

SISTEM REKOMENDASI *CROSS-SELLING* BERBASIS *ITEM SIMILARITY* MENGGUNAKAN *COSINE SIMILARITY* PADA TRANSAKSI PENJUALAN

Erni Seniwati^{1*}, Acihmah Sidauruk², Ninik Tri Hartanti³, Rina Pramitasari⁴

^{1*}Univeritas Langlangbuana Bandung

^{2,3,4}Univeritas Amikom Yogyakarta

Email: ^{1*}erni.seniwati@unla.ac.id, ²acihmah@amikom.ac.id, ³ninik.t@amikom.ac.id,
⁴rina.pramitasari@amikom.ac.id

Submit: 05 Juni 2026, Diterima: 02 Juli 2026, Diterbitkan: 30 Juli 2026

Abstrak: Data transaksi ritel dapat dimanfaatkan untuk mengidentifikasi keterkaitan antarproduk dan mendukung strategi cross-selling, terutama ketika data rating, profil, atau identitas pelanggan tidak tersedia. Penelitian ini bertujuan membangun baseline sistem rekomendasi cross-selling berbasis Item Similarity menggunakan Cosine Similarity dari pola co-occurrence pada transaksi penjualan. Dataset terdiri atas 77 transaksi dan 296 produk. Setelah pembersihan data, transaksi ditransformasikan menjadi matriks transaksi-item biner dan ditransposisikan menjadi vektor item-transaksi untuk menghitung kemiripan antarproduk dan menghasilkan rekomendasi Top-N. Evaluasi menggunakan protokol dua tingkat: leave-one-transaction-out, yaitu mengeluarkan seluruh transaksi uji dari data pembentuk matriks similarity untuk mencegah kebocoran informasi, kemudian leave-one-item-out di dalam transaksi uji, yaitu menahan satu item sebagai target dan menggunakan item lainnya sebagai masukan. Sebanyak 71 transaksi memenuhi syarat dan menghasilkan 368 kasus uji. Pasangan B148 dan B151 memperoleh Cosine Similarity 0,894. Pada Top-10, model menghasilkan Hit Rate/Recall@10 sebesar 8,15%, Precision@10 sebesar 0,82%, dan Mean Reciprocal Rank sebesar 0,040. Performa yang relatif rendah berkaitan dengan sparsity matriks sebesar 98,36% dan dominasi produk berfrekuensi rendah. Hasil ini menunjukkan bahwa Item Similarity berbasis transaksi dapat digunakan sebagai baseline yang sederhana dan interpretable untuk cross-selling pada ritel dengan data terbatas, sekaligus menegaskan perlunya data interaksi yang lebih kaya untuk meningkatkan kualitas rekomendasi.

Kata kunci: Cosine Similarity, cross-selling, Item Similarity, sistem rekomendasi, transaksi penjualan.

Abstract: Retail transaction data can be exploited to identify inter-product relationships and support cross-selling, particularly when customer ratings, profiles, or identities are unavailable. This study develops a baseline cross-selling recommender based on Item Similarity and Cosine Similarity using product co-occurrence in sales transactions. The dataset contains 77 transactions and 296 products. After data cleaning, transactions were transformed into a binary transaction-item matrix and transposed into item-transaction vectors to calculate inter-product similarity and generate Top-N recommendations. Evaluation followed a two-level protocol: leave-one-transaction-out, in which the complete test transaction was excluded from the data used to construct the similarity matrix to prevent information leakage, followed by leave-one-item-out within the held-out transaction, in which one item was treated as the target and the remaining items as the recommendation input. Seventy-one eligible transactions produced 368 test cases. Products B148 and B151 achieved a Cosine Similarity of 0.894. At Top-10, the model obtained a Hit Rate/Recall@10 of 8.15%, Precision@10 of 0.82%, and Mean Reciprocal Rank of 0.040. The relatively low performance is associated with 98.36% matrix sparsity and the dominance of low-frequency products. These findings indicate that transaction-based Item Similarity provides a simple and interpretable baseline for cross-selling in small retail datasets while highlighting the need for richer interaction data to improve recommendation quality.

Keywords: Cosine Similarity, cross-selling, Item Similarity, recommendation system, sales transaction data.

1. PENDAHULUAN

Perkembangan teknologi informasi telah mengubah cara pelaku usaha mengelola data transaksi penjualan menjadi sumber informasi yang bernilai dalam mendukung pengambilan keputusan bisnis. Data transaksi yang tersimpan pada sistem penjualan tidak hanya berfungsi sebagai dokumentasi aktivitas penjualan, tetapi juga dapat dimanfaatkan untuk menggali pola pembelian konsumen sehingga menghasilkan informasi yang mendukung strategi pemasaran. Namun, pada sebagian besar usaha ritel, khususnya usaha kecil dan menengah, data transaksi masih dimanfaatkan sebatas pelaporan penjualan dan pengelolaan stok

barang. Rekomendasi produk kepada pelanggan umumnya masih dilakukan berdasarkan pengalaman penjual atau intuisi, sehingga sering kali kurang sesuai dengan pola pembelian konsumen yang sebenarnya. Kondisi tersebut menyebabkan peluang penerapan strategi *cross-selling* belum dimanfaatkan secara optimal, padahal strategi ini berpotensi meningkatkan nilai transaksi sekaligus memberikan pengalaman berbelanja yang lebih baik bagi pelanggan [1], [2].

