

# Machine Learning Project Making Regression and Clustering models

Data Science – Kalbe

Nutritional  
Presented by

Muhammad Irfan Karim



## Muhammad Irfan Karim

### About Me

Irfan adalah mahasiswa sistem komputer tahun ketiga di Universitas Sriwijaya. Memiliki motivasi yang tinggi dan proaktif, mahir dalam Excel, SQL, Machine Learning, Deep Learning, dan Data Science. Memiliki kemampuan analisis dan pemecahan masalah yang kuat serta keterampilan menengah dalam bidang komputer dan teknologi. Pribadi yang mempunyai kecenderungan yang kuat untuk belajar dan menyerap informasi baru, baik sebagai tim atau kontributor independen.

## My Experience

Data Science Independent Study in Rakamin Academy

August 2023 - Present

Data Scientist in Intelligent System Research Group (ISysRG)

February 2023 - Present

Awardee in Bank Syariah Indonesia

August 2021 - Present

# Case Study 1

Peserta dapat melakukan exploratory data analysis di dbeaver

Query 1 : Berapa rata-rata umur customer jika dilihat dari marital statusnya ?

Query 2 : Berapa rata-rata umur customer jika dilihat dari gender nya ?

Query 3 : Tentukan nama store dengan total quantity terbanyak!

Query 4 : Tentukan nama produk terlaris dengan total amount terbanyak!

# Case Study 2

Peserta dapat membuat dashboard di tableau

Sebelum membuat dashboard terlebih dahulu membuat worksheet sebanyak 4.

1. Worksheet 1 Jumlah qty dari bulan ke bulan
2. Worksheet 2 Jumlah total amount dari hari ke hari
3. Worksheet 3 Jumlah penjualan (qty) by product
4. Worksheet 4 Jumlah penjualan (total amount) by store name

Setelah itu bisa membuat dashboard dengan menggabungkan 4 worksheet.

# Case Study 3

Peserta dapat membuat model prediktif menggunakan regresi dan membuat clustering

1. Membaca data csv
2. Melakukan data cleansing
3. Menggabungkan semua data menjadi 1 data
4. Membuat model machine learning regression (time series) menggunakan ARIMA
5. Membuat model machine learning clustering menggunakan KMeans
6. Membuat data baru untuk clustering, yaitu groupby by customerID lalu yang di aggregasi adalah
  1. Transaction id count
  2. Qty sum
  3. Total amount sum

# Result Case Study 1

	ABC Marital_Status	123 average_age
1	S	29.3846153846
2	M	43.1720116618

1. Rata-rata umur customer berdasarkan marital statusnya yaitu single 29 dan married 43

	ABC gender_type	123 average_age
1	Female	40.326446281
2	Male	39.5365853659

2. Rata-rata umur customer berdasarkan jenis kelamin yaitu female 40 dan male 39

	ABC storename	123 total_quantity
1	Lingga	2,777

3. Store dengan total quantity terbanyak adalah Lingga dengan total quantity sebanyak 2,777

	ABC product_name	123 total_amount
1	Cheese Stick	27,615,000

4. Produk terlaris berdasarkan total amount nya adalah Cheese Stick dengan total amount sebesar 27,615



# Result Case Study 2

Penjelasan :

## 1. Quantity by Month

Quantity terbesar didapat dari bulan Januari sebesar 1,763K

## 2. Total Amount by Day

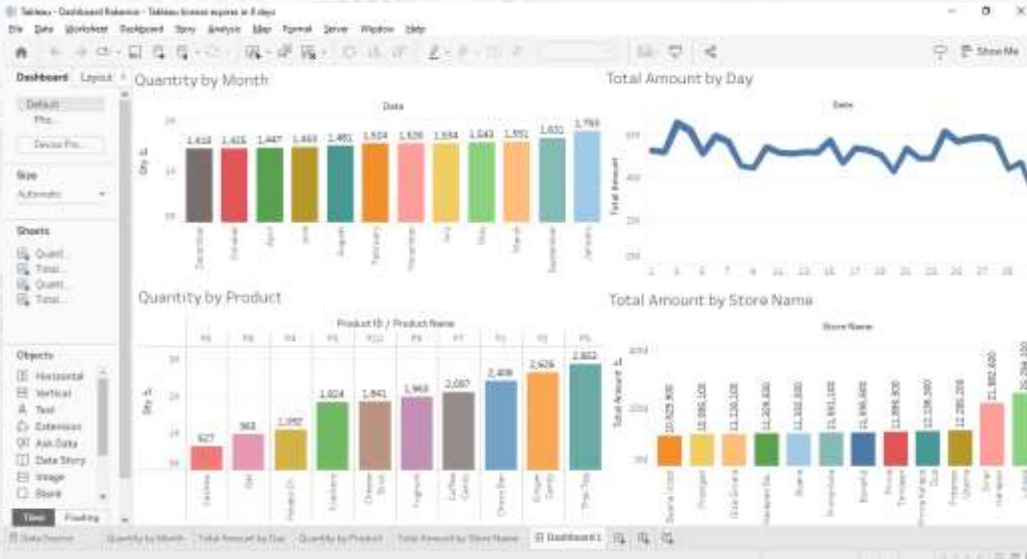
Total amount terbesar didapat pada hari ke 3 yaitu sebesar 6,542,100M

## 3. Quantity by Product

Quantity terbesar didapat pada Product ID P5 dengan nama Product Thai Tea yaitu sebesar 2,853K

## 4. Total Amount by Store Name

Total Amount terbesar didapat pada Store Name Lingga yaitu sebesar 25,294,100M



# Result Case Study 3



Penjelasan :

Saya menggunakan Library dari Scalecast untuk melakukan forecast dengan menggunakan model ARIMA.

Lalu saya menggunakan grid search pada model ARIMA saya agar dapat mengotomasi nilai dan mencari nilai terbaik untuk diterapkan pada model ARIMA.

Gambar disamping merupakan hasil dari forecast menggunakan model ARIMA dengan jarak forecast 1 bulan kedepan.

# Result Case Study 3



Penjelasan : Pada Clustering pertama saya mengagregasi data dengan ketentuan yang sudah diberikan, lalu mencari nilai K terbaik dengan metode elbow method dan didapat bahwa nilai K terbaik adalah 3. Setelah itu saya melakukan training untuk model Kmeans tersebut.

Lalu saya membuat plot berdasarkan 3 case yang sudah diberikan dan mendapatkan hasil seperti gambar diatas.



# Link Folder/File Hasil Pengerjaan

[https://drive.google.com/drive/folders/1xMb6S3CH4ZwX5ix9TDsHTQ\\_QLO8ofvLK?usp=sharing](https://drive.google.com/drive/folders/1xMb6S3CH4ZwX5ix9TDsHTQ_QLO8ofvLK?usp=sharing)

# Link Github Here

[https://github.com/Muhammadirfankarim/Rakamin-VIX\\_Data-Science\\_Kalbe-Nutritionals](https://github.com/Muhammadirfankarim/Rakamin-VIX_Data-Science_Kalbe-Nutritionals)

# Video Presentation Here

<https://youtu.be/A4d-qKtISNA>

# Thank You

