

Sistem Rekomendasi Destinasi Wisata di Kota Bandung dengan *Collaborative Filtering* Menggunakan *K-Nearest Neighbors*

Muhammad Nuril Adlan¹, Erwin Budi Setiawan²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹mnadlan@students.telkomuniversity.ac.id, ²erwinbudisetiawan@telkomuniversity.ac.id,

Abstrak

Kota Bandung adalah salah satu destinasi wisata populer di Indonesia. Banyaknya jumlah destinasi wisata di Kota Bandung, ditambah dengan kurangnya informasi tentang pariwisata, menimbulkan hambatan bagi kebutuhan masyarakat dalam memilih destinasi wisata. Oleh karena itu, diperlukan sebuah sistem rekomendasi untuk membantu wisatawan dalam menentukan destinasi mereka. Penelitian ini mengembangkan sistem rekomendasi destinasi wisata di Kota Bandung dengan menerapkan algoritma *user-based collaborative filtering* dan *K-Nearest Neighbors* untuk membantu wisatawan memutuskan destinasi mereka berdasarkan tempat-tempat yang sebelumnya telah mereka kunjungi. Dua metode kesamaan yang digunakan adalah *cosine similarity* dan *pearson correlation*. *Mean Absolute Error* (MAE) dan hasil rekomendasi digunakan untuk mengevaluasi kinerja sistem. Hasil penelitian menunjukkan bahwa sistem rekomendasi yang dibangun cukup memberikan rekomendasi kepada *user*, dengan nilai MAE sebesar 2.59 untuk metode *cosine similarity* dan nilai MAE sebesar 2.67 untuk metode *Pearson correlation*. Selain itu, hasil rekomendasi wisata yang diberikan dianggap memadai karena sesuai dengan profil wisatawan.

Kata kunci : Sistem Rekomendasi, *Collaborative Filtering*, *User-Based*, *Cosine Similarity*, *Pearson Correlation*, *K-Nearest Neighbors*

Abstract

The city of Bandung is one of the popular tourist destinations in Indonesia. The large number of tourist destinations in the city of Bandung, coupled with the lack of information about tourism, poses obstacles to people's needs in choosing tourist destinations. Therefore, a recommendation system is needed to assist tourists in determining their destinations. This research developed a tourist destination recommender system in Bandung City by implementing *user-based collaborative filtering* and *K-Nearest Neighbors* algorithms to help tourists decide on their destinations based on places they have previously visited. Two similarity methods were utilized: *cosine similarity* and *pearson correlation*. *Mean Absolute Error* (MAE) and recommendation results were employed to evaluate the system's performance. The research results indicate that the built recommendation system adequately provides recommendations to users, with an MAE value of 2.59 for the *cosine similarity* method and an MAE value of 2.67 for the *Pearson correlation* method. Furthermore, the tourist recommendation results provided are deemed adequate as they align with the tourist profiles.

Keywords: Recommender System, *Collaborative Filtering*, *User-Based*, *Cosine Similarity*, *Pearson Correlation*, *K-Nearest Neighbors*

1. Pendahuluan

Latar Belakang

Kota Bandung merupakan salah satu destinasi wisata yang populer di Indonesia, dikenal dengan keindahan alamnya, warisan budayanya, dan beragam atraksi wisata [1]. Oleh karena itu, tidak jarang juga calon wisatawan yang berencana untuk bepergian ke Bandung kebingungan untuk menentukan destinasi wisata yang menarik saat mengunjungi Kota Bandung. Solusi atas permasalahan tersebut bisa diselesaikan dengan adanya sistem rekomendasi pada destinasi wisata di Bandung yang dapat memberikan rekomendasi destinasi wisata secara efektif [2].

Sistem rekomendasi memiliki beberapa teknik diantaranya yaitu *collaborative filtering*, *content-based filtering* dan gabungan dari keduanya yang disebut *hybrid recommenders* [3]. Diantara beberapa teknik tersebut, pendekatan yang paling berhasil adalah rekomendasi berdasarkan teknik *collaborative filtering* [4]. Ada dua pendekatan utama dalam metode *collaborative filtering*: *user-based collaborative filtering* dan *item-based collaborative filtering*. *User-based collaborative* memberikan rekomendasi berdasarkan *item-item* berdasarkan *item-item* yang dipilih oleh berbagai anggota kelompok. Berbeda dengan *item-based collaborative filtering* yang memberikan rekomendasi berdasarkan seorang *user* cenderung memilih *item* yang serupa dengan *item-item* yang telah dipilihnya sebelumnya [5].

Adapun metode yang digunakan dalam pembuatan sistem ini adalah *user-based collaborative filtering*. *User-based collaborative filtering* digunakan karena rekomendasi *item* yang dihasilkan merupakan hasil dari *item* yang diberi *rating* oleh *user* lain atau *user* dengan produk serupa yang telah mereka *rating*, sehingga mendefinisikan setiap *user*-nya merupakan bagian dari sekelompok *user* dengan minat yang sama dengan beberapa destinasi [6]. Namun *collaborative filtering* mempunyai kekurangan yaitu data *sparsity* [7]. Data *sparsity* adalah kondisi dimana *user* hanya menilai sebagian kecil dari total *item* [8]. Untuk mengatasi data *sparsity*, dilakukan proses klasifikasi menggunakan *K-Nearest Neighbors* pada sistem rekomendasi karena *K-Nearest Neighbors* dapat digunakan untuk mengatasi data *sparsity* berdasarkan pada penelitian yang dilakukan oleh Anwar, T et al [9]. Kami menggunakan dua metode untuk menghitung *similarity* yaitu *cosine similarity* dan *pearson correlation* sebagai perbandingan karena kedua metode tersebut adalah metode yang populer digunakan pada sistem rekomendasi [10]. Pengukuran performa akan dilihat berdasarkan MAE apakah sistem ini mampu untuk mengatasi data *sparsity* atau tidak.

