

## PENERAPAN ALGORITMA NAIVE BAYES PADA SISTEM REKOMENDASI FILM BERBASIS KONTEN

Deprinico Riyadsyah <sup>1\*</sup>, Erika Oktaviana Az Zahra <sup>2</sup>, Pipin Anjarwati <sup>3</sup>, Vincensius Gilang Pramudito <sup>4</sup>, Yohanes Indra Kristiawan <sup>5</sup>, Sofyan Nur Rohman <sup>6</sup>

<sup>1,2,3,4,5,6</sup>Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Duta Bangsa Surakarta, Surakarta, Indonesia

Email: <sup>1\*</sup>[210103129@mhs.udb.ac.id](mailto:210103129@mhs.udb.ac.id), <sup>2</sup>[210103130@mhs.udb.ac.id](mailto:210103130@mhs.udb.ac.id), <sup>3</sup>[210103140@mhs.udb.ac.id](mailto:210103140@mhs.udb.ac.id), <sup>4</sup>[210103144@mhs.udb.ac.id](mailto:210103144@mhs.udb.ac.id), <sup>5</sup>[210103145@mhs.udb.ac.id](mailto:210103145@mhs.udb.ac.id), <sup>6</sup>[210103171@mhs.udb.ac.id](mailto:210103171@mhs.udb.ac.id)

### Abstract (English)

Film recommendation systems are a crucial part of the digital entertainment industry to assist users in finding content according to their preferences. This research discusses the implementation of the Naive Bayes algorithm in the content-based filtering method to develop a film recommendation system. Film data are extracted from various sources, including metadata such as genre, director, and actors. This system allows users to search for films based on desired genres and predict films that may be liked based on their preference history. Experimental results show that the Naive Bayes method is capable of generating relevant recommendations and significantly enhancing the user experience.

### Article History

Submitted: 2 Juli 2024  
Accepted: 8 Juli 2024  
Published: 9 Juli 2024

### Key Words

Naive Bayes; Recommendation System; Movie; Content-Based Filtering; Genre

### Abstrak (Indonesia)

Sistem rekomendasi film merupakan bagian krusial dalam industri hiburan digital untuk membantu pengguna menemukan konten sesuai preferensi mereka. Penelitian ini membahas penerapan algoritma Naive Bayes pada metode *content-based filtering* untuk membangun sistem rekomendasi film. Data film diekstraksi dari berbagai sumber, termasuk metadata seperti genre, sutradara, dan aktor. Sistem ini memungkinkan pengguna mencari film berdasarkan genre yang diinginkan dan memprediksi film yang mungkin disukai berdasarkan histori preferensi mereka. Hasil eksperimen menunjukkan bahwa metode Naive Bayes mampu menghasilkan rekomendasi yang relevan dan meningkatkan pengalaman pengguna secara signifikan.

### Sejarah Artikel

Submitted: 2 Juli 2024  
Accepted: 8 Juli 2024  
Published: 9 Juli 2024

### Kata Kunci

Naive Bayes; Sistem Rekomendasi; Film; Content-Based Filtering; Genre

## PENDAHULUAN

Dalam era yang terus maju dalam dunia digital, volume data yang diperoleh di berbagai platform online menghadirkan tantangan bagi pengguna untuk menemukan konten yang relevan dan bermakna (Mohammad Amir Fanani, 2024). Sistem rekomendasi film telah menjadi solusi populer belakangan ini untuk membantu pengguna menemukan film sesuai dengan preferensi mereka (Lestari & Ramdhani, 2023). Namun, dalam melakukannya, pengguna sering menghadapi kesulitan dalam menentukan pilihan film yang tepat.

