



**Penerapan Metode Clustering dalam Analisis Data Produk Laptop di  
Tokopedia untuk Membuat Rekomendasi Produk: Studi Perbandingan  
K-Means, Hierarchical, dan DBSCAN**

**LAPORAN PENELITIAN**

**ARYA RIZKI YUDHA PRATAMA**

**191420065**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS SAINS TEKNOLOGI**

**UNIVERSITAS BINA DARMA**

**PALEMBANG**

**2023**



**Penerapan Metode Clustering dalam Analisis Data Produk Laptop di  
Tokopedia untuk Membuat Rekomendasi Produk: Studi Perbandingan  
K-Means, Hierarchical, dan DBSCAN**

**ARYA RIZKI YUDHA PRATAMA**

**191420065**

**Laporan Penelitian ini diajukan sebagai syarat memperoleh gelar  
Sarjana Komputer**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS SAINS TEKNOLOGI**

**UNIVERSITAS BINA DARMA**

**PALEMBANG**

**2023**

**HALAMAN PENGESAHAN**

**Penerapan Metode Clustering Dalam Analisis Data Produk Laptop di  
Tokopedia untuk Membuat Rekomendasi Produk: Studi Perbandingan  
K-Means, Hierarchical, dan DBSCAN**

**ARYA RIZKI YUDHA PRATAMA**

**191420065**

**Telah diterima sebagai salah satu syarat untuk memperoleh gelar  
Sarjana Komputer pada Program Studi Teknik Informatika**

**Palembang, 2023  
Program Studi Teknik Informatika  
Fakultas Sains Teknologi  
Universitas Bina Darma  
Dekan,**

**Pembimbing,**



**A. Haidar Mirza, S.T., M.Kom.**



Universitas  
Fakultas Sains Teknologi

**Dr. Tata Sutabri, S.Kom., MMSI., M.KM**

## HALAMAN PERSETUJUAN

Penelitian Berjudul "Penerapan Metode Clustering Dalam Analisis Data Produk Laptop di Tokopedia Untuk Membuat Rekomendasi Produk: Studi Perbandingan K-Means, Hierarchical, dan DBSCAN" Oleh "ARYA RIZKI YUDHA PRATAMA" telah dipertahankan didepan komisi penguji pada hari SABTU tanggal 16 SEPTEMBER 2023.

### Komisi Penguji

1. Ketua : A. Haidar Mirza, S.T., M.Kom.
2. Anggota : M. Nasir, M.M., M.Kom.
3. Anggota : Devi Udariansyah, M.Kom.

(.....)

(.....)

(.....)

Mengetahui  
Program Studi Teknik Informatika  
Fakultas Sains Teknologi  
Universitas Bina Darma  
Ketua,

Alek Wijaya, S.Kom., M.I.T.

## SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Arya Rizki Yudha Pratama

NIM : 191420065

Dengan ini menyatakan bahwa :

1. Karya akhir saya adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) di Universitas Bina Darma atau perguruan tinggi lainnya ;
2. Karya tulis ini murni gagasan, rumusan dan penelitian saya dengan arahan dari tim pembimbing ;
3. Di dalam karya tulis ini tidak terdapat karya atau pendapat yang telah di tulis atau di publikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar rujukan ;
4. Saya bersedia karya tulis ini di cek keasliannya menggunakan plagiarism checker serta di unggah di internet, sehingga dapat diakses secara daring ;
5. Surat pernyataan ini saya tulis dengan sungguh-sungguh dan apabila terbukti melakukan penyimpangan atau ketidakbenaran dalam pernyataan ini maka saya bersedia menerima sanksi dengan peraturan dan perundang-undangan yang berlaku;

Demikian surat pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, Agustus 2023  
Yang membuat pernyataan



