

Implementasi Metode *K-Nearest Neighbor* dalam Sistem Rekomendasi Film Berdasarkan Preferensi Pengguna

Siti Nuril Maghfiratus Sholeha, Zaehol Fatah

Universitas Ibrahimy, Situbondo, Indonesia

E-mail: nurilmaghfira82@gmail.com, zaeholfatah@gmail.com

KEYWORD

movie
recommendations; k-
nearest neighbor;
recommendation
systems; user
preferences

ABSTRACT

The increase in film production and availability on streaming platforms has made it challenging for users to find movies that match their preferences. A recommendation system can help users by suggesting movies that align with their interests. This study aims to implement the *K-Nearest Neighbor* (KNN) algorithm in a film recommendation system to provide suggestions based on user preferences. The KNN algorithm was used to classify films by analyzing user preferences from given ratings. Data was collected, preprocessed, and divided into training and testing sets to optimize the model. Cosine similarity was employed to measure similarity between films. The recommendation system achieved an accuracy of 44.83% with a precision of 100% at $K=10$, indicating high relevance but highlighting potential improvements in overall accuracy. The KNN-based recommendation system effectively aligns with user preferences, though additional techniques may enhance accuracy further. This system shows promise in providing relevant movie suggestions but requires further optimization for broader application.

KATA KUNCI

rekomendasi film; k-
nearest neighbor;
sistem rekomendasi;
preferensi pengguna

ABSTRAK

Peningkatan produksi film dan ketersediaan pada platform streaming telah membuat pengguna kesulitan menemukan film yang sesuai dengan preferensi mereka. Sistem rekomendasi dapat membantu pengguna dengan menyarankan film yang sesuai dengan minat mereka. Studi ini bertujuan untuk mengimplementasikan algoritma *K-Nearest Neighbor* (KNN) dalam sistem rekomendasi film untuk memberikan saran berdasarkan preferensi pengguna. Algoritma KNN digunakan untuk mengklasifikasikan film dengan menganalisis preferensi pengguna dari peringkat yang diberikan. Data dikumpulkan, diproses terlebih dahulu, dan dibagi menjadi set pelatihan dan pengujian untuk mengoptimalkan model. Kesamaan kosinus digunakan untuk mengukur kesamaan antar film. Sistem rekomendasi mencapai akurasi 44,83% dengan presisi 100% pada $K=10$, yang menunjukkan relevansi tinggi tetapi menyoroti potensi peningkatan akurasi keseluruhan. Sistem rekomendasi berbasis KNN secara efektif selaras dengan preferensi

pengguna, meskipun teknik tambahan dapat meningkatkan akurasi lebih jauh. Sistem ini menunjukkan janji dalam memberikan saran film yang relevan tetapi memerlukan pengoptimalan lebih lanjut untuk aplikasi yang lebih luas.

PENDAHULUAN

Metode K-Nearest Neighbor (KNN) memiliki keunggulan khusus dibandingkan dengan metode lain dalam konteks sistem rekomendasi, terutama dalam hal kesederhanaan dan kemampuannya mengadaptasi preferensi pengguna secara langsung tanpa perlu pelatihan model yang kompleks (Hidayat et al., 2024). KNN bekerja dengan cara membandingkan kesamaan preferensi pengguna berdasarkan data historis, seperti genre dan rating film, serta menentukan rekomendasi berdasarkan film-film yang mirip dengan yang telah ditonton pengguna sebelumnya (Rochmad Wahono et al., 2024).

