

Predict Customer Segmentation to boost marketing campaign by using Machine Learning

Supported by:
Rakamin Academy
Career Acceleration School
www.rakamin.com



Created by:
Stephen James
ign.james4@gmail.com
[https:// www.linkedin.com/ in/ stephenjames23/](https://www.linkedin.com/in/stephenjames23/)
[https:// github.com/ ignjames](https://github.com/ignjames)

“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 ”

Goal yang ingin dicapai ialah melakukan **segmentasi** customers dan melakukan **analisa** terhadap masing-masing cluster. Untuk mencapai hal tersebut, proses pengerjaan dilakukan dengan langkah-langkah berikut:

- Feature Engineering : Melakukan penambahan fitur (feature engineering) yang mewakili beberapa fitur
- EDA: Mencari insight dari hubungan antar beberapa kolom hasil feature engineering dengan conversion rate.
- Cleaning Data: Membersihkan data dari missing values.
- Feature Encoding: Mengubah fitur kategorikal menjadi fitur numerikal agar dapat dimodelling.
- Feature Transformation: Melakukan penyamaan skala untuk tiap fitur.
- Data Modelling: Melakukan clustering dengan k means.
- Analysis: menganalisa hasil clustering.

```
df1 = df1.loc[df1['NumWebVisitsMonth']!=0]
```

→ **Syarat:** Webiste visit tidak boleh 0, karena campaign dilakukan pada website

- Conversion Rate

Menghitung **persentase** jumlah respon customer yang menerima campaign di web dengan perbandingan terhadap jumlah web visit customer pada bulan tersebut.

```
df1['conv_rate'] = (df1['Response'] / df1['NumWebVisitsMonth'])*100
```

- Umur

Selisih waktu antara tahun lahir dengan waktu customer mendaftar

```
df1['Umur'] = round(((df1['Dt_Customer'] - df1['Year_Birth']).dt.days)/365,0)
```

- Pengelompokkan Umur

Pengelompokkan umur menjadi:

15-24 tahun: Kelompok usia muda
25-34 tahun: Kelompok usia pekerja awal
35-44 tahun: Kelompok usia paruh baya
45-54 tahun: Kelompok usia pra-pensiun
55-64 tahun: Kelompok usia pensiun
65 tahun ke atas: Kelompok usia lanjut

```
def kelompok_usia(i):  
    if i < 25:  
        return 'Muda'  
    elif i < 35:  
        return 'Pekerja Awal'  
    elif i < 45:  
        return 'Paruh Baya'  
    elif i < 55:  
        return 'Pra Pensiun'  
    elif i < 65:  
        return 'Pensiun'  
    else:  
        return 'Lanjut'
```

```
df1['kelompok_usia'] = df1.Umur.apply(kelompok_usia)
```

- Jumlah Anak

Menambahkan jumlah kid dan teen sebagai fitur jumlah anak.

```
df1['jumlah_anak'] = df1['Kidhome'] + df1['Teenhome']
```

- Total Pembelian

Total pembelian customer terhadap berbagai kategori barang.

```
df1['total_pembelian'] = df1['MntCoke'] + df1['MntFruits'] + df1['MntMeatProducts']  
+ df1['MntFishProducts'] + df1['MntSweetProducts'] + df1['MntGoldProds']
```

- Total Transaksi

Berapa kali customer sudah bertransaksi.

```
df1['total_transaksi'] = df1['NumDealsPurchases'] + df1['NumWebPurchases'] + df1['NumCatalogPurchases'] + df1['NumStorePurchases']
```

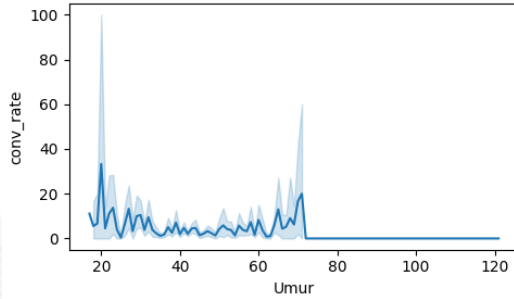
- Rata-Rata Pengeluaran per Transaksi

```
df1['rata2_pt'] = df1['total_pembelian'] / df1['total_transaksi']
```