Kontribusi utama dari penelitian ini terletak pada pemanfaatan dua metode kesamaan yang sudah mapan, yaitu *Pearson correlation* dan *cosine similarity*, untuk *user-based collaborative filtering*, serta klasifikasi KNN untuk mengatasi masalah *sparsity*. Selain itu, studi ini juga menggabungkan hasil dari sistem rekomendasi, termasuk detail tentang rekomendasi *user* untuk destinasi wisata, untuk mengukur kinerja dan memastikan apakah rekomendasi tersebut sesuai dengan preferensi *user*. Hal ini berbeda dengan penelitian sebelumnya yang umumnya hanya melaporkan data statistik dari sistem rekomendasi tanpa menyertakan hasil aktual dari rekomendasi tersebut.

Topik dan Batasannya

Topik pada penelitian ini menjelaskan bagaimana mengimplementasikan metode *collaborative filtering* dengan menggunakan klasifikasi *K-Nearest Neighbour* (KNN) yang nantinya akan dilakukan dua pendekatan *similarity* yaitu *cosine similarity* dan *pearson correlation*. Keterbatasan dari penelitian ini yaitu dataset yang digunakan diperoleh dari situs kaggle.com dan pemrosesan data tidak menggunakan semua atribut yang terdapat pada data.

Tujuan

Tujuan penelitian ini adalah untuk menerapkan *collaborative filtering* menggunakan klasifikasi KNN yang diharapkan dapat membuat sistem rekomendasi pemilihan destinasi wisata dan mengevaluasi performa metode kesamaan antara *Cosine Similarity* dan *Pearson Correlation*

Organisasi Tulisan

Bab 2 menjelaskan teori dasar dalam konteks penelitian yang dilakukan. Bab 3 menjelaskan tentang desain dari sistem pemberi rekomendasi yang kami buat. Bab 4 menjelaskan tentang evaluasi sistem yang dibangun, dan Bab 5 menjelaskan tentang kesimpulan dari penelitian ini.

2. Studi Terkait

Penelitian sebelumnya dalam bidang sistem rekomendasi telah banyak mengusulkan dan membandingkan berbagai teknik. Misalnya, Cherlina Helena P Panjaitan et al pada tahun 2022 yang berjudul "Analisis Metode K-Nearest Neighbor Menggunakan Rapid Miner Untuk Sistem Rekomendasi Tempat Wisata Labuan Bajo" [11]. Berdasarkan penelitian ini, penggunaan KNN pada sistem rekomendasi tempat wisata memberikan akurasi sebesar 83,33%. Hasil tersebut membuktikan bahwa KNN mampu digunakan untuk merekomendasikan tempat wisata pada sistem rekomendasi.

Pada penelitian selanjutnya yang ditulis oleh R.A. Nugroho et al pada tahun 2020 yang berjudul "Tourism Site Recommender System Using Item-Based Collaborative Filtering Approach" [12]. Pada penelitian ini penulis menggunakan *item-based collaborative filtering* untuk membuat model rekomendasi wisatanya. Berdasarkan penelitian ini, metode *collaborative filtering* dapat membuat sistem rekomendasi wisata yang sesuai dan dapat menghasilkan MAE senilai 0,62.

Pada penelitian selanjutnya yang ditulis oleh Alif Azhar Fakhri et al pada tahun 2019 yang berjudul "Restaurant Recommender System Using User-Based Collaborative Filtering Approach: A Case Study at Bandung Raya Region" [13]. Penelitian ini menggunakan *user-based collaborative filtering* beserta KNN yang membandingkan 2 metode *similarity* yaitu, *pearson correlation* dan *similarity attribute*. Dimana *similarity attribute* ini adalah kesamaan jenis kelamin dan usia antar *user*. Berdasarkan penelitian ini, metode *similarity* menggunakan *pearson correlation* lebih unggul dibandingkan dengan menggunakan *similarity attribute*.

Pada penelitian selanjutnya yang ditulis oleh Putra Cahya Purnama et al pada tahun 2020 yang berjudul "Analisis Perbandingan Metode Similarity Pearson dan Cosine pada Sistem Rekomendasi Film dengan Pendekatan User-Based Collaborative Filtering" [14]. Pada penelitian tersebut membahas tentang perbandingan metode *cosine*

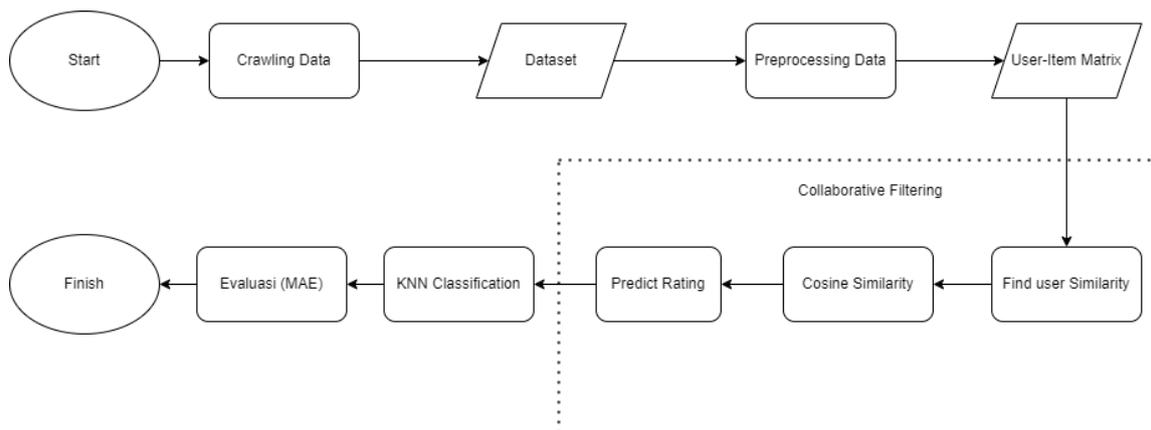
similarity dan *pearson correlation* sebagai metode *similarity* nya. Kedua metode tersebut memiliki hasil RMSE yang hampir identik yaitu, $RMSE\ pearson\ correlation = 1,0206$ dan $RMSE\ cosine\ similarity = 1,0211$. Nilai tersebut menunjukkan bahwa metode *pearson correlation* memiliki sedikit keunggulan dari *cosine similarity*.

