

# Predict Customer Personality to boost marketing campaign by using Machine Learning



**Created by:**

**Naufal Iffa Maulana Ramadhan**

naufalimr12a7@gmail.com

<https://www.linkedin.com/in/naufal-iffa-mr/>

Fresh graduate in Informatics Engineering with strong foundation in data analytics, machine learning, and business problem-solving. Experienced in data processing, predictive modeling, and dashboard development using Python, SQL, Excel, and Tableau.

# Agenda

- Latar Belakang & Tujuan Proyek
- Eksplorasi & Persiapan Data
- Metodologi & Pemodelan Segmentasi
- **Hasil: Mengungkap 4 Persona Pelanggan**
- Rekomendasi Bisnis & Dampaknya
- Kesimpulan

- **“Latar Belakang:** Saat ini, strategi pemasaran kita bersifat *one-size-fits-all*, mengakibatkan pemborosan anggaran dan tingkat konversi yang belum optimal.
- **Konteks Masalah:** Tanpa pemahaman mendalam tentang siapa pelanggan kita, pesan pemasaran menjadi tidak relevan dan kurang personal.
- **Tujuan Proyek:** Mengidentifikasi kelompok-kelompok (segmen) pelanggan yang berbeda untuk memungkinkan strategi pemasaran yang lebih personal, efisien, dan efektif.

- **Masalah Utama:** Bagaimana kita bisa beralih dari pemasaran massal ke pemasaran bertarget untuk meningkatkan *engagement* dan ROI?
- **Pertanyaan :**
  - Siapa saja tipe-tipe pelanggan utama kita?
  - Apa yang membedakan mereka satu sama lain?
  - Bagaimana cara terbaik untuk berkomunikasi dengan setiap tipe pelanggan?

- **Relevansi Bisnis:** Perusahaan yang menggunakan pemasaran berbasis data mengalami **peningkatan pendapatan 5-10%** dan **ROI 8x lebih tinggi** (McKinsey).
- **Dampak:** Proyek ini secara langsung mendukung tujuan bisnis untuk:
  - Meningkatkan *Customer Lifetime Value* (CLV).
  - Menurunkan *Customer Acquisition Cost* (CAC).
  - Memperkuat loyalitas merek.

```
[5] 1 df.head()
```