ARYA RIZKI YUDHA PRATAMA  
NIM : 191420065

## ABSTRAK

Penelitian ini bertujuan untuk menerapkan metode *clustering* dalam analisis data produk laptop di Tokopedia dengan tujuan membuat rekomendasi produk. Penelitian ini juga akan melakukan studi perbandingan antara tiga metode clustering yang umum digunakan, yaitu K-Means, Hierarchical, dan DBSCAN. Data produk laptop akan diambil melalui teknik *Web Scraping* dari situs E-Commerce Tokopedia. Setelah itu, akan dilakukan *preprocessing* data untuk menambahkan kolom baru yaitu atribut spesifikasi dari produk laptop, mengisi nilai pada kolom baru menggunakan regex lalu dilanjutkan secara manual dan menghapus nilai duplikat. Kemudian, akan dilakukan pemilihan fitur yang relevan untuk clustering. Tiga metode clustering yang akan dibandingkan akan diterapkan pada data produk laptop yang telah dipreprocessing. Evaluasi dan perbandingan kinerja ketiga metode *clustering* akan dilakukan berdasarkan matriks evaluasi seperti Silhouette, Calinski-Harabasz dan Davies-Bouldin. Hasil dari penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik mengenai metode *clustering* dalam analisis data produk dan memberikan rekomendasi produk yang lebih akurat dan personalisasi bagi pengguna.

**Kata kunci :** *Clustering, Rekomendasi, Laptop, Tokopedia.*

## KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah Swt. Yang telah melimpahkan rahmat dan karunia-Nya sehingga Karya Akhir ini dapat diselesaikan guna memenuhi salah satu syarat untuk menyelesaikan studi untuk program Sarjana, Teknik Informatika di Universitas Bina Darma.

Pada Kesempatan yang baik ini, tak lupa penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan bimbingan, arahan, nasehat, dan pemikiran dalam penulisan karya akhir ini terutama kepada:

- Prof. Dr. Sunda Ariana, M.Pd., M.M selaku Rektor Universitas Bina Darma Palembang
- Dr. Tata Sutabri, S.Kom., M.MSI., M.KM selaku Dekan Fakultas Sains Teknologi Universitas Bina Darma Palembang
- Alek Wijaya, S.Kom., M.IT selaku Ketua Program Studi Teknik Informatika Universitas Bina Darma
- A. Haidar Mirza, S.T., M.Kom. selaku Dosen Pembimbing yang telah memberikan bimbingan dan arahan dalam penulisan Laporan Karya Akhir ini
- Muhammad Nasir, M.M., M.Kom. selaku dosen penguji yang telah memberikan bimbingan dan arahan dalam penulisan Laporan Karya Akhir ini
- Devi Udariansyah, M.Kom. selaku dosen penguji yang telah memberikan bimbingan dan arahan dalam penulisan Laporan Karya Akhir ini
- Ibu, Ayah, Nenek dan Adik-Adik saya yang telah memberikan semangat dan motivasi untuk dapat mengerjakan dan menyelesaikan penulisan Laporan Karya Akhir ini
- Mas Aldean Tegar Gemilang dan Bala-Bala Esport yang telah menemani sepanjang malam dalam pengerjaan penelitian ini
- Untuk diri saya sendiri. Terima kasih telah berjuang sekuat, sejauh dan sekeras ini untuk mengerjakan dan menyelesaikan laporan penelitian ini

## DAFTAR ISI

### Halaman

<b>HALAMAN PENGESAHAN .....</b>	<b>i</b>
<b>HALAMAN PERSETUJUAN .....</b>	<b>ii</b>
<b>SURAT PERNYATAAN .....</b>	<b>iii</b>
<b>ABSTRAK.....</b>	<b>iv</b>
<b>KATA PENGANTAR.....</b>	<b>v</b>
<b>DAFTAR ISI .....</b>	<b>vi</b>
<b>DAFTAR GAMBAR.....</b>	<b>viii</b>
<b>DAFTAR TABEL .....</b>	<b>ix</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Perumusan Masalah.....	2
1.3 Tujuan Penelitian.....	2
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian.....	3
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>5</b>
2.1 Penelitian Terdahulu.....	5
2.2 <i>Clustering</i> .....	5
2.3 <i>Web Scraping</i> .....	6
2.4 hidupCuan.....	7
2.5 Analisis Data .....	7
2.6 Anaconda.....	8
2.7 Jupyter Notebook.....	8
2.8 Python .....	9
2.9 PyCaret.....	9
2.10 K-Means Clustering .....	10
2.11 Hierarchical Clustering .....	10
2.12 DBSCAN Clustering.....	11
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>12</b>
3.1 Metode Penelitian .....	12
3.2 Teknik Pengumpulan Data.....	13



