Technologia : Jurnal Ilmiah
p-ISSN: 2086-6917, e-ISSN: 2656-8047
Homepage: https://ojs.uniska-bjm.ac.id/index.php/JIT/index
Vol.15, No.4, Oktober 2024

Pemanfaatan Sistem Rekomendasi Menggunakan Content-Based Filtering pada Hotel di Palangka Raya

Nova Noor Kamala Sari[1*], Licantik[2] , Marsha Zahra[3]
[123] Jurusan Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya, Indonesia.

Informasi Artikel:
Dikirim: 09-08-2024; Diterima: 25-09-2024; Diterbitkan: 10-10-2024
Doi : http://dx.doi.org/10.31602/tji.v15i4.16282

ABSTRAK

Masalah: Industri perhotelan di Palangka Raya mengalami pertumbuhan pesat, namun wisatawan sering kesulitan menemukan hotel sesuai preferensi mereka dengan cepat. Oleh karena itu, diperlukan sistem rekomendasi yang efisien untuk membantu pengguna menemukan hotel yang tepat.
Tujuan: Penelitian ini bertujuan untuk mengembangkan model sistem rekomendasi hotel di Palangka Raya menggunakan pendekatan Content-Based Filtering.
Metode: Metode yang digunakan adalah Content-Based Filtering dengan mengkombinasikan algoritma TF-IDF dan cosine similarity. Data hotel diperoleh melalui teknik web scraping dari Google Maps, dengan atribut seperti nama hotel, lokasi, jarak, harga, penilaian, dan fasilitas hotel
Hasil: Hasil evaluasi sistem rekomendasi hotel yang dikembangkan memiliki kinerja yang sangat baik dengan metrik akurasi model mencapai 88.89%, Presisi model 88.89%, Recall model 100.00%, F1-score yang menggabungkan presisi dan recall adalah 94.12%.
Kesimpulan: Hasil penelitian menunjukkan bahwa sistem rekomendasi yang dikembangkan mampu memberikan rekomendasi hotel yang relevan dan pentingnya pemilihan fitur seperti fasilitas, lokasi, dan harga, sangat berpengaruh terhadap kualitas rekomendasi.
Kata Kunci: Sistem Rekomendasi, Content Based Filtering, Hotel, Palangka Raya

This is an open-access article under a Creative Commons Attribution 4.0 International (CC-BY 4.0) License. Copyright © 2024 by authors.

Pendahuluan

Industri perhotelan di Palangka Raya mengalami pertumbuhan yang signifikan seiring dengan meningkatnya kunjungan wisatawan dan aktivitas bisnis di kota ini. Pembangunan hotel dan wisma menurut data dari Dinas Pariwisata dan Ekonomi Kreatif Pemerintah Kota Palangka Raya, pada tahun 2021 terdapat 40 hotel dan 48 wisma (Obajha et al., 2023). Namun, dengan banyaknya pilihan hotel yang tersedia, wisatawan seringkali kesulitan menemukan hotel yang sesuai dengan preferensi mereka dalam waktu yang singkat. Hal ini menimbulkan kebutuhan akan sistem rekomendasi yang dapat membantu pengguna menemukan hotel yang tepat dengan cepat dan efisien. Sistem Rekomendasi membantu pengguna memilih produk atau layanan yang sesuai dari berbagai pilihan (Patel et al., 2023) Sistem rekomendasi hadir sebagai solusi dengan menyaring dan mempersonalisasi informasi, sehingga pengguna dapat lebih mudah menemukan konten dan layanan yang sesuai dengan preferensi mereka (Isinkaye et al., 2015). Content-Based Filtering adalah salah satu metode dalam sistem rekomendasi yang menggunakan informasi detail hotel dan preferensi pengguna untuk memberikan rekomendasi yang personal dan relevan (Rymmai & Js, n.d.). Pendekatan ini memungkinkan sistem untuk memahami kebutuhan dan preferensi pengguna secara

mendalam, sehingga dapat memberikan rekomendasi hotel yang paling sesuai berdasarkan filter radius dan rentang harga.

