

Data Modeling

Data Scientist Kalbe Nutrisionals

Presented by
Muhammad Zaki



Muhammad Zaki

About Me

Lulusan Teknik Komputer Universitas Telkom tahun 2022.

Memiliki minat terhadap Data dan Pengembangan Machine Learning

Experience

Data Input Clerk

Juli – Agustus 2020

Verifying data and digitizing proof of receipt of financial assistance “Bantuan Langsung Tunai”

Research Assistant

Juli – Agustus 2021

Research Assistant on Web Scraping and Data Analyst Project (PT PROCODECG)

Data Scientist ID/X Partners, Project-Based Virtual Intern

Mai – Juni 2023

Finishing various task related with the activity of Data Scientist from ID/X Partners

Case Study

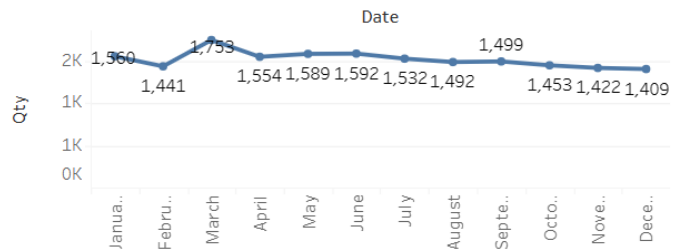
Project yang diminta terbagi 2 yaitu:

Prediksi jumlah penjualan harian. Tujuan dibuatkannya prediksi tersebut agar tim inventory bisa membuat stock persediaan harian yang cukup. Model prediksi akan menggunakan model Time Series Arima.

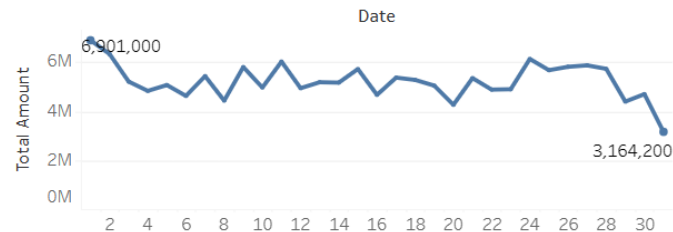
Klasterisasi atau segmentasi konsumen, dengan melihat dari jumlah pembelian dan total belanja konsumen. Hal ini bertujuan untuk memberikan promo yang tepat kepada konsumen. Model klasterisasi yang digunakan yaitu Kmeans.

Visualisasi

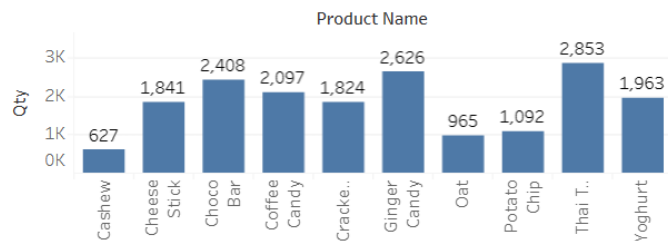
Total Penjualan Perbulan



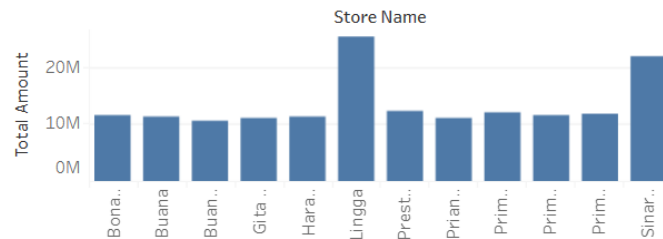
Total Pendapatan Perhari



Total Penjualan Perproduk dalam 1 tahun



Jumlah Penjualan Pertoko dalam 1 Tahun



Visualisasi hubungan antar data dengan menggunakan tablau public.

