

**ANALISIS KUANTITATIF SISTEM REKOMENDASI
PRODUK PADA *PLATFORM E-COMMERCE* BERBASIS
*MODEL EMBEDDING GEMINI***

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

YAN ABELLINO

22.11.4673

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS KUANTITATIF SISTEM REKOMENDASI
PRODUK PADA *PLATFORM E-COMMERCE* BERBASIS
MODEL *EMBEDDING GEMINI***

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

YAN ABELLINO

22.11.4673

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS KUANTITATIF SISTEM REKOMENDASI
PRODUK PADA *PLATFORM E-COMMERCE* BERBASIS
MODEL *EMBEDDING GEMINI***

yang disusun dan diajukan oleh

Yan Abellino

22.11.4673

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 27 Februari 2026

Dosen Pembimbing,

Mulia Sulistiyono, M.Kom.

NIK. 190302248

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS KUANTITATIF SISTEM REKOMENDASI
PRODUK PADA *PLATFORM E-COMMERCE* BERBASIS
MODEL *EMBEDDING GEMINI*

yang disusun dan diajukan oleh

Yan Abellino

22.11.4673

Telah dipertahankan di depan Dewan Penguji
pada tanggal 27 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Rumini, S.Kom., M.Kom.
NIK. 190302246

Rifda Faticha Alfa Aziza, S.Kom., M.Kom.
NIK. 190302392

Mulia Sulistiyono, S.Kom., M.Kom.
NIK. 190302248



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 27 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Yan Abellino
NIM : 22.11.4673

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Kuantitatif Sistem Rekomendasi Produk Pada Platform E-Commerce Berbasis Model *Embedding* Gemini

Dosen Pembimbing: Mulia Sulistiyono, M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 27 Februari 2026

Yang Menyatakan,



MANX363518559

Yan Abellino

HALAMAN PERSEMBAHAN

Puji dan syukur penulis panjatkan ke hadirat Allah SWT, atas segala limpahan rahmat, karunia, dan hidayah-Nya. Atas izin dan kehendak-Nya, penulis diberikan kesehatan, kekuatan, kesabaran, serta kemudahan sehingga mampu menyelesaikan skripsi ini dengan baik. Setiap proses yang dilalui merupakan pembelajaran berharga yang tidak terlepas dari pertolongan-Nya.

Skripsi ini penulis persembahkan dengan penuh rasa hormat dan cinta kepada kedua orang tua tercinta, Bapak Syutami dan Mama Ewi, serta Kakak Icha, yang senantiasa memberikan doa, kasih sayang, dukungan moral maupun materi, serta pengorbanan yang tidak pernah terhitung nilainya. Ketulusan, kesabaran, dan kepercayaan yang diberikan menjadi sumber motivasi utama bagi penulis dalam menempuh perjalanan akademik ini.

Persembahan ini juga penulis tujukan kepada Ara pasangan tercinta, yang telah menjadi pendamping setia dalam berbagai situasi. Dukungan, pengertian, perhatian, serta semangat yang diberikan telah menjadi penguat ketika penulis berada dalam masa-masa sulit, sekaligus menjadi penyemangat untuk terus berusaha dan tidak menyerah.

Terakhir, karya tulis ini penulis persembahkan kepada Aji dan Ino serta teman-teman seperjuangan selama masa studi, sebagai bentuk apresiasi atas semangat, dukungan, dan kebersamaan yang telah menjadi bagian penting dalam perjalanan akademik penulis.

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT yang telah melimpahkan rahmat, hidayah, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Kuantitatif Sistem Rekomendasi Produk Pada *Platform E-Commerce* Berbasis Model *Embedding Gemini*”

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta. Penulis menyadari bahwa dalam penyusunan skripsi ini tidak lepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih yang sebesar besarnya kepada:

1. Prof. Dr. Mohammad Suryanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
2. Prof. Dr. Kusri, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Eli Pujastuti, M.kom selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Bapak Mulia Sulistiyono, M.Kom., selaku Dosen Pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan serta arahan yang sangat berharga selama proses pengerjaan skripsi ini.
5. Tim Dosen Penguji, yang telah memberikan saran, kritik, dan masukan berharga dalam penyempurnaan skripsi ini.
6. Kedua orang tua tercinta dan keluarga besar yang senantiasa memberikan doa, kasih sayang, dukungan moral, maupun materiil yang tiada henti.