Pada penelitian selanjutnya yang ditulis oleh Putra Cahya Purnama et al pada tahun 2019 yang berjudul "The User Personalization with KNN for Recommender System" [15]. Pada penelitian tersebut membahas tentang perbandingan metode *cosine similarity* dan *pearson correlation* sebagai metode *similarity* nya. Kedua metode tersebut memiliki hasil RMSE yaitu, $RMSE\ pearson\ correlation = 1,34$ dan $RMSE\ cosine\ similarity = 1,23$. Berbeda dengan hasil pada penelitian [14], penelitian ini memiliki nilai yang menunjukkan bahwa metode *cosine similarity* memiliki sedikit keunggulan dari *pearson correlation*

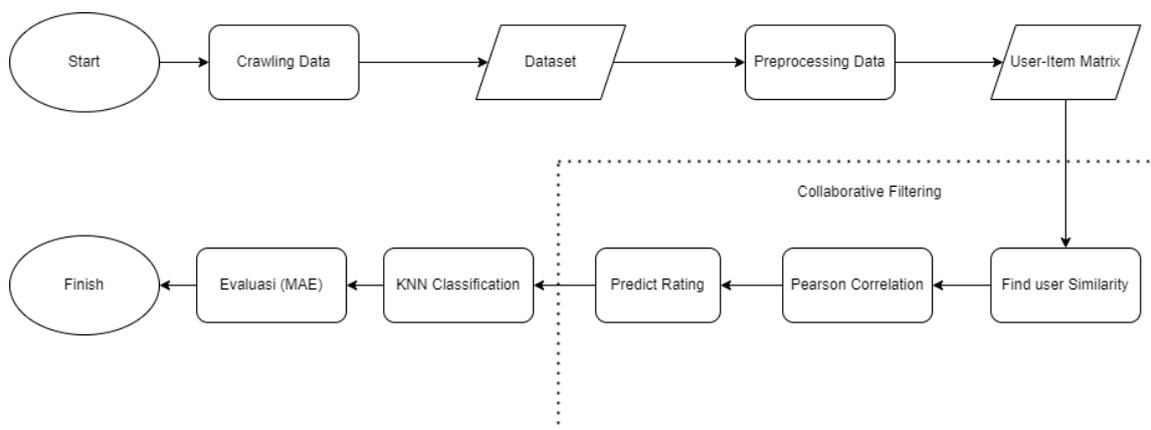
3. Sistem yang Dibangun

3.1 Figuran Sistem

Dalam penelitian ini, dua pendekatan akan digunakan: *Collaborative filtering* menggunakan *Pearson correlation* dan *collaborative filtering* menggunakan *cosine similarity*. Penelitian akan berlangsung melalui beberapa fase, dimulai dengan *data crawling*, diikuti dengan *data preprocessing*. Selanjutnya, kesamaan akan dinilai menggunakan *Pearson correlation* dan *cosine similarity*. Setelah proses kesamaan, data akan diubah menjadi rating. Terakhir, *KNN Classification* akan dijalankan untuk mengurangi *sparsity* data [9], diikuti dengan evaluasi menggunakan *Mean Absolute Error (MAE)*. Desain sistem dapat dilihat pada Gambar 1 dan Gambar 2. Gambar 1 menggambarkan desain sistem yang menggunakan *cosine similarity* dalam metode kesamaannya.



Gambar 1. Perancangan Sistem Metode 1



Gambar 2. Perancangan Sistem Metode 2

3.2 Crawling Data

Pengumpulan dan penyajian data adalah komponen penting dalam penelitian [16]. Dalam penelitian ini, data yang digunakan adalah data pariwisata yang bersumber dari situs Kaggle, yang dibagi menjadi beberapa dataset. Dataset tersebut meliputi:

1. “tourism_rating”: Dataset ini terdiri dari data penilaian wisata yang diberikan oleh *user*.
2. “tourism_with_id”: Dataset ini berisi data pariwisata untuk lima kota di Indonesia, masing-masing diidentifikasi dengan pengenal unik.
3. “data_user”: Dataset ini menyediakan informasi tentang profil *user*.

Dataset-dataset ini secara kolektif membentuk dasar untuk melakukan analisis dan mengembangkan sistem rekomendasi untuk destinasi wisata.

3.3 Data Preprocessing

Data *preprocessing* adalah tahap awal dari pemrosesan data di mana tujuannya adalah untuk menghilangkan data yang berisik atau tidak konsisten [17]. Pada tahap ini, ketiga dataset diintegrasikan, dan filter data diterapkan untuk mengekstrak informasi khusus Bandung. Selanjutnya, data diubah menjadi *user-item matrix*, yang mencakup hanya tempat-tempat dan *user* yang telah dinilai, memudahkan penilaian kesamaan di antara mereka.

3.4 Collaborative Filtering

Collaborative filtering merupakan salah satu metode pada penerapan sistem rekomendasi. Sistem rekomendasi *collaborative filtering* merekomendasikan berdasarkan informasi dari *user* atau preferensi *user* [18]. Pada *collaborative filtering* dibagi dua dalam melakukan rekomendasinya yaitu berdasarkan suatu *item* (*item-based*) dan kemiripan berdasarkan *user* (*user-based*) [19]. Pada tahap *collaborative filtering* ini dilakukan perhitungan *similarity* dan prediksi rating.

Perhitungan nilai *similarity* pada penelitian ini dilakukan dua metode yaitu *cosine similarity* dan *pearson correlation*.

Cosine Similarity merupakan salah satu metode untuk menghitung nilai kemiripan antara *user i* dan *j* yang digambarkan sebagai suatu sudut yang terbentuk diantara 2 buah *vector* yang dapat dihitung menggunakan rumus 1 [20]:

$$sim(i, j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{||\vec{i}|| \cdot ||\vec{j}||} \quad (1)$$

dengan $sim(i, j)$ adalah *similarity* antara *user i* dan *user j*.

