

Implementasi Metode K-Means Clustering Untuk Rekomendasi Film

¹Adrian Jala Putra Mahardika, ²Danang Wahyu Widodo, ³ Ardi Sanjaya

^{1,2,3}Teknik Informatika, Universitas Nusantara PGRI Kediri

E-mail: ¹adriansekolah99@gmail.com, ²danangwahyuwidodo@unpkediri.ac.id,
³dersky@gmail.com

Penulis Korespondens : Adrian Jala Putra Mahardika

Abstrak— Sistem rekomendasi film sering kali menghadapi kesulitan dalam menggabungkan berbagai elemen seperti genre, rating, tahun rilis, dan sutradara untuk memberikan rekomendasi yang tepat. Penelitian ini merancang sebuah sistem rekomendasi berbasis web yang memanfaatkan algoritma K-Means Clustering untuk mengelompokkan film berdasarkan kesamaan atribut tersebut. Proses ini meliputi pengumpulan dan pra-pemrosesan data dari Kaggle, penerapan K-Means untuk pengelompokan, serta evaluasi model dengan menggunakan metrik Silhouette Score untuk memastikan kualitas dari setiap cluster yang terbentuk. Sistem ini mampu menyajikan rekomendasi film yang lebih sesuai dengan preferensi pengguna, serta memberikan informasi yang dapat digunakan oleh pengelola bioskop dalam menentukan film yang layak untuk dipromosikan dan ditayangkan secara lebih strategis.

Kata Kunci— Film, K-Means Clustering, Pengelompokan data, Sistem rekomendasi, Web

Abstract— *Film recommendation systems often face challenges in combining various elements such as genre, rating, release year, and director to provide accurate recommendations. This research designs a web-based recommendation system that utilizes the K-Means Clustering algorithm to group films based on the similarity of these attributes. The process includes data collection and preprocessing from Kaggle, applying K-Means for clustering, and evaluating the model using the Silhouette Score metric to ensure the quality of each formed cluster. This system is capable of presenting film recommendations that are more aligned with user preferences, as well as providing information that can be used by cinema managers to determine which films are worth promoting and screening more strategically.*

Keywords— *Film, K-Means Clustering, Data clustering, Recommendation system, Web*

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

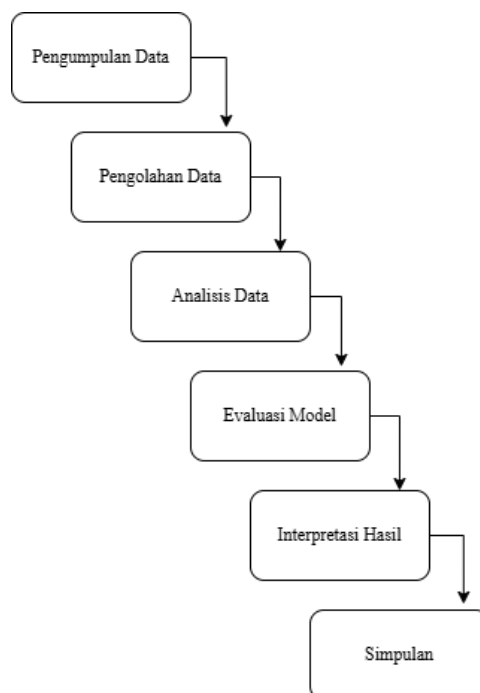
Kemajuan dalam teknologi digital telah membawa perubahan signifikan bagi industri hiburan, terutama dalam cara penonton menemukan dan memilih film. Platform digital dan bioskop kini banyak mengandalkan sistem rekomendasi film untuk menyajikan konten yang sesuai dengan preferensi pengguna. Namun, banyak dari sistem ini masih mengalami kesulitan dalam menggabungkan berbagai atribut penting, seperti genre, rating, tahun rilis, dan sutradara, dengan cara yang efektif. Sistem yang hanya menggunakan satu metode, seperti content-based filtering atau collaborative filtering, sering kali menghasilkan rekomendasi yang tidak sesuai dengan preferensi pengguna [1] [2] [3].