Yogyakarta, <tanggal bulan tahun>

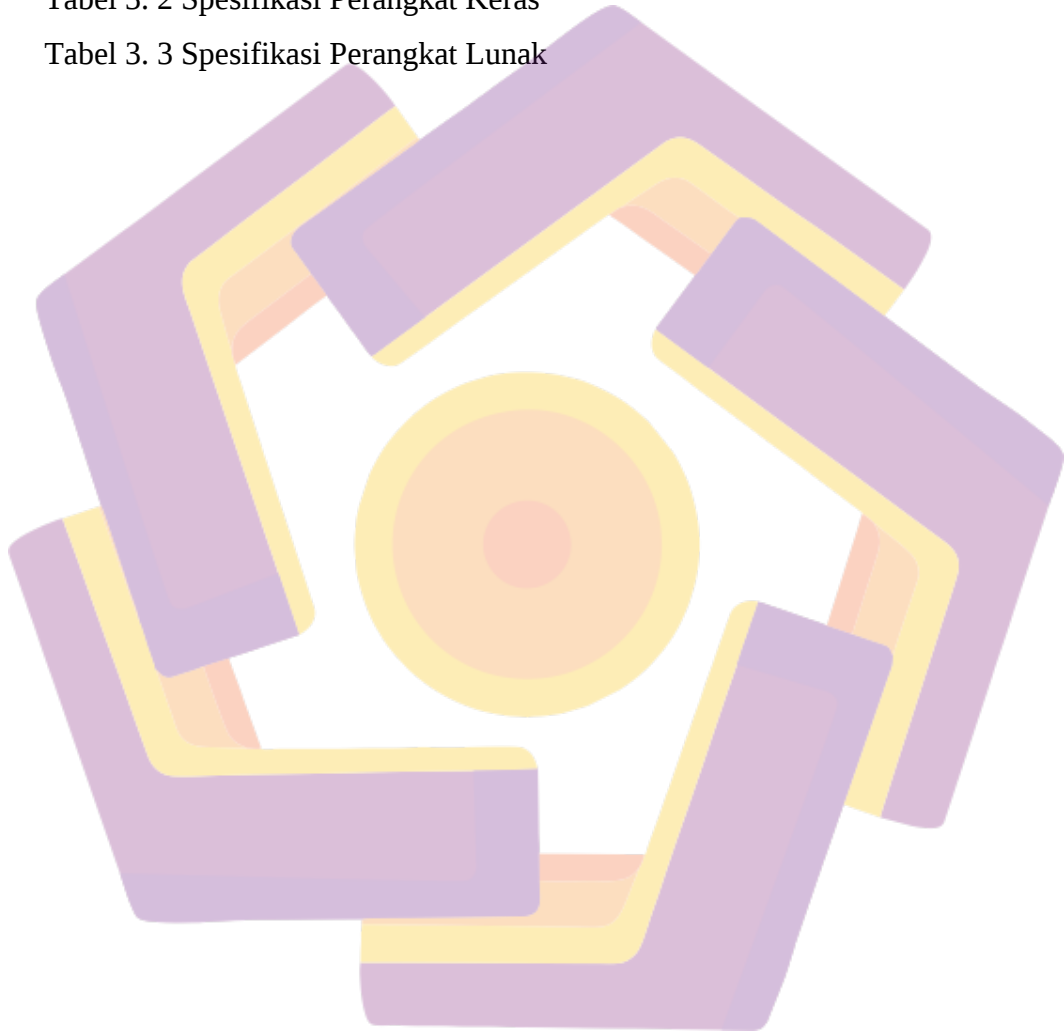
DAFTAR ISI

HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH.....	xiv
INTISARI	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	10

