

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



**Created by:**

**Theofilus Arifin**

theofilus.dev@gmail.com

[.linkedin.com/in/theofilusarifin](https://www.linkedin.com/in/theofilusarifin)

Results-driven data science and artificial intelligence graduate with a degree in Computer Science, demonstrating a solid foundation in programming and statistical concepts. Passionate about leveraging machine learning techniques to solve business problems within the e-commerce industry. Proficient in SQL for efficient data manipulation and analysis. Highly skilled in using Python for data cleaning, preprocessing, statistical analysis, and implementing machine learning and deep learning algorithms.

Supported by:  
**Rakamin Academy**  
Career Acceleration School  
[www.rakamin.com](http://www.rakamin.com)

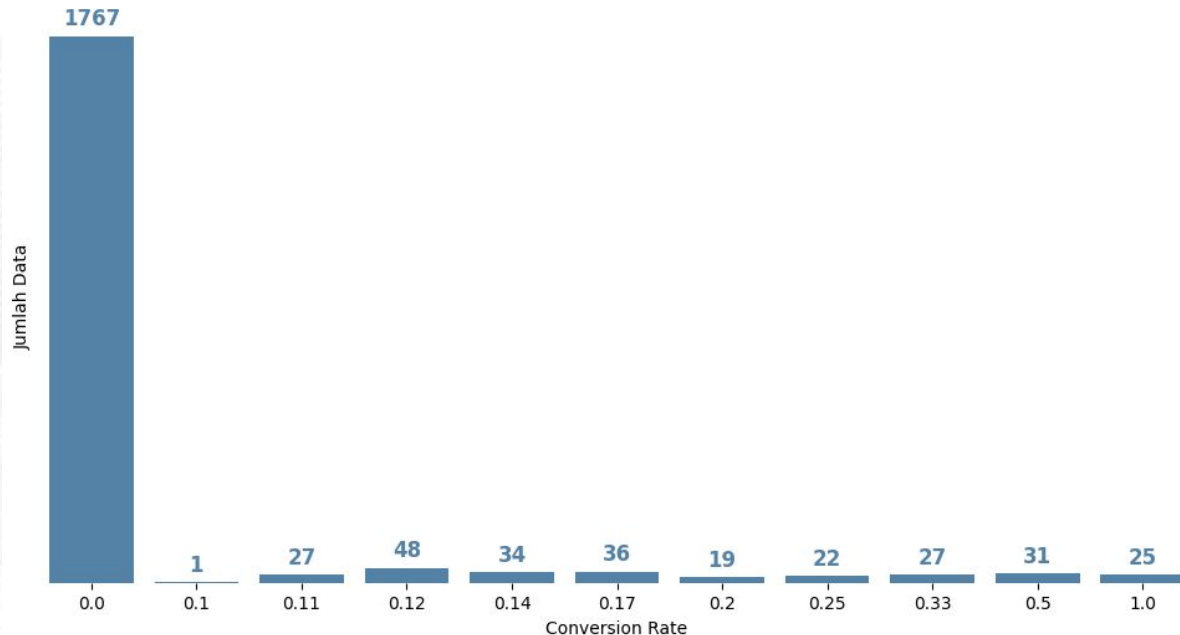
“Sebuah perusahaan dapat berkembang dengan pesat saat mengetahui perilaku customer personality nya, sehingga dapat memberikan layanan serta manfaat lebih baik kepada customers yang berpotensi menjadi loyal customers. Dengan mengolah data historical marketing campaign guna menaikkan performa dan menyasar customers yang tepat agar dapat bertransaksi di platform perusahaan, dari insight data tersebut fokus kita adalah membuat sebuah model prediksi kluster sehingga memudahkan perusahaan dalam membuat keputusan ”

Fitur	Penjelasan
Customer Age	Umur pelanggan yang diekstraksi dari fitur 'Year_Birth'.
Age Group	Pembagian usia pelanggan berdasarkan kategori umur yang didasarkan pada data dari sumber resmi.
First Purchase Age	Umur pelanggan saat pertama kali mendaftar sebagai pelanggan.
Total Children	Jumlah anak yang dimiliki oleh pelanggan.
Membership Duration	Jumlah tahun pelanggan telah menjadi anggota.

