

**ANALISIS PRODUK PADA APLIKASI TOKOPEDIA
MENGUNAKAN ALGORITMA K-MEANS
CLUSTERING DAN HIERARCHICAL AGGLOMERATIVE
CLUSTERING (HAC) BERDASARKAN TREN
PENJUALAN ONLINE
SKRIPSI SARJANA INFORMATIKA**

Oleh

Puspa Ayu Soleha
(217064516020)



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL
2024**

**ANALISIS PRODUK PADA APLIKASI TOKOPEDIA
MENGUNAKAN ALGORITMA K-MEANS
CLUSTERING DAN HIERARCHICAL AGGLOMERATIVE
CLUSTERING (HAC) BERDASARKAN TREN
PENJUALAN ONLINE**

SKRIPSI SARJANA INFORMATIKA

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar Sarjana Program Studi
Informatika Fakultas Teknologi Komunikasi dan Informatika

Oleh
Puspa Ayu Soleha
(217064516020)



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL**

2024

HALAMAN PENGESAHAN
TUGAS AKHIR

**ANALISIS PRODUK PADA APLIKASI TOKOPEDIA MENGGUNAKAN
ALGORITMA K-MEANS CLUSTERING DAN HIERARCHICAL
AGGLOMERATIVE CLUSTERING (HAC) BERDASARKAN TREN PENJUALAN
ONLINE**



PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

**ANALISIS PRODUK PADA APLIKASI TOKOPEDIA MENGGUNAKAN
ALGORITMA K-MEANS CLUSTERING DAN HIERARCHICAL
AGGLOMERATIVE CLUSTERING (HAC) BERDASARKAN TREN PENJUALAN
ONLINE**

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.

Jakarta, 3 Maret, 2025



Puspa Ayu Soleha

(217064516020)

LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

**ANALISIS PRODUK PADA APLIKASI TOKOPEDIA MENGGUNAKAN
ALGORITMA K-MEANS CLUSTERING DAN HIERARCHICAL
AGGLOMERATIVE CLUSTERING (HAC) BERDASARKAN TREN PENJUALAN
ONLINE**

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Ganjil 2024-2025 pada tanggal 25 Februari Tahun 2025

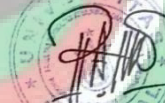
Dosen Pembimbing 1



Dr. Agung Triavudi, S.Kom., M.Kom

NIDN. 0419068604

Ketua Program Studi



Ratih Titi Komala Sari, S.T., MMSI

NIDN. 0301038302

UNIVERSITAS NASIONAL

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

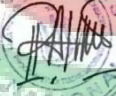
Nama : Puspa Ayu Solcha
 NPM : 217064516020
 Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
 Program Studi : Informatika
 Tanggal Sidang : 25 Februari, 2025

JUDUL DALAM BAHASA INDONESIA :

ANALISIS PRODUK PADA APLIKASI TOKOPEDIA MENGGUNAKAN
 ALGORITMA K-MEANS CLUSTERING DAN HIERARCHICAL
 AGGLOMERATIVE CLUSTERING (HAC) BERDASARKAN TREN PENJUALAN
 ONLINE

JUDUL DALAM BAHASA INGGRIS :

PRODUCT ANALYSIS ON THE TOKOPEDIA APPLICATION USING K-MEANS
 CLUSTERING AND HIERARCHICAL AGGLOMERATIVE CLUSTERING (HAC)
 ALGORITHM BASED ON ONLINE SALES TRENDS

TANDA TANGAN DAN TANGGAL		
Pembimbing I	Ka. Prodi	Mahasiswa
TGL : 4 Maret 2025	TGL : 4 Maret 2025	TGL : 4 Maret 2025
		

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Analisis Produk pada Aplikasi Tokopedia Menggunakan Algoritma K-Means Clustering dan Hierarchical Agglomerative Clustering (HAC) Berdasarkan Tren Penjualan Online". Skripsi ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan program Sarjana di Fakultas Teknologi Komunikasi dan Informatika:

Penulisan skripsi ini tidak lepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr. Agung Triayudi, S.Kom., M.Kom, Selaku pembimbing skripsi yang telah memberikan bimbingan, saran, dan motivasi selama proses penulisan skripsi ini.
2. Bapak Dr. Agung Triayudi, S.Kom., M.Kom, Selaku Dekan Fakultas Teknologi Komunikasi dan Informatika di Universitas Nasional.
3. Ibu Ir. Endah Tri Esti Handayani, M.M.S.I, selaku Wakil Dekan Fakultas Teknologi Komunikasi dan Informatika di Universitas Nasional.
4. Ibu Ratih Titi Komalasari, S.T., M.M., M.M.S.I, selaku Ketua Program Studi Informatika di Universitas Nasional
5. Kepada orang tua penulis (Maman dan Farihah) Terima kasih atas setiap doa yang kalian panjatkan kepada penulis, bahkan sebelum penulis sendiri menyadari bahwa penulis membutuhkannya. Terima kasih atas setiap pengorbanan yang telah kalian lakukan, atas kerja keras, dan atas segala upaya yang kalian tempuh demi memastikan penulis mendapatkan

pendidikan terbaik. Tidak ada kata yang cukup untuk menggambarkan betapa bersyukur memiliki orang tua seperti kalian. Kalian adalah alasan utama penulis untuk terus berjuang, dan skripsi ini persembahkan untuk kalian sebagai bentuk kecil dari rasa terima kasih penulis.