BAB III METODE PENELITIAN	16
3.1 Objek Penelitian	16
3.2 Alur Penelitian.....	17
BAB IV HASIL DAN PEMBAHASAN	25
4.1. Detail Dataset	25
4.2. Hasil Representasi <i>Embedding</i> Pengguna.....	26
4.3. Penentuan Jumlah <i>Cluster</i>	27
4.4. Hasil <i>Clustering</i> Menggunakan Algoritma <i>K-Means</i>	28
4.5. Implikasi terhadap Sistem Rekomendasi Produk.....	30
4.6. Implementasi dan Antarmuka Sistem (Deployment)	31
4.7. Pembahasan Hasil Penelitian.	31
4.8. Keterbatasan Penelitian.	32
4.9. Ringkasan Bab.....	32
BAB V PENUTUP	33
5.1 Kesimpulan.....	33
5.2 Saran.....	34
REFERENSI	36
LAMPIRAN.....	39

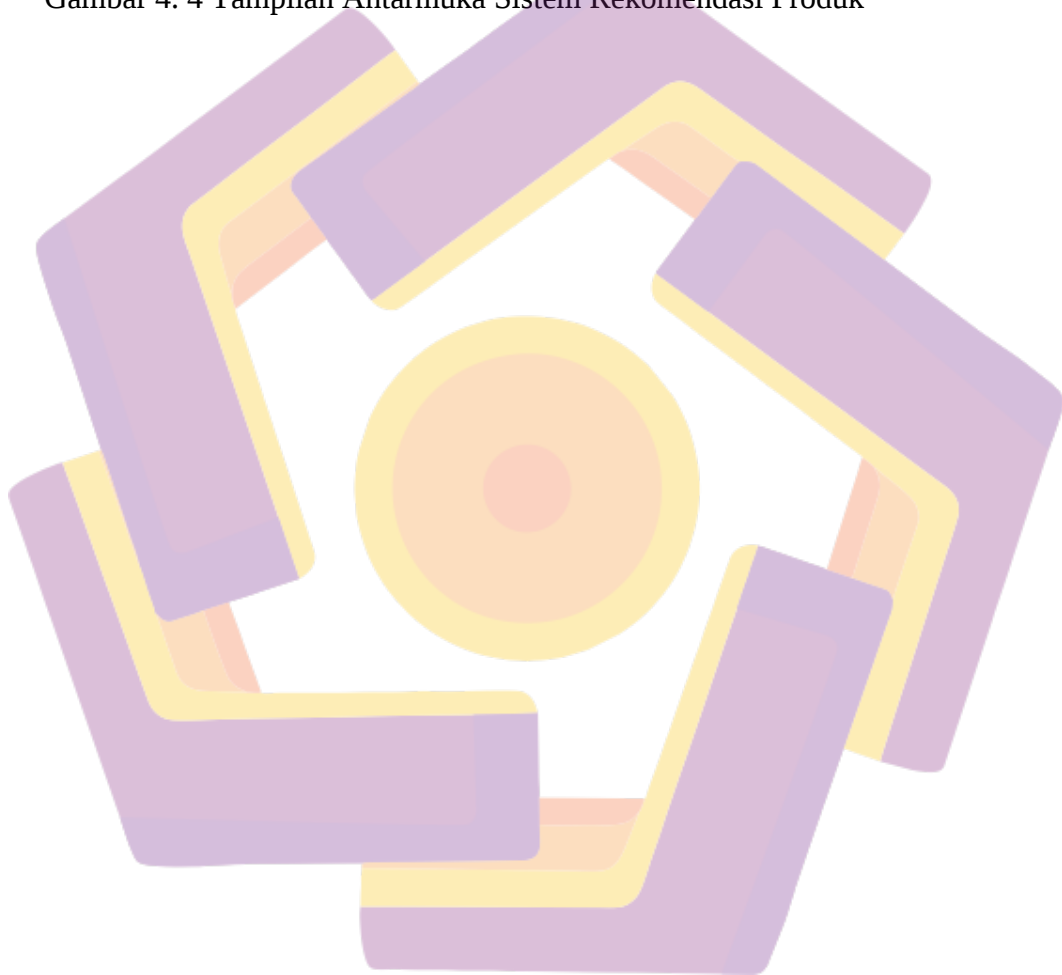
DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	8
Tabel 2. 2 Keaslian Penelitian (Lanjutan)	9
Tabel 3. 1 Variabel Penelitian	16
Tabel 3. 2 Spesifikasi Perangkat Keras	23
Tabel 3. 3 Spesifikasi Perangkat Lunak	24