Masalah utama yang ingin diatasi dalam penelitian ini adalah ketidakefektifan pengelompokan film berdasarkan kesamaan atribut dalam sistem rekomendasi yang ada. Akibatnya, rekomendasi yang diberikan sering kali tidak sesuai dengan preferensi sebenarnya dari pengguna, yang pada gilirannya mempengaruhi tingkat kepuasan dan keterlibatan mereka. Untuk mengatasi isu ini, diperlukan suatu pendekatan yang dapat secara objektif mengelompokkan film berdasarkan kemiripan atribut, sehingga sistem mampu memberikan rekomendasi yang lebih sesuai dengan keinginan pengguna.

Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi film berbasis web menggunakan algoritma *K-Means Clustering* dengan memanfaatkan atribut-atribut seperti genre, rating, tahun rilis, dan sutradara. Tujuan umum dari penelitian ini adalah untuk meningkatkan relevansi rekomendasi film sesuai preferensi pengguna. Tujuan spesifiknya adalah mengembangkan sistem yang mampu melakukan pra-pemrosesan data film, mengelompokkannya ke dalam cluster yang relevan, serta menyajikan hasil rekomendasi melalui antarmuka web yang interaktif. Evaluasi model dilakukan menggunakan metrik *Silhouette Score* untuk menilai efektivitas pengelompokan.

II. METODE

2.1 Alur Penelitian



Gambar 1. Tahapan Penelitian

Tahapan penelitian sesuai Gambar 1 terdiri dari:

1. Pengumpulan Data: Pada tahap ini, dataset film diunduh dari IMDB Movies Dataset yang dapat diakses di platform Kaggle. Dataset ini dipilih karena menyediakan informasi yang

komprehensif, termasuk judul film, genre, rating, tahun rilis, dan sutradara, yang sangat relevan untuk analisis dalam pengembangan sistem rekomendasi berbasis clustering.

2. **Pengolahan Data:** Setelah data dikumpulkan, langkah selanjutnya adalah memprosesnya melalui tahap pra-pemrosesan. Proses ini meliputi penghapusan nilai yang hilang, normalisasi untuk atribut numerik seperti rating dan tahun rilis, serta penerapan one-hot encoding pada atribut kategorikal seperti genre dan sutradara. Semua ini bertujuan untuk mempersiapkan data agar dapat digunakan dalam proses pengelompokan dengan algoritma K-Means.
3. **Analisis Data:** Proses analisis ini menggunakan algoritma K-Means Clustering untuk mengelompokkan film berdasarkan kesamaan atribut yang dimiliki. Nilai k ditentukan melalui pendekatan eksploratif, dan setiap film ditempatkan dalam cluster yang memiliki centroid terdekat, sehingga menghasilkan kelompok film dengan karakteristik yang mirip.
4. **Evaluasi Model :** Setelah proses pengelompokan selesai, hasilnya dianalisis dengan menggunakan metrik Silhouette Score untuk menilai kualitas dan konsistensi masing-masing cluster. Tujuan dari evaluasi ini adalah untuk memastikan bahwa film-film dalam satu cluster memiliki kesamaan yang nyata, serta berbeda secara signifikan dari film-film yang berada di cluster lainnya.
5. **Interpretasi Hasil :** Hasil dari proses pengelompokan dianalisis untuk mengevaluasi seberapa relevan pengelompokan tersebut dengan preferensi pengguna. Kluster yang dihasilkan menjadi landasan bagi sistem rekomendasi film, yang ditampilkan melalui antarmuka web interaktif. Dengan cara ini, pengguna dapat menjelajahi film berdasarkan genre dan tahun rilis dengan pendekatan yang lebih personal dan sesuai dengan selera mereka.
6. **Simpulan :** Berdasarkan analisis dan hasil evaluasi, dapat disimpulkan bahwa penggunaan algoritma K-Means Clustering berhasil mengelompokkan film dengan efektif. Hal ini menunjukkan potensi algoritma tersebut sebagai fondasi untuk mengembangkan sistem rekomendasi yang lebih personal dan relevan.

