



## *Implementation of Decision Tree Algorithm for Movie Recommendation and Rating Classification on the Netflix Platform*

### **Implementasi Algoritma Decision Tree untuk Rekomendasi Film dan Klasifikasi Rating pada Platform Netflix**

**Dimas Aditya Mukhsinin<sup>1</sup>, M Rafliansyah<sup>2</sup>, Sang Adji Ibrahim<sup>3</sup>,  
Rahmadden<sup>4\*</sup>, Denok Wulandari<sup>5</sup>**

<sup>1,4</sup>Program Studi Teknik Informatika, STMIK Amik Riau, Indonesia

<sup>5</sup>Program Studi Teknik Komputer, Institut Az-Zuhra, Indonesia

E-Mail: <sup>1</sup>[dimasadityamukhsinin@gmail.com](mailto:dimasadityamukhsinin@gmail.com), <sup>2</sup>[raxaraxa61@gmail.com](mailto:raxaraxa61@gmail.com), <sup>3</sup>[mr.adji00@gmail.com](mailto:mr.adji00@gmail.com),  
<sup>4</sup>[rahmadden@sar.ac.id](mailto:rahmadden@sar.ac.id), <sup>5</sup>[denokwulandari18@gmail.com](mailto:denokwulandari18@gmail.com)

*Received Dec 20th 2023; Revised Feb 10th 2024; Accepted Mar 5th 2024*

*Corresponding Author: Rahmadden*

#### **Abstract**

*As one of the world's largest video streaming platforms, Netflix has experienced rapid growth since its founding in 1997, initially focusing on DVD rentals before transitioning to online streaming services in 2007. With millions of global subscribers, Netflix continues to innovate with subscription packages, exclusive content production, and the utilization of data analysis technology and machine learning to enhance user experience. This research implements the Decision Tree algorithm to improve recommendation systems and rating classification on Netflix. Utilizing two primary datasets, *movies\_df* and *ratings\_df*, the study involves steps such as data collection and merging, feature and target variable determination, data partitioning, model training, and evaluation. Results include Decision Tree model evaluation using accuracy, precision, recall, and F1-score metrics for each rating category, along with bar chart visualization depicting the number of film ratings and the percentage of ratings from 1-5. Film recommendations based on the Decision Tree model are also presented, providing insights into enhancing recommendation systems on Netflix. The conclusion demonstrates that implementing the Decision Tree algorithm can enhance film recommendation accuracy and rating classification on Netflix, contributing to a more personalized user experience in the era of online streaming services.*

*Keywords: Decision Tree, Film Recommendations, Netflix, Rating Classification*

#### **Abstrak**

Sebagai salah satu platform video streaming terbesar di dunia, Netflix telah berkembang pesat sejak pendiriannya pada tahun 1997, awalnya berfokus pada penyewaan DVD, namun kemudian beralih ke layanan streaming online pada tahun 2007. Dengan jutaan pelanggan global, Netflix terus berinovasi dengan paket langganan, produksi konten eksklusif, dan teknologi analisis data serta machine learning untuk meningkatkan pengalaman pengguna. Penelitian ini menerapkan algoritma Decision Tree untuk meningkatkan sistem rekomendasi dan klasifikasi rating di Netflix. Menggunakan dua dataset utama, *movies\_df* dan *ratings\_df*, penelitian melibatkan langkah-langkah pengumpulan dan penggabungan data, penentuan fitur dan variabel target, pembagian data, pelatihan model, serta evaluasi. Hasilnya mencakup evaluasi model Decision Tree dengan metrik akurasi, precision, recall, dan F1-score untuk setiap kategori rating, serta visualisasi grafik batang tentang jumlah rating film dan presentase rating dari 1-5. Rekomendasi film berdasarkan model Decision Tree juga disajikan, memberikan wawasan tentang peningkatan sistem rekomendasi di Netflix. Kesimpulan menunjukkan bahwa implementasi algoritma Decision Tree dapat meningkatkan akurasi rekomendasi film dan klasifikasi rating di Netflix, berkontribusi pada pengalaman pengguna yang lebih personal di era layanan streaming online.

Kata Kunci: Decision Tree, Klasifikasi Rating, Netflix, Rekomendasi Film

#### **1. PENDAHULUAN**

Netflix merupakan salah satu platform video *streaming* terbesar di dunia, yang didirikan di Amerika Serikat pada tahun 1997. Pada awalnya, platform ini hanya berfokus pada penyewaan DVD melalui media pos. Namun, sejak tahun 2007, Netflix sudah mudah berkembang dan memperluas bisnisnya ke layanan *streaming*

video *online*. Netflix adalah layanan *online* berbasis langganan yang memungkinkan para member yang ada menonton acara TV dan film tanpa iklan di perangkat yang terhubung ke Internet atau dengan kata lain adalah layanan yang memungkinkan pengguna atau penonton, menonton tayangan kesukaan mereka di mana pun, kapan pun, bisa menggunakan *smartphone*, *smartTV*, *tablet*, *PC*, dan *laptop* [1][2].

