

PREDIKSI REKOMENDASI WISATA DI INDONESIA MENGGUNAKAN METODE HYBRID CBF BERBASIS TF-IDF, COSINE SIMILARITY DAN NEURAL NETWORK

Anggi Febrina¹, Dwi Elshie², Sabily Almuhtadi Billah³, Yuly Asterina⁴, Muhammad Sony Maulana⁵

^{1,2,3,4,5}Universitas Bina Sarana Informatika, Pontianak, Indonesia

Email: ¹ anggifebrinaaa@gmail.com, ² dwielshie0@gmail.com, ³ sabilialmuhtadi@gmail.com, ⁴ yullyasterina5@gmail.com, ⁵ sony@bsi.ac.id

ABSTRAK

Penelitian ini mengembangkan model prediksi rekomendasi destinasi wisata di Indonesia dengan pendekatan *hybrid* yang mengintegrasikan analisis konten berbasis *TF-IDF* dan *Cosine Similarity* dengan pemodelan *Neural Network*. Permasalahan utama yang diangkat adalah tingginya kompleksitas dalam menentukan destinasi wisata yang relevan akibat banyaknya pilihan dan keragaman karakteristik antar destinasi. Dataset yang digunakan terdiri dari 935 data destinasi wisata yang mencakup atribut teks dan numerik. Metode *TF-IDF* digunakan untuk mengekstraksi fitur dari deskripsi destinasi, kemudian *Cosine Similarity* menghitung tingkat kesesuaiannya dengan preferensi pengguna. Sementara itu, *Neural Network* memproses fitur numerik seperti kategori, provinsi, dan panjang deskripsi untuk menghasilkan skor prediksi relevansi tambahan. Kedua skor digabungkan melalui pendekatan *hybrid* untuk memperoleh hasil rekomendasi yang lebih akurat. Hasil evaluasi menunjukkan bahwa model *hybrid* memiliki performa yang sangat baik dengan *Precision@250* sebesar 0,90 dan *Recall@250* sebesar 0,97, menandakan bahwa model tidak hanya mampu memilih destinasi yang relevan, tetapi juga berhasil menangkap hampir seluruh destinasi yang sesuai dengan preferensi pengguna. Secara keseluruhan, pendekatan *hybrid* ini memberikan rekomendasi yang lebih komprehensif dibandingkan metode tunggal dan berpotensi diterapkan dalam sistem rekomendasi wisata skala nasional.

Keywords: Rekomendasi, Wisata, TF-IDF, Hybrid, Cosine

ABSTRACT

This research aims to develop a predictive recommendation model for tourist destinations in Indonesia using a hybrid approach that integrates TF-IDF and Cosine Similarity-based content analysis with Neural Network modeling. The main problem addressed is the high complexity in determining relevant tourist destinations due to the numerous choices and diversity of characteristics among destinations. The dataset used consists of 935 tourist destination data points, including text and numeric attributes. The TF-IDF method is used to extract features from destination descriptions, followed by Cosine Similarity to calculate their compatibility with user preferences. Meanwhile, the Neural Network processes numeric features such as category, province, and description length to generate additional relevance prediction scores. These two scores are combined through a hybrid approach to achieve more accurate recommendation results. Evaluation results show that the hybrid model performs very well with *Precision@250* of 0.90 and *Recall@250* of 0.97, indicating that the model not only selects relevant destinations but also successfully captures nearly all destinations matching user preferences. Overall, this hybrid approach provides more comprehensive recommendations compared to single methods and has the potential to be implemented in a national-scale tourism recommendation system.

Keywords: Recommendation, Tourism, TF-IDF, Hybrid, Cosine

1. PENDAHULUAN

Indonesia sebagai negara kepulauan memiliki kekayaan alam, budaya, dan sejarah yang tersebar di ribuan destinasi wisata. Keragaman tersebut menjadikan proses pencarian destinasi yang sesuai dengan minat pribadi wisatawan menjadi kompleks dan tidak mudah. Menurut Novianti et al. (2020) [1] banyaknya pilihan serta perbedaan karakteristik setiap lokasi kerap menyulitkan wisatawan dalam memilih tempat yang paling relevan dengan preferensi mereka [2]. Menurut data Badan Pusat Statistik, jumlah perjalanan wisatawan domestik pada tahun 2024 mencapai 1,2 miliar kunjungan merupakan rekor sepanjang sejarah dan pertama kalinya melampaui angka miliar perjalanan dalam satu tahun [3].

Peningkatan jumlah perjalanan tersebut mengindikasikan bahwa kebutuhan terhadap prediksi relevansi destinasi wisata yang akurat dan personal menjadi semakin penting.

