

Predict Customer Personality to boost marketing campaign by using Machine Learning

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



Created by:

Muhamad Zamzam Istimaqom

zamzamistimaqom@gmail.com

<https://www.linkedin.com/in/zamzamistimaqom>

Hi. I am Zamzam Istimaqom,

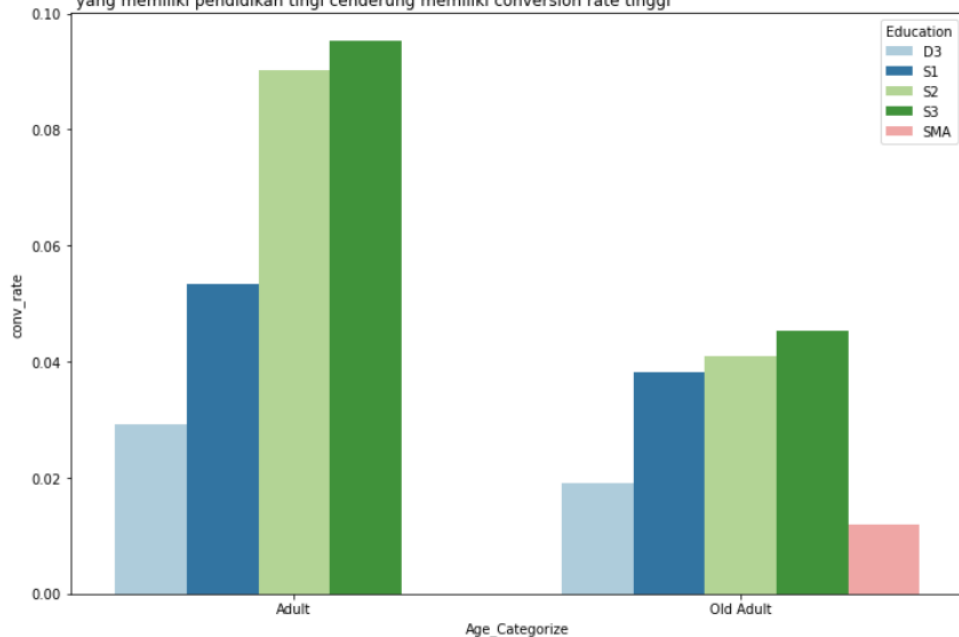
I have the responsibility to process historical marketing campaign data to improve performance and target the right customers so that they can transact on the company's platform. Here I will learn how to process data, clean data, create machine learning models, and draw conclusions and business insights that can be recommended regarding ongoing marketing campaigns.

“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 ”

- Membuat Feature Engineering berupa Conversion rate, umur, pengelompokan kelompok umur, jumlah anak, apakah orangtua, total spending, dan total Transaksi
- Melakukan EDA untuk menarik insight awal untuk dijadikan rekomendasi
- Melihat korelasi antar fitur menggunakan heatmap correlation

Perbedaan tingkat pendidikan terhadap conversion rate berdasarkan kelompok umur

Secara umum, semakin tinggi pendidikan mempengaruhi tingkat conversion rate dalam merespon campaign yang dibuat. Hal ini bisa jadi karena produk fit dengan market fit sehingga terdapat kecenderungan pelanggan yang memiliki pendidikan tinggi cenderung memiliki conversion rate tinggi.