Sistem rekomendasi merupakan salah satu teknologi yang berkembang pesat dalam bidang *data mining* dan *artificial intelligence* untuk membantu pengguna memperoleh informasi atau produk yang relevan berdasarkan karakteristik data yang dimiliki. Berbagai pendekatan telah dikembangkan dalam sistem rekomendasi, di antaranya *Content-Based Filtering*, *Collaborative Filtering*, *Knowledge-Based Recommendation*, *Hybrid Recommendation*, serta pendekatan berbasis kemiripan (*Similarity-Based Recommendation*) [3]. Pada lingkungan ritel yang hanya memiliki data transaksi penjualan, pendekatan berbasis kemiripan item (*Item Similarity*) menjadi salah satu metode yang sesuai karena tidak memerlukan informasi profil maupun preferensi pengguna secara eksplisit. Hubungan antarproduk diperoleh dari pola kemunculan produk dalam transaksi sehingga rekomendasi dapat dihasilkan berdasarkan keterkaitan antaritem yang pernah dibeli secara bersamaan [4], [5].

Metode *Cosine Similarity* merupakan salah satu teknik pengukuran kemiripan yang banyak digunakan dalam sistem rekomendasi karena mampu mengukur tingkat kesamaan antarobjek melalui representasi vektor. Keunggulan metode ini terletak pada kemampuannya mengidentifikasi hubungan antaritem berdasarkan arah vektor tanpa dipengaruhi oleh besarnya nilai frekuensi secara langsung, sehingga efektif diterapkan pada data transaksi biner maupun data dengan dimensi yang tinggi [6], [3]. Selain memiliki kompleksitas komputasi yang relatif rendah, metode ini juga mudah diimplementasikan dan mampu menghasilkan rekomendasi yang relevan pada berbagai domain, seperti e-commerce, rekomendasi dokumen ilmiah, maupun sistem rekomendasi berbasis *collaborative filtering* [7], [5].

Berbagai penelitian terdahulu menunjukkan bahwa metode *Cosine Similarity* memiliki kinerja yang baik dalam membangun sistem rekomendasi. Pratama *et al.* [1] mengembangkan sistem rekomendasi produk e-commerce menggunakan pendekatan *Content-Based Filtering* dan menunjukkan bahwa metode tersebut mampu memberikan rekomendasi berdasarkan kemiripan atribut produk. Adam dan Mokodaser [8] juga menerapkan *Content-Based Filtering* berbasis *Cosine Similarity* pada sistem rekomendasi produk e-commerce dengan hasil yang mampu meningkatkan relevansi rekomendasi. Penelitian Pratomo dan Utami [9] menerapkan metode serupa pada sistem rekomendasi topik tugas akhir mahasiswa dan menunjukkan bahwa *Cosine Similarity* efektif dalam mengukur kedekatan antarobjek. Sementara itu, Putra *et al.* [7] memanfaatkan *Sentence Transformers* yang dikombinasikan dengan *Cosine Similarity* untuk rekomendasi artikel ilmiah berbasis data arXiv. Pada sisi lain, Abdalla *et al.* [4] mengembangkan pendekatan *Item-Based Collaborative Filtering* menggunakan ukuran kemiripan baru untuk meningkatkan kualitas rekomendasi, sedangkan Wang *et al.* [6] mengembangkan pendekatan berbasis pola *Cosine* guna meningkatkan rekomendasi pada produk *long-tail*. Penelitian Bellogin *et al.* [3] juga menunjukkan bahwa pemilihan ukuran kemiripan memiliki pengaruh signifikan terhadap kualitas sistem rekomendasi. Selain itu, Kim *et al.* [2] mengembangkan sistem rekomendasi berbasis urutan (*sequence-aware recommendation*) pada e-commerce fesyen, sedangkan Alsobhi dan Amare [10] mengusulkan sistem rekomendasi berbasis ontologi untuk meningkatkan hubungan semantik antarproduk.

Kajian sistem rekomendasi ritel menunjukkan bahwa transaksi pembelian telah banyak dipelajari sebagai implicit feedback. Pada market basket dan next-basket recommendation, riwayat keranjang digunakan untuk memprediksi produk yang berpotensi dibeli bersama atau pada transaksi berikutnya; pendekatan mutakhir bahkan memodelkan urutan keranjang, pengulangan pembelian, konteks waktu, dan hubungan antarpaket item [11], [12]. Pada session-based recommendation, hubungan item juga dapat dibentuk langsung dari co-occurrence antarproduk dan digunakan untuk menghasilkan kandidat rekomendasi tanpa rating eksplisit [13], [14]. Dengan demikian, penggunaan transaksi sebagai sumber informasi bukan merupakan novelty yang berdiri sendiri. Ruang yang dikaji penelitian ini lebih spesifik, yaitu kondisi ritel berukuran kecil ketika data yang tersedia hanya berupa ID transaksi dan daftar produk, tanpa customer ID, rating, urutan sesi yang kaya, maupun atribut konten produk. Dalam kondisi tersebut, model yang kompleks dan personal sulit diterapkan, sementara baseline Item Similarity yang sederhana, transparan, dan mudah direplikasi tetap diperlukan untuk mengukur seberapa jauh sinyal co-occurrence pada data point-of-sale dapat mendukung cross-selling.

