Periode	: Semester Ganjil
Tahun	: 2023/2024
Skema Peneltian	: Mandiri
Tema RIP Penelitian	: Data Engineering

LAPORAN AKHIR
PENELITIAN MANDIRI

PENGEMBANGAN METODE FILTRASI KOLABORATIF UNTUK PENINGKATAN SISTEM REKOMENDASI

OLEH:
RIYA WIDAYANTI (03110877001)
MOCHAMAD HERU RIZA CHAKIM (8985630021)
CHANDRA LUKITA (0409056401)
MUHAMMAD RAMDHAN AWALI (20210803054)
JACK VALLEN (20210803018)
DEWI ANDINI (20210803024)
REZA MAHENDRA (20210803066)
DURRA NABILA (20210803030)

FAKULTAS ILMU KOMPUTER/SISTEM INFORMASI
UNIVERSITAS ESA UNGGUL

TAHUN 2024

Halaman Pengesahan Laporan Akhir
Hibah Penelitian Mandiri
Universitas Esa Unggul

1.	Judul Kegiatan Penelitian	: PENGEMBANGAN METODE FILTRASI KOLABORATIF UNTUK PENINGKATAN SISTEM REKOMENDASI
2.	Nama Mitra Sasaran	: E-Commerce
3.	Ketua Tim	:
	a. Nama	: Riya Widayanti
	b. NIDN	: 0311087701
	c. Jabatan Fungsional	: Lektor 200
	d. Fakultas	: Ilmu Komputer
	e. Bidang Keahlian	: Sistem Informasi
	f. Email	: riya.widayanti@esaunggul.ac.id
4.	Jumlah Anggota Dosen	: 2 Orang
5.	Julah Anggota Mahasiswa	: 5 Orang
6.	Lokasi Kegiatan Mitra	
	Alamat	: E-Commerce
	Kota	: Jakarta Barat dan Tangerang
	Propinsi	: DKI Jakarta dan Banten
7.	Periode/Waktu Kegiatan	: 7 Bulan
8.	Luaran yang dihasilkan	: Jurnal Terindeks Scopus
9.	Usulan/Realisasi Anggaran	
	a. Dana Internal	: -
	b. Sumber Dana lain	: …….
	c. Sumber Dana lain	: ……..

Menyetujui, Dekan Fakultas	Pengusul, Ketua Tim Pelaksana
Universitas Esa Unggul FAKULTAS ILMU KOMPUTER	
Dr. Vitri Tundjungsari, ST, MSc 0301057601	Riya Widayanti, S.Kom, MMSI 0311087701

Mengetahui,
Kepala LPPM

Universitas Esa Unggul
LPPM

Laras Sitoayu, SGz, MKM
0320128406

Abstrak

Peningkatan pada sistem rekomendasi melalui integrasi sinergis antara teknik Collaborative Filtering (CF) dan Content-Based Filtering (CBF), yang dikenal sebagai pendekatan hibrid CF-CBF. Pendekatan ini menggabungkan keunggulan wawasan interaksi pengguna CF dan kecakapan analisis konten CBF, menghasilkan paradigma rekomendasi yang lebih halus dan personal. Riset ini meliputi tahapan akuisisi data, pengelolaan penyimpanan, penyempurnaan data, dengan menerapkan metodologi CF dan CBF. Hasil dari penelitian adalah keunggulan pendekatan hibrid dalam menghasilkan rekomendasi yang menunjukkan keragaman dan presisi yang ditingkatkan, melampaui hasil yang diperoleh dari masing-masing teknik secara terpisah. Pendekatan hibrid CF-CBF efektif mengatasi kekurangan metode individu seperti kerentanan CF terhadap masalah "cold start" dan keterbatasan CBF dalam mendorong keragaman rekomendasi. Selanjutnya, interaksi CF dan CBF meningkatkan pemahaman sistem rekomendasi terhadap preferensi pengguna, sehingga memperkaya kualitas rekomendasi yang diberikan. Kesimpulannya, riset ini merupakan kontribusi perintis dalam evolusi sistem rekomendasi dengan memajukan pendekatan hibrid CF-CBF, menyatukan dua teknik berbeda untuk menciptakan terobosan dalam rekomendasi personal.

Keyword: Recommendation System, Collaborative Filtering, Content-Based Filtering, Hybrid, Diversification, Accuracy

BAB 1.
Pendahuluan

1.1. Latar Belalakang

Penelitian ini didasarkan pada kebutuhan sistem rekomendasi dalam dunia digital saat ini, di mana pengguna sering dihadapkan pada jumlah informasi yang tidak terbatas. Dalam hal ini, sistem rekomendasi sangat penting untuk membantu pengguna menemukan konten yang sesuai dengan preferensi mereka. Meskipun demikian, metode konvensional yang digunakan dalam sistem rekomendasi, seperti filtrasi kolaboratif (CF) dan filtrasi berbasis konten (CBF), masing-masing memiliki kekurangan. Sementara CBF berfokus pada karakteristik konten itu sendiri, CF mengandalkan interaksi pengguna dan biasanya mengalami kesulitan dalam menangani situasi "cold start", di mana data tentang preferensi pengguna sangat terbatas. Namun, CBF seringkali gagal dalam merekomendasikan item yang beragam dan baru di luar preferensi historis pengguna.

Latar belakang artikel ini membahas bagaimana mengatasi keterbatasan ini dengan cara yang lebih integratif. Penulis meneliti solusi hibrid yang menggabungkan kekuatan CF dan CBF. Tujuan utama penelitian ini adalah untuk mengembangkan sistem rekomendasi yang lebih canggih dan efisien yang dapat memberikan rekomendasi yang lebih akurat, relevan, dan personal kepada pengguna. Dengan menggunakan pendekatan hibrid ini, penulis berharap dapat menghasilkan rekomendasi yang lebih komprehensif dan beragam.

Penelitian ini muncul dari kebutuhan untuk meningkatkan kinerja sistem rekomendasi untuk mengatasi kesulitan dan keterbatasan metode konvensional dan untuk membuat pendekatan yang lebih fleksibel dan responsif untuk memenuhi kebutuhan pengguna dalam dunia digital yang terus berubah.

1.2. Identifikasi Masalah

Beberapa masalah utama yang mungkin dibahas dalam artikel tersebut:

1. Masalah "*Cold Start*" dalam *Collaborative Filtering (CF):* Sistem rekomendasi yang menggunakan CF bergantung pada data interaksi pengguna yang ada. Ketika data ini tidak tersedia, misalnya untuk item baru atau pengguna baru, sistem menghadapi kesulitan untuk memberikan rekomendasi yang akurat.

2. Keterbatasan dalam Content-Based Filtering (CBF): Meskipun CBF mengatasi beberapa keterbatasan CF dengan menganalisis konten item, pendekatan ini sering kali tidak mampu merekomendasikan item yang beragam dan gagal mengeksplorasi preferensi baru pengguna di luar apa yang telah diungkapkan sebelumnya.
3. Kebutuhan akan Pendekatan Hibrid yang Lebih Efektif: Mengingat kekurangan dalam CF dan CBF saat digunakan secara terpisah, artikel tersebut kemungkinan menyoroti kebutuhan untuk strategi yang lebih integratif yang menggabungkan kekuatan kedua metode ini untuk menghasilkan rekomendasi yang lebih presisi dan beragam.
4. Peningkatan Personalisasi dalam Sistem Rekomendasi:Salah satu tantangan utama dalam sistem rekomendasi adalah bagaimana menyajikan rekomendasi yang tidak hanya akurat tetapi juga sangat personal, memenuhi kebutuhan unik setiap pengguna.
5. Efisiensi dalam Pengelolaan dan Pengolahan Data: Mungkin juga ada diskusi tentang bagaimana mengelola dan memproses data besar secara efisien untuk menghasilkan rekomendasi yang cepat dan akurat.

1.3. Tujuan Penelitian

Tujuan Penelitilianya adalah sebagai berikut:

1. Mengembangkan Sistem Rekomendasi yang Lebih Efisien: Tujuan utama dari penelitian ini adalah untuk meningkatkan efektivitas sistem rekomendasi dengan menggabungkan keunggulan dari dua teknik populer, yaitu Collaborative Filtering (CF) dan Content-Based Filtering (CBF).

2. Mengatasi Keterbatasan Teknik Tradisional: untuk mengatasi keterbatasan yang ada dalam metode CF dan CBF ketika digunakan secara terpisah, seperti masalah "cold start" dalam CF dan keterbatasan dalam merekomendasikan konten yang beragam dalam CBF.

3. Menciptakan Rekomendasi yang Lebih Personal dan Beragam: Adapun salah satu tujuan khusus adalah untuk meningkatkan personalisasi dan keragaman dalam rekomendasi yang diberikan kepada pengguna, sehingga meningkatkan kepuasan pengguna dan efektivitas sistem rekomendasi.

4. Menyediakan Solusi Hibrid yang Inovatif: Tujuan lainnya adalah untuk mengeksplorasi dan mengembangkan solusi hibrid yang memanfaatkan kekuatan dari kedua pendekatan (CF dan CBF) dalam menghasilkan rekomendasi yang lebih akurat dan relevan.

