

EVALUASI PERFORMA ALGORITMA SIMILARITY PADA DATASET FILM INDONESIA MENGGUNAKAN METRIK PRECISION DAN RECALL

Mubtaghi Ridho Robbi, Apriade Voutama

Sistem Informasi, Universitas Singaperbangsa Karawang
Jalan HS. Ronggo Waluyo, Telukjambe Timur, Karawang, Indonesia
2310631250065@student.unsika.ac.id

ABSTRAK

Pertumbuhan pesat industri film Indonesia memicu fenomena *information overload*, yang menyulitkan pengguna menemukan konten tontonan yang relevan di antara ribuan judul yang tersedia. Permasalahan utama dalam mengatasi hal ini adalah perlunya sistem rekomendasi yang mampu menyaring informasi secara akurat berdasarkan karakteristik unik dataset film lokal. Penelitian ini bertujuan untuk mengevaluasi performa algoritma kemiripan pada sistem rekomendasi film Indonesia guna memastikan efektivitas rekomendasi yang diberikan kepada pengguna. Metode yang diterapkan adalah *Content-Based Filtering* menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan perhitungan *Cosine Similarity* pada dataset riil *IMDb Indonesian Movies*. Evaluasi performa diukur secara komprehensif menggunakan metrik *Precision at K*, *Recall at K*, dan *Mean Absolute Error* (MAE). Hasil pengujian menunjukkan performa sistem yang cukup menjanjikan sebagai *baseline*, dengan nilai *Precision* mencapai 46.70% pada daftar 5 rekomendasi teratas (*Top-5*). Selain itu, sistem mencatat tingkat simpangan tebakan yang rendah dengan nilai MAE sebesar 0.9177 pada rentang skala rating 1-10. Hasil ini membuktikan bahwa algoritma *Cosine Similarity* dapat diandalkan sebagai landasan awal dalam memberikan rekomendasi film lokal, meskipun masih membutuhkan kombinasi metode lain untuk meningkatkan presisi.

Kata kunci : Sistem Rekomendasi, Film Indonesia, *Content-Based Filtering*, *Cosine Similarity*, *Precision*, MAE

1. PENDAHULUAN

Pertumbuhan industri perfilman di Indonesia menunjukkan tren positif yang sangat signifikan seiring dengan kemajuan platform penayangan digital. Ribuan judul film lokal dari berbagai genre kini dapat diakses dengan mudah oleh penonton di mana saja. Sayangnya, kemudahan akses dan ledakan jumlah konten ini justru memicu fenomena *information overload*. Pengguna seringkali merasa kewalahan dan menghabiskan terlalu banyak waktu hanya untuk menyeleksi tontonan yang sesuai dengan selera mereka. Untuk mengatasi masalah penumpukan informasi tersebut, pengembangan sistem rekomendasi (*recommender system*) yang cerdas dan adaptif menjadi sebuah urgensi guna menyaring konten secara personal [1], [2].

Sebagai solusi teknis, pendekatan *Content-Based Filtering* sangat umum diterapkan dalam menyaring informasi konten hiburan. Metode ini bekerja dengan menganalisis karakteristik item, seperti genre dan sinopsis, untuk mencari nilai kemiripan (*similarity*) dengan preferensi pengguna di masa lalu. Dalam proses perhitungannya, algoritma *Cosine Similarity* sering menjadi pilihan utama karena dinilai sangat efisien dalam mengukur kedekatan antar vektor fitur teks [3], [4]. Walaupun terbukti handal dalam banyak kasus, tidak semua sistem mampu menyajikan urutan rekomendasi teratas (*top-ranking*) yang presisi. Tantangan ini semakin nyata ketika algoritma dihadapkan pada dataset lokal yang memiliki distribusi fitur yang unik, sehingga menuntut adanya pengujian dan pembobotan yang lebih adaptif [5].