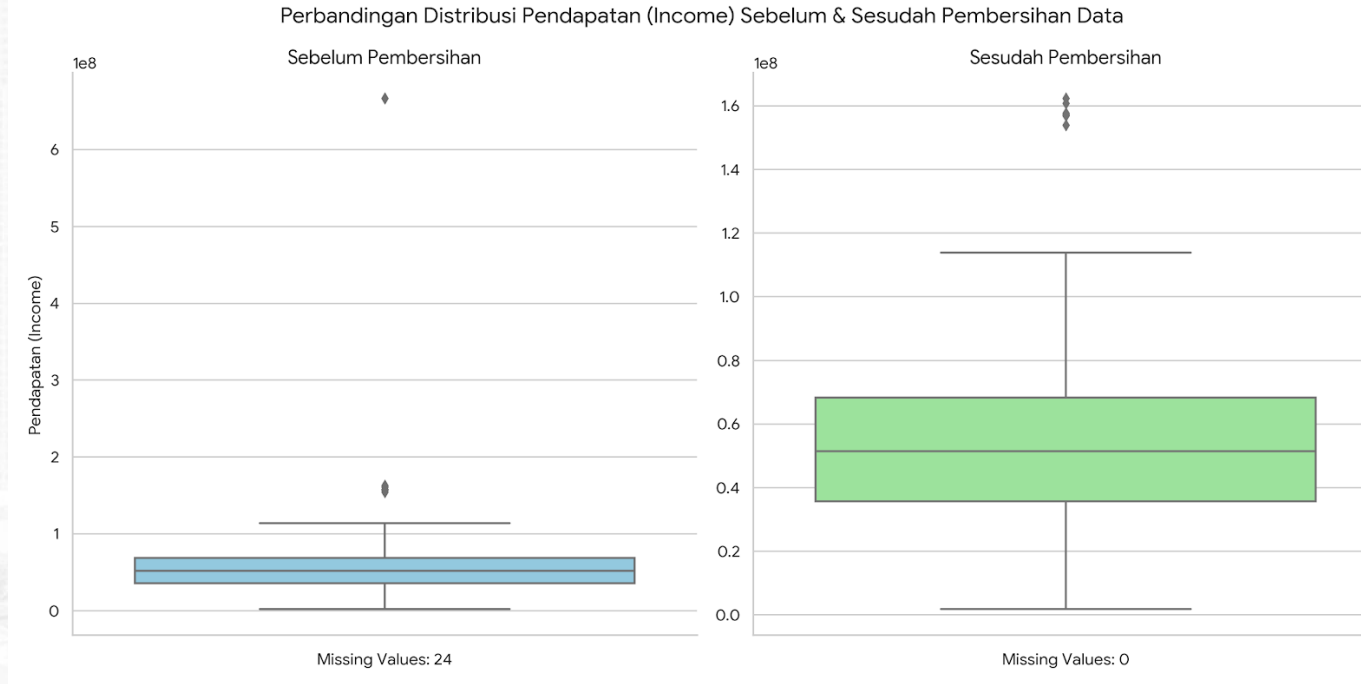


	Unnamed: 0	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	...	NumWebVisitsMonth	AcceptedCmp3	AcceptedCmp4	AcceptedCmp5	AcceptedCmp6
0	0	5524	1957	S1	Lajang	58138000.0	0	0	04-09-2012	58	...	7	0	0	0	0
1	1	2174	1954	S1	Lajang	46344000.0	1	1	08-03-2014	38	...	5	0	0	0	0
2	2	4141	1965	S1	Bertunangan	71613000.0	0	0	21-08-2013	26	...	4	0	0	0	0
3	3	6182	1984	S1	Bertunangan	26646000.0	1	0	10-02-2014	26	...	6	0	0	0	0
4	4	5324	1981	S3	Menikah	58293000.0	1	0	19-01-2014	94	...	5	0	0	0	0

- **Deskripsi Dataset:** Menggunakan dataset marketing\_campaign\_data.csv yang berisi:
  - **2,240 baris** data pelanggan.
  - **29 fitur** mencakup demografi, perilaku pembelian, dan respons kampanye.
- **Kesesuaian Data:** Dataset ini sangat relevan karena mengandung semua variabel kunci yang dibutuhkan untuk memahami profil dan perilaku pelanggan secara 360 derajat.

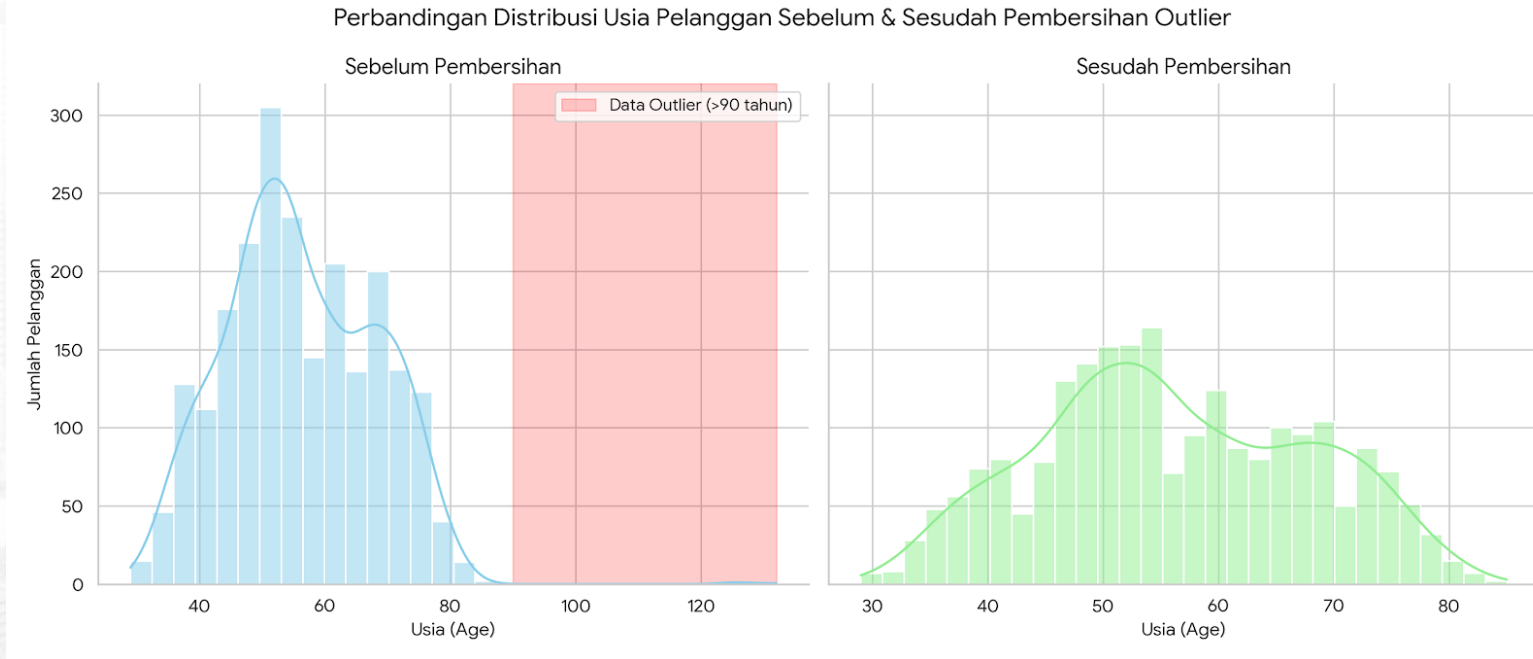
[Untuk selengkapnya, dapat melihat jupyter notebook disini](#)





