

## Analisis Perbandingan Pearson Correlation dan Cosine Similarity pada Rekomendasi Musik berbasis Collaborative Filtering

Fatma Yuniardini<sup>1</sup>, Triyanna Widiyaningtyas<sup>1,\*</sup>

<sup>1</sup> Program Studi Teknik Informatika, Universitas Negeri Malang, Indonesia

\* Correspondence: triyannaw.ft@um.ac.id

**Copyright:** © 2024 by the authors

Received: 12 Oktober 2024 | Revised: 18 Oktober 2024 | Accepted: 5 November 2024 | Published: 19 Desember 2024

### Abstrak

Kemajuan teknologi digital telah merevolusi dunia musik, membuat akses terhadap berbagai *genre* dan musisi menjadi lebih mudah dan tak terbatas, namun pengguna tetap kesulitan menemukan musik yang sesuai selera mereka. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja metode *pearson correlation* dan *cosine similarity* dalam rekomendasi musik yang dipersonalisasi berbasis *collaborative filtering*, dengan fokus pada *item-based filtering*, diukur menggunakan *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE). Dataset yang digunakan adalah *rating* musik metal dari Amazon yang bersifat publik dan diambil dari Kaggle, dengan total 19.065 sampel, serta menggunakan algoritma *k-Nearest Neighbors* (KNN) untuk prediksi rekomendasi. Tahapan penelitian meliputi pengumpulan data, *pre-processing* untuk menangani *missing value*, duplikasi, normalisasi, dan pengecekan *outlier*, dilanjutkan dengan prediksi menggunakan algoritma KNN, dan pengukuran akurasi menggunakan MAE serta RMSE. Evaluasi menunjukkan bahwa *pearson correlation* menghasilkan MAE sebesar 0,066538 dan RMSE sebesar 0,086698, sementara *cosine similarity* menghasilkan MAE sebesar 0,066559 dan RMSE sebesar 0,086709. Hasil ini menunjukkan bahwa *pearson correlation* lebih efektif dalam menangkap hubungan linear pada data *rating*, sehingga rekomendasi yang dihasilkan lebih relevan dan sesuai dengan preferensi pengguna. *Pearson correlation* mampu mempertimbangkan variabilitas *rating* setiap pengguna, membuat rekomendasi lebih akurat dalam menyesuaikan pola penilaian individu.

**Kata kunci:** *cosine similarity; collaborative filtering; pearson correlation; rekomendasi musik*

### Abstract

*Advances in digital technology have revolutionized the world of music, making access to various genres and musicians easier and unlimited, but users still have difficulty finding music that suits their tastes. This research aims to analyze and compare the performance of the pearson correlation and cosine similarity methods on personal music recommendations based on Collaborative Filtering, with a focus on Item-Based Filtering, measured using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The dataset utilized comprises public metal music ratings from Amazon, sourced from Kaggle, totaling 19,065 samples. The k-Nearest Neighbors (KNN) algorithm was employed for recommendation prediction. The research steps included data collection, pre-processing to address missing values, duplicates, normalization, and outlier detection, followed by prediction using the KNN algorithm, and accuracy measurement using MAE and RMSE. Evaluation results indicated that Pearson Correlation produced an MAE of 0.066538 and an RMSE of 0.086698, while cosine similarity yielded an MAE of 0.066559 and an RMSE of 0.086709. These findings suggest that pearson correlation is more effective in capturing linear relationships within the rating data, leading to recommendations that are more relevant and aligned with user preferences. Pearson correlation considers the variability in each user's ratings, resulting in more accurate recommendations that align with individual rating patterns.*

**Keywords:** *cosine similarity; collaborative filtering; pearson correlation; music recommendations*



## PENDAHULUAN

Kemajuan teknologi memainkan peran penting dalam berbagai aspek kehidupan. Terlebih lagi (Mustaqim et al., 2024), kemajuan teknologi di era digital telah merevolusi dunia music, menghadirkan era baru di mana akses terhadap berbagai genre dan musisi menjadi semakin mudah dan tak terbatas. Terlebih lagi generasi digital lebih cenderung mendengarkan musik secara online melalui perangkat digital mereka (Yollis & Netti, 2018). Perubahan ini mengubah cara kita menikmati musik, terutama melalui platform streaming seperti Spotify, Apple Music, dan YouTube, yang memungkinkan pengguna untuk mengakses jutaan lagu hanya dalam hitungan detik. Platform-platform ini tidak hanya memudahkan akses, tetapi juga memanfaatkan algoritma canggih untuk merekomendasikan musik yang relevan bagi setiap pendengar. Menurut (Putra & Santika, 2020), kemudahan ini memberi kesempatan bagi pendengar untuk menemukan musik yang sesuai dengan selera pribadi, yang dapat memberikan ketenangan bagi hati dan pikiran. Namun, beragamnya pilihan musik di era modern sering kali membuat pendengar bingung dalam menentukan lagu yang sesuai dengan suasana hati dan selera mereka. Masalah ini menciptakan kebutuhan akan rekomendasi musik yang efektif, yang dapat memberikan rekomendasi produk yang sesuai dengan preferensi pengguna (Suharya et al., 2021).