Keunggulan utama KNN adalah kemampuannya untuk memberikan hasil rekomendasi yang lebih personal dan akurat secara instan, karena tidak membutuhkan pemrosesan berlebihan yang sering kali diperlukan oleh metode berbasis faktor seperti Matrix Factorization. Dalam sistem rekomendasi film, metode ini efektif karena dapat mengelompokkan pengguna atau film berdasarkan kesamaan preferensi, menghasilkan saran yang sesuai secara langsung dengan minat pengguna (Zakharia et al., 2024). Selain itu, KNN sangat adaptif terhadap perubahan preferensi pengguna, karena setiap rekomendasi dihitung ulang berdasarkan data terbaru yang dimasukkan, membuatnya lebih responsif terhadap selera yang berubah. Hal ini memberikan pengalaman pengguna yang lebih dinamis, terutama bagi pengguna yang memiliki preferensi film yang beragam. Namun, untuk meningkatkan performa akurasi dan presisi lebih jauh, KNN juga dapat dikombinasikan dengan pendekatan hybrid atau tambahan, seperti cosine similarity, yang meningkatkan relevansi rekomendasi tanpa mengorbankan efisiensi. Untuk membantu pengguna dalam menemukan film yang relevan dengan preferensi mereka, berbagai sistem rekomendasi telah dikembangkan.

Sistem rekomendasi memanfaatkan teknik – teknik tertentu untuk menganalisis preferensi pengguna berdasarkan data yang ada, seperti riwayat tontonan, rating, dan genre film yang disukai (Azmi et al., 2024). Salah satu teknik yang umum digunakan dalam sistem rekomendasi adalah metode *K-Nearest Neighbor* (KNN). KNN merupakan algoritma yang dapat digunakan untuk mengklasifikasikan objek baru berdasarkan kesamaan atribut dengan objek – objek lain yang telah ada dalam dataset. Pada sistem rekomendasi film, KNN dapat digunakan untuk menemukan film – film yang mirip dengan film yang telah ditonton atau disukai oleh pengguna sebelumnya (Fitrianti et al., 2020). Pendekatan berbasis KNN dalam sistem rekomendasi memungkinkan pengguna untuk menerima saran film yang sesuai dengan preferensi pribadi mereka. Algoritma ini bekerja dengan mengukur kemiripan antara film yang telah dinilai oleh pengguna dengan film – film lain yang tersedia, menggunakan metrik seperti *cosine similarity* atau *euclidean distance*. Sistem rekomendasi film berbasis KNN telah terbukti mampu meningkatkan akurasi rekomendasi dan memberikan pengalaman menonton yang lebih personal (Ciaputra & Hansun, 2020).

Penelitian ini bertujuan untuk mengimplementasikan metode K-Nearest Neighbor dalam sistem rekomendasi film berbasis preferensi pengguna (Hutagalung et al., 2024). Sistem ini dirancang untuk membantu pengguna menemukan film yang sesuai dengan minat mereka dengan menggunakan data rating dan genre sebagai parameter utama dalam analisis kemiripan.

Diharapkan implementasi sistem ini dapat memberikan rekomendasi yang lebih relevan serta meningkatkan kepuasan pengguna dalam memilih film yang mereka sukai.

Manfaat penelitian ini mencakup berbagai aspek yang berkontribusi pada industri streaming film, peningkatan kualitas rekomendasi, kontribusi akademis, analisis data preferensi, dan efisiensi teknologi rekomendasi. Pertama, dari sisi praktis, penelitian ini memberikan solusi yang dapat diimplementasikan oleh platform streaming film untuk meningkatkan pengalaman pengguna. Sistem rekomendasi yang akurat dapat membantu pengguna menemukan film sesuai preferensi mereka, yang pada akhirnya meningkatkan kepuasan dan loyalitas pengguna terhadap platform tersebut. Kedua, penelitian ini menggunakan metode K-Nearest Neighbor (KNN) yang diharapkan mampu meningkatkan relevansi dan akurasi rekomendasi film. Dengan demikian, pengguna dapat lebih mudah menemukan film yang diminati, tanpa harus merasa bingung di tengah banyaknya pilihan. Selain manfaat praktis, penelitian ini juga berkontribusi dalam ranah akademis dengan menambah literatur terkait penggunaan algoritma KNN dalam sistem rekomendasi, khususnya dalam konteks rekomendasi film. Hal ini dapat menjadi referensi bagi penelitian lebih lanjut di bidang kecerdasan buatan dan machine learning, terutama dalam mengembangkan sistem rekomendasi yang efisien dan adaptif. Lebih lanjut, sistem ini memungkinkan analisis data preferensi pengguna secara mendalam, yang bisa dimanfaatkan untuk memahami pola menonton serta tren minat pengguna. Dengan informasi ini, industri film dapat mengembangkan konten yang lebih sesuai dengan preferensi pasar (Fanani, 2024).

