Rakamin Academy X KALBE Nutritionals

Machine Learning Project

Kalbe Nutritionals Data Scientist Project Based Internship Program

Presented by
Kristy Natasha Yohanes

Kristy Natasha Yohanes

About Me

Alumni Meteorology ITB
(Bachelor of Science - B.S.)
Year of Graduation: 2023

Rakamin Academy

Work Experience

- **Researcher (Machine Learning Weather Forecaster)** - Atmospheric Department ITB
- **Data Science Research Intern** - BRIN Aviation and Space Research Organization
- **Data Analyst** - Community Service Program (PPM) by Garda Caah

Case Study

Membuat model Regression dan Clustering

○ **Dari tim inventory**: diminta untuk dapat membantu memprediksi jumlah penjualan (quantity) dari total keseluruhan product Kalbe
■ Tujuan dari project ini adalah untuk mengetahui perkiraan quantity product yang terjual sehingga tim inventory dapat membuat stock persediaan harian yang cukup.
■ Prediksi yang dilakukan harus harian.

○ **Dari tim marketing**: diminta untuk membuat cluster/segment customer berdasarkan beberapa kriteria.
■ Tujuan dari project ini adalah untuk membuat segment customer.
■ Segment customer ini nantinya akan digunakan oleh tim marketing untuk memberikan personalized promotion dan sales treatment

Exploratory Data Analysis

using DBeaver with a PostgreSQL database

```sql
/* Query 1: Berapa rata-rata umur customer jika dilihat dari marital statusnya? */
SELECT "Marital Status", AVG(Age) AS AverageAge
FROM customer
GROUP BY "Marital Status";

/* Query 2: Berapa rata-rata umur customer jika dilihat dari gender nya? */
SELECT Gender, AVG(Age) AS AverageAge
FROM customer
GROUP BY Gender;

/* Query 3: Tentukan nama store dengan total quantity terbanyak! */
SELECT StoreName, SUM(Qty) AS TotalQuantity
FROM transactiontable
INNER JOIN store ON transactiontable.StoreID = store.StoreID
GROUP BY StoreName
ORDER BY TotalQuantity DESC
LIMIT 1;

/* Query 4: Tentukan nama produk terlaris dengan total amount terbanyak! */
SELECT p."Product Name", SUM(t.totalamount) AS totalsales
FROM transactiontable t
INNER JOIN product p ON t.productid = p.productid
GROUP BY p."Product Name"
ORDER BY totalsales DESC
LIMIT 1;
```

customer 1

```
SELECT "Marital Status", AVG(Age)
```

	Marital Status	averageage
1		31.3333333333
2	Married	43.0382352941
3	Single	29.3846153846

customer 1

```
SELECT Gender, AVG(Age) AS Aver
```

	gender	averageage
1	0	40.326446281
2	1	39.1414634146

store 1

```
SELECT StoreName, SUM(Qty) AS T
```

	storename	totalquantity
1	Lingga	2,777

product 1

```
SELECT p."Product Name", SUM(t.t
```

	Product Name	totalsales
1	Cheese Stick	27,615,000

Exploratory Data Analysis

Marital Status	Gender
Married (avg. 43 years)	Man (avg. 39 years)
Single (avg. 29 years)	Woman (avg. 40 years)

Store	Quantity
Lingga	2,78K
Sinar Harapan	2,59K
Prima Kota	1,40K

Product	Total Amount
Cheese Stick	27,6M IDR
Choco Bar	21,2M IDR
Coffee Candy	19,7M IDR

