

# PORTOFOLIO APLIKASI KLUSTERING CUSTOMER

---

Proyek: Aplikasi Klustering Customer

Bahasa Pemrograman: Python

Framework: Streamlit



## Analisa Klusterisasi Customer

## 1. Deskripsi Proyek

Aplikasi Klustering Customer ini dikembangkan menggunakan Python dan Streamlit. Tujuan aplikasi adalah melakukan segmentasi pelanggan (customer segmentation) berdasarkan data karakteristik tertentu seperti usia, pendapatan, dan perilaku belanja. Dengan teknik unsupervised learning seperti K-Means, aplikasi membantu memahami kelompok pelanggan yang memiliki kesamaan karakteristik sehingga dapat digunakan untuk strategi bisnis dan pemasaran.

## 2. Struktur Folder dan File

Berikut struktur folder dari proyek:

- extracted\_customer\_project/
  - App\_Klustering\_Customer/
    - customer\_shopping\_data.csv
    - customer\_shopping\_data.xlsx
    - Klustering\_Customer.py
    - requirements.txt
    - utils.py
    - assets/
      - logo.png
    - pages/
      - 1\_Upload Dataset.py
      - 2\_Analisis Algoritma.py
      - 3\_Informasi Dataset.py
    - \_\_pycache\_\_/
      - utils.cpython-38.pyc

## 3. Kebutuhan Sistem (Requirements)

Untuk menjalankan aplikasi ini, diperlukan:

- Python 3.8 atau lebih baru
- Streamlit
- Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Plotly
- File requirements.txt berisi daftar dependensi lengkap

## 4. Langkah Instalasi & Menjalankan Aplikasi

1. Ekstrak file ZIP proyek ke dalam direktori pilihan Anda.

2. Buka terminal atau command prompt, lalu masuk ke folder proyek:

```
cd App_Klustering_Customer
```

3. Instal semua dependensi menggunakan perintah:

```
pip install -r requirements.txt
```

4. Jalankan aplikasi Streamlit dengan perintah:

```
streamlit run main.py (atau file utama aplikasi Anda)
```

5. Aplikasi akan terbuka di browser default, biasanya di alamat `http://localhost:8501`

## 5. Penjelasan Modul dan File Utama

- `main.py`: File utama Streamlit yang menjalankan aplikasi dan menampilkan halaman utama.
- `utils.py`: Berisi fungsi bantu seperti preprocessing data, perhitungan jarak antar data, dan penerapan algoritma klustering.
- `pages/`: Folder berisi halaman tambahan Streamlit seperti upload dataset, hasil klustering, dan visualisasi data.
- `assets/`: Berisi file aset seperti logo atau gambar tambahan.
- `requirements.txt`: Berisi daftar pustaka Python yang dibutuhkan.
- `dataset.csv`: Dataset utama yang digunakan untuk analisis dan klustering pelanggan.

## 6. Screenshot Tampilan Aplikasi

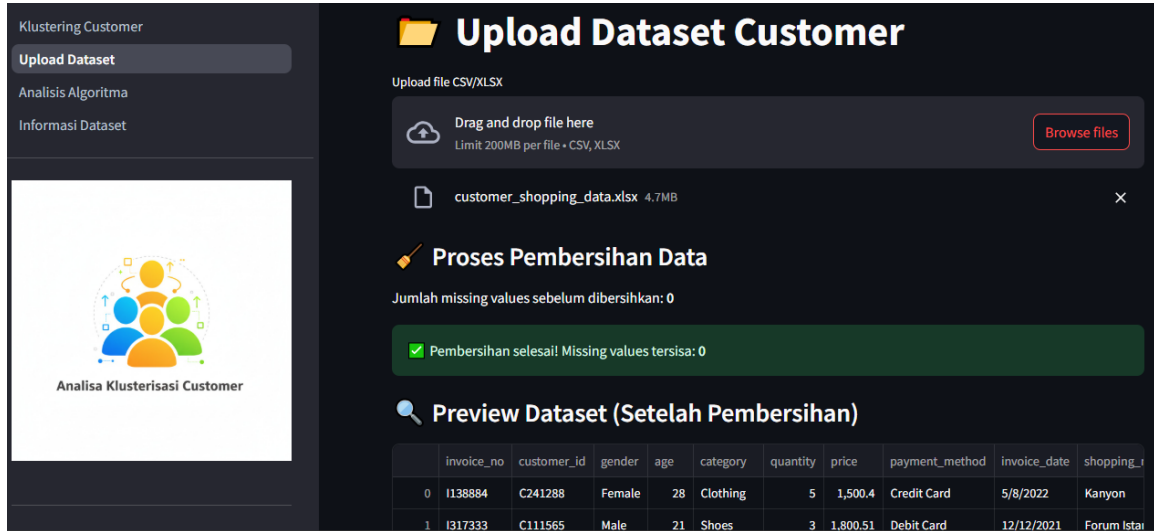
Bagian ini disediakan untuk menambahkan screenshot masing-masing fitur aplikasi.

