

Penerapan Metode K-Nearest Neighbor Untuk Pemilihan Rekomendasi Film

Wisnu Setyawan¹, Rahmat Hidayat², Kurnia Nurcahya³, Farhan Hidayat⁴

Ilmu Komputer, Universitas Putra Bangsa

setyawanwisnu1903@gmail.com , ayat150190@gmail.com

Page | 1

ABSTRAK

Di era digital, semakin banyak orang yang menikmati film sebagai salah satu bentuk hiburan. Hal ini membuat masyarakat semakin sulit menemukan film yang sesuai dengan kesukaannya. Oleh karena itu, diperlukan sistem rekomendasi film yang dapat membantu masyarakat menemukan film yang sesuai dengan kesukaannya. Penelitian ini menerapkan metode K-Nearest Neighbor (KNN) untuk membangun sistem rekomendasi film. KNN merupakan metode klasifikasi yang bekerja dengan cara mencari data yang paling mirip dengan data baru berdasarkan jarak Euclidean. Dalam penelitian ini jarak antar data film diukur berdasarkan atribut film seperti genre, sutradara, aktor, dan rating. Data yang digunakan dalam penelitian ini adalah data film dari website IMDb. Datanya terdiri dari 8 data film dengan 5 atribut. Hasil pengujian algoritma K-NN berhasil diterapkan dengan kriteria seperti rating, review, tahun rilis, durasi, dan genre

Kata Kunci: Kata Kunci 1; K-Nearest Neighbor, System Rekomendasi, Movie

ABSTRACT

In the digital era, more and more people enjoy movies as a form of entertainment. This makes it increasingly difficult for people to find movies that match their preferences. Therefore, a movie recommendation system is needed to help people find movies that match their preferences. This study applies the K-Nearest Neighbor (KNN) method to build a movie recommendation system. KNN is a classification method that works by finding the most similar data to the new data based on the Euclidean distance. In this study, the distance between film data is measured based on film attributes, such as genre, director, actor, and rating. The data used in this study is movie data from the IMDb website. The data consists of 8 movie data with 5 attributes. The results of the testing the K-NN algorithm was successfully applied with criteria such as rating, reviews, year of release, duration, and genre

Keywords: K-Nearest Neighbor, recommendation system, movie

PENDAHULUAN

Sistem rekomendasi film dan TV series telah menjadi bagian penting dari industri hiburan digital. Dengan pertumbuhan konten yang begitu cepat, pengguna sering kali menghadapi kesulitan dalam menemukan konten yang sesuai dengan preferensi mereka. Dalam konteks ini, penerapan algoritma K-Nearest Neighbor (K-NN) dapat menjadi solusi yang efektif untuk menyajikan rekomendasi yang lebih personal dan sesuai dengan preferensi masing-masing pengguna.

Algoritma K-NN bekerja berdasarkan prinsip kesamaan, di mana ia mencari objek yang mirip berdasarkan jarak antar objek dalam ruang fitur. Dalam hal ini, fitur-fitur yang digunakan dapat mencakup genre, aktor, sutradara, atau ulasan pengguna. Dengan pendekatan ini, algoritma K-

NN dapat membantu dalam menyajikan rekomendasi film dan TV series yang lebih sesuai dengan preferensi pengguna (Raharjo, 2021).

Dalam konteks rekomendasi film dan TV series, penerapan algoritma K-NN melibatkan langkah-langkah seperti pemilihan berdasarkan kesamaan, pengklasifikasian berbasis data, dan penggunaan data pelatihan. Langkah pertama melibatkan pencarian film atau series lain yang mirip berdasarkan fitur-fitur tertentu, sedangkan langkah kedua melibatkan pengelompokan film dan TV series berdasarkan preferensi pengguna. Langkah terakhir melibatkan pemanfaatan data pelatihan untuk membuat prediksi film atau series mana yang mungkin disukai oleh seorang pengguna berdasarkan kesamaannya dengan data pelatihan (Irfan, 2020).

Dengan pendekatan ini, algoritma K-NN dapat membantu dalam menyajikan rekomendasi film dan TV series yang lebih personal dan sesuai dengan preferensi masing-masing pengguna. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan sistem rekomendasi yang lebih baik di masa depan.