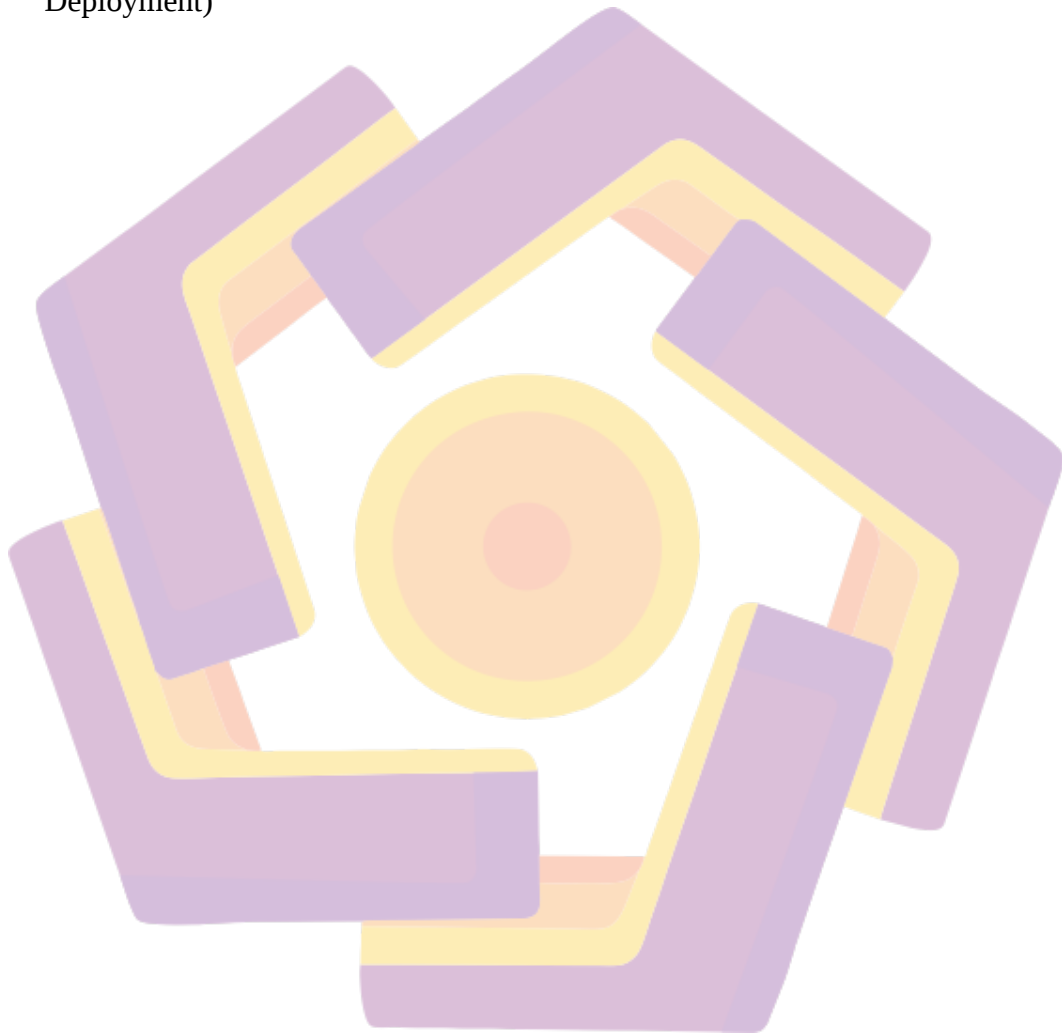
DAFTAR GAMBAR

Gambar 3. 1 Diagram Alir Penelitian	17
Gambar 4. 1 Visualisasi Embedding Pengguna Menggunakan PCA	26
Gambar 4. 2 Grafik Elbow Method Penentuan Jumlah Cluster	27
Gambar 4. 3 Visualisasi Hasil Clustering K-Means	29
Gambar 4. 4 Tampilan Antarmuka Sistem Rekomendasi Produk	31