Penelitian sebelumnya telah menunjukkan potensi besar dari sistem rekomendasi Content Based Filtering dalam berbagai domain, termasuk e-commerce dan pariwisata. Nastiti (2019) berhasil menerapkan metode ini untuk merekomendasikan petani dan pedagang dengan profil serupa. Dalam pengujian presisi, sistem
ini menunjukkan hasil rata-rata presisi sebesar 78.40% dari 10 profil pedagang yang diuji dengan 15 rekomendasi kelompok tani teratas (Nastiti, 2019). Putri, Muchayan, & Kamisutara (2020) yang membahas tentang pembangunan sistem rekomendasi produk menggunakan metode Content-Based Filtering dan Term Frequency Inverse Document Frequency (TF-IDF) untuk meningkatkan Customer Relationship Management (CRM). juga membuktikan efektivitas content-based filtering dalam merekomendasikan produk, meningkatkan kepuasan pelanggan dan brand awareness (Putri et al., 2020).
Penelitian ini bertujuan untuk mengembangkan sebuah model sistem rekomendasi hotel di Palangka Raya menggunakan pendekatan Content-Based Filtering dengan mengkombinasikan algoritma TF-IDF dan cosine similarity.

Landasan Teori

1. Hotel

Berdasarkan Peraturan Pemerintah Republik Indonesia No. 65 Tahun 2021, hotel didefinisikan sebagai bangunan yang secara khusus diperuntukkan untuk menginap ataupun beristirahat dan memperoleh pelayanan serta fasilitas dengan dipungut biaya. Hotel memainkan peran penting dalam industri pariwisata dan perhotelan dengan menyediakan tempat menginap yang nyaman dan aman bagi wisatawan, pelancong bisnis, dan tamu lainnya.

2. Sistem Rekomendasi

Sistem rekomendasi merupakan salah satu cabang dalam bidang information retrieval dan kecerdasan buatan. Sistem rekomendasi adalah sebuah perangkat lunak yang berfungsi untuk memprediksi pilihan item berdasarkan karakteristik pengguna (Gopalachari, 2018). Tujuan dari Sistem rekomendasi adalah untuk mengembangkan model matematis atau fungsi objektif yang dapat memprediksi seberapa besar kemungkinan pengguna menyukai suatu item (Arif et al., 2020; Gorakala, 2016). Sistem Rekomendasi menggunakan informasi yang menggambarkan situasi pengguna seperti lokasi, waktu, dan aktivitas untuk menghasilkan rekomendasi yang lebih relevan dan dapat menyesuaikan kebutuhan pengguna (Adomavicius & Tuzhilin, 2015; Asabere, 2013).

3. Content Based Filtering

Content based filtering merupakan metode untuk memilah konten yang sudah ada dan disesuaikan dengan konteks studi kasus sebelum dilakukan pembobotan terhadap algoritma TF- IDF. Secara garis besar Content-based filtering menggunakan konten atribut yang sudah tersedia dan tidak mengambil opini lain untuk merekomendasikan konten serupa (Putri et al., 2020).

Pada penelitian ini Content-based filtering menggunakan TF-IDF sebagai algoritma untuk melakukan perhitungan pembobotan nilai terhadap atribut konten dan pengguna (Misal et al., 2015).

4. TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) adalah metode yang digunakan untuk memberikan bobot pada hubungan sebuah term dalam sebuah dokumen dengan cara mengkombinasikan konsep perhitungan bobot, yaitu *Term Frequency* (TF) yang merupakan frekuensi kemunculan kata dalam dokumen dan *Inverse Document Frequency* (IDF) yang merupakan kebalikannya, yaitu frekuensi kebalikan dari dokumen yang mengandung kata (Cheryl Ayu Melyani, 2022).

$$W_{ij} = TF_{ij} \times \left(\ln\left(\frac{D}{df_j}\right) + 1\right) \qquad (1)$$

dengan :

Wij : Bobot term j terhadap dokumen i
$TFij$: Frekuensi kemunculan term j pada dokumen i
$IDFj$: Frekuensi invers dokumen dari term j
D : Total Dokumen
dfj : Total dokumen dari term j

