

PENGEMBANGAN SISTEM REKOMENDASI FILM MENGGUNAKAN METODE DATA MINING: STUDI KASUS PADA DATABASE IMDB DENGAN MENGGUNAKAN METODE CLUSTERING (K-MEANS)

Grediyanto¹, Yusuf Kurnia^{2*}

^{1,2} Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Buddhi Dharma

*Corresponding Author, email: yusuf.kurnia@ubd.ac.id

ABSTRAK

Sistem rekomendasi film menjadi semakin penting dalam membantu pengguna menemukan film yang sesuai dengan preferensi mereka, terutama di tengah melimpahnya jumlah film yang tersedia secara daring. Penelitian ini mengangkat topik pengembangan sistem rekomendasi film berbasis data mining dengan menggunakan kombinasi metode Mini-Batch K-Means Clustering, Collaborative Filtering, dan Content-Based Filtering. Permasalahan yang dihadapi adalah bagaimana meningkatkan akurasi rekomendasi agar sesuai dengan kebutuhan pengguna. Penelitian ini menerapkan pendekatan CRISP-DM yang terdiri dari enam tahap: business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Data yang digunakan berasal dari database IMDb, difilter berdasarkan tahun rilis 2014–2024 dan hanya mencakup film berbahasa Inggris. Metode Mini-Batch K-Means digunakan untuk mengelompokkan film berdasarkan kesamaan fitur seperti genre, rating, dan pemeran, sementara Collaborative Filtering dan Content-Based Filtering digunakan untuk menghasilkan rekomendasi yang relevan. Evaluasi sistem dilakukan dengan metrik precision, recall, dan F1-score untuk mengukur kualitas rekomendasi. Hasil evaluasi menunjukkan bahwa penggabungan ketiga metode tersebut mampu meningkatkan akurasi dan relevansi rekomendasi dibandingkan penggunaan metode tunggal. Sistem diuji menggunakan metode Blackbox dan User Acceptance Testing (UAT), dan dinyatakan berjalan dengan baik. Diharapkan penelitian ini memberikan kontribusi dalam pengembangan solusi sistem rekomendasi yang efektif, efisien, dan aplikatif dalam mendukung pengambilan keputusan pengguna dalam memilih film yang sesuai dengan preferensi mereka.

Kata kunci: *Collaborative Filtering, Content-Based Filtering, Data Mining, Mini-Batch K-Means, Sistem Rekomendasi.*

I. PENDAHULUAN

Menurut (Furtado, 2020), industri film telah mengalami pertumbuhan yang signifikan dari tahun ke tahun, dengan lebih dari tiga juta film dirilis hingga 2015. Mengurangi jumlah film tersebut membuatnya lebih sulit bagi pelanggan untuk memilih tontonan yang sesuai dengan preferensi mereka. Dalam hal ini, sistem rekomendasi film memainkan peran penting dalam membantu pengguna menemukan film yang menarik dan relevan. (Ryana Agustian & Prasetyo Nugroho, 2020), keberadaan sistem ini meningkatkan kepuasan pengguna dan menghemat

waktu. Collaborative Filtering dan Content-Based Filtering adalah dua metode yang umum digunakan dalam sistem rekomendasi. Content-Based Filtering menggunakan fitur yang serupa dari film itu sendiri, sedangkan collaborative filter menggunakan kesamaan perilaku antar pengguna untuk menghasilkan rekomendasi (Airen & Agrawal, 2023; (Widiyaningtyas et al., 2021).

Studi ini menyarankan metode yang menggunakan algoritma Mini-Batch K-Means sebagai metode clustering. Metode ini menggabungkan Collaborative dan Content-Based untuk mengatasi keterbatasan ini. Studi sebelumnya menunjukkan bahwa mengelompokkan pengguna atau item ke dalam segmen tertentu yang lebih terfokus dapat meningkatkan akurasi sistem rekomendasi (Gohzali & Panjaitan, 2024) Penelitian oleh (Nellie et al., 2022) juga membuktikan bahwa metode hierarchical clustering dapat memberikan hasil kluster yang akurat dalam pengelompokan film. Selain itu, pendekatan multi-metode ini telah terbukti lebih efisien dan akurat dalam beberapa studi terkini (Wang et al., 2023; (Shaikh, 2020). Penelitian oleh (Sarker et al., 2022) menambahkan bahwa kombinasi filtering dan clustering dapat mengatasi masalah cold start dengan lebih efektif dibanding metode tunggal.

