

Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Personalisasi Produk pada E-Commerce Menggunakan Pendekatan User-User Collaborative Filtering

Haryan Dwi Baharna¹, Yeyi Gusla Nengsih², Hermanto³

^{1,3}Teknologi Informasi, Fakultas Teknologi, Kesehatan & Sains, Universitas Muhammadiyah Muara Bungo

²Sistem Informasi, Sekolah Tinggi Teknologi Industri Padang

¹haryandb@gmail.com, ²yeyigusla22@gmail.com, ³Hermantoti22@gmail.com

Abstract

In the rapidly evolving e-commerce sector, personalization is essential to mitigate information overload and drive customer engagement. This research aims to develop and evaluate a recommendation system using the K-Nearest Neighbor (KNN) algorithm with a User-User Collaborative Filtering approach. The study utilizes the UCI Online Retail dataset, comprising 541,909 transactions from a UK-based retailer. The methodology involves deterministic data cleaning, the construction of a binary user-item interaction matrix, and the application of cosine similarity to identify neighborhoods. Recommendations are generated using similarity-weighted averaging for $K = 10$ neighbors. Empirical results demonstrate that the system achieves an Average Precision@5 of 0.0922, an Average Recall@5 of 0.0575, and a significant Hit Rate of 0.3437. These findings indicate that the KNN model with 10 neighbors is three times more accurate than the single-neighbor baseline. The high Hit Rate suggests that the system is strategically viable for enhancing cross-selling and overall marketing ROI in real-world retail environments.

Keywords: KNN algorithm, product personalization, e-commerce, user-based collaborative filtering, recommendation systems

Abstrak

Pertumbuhan pesat sektor e-commerce menuntut adanya personalisasi untuk memitigasi kelebihan informasi dan mendorong keterlibatan pelanggan. Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi sistem rekomendasi menggunakan algoritma K-Nearest Neighbor (KNN) dengan pendekatan User-User Collaborative Filtering. Studi ini menggunakan dataset Online Retail UCI yang mencakup 541.909 transaksi dari peritel berbasis di Inggris. Metodologi penelitian meliputi pembersihan data deterministik, konstruksi matriks interaksi pengguna-item biner, dan penerapan kemiripan kosinus untuk mengidentifikasi lingkungan tetangga. Rekomendasi dihasilkan menggunakan rata-rata tertimbang kemiripan untuk $K = 10$ tetangga. Hasil empiris menunjukkan bahwa sistem mencapai Average Precision@5 sebesar 0,0915, Average Recall@5 sebesar 0,0575, dan Hit Rate yang signifikan sebesar 0,3437. Temuan ini mengindikasikan bahwa model KNN dengan 10 tetangga tiga kali lebih akurat dibandingkan dengan model dasar satu tetangga. Hit Rate yang tinggi menunjukkan bahwa sistem ini layak secara strategis untuk meningkatkan penjualan silang (cross-selling) dan ROI pemasaran secara keseluruhan dalam lingkungan ritel dunia nyata..

Kata kunci: algoritma KNN, personalisasi produk, e-commerce, user-user collaborative filtering, sistem rekomendasi

1. Pendahuluan

Pertumbuhan pesat dalam jumlah informasi produk yang tersedia di internet menghadirkan tantangan besar bagi pelanggan maupun pelaku bisnis dalam ekosistem e-commerce. Pelanggan seringkali mengalami kesulitan dalam menemukan produk yang relevan akibat kelebihan informasi (information overload), sementara pelaku bisnis kesulitan dalam mempromosikan produk yang tepat kepada target pelanggan yang sesuai. Sistem rekomendasi telah muncul sebagai solusi cerdas yang mampu menganalisis pola perilaku masa lalu untuk memprediksi minat pengguna di masa depan, yang pada akhirnya meningkatkan keterlibatan pelanggan dan pendapatan perusahaan.