5. Cosine Similarity

Cosine similarity adalah metode untuk mengukur kemiripan antara dua vektor n-dimensi yang biasanya digunakan dalam bidang pencarian informasi untuk membandingkan dua buah teks atau dokumen. Jika dua buah teks atau dokumen semakin mirip, maka nilai cosine similarity akan semakin mendekati 1, sedangkan jika nilai cosine similarity mendekati 0 maka kedua teks atau dokumen tersebut semakin tidak mirip. Persamaan yang digunakan untuk menghitung cosine similarity:

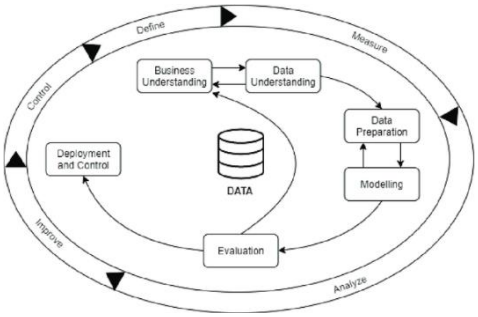
$$Cos \propto = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \qquad (2)$$

dengan :

A : Vektor A
B : Vektor B
Ai : Term I dalam dokumen A
Bi : Term I dalam dokumen B
A.B : Hasil perkalian titik antara vektor A dan vektor B
|A| : Panjang vektor A
|B| : Panjang vektor B
|A||B| : Perkalian silang antara |A|&|B|

Metode

Metode penelitian ini mengikuti tahapan model *Cross-Industry Standard Process for Data Mining* (CRISP-DM). CRISP-DM adalah kerangka kerja yang banyak digunakan dalam proyek data mining dan analisis data. Model ini terdiri dari enam fase utama yang saling berhubungan, yaitu: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment.



Gambar 1 Diagram Alur Penelitian (Huber et al., 2019)

1. Business Understanding

Tujuan bisnis dari penelitian adalah meningkatkan kepuasan pelanggan dengan menyediakan rekomendasi hotel yang lebih relevan dan personal. Tujuan data mining dalam konteks ini meliputi pengembangan model rekomendasi berbasis content-based filtering. Selain itu, data mining bertujuan untuk mengidentifikasi preferensi pelanggan dengan menganalisis data untuk menemukan pola dan preferensi yang dapat digunakan dalam model rekomendasi.

2. Data Understanding

Pada tahap ini, fokus utama adalah mengumpulkan data awal dan memahami data yang tersedia secara mendalam. Data yang dikumpulkan meliputi informasi tentang hotel di Palangka Raya kemudian mendeskripsikan data tersebut untuk memahami struktur dan karakteristiknya. Ini termasuk mengidentifikasi atribut-atribut penting dalam dataset, seperti nama hotel, fasilitas, rating, ulasan, dan preferensi pelanggan. Langkah terakhir dalam tahap data understanding adalah memverifikasi kualitas data untuk memastikan keakuratan dan keandalan data yang melibatkan pembersihan data untuk menangani data yang hilang, duplikat, atau tidak konsisten; validasi data untuk memastikan kesesuaian dengan sumbernya; dan peningkatan kualitas data menggunakan teknik seperti normalisasi dan standardisasi.

3. Data Preparation

Pada tahap ini, data yang telah dikumpulkan akan diproses dan disiapkan untuk dianalisis. Persiapan data melibatkan pengolahan data yang diperoleh secara real-time sebagai dataset. Ketika program dijalankan, data akan diperoleh melalui teknik web scraping. Web scraping dilakukan dengan menggunakan Selenium WebDriver untuk mengakses dan mengontrol browser, menunggu konten dinamis dimuat, dan mengekstraksi data dengan Beautiful Soup. Data yang dikumpulkan mencakup informasi tentang harga, fasilitas, ulasan, dan lokasi hotel. Setelah data diekstraksi, dilakukan filter untuk menghapus entri yang tidak relevan, perhitungan jarak dengan API Geoapify, penyaringan berdasarkan kriteria tertentu, dan penyimpanan data dalam format yang diinginkan.