Menyadari adanya potensi kelemahan pada akurasi peringkat tersebut, penelitian ini berfokus untuk melakukan evaluasi performa algoritma secara komprehensif. Pengukuran keberhasilan sistem ditekankan pada metrik *Precision at K* ($P@K$) dan *Recall at K* ($R@K$) guna mengukur seberapa relevan dan lengkap daftar item pada urutan teratas yang disajikan kepada pengguna [6]. Sebagai pengujian tambahan untuk melihat simpangan prediksi, evaluasi juga melibatkan metrik *Mean Absolute Error* (MAE) [7]. Kebaruan (*novelty*) dari penelitian ini terletak pada fokus pengujian algoritma secara spesifik pada dataset riil *IMDb Indonesian Movies*. Melalui analisis ini, diharapkan hasil penelitian dapat menjadi acuan teknis untuk meminimalkan tingkat kesalahan sistem dan mendorong pengembangan platform informasi perfilman yang jauh lebih akurat di Indonesia. Penelitian ini memiliki kontribusi utama pada evaluasi performa algoritma *Content-Based Filtering* menggunakan metrik berbasis peringkat (*Top-K*), yaitu *Precision at K* dan *Recall at K*, pada dataset film Indonesia. Pendekatan ini dipilih karena lebih merepresentasikan kondisi nyata sistem rekomendasi, di mana pengguna hanya memperhatikan sejumlah item teratas yang disarankan.

2. TINJAUAN PUSTAKA

Sistem rekomendasi merupakan perangkat lunak penunjang keputusan yang dirancang untuk mengatasi masalah *information overload* dengan menyarankan item yang paling relevan bagi pengguna. Dalam perkembangannya, pendekatan *Content-Based Filtering* (CBF) menjadi salah satu metode yang

paling andal secara teori. CBF bekerja dengan mengekstraksi fitur intrinsik suatu item (seperti genre dan sinopsis), kemudian merekomendasikan item lain yang memiliki karakteristik serupa dengan preferensi pengguna sebelumnya.

Untuk memproses fitur teks seperti sinopsis, diperlukan pembobotan kata menggunakan landasan teori *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF berfungsi mengukur seberapa penting sebuah kata dalam suatu dokumen relatif terhadap seluruh korpus data. Setelah dokumen diubah menjadi representasi vektor numerik, kedekatan antar film dihitung menggunakan algoritma *Cosine Similarity*. Secara matematis, algoritma ini mengukur kosinus sudut antara dua vektor berdimensi banyak, di mana nilai yang mendekati angka 1 menandakan tingkat kemiripan konten yang sangat tinggi.

Kualitas dari sebuah sistem rekomendasi diukur menggunakan evaluasi metrik berbasis peringkat (*ranking metrics*) dan metrik prediksi. *Precision at K* ($P@K$) mengukur persentase item relevan di dalam daftar K rekomendasi teratas yang disajikan kepada pengguna, sedangkan *Recall at K* ($R@K$) mengukur seberapa banyak item relevan yang berhasil ditemukan dari total keseluruhan item yang relevan. Sementara itu, untuk evaluasi prediksi skala, *Mean Absolute Error* (MAE) digunakan untuk mengukur simpangan rata-rata (tingkat *error*) antara nilai prediksi sistem dengan nilai *rating* aktual yang diberikan oleh pengguna.

Penelitian mengenai sistem rekomendasi dengan pemanfaatan teori-teori tersebut telah banyak dilakukan. Beberapa studi pada [4], [5], dan [8] membuktikan bahwa algoritma TF-IDF dan *Cosine Similarity* memiliki akurasi tinggi dalam merekomendasikan film, baik pada dataset global seperti *The Movie Database* (TMDb) maupun implementasi pada aplikasi web. Fleksibilitas metode ini juga dibuktikan penerapannya pada sistem pencarian buku perpustakaan [9], serta dikombinasikan dengan *Levenshtein Distance* pada penelitian [10].