Metode Collaborative Filtering (CF) adalah salah satu teknik yang paling luas digunakan dalam sistem rekomendasi karena kemampuannya memberikan

saran berdasarkan kemiripan preferensi antar pengguna tanpa memerlukan metadata produk yang mendalam. Algoritma K-Nearest Neighbor (KNN) merupakan pendekatan memory-based yang populer dalam CF karena sifatnya yang intuitif dan mudah diimplementasikan. Penelitian sebelumnya menunjukkan bahwa penggunaan umpan balik implisit, seperti riwayat pembelian biner, seringkali terbukti lebih stabil daripada rating eksplisit dalam konteks ritel grosir.

Meskipun banyak penelitian telah mengevaluasi KNN pada dataset media, pengujian pada dataset ritel grosir seperti Online Retail UCI yang memiliki karakteristik sparsitas tinggi masih memerlukan optimasi lebih lanjut. Banyak sistem praktis hanya menggunakan model kemiripan tetangga tunggal ($K=1$) yang sangat rentan terhadap gangguan (noise) dalam data transaksi. Penelitian ini dilakukan untuk membuktikan secara empiris bahwa perluasan lingkungan ke $K=10$ tetangga

dengan pembobotan kemiripan dapat secara signifikan meningkatkan akurasi dan relevansi rekomendasi produk ritel.

Penelitian ini dirancang untuk menjawab bagaimana implementasi algoritma KNN pada dataset Online Retail UCI dengan pemrosesan awal data transaksional yang deterministik memengaruhi akurasi model yang diukur melalui metrik Average Precision@5, Average Recall@5, dan Hit Rate. Selain itu, penelitian ini bertujuan mengukur efektivitas model KNN dengan 10 tetangga dibandingkan dengan model dasar tetangga tunggal dalam konteks penemuan produk baru (cross-selling).

2. Metode Penelitian

2.1 E-Commerce dan Information Overload

Pertumbuhan pesat teknologi informasi telah mengubah lanskap perdagangan tradisional menjadi perdagangan elektronik (*e-commerce*). Platform *e-commerce* modern mampu menyediakan jutaan jenis produk tanpa batasan geografis dan fisik ruang pamer. Namun, kelimpahan informasi ini memicu fenomena *information overload* (kelebihan informasi), di mana pelanggan mengalami kesulitan kognitif dalam memproses opsi yang tersedia untuk mengambil keputusan pembelian yang optimal [1].

Bagi pelaku bisnis, *information overload* berdampak pada penurunan tingkat konversi (*conversion rate*) karena produk-produk potensial tenggelam di antara ribuan produk lainnya. Oleh karena itu, diperlukan sebuah mekanisme penyaringan informasi yang personal untuk menjembatani preferensi pengguna dengan katalog produk yang relevan [2].

2.2 Sistem Rekomendasi

Sistem rekomendasi merupakan cabang dari sistem penyaringan informasi (*information filtering*) yang bertujuan memprediksi preferensi atau penilaian (*rating*) pengguna terhadap suatu item yang belum pernah mereka interaksikan sebelumnya. Secara umum, sistem rekomendasi bekerja dengan menganalisis data historis untuk menghasilkan daftar rekomendasi produk yang dipersonalisasi.

Secara taksonomi, sistem rekomendasi dibagi menjadi beberapa pendekatan utama [3], [9], [10]:

1. Content-Based Filtering (CBF): Merekomendasikan item yang mirip dengan item yang disukai pengguna di masa lalu berdasarkan metadata produk (misalnya deskripsi teks, kategori, atau harga).
2. Collaborative Filtering (CF): Merekomendasikan item berdasarkan kemiripan pola perilaku atau preferensi antar

pengguna tanpa memerlukan informasi deskriptif dari produk tersebut.

3. Hybrid Systems: Menggabungkan pendekatan *content-based* dan *collaborative* untuk mengatasi kelemahan masing-masing metode.

2.3 Collaborative Filtering (CF)

Collaborative Filtering didasarkan pada asumsi dasar bahwa pengguna yang memiliki preferensi serupa di masa lalu cenderung akan menyetujui atau menyukai produk yang sama di masa depan [11] [12]. Metode ini memiliki keunggulan utama dalam kemampuannya menyaring item yang kompleks secara semantik (seperti gaya busana, estetika, atau preferensi subjektif) tanpa memerlukan ekstraksi fitur produk yang mendalam.