- Halaman Utama Aplikasi (Home Page)



- Upload Dataset Pelanggan

User dapat mengupload dataset dalam format excel dan csv. Pada halaman ini juga langsung dilakukan pengecekan otomatis apakah terdapat data missing value atau data kotor. Jika ada maka akan dibersihkan datanya menggunakan konsep median. Dan apabila data duplikat maka akan dihapus data tersebut.



	invoice_no	customer_id	gender	age	category	quantity	price	payment_method	invoice_date	shopping_place
0	I138884	C241288	Female	28	Clothing	5	1,500.4	Credit Card	5/8/2022	Kanyon
1	I317333	C111565	Male	21	Shoes	3	1,800.51	Debit Card	12/12/2021	Forum Istai

- Halaman Analisis Algoritma

Pada halaman ini user dipermudah dengan ada fitur tanggal, serta pengaturan data juga terdapat aturan pilihan ataupun input manual data. Kemudian user dipermudah dalam memiliki nilai K. Pada kasus ini saya mendefinisikan nilai K hanya sampai 5:

- K=1 → High Value Customers (Belanja sering, jumlah besar, nilai transaksi tinggi)
- K=2 → Medium Value Customers (Belanja cukup sering, jumlah sedang)
- K=3 → Low Value Customers (Belanja jarang, jumlah kecil, nilai transaksi rendah)
- K=4 → Bargain Shoppers (Belanja quantity tinggi tapi dengan harga murah)
- K=5 → Occasional Buyers (Belanja sesekali, jumlahnya random)


Silhouette Score digunakan untuk menilai kualitas kluster (cluster quality) pada algoritma unsupervised learning seperti K-Means, DBSCAN, atau Agglomerative Clustering.

Silhouette Score selalu berada dalam rentang -1 hingga 1:

- a) +1 Data sangat dekat dengan klasternya sendiri dan jauh dari kluster lain (Kluster sangat jelas)
- b) =0 Data berada di antara dua kluster (tidak jelas masuk ke mana) (Kluster tumpang tindih)
- c) -1 Data lebih dekat ke kluster lain dibanding kluster sendiri (Kluster salah terbentuk)

Nilai Silhouette Score sebesar 0.346 menunjukkan bahwa hasil pengelompokan model cukup baik, meskipun masih terdapat tumpang tindih antar klaster. Nilai ini menandakan bahwa sebagian besar data sudah berada di klaster yang sesuai, namun beberapa titik data masih berada di area perbatasan antar klaster. Model masih dapat ditingkatkan dengan menyesuaikan jumlah klaster atau fitur yang digunakan.

Klustering Customer  
Upload Dataset  
**Analisis Algoritma**  
Informasi Dataset

  
Analisa Klusterisasi Customer

## Analisis Clustering Customer

### Filter Tanggal Invoice

Pilih rentang tanggal:

2021/01/01 - 2021/02/28

### Pengaturan Clustering

Pilih cara menentukan jumlah data:

☒ Pilih dari daftar

☐ Input manual

Pilih jumlah data untuk clustering:

100

Pilih jumlah cluster (k)

2

Silhouette Score: 0.346

Preview hasil clustering dengan K=5

	invoice_no	customer_id	gender	age	category	quantity	price	payment_method	invoice_date	shopping_mall	clust
7,477	I225263	C654298	Female	57	Technology	5	5,250	Credit Card	2021-01-01 00:00:00	Kanyon	
57,173	I202423	C138908	Female	59	Clothing	4	1,200.32	Credit Card	2021-01-17 00:00:00	Cevahir AVM	
51,894	I815829	C178786	Female	22	Clothing	1	300.08	Cash	2021-02-07 00:00:00	Istinye Park	
20,670	I282779	C135632	Female	18	Toys	3	107.52	Cash	2021-02-17 00:00:00	Kanyon	
41,445	I984706	C248901	Female	30	Clothing	5	1,500.4	Cash	2021-01-18 00:00:00	Metrocity	

Visualisasi Clustering.