Kajian komprehensif pada [1], [2], dan [7] menegaskan bahwa *Cosine Similarity* tetap menjadi landasan efisien untuk mengatasi masalah *Cold-Start* dan pemrosesan data tekstual, sejalan dengan arsitektur rekomendasi yang dibangun pada [3]. Kendati sistem-sistem tersebut terbukti fungsional, studi pada [6] menyoroti pentingnya evaluasi performa secara mendalam menggunakan metrik berbasis peringkat agar hasil lebih merepresentasikan akurasi sistem terhadap selera pengguna.

Dari tinjauan literatur terdahulu, ditemukan celah penelitian (*research gap*), yakni sebagian besar pengujian masih menggunakan dataset global dan hanya berfokus pada fungsionalitas sistem. Penelitian ini hadir untuk melengkapi celah tersebut dengan menguji performa algoritma *Cosine Similarity* secara spesifik dan mendalam menggunakan metrik

Precision, *Recall*, dan MAE dengan fokus utama pada dataset *IMDb Indonesian Movies*.

3. METODE PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan sistematis untuk memastikan hasil evaluasi yang valid dan komprehensif. Data yang digunakan adalah dataset *IMDb Indonesian Movies* yang memuat informasi lebih dari 1.200 film lokal. Tahapan tersebut meliputi pengumpulan data, pra-pemrosesan data, perhitungan kemiripan dengan algoritma *Cosine Similarity* menggunakan pendekatan *Content-Based Filtering*, dan diakhiri dengan evaluasi model.

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari repositori publik Kaggle dengan nama dataset *IMDb Indonesian Movies*. Dataset ini berisi informasi mengenai film-film produksi Indonesia, yang mencakup beberapa atribut penting seperti *Title* (judul film), *Genre* (kategori film), *Synopsis* (alur cerita), dan *User Rating* (penilaian pengguna). Atribut *Genre* dan *Rating* akan menjadi parameter utama dalam proses pembentukan vektor fitur untuk sistem rekomendasi.

Berbeda dengan penelitian pada [8] yang menggunakan metrik evaluasi global seperti *precision* dan *recall* secara keseluruhan, penelitian ini menggunakan pendekatan berbasis peringkat (Top-K). Pendekatan ini dinilai lebih relevan dalam konteks sistem rekomendasi karena fokus pada kualitas item yang muncul pada posisi teratas.

3.2. Pra-pemrosesan Data

Sebelum data diolah oleh algoritma, dilakukan tahapan pra-pemrosesan untuk membersihkan noise dan menstandarisasi format data. Tahapan ini meliputi:

- Data Cleansing*: Menghapus baris data yang memiliki nilai kosong (*missing values*) pada kolom esensial seperti *genre* dan *rating*.
- Case Folding*: Mengubah seluruh karakter teks menjadi huruf kecil (*lowercase*) agar konsisten.
- Tokenization*: Memecah string *genre* (misalnya: "Action, Comedy") menjadi token-token kata yang terpisah.
- Vectorization* (TF-IDF): Mengubah token teks tersebut menjadi representasi numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) agar dapat dihitung jaraknya oleh algoritma.

3.3. Algoritma *Cosine Similarity*

Sistem rekomendasi pada penelitian ini dibangun menggunakan pendekatan *Content-Based Filtering*. Untuk menghitung tingkat kemiripan antar film, digunakan algoritma *Cosine Similarity*. Algoritma ini bekerja dengan mengukur kosinus sudut antara dua vektor multidimensi (dalam hal ini, vektor fitur *genre* film). Semakin kecil sudut antar vektor, semakin tinggi

tingkat kemiripannya (mendekati nilai 1). Perhitungan matematisnya dirumuskan sebagai berikut:

$$\text{Cos}(\theta) = \frac{A \cdot B}{||A|| \times ||B||} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)$$

Keterangan:

- A dan B adalah Vektor dari dua film yang dibandingkan.
- A_i dan B_i adalah nilai fitur (TF-IDF dari *genre*) pada indeks ke- i .