Pearson Correlation adalah salah satu metode untuk menghitung seberapa besar hubungan linear antara dua variable yang dapat dihitung menggunakan rumus 2 [21]:

$$pearson(u, v) = \frac{\sum (R_{ui} - \bar{R}_u)(R_{vi} - \bar{R}_v)}{\sqrt{\sum (R_{ui} - \bar{R}_u)^2 \sum (R_{vi} - \bar{R}_v)^2}} \quad (2)$$

dengan $pearson(u, v)$ adalah *similarity* antara *user u* dan *user v*. R_{ui} dan R_{vi} adalah rating dari *user u* dan *v* untuk *item i*, sedangkan \bar{R}_u dan \bar{R}_v adalah rata-rata rating untuk *user u* dan *user v*.

Selanjutnya setelah didapatkan nilai *similarity* nya kita dapat mengisi rating kosong menggunakan metode user based dengan menggunakan rumus 3:

$$P(u, i) = \bar{r}_u + \frac{\sum_1^n sim(u, v) * (r_{vi} - \bar{r}_v)}{\sum_1^n sim(u, v)} \quad (3)$$

Dimana pada rumus (3), $P(u, i)$ adalah prediksi rating dari user u untuk item i . \bar{r}_u adalah rata-rata rating dari user u . $sim(u, v)$ adalah *similarity* antara user u dan user v . \bar{r}_v adalah rata-rata rating dari user v . r_{vi} adalah rating dari user v untuk item i , dimana n adalah jumlah user [8].

3.5 K-Nearest Neighbors

K-Nearest Neighbors merupakan suatu algoritma supervised learning yang digunakan untuk klasifikasi berdasarkan mayoritas dari kategori tetangga terdekat [22]. KNN dipilih karena mempunyai keunggulan yaitu: Pelatihan sangat cepat, sederhana, efektif jika data pelatihan besar dan mudah dipelajari [23]. Tujuan dari KNN adalah mencari jarak minimum antara data yang dievaluasi oleh tetangga terdekat (k). Pada penelitian ini, KNN digunakan untuk mencari topN terbaiknya dan mengklasifikasi hasil dari *collaborative filtering* untuk memberikan hasil yang lebih baik.

3.6 Evaluasi Performa

Pada penelitian ini menggunakan *Mean Absolute Error* untuk pengukuran performanya. *Mean Absolute Error* mengevaluasi perbedaan antara rating yang diprediksi oleh sistem rekomendasi atau *predicted rating* dan rating yang diberikan oleh user atau *actual rating* [24]. *Mean Absolute Error* dapat dihitung menggunakan rumus 4:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y - \hat{y}| \quad (4)$$

dengan y adalah *actual rating* dan \hat{y} adalah *predicted rating*, sedangkan n adalah jumlah user.

Pada penelitian ini nilai *Mean Absolute Error* memiliki jarak antara 0 sampai 5 berdasarkan nilai *rating*, dimana nilai lebih kecil nilai maka lebih baik juga performanya.

4. Evaluasi

Pada penelitian ini langkah yang pertama dilakukan yaitu data yang telah di crawling dari kaggle dan digabungkan akan dilakukan penyaringan data untuk mengambil data yang berasal dari Kota Bandung. Lalu, data akan dihitung *similarity* nya menggunakan metode *cosine similarity* dan *pearson correlation*. Setelah mendapatkan hasil *similarity*, akan dilakukan prediksi rating untuk data yang tidak memiliki nilai rating atau 0. Lalu, data akan diproses menggunakan KNN untuk menentukan topN dan MAE setiap N nya. Setelah mendapatkan topN, data akan menggunakan topN tersebut untuk melakukan rekomendasi terhadap suatu user dan menghasilkan rekomendasi wisata untuk user tersebut.

4.1 Data

Pada penelitian ini, data dari Kaggle dibagi menjadi 3 dataset, yaitu *tourism_id*, *tourism_rating*, dan *user*. Selanjutnya, dataset tersebut digabungkan menjadi satu dataset. Setelah penggabungan dataset, data tersebut akan berisi *user ID*, *place ID*, *place ratings*, *place names*, *place descriptions*, *place cities*, dan *place categories*. Dataset asli mencakup informasi pariwisata di seluruh Indonesia tetapi akan difilter untuk fokus secara khusus pada data pariwisata yang berada di kota Bandung. Berikut adalah contoh isi dari dataset, seperti yang ditunjukkan pada Tabel 1.

Tabel 1. Data Wisata

User_Id	Place_Id	Place_Ratings	Place_Name	Description	City	Category
1	179	3	Candi Ratu Boko	Situs Ratu Bana atau	Yogyakarta	Budaya

...	Candi Boko...
300	20	2	Watu Mabur Mangunan	Kawasan Tebing Watu Mabur...	Yogyakarta	Budaya

Setelah penggabungan data maka selanjutnya data akan disaring untuk mengambil data yang berasal dari Kota Bandung. Sehingga dapat dilihat pada Tabel 2.

Tabel 2. Data Wisata Kota Bandung

User_Id	Place_Id	Place_Ratings	Place_Name	Description	City	Category
1	312	2	Taman Hutan Raya Ir. H. Juanda	Taman Hutan Raya Ir. H. Juanda...	Bandung	Cagar Alam
...
300	279	4	Masjid Agung Trans Studio Bandung	Masjid Agung Trans Studio Bandung (TSB)...	Bandung	Tempat Ibadah

Kemudian dibuat *user-item matrix* yang berupa *pivot table* 2 dimensi yang berisi *User_Id*, *Place_Id*, dan *Place_Ratings* yang memuat 2738 data terisi rating dan 34462 rating kosong sehingga memiliki *data sparsity* 92,64% seperti yang terlihat pada Tabel 3. *user-item matrix* juga merupakan rating yang diberikan oleh *user* atau *actual rating* yang nantinya akan digunakan untuk menghitung MAE setelah mendapatkan *predicted rating*.