Rekomendasi musik bekerja dengan menganalisis data musik dan preferensi pendengar, lalu memberikan rekomendasi musik yang dipersonalisasi. Tujuan utama rekomendasi ini adalah memprediksi preferensi pengguna atau menyarankan *item* tertentu dari berbagai pilihan yang sesuai dengan minat mereka (Afoudi et al., 2021). Dengan menganalisis preferensi pengguna secara mendetail, rekomendasi dapat mengidentifikasi *item* yang serupa, sehingga memberikan pengalaman yang lebih personal (Munawar et al., 2021).

Rekomendasi banyak digunakan untuk membuat prediksi pada berbagai produk, seperti buku, musik, film, tempat wisata, dan lain-lain (Jaja et al., 2020). Metode yang umum digunakan dalam rekomendasi meliputi *collaborative filtering*, *content-based filtering*, dan *hybrid* (Muliawan et al., 2022). *Collaborative filtering* sendiri terbagi menjadi dua pendekatan: *user-based collaborative filtering*, yang mencari pengguna lain dengan preferensi serupa untuk merekomendasikan *item*, dan *item-based collaborative filtering*, yang mencari hubungan antar *item* yang diminati pengguna lain dalam rekomendasi (Muliadi & Lestari, 2019). Selain itu, (Indriawan et al., 2020) menyatakan bahwa menggunakan metode *item-based collaborative filtering* dapat menghasilkan rekomendasi yang lebih akurat. Keunggulan dari *item-based collaborative filtering* adalah tidak bergantung pada konten yang ada, sehingga rekomendasi kepada setiap pelanggan dapat bervariasi berdasarkan *rating* yang diberikan oleh pelanggan terhadap produk tersebut (Geetha et al., 2018). Selain itu, metode ini menjadi solusi untuk sejumlah permasalahan dalam *user-based collaborative filtering*, khususnya terkait keterbatasan (*sparsity*), skalabilitas, serta tantangan dalam waktu dan penggunaan memori (Agustian & Nugroho, 2020). Metode ini juga lebih efektif digunakan pada data yang memiliki perbedaan jumlah *item* yang signifikan antara satu pengguna dan pengguna lainnya (Fernanto et al., 2019). Dengan pendekatan ini, dapat menghasilkan rekomendasi yang lebih akurat berdasarkan peringkat yang diberikan pengguna terhadap beberapa *item* (Widiyaningtyas et al., 2021).

Pada konteks *item-based collaborative filtering*, pola penilaian terhadap suatu *item* digunakan untuk memprediksi penilaian pengguna terhadap *item* lain (Hartatik et al., 2021). Hubungan antara *item* yang diminati oleh pengguna diidentifikasi, dan *item-item* yang memiliki korelasi kuat akan direkomendasikan (Februariyanti et al., 2021). Kemiripan antar *item* biasanya dihitung menggunakan metrik seperti *cosine similarity* atau *pearson correlation*.

Penelitian yang relevan dengan penerapan metode ini, diantaranya, penelitian oleh (Anggoro & Izzatillah, 2022) menggunakan *cosine similarity* dalam aplikasi pemutar musik online untuk menampilkan rekomendasi lagu yang lebih personal. Hasilnya, penelitian tersebut

mendapatkan skor MAE sebesar 0,0964 dan RMSE sebesar 0,0247. Penelitian lain oleh (Puspita et al., 2021a) menerapkan beberapa metode pengukuran kemiripan seperti *cosine similarity* dan *pearson correlation coefficient*. Prediksi dilakukan menggunakan algoritma KNN, Hasil evaluasi untuk *cosine similarity* menunjukkan nilai RMSE sebesar 0,9677 dan MAE sebesar 0,5998, sedangkan untuk *pearson correlation* menunjukkan nilai RMSE sebesar 0,6360 MAE sebesar 0,9010, dengan hasil tersebut menunjukkan metode ini memiliki akurasi yang baik dalam memberikan rekomendasi yang sesuai dengan preferensi pengguna.

*Cosine similarity* merupakan metrik yang umum digunakan dalam rekomendasi untuk menghitung kemiripan vektor fitur antara dua *item* (Mahendra et al., 2024). Metode ini mengukur seberapa mirip dua titik dengan menghitung sudut yang terbentuk terhadap koordinat (0,0), sehingga memberikan gambaran yang jelas mengenai hubungan antara *item* dalam konteks rekomendasi. Di sisi lain, *pearson correlation* digunakan untuk mengukur kedekatan antar pengguna berdasarkan penilaian mereka (Dewi, 2022) *pearson correlation* juga meningkatkan akurasi ketika diterapkan pada dataset berukuran besar (Al-Bakri & Hashim, 2018).