3.4. Evaluasi Model

Evaluasi dalam penelitian ini tidak hanya menggunakan metrik umum, tetapi difokuskan pada metrik berbasis peringkat (Top-K) yang lebih mencerminkan pengalaman pengguna dalam sistem rekomendasi.

Tahap krusial dalam penelitian ini adalah mengevaluasi seberapa baik algoritma *Cosine Similarity* dalam memberikan rekomendasi yang relevan. Evaluasi dilakukan menggunakan tiga metrik utama:

- Precision at K (P@K):** Mengukur persentase film yang benar-benar relevan dari K jumlah film teratas yang direkomendasikan kepada pengguna.
- Recall at K (R@K):** Mengukur proporsi film relevan yang berhasil ditemukan oleh sistem dan masuk ke dalam daftar *Top-K* rekomendasi, dibandingkan dengan total film relevan di seluruh dataset.
- Mean Absolute Error (MAE):** Mengukur rata-rata selisih absolut antara nilai kedekatan/rating yang diprediksi oleh sistem dengan nilai aktual dari dataset. Semakin kecil nilai MAE, semakin akurat prediksi sistem. Rumus MAE didefinisikan sebagai:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i| \quad (2)$$

Di mana n adalah jumlah data, y_i adalah nilai aktual, dan y'_i adalah nilai prediksi kemiripan.

4. HASIL DAN PEMBAHASAN

Memuat hasil, Pengujian dan pembahasan tentang skripsi yang telah dilakukan

4.1. Persiapan Data

Data yang digunakan bersumber dari repositori publik Kaggle dengan nama dataset *indonesian_movies.csv*. Dataset ini terdiri dari 1.200 entri judul film dengan beberapa atribut utama yang digunakan dalam penelitian, yaitu:

- Title (Judul film)
- Genre (Kategori film)
- Description (Sinopsis singkat)
- Users_Rating (Rating rata-rata dari pengguna)

Tabel 1. Contoh Data Mentah (Sebelum *Preprocessing*)

Title	Genre	Description	Users_Rating
Pengabdi Setan	Horror, Drama	Rini and her family are haunted...	7.6
Warkop DKI Reborn	Comedy	Dono, Kasino, and Indro are back..	6.5

4.2. Pra-pemrosesan Data (*Preprocessing*)

Sebelum diolah, data teks harus dibersihkan dan distandarisasi. Kolom *genre* dan *description* digabungkan menjadi satu variabel fitur teks utuh yang disebut *tags*.

a. Data Cleansing

Menghapus data dengan nilai kosong pada kolom penting seperti *genre* dan deskripsi agar tidak menimbulkan error saat perhitungan..

b. Case Folding

Mengubah seluruh teks menjadi huruf kecil (*lowercase*) untuk menjaga konsistensi dan menghindari perbedaan akibat kapitalisasi.

Tabel 2. Contoh Teks Setelah *Case Folding*

Judul Film	Fitur Teks (<i>Tags</i>)
Pengabdi Setan	horror drama rini and her family are haunted...

c. Tokenization

Memecah string fitur teks yang utuh tersebut menjadi token-token kata yang terpisah atau berupa *array* sehingga mesin dapat membaca setiap kata sebagai satu entitas unik.

Tabel 3. Contoh Hasil Tokenisasi

Judul Film	Hasil Tokenisasi
Pengabdi Setan	["horror", "drama", "rini", "and", "her", "family", "are", "haunted"]

d. Vectorization (TF-IDF)

Mengubah teks menjadi representasi numerik menggunakan metode TF-IDF, di mana kata yang sering muncul pada satu film namun jarang pada film lain diberi bobot lebih tinggi.