Langkah pembersihan data melibatkan mengatasi data yang hilang dengan mengidentifikasi dan menangani nilai yang hilang, menghapus data yang duplikat untuk memastikan setiap entri unik dan valid, serta memastikan konsistensi format data seperti format tanggal, satuan harga, dan jenis fasilitas yang tersedia.

Selanjutnya melakukan konstruksi data, yaitu pembuatan fitur yang relevan untuk analisis, seperti fasilitas (Wi-Fi, kolam renang, parkir), harga, dan lokasi. Selain itu, data

dari berbagai sumber digabungkan untuk membentuk satu dataset yang lengkap dan terintegrasi.

4. Modelling

Proses modelling adalah tahapan dalam analisis data dan pengembangan sistem yang bertujuan untuk membangun dan memvalidasi model prediktif atau deskriptif menggunakan data yang telah dikumpulkan dan diproses. Proses ini melibatkan pemilihan algoritma yang sesuai, pelatihan model dengan menggunakan data latih, dan pengujian model untuk memastikan keakuratannya. Dalam konteks sistem rekomendasi, proses modelling melibatkan penggunaan teknik seperti Content Based Filtering untuk mengidentifikasi pola dan hubungan dalam data pengguna dan item, sehingga dapat memberikan rekomendasi yang relevan dan personal. Tahap ini juga mencakup evaluasi performa model menggunakan metrik tertentu, seperti akurasi, presisi, recall dan F-1 score. Proses modeling yang efektif dan tepat sangat penting untuk menghasilkan sistem yang andal dan mampu memberikan nilai tambah bagi pengguna.

5. Evaluation

Proses evaluasi adalah tahap penting dalam pengembangan model, yang bertujuan untuk menilai kinerja model dan memastikan bahwa model tersebut memenuhi tujuan dan persyaratan yang telah ditetapkan.

6. Deployment

Pada penelitian ini, topik pembahasan hanya memberikan hasil yang sebatas suatu pemodelan atau perancangan. Sehingga belum sampai ke tahap deployment

Hasil dan Pembahasan

Proses ini dimulai dengan mengumpulkan data dari Google Maps menggunakan web scraping. Data hotel yang diperoleh, termasuk fasilitas yang disediakan, diekstrak dan diubah menjadi representasi numerik. Harga kamar hotel diubah menjadi nilai numerik dengan fungsi parse_price untuk memungkinkan perbandingan langsung. Fasilitas hotel diubah menjadi teks dan diolah menggunakan TF-IDF (Term Frequency-Inverse Document Frequency) untuk menghitung pentingnya setiap kata kunci dalam deskripsi fasilitas hotel. Cosine similarity digunakan untuk mengukur kemiripan antara deskripsi fasilitas yang diinginkan pengguna dengan fasilitas yang dimiliki hotel. Hotel dengan nilai similarity di atas ambang batas akan direkomendasikan kepada pengguna sebagai pilihan akomodasi yang sesuai dengan preferensi mereka. Pendekatan Content-Based Filtering (CBF) ini memungkinkan rekomendasi personal dan relevan berdasarkan fitur-fitur spesifik hotel yang diperoleh dari Google Maps melalui web scraping.

1) Implementasi API Geoapify

```
# Tentukan radius maksimum dalam meter
max_radius = float(input("Masukkan radius maksimum (dalam meter): "))

# Filter hotel berdasarkan jarak dari lokasi pengguna
filtered_hotels_location = []
for hotel in hotels_data:
    if hotel['latitude'] != "Not Available" and hotel['longitude'] != "Not Available":
        distance = calculate_distance(user_location[0], user_location[1], float(hotel['latitude']), float(hotel['longitude']), api_key)
        if distance <= max_radius:
            hotel['distance'] = distance
            filtered_hotels_location.append(hotel)

# Tampilkan hotel yang telah difilter dalam radius dari lokasi pengguna
print("Hotel yang difilter dalam radius dari lokasi Anda:")
print_hotels_data(filtered_hotels_location)
print()
```