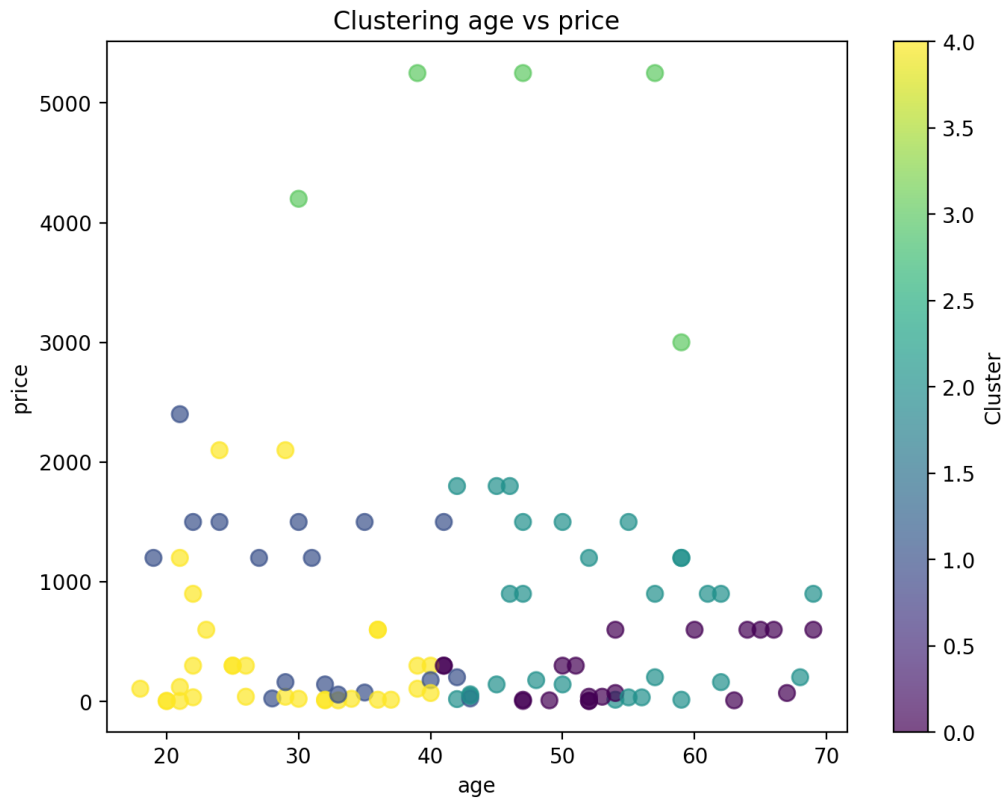
## Visualisasi Clustering

Pilih sumbu X:

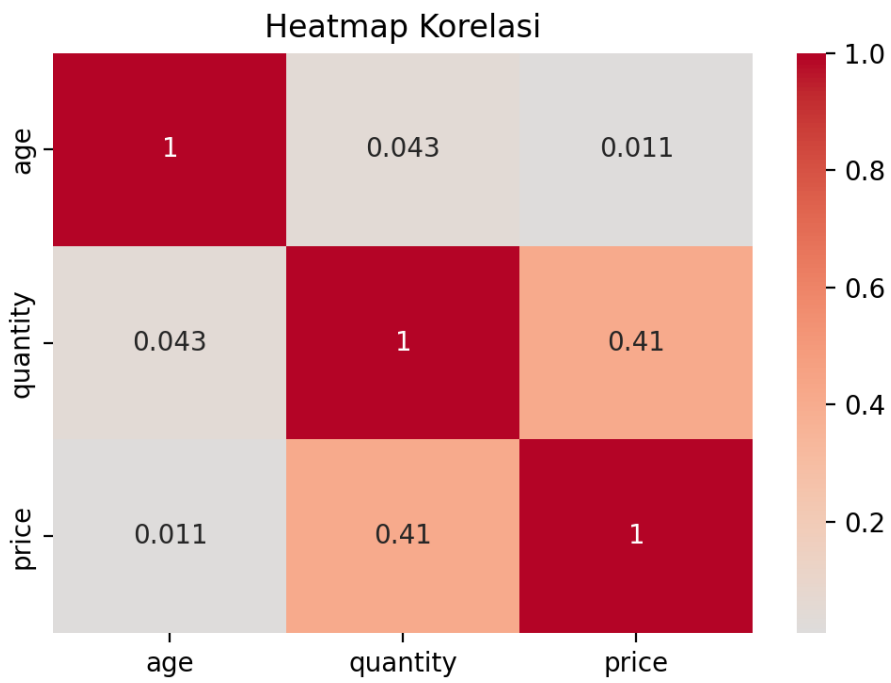
age

Pilih sumbu Y:

price



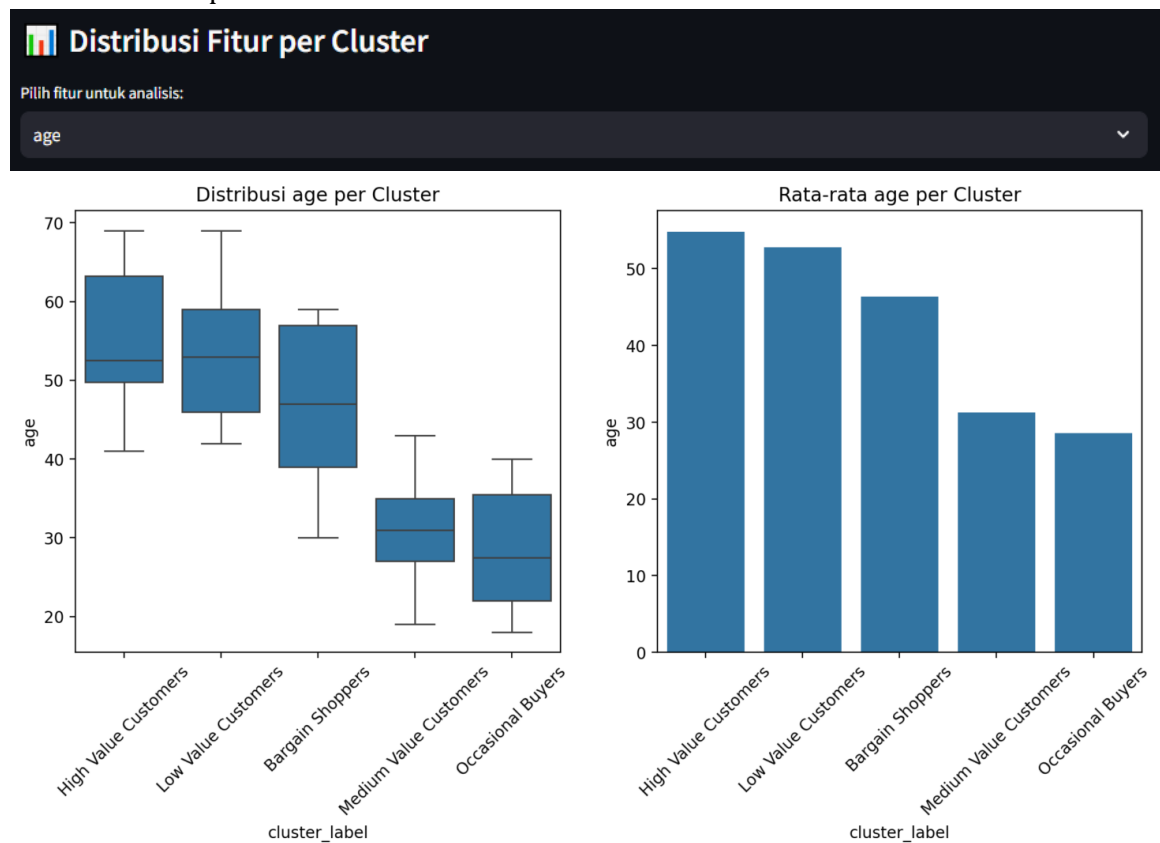
Korelasi antar variabel (Heatmap)