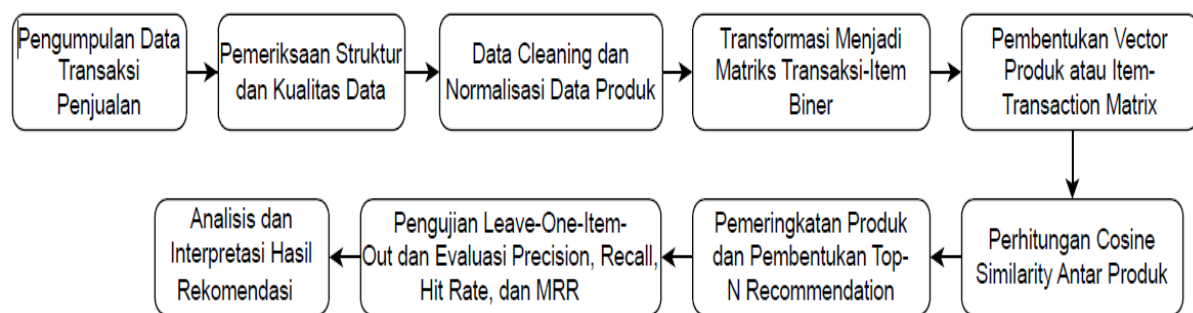
Berdasarkan permasalahan tersebut, penelitian ini bertujuan membangun dan mengevaluasi baseline sistem rekomendasi cross-selling berbasis Item Similarity menggunakan Cosine Similarity pada data transaksi penjualan. Tahapan penelitian meliputi preprocessing, pembentukan matriks transaksi-item biner, pembentukan vektor item-transaksi, perhitungan kemiripan antarproduk, pemeringkatan Top-N, serta evaluasi offline. Penelitian dibatasi pada data transaksi tanpa identitas pelanggan, data demografi, rating,

maupun fitur konten produk sehingga kemampuan model benar-benar ditentukan oleh pola co-occurrence yang tersedia.

Kontribusi penelitian ini tidak ditempatkan pada klaim bahwa pemanfaatan transaksi atau co-occurrence merupakan pendekatan baru, melainkan pada tiga aspek. Pertama, penelitian menyajikan baseline cross-selling yang dapat diterapkan ketika ritel hanya memiliki data point-of-sale sederhana. Kedua, penelitian mengevaluasi baseline tersebut dengan protokol dua tingkat yang memisahkan transaksi uji dari pembentukan similarity dan kemudian menahan satu item sebagai target, sehingga evaluasi lebih ketat terhadap kebocoran informasi. Ketiga, penelitian mengukur secara empiris keterbatasan baseline pada dataset sangat sparse melalui Hit Rate/Recall@K, Precision@K, MRR, dan analisis frekuensi item. Posisi ini membedakan penelitian dari next-basket atau session-based recommender yang memanfaatkan histori pelanggan, urutan interaksi, maupun model yang lebih kompleks [13], [11]–[14].

2. METODE

Pada penelitian ini memiliki alur penelitian yang dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian Sistem Rekomendasi Produk

2.1 Pengumpulan Data

Pengumpulan data dilakukan pada satu objek yaitu koperasi Citra Mart di Universitas Amikom Yogyakarta. Dataset yang digunakan bernama Transaksi Penjualan yang terdiri dari 77 data transaksi penjualan dari arsip transaksi penjualan pada periode Januari 2019 sampai November 2019. Dataset didukung dengan data produk berjumlah 296. Dokumentasi dataset tidak memuat informasi mengenai kerangka sampel, algoritma pengacakan, proporsi sampel terhadap seluruh transaksi, maupun seed pengacakan. Oleh karena itu, penelitian ini tidak mengklaim 77 transaksi tersebut sebagai sampel acak probabilistik, melainkan sebagai sampel transaksi yang tersedia (available transaction sample) dari periode pengamatan. Keterbatasan ini dipertimbangkan dalam interpretasi hasil karena ukuran sampel yang kecil dan mekanisme pemilihan yang tidak terdokumentasi dapat membatasi generalisasi.

2.2 Pemeriksaan Struktur

Pada dataset Transaksi Penjualan memiliki 4 kolom data yaitu No (nomor urutan data), Kode (kode transaksi), Tanggal (tanggal terjadinya transaksi), Urutan Produk (daftar kode produk yang dibeli) sedangkan struktur data barang terdiri dari No (nomor urutan data), Kode (kode barang), Nama Barang (nama barang yang terdaftar di Citra Mart Amikom).