Meskipun informasi tentang destinasi wisata tersedia secara melimpah tidak semua informasi tersebut mampu memberikan pemahaman yang mendalam mengenai sejauh mana suatu destinasi sesuai dengan kebutuhan atau preferensi pengunjung. Banyak wisatawan masih bergantung pada ulasan umum atau popularitas destinasi. Padahal pendekatan semacam ini tidak mencerminkan karakteristik konten destinasi secara komprehensif. Studi terdahulu mengungkapkan bahwa metode berbasis popularitas kerap menghasilkan rekomendasi yang kurang tepat, karena cenderung memprioritaskan destinasi yang sudah populer dan mengesampingkan destinasi lain yang sebenarnya relevan namun belum banyak dikenal.

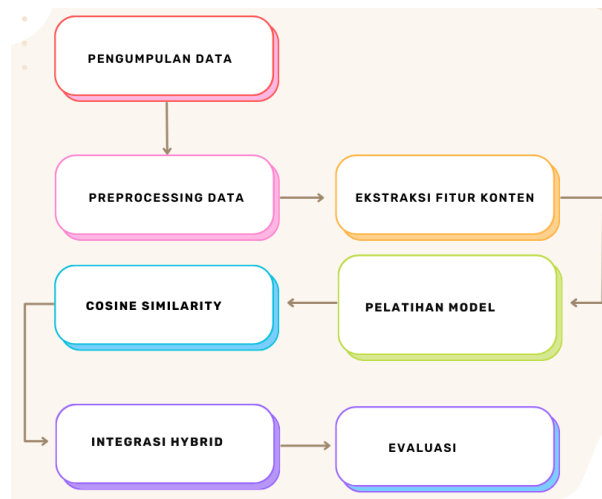
2. TINJAUAN PUSTAKA

Penelitian pada bidang prediksi relevansi wisata sebelumnya menunjukkan penggunaan beragam metode. Pada penelitian Yuliana dkk. (2024) [4] memprediksi kesesuaian destinasi wisata dengan menggunakan *Neural Collaborative Filtering* dengan cakupan studinya hanya Kota Batu dan tidak melibatkan perbandingan dengan pendekatan lainnya. Pendekatan berbasis *Content-Based Filtering* yang memanfaatkan *TF-IDF* dan *Cosine Similarity* juga dilakukan oleh penelitian Sifa'ul Husna dkk. (2024) [5] namun metode tersebut hanya berfokus pada kesesuaian kata kunci, sehingga kurang mampu menangkap makna variasi fitur non-teks secara komprehensif. Beberapa penelitian lain menggunakan metode berbasis analisis teks, seperti penggunaan algoritma C4.5 untuk memprediksi tren wisata [6], *text mining* dalam mengevaluasi pengelolaan destinasi wisata [7], serta algoritma *Bellman-Ford* guna menentukan rute wisata terbaik [8]. Meskipun efektif dalam konteks spesifik, pendekatan-pendekatan tersebut tidak dirancang untuk menangkap hubungan semantik antardestinasi dan tidak mampu memprediksi relevansi destinasi berdasarkan fitur teks maupun non-teks.

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model prediksi rekomendasi destinasi wisata di Indonesia dengan menggabungkan metode *TF-IDF* dan *Cosine Similarity* dalam ekstraksi fitur teks, serta memanfaatkan *Neural Network* untuk memproses fitur non-teks. Model yang diusulkan diharapkan dapat menghasilkan prediksi relevansi yang lebih akurat dibandingkan dengan pendekatan tunggal yang digunakan dalam studi-studi sebelumnya [9]. Dengan menggabungkan kedua metode ini, model rekomendasi yang dihasilkan tidak hanya bergantung pada kesamaan kata kunci, tetapi juga memperhitungkan aspek struktural destinasi. Pendekatan *Hybrid* ini diharapkan mampu menghasilkan prediksi relevansi destinasi yang lebih menyeluruh dan selaras dengan preferensi wisatawan dalam cakupan nasional.

3. METODE PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif berbasis pemodelan *machine learning* untuk melakukan prediksi rekomendasi destinasi wisata di Indonesia. Metode yang diterapkan merupakan kombinasi *Content-Based Filtering* menggunakan *TF-IDF* dan *Cosine Similarity* serta model *Neural Network* untuk memproses fitur non-teks. Secara umum, tahapan penelitian digambarkan pada Gambar 1. Proses penelitian terdiri dari tujuh langkah utama, yaitu pengumpulan data, *preprocessing* data, ekstraksi fitur konten, perhitungan *cosine similarity*, pelatihan model, integrasi *hybrid*, dan evaluasi.