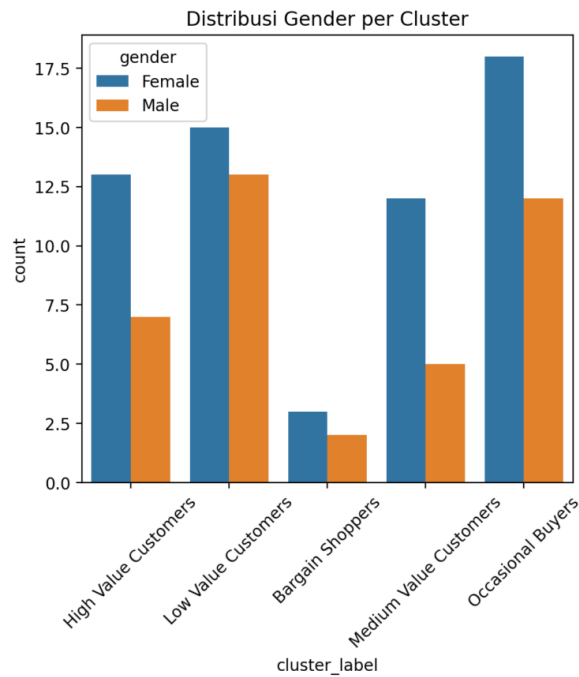
## Ringkasan Cluster

	age		quantity		price	
cluster_label	mean	count	mean	count	mean	count
Bargain Shoppers	46.4	5	4.8	5	4,590.17	5
High Value Customers	54.85	20	1.3	20	253.888	20
Low Value Customers	52.8214	28	3.75	28	720.3746	28
Medium Value Customers	31.2941	17	4.6471	17	845.8141	17
Occasional Buyers	28.6	30	1.8333	30	351.9333	30

