

## **Pengembangan Sistem Rekomendasi Musik dengan K-Means dan KNN Berbasis *Cosine Similarity***

*Development of Music Recommendation System Using K-Means and KNN  
Based on Cosine Similarity*

**Moh Akbar Munajad<sup>1\*</sup>, Achmad Ridwan<sup>2</sup>, Taftazani Ghazi Pratama<sup>3</sup>**

<sup>1,3</sup>*Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi,*

<sup>2</sup>*Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,  
Universitas Muhammadiyah Kudus*

\*corr\_author: [32021110009@std.umku.ac.id](mailto:32021110009@std.umku.ac.id)

### **ABSTRAK**

Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi musik berdasarkan data pengguna dengan menerapkan algoritma K-Means dan K-Nearest Neighbors (K-NN) yang dipadukan dengan *Cosine Similarity*. Pendekatan ini menawarkan kebaruan melalui integrasi metode clustering dan klasifikasi berbasis kemiripan vektor untuk menghasilkan rekomendasi yang lebih kontekstual dan presisi. Pengumpulan data dilakukan melalui platform Spotify yang mencakup berbagai fitur lagu seperti danceability, energy, acousticness, dan lain-lain. Data kemudian diproses melalui tahapan preprocessing, visualisasi, dan reduksi dimensi menggunakan t-SNE untuk mengidentifikasi pola-pola kluster secara visual. Algoritma K-Means digunakan untuk mengelompokkan lagu-lagu berdasarkan kesamaan fitur, sementara K-NN digunakan untuk merekomendasikan lagu kepada pengguna berdasarkan preferensi musik mereka. Evaluasi dilakukan dengan menguji model pada beberapa nilai k (1, 5, dan 10), dan diperoleh akurasi terbaik sebesar 91% pada nilai k = 1. Temuan ini menunjukkan bahwa kombinasi K-Means, K-NN, dan *Cosine similarity* tidak hanya meningkatkan akurasi, tetapi juga memperkuat relevansi rekomendasi terhadap karakteristik musik yang kompleks. Penelitian ini memberikan kontribusi dalam pengembangan sistem rekomendasi berbasis konten, membuka peluang pengembangan lebih lanjut dengan memperluas jenis data dan metode evaluasi, serta menawarkan pendekatan yang efisien untuk personalisasi musik di era big data.

**Kata-kata kunci:** *Cosine Similarity*, K-Means, K-Nearest Neighbors, Sistem Rekomendasi, t-SNE

### **ABSTRACT**

*This study aims to develop a music recommendation system based on user data by applying the K-Means and K-Nearest Neighbors (K-NN) algorithms, combined with Cosine Similarity. This approach offers novelty through the integration of clustering and classification methods based on vector similarity to generate more contextual and precise recommendations. Data collection was carried out through the Spotify platform, which includes various song features such as danceability, energy, acousticness, and others. The*

---

*data was then processed through preprocessing, visualization, and dimensionality reduction using t-SNE to visually identify cluster patterns. The K-Means algorithm was used to group songs based on feature similarities, while K-NN was employed to recommend songs to users based on their music preferences. Evaluation was conducted by testing the model with several values of k (1, 5, and 10), with the best accuracy achieved being 91% at k = 1. These findings indicate that the combination of K-Means, K-NN, and Cosine similarity not only improves accuracy but also enhances the relevance of recommendations to the complex characteristics of music. This research contributes to the development of content-based recommendation systems, opens opportunities for further development by expanding data types and evaluation methods, and offers an efficient approach for music personalization in the era of big data.*

**Keywords:** *Cosine Similarity, K-Means, K-Nearest Neighbors, Recommendation System, t-SNE*

## PENDAHULUAN

Di era digital saat ini, pertumbuhan industri musik mengalami akselerasi signifikan seiring dengan kemajuan teknologi informasi dan komunikasi (Sahel *et al.*, 2024). Transformasi digital telah mengubah cara manusia mengakses, mengonsumsi, dan berinteraksi dengan musik. Dalam hal tersebut, sistem rekomendasi memainkan peran penting dalam memberikan pengalaman personalisasi kepada pengguna. Sistem rekomendasi dapat membantu pengguna menemukan musik baru yang sesuai dengan selera mereka tanpa harus mencarinya secara manual (Amin *et al.*, 2023).

Sistem rekomendasi tidak hanya terbatas pada kenyamanan pengguna, tetapi juga berdampak terhadap promosi dan distribusi konten musik. Algoritma rekomendasi yang akurat dapat meningkatkan retensi pengguna, memperluas eksposur artis, dan meningkatkan pendapatan platform melalui peningkatan *engagement* (Fadlila, Aranta and Bimantoro, 2023). Namun, permasalahan dalam pengembangan sistem rekomendasi adalah bagaimana menyesuaikan rekomendasi terhadap preferensi pengguna yang kompleks dan dinamis. Dalam beberapa kasus, sistem rekomendasi masih belum optimal dalam mengatasi masalah seperti *cold-start user*, *over-specialization*, serta kurangnya adaptasi terhadap perubahan selera pengguna seiring waktu.

Berbagai pendekatan algoritma telah diterapkan untuk membangun sistem rekomendasi, seperti *content-based filtering*, *collaborative filtering*, dan *hybrid recommendation system* (Saputra *et al.*, 2025). Salah satu metode yang populer adalah *collaborative filtering*, yang memanfaatkan kesamaan preferensi antar pengguna, tetapi, metode ini memiliki kelemahan ketika data pengguna tidak lengkap atau belum tersedia (Permadi and Raharjo, 2023). Oleh karena itu, pendekatan berbasis *unsupervised learning* dan *distance-based methods*, seperti algoritma K-Means dan K-Nearest Neighbors (K-NN), menjadi alternatif yang menjanjikan. K-Means digunakan untuk mengelompokkan pengguna atau lagu berdasarkan kesamaan fitur, sedangkan K-NN dapat mengidentifikasi item serupa berdasarkan kedekatan dalam ruang vektor.

Meskipun banyak penelitian yang telah mengeksplorasi metode collaborative atau model berbasis deep learning, terdapat celah praktis pada skenario dengan data pengguna yang sparsity tinggi, keterbatasan sumber daya komputasi, dan kebutuhan akan model yang mudah diinterpretasi. K-Means mampu menemukan struktur kluster alami dalam fitur audio sehingga mengurangi ruang pencarian, sementara K-NN, ketika dipadukan dengan *cosine similarity* memberikan ukuran kemiripan yang stabil pada vektor berdimensi tinggi. Kombinasi ini mengatasi kelemahan *collaborative filtering* yang membutuhkan banyak

---

interaksi pengguna dan menghindari kompleksitas serta kebutuhan data besar pada metode *deep learning*, sehingga cocok untuk aplikasi skala kecil hingga menengah.