Dalam rangka mendukung penelitian ini, akan dilakukan pengumpulan data dari pengguna untuk menentukan preferensi mereka. Data ini akan dianalisis dan dioptimalkan menggunakan algoritma K-NN untuk menghasilkan rekomendasi yang sesuai dengan preferensi pengguna. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan sistem rekomendasi yang lebih baik di masa depan.

2. Metodologi Penelitian

Sistem pendukung keputusan (SPK) adalah sistem informasi yang digunakan untuk membantu pengambilan keputusan dengan menggunakan data, model matematika, dan teknik analisis tertentu (Hutahean et al., 2023). Menurut Budi Sutedjo Darma Oetomo, 2002, SPK mulai dikembangkan pada tahun 1960-an dan 1970-an sebagai respons terhadap kebutuhan pengambilan keputusan yang semakin kompleks dan membutuhkan dukungan teknologi informasi (Rahmawati, 2019).

Karakteristik dari sistem pendukung keputusan adalah sebagai berikut:

1. Sistem pendukung keputusan memberikan dukungan bagi pengambil keputusan pada situasi semi terstruktur.
2. Dukungan untuk semua level manajerial, dari eksekutif puncak sampai manajer lini.
3. Dukungan untuk individu dan kelompok.
4. Dukungan untuk keputusan independen dan sekuensial.
5. Dukungan di semua fase proses pengambilan keputusan, yaitu intelligence, design, choice, dan implementation
6. Dukungan di berbagai proses dan gaya yang berbeda-beda.
7. Adaptivitas sepanjang waktu.
8. Mudah untuk digunakan user.
9. Peningkatan efektivitas dari pengambilan keputusan daripada efisiensi.

10. Kontrol penuh oleh pengambil terhadap semua langkah proses pengambilan keputusan.
11. Pengguna akhir bisa mengembangkan dan memodifikasi sendiri system (Sistem et al., 2013).

Dalam konteks pengambilan keputusan dalam industri hiburan digital, seperti rekomendasi film dan TV series, SPK dapat membantu dalam menyajikan rekomendasi yang lebih personal dan sesuai dengan preferensi masing-masing pengguna. Dengan memanfaatkan data historis tentang preferensi pengguna, SPK dapat mengidentifikasi pola-pola dalam preferensi tersebut dan menggunakan informasi tersebut untuk merekomendasikan film dan TV series yang sesuai (Rahmawati, 2019)

Dalam rangka mendukung pengembangan sistem rekomendasi yang lebih baik di masa depan, SPK dapat menjadi solusi yang efektif untuk menyajikan rekomendasi yang lebih personal dan sesuai dengan preferensi masing-masing pengguna. Dengan memanfaatkan teknologi informasi dan data historis tentang preferensi pengguna, SPK dapat membantu dalam meningkatkan akurasi rekomendasi film dan TV series kepada pengguna (Hutahean et al., 2023).

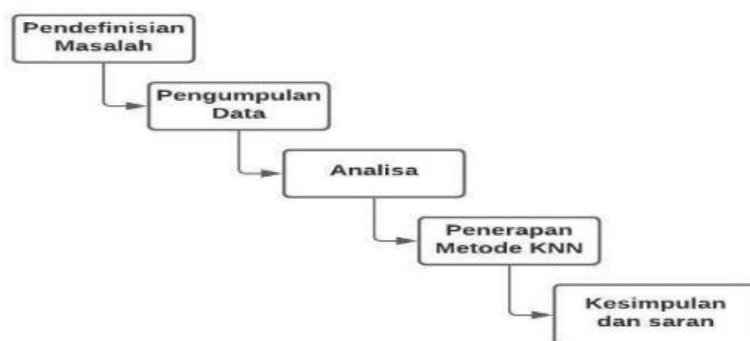
Algoritma K-Nearest Neighbor (K-NN) adalah metode yang digunakan untuk melakukan klasifikasi terhadap objek yang berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek yang ingin diklasifikasi (Asahar et al., 2016). Algoritma ini bekerja dengan mencari objek yang mirip berdasarkan jarak antar objek dalam ruang fitur. Dalam konteks rekomendasi film dan TV series, K-NN dapat digunakan untuk mengelompokkan film dan TV series berdasarkan preferensi pengguna (ADMINLP2M, 2023).