Netflix menawarkan berbagai banyak fitur terbaru salah satunya yaitu jenis paket langganan. Fitur ini menawarkan paket dasar, standar, dan premium dengan harga yang berbeda-beda tergantung pada kualitas video dan jumlah perangkat yang dapat digunakan untuk *streaming*. Selain itu, Netflix juga mampu mengembangkan berbagai jenis konten eksklusif, seperti film dan serial yang diproduksi oleh pihak Netflix sendiri. Netflix juga telah mengubah cara bercerita dengan menerapkan strategi produksi yang berbeda dalam serial. Dalam beberapa tahun terakhir, Netflix dan platform video *streaming* lainnya semakin mengadopsi teknologi analisis data dan *machine learning* untuk meningkatkan pengalaman pengguna. Netflix menggunakan data analisis untuk memahami perilaku pengguna dan preferensi pengguna, sehingga pengguna dapat memberikan rekomendasi konten yang lebih relevan dan personal. Salah satu teknik analisis data yang digunakan oleh Netflix adalah klasifikasi atau pengelompokan. Klasifikasi dapat digunakan untuk mengelompokkan konten berdasarkan preferensi tontonan pengguna, sehingga Netflix dapat memberikan rekomendasi konten yang lebih spesifik dan relevan. Selain itu, klasifikasi juga dapat digunakan untuk mengelompokkan konten berdasarkan karakteristik, sehingga pengguna dapat lebih mudah menemukan konten yang sesuai dengan preferensi pengguna [3][4][5].

Netflix telah menggunakan berbagai teknik untuk memberikan rekomendasi kepada penggunanya, termasuk pemanfaatan algoritma *Decision Tree*. Algoritma ini, yang telah terbukti efektif dalam berbagai aplikasi klasifikasi, memiliki potensi untuk meningkatkan sistem rekomendasi Netflix. Penerapan algoritma *Decision Tree* dalam sistem rekomendasi film Netflix menjadi pilihan yang menarik karena keputusan yang dihasilkan oleh algoritma ini dapat dijelaskan dengan mudah melalui struktur pohon yang intuitif [6] [7][8].

Salah satu aspek penting dari sistem rekomendasi adalah klasifikasi, yang membantu pengguna membuat keputusan tentang film apa yang ingin ditonton. Rating film juga memberikan gambaran tentang kualitasnya. Namun, tantangan terbesar dalam implementasi sistem rekomendasi film adalah jumlah data film yang besar dan beragam di dataset Netflix [9][10][11].

Penggunaan algoritma *Decision Tree* dalam klasifikasi rating film untuk sistem rekomendasi Netflix menawarkan beberapa keunggulan. Keuntungan-keuntungan tersebut meliputi interpretabilitas algoritma untuk membuat visualisasi yang mudah dipahami serta skalabilitasnya dalam mengelola dataset yang besar. Dengan demikian, penggabungan antara klasifikasi rating dan penggunaan algoritma *Decision Tree* dapat menjadi langkah yang bermanfaat dalam memperbaiki sistem rekomendasi Netflix [12][13][14].

Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan algoritma *Decision Tree* untuk rekomendasi film dan klasifikasi rating pada platform Netflix. Rekomendasi film akan diambil berdasarkan nama film, sementara klasifikasi rating akan dilakukan dari id film dan tahun. Penelitian ini diharapkan dapat memberikan wawasan baru tentang bagaimana meningkatkan sistem rekomendasi dan klasifikasi rating di Netflix menggunakan algoritma *Decision Tree*.

## 2. BAHAN DAN METODE

### 1.1 Netflix

Netflix merupakan suatu layanan *streaming* video yang memberikan kemampuan bagi pengguna untuk menyaksikan film, serial televisi, dan berbagai program lainnya secara *online* melalui internet. Selain itu, Netflix juga memanfaatkan teknologi *streaming* agar pengguna dapat menonton video secara langsung tanpa harus menunggu sampai proses pengunduhan selesai. Secara finansial, Netflix merupakan salah satu perusahaan teknologi terbesar di dunia dan memiliki jutaan pelanggan di seluruh dunia. Netflix telah berkembang menjadi salah satu *platform* yang *streaming* terbesar di dunia dengan jangkauan global dan keanggotaan yang meliputi jutaan orang diberbagai negara. Perusahaan ini juga terus berinvestasi dalam produksi secara eksklusif untuk netflix. Oleh karena itu, Netflix akan terus mengoptimalkan dan meningkatkan kualitas produk dan layanan mereka yang dapat diakses dengan mudah dan fleksibel ke berbagai jenis konten melalui platform yang inovatif dan populer di seluruh dunia [2][4][9].