*Cosine similarity* juga merupakan metrik yang banyak digunakan dalam sistem rekomendasi untuk mengukur tingkat kemiripan antar item atau antar pengguna. Dibandingkan dengan euclidean distance, *cosine similarity* lebih sesuai untuk data yang bersifat *sparse* dan *high-dimensional* seperti data musik, karena metrik ini berfokus pada arah vektor ketimbang magnitudonya (Yoshua and Bunyamin, 2021). Dengan menempatkan *cosine similarity* sebagai metrik dalam K-NN, sistem lebih sensitif terhadap pola fitur, seperti pada distribusi energy, danceability, valence, daripada perbedaan skala antar fitur

Penelitian mengenai sistem rekomendasi musik telah banyak dilakukan. Salah satu penelitian yang relevan adalah penelitian berjudul “Sistem Rekomendasi Musik Spotify Berdasarkan Listening History Pengguna” oleh Ichsan Madani dan Hasrullah. Hasil evaluasi menunjukkan tingkat presisi sebesar 87%, recall sebesar 82%, serta *Mean Absolute Error* (MAE) sebesar 0,15 (Madani and Hasrullah, 2023). Nilai-nilai metrik tersebut menunjukkan bahwa sistem rekomendasi yang dibangun mampu memberikan hasil rekomendasi yang cukup akurat dan sesuai.

Penelitian lain dilakukan oleh Melany Mustika Dewi, dkk yang berjudul "Implementasi K-Means dan Collaborative Filtering untuk Sistem Rekomendasi". Evaluasi kinerja sistem pada penelitian tersebut menggunakan metrik *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE), di mana hasilnya menunjukkan nilai error MAE sebesar 0,54, sedangkan RMSE memiliki nilai error lebih tinggi yaitu 0,71 (Dewi, Farida and Dahlan, 2024). Temuan tersebut memperlihatkan bahwa KNN dapat dioptimalkan untuk memberikan rekomendasi yang lebih akurat dengan pemilihan parameter k yang tepat.

Berdasarkan tinjauan dari penelitian terdahulu, dapat disimpulkan bahwa sistem rekomendasi musik telah banyak dikembangkan dengan berbagai pendekatan. Namun, studi yang mengeksplorasi integrasi eksplisit antara clustering sebagai pra-segmentasi, dan K-NN berbasis cosine untuk sistem rekomendasi musik masih relatif terbatas, khususnya pada pengujian parameter k di dalam cluster yang berbeda dan penerapan pada dataset lokal. Penelitian ini berbeda karena menggabungkan metode clustering menggunakan K-Means dengan klasifikasi K-Nearest Neighbors (KNN) yang menggunakan *cosine similarity* sebagai ukuran kesamaan. Secara konseptual, pendekatan yang dikembangkan dalam penelitian ini mengasumsikan bahwa pengguna dengan pola interaksi atau preferensi serupa akan memiliki kecenderungan yang sama dalam menyukai jenis lagu tertentu (Pratama and Darmawan, 2025). Oleh karena itu, pengguna dikelompokkan terlebih dahulu menggunakan K-Means untuk menemukan kelompok homogen, lalu sistem akan merekomendasikan lagu berdasarkan kemiripan pengguna dalam kelompok tersebut menggunakan K-NN dan *cosine similarity* sebagai metrik pengukuran jarak. Dengan demikian, sistem tidak hanya mampu memberikan rekomendasi yang tepat sasaran, tetapi juga mampu beradaptasi dengan kompleksitas data dan selera pengguna yang bervariasi.

Implementasi sistem rekomendasi musik yang mengintegrasikan K-Means dan K-NN dengan *cosine similarity* secara bersamaan masih tergolong minim, khususnya dalam konteks pengembangan lokal atau aplikasi skala kecil (Mustafidah, Mahmud and Suwarsito, 2024). Oleh karena itu, penelitian ini tidak hanya menawarkan kontribusi pada aspek teknis pengembangan algoritma, tetapi juga memberikan evaluasi praktis mengenai trade-off antara akurasi, efisiensi komputasi, dan interpretabilitas model. Hal ini membuka peluang penerapan praktis yang lebih luas, seperti dalam aplikasi berbasis edukasi musik, promosi musisi independen, dan kurasi konten audio berbasis preferensi pengguna.

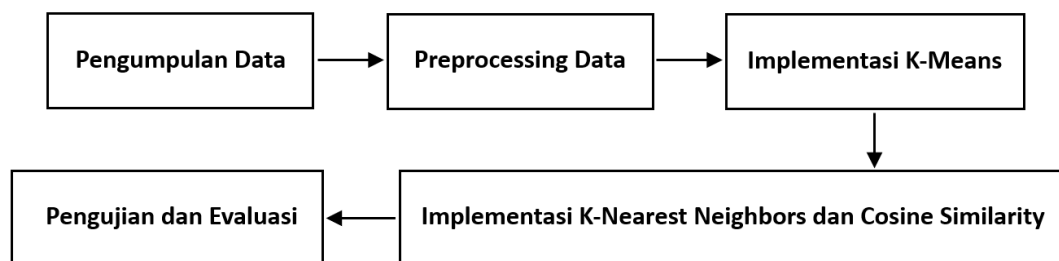
Rumusan masalah yang diangkat dalam penelitian ini yaitu bagaimana cara membangun sistem rekomendasi musik yang mampu memberikan rekomendasi lagu secara akurat dan relevan, bagaimana performa algoritma K-Means dalam mengelompokkan pengguna berdasarkan data interaksi mereka dengan musik, dan seberapa efektif algoritma K-Nearest Neighbors dengan *cosine similarity* dalam merekomendasikan lagu yang sesuai berdasarkan kedekatan vektor.

Tujuan dari penelitian ini yaitu untuk mengembangkan sistem rekomendasi musik dengan menggabungkan algoritma K-Means dan K-Nearest Neighbors berbasis *cosine similarity* guna meningkatkan kualitas personalisasi rekomendasi. Penelitian ini bertujuan untuk mengimplementasikan sistem rekomendasi musik berbasis machine learning yang adaptif terhadap data preferensi pengguna dan mengevaluasi performa kombinasi algoritma K-Means dan K-NN dalam menghasilkan rekomendasi yang relevan.

Dengan adanya penelitian ini, diharapkan dapat memberikan kontribusi terhadap pengembangan sistem rekomendasi berbasis kecerdasan buatan di Indonesia, sekaligus memperluas pemahaman terhadap implementasi teknik klusterisasi dan *similarity measure* dalam domain hiburan digital. Penelitian ini juga diharapkan dapat menjadi dasar untuk pengembangan sistem rekomendasi serupa di bidang lain seperti film, buku, dan *e-learning*. Selain itu, pendekatan yang digunakan diharapkan dapat memberikan gambaran praktis mengenai integrasi metode *unsupervised* dan *supervised learning* dalam menciptakan sistem rekomendasi yang andal dan responsif terhadap kebutuhan pengguna.

## METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dan eksploratif untuk mengembangkan sistem rekomendasi musik berbasis algoritma K-Means clustering dan K-Nearest Neighbors (K-NN) dengan bantuan Cosine Similarity. Pendekatan ini dipilih karena mampu mengelompokkan dan merekomendasikan item (lagu) berdasarkan kemiripan karakteristik yang dimiliki oleh pengguna dan item itu sendiri.



**Gambar 1. Alur metode penelitian**

Gambar 1 menjelaskan tentang alur metode penelitian sebagai berikut:

### 1. Pengumpulan Data

Langkah awal dalam proses pengembangan sistem rekomendasi musik adalah pengumpulan data. Data yang digunakan dalam penelitian ini merupakan data yang bersumber dari dataset daring, baik dari repositori publik. Dataset yang digunakan bersumber dari Spotify API, berisi 5.000 lagu populer dengan 16 fitur numerik seperti danceability, energy, acousticness, tempo, valence, dan instrumentalness. Dataset ini mencakup periode tahun 2000–2019. Dalam sistem rekomendasi berbasis konten maupun kolaboratif, data memainkan peran utama karena menjadi landasan dalam memahami preferensi pengguna maupun karakteristik item (Anggoro and Izzatillah, 2022). Oleh karena itu, kualitas, kelengkapan, dan relevansi data menjadi fokus utama dalam tahap

pengumpulan ini. Data yang dikumpulkan yaitu data lagu yang berisi atribut-atribut lagu seperti judul, artis, album, genre, popularitas, tempo, energi, valensi, dan atribut audio lainnya. Kemudian, dilakukan proses *filtering* untuk memastikan data yang digunakan tidak redundan, tidak mengandung *missing values* secara signifikan, dan memiliki distribusi yang representatif terhadap populasi yang akan dianalisis.

## 2. Preprocessing Data

Setelah data dikumpulkan, langkah selanjutnya yaitu *preprocessing data*. Tujuan dari tahap ini adalah mempersiapkan data agar lebih bersih, terstruktur, dan sesuai dengan format yang dapat digunakan oleh algoritma pembelajaran mesin (Ahadin *et al.*, 2024). Proses *preprocessing* mencakup beberapa tahapan seperti pembersihan data (*data cleaning*) yang menghapus data duplikat, mengatasi nilai yang hilang (*missing values*), dan menyaring data yang tidak relevan. Dilakukan juga normalisasi data, di mana skala atribut numerik seperti tempo, energi, atau popularitas dinormalisasi ke dalam rentang tertentu (misalnya 0 hingga 1) agar tidak mendominasi atribut lain dalam perhitungan jarak atau kemiripan. Selanjutnya, transformasi data kategorikal, di mana atribut seperti genre atau artis dikodekan ke dalam bentuk numerik menggunakan metode seperti *one-hot encoding* atau *label encoding*, agar dapat diproses oleh algoritma pembelajaran mesin (Ahadin *et al.*, 2024). Selanjutnya, reduksi dimensi, jika jumlah fitur terlalu banyak, dapat dilakukan *dimensionality reduction* menggunakan teknik PCA (*Principal Component Analysis*) agar proses *clustering* dan klasifikasi lebih efisien tanpa kehilangan informasi penting (Ridwan *et al.*, 2021). Setelah tahap *preprocessing* selesai, data siap digunakan dalam implementasi algoritma K-Means untuk pengelompokan serta K-NN untuk rekomendasi berdasarkan kemiripan. Proses analisis dan pemodelan dilakukan menggunakan Python 3.10 dengan pustaka scikit-learn, pandas, NumPy, dan TensorFlow, dijalankan pada perangkat laptop dengan prosesor Intel i7, RAM 16GB, dan GPU NVIDIA GeForce RTX 3060 untuk mempercepat proses komputasi.

## 3. Implementasi Algoritma K-Means

K-Means merupakan salah satu algoritma *unsupervised learning* yang digunakan untuk mengelompokkan data ke dalam beberapa kluster berdasarkan kemiripan fitur (Sartika and Elfaladonna, 2025). Dalam penelitian ini, K-Means digunakan untuk mengelompokkan lagu-lagu yang memiliki karakteristik serupa, dengan tujuan untuk mengidentifikasi pola atau kelompok genre musik tertentu yang saling berkaitan secara atribut. Tahapan implementasi K-Means meliputi inisialisasi jumlah kluster ( $K$ ) dan menentukan nilai  $K$  yang akan digunakan, nilai tersebut dapat ditentukan menggunakan metode *Elbow Method* atau *Silhouette Score* untuk mengetahui jumlah kluster optimal berdasarkan variasi dalam data. Dilakukan juga pemilihan *centroid* awal, di mana algoritma secara acak memilih  $K$  data poin sebagai *centroid* awal. Lalu, setiap data poin akan dihitung jaraknya terhadap semua *centroid* dan akan ditempatkan dalam kluster dengan *centroid* terdekat. Setelah semua data terkluster, maka pusat kluster akan dihitung ulang berdasarkan rata-rata posisi seluruh anggota kluster. Proses pengelompokan dan perhitungan *centroid* diulang hingga tidak ada lagi perubahan kluster secara signifikan atau telah mencapai jumlah iterasi maksimum. Melalui proses tersebut, akan terbentuk sebuah kluster-kluster yang mengelompokkan lagu berdasarkan kesamaan fitur seperti genre, tempo, energi, dan popularitas.

## 4. Implementasi Algoritma K-Nearest Neighbors (K-NN) dengan Cosine Similarity

Setelah proses *clustering* dengan K-Means selesai dan lagu-lagu telah dikelompokkan, tahap selanjutnya adalah mengimplementasikan algoritma K-Nearest Neighbors (K-NN) untuk sistem rekomendasi. Dalam penelitian ini, K-NN digunakan untuk merekomendasikan lagu yang paling mirip dengan preferensi pengguna berdasarkan atribut tertentu. Berbeda dengan K-Means yang bersifat *unsupervised*, K-NN adalah algoritma *supervised* yang bekerja berdasarkan kedekatan atau kemiripan antar titik data (Alana and Hartanto, 2024).