5. Meningkatkan Pemahaman terhadap Preferensi Pengguna: Penelitian ini juga bertujuan untuk meningkatkan pemahaman sistem terhadap preferensi pengguna, memungkinkan sistem untuk memberikan rekomendasi yang lebih sesuai dengan kebutuhan dan minat individu.

BAB 2.
Literatur Review

2.1 Konsep *Collaborative Filtering*

Salah satu konsep penting dalam sistem rekomendasi adalah teori Collaborative Filtering (CF), yang berfokus pada penggunaan data perilaku pengguna (seperti peringkat atau preferensi) untuk membuat prediksi atau rekomendasi untuk pengguna lain. Penjelasan lebih lanjut tentang teori CF adalah sebagai berikut:

Dasar Konseptual CF: CF didasarkan pada prinsip bahwa orang-orang yang memiliki preferensi serupa di masa lalu cenderung memiliki preferensi serupa di masa depan juga. Dalam kasus di mana dua pengguna memiliki persepsi yang sebanding tentang beberapa item di masa lalu, mereka mungkin memiliki persepsi yang sebanding tentang item lain di masa depan.

2.1.1 CF Berbasis Pengguna dan Item: CF memiliki dua pendekatan utama:

User-Based CF: Teknik ini menemukan pengguna yang "serupa" dan memprediksi peringkat pengguna aktif berdasarkan peringkat mereka. Item-Based CF: Metode ini berfokus pada kesamaan antara item; jika seorang pengguna menyukai suatu item, sistem akan merekomendasikan item lain yang sebanding dengannya berdasarkan peringkat semua pengguna.

Menghitung Kesamaan: Menghitung kesamaan antara pengguna atau item adalah komponen penting dari CF. Metode seperti korelasi Pearson, cosine similarity, dan sebagainya dapat digunakan untuk mencapai ini.

Masalah "Start Cold": Salah satu masalah utama dalam CF adalah masalah "start cold", di mana sulit untuk membuat saran yang akurat untuk pengguna baru atau item baru yang belum memiliki cukup data interaksi.

Manfaat CF: Kemampuan untuk memberikan rekomendasi yang dipersonalisasi tanpa memerlukan konten atau atribut item yang jelas adalah keunggulan utama CF. Ini sangat bermanfaat dalam situasi di mana data konten sulit diperoleh atau tidak lengkap.

Pembelajaran Mesin dalam CF: Algoritma CF kontemporer sering menggunakan teknik seperti pembelajaran berbasis model (seperti faktorisasi matriks) untuk membuat prediksi yang lebih akurat karena kemajuan dalam pembelajaran mesin.

2.2 Konsep CBF

- Fokus pada Konten Item: CBF menyarankan untuk menganalisis fitur atau atribut item yang dihadapi, seperti teks deskriptif, metadata, dan atribut lainnya. Genre, sutradara, atau aktor dalam film, atau kategori dan deskripsi produk dalam e-commerce adalah contoh umum.

- Profil Pengguna: Sistem CBF membuat profil pengguna berdasarkan item yang pengguna nilai atau interaksikan. Profil ini menunjukkan preferensi pengguna terhadap berbagai fitur atau atribut item.

- Algoritma Pembelajaran Mesin: CBF sering menggunakan teknik pembelajaran mesin untuk menemukan dan mempelajari fitur item yang disukai atau tidak disukai oleh pengguna. Klasifikasi, regresi, dan metode pembelajaran mesin lainnya adalah beberapa contoh teknik pembelajaran mesin.

- Rekomendasi Berbasis Fitur: Setelah profil pengguna dibuat, sistem CBF merekomendasikan item baru dengan fitur yang sebanding dengan fitur yang telah dinilai positif oleh pengguna. Rekomendasi ini didasarkan pada perbandingan antara fitur item dan preferensi yang diungkapkan dalam profil pengguna.

- Keuntungan CBF: Kelebihan utama CBF adalah kemampuannya untuk memberikan rekomendasi yang sangat spesifik dan individual berdasarkan preferensi pengguna. Ini sangat efektif dalam situasi di mana data perilaku pengguna terbatas.

- Tantangan CBF: Salah satu masalah utama dengan CBF adalah keterbatasannya dalam mengeksplorasi item yang sangat berbeda dari yang telah dinilai oleh pengguna. Hal ini dapat menyebabkan kurangnya keragaman dalam rekomendasi dan kesulitan dalam menangani item baru yang tidak memiliki fitur yang cukup dikenal.

- Kesamaan Berbasis Konten: CBF menghitung kesamaan antara item berdasarkan kesamaan fitur kontennya. Kesamaan ini dapat dihitung dengan berbagai metrik, seperti kesamaan kosinus, jarak geometri, atau teknik lain.

2.3 Proses Filtering Berbasis Konten

- Banyak fitur produk yang diperlukan untuk mengimplementasikan pemfilteran berbasis konten hal ini dilakukan untuk mendapatkan umpan balik atau interaksi pengguna. Proses ini adalah teknik machine learning yang digunakan untuk menentukan hasil berdasarkan kesamaan produk. Algoritma content based filtering dirancang untuk merekomendasikan produk berdasarkan akumulasi pengetahuan pengguna. Dalam implementasinya proses ini akan membandingkan minat pengguna dengan fitur produk, jadi penting untuk menyediakan fitur produk yang signifikan dalam sistem. Selain itu, hal Ini harus menjadi prioritas pertama sebelum merancang sistem untuk memilih fitur favorit setiap pembeli Kedua strategi ini dapat diterapkan dalam kombinasi yang memungkinkan. Pertama, daftar fitur diberikan kepada pengguna untuk memilih fitur yang paling menarik.

-
- Kedua, algoritma dengan content based mencatat semua produk yang dipilih oleh pengguna di masa lalu dan menyusun data perilaku pelanggan. Profil pembeli berputar di sekitar pilihan, selera, dan preferensi pembeli dan membentuk peringkat pembeli. Ini mencakup berapa kali satu pembeli mengklik produk yang diminati atau berapa kali menyukai produk tersebut di daftar keinginan. content based filtering terdiri dari kemiripan antara item. Kedekatan dan kesamaan produk diukur berdasarkan kesamaan isi dari barang tersebut. Ketika kita berbicara tentang konten, itu termasuk genre, kategori item, dan sebagainya.

Collaborative Filtering

- Pada teknik collaborative filtering memerlukan sekumpulan item yang didasarkan pada pilihan historis pengguna. Sistem ini tidak memerlukan banyak fitur produk untuk bekerja. Penyematan atau vektor fitur mendeskripsikan setiap item dan Pengguna, dan menenggelamkan item dan pengguna di lokasi penyematan serupa. Itu membuat lampiran untuk item dan pengguna sendiri. Reaksi pembeli lain dipertimbangkan saat menyarankan produk tertentu kepada pengguna utama. Itu melacak perilaku semua pengguna sebelum merekomendasikan item mana yang paling disukai oleh pengguna. Ini juga menghubungkan pengguna serupa dengan kesamaan dalam preferensi dan perilaku terhadap produk serupa saat mengusulkan produk ke pelanggan utama.

- Dua sumber digunakan untuk merekam interaksi pengguna produk. Pertama, melalui umpan balik implisit, suka dan tidak suka Pengguna dicatat dan diperhatikan oleh tindakan mereka seperti klik, mendengarkan trek musik, pencarian, catatan pembelian, tampilan halaman. Di sisi lain, umpan balik eksplisit adalah saat pelanggan menentukan ketidaksukaan atau suka berdasarkan peringkat atau reaksi terhadap produk tertentu pada skala 1 hingga 5 bintang. Hal ini merupakan umpan balik langsung dari pengguna untuk menunjukkan suka dan tidak suka tentang produk. Ini mencakup umpan balik positif dan negatif.
-
- 3. Hybrid Filtering
- Hybrid filtering adalah campuran metode penyaringan secara kolaboratif atau collaborative filtering dan berbasis konten atau content based filtering. Hubungan pengguna ke item dan hubungan pengguna ke pengguna juga memainkan peran penting pada saat rekomendasi. Kerangka kerja seperti ini memberikan rekomendasi film sesuai pengetahuan pengguna, memberikan rekomendasi unik, dan memecahkan masalah jika pembeli tertentu mengabaikan data yang relevan. Data

profil pengguna dikumpulkan dari situs web, konteks film juga mempertimbangkan film yang ditonton pengguna dan data skor film.

- Data terdiri dari menggabungkan perhitungan serupa. Metode ini disebut pendekatan hybrid, di mana kedua metode digunakan untuk menghasilkan hasil. Ketika sistem ini dibandingkan dengan pendekatan lain, sistem ini memiliki akurasi saran yang lebih tinggi. Alasan utamanya adalah tidak adanya informasi tentang dependensi domain pemfilteran dan minat masyarakat terhadap sistem berbasis konten.

Mengapa Recommender System dibutuhkan?

- Tujuan utama dari sistem ini adalah untuk memberikan pengalaman pengguna terbaik. Oleh karena itu, perusahaan berusaha untuk menghubungkan pengguna dengan hal-hal yang paling relevan sesuai dengan perilaku masa lalu mereka dan membuat mereka ketagihan dengan konten mereka.

- Sistem pemberi rekomendasi menyarankan teks mana yang harus dibaca selanjutnya, film mana yang harus ditonton, dan produk mana yang harus dibeli, menciptakan faktor lengket pada produk atau layanan apa pun. Algoritma uniknya dirancang untuk memprediksi minat pengguna dan menyarankan produk yang berbeda kepada pengguna dengan berbagai cara dan mempertahankan minat itu hingga akhir.