- Total Accepted Campaign

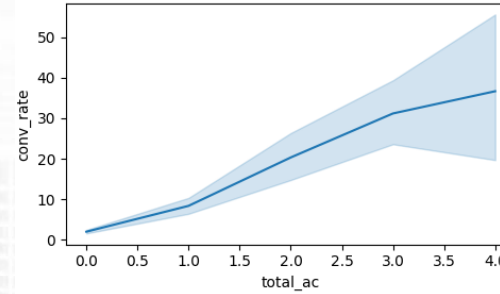
Berapa campaign yang customer tersebut terima.

```
df1['total_ac'] = df1['AcceptedCmp1'] + df1['AcceptedCmp2'] + df1['AcceptedCmp3'] + df1['AcceptedCmp4'] + df1['AcceptedCmp5']
```



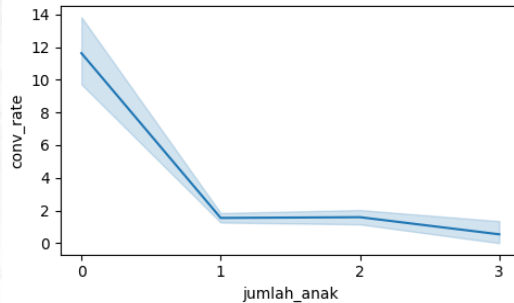
Umur

- Dapat dilihat bahwa tidak ada korelasi antara pertambahan umur dengan kenaikan conversion rate.



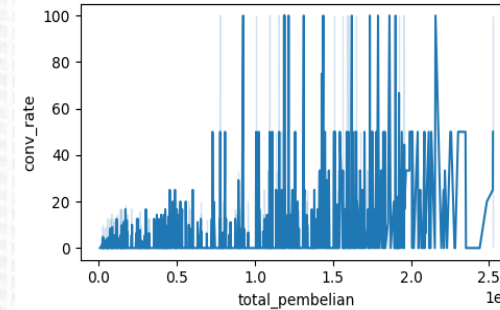
Total Accepted Campaign

- Conversion rate semakin tinggi seiring dengan banyaknya jumlah campaign yang diterima sebelumnya.



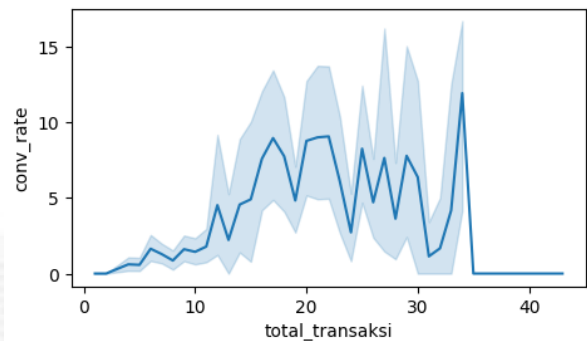
Jumlah Anak

- Jumlah anak yang sedikit menunjukkan adanya conversion rate yang lebih besar



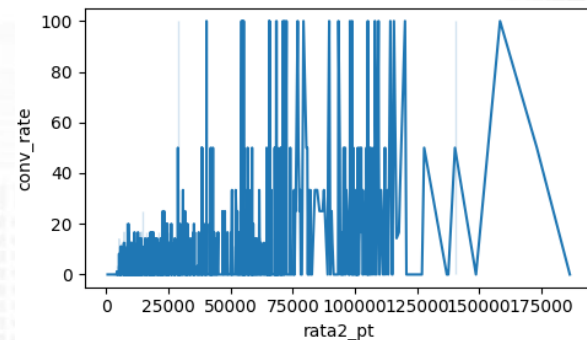
Total Pembelian

- Tidak terdapat korelasi diantara total pembelian dengan conversion rate

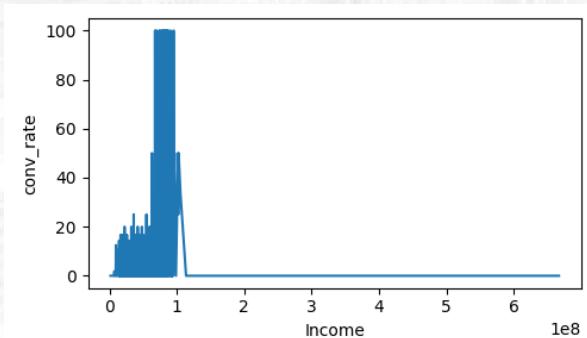


Total Transaksi

- Tidak terdapat korelasi antara kenaikan conversion rate dengan kenaikan jumlah transaksi