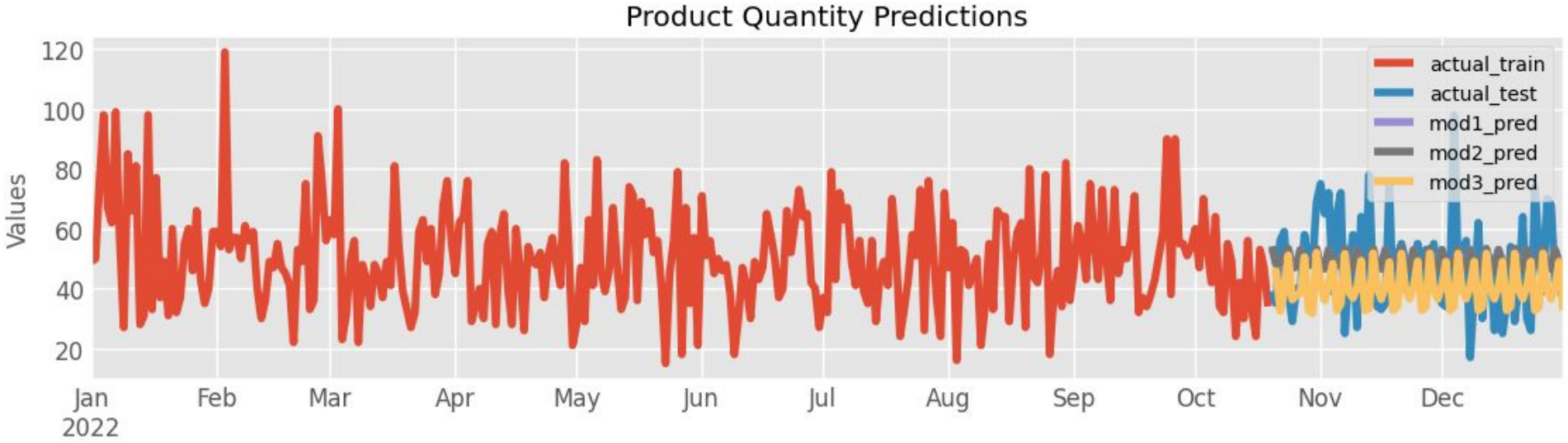
Data Visualization

using Tableau Public

https://public.tableau.com/app/profile/kristy.natasha/viz/KalbeDSIntenship/Dashboard#1

Predictive Analytics

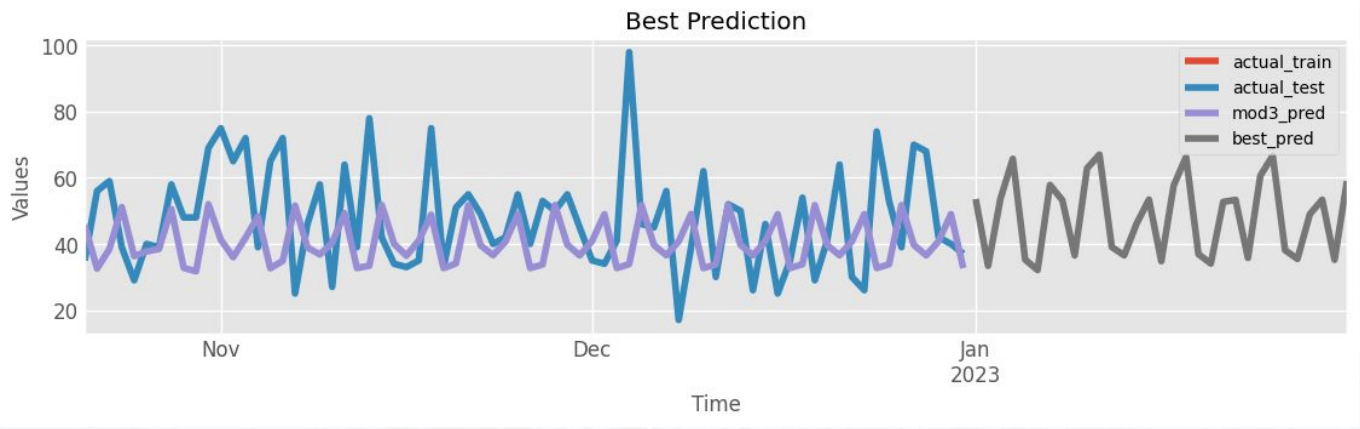
using machine learning regression (time series model ARIMA) with Python