Fitur	Penjelasan
Total Transaction	Total jumlah transaksi yang dilakukan oleh pelanggan, termasuk pembelian paket, web, katalog, dan di toko.
Total Spending	Total pengeluaran pelanggan pada berbagai kategori produk, termasuk Coke, buah-buahan, produk daging, produk ikan, produk manis, dan produk emas.
Total Accepted Campaign	Jumlah total kampanye pemasaran yang diterima oleh pelanggan, dihitung berdasarkan penerimaan terhadap lima kampanye pemasaran yang berbeda (AcceptedCmp1 hingga AcceptedCmp5).
Conversion Rate	Tingkat konversi dari respons pelanggan dibandingkan dengan jumlah kunjungan bulanan ke situs web.

## Distribusi Tingkat Konversi Pembelian pada Website

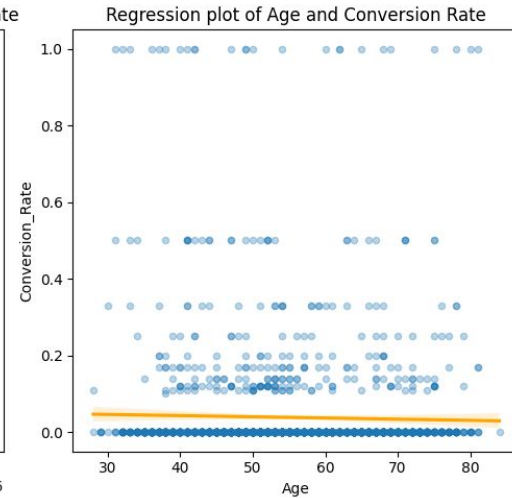
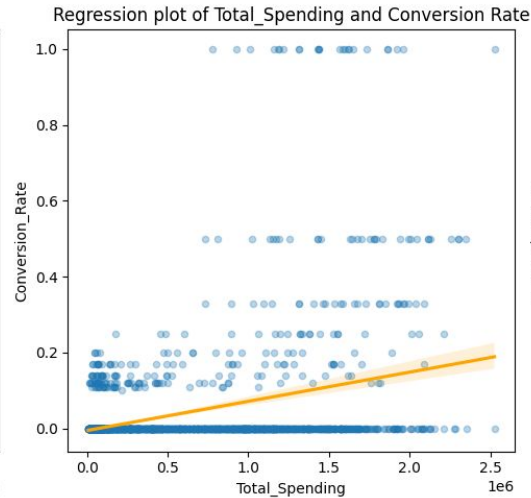
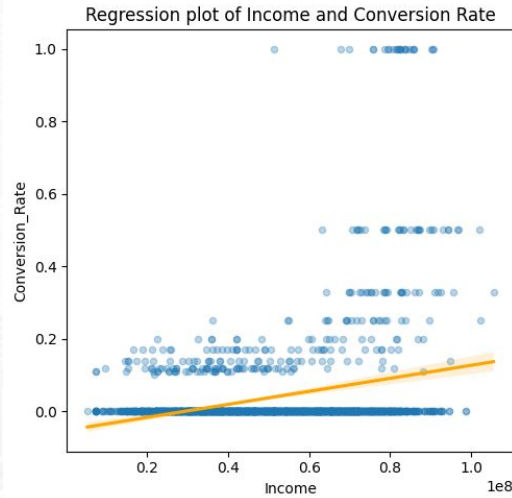
Terdapat sejumlah transaksi dengan tingkat konversi tinggi (0.11 hingga 1.00) menunjukkan peluang untuk meningkatkan efisiensi dengan menganalisis faktor-faktor pembelian



Meskipun banyaknya conversion rate yang mencapai 0%, adanya sejumlah transaksi dengan tingkat konversi tinggi (0.11 hingga 1.00) menunjukkan potensi untuk meningkatkan efisiensi melalui analisis faktor-faktor pembelian.



# Conversion Rate Analysis Based on Income, Spending and Age



Fitur 'Income' dan 'Total Spending' menunjukkan korelasi positif yang signifikan dengan 'Conversion Rate', menandakan bahwa pelanggan dengan pendapatan dan pengeluaran yang lebih tinggi cenderung memiliki tingkat konversi yang lebih tinggi. Sebaliknya, tidak ada korelasi yang signifikan antara fitur 'Age' dengan 'Conversion Rate', menunjukkan bahwa usia tidak mempengaruhi tingkat konversi pelanggan secara signifikan.

## 1. Null Values

- Terdapat 1 fitur yang memiliki missing value yaitu `Income`. Handling missing value dilakukan dengan melakukan imputasi nilai menggunakan nilai median.

## 2. Duplicated Value

- Tidak terdapat data duplicate pada dataset.