Gambar 1 Program Geoapify

Program menggunakan Geoapify API untuk menghitung jarak antara lokasi pengguna dan setiap hotel dengan mengirim permintaan HTTP berisi koordinat latitude dan longitude. API ini memerlukan kunci API untuk autentikasi dan menggunakan mode transportasi 'drive'. Setelah menerima respons dari API, jarak dalam meter dihitung untuk setiap hotel, memperkaya informasi pencarian hotel dengan detail lokasi relatif. Fitur tambahan memungkinkan pengguna memasukkan jarak tempuh maksimum dalam meter untuk menyaring hotel-hotel yang dapat ditempuh. Setelah proses penyaringan menggunakan nilai jarak maksimum yang dimasukkan pengguna, program menampilkan daftar hotel yang memenuhi kriteria jarak yang ditentukan.

2) Filter Data Hotel Dengan Rentang Harga

```
# Tampilkan hotel yang telah difilter dengan jarak tempuh maksimum dari lokasi pengguna
print("Hotel yang difilter dengan jarak tempuh maksimum dari lokasi Anda:")
print_hotels_data(filtered_hotels_location)
print()

# Meminta pengguna untuk memasukkan rentang harga
min_price = float(input("Masukkan harga minimum : "))
max_price = float(input("Masukkan harga maksimum: "))

# Menyaring hotel dengan rentang harga yang dimasukkan pengguna
filtered_hotels_price = filter_hotels_by_price_range(filtered_hotels_location, min_price, max_price)

# Menampilkan hotel yang telah disaring
print("Hotel yang difilter dalam jarak dan rentang harga dari lokasi Anda:")
print_hotels_data(filtered_hotels_price)
print()
```

Gambar 2 Program Filter Dengan Rentang Harga

Setelah pengguna memasukkan rentang harga minimum dan maksimum melalui input, program menyaring data hotel yang telah difilter berdasarkan jarak dari lokasi pengguna. Tujuannya adalah memastikan hanya hotel-hotel dengan harga sesuai yang ditampilkan. Setelah proses penyaringan, program mencetak daftar hotel yang memenuhi kriteria jarak dan rentang harga. Output mencakup informasi lengkap seperti nama hotel, rating, fasilitas, harga, koordinat lokasi, jarak dari lokasi pengguna, deskripsi, dan nilai akurasi jika tersedia. Dengan informasi yang terstruktur ini, pengguna dapat dengan mudah mengevaluasi dan memilih hotel yang sesuai dengan preferensi harga dan lokasi mereka, membantu mereka membuat keputusan yang tepat untuk perjalanan mereka.

3) Filter Data Hotel Dengan Nilai Rating

```
# Meminta pengguna untuk memasukkan rating minimum
min_rating = float(input("Masukkan rating minimum (1-5): "))

# Menyaring hotel berdasarkan rating minimum
filtered_hotels_rating = []

for hotel in filtered_hotels_price:
    rating_str = hotel.get('rating', '0').replace(',', '.')
    rating = float(rating_str) if rating_str.replace('.', '', 1).isdigit() else 0
    if rating >= min_rating:
        hotel['rating'] = rating  # Update the rating to float
        filtered_hotels_rating.append(hotel)


# Menampilkan hotel yang telah disaring berdasarkan rating
print("Hotel yang difilter dalam jarak, rentang harga, dan rating dari lokasi Anda:")
print_hotels_data(filtered_hotels_rating)
```

Gambar 3 Program Filter Dengan Nilai Rating

Program ini berfungsi untuk menyaring hotel berdasarkan rating minimum yang

dimasukkan oleh pengguna. Pengguna diminta untuk memasukkan nilai rating minimum dalam rentang 1 hingga 5, yang kemudian diubah menjadi tipe data float dan disimpan dalam variabel min_rating. Selanjutnya, program menginisialisasi list kosong filtered_hotels_rating untuk menyimpan hotel-hotel yang memenuhi kriteria rating minimum.