2.2 Subjek dan Bahan Penelitian

Fokus utama dari penelitian ini adalah pengembangan sistem rekomendasi film yang memanfaatkan algoritma K-Means Clustering. Sistem ini dirancang untuk mengelompokkan film berdasarkan kesamaan atribut, sehingga dapat memberikan rekomendasi yang sesuai dengan preferensi pengguna. Data yang digunakan dalam penelitian ini adalah IMDB Movies Dataset, yang diambil dari platform Kaggle. Dataset ini dipilih karena menyediakan informasi yang komprehensif dan terstruktur, mencakup atribut-atribut film seperti judul, genre, tahun rilis, rating, dan sutradara. Data ini dianggap representatif untuk menciptakan sistem rekomendasi yang efektif dalam mengidentifikasi pola preferensi penonton berdasarkan data historis dan metadata film yang tersedia.

2.3 Alat yang Digunakan

Penelitian ini menggunakan beberapa perangkat keras dan lunak sebagai alat bantu dalam pengembangan dan pengujian sistem. Dari sisi perangkat keras, digunakan laptop dengan spesifikasi prosesor Intel Celeron N4000, RAM 4 GB, dan penyimpanan SSD 256 GB dengan sistem operasi Windows 11. Dalam hal perangkat lunak, pemrograman dilakukan dengan bahasa Python, memanfaatkan pustaka seperti pandas dan scikit-learn untuk pengolahan data serta penerapan algoritma K-Means Clustering. Untuk pengembangan antarmuka web, HTML dan CSS digunakan, dengan Visual Studio Code sebagai editor. Jupyter Notebook berperan dalam eksplorasi data dan tahap preprocessing.

2.4 Desain Penelitian

Penelitian ini mengadopsi desain pengembangan sistem, bertujuan untuk menciptakan sebuah sistem rekomendasi yang efektif dan dapat diterapkan. Proses penelitian dilaksanakan secara terstruktur, dimulai dengan pengumpulan data, dilanjutkan dengan pra-pemrosesan, penerapan algoritma pengelompokan, evaluasi model, dan diakhiri dengan perancangan antarmuka pengguna. Studi ini menerapkan pendekatan total sampling, di mana semua data yang ada dalam dataset digunakan tanpa pengecualian, sehingga tidak ada teknik sampling yang diterapkan. Variabel yang dianalisis mencakup variabel independen yang terdiri dari atribut film seperti genre, rating, tahun rilis, dan sutradara. Selain itu, terdapat variabel dependen yang berfokus pada hasil rekomendasi film, serta variabel kontrol yang meliputi spesifikasi alat dan ukuran dataset yang konsisten.

2.5 Metode K-Means Clustering

Algoritma K-Means Clustering adalah metode pembelajaran tanpa pengawasan yang digunakan untuk mengelompokkan data ke dalam beberapa cluster berdasarkan kesamaan atribut [4] [5] [6]. Dalam studi ini, algoritma K-Means digunakan untuk mengelompokkan film berdasarkan berbagai atribut, termasuk genre, rating, tahun rilis, dan sutradara. Proses K-Means dimulai dengan menetapkan jumlah cluster yang diinginkan, yang dikenal sebagai nilai k . Nilai ini dipilih secara eksploratif untuk mencapai pengelompokan yang paling efektif. Setelah k ditentukan, algoritma secara acak menginisialisasi centroid awal dari data yang ada. Selanjutnya, sistem menghitung jarak antara setiap data dan semua centroid dengan menggunakan rumus jarak Euclidean [7] [8].

Setiap data akan dikelompokkan ke dalam cluster yang memiliki centroid terdekat, berdasarkan perhitungan jarak yang telah dilakukan. Setelah semua data terorganisir, centroid akan diperbarui dengan menghitung rata-rata dari semua data dalam setiap cluster. Proses ini dikenal sebagai pembaruan centroid. Proses perhitungan jarak dan pembaruan centroid dilakukan secara berulang hingga tidak ada perubahan signifikan pada posisi centroid atau sampai batas maksimum iterasi tercapai. Ketika kondisi ini terpenuhi, itu menunjukkan bahwa algoritma telah mencapai konvergensi, yaitu keadaan stabil di mana hasil pengelompokan tetap konsisten [9], [10].

