

SKRIPSI

**PENERAPAN METODE *COLLABORATIVE
FILTERING*
UNTUK SISTEM REKOMENDASI PRODUK
PADA PLATFORM E-DAGANG NIMBUS**

Diajukan sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer

RANGGA ADITYA NUGROHO
NPM 2006589471

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS INDONESIA
DEPOK
JULI 2026**

*Unofficial template — not affiliated with or endorsed by the university.
Sample content only.*

HALAMAN PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar.

Depok, 14 Juli 2026

Rangga Aditya Nugroho
NPM 2006589471

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Rangga Aditya Nugroho
NPM : 2006589471
Program Studi : Ilmu Komputer
Judul Skripsi : Penerapan Metode *Collaborative Filtering* untuk Sistem Rekomendasi Produk pada Platform E-Dagang Nimbus

telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Indonesia.

DEWAN PENGUJI

Jabatan	Nama	Tanda Tangan
Pembimbing	Dr. Made Wirawan, S.Kom., M.T.	_____
Ketua Penguji	Prof. Dr. Ratna Kusuma Dewi, M.Sc.	_____
Anggota Penguji	Ir. Bayu Setiawan, Ph.D.	_____

Ditetapkan di: Depok
Tanggal: 14 Juli 2026

KATA PENGANTAR

Puji syukur penulis panjatkan atas selesainya penyusunan skripsi ini yang berjudul “Penerapan Metode *Collaborative Filtering* untuk Sistem Rekomendasi Produk pada Platform E-Dagang Nimbus”. Penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer pada Program Studi Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Indonesia.

Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak, sangat sulit bagi penulis untuk menyelesaikan skripsi ini. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Dr. Made Wirawan, S.Kom., M.T., selaku dosen pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan penulis dalam penyusunan skripsi ini;
2. Tim riset data di PT Nexa Digital Nusantara yang telah memberikan akses terhadap data transaksi tersamar (*anonymized*) selama program kerja praktik penulis;
3. Orang tua dan keluarga penulis yang senantiasa memberikan dukungan moral, material, dan doa yang tiada henti;
4. Sahabat seperjuangan di Laboratorium Ilmu Data Fasilkom UI, terutama Dimas Prakoso dan Aulia Rahmawati, atas diskusi dan masukan yang berharga.

Akhir kata, penulis berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga skripsi ini membawa manfaat bagi pengembangan ilmu pengetahuan.

Depok, 14 Juli 2026

Penulis

ABSTRAK

Nama : Rangga Aditya Nugroho
Program Studi : Ilmu Komputer
Judul : Penerapan Metode *Collaborative Filtering* untuk Sistem Rekomendasi Produk pada Platform E-Dagang Nimbus

Pertumbuhan katalog produk pada platform e-dagang Nimbus yang melampaui 40.000 SKU aktif menyebabkan pengguna kesulitan menemukan produk yang relevan dengan preferensinya. Penelitian ini mengusulkan sistem rekomendasi berbasis *collaborative filtering* dengan pendekatan *matrix factorization* (*Alternating Least Squares*) yang dibandingkan dengan *user-based* dan *item-based nearest neighbor* sebagai baseline. Data yang digunakan berupa 1.248.302 catatan interaksi implisit (klik, tambah keranjang, dan pembelian) dari 58.114 pengguna terhadap 41.276 produk selama periode Januari–Desember 2025. Evaluasi dilakukan menggunakan metrik *Root Mean Square Error* (RMSE), *Precision@10*, dan *Normalized Discounted Cumulative Gain* (NDCG@10). Hasil eksperimen menunjukkan bahwa model ALS dengan 64 faktor laten memberikan performa terbaik dengan RMSE sebesar 0,742, *Precision@10* sebesar 0,381, dan NDCG@10 sebesar 0,447, mengungguli baseline *item-based* sebesar 21,4% pada metrik NDCG@10. Penelitian ini juga menemukan bahwa penambahan mekanisme peluruhan waktu (*time decay*) pada bobot interaksi meningkatkan relevansi rekomendasi untuk pengguna dengan riwayat interaksi pendek.