## 3. Handling Outliers

- Terdapat 203 data outlier (9%) yang dihapus dari dataset

## 4. Feature Encoding

- Education

Tipe Kategorikal : Data Ordinal

Metode Encoding: Label Encoding

SMA=0, dan D3=1, S1=2, S2=3, S3=4.

- Marital Status

Tipe Kategorikal : Data Ordinal

Metode Encoding: Binary Encoding

- Age Group

Tipe Kategorikal : Data Ordinal

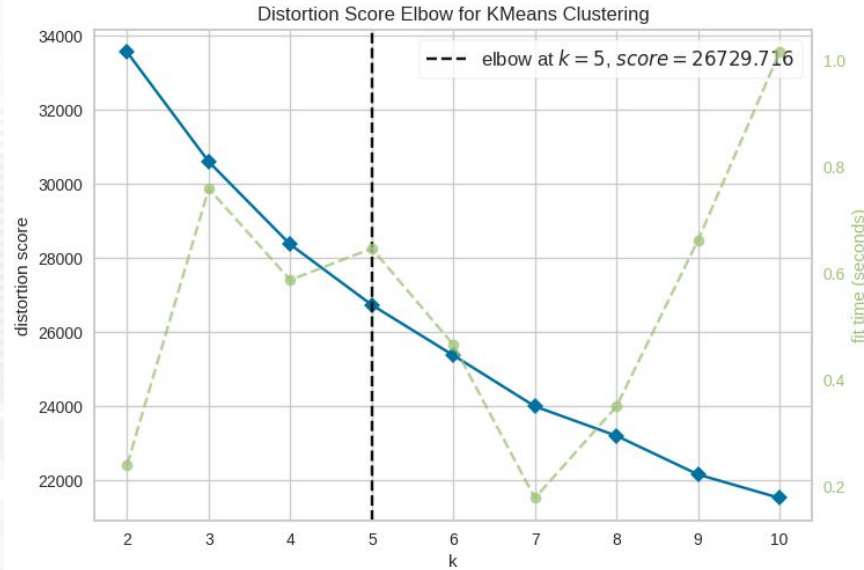
Metode Encoding: Label Encoding

seniors=0, middle-aged adults=1, young adults=2.