- **Proses Pembersihan:**

- **Penanganan Missing Values:** Mengisi 24 data Income yang kosong dengan nilai median untuk menghindari bias



- **Proses Pembersihan:**

- **Penanganan Outliers:** Menghapus data pelanggan dengan usia tidak wajar ( $>90$  tahun) dan pendapatan ekstrim ( $>600$  juta) untuk menjaga kestabilan model.

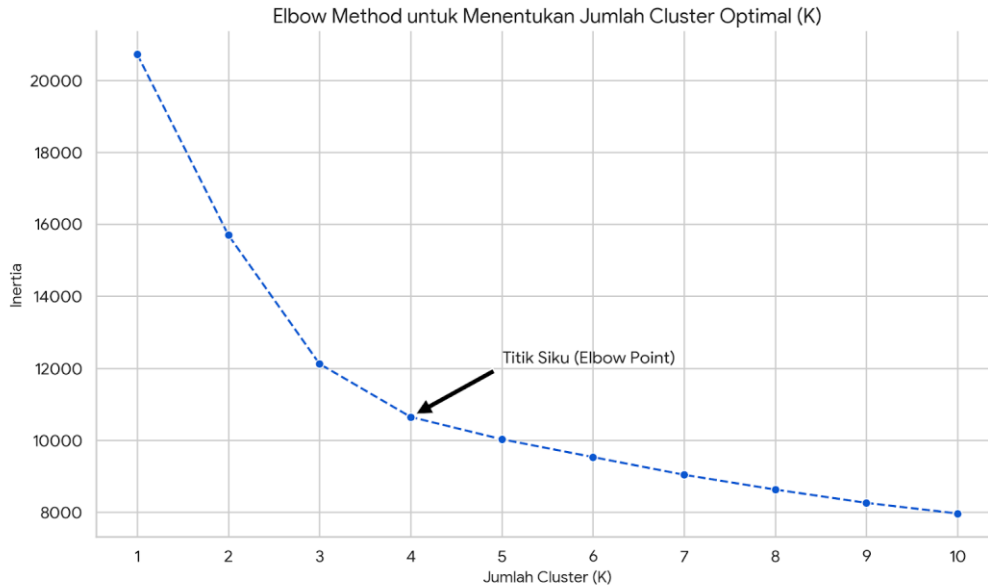
[Untuk selengkapnya, dapat melihat jupyter notebook disini](#)



- **Tujuan:** Mengubah data mentah menjadi fitur yang lebih bermakna untuk model.
- **Fitur Baru yang Dibuat:**
  - Age: Usia pelanggan (lebih intuitif dari tahun lahir).
  - Total\_Children: Total anak (menggabungkan Kidhome & Teenhome).
  - Total\_Spending: Total pengeluaran (menjumlahkan semua produk ).
  - Customer\_Lifetime: Lama menjadi pelanggan.
- **Alasan:** Fitur-fitur ini secara langsung merepresentasikan **nilai demografis** dan **nilai perilaku** pelanggan, yang merupakan inti dari segmentasi.

- **Model yang Dipilih: K-Means Clustering.**
- **Alasan Pemilihan:**
  - Merupakan algoritma *unsupervised learning* yang paling populer dan efisien untuk masalah segmentasi.
  - Mudah diinterpretasikan dan dijelaskan kepada stakeholder bisnis.
  - Sangat cocok untuk menemukan kelompok-kelompok berbeda dalam data tanpa label awal.
- **Proses Penentuan Jumlah Cluster:** Menggunakan **Elbow Method** untuk menemukan jumlah segmen (K) yang paling optimal.