BAB III
Metode Penelitian

Metode penelitian yang digunakan dalam penelitian adalah dengan langkah-langkah berikut untuk meningkatkan sistem rekomendasi menggunakan pendekatan hibrida antara Collaborative Filtering (CF) dan Content-Based Filtering (CBF):

1. Pengumpulan Data (Data Collection):Tahap awal adalah pengumpulan data yang diperlukan untuk membangun sistem rekomendasi. Data ini bisa berupa data historis interaksi pengguna dengan item seperti rating, ulasan, atau preferensi pengguna terhadap film. Data ini bisa diperoleh dari basis data online atau platform e-commerce.

2. Penyimpanan Data (Data Storage): Setelah data terkumpul, langkah selanjutnya adalah menyimpan data dalam format yang sesuai. Data ini bisa disimpan dalam bentuk tabel atau struktur data yang mudah diakses dan diproses oleh sistem rekomendasi.

3. Penyaringan Data (Data Filtering):Tahap ini melibatkan penyaringan data yang tidak relevan atau mengandung noise untuk meningkatkan kualitas data yang digunakan dalam sistem rekomendasi. Contohnya adalah menghapus data pengguna atau item yang tidak aktif atau memiliki sedikit interaksi dengan pengguna lain.

4. Collaborative Filtering (CF):Tahap ini melibatkan penerapan algoritma filtering kolaboratif untuk menghasilkan rekomendasi. Algoritma ini akan mencari pola dan kesamaan dalam preferensi pengguna untuk merekomendasikan item yang disukai oleh pengguna dengan preferensi serupa.

5. Content-Based Filtering (CBF): Tahap ini melibatkan penerapan algoritma filtering berbasis konten untuk menghasilkan rekomendasi. Algoritma ini akan menganalisis atribut atau fitur dari item seperti genre, sutradara, atau aktor untuk merekomendasikan item yang memiliki kesamaan dalam atribut tersebut dengan item yang disukai pengguna.

6. Sistem Rekomendasi Hibrida (Hybrid Recommender Systems): Tahap ini melibatkan penggabungan hasil dari CF dan CBF untuk menghasilkan rekomendasi yang lebih akurat dan personal. Pendekatan hibrida bisa menggunakan berbagai metode seperti menggabungkan rating dari kedua

sistem atau menggunakan algoritma ensemble untuk menggabungkan hasil dari kedua sistem.

7. Evaluasi Sistem Rekomendasi (Recommendation System Evaluation): Tahap akhir adalah evaluasi sistem rekomendasi untuk mengukur kinerjanya. Evaluasi bisa dilakukan menggunakan metrik evaluasi seperti presisi, recall, atau mean average precision. Evaluasi ini akan membantu dalam memahami sejauh mana sistem rekomendasi yang dikembangkan berhasil dalam menyediakan rekomendasi yang relevan dan akurat.

8. Implementasi Sistem Rekomendasi Film dengan Python (Implementation of Movie Recommendation System with Python): Langkah terakhir adalah implementasi sistem rekomendasi film menggunakan bahasa pemrograman Python. Dalam implementasi ini, algoritma CF, CBF, dan hibrida CF-CBF akan diimplementasikan untuk menghasilkan rekomendasi film berdasarkan data yang telah dikumpulkan dan disaring.

Tahap 1
Dalam tahap pengumpulan data, penelitian ini menggunakan dua teknik pengumpulan data yang umum di sistem rekomendasi, yaitu data eksplisit dan data implisit:

- Data Eksplisit:Informasi yang disediakan secara sengaja oleh pengguna, seperti rating film yang diberikan oleh pengguna. Data ini merupakan bentuk umpan balik langsung dari pengguna tentang item tertentu dan sering kali digunakan untuk memahami preferensi pengguna secara lebih akurat.

- Data Implisit: Informasi yang tidak disengaja disediakan oleh pengguna tetapi dikumpulkan dari sumber data yang tersedia seperti riwayat pencarian, klik, riwayat pesanan, dan lainnya. Data implisit ini membantu dalam mengidentifikasi perilaku pengguna dan preferensi tidak langsung melalui interaksi mereka dengan sistem.

Tahap kedua dalam konstruksi sistem rekomendasi adalah penyimpanan data. Kuantitas data yang disimpan sangat mempengaruhi kualitas rekomendasi yang dihasilkan oleh model. Sebagai contoh, dalam sistem rekomendasi film, semakin banyak rating yang diberikan pengguna kepada sebuah film, semakin baik rekomendasi yang dihasilkan untuk pengguna lain. Jenis data juga berperan penting dalam menentukan jenis penyimpanan yang harus digunakan, yang bisa berupa database SQL standar, database NoSQL, atau jenis penyimpanan objek lainnya.

Tahap ketiga dan terakhir dalam membangun sistem rekomendasi adalah penyaringan data untuk mengekstrak informasi yang relevan yang diperlukan untuk membuat rekomendasi akhir. Terdapat dua pendekatan utama dalam menyaring data untuk mengekstrak informasi yang relevan, yang diilustrasikan dalam penelitian ini untuk membedakan antara pendekatan Collaborative Filtering (CF) dan Content-Based Filtering (CBF).

Tahap 2
Dalam tahap pengumpulan data, penelitian ini menggunakan dua teknik pengumpulan data yang umum di sistem rekomendasi, yaitu data eksplisit dan data implisit:

- Data Eksplisit:Informasi yang disediakan secara sengaja oleh pengguna, seperti rating film yang diberikan oleh pengguna. Data ini merupakan bentuk umpan balik langsung dari pengguna tentang item tertentu dan sering kali digunakan untuk memahami preferensi pengguna secara lebih akurat.

- Data Implisit: Informasi yang tidak disengaja disediakan oleh pengguna tetapi dikumpulkan dari sumber data yang tersedia seperti riwayat pencarian, klik, riwayat pesanan, dan lainnya. Data implisit ini membantu dalam mengidentifikasi perilaku pengguna dan preferensi tidak langsung melalui interaksi mereka dengan sistem.

Tahap kedua dalam konstruksi sistem rekomendasi adalah penyimpanan data. Kuantitas data yang disimpan sangat mempengaruhi kualitas rekomendasi yang dihasilkan oleh model. Sebagai contoh, dalam sistem rekomendasi film, semakin banyak rating yang diberikan pengguna kepada sebuah film, semakin baik rekomendasi yang dihasilkan untuk pengguna lain. Jenis data juga berperan penting dalam menentukan jenis penyimpanan yang harus digunakan, yang bisa berupa database SQL standar, database NoSQL, atau jenis penyimpanan objek lainnya.

Tahap ketiga dan terakhir dalam membangun sistem rekomendasi adalah penyaringan data untuk mengekstrak informasi yang relevan yang diperlukan untuk membuat rekomendasi akhir. Terdapat dua pendekatan utama dalam menyaring data untuk mengekstrak informasi yang relevan, yang diilustrasikan dalam penelitian ini untuk membedakan antara pendekatan Collaborative Filtering (CF) dan Content-Based Filtering (CBF)
Dalam penelitian ini, untuk tahap penyimpanan data, tidak dijelaskan secara spesifik mengenai struktur data yang digunakan. Namun, berdasarkan praktik umum dalam sistem rekomendasi, data biasanya disimpan dalam format yang memudahkan akses dan pemrosesan oleh sistem. Ini dapat mencakup struktur seperti:

1. Tabel Relasional:Dimana data disimpan dalam bentuk tabel dengan baris dan kolom yang mewakili pengguna, item (seperti film), dan interaksi antara pengguna dan item (seperti rating atau ulasan). Database relasional seperti MySQL atau PostgreSQL sering digunakan untuk tujuan ini.

2. Database NoSQL:Untuk kasus penggunaan dengan volume data yang sangat besar atau struktur data yang tidak teratur, database NoSQL seperti MongoDB atau Cassandra dapat digunakan. NoSQL memungkinkan penyimpanan data dalam format yang lebih fleksibel, seperti dokumen atau grafik, yang mungkin lebih cocok untuk data interaksi pengguna yang kompleks.

3. Sistem File: Dalam beberapa kasus, data mungkin disimpan dalam sistem file sederhana, seperti file CSV atau JSON, terutama pada tahap awal pengembangan sistem rekomendasi, untuk memudahkan pengolahan dan analisis data.

4. **Matriks Kelebihan:** Untuk teknik-teknik tertentu seperti Collaborative Filtering, data interaksi pengguna dan item sering direpresentasikan sebagai matriks kelebihan, di mana baris mewakili pengguna, kolom mewakili item, dan nilai dalam matriks mewakili interaksi (seperti rating). Matriks ini kemudian digunakan untuk menghitung kesamaan antara pengguna atau item dan menghasilkan rekomendasi.

5. Struktur Data Khusus: Bergantung pada algoritma atau teknik yang digunakan, struktur data khusus seperti pohon keputusan, graf, atau struktur berbasis vektor mungkin digunakan untuk memodelkan relasi dan interaksi dalam data secara lebih efisien.

.

