

# Sistem Rekomendasi Produk pada E-commerce Menggunakan Metode *User-Based Collaborative Filtering*

Devon Vivian Gunawan<sup>a1</sup>, I Komang Ari Mogi<sup>a2</sup>

<sup>a</sup>Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Udayana  
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia  
<sup>1</sup>devonvivian34@gmail.com  
<sup>2</sup>arimogi@unud.ac.id

## Abstract

*In an increasingly advanced digital era, the demand for recommendation systems that can provide products that match user preferences is increasingly high. The method used in this research is User-Based Collaborative Filtering, an approach that utilizes the purchasing patterns of other users to provide recommendations to active users. The research process began with collecting transaction datasets and product user preferences from leading e-commerce platforms. This data is then processed through the loading, preprocessing and analysis stages to prepare good data. Model performance is carried out using Mean Absolute Error (MAE) through a 5-fold cross validation process. The evaluation results show that the model has a satisfactory level of accuracy, with an average MAE value of 0.31. This research contributes to the development of a product recommendation system that can help improve users' online shopping experience by providing relevant and personalized recommendations.*

**Keywords:** *User-Based Collaborative Filtering, Recommendation system, Mean Absolute Error (MAE), e-commerce*

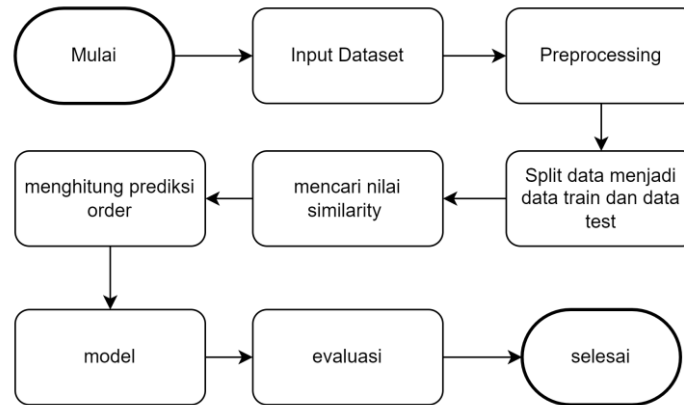
## 1. Pendahuluan

Dalam era digital yang semakin maju, industri e-commerce telah mengalami pertumbuhan pesat. Masyarakat kini lebih cenderung untuk melakukan pembelian secara online, mengingat kenyamanan dan kemudahan yang ditawarkan. Namun, dengan jumlah produk yang tersedia di platform e-commerce yang sangat besar, sering kali konsumen mengalami kesulitan dalam menemukan produk yang sesuai dengan preferensi dan kebutuhan mereka [1]. Untuk mengatasi tantangan ini, sistem rekomendasi produk telah menjadi salah satu solusi yang efektif. Sistem rekomendasi produk memanfaatkan teknik-teknik analisis data untuk memberikan rekomendasi kepada konsumen berdasarkan preferensi mereka [2]. Salah satu metode yang populer digunakan dalam sistem rekomendasi adalah Collaborative Filtering. Collaborative Filtering adalah pendekatan yang menggunakan informasi dari pengguna lain dalam sistem untuk memberikan rekomendasi. Metode ini didasarkan pada asumsi bahwa pengguna yang memiliki preferensi yang serupa di masa lalu cenderung memiliki preferensi yang serupa pula di masa depan. Terdapat dua jenis utama dari Collaborative Filtering: User-Based Collaborative Filtering (UBCF) dan Item-Based Collaborative Filtering (IBCF) [3]. Penelitian ini bertujuan untuk mengembangkan sebuah sistem rekomendasi produk di platform e-commerce dengan menggunakan metode User-Based Collaborative Filtering (UBCF). Pendekatan UBCF efektif dalam menemukan produk yang menarik bagi pengguna dengan mencocokkan preferensi mereka dengan pengguna lain yang memiliki minat serupa. Proses dimulai dengan mengidentifikasi pengguna terdekat atau tetangga yang memiliki preferensi serupa. Selanjutnya, nilai-nilai rating dari tetangga ini digunakan sebagai dasar untuk memberikan rekomendasi kepada pengguna yang sedang aktif [4]. Dalam penelitian ini, penulis akan menggunakan dataset transaksi dan preferensi pengguna produk yang telah dikumpulkan dari platform e-commerce tokopedia. Data ini akan diolah dan dianalisis menggunakan teknik User-Based Collaborative

Filtering untuk menghasilkan rekomendasi produk yang akurat. Selanjutnya, evaluasi kinerja sistem rekomendasi akan diuji berdasarkan performansi nilai *error* MAE.

