

**PENERAPAN *CONTENT-BASED FILTERING* PADA SISTEM
REKOMENDASI NOVEL UNTUK PENINGKATAN
PENGALAMAN PENGGUNA**

SKRIPSI

Yudistira Pratama



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK, KOMPUTER DAN DESAIN
UNIVERSITAS NUSA PUTRA
SUKABUMI
2025**

**PENERAPAN *CONTENT-BASED FILTERING* PADA SISTEM
REKOMENDASI NOVEL UNTUK PENINGKATAN
PENGALAMAN PENGGUNA**

SKRIPSI

*Diajukan Untuk Memenuhi Salah Satu Syarat Dalam Menempuh gelar sarjana
Teknik Informatika*



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK, KOMPUTER DAN DESAIN
UNIVERSITAS NUSA PUTRA
SUKABUMI
2025**

PERNYATAAN PENULIS

JUDUL : PENERAPAN *CONTENT-BASED FILTERING* PADA SISTEM
REKOMENDASI NOVEL UNTUK PENINGKATAN
PENGALAMAN PENGGUNA

NAMA : YUDISTIRA PRATAMA

NIM 20210040108

”Saya menyatakan dan bertanggungjawab dengan sebenarnya bahwa Skripsi ini adalah hasil karya sendiri kecuali cuplikan dan ringkasan yang masing-masing telah dijelaskan sumbernya. Jika pada waktu selanjutnya ada pihak lain yang mengklaim bahwa Skripsi ini sebagai karyanya, yang disertai dengan bukti- bukti yang cukup, maka penulis bersedia untuk dibatalkan gelar Sarjana Komputer beserta segala hak dan kewajiban yang melekat pada gelar tersebut”.



Sukabumi, 17 Juli 2025

Yudistira Pratama

PENGESAHAN SKRIPSI

JUDUL : PENERAPAN *CONTENT-BASED FILTERING* PADA SISTEM
REKOMENDASI NOVEL UNTUK PENINGKATAN
PENGALAMAN PENGGUNA

NAMA : YUDISTIRA PRATAMA

NIM 20210040108

Skripsi ini telah diujikan dan dipertahankan di depan Dewan Penguji pada Sidang
Skripsi tanggal 15 juli 2025 Menurut pandangan kami, Skripsi ini memadai dari
segi kualitas untuk tujuan penganugerahan gelar Sarjana Komputer (S.Kom)

Sukabumi, 15 Juli 2025

Pembimbing I

Pembimbing II

Imam Sanjaya, S.P., M.Kom

NIDN . 0427087802

Alun Sujjada, S.Kom, M.T

NIDN . 0718108001

Ketua Penguji

Ketua Program Studi Teknik Informatika

Muhammad Ikhsan Thohir, M.Kom

NIDN . 0415049302

Ir. Somantri, S.T., M.Kom

NIDN . 0419128801

Dekan Fakultas Teknik Komputer dan Desain

Ir. Paikun S.T., M.T., IPM., ASEAN Eng

NIDN. 0402037401

HALAMAN PERUNTUKAN

Dengan segala kerendahan hati dan rasa syukur, kami ucapkan terima kasih kepada Allah SWT yang senantiasa memberikan segala keberanian, ketabahan, dan semangat yang tak kenal rasa menyerah dalam diri kami. Atas segala izin nya selalu menjawab doa-doa baik yang kami panjatkan dan memberikan takdir yang terbaik untuk kami. Berkat rahmat dan anugerah-nya, akhirnya kami bisa menyelesaikan tugas ini dengan baik.

Dengan segala kerendahan hati, karya skripsi ini saya persembahkan kepada:

1. Allah SWT, Segala puji bagi Allah SWT atas limpahan rahmat, hidayah, dan kemudahan-Nya dalam setiap langkah perjalanan hingga skripsi ini dapat terselesaikan.
2. Kepada Kedua Orang Tua Saya tercinta yang selalu memberikan doa, dan dukungan yang tiada henti. Semangat dan pengorbanan kalian menjadi sumber motivasi terbesar dalam menyelesaikan skripsi ini.
3. Kepada Kedua Saudara Saya, dan Seluruh Keluarga saya tercinta yang selalu memberikan doa, dan dukungan yang tiada henti, menjadi sumber motivasi terbesar dalam menyelesaikan skripsi ini.
4. Dosen Pembimbing, terima kasih kepada Bapak Imam Sanjaya, M.Kom dan Bapak Alun sujada, S.Kom, M.T atas bimbingan, arahan, serta ilmu yang diberikan dengan sabar dan penuh perhatian selama proses penyusunan skripsi ini .
5. Dosen-Dosen Program Studi Teknik Informatika, terima kasih yang sebesar-besarnya kami sampaikan kepada seluruh keluarga besar Teknik Informatika, termasuk Kaprodi beserta jajarannya, serta semua dosen yang tidak dapat kami sebutkan satu per satu. Ilmu dan pengalaman berharga yang kalian berikan selama masa perkuliahan telah menjadi fondasi yang kuat dan sangat bermanfaat dalam penelitian ini.
6. Untuk semua teman yang telah setia kebersamaan kami dan menjadi motivasi terbesar, terutama rekan-rekan Teknik Informatika 21A, Semoga kita dapat terus

bersama dan saling mendukung ke depannya. Terima kasih atas kebersamaan dan dukungan selama masa perkuliahan ini.