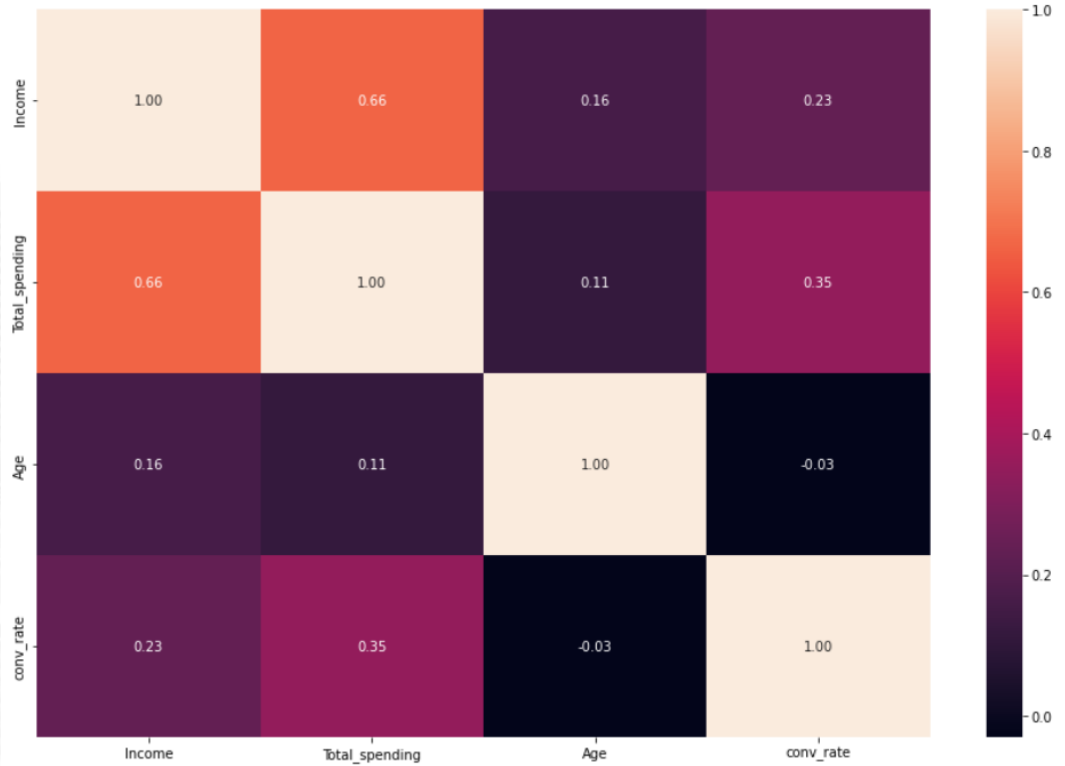
Insight :

Kelompok dewasa (berumur 25 sampai 45)

Memiliki kecenderungan conversion rate tinggi. Daripada yang berumur diatas 45 tahun.

Tingkat Pendidikan memiliki potensi conversion rate tinggi pada masing-masing umur.

Heatmap Correlation (Summary)



Income memiliki hubungan korelasi cukup kuat yaitu 0.66 terhadap Total Spending.

Insight :
Semakin tinggi income pelanggan, maka total spending yang dihabiskan juga semakin besar.

- Pada tahap **cleaning data**, tunjukkan **null** atau **missing value** serta **duplicated value** pada dataset, serta cara penyelesaiannya.
- Selanjutnya untuk data preprocessing, tunjukkan bahwa data sudah dilakukan proses **feature encoding** dan **feature standardisation**.
- **Source code** yang sudah kamu buat, dapat ditampilkan dan berikan link untuk mengakses file tersebut. Contohnya seperti di pojok kanan bawah.



0
Data Duplicate

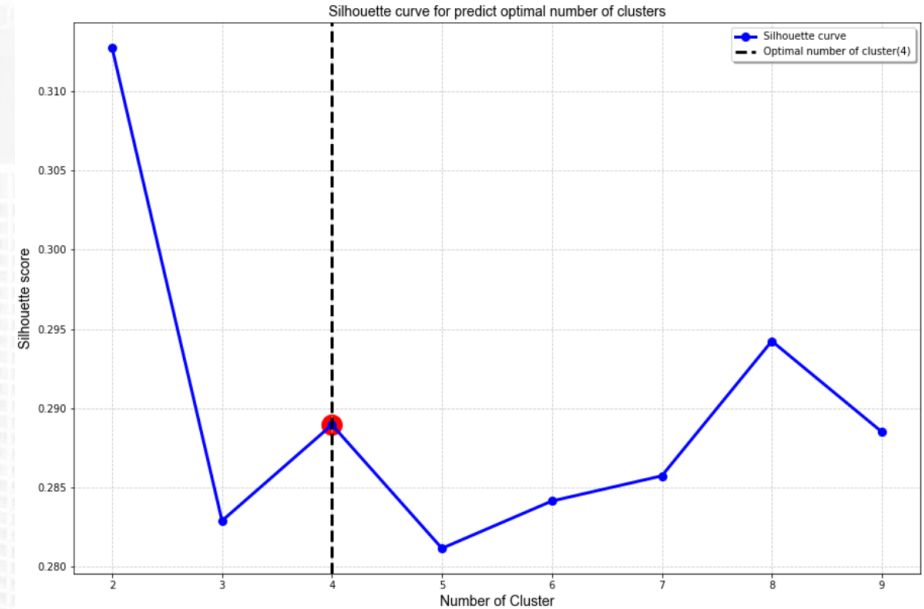
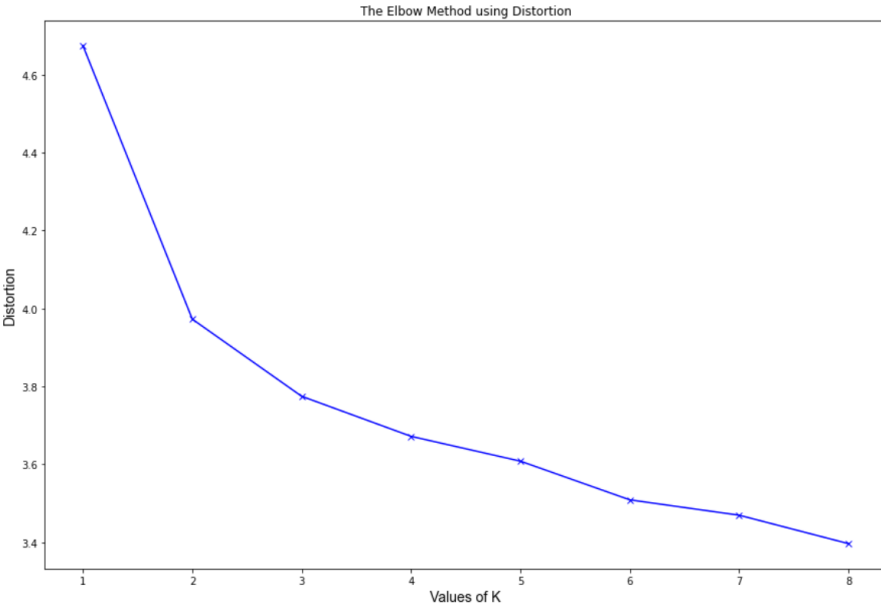
- Missing Value
 - Income (24)
 - Conversion Rate (11)
- Treatment pada kedua datanya handle menggunakan Median dengan alasan data numerikal dan tidak robust terhadap outlier