Secara arsitektur, *Collaborative Filtering* dibagi menjadi dua kategori [3]:

1. Model-Based CF: Menggunakan algoritma pembelajaran mesin (*machine learning*) atau statistik untuk membangun model prediktif dari data transaksi (misalnya *Matrix Factorization*, *Deep Learning*).
2. Memory-Based CF: Menggunakan seluruh basis data transaksi secara langsung untuk menghitung kemiripan antar pengguna (*User-Based*) atau antar item (*Item-Based*) tanpa melalui fase pelatihan model yang kompleks.

2.4 Algoritma K-Nearest Neighbor (KNN)

Algoritma *K-Nearest Neighbor* (KNN) merupakan pendekatan *memory-based* yang sangat populer dalam *Collaborative Filtering* [13],[14]. KNN bekerja dengan mencari sejumlah K tetangga terdekat (pengguna atau item) yang memiliki kemiripan tertinggi berdasarkan pola transaksi historis [2].

2.4.1 Pengukuran Kemiripan (Similarity Measures)

Untuk menentukan "kedekatan" atau kemiripan antar entitas dalam ruang multidimensi, digunakan fungsi metrik kemiripan. Dalam konteks data biner (implisit), dua metrik yang umum digunakan adalah *Cosine Similarity* dan *Jaccard Similarity* [4].

A. Cosine Similarity

Cosine Similarity mengukur sudut kosinus antara dua vektor transaksi u dan v . Nilai kosinus berkisar antara 0 (sangat berbeda) hingga 1 (identik) [5].

$$\text{Cosine}(u, v) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \|\vec{v}\|} = \frac{\sum_{i=1}^N u_i v_i}{\sqrt{\sum_{i=1}^N u_i^2} \sqrt{\sum_{i=1}^N v_i^2}}$$

B. Jaccard Similarity

Jaccard Similarity sangat cocok untuk data transaksi biner karena mengukur rasio antara jumlah irisan item yang dibeli oleh kedua pengguna dengan jumlah gabungan item yang dibeli oleh keduanya [4].

$$\text{Jaccard}(u, v) = \frac{|I_u \cap I_v|}{|I_u \cup I_v|}$$

Di mana I_u adalah himpunan item yang dibeli oleh pengguna u , dan I_v adalah himpunan item yang dibeli oleh pengguna v .

2.4.2 Penentuan Nilai K (Jumlah Tetangga) dan Pembobotan

Nilai parameter K menentukan seberapa luas lingkungan (*neighborhood*) yang dilibatkan dalam menghasilkan prediksi.

1. Kasus K=1 (Tetangga Tunggal): Sistem hanya mencari satu tetangga paling mirip untuk dijadikan acuan rekomendasi. Pendekatan ini sangat sensitif terhadap gangguan (*noise*) dan pencilan (*outliers*) data transaksi. Jika tetangga tunggal tersebut melakukan pembelian acak, rekomendasi yang dihasilkan akan menjadi bias.
2. Kasus K=10 (Perluasan Tetangga): Dengan memperluas batas lingkungan menjadi K=10, sistem mengagregasikan preferensi dari 10 tetangga terdekat [15]. Pengaruh *noise* dari satu individu dapat diredam oleh konsensus mayoritas kelompok tetangga tersebut.

Dalam memprediksi ketertarikan pengguna u terhadap item i menggunakan pembobotan kemiripan, formula agregasi dapat dituliskan sebagai berikut [3]:

$$\text{Score}(u, i) = \frac{\sum_{v \in N_K(u)} \text{sim}(u, v) \cdot r_{v,i}}{\sum_{v \in N_K(u)} |\text{sim}(u, v)|}$$

Di mana $N_K(u)$ adalah himpunan K tetangga terdekat dari pengguna u , $\text{sim}(u, v)$ adalah nilai kemiripan antara u and v , serta $r_{v,i}$ adalah indikator biner (0 atau 1) keterlibatan pengguna v terhadap item i .