Kata kunci: sistem rekomendasi, *collaborative filtering*, *matrix factorization*, e-dagang, *implicit feedback*

ABSTRACT

Name : Rangga Aditya Nugroho
Study Program : Computer Science
Title : Application of Collaborative Filtering for a Product Recommendation System on the Nimbus E-Commerce Platform

The rapid growth of the Nimbus e-commerce catalog, now exceeding 40,000 active SKUs, makes it increasingly difficult for users to discover products relevant to their preferences. This research proposes a collaborative filtering recommendation system based on matrix factorization (Alternating Least Squares), benchmarked against user-based and item-based nearest neighbor baselines. The dataset consists of 1,248,302 implicit interaction records (clicks, cart additions, and purchases) from 58,114 users across 41,276 products collected between January and December 2025. Evaluation used Root Mean Square Error (RMSE), Precision@10, and Normalized Discounted Cumulative Gain (NDCG@10). Experimental results show that the ALS model with 64 latent factors achieves the best performance, with an RMSE of 0.742, a Precision@10 of 0.381, and an NDCG@10 of 0.447, outperforming the item-based baseline by 21.4% on NDCG@10. The study further finds that incorporating a time-decay weighting scheme on interactions improves recommendation relevance for users with short interaction histories.

Keywords: recommendation system, collaborative filtering, matrix factorization, e-commerce, implicit feedback

Contents

I	PENDAHULUAN	1
I.1	Latar Belakang	1
I.2	Rumusan Masalah	1
I.3	Tujuan Penelitian	2
I.4	Manfaat Penelitian	2
I.5	Batasan Masalah	2
I.6	Sistematika Penulisan	2
II	TINJAUAN PUSTAKA	4
II.1	E-Dagang dan Personalisasi	4
II.2	Sistem Rekomendasi	4
II.3	Collaborative Filtering	4
II.4	Penelitian Terkait	5
II.5	Kerangka Pemikiran	5
III	METODE PENELITIAN	7
III.1	Desain Penelitian	7
III.2	Data dan Sumber Data	7
III.3	Tahapan Penelitian	7
III.4	Metode Evaluasi	8
IV	HASIL DAN PEMBAHASAN	9
IV.1	Karakteristik Hasil Praproses	9
IV.2	Hasil Eksperimen	9
IV.3	Perbandingan dengan Baseline	9
IV.4	Analisis Pengaruh Peluruhan Waktu	10
V	KESIMPULAN DAN SARAN	11
V.1	Kesimpulan	11
V.2	Saran	11

List of Figures

II.1	Kerangka pemikiran penelitian	6
III.1	Tahapan penelitian	8

List of Tables

II.1	Perbandingan pendekatan sistem rekomendasi	4
II.2	Ringkasan penelitian terkait	5
III.1	Karakteristik data penelitian	7
IV.1	Hasil evaluasi model ALS pada berbagai jumlah faktor laten	9
IV.2	Perbandingan model ALS dengan baseline	10

BAB I

PENDAHULUAN

I.1 Latar Belakang

Platform e-dagang Nimbus telah berkembang dari sebuah toko daring niche menjadi salah satu penyedia belanja daring dengan lebih dari 40.000 SKU aktif dan rata-rata 58.000 pengguna aktif bulanan pada akhir tahun 2025. Pertumbuhan katalog yang pesat ini menimbulkan fenomena *choice overload*, yaitu kondisi di mana pengguna kesulitan menentukan pilihan karena terlalu banyak alternatif yang tersedia (Kusuma & Prabowo, 2022). Tanpa mekanisme personalisasi yang memadai, tingkat konversi cenderung menurun dan waktu penelusuran pengguna meningkat secara signifikan.

Sistem rekomendasi merupakan salah satu solusi yang paling banyak diadopsi industri e-dagang untuk mengatasi permasalahan tersebut. Di antara berbagai pendekatan yang tersedia, *collaborative filtering* (CF) menjadi pendekatan yang populer karena tidak memerlukan atribut konten produk yang lengkap dan mampu menangkap pola preferensi laten dari interaksi pengguna secara kolektif (Hardjono, Wibisono, & Santoso, 2021). Namun demikian, penerapan CF pada data interaksi implisit — seperti klik dan tambah keranjang, alih-alih rating eksplisit — menghadirkan tantangan tersendiri karena tidak adanya sinyal preferensi negatif yang eksplisit.

