

Sistem Rekomendasi Film menggunakan *Neighborhood-Based Collaborative Filtering*

TUGAS AKHIR

**Rafi Athallah
120450038**



**PROGRAM STUDI SAINS DATA
FAKULTAS SAINS
INSTITUT TEKNOLOGI SUMATERA
LAMPUNG SELATAN
2025**

Sistem Rekomendasi Film menggunakan *Neighborhood-Based Collaborative Filtering*

TUGAS AKHIR

Diajukan sebagai syarat untuk memperoleh gelar sarjana

**Rafi Athallah
120450038**



**PROGRAM STUDI SAINS DATA
FAKULTAS SAINS
INSTITUT TEKNOLOGI SUMATERA
LAMPUNG SELATAN
2025**

LEMBAR PENGESAHAN

Tugas Akhir dengan judul "Sistem Rekomendasi Film menggunakan *Neighborhood-Based Collaborative Filtering*" adalah benar dibuat oleh saya sendiri dan belum pernah dibuat dan diserahkan sebelumnya, baik Sebagian ataupun seluruhnya, baik oleh saya ataupun orang lain, baik di Institut Teknologi Sumatera maupun di Institusi pendidikan lainnya.

Lampung Selatan, 18 Juni 2025

Penulis,



Rafi Athallah
NIM 120450038



Pembimbing I



Ahmad Luky Ramdani, S.Kom., M.Kom
NIP.198804242020121004

Pembimbing II



Luluk Muthoharoh, M.Si
NIP.199504112022032014

Disahkan Oleh,
Koordinator Program Studi Sains Data
Fakultas Sains


Institut Teknologi Sumatera



Tirta Setiawan, S.Pd., M.Si
NIP.199008222022031003

HALAMAN PERNYATAAN ORISINALITAS

Skripsi ini adalah karya saya sendiri dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan benar.

| | | |
|--------------|---|---|
| Nama | : | Rafi Athallah |
| NIM | : | 120450038 |
| Tanda Tangan | : |  |
| Tanggal | : | 18 Juni 2025 |

CS: Original dengan Cap

HALAMAN PERSETUJUAN PUBLIKASI

TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Institut Teknologi Sumatera, saya yang bertanda tangan dibawah ini:

Nama : Rafi Athallah
NIM : 120450038
Program Studi : Sains Data
Fakultas : Sains
Jenis Karya : Tugas Akhir

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Institut Teknologi Sumatera Hak Bebas Royalti Noneksklusif (*Non-Exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

Sistem Rekomendasi Film menggunakan *Neighborhood-Based Collaborative Filtering*

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Institut Teknologi Sumatera berhak menyimpan, mengalihmedia/formatkan, mengelola dalam pangkalan data (*database*), merawat dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian surat pernyataan ini saya buat dengan sebenarnya

Dibuat di : Institut Teknologi Sumatera
Tanggal : 18 Juni 2025

Yang menyatakan,



Rafi Athallah

Sistem Rekomendasi Film menggunakan *Neighborhood-Based Collaborative Filtering*

Rafi Athallah (120450038)

Pembimbing I: Ahmad Luky Ramdani, S.Kom., M.Kom

Pembimbing II: Luluk Muthoharoh, M.Si

ABSTRAK

Sistem rekomendasi telah menjadi alat penting dalam membantu pengguna menemukan konten yang sesuai dengan preferensi mereka. Namun, dengan banyaknya film yang diproduksi dan beragamnya genre yang tersedia, penonton seringkali kesulitan dalam memilih film yang akan ditonton. Kondisi ini menciptakan kebutuhan akan sistem rekomendasi yang dapat membantu pengguna menemukan film yang relevan. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi berbasis *Neighborhood-Based Collaborative Filtering*. Model yang dibangun menggunakan pendekatan *User-Based Collaborative Filtering* dan menghitung kemiripan dengan *Cosine Similarity*. Pemilihan jumlah tetangga terbaik (K) dilakukan menggunakan *GridSearchCV*, sedangkan akurasi prediksi dievaluasi menggunakan *Mean Absolute Error (MAE)*. Hasil penelitian menunjukkan bahwa model mampu memberikan rekomendasi yang akurat. Setelah mengeksplorasi berbagai skema *splitting data*, skema 80:20 dengan nilai $K = 4$ terbukti menjadi konfigurasi terbaik dengan *MAE 0.7236*. Nilai *MAE* yang rendah mengindikasikan bahwa sistem rekomendasi yang dikembangkan mampu memprediksi kesalahan kecil dan relevan dengan preferensi pengguna. Dengan sistem ini, pengguna dapat memperoleh pengalaman menonton film yang lebih personal dan memuaskan berdasarkan preferensi.

Kata Kunci: Sistem Rekomendasi, *Neighborhood-Based Collaborative Filtering*, *Collaborative Filtering*, *Cosine Similarity*, *Mean Absolute Error*

Movie Recommender System with Neighborhood-Based Collaborative Filtering Method

Rafi Athallah (120450038)

Advisor I: Ahmad Luky Ramdani, S.Kom., M.Kom

Advisor II: Luluk Muthoharoh, M.Si

ABSTRACT

Recommendation systems have become an important tool in helping users find content that matches their preferences. However, with the large number of movies produced and the variety of genres available, viewers often find it difficult to choose a movie to watch. This condition creates the need for a recommendation system that can help users find relevant movies. This research aims to develop a recommendation system based on Neighborhood-Based Collaborative Filtering. The model built uses the User-Based Collaborative Filtering approach and calculates similarity with Cosine Similarity. The selection of the best number of neighbors (K) is done using GridSearchCV, while the prediction accuracy is evaluated using Mean Absolute Error (MAE). The results show that the model is able to provide accurate recommendations. After exploring various data splitting schemes, the 80:20 scheme with $K = 4$ proved to be the best configuration with an MAE of 0.7236. The low MAE value indicates that the developed recommendation system is able to predict small errors and is relevant to user preferences. With this system, users can have a more personalized and satisfying movie watching experience based on their preferences.

Keywords: Recommendation System, Neighborhood-Based Collaborative Filtering, Collaborative Filtering, Cosine Similarity, Mean Absolute Error

MOTTO

*“Janganlah engkau bersedih, sesungguhnya Allah bersama kita”
(Q.S At-Taubah : 40)*

*“Maka, sesungguhnya beserta kesulitan ada kemudahan.
Sesungguhnya beserta kesulitan ada kemudahan”.
(Q.S Al-Insyirah : 5-6)*

*“Sesungguhnya Allah tidak mengubah keadaan suatu kaum hingga
mereka mengubah apa yang ada pada diri mereka”.
(Q.S Ar-Ra'd : 11)*

PERSEMBAHAN

Dengan penuh rasa cinta dan hormat, karya ini saya persembahkan kepada :

Orang Tua tercinta,

Terima kasih atas kesempatan, kepercayaan, kasih sayang, dan perjuangan tiada henti yang diberikan berikan untuk memastikan saya mendapatkan pendidikan yang layak. Kalian adalah cahaya dalam setiap langkah hidup saya, sumber kekuatan yang tak pernah padam, dan inspirasi terbesar yang mengajarkan arti ketulusan, kerja keras, kesabaran, dan keikhlasan.

Doa-doa mu adalah pelindung saya, cinta kalian yang tanpa syarat mengajarkan saya untuk terus berjuang, meskipun dalam kesulitan, dan tidak pernah menyerah dalam menggapai mimpi.

Semoga karya ini dapat menjadi bukti kecil dan rasa terima kasih saya yang tak terhingga. Teruslah menjadi pelita yang membimbing anakmu menjadi pribadi yang lebih baik, yang mampu membanggakan kalian, dan yang suatu hari dapat memberikan kebahagiaan seperti yang telah diberikan kepada saya.

KATA PENGANTAR

Segala puji hanya bagi Allah SWT, Tuhan semesta alam yang Maha Pengasih lagi Maha Penyayang, atas limpahan ramhat, taufik serta hidayah-Nya sehingga saya dapat menyelesaikan penulisan Tugas Akhir Skripsi ini. Saya menyadari bahwa Tugas Akhir Skripsi ini tidak akan terselesaikan tanpa bantuan, dukungan dan doa dari berbagai pihak. Oleh karena itu, dengan tulus saya menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua tercinta, Bapak Yurizalwis dan Ibu Nur Yeni, Adik-adik Shafen Asyfa dan Lislahiya Nisa, dan *Om* serta *Etek* keluarga besar *Nurmin Fam's* yang telah memberikan dukungan moril dan materil kepada saya.
2. Bapak Ahmad Luky Ramdhani, S.Kom., M.Kom selaku Pembimbing I yang telah memberikan banyak pengetahuan, arahan, bimbingan dan motivasi serta meluangkan waktu diskusi sehingga saya dapat menyelesaikan tugas akhir skripsi ini.
3. Ibu Luluk Muthoharoh, M.Si selaku Pembimbing II yang telah membantu banyak dalam proses penyusunan laporan, serta memberikan motivasi dan arahan sehingga saya dapat menyelesaikan tugas akhir skripsi ini.
4. Seluruh Dosen Prodi Sains Data, dan Civitas akademika Institut Teknologi Sumatera yang telah memberikan ilmu yang sangat berharga di tempat ini.
5. Teman-teman Prodi Sains Data Angkatan 2020, dan teman seperbimbingan, yang telah memberikan bantuan berupa dukungan satu sama lain untuk menyelesaikan tugas akhir skripsi ini.
6. Teman-teman *Kantin Receh*, *Pejuang Skripsi*, *Banner 24 Jam*, dan *Full Senyum* yang telah memberikan warna dalam kehidupan kampus

Saya menyadari bahwa karya ini terdapat banyak kekurangan dalam penulisan tugas akhir skripsi ini. Oleh karena itu, saya mengharapkan kritik dan saran yang membangun demi perbaikan. Akhir kata, semoga tugas akhir skripsi ini dapat memberikan manfaat bagi kita semua.

Lampung Selatan, 18 Juni 2025



Rafi Athallah

DAFTAR ISI

| | |
|--|-------------|
| LEMBAR PENGESAHAN..... | i |
| HALAMAN PERNYATAAN ORISINALITAS | ii |
| HALAMAN PERSETUJUAN PUBLIKASI | iii |
| TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS..... | iii |
| ABSTRAK | iv |
| ABSTRACT | v |
| MOTTO | vi |
| PERSEMBAHAN | vii |
| KATA PENGANTAR..... | viii |
| DAFTAR ISI..... | ix |
| DAFTAR GAMBAR..... | xii |
| DAFTAR TABEL | xiii |
| BAB I..... | 1 |
| PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 3 |
| 1.3 Tujuan Penelitian | 3 |
| 1.4 Batasan Masalah | 3 |
| 1.5 Manfaat Penelitian | 3 |
| 1.5.1 Manfaat untuk Peneliti..... | 3 |
| 1.5.2 Manfaat untuk Perguruan Tinggi..... | 4 |
| BAB II | 5 |
| TINJAUAN PUSTAKA | 5 |
| 2.1 Penelitian Rujukan | 5 |
| 2.2 Sistem Rekomendasi | 6 |
| 2.2 Collaborative Filtering..... | 8 |
| 2.2.1 Neighborhood-Based Collaborative Filtering | 8 |

| | |
|--|-----------|
| 2.4 Cosine Similarity..... | 10 |
| 2.5 Prediksi Rating..... | 11 |
| 2.6 Mean Absolute Error..... | 11 |
| 2.7 Pencarian <i>Hyperparameter</i> Optimal dengan <i>GridSearchCV</i> ... | 12 |
| BAB III..... | 14 |
| METODOLOGI PENELITIAN..... | 14 |
| 3.1 Waktu dan Tempat Penelitian..... | 14 |
| 3.2 Alat dan Bahan..... | 14 |
| 3.2.1 Spesifikasi Perangkat Keras..... | 14 |
| 3.2.2 Spesifikasi Perangkat Lunak..... | 14 |
| 3.3 Deskripsi Data..... | 14 |
| 3.4 Tahapan Penelitian..... | 15 |
| 3.5 Diagram Alir Pemodelan User-Based Collaborative Filtering..... | 16 |
| BAB IV..... | 19 |
| HASIL DAN PEMBAHASAN..... | 19 |
| 4.1 Dataset..... | 19 |
| 4.2 Praproses Data..... | 20 |
| 4.2.1 Menggabungkan Dataset <i>movies_csv</i> dan <i>ratings_csv</i> | 20 |
| 4.3 Hasil Pemodelan Sistem Rekomendasi <i>User-Based Collaborative Filtering</i> | 21 |
| 4.3.1 <i>Splitting</i> Data..... | 21 |
| 4.3.2 Matriks Rating <i>User-Item</i> | 21 |
| 4.3.4 Hasil Pencarian <i>Hyperparameter</i> Optimal dengan <i>GridSearchCV</i> | 22 |
| 4.3.5 Pelatihan Model <i>User-Based Collaborative Filtering</i> | 26 |
| 4.3.6 Nilai <i>Cosine Similarity User-Based Collaborative Filtering</i> | 26 |
| 4.3.7 Prediksi Rating <i>User-Based Collaborative Filtering</i> | 28 |
| 4.4 Hasil Evaluasi Model User-Based Collaborative Filtering..... | 31 |