Terakhir, penelitian ini memperkenalkan KNN sebagai teknologi yang tidak hanya efisien dalam memproses data tetapi juga mampu memberikan rekomendasi yang personal tanpa pelatihan model yang kompleks. Teknologi ini sangat cocok untuk aplikasi yang membutuhkan rekomendasi real-time, memberikan hasil yang relevan sesuai perubahan selera pengguna. Secara keseluruhan, penelitian ini berpotensi meningkatkan efisiensi dan personalisasi dalam teknologi rekomendasi, mendukung pengembangan platform yang lebih responsif terhadap kebutuhan pengguna.

METODE

Penelitian ini menggunakan metode *K-Nearest Neighbor* (KNN) untuk membangun sistem rekomendasi film berbasis preferensi pengguna. KNN dipilih karena kemampuannya untuk mengklasifikasikan objek baru berdasarkan kesamaan atribut dengan objek yang sudah ada dalam dataset. Pada sistem rekomendasi film, KNN bekerja dengan menganalisis kesamaan antara film yang telah ditonton pengguna dan film lain yang tersedia berdasarkan genre, rating, dan atribut lainnya. (Rochmad Wahono et al., 2024)

Langkah – langkah penelitian ini adalah sebagai berikut :

1. Pengumpulan Data

Dataset film dan rating pengguna diambil dari website kaggle. Dataset ini terdiri dari beberapa atribut penting seperti ID film, judul, genre, dan rating yang diberikan oleh pengguna. (Fitrianti et al., 2020) Pengumpulan data dilakukan untuk mengumpulkan informasi yang relevan terkait film dan riwayat rating dari pengguna.

2. Preprocessing Data

Data yang dikumpulkan melalui tahap preprocessing, yaitu mencakup pembersihan data (data cleaning), penghapusan data yang tidak relevan atau duplikat, dan normalisasi data untuk memastikan konsistensi format. Pada tahap ini, atribut – atribut yang tidak diperlukan seperti deskripsi film atau poster dihapus, sementara atribut yang relevan seperti genre dan rating tetap dipertahankan untuk analisis lebih lanjut.

3. Pembagian Data

Dataset dibagi menjadi dua bagian yaitu data training dan data testing. Data training digunakan untuk melatih algoritma KNN, sedangkan data testing digunakan untuk menguji akurasi dan performa algoritma. Pembagian data dilakukan dengan rasio 80:20, di mana 80% data digunakan sebagai data training dan 20% sebagai data testing (Rochmad Wahono et al., 2024)

4. Implementasi Algoritma *K-Nearest Neighbor* (KNN).

Algoritma KNN diterapkan untuk menghitung kesamaan antara film yang sudah ditonton pengguna dengan film lain yang belum ditonton. Dalam penelitian ini, metrik yang digunakan untuk mengukur kesamaan antar film adalah cosine similarity. Algoritma KNN akan mencari sejumlah tetangga terdekat dari film yang telah ditonton berdasarkan kemiripan genre dan rating, kemudian memberikan rekomendasi film yang paling mirip (Ciaputra & Hansun, 2020).

Penentuan Nilai K yaitu dimana nilai k yang optimal perlu ditentukan untuk memastikan akurasi prediksi. Nilai k dipilih berdasarkan percobaan dengan beberapa nilai dan mengukur performa algoritma dalam memberikan rekomendasi yang tepat.