I.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah pada penelitian ini adalah sebagai berikut.

1. Bagaimana merancang model *collaborative filtering* berbasis *matrix factorization* yang sesuai untuk data interaksi implisit pada platform Nimbus?

2. Bagaimana performa model yang diusulkan dibandingkan dengan pendekatan *neighborhood-based* sebagai baseline?
3. Bagaimana pengaruh mekanisme peluruhan waktu terhadap kualitas rekomendasi bagi pengguna dengan riwayat interaksi pendek?

I.3 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. merancang dan mengimplementasikan model rekomendasi berbasis *Alternating Least Squares* (ALS) untuk data implisit;
2. mengevaluasi dan membandingkan performa model yang diusulkan terhadap baseline *user-based* dan *item-based nearest neighbor*; dan
3. menganalisis pengaruh peluruhan waktu terhadap relevansi rekomendasi pada segmen pengguna *cold-start*.

I.4 Manfaat Penelitian

Secara praktis, hasil penelitian ini diharapkan dapat menjadi rujukan bagi tim produk Nimbus dalam mengembangkan modul rekomendasi produksi. Secara akademis, penelitian ini memperkaya kajian penerapan *collaborative filtering* pada konteks data implisit berskala menengah di ekosistem e-dagang Indonesia.

I.5 Batasan Masalah

Penelitian ini dibatasi pada beberapa hal berikut.

- Data yang digunakan adalah data interaksi tersamar (*anonymized*) periode Januari–Desember 2025.
- Model yang dibandingkan terbatas pada *matrix factorization* (ALS) dan *neighborhood-based collaborative filtering*; metode berbasis *deep learning* berada di luar cakupan penelitian ini.
- Evaluasi dilakukan secara *offline* menggunakan data historis, tanpa pengujian A/B pada pengguna produksi.

I.6 Sistematika Penulisan

Skripsi ini disusun dalam lima bab. Bab I menjelaskan latar belakang dan rumusan masalah. Bab II membahas tinjauan pustaka terkait sistem rekomendasi dan *col-*

laborative filtering. Bab III menguraikan metode penelitian. Bab IV memaparkan hasil eksperimen dan pembahasan. Bab V menyajikan kesimpulan dan saran.

BAB II

TINJAUAN PUSTAKA

II.1 E-Dagang dan Personalisasi

Personalisasi dalam konteks e-dagang merujuk pada penyesuaian tampilan maupun konten yang disajikan kepada pengguna berdasarkan riwayat perilaku mereka (Santoso, 2020). Studi Hardjono dkk. (2021) menunjukkan bahwa platform yang menerapkan personalisasi berbasis rekomendasi mengalami peningkatan rata-rata nilai transaksi sebesar 12–18% dibandingkan platform tanpa personalisasi.

II.2 Sistem Rekomendasi

Secara umum, sistem rekomendasi dapat dikategorikan menjadi tiga pendekatan utama, sebagaimana dirangkum pada Tabel II.1.

Table II.1: Perbandingan pendekatan sistem rekomendasi

Pendekatan	Karakteristik Utama
<i>Content-based filtering</i>	Merekomendasikan item dengan atribut serupa dengan item yang disukai pengguna sebelumnya.
<i>Collaborative filtering</i>	Memanfaatkan pola interaksi kolektif antarpengguna dan item tanpa memerlukan atribut konten.
<i>Hybrid</i>	Menggabungkan sinyal konten dan kolaboratif untuk mengurangi kelemahan masing-masing pendekatan.

II.3 Collaborative Filtering

Collaborative filtering dibagi menjadi dua kelompok besar: *neighborhood-based* dan *model-based*. Pendekatan *neighborhood-based* menghitung kemiripan antarpengguna atau antaritem menggunakan metrik seperti *cosine similarity*. Pendekatan *model-based*, khususnya *matrix factorization*, memetakan pengguna

dan item ke dalam ruang faktor laten berdimensi k sedemikian rupa sehingga preferensi $\hat{r}_{ui}$ diaproksimasi sebagai:

$$\hat{r}_{ui} = \mathbf{p}_u^\top \mathbf{q}_i \quad (\text{II.1})$$

dengan $\mathbf{p}_u \in \mathbb{R}^k$ merupakan vektor faktor laten pengguna u dan $\mathbf{q}_i \in \mathbb{R}^k$ merupakan vektor faktor laten item i (Hu, Koren, & Volinsky, 2008). Untuk data *implicit feedback*, fungsi kerugian yang diminimalkan mengikuti formulasi berbobot:

$$\min_{\mathbf{p}_*, \mathbf{q}_*} \sum_{u,i} c_{ui} \left(p_{ui} - \mathbf{p}_u^\top \mathbf{q}_i \right)^2 + \lambda \left(\sum_u \|\mathbf{p}_u\|^2 + \sum_i \|\mathbf{q}_i\|^2 \right) \quad (\text{II.2})$$

dengan c_{ui} merupakan bobot keyakinan (*confidence*) yang diturunkan dari frekuensi interaksi dan λ merupakan koefisien regularisasi.

II.4 Penelitian Terkait

Tabel II.2 merangkum beberapa penelitian terdahulu yang relevan dengan topik penelitian ini.

Table II.2: Ringkasan penelitian terkait

Peneliti (Tahun)	Metode	Temuan Utama
Hu, Koren, & Volinsky (2008)	ALS berbobot untuk data implisit	Memperkenalkan formulasi <i>confidence-weighted</i> untuk data implisit.
Wibisono & Ananta (2019)	<i>Item-based</i> kNN	Efektif untuk katalog kecil namun sulit <i>scale</i> pada jutaan item.
Hardjono dkk. (2021)	<i>Hybrid</i> CF + konten	Meningkatkan <i>recall</i> 9% dibanding CF murni pada domain fesyen.
Prasetyo & Ramadhani (2023)	<i>Time-aware</i> CF	Peluruhan waktu memperbaiki relevansi pada pengguna baru.

II.5 Kerangka Pemikiran

Gambar II.1 menggambarkan kerangka pemikiran penelitian ini, mulai dari data interaksi mentah hingga evaluasi model rekomendasi.

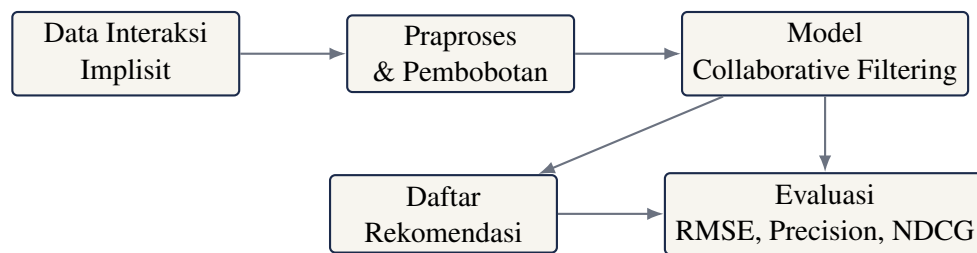


Figure II.1: Kerangka pemikiran penelitian

BAB III

METODE PENELITIAN

III.1 Desain Penelitian

Penelitian ini menggunakan pendekatan eksperimental kuantitatif dengan tahapan yang mengikuti kerangka *Cross-Industry Standard Process for Data Mining* (CRISP-DM), disesuaikan dengan konteks sistem rekomendasi.

III.2 Data dan Sumber Data

Data penelitian bersumber dari log interaksi tersamar platform Nimbus periode Januari–Desember 2025. Ringkasan karakteristik data disajikan pada Tabel III.1.

Table III.1: Karakteristik data penelitian

Karakteristik	Nilai
Jumlah pengguna unik	58.114
Jumlah produk unik	41.276
Jumlah catatan interaksi	1.248.302
Kepadatan matriks interaksi (<i>density</i>)	0,052%
Rentang waktu data	Jan 2025 – Des 2025

Setiap catatan interaksi diberi bobot berdasarkan jenis peristiwa: klik bernilai 1, tambah keranjang bernilai 3, dan pembelian bernilai 5, mengikuti skema pembobotan pada Prasetyo & Ramadhani (2023).

III.3 Tahapan Penelitian

Tahapan penelitian yang dilakukan ditunjukkan pada Gambar III.1.