| | |
|---|-----------|
| 4.5 Hasil Rekomendasi User-Based Collaborative Filtering | 32 |
| BAB V | 37 |
| KESIMPULAN DAN SARAN..... | 37 |
| 5.1 Kesimpulan | 37 |
| 5.2 Saran..... | 37 |
| DAFTAR PUSTAKA..... | 39 |
| LAMPIRAN | 45 |
| LAMPIRAN A. Contoh Perhitungan <i>Cosine Similarity</i> | 45 |
| LAMPIRAN B. Contoh Perhitungan Prediksi Rating <i>Collaborative Filtering</i> menggunakan <i>User-Based Collaborative Filtering</i> | 47 |
| LAMPIRAN C. Contoh Perhitungan <i>Mean Absolute Error</i> | 48 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 2.1 Proses Sistem Rekomendasi [21]. | 8 |
| Gambar 2.2 <i>User Based Collaborative Filtering</i> [28]. | 10 |
| Gambar 2.3 Ilustrasi Proses <i>GridSearchCV</i> [34]. | 13 |
| Gambar 3.1 Tahapan Penelitian. | 16 |
| Gambar 3.2 Pemodelan <i>User-Based Collaborative Filtering</i> | 18 |
| Gambar 4.1 Hasil Pencarian <i>Hyperparameter K</i> dengan Skema <i>Splitting 80:20</i> | 25 |
| Gambar 4.2 Hasil Pencarian <i>Hyperparameter K</i> dengan Skema <i>Splitting 75:25</i> | 25 |
| Gambar 4.3 Hasil Pencarian <i>Hyperparameter K</i> dengan Skema <i>Splitting 95:5</i> | 26 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 2.1 Daftar Penelitian Rujukan | 5 |
| Tabel 3.1 Deskripsi Data | 15 |
| Tabel 4.1 Data movies_csv | 19 |
| Tabel 4.2 Data ratings_csv | 20 |
| Tabel 4.3 Hasil Penggabungan Data | 20 |
| Tabel 4.4 Matriks Rating <i>User-Item User-Collaborative Filtering</i> .. | 22 |
| Tabel 4.5 Hasil Rata-Rata Nilai <i>K</i> dalam <i>GridSearchCV</i> | 23 |
| Tabel 4.6 <i>Cosine Similarity User-Based Collaborative Filtering</i> <i>Skema Splitting 80:20</i> | 27 |
| Tabel 4.7 <i>Cosine Similarity User-Based Collaborative Filtering</i> <i>Skema Splitting 75:25</i> | 27 |
| Tabel 4.8 <i>Cosine Similarity User-Based Collaborative Filtering</i> <i>Skema Splitting 95:5</i> | 27 |
| Tabel 4.9 Prediksi Rating <i>User-Based Collaborative Filtering</i> Skema <i>Splitting 80:20</i> | 29 |
| Tabel 4.10 Prediksi Rating <i>User-Based Collaborative Filtering</i> <i>Skema Splitting 75:25</i> | 29 |
| Tabel 4.11 Prediksi Rating <i>User-Based Collaborative Filtering</i> <i>Skema Splitting 95:25</i> | 30 |
| Tabel 4.12 Hasil Rekomendasi <i>User-Based Collaborative Filtering</i> <i>Skema Splitting 80:20</i> | 33 |
| Tabel 4.13 Hasil Rekomendasi <i>User-Based Collaborative Filtering</i> <i>Skema Splitting 75:25</i> | 34 |
| Tabel 4.14 Hasil Rekomendasi <i>User-Based Collaborative Filtering</i> <i>Skema Splitting 95:5</i> | 34 |
| Tabel Lampiran 1 Contoh Matriks Rating <i>User-Item</i> | 45 |
| Tabel Lampiran 2 Nilai <i>Cosine Similarity</i> antar <i>User</i> | 47 |

BAB I

PENDAHULUAN

1.1 Latar Belakang

Industri film menjadi salah satu sektor kreatif strategis yang mengalami pertumbuhan positif. Deputy Bidang Kebijakan Strategis Kemenparekraf, Dessy Ruhati menjelaskan kondisi makro di sektor pariwisata dan ekonomi kreatif tahun 2023, dari seluruh subsektor pariwisata dan ekonomi kreatif, film menjadi salah satu yang memiliki pertumbuhan paling positif, di sepanjang tahun 2023 total jumlah penonton bioskop tanah air mencapai 55 juta penonton [1]. Hal ini menunjukkan bahwa sektor pariwisata dan ekonomi kreatif dapat meningkatkan ekonomi dan pariwisata Indonesia. Berdasarkan laporan perusahaan *PricewaterhouseCoopers (PWC)* dan Lembaga Penyelidikan Ekonomi dan Masyarakat Fakultas Ekonomi dan Bisnis (LPEM FEB) Universitas Indonesia, total pendapatan industri layar (film, animasi, dan video) diprediksi tumbuh dari Rp.90,9 triliun pada 2022 menjadi Rp.109,6 triliun pada 2027 mendatang [2]. Menurut Kemenparekraf, pertumbuhan industri film bukan hanya karena tingginya antusiasme penonton, melainkan meningkatnya jumlah film yang berkualitas sehingga masuk ke dalam jajaran *Box Office* [3].

Banyaknya film yang diproduksi, seringkali penonton kesulitan dalam menentukan film yang akan ditonton [4]. Hal ini disebabkan berbagai macam genre film yang tersedia, menyebabkan penonton kesulitan dalam memilih film yang sesuai preferensi penonton. Selain itu, meningkatnya *platform streaming* dan melimpahnya ketersediaan film dari berbagai sumber, penonton sering merasa kewalahan akan pilihan yang ada [5]. Penonton juga tidak memiliki informasi yang cukup untuk melihat setiap film sehingga penonton kebingungan untuk memilih film. Setiap orang memiliki preferensi berbeda akan film yang ingin ditonton, seseorang bisa menyukai film berdasarkan rating, genre, aktor, atau perusahaan produksi film. Preferensi yang beragam ini membuat semakin sulit untuk memilih film yang sesuai dengan ekspektasi [6]. Oleh karena itu, diperlukan sebuah sistem yang dapat membantu penonton menemukan film yang sesuai dengan preferensi mereka. Sistem Rekomendasi dapat membantu penonton dalam memilih film

yang tepat sesuai dengan preferensi mereka [5]. Sistem ini bekerja dengan menggunakan berbagai metode, seperti *Content-Based Filtering*, *Collaborative Filtering*, dan *Hybrid Filtering*.

Sistem rekomendasi dalam konteks film dapat menggunakan berbagai macam pendekatan, salah satu pendekatan yang bisa dilakukan adalah dengan menggunakan *Neighborhood-Based Collaborative Filtering*. Pendekatan ini didasarkan pada asumsi bahwa pengguna yang memiliki preferensi serupa akan menunjukkan pola perilaku yang mirip pula. Oleh karena itu, metode ini dapat digunakan untuk menemukan pengguna dengan selera film yang serupa guna memberikan rekomendasi yang relevan [7]. Penelitian sebelumnya telah menunjukkan bahwa metode ini efektif digunakan untuk memberikan rekomendasi, *Grouplens* [8] merupakan pelopor yang mengenalkan *Collaborative Filtering*, selanjutnya penelitian *The Bellcore Video Recommender* oleh Hill et al. (1995) [9] mengembangkan algoritma yang pertama kali diperkenalkan oleh *Grouplens*, Selain itu, Herlocker et al. (2002) [10] menekankan pentingnya jumlah tetangga (K) dalam menentukan kualitas dan akurasi rekomendasi. Penelitian yang dilakukan Roziqiin et al. (2024) [11] berhasil membuat sistem rekomendasi pemilihan anime yang sesuai dengan preferensi pengguna dengan hasil skor *MAE* sebesar 2.9449 yang menunjukkan pendekatan ini efektif dalam sistem rekomendasi, Selain itu, penelitian oleh Fakhri et al. (2019) [12] membangun sistem rekomendasi restoran di wilayah bandung raya dengan menggunakan pendekatan yang sama, evaluasi yang digunakan menggunakan *MAE* menghasilkan nilai terbaik 1.492, menegaskan efektivitas metode ini dalam berbagai domain.

Dalam penelitian ini, *Neighborhood-Based Collaborative Filtering* akan diterapkan dengan menggunakan *User-Based Collaborative Filtering*. Pendekatan ini dipilih karena mampu mengidentifikasi kesamaan preferensi antar pengguna, Dengan memanfaatkan data rating dari pengguna lain yang memiliki kemiripan preferensi, diharapkan sistem rekomendasi yang dikembangkan tidak hanya akurat, tetapi juga relevan dengan kebutuhan pengguna. Hal ini diharapkan dapat meningkatkan pengalaman menonton yang lebih menyenangkan dan sesuai dengan selera masing-masing individu.

1.2 Rumusan Masalah

1. Bagaimana implementasi *Neighborhood-Based Collaborative Filtering* dalam sistem rekomendasi film?
2. Bagaimana hasil akurasi model yang dihasilkan *Neighborhood-Based Collaborative Filtering* dalam memberikan sistem rekomendasi film?

1.3 Tujuan Penelitian

1. Mengimplementasikan menggunakan model *User-Based Collaborative Filtering* dalam sistem rekomendasi film.
2. Mengevaluasi akurasi sistem rekomendasi yang dihasilkan dari model *User-Based Collaborative Filtering* menggunakan *Mean Absolute Error (MAE)*.

1.4 Batasan Masalah

Batasan Masalah Dalam Penelitian ini adalah:

1. Teknik *Neighborhood-Based Collaborative Filtering* yang digunakan adalah *User-Based Filtering*.
2. Data yang digunakan berisi total 610 pengguna dengan 100.836 rating pengguna dengan jumlah 9742 film. Dengan Atribut data yang digunakan berupa *userId*, *movieId*, *rating*, dan *title*.

1.5 Manfaat Penelitian

Berdasarkan rumusan masalah dan tujuan penelitian diatas, berikut manfaat dari penelitian ini:

1.5.1 Manfaat untuk Peneliti

1. Mendapatkan pengetahuan baru dan pengalaman dalam mengolah data teks untuk menghasilkan rekomendasi.
2. Berkontribusi dalam pengetahuan akan sistem rekomendasi film.

1.5.2 Manfaat untuk Perguruan Tinggi

1. Mengembangkan literatur ilmiah di bidang sistem rekomendasi film.
2. Memberikan kesempatan untuk publikasi penelitian dalam jurnal ilmiah.

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Rujukan

Penelitian rujukan memainkan peran penting dalam memahami dan mengembangkan landasan teoritis serta metodologi untuk penelitian yang sedang dilakukan. Tabel 2.1 dibawah ini menyajikan ringkasan dari beberapa penelitian yang relevan dengan topik yang sedang dieeliti, yaitu penggunaan *Neighborhood-Based Collaborative Filtering* dalam Sistem Rekomendasi Film. Melalui penelitian terdahulu dibawah dapat mengidentifikasi pengetahuan yang ada dan serta kontribusi dalam bidang ini.

Tabel 2.1 Daftar Penelitian Rujukan.

| No | Judul Penelitian | Nama Peneliti, Tahun | Data dan Metode | Hasil |
|----|---|---|--|--|
| 1. | <i>Restaurant Recommender System Using User-Based Collaborative Filtering Approach: A Case Study at Bandung Raya Region</i> | Alif Azhar Fakhri et al., 2019 [12] | Data yang diperoleh dari kuisisioner,di mana pengguna membagikan pengalaman dan tingkat kepuasan mereka mengenai restoran di daerah bandung. | Penelitian ini menggunakan <i>User-Based Collaborative Filtering</i> dalam mengembangkan sistem rekomendasi restoran. Penelitian ini menekankan pemilihan tetangga (K) berpengaruh pada hasil rekomendasi. Hasil evaluasi menunjukkan nilai MAE sebesar 1.492 yang menunjukkan hasil rekomendasi cukup baik. |
| 2. | <i>Sistem Rekomendasi Pemilihan Anime menggunakan User-Based Collaborative Filtering</i> | Nazhif Muafa Roziqiin et al., 2024 [11] | Dataset penelitian ini diambil dari platform Kaggle dengan beberapa file seperti Anime_csv yang berisi judul,genre,dan judul, lalu | Penelitian ini berhasil mengunakan <i>User-Based Collaborative Filtering</i> yang mengandalkan preferensi pengguna untuk merekomendasikan anime. Metode ini |

| No | Judul Penelitian | Nama Peneliti, Tahun | Data dan Metode | Hasil |
|----|---|---|--|--|
| | | | Rating_csv berisi rating user terhadap anime yang telah ditonton | mengidentifikasi pengguna dengan selera yang serupa dan menyarankan anime berdasarkan apa yang telah dirating oleh pengguna tersebut. Hasil evaluasi dengan MAE 2.944 menunjukkan kualitas rekomendasi yang baik. |
| 3. | <i>Sistem Rekomendasi Produk UMKM menggunakan Algoritma User-Based Collaborative Filtering berbasis website</i> | Esy Anugerah Rahayu Kasim et al., 2024 [13] | <i>Dataset penelitian ini diambil dari rating yang didapatkan dari halaman riwayat dari sistem yang digunakan, dan survey kepada pelaku UMKM terkait produk UMKM tersebut.</i> | Penelitian ini menggunakan metode populer <i>User-Based Collaborative Fitering</i> dengan menganalisis perilaku pengguna untuk merekomendasikan produk. Metode ini menggunakan nilai kemiripan dalam menghitung kesamaan pengguna. Hasil Evaluasi menggunakan MAE dengan nilai 1.11 menunjukkan bahwa sistem rekomendasi ini efektif dalam memberikan rekomendasi kepada pengguna. |

2.2 Sistem Rekomendasi

Sistem rekomendasi merupakan sistem yang dirancang untuk memberikan saran terkait produk atau layanan tertentu yang dapat dimanfaatkan oleh pengguna [14]. Saran yang diberikan bertujuan untuk memudahkan pengguna dalam pengambilan keputusan menemukan item yang menarik dan relevan bagi pengguna. Saat ini, sistem rekomendasi banyak digunakan di berbagai produk digital seperti *e-commerce* [15], media sosial [16], layanan musik digital [18],

layanan *streaming* film [18], dan lain-lain. Sistem rekomendasi mempelajari pengalaman, opini, dan perilaku pengguna lalu merekomendasikan produk yang mereka anggap paling relevan sesuai dengan preferensi pengguna.

Terdapat beberapa teknik yang digunakan untuk membangun sistem rekomendasi, diantaranya *Content-Based Filtering*, *Collaborative Filtering*, *Hybrid Filtering*, dan lain-lain. *Content-Based Filtering* bekerja dengan menganalisis karakteristik atau fitur dari item yang pernah disukai oleh pengguna, kemudian merekomendasikan item lain yang memiliki kemiripan dengan item-item tersebut. [19]. Sementara itu *Collaborative Filtering* memanfaatkan interaksi pengguna dalam bentuk penilaian atau aktivitas terhadap item, kemudian sistem mencari pengguna lain yang memiliki preferensi serupa (*User-Based*), atau mencari item yang sering dinilai serupa oleh pengguna yang sama (*Item-Based*), untuk memberikan rekomendasi [20]. Sedangkan *Hybrid Filtering* merupakan gabungan dari teknik rekomendasi untuk menghasilkan rekomendasi.

Gambar 2.1 merupakan proses bagaimana Sistem Rekomendasi Film bekerja, dengan mengumpulkan data perilaku pengguna, seperti tindakan membeli, menyukai, menonton, atau tidak menyukai suatu konten atau produk melalui berbagai aplikasi (dalam kasus ini adalah film). Data tersebut dikirim ke sistem rekomendasi sebagai masukan pengguna (*user feedback*), yang kemudian dianalisis menggunakan algoritma pembelajaran mesin untuk menghasilkan daftar rekomendasi yang relevan sesuai preferensi pengguna. Sistem ini berfungsi secara berkelanjutan, di mana setiap interaksi pengguna memperbaiki akurasi rekomendasi berikutnya, menciptakan pengalaman yang lebih personal dan tepat sasaran.



