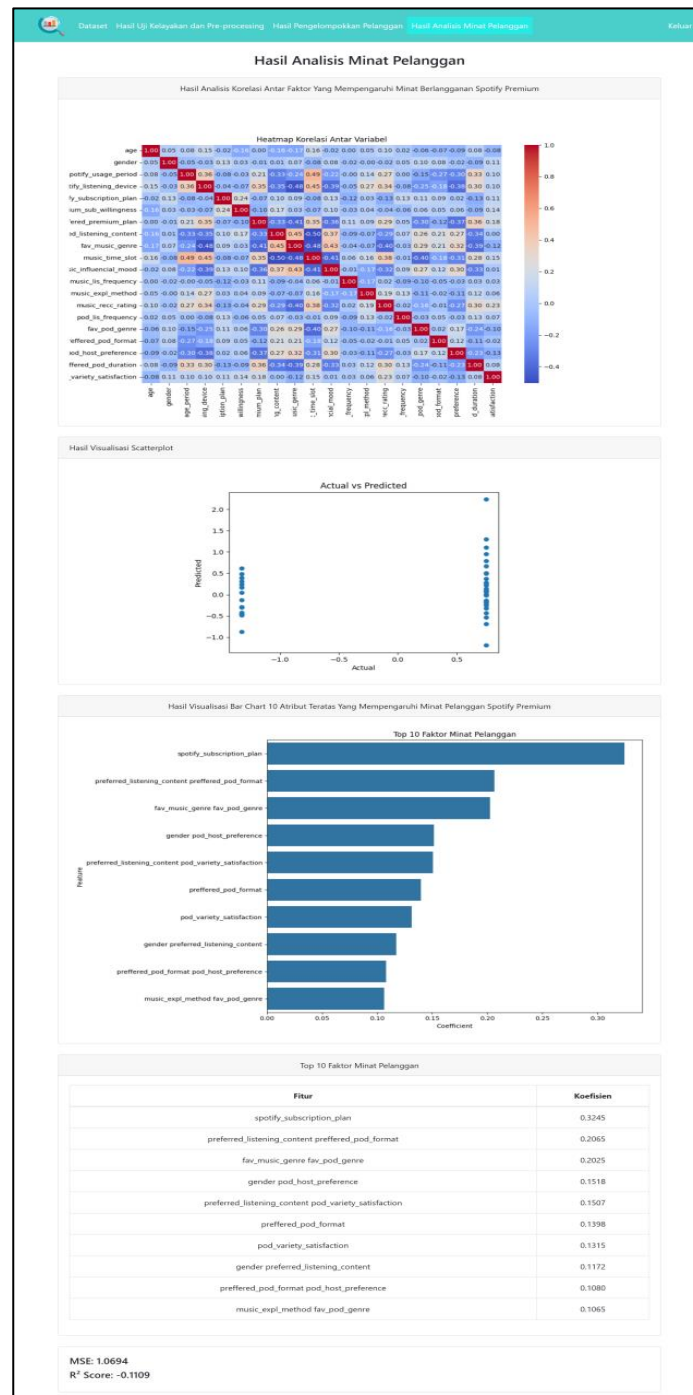


Gambar 8. Tampilan Lihat Hasil Pengelompokan Pelanggan.

Selain itu, sistem juga menyediakan halaman analisis yang menampilkan hasil pengolahan data dalam bentuk visualisasi yang lebih lanjut. Halaman ini digunakan untuk memberikan gambaran mengenai hubungan antar variabel yang dianalisis serta membantu dalam proses pengambilan keputusan. Informasi yang disajikan memungkinkan pengguna untuk memahami faktor-faktor yang berpengaruh terhadap data yang diolah dalam sistem.

Berdasarkan hasil analisis korelasi, variabel *spotify subscription plan* memiliki hubungan paling kuat terhadap minat berlangganan dengan nilai korelasi sebesar 0.24, sedangkan sebagian besar variabel lain menunjukkan hubungan yang lemah. Hal ini mengindikasikan bahwa strategi peningkatan pelanggan premium dapat difokuskan pada

optimalisasi penawaran paket langganan, seperti pemberian promosi atau diferensiasi fitur pada setiap paket untuk meningkatkan daya tarik pengguna.