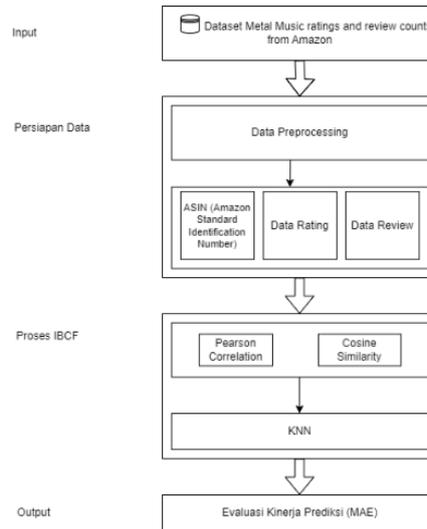
Seiring berkembangnya metode-metode ini, rekomendasi musik yang dipersonalisasi menjadi kebutuhan mendesak untuk membantu pengguna menemukan lagu-lagu yang relevan dan menarik. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja model rekomendasi musik menggunakan metode *collaborative filtering*, khususnya pendekatan *item-based filtering*. Meskipun berbagai metode telah diterapkan sebelumnya, masih terdapat ruang untuk pengembangan dalam meningkatkan akurasi dan relevansi rekomendasi. Oleh karena itu, tujuan penelitian ini adalah untuk menilai efektivitas rekomendasi dalam membantu pengguna menemukan lagu-lagu yang sesuai dengan minat mereka, dengan menggunakan algoritma *pearson correlation* dan *cosine similarity*. Penelitian ini juga berupaya meningkatkan akurasi rekomendasi musik yang dipersonalisasi, yang diukur menggunakan MAE dan RMSE. Dataset yang digunakan, yaitu *metal music ratings and review counts from amazon*, memberikan konteks yang relevan untuk analisis. Akurasi dari rekomendasi ini akan diukur menggunakan metrik MAE dan RMSE. Dengan fokus pada peningkatan metode yang belum banyak diterapkan dalam penelitian sebelumnya, diharapkan dapat memberikan rekomendasi yang lebih akurat dan memuaskan bagi pengguna.

## METODE

Penelitian ini bertujuan untuk membandingkan dua metode similaritas, yaitu *pearson correlation* dan *cosine similarity*, dalam konteks rekomendasi musik. Analisis dilakukan menggunakan dataset "*Metal Music Ratings and Review Counts*" dari Amazon, yang terdiri dari penilaian dan ulasan berbagai subkategori musik metal. Proses penelitian dijelaskan melalui tahapan *Item-Based Collaborative Filtering* (IBCF) yang disajikan pada gambar 1.

Tahap pertama adalah pengumpulan dataset, yang berisi informasi tentang artis, ASIN (Nomor Identifikasi Standar Amazon), media, jumlah ulasan, *rating* bintang, judul lagu, dan tahun. Dataset ini, yang dikumpulkan oleh Patrick Klein, mencakup total 37.748 sampel yang tersebar dalam delapan file berdasarkan subgenre musik metal, seperti *alternative\_metal.csv*, *black\_metal.csv*, dan lain-lain. Dalam penelitian ini, fokus diarahkan pada 19.065 sampel dari tiga file: *alternative\_metal.csv*, *metal.csv*, dan *pop\_metal.csv*.

Tahap kedua adalah persiapan data, termasuk pemrosesan data untuk memastikan kualitas data yang akan dianalisis. Proses ini melibatkan penghapusan nilai yang hilang, penghapusan data duplikat, normalisasi *rating* menjadi skala 0-1, serta pemeriksaan dan penghapusan outlier. Langkah-langkah ini penting agar data siap untuk perhitungan similaritas. IBCF dimulai dengan menghitung korelasi antara *item* menggunakan algoritma *pearson correlation* dan *cosine similarity*. Perhitungan *cosine similarity* menggunakan persamaan 1.



Gambar 1. Metode IBCF

$$C(i, j) = \frac{\sum_{u \in U_{ij}} r_{ui} \cdot r_{uj}}{\sqrt{\sum_{u \in U_{ij}} r_{ui}^2} \cdot \sqrt{\sum_{u \in U_{ij}} r_{uj}^2}} \quad (1)$$

Pada persamaan 1 digunakan untuk menghitung kemiripan antara dua item  $i$  dan  $j$  menggunakan *Cosine Similarity*. Bagian atas,  $\sum_{u \in U_{ij}} r_{ui} \cdot r_{uj}$  adalah jumlah perkalian *rating* yang diberikan oleh pengguna terhadap kedua item tersebut. Bagian bawah terdiri dari norma vektor *rating* masing-masing item:  $\sqrt{\sum_{u \in U_{ij}} r_{ui}^2}$  untuk item  $i$  dan  $\sqrt{\sum_{u \in U_{ij}} r_{uj}^2}$  untuk item  $j$ . Hasil akhir menunjukkan tingkat kemiripan dalam rentang 0 hingga 1, di mana nilai lebih tinggi berarti kedua *item* lebih mirip.