Gambar 2.1 Proses Sistem Rekomendasi [21].

2.2 Collaborative Filtering

Collaborative Filtering adalah salah satu teknik populer yang digunakan dalam sistem rekomendasi yang bekerja dengan mengandalkan kolaborasi (*Collaborative*) antar pengguna atau antar item untuk memberikan rekomendasi. Berbeda dengan *Content-Based Filtering* yang berfokus pada karakteristik item, *Collaborative Filtering* memanfaatkan data interaksi seperti rating, klik, atau histori dalam memberikan rekomendasi. Sebagai contoh jika pengguna U dan V memiliki penilaian yang sama dalam menonton suatu film, besar kemungkinan mereka berdua memiliki preferensi yang serupa terhadap item lain [22].

2.2.1 Neighborhood-Based Collaborative Filtering

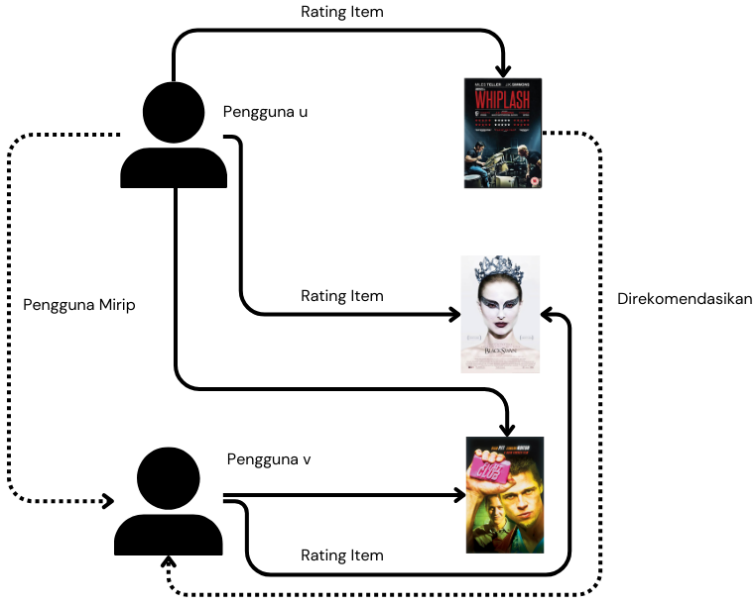
Neighborhood-Based Collaborative Filtering atau yang biasa disebut *Memory-Based Filtering* merupakan awal mula dari perkembangan Sistem Rekomendasi, yang mana prediksi rating rekomendasi didasarkan pada tetangga-nya [23]. Dalam hal ini *Neighborhood-Based Collaborative Filtering* dapat dibagi menjadi dua teknik yaitu:

- ***User-Based Collaborative Filtering*** bekerja dengan menganalisis perilaku dan preferensi pengguna lain yang memiliki kemiripan dengan pengguna target untuk menghasilkan rekomendasi [24]. Sistem akan membandingkan histori interaksi pengguna target dan pengguna lain untuk

mengidentifikasi sekelompok pengguna yang disebut sebagai *neighbors* atau tetangga. Kemiripan antar pengguna dihitung menggunakan nilai kemiripan seperti *Cosine Similarity*, *Pearson Correlation* dll [25]. Setiap pengguna akan mendapatkan nilai kemiripan terhadap pengguna target, lalu sistem akan memilih pengguna dengan nilai kemiripan tertinggi yang akan menjadi K tetangga. Prediksi rating akan didasarkan pada tetangga tersebut, sebagai contoh pengguna U dan pengguna V yang memiliki perilaku serupa, maka rating pengguna U terhadap item yang belum dinilai oleh pengguna V bisa dijadikan sebagai dasar untuk prediksi rating [26].

- ***Item-Based Collaborative Filtering*** bekerja dengan menganalisis perilaku dan pola interaksi pengguna dengan item yang menghasilkan rekomendasi dari kemiripan antar item [27]. Berbeda dengan *Content-Based Filtering* yang membandingkan fitur dari item, *Item-Based Collaborative Filtering* hanya berdasarkan pola interaksi dengan item tersebut. Sebagai contoh pengguna yang memberi rating tinggi pada item A dan B namun belum menonton film C, lalu sebagian besar pengguna lain yang menyukai film A dan film B juga menyukai C, maka film C akan direkomendasikan kepada pengguna tersebut.

Gambar 2.2 merupakan ilustrasi dari *User-Based Collaborative Filtering* yang bekerja dengan memanfaatkan kemiripan antar pengguna untuk memberikan rekomendasi. Dalam ilustrasi tersebut, terdapat dua pengguna U dan pengguna V yang memberi rating pada film yang sama, seperti *Black Swan* dan *Fight Club*, dalam hal ini kedua pengguna memberikan rating mereka terhadap film yang sama sehingga dianggap pengguna yang serupa. Karena pengguna U memberikan rating film *Whiplash*, yang mana pengguna V belum pernah menonton film tersebut, maka film *Whiplash* akan direkomendasikan kepada Pengguna V karena kedua pengguna memiliki preferensi film serupa sebelumnya sehingga dianggap pengguna yang mirip satu sama lain.



Gambar 2.2 *User Based Collaborative Filtering* [28].

2.4 Cosine Similarity

Cosine Similarity adalah salah satu metode yang digunakan untuk mengukur kemiripan antara dua vektor dalam ruang berdimensi tinggi [29]. Dalam konteks sistem rekomendasi, vektor ini sering kali mewakili pengguna atau item, dengan elemen-elemen vektor menunjukkan rating yang diberikan pengguna terhadap berbagai item atau sebaliknya. *Cosine Similarity* mengukur sudut kosinus antara dua vektor, yang nilainya berkisar antara 0 hingga 1. Nilai 1 menunjukkan bahwa dua vektor tersebut memiliki arah yang sama, atau dengan kata lain sangat mirip. Nilai 0 menunjukkan bahwa dua vektor tersebut tegak lurus yang berarti tidak memiliki kemiripan [30]. Pada persamaan 2.1 ini adalah rumus untuk menghitung *Cosine Similarity* antara dua vektor:

$$sim(u, v) = \frac{\sum_{i \in I_{uv}} r_{ui} \cdot r_{vi}}{\sqrt{\sum_{i \in I_{uv}} r_{ui}^2} \cdot \sqrt{\sum_{i \in I_{uv}} r_{vi}^2}} \quad (2.1)$$

Keterangan :

$sim(u, v)$: nilai kemiripan antara pengguna u dan pengguna v

- I_{uv} : himpunan dari semua item i yang telah diberi rating oleh pengguna u dan v
 r_{ui} : rating pengguna u terhadap item i
 r_{vi} : rating pengguna v terhadap item i

2.5 Prediksi Rating

Prediksi Rating adalah proses dalam sistem rekomendasi yang digunakan untuk memperkirakan nilai rating yang akan diberikan oleh pengguna terhadap item yang belum dirating [31]. Tujuan prediksi rating adalah memberikan prediksi rating tertinggi, yang mencerminkan preferensi pengguna terhadap item tersebut. Prediksi rating menjadi inti dari sistem rekomendasi *Collaborative Filtering*, yang bergantung pada pola interaksi antara pengguna dan item.

$$\hat{r}_{ui} = \frac{\sum_{v \in N_i^k(u)} \mathbf{sim}(u, v) \cdot r_{vi}}{\sum_{v \in N_i^k(u)} \mathbf{sim}(u, v)} \quad (2.2)$$

Keterangan:

- \hat{r}_{ui} : rating yang diprediksi untuk pengguna u pada item i
 $N_i^k(u)$: himpunan tetangga dari pengguna u yang telah memberi rating pada item i
 $\mathbf{sim}(u, v)$: nilai kemiripan antara pengguna u dan v
 r_{vi} : rating pengguna v terhadap item i

2.6 Mean Absolute Error

Mean Absolute Error adalah metrik evaluasi yang digunakan untuk mengukur seberapa baik suatu model prediksi bekerja dengan menghitung rata-rata perbedaan nilai absolut antara nilai yang diprediksi oleh model dan nilai sebenarnya [32]. Dalam konteks sistem rekomendasi, mean absolute error menghitung rata-rata dari nilai absolut kesalahan antara rating yang diprediksi oleh model dan rating sebenarnya yang diberikan oleh pengguna. Berikut ini adalah rumus *Mean Absolute Error*:

$$MAE = \frac{1}{|\hat{R}|} \sum_{\hat{r}_{ui} \in \hat{R}} |r_{ui} - \hat{r}_{ui}| \quad (2.3)$$

Keterangan:

$|\hat{R}|$: jumlah total prediksi

r_{ui} : rating sebenarnya yang diberikan pengguna u untuk item i

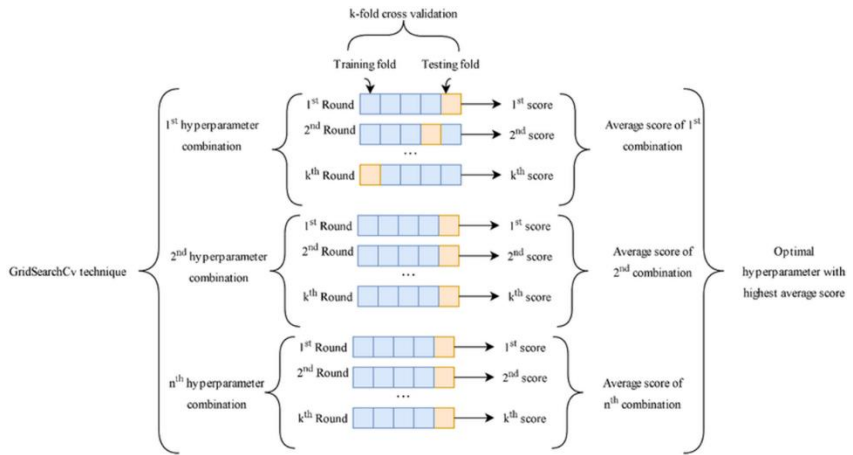
\hat{r}_{ui} : rating yang diprediksi model untuk pengguna u terhadap

2.7 Pencarian *Hyperparameter* Optimal dengan *GridSearchCV*

GridSearchCV adalah metode yang digunakan untuk menemukan kombinasi *Hyperparameter* yang optimal dalam model *machine learning* [33]. Pada penelitian ini, *GridSearchCV* diterapkan untuk menemukan *hyperparameter* K tetangga terbaik dalam sistem rekomendasi. Metode ini bekerja dengan mengeksplorasi setiap kemungkinan kombinasi *Hyperparameter* yang telah ditentukan sebelumnya, kemudian mengevaluasi kinerja model menggunakan metrik tertentu, misalnya akurasi.

GridSearchCV secara otomatis mengintegrasikan proses pencarian *Hyperparameter* dengan teknik *Cross Validation*. Pada tahap awal, dataset dibagi menjadi beberapa K bagian (*fold*). Dalam setiap iterasi *Cross Validation*, $K-1$ *fold* digunakan sebagai data latih, sementara satu *fold* sisanya digunakan sebagai data uji. Untuk setiap kombinasi *Hyperparameter* yang diuji, *GridSearchCV* melatih model menggunakan data latih dari setiap $K-1$ *fold* tersebut dan mengevaluasi performanya pada data uji dari satu *fold* yang sama. Skor rata-rata dari putaran *Cross Validation* kemudian digunakan untuk menentukan kinerja akhir dari kombinasi *Hyperparameter* tersebut.

Gambar 2.5 menunjukkan ilustrasi proses *GridSearch* biasanya dikombinasikan dengan *Cross Validation* (*GridSearchCV*) untuk memastikan performa model yang diuji dapat digeneralisasi dengan baik dan memungkinkan pemilihan *hyperparameter* yang baik sehingga menghasilkan model yang optimal [34].



Gambar 2.3 Ilustrasi Proses *GridSearchCV* [34].

BAB III

METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian yang berjudul “**Sistem Rekomendasi Film menggunakan *Neighborhood-Based Collaborative Filtering***” yang akan dilaksanakan dalam waktu satu semester Institut Teknologi Sumatera dan data diolah di Institut Teknologi Sumatera.

3.2 Alat dan Bahan

Alat dan Bahan yang digunakan pada penelitian ini mencakup kebutuhan akan perangkat keras dan lunak yang dibutuhkan dalam pembuatan Sistem Rekomendasi Film.

3.2.1 Spesifikasi Perangkat Keras

Berikut spesifikasi perangkat keras yang digunakan adalah Processor Intel(R) Core (TM) i3-6006U PU @ 2.00GHz 1.99 GHz, RAM 4,00 GB dan Penyimpanan SSD dengan kapasitas memori 512 GB

3.2.2 Spesifikasi Perangkat Lunak

Berikut spesifikasi perangkat lunak yang digunakan adalah Sistem Operasi Windows 10 menggunakan Bahasa Pemrograman *Python* dan *Google Collaboratory*

3.3 Deskripsi Data

Pada penelitian ini, data diperoleh dari *website MovieLens* yang berasal dari *GroupLens*¹, laboratorium penelitian di Departemen Ilmu Komputer dan Teknik Universitas Minesota. Data yang digunakan berupa dataset *MovieLens (ml-latest-small)* dengan total 610 pengguna dengan 100.836 rating pengguna dengan jumlah 9742 film. Informasi yang digunakan antara lain *userId*, *movieId*, *rating*, *timestamp* dan *Genre* secara konsisten dapat dilihat pada Tabel 3.1 berikut:

¹ <https://grouplens.org/datasets/movielens/>

Tabel 3.1 Deskripsi Data

| Atribut | Deskripsi |
|----------------|---|
| <i>userId</i> | Pengguna <i>MovieLens</i> |
| <i>movieId</i> | Film dengan rating dalam kumpulan data |
| <i>rating</i> | Rating pengguna terhadap film rentang (1-5) |
| <i>title</i> | Judul film |
| <i>genres</i> | Genre film |

3.4 Tahapan Penelitian

Pada bagian ini akan ditampilkan tahapan dalam penelitian ini sesuai dengan gambar di bawah untuk memudahkan pembaca dalam memahami penelitian ini, tahapan penelitian dimulai dari Pengumpulan Data, Praproses Data, Implementasi *User-Based Collaborative Filtering*, Evaluasi Model dan Hasil Rekomendasi. Tahapan-tahapan ini dapat dilihat secara lengkap pada Gambar 3.1, yaitu:

1. Mulai

Penelitian diawali dengan tahap perencanaan dan perumusan masalah. Pada tahap fundamental ini, ditentukan tujuan dari penelitian ini. Tahapan awal ini merupakan tahapan penting sebelum memulai implementasi Sistem rekomendasi menggunakan *User-Based Collaborative Filtering*.