Gambar 1. Alur Penelitian

3.1 Pengumpulan Data

Data diperoleh dari dataset publik yang tersedia di *platform Kaggle*, berupa kumpulan informasi mengenai destinasi wisata di seluruh Indonesia yang terdiri atas 935 data. Dataset tersebut mencakup sejumlah atribut, antara lain (*name, city, province, island, category, rating, lat, lon, address, price, description, image, source*). Data ini masih berupa data mentah sehingga memerlukan proses pembersihan data.

3.2 Preprocessing Data

Tahap *preprocessing* dilakukan untuk menyiapkan data agar siap diproses oleh model secara optimal. Langkah pertama adalah menghapus kolom yang tidak relevan [10] seperti, *rating, city, image, price, dan source*. Selanjutnya, nilai yang hilang (*missing values*) pada kolom kategorikal diisi menggunakan nilai modus agar tidak mengganggu proses analisis. Setelah itu, dilakukan standarisasi nama kolom untuk menyatukan variasi penamaan atribut dalam dataset. Penyesuaian yang diterapkan meliputi *name* menjadi *nama_wisata*, *province* menjadi *provinsi*, *island* menjadi *kota_kabupaten*, *address* menjadi *alamat*, dan *description* menjadi *deskripsi*. Tahap *preprocessing* ditutup dengan penghapusan data duplikat guna mencegah bias dalam pelatihan dan evaluasi model. Hasil akhirnya yaitu dataset yang bersih, konsisten, dan siap untuk ekstraksi fitur.

3.3 Ekstraksi Fitur Konten TF-IDF

Ekstraksi fitur dilakukan dengan menggunakan metode *TF-IDF* (*Term Frequency-Inverse Document Frequency*) untuk mengubah teks deskripsi destinasi menjadi representasi numerik. Langkah-langkah yang dilakukan dalam proses ini meliputi penggabungan beberapa atribut teks menjadi satu kolom analisis tunggal yaitu, (*nama_wisata + kategori + provinsi + alamat + deskripsi_bersih*). Selanjutnya yaitu menerapkan *TF-IDF Vectorizer* dengan konfigurasi seperti menggunakan daftar *stopwords* bahasa Indonesia dari NLTK (*Natural Language Toolkit*) [11], memanfaatkan *n-gram* rentang 1-2, dan menetapkan parameter *min_df* = 2 dan *max_df* = 0.9 untuk menyaring istilah yang terlalu jarang atau terlalu umum sehingga mengurangi noise dalam data teks. Hasil dari tahap ini berupa *matriks TF-IDF* yang berisi bobot numerik setiap kata untuk setiap destinasi wisata serta siap digunakan dalam perhitungan kesamaan konten.

3.4 Perhitungan Cosine Similarity

Cosine Similarity digunakan untuk mengukur tingkat kesamaan antara preferensi wisatawan dan setiap destinasi wisata. Preferensi wisatawan direpresentasikan dalam bentuk daftar kategori yang diinginkan. Berdasarkan representasi teks dari destinasi (hasil ekstraksi *TF-IDF*) dan vektor preferensi pengguna, dihitung nilai *cosine similarity* untuk setiap destinasi [12]. Hasil dari tahapan ini berupa skor kesamaan yang mencerminkan sejauh mana masing-masing destinasi sesuai dengan preferensi pengguna berdasarkan konten teks sebagai skor awal rekomendasi berbasis kemiripan konten.

3.5 Pelatihan Model Neural Network

Selain fitur teks, penelitian ini juga memanfaatkan fitur numerik melalui *Neural Network* untuk menangkap pola non-teks [13]. Fitur yang digunakan meliputi *kategori_enc*, *provinsi_enc*, *desc_len*, dan *is_preferred_category*. Prosesnya meliputi penskalaan fitur menggunakan *MinMaxScaler*, pembagian data menjadi 80% data latih dan 20% data uji, serta pelatihan model *Neural Network* dengan tiga lapisan tersembunyi, aktivasi ReLU, *dropout* untuk mengurangi *overfitting*, dan fungsi aktivasi *sigmoid* pada lapisan output. Model dilatih selama 30 *epoch* menggunakan *optimizer* Adam dan fungsi loss MSE, menghasilkan *predicted score* yaitu skor relevansi berdasarkan fitur non-teks [4].

3.6 Integrasi Hybrid

Tahap *Hybrid* menggabungkan dua skor utama yaitu *cosine score* (yang mencerminkan kemiripan berdasarkan konten teks) dan *model score* (yang merupakan prediksi dari *Neural Network* berdasarkan fitur numerik) [14]. Kedua skor digabungkan dengan bobot 70% untuk *cosine score* dan 30% untuk *model score*. Pendekatan *Hybrid* ini menghasilkan prediksi yang lebih akurat karena memadukan aspek semantik dari teks dengan pola terstruktur dari fitur non-teks.