Evaluasi pengelompokan dilakukan dengan menggunakan metrik Silhouette Score [11], yang menilai konsistensi internal data dalam suatu cluster serta jarak data tersebut dari cluster lain. Nilai Silhouette Score diperoleh dengan menghitung selisih antara rata-rata jarak data ke semua anggota cluster yang sama ($a(i)$) dan rata-rata jarak data ke anggota cluster terdekat lainnya ($b(i)$), kemudian membagi hasilnya dengan nilai maksimum dari kedua rata-rata tersebut. Rumusnya dituliskan sebagai berikut:

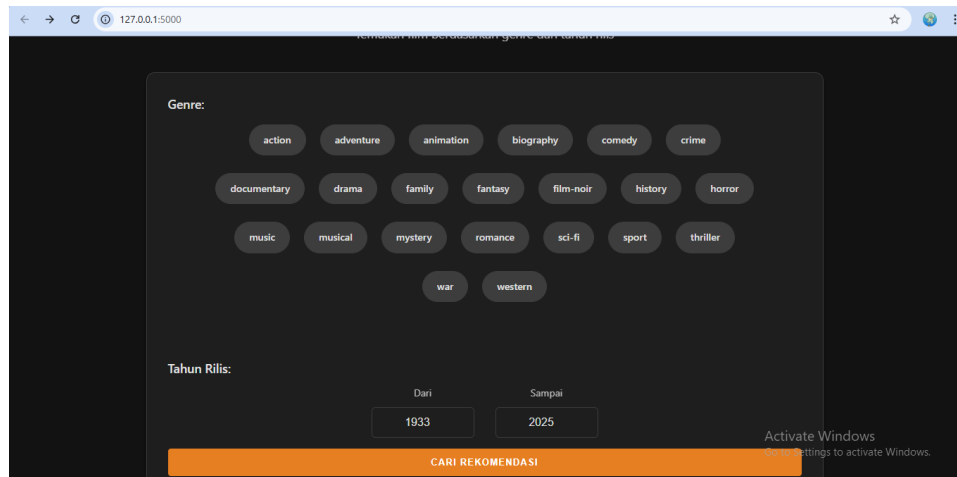
$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (1)$$

Nilai $s(i)$ yang mendekati 1 mengindikasikan bahwa data telah terklasifikasi dengan baik. Sebaliknya, nilai yang mendekati 0 menunjukkan bahwa data berada di perbatasan antara dua cluster, sementara nilai negatif menandakan kemungkinan kesalahan dalam pengelompokan data. Dengan menggunakan metode ini, sistem rekomendasi dapat mengelompokkan film dengan lebih efektif, sehingga hasil rekomendasi yang diberikan sesuai dengan preferensi pengguna.

III. HASIL DAN PEMBAHASAN

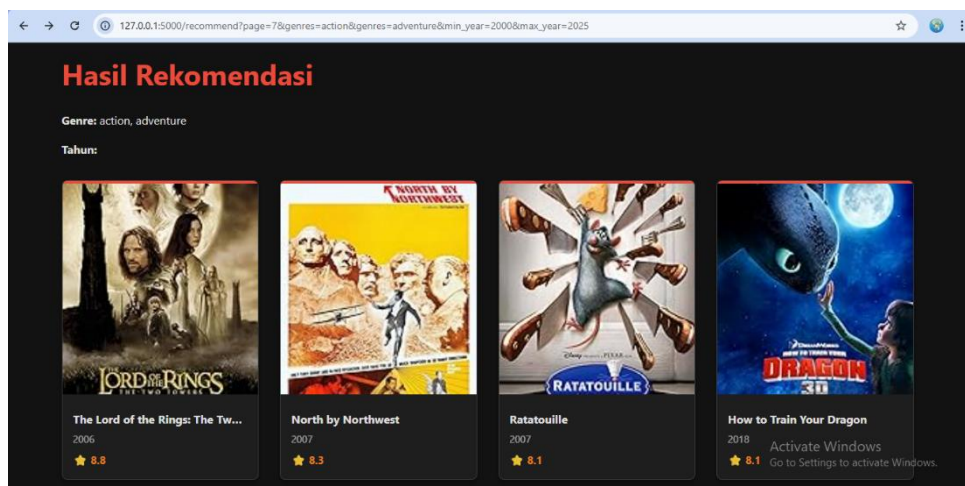
3.1 Pengujian Sistem

Sistem rekomendasi ini dikembangkan dengan menggunakan algoritma K-Means Clustering, pengujian dilakukan untuk memberikan rekomendasi film kepada pengguna. Fokus dari pengujian ini adalah untuk melihat hasil rekomendasi berdasarkan rating dan rating sutradara. Pengujian dilakukan dengan menggunakan aplikasdi berbasis Flask secara lokal melalui browser dengan alamat 127.0.0.1:5000.