7. Terakhir kepada pihak lainnya yang tidak bisa disebutkan satu persatu, untuk semua yang telah memberikan bantuan, dukungan, dan doa dalam berbagai bentuk selama masa studi dan penyusunan skripsi kami, baik yang secara langsung maupun tidak langsung, kami ucapkan terima kasih yang sebesar-besarnya.
8. Akhir kata, kami berharap semoga karya yang sederhana ini dapat memberikan manfaat bagi kami sebagai penulis, pembaca, dan juga bagi perkembangan ilmu pengetahuan. Semoga segala kesalahan dan kekurangan yang ada dalam karya ini menjadi pelajaran berharga bagi kami di masa mendatang. Aamiin.



ABSTRAK

Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi novel yang mampu memberikan saran bacaan yang relevan dan personal kepada pengguna berdasarkan preferensi mereka. Metode yang digunakan adalah *Content-Based Filtering*, yang menganalisis karakteristik konten novel seperti judul, genre, dan penulis. Untuk meningkatkan kualitas rekomendasi, *algoritma Random Forest* diintegrasikan sebagai alat prediksi terhadap tingkat popularitas novel berdasarkan skor dan atribut lainnya. Data yang digunakan berasal dari dataset publik berisi 4.500 judul novel dengan atribut seperti genre, skor, popularitas, dan tahun rilis. Proses penelitian meliputi pembersihan data, ekstraksi fitur menggunakan *TF-IDF*, perhitungan kemiripan teks dengan *cosine similarity*, serta pelatihan model klasifikasi menggunakan *Random Forest*. Evaluasi model dilakukan menggunakan *Confusion Matrix*, menghasilkan nilai akurasi sebesar 94,2%, *precision* 78,3%, *recall* 87,4%, dan *F1-score* 82,6%. Hasil ini menunjukkan bahwa integrasi metode *Content-Based Filtering* dan *algoritma Random Forest* mampu memberikan rekomendasi yang lebih akurat, relevan, dan sesuai dengan minat pengguna. Sistem ini kemudian dikembangkan menjadi aplikasi berbasis web menggunakan *Streamlit* yang memungkinkan pengguna mengakses rekomendasi secara interaktif. Aplikasi menampilkan tiga fitur utama yaitu rekomendasi berdasarkan popularitas, skor/rating, serta kesamaan genre dan judul. Dengan pendekatan ini, diharapkan pengalaman membaca pengguna dapat ditingkatkan melalui rekomendasi yang lebih tepat sasaran dan memuaskan.

Kata Kunci: Sistem Rekomendasi, *Content-Based Filtering*, *Random Forest*, *TF-IDF*, *Cosine Similarity*, *Streamlit*, Novel.

ABSTRACT

This research aims to develop a novel recommendation system capable of delivering relevant and personalized reading suggestions to users based on their preferences. The method employed is Content-Based Filtering, which analyzes novel content attributes such as title, genre, and author. To improve the accuracy and quality of recommendations, the Random Forest algorithm is integrated to predict a novel's popularity based on its rating and other attributes. The dataset used in this study consists of 4,500 publicly available novel titles, with features such as genre, score, popularity, and release year. The research process involves data cleaning, feature extraction using TF-IDF, text similarity computation via cosine similarity, and model training using Random Forest classification. Model evaluation is conducted using a Confusion Matrix, achieving 94,2% accuracy, 78,3% precision, 87.4% recall, and an F1-score of 82.6%. These results demonstrate that the integration of Content-Based Filtering and Random Forest can produce more accurate and user-aligned recommendations. The system is then implemented as a web-based application using Streamlit, allowing users to interactively access the recommendation features. The application includes three main features: recommendations based on popularity, rating score, and similarity in genre and title. This approach is expected to enhance user reading experiences by offering more relevant, precise, and satisfying novel suggestions, especially in the digital platform ecosystem.

Keywords: Recommendation System, Content-Based Filtering, Random Forest, TF-IDF, Cosine Similarity, Streamlit, Novel.

KATA PENGANTAR

Puji syukur kami panjatkan ke hadirat Allah SWT atas rahmat dan karunia-Nya. Shalawat dan salam kami haturkan kepada Nabi Muhammad SAW. Dengan penuh rasa syukur, penulis dapat menyelesaikan skripsi ini yang berjudul “*PENERAPAN CONTENT-BASED FILTERING PADA SISTEM REKOMENDASI NOVEL UNTUK PENINGKATAN PENGALAMAN PENGGUNA*” dengan baik.

Penulis menyadari bahwa penulisan skripsi ini masih jauh dari kata sempurna, baik dari segi penyajian, permasalahan, maupun pemecahan masalah. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang konstruktif untuk perbaikan di masa mendatang. Semoga skripsi ini dapat bermanfaat, khususnya bagi penulis dan umumnya bagi semua pihak yang membutuhkan.