2.3 Data Cleaning

Tahapan Cleaning ini memeriksa kelengkapan atribut, konsistensi pemisah produk, format kode produk, nilai kosong, dan kemungkinan duplikasi. Kolom kosong yang tidak memiliki informasi dihapus karena tidak memberikan kontribusi terhadap pembentukan model. Normalisasi juga dilakukan pada inkonsistensi penulisan yang dapat mengubah pembacaan item, misalnya pemisah titik pada transaksi T052 ('B187. B188') dinormalisasi menjadi dua kode produk B187 dan B188, serta kode '73' dinormalisasi menjadi B73 agar konsisten dengan data master. Kolom yang tidak berkontribusi terhadap pembentukan model tidak digunakan. Setelah pembersihan, setiap transaksi direpresentasikan sebagai himpunan kode produk yang unik.

2.4 Transformasi

Setelah proses pembersihan, daftar produk pada setiap transaksi diubah menjadi matriks transaksi-item biner. Baris merepresentasikan transaksi, sedangkan kolom merepresentasikan produk. Nilai 1 menunjukkan bahwa produk terdapat dalam transaksi, sedangkan nilai 0 menunjukkan produk tidak terdapat dalam transaksi.

2.5 Pembentukan Vector

Pada tahap ini melakukan representasi setiap produk ke dalam bentuk vektor yang mana vektor ini menunjukkan pola kemunculannya pada seluruh transaksi. Sebagai contoh, produk (A) dan (B) dapat direpresentasikan menjadi:

$$\begin{aligned} [A] &= [1, 1, 0, 1, 0] \\ [B] &= [1, 1, 0, 0, 1] \end{aligned}$$

2.6 Perhitungan Cosine Similarity

Tahapan ini melakukan perhitungan nilai kemiripan antar produk dengan menggunakan persamaan Cosine Similarity(1):

$$\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)$$

Keterangan persamaan (1):

A = vektor pertama

B = vektor kedua

A_i = nilai elemen ke- i pada vektor A

B_i = nilai elemen ke- i pada vektor B

n = jumlah fitur/dimensi

$A \cdot B$ = **dot product** antara vektor A dan B

$\|A\|$ = panjang atau magnitude vektor A

$\|B\|$ = panjang atau magnitude vektor B

2.7 Pemeringkatan Produk

Setelah matriks kemiripan diperoleh, sistem membentuk rekomendasi Top-N. Untuk query yang hanya berisi satu item q , kandidat i diurutkan berdasarkan $\text{Sim}(q, i)$. Untuk keranjang yang berisi lebih dari satu item, $C = \{q_1, q_2, \dots, q_m\}$, skor setiap kandidat j dihitung dengan strategi maximum aggregation, yaitu mengambil nilai kemiripan tertinggi antara j dan seluruh item di dalam keranjang. Kandidat yang telah terdapat di dalam C dikeluarkan, kemudian kandidat lainnya diurutkan menurun berdasarkan skor dan N kandidat teratas dikembalikan sebagai rekomendasi. Adapun formula rekomendasi Top-N dapat dilihat pada persamaan (2):

$$R_N(q) = \text{Top}_{N_{i \in I, i \neq q}(\text{sim}(q, i))} \quad (2)$$

Keterangan persamaan(2):

$R_N(q)$ = himpunan N item rekomendasi teratas untuk item/query q

q = item yang digunakan sebagai acuan

i = kandidat item yang akan direkomendasikan

I = seluruh kumpulan item

N = jumlah rekomendasi yang ditampilkan, misalnya Top-5 atau Top-10

$\text{Sim}(q, i)$ = nilai Cosine Similarity antara item q dan kandidat i

2.8 Pengujian

Evaluasi dilakukan menggunakan protokol dua tingkat leave-one-transaction-out dan leave-one-item-out. Pertama, dipilih satu transaksi uji t yang memiliki sedikitnya dua produk. Transaksi t dikeluarkan seluruhnya dari data pelatihan; matriks transaksi-item, frekuensi item, co-occurrence, dan matriks similarity kemudian dihitung ulang hanya dari 76 transaksi lainnya. Dengan demikian, tidak ada informasi dari transaksi uji yang digunakan dalam pembentukan skor similarity. Kedua, untuk setiap item y di dalam t , item y ditahan sebagai target, sedangkan $C = t \setminus \{y\}$ menjadi keranjang query. Untuk setiap kandidat j yang tidak berada dalam C , skor rekomendasi dihitung sebagai $\text{Score}(j|C) = \max_{i \in C} \text{Sim}(i, j)$. Kandidat diurutkan berdasarkan skor tersebut dan dievaluasi pada $K=1, 3, 5$, dan 10 . Prosedur ini diulang hingga setiap item pada setiap transaksi yang memenuhi syarat pernah menjadi target. Dari 77 transaksi, 71 transaksi memiliki sedikitnya dua produk dan menghasilkan 368 kasus uji.