6. Teman-teman Prodi Informatika, yang telah menjadi bagian dari perjalanan akademik penulis. Terima kasih atas kebersamaan, dukungan, diskusi, serta semangat yang diberikan selama masa kuliah hingga penyusunan skripsi ini.
7. Sahabat dan teman terkasih penulis Saidah Fatma, Mulki Muhammad Zidane, dan Muhammad Naufal Rifqi yang telah menjadi partner bertumbuh di segala situasi, Membantu penulis agar bisa menyelesaikan skripsi ini dan tak henti saling menyemangati.
8. Seseorang yang tak kalah penting kepada pemilik tanggal lahir 10-10-2000 (MNR) Terima Kasih telah menjadi bagian perjalanan dari hidup penulis. Berkontribusi dalam penulisan skripsi ini, Terima Kasih telah memberikan motivasi ketika penulis merasa tidak mampu dan meyakinkan penulis untuk tidak menyerah hingga penyusunan skripsi ini terselesaikan.
9. ***Last but not least***, Terima kasih untuk diri sendiri telah berjuang sejauh ini bertanggung jawab menyelesaikan apa yang telah dimulai. Terima Kasih telah bertahan dan tidak pernah menyerah.

Jakarta, 18 Februari 2025

Penulis

ABSTRAK

Perkembangan e-commerce yang pesat telah mendorong perusahaan untuk memahami tren penjualan guna meningkatkan strategi pemasaran dan efisiensi bisnis. Tokopedia, sebagai salah satu platform e-commerce terbesar di Indonesia, memiliki jumlah produk yang sangat besar, sehingga diperlukan metode analisis yang efektif untuk mengelompokkan produk berdasarkan kategori pola penjualan. Penelitian ini menerapkan algoritma K-Means Clustering dan Hierarchical Agglomerative Clustering (HAC) dalam mengelompokkan produk di Tokopedia berdasarkan kategori tren penjualan. Data yang digunakan diperoleh melalui web scraping dari platform Tokopedia dengan total 2.209 data produk dari berbagai kategori. Setelah dilakukan preprocessing, data dianalisis menggunakan kedua algoritma clustering untuk menemukan pola pengelompokan produk berdasarkan atribut seperti harga, jumlah ulasan, rating, dan jumlah penjualan. Evaluasi dilakukan menggunakan Silhouette Score dan Within-Cluster Sum of Squares (WCSS) untuk menentukan performa dan akurasi masing-masing algoritma. Hasil penelitian menunjukkan bahwa algoritma HAC lebih optimal dalam membentuk cluster yang lebih kohesif dan sesuai dengan tren penjualan dibandingkan algoritma K-Means. Cluster yang dihasilkan dapat membantu penjual dan pihak Tokopedia dalam memahami pola penjualan, mengelola stok, serta meningkatkan strategi pemasaran berdasarkan kategori produk yang lebih spesifik. Dengan demikian, penelitian ini memberikan kontribusi dalam pemanfaatan metode clustering untuk analisis data e-commerce dalam skala besar.

Kata Kunci: *E-commerce, Tokopedia, Clustering, K-Means, Hierarchical Agglomerative Clustering (HAC), Web Scraping, Tren Penjualan.*

ABSTRACT

The rapid development of e-commerce has led companies to understand sales trends in order to improve marketing strategies and business efficiency. Tokopedia, as one of the largest e-commerce platforms in Indonesia, has a very large number of products, so it applies effective analysis methods to organize products by category sales patterns. This study examines the K-Means Clustering algorithm and Hierarchical Agglomerative Clustering (HAC) in arranging products on Tokopedia based on sales trend categories. The Data used was obtained through web scraping from the Tokopedia platform with a total of 2,209 product data from various categories. After preprocessing, the data were analyzed using both clustering algorithms to find product grouping patterns based on attributes such as price, number of reviews, rating, and number of sales. Evaluation was conducted using Silhouette Score and Within-Cluster Sum of Squares (WCSS) to establish the performance and accuracy of each algorithm. The results showed that the HAC algorithm is more optimal in forming clusters that are more cohesive and in accordance with research trends compared to the k-Means algorithm. The resulting Cluster can help increase Tokopedia's sales and profits in understanding sales patterns, managing stock, and improving marketing strategies based on more specific product categories. Thus, this study contributes to the application of clustering methods for the analysis of e-commerce data on a large scale.