### 1.2 Klasifikasi

Klasifikasi merupakan salah satu topik utama dalam *data mining* atau *machine learning*. Klasifikasi adalah suatu pengelompokan data dimana data yang digunakan tersebut mempunyai kelas label atau target. Sehingga algoritma-algoritma untuk menyelesaikan masalah klasifikasi dikategorisasikan ke dalam *supervised learning* atau pembelajaran yang diawasi. Maksud dari pembelajaran yang diawasi adalah data label atau target ikut berperan sebagai ‘supervisor’ atau ‘guru’ yang mengawasi proses pembelajaran dalam mencapai tingkat akurasi atau presisi tertentu [15][16][17].

### 1.3 Decision Tree

Decision tree adalah metode pemodelan prediktif dalam analisis data yang menggunakan struktur pohon. Tujuan decision tree adalah untuk menggambarkan serta membuat keputusan berdasarkan serangkaian aturan dan kondisi. Dikutip dari pengertian lainnya, decision tree juga disebut sebagai alat yang kuat dalam machine learning, dan dapat digunakan untuk memprediksi hasil berdasarkan serangkaian fitur atau variabel input [18]. Persamaan umum dari Decision tree adalah sebagai berikut:

$$f(x) = \sum_{i=1}^N c_i \cdot I(x \in R_i) \quad (1)$$

di mana:

$f(x)$  = prediksi yang dihasilkan oleh decision tree untuk input  $x$ .

$N$  = jumlah simpul daun.

$c_i$  = nilai kelas yang diberikan pada simpul daun ke- $i$ .

$R_i$  = wilayah yang diwakili oleh simpul daun ke- $i$ .

Struktur data decision tree yang terdiri dari *node* dan rusuk (tepi). Ada tiga jenis *node* pada decision tree yaitu simpul akar (*root/node*), simpul percabangan/internal (*branch/simpul internal*), dan simpul daun (*leaf/node*).

Decision tree merupakan penyederhanaan teknik klasifikasi untuk sejumlah kelas yang terbatas, dengan internal simpul dan simpul akar diberi label sebagai nama atribut, rusuk diberi label sebagai nilai atribut yang mungkin, dan simpul daun diberi label sebagai kelas yang berbeda [19][20].

Decision tree merupakan salah satu teknik pembelajaran mesin (*machine learning*) yang menggunakan hierarki aturan klasifikasi struktur berurutan dengan mempartisi dataset pelatihan secara rekursif [21]. Decision tree adalah astruktur diagram alur berbentuk seperti pohon, dimana setiap simpul bagian dalam menandakan pengujian suatu atribut, dengan hasilnya cabang menunjukkan hasil pengujian, dan simpul daun mewakili distribusi kelas [22][23][24].

### 1.4 Metode Penelitian

Penelitian ini menggunakan dataset dari Netflix yang berisi informasi tentang film dan rating yang diberikan oleh pengguna. Algoritma Decision Tree diimplementasikan dengan menggunakan bahasa pemrograman Python dan *library* Scikit-learn. Proses ini melibatkan beberapa tahap, termasuk pra-pemrosesan data, membersihkan data dan mengubahnya menjadi format yang dapat digunakan oleh model dan pelatihan model, lalu menggunakan sebagian data untuk mengajarkan model bagaimana membuat prediksi dan evaluasi model, setelah itu dilakukan pengujian kinerja model pada data yang tidak digunakan selama pelatihan.

Pada penelitian ini dilakukan beberapa tahapan sebelum melakukan proses pengaplikasian program Algoritma Decision Tree untuk sistem rekomendasi film dengan tahapan-tahapannya antara lain yaitu perumusan masalah, studi literatur, tujuan, pengumpulan data dalam hal ini adalah data primer yaitu dataset film dan data sekunder yaitu data *user*, lalu proses pengolahan data yang mencakup penggabungan, pelatihan dan visualisasi, selain itu dari hasil di lakukan Analisa dan kesimpulan. Secara detail dapat ditunjukkan pada gambar 1.

#### 2.1. Pengumpulan Data

Pertama-tama, penelitian dimulai dengan pengumpulan data. Dua dataset utama yang digunakan adalah *movies\_df* dan *ratings\_df*. Dataset ini mencakup informasi penting tentang film-film, termasuk judul, tahun rilis, dan ID film. Selain itu, dataset *ratings\_df* memberikan wawasan tentang bagaimana pengguna memberi rating kepada film-film tersebut. Pengumpulan data yang cermat dan representatif merupakan fondasi yang kuat untuk analisis lebih lanjut.