Kami mengharapkan penelitian ini dapat bermanfaat bagi berbagai pemangku kepentingan, termasuk: para pengguna, yang pengalamannya pasti akan meningkat karena ketepatan dan kecepatan mereka dalam menemukan film yang sesuai dengan minat akan meningkat; para peneliti dan pengembang, yang kini memiliki referensi tentang pengembangan sistem hibrida setelah menggunakan penelitian ini serta memperdalam pemahaman tentang penerapan algoritma pengelompokan di bidang tersebut; dan industri layanan streaming seperti Netflix atau Amazon Prime di mana sistem yang lebih personal dan efisien dapat membantu mempertahankan pengguna melalui pengalaman menonton yang sangat relevan. (Rai et al., 2024; Alsekait et al., 2024).

II. METODOLOGI

Pembahasan fokus pada pengembangan sistem rekomendasi film dengan pendekatan Top-N dekat neighbourhood menggunakan teknik rekomendasi alternatif. Oleh karena itu, metode pendekatan CRISP-DM (Cross Industry

Standard Process for Data Mining) diterapkan dalam proyek penelitian ini (Duque et al., 2024).



Gambar 1. Tahapan Crisp-DM

Algoritma Mini-Batch K-Means dan Hierarchical Clustering digunakan sebagai metode pengelompokan (clustering) untuk meningkatkan akurasi sistem rekomendasi film. Mini-Batch K-Means dipilih karena kemampuannya dalam menangani dataset besar secara efisien dengan menggunakan subset data (mini-batch) dalam proses iterasinya, sehingga mempercepat proses komputasi dibanding K-Means konvensional (Shaikh, 2020). Sementara itu, Hierarchical Clustering—khususnya metode Agglomerative—digunakan sebagai pembanding karena mampu memberikan visualisasi dendrogram dan mendeteksi struktur alami data tanpa perlu menentukan jumlah cluster di awal (Nellie et al., 2022). Hasil evaluasi menunjukkan bahwa Mini-Batch K-Means memiliki nilai F1-score yang lebih tinggi dan performa yang lebih stabil dibandingkan dengan Hierarchical Clustering dalam konteks data film IMDb (Gohzali & Panjaitan, 2024).

III. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Masalah yang dihadapi adalah ketidakakuratan sistem rekomendasi konvensional dalam menghadapi data berjumlah besar serta masalah cold start. Metode seperti Content-Based Filtering dan Collaborative Filtering sering kali memiliki keterbatasan dalam menghadapi variasi pengguna dan item baru yang belum memiliki interaksi historis, serta rentan terhadap masalah data sparsity dan over-specialization (Datta et al., 2020) Oleh karena itu, penelitian ini mengadopsi pendekatan hybrid yang menggabungkan metode Mini-Batch K-Means, Collaborative Filtering, dan Content-Based Filtering (Gohzali & Panjaitan, 2024). Pendekatan seperti ini telah terbukti dalam penelitian sebelumnya dapat meningkatkan akurasi dan personalisasi sistem rekomendasi dengan memanfaatkan

kekuatan masing-masing metode filtering dan clustering secara sinergis (Ruchika et al., 2023).

3.2 Data Understanding

Data dikumpulkan dari website Bernama Kaggle IMDb Dataset (2024) updated, mencakup atribut seperti judul film, tahun rilis (2014–2024), genre, rating IMDb, meta score, sutradara, dan pemeran utama. Analisis awal dilakukan untuk memahami distribusi data, mendeteksi nilai kosong, serta mengevaluasi fitur-fitur penting yang relevan untuk sistem rekomendasi.

Tabel 1. Keterangan Dataset

Nama Variabel	Tipe	Deksripsi	Satuan
<i>Title</i>	<i>Binary</i>	Judul film	
<i>Score</i>	<i>Integer</i>	Score dari IMDB	
<i>Year</i>	<i>Integer</i>	Tahun film	Tahun
<i>Genre</i>	<i>Binary</i>	Jenis-jenis film	
<i>Director</i>	<i>Binary</i>	Pembuat film / sutradara film	
<i>Casting</i>	<i>Binary</i>	Pemain yang bermain dalam film	
<i>Duration (Minutes)</i>	<i>Integer</i>	Waktu film berlangsung	minutes

3.3 Data Preparation

Yang pertama adalah perbaikan data; analisis ini akan menghapus nilai-nilai ceroboh dan duplikat dari atribut genre, rating, tahun, dan aktor. Selanjutnya adalah transformasi data, di mana kategori genre dan karakter utama diubah menjadi kolom angka dengan menggunakan satu hot encoding dan TF-IDF yang diimbangi untuk teks. Untuk fitur numerik lainnya, seperti rating IMDb dan durasi film, normalisasi Z-Score akan digunakan untuk memastikan bahwa semua nilai berada pada skala yang sama tanpa ada yang mendominasi selama clustering. Untuk mempercepat proses pemodelan dan mengurangi kompleksitas perhitungan, sejumlah atribut seperti genre, peringkat, skor meta, dan pemeran utama dikurangi dari data. Tujuan dari semua ini adalah untuk mempersiapkan seluruhnya dalam kondisi yang konsisten dan ideal sebelum diproses dengan Mini-Batch K-Means, Filter Berbasis Konten, dan Filter Berkolaborasi.