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}} \quad (2)$$

Pada persamaan 2 adalah untuk melakukan perhitungan *pearson correlation*, dimana  $r$  digunakan untuk mengukur derajat hubungan antara dua variabel  $X$  dan  $Y$ . Selain itu,  $n$  mencerminkan jumlah pasangan data atau pengamatan yang dianalisis. Komponen  $\sum XY$  adalah total hasil perkalian setiap pasangan nilai  $X$  dan  $Y$ , sementara  $(\sum X)$  dan  $(\sum Y)$  adalah jumlah nilai masing-masing variabel. Terakhir,  $\sum X^2$  dan  $\sum Y^2$  adalah jumlah kuadrat dari nilai-nilai variabel, yang membantu dalam menghitung variansi dan mendukung analisis hubungan antara variabel. Semua komponen ini berkontribusi pada pemahaman hubungan antara dua variabel dalam data.

Setelah menghitung kemiripan antar *item*, selanjutnya dilakukan perhitungan prediksi menggunakan algoritma KNN untuk menentukan rekomendasi yang relevan bagi pengguna, berdasarkan *item* yang telah dinilai sebelumnya. Rekomendasi yang dihasilkan dievaluasi menggunakan MAE dan RMSE untuk mengukur akurasi dan efektivitasnya. MAE memberikan ukuran seberapa baik rekomendasi sesuai dengan *rating* aktual pengguna, sementara RMSE lebih sensitif terhadap kesalahan besar. Hasil penelitian ini mencakup rekomendasi musik yang relevan untuk pengguna yang dihasilkan melalui proses IBCF dengan

menggunakan Algoritma *pearson correlation* dan *cosine similarity*, serta evaluasi keakuratan rekomendasi yang diukur dengan MAE dan RMSE.

## HASIL DAN PEMBAHASAN

### Hasil

Penelitian ini dilakukan menggunakan dataset yang digunakan adalah *metal music ratings and review counts from amazon*, yang dipublikasikan oleh Patrick Klein di platform Kaggle. Dataset ini berisi informasi mengenai *rating* dan ulasan musik metal dari berbagai subgenre, seperti *alternative\_metal.csv*, *black\_metal.csv*, *british\_metal.csv*, *death\_metal.csv*, *metal.csv*, *pop\_metal.csv*, *progressiv\_metal.csv*, dan *thrash\_and\_speed\_met.csv*. Dari keseluruhan dataset yang tersedia, penelitian ini hanya memfokuskan pada tiga subkategori: *alternative\_metal.csv*, *metal.csv*, dan *pop\_metal.csv*, karena tiga genre tersebut termasuk banyak didengarkan oleh pendengar umum dengan total 19.065 sampel. Tabel 1 menampilkan kolom yang digunakan untuk analisis, yaitu *asin* (*Amazon Standard Identification Number*), *review\_count* (jumlah *review* dari satu lagu), dan *star\_rating* (rata-rata *rating* pada lagu yang *direview*). Kolom *asin* (*Amazon Standard Identification Number*), *review\_count*, dan *star\_rating* dipilih karena mereka berperan penting dalam analisis tingkat popularitas lagu berdasarkan ulasan pengguna. Kolom-kolom tersebut memberikan informasi yang relevan untuk mengidentifikasi lagu mana yang mendapatkan perhatian dan evaluasi lebih tinggi dari pengguna.

Pengambilan sampel dilakukan secara acak menggunakan fungsi *sample* dari *library pandas*, proses pengambilan sampel dilakukan secara acak, artinya setiap data memiliki peluang yang sama untuk dipilih, tanpa adanya pengaruh atau preferensi tertentu. Untuk melakukannya, digunakan fungsi sampel yang merupakan salah satu fungsi dari *library pandas*. Fungsi ini membantu memilih sejumlah data secara acak dari kumpulan data yang lebih besar. Tujuannya adalah untuk memastikan bahwa sampel yang dipilih dapat mewakili keseluruhan data dengan adil, sehingga hasil analisis nantinya tidak dipengar. Pada penelitian ini pula menggunakan dua *library* yaitu, *library pandas* dan *library surprise* untuk menjalankan implementasi algoritma KNN.

**Tabel 1.** Kolom dataset yang digunakan

No	<i>asin</i>	<i>review_count</i>	<i>star_rating</i>
1	B000A14OI8	136,0	4,9
2	B08KMFYQMV	134,0	4,9
3	B09JVM31QY	130,0	4,9
4	B09RKMZDKR	126,0	4,9
5	B00190KZBO	125,0	4,9

Tahap selanjutnya adalah *pre-processing*, kami melakukan beberapa langkah untuk membersihkan dan mempersiapkan data sebelum digunakan dalam analisis. Pada Tabel 2 merupakan kolom dataset yang telah dilakukan *pre-processing*. *Pre-processing* yang telah dilakukan berupa menghilangkan missing value, duplikat data, mengecek dan menghilangkan outlier serta normalisasi data *rating*.