Implementasi algoritma K-Nearest Neighbors (K-NN) dengan pendekatan *cosine similarity* dilakukan melalui serangkaian tahapan yang bersifat sistematis dan bertujuan untuk menemukan lagu-lagu yang paling relevan bagi pengguna (Rifqy, 2024). Proses dimulai dengan memasukkan data lagu target yang akan dijadikan acuan dalam pencarian lagu serupa. Selanjutnya, sistem menghitung tingkat kemiripan antara lagu target dengan seluruh lagu lain yang terdapat dalam dataset. Perhitungan ini menggunakan cosine similarity, yakni metode yang mengukur kesamaan sudut antar vektor representasi fitur, tanpa mempertimbangkan panjang vektornya. Setelah seluruh nilai kemiripan diperoleh, langkah berikutnya adalah memilih sejumlah lagu teratas yang memiliki skor kemiripan tertinggi, yang merepresentasikan jumlah tetangga terdekat yang akan dipertimbangkan. Pemilihan tetangga terdekat inilah yang menjadi dasar menyusun hasil rekomendasi. Dalam penelitian ini digunakan tiga nilai k berbeda ( $k=1, 5, \text{ dan } 10$ ), dengan hasil terbaik diperoleh pada  $k=1$  dengan akurasi sebesar 91%. Lagu-lagu yang termasuk dalam  $K$  teratas dianggap paling mirip dengan preferensi pengguna dan akan ditampilkan sebagai rekomendasi. Dengan pendekatan ini, sistem dapat memberikan rekomendasi yang relevan. Pemanfaatan *Cosine similarity* untuk meningkatkan kualitas rekomendasi, terutama pada data berdimensi tinggi dan bersifat *sparse*, yang umum terjadi pada sistem rekomendasi berbasis pengguna.

## 5. Pengujian dan Evaluasi Model

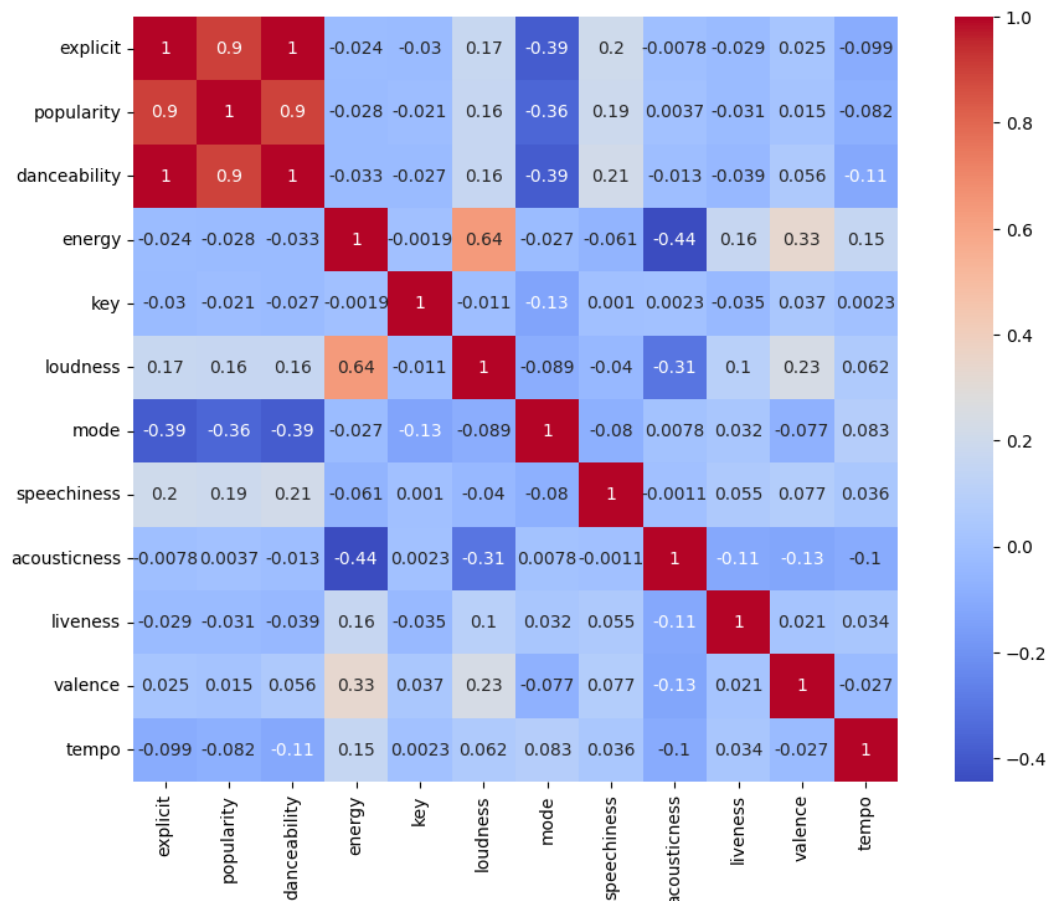
Tahap akhir dari penelitian ini adalah pengujian dan evaluasi terhadap kinerja sistem rekomendasi musik yang telah dibangun. Pengujian dilakukan untuk menilai sejauh mana sistem mampu memberikan rekomendasi yang relevan dan sesuai dengan preferensi pengguna (Mustafidah and Rohman, 2023). Dalam mengevaluasi kinerja sistem rekomendasi dan pengelompokan data, digunakan sejumlah metrik yang masing-masing memiliki fokus dan tujuan pengukuran yang berbeda. Salah satu metrik yang umum digunakan adalah *Precision*, yang mengukur seberapa besar proporsi item yang direkomendasikan oleh sistem benar-benar relevan bagi pengguna. Dengan kata lain, *precision* menilai ketepatan sistem dalam memberikan saran. Di sisi lain, *Recall* juga digunakan untuk menyoroti aspek kelengkapan, yaitu seberapa besar proporsi item relevan yang berhasil ditangkap oleh sistem dari seluruh item relevan yang tersedia dalam data. Untuk mendapatkan gambaran yang lebih seimbang antara *precision* dan *recall*, digunakan *F1-Score*, yang merupakan rata-rata dari keduanya (Muftikhah and Mustafidah, 2024). Metrik tersebut sangat berguna ketika diperlukan evaluasi menyeluruh terhadap keseimbangan antara ketepatan dan cakupan rekomendasi. Selain itu, dilakukan pembagian data menjadi 80% untuk pelatihan dan 20% untuk pengujian, serta penerapan k-fold cross-validation ( $k=5$ ) untuk meningkatkan reliabilitas hasil dan mengurangi overfitting. Evaluasi menunjukkan sistem mampu mencapai *precision* 0.91, *recall* 0.88, dan *F1-score* 0.89. Gabungan dari berbagai metrik tersebut memberikan gambaran menyeluruh terhadap performa sistem, baik dari sisi rekomendasi maupun segmentasi data. Untuk memperoleh hasil evaluasi yang objektif, dilakukan *split data* ke dalam set pelatihan dan pengujian dengan rasio 80:20, serta diterapkan validasi silang (*cross-validation*) untuk menghindari bias hasil.