## 2. Metode Penelitian

### 2.1. Desain Penelitian



**Gambar 1.** Tahapan Penelitian

Penelitian ini dimulai dengan pengumpulan dataset yang diperlukan, diikuti oleh proses pemuatan data menggunakan *library* pandas. Tahap *preprocessing* dilakukan dengan cermat untuk menangani nilai yang hilang. Setelah itu, dilakukan *5-fold cross validation* pada data transaksi untuk memisahkan dataset menjadi data pelatihan dan pengujian. Data pelatihan digunakan untuk membentuk model, sementara data pengujian digunakan untuk mengevaluasi kinerja model. Selanjutnya, dilakukan perhitungan matriks kemiripan menggunakan *cosine similarity* untuk mengukur kesamaan pola pembelian pengguna. Metode *user-based similarity* digunakan untuk memprediksi jumlah pembelian produk oleh seorang pelanggan berdasarkan perilaku pembelian pengguna lain yang serupa. Model berbasis kemiripan pengguna diuji menggunakan *5-fold cross validation* untuk menguji keakuratannya dalam memberikan rekomendasi produk kepada pengguna.

### 2.2. Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data sekunder yang diambil dari laman medium yang berisi data penjualan produk pada e-commerce tokopedia pada tahun 2021-2022. terdapat 4 file (.csv) yaitu *order\_detail*, *customer\_detail*, *sku\_detail* dan *payment\_detail*. keempat dataset saling berelasi. dataset yang digunakan mencakup data customer, data produk dan data transaksi. Cuplikan data ditunjukkan pada Gambar 2.

	id	customer_id	sku_id	qty_ordered	payment_id	payment_method	sku_name	base_price	cogs	category
0	ODR9939707760w	C713589L	P858068	200	5	jazzwallet	RB_Dettol Germ Busting Kit-bf	26100.0	18270	Others
1	ODR7448356649d	C551551L	P886455	5	5	jazzwallet	PS4_Slim-500GB	1971942.0	1321182	Entertainment
2	ODR4011281866z	C685596L	P678648	1	4	Payaxis	Changhong Ruba 55 Inches UD55D6000i Ultra HD T...	7482000.0	5162580	Entertainment
3	ODR3378927994s	C830683L	P540013	1	5	jazzwallet	dawlance_Inverter 30	3593680.0	3054628	Appliances
4	ODR4904430099k	C191766L	P491032	1	4	Payaxis	Dawlance_Inverter-45 2.0 ton	4413220.0	3177472	Appliances
...	...	...	...	...	...	...	...	...	...	...
5879	ODR4312238651y	C295679L	P194155	2	1	cod	Bahr-e-Shifa	0.0	0	Books
5880	ODR3930118778w	C368658L	P194155	1	1	cod	Bahr-e-Shifa	0.0	0	Books
5881	ODR7628482404c	C265450L	P194155	3	1	cod	Bahr-e-Shifa	0.0	0	Books
5882	ODR2809405363p	C790722L	P194155	2	1	cod	Bahr-e-Shifa	0.0	0	Books
5883	ODR6666616009w	C676393L	P100134	1	4	Payaxis	akl-GFSU1265-M	347710.0	302470	Women Fashion

**Gambar 2.** Gambaran Dataset

### 2.3. Preprocessing Data

Dalam proses preprocessing, perhatian khusus diberikan untuk memastikan bahwa data yang digunakan dalam penelitian telah dipersiapkan dengan baik dan siap untuk digunakan dalam analisis lebih lanjut. Langkah-langkah preprocessing ini bertujuan untuk membersihkan, mengorganisir, dan menyiapkan data agar sesuai dengan kebutuhan analisis. Beberapa langkah yang dilakukan dalam preprocessing dataset, seperti membersihkan data dari nilai yang hilang, menghapus atau mengisi nilai yang hilang, normalisasi data, dan mengubah format data menjadi format yang sesuai.