DAFTAR LAMPIRAN

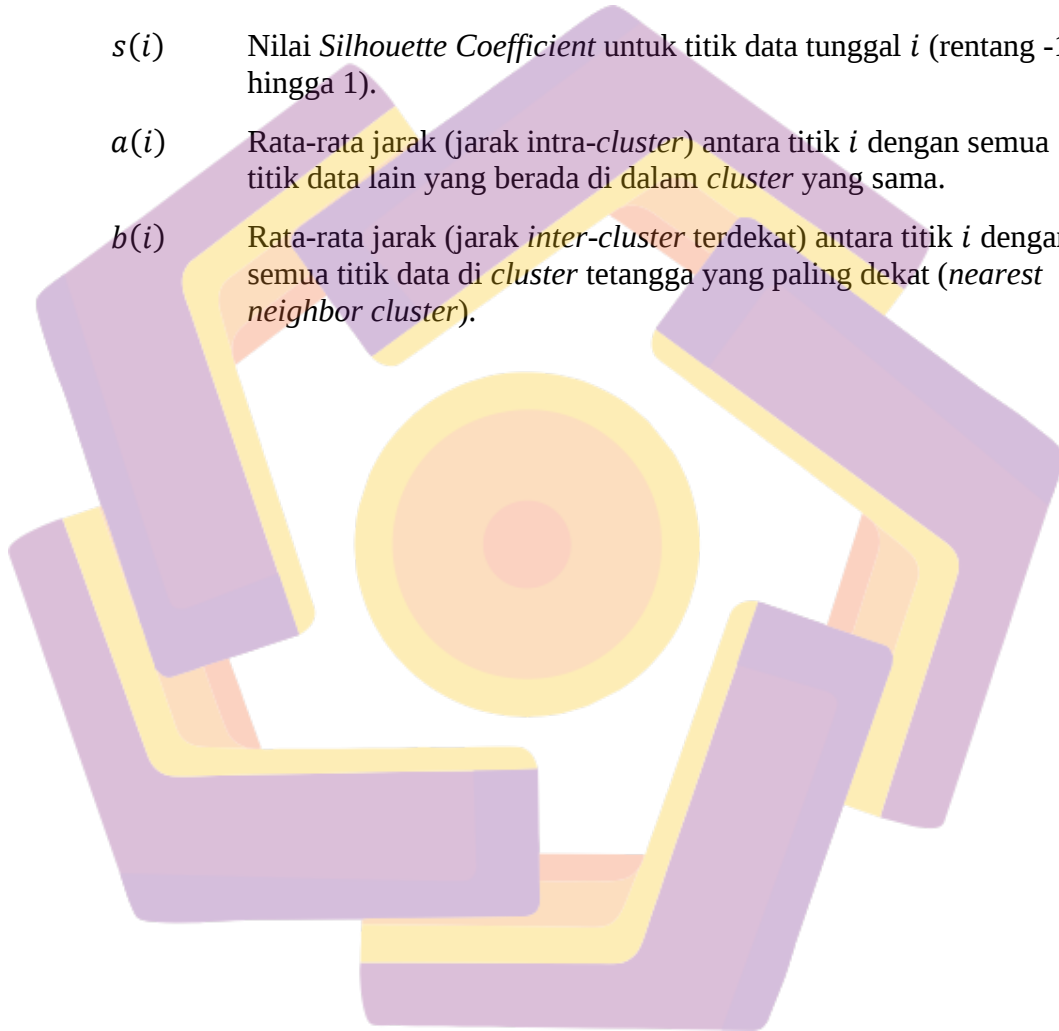
Lampiran 1 Dataset	39
Lampiran 2 Source Code	39
Lampiran 3 Tampilan Antarmuka Sistem Rekomendasi E-Commerce (Hasil Deployment)	41