Tabel 4. Contoh Hasil Pembobotan TF-IDF untuk Film "Pengabdi Setan"

Term (Kata)	Bobot TF-IDF
horror	0.3542
rini	0.6120
haunted	0.5431
family	0.2815

4.3. Implementasi Algoritma *Cosine Similarity*

Setelah nilai bobot TF-IDF didapatkan, algoritma *Cosine Similarity* diterapkan untuk menghitung tingkat kemiripan antar dokumen film. Algoritma ini membandingkan sudut antar vektor fitur; nilai yang mendekati angka 1.0 menunjukkan tingkat kemiripan

yang sangat tinggi. Perhitungan ini menghasilkan matriks korelasi (*similarity matrix*) berukuran $N \times N$ (di mana N adalah jumlah film).

4.4. Tampilan Hasil Rekomendasi

Sebagai simulasi fungsional sistem, dilakukan pengujian dengan memberikan kueri judul film "Pengabdi Setan". Sistem kemudian mengurutkan nilai *Cosine Similarity* tertinggi dan menyajikan 5 film teratas (*Top-5*) sebagai rekomendasi.

Tabel 5. Contoh Hasil Rekomendasi untuk Film "Pengabdi Setan"

Film Rekomendasi	Genre	Cosine Similarity
Rindu Kami Padamu	Drama	0.1687
Jeritan Kuntulanak	Horror	0.1611
Rumah dan Musim Hujan	Comedy	0.1112
Aku Ingin Ibu Pulang	Drama	0.0879
27 Steps of May	Drama	0.0846

Dari Tabel 5 dapat dilihat bahwa sistem berhasil menampilkan daftar film rekomendasi berdasarkan kedekatan fitur teks secara keseluruhan. Menariknya, hasil rekomendasi tidak hanya memunculkan film bergenre *Horror* (seperti *Jeritan Kuntulanak*), melainkan juga film dengan genre *Drama* dan *Comedy*. Hal ini terjadi karena algoritma TF-IDF dan *Cosine Similarity* mencocokkan kemiripan *term* atau kata pada sinopsis. Jika film "Pengabdi Setan" memiliki dominasi kata kunci terkait "keluarga" atau "rumah", maka film dengan genre lain yang memiliki fokus cerita serupa berpotensi direkomendasikan. Ini membuktikan bahwa sistem berbasis konten bekerja lebih detail melampaui batas klasifikasi genre semata.

4.5. Evaluasi Performa Sistem

Pengujian akurasi daftar rekomendasi secara keseluruhan dilakukan menggunakan metrik *Precision at K* ($P@K$) dan *Recall at K* ($R@K$) dengan variasi K (jumlah daftar rekomendasi yang ditampilkan) sebesar 5, 10, dan 20. Film direpresentasikan "relevan" apabila memiliki minimal satu irisan target genre yang sepadan. Hasil pengujian ditunjukkan pada Tabel 6.

Tabel 6. Hasil Pengujian *Precision* dan *Recall*

Nilai K	Precision (%)	Recall (%)
K = 5	46.70	2.20
K = 10	43.67	3.87
K = 20	41.57	7.07

Berdasarkan Tabel 6, terlihat bahwa nilai *Precision* tertinggi didapatkan pada rekomendasi *Top-5* ($K=5$) sebesar 46.70%. Angka ini menunjukkan bahwa rata-rata hampir separuh dari daftar film teratas yang disajikan sistem memiliki kecocokan (*relevance*) dengan profil film yang dicari. Seiring dengan bertambahnya daftar K menjadi 20, *Precision* menurun secara gradual menjadi 41.57%, sementara nilai *Recall* meningkat menjadi 7.07%. Walaupun tidak mencapai

angka absolut di atas 80%, hasil presisi di kisaran 41-46% ini merupakan kinerja yang wajar dan realistis untuk sistem yang sepenuhnya mengandalkan metode murni *Content-Based Filtering*. Sistem ini memiliki batasan dalam memahami konteks semantik yang lebih dalam dan tidak memiliki riwayat preferensi *rating* pengguna (hanya mencocokkan teks sinopsis dan genre). Fenomena *trade-off* ini merupakan kondisi lumrah dalam *Information Retrieval*, di mana menyajikan lebih banyak item akan memperbesar kemungkinan masuknya film yang kurang relevan [6].