2. Pengumpulan Data

Data Penelitian ini diperoleh dari sumber sekunder, yaitu dataset publik *MovieLens* oleh *GroupLens*. Pengumpulan data dilakukan untuk mendapatkan informasi yang dibutuhkan dalam penelitian untuk mencapai tujuan dalam penelitian.

3. Praproses Data

Tahap praproses data dilakukan untuk membersihkan dan mentransformasi dataset mentah. Tujuannya adalah mempersiapkan data agar siap digunakan dalam proses implementasi pemodelan *User-Based Collaborative Filtering*.

4. Implementasi *User-Based Collaborative Filtering*

Pada tahap implementasi, model sistem rekomendasi dibangun menggunakan metode *User-Based Collaborative Filtering*. Tahapan pembuatan model merupakan bagian penting dalam penelitian ini, pembuatan model dilakukan dengan untuk

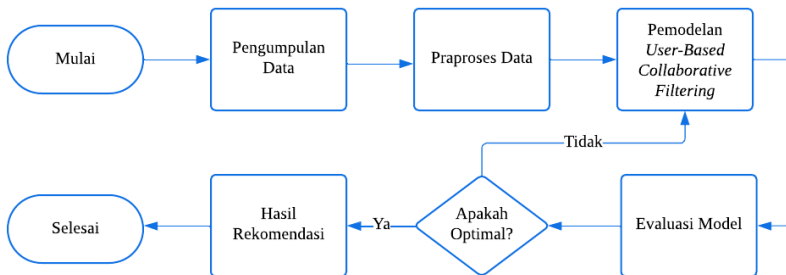
membangun sistem rekomendasi film berdasarkan data rating pengguna.

5. Evaluasi Model

Evaluasi Model dilakukan untuk menentukan performa sistem rekomendasi, proses evaluasi dilakukan menggunakan *Mean Absolute Error (MAE)* untuk mengukur kesalahan prediksi rating. Hal ini memastikan sistem rekomendasi *User-Based Collaborative Filtering* memiliki performa yang baik.

6. Hasil Rekomendasi

Tahapan ini adalah tahap terakhir dalam penelitian ini, yang bertujuan untuk memberikan rekomendasi kepada pengguna berdasarkan hasil prediksi rating model.



Gambar 3.1 Tahapan Penelitian.

3.5 Diagram Alir Pemodelan User-Based Collaborative Filtering

Diagram alir ini berfungsi sebagai peta konseptual yang menjelaskan bagaimana proses di dalam pemodelan *User-Based Collaborative Filtering* yang mana pada tahap ini dimulai dengan *Splitting Data*, pembuatan Matriks Rating *User-Item*, Pencarian *Hyperparameter* terbaik dengan *GridSearchCV*, Pelatihan Model, Menghitung *Cosine Similarity*, dan Prediksi Rating yang berjalan secara sistematis dan teratur. Adapun tahapan yang dilakukan adalah sesuai dengan Gambar 3.2 berikut:

1. Mulai

Tahapan awal sebelum memulai proses pemodelan *User-Based Collaborative Filtering*

2. *Splitting Data*

Dataset yang telah melalui tahap praproses dibagi menjadi data latih (training set) dan data uji (testing set). Penelitian ini mengeksplorasi pembagian dengan skema proporsi 80:20, 75:25, dan 95:5. Hasil dari ketiga skema tersebut selanjutnya akan ditampilkan pada sub-bab selanjutnya.

3. Matriks Rating *User-Item*

Matriks rating *User-Item* dibentuk untuk merepresentasikan interaksi rating antar pengguna dan item. Matriks ini menjadi dasar untuk analisis selanjutnya, dengan pengguna sebagai baris, item sebagai kolom.

4. Pencarian *Hyperparameter* Optimal dengan *GridSearchCV*

Mencari *Hyperparameter K* (Jumlah tetangga) Optimal untuk model *User-Based Collaborative Filtering* dilakukan menggunakan metode *GridSearchCV*. Proses ini memanfaatkan *10-fold Cross Validation* pada data latih untuk mengevaluasi berbagai nilai *K* dan memilih *Hyperparameter* yang menghasilkan performa terbaik berdasarkan metrik evaluasi *Mean Absolute Error (MAE)*.

5. Pelatihan Model *User-Based Collaborative Filtering*

Model *User-Based Collaborative Filtering* kemudian dilatih menggunakan seluruh data latih dengan konfigurasi *Hyperparameter K* terbaik yang diperoleh dari tahap *GridSearchCV*. Pelatihan model *User-Based Collaborative Filtering* setelah mendapatkan *Hyperparameter* terbaik.

6. Perhitungan *Cosine Similarity*

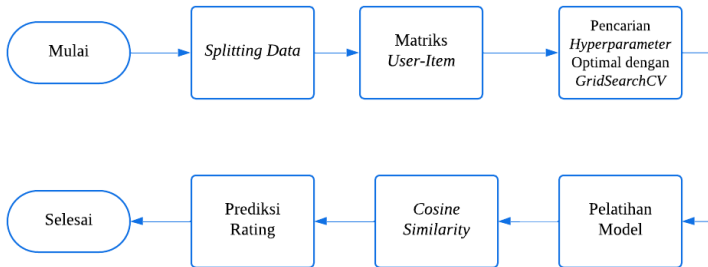
Kemiripan antar pengguna dihitung berdasarkan matriks rating *user-item*. Penelitian ini *Cosine Similarity* digunakan untuk mengukur seberapa mirip preferensi antara satu pengguna dengan pengguna lainnya.

7. Prediksi Rating *User-Based Collaborative Filtering*

Pada tahap ini, model *User-Based Collaborative Filtering* yang telah dilatih digunakan untuk membuat prediksi rating terhadap item-item yang ada di dalam data uji. Data uji berisi pasangan *User-Item* beserta rating aktualnya yang tidak digunakan selama proses pelatihan model. Prediksi rating yang dihasilkan kemudian dibandingkan dengan rating aktual pada data uji untuk mengukur performa dan akurasi model.

8. Selesai

Proses pemodelan inti *User-Based Collaborative Filtering* selesai. Model yang telah diprediksi ratingnya pada data uji ini kemudian siap untuk dilakukan evaluasi model.



Gambar 3.2 Pemodelan *User-Based Collaborative Filtering*

BAB IV

HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi *User-Based Collaborative Filtering* dalam sistem rekomendasi film. Penelitian ini bertujuan untuk memberikan rekomendasi yang relevan dengan preferensi pengguna berdasarkan rating yang diberikan oleh pengguna lain yang memiliki pola perilaku serupa.

Proses implementasi mencakup pemilihan dan praproses dataset, pembuatan matriks rating *user-item*, pencarian *hyperparameter* terbaik, pelatihan model, menghitung *Cosine Similarity*, prediksi rating, dan evaluasi performa model disajikan berdasarkan metrik *Mean Absolute Error (MAE)*. Pembahasan dilakukan untuk menganalisis hasil tersebut dan menghubungkannya dengan literatur dan tujuan penelitian.

4.1 Dataset

Dataset yang digunakan dalam penelitian ini adalah data film yang telah dirating oleh pengguna yang berasal dari *website MovieLens*, dengan total 610 pengguna dengan 100.836 rating pengguna dengan jumlah 9742 film. Penelitian ini menggunakan dua file utama yaitu *movies_csv* dan *ratings_csv*. Berikut adalah data *movies_csv* dan *ratings_csv* yang akan digunakan terlihat dalam Tabel 4.1 dan Tabel 4.2

Tabel 4.1 Data *movies_csv*

| <i>movieId</i> | <i>title</i> | <i>genres</i> |
|----------------|---|---|
| 1 | <i>Toy Story (1995)</i> | <i>Adventure/Animation/Children/Comedy/ Fantasy</i> |
| 2 | <i>Jumanji (1995)</i> | <i>Adventure/Children/Fantasy</i> |
| 3 | <i>Grumpier Old Men (1995)</i> | <i>Comedy/Romance</i> |
| 4 | <i>Waiting to Exhale (1995)</i> | <i>Comedy/Drama/Romance</i> |
| 5 | <i>Father of the Bride Part II (1995)</i> | <i>Comedy</i> |
| ... | ... | |

Tabel 4.2 Data ratings_csv

| <i>userId</i> | <i>movieId</i> | <i>rating</i> | <i>timestamp</i> |
|---------------|----------------|---------------|------------------|
| 1 | 1 | 4 | 964982703 |
| 1 | 3 | 4 | 964981247 |
| 1 | 6 | 4 | 964982224 |
| 1 | 47 | 5 | 964983815 |
| 1 | 50 | 5 | 964982931 |
| ... | ... | ... | ... |

4.2 Praproses Data

Praproses data adalah proses penting yang dilakukan sebelum melakukan pemodelan. Tahapan ini bertujuan untuk memastikan bahwa data yang digunakan dalam penelitian ini bersih, relevan dan siap digunakan dalam pemodelan. Berikut adalah tahapan praproses data pada penelitian ini adalah sebagai berikut:

4.2.1 Menggabungkan Dataset *movies_csv* dan *ratings_csv*

Tahap pertama dalam praproses adalah menggabungkan dataset dari *movie_csv* dan *ratings_csv*. Proses penggabungan dilakukan untuk mendapatkan satu dataset yang mencakup informasi seperti *userId*, *movieId*, *rating*, *timestamp*, dan *title*. Penggabungan dilakukan berdasarkan *movieId*, sehingga menghasilkan dataset yang lengkap. Hasil penggabungan tabel dapat dilihat dalam Tabel 4.3 berikut:

Tabel 4.3 Hasil Penggabungan Data

| <i>userId</i> | <i>movieId</i> | <i>rating</i> | <i>title</i> | <i>genres</i> |
|---------------|----------------|---------------|---------------------------------------|-------------------------------------|
| 1 | 1 | 4 | <i>Toy Story (1995)</i> | <i>Adventure/Animation/Children</i> |
| 1 | 3 | 4 | <i>Grumpier Old Men (1995)</i> | <i>Comedy/Romance</i> |
| 1 | 6 | 4 | <i>Heat (1995)</i> | <i>Action/Crime/Thriller</i> |
| ... | ... | ... | ... | ... |
| 610 | 168250 | 5 | <i>Get Out (2017)</i> | <i>Horror</i> |
| 610 | 168252 | 5 | <i>Logan (2017)</i> | <i>Action/Sci-Fi</i> |
| 610 | 170875 | 3 | <i>The Fate of The Furious (2017)</i> | <i>Action/Crime/Drama</i> |

4.3 Hasil Pemodelan Sistem Rekomendasi *User-Based Collaborative Filtering*

Pada sub-bab ini, akan dijelaskan hasil pemodelan sistem rekomendasi yang dikembangkan menggunakan *User-Based Collaborative Filtering*. Implementasi ini bertujuan untuk memberikan rekomendasi film kepada pengguna yang berguna untuk meningkatkan relevansi hasil rekomendasi.

4.3.1 *Splitting Data*

Pada tahap pemodelan, data yang sudah di proses akan dibagi menjadi data latih (*training set*) dan data uji (*testing set*). Untuk menentukan proporsi yang optimal, beberapa skema pembagian data dilakukan sebagai eksplorasi dan eksperimen. Data latih yang besar memungkinkan model untuk mempelajari pola yang kompleks, tapi memiliki data uji yang sedikit menyebabkan evaluasi yang kurang diandalkan terhadap kemampuan generalisasi model [35]. Penelitian ini mengeksplorasi pembagian dengan skema proporsi 80:20, 75:25, dan 95:5. Hasil dari ketiga skema tersebut selanjutnya akan ditampilkan pada sub-bab selanjutnya. Dari hasil evaluasi terhadap skema-skema tersebut, pembagian data dengan rasio 80:20 (80% data latih dan 20% data uji) memberikan performa model terbaik dan selanjutnya akan digunakan dalam penelitian ini.

4.3.2 Matriks Rating *User-Item*

Pada tahap ini, akan dibuat Matriks Rating *User-Item* dari model *User-Based Collaborative Filtering*. Matriks ini adalah representasi yang menggambarkan interaksi rating antara *user* dan *item*, dimana setiap kolom mewakili item dan baris mewakili *user*. Secara teknis, pembuatan matriks ini dilakukan dalam Bahasa pemrograman Python menggunakan fungsi *pivot_table* dari library *Pandas*, fungsi ini mentransformasi data rating dengan *userId* sebagai indeks baris, *movieId* sebagai nama kolom dan nilai rating sebagai isi sel. Kombinasi *user-item* yang tidak memiliki nilai rating akan menghasilkan nilai kosong (*NaN*) pada matriks awa, sehingga nilai *NaN* tersebut diisi dengan 0 yang mengindikasikan bahwa pengguna tersebut belum memberikan rating untuk item terkait. Matriks *user-item* ini digunakan

sebagai dasar sebelum menghitung kemiripan antar pengguna menggunakan *Cosine Similarity*. Matriks yang berisi nilai 0 berarti pengguna belum memberikan rating untuk item tersebut, hal ini dilakukan untuk memastikan bahwa hanya interaksi yang memiliki rating yang nantinya akan dipertimbangkan untuk menghitung kemiripan (*Cosine Similarity*) [36]. Berikut ini adalah Matriks Rating *User-Item User-Based Collaborative Filtering*:

Tabel 4.4 Matriks Rating *User-Item User-Collaborative Filtering*

| User | Item | | | | | | | | | | | |
|------|------|-----|-----|-----|-----|-----|-----|--------|--------|--------|--------|--------|
| | | 1 | 2 | 3 | 4 | 5 | ... | 193579 | 193581 | 193583 | 193585 | 193587 |
| | 1 | 4.0 | 0.0 | 4.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 5 | 4.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| | 606 | 3.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 607 | 4.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 608 | 0.0 | 2.0 | 2.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 609 | 3.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 610 | 5.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

4.3.4 Hasil Pencarian *Hyperparameter* Optimal dengan *GridSearchCV*

Pada tahap ini, model *User-Based Collaborative Filtering* akan dilakukan proses pencarian parameter terbaik menggunakan *GridSearchCV* pada data latih menggunakan skema ketiga skema pembagian data 80:20, 75:25 dan 95:5. dalam proses pencarian *hyperparameter* K terbaik, pemilihan nilai K menjadi penting untuk menentukan apakah model yang dibuat akan menghasilkan akurasi yang baik atau tidak. Pemilihan nilai K yang biasa digunakan oleh penelitian lain adalah $K=\{3,5,7,10\}$ [37] [38] [39] [40]. Nilai K tersebut digunakan dalam penelitian ini bertujuan untuk mengeksplorasi performa model pada rentang nilai yang sering digunakan dalam penelitian. Pada penelitian ini, rentang nilai K diperluas menjadi $K = \{1,2,3,...15\}$ untuk mengeksplorasi efek nilai K pada performa model secara lebih menyeluruh. Adapun berdasarkan penelitian sebelumnya

yang menggunakan rentang nilai untuk mencari *Hyperparameter K* dilakukan untuk melihat perbedaan penggunaan nilai *K* yang terlalu besar atau kecil [33].