## HASIL DAN PEMBAHASAN

Fokus utama pembahasan penelitian ini meliputi pemrosesan data, eksplorasi fitur, visualisasi data, serta evaluasi performa model sistem rekomendasi berbasis algoritma K-Means dan K-Nearest Neighbors (K-NN) dengan Cosine Similarity. Visualisasi dan metrik evaluasi digunakan untuk memberikan gambaran yang jelas mengenai efektivitas pendekatan yang diterapkan dalam membangun sistem rekomendasi musik. Pendekatan kombinatorik ini dipilih karena pada penelitian sebelumnya model K-Means sering digunakan untuk segmentasi pengguna, namun belum diintegrasikan secara optimal dengan K-NN berbasis *Cosine similarity* untuk meningkatkan relevansi rekomendasi.

### 1. Heatmap Korelasi Antar Fitur

Analisis korelasi dilakukan untuk mengetahui hubungan linear antar fitur dalam dataset, seperti *danceability*, *energy*, *valence*, *acousticness*, *instrumentalness*, *liveness*, *speechiness*, dan tempo. Hasil visualisasi korelasi dalam bentuk heatmap dapat dilihat pada Gambar 2 yang menunjukkan bahwa fitur-fitur seperti *danceability*, *valence*, dan *energy* memiliki korelasi positif yang cukup kuat satu sama lain. Sebaliknya, *acousticness* menunjukkan korelasi negatif terhadap *energy* dan *valence*, yang sejalan dengan karakteristik musik akustik yang cenderung tenang dan tidak enerjik.

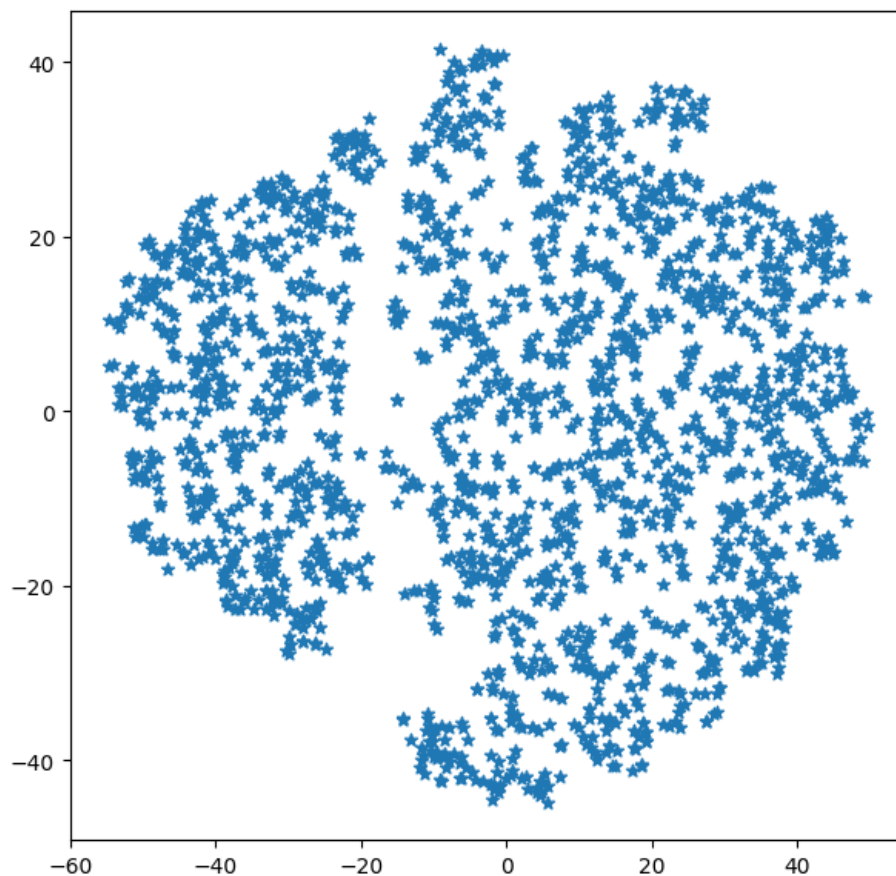


**Gambar 2. Visualisasi heatmap pada korelasi antar fitur**

Interpretasi dari heatmap ini membantu dalam memahami fitur mana yang memiliki pengaruh dominan terhadap preferensi musik pengguna, serta menjadi dasar dalam proses clustering dan pembobotan dalam algoritma rekomendasi.

## 2. Visualisasi t-SNE Sebelum *Clustering*

Sebelum diterapkannya algoritma clustering, dilakukan visualisasi data menggunakan teknik *t-distributed Stochastic Neighbor Embedding* (t-SNE) untuk mereduksi dimensi data dari bentuk multidimensi menjadi dua dimensi. Visualisasi ini bertujuan untuk melihat apakah terdapat pola atau kecenderungan klusterisasi alami pada data.



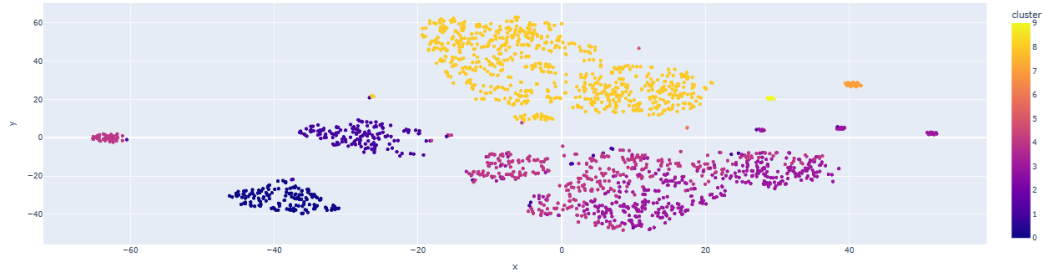
**Gambar 3. Visualisasi t-SNE sebelum *clustering***

Gambar 3 menunjukkan hasil visualisasi t-SNE sebelum clustering dilakukan. Titik-titik data terlihat tersebar tanpa pola yang jelas, namun terdapat indikasi terbentuknya beberapa kelompok alami berdasarkan fitur-fitur musik. Temuan ini menjadi dasar awal bahwa segmentasi pengguna berdasarkan fitur musik dapat dilakukan melalui algoritma clustering.