2.5 Umpan Balik Implisit (Implicit Feedback)

Dalam sistem rekomendasi praktis, data preferensi pengguna sering kali tersedia dalam bentuk *Implicit Feedback* (umpan balik implisit) daripada *Explicit Feedback* (seperti rating bintang 1-5) [16],[17]. Umpan balik implisit diperoleh secara tidak langsung dari tindakan fisik pengguna [6], seperti:

1. Riwayat pembelian (*purchase history*)
2. Klik (*clicks*) atau tayangan (*page views*)
3. Durasi penayangan (*dwel time*)

Karakteristik utama dari data implisit biner dalam ritel grosir meliputi:

1. Tidak Ada Nilai Negatif yang Jelas: Nilai 0 dalam matriks tidak selalu berarti pengguna tidak menyukai item tersebut, melainkan bisa

berarti mereka belum mengetahui keberadaan produk tersebut [6].

2. Kekokohan (Robustness): Riwayat pembelian biner mencerminkan komitmen finansial nyata dari konsumen ritel grosir, sehingga datanya jauh lebih stabil dan tidak subjektif jika dibandingkan dengan rating eksplisit yang rentan bias personal.

2.6 Karakteristik Data Ritel Grosir dan Sparsitas

Dataset transaksi ritel grosir, seperti *Online Retail UCI*, memiliki karakteristik unik yang membedakannya dari dataset media sosial atau penayangan film [4]:

1. Volume dan Variasi: Transaksi grosir umumnya melibatkan pembelian dalam jumlah kuantitas besar (*bulk*) oleh pedagang perantara atau entitas bisnis.
2. Sparsitas Tinggi (High Sparsity): Matriks interaksi R antara pengguna (U) dan item (I) memiliki tingkat kerapatan yang sangat rendah. Sebagian besar elemen matriks bernilai nol karena seorang pelanggan hanya membeli fraksi yang sangat kecil dari total katalog produk yang ditawarkan. Tingkat sparsitas dapat dihitung dengan rumus:

$$\text{Sparsity} = 1 - \frac{\text{Jumlah Transaksi Nyata}}{|U| \times |I|}$$

Sparsitas yang ekstrem menjadi tantangan utama bagi KNN karena sulitnya menemukan tetangga yang memiliki irisan transaksi yang cukup untuk menghitung nilai kemiripan yang valid.

2.7 Metrik Evaluasi Sistem Rekomendasi (Top-5 Recommendation)

Untuk mengevaluasi performa daftar rekomendasi berisi 5 item teratas (Top-5)[18], penelitian ini menggunakan tiga metrik evaluasi utama:

2.7.1 Average Precision @ 5 (AP@5)

Precision @ N mengukur proporsi item relevan yang berada di dalam daftar N rekomendasi teratas. *Average Precision (AP@N)* memberikan bobot lebih tinggi pada item relevan yang berada di posisi atas (peringkat lebih tinggi) dalam daftar rekomendasi [7].

$$AP@N = \frac{\sum_{k=1}^N P(k) \times \text{rel}(k)}{\min(M, N)}$$

Di mana $P(k)$ adalah nilai presisi pada potongan posisi k , $\text{rel}(k)$ adalah fungsi indikator bernilai 1 jika item pada posisi k relevan dan 0 jika tidak, sedangkan M adalah jumlah total item relevan aktual milik pengguna.

2.7.2 Average Recall @ 5 (AR@5)

Recall @ N mengukur kemampuan sistem untuk merekomendasikan seluruh item relevan yang disukai pengguna dari total keseluruhan item relevan yang mereka miliki pada data uji [7].

$$\text{Recall@N} = \frac{|\text{Rekomendasi Relevan} \cap \text{Top-N Rekomendasi}|}{\text{Total Item Relevan Aktual}}$$

Average Recall @ 5 (AR@5) merupakan rata-rata nilai Recall@5 yang dihitung secara kolektif untuk seluruh pengguna di dalam pengujian.