GridSearchCV digunakan untuk mempermudah proses pencarian hyperparameter, dengan memanfaatkan *10 - fold Cross-Validation* ($CV=10$) untuk mengevaluasi performa model pada setiap kombinasi parameter dalam grid. Untuk mengevaluasi performa model, metrik *Mean Absolute Error (MAE)* yang akan digunakan dalam penelitian ini. *MAE* dipilih karena sifatnya yang intuitif, yakni memberikan gambaran rata-rata besarnya kesalahan absolut antara prediksi model dan nilai aktual. Nilai *MAE* yang lebih kecil menunjukkan performa model yang lebih baik, karena menunjukkan kesalahan prediksi yang lebih kecil.

Detail hasil evaluasi dari proses hasil pencarian *Hyperparameter K* pada model *User-Based Collaborative Filtering* untuk setiap nilai *K* dan skema *splitting data* yang diuji, disajikan pada Tabel 4.5 berikut:

Tabel 4.5 Hasil Rata-Rata Nilai *K* dalam *GridSearchCV*

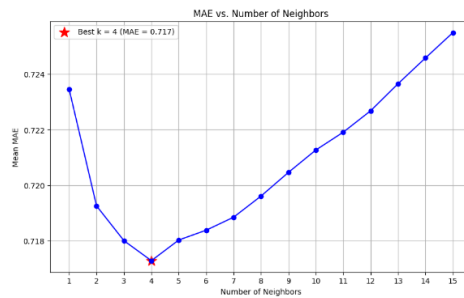
| Nilai <i>K</i> | Skema <i>Splitting</i> 80:20 | | Skema <i>Splitting</i> 75:25 | | Skema <i>Splitting</i> 95:5 | |
|----------------|------------------------------|-------------------------|------------------------------|-------------------------|-----------------------------|-------------------------|
| | Peringkat | Rata-Rata <i>MAE</i> | Peringkat | Rata-Rata <i>MAE</i> | Peringkat | Rata-Rata <i>MAE</i> |
| 1 | 12 | 0.723 | 13 | 0.724 | 10 | 0.720 |
| 2 | 6 | 0.719 | 7 | 0.719 | 5 | 0.716 |
| 3 | 2 | 0.718 | 2 | 0.717 | 2 | 0.715 |
| 4 | 1 | 0.717 | 1 | 0.717 | 3 | 0.715 |
| 5 | 3 | 0.718 | 3 | 0.716 | 1 | 0.715 |
| 6 | 4 | 0.718 | 4 | 0.717 | 4 | 0.716 |
| 7 | 5 | 0.718 | 5 | 0.718 | 6 | 0.717 |
| 8 | 7 | 0.719 | 6 | 0.719 | 7 | 0.718 |
| 9 | 8 | 0.720 | 8 | 0.721 | 8 | 0.718 |
| 10 | 9 | 0.721 | 9 | 0.722 | 9 | 0.719 |

| Nilai K | Skema <i>Splitting</i> 80:20 | | Skema <i>Splitting</i> 75:25 | | Skema <i>Splitting</i> 95:5 | |
|-----------|------------------------------|-------------------------|------------------------------|-------------------------|-----------------------------|-------------------------|
| | Peringkat | Rata-Rata <i>MAE</i> | Peringkat | Rata-Rata <i>MAE</i> | Peringkat | Rata-Rata <i>MAE</i> |
| 11 | 10 | 0.721 | 10 | 0.722 | 11 | 0.720 |
| 12 | 11 | 0.722 | 11 | 0.722 | 12 | 0.721 |
| 13 | 13 | 0.723 | 12 | 0.723 | 13 | 0.721 |
| 14 | 14 | 0.724 | 14 | 0.724 | 14 | 0.722 |
| 15 | 15 | 0.725 | 15 | 0.725 | 15 | 0.723 |

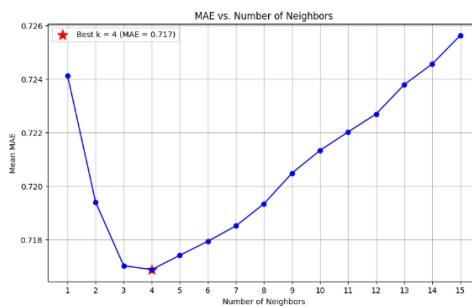
Berdasarkan hasil yang disajikan pada Tabel 4.5 analisis komparatif performa model *User-Based Collaborative Filtering* menunjukkan bahwa nilai K optimal cenderung berada pada rentang yang konsisten di ketiga skema *Splitting* data yang diuji. Secara spesifik, nilai $K = 4$ atau $K = 5$ menempati peringkat teratas dengan nilai *MAE* terendah untuk setiap skema. Meskipun demikian, skema pembagian data 95:5 secara keseluruhan menghasilkan *MAE* rata-rata yang sedikit lebih rendah (0.716) dibandingkan skema 80:20 dan 75:25 (0.717). namun nilai K yang didapatkan dari skema 95:5 sedikit lebih besar $K = 5$ yang mempengaruhi hasil prediksi rating. Hal ini mengindikasikan adanya potensi overfitting pada model yang dilatih dengan skema 95:5, dengan data latih yang sangat besar dan data uji yang sangat kecil model cenderung belajar terlalu spesifik pada pola yang ada di data latih, sehingga performanya terlihat sangat baik pada data uji yang kecil tersebut. Namun, data uji yang kecil ini mungkin tidak cukup representatif untuk mengukur kemampuan generalisasi model pada data sesungguhnya yang belum pernah dilihat [41]. Oleh karena itu, meskipun memberikan *MAE* yang sedikit lebih rendah, model dari skema 95:5 belum tentu menunjukkan kualitas prediksi yang lebih baik.

Untuk memvisualisasikan performa model, berdasarkan hasil yang ditampilkan pada Gambar 4.1, Gambar 4.2, dan Gambar 4.3 terlihat hasil pencarian *Hyperparameter K* pada model *User-Based Collaborative Filtering* dengan Skema *Splitting* 80:20, 75:25 dan 95:5. Ketiga grafik menunjukkan pola yang sama, yaitu nilai *MAE* cenderung

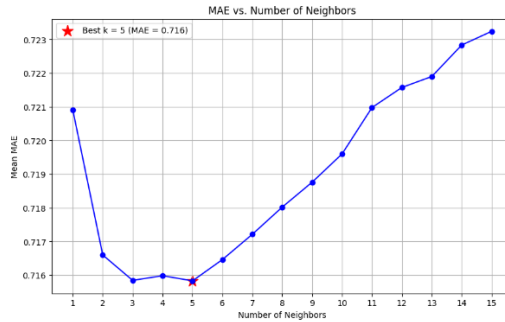
menurun seiring dengan peningkatan nilai K pada awalnya, lalu mencapai titik optimal, kemudian mulai meningkat kembali setelah melewati titik tersebut. Sesuai dengan Gambar 4.1 dengan Skema *Splitting* 80:20 menunjukkan bahwa nilai K optimal adalah $K = 4$, dengan nilai MAE terendah sebesar 0.717 pada nilai $K = 4$, begitu juga dengan Skema *Splitting* 75:25 pada nilai $K = 4$ dan Skema *Splitting* 95:5 pada nilai $K = 5$ model menunjukkan kesalahan rata-rata prediksi terendah dibandingkan dengan nilai K lainnya. Dari hasil grafik juga menunjukkan bahwa seiring bertambahnya nilai K , performa model membaik hingga mencapai nilai minimal K , setelah titik tersebut, nilai MAE cenderung meningkat, yang mengindikasikan bahwa jumlah penambahan K akan mengurangi performa model.



Gambar 4.1 Hasil Pencarian *Hyperparameter* K dengan Skema *Splitting* 80:20



Gambar 4.2 Hasil Pencarian *Hyperparameter* K dengan Skema *Splitting* 75:25



Gambar 4.3 Hasil Pencarian *Hyperparameter K* dengan Skema *Splitting 95:5*

4.3.5 Pelatihan Model *User-Based Collaborative Filtering*

Setelah diperoleh *Hyperparameter K* optimal dari proses *GridSearchCV*, langkah selanjutnya adalah melakukan pelatihan ulang dari model *User-Based Collaborative Filtering* menggunakan data latih dari Skema *Splitting 80:20*, *75:25* dan *95:5*. Nilai *K* optimal yang diperoleh melalui *GridSearchCV* dengan skema Skema *Splitting 80:20* yaitu $K = 4$, Skema *Splitting 75:25* dengan $K = 4$ dan Skema *Splitting 95:5* dengan $K = 5$. Pelatihan ulang model ini dilakukan pada seluruh data latih untuk memanfaatkan dan menangkap pola yang lebih baik dari data latih, sehingga dapat memastikan kemampuan generalisasi yang lebih baik. Model yang telah dilatih ini selanjutnya akan dievaluasi menggunakan data uji dengan *Mean Absolute Error*.

4.3.6 Nilai *Cosine Similarity User-Based Collaborative Filtering*

Pada tahap ini, nilai *Cosine Similarity* digunakan untuk menghitung kemiripan antar user berdasarkan rating yang diberikan terhadap item. Tujuan dari *Cosine Similarity* adalah untuk mengidentifikasi pengguna yang memiliki pola preferensi yang mirip berdasarkan data rating. Nilai *Cosine Similarity* mendekati 1 menunjukkan bahwa pengguna memiliki preferensi yang mirip berdasarkan pola rating mereka terhadap item. Sebaliknya, nilai mendekati 0 menunjukkan bahwa pengguna memiliki preferensi yang berbeda atau bertentangan [30]. Hasil *Cosine Similarity* untuk model *User-Based Collaborative Filtering* dari masing-masing skema *splitting 80:20*, *75:25* dan *95:5* dapat dilihat dalam Tabel 4.6, Tabel 4.7, dan Tabel 4.8 berikut:

Tabel 4.6 *Cosine Similarity User-Based Collaborative Filtering*
Skema Splitting 80:20

| | 1 | 2 | 3 | ... | 608 | 609 | 610 |
|-----|----------|----------|----------|-----|----------|----------|----------|
| 1 | 1.000000 | 0.000000 | 0.914576 | ... | 0.964554 | 0.978232 | 0.981634 |
| 2 | 0.000000 | 1.000000 | 0.000000 | ... | 0.978709 | 0.000000 | 0.989127 |
| 3 | 0.914576 | 0.000000 | 1.000000 | ... | 0.759072 | 0.000000 | 0.903327 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 608 | 0.964554 | 0.978709 | 0.759072 | ... | 1.000000 | 0.977049 | 0.958838 |
| 609 | 0.978232 | 0.000000 | 0.000000 | ... | 0.977049 | 1.000000 | 0.966782 |
| 610 | 0.981634 | 0.989127 | 0.903327 | ... | 0.958838 | 0.966782 | 1.000000 |

Tabel 4.7 *Cosine Similarity User-Based Collaborative Filtering*
Skema Splitting 75:25

| | 1 | 2 | 3 | ... | 608 | 609 | 610 |
|-----|----------|----------|----------|-----|----------|----------|----------|
| 1 | 1.000000 | 0.000000 | 0.914576 | ... | 0.963786 | 0.992915 | 0.980067 |
| 2 | 0.000000 | 1.000000 | 0.000000 | ... | 0.972229 | 0.000000 | 0.987215 |
| 3 | 0.914576 | 0.000000 | 1.000000 | ... | 0.759072 | 0.000000 | 0.842712 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 608 | 0.963786 | 0.978709 | 0.759072 | ... | 1.000000 | 0.977049 | 0.958193 |
| 609 | 0.992915 | 0.000000 | 0.000000 | ... | 0.977049 | 1.000000 | 0.956681 |
| 610 | 0.980067 | 0.987215 | 0.842712 | ... | 0.958193 | 0.956681 | 1.000000 |

Tabel 4.8 *Cosine Similarity User-Based Collaborative Filtering*
Skema Splitting 95:5

| | 1 | 2 | 3 | ... | 608 | 609 | 610 |
|-----|----------|----------|----------|-----|----------|----------|----------|
| 1 | 1.000000 | 0.000000 | 0.844826 | ... | 0.962036 | 0.978232 | 0.980541 |
| 2 | 0.000000 | 1.000000 | 0.000000 | ... | 0.980101 | 1.000000 | 0.989683 |
| 3 | 0.844826 | 0.000000 | 1.000000 | ... | 0.763542 | 0.000000 | 0.903327 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 608 | 0.962036 | 0.980101 | 0.763542 | ... | 1.000000 | 0.978922 | 0.962687 |
| 609 | 0.975417 | 1.000000 | 0.000000 | ... | 0.978922 | 1.000000 | 0.960322 |
| 610 | 0.980541 | 0.989683 | 0.903327 | ... | 0.962687 | 0.960322 | 1.000000 |

Analisis terhadap Tabel 4.6, Tabel 4.7 dan Tabel 4.8, yang menampilkan sebagian dari representasi dari nilai *Cosine Similarity* menunjukkan adanya tingkat kemiripan antar pengguna. Nilai yang bervariasi dari 0-1 menunjukkan luasnya preferensi pengguna. Nilai 1 pada diagonal pada setiap tabel menunjukkan identitas pengguna dengan dirinya sendiri. Kemunculan nilai 0 untuk beberapa pasangan pengguna seperti pengguna 2 dengan pengguna 1, pengguna 3 dan

pengguna 609 secara konsisten sama di ketiga skema yang menunjukkan bahwa pasangan pengguna tersebut memiliki preferensi yang sangat berbeda, yang mencerminkan tidak adanya kesamaan pola rating yang teridentifikasi. Pola kemiripan antar pengguna di ketiga skema pembagian data cenderung konsisten. Meskipun terdapat sedikit variasi pada nilai *Cosine Similarity* untuk beberapa pasangan pengguna antar skema. misalnya pengguna 1 dan pengguna 3 pada skema 95:5 sedikit berbeda dari skema lainnya. Hal ini dikarenakan jumlah data latih yang digunakan berbeda, sehingga memungkinkan adanya perbedaan dari hasil *Cosine Similarity* yang disajikan. Namun hal ini juga menegaskan bahwa konsistensi dari nilai *Cosine Similarity* dari ketiga skema menunjukkan bahwa *Cosine Similarity* mampu mengidentifikasi hubungan antar pengguna.