Tabel 3. User-Item Matrix

Place_Id	211	220	328	334
User_Id				
1	0.00	0.00	2.00	0.00
...
299	0.00	1.00	2.00	0.00
300	0.00	0.00	0.00	0.00

4.2 Collaborative Filtering

Menggunakan *user-item matrix* dilakukan penghitungan nilai *similarity* menggunakan *cosine similarity* dan *pearson correlation*. Setelah mendapatkan data *pearson correlation* selanjutnya dilakukan normalisasi menggunakan *min-max scaler* untuk menyamakan nilainya dengan *cosine similarity*. Data dapat dilihat pada Tabel 4 untuk *cosine similarity* dan Tabel 5 untuk *pearson correlation*.

Tabel 4. Data Cosine Similarity

User_Id	1	2	...	299	300
User_Id					
1	1.00	0.00	...	0.03	0.00
2	0.00	1.00	...	0.09	0.00
...
299	0.03	0.09	...	1.00	0.00
300	0.00	0.00	...	0.00	1.00

Nilai *cosine similarity* berkisar antara 0 hingga 1. Jika nilai kesamaan mendekati 0, ini menunjukkan bahwa dua *user* sangat tidak mirip satu sama lain, sedangkan jika mendekati 1, ini menunjukkan kesamaan yang tinggi. Sedangkan pada *Pearson correlation*, nilainya berkisar dari -1 hingga 1. Nilai -1 menunjukkan bahwa dua *user* sangat tidak mirip satu sama lain. Sebaliknya, nilai 1 menunjukkan hubungan linier positif yang kuat, yang menandakan kesamaan yang tinggi. Ketika koefisien korelasi mendekati 0, ini menunjukkan hubungan linier yang

lemah atau tidak ada antara variabel, yang menyiratkan tidak ada kesamaan atau ketidakmiripan yang kuat. Tabel 5 menampilkan data *pearson correlation* yang diperoleh dari *collaborative filtering*.

Tabel 5. Data *Pearson Correlation*

User_Id	1	2	...	299	300
User_Id					
1	1.00	-0.04	...	-0.03	-0.06
2	0.04	1.00	...	0.04	-0.04
...
299	-0.03	0.04	...	1.00	-0.06
300	-0.06	-0.04	...	-0.06	1.00

Setelah mendapatkan *pearson correlation*, normalisasi dilakukan menggunakan *min-max scaler* untuk menstandarkan nilai dari 0 hingga 1 setelah memperoleh data *Pearson correlation*, karena rumus prediksi dapat memberikan hasil yang salah jika nilainya berkisar dari -1 hingga 1, seperti halnya pada nilai dari *pearson correlation*. Data setelah normalisasi dapat ditunjukkan pada Tabel 6.

Tabel 6. Data *Pearson Correlation* Setelah Normalisasi

User_Id	1	2	...	299	300
User_Id					
1	1.00	0.01	...	0.05	0.02
2	0.04	1.00	...	0.12	0.03
...
299	0.05	0.10	...	1.00	0.01
300	0.03	0.02	...	0.02	1.00

Setelah memperoleh nilai *similarity*, prediksi rating akan dilakukan untuk mengisi rating yang hilang pada dataset. Dataset kemudian dibagi menjadi set pelatihan dan pengujian untuk KNN, dengan rasio pembagian 20% untuk data pengujian dan 80% untuk data pelatihan. Tabel 7 menunjukkan prediksi rating dengan metode *cosine similarity* dan Tabel 8 menunjukkan prediksi rating dengan metode *pearson correlation*.

Tabel 7. Hasil Prediksi Rating *Cosine Similarity*

Place_Id	211	...	333	334
User_Id				
1	2.44	...	4.04	3.35
...
299	3.67	...	3.78	2.89
300	2.33	...	3.38	3.84

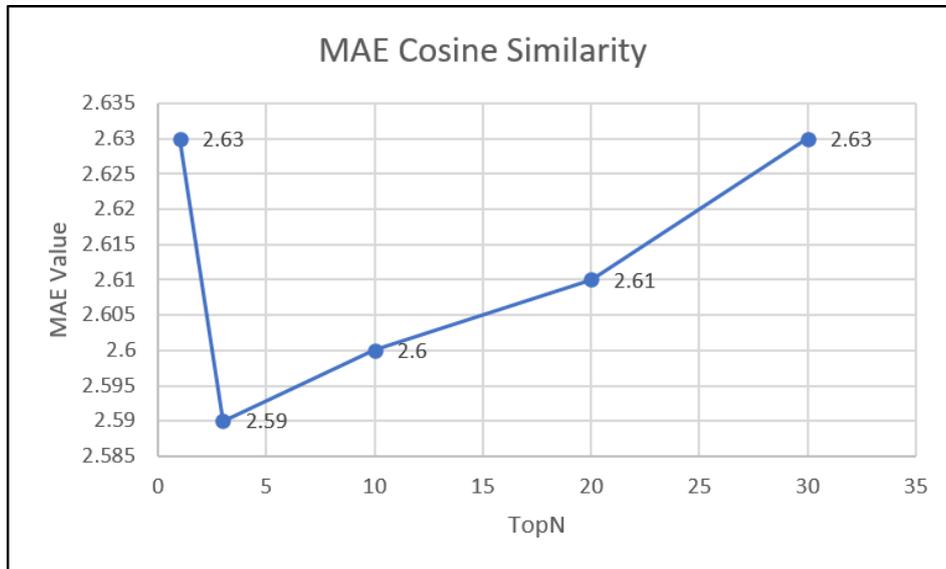
Tabel 8. Hasil Prediksi Rating *Pearson Correlation*

Place_Id	211	...	333	334
User_Id				
1	2.68	...	3.75	3.27
...
299	3.59	...	3.71	2.94
300	2.66	...	3.39	3.62

Dapat diamati bahwa nilai dari kedua metode hampir identik, meskipun *pearson correlation* menghasilkan hasil yang sedikit lebih tinggi dibandingkan dengan *cosine similarity*. Setelah mengisi rating yang kosong dengan nilai rating yang diproyeksikan, terdapat 34.462 data rating yang terisi dan 2.738 rating kosong, menghasilkan *sparsity data* sebesar 7,36%. Selanjutnya, data akan diproses menggunakan klasifikasi KNN.