Gambar 9. Tampilan Lihat Hasil Pengelompokkan Pelanggan.

Berdasarkan hasil implementasi, sistem yang dibangun mampu meningkatkan efisiensi dalam proses analisis data pelanggan Spotify Premium. Proses analisis data pelanggan yang sebelumnya dilakukan secara terpisah kini dapat dilakukan dalam satu sistem terintegrasi, sehingga memudahkan pengolahan, analisis, dan visualisasi data secara lebih efisien.

Hasil pengujian sistem menggunakan metode blackbox menunjukkan bahwa seluruh fungsi utama sistem telah berjalan sesuai dengan yang diharapkan. Proses login, pengelolaan data, pengajuan layanan, serta monitoring layanan dapat dilakukan tanpa mengalami kendala yang signifikan. Hal ini menunjukkan bahwa sistem yang dibangun telah memenuhi kebutuhan pengguna dan dapat digunakan sebagai solusi dalam analisis segmentasi pelanggan serta pengambilan keputusan strategis pada layanan Spotify Premium.

Nilai *Silhouette Score* yang relatif rendah menunjukkan bahwa pemisahan antar cluster belum optimal, yang kemungkinan disebabkan oleh kemiripan karakteristik antar pengguna. Selain itu, nilai R^2 yang negatif mengindikasikan bahwa model regresi belum mampu menangkap hubungan kompleks antar variabel secara maksimal, sehingga diperlukan pendekatan model lain atau seleksi fitur yang lebih optimal pada penelitian selanjutnya.

5. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, maka diperoleh kesimpulan sebagai berikut:

1. Algoritma K-Medoids berhasil digunakan untuk melakukan clustering pelanggan Spotify Premium berdasarkan pola perilaku berlangganan, dengan jumlah kluster optimal sebanyak $k=9$ berdasarkan metode Elbow. Namun, hasil evaluasi menunjukkan nilai *Silhouette Score* sebesar 0,1659, yang mengindikasikan bahwa kualitas pemisahan antar cluster masih berada pada kategori lemah hingga sedang.
2. Algoritma Regresi Polinomial berhasil digunakan untuk menganalisis hubungan antara faktor-faktor perilaku pelanggan dengan minat berlangganan Spotify Premium, dengan hasil evaluasi model menunjukkan nilai Mean Squared Error (MSE) sebesar 1,0694 dan R^2 score sebesar -0,1109, yang menandakan bahwa model belum mampu menjelaskan variasi data dengan baik dan masih memiliki performa yang kurang optimal.

Saran

Berdasarkan hasil penelitian, maka saran yang dapat diberikan adalah sebagai berikut:

1. Pada penggunaan dataset yang berasal dari Kaggle yaitu Spotify User Behaviour Dataset, disarankan untuk menggunakan dataset yang lebih terbaru atau menambahkan data dari sumber lain agar hasil analisis lebih representatif dan dapat mencerminkan kondisi pengguna yang lebih luas.
2. Atribut-atribut yang digunakan dalam penelitian ini dapat dikembangkan lebih lanjut dengan menambahkan variabel lain di luar 20 atribut yang telah digunakan, seperti data perilaku pengguna secara real-time atau data interaksi aplikasi, sehingga dapat

meningkatkan akurasi clustering maupun prediksi model.

3. Hasil analisis yang diperoleh seperti jumlah kluster, clustering segmentasi pelanggan, heatmap korelasi, grafik Regresi Polinomial, dan 10 faktor teratas sebaiknya dapat dikembangkan lebih lanjut dengan metode lain seperti validasi clustering tambahan (Davies-Bouldin Index) atau model prediksi lain seperti Random Forest dan Gradient Boosting agar hasil analisis lebih akurat dan komprehensif.