5. Evaluasi Sistem

Kinerja algoritma KNN dievaluasi dengan menggunakan matriks confusion matrix untuk menghitung nilai presisi, recall, dan akurasi sistem.

- a. Presisi adalah proporsi rekomendasi yang benar – benar relevan dengan preferensi pengguna dari total rekomendasi yang dihasilkan.
- b. Recall adalah proporsi film yang relevan yang berhasil direkomendasikan oleh sistem dari total film yang seharusnya direkomendasikan.

Uji coba dilakukan dengan menggunakan data testing untuk membandingkan hasil rekomendasi yang dihasilkan oleh sistem dengan preferensi pengguna sesungguhnya. Pada akhir pengujian, sistem diharapkan mampu memberikan rekomendasi dengan presisi dan recall yang tinggi, sesuai dengan preferensi pengguna.

6. Pengembangan Sistem Prototipe

Sistem rekomendasi film berbasis web dibangun menggunakan metode prototyping. Proses pengembangan ini melibatkan beberapa tahap, yaitu komunikasi dengan pengguna untuk mengidentifikasi kebutuhan, perancangan model sistem, pembangunan sistem, serta pengujian dan evaluasi sistem secara langsung oleh pengguna (Fitrianti et al., 2020).

HASIL DAN PEMBAHASAN

Dalam penelitian ini, metode K-Nearest Neighbors (KNN) digunakan untuk membangun sistem rekomendasi film berdasarkan dataset ratings_small.csv, yang terdiri dari 671 pengguna dan 9066 film. Dataset ini diolah untuk membuat matriks user-item, di mana setiap baris mewakili pengguna dan setiap kolom mewakili film, dengan rating sebagai nilai. Matriks ini membantu dalam mengidentifikasi kesamaan preferensi film diantara pengguna.

1. Model KNN

Model KNN dibangun dengan menghitung Cosine Similarity antara pengguna, yang mengukur tingkat kesamaan preferensi antara mereka. Berdasarkan kesamaan ini, kami mengidentifikasi K tetangga terdekat (nearest neighbors) untuk setiap pengguna. Pada penelitian ini, beberapa nilai K diuji, seperti K=5, K=8, dan K=10, untuk menentukan pengaruh jumlah tetangga terhadap kualitas rekomendasi.

2. Pengujian dan Evaluasi

Proses pengujian dilakukan dengan membagi dataset menjadi 70% data pelatihan dan 30% data pengujian. Pembagian ini bertujuan untuk melatih algoritma KNN pada data pelatihan dan

mengevaluasi performanya menggunakan data pengujian. Evaluasi kinerja sistem dilakukan dengan menghitung tiga metrik utama, yaitu :

- a. Akurasi yaitu mengukur seberapa tepat sistem dalam memberikan rekomendasi film.
- b. Precision yaitu mengukur proporsi rekomendasi yang relevan dari total rekomendasi yang diberikan oleh sistem.
- c. Recall yaitu mengukur seberapa banyak film yang relevan berhasil direkomendasikan oleh sistem dari total film yang relevan.

Berikut adalah hasil evaluasi untuk beberapa nilai K :

Tabel 1. Hasil Evaluasi

Nilai K	Akurasi (%)	Recall (%)	Precision (%)
K=5	42.81	42.81	100
K=8	43.86	43.86	100
K=10	44.83	44.83	100

Dari tabel di atas, terlihat bahwa nilai akurasi tertinggi dicapai saat K=10, dengan nilai akurasi sebesar 44.83%. Selain itu, nilai precision tetap 100% untuk semua nilai K, yang menunjukkan bahwa semua rekomendasi yang dihasilkan oleh sistem adalah benar sesuai dengan preferensi pengguna. Namun, akurasi dan recall menunjukkan bahwa ada ruang untuk peningkatan.