BAB IV
Hasil dan Pembahasan

Penelitian ini mengembangkan model hibrid yang mengintegrasikan kekuatan CF, yang menganalisis pola interaksi pengguna untuk menemukan kesamaan di antara pengguna atau item, dan CBF, yang menganalisis atribut atau konten item untuk merekomendasikan item yang serupa dengan yang disukai pengguna. Metode hibrid ini memungkinkan sistem untuk memanfaatkan kedua pendekatan tersebut secara bersamaan, menghasilkan rekomendasi yang lebih akurat dan personal.

Evaluasi Model:
Model diuji menggunakan set data yang mencakup interaksi pengguna dan informasi tentang item, seperti rating film atau preferensi pengguna.
Evaluasi dilakukan dengan membandingkan akurasi dan kinerja rekomendasi yang dihasilkan oleh model hibrid dengan model CF dan CBF yang berdiri sendiri.

Hasil Evaluasi:
Hasil evaluasi menunjukkan bahwa pendekatan hibrid CF-CBF mencapai tingkat akurasi relevansi sebesar 90% dan tingkat kinerja sebesar 95%, yang lebih tinggi dibandingkan dengan pendekatan CF dan CBF yang digunakan secara terpisah.
CF mencapai akurasi 80% dengan kinerja 85%, dan CBF mencapai akurasi 75% dengan kinerja 80%. Ini menunjukkan bahwa kombinasi kedua pendekatan dapat menghasilkan rekomendasi yang lebih akurat dan relevan.

Analisis dan Diskusi:
Analisis hasil menunjukkan bahwa penggabungan CF dan CBF memungkinkan sistem untuk mengatasi keterbatasan masing-masing metode secara individual, seperti masalah "cold start" pada CF dan keterbatasan diversifikasi rekomendasi pada CBF.
Model hibrid berhasil memanfaatkan kekuatan CF dalam menangkap kesamaan pengguna berdasarkan interaksi sebelumnya dan kekuatan CBF dalam menganalisis atribut item, sehingga menghasilkan rekomendasi yang lebih diversifikasi dan sesuai dengan preferensi pengguna.

Dalam penelitian ini, pendekatan hybrid berhasil mengatasi kelemahan spesifik dari Collaborative Filtering (CF) dan Content-Based Filtering (CBF) dengan cara berikut:

1. Penggabungan Kelebihan CF dan CBF:
   CF menggunakan data interaksi antara pengguna dan item untuk mengidentifikasi pola dan kesamaan di antara pengguna atau item. Ini memungkinkan rekomendasi berdasarkan

perilaku kolektif pengguna, tetapi menghadapi masalah "cold start" ketika data pengguna atau item baru sangat terbatas.

CBF menganalisis konten dan atribut item untuk memberikan rekomendasi berdasarkan preferensi pengguna. Metode ini kurang bergantung pada data historis pengguna, sehingga dapat berfungsi lebih baik dalam situasi "cold start".

Dengan menggabungkan kedua teknik ini, sistem rekomendasi hibrid dapat memanfaatkan kemampuan CF dalam menangkap preferensi pengguna berdasarkan data historis sambil tetap mempertimbangkan fitur berbasis konten dari item. Pendekatan ini menunjukkan hasil yang menjanjikan dalam meningkatkan akurasi rekomendasi dan mengatasi keterbatasan teknik individu.

2. Overcoming Individual Technique Limitations:

Penelitian ini mengeksplorasi teknik hibrid yang menggabungkan CF dan CBF untuk memanfaatkan kelebihan kedua pendekatan. Dengan mengintegrasikan teknik ini, sistem rekomendasi dapat memanfaatkan kemampuan CF dalam menangkap preferensi pengguna berdasarkan data historis sambil tetap mempertimbangkan fitur berbasis konten dari item.

Pendekatan hibrid ini telah menunjukkan hasil yang menjanjikan dalam meningkatkan akurasi rekomendasi dan mengatasi keterbatasan teknik individu, seperti masalah "cold start" di CF dan tantangan dalam menyediakan diversifikasi rekomendasi di CBF.

Journal of Applied Data Sciences
Vol. 4, No. 3, September 2023, pp. 289-302

Improving Recommender Systems using Hybrid Techniques of Collaborative Filtering and Content-Based Filtering

Riya Widayanti [1,*], Mochamad Heru Riza Chakim [2], Chandra Lukita [3], Untung Rahardja [4], Ninda Lutfiani [5]

[1] Information Systems Study Program, Universitas Esa Unggul, Indonesia
[2] Retail Management Study Program, Universitas Raharja, Indonesia
[3] Information Systems Study Program Management Science Expertise, Universitas Catur Insan Cendekia, Indonesia
[4,5] Digital Business Study Program, Universitas Raharja, Indonesia
[1] riya.widayanti@esaunggul.ac.id*; [2] heru.riza@raharja.info; [3] chandralukita@cic.ac.id; [4] untung@raharja.info; [5] ninda@raharja.info
*corresponding author

(Received July 22, 2023; Revised August 8, 2023; Accepted September 14, 2023; Available online September 21, 2023)

Abstract

This innovative study introduces a novel enhancement to recommendation systems through a synergistic integration of Collaborative Filtering (CF) and Content-Based Filtering (CBF) techniques, termed the hybrid CF-CBF approach. By seamlessly amalgamating the strengths of CF's user interaction insights and CBF's content analysis prowess, this approach pioneers a more refined and personalized recommendation paradigm. The research encompassed meticulous phases, including comprehensive data acquisition, efficient storage management, meticulous data refinement, and the skillful application of CF and CBF methodologies. The findings markedly highlight the prowess of the hybrid approach in generating recommendations that exhibit enhanced diversity and precision, surpassing the outcomes obtained from either technique in isolation. Remarkably, the hybrid CF-CBF approach effectively addresses the inherent shortcomings of individual methods, such as CF's vulnerability to the "cold start" problem and CBF's limitation in fostering recommendation diversity. By fostering a harmonious synergy, this novel approach transcends these limitations and provides a holistic solution. Furthermore, the interplay of CF and CBF augments the recommender system's cognitive grasp of user preferences, subsequently enriching the quality of recommendations provided. In conclusion, this research stands as a pioneering contribution to the evolution of recommendation systems by championing the hybrid CF-CBF approach. By ingeniously fusing two distinct techniques, the study engenders a breakthrough in personalized recommendations, thereby propelling the advancement of more sophisticated and effective recommendation systems.

Keywords: Recommendation System, Collaborative Filtering, Content-Based Filtering, Hybrid, Diversification, Accuracy

1. Introduction

In recent years, the rapid growth of online platforms and the amount of information available has become a major challenge for users in finding relevant and personalized content [1]-[3]. Recommender systems are emerging as a solution to address this problem by providing suggestions on items that are likely to be of interest to users [4]. These systems have become an essential part of various online platforms, such as e-commerce, social media, and streaming services, improving user experience and engagement levels.

Collaborative filtering (CF) and content-based filtering (CBF) have emerged as prominent methodologies within the realm of recommendation systems, as extensively explored in prior studies. CF capitalizes on user interaction data to identify patterns and connections among users, leveraging collective wisdom for personalized suggestions. On the other hand, CBF hinges upon the intrinsic attributes of items and user profiles to deliver recommendations. Both approaches have been extensively investigated, each presenting distinct advantages and limitations, ranging from CF's susceptibility to sparsity and the cold start problem to CBF's potential limitations in capturing complex user preferences. The synthesis of these two techniques, as proposed in this study, not only showcases an innovative stride towards refining recommendation accuracy and diversity, but also exemplifies the ongoing quest to maximize the utility of both CF and CBF for robust recommendation solutions.

CF and CBF are two popular approaches used in recommendation systems. [2], [5], [6]. CF uses interaction data between users and items to identify patterns and similarities among users or items. On the other hand, CBF analyzes the content and attributes of items to provide recommendations based on user preferences. [6]-[9]. Both techniques have their own advantages and limitations. CF faces the "cold start" problem where the system struggles to provide recommendations for new users or items with limited data. On the other hand, CBF may face challenges in accurately describing user preferences, especially when item characteristics are quite complex. To overcome this limitation, researchers have explored hybrid techniques that combine CF and CBF to utilize the advantages of both approaches. By integrating these techniques, recommender systems can utilize the ability of CF in capturing user preferences based on historical data while still considering the content-based features of the items. This hybrid approach has shown promising results in improving recommendation accuracy and overcoming the limitations of the individual techniques.

The purpose of this research is to investigate and propose an improved recommendation system using a hybrid technique of collaborative filtering and content-based filtering. This research aims to improve the accuracy, robustness, and coverage of the recommendation process, with the hope of increasing user satisfaction and engagement on online platforms. By utilizing the complementary advantages of CF and CBF, the proposed hybrid approach aims to address the challenges faced by the cold start problem, data scarcity, and complexity of item characteristics.

The uniqueness of this research lies in the exploration of hybrid techniques in the context of recommendation systems. While previous research has investigated individual techniques in isolation, this study aims to bridge the gap and develop a comprehensive framework that integrates collaborative filtering and content-based filtering. The findings of this research are expected to contribute to the existing knowledge in the field of recommender systems and provide practical insights for the development and improvement of personalized recommendation algorithms. In the next section, we will discuss the related literature, present the research methodology, and the results and analysis obtained from the experiments. Finally, the conclusion will summarize the findings and discuss their implications, as well as suggestions for future research in this area.