3.7 Evaluasi

Evaluasi dilakukan menggunakan *Precision@K* dan *Recall@K*, sesuai praktik umum pada prediksi rekomendasi. Evaluasi dilakukan dengan menentukan label “relevan” berdasarkan kesesuaian dengan preferensi wisatawan, mengambil top-K prediksi, serta menghitung *Precision@K* dan *Recall@K*. Tahapan evaluasi ini digunakan untuk menilai seberapa baik model mengenali destinasi wisata yang cocok dengan preferensi wisatawan [15].

4. PEMBAHASAN

Bagian ini menyajikan hasil pengolahan data dan analisis kinerja model rekomendasi sesuai alur metodologi. Pembahasan disusun secara terstruktur, mencakup kondisi dataset awal, hasil tahap *preprocessing*, proses ekstraksi fitur, hingga evaluasi performa model *Hybrid* yang menggabungkan *TF-IDF* dan *Neural Network*.

4.1 Hasil Pengumpulan Data

Dataset awal yang digunakan dalam penelitian ini terdiri dari 935 destinasi wisata yang tersebar di berbagai provinsi di Indonesia. Dataset tersebut mencakup:

Tabel 1. Data Katalog Wisata

Dataset loaded from	Shape	Columns
Data Rekomendasi Wisata.xlsx	(935, 13)	['name', 'city', 'province', 'island', 'category', 'rating', 'lat', 'lon', 'address', 'price', 'description', 'image', 'source']

Atribut-atribut ini menyediakan informasi dasar mengenai identitas destinasi, lokasi geografis, kategori wisata, serta deskripsi yang menjadi komponen utama dalam analisis berbasis konten. Atribut-atribut tersebut menjadi dasar dalam proses selanjutnya, termasuk pembersihan data, standarisasi nama kolom, serta ekstraksi fitur, dengan demikian, kondisi data mentah ini menjadi titik awal yang menentukan alur pemodelan prediksi relevansi destinasi wisata dalam penelitian ini.

4.2 Hasil Preprocessing Data

Tahap *preprocessing* menghasilkan dataset yang lebih bersih, konsisten, dan siap digunakan untuk ekstraksi fitur. Langkah pertama adalah menghapus kolom yang tidak relevan dengan pemodelan seperti *rating*, *city*, *image*, *price*, dan *source*. Proses penghapusan kolom tersebut ditunjukkan pada **Gambar 2**, yang memperlihatkan kode yang digunakan dalam pembersihan data.

```
# Menghapus kolom yang tidak diperlukan
df = df.drop(columns=['rating', 'city', 'image', 'price', 'source'], errors='ignore')

# Menampilkan kolom setelah penghapusan
print("Kolom setelah dihapus:", df.columns.tolist())
```

Gambar 2. Proses Penghapusan Data

Setelah kolom dihapus, struktur dataset berubah dan hanya menyisakan atribut inti yang relevan untuk pemodelan. Kolom yang tersisa ditampilkan pada Tabel 2, yang menunjukkan atribut-atribut penting seperti *name*, *province*, *island*, *category*, *lat*, *lon*, *address*, dan *description*. Atribut-atribut inilah yang kemudian digunakan dalam proses ekstraksi fitur dan pemodelan selanjutnya.

Tabel 2. Hasil Penghapusan Data

Kolom setelah dihapus
name
province
island
category
lat
lon
address
description

Selanjutnya dilakukan pemeriksaan terhadap nilai hilang (*missing value*). Berdasarkan keluaran kode yang ditampilkan pada Tabel 3, terlihat bahwa seluruh kolom memiliki jumlah nilai hilang sebesar nol (0).

Tabel 3. Jumlah *Missing Value*

Nama Kolom	Jumlah MV
name	0
province	0
island	0
category	0
lat	0
lon	0
address	0
description	0

Hasil pemeriksaan menunjukkan bahwa dataset berada dalam kondisi lengkap, sehingga tidak diperlukan proses imputasi untuk menangani nilai hilang setelah memastikan tidak terdapat *missing value*. Langkah selanjutnya adalah standarisasi nama kolom guna menyatukan variasi penamaan atribut dalam dataset, penyesuaian yang dilakukan meliputi dapat dilihat pada Gambar 3.

```
# mapping common variants (standarisasi nama kolom)
if 'name' in df.columns:
    col_map['name'] = 'nama_wisata'

if 'province' in df.columns:
    col_map['province'] = 'provinsi'

if 'island' in df.columns:
    col_map['island'] = 'kota_kabupaten'

if 'category' in df.columns:
    col_map['category'] = 'kategori'

if 'lat' in df.columns:
    col_map['lat'] = 'garis_lintang'

if 'lon' in df.columns:
    col_map['lon'] = 'garis_bujur'

if 'address' in df.columns:
    col_map['address'] = 'alamat'

if 'description' in df.columns:
    col_map['description'] = 'deskripsi'
```