Untuk mengatasi tantangan ini, diperlukan pengembangan sistem rekomendasi yang dapat membantu pengguna dalam menentukan pilihan film yang tepat (Rosetya Wardhana & Kembang Hapsari, 2023). Sistem rekomendasi adalah alat yang berguna dalam menyaring dan mengidentifikasi item seperti produk, layanan, atau informasi yang berpotensi besar untuk dipilih, dibeli, atau digunakan oleh pengguna (Arfisko, 2022). Di antara berbagai metode yang ada, *content-based filtering* menonjol sebagai pendekatan yang fokus pada atribut, karakteristik, atau fitur yang dimiliki oleh sebuah item (Putri dkk., 2022). Metode ini memungkinkan pengguna untuk mencari film berdasarkan genre yang diinginkan dan memprediksi film yang mungkin disukai berdasarkan histori preferensi mereka, serta menggunakan metode Naive Bayes dalam *content-based filtering* untuk menghitung *similarity*.

Selain itu, *content-based filtering* juga mempertimbangkan perilaku pengguna dari masa lalu yang kemudian diidentifikasi pola perilakunya untuk merekomendasikan barang yang

sesuai dengan pola perilaku tersebut (Fajriansyah dkk., 2021). Namun, metode ini tidak lepas dari tantangan dalam mengelola dan menganalisis data secara efektif untuk menghasilkan rekomendasi yang relevan dan berkualitas.

Metode Naive Bayes, yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, menawarkan pendekatan yang menarik dalam konteks sistem rekomendasi film (Yulia Hayuningtyas, 2019). Metode ini menggunakan teori probabilitas dan statistik untuk menentukan kemungkinan suatu kejadian berdasarkan informasi yang tersedia. Dalam konteks sistem rekomendasi film, Naive Bayes dapat digunakan untuk menghitung probabilitas bahwa seorang pengguna akan menyukai suatu film berdasarkan atribut-atribut yang dianalisis, seperti genre favorit atau aktor yang disukai. Dengan mengintegrasikan Naive Bayes dalam *content-based filtering*, diharapkan sistem rekomendasi ini dapat memberikan hasil yang lebih personal dan relevan bagi pengguna.

Penelitian ini bertujuan untuk mengembangkan serta menerapkan sistem rekomendasi film berbasis *content-based* menggunakan metode Naive Bayes, dengan fokus pada peningkatan akurasi dan relevansi rekomendasi film bagi pengguna. Dengan menerapkan pendekatan analisis data dan perhitungan similarity yang mendasari Naive Bayes, diharapkan penelitian ini dapat memberikan kontribusi penting dalam pengembangan sistem rekomendasi film yang lebih efektif dan dapat diandalkan, sehingga membantu pengguna menemukan genre film yang mereka sukai.

## METODE

### Studi Literatur

Studi literatur ini dilakukan untuk menelaah dan menganalisis informasi terbaru mengenai penerapan sistem rekomendasi film menggunakan metode Naive Bayes berbasis *content-based*. Pencarian literatur dilakukan melalui berbagai sumber akademis dan teknis, termasuk jurnal ilmiah, konferensi, dan buku teks terkait. Fokus utama dari studi pustaka ini adalah untuk memahami konsep dasar tentang metode Naive Bayes dalam konteks sistem rekomendasi, serta aplikasi *content-based filtering* dalam pengenalan dan prediksi preferensi pengguna terhadap konten film. Selain itu, studi ini juga memperdalam pemahaman tentang teknik-teknik penghitungan similarity antar genre film, yang merupakan elemen kunci dalam penentuan rekomendasi yang relevan dan personal bagi pengguna. Informasi yang diperoleh dari literatur ini akan menjadi dasar yang kuat untuk merancang dan mengimplementasikan sistem rekomendasi yang efektif dan akurat.

### Pengembangan Sistem

Penelitian ini mengadopsi metode pengembangan *Rapid Application Development* (RAD). Metode RAD dikenal dengan siklus pengembangan sistem yang singkat dan cepat, dimana dapat mengurangi waktu normal yang dibutuhkan untuk pembangunan perangkat lunak (Arda dkk., 2023). Tahapan pengembangan RAD dalam penelitian ini terfokus pada tahap pemodelan, termasuk pemodelan bisnis, pemodelan data, dan pemodelan proses.

- a. Pemodelan bisnis dilakukan dengan mengumpulkan data secara komprehensif untuk memahami kebutuhan sistem, alur kerja, peran yang terlibat, dan proses terkait. Tahap ini melibatkan pengumpulan informasi mengenai kebutuhan sistem yang akan direpresentasikan melalui diagram alur kerja.
- b. Pemodelan data dilakukan dengan menerapkan metode *content-based recommendation* untuk membangun model data berdasarkan informasi yang diperoleh dari tahap bisnis.