Selama mengerjakan skripsi ini, penulis tidak lepas dari bantuan dan dorongan dari berbagai pihak, sehingga skripsi ini dapat terselesaikan. Pada kesempatan ini, penulis ingin mengucapkan terima kasih yang setulus-tulusnya kepada:

1. Bapak Dr. Kurniawan, S.T., M.Si, M.M, selaku Rektor Universitas Nusa Putra.
2. Bapak Ir. Paikun, S.T., IPM., Asean Eng, selaku Dekan Fakultas Univeristas Nusa Putra.
3. Bapak Ir.Somantri, ST., M.Kom, selaku ketua program studi Teknik Informatika Universitas Nusa Putra.
4. Bapak Imam Sanjaya, S.P, M.Kom, dan Bapak Alun sujjada, S.Kom, M.T selaku dosen pembimbing skripsi atas segala bimbingan, arahan serta saran yang diberikan kepada kami sehingga skripsi ini dapat diselesaikan dengan baik.
5. Ibu Anggun Fergina, S.Kom., M.Kom selaku pembimbing akademik kami, yang selalu memberikan bimbingan, arahan, dan dukungan selama proses perkuliahan.
6. Seluruh dosen/staff pengajar Fakultas Komputer Teknik dan desain program studi Teknik Informatika yang telah memberikan ilmu pengetahuan yang tak ternilai selama kami menempuh pendidikan di Universitas Nusa Putra.
7. Kepada seluruh keluarga, orang tua, adik, kakek, nenek, paman, bibi, dan sepupu saya yang selalu memberikan dukungan penuh agar kami bisa

menyelesaikan skripsi ini. Dorongan dan motivasi dari mereka sangat berarti, dan menjadi sumber semangat yang tak ternilai selama proses penulisan.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu kritik dan saran yang membangun dari berbagai pihak sangat kami harapkan demi perbaikan. Amin Yaa Rabbal 'Alamin.

Sukabumi, Juli 2025

Penulis

Yudistira Pratama



HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik UNIVERSITAS NUSA PUTRA, kami yang bertanda tangan dibawah ini :

Nama : Yudistira Pratama
NIM : 20210040108
Program Studi : Teknik Informatika
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Nusa Putra **Hak Bebas Royalti Noneksklusif** (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

“PENERAPAN CONTENT-BASED FILTERING PADA SISTEM REKOMENDASI NOVEL UNTUK PENINGKATAN PENGALAMAN PENGGUNA”, beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Noneksklusif ini, Universitas Nusa Putra berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di: Sukabumi

Pada tanggal: 17 Juli 2025

Yang Menyatakan,

Yudistira Pratama

DAFTAR ISI

PERNYATAAN PENULIS.....	ii
PENGESAHAN SKRIPSI	iii
ABSTRAK.....	vi
ABSTRACT	vii
KATA PENGANTAR	viii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	x
DAFTAR ISI.....	xi
DAFTAR TABLE.....	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat penelitian.....	5
1.6 Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA.....	8
2.1 Penelitian Terkait.....	8
2.2 Landasan teori.....	13
2.2.1 Sistem rekomendasi	13
2.2.2 Content-based filtering	13
2.2.3 Algoritma Random Forest.....	14
2.2.4 Frequency-Inverse Document Frequency (TF-IDF).....	16
2.2.5 Confusion Matrix	17
2.3 Kerangka berpikir.....	19
BAB III METODOLOGI PENELITIAN.....	21

3.1 Metode Penelitian	21
3.2 Identifikasi masalah	23
3.3 Pengumpulan data.....	23
3.4 Pre-Processing Data	24
3.4.1 Load data.....	25
3.4.2 Duplikat data.....	25
3.4.3 Mengubah Data CSV Ke Data Frame	25
3.4.4 Isi Data Kosong (fillna).....	26
3.4.5 Ekstraksi Fitur	27
3.5 Pemodelan.....	29
3.5.1 Content-Based Filtering	30
3.5.2 Algoritma.....	31
3.6 Evaluasi.....	33
3.7 Deployment	34
3.8 Skenario	37
BAB IV HASIL DAN PEMBAHASAN.....	39
4.1 Pre-processing data	39
4.2 Pemodelan.....	41
4.3 Evaluasi.....	42
4.4 EDA (Exploratory Data Analysis).....	47
4.5 WebApp (Black box).....	52
BAB V KESIMPULAN DAN SARAN	55
5.1 Kesimpulan	55
5.2 Saran	56
DAFTAR PUSTAKA	57
LAMPIRAN.....	60
A. Bagian Pengalaman Pengguna	64
B. Kualitas Rekomendasi	66

DAFTAR TABLE

Tabel 2. 1 Hasil Penelitian Terdahulu.....	8
Tabel 2. 2 Confution matrix.....	17
Tabel 3. 1 Data novel.....	24
Tabel 3. 2 contoh duplikat data.....	25
Tabel 3. 3 Dokumen Teks TF-IDF.....	28
Tabel 3. 4 RingkasanTF-IDF (Untuk D3).....	29
Tabel 3. 5 Skenario Pengujian.....	37
Tabel 4. 1 Hasil Pengujian Black Box.....	52