2.9 Analisis dan Interpretasi

Analisis mencakup pola co-occurrence, nilai similarity, pemeringkatan Top-N, serta performa offline menggunakan Hit Rate/Recall@K, Precision@K, dan Mean Reciprocal Rank (MRR). Karena setiap kasus hanya memiliki satu target relevan, Recall@K identik dengan Hit Rate@K, sedangkan Precision@K adalah $1/K$ ketika target ditemukan dan 0 jika tidak ditemukan. Untuk menilai apakah Cosine Similarity memberikan keuntungan terhadap pendekatan sederhana, protokol pengujian yang sama juga diterapkan pada empat pembandingan: Most-Popular, raw co-occurrence, Jaccard Similarity, dan Lift. Most-Popular mengurutkan kandidat berdasarkan frekuensi kemunculan pada data pelatihan; raw co-occurrence menggunakan jumlah kemunculan bersama; Jaccard menormalisasi irisan terhadap gabungan transaksi; dan Lift menilai kekuatan asosiasi relatif terhadap frekuensi marginal item. Popularity digunakan sebagai baseline referensi yang lazim dalam evaluasi recommender system [19].

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Dataset barang dapat dilihat pada Tabel 1 yang terdiri dari 296 barang. Jumlah 296 barang belum mencakup keseluruhan barang yang ada di Citra Mart Amikom tetapi sudah dapat digunakan untuk menghasilkan eksperimen pada penelitian ini.

Tabel 1. Data Barang

No	Kode	Nama Barang
1	B01	Mony Cincau Can 30
2	B02	Energen Coklat 30
3	B03	Energen Jagung 30
4	B04	Taro Italian Pizza
5	B05	Taro Seaweed 10GR
...	...	...
294	B294	Hilo Choc Avoc PET
295	B295	Grd Pilus Mi GRG 9
296	B296	Milo 3 In 1 1KG

Pada Tabel 2 menampilkan contoh dari 77 transaksi yang tersedia pada periode pengamatan. Dataset ini bukan keseluruhan transaksi Citra Mart Amikom dan mekanisme pemilihannya dari populasi transaksi tidak terdokumentasi secara rinci. Karena itu, hasil penelitian diinterpretasikan sebagai eksperimen baseline pada sampel transaksi yang tersedia, bukan sebagai estimasi performa yang dapat langsung digeneralisasikan ke seluruh transaksi ritel.

Tabel 2. Data Transaksi

No	Kode	Tanggal	Urutan Produk
1	T001	29/3/2019	B278, B279, B280, B281, B282, B191
2	T002	29/3/2019	B283, B284, B285, B148, B221, B149, B151
3	T003	29/3/2019	B286, B287, B139, B239, B196
4	T004	29/3/2019	B288, B289
5	T005	28/3/2019	B290, B291
...	...	...	...
75	T075	07/5/2019	B43
76	T076	07/5/2019	B123, B124, B125, B126, B127, B128, B129, B130, B131
77	T077	07/1/2019	B132, B133, B134, B135

3.2 Transformasi Data Transaksi

Atribut tanggal tidak digunakan sebagai fitur utama dalam perhitungan kemiripan karena tujuan penelitian adalah mengukur hubungan antarproduk berdasarkan kemunculan dalam transaksi. Setelah proses pembersihan data dilakukan kemudian daftar produk pada setiap transaksi diubah menjadi matriks transaksi-item biner. Baris merepresentasikan transaksi, sedangkan kolom merepresentasikan produk. Nilai 1 menunjukkan bahwa produk terdapat dalam transaksi, sedangkan nilai 0 menunjukkan produk tidak terdapat dalam transaksi. Pada Tabel 3 dapat dilihat hasil ilustrasi transformasi data.

Tabel 3. Ilustrasi Matriks Transaksi-item

Transaksi	Kode					
	B148	B149	B151	B221	B278	B279
T001	0	0	0	1	1	1
T002	1	1	1	0	0	0
T003	0	0	0	0	0	0
T004	0	0	0	0	0	0
T005	0	0	0	0	0	0

Tabel 3 hanya merupakan ilustrasi dari matriks yang akan terbentuk dari pemetaan data transaksi dan data barang yang mengalami pembelian ataupun yang tidak mengalami pembelian. Pada implementasi nya, matriks yang terbentuk berukuran 77 baris dan 296 kolom produk. Setelah matriks terbentuk maka selanjutnya untuk menghitung kemiripan antar produk, matriks transaksi-item ditransposisikan menjadi matriks item-transaksi berukuran 296.

3.3 Pembentukan Vektor Item

Setiap produk direpresentasikan dalam bentuk vektor yang menunjukkan pola kemunculannya pada seluruh transaksi. Sebagai contoh, produk (A) dan (B) dapat direpresentasikan sebagai:

$$[A = [1, 1, 0, 1, 0]]$$

$$[B = [1, 1, 0, 0, 1]]$$

Nilai 1 menunjukkan bahwa produk muncul dalam transaksi tertentu. Kemiripan antara dua produk ditentukan berdasarkan kesamaan arah kedua vektor tersebut. Pendekatan ini memungkinkan produk dengan kategori berbeda tetap memperoleh nilai kemiripan tinggi apabila sering muncul pada transaksi yang sama. Dengan demikian, sistem tidak bergantung pada kesamaan nama, deskripsi, atau kategori produk.