Alat bantu model digunakan untuk menggambarkan secara sistematis data film beserta atribut-atribut yang relevan.

- c. Pemodelan proses melibatkan penjelasan rinci tentang proses bisnis dan pemodelan data yang telah dikembangkan. *Unified Modeling Language* (UML) digunakan untuk memvisualisasikan dengan jelas antarmuka pengguna sistem, termasuk desain input dan output yang disesuaikan dengan kebutuhan penggunaan sistem.

Dengan pendekatan ini, diharapkan sistem rekomendasi film yang dikembangkan tidak hanya memberikan pengalaman pengguna yang lebih interaktif, tetapi juga meningkatkan akurasi dalam menyarankan film-film yang cocok dengan preferensi pengguna melalui analisis mendalam terhadap data dan proses yang terlibat.

## ANALISA DAN PEMBAHASAN

### Pemodelan Bisnis

Dalam industri hiburan, jumlah film yang tersedia bagi konsumen meningkat setiap hari. Menyediakan rekomendasi yang akurat dapat meningkatkan kepuasan pengguna dan waktu yang mereka habiskan di platform, sehingga dapat meningkatkan pendapatan. Model bisnis ini bertujuan menjelaskan bagaimana sistem rekomendasi film berbasis Naive Bayes dapat diimplementasikan dan dimonetisasi.

- a. Nilai Proposisi
 

Memberikan rekomendasi film yang dipersonalisasi berdasarkan preferensi pengguna, meningkatkan kepuasan dan keterlibatan mereka. Ini dapat meningkatkan retensi pengguna dan waktu menonton, yang dapat diterjemahkan menjadi peningkatan pendapatan melalui langganan dan iklan.
- b. Segmen Pelanggan
  1. Pengguna platform streaming: individu yang mencari rekomendasi film sesuai dengan preferensi pribadi mereka.
  2. Penyedia layanan streaming: perusahaan yang ingin meningkatkan pengalaman pengguna mereka dengan menawarkan rekomendasi film yang relevan.
  3. Pembuat konten dan studio film: pihak yang ingin meningkatkan visibilitas konten mereka.
- c. Sumber Pendapatan
  1. Langganan Premium: Pengguna dapat membayar langganan bulanan untuk fitur rekomendasi lanjutan.
  2. Iklan: Menampilkan iklan yang dipersonalisasi berdasarkan preferensi film pengguna.
  3. Kemitraan dengan Studio: Studio film dapat membayar untuk peningkatan visibilitas konten mereka melalui sistem rekomendasi.

### Pemodelan Data

Penelitian ini menggunakan data film yang diambil sampelnya dari film-film Hollywood. Sebanyak 25 sampel data film dengan 3 atribut digunakan untuk membuat pemodelan *Content-Based Filtering*. Berikut adalah tabel yang memuat informasi mengenai data film yang digunakan dalam analisis ini.

Tabel 1. Data film

Id Film	Judul	Genre
1	Toy Story	Animation, Comedy, Fantasy, Children, Adventure

2	Jumanji	Children, Adventure, Fantasy
3	Grumpier Old Men	Romance, Comedy
4	Waiting to Exhale	Drama, Comedy, Romance
5	Father of the Bride Part II	Comedy
6	Heat	Thriller, Action, Crime
7	Sabrina	Romance, Comedy
8	Tom and Huck	Children, Adventure
9	Sudden Death	Action
10	GoldenEye	Thriller, Action, Adventure
11	American President, The	Comedy, Romance, Drama
12	Dracula: Dead and Loving It	Horror, Comedy
13	Balto	Adventure, Children, Animation
14	Nixon	Drama
15	Cutthroat Island	Action, Romance, Adventure
16	Casino	Drama, Crime
17	Sense and Sensibility	Romance, Drama
18	Four Rooms	Comedy
19	Ace Ventura: When Nature	Comedy
20	Calls	Comedy, Action, Drama, Crime, Thriller
21	Money Train	Thriller, Crime, Comedy
22	Get Shorty	Horror, Drama, Mystery, Crime, Thriller
23	Copycat	Action, Thriller, Crime
24	Assassins	Sci-Fi, Drama
25	Powder	Drama, Romance
	Leaving Las Vegas	