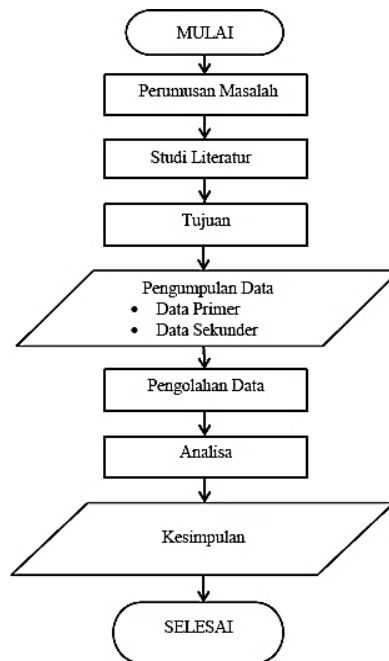
#### 2.2. Penggabungan Data

Setelah data terkumpul, langkah selanjutnya adalah menggabungkan dataset. Dalam kasus ini, penggabungan dilakukan dengan menggunakan fungsi *pd.merge* untuk menggabungkan dataset film dan dataset rating. Melalui langkah ini, akan menciptakan hubungan yang erat antara informasi film dan penilaian yang diberikan oleh pengguna. Penggabungan data ini menjadi landasan penting untuk analisis lebih lanjut, menggabungkan konteks film dan ulasan pengguna.

#### 2.3. Fitur dan Variabel Target

Penentuan fitur dan variabel target adalah tahap yang penting dalam pengembangan model klasifikasi rating. Dalam penelitian ini, dilakukan pemilihan 'Movie\_ID' dan 'Year' sebagai fitur yang digunakan untuk melatih model, sementara 'Rating' dijadikan variabel target. Pemilihan fitur ini didasarkan pada asumsi bahwa

tahun dan ID film dapat memberikan informasi kritikal terkait faktor-faktor yang memengaruhi rating pengguna.



**Gambar 1.** Flowchart metodologi

#### 2.4. Pembagian Data

Pembagian data menjadi set pelatihan dan pengujian merupakan langkah esensial dalam mengembangkan model klasifikasi yang efektif. Dalam implementasi program ini, penggunaan fungsi `train_test_split` dilakukan untuk memisahkan data menjadi dua bagian, dengan 80% digunakan sebagai set pelatihan dan 20% sebagai set pengujian. Proporsi ini dipilih untuk memastikan model dapat belajar dari sebagian besar data tetapi juga diuji pada data yang belum pernah dilihat.

#### 2.5. Pelatihan Model

Langkah berikutnya adalah pelatihan model, yang dalam kasus ini menggunakan Decision Tree Classifier. Decision Tree dipilih karena kemampuannya untuk menangani klasifikasi dengan baik dan kemampuan mudah dipahami. Proses pelatihan melibatkan fitting model pada set pelatihan, di mana model belajar untuk mengidentifikasi pola dan hubungan antara fitur dan variabel target.

#### 2.6. Evaluasi Model

Setelah pelatihan, model dievaluasi pada set pengujian. Metrik evaluasi seperti akurasi, *confusion matrix*, dan *classification report* digunakan untuk mengukur sejauh mana model dapat mengklasifikasikan rating dengan akurat. Hasil evaluasi ini menjadi landasan untuk menganalisis keefektifan model dan mengidentifikasi area yang mungkin perlu ditingkatkan.

Dengan proses setiap tahap secara rinci, penelitian ini memastikan bahwa setiap aspek metodologi diperlakukan secara komprehensif. Dengan kombinasi langkah-langkah ini, analisis klasifikasi rating dapat memberikan rekomendasi film berdasarkan rating film di platform Netflix.

### 3. Hasil dan Pembahasan

Dataset yang digunakan adalah `movies_df` dan `ratings_df`, yang akan di proses untuk digunakan dalam melakukan klasifikasi rating dan rekomendasi film netflix dimana pada dataset `movies_df` mencakup 3 kolom data sebagai berikut.

1. `Movie_ID` : pengidentifikasi unik untuk setiap film.
2. `Year` : tahun ketika film dirilis.
3. `Name` : judul film.

Dan `ratings_df` yang juga mencakup 3 kolom data sebagai berikut.

1. `User_ID` : pengidentifikasi unik untuk pengguna yang memberi peringkat.
2. `Rating` : rating yang diberikan pengguna untuk film tertentu - mungkin merupakan variabel target

yang baik untuk klasifikasi.

3. `Movie_ID` : pengidentifikasi unik untuk setiap film yang diberi rating.