https://colab.research.google.com/drive/14rTMcSWFT9lCfjn_bCG-zvqdR2chq9sB

Predictive Analytics

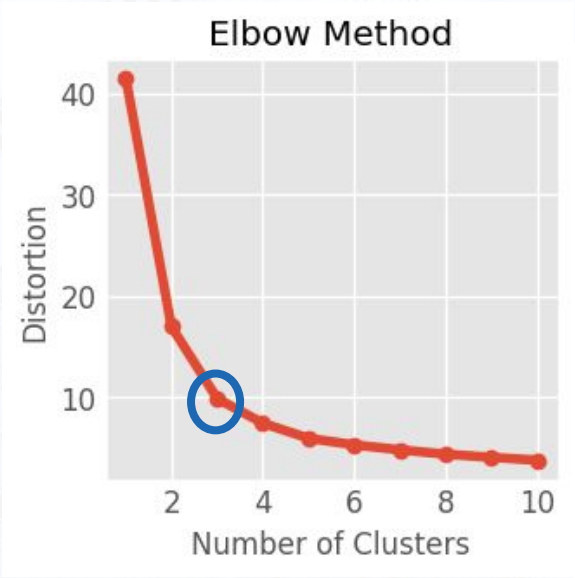
using machine learning regression (time series model ARIMA) with Python

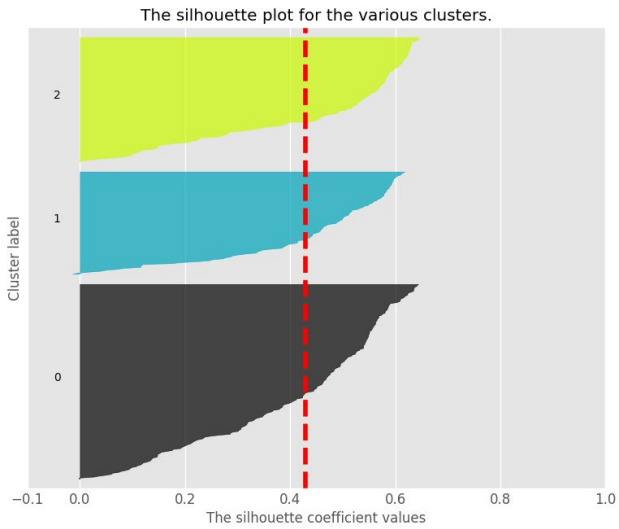


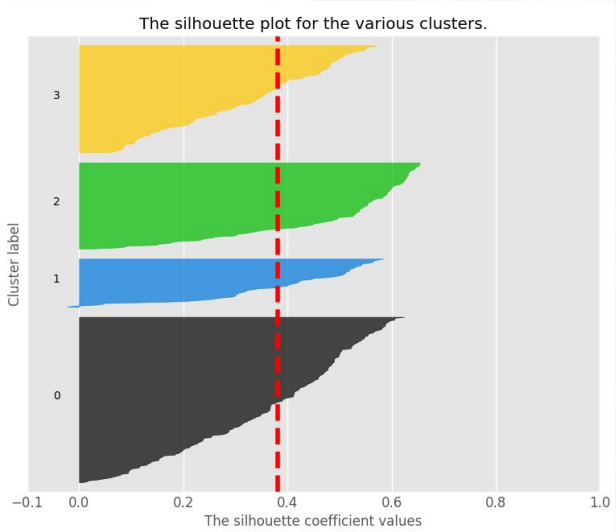
	predicted_mean
count	31.000000
mean	48.233205
std	11.955421
min	32.113562
25%	36.149794
50%	52.715606
75%	57.762546
max	67.016036