### 2.4. Cosine Similarity

Cosine similarity digunakan untuk mengevaluasi sejauh mana kemiripan pola pembelian atau preferensi antara pengguna atau item dalam sistem rekomendasi. Dengan memperhitungkan cosine similarity antara vektor yang menggambarkan pola pembelian, sistem mampu mengidentifikasi pengguna atau item yang memiliki pola pembelian yang serupa. Dengan demikian, sistem dapat menyarankan produk kepada pengguna berdasarkan preferensi dari pengguna lain yang memiliki pola serupa, sehingga meningkatkan peluang rekomendasi yang relevan dan bermanfaat [6]. Untuk mencari *cosine similarity* antara dua dokumen ditunjukkan pada persamaan berikut.

$$SIM(\vec{t}_a, \vec{t}_b) = \frac{\vec{t}_a \cdot \vec{t}_b}{|\vec{t}_a| \times |\vec{t}_b|} \quad (1)$$

Cosine similarity menghitung kesamaan antara dua vektor multidimensional,  $t_a$  dan  $t_b$ , yang mewakili kumpulan istilah  $T = \{t_1, t_2, t_3, \dots, t_n\}$ . Setiap dimensi vektor mewakili sebuah istilah dengan bobot non-negatif yang mencerminkan kehadiran istilah tersebut dalam dokumen. Hasil cosine similarity adalah nilai non-negatif yang berkisar antara 0 dan 1 [1].

### 2.5. Collaborative Filtering

Collaborative filtering adalah teknik yang digunakan untuk melakukan prediksi dengan memanfaatkan informasi tentang item yang disukai atau dipilih oleh pengguna lain [7]. Metode ini bergantung pada sekumpulan pendapat, preferensi, dan minat dari beberapa pengguna terhadap suatu item, biasanya dalam bentuk penilaian atau rating. Terdapat dua pendekatan utama dalam collaborative filtering, yaitu user-based collaborative filtering, yang juga dikenal sebagai pendekatan memory-based, dan item-based collaborative filtering, yang sering disebut sebagai pendekatan model-based [8]. User-based collaborative filtering mengacu pada konsep di mana untuk menemukan item yang menarik bagi suatu pengguna, digunakan pendekatan yang mempertimbangkan kesamaan minat dengan pengguna lain. Dalam pendekatan ini, langkah awalnya adalah mencari pengguna lain yang memiliki minat yang serupa, yang disebut sebagai tetangga pengguna (user neighbor). Nilai rating dari tetangga pengguna ini kemudian dijadikan sebagai dasar untuk memberikan rekomendasi kepada pengguna yang sedang aktif [4].

### 2.6. Evaluasi

Evaluasi kinerja sistem rekomendasi melibatkan penggunaan prediksi yang dihasilkan untuk menghitung seberapa baik kualitas rekomendasi yang diberikan. Evaluasi dilakukan dengan menggunakan metrik Mean Absolute Error (MAE), yang mengukur seberapa besar selisih antara nilai prediksi dan nilai sebenarnya dalam model [9]. MAE adalah metrik yang digunakan untuk mengevaluasi perbedaan absolut rata-rata antara nilai sebenarnya dan nilai yang diprediksi. Perhitungan MAE melibatkan mengambil rata-rata dari selisih absolut antara nilai yang diprediksi dan nilai aktual [10]. Berikut merupakan persamaan matematis dari MAE.

$$MAE = \frac{1}{n} \sum_i^n |y_i - \hat{y}| \quad (2)$$

Keterangan:

- $n$  = Banyaknya sampel
- $y_i$  = nilai aktual pada sampel ke  $i$
- $\hat{y}_i$  = nilai prediksi pada sampel ke  $i$

MAE yang mendekati 0 menandakan bahwa prediksi dari suatu metode memiliki tingkat akurasi yang tinggi, karena kesalahan prediksinya hampir tidak ada [6].