Gambar 2. Tampilan Awal

Pada gambar 2 ditampilkan halaman awal pada system rekomendasi. Pengguna bisa memilih genre dan juga tahun rilis film. Pengguna bisa memilih lebih dari satu genre film. Kemudian setelah memilih semua itu, pilih tombol cari rekomendasi untuk menampilkan hasil rekomendasi film.



Gambar 3. Tampilan Rekomendasi

Pada tampilan gambar 3 ini, sistem akan menunjukkan pengelompokan hasil rekomendasi berdasarkan genre dan tahun sesuai pilihan pengguna. Sistem akan merekomendasikan film dengan rating sutradara paling tinggi terlebih dahulu.



Gambar 4. Tampilan Detail Film

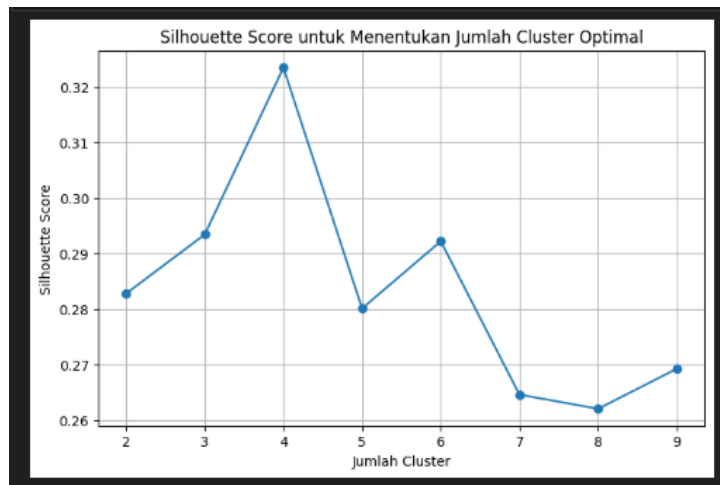
Pada tampilan gambar 4 ini, pengguna bisa melihat detail dari film yang dipilih pengguna dengan lebih jelas. Ada Tahun, Rating Film, Rating Sutradara, Sutradara, Cluster dan Genre film. Contohnya seperti pada gambar, film “The Lord of the Rings: The Two Towers” mempunyai rating film 8,8 dan rating sutradara 7,6. Dan juga film tersebut masuk pada cluster 2.



Gambar 5. Tampilan Detail Film Lainnya

Pada gambar 5, film berjudul “In the Name of the King: A Dungeon Siege Tale” mendapatkan rating film 3,8 dan rating sutradara 3,2. Film tersebut masuk pada cluster 3

3.2 Analisis Hasil



Gambar 6. Menentukan Jumlah Cluster Optimal

Gambar 6 yaitu menentukan cluster paling optimal berdasarkan nilai silhouette score yang paling tinggi. Pada gambar jumlah cluster yang paling optimal itu bernilai 4, dengan hasil silhouette score 0,32 lebih. Jadi sistem pada program menggunakan `n_cluster` 4 untuk pengelompokkannya.

IV. KESIMPULAN

Penerapan algoritma K-Means Clustering dalam pembuatan sistem rekomendasi film berbasis web telah terbukti efektif dalam meningkatkan kualitas rekomendasi yang sesuai dengan selera pengguna. Dengan memanfaatkan berbagai atribut seperti genre, rating, tahun rilis, dan sutradara, algoritma ini mampu mengelompokkan film secara efisien, sehingga menghasilkan rekomendasi yang lebih relevan bagi pengguna. Evaluasi menggunakan Silhouette Score menunjukkan bahwa kualitas pengelompokan yang dihasilkan cukup baik, yang mengindikasikan potensi algoritma ini untuk meningkatkan sistem rekomendasi. Untuk penelitian selanjutnya, disarankan agar peneliti mengeksplorasi penggabungan K-Means Clustering dengan metode lain, seperti Collaborative Filtering. Hal ini bertujuan untuk mengatasi keterbatasan dalam menangani preferensi pengguna yang tidak jelas, serta untuk memperluas atribut data dengan menambahkan elemen seperti ulasan pengguna atau metadata tambahan. Langkah ini diharapkan dapat meningkatkan akurasi rekomendasi dan pemahaman sistem terhadap preferensi pengguna.