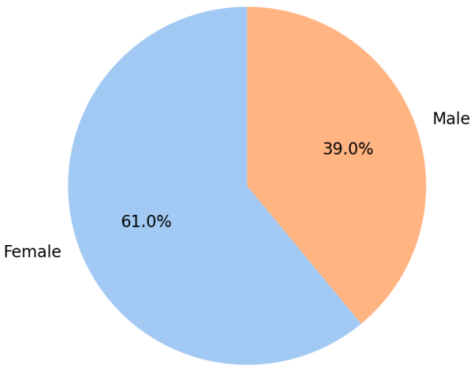
## Distribusi Fitur per Cluster



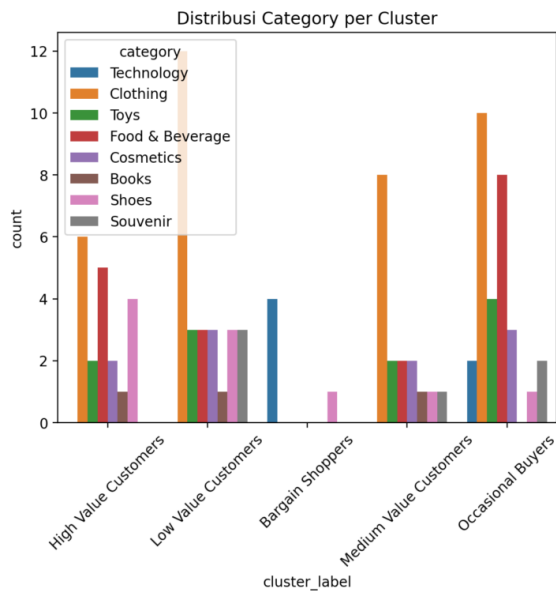
Distribusi Gender per Cluster



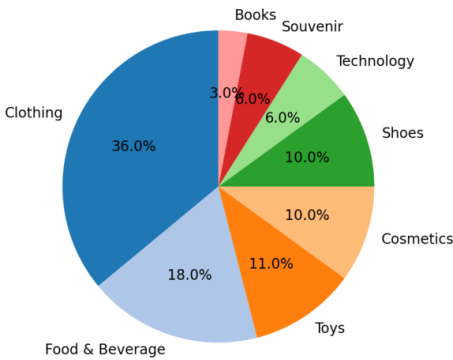
Persentase Gender Keseluruhan



Distribusi Category per Cluster

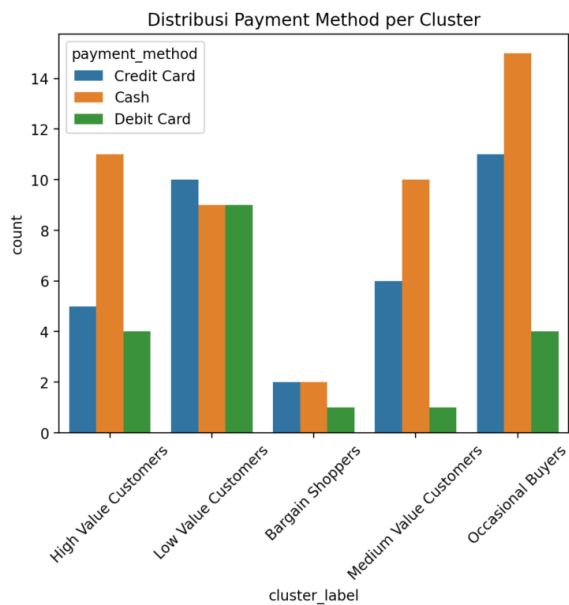


Persentase Category Keseluruhan

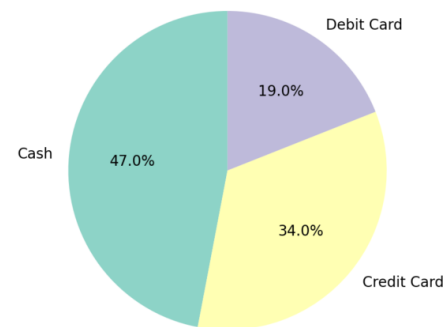




# Distribusi Payment Method per Cluster

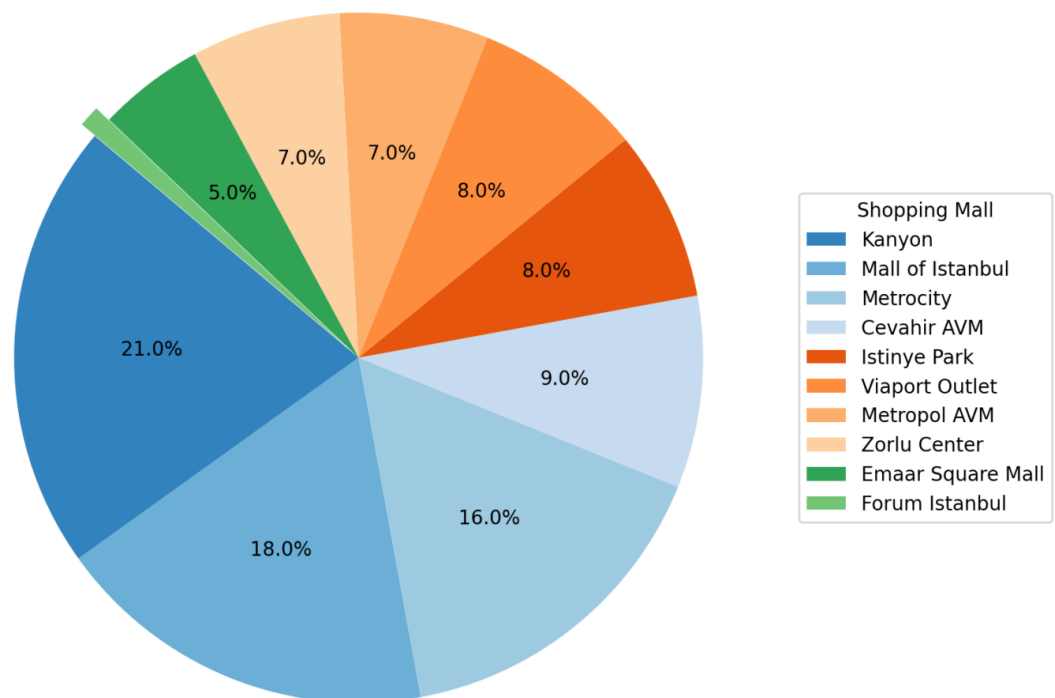


Persentase Payment Method Keseluruhan



# Persentase Shopping Mall

Persentase Shopping Mall Keseluruhan



Simpan hasil clustering dalam bentuk CSV dan Excel

The image shows a web interface for downloading clustering results. It has two buttons: "Download Hasil Clustering (CSV)" and "Download Hasil Clustering (Excel + Summary)". Below this is a screenshot of an Excel spreadsheet titled "hasil\_clustering - Microsoft Excel". The spreadsheet contains a table with columns: invoice\_no, customer\_id, gender, age, category, quantity, price, ment, meth, voice, da, lopping, m, cluster, and luster\_label. The data is organized into rows, with the first row being a header. The table is divided into two sections: "ClusteringResult" and "ClusterSummary".

invoice_no	customer_id	gender	age	category	quantity	price	ment	meth	voice	da	lopping	m	cluster	luster_label
1	I225263	C654298	Female	57	Technology	5	5250	Credit Car	#####	Kanyon			3	Bargain Shoppers
2	I202423	C138908	Female	59	Clothing	4	1200,32	Credit Car	#####	Cevahir A			2	Low Value Customers
3	I815829	C178786	Female	22	Clothing	1	300,08	Cash	#####	Istinye Pa			4	Occasional Buyers
4	I282779	C135632	Female	18	Toys	3	107,52	Cash	#####	Kanyon			4	Occasional Buyers
5	I984706	C248901	Female	30	Clothing	5	1500,4	Cash	#####	Metrocity			1	Medium Value Customers
6	I208252	C572891	Male	54	Food & Be	3	15,69	Cash	#####	Emaar Squ			2	Low Value Customers