DAFTAR GAMBAR

Gambar 2. 1 visualisasi metode content-based filtering.....	14
Gambar 2. 2 Ilustrasi Algoritma random forest.....	15
Gambar 2. 3 Kerangka Berpikir.....	20
Gambar 3. 1 Alur diagram penelitian.....	21
Gambar 3. 2 Alur diagram Sistem Rekomendasi	22
Gambar 3. 3 Karakteristik Data.....	24
Gambar 3. 4 Visualisasi Data Frame.....	26
Gambar 3. 5 Visualisasi Standardisasi kolom dan pengisian nilai.....	27
Gambar 3. 6 Visual Content-based filtering.....	30
Gambar 3. 7 Visualisasi pohon keputusan	32
Gambar 3. 8 Alur Diagram Deployment.....	35
Gambar 3. 9 Alur Diagram Sistem	36
Gambar 4. 1 Cek Duplikat Data.....	39
Gambar 4. 2 Standardisasi Nama.....	39
Gambar 4. 3 Hasil mengubah data csv ke dataframe.....	40
Gambar 4. 4 Isi Data Kosong Tanpa Menghapus Baris (fillna)	40
Gambar 4. 5 Info Data.....	41
Gambar 4. 6 Latih Random forest.....	42
Gambar 4. 7 TF-IDF.....	42
Gambar 4. 8 Evaluasi Confusion Metrix.....	43
Gambar 4. 9 Visual distribusi novel.....	47
Gambar 4. 10 Visual Distribusi mulai.....	48
Gambar 4. 11 Visual Distribusi tahun selesai.....	49
Gambar 4. 12 Tampilan Home.....	50
Gambar 4. 13 Tampilan Rekomendasi Novel Berdasarkan Score/Rating.....	51
Gambar 4. 14 Tampilan Rekomendasi Novel Berdasarkan genre.....	51
Gambar 4. 15 Tampilan Distribusi Genre, Status Novel, dan tahun.....	52

DAFTAR LAMPIRAN

LAMPIRAN 1 HASIL PENGUJIAN BLACKBOX.....	60
LAMPIRAN 2 HASIL DATA PENGALAMAN PENGGUNA.....	64



BAB I

PENDAHULUAN

1.1 Latar Belakang

Novel adalah salah satu bentuk karya sastra berupa prosa naratif panjang yang menceritakan rangkaian peristiwa yang dialami oleh tokoh-tokoh tertentu, dengan struktur cerita yang kompleks mencakup pengembangan karakter, latar, konflik, dan tema. Sejarah novel dimulai sejak abad ke-18 dengan karya-karya awal seperti *Don Quixote* oleh Miguel de Cervantes dan *Robinson Crusoe* oleh Daniel Defoe, yang dianggap sebagai tonggak awal genre ini. Seiring waktu, novel berkembang menjadi media penting untuk menyampaikan pandangan sosial, budaya, hingga psikologis masyarakat dari berbagai zaman. Jepang menganut penulisan novel yang ringan dan mudah diakses mulai populer pada akhir abad ke-20, bersamaan dengan meningkatnya minat terhadap budaya pop seperti anime dan manga[1]. Jenis novel seperti ini mulai mendapat perhatian pada awal 2000-an di Indonesia, mendorong banyak penulis lokal untuk menciptakan karya dengan gaya serupa yang menekankan pada cerita yang menarik, ilustrasi pendukung, serta kedekatan dengan pembaca remaja dan dewasa muda. Kemunculan novel digital sebagai bentuk modern dari novel tradisional memberikan berbagai keunggulan, seperti *aksesibilitas* tinggi, format ringkas dan interaktif. Mayoritas pengguna menunjukkan bahwa pembaca novel digital di Indonesia umumnya berasal dari kalangan remaja dan dewasa muda, dengan ketertarikan pada genre fantasi, petualangan, dan romansa. Mereka umumnya menggunakan platform seperti WebNovel dan Webtoon, serta aktif dalam komunitas daring untuk berbagi rekomendasi dan ulasan[2]. Pada penelitian ini, sistem rekomendasi pada platform pembaca novel menjadi sangat relevan karena membantu pengguna menemukan karya yang sesuai dengan minat pengguna, sehingga dapat meningkatkan pengalaman membaca sekaligus mendorong perkembangan ekosistem penulisan novel digital di Indonesia[3].

Sistem rekomendasi pada dasarnya bertujuan untuk mempermudah pengguna dalam menemukan item yang relevan di tengah banyaknya pilihan yang tersedia. Salah satu pendekatan utama yang sering digunakan dalam sistem

rekomendasi adalah *content-based filtering* [4]. Pendekatan ini menekankan pada karakteristik konten dari item yang disukai pengguna, seperti genre, penulis, atau tema novel. Dengan menganalisis preferensi pengguna terhadap konten-konten sebelumnya, sistem dapat merekomendasikan novel lain yang memiliki kemiripan karakteristik. Pendekatan ini sangat berguna dalam membantu pengguna menemukan novel yang sesuai dengan minat dan selera mereka, terutama ketika data interaksi dengan pengguna lain masih terbatas[5].

Penerapan metode *content-based filtering* dalam sistem rekomendasi novel pada platform digital dilakukan dengan menganalisis karakteristik dari novel-novel yang sebelumnya disukai atau dibaca oleh pengguna, seperti genre, penulis, atau tema cerita. Berdasarkan data ini, sistem dapat merekomendasikan novel-novel lain yang memiliki kemiripan dalam konten, meskipun pengguna belum pernah membaca karya-karya tersebut sebelumnya[6]. Untuk meningkatkan akurasi dan kinerja sistem rekomendasi, algoritma *Random forest* dapat dimanfaatkan sebagai alat prediksi preferensi pengguna berdasarkan atribut konten dari novel yang telah diinteraksikan, seperti genre yang sering dibaca, waktu membaca, serta tingkat keterlibatan pengguna[7].