DAFTAR PUSTAKA

- Amna, Wahyuddin, S., Sudipa, I. G. I., Putra, T. A. E., Wahidin, A. J., Syukrilla, W. A., Wardhani, A. K., Heryana, N., Indriyani, T., & Santoso, L. W. (2023). *Data Mining*. PT Global Eksekutif Teknologi.
- Damayanti, S. E., Fajriana, M. I., Meilani, D., & Fatimah, S. S. (2024). Menjelajahi Wawasan Industri Musik: Klusterisasi Lagu Terpopuler di Spotify 2024 dengan Metode K-Means Clustering. *Prosiding Seminar Nasional*, 562–567.
- Faran, J., & Aldisa, R. T. (2024a). Perbandingan Algoritma K-Means dan K-Medoids dalam Pengelompokan Kelas untuk Mahasiswa Baru Program Magister. *Journal of Information System Research*, 5(2), 509–519. <https://doi.org/10.47065/josh.v5i2.4753>
- Faran, J., & Aldisa, R. T. (2024b). Perbandingan Algoritma K-Means dan K-Medoids Dalam Pengelompokan Kelas Untuk Mahasiswa Baru Program Magister. *Journal of Information System Research*, 5(2), 509–519. <https://doi.org/10.47065/josh.v5i2.4753>
- Gomathy, C. K., Hasitha, N. S., Sampadaa, G. S., & Srilalitha, S. (2022). Big Data Analytics in Spotify. *International Journal of Scientific Research in Engineering and Management*, 6(11), 1–9. <https://doi.org/10.55041/ijssrem16591>
- Hidayat, A. M. P., & Saputra, D. I. S. (2026). Perancangan Website Pelayanan dan Edukasi Satres PPA PPO Polresta Banyumas Berbasis Laravel. *Bridge: Jurnal Publikasi Sistem Informasi Dan Telekomunikasi*, 4(2). <https://doi.org/10.62951/bridge.v4i2.876>
- Hidayat, H., Pendi, H. Z., & Warlizasusi, J. (2026). Faith-Based Ecological Counselling: Integrating Spirituality and Sustainability for Muslim Students; [Poradnictwo ekologiczne oparte na wierze. Integracja duchowości i zrównoważonego rozwoju w pracy ze studentami muzułmańskimi]. *Studia Ecologiae et Bioethicae*, 24(1), 45 – 60. <https://doi.org/10.21697/seb.5873>
- Kaligis, & others. (2022). Analisa Perbandingan Algoritma K-Means, K-Medoids, dan X-Means untuk Pengelompokan Kinerja Pegawai. *IT-Explore: Jurnal Penerapan Teknologi Informasi Dan Komunikasi*, 1(3). <https://doi.org/10.24246/itexplore.v1i3.2022.pp179-193>
- Laksono, S. A., Pratama, A. R., & Rahmat. (2023). Perbandingan Metode Linear Regresi dan Polynomial Regresi Untuk Memprediksi Harga Saham Studi Kasus Bank BCA. *INFOTECH: Jurnal Informatika & Teknologi*, 4(1), 59–70. <https://doi.org/10.37373/infotech.v4i1.602>
- Matulessy, V. E., Soegoto, A. S., & Samadi, R. L. (2025). Pengaruh User Experience (UX), Personalisasi Artificial Intelligence (AI) dan Perceived Usefulness Terhadap Loyalitas Pengguna Aplikasi Spotify Pada Generasi Z di Kelurahan Bahu. *Neraca Manajemen*,

Ekonomi, 24(10).