Program melakukan iterasi melalui list filtered_hotels_price, yang berisi hotel-hotel yang sudah difilter berdasarkan jarak dan harga. Untuk setiap hotel dalam list ini, program memeriksa nilai rating dari key rating. Jika nilai rating tersebut lebih besar atau sama dengan min_rating, hotel tersebut ditambahkan ke dalam list filtered_hotels_rating.

4) TF-IDF dan Cosine *Similiarity*

```
# Fasilitas yang diinginkan pengguna
user_input = str(input("Masukkan fasilitas yang Anda inginkan (dipisahkan dengan spasi): "))

# Praproces data untuk membuat korpus deskripsi fasilitas
corpus = [' '.join(hotel['facilities']) for hotel in filtered_hotels_rating]

# Buat TF-IDF vectorizer dan transformasikan korpus
vectorizer = TfidfVectorizer()
tfidf_matrix = vectorizer.fit_transform(corpus)

# Transformasikan input pengguna menggunakan vectorizer yang sama
user_tfidf = vectorizer.transform([user_input])

# Hitung kesamaan kosinus antara input pengguna dan fasilitas hotel
similarity_scores = cosine_similarity(user_tfidf, tfidf_matrix).flatten()

# Tetapkan ambang batas kesamaan untuk menyaring hotel
similarity_threshold = 0.3
filtered_hotels_facilities = [filtered_hotels_rating[i] for i in np.where(similarity_scores >= similarity_threshold)[0]]

print()
# Menampilkan hotel yang telah disaring berdasarkan radius, rentang harga, dan fasilitas
print("Hotel yang difilter dalam radius, rentang harga, rating dan fasilitas dari lokasi Anda:")
print_hotels_data(filtered_hotels_facilities)
```

Gambar 4 Program *TF-IDF* dan *Cosine SImilarity*

Program ini memulai proses filtering hotel berdasarkan preferensi pengguna dengan meminta pengguna untuk memasukkan radius maksimum dalam meter untuk pencarian lokasi hotel dari lokasi pengguna. Hotel-hotel yang memiliki data koordinat yang valid dihitung jaraknya menggunakan fungsi calculate_distance, dan yang berjarak kurang dari atau sama dengan radius maksimum disaring ke dalam filtered_hotels_location.

Selanjutnya, pengguna diminta memasukkan rentang harga yang diinginkan. Hotel-hotel yang telah difilter berdasarkan lokasi kemudian disaring berdasarkan rentang harga yang dimasukkan pengguna, yang disimpan dalam filtered_hotels_price.

Pengguna juga diminta memasukkan rating minimum yang diinginkan untuk hotel. Hotel-hotel yang telah difilter berdasarkan radius dan rentang harga kemudian disaring kembali berdasarkan rating minimum yang dimasukkan pengguna, dan hasilnya disimpan dalam filtered_hotels_rating.

Selain itu, pengguna diminta memasukkan fasilitas yang diinginkan, yang diubah menjadi korpus deskripsi fasilitas hotel. Daftar fasilitas dari setiap hotel yang telah difilter sebelumnya berdasarkan rating diolah menggunakan TF-IDF (Term Frequency-Inverse Document Frequency) vectorizer untuk menghasilkan matriks TF-IDF. Input fasilitas yang dimasukkan pengguna juga diubah menggunakan vectorizer yang sama. Dilakukan perhitungan kesamaan kosinus antara vektor TF-IDF input pengguna dan vektor TF-IDF fasilitas hotel. Hotel-hotel yang memiliki kesamaan kosinus di atas ambang batas yang ditetapkan (0.3 dalam hal ini) akan disaring dan disimpan dalam filtered_hotels_facilities.