Setelah tahap *pre-processing* selesai, dilakukan tahapan pemodelan. Pada tahap ini, algoritma KNN digunakan untuk memprediksi kesamaan antar *item*. Nilai *k*, yang merepresentasikan jumlah tetangga terdekat yang dipertimbangkan, ditetapkan secara eksperimen untuk menemukan hasil prediksi yang optimal. Metode penentuan tetangga terdekat dilakukan dengan menghitung similaritas antar *item* menggunakan dua pendekatan: *cosine similarity* dan *pearson correlation*. *Cosine similarity* digunakan untuk mengukur sudut antara dua vektor dalam ruang multi-dimensional, sementara *pearson correlation* menghitung

hubungan linier antar variabel. Kedua metode ini memungkinkan identifikasi *item* yang paling mirip untuk digunakan dalam prediksi lebih lanjut.

**Tabel 2.** Kolom dataset yang setelah dilakukan *pre-processing*

No	asin	review_count	star_rating
1	B000A14OI8	136,0	4,9
2	B08KMFYQMV	134,0	4,9
3	B09JVM31QY	130,0	4,9
4	B09RKMZDKR	126,0	4,9
5	B00190KZBO	125,0	4,9

Teknik *cross-validation* juga diterapkan untuk meminimalkan risiko *overfitting*, sehingga model dapat bekerja lebih baik saat dihadapkan dengan data baru. Penelitian ini menerapkan *5-fold cross validation* untuk memperoleh estimasi akurasi yang lebih akurat dan minim bias. Dalam *k-fold cross validation*, model dilatih menggunakan sebagian *data-training set* dan diuji menggunakan sebagian lain-*testing set*. Dalam *5-fold cross validation*, data dibagi dalam 5 bagian atau *fold* dengan ukuran yang sama, sehingga tersedia 5 subset data untuk mengevaluasi kinerja model. Proses pengujian ini dilakukan berulang sebanyak 5 kali, masing-masing dengan *fold* yang berbeda sebagai data uji. *Cosine similarity* menghitung kemiripan antar *item* berdasarkan sudut di antara vektor *rating* yang diberikan pengguna, sedangkan *pearson correlation* digunakan untuk menemukan pola *rating* serupa antara *item* yang berbeda, memungkinkan rekomendasi yang lebih akurat dan relevan.

```

# Menghitung Similaritas menggunakan Cosine Similarity
sim_options = {
    'name': 'cosine', # Menggunakan Cosine Similarity
    'user_based': False # Similaritas antar item
}

# Menggunakan KNN
algo = KNNBasic(sim_options=sim_options)

#perform cross validation
cv_results = cross_validate(algo, data, measures=['RMSE', 'MAE'], cv=5, verbose=True)
results.append({
    'Algorithm': 'KNNBasic (cosine)',
    'MAE': cv_results['test_mae'].mean(),
    'RMSE': cv_results['test_rmse'].mean()
})

```

**Gambar 2.** Proses pemodelan menggunakan *cosine similarity*

```

[19] # Menghitung Similaritas menggunakan Pearson Correlation
sim_options = {
    'name': 'pearson', # Menggunakan Pearson Correlation
    'user_based': False # Similaritas antar item
}

# Menggunakan KNN
algo = KNNBasic(sim_options=sim_options)

#perform cross validation
cv_results = cross_validate(algo, data, measures=['RMSE', 'MAE'], cv=5, verbose=True)
results.append({
    'Algorithm': 'KNNBasic (pearson)',
    'MAE': cv_results['test_mae'].mean(),
    'RMSE': cv_results['test_rmse'].mean()
})

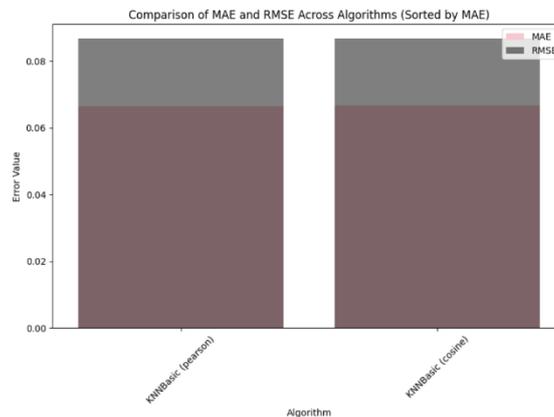
```