Gambar 3. Proses Standarisasi Nama Kolom

Tahap *preprocessing* ditutup dengan pengecekan duplikasi menggunakan fungsi *df.duplicated()* pada seluruh kolom untuk memastikan tidak ada entri yang muncul lebih dari satu kali. Tujuannya adalah menghindari bias dalam pelatihan model yang bisa muncul akibat data berulang. Hasil pengecekan menunjukkan bahwa jumlah baris duplikat yaitu nol (0).

```
# Duplikat
print("\nJumlah baris duplikat:", df.duplicated().sum())
dup = df[df.duplicated(keep=False)]
print("Jumlah duplikat berdasarkan semua kolom:", dup.shape[0])
display(dup.head())

Jumlah baris duplikat: 0
Jumlah duplikat berdasarkan semua kolom: 0
```

Gambar 4. Output Pengecekan Data Duplikat

Artinya, data bebas duplikat dan semua entri destinasi wisata bersifat unik. Hal ini menjamin bahwa tidak ada destinasi yang diberi bobot berlebih karena kemunculan ganda.

4.3 Hasil Ekstraksi Fitur *TF-IDF*

Untuk mencapai representasi konten yang lebih utuh, beberapa atribut teks seperti *nama_wisata*, *kategori*, *provinsi*, *alamat*, dan *deskripsi_bersih* digabungkan menjadi satu model tunggal. Penggabungan ini memungkinkan model menangkap informasi deskriptif secara komprehensif. Proses tersebut dilihat pada Gambar 5, yang menampilkan kode pemrosesan teks sebelum dilanjutkan ke tahap vektorisasi.

```
# Combine relevant text columns for TF-IDF (EKSTRAKSI FITUR KONTEN)
filtered_df['combined_text'] = filtered_df[[
    'nama_wisata', 'kategori', 'provinsi', 'alamat', 'deskripsi_bersih'
]].astype(str).agg(' '.join, axis=1)
```

Gambar 5. Proses Penggabungan Atribut Teks

Selanjutnya, dilakukan proses vektorisasi menggunakan *TF-IDF Vectorizer* dengan konfigurasi seperti pemanfaatan *stopwords* bahasa Indonesia dari NLTK, penggunaan rentang *n-gram* 1–2 (unigram dan bigram), serta penyetelan parameter *min_df* = 2 dan *max_df* = 0.9 untuk menyaring kata yang terlalu jarang atau terlalu umum. Hasil dari proses ini berupa *matriks TF-IDF* yang merepresentasikan bobot setiap kata atau frasa untuk setiap destinasi wisata.

```
tfidf = TfidfVectorizer(
    stop_words=stopwords_indonesia,
    ngram_range=(1,2), # menambah bigram
    max_df=0.9,
    min_df=2
)
tfidf_matrix = tfidf.fit_transform(filtered_df['combined_text'])
```

Gambar 6. Proses *TF-IDF Vectorizer*

Matriks TF-IDF tersebut kemudian digunakan sebagai dasar perhitungan *Cosine Similarity* untuk mengukur tingkat kesamaan antar destinasi. Semakin tinggi nilai kesamaan yang dihasilkan, semakin relevan destinasi tersebut terhadap preferensi wisatawan berdasarkan konten teksnya.

4.4 Hasil Perhitungan *Cosine Similarity*

Tahap perhitungan *Cosine Silimarity* dilakukan untuk mengukur seberapa mirip preferensi wisatawan dengan masing-masing destinasi wisata berdasarkan representasi teks yang telah diubah ke dalam bentuk vektor menggunakan *TF-IDF*. Hasilnya berupa skor kemiripan dalam rentang 0 hingga 1, di mana skor yang lebih mendekati 1 mengindikasikan kesamaan konten yang lebih tinggi antara preferensi wisatawan dan deskripsi destinasi. Sepuluh destinasi dengan skor *Cosine Similarity* tertinggi ditampilkan dalam tabel berikut:

Tabel 4. Top 10 Score *Cosine Similarity*

No.	nama_wisata	score
1	Gunung Sumbing	0.360635
2	Pantai Rajegwesi	0.360093
3	Gunung Sindoro	0.344188
4	Pantai Tambakerjo	0.321863
5	Pantai Popoh	0.315489
6	Gunung Gede	0.309011
7	Pantai Plengkung	0.307481
8	Pantai Kartini	0.307024
9	Pantai Soge	0.305397
10	Taman Nasional Baluran	0.303838

Berdasarkan Tabel 4, nilai *Cosine Similarity* menunjukkan bahwa destinasi dengan skor tertinggi adalah Gunung Sumbing (0,3606), diikuti oleh Pantai Rajegwesi (0,3601) dan Gunung Sindoro (0,3442) yang mengindikasikan bahwa konten teks maupun kategori dari destinasi-destinasi tersebut cukup selaras dengan preferensi yang diberikan. Secara keseluruhan sepuluh destinasi wisata yang paling relevan berdasarkan kedekatan semantik yang diukur melalui *Cosine Similarity*.