Pemodelan proses dalam *content-based filtering* menggunakan metode Naive Bayes untuk menghitung kesamaan (similarity). Dalam konteks ini, perhitungan similarity dari data yang diperoleh menggunakan Koefisien Kesamaan Dice dengan menghitung jumlah K-Gram. K-Gram adalah serangkaian term dengan panjang K yang digunakan untuk mengambil potongan karakter sebanyak K dari sebuah kata secara berkelanjutan dari teks sumber hingga akhir dokumen (Mondi & Wijayanto, 2019). Rumus similarity  $S$  dapat dihitung dengan menggunakan persamaan berikut:

$$S = \frac{K \cdot C}{(A + B)}$$

Dengan keterangan sebagai berikut:

$S$  = nilai similarity antara dua teks.

$K$  = jumlah bigrams yang digunakan (dalam konteks ini,  $K=2$ ).

$C$  = jumlah k-gram yang sama antara teks 1 dan teks 2.

$A$  = jumlah total k-gram dalam teks 1.

$B$  = jumlah total k-gram dalam teks 2.

Setelah memahami rumus di atas, selanjutnya akan dihitung similarity berdasarkan genre yang dicari pengguna yaitu 'action comedy'. Data ini terdiri dari 25 judul film yang diklasifikasikan ke dalam 13 jenis genre film yang meliputi Action, Adventure, Drama, Sci-Fi,

Animation, Horror, Children, Comedy, Crime, Fantasy, Mystery, Romance, dan Thriller, sebagaimana tercantum dalam Tabel 1.

<b>Film 1</b> $Sim = \frac{2.1}{2+5} = \frac{2}{7} = 0,285$	<b>Film 2</b> $Sim = \frac{2.0}{2+3} = \frac{0}{5} = 0$	<b>Film 3</b> $Sim = \frac{2.1}{2+2} = \frac{2}{4} = 0,5$
<b>Film 4</b> $Sim = \frac{2.1}{2+3} = \frac{2}{5} = 0,4$	<b>Film 5</b> $Sim = \frac{2.1}{2+1} = \frac{2}{3} = 0,666$	<b>Film 6</b> $Sim = \frac{2.1}{2+3} = \frac{2}{5} = 0,4$
<b>Film 7</b> $Sim = \frac{2.1}{2+2} = \frac{2}{4} = 0,5$	<b>Film 8</b> $Sim = \frac{2.0}{2+2} = \frac{0}{4} = 0$	<b>Film 9</b> $Sim = \frac{2.1}{2+1} = \frac{2}{3} = 0,666$
<b>Film 10</b> $Sim = \frac{2.1}{2+3} = \frac{2}{5} = 0,4$	<b>Film 11</b> $Sim = \frac{2.1}{2+3} = \frac{2}{5} = 0,4$	<b>Film 12</b> $Sim = \frac{2.1}{2+2} = \frac{2}{4} = 0,5$
<b>Film 13</b> $Sim = \frac{2.0}{2+3} = \frac{0}{5} = 0$	<b>Film 14</b> $Sim = \frac{2.0}{2+1} = \frac{0}{3} = 0$	<b>Film 15</b> $Sim = \frac{2.1}{2+3} = \frac{2}{5} = 0,4$
<b>Film 16</b> $Sim = \frac{2.0}{2+2} = \frac{0}{4} = 0$	<b>Film 17</b> $Sim = \frac{2.0}{2+2} = \frac{0}{4} = 0$	<b>Film 18</b> $Sim = \frac{2.1}{2+1} = \frac{2}{3} = 0,666$
<b>Film 19</b> $Sim = \frac{2.1}{2+1} = \frac{2}{3} = 0,666$	<b>Film 20</b> $Sim = \frac{2.2}{2+5} = \frac{4}{7} = 0,571$	<b>Film 21</b> $Sim = \frac{2.1}{2+3} = \frac{2}{5} = 0,4$
<b>Film 22</b> $Sim = \frac{2.0}{2+5} = \frac{0}{7} = 0$	<b>Film 23</b> $Sim = \frac{2.1}{2+3} = \frac{2}{5} = 0,4$	<b>Film 24</b> $Sim = \frac{2.0}{2+2} = \frac{0}{4} = 0$
<b>Film 25</b> $Sim = \frac{2.0}{2+2} = \frac{0}{4} = 0$		