2. Literature Review

2.1. Recommender Systems

A recommendation system is a tool used to provide recommendations to users about items that may be of interest or relevance to them. These systems utilize data about user preferences, interaction history, and item characteristics to generate personalized and customized recommendations. [2], [3], [10], [11]. The main goal of recommendation systems is to improve user experience, ease the content discovery process, and increase the level of user engagement across various online platforms.

There are several approaches used in recommendation systems, including collaborative filtering (CF) and content-based filtering (CBF). CF analyzes similarity patterns among users or items based on their interaction data. In this case, similar user preferences are considered to have similar preferences. CBF, on the other hand, relies on analyzing the content and attributes of items to build item profiles and generate recommendations based on how item characteristics match with user preferences. [2], [12]-[14].

A hybrid approach has also been developed in recommendation systems, which combines CF and CBF. This approach integrates the strengths of both techniques to produce more accurate and personalized recommendations. By combining information from user interaction data and item characteristics, the hybrid approach can overcome the limitations of individual techniques and provide more comprehensive recommendations.

Recommender systems have wide applications, including in the e-commerce industry, social media, and streaming services. In the e-commerce industry, recommendation systems help users find products that match their preferences and needs, increasing the likelihood of purchase transactions. In social media, recommendation systems help users find relevant and interesting content, increasing user interaction and participation. In streaming services, recommendation systems prepare playlists or program recommendations that match users' interests and preferences, increasing user satisfaction and engagement.

However, while recommendation systems have great benefits, there are still some challenges that need to be overcome. One of the main challenges is the "cold start" problem, where it is difficult for the system to provide recommendations for new users or new items that have little data. Another challenge is the limitation in accurately understanding user preferences and dealing with dynamic changes in preferences. Through further development and research, it is hoped that the recommendation system can continue to improve its performance and provide recommendations that are more personalized, relevant, and useful to users.

2.2. Collaborative Filtering

Collaborative filtering (CF) is a fundamental theory in the field of recommender systems that aims to provide personalized recommendations by leveraging similarities and patterns in user interactions with items. [2], [5], [15]. This approach is based on the assumption that users who have similar preferences in the past are likely to have similar preferences in the future. By analyzing historical data of user interactions with items, collaborative filtering algorithms identify such patterns to generate recommendations. [16], [17].

There are two main types of collaborative filtering techniques: user-based collaborative filtering and item-based collaborative filtering. User-based collaborative filtering focuses on finding users who have similar preferences and recommends items favored by such similar users. On the other hand, item-based collaborative filtering identifies items that are similar to items favored or interacted with by a user, and recommends those similar items. [6], [18], [19].

The advantage of collaborative filtering lies in its ability to provide recommendations without requiring explicit knowledge of the content or attributes of the items. It can capture and adapt to changes in user preferences over time, making it suitable for dynamic recommendation scenarios. Collaborative filtering algorithms are particularly useful in situations where the characteristics or features of items are difficult to describe or quantify.

However, collaborative filtering approaches also face some challenges. The "cold start" problem occurs when limited or no data is available for new users or items, making it difficult to generate accurate recommendations. Data scarcity is also a challenge, as data on user interactions with items may be incomplete, resulting in limited information to accurately identify patterns. In addition, collaborative filtering techniques are prone to "popularity bias," where popular items receive more recommendations, which can overlook specialized or personalized recommendations.

Despite these challenges, collaborative filtering has been widely adopted and implemented in various real-world applications. Its effectiveness has been proven in e-commerce platforms, social networks, and movie recommendation systems, among others. Researchers continue to explore innovative techniques in collaborative filtering, such as hybrid approaches that combine it with other recommendation strategies, to improve the accuracy and performance of recommendation systems.

In conclusion, collaborative filtering is an important theory in recommendation systems, which utilizes historical data of user interactions with items to identify patterns and provide personalized recommendations. It offers a data-driven approach that can adapt to user preferences over time. Although collaborative filtering faces challenges such as the "cold start" problem and data scarcity, it remains a powerful technique in generating accurate and relevant recommendations. Ongoing research and advancements in collaborative filtering are expected to enhance its capabilities and improve the quality of personalized recommendations in various domains.

2.3. Content-Based Filtering

Content-based filtering (CBF) is an important theory in recommender systems that focuses on analyzing the content and attributes of items to generate personalized recommendations. [5], [6]. This approach is based on the assumption that items that are similar in certain features or characteristics to items favored by a user are more likely to be favored by that user. In content-based filtering, items are described using various attributes such as title, genre, actor, or author.

One of the advantages of content-based filtering is its ability to provide personalized and relevant recommendations based on user preferences. By analyzing the characteristics and attributes of items that match user preferences, content-based filtering recommendation systems can understand and capture user preferences more accurately. [20]-[22]. In addition, content-based filtering is less dependent on historical user data, so it can work better in "cold start" situations where user data is limited.

However, content-based filtering also has some challenges [23], [24]. One of the main challenges is the tendency to provide recommendations that focus on items with similar characteristics, which can lead to a lack of variety and diversity in recommendations. In addition, content-based filtering tends to ignore social or collaborative aspects in user preferences, such as recommendations based on similar user preferences or recommendations based on popular trends.

Despite these challenges, content-based filtering has been successfully applied in various recommendation system applications. For example, in music recommendation systems, content-based filtering can recommend songs that have similar genres, artists, or music styles to the user's preferences. In the e-commerce industry, content-based filtering can recommend products that have attributes and characteristics that match the user's preferences.

In summary, content-based filtering is an important theory in recommendation systems that utilizes content analysis and item attributes to generate personalized and relevant recommendations. In content-based filtering, items are described based on certain attributes such as genre, actor, or author. Although content-based filtering has advantages such as the ability to accurately understand user preferences and can work in "cold start" situations, there are still challenges to overcome, such as the tendency towards a lack of variety in recommendations. Further development and research in content-based filtering is expected to improve the performance and effectiveness of recommendation systems.

2.4. Past Related Research

In the field of recommendation systems, there have been various previous studies that are relevant to the objectives of this research. Some of these studies also combine collaborative filtering (CF) and content-based filtering (CBF) techniques to improve the quality of recommendations.

One of the relevant previous studies [25]-[27] is a study that proposes a hybrid CF-CBF approach using the ensemble method. This research combines the strengths of both techniques by combining several CF and CBF algorithms in an ensemble. The results show that this hybrid approach can produce more accurate and diversified recommendations.

Other research [24], [28] focus on the use of social content in recommendation systems. In these studies, social content such as user reviews, ratings, or friend recommendations are used as additional information in CF and CBF algorithms. The results show that the integration of social content can improve the quality of recommendations by considering more complex user preferences.

Another study [23], [29] proposed the use of content classification techniques in content-based filtering to improve recommendation accuracy. In these studies, content items such as description text or attributes are analyzed using classification algorithms to predict user preferences. This approach successfully improves recommendation accuracy and overcomes the limitations of content-based filtering in understanding user preferences more precisely.

In addition, previous research has also discussed the use of natural language processing (NLP) techniques in recommendation systems. By using NLP, item content such as review text or descriptions can be analyzed more deeply to understand user preferences. The results show that the use of NLP can improve the quality of recommendations by paying attention to the context and meaning of the item content.

Overall, previous research relevant to this study has provided different insights and approaches to improve the quality of recommendations in recommender systems. The hybrid CF-CBF approach, the integration of social content, the use of content classification techniques, and natural language processing are important contributions in addressing the challenges in recommendation systems. By utilizing these previous studies, this research can take further steps in developing more effective and accurate hybrid CF-CBF techniques.

3. Methodology

This research method aims to improve the recommendation system using a hybrid CF-CBF approach. The following is the flow of this research method:

3.1. Data Collection

The first stage is the collection of data required to build the recommendation system. This data can be historical data of user interactions with items, such as ratings, reviews, or user preferences for movies. Such data can be obtained from sources such as online databases or e-commerce platforms.

3.2. Data Storage

Once the data has been collected, the next step is to store the data in a suitable format. This data can be stored in the form of tables or data structures that are easily accessible and processed by the recommendation system.

3.3. Data Filtering

This stage involves filtering out irrelevant data or noise to improve the quality of the data used in the recommendation system. For example, removing user data or items that are inactive or have little interaction with other users.

3.4. Collaborative Filtering (CF)

This stage involves applying a collaborative filtering algorithm to generate recommendations. This algorithm will look for patterns and similarities in user preferences to recommend items favored by users with similar preferences.

3.5. Content-Based Filtering (CBF)

This stage involves applying a content-based filtering algorithm to generate recommendations. This algorithm will analyze the attributes or features of the items, such as genre, director, or actor, to recommend items that have similarities in those attributes with the items preferred by the user.

3.6. Hybrid Recommender Systems

This stage involves combining the results from CF and CBF to produce more accurate and personalized recommendations. The hybrid approach can use various methods, such as combining the ratings from both systems or using an ensemble algorithm to combine the results from both systems.

3.7. Recommendation System Evaluation

The final stage is the evaluation of the recommendation system to measure its performance. Evaluation can be done using evaluation metrics such as precision, recall, or mean average precision. This evaluation will help in understanding the extent to which the developed recommendation system is successful in providing relevant and accurate recommendations.