4.3.7 Prediksi Rating *User-Based Collaborative Filtering*

Setelah menghitung kemiripan antar pengguna menggunakan *Cosine Similarity* dan mendapatkan *Hyperparameter K* optimal, Tahap selanjutnya adalah melakukan prediksi rating menggunakan data uji pada model *User-Based Collaborative Filtering*. Hasil prediksi rating dibandingkan dengan nilai rating aktual dari data uji untuk mengevaluasi performa model dalam memprediksi preferensi *user* [42].

Dalam memprediksi rating, prediksi dilakukan menggunakan rumus prediksi rating yang memanfaatkan rating dari pengguna terdekat [43]. Prediksi rating dalam sistem rekomendasi dihitung menggunakan kontribusi dari tetangga terdekat (K) yang dipilih berdasarkan nilai kemiripan (*Cosine Similarity*) antar pengguna [44]. Nilai K merepresentasikan jumlah tetangga terdekat yang digunakan untuk memprediksi rating pengguna [45]. Tetangga dengan nilai kemiripan (*Cosine Similarity*) yang lebih tinggi memiliki bobot lebih besar dalam perhitungan prediksi, sehingga preferensi mereka lebih memengaruhi hasil prediksi rating pengguna.

Contoh hasil prediksi rating untuk pengguna dengan *userId* 1 pada ketiga skema *splitting* disajikan dalam Tabel 4.9 (skema *splitting* 80:20), Tabel 4.10 (skema *splitting* 75:25), dan Tabel 4.11 (skema *splitting* 95:5). Dari tabel-tabel yang disajikan ini, dapat diamati

perbandingan antara rating aktual dengan rating yang diprediksi oleh model untuk item-item yang telah dinilai pengguna 1. Pada Tabel 4.9, terlihat hasil prediksi rating yang dilakukan ketika dibandingkan dengan rating aktual, sebagai contoh *user* 1 memberi rating 5.0 pada film *Fight Club (1999)* dengan genre *Action/Crime/Drama/Thriller*, hasil prediksi rating yang dilakukan model adalah 4.700324. Begitupun dengan hasil prediksi dengan film yang memiliki genre serupa menunjukkan bahwa Perbandingan antara rating aktual dan prediksi pada *User-Based Collaborative Filtering* tersebut mendekati nilai aktual yang mana model ini mampu memberikan prediksi yang cukup akurat.

Tabel 4.9 Prediksi Rating *User-Based Collaborative Filtering* Skema *Splitting 80:20*

| <i>userId</i> | <i>movieId</i> | <i>actual_rating</i> | <i>predicted_rating</i> | <i>title</i> | <i>genres</i> |
|---------------|----------------|----------------------|-------------------------|---|------------------------------------|
| 1 | 2959 | 5.0 | 4.700324 | <i>Fight Club (1999)</i> | <i>Action/Crime/Drama/Thriller</i> |
| 1 | 1213 | 5.0 | 4.650256 | <i>Goodfellas (1990)</i> | <i>Crime/Drama/</i> |
| 1 | 527 | 5.0 | 4.576333 | <i>Schinder's List (1993)</i> | <i>Drama\War</i> |
| 1 | 260 | 5.0 | 4.549563 | <i>Star Wars : Episode IV – A New Hope (1997)</i> | <i>Action/Adventure/Sci-Fi</i> |
| 1 | 2502 | 5.0 | 4.448098 | <i>Office Space (1999)</i> | <i>Comedy/Crime</i> |

Tabel 4.10 Prediksi Rating *User-Based Collaborative Filtering* Skema *Splitting 75:25*

| <i>userId</i> | <i>movieId</i> | <i>actual_rating</i> | <i>predicted_rating</i> | <i>title</i> | <i>genres</i> |
|---------------|----------------|----------------------|-------------------------|--------------------------|------------------------------------|
| 1 | 2959 | 5.0 | 4.725102 | <i>Fight Club (1999)</i> | <i>Action/Crime/Drama/Thriller</i> |

| <i>userId</i> | <i>movieId</i> | <i>actual_rating</i> | <i>predicted_rating</i> | <i>title</i> | <i>genres</i> |
|---------------|----------------|----------------------|-------------------------|-------------------------------|------------------------------------|
| 1 | 1213 | 5.0 | 4.600103 | <i>Goodfellas (1990)</i> | <i>Crime/Drama/</i> |
| 1 | 527 | 5.0 | 4.600013 | <i>Schinder's List (1993)</i> | <i>Drama\War</i> |
| 1 | 608 | 5.0 | 4.550087 | <i>Fargo (1996)</i> | <i>Comedy/Crime/Drama/Thriller</i> |
| 1 | 2858 | 5.0 | 4.499951 | <i>American Beauty (1999)</i> | <i>Drama/Romance</i> |

Tabel 4.11 Prediksi Rating *User-Based Collaborative Filtering*
Skema *Splitting 95:25*

| <i>userId</i> | <i>movieId</i> | <i>actual_rating</i> | <i>predicted_rating</i> | <i>title</i> | <i>genres</i> |
|---------------|----------------|----------------------|-------------------------|------------------------------------|------------------------------------|
| 1 | 5060 | 5.0 | 4.027641 | <i>MASH (1970)</i> | <i>Comedy/Drama/War</i> |
| 1 | 2353 | 5.0 | 4.025953 | <i>Enemy of the States (1998)</i> | <i>Action/Thriller</i> |
| 1 | 3448 | 5.0 | 4.001142 | <i>Vietnam Good Morning (1987)</i> | <i>Comedy/Drama/War</i> |
| 1 | 3729 | 5.0 | 3.758054 | <i>Shaft (1971)</i> | <i>Action/Crime/Drama/Thriller</i> |
| 1 | 3386 | 5.0 | 3.620378 | <i>JFK (1991)</i> | <i>Drama/Mystery/Thriller</i> |

Pada Tabel 4.9 dan Tabel 4.10 yang masing-masing merepresentasikan skema *splitting* 80:20 dan 75:25, menunjukkan bahwa prediksi rating berada dalam rentang yang relatif dekat dengan nilai aktual 5.0. Hal ini

mengindikasikan bahwa model pada kedua skema ini mampu memberikan prediksi yang cukup akurat dan mendekati preferensi aktual pengguna. Namun, pada Tabel 4.11 yang merepresentasikan skema *splitting* 95:5, menunjukkan pola yang berbeda, nilai prediksi rating untuk film-film dengan rating aktual 5.0 pada skema ini cenderung lebih rendah, yang berada dalam rentang 3.6 – 4.0. Hal ini menegaskan analisis sebelumnya mengenai potensi *overfitting* pada model yang dilatih dengan skema 95:5, meskipun skema ini menghasilkan nilai *MAE* terendah dalam proses *GridSearchCV* seperti yang ditunjukkan pada Tabel 4.5, hasil prediksi ketika dibandingkan nilai aktualnya menunjukkan akurasi yang kurang baik. ini mengindikasikan bahwa ketika model dilatih dengan 95% data yang besar dan diuji pada data uji yang kecil 5%, mungkin terlalu spesifik mempelajari pola dari data latih sehingga kemampuannya untuk menggeneralisasi data yang belum pernah dilihat.

Dapat disimpulkan bahwa dari hasil prediksi ini, skema *splitting* 80:20, dan 75:25 yang menunjukkan pola serupa, meskipun memiliki *MAE* yang sedikit lebih tinggi pada *GridSearchCV* dibandingkan skema 95:5, secara efektif mampu menghasilkan model dengan prediksi rating yang lebih baik dan lebih mendekati nilai aktual. Hal ini memperkuat justifikasi skema 80:20 sebagai skema yang optimal untuk model *User-Based Collaborative Filtering* dalam penelitian ini.

4.4 Hasil Evaluasi Model User-Based Collaborative Filtering

Evaluasi model dilakukan untuk mengukur performa sistem rekomendasi dalam memprediksi rating pengguna [42]. Pada penelitian ini, pengujian dilakukan menggunakan data uji dari *User-Based Collaborative Filtering* dalam pemodelan berdasarkan ketiga skema *splitting* (80:20, 75:25 dan 95:5) yang telah dilakukan di tahap sebelumnya. Kinerja model diukur menggunakan metrik *Mean Absolute Error (MAE)* yang menunjukkan rata-rata kesalahan prediksi rating dengan rating aktual [32]. Hasil evaluasi *MAE* untuk model *User-Based Collaborative Filtering* pada nilai *K* optimal dari masing masing skema *splitting* didapatkan hasil bahwa dari skema *splitting* 80:20 dengan nilai $K = 4$, didapatkan *MAE* 0.7236, sementara itu, skema *splitting* 75:25 menghasilkan *MAE* 0.7256 dengan nilai $K = 4$, Adapun

skema 95:5 menunjukkan nilai *MAE* terendah dengan nilai *MAE* 0.7099 dengan nilai $K = 5$.

Umumnya nilai *MAE* yang lebih kecil dari 1.0 dianggap cukup baik dalam sistem rekomendasi berbasis rating dengan skala 1-5, karena kesalahan prediksi rata-rata masih berada dalam rentang kurang dari satu poin rating. Nilai *MAE* dalam penelitian ini, khususnya pada skema *splitting* 80:20 menunjukkan bahwa rata-rata kesalahan prediksi hanya sekitar 0.72-0.73 poin dari rating aktual pengguna. Ini berarti bahwa jika seorang pengguna memberikan rating 4 untuk suatu item, maka prediksi model rata-rata berada dalam rentang 3.3 hingga 4.7, yang masih cukup mendekati nilai aslinya.

Berdasarkan Tabel 4.11 sebagaimana yang telah dijelaskan sebelumnya, meskipun skema 95:5 menghasilkan nilai *MAE* terendah, skema tersebut terindikasi memiliki potensi overfitting. Hal ini dibuktikan oleh kualitas prediksi rating yang dihasilkan kurang optimal dibandingkan dengan skema *splitting* 80:20. Model yang dilatih dengan 95% data latih dan hanya diuji pada 5% mungkin bekerja sangat baik pada data uji yang sedikit tersebut, tapi gagal menggeneralisasi dengan baik. Walaupun skema *splitting* 80:20 dan 75:25 menunjukkan nilai *MAE* sedikit lebih tinggi, dari skema *splitting* 95:25, namun memberikan prediksi rating yang baik dan dapat diandalkan. Adapun nilai K yang didapatkan juga berpengaruh pada skema *splitting* 95:5, dikarenakan nilai K yang didapatkan sedikit lebih tinggi daripada skema *splitting* 80:20 dan 75:25. Model *User-Based Collaborative Filtering* untuk skema *splitting* 80:20 sudah efektif untuk model *User-Based Collaborative Filtering*. Hal ini dibuktikan dengan nilai *MAE* yang rendah yaitu $K = 4$ *MAE* 0.7236 pada skema *splitting* 80:20. Nilai *MAE* yang mendekati 0 menunjukkan bahwa rata-rata kesalahan prediksi cukup baik, sehingga sistem rekomendasi mampu memberikan prediksi rating yang baik dalam memberikan rekomendasi [42].

4.5 Hasil Rekomendasi User-Based Collaborative Filtering

Hasil rekomendasi menunjukkan bahwa setiap pengguna menerima daftar film berbeda berdasarkan prediksi rating yang paling tinggi [42]. Hasil Rekomendasi yang dihasilkan merupakan film yang belum pernah dilihat dan diberi rating oleh pengguna pada data uji. Hasil

rekomendasi menunjukkan bahwa model mampu menangkap pola preferensi pengguna berdasarkan kontribusi tetangga terdekat [46]. Dari hasil distribusi prediksi rating, menunjukkan bahwa model cenderung memberikan prediksi tinggi untuk item yang direkomendasikan.

Contoh hasil rekomendasi film untuk pengguna dengan *userId* 1, berdasarkan model *User-Based Collaborative Filtering* dari masing-masing skema *splitting*, disajikan dalam Tabel 4.12 (skema *splitting* 80:20), Tabel 4.13 (skema *splitting* 75:25, dan Tabel 4.12 (skema *splitting* 95:5). Berikut adalah masing-masing tabel hasil rekomendasi berdasarkan skema *splitting* 80:20,75:25, dan 95:5:

Tabel 4.12 Hasil Rekomendasi *User-Based Collaborative Filtering*
Skema *Splitting* 80:20.

| <i>userId</i> | <i>movieId</i> | <i>predicted_ratings</i> | <i>title</i> | <i>genres</i> |
|---------------|----------------|--------------------------|---|--|
| 1 | 177593 | 5.000000 | <i>Missouri, Three Billboards Outside Ebbing</i> (2017) | Crime/Drama |
| 1 | 158966 | 5.000000 | <i>Captain Fantastic</i> (2016) | Drama |
| 1 | 6460 | 5.000000 | <i>The Trial</i> (1962) | Drama |
| 1 | 43376 | 5.000000 | <i>Sophie Scholl: The Final Days</i> (2005) | Drama/War |
| 1 | 1178 | 4.872578 | <i>Paths of Glory</i> (1957) | Drama/War |
| 1 | 106642 | 4.832174 | <i>The Day of the Doctor</i> (2013) | Adventure/Drama/Sci-Fi |
| 1 | 1446 | 4.807197 | <i>Kolya</i> (1996) | Comedy/Drama |
| 1 | 31364 | 4.798842 | <i>Memories of Murder</i> (2003) | Action/Adventure/Animation /Crine/Fantasy |
| 1 | 436 | 4.754507 | <i>Color of Night</i> (1994) | Action/Animation/Drama/Fantasy/Sci-Fi |
| 1 | 55167 | 4.753992 | <i>Tekkonkinkreet</i> (2006) | Comedy |

Tabel 4.13 Hasil Rekomendasi *User-Based Collaborative Filtering*
Skema *Splitting 75:25*.