Hasil urutan film rekomendasi ketika pengguna mencari judul film dengan genre 'action comedy' dengan empat film teratas, yaitu Father of the Bride Part II, Sudden Death, Four Rooms, dan Ace Ventura: When Nature Calls, karena memiliki nilai similarity tertinggi, yaitu 0,666. Di baris kedua, Money Train, Grumpier Old Men, Sabrina, dan Dracula: Dead and Loving It direkomendasikan dengan nilai similarities masing-masing 0,571 dan 0,5. Sementara itu, pada baris berikutnya, film-film seperti Waiting to Exhale, Heat, GoldenEye, dan American President, The direkomendasikan dengan nilai similarities sebesar 0,4. Pengurutan ini berdasarkan nilai similaritiesnya, semakin besar nilai similarities maka akan ditampilkan di paling atas.

### Pemodelan Proses

#### a. Pengumpulan Data

Dataset yang digunakan berasal dari kumpulan data film Hollywood, yang mencakup aktivitas penilaian dan penandaan teks bebas dari layanan rekomendasi film MovieLens. Terdapat 20.000.263 penilaian dan 465.564 aplikasi tag untuk 27.278 film.

Data ini dikumpulkan oleh 138.493 pengguna dari 9 Januari 1995 hingga 31 Maret 2015 dan dataset ini dibuat pada 17 Oktober 2016.

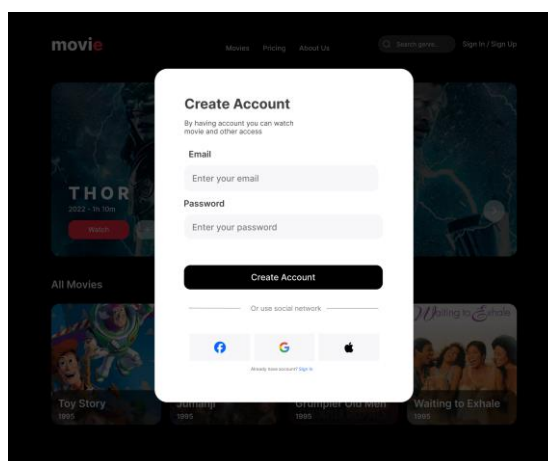
b. Pra-pemrosesan Data

Pra-pemrosesan data merupakan langkah penting untuk memastikan kualitas dan konsistensi data sebelum digunakan dalam model. Langkah pertama adalah menghapus data yang tidak lengkap atau duplikat untuk menjaga integritas dataset. Selanjutnya, dilakukan normalisasi data untuk memastikan semua fitur berada dalam skala yang sama. Kategori film yang berbentuk teks diubah menjadi format numerik menggunakan teknik *one-hot encoding*. Selain itu, fitur-fitur yang tidak relevan atau kurang signifikan dihapus untuk meningkatkan efisiensi model. Proses pra-pemrosesan ini memastikan bahwa data siap digunakan untuk melatih model Naive Bayes berbasis *content-based* dengan akurasi yang optimal.