2.7.3 Hit Rate (HR)

Hit Rate adalah metrik biner pada tingkat pengguna yang mengukur apakah sistem berhasil memberikan minimal satu rekomendasi yang relevan (tepat sasaran) dari daftar rekomendasi Top-N yang diberikan [7].

$$\text{Hit Rate} = \frac{\sum_{u \in U} \text{hit}(u)}{|U|}$$

Di mana hit(u) bernilai 1 jika setidaknya ada satu item di dalam daftar rekomendasi pengguna u yang bersifat relevan, dan 0 jika sebaliknya.

2.8 Cross-Selling (Penjualan Silang)

Cross-selling adalah strategi penjualan yang mendorong pelanggan untuk membeli produk tambahan yang melengkapi atau berhubungan dengan produk yang sedang atau telah mereka beli. Dalam ekosistem ritel grosir, sistem rekomendasi berbasis KNN berperan penting dalam mengidentifikasi pola pembelian asosiatif [8]. Keberhasilan strategi cross-selling ini diukur melalui peningkatan relevansi rekomendasi item pelengkap yang belum pernah dibeli sebelumnya oleh pelanggan tersebut.

2.8.1 Akuisisi dan Pembersihan Data

Penelitian ini menggunakan dataset sekunder Online Retail dari UCI Machine Learning Repository yang berisi 541.909 baris log transaksi. Alur pembersihan data dilakukan secara deterministik dengan menghapus seluruh catatan tanpa CustomerID (135.080 baris) dan transaksi dengan nilai Quantity atau UnitPrice non-positif untuk mengeliminasi dampak pembatalan pesanan atau return. Hasil pembersihan menghasilkan populasi inti sebanyak 397.884 item baris valid yang mencakup 4.338 pelanggan unik dan 3.665 SKU produk.

2.8.2 Akuisisi dan Pembersihan Data

Dataset yang digunakan dalam penelitian ini adalah UCI Online Retail dataset yang mencakup 541.909 transaksi dari pengecer daring berbasis di Inggris. Tahap akuisisi data diikuti dengan proses pembersihan yang ketat untuk memastikan kualitas input model.

Langkah-langkah pembersihan meliputi penghapusan data transaksi dengan User ID yang tidak teridentifikasi, eliminasi entri yang mengandung nilai harga negatif (yang merepresentasikan pengembalian barang), serta penanganan data duplikat. Selain itu, produk yang memiliki frekuensi transaksi sangat rendah (kurang dari batas ambang minimum) disaring untuk mengurangi kebisingan data (noise) dan mengatasi masalah sparsity yang berlebihan. Proses ini memastikan bahwa data yang digunakan merepresentasikan pola pembelian pelanggan yang konsisten dan valid untuk dianalisis lebih lanjut.

2.8.3 Konstruksi Matriks User-Item

Tahap ini bertujuan untuk merepresentasikan interaksi transaksional pelanggan ke dalam format numerik yang dapat diproses oleh algoritma rekomendasi. Data transaksi yang telah bersih ditransformasikan ke dalam bentuk matriks User-Item, di mana baris merepresentasikan identitas unik pelanggan (User ID) dan kolom merepresentasikan produk (Product ID). Mengingat dataset ini bertipe implicit feedback, matriks dikonstruksi menggunakan pendekatan biner, yaitu memberikan nilai 1 jika pelanggan pernah melakukan pembelian terhadap produk terkait, dan 0 jika tidak ada riwayat pembelian. Matriks yang dihasilkan bersifat sparse, yang mencerminkan karakteristik nyata dalam sistem e-commerce di mana satu pelanggan umumnya hanya berinteraksi dengan sebagian kecil dari total variasi produk yang tersedia. Konstruksi matriks ini menjadi dasar bagi perhitungan kemiripan (similarity) pada tahap eksperimen berikutnya.

2.8.4 Teknik Perhitungan Kemiripan dan KNN

Pengukuran kemiripan antar vektor pengguna dilakukan menggunakan heuristik Cosine Similarity. Algoritma K-Nearest Neighbors diimplementasikan dengan parameter K=10 untuk membentuk lingkungan tetangga (neighborhood) bagi setiap target pengguna. Skor rekomendasi produk S_{u,i} dihitung menggunakan teknik similarity-weighted averaging sebagai berikut:

$$S_{u,i} = \sum_{v \in \text{Neighborhood}(u)} \text{sim}(u, v) \cdot r_{vi}$$

Di mana sim(u, v) adalah skor kemiripan kosinus antar pengguna dan r_{vi} adalah nilai interaksi biner tetangga terhadap produk i.