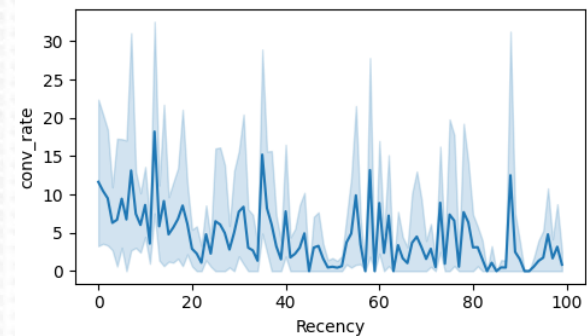
Rata-Rata
Pembelian
per Transaksi

- Tidak terdapat korelasi antara conversion rate dan rata-rata pembelian per transaksi



Income

- Dapat dilihat bahwa conversion rate tinggi disekitar income 100.000.000.



Recency

- Meskipun tidak merata tetapi, dengan peningkatan recency terjadi penurunan tren conversion rate

- Cek Duplikat

Berdasarkan ID customer tidak terdapat adanya duplikat.

```
#cek duplicate  
df2.duplicated(subset=['ID']).sum()
```

- Cek Missing Values

Terdapat missing values pada kolom 'Income' dengan persentase sebesar 1.03%

```
(df3.isna().mean()[df3.isna().mean()>0])*100
```

- Handle Missing Values

Berdasarkan persebaran yang right skewed maka missing values akan diisi dengan **median**.

```
df3['Income'].fillna(df3['Income'].median(),inplace=True)
```


- Marital Status

Menyederhanakan marital status menjadi single dan menikah, lalu melakukan **one hot encoding**.

```
#marital_status
ms = {
    new_key : new_val
    for keys, new_val in [(['Lajang', 'Ceraai', 'Janda', 'Duda', 'Bertunangan'], 'Single')]
    for new_key in keys
}
df3['Marital_Status'] = df3['Marital_Status'].replace(ms)

df3 = pd.get_dummies(df3, columns=['Marital_Status'])
```

- Education

Membuang kolom education karena mayoritas menyelesaikan kuliah dan hanya sedikit yang SMA.

```
#education
df3 = df3.drop(columns=['Education'], axis=1)
```

- Kelompok Usia

Melakukan **one hot encoding** pada kelompok usia

```
#kelompok_usia
df3 = pd.get_dummies(df3, columns=['kelompok_usia'])
```

- Standarisasi

Melakukan **penyamaan skala** untuk semua kolom.

```
#standarisasi

from sklearn.preprocessing import StandardScaler
ss = StandardScaler()

for col in df3:
    df3[col] = ss.fit_transform(df3[col].values.reshape(-1,1))
```

df3.head(3)

| | Income | Recency | NumWebVisitsMonth | Complain | Response | conv_rate | Umur | jumlah_anak | total_pembelian | total_transaksi | rata2_pt | total_ac | Marital |
|---|-----------|-----------|-------------------|-----------|-----------|-----------|----------|-------------|-----------------|-----------------|-----------|----------|---------|
| 0 | 0.247921 | 0.308550 | 0.696627 | -0.097568 | 2.380686 | 0.693198 | 0.939283 | -1.270926 | 1.487379 | 1.331090 | 0.984512 | -0.43999 | |
| 1 | -0.231566 | -0.382530 | -0.140152 | -0.097568 | -0.420047 | -0.305755 | 1.272356 | 1.392935 | -0.918979 | -1.162376 | -0.932149 | -0.43999 | |
| 2 | 0.795750 | -0.797178 | -0.558541 | -0.097568 | -0.420047 | -0.305755 | 0.356405 | -1.270926 | 0.204373 | 0.806150 | 0.097882 | -0.43999 | |

Clustering Customer

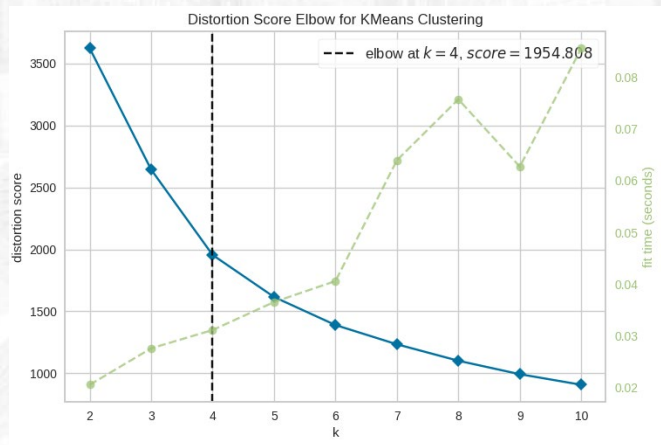
- Melakukan segmentasi berdasarkan **Recency, Frequency, dan Monetary**

Recency = Kapan terakhir kali customer melakukan purchase, diwakilkan oleh kolom 'Recency'.