Pemanfaatan struktur algoritma *random forest* yang membagi data ke dalam banyak pohon keputusan berdasarkan berbagai atribut memungkinkan sistem untuk mengidentifikasi pola-pola spesifik dalam perilaku pengguna yang berkontribusi terhadap keputusan membaca mereka. Hal ini menghasilkan rekomendasi yang lebih personal dan relevan. Kelebihan dari *content-based filtering* termasuk kemampuannya untuk memberikan rekomendasi berdasarkan karakteristik konten dari item yang disukai pengguna, seperti genre, penulis, atau tema, tanpa bergantung pada preferensi pengguna lain[8]. Selain itu, metode ini mampu menangani masalah *cold start* untuk pengguna baru karena hanya memerlukan data interaksi dari satu pengguna saja[9]. Namun, pendekatan ini juga memiliki keterbatasan, seperti risiko menghasilkan rekomendasi yang sempit atau terlalu mirip dengan item yang sudah pernah dikonsumsi sebelumnya. Sementara itu, algoritma *random forest* dapat digunakan untuk menganalisis data konten dan interaksi pengguna secara lebih mendalam, serta mengidentifikasi pola dan faktor-faktor yang memengaruhi preferensi membaca. Dengan mengintegrasikan kedua

pendekatan ini, sistem rekomendasi tidak hanya dapat memberikan hasil yang lebih relevan dan personal, tetapi juga meningkatkan akurasi dalam memprediksi *preferensi* pengguna, sehingga menciptakan pengalaman membaca yang lebih memuaskan di *platform digital* [10].

Sistem rekomendasi yang baik sangat penting untuk membantu menemukan novel yang sesuai dengan minat pengguna, namun banyak pengguna merasa bahwa rekomendasi yang diberikan sering kali tidak memadai dan tidak sesuai dengan selera mereka, kemungkinan disebabkan oleh algoritma yang belum sepenuhnya memahami *preferensi* unik setiap pengguna [11]. Untuk mengembangkan sistem rekomendasi novel yang lebih baik, peneliti dapat mengikuti beberapa langkah menggunakan data yang telah disiapkan, termasuk *file* novel yang berisi 4.500 judul dengan atribut seperti ID, judul, sinopsis, penulis, genre, volume, status, skor, popularitas, favorit, tahun penerbitan, dan tahun selesai. Setelah mengumpulkan data, langkah selanjutnya adalah membersihkan dan mengubah data menjadi format yang konsisten, termasuk menghapus duplikasi dan melakukan normalisasi pada rating.

Metode yang digunakan adalah *content-based filtering*, yang memberikan rekomendasi berdasarkan kesamaan atribut konten dari novel yang telah disukai atau dibaca pengguna sebelumnya, seperti genre, penulis, atau tema cerita [12]. Dengan pendekatan ini, sistem dapat menyarankan novel yang memiliki karakteristik serupa dengan preferensi individu pengguna, tanpa bergantung pada interaksi pengguna lain. Selain itu, algoritma *Random Forest* dapat digunakan untuk memberikan rekomendasi kepada pengguna berdasarkan atribut yang relevan, dengan membangun sejumlah pohon keputusan dan menggabungkan hasilnya untuk menghasilkan saran yang lebih akurat dan andal. Algoritma ini bekerja dengan membagi data menjadi subset berdasarkan nilai atribut yang signifikan, lalu mengombinasikan hasil dari setiap pohon untuk menentukan rekomendasi akhir [13].

Setelah model dikembangkan, penting untuk menguji kinerjanya menggunakan *Confusion Matrix*, yang membantu mengevaluasi performa model klasifikasi dengan membandingkan hasil rekomendasi terhadap preferensi aktual pengguna, serta menghitung metrik seperti *akurasi*, *precision*, *recall*, dan *F1-score*.

Dengan mengikuti langkah-langkah ini, kita dapat membangun sistem rekomendasi novel yang lebih tepat sasaran dan sesuai dengan minat pengguna, sehingga dapat meningkatkan pengalaman membaca di platform digital [14].

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah yang diajukan penulis, yaitu:

1. Bagaimana menerapkan metode *Content-based filtering* untuk sistem rekomendasi novel di platform digital menggunakan Algoritma *Random forest*?
2. Sejauh mana algoritma *Random Forest* dapat diterapkan dalam sistem rekomendasi novel untuk merekomendasi *preferensi* berdasarkan data yang terkumpul dari interaksi pengguna?

1.3 Batasan Masalah

Agar penelitian ini mencapai pada sasaran yang diinginkan, maka penulis membatasi permasalahan dalam penelitian ini sebagai Berikut:

1. Jenis Penelitian hanya fokus pada sistem rekomendasi menggunakan metode Content-Based Filtering, tanpa membahas metode seperti collaborative filtering atau hybrid filtering.
2. Algoritma yang digunakan dalam penelitian ini terbatas pada Random Forest untuk proses klasifikasi.
3. Objek rekomendasi dibatasi hanya pada novel digital, tidak mencakup film, musik, atau media lain.
4. Dataset yang digunakan hanya berasal dari file novel.csv dengan 4500 entri dari platform seperti Kaggle.
5. Fitur yang dianalisis terbatas pada judul, genre, skor/rating, popularitas, dan penulis novel.
6. Evaluasi sistem hanya menggunakan metrik akurasi, precision, recall, dan F1-score.
7. Data pengguna bersifat tidak personal dan tidak mencakup faktor eksternal seperti lokasi geografis atau waktu interaksi.