<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>15</b>
4.1 Pengumpulan Data .....	15
4.2 <i>Preprocessing Data</i> .....	16
4.2.1 <i>Data Cleaning</i> .....	16
4.2.2 <i>Data Selection</i> .....	17
4.3 Perbandingan kinerja metode clustering .....	17
4.3.1 <i>K-Means Clustering</i> .....	18
4.3.2 <i>Hierarchical Clustering</i> .....	21
4.3.3 <i>DBSCAN Clustering</i> .....	23
4.4 Pembuatan Rekomendasi .....	24
4.4.1 Content Based Filtering .....	24
4.4.1 Item Based Collaborative Filtering .....	27
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>29</b>
5.1 Kesimpulan .....	29
5.2 Saran .....	30
<b>Daftar Pustaka .....</b>	<b>31</b>
<b>Lampiran .....</b>	<b>33</b>

## DAFTAR GAMBAR

	<b>Halaman</b>
Gambar 3. 1 Tahapan Metode CRISP-DM .....	12
Gambar 4. 1 Interface Scraping hidupCuan.....	15
Gambar 4. 2 Hasil Scraping dalam bentuk file csv .....	15
Gambar 4. 3 Tampilan keseluruhan data .....	16
Gambar 4. 4 Data yang telah dibersihkan .....	16
Gambar 4. 5 Data yang telah diseleksi.....	17
Gambar 4. 6 Setup data menggunakan pycaret .....	18
Gambar 4. 7 Membangun model K-Means.....	18
Gambar 4. 8 Elbow Plot.....	19
Gambar 4. 9 Membangun ulang model K-Means.....	19
Gambar 4. 10 PCA Plot K-Means.....	20
Gambar 4. 11 Distribution Plot K-Means.....	20
Gambar 4. 12 Distance Plot K-Means.....	21
Gambar 4. 13 Membangun model Agglomerative Hierarchical Clustering.....	21
Gambar 4. 14 PCA Plot Hierarchical.....	22
Gambar 4. 15 Distribution Plot Hierarchical .....	22
Gambar 4. 16 Membangun model DBSCAN Clustering.....	23
Gambar 4. 17 PCA Plot DBSCAN .....	23
Gambar 4. 18 Distribution Plot DBSCAN .....	24
Gambar 4. 19 Import Library Content Based Filtering.....	24
Gambar 4. 20 Membaca data yang digunakan.....	25
Gambar 4. 21 Penggabungan Atribut.....	25
Gambar 4. 22 Menggunakan TF-IDF .....	25
Gambar 4. 23 Menghitung Similarity .....	25
Gambar 4. 24 Membuat fungsi rekomendasi.....	25
Gambar 4. 25 Membuat fungsi input preferensi.....	26
Gambar 4. 26 Input preferensi pengguna .....	26
Gambar 4. 27 Menampilkan hasil rekomendasi .....	26
Gambar 4. 28 Hasil rekomendasi dari kode .....	26
Gambar 4. 29 Import Library .....	27
Gambar 4. 30 Membaca data yang akan digunakan.....	27
Gambar 4. 31 Membuat pivot table .....	27
Gambar 4. 32 Menghitung Similarity .....	27
Gambar 4. 33 Membuat dataframe similarity .....	27
Gambar 4. 34 Membuat fungsi rekomendasi.....	28
Gambar 4. 35 Membuat rekomendasi .....	28
Gambar 4. 36 Hasil rekomendasi Item-Based Collaborative Filtering .....	28

## DAFTAR TABEL

<b>Tabel 2. 1</b> Penelitian Terdahulu .....	5
--	---