Frequency = Seberapa sering customer berbelanja, diwakilkan oleh kolom 'total_transaksi'.

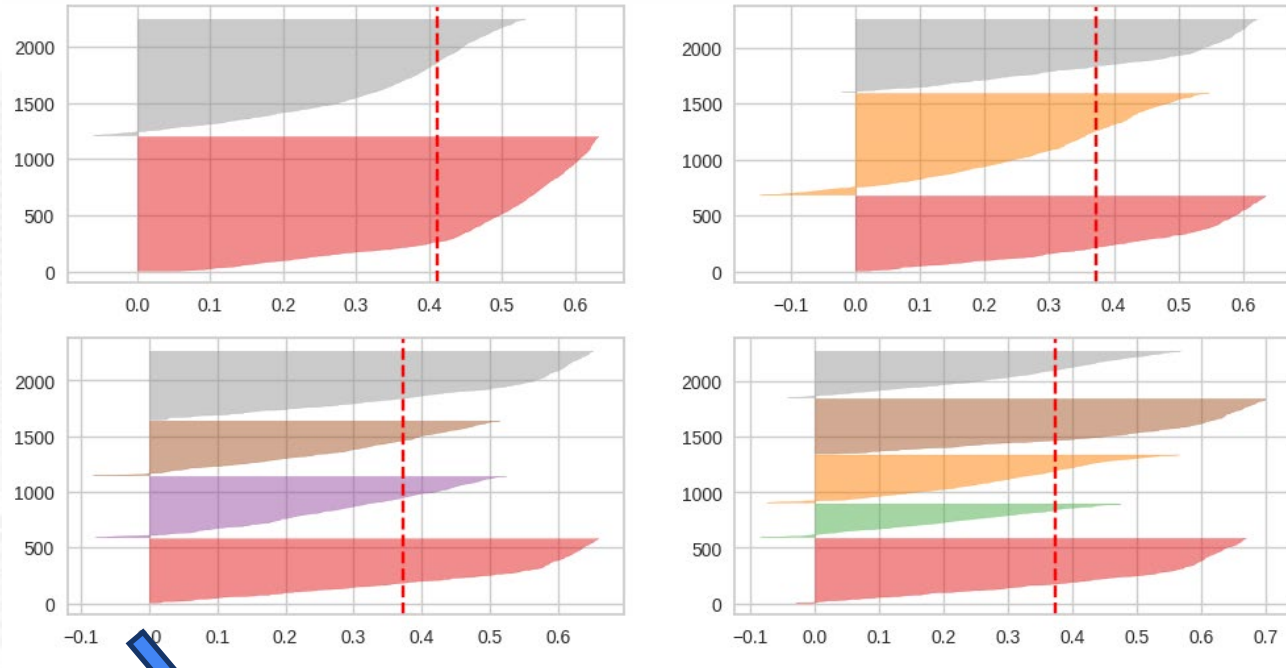
Monetary = Berapa banyak customer berbelanja, diwakilkan oleh kolom 'total_pembelian'.

- Menentukan jumlah cluster dengan Elbow Method



Berdasarkan elbow method, jumlah cluster yang paling optimal adalah 4.