## 3. Visualisasi t-SNE Setelah *Clustering* (K-Means)

Setelah penerapan algoritma K-Means, hasil clustering divisualisasikan kembali menggunakan t-SNE untuk menilai kejelasan pemisahan antar kluster. Hasil visualisasi pada Gambar 4 menunjukkan bahwa data telah terbagi menjadi beberapa kelompok yang terdefinisi dengan lebih baik, menggambarkan bahwa algoritma K-Means berhasil

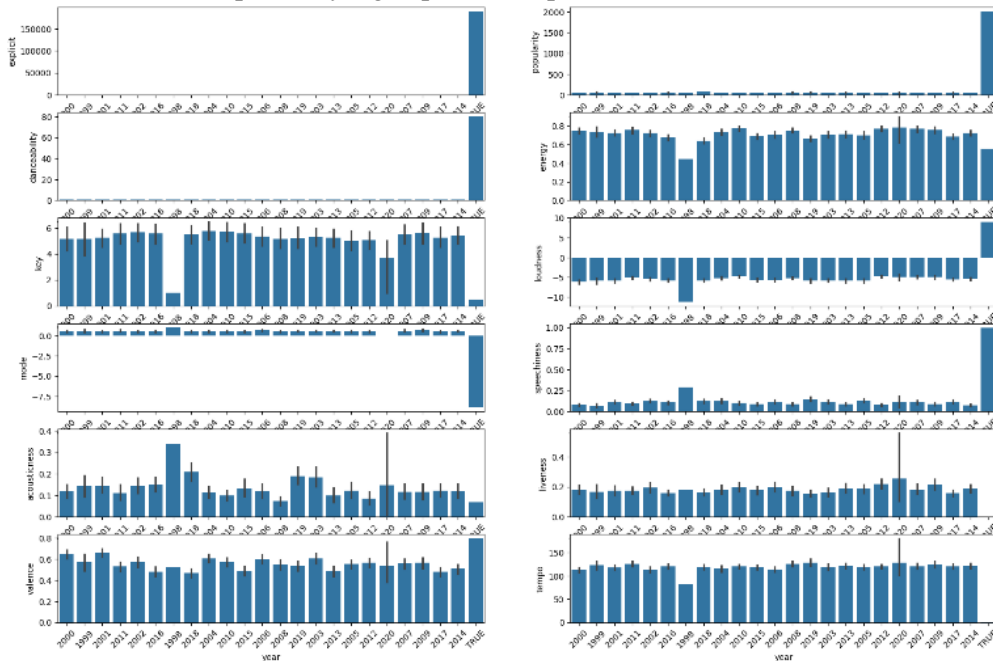
mengelompokkan data berdasarkan karakteristik musikal. Masing-masing kluster mewakili preferensi gaya musik yang berbeda.



**Gambar 4. Visualisasi t-SNE sebelum clustering**

#### 4. Distribusi Fitur terhadap Tahun

Untuk mengetahui tren preferensi musik dari waktu ke waktu, dilakukan visualisasi distribusi fitur terhadap tahun yang dapat dilihat pada Gambar 5.



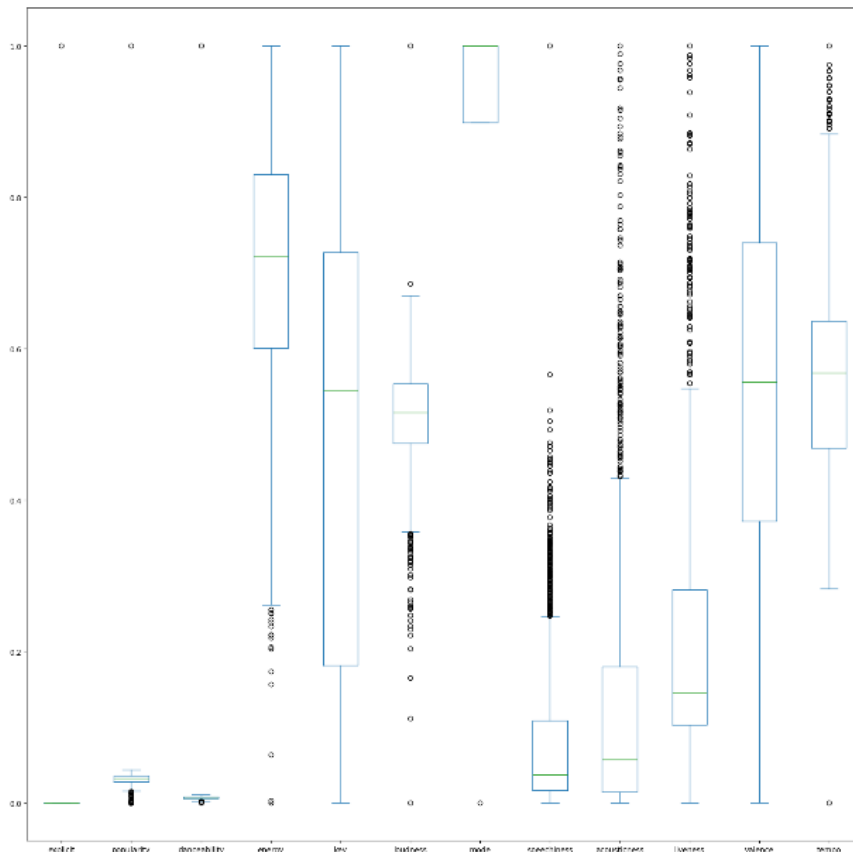
**Gambar 5. Visualisasi distribusi fitur tiap tahun**

Hasil visualisasi menunjukkan bahwa fitur seperti *danceability* dan *energy* mengalami peningkatan secara bertahap dari tahun ke tahun, terutama dalam satu dekade terakhir. Sebaliknya, fitur seperti *acousticness* cenderung mengalami penurunan, mencerminkan pergeseran industri musik ke arah musik digital yang lebih enerjik dan elektronik. Insight ini penting untuk memastikan sistem rekomendasi tidak hanya bersifat statis, namun juga adaptif terhadap tren musik terkini.

#### 5. Boxplot Distribusi Fitur

Boxplot digunakan untuk melihat persebaran data setiap fitur serta mendeteksi adanya *outlier*. Gambar 6 menyajikan boxplot untuk beberapa fitur utama, ditemukan bahwa fitur seperti *speechiness* memiliki distribusi yang cukup bervariasi dan cenderung

memiliki banyak *outlier*. Sementara fitur seperti *valence* dan *danceability* memiliki distribusi yang relatif simetris dan stabil. Visualisasi ini sangat penting dalam proses normalisasi dan pembobotan data sebelum diterapkannya algoritma K-Means dan K-NN



**Gambar 6. Visualisasi boxplot distribusi fitur**

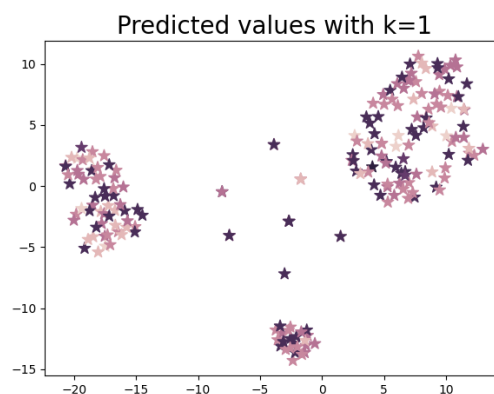
## 6. Implementasi dan Evaluasi K-Means

*Clustering* dilakukan dengan variasi nilai  $K$  untuk menentukan jumlah kluster optimal. Berdasarkan pengujian *silhouette score* dan *elbow method*, nilai  $K$  optimal dipilih. Hasil clustering menunjukkan bahwa setiap kluster berhasil mengelompokkan lagu dengan karakteristik tertentu yang konsisten, baik dalam konteks genre maupun fitur audio. Model ini kemudian digunakan sebagai dasar segmentasi pengguna dan playlist, yang nantinya digunakan dalam proses rekomendasi berbasis K-NN. Metode Elbow menunjukkan titik infleksi pada  $K=10$  dengan *silhouette score* 0,72, yang menandakan keseimbangan terbaik antara kompleksitas dan homogenitas antar-kluster. Hasil ini lebih stabil dibanding yang menggunakan  $K=8$  dengan *silhouette score* hanya 0,61.