Sebelum dilakukan evaluasi Decision Tree, dataset akan dilakukan pemrosesan data lebih lanjut dengan memetakan `User_ID` dan `Movie_ID` ke nilai unik yang lebih kecil.

```
mapping1 = {
    old:new for old, new in list(zip(sorted(merged_df.User_ID.unique()), range(merged_df.User_ID.unique().shape[0])))
}
mapping2 = {
    old:new for old, new in list(zip(sorted(merged_df.Movie_ID.unique()), range(merged_df.Movie_ID.unique().shape[0])))
}
```

Hasil evaluasi model Decision Tree pada dataset "movies.csv" dan "ratings.csv" dapat dilihat pada Tabel 1. Tabel yang disajikan adalah *Classification Report*, yang memberikan metrik evaluasi kinerja model klasifikasi untuk setiap kelas target. Setiap kelas memiliki *precision*, *recall*, dan *f1-score*. Ini adalah:

1. *Precision*: Persentase prediksi positif yang benar dari semua prediksi positif yang dibuat oleh model. Dalam konteks ini, semakin tinggi nilai *precision*, semakin sedikit hasil *false positive* yang dihasilkan oleh model untuk kelas tertentu.
2. *Recall*: Persentase instance positif yang telah diidentifikasi dengan benar dari semua instance positif yang ada dalam dataset. Dalam konteks ini, semakin tinggi nilai *recall*, semakin sedikit hasil *false negative* yang dihasilkan oleh model untuk kelas tertentu.
3. *F1-score*: Rata-rata harmonik dari *precision* dan *recall*. *F1-score* memberikan keseimbangan antara *precision* dan *recall*. Nilai *F1-score* yang tinggi menunjukkan kinerja yang baik dari model.
4. *Support*: Jumlah instance dalam setiap kelas.

Selain itu, terdapat juga metrik evaluasi untuk keseluruhan model:

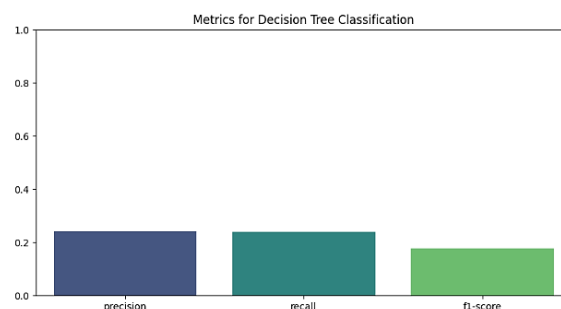
1. *Accuracy*: Persentase prediksi yang benar secara keseluruhan dari model.
2. *Macro avg*: Rata-rata dari metrik evaluasi untuk setiap kelas, tanpa memperhatikan jumlah instance dalam setiap kelas.
3. *Weighted avg*: Rata-rata dari metrik evaluasi untuk setiap kelas, dengan memperhatikan jumlah instance dalam setiap kelas.

Dari tabel 1, dapat dilihat bahwa model memiliki kinerja yang cukup baik dalam mengidentifikasi kelas 3 dan 4 (yang memiliki nilai *precision* dan *recall* yang relatif tinggi), sementara kinerja untuk kelas 1, 2, dan 5 terbilang rendah.

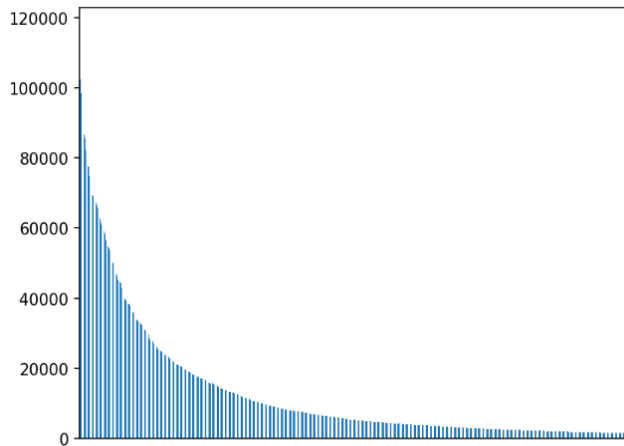
**Tabel 1.** Hasil Evaluasi Model Decision Tree

	precision	recall	f1-score	support
1	0.346121	0.013847	0.026628	1.446540
2	0.322599	0.012711	0.024458	3.515070
3	0.383132	0.465972	0.420511	1.036425
4	0.386423	0.572739	0.461485	1.180680
5	0.463987	0.269075	0.340618	7.542260
accuracy	0.394689	0.394689	0.394689	3.420789
macro avg	0.380453	0.266869	0.254740	3.467492
weighted avg	0.394160	0.394689	0.360505	3.467492

Grafik batang untuk metrik-metrik ini dapat dilihat pada Gambar 2. Gambar ini memberikan visualisasi yang lebih jelas tentang performa model untuk setiap kategori rating.