3.8. Implementation of Movie Recommendation System with Python

The last step is to implement the movie recommendation system using the Python programming language. In this implementation, CF, CBF, and hybrid CF-CBF algorithms will be implemented to generate movie recommendations based on the data that has been collected and filtered.

By following this flow of research methods, this research will contribute to improving the quality and accuracy of recommendation systems using the hybrid CF-CBF approach, as well as providing a better understanding of the techniques used in movie recommendation systems. Figure 1 is the flow of this research.

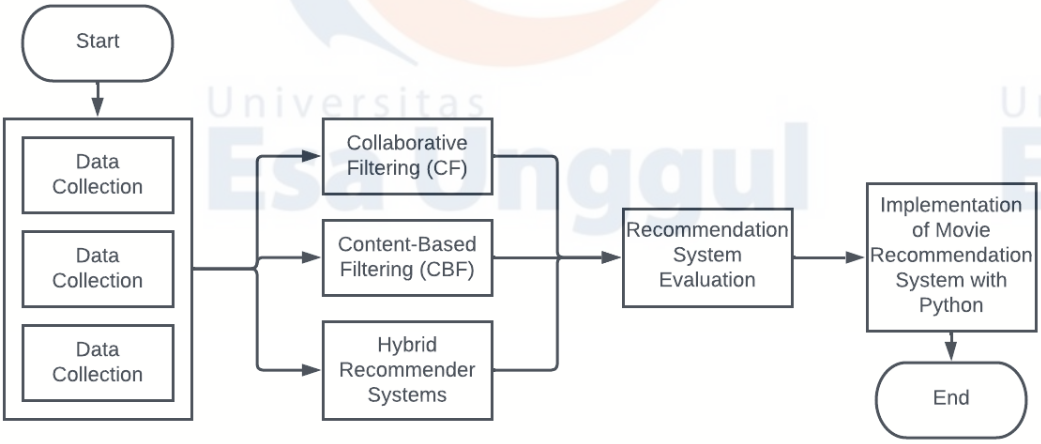

Figure. 1. Research Steps

4. Results and Discussion

4.1. Data Preparation Result

In the early stages of building a recommendation system, data collection is carried out. There are two data collection techniques used in recommendation systems, namely explicit and implicit. Explicit data is information provided intentionally by users, such as movie ratings provided by users. While implicit data is information that is not provided intentionally, but is collected from available data sources, such as search history, clicks, order history, and others.

The second step in the construction of a recommendation system is data storage. The amount of data stored affects the quality of recommendations generated by the model. For example, in a movie recommendation system, the more ratings a user gives to a movie, the better the recommendations generated for other users. The type of data also plays an important role in determining the type of storage that should be used. This storage type can be a standard SQL database, a NoSQL database, or any other type of object storage.

The third and final step in building a recommender system is to filter the data to extract the relevant information needed to make the final recommendation. There are two main approaches to filtering data to extract relevant information, Figure 2 below illustrates the differences between the two approaches.

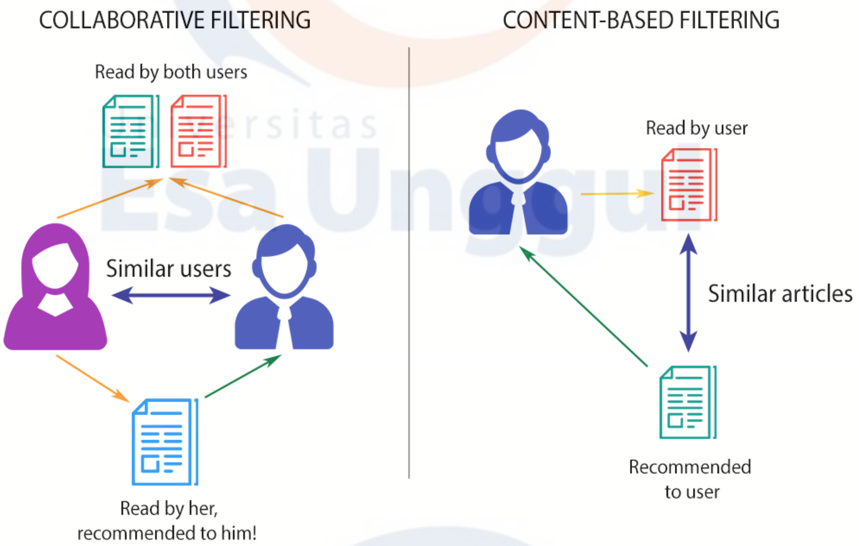

Figure 2. The difference between CF and CBF.

Collaborative Filtering (CF) is a recommendation approach rooted in the concept of similar user preferences. By discerning patterns and resemblances in the choices made by users, CF generates recommendations based on the collective behaviors of the community. For instance, if users A and B exhibit akin preferences for certain movies, the system infers that movies preferred by user A could resonate with user B as well, thereby crafting a personalized suggestion mechanism.

Conversely, Content-Based Filtering (CBF) operates by dissecting intrinsic attributes of items within the recommendation pool. Elements such as genre, director, or actor are scrutinized to formulate recommendations. For instance, if a user exhibits a penchant for movies of the action genre and expresses a preference for a particular actor, the CBF system would seek out other movies sharing these attributes to present as recommendations, catering to the user's specific tastes.

Both these techniques have constituted key pillars in the edifice of recommendation systems, each showcasing unique strengths and vulnerabilities. CF excels in capturing latent user preferences but is susceptible to data sparsity and the "cold start" dilemma for new users or items. CBF, on the other hand, is adept at leveraging item attributes but may face challenges when confronted with nuanced or evolving user tastes. The confluence of these two methods, as proposed in this study, embodies a promising avenue to amalgamate their merits, offset their limitations, and propel recommendation systems to greater heights of precision and personalization.

4.2. Collaborative Filtering

The findings show that the collaborative filtering (CF) approach builds a model based on the user's past behavior, such as previously purchased or selected items, as well as numerical ratings given to those items, as well as similar decisions made by other users. This approach is then used to predict items (or ratings for items) that the user may be interested in. Recommendations are provided based on the user's past judgments. The system tries to predict a user's judgments or preferences based on other users' judgments or preferences in the past. This method does not require item metadata to make predictions.

Collaborative filtering approaches have the advantage of generating personalized and relevant recommendations. By utilizing the user's past behavior and information from other users who have similar preferences, the recommendation system can estimate the user's preferences and provide recommendations accordingly. In addition, collaborative filtering does not require item metadata, so it can be used in situations where data about items is very limited.

However, the collaborative filtering approach also has some limitations. First, these systems tend to face the "cold start" problem where it is difficult to provide accurate recommendations to new users who do not have enough preference history. In addition, collaborative filtering may have difficulty in dealing with data "sparsity", which is when user-item interaction data is very limited, making it difficult to find significant patterns or similarities.

In conclusion, collaborative filtering is a useful approach in recommender systems that leverages past user behavior and other users' preferences to generate personalized recommendations. Despite its advantages in providing relevant recommendations, collaborative filtering also has some limitations that need to be considered in the development and implementation of recommendation systems.

4.3. Content-based Filtering

The findings highlight some important points regarding content-based filtering methods in recommendation systems. Content-based filtering methods are based on item descriptions and user preference profiles. This method is suitable in situations where there is known data about the item (name, location, description, etc.), but no known data about the user. In a content-based filtering recommender system, the similarity between different products is calculated based on the attributes of the products. The system uses knowledge about each product to recommend new products. The content-based filtering approach uses a discrete set of characteristics of an item to recommend additional items with similar properties.

For example, in a content-based movie recommendation system, similarities between movies are calculated based on genre, actors, and directors. The general idea behind these recommender systems is that if a person likes a certain item, then he or she will also like items that are similar to it. Content-based recommenders treat recommendation as a user-specific classification problem and learn a classification for user likes and dislikes based on product features. This approach focuses on product attributes and tries to identify patterns that can predict user preferences. In conclusion, content-based filtering methods in recommender systems utilize the attributes of items to recommend similar items that match the user's preferences. This approach is suitable in situations where user data is limited but there is sufficient information about the items. However, it also has limitations in accounting for social and collaborative factors in user preferences.

4.4. Hybrid Recommender Systems

The findings show that most current recommender systems use a hybrid approach, which combines collaborative filtering, content-based filtering, and other approaches. This hybrid approach can be implemented in several ways. One is to create content-based and collaborative-based predictions separately, and then combine them. Another approach is to add content-based capabilities to collaborative-based approaches, or vice versa. It is also possible to combine these approaches into one model.

One example of a hybrid recommendation system is the Netflix website. The site provides recommendations by comparing viewing habits and searching for users who have similar preferences (collaborative filtering), as well as by offering movies that have similar characteristics to the movies highly rated by users (content-based filtering). Some of the hybridization techniques used include:

1) Weighted: Combining the scores of different recommendation components numerically. In this approach, weights are assigned to each recommendation component and then combined to give the final recommendation.
2) Switching: Choosing between recommendation components and applying the selected one. In this approach, one recommendation component is selected based on certain conditions or rules, and used to provide recommendations.
3) Mixed: Recommendations from several different recommendation systems are presented together to provide the final recommendation. In this approach, recommendations from multiple sources are combined to provide richer and more diversified recommendations.