Parameter ARIMA

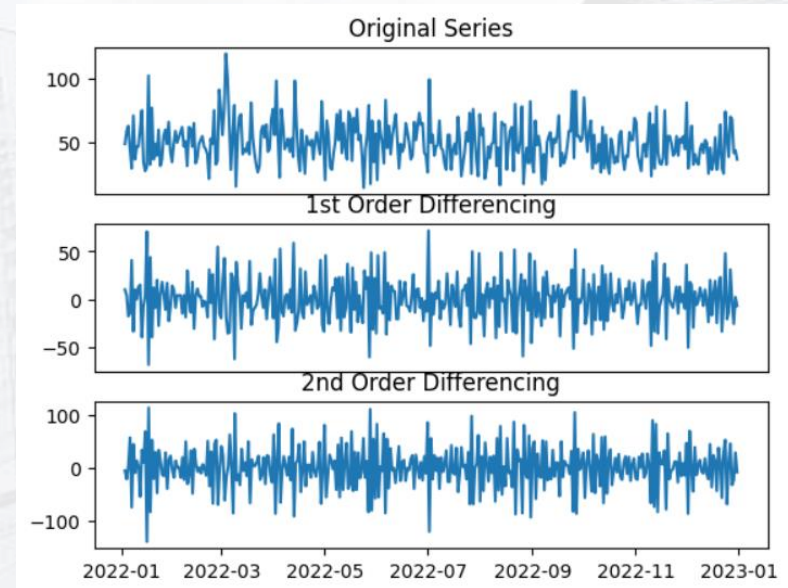
Parameter yang digunakan untuk membuat model ARIMA:

Nilai $p = 59$

Nilai $d = 2$

Nilai $q = 1$

Nilai d didapat dengan melakukan diferensiasi pada data, dilihat pada gambar. Pada diferensiasi ke-2 terlalu banyak noise, sehingga paling aman menggunakan diferensiasi ke-1. Jadi nilai $d = 2$.

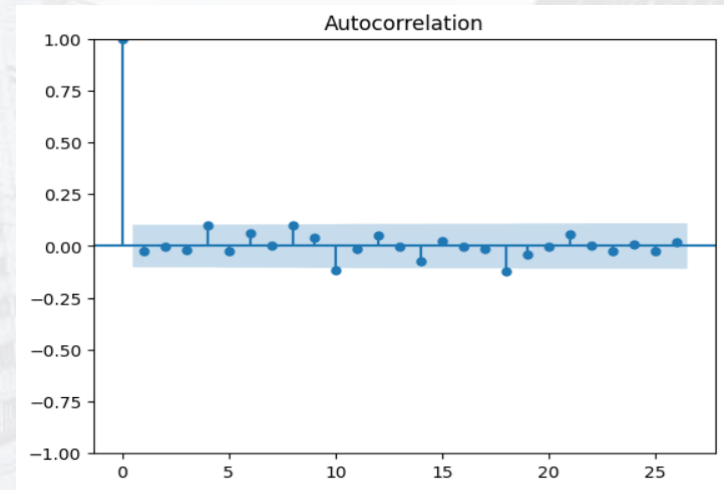


Grafik Data Sebelum dan Setelah diferensiasi

Parameter ARIMA

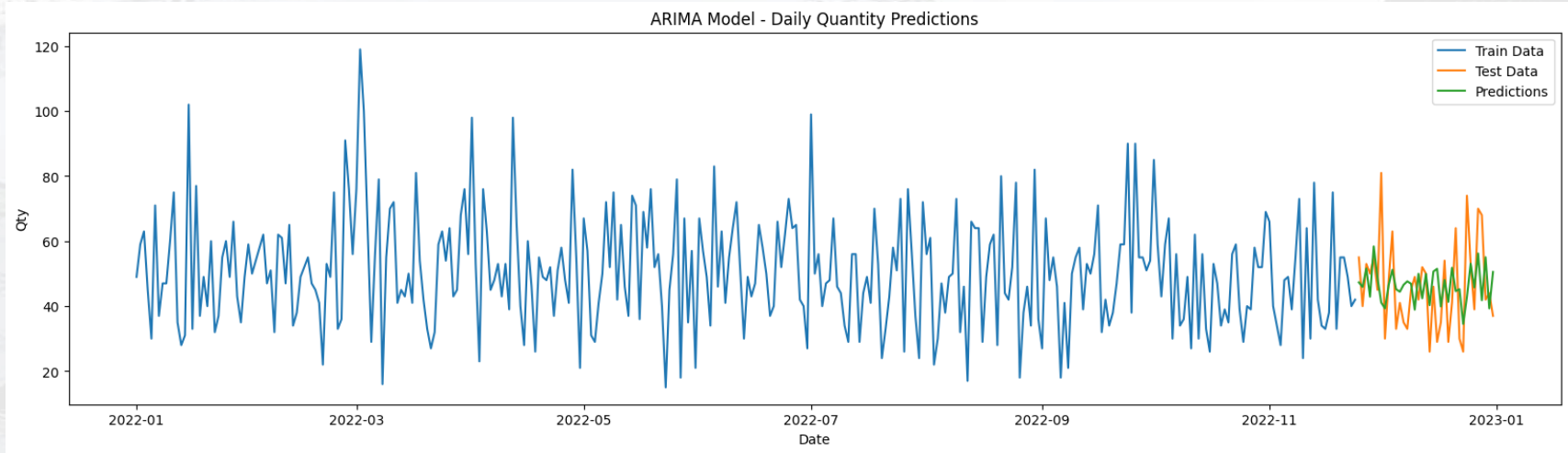
Nilai q didapat dengan menggunakan fungsi auto korelasi, dengan menghitung jumlah lags yang melebihi limit maka nilai $q = 1$.