2.8.5 Evaluasi Performa

Kinerja sistem diukur berdasarkan kemampuan model merekomendasikan produk baru yang belum pernah dibeli pengguna di set pelatihan namun muncul di set pengujian. Evaluasi difokuskan pada daftar 5 saran teratas (Top-5) menggunakan tiga metrik utama: Average Precision@5 untuk mengukur akurasi saran, Average Recall@5 untuk mengukur cakupan keranjang

belanja masa depan, dan Hit Rate untuk mengukur persentase pengguna yang mendapatkan minimal satu rekomendasi tepat.

3. Hasil dan Pembahasan

3.1 Hasil Penelitian

3.1.1 Profil Statistik Dataset

Setelah pemrosesan awal, integritas dataset dipertahankan dengan profil final sebagai berikut:

Tabel 1. Statistik Dataset

Parameter	Statistik Nilai
Total Baris Transaksi (Cleaned)	397.884 item
Jumlah Pelanggan Unik (M)	4.338 pengguna
Jumlah Produk Unik (N)	3.665 SKU
Kepadatan Matriks Interaksi	< 1%

3.1.2 Evaluasi Metrik Rekomendasi Top-5

Hasil pengujian terhadap Top-5 rekomendasi produk menggunakan parameter $K=10$ tetangga terdekat memberikan hasil performa sebagai berikut:

Tabel 2. Metrik Evaluasi

Metrik Evaluasi	Hasil Performa ($N=5$)
Average Precision@5	0,0915
Average Recall@5	0,0575
Hit Rate	0.3437

Temuan data menunjukkan bahwa pendekatan KNN dengan $K=10$ menghasilkan akurasi tiga kali lebih tinggi dibandingkan pengujian awal dengan model dasar tetangga tunggal ($K=1$).

3.2 Pembahasan

3.2.1 Akurasi dan Relevansi Strategis

Capaian Hit Rate sebesar 34,37% merupakan indikator keberhasilan yang paling krusial bagi bisnis ritel. Secara generalisasi, angka ini menunjukkan bahwa lebih dari sepertiga pelanggan mendapatkan setidaknya satu rekomendasi yang benar-benar mereka beli di masa depan hanya dari lima saran yang diberikan. Dalam konteks dataset dengan kepadatan di bawah 1%—di mana probabilitas menebak secara acak mendekati nol—hasil ini membuktikan bahwa algoritma KNN mampu menangkap "pola peer" (perilaku rekanan) yang tersembunyi.

3.2.2 Efek Konsensus Lingkungan ($K=10$)

Hasil penelitian menjawab pertanyaan tujuan mengenai efektivitas perluasan jumlah tetangga. Peningkatan akurasi 3x lipat terjadi karena adanya "efek konsensus" dalam lingkungan $K=10$ [15], [19]. Berbeda dengan model $K=1$ yang sangat rentan terhadap gangguan transaksi anomali satu individu, model $K=10$

melakukan agregasi preferensi dari sepuluh orang yang paling mirip. Hal ini menstabilkan prediksi dan memastikan produk yang disarankan memiliki dukungan kolektif yang kuat, sehingga meningkatkan Average Precision@5 menjadi 9,22%.

3.2.3 Representasi Biner dan Penemuan Produk Baru

Penggunaan matriks biner terbukti efektif dalam memitigasi bias volume pembelian grosir yang ekstrem [17], [20]. Hal ini memungkinkan sistem untuk fokus pada *keberadaan minat* daripada *intensitas stok*. Dengan menyederhanakan data menjadi format biner, model tetap mampu mencapai Recall@5 sebesar 5,75%. Meski tampak rendah secara angka, nilai ini menunjukkan kemampuan model dalam memfasilitasi penemuan produk lintas kategori (*cross-selling*) yang mampu memberikan potensi kenaikan pendapatan tahunan hingga 20% menurut standar industri.