Data Clustering

using KMeans library in Python





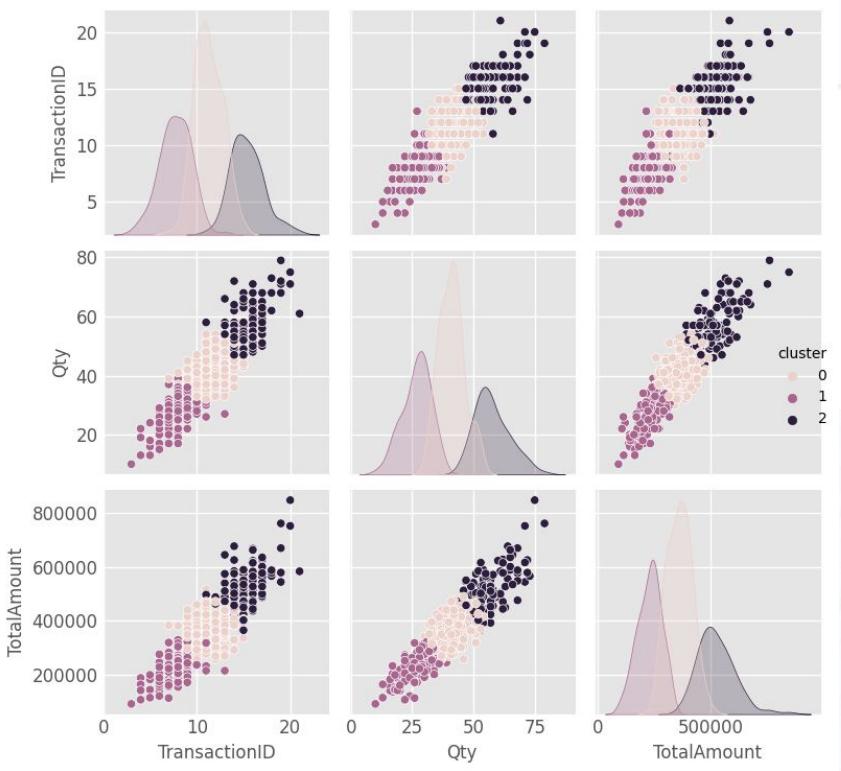


Jumlah cluster optimal = 3

For n_clusters = 3 The average silhouette_score is : 0.4294669050463297
For n_clusters = 4 The average silhouette_score is : 0.38109175331136835

Data Clustering

using KMeans library in Python



Customer Segmentation

cluster	0	1	2
TransactionID	11.253659	7.702290	15.370370
Qty	41.004878	26.725191	57.574074
TotalAmount	360908.292683	228550.381679	524466.666667

Customer Profile

High Spenders

Customers in this cluster are the highest spenders, making a large number of transactions and purchasing substantial quantities of products. They are the most consumptive group.

- VIP Treatment (exclusive perks, early access to promotions)
- Premium Products
- Referral Programs
- Personalization

Moderate Shoppers

This cluster consists of customers with a moderate level of consumption. They make a good number of transactions and purchase reasonably-sized quantities of products.

- Retention and Upselling (loyalty programs, offer exclusive discounts, rewards)
- Cross-selling
- Personalization

Budget Shoppers

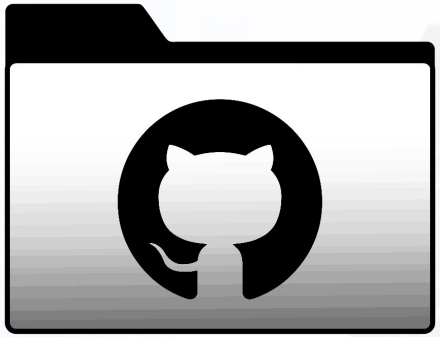
Customers in this cluster are budget-conscious shoppers. They make fewer transactions and opt for smaller quantities. They are the least consumptive group.

- Customer Engagement
- Product Bundles
- Feedback and Surveys

RESULT DOCUMENTATION

Video Presentation

drive.google.com/drive/folders/1eS8P3QKZq6Ipvq6-gVvfz9LvlwK0c3qs



Project Repository

github.com/kristynatasha/FMCG-Data-Modeling

CONTACT INFO

GitHub

github.com/kristynatasha

LinkedIn

linkedin.com/in/kristynatasha/

(+62) 8788-658-3513 | kristynatasha011@gmail.com | Jakarta, Indonesia

Thank You