**Gambar 3.** Proses pemodelan menggunakan *pearson correlation*

Pada gambar 2, *cosine similarity* digunakan untuk menghitung tingkat kemiripan antar *item* dalam rekomendasi berbasis *collaborative filtering*. *Cosine Similarity* mengukur seberapa mirip dua *item* berdasarkan pola *rating* yang diberikan oleh pengguna. Dengan menggunakan algoritme KNN dan *cosine similarity*, rekomendasi dapat menemukan *item-item* yang paling mirip dengan *item* target, yang disebut sebagai tetangga terdekat (*nearest neighbors*). Selanjutnya, dapat merekomendasikan *item* kepada pengguna berdasarkan kemiripan antar

*item*, di mana *item* yang memiliki pola *rating* serupa dengan *item* yang sudah dinilai tinggi oleh pengguna akan direkomendasikan. Dengan pendekatan ini, model dapat menghasilkan rekomendasi *item* yang relevan dengan preferensi pengguna, berdasarkan perhitungan *Cosine Similarity* yang digunakan untuk menemukan *item* dengan kemiripan terbesar.

Selain *cosine similarity*, gambar 3 menunjukkan kode yang menggunakan *pearson correlation*. Dalam penelitian ini, *pearson correlation* digunakan untuk menghitung tingkat kemiripan antar *item* dalam rekomendasi. Dengan memanfaatkan algoritme k-NN, *pearson correlation* membantu menemukan *item-item* yang memiliki pola *rating* serupa dengan *item* target. Hal ini memungkinkan model untuk merekomendasikan *item* yang lebih relevan dengan preferensi pengguna berdasarkan korelasi statistik dari *rating* yang diberikan sebelumnya.



**Gambar 4.** Visualisasi hasil

Seperti yang ditunjukkan pada gambar 4 merupakan hasil performa dari metode *cosine similarity* dan *pearson correlation* yang hampir setara dalam mengukur kesesuaian prediksi. Namun, *pearson correlation* sedikit lebih unggul dibandingkan *cosine similarity*, terutama dalam hal tingkat kesalahan prediksi yang lebih rendah. Pengukuran performa dilakukan dengan menggunakan dua metrik utama, yakni MAE dan RMSE. Berdasarkan hasil pengukuran, *pearson correlation* menghasilkan MAE sebesar 0,066538 dan RMSE sebesar 0,086698, sementara *cosine similarity* menghasilkan MAE sebesar 0,066559 dan RMSE sebesar 0,086709. Meskipun perbedaan antara kedua metode sangat kecil, *pearson correlation* menunjukkan kemampuan prediksi yang lebih akurat pada dataset ini. Hasil akurasi dari kedua metode ini, termasuk nilai MAE dan RMSE masing-masing.

## Pembahasan

Penelitian ini membandingkan performa *pearson correlation* dan *cosine similarity* dalam rekomendasi musik berbasis *collaborative filtering* menggunakan dataset *Metal Music Ratings and Review Counts from Amazon*. Dataset ini difokuskan pada tiga subkategori: *alternative\_metal.csv*, *metal.csv*, dan *pop\_metal.csv*, dengan total 19.065 sampel yang berisi kolom *asin*, *review\_count*, dan *star\_rating*. Setelah melalui proses *pre-processing*, yang mencakup penghapusan *missing values*, duplikasi, dan normalisasi *rating*, data siap digunakan untuk pemodelan.

Pada rekomendasi berbasis *collaborative filtering*, *cosine similarity* menghitung kemiripan antar *item* berdasarkan sudut vektor, sementara *pearson correlation* menghitung korelasi linier antar *rating* pengguna. Keduanya dievaluasi menggunakan metode *5-fold cross validation* untuk mengurangi bias. Berdasarkan hasil penelitian, *pearson correlation* sedikit lebih unggul dengan MAE sebesar 0,066538 dan RMSE sebesar 0,086698 dibandingkan *cosine similarity*, yang memiliki MAE 0,066559 dan RMSE 0,086709. Meskipun perbedaannya kecil, *pearson correlation* lebih efektif dalam menangkap pola *rating*, terutama dalam dataset yang

bervariasi. Hal ini terjadi karena *pearson correlation* mampu mengabaikan perbedaan absolut dalam skala rating pengguna, sehingga pengukurannya lebih stabil dan adil.

Hasil penelitian ini dibandingkan dengan penelitian sebelumnya oleh Anggoro & Izzatillah (2022), yang hanya menggunakan *cosine similarity*. Dalam penelitian mereka, *cosine similarity* diterapkan pada aplikasi pemutar musik online untuk memberikan rekomendasi lagu yang lebih personal, dengan hasil MAE sebesar 0,0964 dan RMSE sebesar 0,0247. Penelitian ini menunjukkan bahwa pendekatan *pearson correlation* memberikan akurasi prediksi yang lebih baik dibandingkan *cosine similarity* dalam konteks rekomendasi musik.