4.5 Hasil Pelatihan Model *Neural Network*

Pada tahap ini dilakukan pelatihan model *Neural Network* untuk memproses fitur numerik yang telah melalui proses *scaling* dan *encoding*. Model dibangun menggunakan tiga lapisan tersembunyi dengan aktivasi *ReLU*, serta dilengkapi dropout untuk mengurangi risiko *overfitting*. Struktur model terdiri dari empat input kemudian dua lapisan *Dense* berturut 124 dan 64 unit, dan satu lapisan output *sigmoid* yang menghasilkan nilai skor relevansi.

```
# Modeling dengan TensorFlow
model = Sequential([
    Input(shape=(X_train.shape[1],)), # gunakan shape dinamis (4 fitur)
    Dense(124, activation='relu'),
    Dropout(0.3),
    Dense(64, activation='relu'),
    Dropout(0.2),
    Dense(1, activation='sigmoid')
])

model.compile(optimizer=Adam(learning_rate=0.001), loss='mse', metrics=['mae'])
print("\nModel summary:")
model.summary()

# Train model
history = model.fit(X_train, y_train, epochs=30, batch_size=16, validation_split=0.2, verbose=1)
```

Gambar 7. Proses Pelatihan Model *Neural Network*

Selama proses pelatihan, model menunjukkan penurunan nilai loss dan MAE secara bertahap, yang mengindikasikan bahwa model mampu mempelajari pola hubungan dalam fitur numerik. Validasi juga menghasilkan tren yang stabil, sehingga model dinilai cukup baik untuk menghasilkan prediksi skor relevansi berbasis numerik. Nilai output model berupa *predicted score*, yaitu skor prediksi yang mencerminkan relevansi destinasi wisata terhadap preferensi pengguna berdasarkan pola fitur non-teks.

4.6 Hasil Integrasi Hybrid

Tahap integrasi *hybrid* dilakukan dengan menggabungkan dua sumber skor utama, yaitu hasil dari *cosine similarity* menggunakan representasi *TF-IDF* dan skor prediksi yang dihasilkan oleh model *Neural Network*. Sebelum digabungkan, kedua skor tersebut dinormalisasi ke dalam rentang yang sama menggunakan *MinMaxScaler* untuk memastikan kesetaraan skala. Skor akhir diperoleh dengan memberikan bobot 60% pada skor berbasis *TF-IDF* dan 40% pada skor prediksi dari *Neural Network*. Nilai gabungan ini disebut juga sebagai *hybrid score*, kemudian digunakan sebagai dasar untuk memeringkat destinasi wisata dalam rekomendasi akhir.

Tabel 5. Top 10 Hybrid Score

nama_wisata	provinsi	kategori	score	predicted_score	final_score
Gunung Sumbing	Jawa Tengah	Gunung	0.3606354583 5378025	0.199034 89947319 03	0.96152845 77945924
Gunung Sindoro	Jawa Tengah	Gunung	0.3441879980 1089395	0.199034 89947319 03	0.93416432 60179248
Pantai Rejengwesi	Jawa Timur	Pantai	0.3600934368 0512307	0.171718 04606914 52	0.91090438 98174745
Gunung Masurai	Jambi	Gunung	0.2967457120 2936806	0.201479 06243801 117	0.85968205 48595672
Gunung Puntang	Jawa Barat	Gunung	0.2791648305 411571	0.216456 56228065 49	0.85769434 60393123
Pantai Soge	Jawa Timur	Pantai	0.3053966816 7878636	0.190893 65005493 164	0.85480730 45729432
Pantai Bandalit	Jawa Timur	Pantai	0.2976392608 8870146	0.190893 65005493 164	0.84190105 15514087
Pantai Pelang	Jawa Timur	Pantai	0.2976041184 5344803	0.190893 65005493 164	0.84184258 40306377
Pantai Jatimalang	Jawa Tengah	Pantai	0.2955600168 115777	0.192160 87460517 883	0.84074836 02503616
Pantai Ciputih	Banten	Pantai	0.2922527768 2974754	0.192083 01603794 098	0.83510428 7557978