DAFTAR LAMBANG DAN SINGKATAN

LLM	<i>Large Language Model</i> (Model Bahasa Besar)
PCA	<i>Principal Component Analysis</i> (Analisis Komponen Utama)
K-Means	Algoritma pengelompokan data non-hirarki (<i>partitioning clustering</i>)
CF	<i>Collaborative Filtering</i> (Penyaringan Kolaboratif)
CBF	<i>Content-Based Filtering</i> (Penyaringan Berbasis Konten)
TF-IDF	<i>Term Frequency-Inverse Document Frequency</i>
API	<i>Application Programming Interface</i> (Antarmuka Pemrograman Aplikasi)
WCSS	<i>Within-Cluster Sum of Squares</i> (Jumlah Kuadrat Dalam Klaster)
SSE	<i>Sum of Squared Errors</i> (Jumlah Kesalahan Kuadrat)
NDCG	<i>Normalized Discounted Cumulative Gain</i> (Metrik evaluasi peringkat)
KDD	<i>Knowledge Discovery in Databases</i>
SSD	<i>Solid State Drive</i> (Media penyimpanan data)
NVMe	<i>Non-Volatile Memory Express</i> (Protokol transfer data penyimpanan)
<i>similarity</i>	Nilai skor kemiripan (rentang biasanya -1 hingga 1, atau 0 hingga 1 untuk data positif).
$\cos \cos (\theta)$	Sudut yang terbentuk antara vektor A dan vektor B
A, B	Vektor data yang sedang dibandingkan kemiripannya.
$ A B $	Dua vektor (data) yang sedang dibandingkan
$A_i B_i$	Komponen ke i dari vektor A dan B .
n	Jumlah total dimensi atau fitur dalam vektor tersebut.
i	Jumlah total dimensi atau fitur dalam vektor tersebut.
J	Nilai fungsi objektif (cost function), merepresentasikan total error kuadrat jarak.

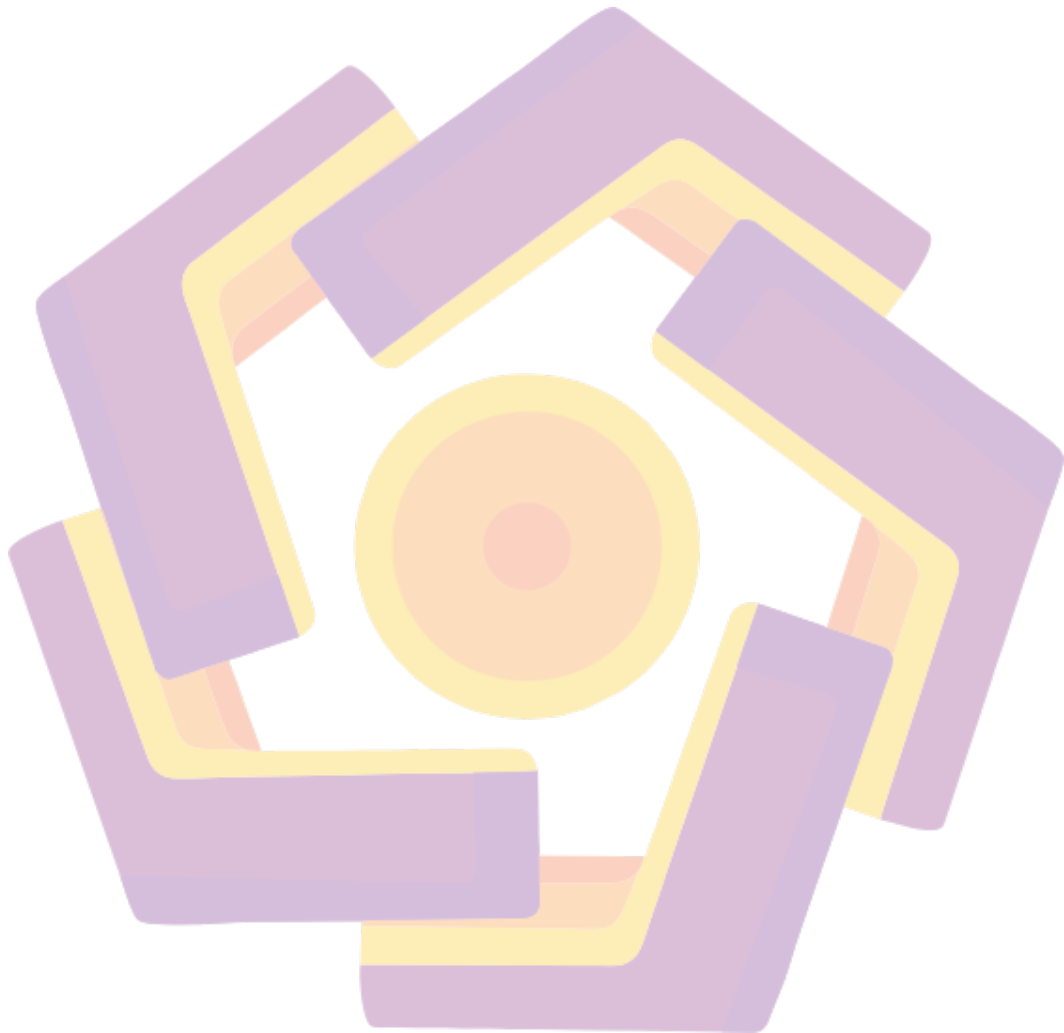
- k Jumlah total *cluster* yang ditentukan.
- $x_i^{(j)}$ Titik data ke- i yang telah dikelompokkan ke dalam *cluster*.
- c_j Centroid (titik pusat rata-rata) dari *cluster*.
- j Indeks untuk *cluster*.
- $s(i)$ Nilai *Silhouette Coefficient* untuk titik data tunggal i (rentang -1 hingga 1).
- $a(i)$ Rata-rata jarak (jarak *intra-cluster*) antara titik i dengan semua titik data lain yang berada di dalam *cluster* yang sama.
- $b(i)$ Rata-rata jarak (jarak *inter-cluster* terdekat) antara titik i dengan semua titik data di *cluster* tetangga yang paling dekat (*nearest neighbor cluster*).