8. Sistem rekomendasi tidak menggunakan interaksi pengguna secara real-time atau riwayat perilaku secara mendalam.
9. Platform implementasi hanya berupa web aplikasi menggunakan Streamlit, bukan versi mobile.
10. Sistem tidak menggunakan data citra, hanya data teks untuk representasi dan analisis konten.
11. Penanganan data kosong hanya dilakukan dengan metode fillna(mean), tidak mencakup teknik imputasi kompleks lainnya.
12. Kemiripan konten novel hanya dihitung menggunakan metode TF-IDF dan cosine similarity.
13. Model klasifikasi dibangun tanpa eksplorasi model pembandingan lain seperti SVM atau KNN.
14. Sistem tidak melakukan analisis sentimen terhadap isi novel atau review pembaca.
15. Sistem rekomendasi tidak melakukan pembobotan preferensi berdasarkan frekuensi atau durasi membaca pengguna.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini sebagai berikut:

1. Menerapkan metode Content-Based Filtering untuk membangun sistem rekomendasi novel berdasarkan kemiripan konten seperti genre dan judul.
2. Mengintegrasikan algoritma Random Forest untuk mengklasifikasikan popularitas novel guna meningkatkan akurasi rekomendasi.
3. Mengolah dan mengekstraksi fitur teks menggunakan TF-IDF dan cosine similarity sebagai dasar perhitungan kesamaan antar novel.
4. Membangun aplikasi sistem rekomendasi berbasis web dengan Streamlit agar pengguna dapat mengakses rekomendasi secara interaktif.
5. Mengevaluasi performa sistem menggunakan confusion matrix untuk memperoleh nilai akurasi, precision, recall, dan F1-score dari model yang dibangun.

1.5 Manfaat penelitian

Manfaat yang diharapkan oleh peneliti dalam penelitian ini adalah sebagai berikut:

1. Bagi peneliti

- 1) Menambah wawasan mengenai implementasi sistem rekomendasi dalam platform pembaca novel.
- 2) Memberikan pengalaman dalam mengembangkan dan mengevaluasi algoritma *content-based filtering*, serta *Random Forest*.
- 3) Dapat menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan sistem rekomendasi berbasis kecerdasan buatan.

2. Bagi Kampus

- 1) Mendukung pengembangan keilmuan di bidang data *science*, kecerdasan buatan, dan sistem rekomendasi.
- 2) Menjadi bahan referensi akademik bagi mahasiswa dan dosen dalam penelitian terkait sistem rekomendasi.
- 3) Meningkatkan kontribusi institusi dalam inovasi teknologi yang dapat diterapkan dalam berbagai bidang, termasuk literasi digital.

3. Bagi Pengguna Platform Pembaca Novel

- 1) Mempermudah pengguna dalam menemukan novel yang sesuai dengan preferensi mereka.
- 2) Meningkatkan pengalaman membaca dengan rekomendasi yang lebih personal dan relevan.
- 3) Menghemat waktu dalam mencari bacaan yang diminati, sehingga meningkatkan keterlibatan dan kepuasan pengguna terhadap platform tersebut.

1.6 Sistematika Penulisan

Untuk memudahkan bagi pembaca dalam menganalisa dan memahami hasil dari penelitian yang dilakukan penulis, maka penulis membuat suatu sistematika penulisan yang dibagi atas beberapa bab sebagai Berikut:

BAB I: PENDAHULUAN

Berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II: TINJAUAN PUSTAKA

Membahas penelitian terkait dan kerangka pemikiran yang mendasari penelitian ini.

BAB III: METODOLOGI PENELITIAN

Membahas tahapan penelitian, metode pengumpulan data, dan skenario pengujian yang digunakan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas tentang implementasi teknik penyaringan berbasis konten dan algoritma tetangga terdekat dalam sistem rekomendasi.

BAB V KESIMPULAN DAN SARAN

Dalam bagian ini akan dibahas mengenai ringkasan dari studi yang telah dilaksanakan.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai penerapan algoritma Random Forest pada sistem rekomendasi musik menggunakan pendekatan content-based filtering, maka dapat disimpulkan beberapa hal sebagai berikut:

1. Penerapan metode Content-Based Filtering dalam sistem rekomendasi novel dilakukan dengan menganalisis atribut konten dari novel, seperti judul, genre, skor/rating, dan popularitas. Pertama-tama, data novel dikumpulkan, dibersihkan, dan diproses ke dalam bentuk numerik menggunakan teknik TF-IDF untuk judul. Kemudian, cosine similarity digunakan untuk mengukur kemiripan antar judul novel.

Agar rekomendasi tidak hanya relevan dari sisi konten, digunakan algoritma Random Forest untuk memprediksi popularitas dari novel berdasarkan skor. Hasil dari prediksi model tersebut dikombinasikan dengan skor kemiripan (hasil TF-IDF) untuk menghasilkan daftar rekomendasi akhir yang bersifat personal, akurat, dan relevan dengan preferensi pengguna.