3.4 Perhitungan Cosine Similarity

Nilai kemiripan berada pada rentang 0 sampai 1. Nilai mendekati 1 menunjukkan pola kemunculan transaksi yang sangat mirip, sedangkan nilai 0 menunjukkan tidak adanya co-occurrence. Berdasarkan dataset, B148 (Lays RPT Laut 18GR) muncul pada 4 transaksi dan B151 (BENG2 Choco 20GR) muncul pada 5 transaksi. Kedua produk muncul bersama pada 4 transaksi. Karena vektor bersifat biner, dot product

$$B148 \cdot B151 = 4, \|B148\| = \sqrt{4} = 2, \text{ dan } \|B151\| = \sqrt{5} = 2,236. \text{ Dengan demikian, maka:}$$

$$\text{Sim}(B148, B151) = 4 / (\sqrt{4} \times \sqrt{5}) = 4 / \sqrt{20} = 0.894.$$

3.5 Rekomendasi Top-N

Setelah matriks kemiripan diperoleh, setiap produk dipasangkan dengan produk lain dan diurutkan berdasarkan nilai kemiripan (similarity) tertinggi. Penelitian menggunakan Top-10 sebagai keluaran utama karena memberikan cakupan kandidat lebih luas, sedangkan K = 1, 3, 5, dan 10 digunakan pada evaluasi untuk melihat perubahan performa terhadap panjang daftar rekomendasi. Untuk contoh query B148, rekomendasi teratas ditunjukkan pada Tabel 4.

Tabel 4. Top-10 Rekomendasi untuk Produk B148

Rank	Kode	Nama Produk	Cosine Similarity	Co-occurrence
1	B151	BENG2 Choco 20GR	0,894	4
2	B146	Lays RPT Laut 40GR	0,866	3
3	B221	Gery Choccolts Dark	0,866	3
4	B149	Gery Choccolts GRAN	0,707	2
5	B152	CHOKI2 ALL var 11GR	0,707	2
6	B220	Mikako Mie GRG 24G	0,707	2
7	B150	Gery Choccolts 14.5	0,577	2
8	B147	GRD KCG KLT 75GR	0,500	1
9	B154	Roma SR Gandum COK	0,500	1
10	B219	Oreo Peanut&CHOC 2	0,500	1

Apabila konsumen membeli lebih dari satu produk, skor kandidat j dihitung menggunakan maximum aggregation, $\text{Score}(j|C) = \max_{i \in C} \text{Sim}(i, j)$, dengan C sebagai himpunan item yang tetap berada di dalam keranjang setelah satu item ditahan sebagai target. Sebagai contoh, apabila $C = \{B148, B149\}$, maka skor kandidat B151 adalah nilai maksimum antara $\text{Sim}(B148, B151)$ dan $\text{Sim}(B149, B151)$. Strategi ini memastikan kandidat memperoleh skor dari hubungan terkuatnya dengan salah satu item yang sudah terdapat di keranjang. Produk yang telah berada di dalam C dikeluarkan dari daftar kandidat sebelum Top-N dibentuk. $\text{Score}(j|C) = \max_{i \in C} \text{Sim}(i, j)$

3.6 Hasil Eksperimen

Eksperimen dilakukan pada 71 dari 77 transaksi yang memiliki minimal dua produk. Setiap transaksi uji dikeluarkan seluruhnya sebelum matriks similarity dihitung ulang (leave-one-transaction-out), kemudian setiap item di dalam transaksi tersebut secara bergantian ditahan sebagai target (leave-one-item-out), sehingga diperoleh 368 kasus uji. Item lain pada transaksi menjadi keranjang query dan skor kandidat digabungkan menggunakan maximum aggregation sebagaimana dijelaskan pada Bagian 3.7 dan 3.8. Tabel 5 menunjukkan performa Cosine Similarity pada beberapa nilai K. Karena setiap kasus hanya memiliki satu target, Recall@K sama dengan Hit Rate@K dan Precision@K sama dengan Hit Rate@K/K

Tabel 5. Hasil Evaluasi Top-K

K	Hit Rate/Recall @K	Precision @K
1	1,36 %	1,36 %
3	2,17 %	0,72 %
5	4,08 %	0,82 %
10	8,15 %	0,82 %

Cosine Similarity memperoleh MRR 0,040. Hit Rate meningkat dari 1,36% pada Top-1 menjadi 8,15% pada Top-10, tetapi tingkat keberhasilan absolut masih rendah. Kondisi ini berkaitan dengan sparsity matriks sebesar 98,36% dan 241 dari 296 produk yang hanya muncul satu kali. Ketika item berfrekuensi sangat rendah terdapat pada transaksi yang sedang diuji, pengeluaran transaksi tersebut dari data pelatihan menghilangkan sebagian atau seluruh informasi co-occurrence yang diperlukan untuk menemukan kembali target. Dengan demikian, hasil ini menunjukkan bahwa estimasi kemiripan pada dataset sangat sparse tidak cukup stabil hanya dengan menambah panjang daftar rekomendasi.