Hasil integrasi *hybrid* menunjukkan bahwa destinasi yang memiliki skor tinggi pada kedua sumber *TF-IDF* dan *Neural Network* cenderung menempati peringkat teratas, artinya destinasi

tersebut mirip secara deskripsi dan relevan secara atribut numerik. Beberapa destinasi yang awalnya memiliki skor *cosine* tinggi namun skor *Neural Network* rendah mengalami penurunan peringkat setelah proses *hybrid*. Model *hybrid* menghasilkan rekomendasi yang lebih stabil, tidak hanya bergantung pada kata-kata yang muncul di deskripsi, melainkan juga mempertimbangkan struktur destinasi. Oleh karena itu, dengan dibandikannya pendekatan tunggal, model *hybrid* memberikan hasil yang lebih seimbang karena memadukan informasi semantik teks oleh *TF-IDF*, dan informasi numerik dan struktural oleh *Neural Network*. Dengan demikian, pendekatan *hybrid* dalam penelitian ini mampu menghasilkan prediksi relevansi destinasi wisata yang lebih akurat dan komprehensif.

4.7 Hasil Evaluasi

Nilai *Precision@250* sebesar 0,90 mengindikasikan bahwa 90% dari 250 rekomendasi teratas memang sesuai dengan preferensi pengguna, artinya mayoritas destinasi yang ditampilkan benar-benar relevan dengan minat pengguna. Di sisi lain, *Recall@250* sebesar 0,97 menunjukkan bahwa 97% dari seluruh destinasi yang sebenarnya relevan berhasil diikutsertakan dalam daftar rekomendasi. Dengan kata lain, hampir seluruh destinasi yang relevan berhasil ditangkap oleh model.

```
# Evaluasi Model (Precision dan Recall)
name_col = 'nama_wisata' if 'nama_wisata' in filtered_df.columns else ('name' if 'name' in filtered_df.columns else filtered_df.columns[0])

filtered_df['label'] = filtered_df['kategori'].apply(lambda x: 1 if any(pc.lower() in str(x).lower() for pc in preferred_categories) else 0)
filtered_df_sorted = filtered_df.sort_values(by='score', ascending=False)
top_k = 250
top_k_recommendations = filtered_df_sorted.head(top_k)
ground_truth = filtered_df[filtered_df['label'] == 1][name_col].values
predicted = top_k_recommendations[name_col].values

def precision_recall_at_k(ground_truth, predicted, k):
    predicted_k = predicted[:k]
    relevant = set(ground_truth)
    retrieved = set(predicted_k)
    true_positive = relevant.intersection(retrieved)
    precision = len(true_positive) / k
    recall = len(true_positive) / len(relevant) if len(relevant) > 0 else 0.0
    return precision, recall

precision, recall = precision_recall_at_k(ground_truth, predicted, top_k)
print(f"\nPrecision@{top_k}: {precision:.2f}")
print(f"Recall@{top_k}: {recall:.2f}")

Precision@250: 0.90
Recall@250: 0.97
```

Gambar 8. Hasil *Precision* dan *Recall*

Tingginya nilai *Precision* dan *Recall* secara bersamaan mencerminkan bahwa model *hybrid* ini sangat efektif dalam menghasilkan rekomendasi yang relevan, terutama di peringkat atas, penggabungan skor *TF-IDF* yang menangkap informasi dari konten teks dengan *Neural Network* yang memproses fitur numerik menciptakan keseimbangan optimal antara ketepatan rekomendasi dan kelengkapan cakupan destinasi yang relevan, serta model tidak hanya mampu memahami kesamaan semantik dari deskripsi destinasi, tetapi juga mempertimbangkan aspek struktural seperti kategori, provinsi, dan alamat.

5. KESIMPULAN

Penelitian ini berhasil mengembangkan model prediksi rekomendasi destinasi wisata di Indonesia melalui pendekatan *hybrid* yang mengintegrasikan *TF-IDF* dan *Cosine Similarity* untuk menganalisis teks deskriptif destinasi, serta *Neural Network* untuk memproses fitur numerik seperti kategori, provinsi, dan alamat. Kombinasi kedua pendekatan ini memungkinkan model menangkap kesamaan semantik dari deskripsi sekaligus mempelajari pola-pola struktural, sehingga menghasilkan rekomendasi yang lebih akurat, relevan, dan seimbang.

Hasil evaluasi menunjukkan performa model yang sangat baik. Nilai *Precision@250* sebesar 0,90 mengindikasikan bahwa 90% dari 250 rekomendasi teratas sesuai dengan preferensi pengguna. Di sisi lain, *Recall@250* sebesar 0,97 menunjukkan bahwa model mampu mengidentifikasi hampir seluruh destinasi yang relevan.