3.4 Modelling

Pada tahap pemodelan menggabungkan tiga algoritma utama: Mini-Batch K-Means, Collaborative Filtering, dan Content-Based Filtering. Mini-Batch K-Means digunakan untuk mengelompokkan film ke dalam beberapa cluster berdasarkan kesamaan fitur seperti genre, rating IMDb, meta score, dan aktor utama, memberikan efisiensi yang lebih tinggi dibandingkan K-Means konvensional untuk set data yang besar. Setelah proses pengelompokan, dua metode penyaringan diterapkan oleh sistem untuk menghasilkan rekomendasi. Pertama, filtrasi kolaboratif melihat kesamaan perilaku pengguna melalui penilaian dan hubungan preferensi penyuka dengan pendekatan berbasis item. Kedua, filtrasi berbasis konten menghitung kesamaan antara film berdasarkan isinya, seperti genre, sutradara, dan pemeran, menggunakan teks TF-IDF yang berbobot untuk menggunakan Cosine Similarity. Berdasarkan preferensi historis dan kesamaan konten film, sistem dapat memberikan rekomendasi yang lebih personal, relevan, dan akurat kepada pengguna setelah kedua metode ini digabungkan.

3.5 Evaluation

Pada bagian evaluasi ini, sistem rekomendasi yang telah dibuat dan menggunakan pendekatan Hybrid Filtering Collaborative, Mini-Batch K-Means serta Filtering Berbasis Konten, akan dievaluasi sesuai dengan kinerja yang dicapai. Untuk menganalisisnya, kerangka informatika dan klasifikasi lainnya digunakan untuk mengukur tiga metrik utama di atas: presisi, recall dan F1 score. Presisi merupakan salah satu metrik yang menilai seberapa relevan subset rekomendasi yang diberikan sistem dibandingkan dengan kesukaan sang pengguna. Recall menghitung seberapa mampu sistem mengambil sebuah film relevan dari kumpulan film-film relevan yang ada. Metrik tersebut diperoleh dari hasil seimbang F1 dimana kedua sisi dari efektivitas recall, dan presisi saling berimbangan. Terdapat 30 data film ground truth sebagai bahan evaluasi yang dikumpulkan dari berbagai genre serta tahun rilis. Hasil evaluasi menunjukkan bahwa kombinasi Mini-Batch K-Means dengan Dua Metode Penyaringan memberikan hasil yang jauh lebih baik dibandingkan single method maupun pada metode lain yang diajukan seperti Clustering Hierarchical.

3.6 Development

Sistem rekomendasi yang telah dibuat diimplementasikan dalam sebuah aplikasi web berbasis Python yang menggunakan framework Flask. Dengan aplikasi ini, pengguna dapat memasukkan judul film yang mereka sukai, dan secara otomatis, sistem akan memberikan rekomendasi film dengan menggunakan gabungan algoritma Mini-Batch K-Means, Collaborative Filtering, dan Content-Based Filtering. Proses backend mencakup pemanggilan model clustering dan filtering, serta perhitungan Cosine Similarity antar item dalam alur sistem yang terintegrasi (Azmi et al., 2024). Antarmuka pengguna (user interface) dirancang dengan tampilan sederhana dan responsif untuk memudahkan interaksi, sedangkan proses rekomendasi dilakukan secara *on-demand* berdasarkan input pengguna. Untuk memastikan sistem berjalan sebagaimana mestinya, dilakukan pengujian Blackbox Testing untuk memverifikasi seluruh fungsi, serta User Acceptance Testing (UAT) untuk mengevaluasi kepuasan pengguna terhadap sistem rekomendasi dan kemudahan penggunaan aplikasi (Asekait et al., 2024). Hasil pengujian menunjukkan bahwa sistem berfungsi dengan baik, dan relevansi rekomendasi yang diberikan sesuai dengan harapan pengguna. Meskipun sistem ini belum terintegrasi secara langsung dengan platform streaming atau sistem real-time, sistem telah siap digunakan sebagai prototipe dan memiliki potensi untuk dikembangkan lebih lanjut pada skala implementasi yang lebih luas (Gohzali & Panjaitan, 2024).

IV. SIMPULAN

Penelitian ini berhasil mengembangkan sistem rekomendasi film berbasis data mining dengan mengintegrasikan algoritma Mini-Batch K-Means, Collaborative Filtering, dan Content-Based Filtering menggunakan data dari IMDb. Hasil evaluasi menunjukkan bahwa kombinasi metode tersebut mampu meningkatkan akurasi dan relevansi rekomendasi film berdasarkan preferensi pengguna. Sistem yang dibangun juga mampu mengatasi keterbatasan metode tunggal dan memberikan pengalaman pencarian film yang lebih efisien dan personal.