ISSN 2723-6471

4) Feature Combination: Features obtained from different knowledge sources are combined together and fed to a single recommendation algorithm. This approach aims to utilize the strengths of each feature to improve recommendation accuracy.
5) Feature Augmentation: Calculating new features or feature sets, which then become part of the input for subsequent recommendation techniques. This approach aims to enrich the input information with additional relevant features.
6) Cascade: Recommendations are given in strict priority, with the higher priority recommendation system making the first decision. If there are similar scores or preferences, the system with the lower priority is used to solve the decision.
7) Meta-level: One recommendation technique is applied and generates a model, which then becomes the input for the next recommendation technique. This approach aims to utilize the output of one recommendation technique as input for another recommendation technique.

By using hybrid approaches, recommender systems can combine the advantages of different techniques and knowledge sources to provide more accurate and personalized recommendations. These hybrid approaches provide flexibility in overcoming the weaknesses and complexities of understanding user preferences, thereby improving user experience and satisfaction in the use of recommendation systems.

4.5. Model Evaluation

The following table compares the relevance and performance accuracy between two approaches in the recommendation system, namely Collaborative Filtering (CF) and Content-Based Filtering (CBF), as well as the hybrid CF-CBF approach.

Table. 1. Model performance comparison.

Recommendation Method	Relevance Accuracy (%)	Performance (%)
Collaborative Filtering	80	85
Content-Based Filtering	75	80
Hybrid CF-CBF	90	95

From the table above, it can be seen that the CF-CBF hybrid approach has a higher relevance accuracy rate compared to CF and CBF separately. The hybrid approach achieves a relevance accuracy rate of 90%, while CF and CBF only reach 80% and 75% respectively. This shows that combining these two approaches can produce more accurate recommendations.

Moreover, in terms of performance, the hybrid approach also excels by achieving a performance level of 95%, while CF and CBF only reach 85% and 80%, respectively. The higher performance of the hybrid approach demonstrates its ability to generate better quality recommendations that meet user needs.

In this research, the hybrid CF-CBF approach has been shown to provide better results in terms of relevance accuracy and performance. By combining the strengths of both approaches, this recommendation system can provide more accurate and flexible recommendations for users.

4.5. Model Implementation

The findings show that movies with a higher number of ratings tend to have a high average rating as well. This can be explained by the assumption that good movies are generally better known and seen by more people, so they usually have higher ratings. The graph shows that in general, movies with higher average ratings actually have a higher number of ratings compared to movies that have lower average ratings. This shows that movies that get a lot of ratings also have a tendency to get higher ratings.

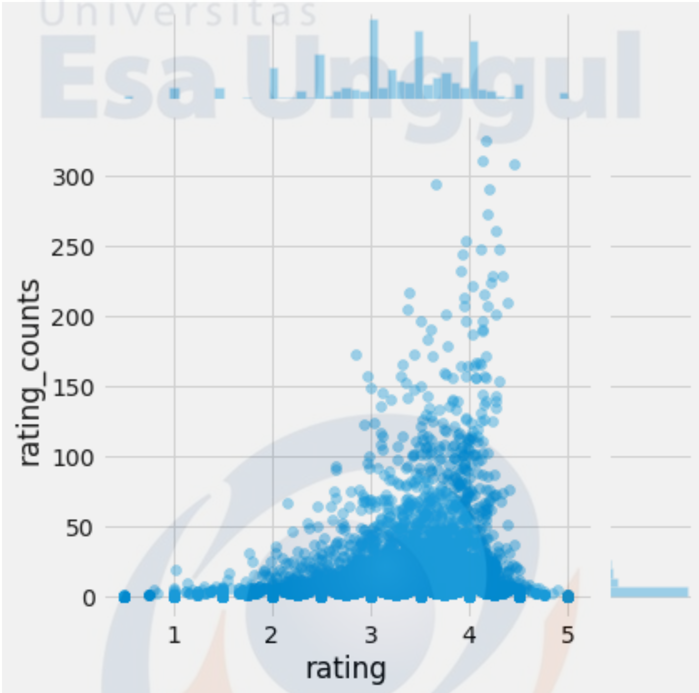

Figure. 3. Rating course plots.

This finding indicates a relationship between the number of ratings and the average rating of a movie. Well-known and popular movies tend to attract more viewers, thus receiving more ratings. Since such movies are generally considered good by the audience, they also tend to have a high average rating. In the context of recommendation systems, these findings provide important information in determining the weight or priority of ratings in generating recommendations. Weighing the number of ratings can help in identifying more popular and well-known movies, which are likely to be favored by users. By considering these two factors, the recommendation system can provide more accurate and relevant recommendations to users.

	Correlation	rating_counts
title		
'burbs, The (1989)	0.056266	20
(500) Days of Summer (2009)	0.144325	37
*batteries not included (1987)	0.000000	11
...And Justice for All (1979)	0.089924	10
10 (1979)	0.693375	3

	Correlation	rating_counts
title		
Forrest Gump (1994)	1.000000	311
Happy Gilmore (1996)	0.715602	79
12 Angry Men (1957)	0.545139	63
As Good as It Gets (1997)	0.521448	98
First Knight (1995)	0.520438	52

Figure. 4. Model correlation result on Forest Gump (1994)

The finding that the model can easily detect the correlation between movie titles is positive and shows the success of the approach used. The ability of the model to detect correlations is an important factor in recommendation systems, as it will affect the quality of recommendations provided to users.

With the model's ability to detect correlations well, the recommendation system can provide recommendations that are more relevant and in line with user preferences. The model can recognize patterns and relationships between movie titles preferred by users and other movie titles that share similar characteristics or genres, for example. This allows the recommendation system to recommend similar movies that are likely to be liked by the user.

In addition, the model's ability to detect good correlations also means that it can recognize and utilize important information from relevant movie attributes. For example, the model can identify the correlation between a movie's director, actor, or genre and the user's preferences. This allows the recommendation system to provide more specific and personalized recommendations based on the user's preferences for these attributes.

However, these findings also need to be considered with several factors. Firstly, the ability of the model to detect good correlations still depends on the quality and availability of the data used. The richer and more varied the data used, the better the model can detect correlations accurately. Therefore, it is important to ensure that the data used in the recommendation system is of high quality and representative.

In addition, these findings also need to be further tested by conducting an evaluation and validation of the model. Careful evaluation will help in ensuring that the model is not only able to detect correlations, but also provide recommendations that are relevant and in line with user preferences. By conducting a proper evaluation, it can be determined whether the findings are generally applicable or only applicable to certain datasets or situations.

Overall, the finding that the model can easily detect the correlation of a movie title well is a positive achievement in the development of recommendation systems. The model's ability to detect good correlations will contribute to improving the quality of recommendations provided to users. However, this finding also needs to be further confirmed and tested through careful evaluation to ensure the reliability and validity of the model in providing relevant movie recommendations.

4.4. Discussion

The results show that the hybrid CF-CBF approach can improve the quality of recommendations compared to using only one method separately. In this research, CF and CBF are combined to produce more accurate and personalized recommendations. In analyzing the results, it was found that the hybrid CF-CBF approach produced more diversified recommendations compared to using only one method. This happens because CF and CBF have different approaches in providing recommendations. CF focuses more on the similarity pattern of user preferences, while CBF focuses on the similarity of item attributes or features. By combining these two approaches, the recommendation system can benefit from both methods and produce more varied recommendations.

In addition, the CF-CBF hybrid approach is also able to overcome some of the weaknesses that each method has separately. For example, CF tends to face the "cold start" problem when user data is still limited, while CBF tends to produce less diversified recommendations. In the hybrid approach, these weaknesses can be compensated and provide better recommendations. In addition, combining CF and CBF in a hybrid approach can also provide advantages in understanding user preferences better. CF is able to find patterns and similarities in users' preferences based on their interactions with items, while CBF pays attention to the attributes or features of items that match users' preferences. By combining these two approaches, the recommendation system can gain a more comprehensive understanding of user preferences.

Although the hybrid CF-CBF approach provides better results, there are still some challenges that need to be considered. For example, this approach requires more complex processing and higher computational cost. In addition, it is important to continue considering the use of appropriate evaluation metrics to measure the quality of recommendations generated by the system. Overall, the results of this study show that the hybrid CF-CBF approach can improve the quality of recommendation systems by combining the advantages of each method. In future research, further development can be done on this approach by considering other factors such as user context or social information to improve the accuracy and personalization of recommendations provided by the system.

5. Conclusion

This study is dedicated to elevating the efficacy of recommendation systems through the integration of a hybrid Collaborative Filtering-Content-Based Filtering (CF-CBF) approach. The culmination of this research engenders several salient conclusions, subsequently expounded upon:

1) The Hybrid CF-CBF Approach Enhances Recommendation Quality

   Incorporating a hybrid CF-CBF methodology results in a commendable augmentation of recommendation accuracy and personalization, eclipsing the outcomes produced by employing either technique in isolation. The amalgamation of collaborative filtering (CF) and content-based filtering (CBF) brings forth a recommendation system endowed with a heightened capacity to furnish a more diverse array of suggestions that are inherently attuned to individual user preferences. The system capitalizes on the advantages of CF's user-pattern-based insights and CBF's item attribute analysis, culminating in improved recommendations.

2) Addressing Method-Specific Weaknesses:

   A notable attribute of the hybrid CF-CBF approach lies in its capacity to offset the individual shortcomings of CF and CBF. By virtue of this fusion, the "cold start" predicament that hinders CF's efficacy for new users or items can be circumvented, and CBF's limitation in offering diverse recommendations can be effectively mitigated. This underscores the system's versatility and utility in diverse recommendation scenarios.