c. Pelatihan Model Naive Bayes

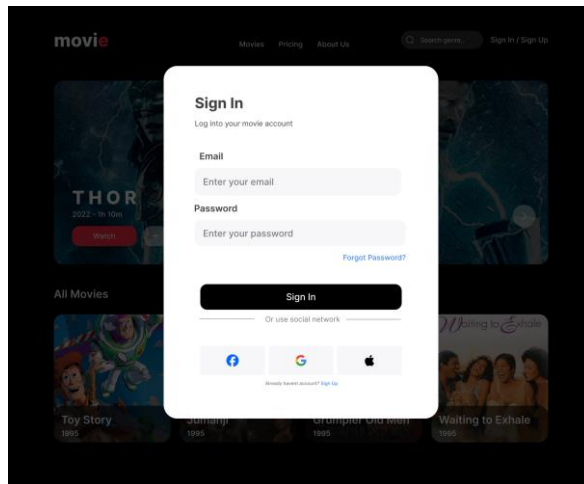
Model Naive Bayes dilatih untuk memungkinkan pengguna mencari film berdasarkan genre yang diinginkan dan memprediksi film yang kemungkinan disukai berdasarkan histori preferensi mereka. Proses ini melibatkan pembelajaran pola dari fitur-fitur seperti genre, rating, dan metadata lainnya untuk menyediakan rekomendasi yang relevan dan personal bagi pengguna.

d. *User-interface* Sistem Rekomendasi



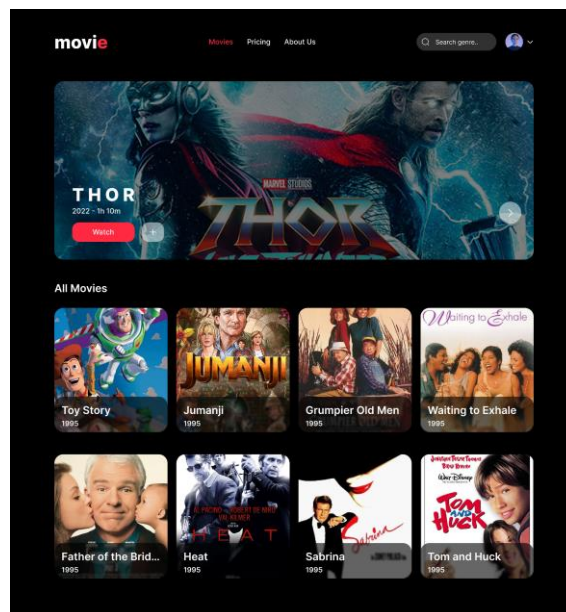
**Gambar 1.** Halaman Sign Up/Registrasi

Gambar 1 menampilkan halaman Registrasi/Sign Up untuk membuat akun, dimana pengguna diminta memasukkan alamat email dan kata sandi. Terdapat juga opsi untuk membuat akun melalui Facebook, Gmail, atau Apple account. Setelah membuat akun, pengguna dapat langsung masuk ke halaman Log In/Masuk (Gambar 2) masukkan alamat email dan kata sandi yang telah terdaftar sebelumnya. Halaman ini juga menyediakan pilihan Log In melalui akun Facebook, Gmail, atau Apple account. Jika pengguna lupa kata sandi, tersedia opsi untuk mengklik "forgot password?".



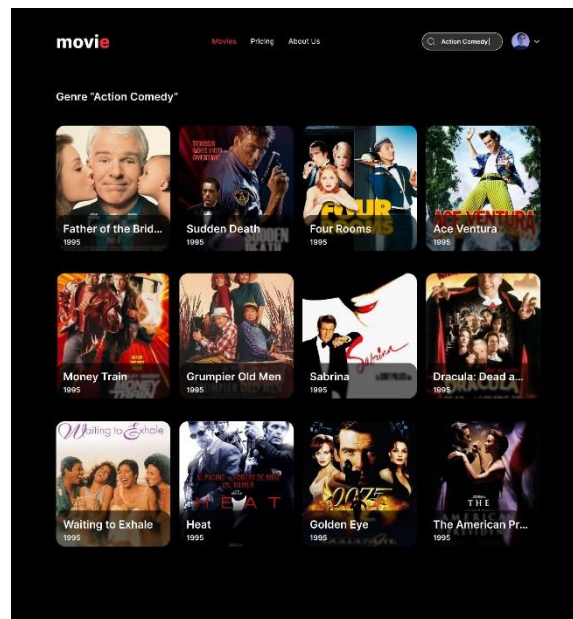
**Gambar 2.** Halaman Log In/Masuk

Gambar 3 menampilkan halaman utama/beranda, dimana pengguna dapat melihat daftar film terkini atau populer saat ini. Pengguna memiliki opsi untuk mengeksplorasi halaman ini untuk mencari film yang diminati. Terdapat navigasi untuk mencari kata kunci genre, sehingga pengguna dapat dengan mudah menemukan film berdasarkan preferensi mereka. Fitur-fitur tambahan seperti kategori film, ulasan pengguna, dan fitur pencarian juga tersedia untuk meningkatkan pengalaman eksplorasi pengguna.