Feature Encoding and Standarization

	Income	Recency	MntCoke	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProds	NumDealsPurchases	
0	0.235696	0.307039	0.983781	1.551577	1.679702	2.462147	1.476500	0.843207	0.349414	
1	-0.235454	-0.383664	-0.870479	-0.636301	-0.713225	-0.650449	-0.631503	-0.729006	-0.168236	
2	0.773999	-0.798086	0.362723	0.570804	-0.177032	1.345274	-0.146905	-0.038766	-0.685887	
3	-1.022355	-0.798086	-0.870479	-0.560857	-0.651187	-0.503974	-0.583043	-0.748179	-0.168236	
4	0.241888	1.550305	-0.389085	0.419916	-0.216914	0.155164	-0.001525	-0.556446	1.384715	
NumWebPurchases			NumCatalogPurchases		NumStorePurchases	NumWebVisitsMonth	AcceptedCmp3	AcceptedCmp4	AcceptedCmp5	AcceptedCmp1
1.409304			2.510890		-0.550785	0.693904	-0.28014	-0.283830	-0.28014	-0.262111
-1.110409			-0.568720		-1.166125	-0.130463	-0.28014	-0.283830	-0.28014	-0.262111
1.409304			-0.226541		1.295237	-0.542647	-0.28014	-0.283830	-0.28014	-0.262111
-0.750450			-0.910898		-0.550785	0.281720	-0.28014	-0.283830	-0.28014	-0.262111
0.329427			0.115638		0.064556	-0.130463	-0.28014	-0.283830	-0.28014	-0.262111

AcceptedCmp2	Complain	Z_CostContact	Z_Revenue	Age	Total_kid	Total_spending	Total_transaction	education_lvl	conv_rate	is_parent	Bertunangan	Cerai	Duda	Janda	Lajang	Menikah	Adult	Old Adult
-0.11651	-0.097282	0.0	0.0	0.985345	-1.264505	1.679417	1.320826	-0.458383	0.142857	0	0	0	0	0	1	0	0	1
-0.11651	-0.097282	0.0	0.0	1.235733	1.396361	-0.961275	-1.154596	-0.458383	0.000000	1	0	0	0	1	0	0	0	1
-0.11651	-0.097282	0.0	0.0	0.317643	-1.264505	0.282673	0.799685	-0.458383	0.000000	0	1	0	0	0	0	0	0	1
-0.11651	-0.097282	0.0	0.0	-1.268149	0.065928	-0.918094	-0.894025	-0.458383	0.000000	1	1	0	0	0	0	0	1	0
-0.11651	-0.097282	0.0	0.0	-1.017761	0.065928	-0.305254	0.539114	1.533425	0.000000	1	0	0	0	0	0	1	1	0