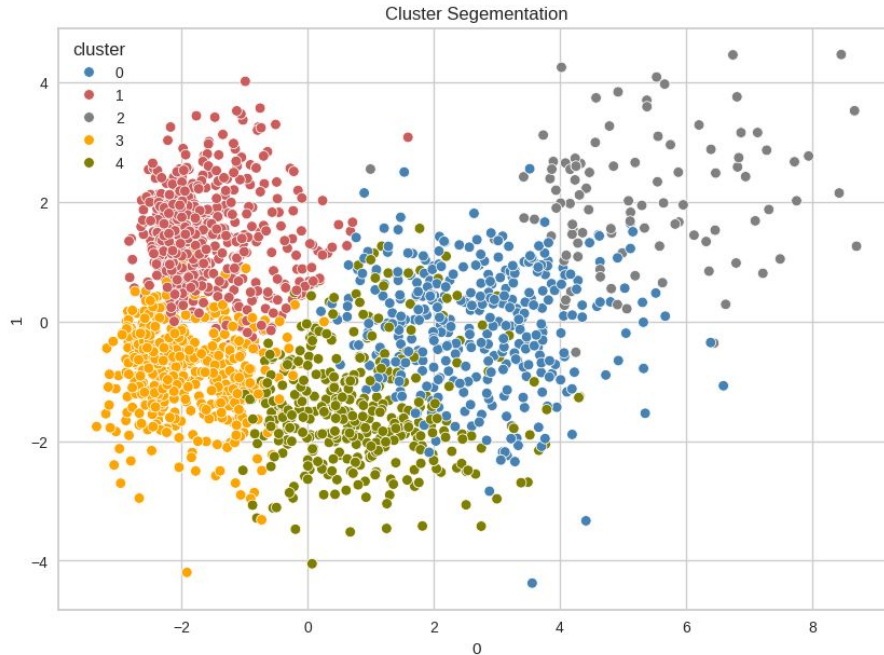


## 5. Feature Standardization

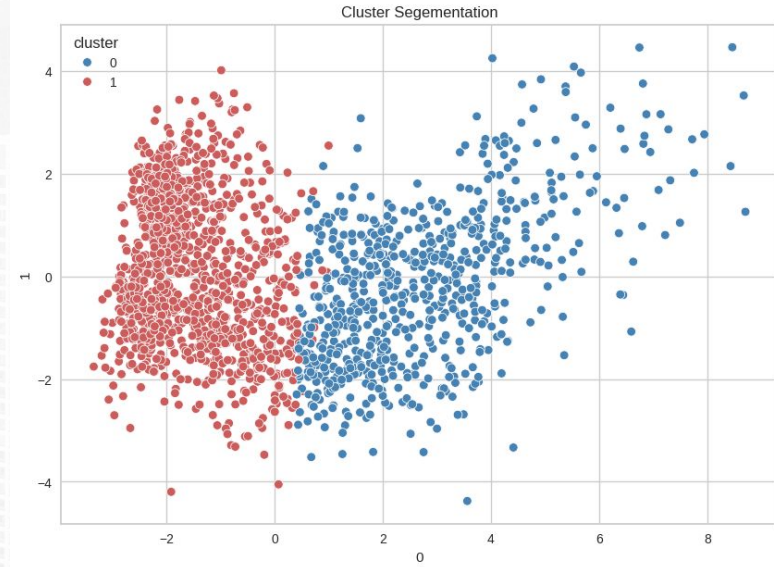
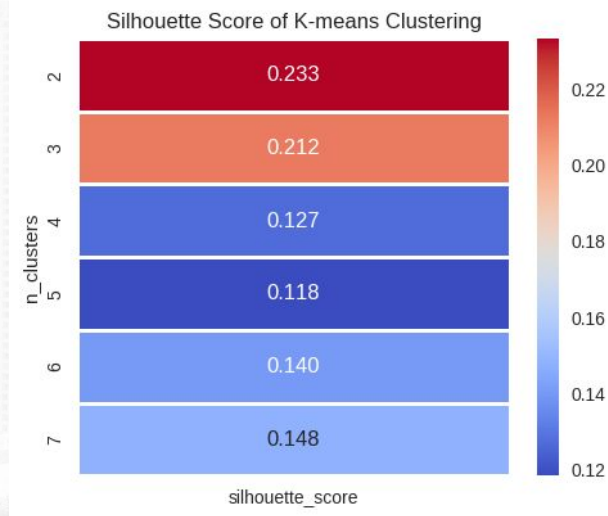
Feature standardization dilakukan menggunakan standard scaller untuk mengubah skala data sedemikian rupa sehingga setiap fitur memiliki rata-rata nol dan varians satu



Berdasarkan Distortion Score dan Elbow method didapatkan jumlah cluster terbaik adalah 5

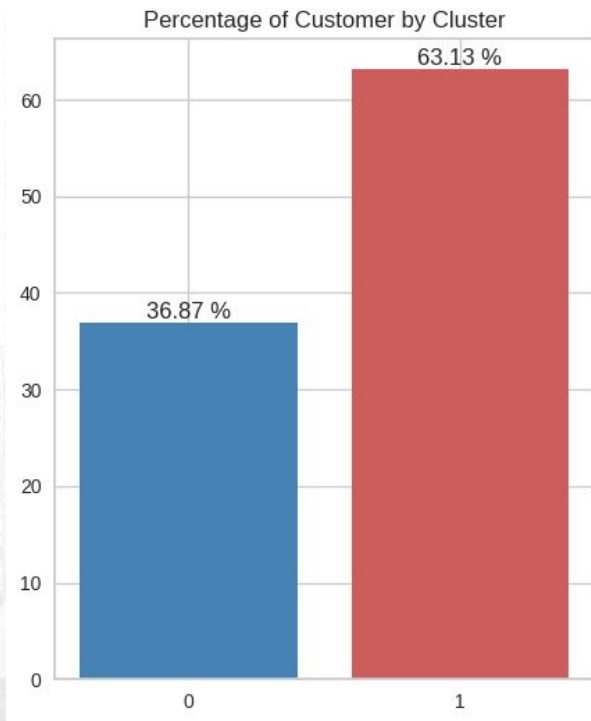


- Visualisasi Cluster dilakukan dengan PC 1 dan 2 yang mewakili varians terbanyak dari Principal Components pada PCA
- Dari hasil clustering, cluster 1 (Merah) dan 3 (Kuning), sudah terlihat cukup terpisah. Namun untuk cluster 0, 2, dan 4 masih belum terlihat terpisah secara baik, evaluasi lebih lanjut perlu dilakukan