Nilai p didapat dengan menggunakan hyperparameter tuning dengan metode grid search dari 1 sampai 100, dengan melihat nilai MSE terendah maka terpilih nilai $p = 59, 73, 85$. setelah dimasukkan ke model ARIMA nilai MAE dan RMSE dari ketiga nilai p hanya memiliki perbedaan yang tipis. Sehingga nilai p yang dipilih adalah $p = 59$, dikarenakan memiliki selisih MAE dan RMSE yang paling sedikit



Grafik Autocorrelation

Prediksi Penjualan



Hasil Prediksi Penjualan menggunakan Metode Time Series ARIMA, dengan nilai MAE dan RMSE adalah 10.57 dan 13.53

Klasterisasi Pelanggan

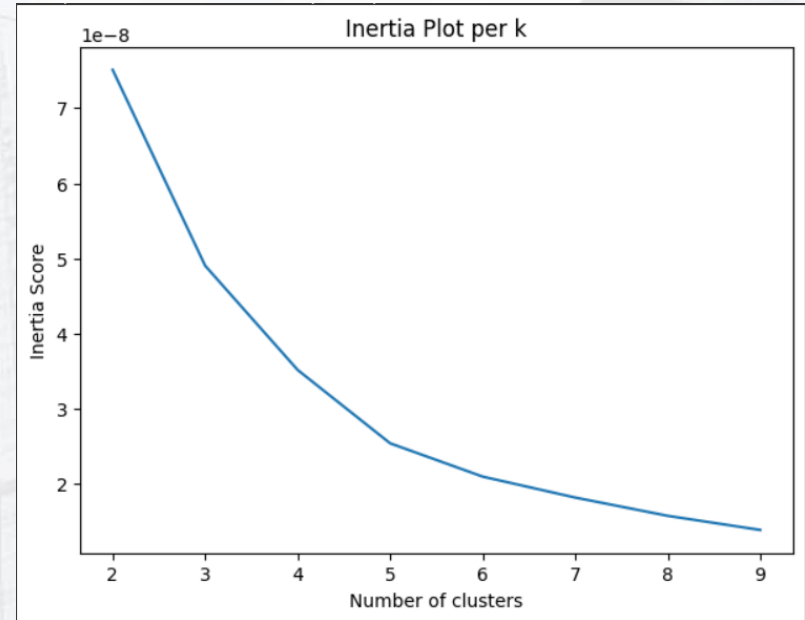
Klasterisasi atau pengelompokan pelanggan ditujukan untuk mengelompokan pelanggan agar Tim marketing bisa memberikan promo dan treatment bisnis yang sesuai. Klasterisasi dilakukan dengan menggunakan variabel banyak barang, jumlah transaksi dan total pembelian pelanggan yang dilakukan selama 1 tahun.

CustomerID	TransactionID	Qty	TotalAmount
1	17	60	623300
2	13	57	392300
3	15	56	446200
4	10	46	302500
5	7	27	268600

Elbow Method

Elbow method digunakan untuk mendapatkan nilai optimal untuk jumlah kelompok pelanggan atau klasterisasi.

Nilai optimal bisa diketahui dengan cara melihat grafik hubungan antara jumlah klaster dengan skor inertia. Jika penurunan nilai inertia tidak signifikan lagi maka pada titik itu adalah nilai optimal jumlah klaster. Maka jumlah klaster yang digunakan yaitu $K = 5$.



KMeans

Kmeans merupakan algoritma untuk melakukan klasterisasi, dengan menggunakan KMeans didapatkan klaster sebagai berikut:

Klaster 0 -> pembelian barang lumayan banyak dan harga barang tinggi

Klaster 1 -> pembelian barang lumayan banyak dan barang yang dibeli memiliki harga yang sedang.