Jika dibandingkan dengan penelitian sebelumnya seperti pada [8], nilai *precision* dalam penelitian ini relatif lebih rendah. Hal ini disebabkan oleh perbedaan pendekatan evaluasi yang digunakan. Penelitian [8] menggunakan metrik evaluasi global, sedangkan penelitian ini menggunakan *Precision at K* yang lebih ketat karena hanya mengevaluasi relevansi pada sejumlah item teratas.

Selain itu, penggunaan dataset film Indonesia yang memiliki variasi bahasa dan struktur teks yang lebih beragam juga mempengaruhi performa model. Kompleksitas ini menyebabkan proses pembobotan TF-IDF menjadi lebih menantang dibandingkan dataset global yang lebih terstandarisasi.

Dengan demikian, nilai *precision* pada kisaran 41%–46% masih dapat dianggap wajar untuk sistem berbasis *Content-Based Filtering*, terutama dalam konteks evaluasi *Top-K* yang lebih realistis.

Di sisi lain, pengujian simpangan prediksi *rating* menggunakan *Mean Absolute Error* (MAE) juga dilakukan untuk mengukur rata-rata kesalahan tebakan sistem terhadap skor *rating* asli pengguna. Hasil pengujian MAE ditunjukkan pada Tabel 7.

Tabel 7. Hasil Pengujian Mean Absolute Error (MAE)

Metrik Evaluasi	Nilai Error
Mean Absolute Error (MAE)	0.9177

Berdasarkan Tabel 7, nilai MAE yang dihasilkan adalah sebesar 0.9177. Perlu diketahui bahwa nilai *rating* film (*users_rating*) pada dataset berada pada skala maksimal 10.0 (misalnya 7.6 atau 8.0). Nilai MAE sebesar 0.9177 membuktikan bahwa selisih tebakan sistem dengan skor *rating* aslinya rata-rata kurang dari 1 poin. Hal ini menunjukkan bahwa pembobotan *Content-Based* terbukti cukup stabil dan adaptif dalam memprediksi ketertarikan penonton film nasional.

Meskipun menunjukkan performa yang cukup baik sebagai *baseline*, sistem ini masih memiliki keterbatasan, terutama dalam memahami konteks semantik yang lebih dalam. Hal ini disebabkan karena metode yang digunakan hanya mengandalkan kesamaan kata (*term-based similarity*) tanpa mempertimbangkan makna kontekstual.

Selain itu, tidak digunakannya teknik *preprocessing* lanjutan seperti *stopword removal* dan *lemmatization* juga dapat mempengaruhi kualitas

representasi fitur, sehingga berdampak pada akurasi hasil rekomendasi.

5. KESIMPULAN DAN SARAN

Penerapan *Content-Based Filtering* menggunakan *Cosine Similarity* pada dataset *IMDb Indonesian Movies* telah berhasil diterapkan sebagai landasan utama dalam merekomendasikan film lokal. Hasil evaluasi menunjukkan bahwa sistem mampu memberikan performa awal yang cukup menjanjikan, dengan nilai *Precision* tertinggi mencapai 46.70% pada daftar 5 rekomendasi teratas (P@5). Meskipun akurasi prediksi item relevan berada di level menengah karena keterbatasan pencocokan murni berbasis teks, sistem berhasil mempertahankan tingkat simpangan prediksi yang sangat baik dengan nilai MAE yang rendah yakni 0.9177 pada skala rating 1-10. Penurunan nilai presisi seiring bertambahnya jumlah *K* menjadi bukti bahwa *Content-Based* murni rentan terhadap masalah *over-specialization*, di mana kemiripan kosakata sinopsis dapat mengabaikan konteks kesukaan *genre* penonton. Oleh karena itu, penelitian selanjutnya sangat disarankan untuk mengintegrasikan pendekatan *Collaborative Filtering* untuk membentuk algoritma *Hybrid* agar mesin dapat belajar dari riwayat *rating* pengguna, sehingga rekomendasi yang dihasilkan menjadi jauh lebih akurat dan terfokus. Secara keseluruhan, penelitian ini menunjukkan bahwa penggunaan metrik evaluasi berbasis peringkat memberikan gambaran yang lebih realistis terhadap performa sistem rekomendasi. Hasil yang diperoleh diharapkan dapat menjadi dasar pengembangan lebih lanjut, khususnya dalam meningkatkan kualitas rekomendasi melalui pendekatan *hybrid*.