- Mei, R. R. (2024). The Influential Factors of Spotify's Business Strategy. *Journal of Education, Humanities and Social Sciences*, 27, 318–322. <https://doi.org/10.54097/frmx5s45>
- Muttaqin, Arafah, M., Jaya, A. K., Suryawan, M. A., Gustiana, Z., Banjarnahor, A. R., Bukidz, D. P., Hazriani, Simanjuntak, M., Saputra, N., & Fajrillah. (2023). *Implementasi Artificial Intelligence (AI) Dalam Kehidupan*. Yayasan Kita Menulis.
- Nita, & others. (2022). Perancangan Website pada Kedai Geulish Menggunakan Metode Rapid Application Development. *Jurnal Elektronika Dan Komputer*, 15(2). <https://doi.org/10.51903/elkom.v15i2.923>
- Nugraha, R. D., Adelia, D. D., & Rivaldi, D. (2025). Segmentasi Pelanggan Retail Berbasis Perilaku menggunakan Algoritma K- Means Clustering. *Digital Transformation Technology (Digitech)*, 5(2), 141–148.
- Panjaitan, R. A. R., & Andarini, S. (2024). The Influence of Brand Experience and Perceived Usefulness on Brand Loyalty in Spotify Premium Services in Surabaya. *COSTING: Journal of Economic, Business and Accounting*, 7(4), 7465–7477.
- Permana, A. A., S, W., Santoso, L. W., Wibowo, G. W. N., Wardhani, A. K., Rahmadden, Wahidin, A. J., Yuliasuti, G. E., Elisawati, Wijayanti, R. R., & Abdurrasyid. (2022). *Machine Learning*. PT Global Eksekutif Teknologi.
- Prayitno, P., & others. (2026). Segmentasi Pelanggan Ritel Global dan Inggris Menggunakan RFM dan K-Means Clustering. *Prosiding Seminar Nasional Ilmu Teknik*, 2(2). <https://doi.org/10.61132/prosemnasproit.v2i2.84>
- Ruddin, I., Santoso, H., & Indrajit, R. E. (2022). Digitalisasi Musik Industri: Bagaimana Teknologi Informasi Mempengaruhi Industri Musik di Indonesia. *Jurnal Pendidikan Sains Dan Komputer*, 2(01), 124–136. <https://doi.org/10.47709/jpsk.v2i01.1395>
- Salwa, A., Asdini, I. S., Amanda, S., Putriku, A. E., & Saragih, L. S. (2024). Strategi Pemasaran Aplikasi Spotify dalam Global Spotify's Global Marketing Strategy. *Student Scientific Creativity Journal (SSCJ)*, 2(4), 100–108.
- Sangging, K. N. D. S., Mitriani, N. W. E., & I Gede Yudhi. (2025). Pengaruh Gaya Hidup, Citra Merek dan Kepercayaan Terhadap Keputusan Pembelian Konsumen Dalam Memilih Spotify Premium (Studi Kasus Pada Konsumen Pengguna Spotify di Kota Denpasar). *Dinamika Kreatif Manajemen Strategis*, 7(1), 63–81.
- Saputro, D. T., & Sucihermayanti, W. P. (2021a). Penerapan Klasterisasi Menggunakan K-Means untuk Menentukan Tingkat Kesehatan Bayi dan Balita di Kabupaten Bengkulu Utara. *Jurnal Buana Informatika*, 12(2), 146–155. <https://doi.org/10.30659/abdimasku.1.3.126-133>
- Saputro, D. T., & Sucihermayanti, W. P. (2021b). Penerapan Klasterisasi Menggunakan K-Means untuk Menentukan Tingkat Kesehatan Bayi dan Balita di Kabupaten Bengkulu Utara. *Jurnal Buana Informatika*, 12(2), 146–155.
- Simorangkir, E. S. A., Siahaan, A. P. U., Marlina, L., Nasution, D., & Sitorus, Z. (2024). Deteksi Outlier Hasil Clustering Algoritma K-Medoids Menggunakan Metode Boxplot Pada Data KIP Kuliah. *Journal of Computer System and Informatics (JoSYC)*, 5(4), 893–902. <https://doi.org/10.47065/josyc.v5i4.5479>
- Wahyuddin, Pasaribu, J. S., Bau, R. T. R. ., Munawar, Z., A, H., Harto, B., Joshua, S. R., Putri, N. I., Safii, M., Amna, Sophian, S., Rukmana, A. Y., & Hariyadi. (2023). *Layanan*

Digital di Era 5.0. PT Global Eksekutif Teknologi.

Yuliani, A. D., & Kurniadi, O. (2023). Peranan Media Streaming dalam Menggantikan Televisi Konvensional di Kalangan Masyarakat. *Jurnal Riset Manajemen Komunikasi*, 3(2), 109–114. <https://doi.org/10.29313/jrmk.v3i2.3140>