Pembahasan

Hasil penelitian ini menunjukkan bahwa semakin besar nilai K, sistem memberikan akurasi dan recall yang lebih baik, meskipun perbedaannya tidak terlalu signifikan antara K=8 dan K=10. Nilai precision yang tinggi menunjukkan bahwa rekomendasi yang dihasilkan sangat relevan, namun nilai akurasi yang masih di bawah 50% menunjukkan bahwa metode KNN mungkin memerlukan penyesuaian lebih lanjut atau penggunaan metode lain, seperti Matrix Factorization atau algoritma hybrid, untuk meningkatkan kualitas rekomendasi.

KESIMPULAN

Penelitian ini berhasil mengimplementasikan metode K-Nearest Neighbors (KNN) untuk membangun sistem rekomendasi film berbasis preferensi pengguna dengan menggunakan dataset *ratings_small.csv* yang berisi 671 pengguna dan 9066 film. Berdasarkan hasil pengujian, sistem KNN dengan nilai K yang lebih besar (K=10) memberikan akurasi dan recall yang lebih tinggi dibandingkan nilai K yang lebih kecil, meskipun peningkatan akurasi tidak terlalu signifikan antara K=8 dan K=10. Precision yang konsisten mencapai 100% menunjukkan bahwa semua rekomendasi yang diberikan sesuai dengan preferensi pengguna, namun tingkat akurasi yang masih berada di bawah 50% mengindikasikan adanya ruang untuk perbaikan. Pemilihan nilai K dalam algoritma KNN memiliki pengaruh signifikan terhadap akurasi rekomendasi. Dalam penelitian ini, berbagai nilai K diuji untuk menemukan titik optimal di mana akurasi, presisi, dan recall sistem berada pada level terbaik. Nilai K yang terlalu kecil dapat menyebabkan sistem terlalu terfokus pada tetangga terdekat, yang dapat mengurangi keragaman rekomendasi. Sebaliknya, nilai K yang terlalu besar dapat membuat rekomendasi menjadi kurang spesifik dan relevan. Berdasarkan uji coba, nilai K=10 dipilih karena memberikan keseimbangan terbaik antara akurasi dan relevansi rekomendasi dalam konteks preferensi pengguna. Penentuan parameter K dapat lebih ditingkatkan melalui teknik seperti validasi silang (cross-validation), yang membagi dataset ke dalam beberapa subset untuk menguji berbagai nilai K pada lebih banyak variasi data. Selain itu,

pendekatan otomatis seperti pencarian grid (grid search) atau algoritma optimasi, seperti algoritma genetika, dapat digunakan untuk menentukan nilai K yang optimal secara adaptif. Dengan pendekatan ini, nilai K dapat disesuaikan secara dinamis berdasarkan karakteristik data pengguna yang berbeda, menghasilkan rekomendasi yang lebih akurat dan personal. Hal ini menunjukkan bahwa meskipun KNN efektif dalam memberikan rekomendasi yang relevan, metode ini mungkin perlu dikombinasikan dengan pendekatan lain, seperti *Matrix Factorization* atau algoritma *hybrid*, untuk meningkatkan akurasi sistem secara keseluruhan. Dengan demikian, meskipun KNN dapat memberikan rekomendasi film yang sesuai dengan preferensi pengguna, masih diperlukan pengembangan lebih lanjut untuk meningkatkan performa sistem rekomendasi, terutama dalam hal akurasi.