Untuk memberikan konteks terhadap nilai tersebut, Cosine Similarity dibandingkan dengan Most-Popular, raw co-occurrence, Jaccard Similarity, dan Lift menggunakan 368 kasus uji dan protokol evaluasi yang identik. Hasil perbandingan pada Tabel 6 menunjukkan bahwa Cosine mengungguli Most-Popular dan Lift pada Hit Rate@10, tetapi tidak menjadi metode terbaik. Jaccard mencapai Hit Rate@10 sebesar 8,42%, sedikit lebih tinggi daripada Cosine 8,15%, sedangkan raw co-occurrence menghasilkan Hit Rate@10 tertinggi sebesar 8,70% dan MRR 0,060. Temuan ini penting karena menunjukkan bahwa pada dataset yang sangat jarang, normalisasi Cosine belum memberikan keuntungan akurasi dibanding hitungan co-occurrence sederhana. Oleh sebab itu, kontribusi Cosine dalam penelitian ini lebih tepat dipahami sebagai baseline yang sederhana, terinterpretasi, dan mudah diimplementasikan, bukan sebagai metode yang terbukti unggul terhadap seluruh pembanding.

Tabel 6. Perbandingan Metode pada Protokol Evaluasi yang Sama

Metode	HR@1	HR@3	HR@5	HR@10	MRR
Most-Popular	0,00 %	2,17 %	2,17 %	4,08 %	0,021 %
Raw co-occurrence	3,26 %	5,43 %	5,43 %	8,70 %	0,060 %
Jaccard Similarity	1,36 %	2,17 %	2,17 %	8,42 %	0,040 %
Lift	1,36 %	1,90 %	1,90 %	7,07 %	0,037 %
Cosine Similarity	1,36 %	2,17 %	2,17 %	8,15 %	0,040 %

Secara keseluruhan, seluruh metode menghasilkan performa yang rendah, yang memperkuat indikasi bahwa keterbatasan utama berasal dari kepadatan data, bukan semata-mata dari pemilihan similarity measure. Popularity baseline tetap penting sebagai titik acuan karena metode tersebut merekomendasikan item berdasarkan frekuensi interaksi pada data pelatihan. Perbandingan ini juga memperlihatkan bahwa raw co-occurrence dapat menjadi baseline yang kompetitif pada data keranjang berukuran kecil. Keterbatasan penelitian mencakup jumlah transaksi yang sedikit, mekanisme sampling yang tidak terdokumentasi dan belum adanya validasi pada periode atau toko lain.

4. KESIMPULAN

Penelitian ini membangun baseline rekomendasi produk berbasis Item Similarity menggunakan Cosine Similarity pada 77 transaksi dan 296 produk. Data ditransformasikan menjadi matriks transaksi-item biner sehingga hubungan antarproduk dapat dihitung dari pola co-occurrence tanpa identitas pelanggan atau rating. Evaluasi dua tingkat leave-one-transaction-out dan leave-one-item-out menghasilkan 368 kasus uji; Cosine Similarity memperoleh Hit Rate/Recall@10 8,15%, Precision@10 0,82%, dan MRR 0,040. Perbandingan dengan metode sederhana menunjukkan bahwa Cosine lebih baik daripada Most-Popular dan Lift pada Hit Rate@10, tetapi sedikit di bawah Jaccard (8,42%) dan raw co-occurrence (8,70%); raw co-occurrence juga memperoleh MRR tertinggi sebesar 0,060. Temuan tersebut menegaskan bahwa pada dataset dengan sparsity 98,36% dan dominasi item berfrekuensi satu kali, kualitas rekomendasi lebih dibatasi oleh ketersediaan

interaksi daripada oleh satu ukuran similarity tertentu. Karena penelitian ini hanya melakukan evaluasi offline dan belum mengukur perubahan nilai transaksi, conversion rate, atau penjualan aktual, model tidak dapat disimpulkan telah meningkatkan cross-selling. Hasilnya menunjukkan potensi implementasi sebagai baseline untuk mendukung eksperimen cross-selling selanjutnya. Penelitian berikutnya perlu menggunakan data transaksi yang lebih besar dan representatif, membandingkan strategi agregasi dan similarity secara lebih luas, serta melakukan A/B test atau evaluasi online untuk mengukur dampak rekomendasi terhadap cross-selling secara nyata.