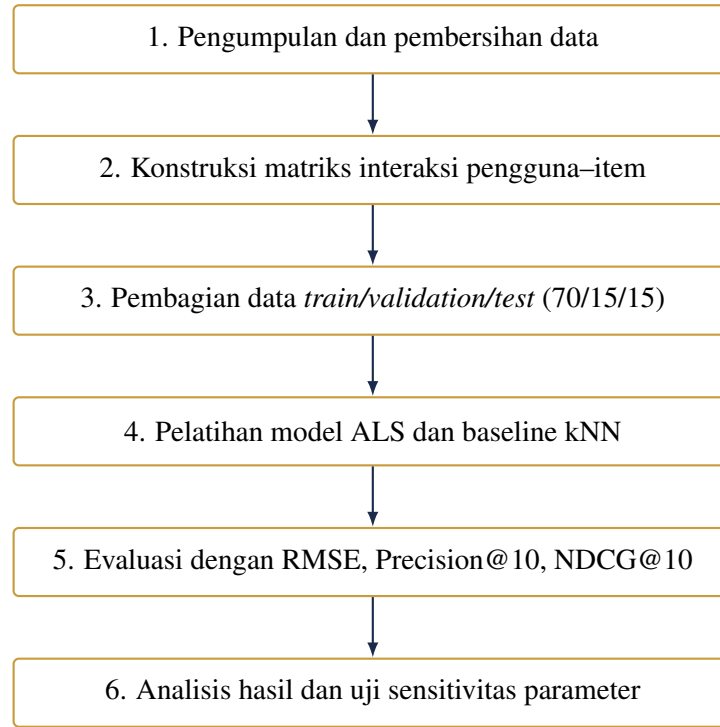


Figure III.1: Tahapan penelitian

III.4 Metode Evaluasi

Performa model dievaluasi menggunakan tiga metrik. *Root Mean Square Error* (RMSE) mengukur deviasi prediksi terhadap preferensi aktual:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{u,i} (r_{ui} - \hat{r}_{ui})^2} \quad (\text{III.1})$$

Precision@K mengukur proporsi item relevan pada K rekomendasi teratas, sedangkan *Normalized Discounted Cumulative Gain* (NDCG@K) memperhitungkan posisi item relevan dalam daftar rekomendasi:

$$\text{NDCG@K} = \frac{1}{\text{IDCG@K}} \sum_{j=1}^K \frac{2^{\text{rel}_j} - 1}{\log_2(j+1)} \quad (\text{III.2})$$

⚙️ Konfigurasi Eksperimen

Model ALS dilatih dengan jumlah faktor laten $k \in \{16, 32, 64, 128\}$, koefisien regularisasi $\lambda = 0, 1$, dan 20 iterasi. Pencarian hiperparameter dilakukan melalui *grid search* dengan validasi silang 5 lipatan pada himpunan validasi.

BAB IV

HASIL DAN PEMBAHASAN

IV.1 Karakteristik Hasil Praproses

Setelah penyaringan pengguna dan item dengan interaksi minimum (*k-core filtering* dengan ambang 5 interaksi), matriks interaksi akhir memiliki dimensi 52.847×38.910 dengan kepadatan 0,061%.

IV.2 Hasil Eksperimen

Tabel IV.1 menyajikan hasil evaluasi untuk setiap konfigurasi jumlah faktor laten pada model ALS.

Table IV.1: Hasil evaluasi model ALS pada berbagai jumlah faktor laten

Faktor Laten (k)	RMSE	Precision@10	NDCG@10
16	0,811	0,329	0,388
32	0,768	0,357	0,419
64	0,742	0,381	0,447
128	0,745	0,376	0,441

Konfigurasi dengan 64 faktor laten memberikan performa terbaik pada ketiga metrik. Peningkatan lebih lanjut ke 128 faktor tidak memberikan perbaikan performa dan justru mengindikasikan gejala *overfitting* ringan.

IV.3 Perbandingan dengan Baseline

Tabel IV.2 membandingkan model ALS terbaik ($k = 64$) terhadap dua baseline *neighborhood-based*.