Secara keseluruhan, pendekatan *hybrid* terbukti lebih unggul dibandingkan metode tunggal karena mampu memadukan kekuatan analisis konten teks dan pemrosesan fitur numerik secara sinergis. Model ini berpotensi menjadi fondasi bagi pengembangan rekomendasi wisata yang personal, andal, dan dapat diimplementasikan dalam cakupan nasional.

DAFTAR PUSTAKA

- [1] E. Novianti *et al.*, “PARIWISATA BERBASIS ALAM : MEMAHAMI PERILAKU WISATAWAN Pendidikan Pariwisata , Universitas Pendidikan Indonesia,” vol. 14, no. 1, pp. 46–52, 2020.
- [2] Y. Mz, S. N. Amalia, P. Mahinnudin, and J. Soares, “PENENTUAN DESA SADAR WISATA (Studi Kasus Desa Taman Yogyakarta),” vol. 8, no. 3, pp. 101–108, 2023.
- [3] T. Sugiyarto *et al.*, *Statistik Pariwisata Domestik 2024*. 2025. [Online]. Available: <https://www.bps.go.id/en/publication/2025/03/26/a7dd5a107d0dac1a41a3e414/domestic-tourism-statistics-2024.html>
- [4] Y. Aristantia, Y. M. Arif, and Z. Abidin, “Sistem Rekomendasi Destinasi Wisata di Kota Batu Menggunakan Neural Collaborative Filtering,” vol. 6, no. 3, pp. 341–352, 2024.
- [5] S. O. S. Husna, I. N. Fajri, and Y. Pristyanto, “SISTEM REKOMENDASI PARIWISATA GUNUNGKIDUL BERBASIS WEB MENGGUNAKAN METODE CONTENT - BASED FILTERING,” vol. 6, no. 3, pp. 534–543, 2024.
- [6] S. Suhendri and M. Karmila, “Penerapan algoritma c4.5 untuk memprediksi tren wisata di kabupaten majalengka 1,2,” vol. 1861, no. 9, pp. 644–652, 2023.
- [7] A. Koswara, “Evaluasi dan Optimasi Pengelolaan Pariwisata Grobogan Berdasarkan Analisis Data-Text Mining pada Google Review Asep Koswara Seiring dengan berkembangnya teknologi digital , data dari platform media sosial dan mengevaluasi kepuasan dan persepsi wisatawan terhadap destinasi wisata . Ulasan daring informasi yang tersembunyi dalam kumpulan data teks tidak terstruktur seperti ulasan bagi pemerintah daerah dan pengelola wisata untuk mengetahui persepsi wisatawan secara real-,” vol. 2, pp. 122–142, 2025.
- [8] E. Sari and A. S. Purnomo, “Rekomendasi Pemilihan Rute Wisata Menggunakan Metode Bellman- Ford (Studi Kasus : Daerah Wisata Kabupaten Luwu Timur),” vol. 1, no. 2, pp. 84–93, 2024.
- [9] P. M. Endraswari, N. Tou, and I. Zaliman, “Analisis Ketergantungan Mahasiswa Terhadap Chatgpt Dalam Konteks Akademik : Pemodelan Klasifikasi Berbasis Naïve Bayes,” vol. 10, no. 3, pp. 199–208, 2025.
- [10] D. Pratiwi and L. Rosnita, “Penerapan Metode Content-Based Filtering dalam Sistem Rekomendasi Objek Wisata di Aceh Tamiang,” vol. 4, no. 2, pp. 85–96, 2024.
- [11] S. Oyadila, D. Abdullah, and A. Razi, “Implementasi Content-Based Filtering Dengan Tf-Idf Dan Cosine Similarity Untuk Sistem Rekomendasi Destinasi,” vol. 10, no. 2, pp. 1329–1339, 2025.
- [12] C. Salsabilla and D. W. Utomo, “Sistem Rekomendasi Wisata Kabupaten Pekalongan dengan Content Based Filtering Pekalongan Regency Tourism Recommendation System with Content based Filtering,” vol. 14, pp. 262–270, 2025.
- [13] A. Utama, W. I. Sabilla, and R. Wakhidah, “Sistem Rekomendasi Tempat Wisata di Malang Raya Menggunakan Metode K-Means Clustering,” vol. 16, no. 1, pp. 1–15, 2024.
- [14] S. Supatminingsih and D. H. Wahyuningsih, “Pengembangan Metode Klasterisasi untuk Pengelompokan Potensi Wisata guna Menunjang Digitalisasi Wisata di Indonesia,” vol. 13, no. 2, pp. 217–230, 2024.
- [15] R. A. Purba, D. J. M. Sembiring, S. M. Perangin, Angin, and R. Buaton, “Implementasi Metode Promethee Untuk Rekomendasi Destinasi Objek Wisata Unggulan,” vol. 13, no. 1, pp. 181–189, 2024.