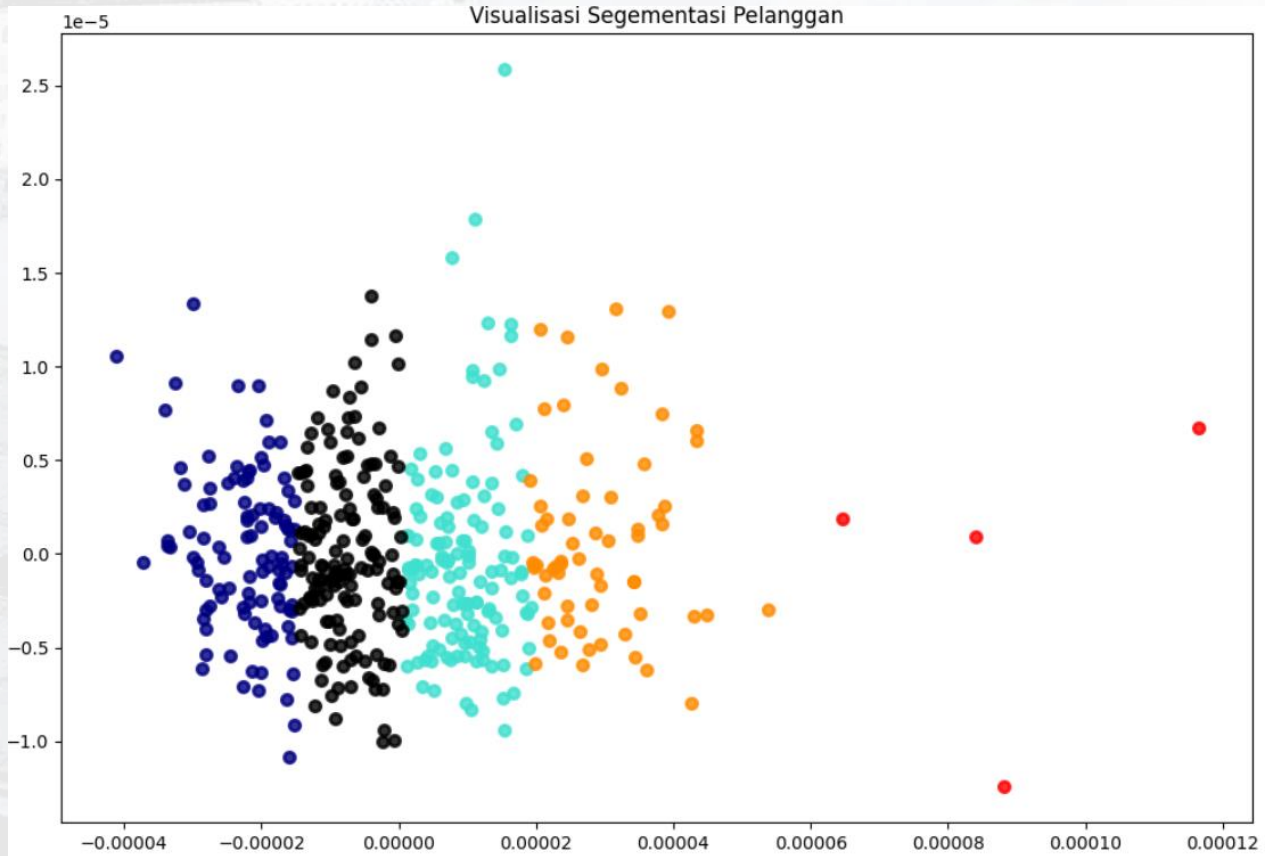
Klaster 2-> pembelian barang sedang, dan harga barang juga kurang dari klaster 1.

	CustomerID	TransactionID	Qty	TotalAmount
cluster_label				
0	103	11.388350	39.174757	417048.543689
1	130	11.092308	41.876923	336334.615385
2	63	10.634921	40.238095	281184.126984
3	4	7.250000	29.500000	150100.000000
4	147	11.605442	41.931973	388085.034014

Klaster 3-> pembelian barang sedikit, harga barang murah

Klaster 4-> pembelian barang banyak, jumlah barang banyak, dan harga barang sedang.

Scatterplot



<https://github.com/kaniang>

Video Presentation

https://drive.google.com/file/d/1LI_g-9VNHkswk8k-5FweyU2ui0_wEBAX/view?usp=sharing

Thank You



Rakamin
Academy



KALBE
Nutrifionals