Penelitian lain oleh Puspita et al. (2021) juga mendukung temuan ini dengan menunjukkan bahwa *pearson correlation* lebih efektif dalam beberapa konteks. Penelitian tersebut membandingkan berbagai metode pengukuran kemiripan, termasuk *cosine similarity* dan *pearson correlation*, menggunakan algoritma KNN. Evaluasi menunjukkan bahwa *cosine similarity* memiliki nilai RMSE sebesar 0,9677 dan MAE sebesar 0,5998, sedangkan *pearson correlation* memiliki nilai RMSE sebesar 0,6360 dan MAE sebesar 0,9010. Temuan ini mengindikasikan bahwa *pearson correlation* lebih akurat dalam memberikan rekomendasi yang sesuai dengan preferensi pengguna dibandingkan *cosine similarity*.

Penelitian ini memberikan kontribusi baru dengan secara langsung membandingkan *pearson correlation* dan *cosine similarity* dalam konteks rekomendasi musik, sesuatu yang belum dilakukan dalam penelitian sebelumnya. Temuan ini berpotensi untuk diterapkan pada rekomendasi lainnya, seperti film atau buku, dan diharapkan dapat meningkatkan kualitas rekomendasi berbasis preferensi pengguna.

Secara rinci, *pearson correlation* dan *cosine similarity* memiliki perbedaan mendasar dalam pendekatan kemiripan. *pearson correlation* unggul karena menghitung deviasi dari rata-rata, sehingga lebih baik dalam menangkap pola preferensi relatif pengguna, sementara *Cosine Similarity* hanya mempertimbangkan kesamaan sudut antar vektor *rating* tanpa memperhitungkan skala *rating* individu. Perbedaan ini memberi *pearson correlation* keunggulan dalam menangkap pola rating yang lebih variatif dan linier, yang krusial dalam rekomendasi berbasis collaborative filtering dengan menggunakan algoritma KNN. Dengan demikian, penelitian ini menambah wawasan mengenai efektivitas *Pearson Correlation* dalam meningkatkan akurasi rekomendasi, menciptakan nilai kebaruan dalam bidang rekomendasi musik berbasis *collaborative filtering* yang lebih personal dan adaptif.

## SIMPULAN

Hasil temuan kami menunjukkan bahwa metode *pearson correlation* sedikit lebih unggul dibandingkan *cosine similarity* dalam konteks rekomendasi musik menggunakan *collaborative filtering*. Hasil analisis menunjukkan bahwa *pearson correlation* menghasilkan MAE sebesar 0,066538 dan RMSE sebesar 0,086698, sedangkan *cosine similarity* mencatat MAE 0,066559 dan RMSE 0,086709. Meskipun perbedaan nilai antara kedua metode ini sangat kecil, temuan ini memberikan wawasan berharga bagi peneliti lain untuk menciptakan rekomendasi yang lebih akurat, sehingga dapat meningkatkan pengalaman pengguna dalam menemukan musik yang sesuai dengan preferensi mereka. Implikasi dari hasil penelitian ini sangat penting dalam industri musik, di mana pemilihan metode similaritas yang lebih efektif dapat mempengaruhi kualitas rekomendasi yang diberikan kepada pendengar. Temuan ini juga membuka peluang untuk menerapkan pendekatan yang sama dalam rekomendasi lain, seperti film atau buku. Namun, penelitian ini memiliki keterbatasan tertentu, termasuk ukuran dataset dan cakupan genre musik yang mungkin memengaruhi hasil. Oleh karena itu, perlu dilakukan penelitian lebih lanjut dengan dataset yang lebih besar dan beragam untuk memperkuat temuan ini.