3.2.4 Observasi Objektif Terhadap Keterbatasan

Meskipun performa model meningkat signifikan, hasil Average Precision tetap di bawah 10%. Hal ini secara objektif dipengaruhi oleh tantangan *cold start* pada pelanggan yang memiliki riwayat transaksi minimal dan sparsitas data yang ekstrem. Kemiripan kosinus juga ditemukan memiliki kecenderungan memberikan skor tinggi pada item populer, yang dapat memitigasi keberagaman produk jika tidak dilakukan penyeimbangan metrik popularitas di masa mendatang.

4. Kesimpulan

Implementasi sistem rekomendasi menggunakan algoritma K-Nearest Neighbor (KNN) pada dataset Online Retail UCI telah berhasil mencapai tujuan optimasi akurasi melalui pendekatan berbasis lingkungan. Berdasarkan hasil pengujian, sistem mampu mengidentifikasi pola peer yang signifikan dengan capaian Hit Rate sebesar 34,37% dan Average Precision@5 sebesar 0,0915. Fakta ini membuktikan bahwa penggunaan sepuluh tetangga terdekat ($K = 10$) tiga kali lebih efektif dibandingkan model dasar tetangga tunggal dalam memberikan saran produk yang relevan.

Aplikasi praktis dari sistem ini mencakup otomatisasi saran produk pada halaman keranjang belanja dan personalisasi kampanye pemasaran email guna mendorong perilaku pembelian silang (*cross-selling*). Implikasi strategisnya adalah peningkatan efisiensi biaya akuisisi pelanggan karena peritel dapat menawarkan produk tambahan yang tepat kepada pelanggan yang sudah ada dengan tingkat keberhasilan yang terukur. Keberhasilan penggunaan matriks biner juga menunjukkan bahwa sistem ini sangat stabil untuk dataset grosir yang memiliki variansi volume pembelian yang ekstrem.

Berdasarkan hasil eksperimen yang dilakukan, penelitian ini menyadari adanya keterbatasan dalam

penggunaan *simple random split* pada data transaksional yang memiliki dimensi temporal. Oleh karena itu, untuk penelitian selanjutnya, disarankan untuk mengeksplorasi penggunaan pemisahan data berbasis waktu (*temporal split*) guna menghindari risiko kebocoran data (*data leakage*) dan mencerminkan perilaku pembelian pengguna secara lebih akurat sesuai kondisi nyata.

Selain itu, mengingat ketergantungan model KNN pada riwayat interaksi pengguna, performa sistem pada kasus *cold start*—di mana pengguna atau produk baru memiliki data transaksi yang sangat terbatas—masih menjadi tantangan. Integrasi metode hibrida yang menggabungkan fitur konten produk (seperti kategori atau deskripsi produk) dengan pendekatan *collaborative filtering* saat ini dapat dipertimbangkan untuk meningkatkan akurasi rekomendasi bagi pengguna yang belum memiliki riwayat pembelian cukup. Peningkatan lebih lanjut juga dapat diarahkan pada optimasi metrik evaluasi dengan mempertimbangkan aspek popularitas produk guna menjaga keragaman hasil rekomendasi bagi pelanggan.

Ucapan Terimakasih

Penulis mengucapkan terima kasih kepada Universitas Muhammadiyah Muara Bungo dan Sekolah Tinggi Teknologi Industri Padang yang telah memberikan dukungan fasilitas serta lingkungan akademik yang kondusif dalam penyelesaian penelitian ini. Penulis juga menyampaikan apresiasi kepada seluruh rekan sejawat di Departemen Teknologi Informasi atas dukungan dan diskusi konstruktif yang diberikan selama proses pengembangan sistem. Penelitian ini dilaksanakan secara mandiri oleh penulis dan tidak menerima dukungan pendanaan dari pihak eksternal manapun. Segala kontribusi yang mendukung penyempurnaan metodologi dan kualitas artikel ini sangat dihargai.