DAFTAR ISTILAH

<i>Embedding</i>	Representasi data (teks) menjadi vektor numerik berdimensi tinggi
<i>Deep Learning</i>	Pembelajaran mendalam; cabang <i>Machine Learning</i> menggunakan jaringan saraf tiruan
<i>Transformer</i>	Arsitektur jaringan saraf yang menjadi dasar model <i>LLM</i> (seperti <i>Gemini</i>)
<i>Information Overload</i>	Fenomena kelebihan informasi yang membingungkan pengguna
<i>Cold-start</i>	Masalah sistem rekomendasi saat menghadapi pengguna/produk baru tanpa data historis
<i>Cosine Similarity</i>	Metode pengukuran kemiripan dua vektor berdasarkan sudut kosinus
<i>Elbow Method</i>	Metode grafik untuk menentukan jumlah kluster optimal (titik siku)
<i>Silhouette Score</i>	Skor validasi kualitas kluster (-1 hingga 1)
<i>Sparsity</i>	Kelangkaan data; kondisi matriks yang didominasi nilai nol/kosong
<i>User Corpus</i>	Gabungan teks profil dan riwayat pengguna menjadi satu dokumen
<i>L2 Normalization</i>	Teknik normalisasi vektor agar memiliki panjang satuan (unit norm)
<i>Euclidean Distance</i>	Jarak garis lurus antara dua titik dalam ruang vektor
<i>Feature Engineering</i>	Rekayasa fitur; proses mengubah data mentah menjadi fitur yang dapat diproses model
<i>Precision@K</i>	Presisi pada peringkat ke- <i>K</i> ; proporsi item relevan dalam daftar rekomendasi teratas
<i>Recall@K</i>	<i>Recall</i> pada peringkat ke- <i>K</i> ; kemampuan sistem menemukan kembali item relevan
<i>DBSCAN</i>	<i>Density-Based Spatial Clustering of Applications with Noise</i> (Algoritma kluster berbasis kepadatan)
<i>Unsupervised Learning</i>	Pembelajaran tanpa pengawasan (tanpa label data)

<i>Data Mining</i>	Penambangan data; proses menemukan pola dalam kumpulan data besar
<i>Generative AI</i>	Kecerdasan buatan yang mampu menghasilkan konten baru (teks, gambar, dll)
<i>Inertia</i>	Jumlah kuadrat jarak sampel ke pusat klaster terdekatnya (terkait WCSS)