### 3. Hasil dan Diskusi

#### 3.1. Preprocessing Data

Untuk memulai, data dari empat file dataset digabungkan menggunakan proses penggabungan (merging) untuk membentuk satu set data yang lengkap. Kemudian, analisis deskriptif dilakukan untuk mendapatkan wawasan tentang dataset, seperti jumlah produk unik, total pembelian, jumlah pelanggan unik, dan jumlah produk unik. Selanjutnya, dilakukan penyaringan data untuk menghilangkan outlier dan membersihkan data dari nilai yang tidak realistis atau tidak diinginkan, dengan menjaga entri hanya untuk jumlah produk yang dipesan kurang dari 20. Langkah terakhir adalah mengganti nilai NaN dengan 0, yang digunakan untuk merepresentasikan ketiadaan interaksi atau pembelian antara pelanggan dan produk. Hal ini umum dilakukan dalam sistem rekomendasi untuk menunjukkan ketiadaan rating atau pembelian. Dengan demikian, proses ini memastikan data yang digunakan untuk analisis lebih akurat dan dapat diandalkan untuk memberikan rekomendasi yang lebih relevan dan berguna kepada pengguna.

	id	customer_id	sku_id	qty_ordered	payment_id	payment_method	sku_name	base_price	cogs	category
1	ODR7448356649d	C551551L	P886455	5	5	jazzwallet	PS4_Slim-500GB	1971942.0	1321182	Entertainment
2	ODR4011281866z	C685596L	P678648	1	4	Payaxis	Changhong Ruba 55 Inches UD55D6000i Ultra HD T...	7482000.0	5162580	Entertainment
3	ODR3378927994s	C830683L	P540013	1	5	jazzwallet	dawlance_Inverter 30	3593680.0	3054628	Appliances
4	ODR4904430099k	C191766L	P491032	1	4	Payaxis	Dawlance_Inverter-45 2.0 ton	4413220.0	3177472	Appliances
5	ODR7618778722h	C299859L	P886455	2	5	jazzwallet	PS4_Slim-500GB	1971942.0	1321182	Entertainment
...	...	...	...	...	...	...	...	...	...	...
5879	ODR4312238651y	C295679L	P194155	2	1	cod	Bahr-e-Shifa	0.0	0	Books
5880	ODR3930118778w	C368658L	P194155	1	1	cod	Bahr-e-Shifa	0.0	0	Books
5881	ODR7628482404c	C265450L	P194155	3	1	cod	Bahr-e-Shifa	0.0	0	Books
5882	ODR2809405363p	C790722L	P194155	2	1	cod	Bahr-e-Shifa	0.0	0	Books
5883	ODR6666616009w	C676393L	P100134	1	4	Payaxis	akl-GFSU1265-M	347710.0	302470	Women Fashion

5827 rows x 10 columns

Gambar 3. Hasil preprocessing

#### 3.2. Split Data

Langkah berikutnya adalah menerapkan *5-fold cross validation* pada dataset transaksi, yang mencakup ID pembeli, ID produk, dan jumlah pesanan produk. Ini dilakukan untuk memisahkan dataset menjadi data pelatihan dan pengujian. Data pelatihan digunakan untuk mengembangkan model, yang merepresentasikan pengetahuan yang akan digunakan untuk memprediksi data baru. Di sisi lain, data pengujian digunakan untuk mengevaluasi seberapa baik model dapat melakukan prediksi dengan tepat.

```
skf = StratifiedKFold((n_splits = 5))
skf.get_n_splits(X, y)

X_train, (X_test = X.loc[train_index]), X.loc[test_index]
y_train, (y_test = y.loc[train_index]), y.loc[test_index]
```

**Gambar 4.** Implementasi 5-fold cross validation

### 3.3. Cosine Similarity

Untuk tahap selanjutnya adalah menghitung kedekatan atau nilai similarity antar Customer menggunakan metode cosine Similarity Berdasarkan pola pembelian atau transaksi yang mereka miliki terhadap produk. Berikut adalah implementasi dari perhitungan cosine similarity.

```
# Menghitung user-user similarity matrix
similarity_matrix_user_user = cosine_similarity(df_final_data2)
similarity_matrix_df_user_user = pd.DataFrame(similarity_matrix_user_user, index=df_final_data.index,
columns=df_final_data.index)

# Menampilkan matriks similaritas user-user
print(similarity_matrix_df_user_user)
```