- Evaluasi hasil Elbow Method dengan silhouette score



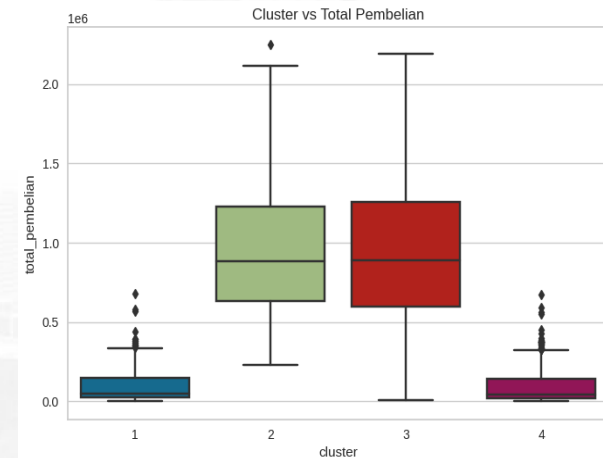
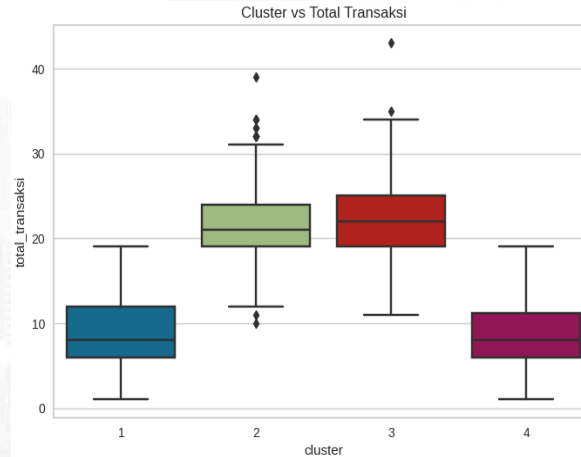
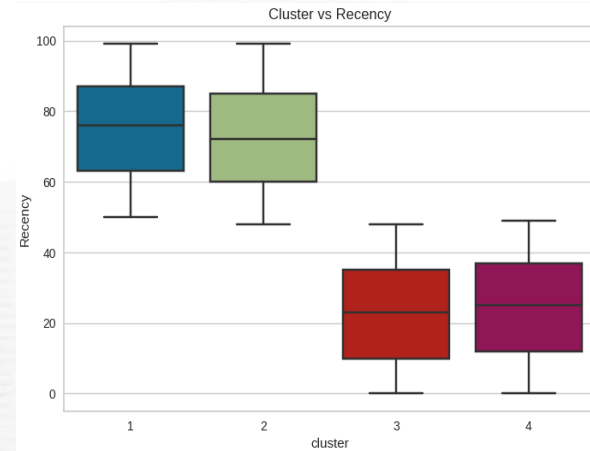
Cluster paling optimal, $K = 4$

Hasil evaluasi silhouette score menunjukkan 4 cluster merupakan yang paling optimal.

Hal tersebut ditunjukkan dengan setiap plot memiliki ketebalan yang hampir sama.

Customer Segmentation Analysis for Marketing Retargeting

● Segmentasi konsumen



Kesimpulan:

- Cluster 1: Pembeli yang sudah lama tidak belanja, transaksi dilakukan sedikit, pembelian sedikit. Pada cluster ini, terdapat customer yang merasa **tidak cocok** dengan produk dan layanan kita sehingga memutuskan untuk tidak berbelanja kembali.
- Cluster 2: Pembeli yang sudah lama tidak belanja, tetapi transaksi dan pembelian yang dilakukan banyak. Hal ini terlihat seperti loyal customer yang akan **churn**.
- Cluster 3: Pembeli yang baru berbelanja, dengan transaksi dan pembelian yang dilakukan banyak. Ini merupakan golongan **loyal** customer.
- Cluster 4: Pembeli yang baru berbelanja, tetapi transaksi dan pembelian yang dilakukan sedikit. Cluster ini terdiri dari customer - customer **baru**.

- Rekomendasi Bisnis

- Cluster 1: Bisa **meningkatkan kualitas layanan dan kualitas produk** terlebih dahulu agar customer pada segmentasi ini bisa kembali berbelanja. Lalu dapat disertai dengan mengadakan promo promo menarik agar mendapat perhatian dari customer cluster 1 ini.
- Cluster 2: Segmentasi ini membutuhkan perhatian khusus karena perusahaan dapat berpotensi kehilangan loyal customer. Hal ini dapat dilakukan dengan melihat **buying habitnya** dan membuat campaign-campaign yang menarik dan sesuai untuk meyakinkan konsumen pada cluster ini untuk kembali.
- Cluster 3: Perusahaan perlu terus memonitor customer di segmentasi ini untuk keberlanjutan perusahaan. Dengan cara **meningkatkan customer experience** seperti membuat loyalty program atau berikan exclusive promo seperti special offers dan discount pada hari special seperti ulang tahun.
- Cluster 4: Untuk customer baru, perusahaan dapat memberikan promo menarik pada **produk-produk unggulan** perusahaan.