2. Berdasarkan hasil evaluasi model (Bab IV), algoritma Random Forest Classifier menunjukkan kinerja yang sangat baik dalam mengklasifikasikan novel populer dan tidak populer. Hasil evaluasi menunjukkan hasil Accuracy: 94.2% ,Precision: 78.3% ,Recall: 87.4% ,F1-Score: 82.6% Angka-angka tersebut mengindikasikan bahwa Random Forest mampu mengidentifikasi pola dari atribut seperti genre dan skor untuk menilai apakah sebuah novel berpotensi disukai oleh pengguna atau tidak. Bahkan tanpa riwayat interaksi yang luas, model ini bisa mengenali dan memprediksi preferensi pengguna secara efisien. Penelitian ini dengan demikian, algoritma Random Forest sangat cocok diterapkan dalam sistem rekomendasi berbasis content, khususnya saat digunakan bersama metode TF-IDF untuk menghasilkan rekomendasi novel

yang tidak hanya relevan dari sisi konten, tetapi juga diprediksi populer secara statistik.

5.2 Saran

Berdasarkan hasil dan keterbatasan penelitian ini, penulis memberikan beberapa saran sebagai berikut:

1. Pengembangan Fitur Eksternal Untuk meningkatkan kualitas rekomendasi, sistem dapat dikembangkan lebih lanjut dengan menambahkan fitur interaksi pengguna secara eksplisit, seperti riwayat pembacaan, waktu baca, dan penilaian langsung terhadap novel.
2. Penggunaan Metode Hybrid Penelitian selanjutnya disarankan menggabungkan metode Content-Based Filtering dengan Collaborative Filtering (metode hybrid) untuk menangani keterbatasan dari masing-masing pendekatan dan meningkatkan keberagaman rekomendasi.
3. Evaluasi dari Sisi Pengguna Nyata Selain evaluasi teknis menggunakan confusion matrix, perlu dilakukan pengujian langsung oleh pengguna (user testing) untuk menilai tingkat kepuasan dan kenyamanan dalam menggunakan sistem rekomendasi yang dibangun.
4. Perluasan Dataset dan Bahasa Dataset dapat diperluas untuk mencakup lebih banyak novel dari berbagai genre dan bahasa, sehingga sistem menjadi lebih inklusif dan dapat diterapkan di berbagai platform dan skenario pengguna.

DAFTAR PUSTAKA

- [1] J. Bahasa, D. Budaya, M. Umam Kau, and A. Hidayanti Ali, “ANALISIS PENDEKATAN PSIKOLOGI SASTRA DALAM NOVEL RE: DAN PEREMPUAN Eka sartika 1*,” vol. 12, no. 2, p. 2022, 2022, [Online]. Available: <http://ejurnal.ung.ac.id/index.php/JBSP/index>
- [2] B. Soni, D. Thakuria, N. Nath, N. Das, and B. Boro, “RikoNet: A Novel Anime Recommendation Engine,” Jun. 2021, [Online]. Available: <http://arxiv.org/abs/2106.12970>
- [3] G. Adomavicius, K. Bauman, A. Tuzhilin, and M. Unger, “Context-Aware Recommender Systems: From Foundations to Recent Developments,” in *Recommender Systems Handbook*, New York, NY: Springer US, 2022, pp. 211–250. doi: 10.1007/978-1-0716-2197-4_6.
- [4] E. Salim, J. Pragantha, and M. D. Lauro, “Perancangan Sistem Rekomendasi Film menggunakan metode Content-based Filtering,” 2022.
- [5] D. A. Putri, D. Pramesti, D. I., and W. Santiyasa, “Penerapan Metode Content-Based Filtering dalam Sistem Rekomendasi Video Game,” 2022.
- [6] B. MULYAWAN SKom and T. SUTRISNO SSi, “Jurnal Ilmu Komputer dan Sistem Informasi PEMBUATAN APLIKASI E-COMMERCE BERBASIS WEB DENGAN FITUR REKOMENDASI MENGGUNAKAN METODE CONTENT-BASED FILTERING,” 2021.
- [7] Y. I. Fajarendra, Y. R. Fauzan, and S. ' Uyun, “KLASIFIKASI CITRA EUROSAT MENGGUNAKAN ALGORITMA KNN, DECISION TREE DAN RANDOM FOREST,” 2024.
- [8] Reynaldi and W. Istiono, “Content-based Filtering and Web Scraping in Website for Recommended Anime,” *Asian Journal of Research in Computer Science*, vol. 15, no. 2, pp. 32–42, Mar. 2023, doi: 10.9734/ajrcos/2023/v15i2318.
- [9] K. R. PUTRA and M. A. RACHMAN, “Perbandingan Metode Content-based, Collaborative dan Hybrid Filtering pada Sistem Rekomendasi Lagu,” *MIND Journal*, vol. 9, no. 2, pp. 179–193, Dec. 2024, doi: 10.26760/mindjournal.v9i2.179-193.
- [10] S. Diajukan, S. Persyaratan, U. Memperoleh, G. S. Kom, G. H. Al, and R. Noer, “IMPLEMENTASI ALGORITMA NAÏVE BAYES DAN TF-IDF DALAM ANALISIS SENTIMEN DATA ULASAN (Studi Kasus: Ulasan Review Aplikasi E-Commerce Shopee Di Situs Google Playstore),” 2020.
- [11] R. Arya Permana, “Sistem Rekomendasi Laptop Berbasis Website Menggunakan Metode Content Based Filtering,” 2021.