**Gambar 5.** Implementasi cosine similarity customer

### 3.4. Prediksi

Proses ini bertujuan untuk memperkirakan jumlah pembelian produk yang belum pernah dibeli oleh pelanggan, berdasarkan pola pembelian pelanggan terdekat. Tujuannya adalah untuk memberikan pemahaman yang lebih baik tentang perilaku pelanggan dan permintaan di pasar. Dalam proses ini, setiap entri dalam data uji dievaluasi, dan prediksi kuantitas pemesanan (qty\_ordered) untuk setiap produk (sku\_id) oleh setiap pelanggan (customer\_id) dihitung dengan menggunakan metode calculate\_final\_data. Nilai sebenarnya dari kuantitas pemesanan juga diambil dari data uji. Setelah mengumpulkan prediksi dan nilai sebenarnya, Mean Absolute Error (MAE) dihitung untuk mengevaluasi kualitas prediksi.

```
Mean Absolute Error (MAE): 0.33549624359287983
  Actual Qty Ordered  Predicted Qty Ordered
0                   5                   2.226287
1                   1                   1.000000
2                   1                   1.000000
3                   1                   1.000000
4                   2                   2.226287
...                 ...                   ...
4648                 3                   1.949735
4649                 1                   1.949735
4650                 2                   1.949735
4651                 3                   1.949735
4652                 2                   1.950293

[4653 rows x 2 columns]
```

**Gambar 6.** Hasil prediksi

### 3.5. Pemodelan

Pertama, fungsi mencari pelanggan lain yang memiliki kesamaan preferensi dengan pelanggan yang ditentukan, dengan memperhitungkan similarity score antara mereka. Kemudian, dari pelanggan-pelanggan yang mirip tersebut, data pesanan dari data latih dipilih. Selanjutnya, rata-rata jumlah pesanan per produk oleh pelanggan-pelanggan tersebut dihitung untuk menghasilkan rekomendasi. Informasi produk seperti nama produk kemudian ditambahkan ke dalam Data Frame rekomendasi.

```
Rekomendasi produk untuk customer dengan ID C551551L
  sku_id      sku_name
0  P886455    PS4_Slim-500GB
1  P170108  Noble_Nobel-32-Inch-Black
2  P354092    lenovo_20DF000GUE
3  P505931    Haier_LE 32B8000
4  P572907    lenovo_20EV0002UE
5  P608043    Sony_DZ 950
```

**Gambar 7.** Hasil rekomendasi

### 3.6. Evaluasi

Tahap evaluasi dilakukan untuk mengetahui seberapa baik model dapat memprediksi jumlah pesanan produk pada data uji. Ini dilakukan dengan membandingkan prediksi model dengan jumlah pesanan sebenarnya dari data uji. Pengukuran kualitas prediksi dilakukan menggunakan metode yang disebut mean absolute error (MAE), yang menghitung rata-rata selisih absolut antara prediksi dan nilai sebenarnya. Dalam praktiknya, kita melihat performa model pada setiap lipatan data uji dalam skema validasi silang. Dalam proses evaluasi ini, kita menggunakan pendekatan berbasis pengguna (user-based), di mana kita melihat kesamaan pola pembelian antar pengguna menggunakan metode cosine similarity. Hasil evaluasi dianalisis untuk mendapatkan gambaran keseluruhan tentang seberapa baik model berfungsi dalam memprediksi pola pembelian pengguna.

```
Fold 1 : 0.33085481916355325
Fold 2 : 0.31605811297313624
Fold 3 : 0.2939323386394641
Fold 4 : 0.3051225702638036
Fold 5 : 0.33549624359287983
Average MAE: 0.3162928169265674
```

**Gambar 8.** Hasil Uji MAE

Variasi nilai MAE dipengaruhi oleh sejumlah faktor, termasuk jumlah tetangga terdekat yang dipertimbangkan dalam perhitungan dan jumlah item yang telah atau belum menerima masukan dari pengguna. Hasil pengujian menunjukkan bahwa nilai MAE yang dicapai dalam penelitian ini dapat dianggap memuaskan, sesuai dengan standar akurasi yang diharapkan, yaitu mendekati nol atau minim kesalahan dengan nilai absolut [9].