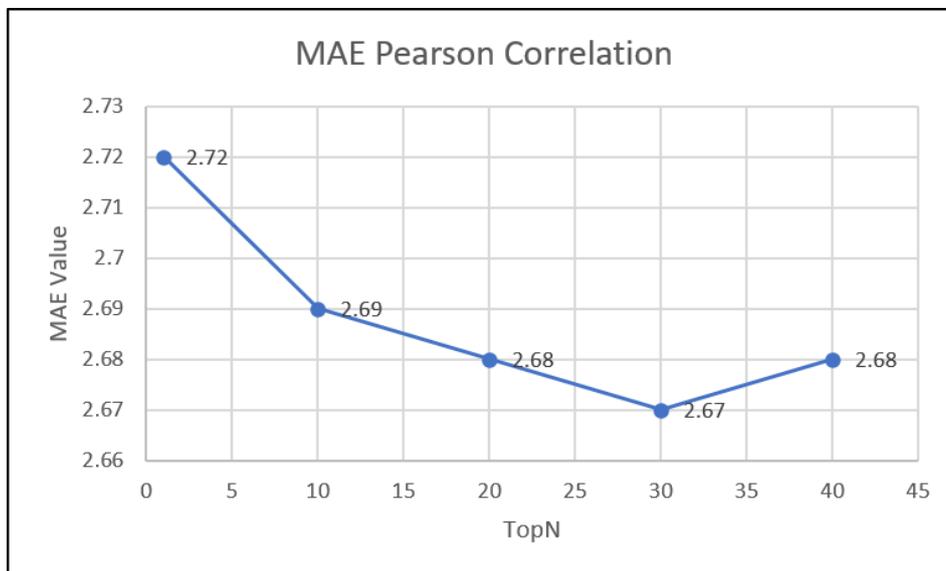
4.3 K-Nearest Neighbors

Sebelum dilakukan klasifikasi KNN, MAE yang dihasilkan dari dataset hasil prediksi rating memiliki nilai MAE sebesar 3.00 untuk *cosine similarity* dan *pearson correlation*. Dataset hasil prediksi rating akan digunakan pada proses klasifikasi KNN untuk mencari TopN dan MAE dengan menggunakan parameter K = 1 sampai 40. Hasil dari KNN dapat dilihat pada Gambar 2 untuk *cosine similarity* dan Gambar 3 untuk *pearson correlation*.



Gambar 3. MAE dan TopN untuk Cosine Similarity

Gambar 3 menunjukkan bahwa hasil MAE terendah untuk KNN dengan *cosine similarity* terjadi pada N = 3, dengan nilai MAE sebesar 2.59. Dapat disimpulkan bahwa N = 3 adalah jumlah optimal untuk KNN dengan *cosine similarity*. Gambar 4 menggambarkan temuan KNN untuk *pearson correlation* menggunakan grafik garis.



Gambar 4. MAE dan TopN untuk Pearson Correlation

Gambar 4 menunjukkan bahwa hasil MAE terendah untuk KNN dengan *pearson correlation* terjadi pada N = 30, dengan nilai MAE sebesar 2.67. Dapat disimpulkan bahwa N = 30 adalah jumlah optimal untuk KNN dengan *pearson correlation*.

Dari kedua gambar ini, dapat disimpulkan bahwa MAE yang dihasilkan oleh KNN memiliki peningkatan dari sebelumnya yang memiliki MAE sebesar 3.00 menjadi 2.59 untuk *cosine similarity* dan 2.67 untuk *pearson correlation* dan hasil MAE menggunakan *cosine similarity* lebih rendah dibandingkan dengan MAE menggunakan *pearson correlation*. Oleh karena itu, *cosine similarity* lebih baik daripada *pearson correlation* untuk kasus ini.

4.3 Hasil Rekomendasi Destinasi Wisata

Pada penelitian ini, *user* 181 akan dipilih untuk menampilkan hasil rekomendasi destinasi wisata. *User* ini dipilih karena memiliki total rating tertinggi dalam data. Rekomendasi wisata dan profil *user* mengenai tempat yang dikunjungi di Bandung ditunjukkan dalam Tabel 9. Tabel tersebut mencantumkan tempat-tempat yang telah dikunjungi oleh *user* 181 di Bandung, beserta kategori masing-masing.

Tabel 9. Data Wisata Kota Bandung yang Pernah Dikunjungi User

No	Place Name	Category
1	Gunung Manglayang	Cagar Alam
2	NuArt Sculpture Park	Taman Hiburan
3	Masjid Daarut Tauhiid Bandung	Tempat Ibadah
4	Orchid Forest Cikole	Taman Hiburan
5	Taman Balai Kota Bandung	Taman Hiburan
6	Batununggal Indah Club	Taman Hiburan
7	Kampung Batu Malakasari	Taman Hiburan
8	Taman Lalu Lintas Ade Irma Suryani Nasution	Taman Hiburan
9	Tafso Barn	Cagar Alam
10	Taman Vanda	Taman Hiburan
11	Trans Studio Bandung	Taman Hiburan
12	Wisata Batu Kuda	Cagar Alam
13	Museum Nike Ardilla	Budaya
14	Museum Barli	Budaya
15	Situ Patenggang	Cagar Alam
16	Kyotoku Floating Market	Budaya
17	The Lodge Maribaya	Cagar Alam
18	Curug Tilu Leuwi Opat	Cagar Alam

Tabel 9 menunjukkan bahwa *user* mengunjungi kategori taman hiburan paling banyak, dengan total delapan kunjungan, diikuti oleh enam kunjungan ke tempat suci, tiga kunjungan ke situs budaya, dan satu kunjungan ke tempat ibadah. Maka dari itu, destinasi yang direkomendasikan untuk *user* adalah yang sesuai dengan kategori-kategori yang paling sering dikunjungi oleh *user*.