Kelebihan K-NN meliputi Sederhana dan mudah diimplementasikan, Tidak memerlukan membangun model, membuat beberapa parameter, atau membuat asumsi tambahan, Serbaguna dan dapat digunakan untuk membuat klasifikasi, regresi, dan pencarian data (Bunga, 2022). Namun, K-NN juga memiliki beberapa kekurangan, seperti Kemungkinan kompleks dalam penentuan nilai K yang mungkin, Biaya komputasinya tinggi karena harus menghitung jarak antara titik data dengan semua sampel K (Bunga, 2022).

Dalam konteks rekomendasi film dan TV series, algoritma K-NN dapat membantu dalam menyajikan rekomendasi yang lebih personal dan sesuai dengan preferensi pengguna. Dengan memanfaatkan data historis tentang preferensi pengguna, algoritma K-NN dapat mengidentifikasi pola-pola dalam preferensi tersebut dan menggunakan informasi tersebut untuk merekomendasikan film dan TV series yang sesuai (ADMINLP2M, 2023).

Alur Penelitian

Penelitian dimulai dengan penjelasan masalah dan penerapan teknik. Penelitian ini melakukan hal-hal berikut: pendefinisian masalah, pengumpulan data, analisis, penerapan metode KNN, dan akhirnya, kesimpulan, dan rekomendasi. seperti yang ditunjukkan pada gambar berikut:



Gambar 1. Alur Penelitian

1. Pendefinisian Masalah. Merupakan tahapan mendefinisikan serta menjabarkan permasalahan yang telah diidentifikasi.
2. Pengumpulan Data. Setelah melakukan pendefinisian masalah, maka dilakukan pengumpulan data Film dan Tv Series dengan melihat rating, review, genre, umur dan judul.
3. Analisa, Pada tahapan ini dilakukannya analisis dari data-data yang telah dikumpulkan serta membuat data ujiberdasarkan pengalaman pengguna dalam memilih Film dan Tv series.
4. Penerapan Metode K-NN, Pada tahapan ini setelah melakukan analisa data, maka dilakukannya penerapan metode K-NN terhadap data-data yang telah dikumpulkan. Data tersebut akan diuji dan menghasilkan jarak yang menjadi parameter utama untuk merekomendasikan aplikasi Film dan Tv series terbaik.
5. Kesimpulan dan Saran. Tahapan ini adalah tahapan akhir yang berisi kesimpulan hasil evaluasi seluruh tahapan penelitian yang sudah dilakukan dalam penelitian penentuan Film dan Tv series terbaik dan tahap untuk memberikan masukan dalam penelitian yang akan dilakukan di masa datang.

Subjek dan Objek Penelitian:

Subjek penelitian ini Website/Bioskop, dan objek penelitian adalah Film dan tv series.

Alat dan Bahan Penelitian:

Bahan penelitian ini adalah data Film dan tv series yang dievaluasi berdasarkan rating, review, genre, umur dan judul.

yang dapat ditemukan di Website atau situs lainnya. K-Nearest Neighbor adalah alat atau instrumen yang digunakan dalam penelitian ini.

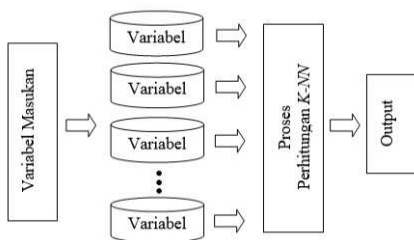
3. Hasil dan Pembahasan

Tahapan pertama dimulai dari inputan user yang terdiri dari variabel-variabel yang di berikan sistem. Setelah user menjawab sesuai kriteria sistem, proses selanjutnya merupakan evaluasi data uji dengan data sampel calon user menggunakan perhitungan K-NN. Dalam hal ini yang sampel data sudah ada . Tahapan dari metode K-NearestNeighbor sebagai berikut:

1. Menentukan jumlah parameter k ($k=3$)
2. Hitung jarak antara data yang akan di evaluasi dengan semua data sample.
3. Urutkan jarak yang terbentuk.
4. Pilih tiga sampel terdekat.
5. Pilih nama film yang ada dalam tiga sampel tersebut
6. Tampilkan output

Tahapan pertama adalah penentuan parameter k, pada peneitian ini memilih $k=3$, yang berarti 3 sampel film terdekan akan dijadikan rekomendasi.