DAFTAR PUSTAKA

- [1] F. Pratama, S. Prastio, E. Seniwati, N. T. Hartanti, dan Sudarmanto, "Content-Based Filtering untuk Sistem Rekomendasi Produk E-Commerce," *The Indonesian Journal of Computer Science Research (IJCSR)*, vol. 5, no. 1, pp. 86–98, 2026, doi: 10.59095/ijcsr.v5i1.254.
- [2] Y. S. Kim, H. Hwangbo, H. J. Lee, *et al.*, "Sequence Aware Recommenders for Fashion E-Commerce," *Electronic Commerce Research*, vol. 24, pp. 2733–2753, 2024, doi: 10.1007/s10660-022-09627-8.
- [3] A. Bellogín, I. Cantador, *et al.*, "Similarity Measures for Collaborative Filtering-Based Recommender Systems: Review and Experimental Comparison," *Journal of King Saud University – Computer and Information Sciences*, vol. 34, no. 9, pp. 7645–7669, 2022.
- [4] H. I. Abdalla, A. A. Amer, Y. A. Amer, L. Nguyen, *et al.*, "Boosting the Item-Based Collaborative Filtering Model with Novel Similarity Measures," *International Journal of Computational Intelligence Systems*, vol. 16, Art. 123, 2023, doi: 10.1007/s44196-023-00299-2.
- [5] Hartatik, R. R. Isnanto, B. Warsito, N. Firdaus, and F. Y. A'La, "Towards the Application of Artificial Intelligence: Cosine Similarity in Recommendation Systems Based on Collaborative Filtering," in *Proceedings of the 7th International Conference of Computer and Informatics Engineering (IC2IE 2024)*, IEEE, 2024, doi: 10.1109/IC2IE63342.2024.10747946.
- [1] Y. Wang, J. Wu, Z. Wu, dan G. Adomavicius, "Efficient and Flexible Long-Tail Recommendation Using Cosine Patterns," *INFORMS Journal on Computing*, vol. 37, no. 2, pp. 446–464, 2024, doi: 10.1287/ijoc.2022.0194.
- [2] A. P. Putra, D. P. S. Putri, dan A. A. K. A. Wiranatha, "Scientific Paper Recommendation System: Application of Sentence Transformers and Cosine Similarity Using arXiv Data," *Journal of Applied Informatics and Computing*, vol. 9, no. 4, 2025.
- [3] S. I. Adam dan W. G. Mokodaser, "Implementasi Sistem Rekomendasi Produk E-Commerce Menggunakan Content-Based Filtering Berbasis Cosine Similarity," *SIMTEK: Jurnal Sistem Informasi dan Teknik Komputer*, vol. 10, no. 2, pp. 418–424, 2025.
- [4] E. W. Pratomo dan E. Utami, "Analisis dan Implementasi Content-Based Filtering dengan Cosine Similarity untuk Sistem Rekomendasi Tugas Akhir Mahasiswa," *AITI*, vol. 23, no. 2, 2026.
- [5] A. Alsobhi dan N. Amare, "Ontology-Based Relational Product Recommendation System," *Computational Intelligence and Neuroscience*, 2022, Article ID 1591044.
- [6] M. Li, S. Jullien, M. Ariannezhad, and M. de Rijke, "A Next Basket Recommendation Reality Check," *ACM Transactions on Information Systems*, vol. 41, no. 4, Art. 116, 2023, doi: 10.1145/3587153.
- [7] A. Romanov, O. Lashinin, M. Ananyeva, and S. Kolesnikov, "Time-Aware Item Weighting for the Next Basket Recommendations," in *Proceedings of the 17th ACM Conference on Recommender Systems (RecSys '23)*, 2023, doi: 10.1145/3604915.3608859.
- [8] R. Esmeli, M. Bader-El-Den, H. Abdullahi, *et al.*, "Implicit Feedback Awareness for Session Based Recommendation in E-Commerce," *SN Computer Science*, vol. 4, Art. 320, 2023, doi: 10.1007/s42979-023-01752-x.
- [9] A. S. G. *et al.*, "Session-Based Recommendations for e-Commerce with Graph-Based Data Modeling," *Applied Sciences*, vol. 13, no. 1, Art. 394, 2023.
- [10] P. Kumar, M. K. Gupta, C. R. S. Rao, M. Bhavsingh, and M. Srilakshmi, "A Comparative Analysis of Collaborative Filtering Similarity Measurements for Recommendation Systems," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 3s, pp. 184–192, 2023, doi: 10.17762/ijritcc.v11i3s.6180.
- [11] E. Zangerle and C. Bauer, "Evaluating Recommender Systems: Survey and Framework," *ACM Computing Surveys*, vol. 55, no. 8, Art. 170, pp. 1–38, 2022, doi: 10.1145/3556536.
- [12] P. Castells and A. Moffat, "Offline Recommender System Evaluation: Challenges and New Directions," *AI Magazine*, vol. 43, no. 2, pp. 225–238, 2022, doi: 10.1002/aaai.12051.
- [13] X. Wu, "Review of Collaborative Filtering Recommendation Systems," *Applied and Computational Engineering*, vol. 43, pp. 76–82, 2024, doi: 10.54254/2755-2721/43/20230811.
- [14] Y. Ji, A. Sun, J. Zhang, and C. Li, "A Re-visit of the Popularity Baseline in Recommender Systems," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval (SIGIR '20)*, 2020, pp. 1749–1752, doi: 10.1145/3397271.3401233.