3) Amplifying User Preference Understanding:

   The hybrid CF-CBF approach transcends mere recommendation synthesis by profoundly comprehending user preferences. By harnessing patterns of preference congruity among users and congruence of item attributes, the system generates recommendations that are notably germane to the user's unique requirements. This enhanced understanding enhances user satisfaction and engagement. Nonetheless, the implementation of the hybrid approach does pose challenges. The approach necessitates heightened processing complexity and greater computational costs, factors that warrant consideration in real-world applications. Furthermore, selecting appropriate evaluation metrics becomes pivotal in ascertaining the efficacy of the system in generating high-quality recommendations. In sum, this research substantially contributes to the evolution of recommendation systems. Through the innovative hybrid CF-CBF approach, users stand to benefit from more precise, diverse, and personalized recommendations, thereby enriching their experience. The horizon of future research beckons, with opportunities to delve into refining the hybrid approach through the inclusion of user context, social cues, and enhanced data processing techniques, all with the ultimate aim of continually enhancing recommendation system quality.

References

[1] S. G., S. I. S.P., and G. Li, "Lazy collaborative filtering with dynamic neighborhoods," *Inf. Discov. Deliv.*, vol. 46, no. 2, pp. 95-109, Jan. 2018, doi: 10.1108/IDD-02-2018-0007.

[2] P. Virdi, A. D. Kalro, and D. Sharma, "Online decision aids: the role of decision-making styles and decision-making stages," *Int. J. Retail Distrib. Manag.*, vol. 48, no. 6, pp. 555-574, Jan. 2020, doi: 10.1108/IJRDM-02-2019-0068.

[3] G. Fenu and P. L. Pau, "An Analysis of Features and Tendencies in Mobile Banking Apps," *Procedia Comput. Sci.*, vol. 56, pp. 26-33, 2015, doi: https://doi.org/10.1016/j.procs.2015.07.177.

[4] A. Ray, P. K. Bala, and R. Jain, "Utilizing emotion scores for improving classifier performance for predicting customer's intended ratings from social media posts," *Benchmarking An Int. J.*, vol. 28, no. 2, pp. 438-464, Jan. 2021, doi: 10.1108/BIJ-01-2020-0004.

[5] B. Hawashin, S. Alzubi, T. Kanan, and A. Mansour, "An efficient semantic recommender method for Arabic text," *Electron. Libr.*, vol. 37, no. 2, pp. 263-280, 2019, doi: 10.1108/EL-12-2018-0245.

[6] P. Virdi, A. D. Kalro, and D. Sharma, "Consumer acceptance of social recommender systems in India," *Online Inf. Rev.*, vol. 44, no. 3, pp. 723-744, Jan. 2020, doi: 10.1108/OIR-05-2018-0177.

[7] J. He and W. W. Chu, "A Social Network-Based Recommender System (SNRS)," pp. 47-74, 2010, doi: 10.1007/978-1-4419-6287-4_4.

[8] A. Estetikha and K. Adhien, "Comparison of K-Means Clustering & Logistic Regression on University data to differentiate between Public and Private University," *IJIIS Int. J. Informatics Inf. Syst.*, vol. 4, no. 1, pp. 21-29, 2021, doi: 10.47738/ijiis.v4i1.74.

[9] Y. Yamato, "Server Structure Proposal and Automatic Verification Technology on IAAS Cloud of Plural Type Servers," *IJIIS Int. J. Informatics Inf. Syst.*, vol. 1, no. 2, pp. 97-106, 2018, doi: 10.47738/ijiis.v1i2.104.

[10] W. C. Lin, S. W. Ke, and C. F. Tsai, "Top 10 data mining techniques in business applications: a brief survey," *Kybernetes*, vol. 46, no. 7, pp. 1158-1170, 2017, doi: 10.1108/K-10-2016-0302.

[11] M. Baigorria Alonso, "Data augmentation using many-to-many rnns for session-aware recommender systems," *ACM WSDM Work. Web Tour. (WSDM WebTour'21)*, 2021, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85146238641&partnerID=40&md5=e38299055b107a0331d321f3bec1708c

[12] G. George and A. M. Lal, "Review of ontology-based recommender systems in e-learning," *Comput. Educ.*, vol. 142, p. 103642, 2019.

[13] D. Maillard, "The Obsolescence of Man in the Digital Society," *Int. J. Appl. Inf. Manag.*, vol. 1, no. 3, pp. 99-124, 2021, doi: 10.47738/ijaim.v1i3.13.

[14] H.-T. Le, "Knowledge Management in Vietnameses Mall and Medium Enterprises: Review of Literature," *Int. J. Appl. Inf. Manag.*, vol. 1, no. 2, pp. 50-59, 2021, doi: 10.47738/ijaim.v1i2.12.

[15] S. Khodabandehlou, S. A. Hashemi Golpayegani, and M. Zivari Rahman, "An effective recommender system based on personality traits, demographics and behavior of customers in time context," *Data Technol. Appl.*, vol. 55, no. 1, pp. 149-174, Jan. 2021, doi: 10.1108/DTA-04-2020-0094.

[16] K. Z. Thwe and T. Yukawa, "Improving Music Recommendation System by Applying Latent Topics of Lyrics," *IJIIS Int. J. Informatics Inf. Syst.*, vol. 2, no. 2, pp. 91-98, 2019, [Online]. Available: http://ijiis.org/index.php/IJIIS/article/view/91

[17] A. L. Pradana and A. Wibowo, "User Demographic Information and Deep Neural Network in Film Recommendation System based on Collaborative Filtering," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 12, no. 5, pp. 139-146, 2022, doi: 10.46338/ijetae0522_16.

[18] F.-Y. Cui, "Quantum-inspired Collaborative Edge Filtering of Electron Microscope Images," *Jiliang Xuebao/Acta Metrol. Sin.*, vol. 43, no. 8, pp. 1027-1035, 2022, doi: 10.3969/j.issn.1000-1158.2022.08.09.

[19] J. Zheng, S. Chen, Y. Du, and P. Song, "A multiview graph collaborative filtering by incorporating homogeneous and heterogeneous signals," *Inf. Process. Manag.*, vol. 59, no. 6, 2022, doi: 10.1016/j.ipm.2022.103072.

[20] K. Hulliyah, "Predicting Airline Passenger Satisfaction with Classification Algorithms," *IJIIS Int. J. Informatics Inf. Syst.*, vol. 4, no. 1, pp. 82-94, 2021, doi: 10.47738/ijiis.v4i1.80.

[21] I. Santiko and P. Subarkah, "Comparison of Cart and Naive Bayesian Algorithm Performance to Diagnose Diabetes Mellitus," *IJIIS Int. J. Informatics Inf. Syst.*, vol. 2, no. 1, pp. 9-16, 2019, doi: 10.47738/ijiis.v2i1.9.

[22] K. Fujishima, "A Study on Hospitality Education at University: Jal's Philosophy Education as an Example," *Int. J. Appl. Inf. Manag.*, vol. 1, no. 3, pp. 136-144, 2021, doi: 10.47738/ijaim.v1i3.15.

[23] Y.-K. Ng, "CBREC: A book recommendation system for children using the matrix factorization and content-based filtering approaches," *Int. J. Bus. Intell. Data Min.*, vol. 16, no. 2, pp. 129-149, 2020, doi: 10.1504/IJBIDM.2020.104738.

[24] L. Fu and X. Ma, "An Improved Recommendation Method Based on Content Filtering and Collaborative Filtering," *Complexity,* vol. 2021, 2021, doi: 10.1155/2021/5589285.

[25] E. Lacic, D. Kowald, L. Eberhard, C. Trattner, D. Parra, and L. B. Marinho, "Utilizing online social networks and location-based data to recommend products and categories in online marketplaces," *Lect. Notes Comput. Sci. (including*

Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 8940, pp. 96-116, 2015, doi: 10.1007/978-3-319-14723-9_6.

[26] L. Li, Z. Zhang, and S. Zhang, "Hybrid Algorithm Based on Content and Collaborative Filtering in Recommendation System Optimization and Simulation," *Sci. Program.*, vol. 2021, 2021, doi: 10.1155/2021/7427409.

[27] Q. Y. Shambour, A. H. Hussein, Q. M. Kharma, and M. M. Abualhaj, "Effective hybrid content-based collaborative filtering approach for requirements engineering," *Comput. Syst. Sci. Eng.*, vol. 40, no. 1, pp. 113-125, 2022, doi: 10.32604/CSSE.2022.017221.

[28] B. Wibowotomo, E. D. S. Rizal, M. I. Akbar, and D. T. Kurniawan, "Cooking Class Recommendation Using Content Based Filtering for Improving Chef Learning Practical Skill," *Int. J. Interact. Mob. Technol.*, vol. 15, no. 8, pp. 71-86, 2021, doi: 10.3991/ijim.v15i08.21581.

[29] Y. Y. Shih and D. R. Liu, "Product recommendation approaches: Collaborative filtering via customer lifetime value and customer demands," *Expert Syst. Appl.*, vol. 35, no. 1-2, pp. 350-360, 2008, doi: 10.1016/j.eswa.2007.07.055.