5) Confusion Matrix

```
# Menghitung akurasi
total_predictions = tp + fp + fn + tn
accuracy = ((tp + tn) / total_predictions) * 100 if total_predictions != 0 else 0

# Menghitung presisi
precision = (tp / (tp + fp)) * 100 if (tp + fp) != 0 else 0

# Menghitung recall
recall = (tp / (tp + fn)) * 100 if (tp + fn) != 0 else 0

# Menghitung F1-score
f1_score = 2 * (precision * recall) / (precision + recall) if (precision + recall) != 0 else 0
```

Gambar 5 Program *Confusion Matrix*

Perhitungan confusion matrix dalam evaluasi sistem rekomendasi hotel bertujuan untuk membandingkan hasil prediksi dengan data aktual. Data asli hotel dan hasil prediksi dimuat dari file Excel terpisah, kemudian kolom 'name' dari kedua file diekstrak dan dikonversi menjadi set untuk memfasilitasi perhitungan. True Positives (TP) dihitung sebagai elemen yang ada di kedua set, False Positives (FP) sebagai elemen yang hanya ada di set prediksi, dan False Negatives (FN) sebagai elemen yang hanya ada di set data asli. True Negatives (TN) diasumsikan nol dalam konteks ini. Confusion matrix 2x2 disusun dengan format [[TP, FP], [FN, TN]], yang mencerminkan performa sistem dalam mengenali hotel yang relevan. Program juga menyajikan evaluasi dengan metrik seperti akurasi, presisi, recall, dan F1-score, yang memberikan gambaran tentang kinerja sistem dalam merekomendasikan hotel berdasarkan preferensi pengguna dengan membandingkan data asli dan hasil prediksi dari sistem rekomendasi.

6) Evaluasi Hasil Perhitungan *Confusion Matrix*

Tabel 1. Hasi Confusion Matrix

No	Nama	Fr
1	Akurasi	88,89%
2	Presisi	88,89%
3	Recall	100%
4	F1- Score	94,12%

Dengan data ini, akurasi model mencapai 88.89%, yang menunjukkan bahwa 88.89% dari total prediksi adalah benar. Presisi model adalah 88.89%, menandakan bahwa setiap prediksi positif yang dibuat oleh model adalah benar. Recall model adalah 100.00%, menunjukkan bahwa model berhasil mendeteksi 100% dari semua kasus positif yang ada. F1-score, yang menggabungkan presisi dan recall, adalah 94.12%, menandakan keseimbangan yang baik antara presisi dan recall.

Sistem rekomendasi hotel menggunakan Content-Based Filtering telah dikembangkan untuk memberikan rekomendasi hotel kepada pengguna berdasarkan preferensi mereka. Untuk mengevaluasi kinerja sistem ini, dilakukan pengujian menggunakan dua set data: data asli (ground truth) dan data hasil CBF.

Tabel 2. *Ground Truth*

name
Luwansa Hotel Palangka Raya
Amaris Hotel Palangkaraya
RedDoorz Plus @ Orange GH Diponogoro
Putra Kahayan Hotel
SUJA Menteng Guest House
Surya Kahayan Hotel
AM HOTEL & AM GUEST HOUSE
Borneo Homestay Palangkaraya

Tabel 3. Data Hasil CBF

name
Luwansa Hotel Palangka Raya
Amaris Hotel Palangkaraya
RedDoorz Plus @ Orange GH Diponogoro
Putra Kahayan Hotel
SUJA Menteng Guest House
Surya Kahayan Hotel
AM HOTEL & AM GUEST HOUSE
RedDoorz near Universitas Palangkaraya 3
Borneo Homestay Palangkaraya

Kesimpulan

Penelitian ini mengimplementasi sistem rekomendasi hotel di Kota Palangka Raya menggunakan Content Based Filtering. Penelitian difokuskan pada eksplorasi model untuk membangun sistem yang memberikan rekomendasi hotel berdasarkan atribut seperti lokasi, harga, dan fasilitas hotel.

Content Based Filtering ini memungkinkan rekomendasi hotel yang personal dan relevan berdasarkan preferensi fasilitas yang diinginkan pengguna, dengan memanfaatkan data yang dikumpulkan dari Google Maps melalui teknik web scraping. Data hotel, termasuk harga dan deskripsi fasilitas, diolah menjadi representasi numerik menggunakan metode seperti parsing harga dan analisis TF-IDF untuk deskripsi fasilitas. Nilai cosine similarity kemudian digunakan untuk mengukur kesamaan antara fasilitas yang diinginkan pengguna dengan yang tersedia di hotel, sehingga hanya hotel-hotel yang relevan dan sesuai dengan preferensi pengguna yang direkomendasikan.