INTISARI

Perkembangan *e-commerce* yang pesat menyebabkan meningkatnya jumlah produk dan informasi yang harus diproses oleh pengguna. Kondisi ini menimbulkan permasalahan *information overload*, di mana pengguna kesulitan menemukan produk yang relevan dengan preferensinya. Dampak dari permasalahan tersebut adalah menurunnya kualitas pengalaman pengguna, rendahnya tingkat kepuasan, serta berpotensi menurunkan konversi dan retensi pelanggan pada *platform e-commerce*. Oleh karena itu, diperlukan sistem rekomendasi yang mampu memahami karakteristik dan preferensi pengguna secara lebih mendalam melalui representasi data yang akurat dan kontekstual.

Penelitian ini mengusulkan pendekatan sistem rekomendasi produk berbasis *embedding* semantik dengan memanfaatkan *Large Language Model (LLM) Google Gemini*. Deskripsi produk diproses menggunakan model *text-embedding-004* untuk menghasilkan representasi vektor berdimensi tinggi. Selanjutnya, algoritma *K-Means* digunakan untuk melakukan pengelompokan produk berdasarkan kemiripan semantik. Penentuan jumlah kluster optimal dilakukan menggunakan *Elbow Method*, sementara kualitas kluster dievaluasi menggunakan *Silhouette Score*. Untuk membantu interpretasi hasil, *Principal Component Analysis (PCA)* diterapkan sebagai alat visualisasi tanpa memengaruhi proses *clustering*.

Hasil penelitian menunjukkan bahwa *embedding Google Gemini* mampu membentuk struktur kluster produk yang koheren secara semantik, meskipun nilai *Silhouette Score* menunjukkan tingkat pemisahan kluster yang moderat. Temuan ini mengindikasikan bahwa pendekatan berbasis *embedding* efektif untuk merepresentasikan preferensi pengguna yang bersifat kompleks dan tumpang tindih. Kontribusi penelitian ini terletak pada penerapan *LLM* dalam sistem rekomendasi berbasis *clustering* semantik pada konteks *e-commerce*. Hasil penelitian ini dapat dimanfaatkan oleh pengembang sistem rekomendasi, pelaku industri *e-commerce*, serta peneliti di bidang *data mining* dan sistem rekomendasi. Penelitian selanjutnya disarankan untuk mengintegrasikan metrik evaluasi rekomendasi seperti *Precision@K* dan *NDCG* guna meningkatkan validitas kinerja sistem.

Kata kunci: Sistem Rekomendasi, *E-Commerce*, *Google Gemini*, *Embedding Semantik*, *K-Means*.

ABSTRACT

The rapid development of e-commerce has led to an increase in the number of products and information that users have to process. This condition causes the problem of information overload, where users have difficulty finding products that are relevant to their preferences. The impact of these problems is a decrease in the quality of user experience, low satisfaction levels, and the potential to decrease customer conversions and retention on e-commerce platforms. Therefore, a recommendation system is needed that is able to understand user characteristics and preferences more deeply through accurate and contextual data representation.

This study proposes an approach to a product recommendation system based on semantic embedding by utilizing Google Gemini's Large Language Model (LLM). Product descriptions are processed using the text-embedding-004 model to generate high-dimensional vector representations. Furthermore, the K-Means algorithm is used to group products based on semantic similarity. The determination of the optimal number of clusters was carried out using the Elbow Method, while the quality of the cluster was evaluated using the Silhouette Score. To help interpret the results, Principal Component Analysis (PCA) is applied as a visualization tool without affecting the clustering process.

The results show that Google Gemini's embedding is capable of forming a semantically coherent product cluster structure, although the Silhouette Score value indicates a moderate level of cluster separation. These findings indicate that embedding-based approaches are effective for representing complex and overlapping user preferences. The contribution of this research lies in the application of LLM in a semantic clustering-based recommendation system in the context of e-commerce. The results of this research can be used by recommendation system developers, e-commerce industry players, and researchers in the field of data mining and recommendation systems. Further research is suggested to integrate recommendation evaluation metrics such as Precision@K and NDCG to improve the validity of system performance.

Keywords: recommendation system, e-commerce, Google Gemini, semantic embedding, K-Means.