Untuk selengkapnya, dapat melihat jupyter notebook disini



Nilai optimal untuk menentukan jenis cluster berdasarkan kedua plot ada 4 cluster

- Pada elbow method diambil nilai k setelah mengalami pembelokan
- Pada silhouette score diambil nilai maksimal, walaupun terdapat score maksimal pada $n = 8$. tetapi n cluster = 8 dinilai terlalu banyak untuk pengelompokan customer

Evaluation Silhouette Score

Cluster	Silhouette Score
2	0.3127516474151014,
3	0.28289565400020106,
4	0.28895486066913784,
5	0.2811570328004259,
6	0.2841429107017782,
7	0.2857304059651237,
8	0.2885100118187077



OPTIMAL

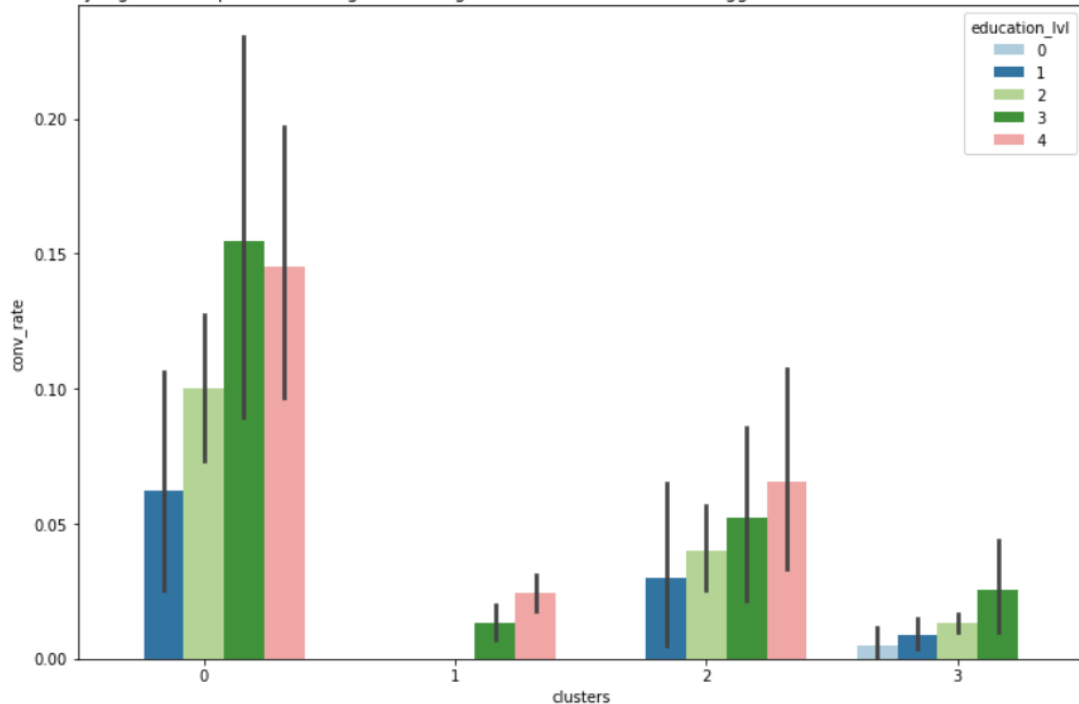
Nilai Silhouette score pada cluster 4
Memiliki score paling tinggi.