Tabel 10 menampilkan hasil dari *cosine similarity*. Tabel tersebut menunjukkan hasil rekomendasi pariwisata yang dihasilkan oleh sistem rekomendasi menggunakan metode *cosine similarity*. Ini menampilkan 10 tempat yang direkomendasikan berdasarkan rating tertinggi, dengan tempat yang paling direkomendasikan adalah Glamping Lakeside Rancabali, yang memiliki rating 4.26.

Tabel 10. Hasil Rekomendasi Destinasi Wisata Cosine Similarity

Place Name	Category	Predicted Rating
Glamping Lakeside Rancabali	Taman Hiburan	4.26
Teras Cikapundung BBWS	Taman Hiburan	4.18
Stone Garden Citatah	Taman Hiburan	3.70
Kampung Korea Bandung	Budaya	3.69
Masjid Agung Trans Studio Bandung	Tempat Ibadah	3.64
Sunrise Point Cukul	Cagar Alam	3.61
Curug Cilengkrang	Cagar Alam	3.60
Kawasan Punclut	Taman Hiburan	3.60
Lawangwangi Creative Space	Taman Hiburan	3.59
Gua Belanda	Cagar Alam	3.59

Tabel 11 menampilkan hasil dari *pearson correlation*. Tabel tersebut menunjukkan hasil rekomendasi pariwisata yang dihasilkan oleh sistem rekomendasi menggunakan metode *pearson correlation*. Ini menampilkan 10 tempat yang direkomendasikan berdasarkan rating tertinggi, dengan tempat yang paling direkomendasikan adalah Teras Cikapundung BBWS, yang memiliki rating 3.58.

Tabel 11. Hasil Rekomendasi Destinasi Wisata Pearson Correlation

Place Name	Category	Predicted Rating
Teras Cikapundung BBWS	Taman Hiburan	3.58
Masjid Agung Trans Studio Bandung	Taman Hiburan	3.54
Rainbow Garden	Taman Hiburan	3.44
Glamping Lakeside Rancabali	Budaya	3.44
Dago Dreampark	Tempat Ibadah	3.40
Sanghyang Heuleut	Cagar Alam	3.39
Bukit Jamur	Cagar Alam	3.39
Stone Garden Citatah	Taman Hiburan	3.38
Ciwangun Indah Camp Official	Taman Hiburan	3.24
Upside Down World Bandung	Cagar Alam	3.23

Berdasarkan kedua tabel tersebut, dapat disimpulkan bahwa hasil untuk tempat-tempat yang diberikan dari *cosine similarity* dan *pearson correlation* hampir identik, dengan empat taman hiburan, tiga tempat suci, satu tempat budaya, dan satu tempat ibadah, meskipun dalam urutan yang berbeda. Selain itu, beberapa tempat direkomendasikan oleh *cosine similarity* tetapi tidak oleh *pearson correlation*, dan sebaliknya. Oleh karena itu, hasil yang disarankan dapat diterima karena kategori destinasi wisata sesuai dengan destinasi wisata sebelumnya *user* berdasarkan baik *pearson correlation* maupun *cosine similarity*.

5. Kesimpulan

Dalam penelitian ini, sistem rekomendasi diterapkan dengan menggunakan metode *user-based collaborative filtering* dan klasifikasi KNN. Dataset yang digunakan berupa dataset yang berasal dari Kaggle. Berdasarkan penelitian yang sudah dilakukan, didapatkan bahwa hasil pengujian untuk metode *cosine similarity* mendapatkan hasil MAE sebesar 2,59 dengan $K = 3$ dan pada hasil pengujian *pearson correlation* mendapatkan hasil MAE sebesar 2,67 dengan $K = 30$. Berdasarkan hasil MAE dinyatakan bahwa *cosine similarity* memiliki performa yang lebih baik dibanding *pearson correlation* pada data yang diuji. Hasil dari rekomendasi yang diberikan kepada *user* sudah cukup memadai karena hasil tersebut sudah sesuai dengan kategori wisata yang sudah pernah dikunjungi oleh *user*. Dapat disimpulkan bahwa penggabungan *collaborative filtering* dan klasifikasi KNN dapat digunakan untuk melakukan rekomendasi destinasi wisata. Untuk penelitian lebih lanjut dapat menggunakan semua atribut yang terdapat pada data dan dapat dikombinasikan dengan metode selain KNN, serta penggunaan metode *hybrid filtering* untuk rekomendasi.

Daftar Pustaka

- [1] D. Gede, A. Pradiva Viveka, and Z. K. Abdurahman Baizal, 'Implementasi Metode Item-based Collaborative Filtering dan Context-aware dalam Sistem Rekomendasi Pariwisata', 2019. [Online]. Available: <https://repository.telkomuniversity.ac.id/pustaka/155430/implementasi-metode-item-based-collaborative-filtering-dan-context-aware-dalam-sistem-rekomendasi-pariwisata.html>. Accessed: April 24, 2024.
- [2] M. Lusmiawati, E. Fatkhiyah, and A. Hamzah, 'Penentuan Objek Wisata Kota Bandung Menggunakan Metode Fuzzy Tsukamoto', *Jurnal SCRIPT*, vol. 9, no. 2, 2021.
- [3] K. Falk, 'Practical Recommender Systems', Manning Publications Co., 2019.
- [4] X. Wang, Z. Dai, H. Li, and J. Yang, 'A New Collaborative Filtering Recommendation Method Based on Transductive SVM and Active Learning', *Discrete Dynamics in Nature and Society*, vol. 2020. Hindawi Limited, 2020. doi: 10.1155/2020/6480273.
- [5] A. Refkrisnatta and D. Handayani, 'Cafe Selection Recommendation System in Semarang City Uses Collaborative Filtering Method with Item Based Filtering Algorithm', *JEEE-U (Journal of Electrical and Electronic Engineering-UMSIDA)*, vol. 6, no. 2, pp. 95–108, Oct. 2022, doi: 10.21070/jeeeu.v6i2.1637.
- [6] F. R. Hariri and L. W. Rochim, 'Sistem Rekomendasi Produk Aplikasi Marketplace Berdasarkan Karakteristik Pembeli Menggunakan Metode User Based Collaborative Filtering', *Teknika*, vol. 11, no. 3, pp. 208–217, Nov. 2022, doi: 10.34148/teknika.v11i3.538.
- [7] F. Ricci, · Lior, R. Bracha, and S. Editors, 'Recommender Systems Handbook Second Edition', Springer, 2015.
- [8] R. Rifaldy and E. B. Setiawan, 'Recommender System Movie Netflix using Collaborative Filtering with Weighted Slope One Algorithm in Twitter', *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 2, pp. 500–506, Sep. 2022, doi: 10.47065/bits.v4i2.1959.