**Gambar 3.** Halaman Utama/Beranda

Gambar 4 menampilkan halaman rekomendasi, dimana pengguna mencari film dengan genre action comedy. Film-film direkomendasikan berdasarkan tingkat kemiripan, dengan yang memiliki nilai similarity tertinggi ditampilkan terlebih dahulu.



Gambar 4. Halaman Rekomendasi

## KESIMPULAN

Studi ini mengimplementasikan sistem rekomendasi film menggunakan metode Naive Bayes berbasis konten. Percobaan dilakukan dengan mencari film genre 'action comedy' yang diminati pengguna. Beberapa film seperti *Father of the Bride Part II*, *Sudden Death*, *Four Rooms*, dan *Ace Ventura: When Nature Calls* memiliki similarity tertinggi sebesar 0,666. Dengan menggunakan teknik pengumpulan data dan pra-pemrosesan menggunakan *one-hot encoding* serta perhitungan similarity dengan Naive Bayes, sistem ini efektif dalam memberikan rekomendasi film yang sesuai dengan preferensi pengguna berdasarkan genre yang dicari.

## REFERENCES

- Arda, K., Editya, S., Khuwarismi, A., Robbuhu, I., & Surya Editya, A. (2023). Nusantara Computer and Design Review Game Edukasi Matematika untuk Anak Sekolah Dasar Menggunakan Metode Rapid Application Development (RAD). *NCDR*, 1(1), 1–8. <https://journal.unusida.ac.id/index.php/ncdr/>
- Arfisko, H. H. (2022). *Sistem Rekomendasi Film Menggunakan Metode Hybrid Collaborative Filtering Dan Content-Based Filtering*.
- Fajriansyah, M., Adikara, P. P., & Widodo, A. W. (2021). *Sistem Rekomendasi Film Menggunakan Content Based Filtering* (Vol. 5, Nomor 6). <http://j-ptiik.ub.ac.id>
- Lestari, S., & Ramdhani, M. M. (2023). SISTEM REKOMENDASI FILM MENGGUNAKAN METODE CONTENT-BASED FILTERING STUDI KASUS MATERI DATA MINING DI SMK IDN BOARDING SCHOOL. *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, 4(3), 1581–1587. <https://doi.org/10.35870/jimik.v4i3.381>
- Mohammad Amir Fanani. (2024). Sistem Rekomendasi Film Menggunakan Metode K-NN. *Jurnal ilmiah Sistem Informasi dan Ilmu Komputer*, 4(1), 178–185. <https://doi.org/10.55606/juisik.v4i1.760>

- Mondi, R. H., & Wijayanto, A. (2019). *RECOMMENDATION SYSTEM WITH CONTENT-BASED FILTERING METHOD FOR CULINARY TOURISM IN MANGAN APPLICATION*.
- Putri, D. A., Pramesti, D., I, D., & Santiyasa, W. (2022). Penerapan Metode Content-Based Filtering dalam Sistem Rekomendasi Video Game. Dalam *JNATIA* (Vol. 1, Nomor 1).
- Rosetya Wardhana, S., & Kembang Hapsari, R. (2023). Sistem Rekomendasi Film dengan Menggunakan Pendekatan Collaborative Filtering Berdasarkan Class. *Prosiding Seminar Implementasi Teknologi Informasi dan Komunikasi*, 2(1).  
<https://doi.org/10.31284/p.semtik.2023-1.4153>
- Yulia Hayuningtyas, R. (2019). Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita. *JURNAL INFORMATIKA*, 6(1), 18–22.  
<http://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/4685>