7	I273872	C349509	Male	39	Clothing	1	300,08	Credit Car	#####	Viaport O			4	Occasional Buyers
8	I249387	C204986	Female	41	Clothing	1	300,08	Cash	#####	Viaport O			0	High Value Customers
9	I278809	C118896	Male	53	Cosmetics	1	40,66	Cash	#####	Kanyon			0	High Value Customers
10	I170004	C860974	Female	43	Books	4	60,6	Debit Carc	#####	Mall of Ist			2	Low Value Customers
11	I259039	C180485	Female	40	Clothing	1	300,08	Debit Carc	#####	Istinye Pa			4	Occasional Buyers
12	I285349	C311167	Female	59	Shoes	5	3000,85	Debit Carc	#####	Metropol			3	Bargain Shoppers
13	I187655	C252533	Female	28	Food & Be	5	26,15	Credit Car	#####	Metropol			1	Medium Value Customers
14	I336418	C108876	Male	36	Clothing	2	600,16	Credit Car	#####	Zorlu Cen			4	Occasional Buyers
15	I803082	C335409	Female	59	Food & Be	3	15,69	Credit Car	#####	Metrocity			2	Low Value Customers
16	I147737	C654510	Female	24	Technology	2	2100	Cash	#####	Metropol			4	Occasional Buyers
17	I242532	C221385	Male	62	Clothing	3	900,24	Credit Car	#####	Istinye Pa			2	Low Value Customers
18	I261197	C228794	Female	36	Clothing	2	600,16	Credit Car	#####	Mall of Ist			4	Occasional Buyers
19	I770125	C705431	Male	50	Clothing	1	300,08	Cash	#####	Forum Ist			0	High Value Customers
20	I140903	C333701	Male	47	Clothing	5	1500,4	Cash	#####	Kanyon			2	Low Value Customers

## 7. Catatan Pengembangan

Aplikasi ini dapat dikembangkan lebih lanjut dengan fitur tambahan seperti:

- Menambahkan algoritma lain seperti DBSCAN atau Hierarchical Clustering
- Menyimpan hasil segmentasi ke database
- Menampilkan dashboard interaktif menggunakan Plotly atau Power BI
- Fitur ekspor hasil analisis ke file Excel atau PDF

## 8. Profil dan Kontak

Nama: Arjun Yuda Firwanda

Role: Team Program Liraa Job

Email: (arjunyudafirwanda@gmail.com)

LinkedIn: (<https://www.linkedin.com/in/arjun-yuda-firwanda-8095261a4/>)

GitHub: (<https://github.com/arjunyudafirwanda/>)