Evaluasi hasil program menggunakan Confusion Matrix menunjukkan bahwa hasil dari Content Based Filtering (CBF) dan data asli memiliki delapan data hasil CBF yang sama dengan delapan data asli dan satu data hasil CBF yang tidak ada di data asli. Ini menghasilkan nilai akurasi model mencapai 88.89%, yang menunjukkan bahwa 88.89% dari total prediksi adalah benar. Presisi model adalah 88.89%, menandakan bahwa setiap prediksi positif yang dibuat oleh model adalah benar. Recall model adalah 100.00%, menunjukkan bahwa model berhasil mendeteksi 100% dari semua kasus positif yang ada. F1-score, yang menggabungkan presisi dan recall, adalah 94.12%. Hal ini menunjukkan bahwa sistem yang dikembangkan sangat akurat dalam memberikan rekomendasi yang sesuai dengan preferensi pengguna.

Referensi

Adomavicius, G., & Tuzhilin, A. (2015). *Context-Aware Recommender Systems, Recommender Systems Handbook*. Springer US.

Arif, Y. M., Nurhayati, H., Kurniawan, F., Nugroho, S. M. S., & Hariadi, M. (2020). Blockchain-Based Data Sharing for Decentralized Tourism Destinations Recommendation System. *International Journal of Intelligent Engineering and Systems*, *13*(6), 472–486. https://doi.org/10.22266/ijies2020.1231.42

Asabere, N. Y. (2013). Towards a Viewpoint of Context-Aware Recommender Systems (CARS) and Services. *International Journal of Computer Science and Telecommunications*, *4*(1).

Cheryl Ayu Melyani. (2022). Hotel Recommendation System with Content-Based Filtering Approach (Case Study: Hotel in Yogyakarta on Nusatrip Website). *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, *15*(1). https://doi.org/10.36456/jstat.vol15.no1.a5375

Gopalachari, M. V. (2018). DBT recommender: Improved trustworthiness of ratings through de-biasing tendency of users. *International Journal of Intelligent Engineering and Systems*, *11*(2), 85–92. https://doi.org/10.22266/IJIES2018.0430.10

Gorakala, S. K. (2016). *Building Recommendation Engines*. Packt Publishing.

Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. *Procedia CIRP*, *79*, 403–408. https://doi.org/10.1016/J.PROCIR.2019.02.106

Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, *16*(3), 261–273. https://doi.org/10.1016/j.eij.2015.06.005

Misal, M., Nagaraj, U., Student, M. E., & Head, P. (2015). Recommender System Methods and Feedback Mechanisms: A Survey. *International Journal of Innovative Research in Computer and Communication Engineering (An ISO*, *3*(11), 11710–11716. https://doi.org/10.15680/IJIRCCE.2015

Nastiti, P. (2019). Penerapan Metode Content Based Filtering Dalam Implementasi Sistem Rekomendasi Tanaman Pangan. *Teknika*, *8*(1), 1–10. https://doi.org/10.34148/teknika.v8i1.139

Obajha, K., Sari, N. N. K., & Pranatawijaya, V. H. (2023). *Implementasi Metode Collaborative Filtering pada Aplikasi Rekomendasi Hotel dan Wisma di Kota Palangka Raya Berbasis Website* (Vol. 3, Issue 2).

Patel, D., Patel, F., & Chauhan, U. (2023). Recommendation Systems: Types, Applications, and Challenges. *International Journal of Computing and Digital Systems*, *13*(1). https://doi.org/10.12785/ijcds/130168

Putri, M. W., Muchayan, A., & Kamisutara, M. (2020). *Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode Content-Based Filtering dan TF-IDF* (Vol. 3, Issue 1).

Rymmai, R. G., & Js, S. (n.d.). Book Recommendation using Cosine Similarity. *International Journal of Advanced Research in Computer Science*, *8*(3). www.ijarcs.info