# Model Evaluation (Elbow Method)

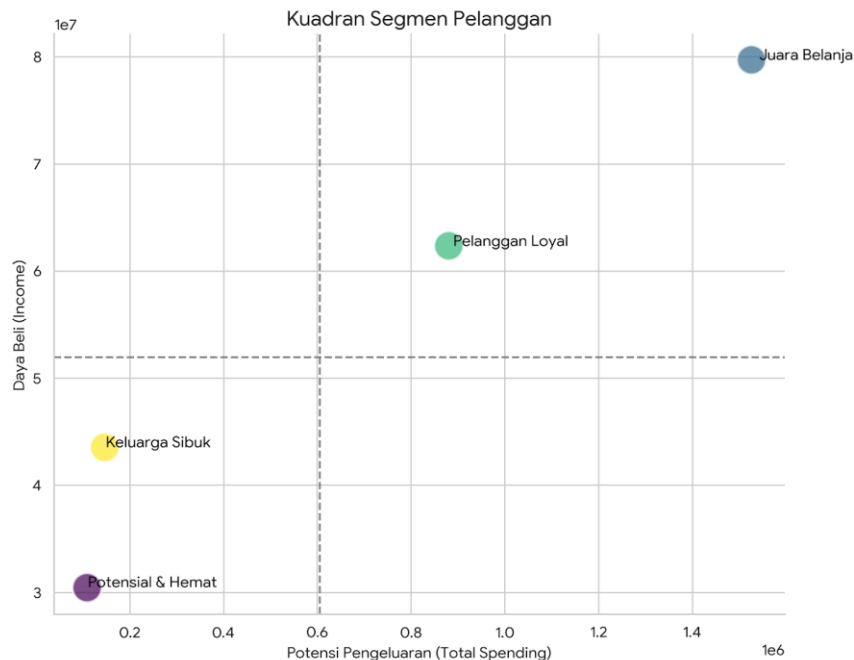


- **Metode:** Menjalankan K-Means untuk  $K=1$  hingga 10 dan memplot nilai *inertia*-nya.
- **Hasil:** Grafik menunjukkan "siku" yang jelas pada **K=4**, menandakan bahwa 4 adalah jumlah segmen yang paling efisien dan informatif.

Perbandingan Karakteristik Antar Segmen Pelanggan



- **Insight Utama:** Pelanggan kita tidak homogen. Mereka terbagi menjadi 4 kelompok dengan kebutuhan dan perilaku yang sangat berbeda.
- **Persona yang Ditemukan:**
  - 🏆 Juara Belanja (High-Spenders)
  - 🛒 Pelanggan Loyal & Mapan
  - 👨‍👩‍👧 Keluarga Sibuk
  - 💰 Potensial & Hemat



- **Rekomendasi per Segmen:**

- **Juara Belanja:** Berikan perlakuan VIP, akses produk eksklusif, dan layanan premium.
- **Pelanggan Loyal:** Fokus pada program retensi, bangun komunitas, dan hargai kesetiaan mereka.
- **Keluarga Sibuk:** Tawarkan paket keluarga, kemudahan belanja online, dan produk relevan.
- **Potensial & Hemat:** Ajak dengan promosi diskon, program referral, dan penawaran *value-for-money*.



- **Dampak pada Industri:** Mengadopsi segmentasi berbasis data memungkinkan perusahaan beralih dari *product-centric* menjadi **customer-centric**.
- **Peningkatan Proses:** Solusi ini memungkinkan otomatisasi personalisasi skala besar, sebuah keunggulan kompetitif yang krusial di era digital saat ini. Perusahaan dapat merespons perubahan pasar dengan lebih cepat dan akurat.

- **Tantangan:**

- Data Income yang tidak lengkap memerlukan imputasi yang cermat.
- Menginterpretasikan hasil clustering menjadi persona yang bermakna membutuhkan kombinasi analisis kuantitatif dan pemahaman bisnis.

- **Keterbatasan:**

- **Segmentasi Statis:** Model ini adalah potret sesaat. Perilaku pelanggan dapat berubah seiring waktu.
- **Keterbatasan Fitur:** Analisis belum mencakup data preferensi produk secara spesifik.



- **Ringkasan Hasil:** Proyek ini berhasil mengubah data mentah menjadi 4 persona pelanggan yang dapat ditindaklanjuti, memberikan peta jalan yang jelas untuk strategi pemasaran.
- **Pesan Utama:** Dengan memahami siapa pelanggan kita, kita dapat berhenti berteriak ke kerumunan dan mulai membangun percakapan yang bermakna. **Personalisasi bukan lagi pilihan, melainkan keharusan.**