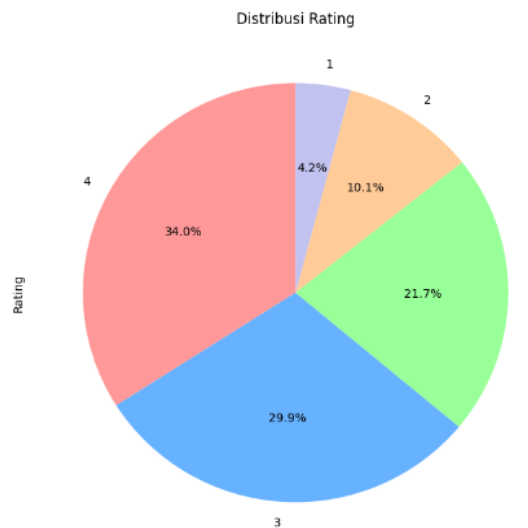


**Gambar 2.** Grafik Metrik Evaluasi Model Decision Tree



**Gambar 3.** Grafik Jumlah Rating Pada Film

Pada keseluruhan dataset ratings, dapat kita lihat bahwa retribusi nilai rating paling umum dilakukan oleh *user* atau terbanyak adalah rating nilai 4 dengan rata-rata 34.0% dari rata-rata *user* yang dapat dilihat pada gambar 4 berikut.



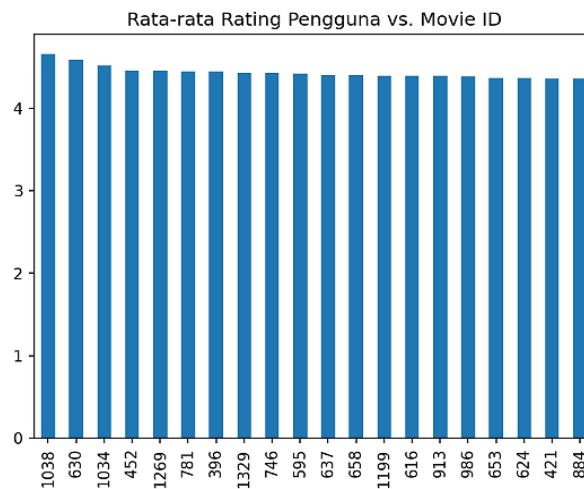
**Gambar 4.** Retribusi Rating

Grafik 4 menunjukkan distribusi rating pada dataset yang telah digabungkan (*merged\_df*). Di dalam grafik, setiap bagian dari *pie chart* mewakili persentase dari masing-masing rating. Dengan menggunakan kode `merged_df['Rating'].value_counts().plot.pie()`, kita dapat melihat seberapa sering masing-masing rating muncul dalam dataset. Sedangkan `autopct='%1.1f%%'` menampilkan persentase dari setiap bagian dengan satu angka di belakang koma. Warna yang digunakan untuk setiap bagian *pie chart* juga didefinisikan, dimana warna yang berbeda membantu membedakan antara kategori rating. Titik awal *angle* untuk menggambar *pie* dimulai dari 90 derajat, yang berarti dimulai dari bagian atas. Dengan demikian, grafik ini memberikan gambaran visual tentang seberapa umum masing-masing rating pada platform Netflix. Selain itu, kode `common_rating = merged_df['Rating'].mode().values[0]` juga memberikan informasi tentang rating yang paling umum dalam bentuk teks yang dicetak, memungkinkan untuk melihat rating yang paling dominan secara langsung di samping grafik. Retribusi rating dapat ditunjukkan pada gambar 5.

Gabungan dari dataset menghasilkan perbandingan berupa 10 film dengan jumlah rata-rata rating pengguna terhadap data film yang dapat dilihat pada gambar 6.



Gambar 5. Retribusi Rating



Gambar 6. Retribusi Rating

Setelah dilakukan penggabungan dataset movies dan ratings, dilakukan penghitungan untuk melihat jumlah rating terbanyak pada top 10 film. Hasilnya, beberapa rekomendasi teratas dapat dilihat pada tabel 2.

Tabel 2. Top 10 Rating Film Terbanyak

Nama Film	Jumlah Rating
Pirates of the Caribbean: The Curse of the Black Pearl	117075
Lord of the Rings: The Fellowship of the Ring	102721
The Sixth Sense	102376
American Beauty	101450
Bruce Almighty	98545
The Silence of the Lambs	95053
Finding Nemo (Widescreen)	94235
The Italian Job	93886
Shrek 2	92893
Braveheart	91502