DAFTAR PUSTAKA

- Azmi, A. H., Mahardika, N. I., PRASETYA, R. S. A., & SARI, A. P. (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), 2135–2145.
- Ciaputra, A. T., & Hansun, D. S. (2020). Rekomendasi Pemilihan Film Dengan Hybrid Filtering Dan K-Nearest Neighbor. *Jurnal Rekayasa Informasi*, 9(2), vol 9-2.
- Fanani, M. A. (2024). Sistem Rekomendasi Film Menggunakan Metode K-NN. *Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer*, 4(1), 178–185.
- Fitrianti, A. R., Rohmani, A., & Widjanarto, W. (2020). Sistem Rekomendasi Film Berbasis Website Dengan Metode Prototype Menggunakan Algoritma K-Nearest Neighbors (KNN). *JOINS (Journal of Information System)*, 5(2), 278–287. <https://doi.org/10.33633/joins.v5i2.4168>
- Hidayat, R., Wisnu, W. S., Kurnia, K. N., & Farhan, F. H. (2024). Penerapan Metode K-Nearest Neighbor Untuk Pemilihan Rekomendasi Film. *Technology and Informatics Insight Journal*, 3(1), 1–9.
- Hutagalung, D. D., Hanifurohman, C., & Baskhara, D. R. (2024). Bot Sistem Rekomendasi Webtoon Dengan Metode User-Collaborative Filtering Menggunakan Algoritma K-Nearest Neighbor (KNN). *LOGIC: Jurnal Ilmu Komputer Dan Pendidikan*, 2(4), 751–758.
- Rochmad Wahono, A., Aji Saputra, B., & Fadlu Rahman, F. (2024). Sistem Rekomendasi Film Menggunakan Metode Content-Based Filtering dan Algoritma K-Nearest Neighbors (KNN). *Prosiding Seminar Nasional Teknologi Informasi Dan Bisnis*, 1–6. <https://doi.org/10.47701/senatib.v4i1.3994>
- Zakharia, A., Ulhaq, A. D., Suryono, A. B., Nugroho, N. C., Hafith, D. F., & Gusmao, N. D. A. (2024). Sistem Rekomendasi Film Indonesia Menggunakan Metode Content-Based Filtering. *LOGIC: Jurnal Ilmu Komputer Dan Pendidikan*, 2(4), 671–678.
- Azmi, A. H., Mahardika, N. I., PRASETYA, R. S. A., & SARI, A. P. (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), 2135–2145.
- Ciaputra, A. T., & Hansun, D. S. (2020). Rekomendasi Pemilihan Film Dengan Hybrid Filtering Dan K-Nearest Neighbor. *Jurnal Rekayasa Informasi*, 9(2), vol 9-2.
- Fanani, M. A. (2024). Sistem Rekomendasi Film Menggunakan Metode K-NN. *Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer*, 4(1), 178–185.
- Fitrianti, A. R., Rohmani, A., & Widjanarto, W. (2020). Sistem Rekomendasi Film Berbasis

- Website Dengan Metode Prototype Menggunakan Algoritma K-Nearest Neighbors (KNN). *JOINS (Journal of Information System)*, 5(2), 278–287. <https://doi.org/10.33633/joins.v5i2.4168>
- Hidayat, R., Wisnu, W. S., Kurnia, K. N., & Farhan, F. H. (2024). Penerapan Metode K-Nearest Neighbor Untuk Pemilihan Rekomendasi Film. *Technology and Informatics Insight Journal*, 3(1), 1–9.
- Hutagalung, D. D., Hanifurohman, C., & Baskhara, D. R. (2024). Bot Sistem Rekomendasi Webtoon Dengan Metode User-Collaborative Filtering Menggunakan Algoritma K-Nearest Neighbor (KNN). *LOGIC: Jurnal Ilmu Komputer Dan Pendidikan*, 2(4), 751–758.
- Rochmad Wahono, A., Aji Saputra, B., & Fadlu Rahman, F. (2024). Sistem Rekomendasi Film Menggunakan Metode Content-Based Filtering dan Algoritma K-Nearest Neighbors (KNN). *Prosiding Seminar Nasional Teknologi Informasi Dan Bisnis*, 1–6. <https://doi.org/10.47701/senatib.v4i1.3994>
- Zakharia, A., Ulhaq, A. D., Suryono, A. B., Nugroho, N. C., Hafith, D. F., & Gusmao, N. D. A. (2024). Sistem Rekomendasi Film Indonesia Menggunakan Metode Content-Based Filtering. *LOGIC: Jurnal Ilmu Komputer Dan Pendidikan*, 2(4), 671–678.