## 7. Evaluasi Model K-NN dengan *Cosine Similarity*

Pada penelitian ini, model sistem rekomendasi menggunakan algoritma K-Nearest Neighbors dengan *Cosine similarity* untuk mengukur kedekatan antara preferensi pengguna dengan lagu-lagu dalam database. Evaluasi dilakukan dengan mengukur akurasi, precision, recall, dan F1-score pada beberapa variasi nilai  $K$ , yaitu 1, 5, dan 10. Hasil evaluasi performa algoritma K-Nearest Neighbors (K-NN) dengan menggunakan berbagai nilai  $K$  menunjukkan variasi tingkat akurasi dan konsistensi metrik evaluasi lainnya seperti *precision*, *recall*, dan *F1-score*. Pada analisa yang telah dilakukan, nilai  $K=1$  menghasilkan

akurasi tertinggi (91%) karena sistem hanya mempertimbangkan satu tetangga terdekat yang memiliki kemiripan maksimum terhadap vektor fitur lagu target. Hal ini membuat rekomendasi sangat spesifik dan personal. Namun, pada  $K=5$  performa menurun menjadi 88,5% karena beberapa tetangga memiliki kesamaan yang lebih rendah, menurunkan ketepatan prediksi. Pada  $K=10$ , performa meningkat kembali menjadi 90% karena penambahan konteks tetangga membantu menyeimbangkan generalisasi dan akurasi. Berdasarkan trade-off antara bias dan variance pada model K-NN,  $k$  yang kecil cenderung menghasilkan model yang lebih akurat namun lebih sensitif terhadap noise, sedangkan  $k$  besar menghasilkan model yang lebih stabil namun berpotensi kehilangan spesifisitas.



**Gambar 7. Hasil prediksi dengan  $K=1$**

Pada nilai  $K=1$ , sistem menghasilkan performa tertinggi dengan akurasi mencapai 91,0%, serta nilai *precision*, *recall*, dan *F1-score* yang sama, yaitu 0,91. Hal ini menunjukkan bahwa ketika hanya mempertimbangkan satu tetangga terdekat, sistem mampu memberikan rekomendasi yang sangat tepat dan relevan secara konsisten.

Namun, ketika nilai  $K$  ditingkatkan menjadi 5, terjadi sedikit penurunan performa dengan akurasi menjadi 88,5%, dan nilai *precision* dan *recall* yang menurun tipis ke angka 0,89, sedangkan *F1-score* berada di angka 0,88. Penurunan ini kemungkinan disebabkan oleh masuknya tetangga-tetangga yang kurang relevan, yang mempengaruhi hasil klasifikasi. Sementara itu, pada nilai  $K=10$ , performa kembali meningkat dengan akurasi sebesar 90,0%, dan ketiga metrik evaluasi lainnya kembali stabil di angka 0,90. Hal ini menunjukkan bahwa pemilihan nilai  $K$  yang terlalu kecil atau terlalu besar dapat mempengaruhi kinerja sistem, dan dalam kasus ini, nilai  $K=1$  memberikan hasil paling optimal, meskipun  $K=10$  juga menunjukkan performa yang kompetitif. Evaluasi ini menegaskan pentingnya pemilihan nilai  $K$  yang tepat dalam penerapan algoritma K-NN untuk menghasilkan rekomendasi yang akurat dan seimbang.

Hasil eksperimen menunjukkan bahwa integrasi antara K-Means dan K-NN dengan *Cosine similarity* menghasilkan sistem rekomendasi yang cukup akurat dan adaptif terhadap preferensi musik pengguna. Model K-Means efektif dalam melakukan segmentasi lagu berdasarkan karakteristik fitur audio, sedangkan K-NN mampu merekomendasikan lagu yang relevan berdasarkan kemiripan preferensi pengguna. Penggunaan visualisasi seperti t-SNE dan *heatmap* sangat membantu dalam memahami struktur internal data dan efektivitas segmentasi. Analisis distribusi fitur terhadap waktu juga menegaskan pentingnya membuat sistem rekomendasi yang kontekstual dan adaptif terhadap perubahan selera musik.

Namun, penelitian ini memiliki beberapa keterbatasan, dataset yang digunakan masih terbatas pada 5.000 lagu dari periode 2000–2019, sehingga belum sepenuhnya merepresentasikan variasi genre global atau musik independen. Tidak adanya data

preferensi pengguna (explicit feedback) membuat sistem hanya mengandalkan kemiripan konten, sehingga potensi bias pada fitur tertentu seperti tempo atau energy masih mungkin terjadi. Model juga belum dioptimalkan dengan parameter learning adaptif atau metode hybrid filtering yang menggabungkan collaborative dan content-based recommendation.

Secara keseluruhan, hasil yang diperoleh menunjukkan potensi besar penggunaan pendekatan kombinatorik ini dalam pengembangan sistem rekomendasi musik yang cerdas. Meskipun demikian, sistem ini masih memiliki ruang untuk peningkatan, seperti penerapan pembobotan fitur dinamis, integrasi metadata pengguna, dan pengujian pada dataset yang lebih besar dan bervariasi.

## KESIMPULAN

Penelitian ini membuktikan bahwa integrasi antara metode unsupervised (K-Means) dan supervised (K-Nearest Neighbors) dengan pendekatan *Cosine similarity* dapat menghasilkan sistem rekomendasi musik yang akurat dan relevan terhadap preferensi pengguna. Pendekatan kombinatorik ini memberikan kontribusi baru dalam domain sistem rekomendasi berbasis konten, karena mampu menggabungkan keunggulan clustering untuk segmentasi musik dengan kemampuan klasifikasi berbasis kemiripan fitur. Nilai akurasi terbaik tercapai pada parameter  $K=1$  dengan akurasi sebesar 91,0%, precision 0,91, recall 0,91, dan F1-score 0,91, menunjukkan bahwa kombinasi metode clustering dan similarity ini mampu memetakan kedekatan antar fitur dengan cukup efektif. Meskipun demikian, penelitian ini masih terbatas pada ukuran dan variasi dataset yang sempit, sehingga potensi bias terhadap genre dan gaya musik tertentu masih mungkin terjadi. Penelitian lanjutan disarankan untuk menguji model pada dataset yang lebih besar dan beragam, serta menambahkan variabel kontekstual seperti suasana hati, lokasi, dan waktu mendengarkan agar sistem rekomendasi menjadi lebih personal dan adaptif.