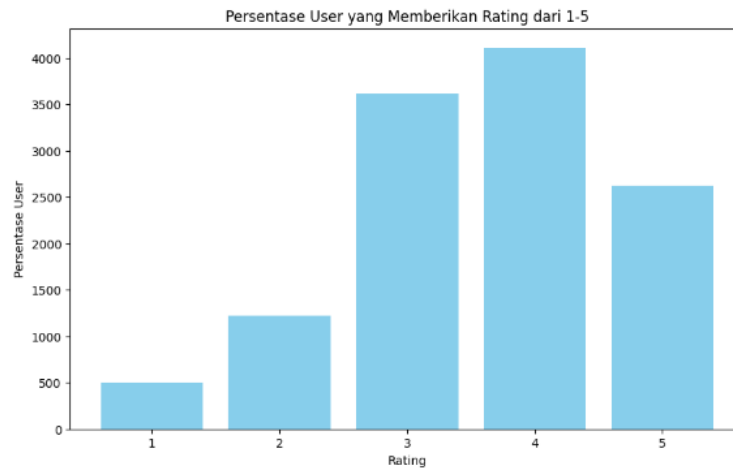
dilakukan kembali perhitungan untuk mencari top 10 film terpopuler yang didapatkan dengan mencari hasil rata-rata rating yang hasilnya dapat dilihat pada tabel 3.

Tabel 3. Top 10 Film Terpopuler

Nama Film	Rata-rata Rating
Lost: Season 1	4.665432
The Simpsons: Season 6	4.589824
Family Guy: Freakin' Sweet Collection	4.520766
Six Feet Under: Season 4	4.461601
Inu-Yasha	4.457774
Stargate SG-1: Season 8	4.456026
The Best of Friends: Vol. 4	4.449168

Nama Film	Rata-rata Rating
The West Wing: Season 3	4.436258
Lord of the Rings: The Fellowship of the Ring	4.431489
Gilmore Girls: Season 3	4.428943

Presentase rata-rata *user* yang memberikan rating antara 1-5 dapat dilihat pada gambar 7 dibawah dimana nilai rating dengan presentase *user* rata-rata berjumlah 4000 memberikan rating 4 lalu di ikuti dengan rating 3, 5, 2 hingga 1.



**Gambar 7.** Presentase user memberikan rating dari 1-5

Hasil rata-rata user yang memberikan rating, klasifikasi dibuat untuk mengidentifikasi siapa user yang paling banyak memberikan rating terhadap film yang ada dimana dapat dilihat pada grafik batang tabel 4.

**Tabel 4.** Top 10 User yang Memberikan Rating Terbanyak

User ID	Jumlah Rating
16592	1344
21168	1339
132092	1324
114829	1305
90366	1257
39086	1169
71577	1132
111457	1072
100658	1053
141109	1039

Hasil pelatihan Decision Tree di jalankan dalam bentuk *function* dimana *function* ini akan bekerja dengan memproses kata kunci yang dituliskan user yang nantinya akan menampilkan rekomendasi film dengan kata kunci serupa, dalam contoh pada tabel 5 bisa dilihat hasil dari rekomendasi dengan keyword *Toy Story*.

**Tabel 5.** Top 10 Rekomendasi Film Hasil Evaluasi Model Decision Tree

	Name	Year
13672	Toy Story	1995
3839	The Toy	1982
12676	Toy Soldiers	1991
4894	Toy Soldiers	1984
1954	The Story of Us	1999
1251	The Story of O	1975
9131	L.A. Story	1991
15560	VeggieTales: The Toy That Saved Christmas	1996
8693	Love Story	1981
9036	Love Story	1970



#### 4. KESIMPULAN

Penelitian ini berhasil mencapai tujuannya untuk meningkatkan sistem rekomendasi dan klasifikasi rating di platform Netflix dengan mengimplementasikan algoritma Decision Tree. Melalui langkah-langkah metodologi yang mencakup pengumpulan data, penggabungan data, pemilihan fitur dan variabel target, pembagian data, pelatihan model, dan evaluasi model, penelitian ini berhasil membangun sistem rekomendasi yang mampu memberikan rekomendasi film yang relevan berdasarkan kata kunci yang dimasukkan oleh pengguna. Evaluasi model menunjukkan performa yang baik dalam klasifikasi rating, dengan metrik seperti akurasi, *precision*, *recall*, dan *F1-score* memberikan gambaran tentang sejauh mana model dapat mengklasifikasikan rating dengan akurat. Selain itu, analisis data juga memberikan wawasan tentang perilaku pengguna dalam memberikan rating, yang dapat memperkaya konten yang ditawarkan oleh platform Netflix. Dengan demikian, implementasi ini berhasil memberikan kontribusi dalam mengoptimalkan pengalaman pengguna dan meningkatkan kualitas layanan yang ditawarkan oleh platform streaming video ini.