setelah semua kriteria tersebut lengkap, barulah dihitung perihitungan jarak antara data uji dengan data sampel film yang sudah ada. Dalam perhitungan jarak nanti akan digunakan rumus Euclidean Distance. Penerapan K-NN bisa dilihat pada gambar berikut ini :



Gambar 2. Penerapan K-NN [23]

Setelah dilakukan identifikasi maka diperoleh data sampel. Sesuai dengan gambar 3 berikut:

Film 1	
Rating	8.3
Review	37.000
Tahun Rilis	2023
Durasi	124 Menit
Genre	Action, Adventure, Drama
Godzilla Minus One	

Film 5	
Rating	7.7
Review	1.500
Tahun Rilis	2013
Durasi	146.4 Menit
Genre	Drama, Romance
Tenggelamnya Kapal Van der Wijck	

Film 2	
Rating	8.6
Review	332.000
Tahun Rilis	2023
Durasi	132 Menit
Genre	Action, Adventure, Animation
Spider-Man : Across the Spider-Verse	

Film 6	
Rating	7.6
Review	2.200
Tahun Rilis	2012
Durasi	120 Menit
Genre	Biography, Drama, Romance
Habibie & Ainun	

Film 3	
Rating	8.4
Review	1.200.000
Tahun Rilis	2019
Durasi	181 Menit
Genre	Action, Adventure, Drama
Avengers: Endgame	

Film 7	
Rating	7.0
Review	4.000
Tahun Rilis	2015
Durasi	130 Menit
Genre	Action, Crime, Drama, Thriller
Inside Men	

Film 4	
Rating	7.9
Review	432.000
Tahun Rilis	2021
Durasi	242 Menit
Genre	Action, Adventure, Fantasy, Sci-Fi
Zack Snyder's Justice League	

Film 8	
Rating	6.5
Review	1.200
Tahun Rilis	2020
Durasi	90 Menit
Genre	Crime, Drama, Thriller
Innocence	

Gambar 3. Data sampel

Gambar 3. Merupakan beberapa sampel data film berupa rating, review, tahun rilis, durasi, dan genre. Yang didapatkan pada website imdb. Dari sekian sampel yang didapatkan data sampel akan dihitung dengan data uji. Setelah mendapatkan data sampel kemudian dibuatlah data uji pada film X. Seperti pada tabel 1 berikut ini :

Table 1. data uji film

Film X	
Rating	8.0
Review	100.000
Tahun Rilis	2020
Durasi	125 Menit
Genre	Action, Romance
?	

Tabel 2 merupakan proses perhitungan jarak data sampel dengan data uji film dilakukan dengan mencocokkan variabel atau kriteria yang diinputkan user dengan data sampel film yang ada. Berikut merupakan rumus perhitungan:

$$d_i = \sqrt{\sum(x_{2i} - x_{1i})^2}$$

Keterangan:

X1 = Sampe data

Film 1		Film x	
Rating	8.3	Rating	8.0
Review	37.000	Review	100.000
Tahun Rilis	2023	Tahun Rilis	2020
Durasi	124 Menit	Durasi	125 Menit
Genre	Action, Adventure, Drama	Genre	Action, Romance
Godzilla Minus One		?	

X2 = Data Uji

i = Variabel Data

d = Jarak

p = Dimensi Data

Uji coba data testing dengan data sampel film pertama yaitu data film 1 dengan data film X, seperti Tabel 2 berikut :

Table 2. data sampel dan data uji

Perhitungan jarak Encledian antara data uji Film X dengan data sampel Film 1 adalah

$$d_1 = \sqrt{(0,3)^2 + (63000)^2 + (1)^2 + (1)^2 + (1)^2} = 63000$$

Tabel 3. Jarak antara data sampel dengan data uji

Data uji	Data sampel	Jarak
Film X	Film 1	63000
	Film 2	232000
	Film 3	1100000
	Film 4	332000
	Film 5	98500
	Film 6	97800
	Film 7	96000
	Film 8	98800

Sesuai pada tabel 4 yaitu merupakan jarak antara data sampel dengan data uji. Data sampel mempunyai 7 Film, masing-masing dari film tersebut mempunyai jarak. Setelah jarak data sampel Film dihitung maka didapatkan urutan jarak terdekat seperti tabel 4 dibawah ini:

Tabel 4. Hasil perhitungan jarak

Data uji	Data sampel	Jarak
Film X	Film 1	63000
	Film 7	96000
	Film 6	97800
	Film 5	98500
	Film 8	98800
	Film 2	232000
	Film 4	332000
Film 3	1100000	

Sesuai pada tabel x merupakan urutan jarak terdekat data uji dengan data sampel yang dimulai dari Film 1 yang mempunyai jarak yaitu 63000 hingga Film 3 yang mempunyai jarak yaitu 1100000. Setelah diurutkan dipilih sebanyak ($k=3$) sampel Film yang mempunyai jarak terdekat dengan data uji yaitu: Film 1, Film 7, dan Film 6. Film tersebut sangat menjadi rekomendasi untuk para user untuk ditonton oleh user.

4. Kesimpulan

Berdasarkan hasil dari penelitian ini, menunjukkan penelitian berhasil menerapkan metode K-NN untuk rekomendasi pemilihan Film terbaik pada imdb dengan menerapkan kriteria seperti rating, review, tahun rilis, durasi, serta genre. Saran untuk penelitian lebih lanjut dapat dikombinasikan dengan metode yang lainnya. Tidak hanya menggunakan metode K-NN, melainkan seperti algoritma KMeans, Taxonomic Matcher dan sebagainya. Metode pada penelitian ini yaitu menggunakan metode kuantitatif yang mana bahan penelitian atau data penelitian berupa data angka atau statistic. Tahapan-tahapan dari metode K-Nearest pada penelitian ini yaitu: menentukan jumlah parameter k ($k=3$), menghitung jarak antara data yang akan di evaluasi dengan semua data sample, mengurutkan jarak yang terbentuk, memilih tiga sampel terdekat, pilih nama film yang ada dalam tiga sampel tersebut dan menampilkan output. Tujuan dari penelitian ini menghasilkan rekomendasi film yang sesuai dengan keinginan user. Kriteria-kriteria tersebut terdiri dari Rating, Review, Tahun Rilis, Durasi, dan Genre dari film.

References

- Hutahean, J., Nugroho, Fi., Abdullah, D., & Aini, Q. (2023). *FullBook Sistem Pendukung Keputusan*.
- Irfan, R. (2020). *ANALISIS PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR DENGAN ALGORITMA SUPPORT VECTOR MACHINE PADA PENGLASIFIKASIAN HADITS SHAHIH MUSLIM*.
- Raharjo, B. (2021). *P Y YAYASAN PRIMA AGUS TEKNIK YAYASAN PRIMA AGUS TEKNIK YAYASAN PRIMA AGUS TEKNIK Pembelajaran Mesin (Machine Learning)*.
- Rahmawati, N. (2019). *SISTEM PENDUKUNG KEPUTUSAN PEMILIHAN GURU TERBAIK MENGGUNAKAN METODE SIMPLE MULTI ATTRIBUTE RATING TECHNIQUE (SMART) STUDY KASUS SMA MUHAMMADIYAH 1 GRESIK*.
- Sistem, J., Fakultas, I., Komputer, I., Sriwijaya, U., Raya, J., Km, P.-P., Ogan Ilir, I., Rohayani, H., Informatika, T., Dinamika, S., & Jambi, B. (2013). Analisis Sistem Pendukung Keputusan Dalam Memilih Program Studi Menggunakan Metode Logika Fuzzy. *Jurnal Sistem Informasi (JSI)*, 5(1), 530–539. <http://ejournal.unsri.ac.id/index.php/jsi/index>
- T, Asahar, J., Yanosmas, D., Anggriani K. (2016). IMPLEMENTASI METODE K-NEAREST NEIGHBOR (KNN) DAN SIMPLE ADDITIVE WEIGHTING (SAW) DALAM PENGAMBILAN KEPUTUSAN SELEKSI PENERIMAAN ANGGOTA PASKIBRAKA. *Jurnal Pseudocode*, 3(2) , 2355-5920
- ADMINLP2M. (2023). Algoritma K-Nearest Neighbors (KNN) – Pengertian dan Penerapan. Di ambil dari <https://lp2m.uma.ac.id/2023/02/16/algoritma-k-nearest-neighbors-knn-pengertian-dan-penerapan/>
- Bunga Dea Laraswati. (2022). Mengenal K-nearest Neighbor dan Pengaplikasiannya. Di ambil dari <https://blog.algorit.ma/k-nearest-neighbor/>