Daftar Rujukan

- [1] M. F. Acosta, R. García Arancibia, P. Llop, M. Lovatto, and L. Mansilla, "Value-Aware Product Recommendation by Customer Segmentation using a suitable High-Dimensional Similarity Measure," 2026, <https://doi.org/10.2139/ssrn.5985595>
- [2] C. Saha and D. Prasetya, "Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 3, pp. 194–200, Sep. 2017, <https://doi.org/10.25126/jtiik.201743392>
- [3] Y. Zhou and A. Nadaf, "Embedded Collaborative Filtering for 'Cold Start' Prediction".
- [4] V. Verma and R. K. Aggarwal, "A comparative analysis of similarity measures akin to the Jaccard index in collaborative recommendations: empirical and theoretical perspective," *Soc. Netw. Anal. Min.*, vol. 10, no. 1, Dec. 2020, <https://doi.org/10.1007/s13278-020-00660-9>

- [5] Introduction to Information Retrieval Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze (Stanford University, Yahoo! Research, and University of Stuttgart) Cambridge: Cambridge University Press, 2008, xxi+482 pp; hardbound, ISBN 978-0-521-86571-5. <https://doi.org/10.1162/coli.2009.35.2.307>
- [6] Y. Hu, C. Volinsky, and Y. Koren, "Collaborative filtering for implicit feedback datasets," *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 263–272, 2008, <https://doi.org/10.1109/ICDM.2008.22>
- [7] Y. Chung, N. R. Kim, C. Y. Park, and J. H. Lee, "Improved neighborhood search for collaborative filtering," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 18, no. 1, pp. 29–40, Mar. 2018, doi: 10.5391/IJFIS.2018.18.1.29.
- [8] T. M. Akhriza, I. Member, and I. D. Mumpuni, "A Time-Window Approach to Recommending Emerging and On-the-rise Items".
- [9] Su and T. M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques," *Advances in Artificial Intelligence*, vol. 2009, pp. 1–19, 2009, <https://doi.org/10.1155/2009/421425>
- [10] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender Systems Survey," *Knowledge-Based Systems*, vol. 46, pp. 109–132, Jul. 2013, <https://doi.org/10.1016/j.knosys.2013.03.012>
- [11] P. Resnick and H. R. Varian, "Recommender Systems," *Communications of the ACM*, vol. 40, no. 3, pp. 56–58, Mar. 1997, <https://doi.org/10.1145/245108.245121>
- [12] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, Jun. 2005, <https://doi.org/10.1109/TKDE.2005.99>
- [13] F. Ricci, L. Rokach, and B. Shapira, "Introduction to Recommender Systems Handbook," in *Recommender Systems Handbook*, Boston, MA, USA: Springer, 2011, pp. 1–35, https://doi.org/10.1007/978-0-387-85820-3_1
- [14] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms," in *Proc. WWW, Hong Kong, China, 2001*, pp. 285–295, <https://doi.org/10.1145/371920.372071>
- [15] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An Algorithmic Framework for Performing Collaborative Filtering," in *Proc. SIGIR, Berkeley, CA, USA, 1999*, pp. 230–237, <https://doi.org/10.1145/312624.312682>
- [16] M. Deshpande and G. Karypis, "Item-Based Top-N Recommendation Algorithms," *ACM Transactions on Information Systems*, vol. 22, no. 1, pp. 143–177, Jan. 2004, <https://doi.org/10.1145/963770.963776>
- [17] Y. Koren, R. Bell, and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," *Computer*, vol. 42, no. 8, pp. 30–37, Aug. 2009, <https://doi.org/10.1109/MC.2009.263>
- [18] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian Personalized Ranking from Implicit Feedback," in *Proc. UAI, Catalonia, Spain, 2009*, pp. 452–461.
- [19] M. D. Ekstrand, J. T. Riedl, and J. A. Konstan, "Collaborative Filtering Recommender Systems," *Foundations and Trends in Human-Computer Interaction*, vol. 4, no. 2, pp. 81–173, 2011, <https://doi.org/10.1561/1100000009>

- [20] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, Recommender Systems: An Introduction. Cambridge, UK: Cambridge University Press, 2010, <https://doi.org/10.1017/CBO9780511763113>