#### REFERENSI

- [1] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, "Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner," *JBASE-Journal of Business and Audit Information Systems*, vol. 4, no. 1, 2021.
- [2] A. D. Lotz, O. Eklund, and S. Soroka, "Netflix, library analysis, and globalization: rethinking mass media flows," *Journal of Communication*, vol. 72, no. 4, pp. 511–521, 2022.
- [3] S. Chris and D. Susilawati, "Analisis Perancangan Sistem untuk Kepuasan Pelanggan pada UD. Shinta Elektronik dengan Menggunakan Metode Algoritma C4. 5," *ALGOR*, vol. 1, no. 2, pp. 52–58, 2020.
- [4] M. L. Wayne, "Netflix audience data, streaming industry discourse, and the emerging realities of 'popular' television," *Media Cult Soc*, vol. 44, no. 2, pp. 193–209, 2022.
- [5] Natalie Sherman, "Netflix: Four things which have driven its success." Accessed: Jan. 14, 2024. [Online]. Available: <https://www.bbc.com/news/business-55723926>
- [6] J. Fadlil and W. F. Mahmudy, "Pembuatan Sistem Rekomendasi Menggunakan Decision Tree dan Clustering," *Jurnal Ilmiah Kursor*, vol. 3, no. 1, 2007.
- [7] H. Steck, L. Baltrunas, E. Elahi, D. Liang, Y. Raimond, and J. Basilico, "Deep learning for recommender systems: A Netflix case study," *AI Mag*, vol. 42, no. 3, pp. 7–18, 2021.
- [8] C. A. Gomez-Urbe and N. Hunt, "The netflix recommender system: Algorithms, business value, and innovation," *ACM Transactions on Management Information Systems (TMIS)*, vol. 6, no. 4, pp. 1–19, 2015.
- [9] Kait Fore, "The Impact of Streaming Services on the Movie Industry." Accessed: Jan. 16, 2024. [Online]. Available: <https://motionpicture.edu/socialcinema/2023/06/19/the-impact-of-streaming-services-on-the-movie-industry/>
- [10] S. Jayalakshmi, N. Ganesh, R. Ćep, and J. Senthil Murugan, "Movie recommender systems: Concepts, methods, challenges, and future directions," *Sensors*, vol. 22, no. 13, p. 4904, 2022.
- [11] Y. Ali *et al.*, "A hybrid group-based movie recommendation framework with overlapping memberships," *PLoS One*, vol. 17, no. 3, p. e0266103, 2022.
- [12] M. R. A. Zayyad and A. Kurniawardhani, "Penerapan Metode Deep Learning pada Sistem Rekomendasi Film," *AUTOMATA*, vol. 2, no. 1, 2021.
- [13] Sarah Juliandiny, "Aplikasi Data Science: Sistem Rekomendasi Netflix." Accessed: Jan. 16, 2024. [Online]. Available: <https://pacmann.io/blog/aplikasi-data-science-pada-recommendation-system-netflix>
- [14] R. Rusito and M. Firmansyah, "Implementasi Metode Decision Tree Dan Algoritma C4. 5 Untuk Klasifikasi Data Nasabah Bank," *Jurnal Ilmiah Infokam*, vol. 12, no. 2, 2016.
- [15] Antoni Wibowo, "Klasifikasi."
- [16] R.-C. Chen, C. Dewi, S.-W. Huang, and R. E. Caraka, "Selecting critical features for data classification based on machine learning methods," *J Big Data*, vol. 7, no. 1, p. 52, 2020.
- [17] T. Sasao, *Classification Functions for Machine Learning and Data Mining*. Springer Nature, 2023.
- [18] "Decision Tree: Pengertian, Manfaat, Cara Buat & Plus Minusnya." Accessed: Jan. 14, 2024. [Online]. Available: <https://www.ocbc.id/id/article/2023/06/20/decision-tree-adalah>
- [19] I. Sutoyo, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik," *Jurnal Pilar Nusa Mandiri*, vol. 14, no. 2, pp. 217–224, 2018.
- [20] G. D. M. Zulma and N. Chamidah, "Perbandingan Metode Klasifikasi Naive Bayes, Decision Tree Dan K-Nearest Neighbor Pada Data Log Firewall," in *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya*, 2021, pp. 679–688.
- [21] J. Eska, "Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4. 5," 2018.

- [22] K. F. L. Simanjuntak, A. C. B. Barus, and A. Anita, "Implementasi Metode Decision Tree Dan Algoritma C4. 5 Untuk Klasifikasi Kepribadian Masyarakat," *JOISIE (Journal Of Information Systems And Informatics Engineering)*, vol. 5, no. 1, pp. 51–59, 2021.
- [23] S. B. Kotsiantis, "Decision trees: a recent overview," *Artif Intell Rev*, vol. 39, pp. 261–283, 2013.
- [24] V. Podgorelec, P. Kokol, B. Stiglic, and I. Rozman, "Decision trees: an overview and their use in medicine," *J Med Syst*, vol. 26, pp. 445–463, 2002.