Sehingga menggunakan 4 kluster dinilai
Cukup untuk pemilihan cluster pelanggan

Perbedaan tingkat Pendidikan terhadap conversion rate

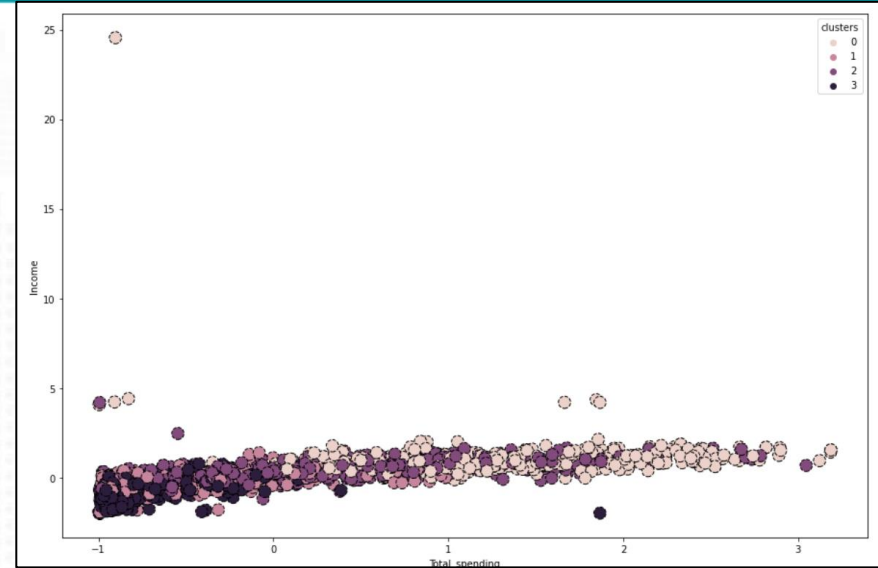
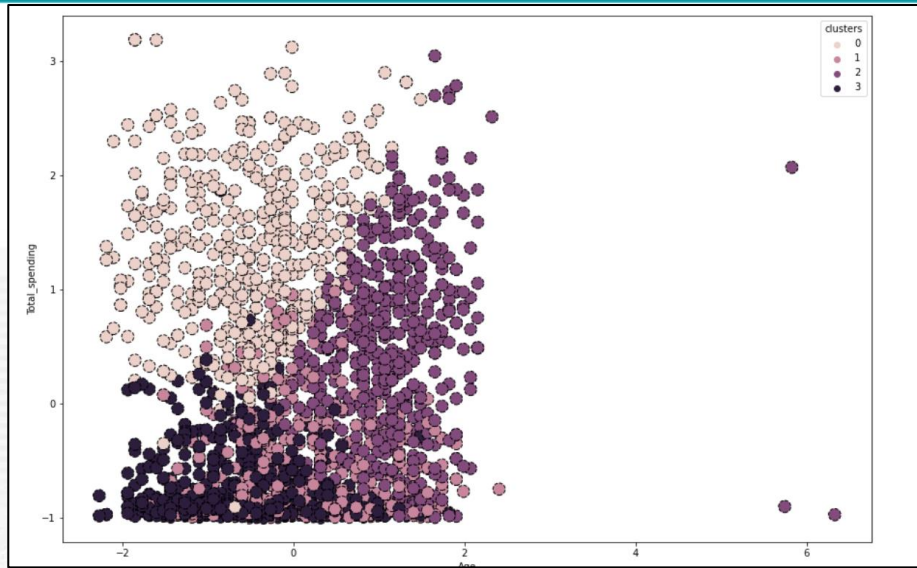
Perbedaan tingkat pendidikan terhadap conversion rate berdasarkan cluster

Secara umum, semakin tinggi pendidikan mempengaruhi tingkat conversion rate dalam merespon campaign yang dibuat. Hal ini bisa jadi karena produk fit dengan market fit sehingga terdapat kecenderungan pelanggan yang memiliki pendidikan tinggi cenderung memiliki conversion rate tinggi.



Untuk selengkapnya, dapat melihat jupyter notebook disini

Clustering



Macam – macam cluster

High Spender : cluster 0

Mid Spender : cluster 2

Low Spender : cluster 1

Risk Churn : cluster 3

Untuk selengkapnya, dapat melihat jupyter notebook disini

Summary Per Cluster

Cluster	Jenis Cluster	Total visit per bulan	Rata -rata conv rate	Jumlah Pelanggan	Total Spending
0	High Spender	1648	0.113252	475	660.567.000
1	Low Spender	2958	0.020568	484	135.307.000
2	Mid Spender	2416	0.045262	540	454.259.000
3	Risk Churn	4887	0.012395	741	106.855.000

Untuk selengkapnya, dapat melihat jupyter notebook disini

1. High spender bisa dikatakan pelanggan setia. Tingkat konversinya memiliki rata-rata paling tinggi sehingga dalam merespon campaign. Pelanggan seperti ini tidak perlu treatment khusus karena justru mereka mencari produk yang mereka inginkan. Yang terpenting adalah menjaga ketersediaan stock agar mereka tetap berbelanja.
2. Mid Spender memiliki jumlah pelanggan kedua dengan konversi 4%. Pelanggan jenis ini perlu diperlu treatment yang lebih personal seperti promo ulang tahun, sehingga mereka bisa naik ke high spender.
3. Low spender perlu dilakukan retargetting paid. Jumlah pelanggannya melebihi high spender, namun daya belinya kurang. Sehingga perlu dilakukan retargetting pada cluster jenis ini.
4. Risk Churn memiliki jumlah pengunjung paling tinggi, namun dengan spending paling kecil. Rekomendasi untuk cluster jenis ini adalah promo baik potongan harga ataupun voucher ongkos kirim. Selain menghemat biaya promo,