| <i>userId</i> | <i>movieId</i> | <i>predicted_ratings</i> | <i>title</i> | <i>genres</i> |
|---------------|----------------|--------------------------|---|--|
| 1 | 177593 | 5.000000 | <i>Missouri, Three Billboards Outside Ebbing</i> (2017) | <i>Crime/Drama</i> |
| 1 | 158966 | 5.000000 | <i>Captain Fantastic</i> (2016) | <i>Drama</i> |
| 1 | 6460 | 5.000000 | <i>The Trial</i> (1962) | <i>Drama</i> |
| 1 | 43376 | 5.000000 | <i>Sophie Scholl: The Final Days</i> (2005) | <i>Drama/War</i> |
| 1 | 1178 | 4.855959 | <i>Paths of Glory</i> (1957) | <i>Drama/War</i> |
| 1 | 106642 | 4.829463 | <i>The Day of the Doctor</i> (2013) | <i>Adventure/Drama/Sci-Fi</i> |
| 1 | 1446 | 4.809049 | <i>Kolya</i> (1996) | <i>Comedy/Drama</i> |
| 1 | 31364 | 4.797646 | <i>Memories of Murder</i> (2003) | <i>Action/Adventure/Animation</i> <i>/Crine/Fantasy</i> |
| 1 | 436 | 4.755841 | <i>Color of Night</i> (1994) | <i>Action/Animation/Drama/</i> <i>Fantasy/Sci-Fi</i> |
| 1 | 55167 | 4.751750 | <i>Tekkonkinkreet</i> (2006) | <i>Comedy</i> |

Tabel 4.14 Hasil Rekomendasi *User-Based Collaborative Filtering*
Skema *Splitting 95:5*.

| <i>userId</i> | <i>movieId</i> | <i>predicted_ratings</i> | <i>title</i> | <i>genres</i> |
|---------------|----------------|--------------------------|---------------------------------|---------------|
| 1 | 158966 | 5.000000 | <i>Captain Fantastic</i> (2016) | <i>Drama</i> |

| <i>userId</i> | <i>movieId</i> | <i>predicted_ratings</i> | <i>title</i> | <i>genres</i> |
|---------------|----------------|--------------------------|---|-------------------------------------|
| 1 | 177593 | 5.000000 | <i>Missouri, Three Billboards Outside Ebbing (2017)</i> | <i>Crime/Drama</i> |
| 1 | 6460 | 5.000000 | <i>The Trial (1962)</i> | <i>Drama</i> |
| 1 | 31364 | 4.800457 | <i>Memories of Murder (2003)</i> | <i>Crime/Drama/Mystery/Thriller</i> |
| 1 | 318 | 4.723987 | <i>The Shawshank Redemption (1994)</i> | <i>Crime/Drama</i> |
| 1 | 8132 | 4.713684 | <i>Gladiator</i> | <i>Action/Drama</i> |
| 1 | 106642 | 4.709896 | <i>The Day of the Doctor (2013)</i> | <i>Adventure/Drama/Sci-Fi</i> |
| 1 | 58559 | 4.700220 | <i>The Dark Knight (2008)</i> | <i>Action/Crime/Drama</i> |
| 1 | 3030 | 4.700206 | <i>Yojimbo (1961)</i> | <i>Action/Adventure</i> |
| 1 | 7234 | 4.669665 | <i>La Strada (1954)</i> | <i>Drama</i> |

Pada Tabel 4.12, Tabel 4.13 dan Tabel 4.14, hasil rekomendasi pada model *User-Based Collaborative Filtering* berdasarkan ketiga skema *splitting*, model merekomendasikan film dengan prediksi rating tinggi pada film *Missouri, Three Billboards Outside Ebbing (2017)*, *Captain Fantastic (2016)*, dan *The Trial (1962)* dengan prediksi 5.0 muncul di ketiga film teratas di masing-masing skema *splitting*, konsisten dengan film yang sudah ditonton oleh *userId* 1 pada Tabel 4.9 yang lebih banyak menonton film bergenre *Drama*. Film yang direkomendasikan menunjukkan relevansi dengan pola rating pengguna serupa, yang mana tetangga terdekat cenderung memberikan rating tinggi pada film tersebut. hal itu tercermin dari hasil prediksi rating yang dilakukan untuk menghasilkan rekomendasi, nilai $K=4$ atau 4 tetangga yang digunakan dalam memprediksi rating lebih baik dibandingkan dengan nilai $K = 5$ seperti yang didapatkan pada skema *splitting* 95:5.

Dengan mempertimbangkan keseluruhan hasil evaluasi, dan hasil rekomendasi yang didapatkan, skema *splitting* 80:20 merupakan skema *splitting* terbaik untuk model *User-Based Collaborative Filtering* dalam penelitian ini. Meskipun skema *splitting* 95:5 menunjukkan hasil evaluasi *MAE* terendah (0.7099), perbedaannya dengan *MAE* skema *splitting* 80:20 (0.7236) relatif kecil, pun juga dengan skema *splitting* 75:25 meskipun hasil evaluasi yang didapatkan tidak jauh berbeda dengan skema *splitting* 80:20 yaitu (0.7256), namun dari hasil prediksi rating dan hasil rekomendasi yang didapatkan masih lebih unggul skema *splitting* 80:20 dalam memberikan hasil rekomendasi. Oleh karena itu, pemilihan skema *splitting* 80:20 dengan $K = 4$ merupakan hasil yang terbaik dalam memberikan prediksi rating dan hasil rekomendasi yang didapatkan.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian “Sistem Rekomendasi Film menggunakan *Neighborhood-Based Collaborative Filtering*” dapat ditarik kesimpulan:

1. Penelitian ini berhasil mengimplementasikan *Neighborhood-Based Collaborative Filtering* untuk membangun sistem rekomendasi film yang relevan. Sistem rekomendasi ini memanfaatkan *User-Collaborative Filtering* dengan menghitung kemiripan antar pengguna menggunakan *Cosine Similarity* untuk menemukan tetangga terdekat, lalu menggunakan rating tetangga terdekat tersebut untuk memprediksi rating film yang belum dirating oleh pengguna untuk memberikan rekomendasi.
2. Evaluasi model menggunakan metrik *Mean Absolute Error (MAE)* menunjukkan kemampuan sistem dalam memberikan prediksi rating yang akurat. Melalui proses *GridSearchCV*, ditemukan *Hyperparameter* optimal K untuk model *User-Based Collaborative Filtering*. Hasil evaluasi *MAE* menunjukkan bahwa untuk skema splitting terbaik dari ketiga skema splitting yang dilakukan adalah skema 80:20 dengan nilai *MAE* yang diperoleh adalah 0.7236 dengan nilai $K = 4$, nilai *MAE* yang rendah menunjukkan bahwa sistem rekomendasi ini berhasil menghasilkan prediksi yang sesuai dengan preferensi pengguna.

5.2 Saran

Berdasarkan kesimpulan dari hasil dan pembahasan pada penelitian ini, berikut ini adalah saran yang dapat digunakan dalam penelitian selanjutnya:

Penelitian ini menggunakan *Neighborhood-Based Collaborative Filtering* dengan *User-Based Collaborative Filtering* dan *Cosine Similarity* untuk menghitung kemiripan antar pengguna dan menghasilkan rekomendasi. Penelitian selanjutnya disarankan untuk mengeksplorasi penggunaan algoritma yang lebih kompleks seperti

Hybrid Filtering, Matrix Factorization, Neural Collaborative Filtering, atau metode berbasis *Deep Learning* lainnya dalam menangani dataset yang lebih besar yang mampu meningkatkan akurasi prediksi dan memberikan rekomendasi yang lebih relevan dan personal.

DAFTAR PUSTAKA

- [1] I. G. A. D. Hendriyani, “Siaran Pers: Film Jadi Salah Satu Subsektor Ekonomi Kreatif yang Tumbuh Positif pada 2023,” Kemenparekraf/Baparekraf RI. Accessed: Jul. 10, 2024. [Online]. Available: <https://kemenparekraf.go.id/berita/siaran-pers-film-jadi-salah-satu-subsektor-ekonomi-kreatif-yang-tumbuh-positif-pada-2023>
- [2] S. Ambarwati, “Kontribusi film pada PDB 2024 diproyeksi capai Rp3,41 triliun,” ANTARA NEWS. Accessed: Jul. 12, 2024. [Online]. Available: <https://www.antaranews.com/berita/4035054/kontribusi-film-pada-pdb-2024-diproyeksi-capai-rp341-triliun>
- [3] K. Baparekraf RI, “Rumah Produksi Film Indonesia yang Sukses Tembus Box Office,” Kemenparekraf/Baparekraf RI. Accessed: Jul. 12, 2024. [Online]. Available: <https://www.kemenparekraf.go.id/ragam-ekonomi-kreatif/rumah-produksi-film-indonesia-yang-sukses-tembus-box-office>
- [4] J. Lemantara, “Sistem Pendukung Keputusan Penentuan Film Berdasarkan Minat dengan Metode Weighted Product,” vol. 8, 2023.
- [5] A. R. Fitrianti, A. Rohmani, and W. Widjanarto, “Sistem Rekomendasi Film Berbasis Website Dengan Metode Prototype Menggunakan Algoritma K-Nearest Neighbors (KNN),” *JOINS J. Inf. Syst.*, vol. 5, no. 2, pp. 278–287, Nov. 2020, doi: 10.33633/joins.v5i2.4168.
- [6] D. R. H. Klavert, “Perbedaan Preferensi Konsumen Terhadap Atribut Film”.
- [7] geeksforgeeks, “Recommender Systems using KNN,” geeksforgeeks. Accessed: Jul. 23, 2024. [Online]. Available: <https://www.geeksforgeeks.org/recommender-systems-using-knn/>
- [8] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, “GroupLens: an open architecture for collaborative filtering of netnews,” in *Proceedings of the 1994 ACM conference on Computer supported cooperative work - CSCW '94*, Chapel

- Hill, North Carolina, United States: ACM Press, 1994, pp. 175–186. doi: 10.1145/192844.192905.
- [9] W. Hill, L. Stead, M. Rosenstein, and G. Furnas, “Recommending and evaluating choices in a virtual community of use,” in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '95*, Denver, Colorado, United States: ACM Press, 1995, pp. 194–201. doi: 10.1145/223904.223929.
 - [10] J. Herlocker, J. A. Konstan, and J. Riedl, “An Empirical Analysis of Design Choices in Neighborhood-Based Collaborative Filtering Algorithms,” *Inf. Retr.*, vol. 5, no. 4, pp. 287–310, 2002, doi: 10.1023/A:1020443909834.
 - [11] N. M. Roziqiin, “SISTEM REKOMENDASI PEMILIHAN ANIME MENGGUNAKAN USER-BASED COLLABORATIVE FILTERING,” *JIPI J. Ilm. Penelit. Dan Pembelajaran Inform.*, vol. 9, no. 1, pp. 299–306, Mar. 2024.
 - [12] A. A. Fakhri, Z. K. A. Baizal, and E. B. Setiawan, “Restaurant Recommender System Using User-Based Collaborative Filtering Approach: A Case Study at Bandung Raya Region,” *J. Phys. Conf. Ser.*, vol. 1192, p. 012023, Mar. 2019, doi: 10.1088/1742-6596/1192/1/012023.
 - [13] E. A. Rahayu Kasim, S. S, N. Ransi, and Isnawaty, “Sistem Rekomendasi Produk UMKM Menggunakan Algoritma User-Based Collaborative Filtering Berbasis Website,” *SISFOTENIKA*, vol. 14, no. 2, pp. 152–162.
 - [14] D. Roy and M. Dutta, “A systematic review and research perspective on recommender systems,” *J. Big Data*, vol. 9, no. 1, p. 59, Dec. 2022, doi: 10.1186/s40537-022-00592-5.
 - [15] P. M. Alamdari, N. J. Navimipour, M. Hosseinzadeh, A. A. Safaei, and A. Darwesh, “A Systematic Study on the Recommender Systems in the E-Commerce,” *IEEE Access*, vol. 8, pp. 115694–115716, 2020, doi: 10.1109/ACCESS.2020.3002803.
 - [16] A. Anandhan, L. Shuib, M. A. Ismail, and G. Mujtaba, “Social Media Recommender Systems: Review and Open Research Issues,” *IEEE Access*, vol. 6, pp. 15608–15628, 2018, doi: 10.1109/ACCESS.2018.2810062.

- [17] M. Schedl, H. Zamani, C.-W. Chen, Y. Deldjoo, and M. Elahi, "Current challenges and visions in music recommender systems research," *Int. J. Multimed. Inf. Retr.*, vol. 7, no. 2, pp. 95–116, Jun. 2018, doi: 10.1007/s13735-018-0154-2.
- [18] C. A. Gomez-Urbe and N. Hunt, "The Netflix Recommender System: Algorithms, Business Value, and Innovation," *ACM Trans. Manag. Inf. Syst.*, vol. 6, no. 4, pp. 1–19, Jan. 2016, doi: 10.1145/2843948.
- [19] F. Ricci, L. Rokach, and B. Shapira, "Introduction to Recommender Systems Handbook," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds., Boston, MA: Springer US, 2011, pp. 1–35. doi: 10.1007/978-0-387-85820-3_1.
- [20] P. Ajeng Swari Sukmawati, L. Hiryanto, and V. Christanti Mawardi, "Implementasi Metode Collaborative Filtering Based Untuk Sistem Rekomendasi Buku Fiksi," *J. Ilmu Komput. Dan Sist. Inf.*, vol. 11, no. 2, Aug. 2023, doi: 10.24912/jiksi.v11i2.25999.
- [21] D. Ferreira, S. Silva, A. Abelha, and J. Machado, "Recommendation System Using Autoencoders," *Appl. Sci.*, vol. 10, no. 16, p. 5510, Aug. 2020, doi: 10.3390/app10165510.
- [22] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds., *Recommender Systems Handbook*. Boston, MA: Springer US, 2011. doi: 10.1007/978-0-387-85820-3.
- [23] C. C. Aggarwal, *Recommender Systems*. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-29659-3.
- [24] D. Nugraha, T. W. Purboyo, and R. A. Nugrahaeni, "SISTEM REKOMENDASI FILM MENGGUNAKAN METODE USER BASED COLLABORATIVE FILTERING".
- [25] T. Arsan, E. Koksall, and Z. Bozkus, "Comparison of Collaborative Filtering Algorithms with Various Similarity Measures for Movie Recommendation," *Int. J. Comput. Sci. Eng. Appl.*, vol. 6, no. 3, pp. 1–20, Jun. 2016, doi: 10.5121/ijcsea.2016.6301.