**REFERENSI**

- Afoudi, Y., Lazaar, M., & Al Achhab, M. (2021). Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network. *Simulation Modelling Practice and Theory*, 113, 102375. <https://doi.org/10.1016/j.simpat.2021.102375>
- Agustian, E. R., & Nugroho, E. P. (2020). Sistem Rekomendasi Film Menggunakan Metode Collaborative Filtering dan K-Nearest Neighbors. *JATIKOM: Jurnal Aplikasi dan Teori Ilmu Komputer*, 3(1), 18-21.
- Al-Bakri, N. F., & Hashim, S. H. (2018). Reducing Data Sparsity in Recommender Systems. *Journal of Al-Nahrain University Science*, 21(2), 138–147. <https://doi.org/10.22401/JNUS.21.2.20>
- Anggoro, M. V., & Izzatillah, M. (2022). Sistem Rekomendasi Musik Dengan Metode Collaborative Filtering Berbasis Android. *Satuan Tulisan Riset Dan Inovasi Teknologi*, 7(1), 1–8.
- Dewi, M. M. (2022). Optimasi Pearson Correlation untuk Sistem Rekomendasi menggunakan Algoritma Firefly. *Jurnal Informatika*, 9(1), 1–5. <https://dx.doi.org/10.31294/inf.v9i1.10209>
- Februariyanti, H., Dwi Laksono, A., Sasongko Wibowo, J., & Siswo Utomo, M. (2021). Implementasi Metode Collaborative Filtering Untuk Sistem Rekomendasi Penjualan Pada Toko Mebel. *Jurnal Khatulistiwa Informatika*, 9(1), 43-50.
- Fernanto, G. F., Intan, R., & Rostianingsih, S. (2019). Sistem rekomendasi mata kuliah pilihan menggunakan metode user based collaborative filtering berbasis algoritma adjusted cosine similarity. *Jurnal Infra*, 7(1), 39-45.
- Geetha, G., Safa, M., Fancy, C., & Saranya, D. (2018). A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System. *Journal of Physics: Conference Series*, 1000, 012101. <https://doi.org/10.1088/1742-6596/1000/1/012101>
- Hartatik, H., Nurhayati, S. D., & Widayani, W. (2021). Sistem Rekomendasi Wisata Kuliner di Yogyakarta dengan Metode Item-Based Collaborative Filtering. *Journal Automation Computer Information System*, 1(2), 55–63. <https://doi.org/10.47134/jacis.v1i2.8>
- Indriawan, W., Irham Gufroni, A., & Informatika Fakultas Teknik Universitas Siliwangi Tasikmalaya, J. (2020). Sistem Rekomendasi Penjualan Produk Pertanian Menggunakan Metode Item Based Collaborative Filtering. *Jurnal Siliwangi*, 6(2), 53-59.
- Jaja, V. L., Susanto, B., & Sasongko, L. R. (2020). Penerapan Metode Item-Based Collaborative Filtering Untuk Sistem Rekomendasi Data MovieLens. *D'CARTESIAN*, 9(2), 78–83. <https://doi.org/10.35799/dc.9.2.2020.28274>
- Mahendra, R. R., Anggraeny, F. T., & Wahanani, H. E. (2024). Implementasi Item-Based Collaborative Filtering Untuk Rekomendasi Film. *Repeater: Publikasi Teknik Informatika Dan Jaringan*, 2(3), 213–221. <https://doi.org/10.62951/repeater.v2i3.140>
- Muliadi, K. H., & Lestari, C. C. (2019). Rancang Bangun Sistem Rekomendasi Tempat Makan Menggunakan Algoritma Typicality Based Collaborative Filtering. *Techno.Com*, 18(4), 275–287. <https://doi.org/10.33633/tc.v18i4.2515>
- Muliawan, A., Badriyah, T., & Syarif, I. (2022). Membangun Sistem Rekomendasi Hotel dengan Content Based Filtering Menggunakan K-Nearest Neighbor dan Haversine Formula. *Technomedia Journal*, 7(2), 231–247. <https://doi.org/10.33050/tmj.v7i2.1893>
- Munawar, Z., Herdiana, Y., Indah Putri, N., Informatika, M., Informatika, T., Bandung, I., & Bale Bandung, U. (2021). Sistem Rekomendasi Hibrid Menggunakan Algoritma Apriori Mining Asosiasi. *Jurnal Teknologi Informasi Dan Komunikasi*, 8(1), 84-95.
- Mustaqim, K., Amaresti, F. A., & Dewi, I. N. (2024). Analisis Sentimen Ulasan Aplikasi PosPay untuk Meningkatkan Kepuasan Pengguna dengan Metode K-Nearest Neighbor

- (KNN). *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 11–20. <https://doi.org/10.29408/edumatic.v8i1.24779>
- Puspita, A. D., Permadi, V. A., Anggani, A. H., & Christy, E. A. (2021). Musical Instruments Recommendation System Using Collaborative Filtering and KNN. *UMYGRACE : Proceedings Universitas Muhammadiyah Yogyakarta Undergraduate Conference*, 1(2), 1–6.
- Putra, A. I., & Santika, R. R. (2020). Implementasi Machine Learning dalam Penentuan Rekomendasi Musik dengan Metode Content-Based Filtering. *Edumatic : Jurnal Pendidikan Informatika*, 4(1), 121–130. <https://doi.org/10.29408/edumatic.v4i1.2162>
- Widiyaningtyas, T., Hidayah, I., & Adji, T. B. (2021). Recommendation Algorithm Using Clustering-Based UPCSim (CB-UPCSim). *Computers*, 10(10), 1-17. <https://doi.org/10.3390/computers10100123>
- Suharya, Y., Herdiana, Y., Putri, N. I., & Munawar, Z. (2021). Sistem Rekomendasi Untuk Toko Online Kecil Dan Menengah. *TEMATIK*, 8(2), 176-185. <https://doi.org/10.38204/tematik.v8i2.683>
- Yollis, S., & Netti, M. (2018). Spotify: Aplikasi Music Streaming untuk Generasi Milenial. *Jurnal Komunikasi*, 10(1), 1-16. <https://doi.org/10.24912/jk.v10i1.1102>