- [12] R. Faurina, E. Sitanggang, J. W. Supratman, and K. Limun, "Implementasi Metode Content-Based Filtering dan Collaborative Filtering pada Sistem Rekomendasi Wisata di Bali Implementation of Content-Based Filtering and Collaborative Filtering Method in the Tourism Recommendation System in Bali," 2023.
- [13] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 640, Apr. 2021, doi: 10.30865/mib.v5i2.2937.
- [14] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, pp. 4023–4031, Nov. 2024, doi: 10.53555/AJBR.v27i4S.4345.
- [15] A. Zakharia *et al.*, "Sistem Rekomendasi Film Indonesia Menggunakan Metode Content-Based Filtering," 2024, [Online]. Available: <https://journal.mediapublikasi.id/index.php/logic>
- [16] M. Zuhair, "SKRIPSI SISTEM REKOMENDASI LAGU DI SPOTIFY BERDASARKAN KEMIRIPAN LIRIK MENGGUNAKAN METODE CONTENT BASED FILTERING Oleh," 2024.
- [17] M. Johari and A. Laksito, "The Hybrid Recommender System of the Indonesian Online Market Products using IMDb weight rating and TF-IDF," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 977–983, Oct. 2021, doi: 10.29207/resti.v5i5.3486.
- [18] G. P. Insany, P. Somantri, and A. Putri, "Implementasi Sistem Rekomendasi dengan Content Based Filtering dan Teknologi Virtual Tour Untuk Strategi Pemasaran Pada Website," *Technology and Science (BITS)*, vol. 6, no. 1, pp. 300–313, 2024, doi: 10.47065/bits.v6i1.5358.
- [19] G. V. R. Pake, V. C. Mawardi, and T. Sutrisno, "PENERAPAN METODE CONTENT BASED FILTERING DALAM IMPLEMENTASI SISTEM REKOMENDASI MUSIK," *Jurnal Serina Sains, Teknik dan Kedokteran*, vol. 1, no. 2, pp. 455–462, Oct. 2023, doi: 10.24912/jsstk.v1i2.31037.
- [20] M. Rizqi, A. Zayyad, A. Kurniawardhani, S. Si, and M. Kom, "Penerapan Metode Deep Learning pada Sistem Rekomendasi Film," 2020.
- [21] F. A. Rasyid and Y. Sibaroni, "Sistem Rekomendasi Anime Menggunakan Content-Based Filtering Berbasis Website," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 2022.
- [22] S. M. Iqbal and I. Pratama, "Penerapan Metode Content Based Filtering pada Sistem Rekomendasi Pemilihan Produk Skincare," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 13, no. 3, Jul. 2024, doi: 10.30591/smartcomp.v13i3.7156.

- [23] Y. I. Fajarendra, Y. R. Fauzan, and S. ' Uyun, "KLASIFIKASI CITRA EUROSAT MENGGUNAKAN ALGORITMA KNN, DECISION TREE DAN RANDOM FOREST," 2024.
- [24] M. Farhan, M. Rifky Andreawan, R. Antonius, and N. Dhiya Ulhag, "Biner : Jurnal Ilmu Komputer, Teknik dan Multimedia Evaluasi Dan Pengembangan Sistem Rekomendasi Game Berbasis Content-Based Filtering Dengan TF-IDF Dan Cosine Similarity," 2020.
- [25] A. Akbar, S. Pratama, J. K. A Dahlan NoKM, K. Pangkal Pinang, and K. Bangka Belitung, "Sistem Rekomendasi Program Studi Sarjana Berbasis Machine Learning Untuk Model Klasifikasi Calon Mahasiswa Baru," *Journal of Information Technology and society (JITS)*, vol. 1, no. 1, p. 11, 2023, doi: 10.35870/jti.
- [26] R. Arya Permana, "Sistem Rekomendasi Laptop Berbasis Website Menggunakan Metode Content Based Filtering," 2023.
- [27] Z. Fadilla, P. Muhammad, Z. Penerbit, M. Zaini, K. A. Lawang, and M. Jannah, "METODOLOGI PENELITIAN KUANTITATIF," 2023. [Online]. Available: <https://www.researchgate.net/publication/370561251>
- [28] N. Sarangpure, V. Dhamde, A. Roge, J. Doye, S. Patle, and S. Tamboli, "Automating the Machine Learning Process using PyCaret and Streamlit," in *2023 2nd International Conference for Innovation in Technology, INOCON 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/INOCON57975.2023.10101357.
- [29] G. Ayu Syafarina and I. Kalimantan Muhammad Arsyad Al Banjari Banjarmasin, "Implementasi Framework Streamlit Sebagai Prediksi Harga Jual Rumah Dengan Linear Regresi," vol. 7, p. 2023, 2023, doi: 10.47002/metik.v7i2.680.
- [30] G. Ubale, B. Deore, S. Deogade, S. Deogade, S. Deokar, and S. Deokate, "ANI-DB (An Enhanced Anime Database Website)," *Int J Res Appl Sci Eng Technol*, vol. 11, no. 11, pp. 2081–2086, Nov. 2023, doi: 10.22214/ijraset.2023.57012.