- [9] T. Anwar, V. Uma, M. I. Hussain, and M. Pantula, 'Collaborative filtering and kNN based recommendation to overcome cold start and sparsity issues: A comparative analysis', *Multimed Tools Appl*, 2022, doi: 10.1007/s11042-021-11883-z.
- [10] S. C. Mana and T. Sasipraba, 'Research on cosine similarity and pearson correlation based recommendation models', in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Apr. 2021. doi: 10.1088/1742-6596/1770/1/012014.
- [11] C. H. P. Panjaitan, L. J. Pangaribuan, and C. I. Cahyadi, 'Analisis Metode K-Nearest Neighbor Menggunakan Rapid Miner untuk Sistem Rekomendasi Tempat Wisata Labuan Bajo', *Remik*, vol. 6, no. 3, pp. 534–541, Aug. 2022, doi: 10.33395/remik.v6i3.11701.
- [12] R. A. Nugroho, A. M. Polina, and Y. D. Mahendra, 'Tourism Site Recommender System Using Item-Based Collaborative Filtering Approach', *International Journal of Applied Sciences and Smart Technologies*, vol. 2, no. 2, pp. 209–216.
- [13] A. A. Fakhri, Z. K. A. Baizal, and E. B. Setiawan, 'Restaurant Recommender System Using User-Based Collaborative Filtering Approach: A Case Study at Bandung Raya Region', in *Journal of Physics: Conference Series*, Institute of Physics Publishing, May 2019. doi: 10.1088/1742-6596/1192/1/012023.
- [14] A. S. Dharma, 'The User Personalization with KNN for Recommender System', *Journal Publications & Informatics Engineering Research*, vol. 3, no. 2, 2019, doi: 10.33395/sinkron.v3i1.10047.
- [15] P. Cahya Purnama and S. Al Faraby, 'Analisis Perbandingan Metode Similarity Pearson dan Cosine pada Sistem Rekomendasi Film dengan Pendekatan User-Based Collaborative Filtering', 2020. [Online]. Available: <https://repository.telkomuniversity.ac.id/pustaka/164225/analisis-perbandingan-metode-similarity-pearson-dan-cosine-pada-sistem-rekomendasi-film-dengan-pendekatan-user-based-collaborative-filtering.html>. Accessed: April 24, 2024.
- [16] B. Widjanarko Otok and Ms. Dewi Juliah Ratnaningsih, 'Konsep Dasar dalam Pengumpulan dan Penyajian Data', Universitas Terbuka, 2016.
- [17] F. Alghifari and D. Juardi, 'Penerapan Data Mining Pada Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes', *Jurnal Ilmiah Informatika*, vol. 9, no. 2, pp. 76-81, 2021.
- [18] M. S. Kabul and E. B. Setiawan, 'Recommender System with User-Based and Item-Based Collaborative Filtering on Twitter using K-Nearest Neighbors Classification', *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 4, pp. 478–484, Sep. 2022, doi: 10.47065/josyc.v3i4.2204.
- [19] H. Februariyanti, A. Dwi Laksono, J. Sasongko Wibowo, and M. Siswo Utomo, 'Implementasi Metode Collaborative Filtering Untuk Sistem Rekomendasi Penjualan Pada Toko Mebel', *Jurnal Khatulistiwa Informatika*, vol. 9, no. 1, pp. 43-50, 2021, doi: 10.31294/jki.v9i1.9859.g4873.
- [20] G. Ferio, R. Intan, and S. Rostianingsih, 'Sistem Rekomendasi Mata Kuliah Pilihan Menggunakan Metode User Based Collaborative Filtering Berbasis Algoritma Adjusted Cosine Similarity'. *Jurnal Infra*, vol. 7, no. 1, pp. 1-7, 2019.
- [21] G. Jain, T. Mahara, and K. N. Tripathi, 'A Survey of Similarity Measures for Collaborative Filtering-Based Recommender System', in *Advances in Intelligent Systems and Computing*, Springer, 2020, pp. 343–352. doi: 10.1007/978-981-15-0751-9_32.
- [22] M. Al-Ghobari, A. Muneer, and S. M. Fati, 'Location-aware personalized traveler recommender system (lapta) using collaborative filtering knn', *Computers, Materials and Continua*, vol. 69, no. 2, pp. 1553–1570, 2021, doi: 10.32604/cmc.2021.016348.
- [23] R. A. Chandra, J. Try, and A. Halim, "Implementasi Algoritma K-Nearest Neighbor Pada Website Rekomendasi Laptop," *Jurnal Buana Informatika*, vol. 10, no. 1, pp. 75-84, 2019, doi: 10.24002/jbi.v10i1.1847.
- [24] M. Al-Ghamdi, H. Elazhary, and A. Mojahed, 'Evaluation of Collaborative Filtering for Recommender Systems', *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 3, pp. 559–564, 2021, doi: 10.14569/IJACSA.2021.0120367.