Model ALS mengungguli baseline *item-based* kNN sebesar 21,4% pada metrik NDCG@10 dan 19,8% pada Precision@10, konsisten dengan temuan Hu, Koren,

Table IV.2: Perbandingan model ALS dengan baseline

Model	RMSE	Precision@10	NDCG@10
User-based kNN	0,893	0,301	0,352
Item-based kNN	0,856	0,318	0,368
ALS ($k = 64$)	0,742	0,381	0,447

& Volinsky (2008) bahwa *matrix factorization* lebih unggul menangkap pola laten pada data berdimensi tinggi dan jarang (*sparse*).

IV.4 Analisis Pengaruh Peluruhan Waktu

Pada segmen pengguna dengan kurang dari 5 interaksi historis (*cold-start*), penambahan mekanisme peluruhan waktu eksponensial pada bobot interaksi meningkatkan NDCG@10 dari 0,291 menjadi 0,334, atau peningkatan relatif sebesar 14,8%. Temuan ini memperkuat argumen Prasetyo & Ramadhani (2023) bahwa interaksi terbaru memberikan sinyal preferensi yang lebih andal dibanding interaksi lampau, khususnya bagi pengguna dengan riwayat pendek.

BAB V

KESIMPULAN DAN SARAN

V.1 Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan sebagai berikut.

1. Model *collaborative filtering* berbasis ALS dengan 64 faktor laten memberikan performa terbaik dengan RMSE 0,742, Precision@10 0,381, dan NDCG@10 0,447 pada data interaksi implisit platform Nimbus.
2. Model ALS secara konsisten mengungguli baseline *user-based* maupun *item-based nearest neighbor* pada seluruh metrik evaluasi.
3. Mekanisme peluruhan waktu terbukti meningkatkan relevansi rekomendasi khususnya bagi pengguna *cold-start* dengan riwayat interaksi pendek.

V.2 Saran

Untuk penelitian dan pengembangan selanjutnya, penulis menyarankan:

1. melakukan pengujian *online* melalui *A/B testing* pada lingkungan produksi untuk memvalidasi temuan *offline*;
2. mengeksplorasi pendekatan *hybrid* yang menggabungkan sinyal konten produk dengan faktor laten kolaboratif; dan
3. mengkaji arsitektur berbasis *deep learning*, seperti *neural collaborative filtering*, sebagai pembanding lanjutan.

DAFTAR PUSTAKA

- Hardjono, T., Wibisono, A., & Santoso, D. (2021). Hybrid recommendation approaches for fashion e-commerce platforms. *Jurnal Sistem Informasi Nusantara*, 9(2), 112–124.
- Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. *Proceedings of the IEEE International Conference on Data Mining*, 263–272.
- Kusuma, S. R., & Prabowo, A. (2022). Choice overload and personalization strategies in online retail. *Jurnal Manajemen Teknologi Informasi*, 14(1), 45–58.
- Prasetyo, B., & Ramadhani, F. (2023). Time-aware collaborative filtering for cold-start users. *Jurnal Ilmu Komputer dan Informasi*, 16(3), 201–215.
- Santoso, W. (2020). Personalisasi pengalaman pengguna pada platform e-dagang: Sebuah tinjauan sistematis. *Jurnal Teknologi dan Sistem Informasi*, 6(1), 33–47.
- Wibisono, A., & Ananta, R. (2019). Scalability challenges of item-based collaborative filtering on large-scale catalogs. *Prosiding Seminar Nasional Ilmu Komputer*, 78–86.

LAMPIRAN

Lampiran A. Cuplikan Kueri Konstruksi Matriks Interaksi

```
SELECT user_id, product_id,
       SUM(CASE
         WHEN event_type = 'click' THEN 1
         WHEN event_type = 'add_to_cart' THEN 3
         WHEN event_type = 'purchase' THEN 5
         ELSE 0
       END) AS interaction_weight
FROM interaction_log
WHERE event_date BETWEEN '2025-01-01' AND '2025-12-31'
GROUP BY user_id, product_id;
```

Lampiran B. Parameter Akhir Model ALS Terpilih

Parameter	Nilai
Jumlah faktor laten (k)	64
Koefisien regularisasi (λ)	0,1
Jumlah iterasi	20
Skema pembobotan	<i>time-decay exponential</i>