DAFTAR PUSTAKA

- [1] I. N. Farida and P. Kasih, "Rancangan Sistem Rekomendasi Bakat Anak dengan Metode AHP dan SAW," vol. 8, pp. 1671–1682, 2024.
- [2] V. No, D. T. Santoso, V. Atina, D. Hartanti, and V. No, "Infotek : Jurnal Informatika dan Teknologi Prototipe Sistem Rekomendasi Film Indonesia Menggunakan Pendekatan Content Based Filtering dan Metode Vector Space Model Infotek : Jurnal Informatika dan Teknologi Film , menurut Pratista dalam bukunya " Memahami," vol. 7, no. 2, 2024.
- [3] A. Halim, H. Gohzali, D. Maria Panjaitan, and I. Maulana, "Sistem Rekomendasi Filmmenggunakan Bisecting K-Means dan Collaborative Filtering," *Citisee*, vol. 4567789, no. 061, pp. 37–41, 2017.
- [4] H. Mutiasari, T. W. Purboyo, and R. A. Nugrahaeni, "Sistem Rekomendasi Film Menggunakan

- Metode K-Means Clustering (Movie Recommendation System Using K-Means Clustering Method),” *e-Proceeding Eng.*, vol. 8, no. 5, pp. 6755–6764, 2021, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/16511>
- [5] M. A. Zartesa and D. S. Prasvita, “Penerapan Collaborative Filtering, PCA dan K-Means dalam Pembangunan Sistem Rekomendasi Film,” *Senamika*, no. April, pp. 606–615, 2021, [Online]. Available: <https://conference.upnvj.ac.id/index.php/senamika/article/view/1318%0Ahttps://conference.upnvj.ac.id/index.php/senamika/article/download/1318/1089>
- [6] V. No, V. No, and D. Dinas, “Infotek : Jurnal Informatika dan Teknologi Implementasi Algoritma K-Means Untuk Rekomendasi Pengadaan Buku Perpustakaan merupakan pusat informasi dan pengetahuan yang sangat penting bagi masyarakat . Salah satu tantangan utama yang dihadapi oleh perpustak,” vol. 8, no. 1, 2025.
- [7] C. Budihartanti, C. I. Ifaru, A. Zahra, and M. H. Aenuddin, “Pengelompokan Film Pada Platform Netflix Menggunakan Metode K-Means Clustering Sebagai Rekomendasi Film,” vol. 5, no. 4, pp. 1392–1402, 2024, doi: 10.47065/josh.v5i4.5482.
- [8] I. Zuhdiansyah and A. Luthfiarta, “Sistem Rekomendasi Pembelian Smartphone berbasis Algoritma K-Means dan Singular Value Decomposition,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 10, no. 1, pp. 45–53, 2024, doi: 10.25077/teknosi.v10i1.2024.45-53.
- [9] M. R. Firmansyah and A. Habib, “MODEL KLASTERISASI DAN ANALISIS SISTEM REKOMENDASI KEPADA PENGGUNA PERPUSTAKAAN UNIVERSITAS 17 AGUSTUS 1945 SURABAYA MENGGUNAKAN K-MEANS,” vol. 2, no. 2, pp. 744–750, 2024.
- [10] A. Dzakiroh and E. Widodo, “Implementasi Algoritma K-Means Untuk Segmentasi Siswa Dalam Pembentukan Kelas Unggulan,” vol. 11, no. 1, pp. 34–43, 2025.
- [11] V. Darsono, Amroni, and A. Andrianti, “Penerapan Data Mining Algoritma K-Means Untuk Rekomendasi Pemilihan Bidang Studi Perguruan Tinggi Pada Siswa SMKN 1 Kota Jambi,” *J. Inform. Dan Rekayasa Komputer(JAKAKOM)*, vol. 2, no. 2, pp. 161–171, 2022, doi: 10.33998/jakakom.2022.2.2.80.