DAFTAR PUSTAKA

- Airen, S., & Agrawal, J. (2023). ScienceDirect Movie Recommender System Parameter Tuning of User and Movie Neighbourhood Movie Recommender System Using via Parameter Tuning of User and Movie Neighbourhood via. *Procedia Computer Science*, 218, 1176–1183. <https://doi.org/10.1016/j.procs.2023.01.096>
- Alsekait, D. M., Shdefat, A. Y., Mostafa, N., Hamdy, A. M. M., Fathi, H., & Abdelminaam, D. S. (2024). Next-Generation Movie Recommenders: Leveraging Hybrid Deep Learning for Enhanced Personalization. *Applied Mathematics and Information Sciences*, 18(5), 957–981. <https://doi.org/10.18576/amis/180504>
- Azmi, A. H., Naufal, I., Mahardika, R., Shaktika, A., Prasetya, A., & Puspita, S. (2024). Sistem Rekomendasi Film Berbasis Konten Menggunakan Teknik Cosine Similarity dan TF-IDF. *Journal of Multidisciplinary Inquiry in Science Technology and Educational Research*, 1(4). <https://doi.org/10.32672/mister.v1i4.2206>
- Datta, D., Navamani, T. M., & Deshmukh, R. (2020). Products And Movie Recommendation System For Social Networking Sites. *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, 9(10). www.ijstr.org
- Duque, J., Godinho, A., Moreira, J., & Vasconcelos, J. (2024). Data Science with Data Mining and Machine Learning A design science research approach. *Procedia Computer Science*, 237, 245–252. <https://doi.org/10.1016/j.procs.2024.05.102>
- Furtado, F. (2020). *Movie Recommendation System Using Machine Learning*. 9(1), 84–98. <https://doi.org/10.22105/riej.2020.226178.1128>
- Gohzali, H., & Panjaitan, D. M. (2024). Movie Recommendation System Model using Bisecting K-Means Technique and Collaborative Filtering. In *Journal of Multimedia Trend and Technology-JMTT* (Vol. 3, Issue 2). <https://journal.educollabs.org/index.php/jmtt/>
- Nellie, V., Viny,), Mawardi, C., Novario,), & Perdana, J. (2022). *Jurnal Ilmu Komputer dan Sistem Informasi IMPLEMENTASI METODE*

AGGLOMERATIVE HIERARCHICAL CLUSTERING UNTUK SISTEM REKOMENDASI FILM.

- Rai, A. K., Khanna, P., Singh, P., Rai, A. K., Khanna, P., & Singh, P. (2024). Movie Recommender System: Collaborative Filtering Methods "Movie Recommender System: Collaborative Filtering Methods". *Journal of Management and Service Science*, 04(066), 1–11. <https://doi.org/10.54060/a2zjourna>
- Ruchika, Sharma, M., & Hossain, S. A. (2023). Efficient Machine Learning Algorithms in Hybrid Filtering Based Recommendation System. *Journal of Information Technology Management*, 15(3), 134–161. <https://doi.org/10.22059/jitm.2023.93631>
- Ryana Agustian, E., & Prasetyo Nugroho, E. (2020). *Sistem Rekomendasi Film Menggunakan Metode Collaborative Filtering dan K-Nearest Neighbors Film Recommendation System Using Collaborative Filtering Method and K-Nearest Neighbors* (Vol. 3, Issue 1). <https://ejournal.upi.edu/index.php/JATIKOM>
- Sarker, K. U., Saqib, M., Hasan, R., Mahmood, S., Hussain, S., Abbas, A., & Deraman, A. (2022). A Ranking Learning Model by K-Means Clustering Technique for Web Scraped Movie Data. *Computers*, 11(11). <https://doi.org/10.3390/computers11110158>
- Shaikh, M. I. (2020). *Top-N Nearest Neighbourhood based Movie Recommendation System using different Recommendation Techniques MSc Research Project Data Analytics*.
- Wang, L., Mistry, S., Hasan, A. A., Hassan, A. O., Islam, Y., & Junior Osei, F. A. (2023). Implementation of a Collaborative Recommendation System Based on Multi-Clustering. *Mathematics*, 11(6). <https://doi.org/10.3390/math11061346>
- Widiyaningtyas, T., Hidayah, I., & Adji, T. B. (2021). User profile correlation - based similarity (UPCSim) algorithm in movie recommendation system. *Journal of Big Data*. <https://doi.org/10.1186/s40537-021-00425-x>