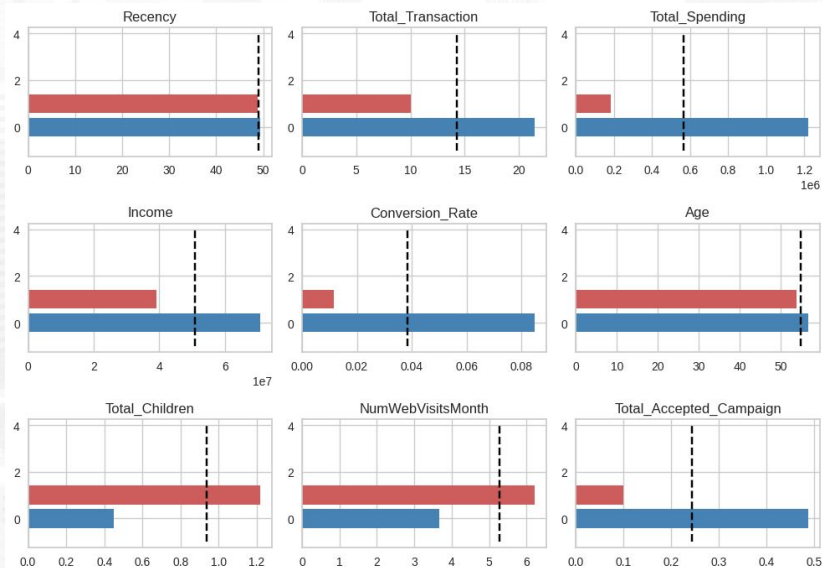
Meskipun elbow method menunjukkan cluster terbaik adalah 5, namun nilai silhouette terbaik ada pada cluster = 2. Berdasarkan visualisasi 2 cluster juga terpisah dengan jelas. Maka dari itu, cluster yang akan dipilih adalah 2





Cluster 0 memiliki jumlah yang lebih sedikit yaitu 36.87% dibandingkan cluster 1 yang memiliki jumlah 63.13%

# Customer Personality Analysis for Marketing Retargeting



visualisasi ini menampilkan bagaimana rata-rata nilai fitur dalam dataset berbeda-beda di antara kelompok pelanggan yang berbeda, sehingga memungkinkan untuk melihat pola atau perbedaan yang mungkin ada di antara kelompok tersebut.

## Cluster

0:

## Pelanggan

## Berkelas

### Key

- 
- 
- 

Personality:

Loyal

Exclusive

Regular

### Deskripsi:

Pelanggan dalam cluster ini cenderung merupakan individu yang mapan secara finansial, dengan usia rata-rata 57 tahun. Mereka memiliki pendapatan yang tinggi dan cenderung melakukan transaksi dengan nilai yang signifikan (rata-rata 21 transaksi dengan total pengeluaran rata-rata sebesar 1.220.860). Meskipun mereka tidak menerima kampanye secara keseluruhan, kemungkinan besar mereka adalah pelanggan setia yang telah lama berlangganan dan memiliki preferensi produk atau layanan tertentu.

Cluster	1:	Pelanggan	Berpotensi
Key			Personality: Exploratory Potential Online-centric
-			
-			
-			

## Deskripsi:

Pelanggan dalam cluster ini lebih muda dengan usia rata-rata 54 tahun dan cenderung memiliki pendapatan yang lebih rendah. Meskipun memiliki pendapatan yang lebih rendah, mereka memiliki tingkat kunjungan situs web yang lebih tinggi (rata-rata 6 kunjungan per bulan). Namun, mereka cenderung melakukan transaksi dengan nilai yang lebih rendah (rata-rata 10 transaksi dengan total pengeluaran rata-rata sebesar 184.922). Meskipun mereka tidak menerima kampanye secara keseluruhan, mereka memiliki potensi untuk lebih terlibat dalam program pemasaran jika disasar dengan strategi yang tepat.



## Pelanggan Berkelas (Cluster 0)

- Fokus pada pelayanan dan pengalaman pelanggan yang eksklusif untuk mempertahankan loyalitas pelanggan yang sudah mapan.
- Gunakan strategi pemasaran yang menekankan kualitas, eksklusivitas, dan penghargaan kesetiaan untuk memperkuat hubungan dengan pelanggan.
- Pertimbangkan program insentif atau keanggotaan eksklusif untuk meningkatkan keterlibatan pelanggan yang sudah setia.

## Pelanggan Berpotensi (Cluster 1)

- Tingkatkan visibilitas merek secara online dan berikan pengalaman belanja yang menyenangkan dan mudah diakses.
- Gunakan strategi pemasaran yang menarik dan relevan untuk menarik perhatian pelanggan potensial dan mendorong konversi.
- Pertimbangkan untuk menawarkan diskon atau insentif khusus untuk mendorong transaksi dan meningkatkan keterlibatan pelanggan baru.