## 4. Kesimpulan

Dalam penelitian ini, sistem rekomendasi produk dikembangkan menggunakan pendekatan User-Based Collaborative Filtering. Evaluasi kinerja model dilakukan dengan menggunakan metode Mean Absolute Error (MAE). Hasil evaluasi menunjukkan bahwa model mencapai tingkat akurasi yang memuaskan. Pada setiap lipatan (fold) dalam validasi silang, nilai MAE yang tercatat adalah 0,33, 0,31, 0,29, 0,30, dan 0,33. Dengan demikian, rata-rata MAE yang dihasilkan adalah 0,31.

Penelitian berikutnya dapat mempertimbangkan penggabungan metode Item-Based dengan User-Based serta penambahan algoritma dalam implementasinya.

#### Daftar Pustaka

- [1] C. S. D. Prasetya, "Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 3, p. 194, Sep. 2017, doi: <https://doi.org/10.25126/jtiik.201743392>.
- [2] A. Kurniawan, "Sistem Rekomendasi Produk Sepatu Dengan Menggunakan Metode Collaborative Filtering," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2016, no. Sentika, pp. 610–614, 2016, [Online]. Available: <https://fti.uaij.ac.id/sentika/publikasi/makalah/2016/92.pdf>.
- [3] K. H. Muliadi and C. C. Lestari, "Rancang Bangun Sistem Rekomendasi Tempat Makan Menggunakan Algoritma Typicality Based Collaborative Filtering," *Techno.Com*, vol. 18, no. 4, pp. 275–287, 2019, doi: 10.33633/tc.v18i4.2515.
- [4] H. Februariyanti, A. D. Laksono, J. S. Wibowo, and M. S. Utomo, "Implementasi Metode Collaborative Filtering Untuk Sistem Rekomendasi Penjualan Pada Toko Mebel," *Jurnal Khatulistiwa Informatika*, vol. 9, no. 1, Jun. 2021, doi: <https://doi.org/10.31294/jki.v9i1.9859>.
- [5] D. R. Pradana, S. Sa'adah, and D. Nurjanah, *Sistem Rekomendasi Sepatu Lokal Menggunakan Metode Collaborative Filtering Pada Toko Sepatu Tarsius Store*, vol. 9, Jun. 2022, ISSN 2355-9365
- [6] Hartatik, S.D.Nurhayati, and W.Widayani, "Sistem Rekomendasi Wisata Kuliner di Yogyakarta dengan Metode Item-Based Collaborative Filtering," *Journal Automation Computer Information System*, vol. 1, no. 2, pp. 55–63, Nov. 2021, doi: <https://doi.org/10.47134/jacis.v1i2.8>.
- [7] V. L. Jaja, B. Susanto, and L. R. Sasongko, "Penerapan Metode Item-Based Collaborative Filtering Untuk Sistem Rekomendasi Data MovieLens," *d'CARTESIAN*, vol. 9, no. 2, p. 78, Sep. 2020, doi: <https://doi.org/10.35799/dc.9.2.2020.28274>.
- [8] A. Wijaya and Deni Alfian, "Sistem Rekomendasi Laptop Menggunakan Collaborative Filtering Dan Content-Based Filtering," *JURNAL COMPUTECH & BISNIS*, vol. 12, no. 1, pp. 11–27, Jun. 2018.
- [9] A. A. Suryanto, "Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi," *SAINTEKBU*, vol. 11, no. 1, pp. 78–83, Feb. 2019, doi: <https://doi.org/10.32764/saintekbu.v11i1.298>.
- [10] C. Willmott and K. Matsuura, "Advantages of the Mean Absolute Error (Mae) Over The Root Mean Square Error (Rmse) In Assessing Average Model Performance," *Clim. Res.*, no. 30, pp. 79–82, 2005. doi:10.3354/cr030079

Halaman ini sengaja dibiarkan kosong