- [26] IBM, “What is collaborative filtering?” Accessed: Apr. 21, 2025. [Online]. Available: <https://www.ibm.com/think/topics/collaborative-filtering>
- [27] geeksforgeeks, “Item-to-Item Based Collaborative Filtering,” geeksforgeeks. Accessed: Jul. 23, 2024. [Online]. Available: <https://www.geeksforgeeks.org/item-to-item-based-collaborative-filtering/>
- [28] L. V. Nguyen, Q.-T. Vo, and T.-H. Nguyen, “Adaptive KNN-Based Extended Collaborative Filtering Recommendation Services,” *Big Data Cogn. Comput.*, vol. 7, no. 2, p. 106, May 2023, doi: 10.3390/bdcc7020106.
- [29] D. A. R. Ariantini, A. S. M. Lumenta, and A. Jacobus, “PENGUKURAN KEMIRIPAN DOKUMEN TEKS BAHASA INDONESIA MENGGUNAKAN METODE COSINE SIMILARITY,” *J. Tek. Inform.*, vol. 9, no. 1, Oct. 2016, doi: 10.35793/jti.9.1.2016.13752.
- [30] D. S. A. Tambunan and D. A. Dermawan, “Implementasi Content Based Filtering dalam Sistem Rekomendasi Jurnal Scopus Berbasis Web Untuk Menunjang Pelaksanaan Penelitian dan Tugas Mahasiswa Program Studi Bisnis Digital”.
- [31] I. M. N. Yudiskara, I. G. S. Astawa, L. G. Astuti, M. A. Raharja, I. M. Widiartha, and I. W. Supriana, “Sistem Rekomendasi Film Dengan Item-Based Collaborative Filtering Menggunakan Flask Framework,” *JELIKU J. Elektron. Ilmu Komput. Udayana*, vol. 11, no. 3, p. 587, Jul. 2022, doi: 10.24843/JLK.2023.v11.i03.p15.
- [32] A. A. Suryanto, “PENERAPAN METODE MEAN ABSOLUTE ERROR (MEA) DALAM ALGORITMA REGRESI LINEAR UNTUK PREDIKSI PRODUKSI PADI,” *SAINTEKBU*, vol. 11, no. 1, pp. 78–83, Feb. 2019, doi: 10.32764/saintekbu.v11i1.298.
- [33] S. T. Kusuma and T. B. Sasongko, “Optimasi K-Nearest Neighbor dengan Grid Search CV pada Prediksi Kanker Paru-Paru,” *Indones. J. Comput. Sci.*, vol. 12, no. 4, Aug. 2023, doi: 10.33022/ijcs.v12i4.3267.

- [34] Z. M. Alhakeem, Y. M. Jebur, S. N. Henedy, H. Imran, L. F. A. Bernardo, and H. M. Hussein, "Prediction of Ecofriendly Concrete Compressive Strength Using Gradient Boosting Regression Tree Combined with GridSearchCV Hyperparameter-Optimization Techniques," *Materials*, vol. 15, no. 21, p. 7432, Oct. 2022, doi: 10.3390/ma15217432.
- [35] M. Sivakumar, S. Parthasarathy, and T. Padmapriya, "Trade-off between training and testing ratio in machine learning for medical image processing," *PeerJ Comput. Sci.*, vol. 10, p. e2245, Sep. 2024, doi: 10.7717/peerj-cs.2245.
- [36] N. Dalai, "Single User-Item Profile Matrix in Recommendations System," *Tuijin Jishu Journal Propuls. Technol.*, vol. 44, no. 3, pp. 220–224, Sep. 2023, doi: 10.52783/tjjpt.v44.i3.260.
- [37] N. A. Widiastuti, M. Azhar, and H. Mulyo, "Implementasi Algoritma K-Nearest Neighbor untuk Klasifikasi Jurusan pada Peserta Didik Baru," *Simetris J. Tek. Mesin Elektro Dan Ilmu Komput.*, vol. 14, no. 2, pp. 195–208, Nov. 2023, doi: 10.24176/simet.v14i2.10092.
- [38] P. Yuli Santoso, "ALGORITMA K-NEAREST NEIGHBOR DENGAN MENGGUNAKAN METODE EUCLIDEAN DISTANCE UNTUK MEMPREDIKSI KELULUSAN UJIAN NASIONAL BERBASIS DESKTOP PADA SMA NEGERI 12 TANGERANG," *SKANIKA Sist. Komput. Dan Tek. Inform.*, vol. 1, no. 1, pp. 123–129, Mar. 2018, no. no.1, p. pp.123-129, Mar. 2018.
- [39] Rubangi, "Sistem Rekomendasi Pada Tokopedia Menggunakan Algoritma K-Nearest Neighbor," *Httpjournalbsiacidejurnalindexphpjtk103This Work Licens. A Creative Commons Attrib.-ShareAlike 40 Int. Licens. Tek. Komputer AMIK BSI*, vol. Volume 8, no. No.1, Jan. 2022, doi: DOI:10.31294/jtk.v4i2.
- [40] I. G. Gusti, M. Nasrun Hasibuan, and R. Astuti Nugrahaeni, "REKOMENDASI SISTEM PEMILIHAN MOBIL MENGGUNAKAN K-NEAREST NEIGHBOR (KNN) COLLABORATIVE FILTERING," *TEKTRIKA - J. Penelit. Dan Pengemb. Telekomun. Kendali Komput. Elektr. Dan*

- Elektron.*, vol. 4, no. 1, p. 26, Sep. 2019, doi: 10.25124/tektrika.v4i1.1846.
- [41] A. Metanosa Afinda, “Overfitting vs. Underfitting, Apa Bedanya?” Accessed: May 31, 2024. [Online]. Available: <https://www.dicoding.com/blog/overfitting-vs-underfitting-apa-bedanya/>
 - [42] F. R. Hariri and L. W. Rochim, “Sistem Rekomendasi Produk Aplikasi Marketplace Berdasarkan Karakteristik Pembeli Menggunakan Metode User Based Collaborative Filtering,” *Teknika*, vol. 11, no. 3, pp. 208–217, Nov. 2022, doi: 10.34148/teknika.v11i3.538.
 - [43] A. Theodorus, “USER-BASED COLLABORATIVE FILTERING DENGAN MEMANFAATKAN PEARSON-CORRELATION UNTUK MENCARI NEIGHBORS TERDEKAT DALAM SISTEM REKOMENDASI,” 2016. [Online]. Available: <https://api.semanticscholar.org/CorpusID:130415603>
 - [44] A. Komaruddin, B. Susilo, and Y. Setiawan, “Perancangan Sistem Rekomendasi Pemilihan Cinderamata Khas Bengkulu Berbasis E-Marketplace,” *Rekursif J. Inform.*, vol. 7, no. 1, Jul. 2019, doi: 10.33369/rekursif.v7i1.6132.
 - [45] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, “Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa,” *IJCIT Indones. J. Comput. Inf. Technol.*, vol. 6, no. 2, Dec. 2021, doi: 10.31294/ijcit.v6i2.10438.
 - [46] D. R. Anamisa, A. Jauhari, and F. Ayu Mufarroha, “K-Nearest Neighbors Method for Recommendation System in Bangkalan’s Tourism,” *ComTech Comput. Math. Eng. Appl.*, vol. 14, no. 1, pp. 33–44, May 2023, doi: 10.21512/comtech.v14i1.7993.

LAMPIRAN

LAMPIRAN A. Contoh Perhitungan *Cosine Similarity*

$$\text{cosine sim}(u, v) = \frac{\sum_{i \in I_{uv}} r_{ui} \cdot r_{vi}}{\sqrt{\sum_{i \in I_{uv}} r_{ui}^2} \cdot \sqrt{\sum_{i \in I_{uv}} r_{vi}^2}}$$

Dibawah ini adalah Tabel Lampiran A yang merupakan contoh *Matriks Rating User-Item* yang akan digunakan dalam contoh perhitungan.

Tabel Lampiran 1 Contoh *Matriks Rating User-Item*

| | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | Item 6 | Item 7 | Item 8 | Item 9 | Item 10 |
|----------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|
| <i>User 1</i> | 5 | 4 | 3 | 0 | 4 | 5 | 0 | 3 | 5 | 4 |
| <i>User 2</i> | 4 | 3 | 2 | 5 | 4 | 5 | 3 | 3 | 0 | 5 |
| <i>User 3</i> | 5 | 4 | 3 | 4 | 5 | 4 | 4 | 0 | 2 | 0 |
| <i>User 4</i> | 0 | 0 | 5 | 5 | 4 | 0 | 3 | 0 | 3 | 4 |
| <i>User 5</i> | 0 | 0 | 3 | 0 | 5 | 0 | 2 | 3 | 0 | 5 |
| <i>User 6</i> | 3 | 0 | 4 | 5 | 3 | 5 | 0 | 5 | 0 | 3 |
| <i>User 7</i> | 3 | 3 | 0 | 2 | 0 | 2 | 0 | 0 | 0 | 3 |
| <i>User 8</i> | 0 | 0 | 0 | 5 | 0 | 4 | 0 | 0 | 0 | 3 |
| <i>User 9</i> | 0 | 4 | 0 | 2 | 2 | 4 | 0 | 3 | 0 | 3 |
| <i>User 10</i> | 3 | 3 | 0 | 5 | 4 | 0 | 0 | 3 | 0 | 5 |

Dalam contoh ini, akan dihitung nilai *Cosine Similarity* antara *User 1* dan *User 3* berdasarkan Tabel Lampiran 1. Langkah pertama dalam menghitung *Cosine Similarity* antara kedua Pengguna adalah menghitung *dot product* antara *User 1* dan *User 3* menggunakan rumus

$$\sum_{i \in I_{uv}} r_{ui} \cdot r_{vi}$$

$$\begin{aligned} \text{Dot Product} &= (5 \cdot 5) + (4 \cdot 4) + (3 \cdot 3) + (0 \cdot 4) + (4 \cdot 5) + (5 \cdot 4) \\ &\quad + (0 \cdot 4) + (3 \cdot 0) + (5 \cdot 2) + (4 \cdot 0) = 100 \end{aligned}$$

Selanjutnya menghitung Panjang dari masing-masing vektor rating

$$\sqrt{\sum_{i \in I_{uv}} r_{ui}^2} \cdot \sqrt{\sum_{i \in I_{uv}} r_{vi}^2}$$

Untuk *User 1*:

$$|r_1| = \sqrt{5^2 + 4^2 + 3^2 + 4^2 + 5^2 + 3^2 + 5^2} = \sqrt{116}$$

Untuk *User 3*:

$$|r_5| = \sqrt{5^2 + 4^2 + 3^2 + 5^2 + 4^2 + 3^2 + 2^2} = \sqrt{157}$$

$$\text{cosine sim}(1, 3) = \frac{\sum_{i \in I_{uv}} r_{1i} \cdot r_{5i}}{\sqrt{\sum_{i \in I_{uv}} r_{1i}^2} \cdot \sqrt{\sum_{i \in I_{uv}} r_{5i}^2}} = \frac{100}{\sqrt{116} \cdot \sqrt{157}} = 0.747$$

Berikut ini adalah tabel lengkap dari hasil perhitungan *Cosine Similarity* antar *User*:

Tabel Lampiran 2 Nilai *Cosine Similarity* antar *User*

| | <i>User 1</i> | <i>User 2</i> | <i>User 3</i> | <i>User 4</i> | <i>User 5</i> | <i>User 6</i> | <i>User 7</i> | <i>User 8</i> | <i>User 9</i> | <i>User 10</i> |
|----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|
| <i>User 1</i> | 1 | 0,9832 | 0,9525 | 0,9393 | 0,9947 | 0,9436 | 0,9719 | 0,9995 | 0,9768 | 0,9642 |
| <i>User 2</i> | 0,9833 | 1 | 0,9725 | 0,9435 | 0,9800 | 0,9412 | 0,9466 | 0,9798 | 0,9307 | 0,9955 |
| <i>User 3</i> | 0,9526 | 0,9726 | 1 | 0,9519 | 0,9635 | 0,9379 | 0,9870 | 0,9939 | 0,9253 | 0,9633 |
| <i>User 4</i> | 0,9394 | 0,9435 | 0,9520 | 1 | 0,9460 | 0,9920 | 0,9529 | 0,9910 | 0,9637 | 0,9945 |
| <i>User 5</i> | 0,9947 | 0,9800 | 0,9635 | 0,9460 | 1 | 0,8999 | 1 | 1 | 0,9437 | 0,9942 |
| <i>User 6</i> | 0,9436 | 0,9412 | 0,9379 | 0,9920 | 0,8999 | 1 | 0,9037 | 0,9942 | 0,9600 | 0,9450 |
| <i>User 7</i> | 0,9719 | 0,9466 | 0,9870 | 0,9529 | 1 | 0,9037 | 1 | 0,9261 | 0,9648 | 0,9366 |
| <i>User 8</i> | 0,9995 | 0,9798 | 0,9939 | 0,9910 | 1 | 0,9942 | 0,9261 | 1 | 0,9191 | 0,9701 |
| <i>User 9</i> | 0,9768 | 0,9307 | 0,9253 | 0,9637 | 0,9437 | 0,9600 | 0,9648 | 0,9191 | 1 | 0,9091 |
| <i>User 10</i> | 0,9642 | 0,9955 | 0,9633 | 0,9945 | 0,9942 | 0,9450 | 0,9366 | 0,9701 | 0,9091 | 1 |

LAMPIRAN B. Contoh Perhitungan Prediksi Rating *Collaborative Filtering* menggunakan *User-Based Collaborative Filtering*

Pada bagian ini, akan dilakukan perhitungan prediksi rating menggunakan *User-Based Collaborative Filtering*, yang mana hanya pengguna terdekat yang memiliki nilai *Cosine Similarity* tertinggi saja yang digunakan dalam perhitungan prediksi rating, pada contoh ini akan dipilih sebanyak $K=3$ yang mana hanya *User 2*, *User 3* dan *User 6* yang digunakan dalam perhitungan ini.

$$\hat{r}_{11} = \frac{\sum_{v \in N_i^k(u)} sim(1, v) \cdot r_{v1}}{\sum_{v \in N_i^k(u)} sim(1, v)}$$

$$= \frac{(0.9832 \cdot 4) + (0.9525 \cdot 5) + (0.9436 \cdot 3)}{0.9832 + 0.9525 + 0.9436}$$

$$= 4,00309$$

Maka didapatkan rating Item 1 untuk *User 1* adalah **4,00309**

LAMPIRAN C. Contoh Perhitungan *Mean Absolute Error*

$$MAE = \frac{1}{|\hat{R}|} \sum_{\hat{r}_{ui} \in \hat{R}} |r_{ui} - \hat{r}_{ui}|$$

Perhitungan *Mean Absolute Error* dilakukan untuk mengukur kinerja model menggunakan *data test*. Untuk *User 1* Item 1, berdasarkan perhitungan didapatkan nilai Prediksi Rating di tahap sebelumnya, didapatkan nilai Prediksi Rating sebesar **4,035154** dan data rating aktual sebesar **5**. Langkah selanjutnya adalah menghitung nilai *MAE* menjadi :

$$MAE = \frac{|5 - 4,00309|}{1} = 0,99691$$

Maka didapatkan Nilai *MAE* dari Prediksi Rating Pengguna 1 terhadap Item 1 adalah **0,964846**