DAFTAR PUSTAKA

- [1] U. Javed, K. Shaukat, I. A. Hameed, F. Iqbal, T. M. Alam, and S. Luo, "A Review of Content-Based and Context-Based Recommendation Systems," *Int. J. Emerg. Technol. Learn.*, vol. 16, no. 3, pp. 274–306, 2021, doi: 10.3991/ijet.v16i03.18851.
- [2] C. Li, I. Ishak, H. Ibrahim, M. Zolkepli, F. Sidi, and C. Li, "Deep Learning-Based Recommendation System: Systematic Review and Classification," *IEEE Access*, vol. 11, no. September, pp. 113790–113835, 2023, doi: 10.1109/ACCESS.2023.3323353.
- [3] E. Salim, J. Pragantha, and D. L. Manatap, "Perancangan Sistem Rekomendasi Film menggunakan metode Content- based Filtering," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 6, pp. 2188–2199, 2021, [Online]. Available: https://linter.untar.ac.id/repository/penelitian/bukti/penelitian_10390001_7A281222103549.pdf
- [4] M. Fajriansyah, P. P. Adikara, and A. W. Widodo, "Sistem Rekomendasi Film Menggunakan Content Based Filtering," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 6, pp. 2188–2199, 2021, [Online]. Available: https://linter.untar.ac.id/repository/penelitian/bukti/penelitian_10390001_7A281222103549.pdf
- [5] Z. F. Meillano, D. Richasdy, and H. Hasmawati, "Web-Based Movie Recommendation System Using Content-Based Filtering and Cosine Similarity," *Indones. J. Artif. Intell. Data Min.*, vol. 7, no. 1, p. 157, 2024, doi: 10.24014/ijaidm.v7i1.27390.
- [6] A. H. Juni Permana and Agung Toto Wibowo, "Movie Recommendation System Based on Synopsis Using Content-Based Filtering with TF-IDF and Cosine Similarity," *Int. J. Inf. Commun. Technol.*, vol. 9, no. 2, pp. 1–14, 2023, doi: 10.21108/ijoict.v9i2.747.
- [7] H. Ko and A. Lee, Suyeon; Park, Yoonseo; Choi, "A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields," *Inf. Resour. Manag. J.*, vol. 11, no. 1, pp. 53–73, 2024, doi: 10.4018/IRMJ.2020100104.
- [8] D. Velamentosa and E. Zuliarso, "Sistem Rekomendasi Film Menggunakan Metode Content-Based Filtering," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 2, pp. 2918–2922, 2025, doi: 10.36040/jati.v9i2.13251.
- [9] R. Hanun, D. Reswara, and A. Susanto, "Implementasi Content-Based Filtering pada Sistem Rekomendasi Buku Perpustakaan," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 2, pp. 3243–3250, 2025.
- [10] M. M. Kunde, Omkar; Gaikwad, Omkar; Kelgandre, Prathamesh; Damodhar, Rohan; Swami, "The Movie Recommendation System using Content Based Filtering with TF-IDF-Vectorization and Levenshtein Distance," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 2, no. 2, pp. 257–263, 2022, doi: 10.48175/ijarsct-3648.