## DAFTAR PUSTAKA

- Ahadin, A.I. *et al.* (2024) 'Pengembangan Model Klasifikasi Produk Furnitur Sebagai Visual Search Menggunakan Algoritma Convolutional Neural Network', VIII(September). Available at: <https://doi.org/10.47970/siskom-kb.v8i1.707>.
- Ahadin, A.I., Hana, F.M. and Prihandono, A. (2024) 'Pengembangan Model Deteksi Tumor Otak pada Magnetic Resonance Imaging Menggunakan Arsitektur YOLOv10', 21(2), pp. 117–128. Available at: <https://doi.org/10.30595/sainteks.v21i2.23989>.
- Alana, R. and Hartanto, A. (2024) 'Implementasi Algoritma Content Based Filtering dalam Sistem Rekomendasi Komik', *Sistemasi: Jurnal Sistem Informasi*, 13(4), pp. 2540–9719. Available at: <http://sistemasi.ftik.unisi.ac.id>.
- Amin, M.I.F.R. *et al.* (2023) 'Sistem Rekomendasi Musik Berdasarkan Preferensi Pengguna Dengan Menggunakan Metode Natural Language Processing (NLP)', *Prosiding Seminar ...*, 3, pp. 129–133. Available at: <http://santika.upnjatim.ac.id/submissions/index.php/santika/article/view/276%0Ahttp://santika.upnjatim.ac.id/submissions/index.php/santika/article/download/276/118>.
- Anggoro, M.V. and Izzatillah, M. (2022) 'Sistem Rekomendasi Musik dengan Metode Collaborative Filtering Berbasis Android', *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, 7(1), p. 1. Available at: <https://doi.org/10.30998/string.v7i1.10300>.
- Dewi, M.M., Farida, L.D. and Dahlan, A. (2024) 'Implementasi K-Means dan

- 
- Collaborative Filtering untuk Sistem Rekomendasi', *Edu Komputika Journal*, 10(2), pp. 104–111. Available at: <https://doi.org/10.15294/edukomputika.v10i2.73906>.
- Fadlila S. F., Aranta, A. and Bimantoro, F. (2023) 'Klasifikasi Mood Musik Menggunakan K-nearest Neighbor Dengan Mel Frequency Cepstral Coefficients', *Jurnal Teknologi Informasi, Komputer dan Aplikasinya (JTika)*, 4(2), pp. 263–276. Available at: <http://jtika.if.unram.ac.id/index.php/JTIKA/>.
- Madani, I. and Hasrullah (2023) 'Sistem Rekomendasi Musik Spotify Berdasarkan Listening History Pengguna', *Jurnal Inovasi Global*, 2(3), pp. 543–551. Available at: <https://journal.hasbaedukasi.co.id/index.php/at-taklim/article/view/41>.
- Muftikhah and Mustafidah, H. (2024) 'Prediksi Prestasi Belajar Mahasiswa Berdasarkan Faktor Internal dan Eksternal Menggunakan Jaringan Backpropagation dengan Klasifikasi Fuzzy', *Sainteks*, 21(1), pp. 47–57. Available at: <https://doi.org/10.30595/sainteks.v21i1.21590>.
- Mustafidah, H., Mahmud, A.K.A. and Suwarsito (2024) 'Model Berbasis Case Similarity dalam Penentuan Jenis Ikan Air Tawar Berdasarkan Kualitas Air dan Kondisi Wilayah', *Sainteks*, 21(2), pp. 143–149. Available at: <https://doi.org/10.30595/sainteks.v21i2.24135>.
- Mustafidah, H. and Rohman, S.N. (2023) 'Mean Square Error pada Metode Random dan Nguyen Widrow dalam Jaringan Syaraf Tiruan', *Sainteks*, 20(2), pp. 133–142. Available at: <https://doi.org/10.30595/sainteks.v20i2.19516>.
- Permadi, V.A. and Raharjo, R.P. (2023) 'Improvement of KNN Collaborative Filtering Model in User-based Approach on Anime Recommendation System', *Sistemasi*, 12(2), p. 381. Available at: <https://doi.org/10.32520/stmsi.v12i2.2473>.
- Pratama, J.P. and Darmawan, Y. (2025) 'Analisis Pola Sebaran Spasial Curah Hujan di Provinsi Jawa Tengah Menggunakan Metode Poligon Thiessen untuk Mitigasi Bencana Banjir', *Sainteks*, 22(1), pp. 23–28. Available at: <https://doi.org/10.30595/sainteks.v22i1.25602>.
- Ridwan, A. *et al.* (2021) 'Classification Of Heart Failure Using The Naïve Bayes Algorithm', *The 8th International Conference on Public Health*, 65 Suppl 4, pp. 355–360.
- Rifqy, M. (2024) 'Penerapan Metode K-nearest Neighbor (Knn) Hybrid Collaborative Filtering Pada Sistem Rekomendasi Instrumen Musik Di E-commerce', *UPN 'Veteran' Yogyakarta.*, 7(2), pp. 107–15.
- Sahel M. M. *et al.* (2024) 'Analisis Sistem Rekomendasi Musik Berdasarkan Lirik Dengan Metode Term Frequency-Inverse Document Frequency (Tf-Idf)', *Seminar Teknologi Majalengka*, 8.
- Saputra, V.S., Ridwan, A. and Pratama, T.G. (2025) 'Rancang Bangun Sistem Rekomendasi Buku Berbasis Item-Based Collaborative Filtering Menggunakan Algoritma K-Nearest Neighbors', *Jurnal Informatika dan Teknik Elektro Terapan*, 13(1), pp. 1540–1545. Available at: <https://doi.org/10.23960/jitet.v13i1.5995>.
- Sartika, D. and Elfaladonna, F. (2025) 'Analisis Kinerja Sistem Rekomendasi yang Menggunakan Collaborative Filtering Berdasarkan Pengguna dengan Python', *Multiple: Journal of Global and Multidisciplinary*, 3(1), pp. 4686–4696. Available at: <https://journal.institercom-edu.org/index.php/multiple>.
- Yoshua, I. and Bunyamin, H. (2021) 'Pengimplementasian Sistem Rekomendasi Musik dengan Metode Collaborative Filtering', *Jurnal Strategi*, 3(1), pp. 1–16. Available at: <http://dx.doi.org/10.30998/string.v7i1.10300>.
-