Keywords: *E-commerce, Tokopedia, clustering, K-Means, hierarchical agglomerative clustering (HAC), Web Scraping, sales trends.*

DAFTAR ISI

KATA PENGANTAR.....	vii
ABSTRAK.....	ix
ABSTRACT.....	x
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
BAB I.....	15
PENDAHULUAN.....	15
1.1 Latar Belakang.....	15
1.2 Identifikasi Masalah.....	17
1.3 Tujuan.....	18
1.4 Batasan Masalah.....	18
1.5 Kontribusi.....	19
BAB II TINJAUAN PUSTAKA.....	20
2.1 Landasan Teori.....	20
2.1.1 E-Commerce.....	20
2.1.2 Tokopedia.....	20
2.1.3 Clustering.....	21
2.1.4 Algoritma K-Means.....	21
2.1.5 Algoritma HAC (Hierarchical Agglomerative Clustering).....	23
2.1.6 Machine Learning.....	25
2.1.7 Streamlit.....	26
2.2 Studi Literatur.....	26
BAB III.....	38
METODOLOGI PENELITIAN.....	38
3.1 Waktu Penelitian.....	38
3.2 Alat dan Bahan.....	38
3.3 Penentuan Objek Penelitian.....	39
3.4 Fokus Penelitian.....	39

3.5	Sumber Data.....	39
3.6	Teknik Pengumpulan Data	40
3.7	Desain Penelitian.....	41
3.7.1	Tahap Penelitian	41
3.7.2	Identifikasi Masalah	41
3.7.3	Studi Literatur	42
3.7.4	Pengumpulan Data	42
3.7.5	Preprocessing	42
3.7.6	Penerapan Algoritma K-means Clustering.....	43
3.7.7	Penerapan Hierarchical Agglomerative Clustering.....	44
3.7.8	Pengujian Algoritma	44
3.7.9	Evaluasi Hasil Clustering	45
BAB IV		46
HASIL DAN PEMBAHASAN		46
4.1	Implementasi Algoritma K-Means Clustering	46
4.2.1	Perhitungan Manual Algoritma K-means Clustering.....	46
4.2.2	Tahap Pemodelan Machine Learning Algoritma K-Means Clustering	49
4.2	Implementasi Algoritma Hierarchical Agglomerative Clustering.....	55
4.2.1	Perhitungan Manual Algoritma Hierarchical Agglomerative Clustering.....	55
4.2.2	Tahap Pemodelan Machine Learning Algoritma Hierarchical Agglomerative Clustering.....	58
4.3	Tampilan Interface	66
4.3.1	Visualisasi Data.....	66
4.4	Hasil Perbandingan	73
4.5	Analisis Tambahan	74
BAB V KESIMPULAN DAN SARAN		75
5.1	Kesimpulan	75
5.2	Saran.....	75
DAFTAR PUSTAKA		76
LAMPIRAN		78

DAFTAR TABEL

Tabel 2. 1 Studi Literatur	27
Tabel 3. 1 Waktu Penelitian.....	38
Tabel 4. 1 Sampel Data K-Means	46
Tabel 4. 2 Penentuan Awal Centroid	47
Tabel 4. 3 Perhitungan Jarak Euclidean	48
Tabel 4. 4 Data Konvergen.....	48
Tabel 4. 5 Sampel Data HAC	55
Tabel 4. 6 Normalisasi Data	56
Tabel 4. 7 Matriks Jarak Awal.....	57
Tabel 4. 8 Iterasi Pertama.....	57
Tabel 4. 9 Iterasi Terakhir	57



DAFTAR GAMBAR

Gambar 3. 1 Tahap Penelitian	41
Gambar 3. 2 Algoritma K-Means	43
Gambar 3. 3 Hierarchical Agglomerative Clustering.....	44
Gambar 4. 1 Data Set dan Missing Value K-Means.....	49
Gambar 4. 2 Grafik Elbow Method K-Means.....	50
Gambar 4. 3 Correlation Heatmap K-Means	51
Gambar 4. 4 K-Means Clustering (PCA).....	51
Gambar 4. 5 K-Means: Number Sold vs Price.....	52
Gambar 4. 6 K-Means: Number Sold vs Total Review.....	53
Gambar 4. 7 K-Means: Price vs Total Review.....	53
Gambar 4. 8 Data Set dan Missing Value HAC	58
Gambar 4. 9 Grafik Elbow Method HAC	58
Gambar 4. 10 Correlation Heatmap HAC.....	59
Gambar 4. 11 HAC (PCA).....	60
Gambar 4. 12 HAC: Number Sold vs Price.....	61
Gambar 4. 13 HAC: Number Sold vs Price.....	62
Gambar 4. 14 Single Linkage: Number Sold by Category	62
Gambar 4. 15 Single Linkage: Price by Category.....	63
Gambar 4. 16 Single Linkage: Total Review by Category.....	64
Gambar 4. 17 Dendogram HAC: Single Linkage	65
Gambar 4. 18 Tampilan Streamlit	73
Gambar 4. 19 Silhouette Score: K-Means vs HAC	74