



Virtual Internship Experience

Machine Learning Project

Membuat model Regression dan
Clustering

Table of Content

- 1 Challenge**
- 2 Prerequisite**
- 3 Hints**
- 4 Submission**

Challenge

- **Background Story**

- Kamu adalah seorang Data Scientist di Kalbe Nutritional dan sedang mendapatkan project baru dari tim inventory dan tim marketing.
- Dari tim inventory, kamu 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 kamu 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.
- Tools yang akan kamu gunakan dalam project ini adalah
 - Python
 - Jupyter Notebook
 - Tableau
 - Dbeaver
 - PostgreSQL

Challenge

- **Peserta dapat melakukan data ingestion ke dalam dbeaver**
 - Pilih connect to database di pojok kiri dan pilih postgresSQL
 - Test connection ke postgresSQL
 - Di menu drop down postgresSQL pilih table dan klik kanan (import data)
 - Pilih CSV file dan upload 4 file dari Kalbe
https://drive.google.com/drive/folders/1_rQrauVW20vLle2zd54Vcwnr2EY-vnnR?usp=sharing
- **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!
- **Peserta dapat melakukan data ingestion ke dalam tableau public**
 - Pertama buka tableau public (<https://public.tableau.com/app/discover>) . Sign in jika sudah punya akun dan sign up jika belum punya akun tableau public.

Challenge

- Setelah itu klik create dan pilih web authoring
- Setelah itu akan di redirect ke halaman baru dan bisa langsung upload 4 data dari Kalbe
- **Peserta dapat membuat dashboard di tableau**
 - Sebelum membuat dashboard terlebih dahulu membuat worksheet sebanyak 4.
 - Worksheet 1 Jumlah qty dari bulan ke bulan
 - Worksheet 2 Jumlah total amount dari hari ke hari
 - Worksheet 3 Jumlah penjualan (qty) by product
 - Worksheet 4 Jumlah penjualan (total amount) by store name
 - Setelah itu bisa membuat dashboard dengan menggabungkan 4 worksheet.
- **Peserta dapat membuat model prediktif menggunakan regresi dan membuat clustering**
 - Membaca data csv
 - Melakukan data cleansing
 - Menggabungkan semua data menjadi 1 data
 - Membuat model machine learning regression (time series)
 - Membuat model machine learning clustering

Prerequisite

Sebelum mengerjakan challenges, silakan untuk menyelesaikan pre-requisite di bawah ini:

- Download dbeaver (<https://dbeaver.io/download/>)
- Download postgresql (<https://www.postgresql.org/download/>)
- Download anaconda (<https://www.anaconda.com/download>)
 - Install python
 - Install jupyter
 - Install library yang dibutuhkan seperti pandas, numpy, scikit-learn
- Mendaftar di tableau public (<https://public.tableau.com/app/discover>)

Hints

Dbeaver Connection with PostgreSQL

- Sebelum menggunakan dbeaver, pastikan sudah melakukan instalasi postgresQL terlebih dahulu.
- Jangan lupa test connection terlebih dahulu ketika mau connect ke PostgreSQL
- Delimiter dari data adalah ; bukan ,
- Untuk melakukan query pada data klik icon SQL pada dbeaver

Hints

Tableau Public Create Dashboard

- Harus sign up terlebih dahulu jika belum punya akun
- Tidak perlu install tableau di local laptop
- Bisa langsung menggunakan 4 data
- Saat membuat dashboard bisa menggabungkan 4 worksheet yang telah dibuat
- Saat membuat dashboard, setiap worksheet jangan lupa dibikin floating supaya mudah untuk di drag and drop

Hints

Machine Learning Regression (Time Series)

- Tujuan dari pembuatan model machine learning ini adalah untuk dapat memprediksi total quantity harian dari product yang terjual.
- Data cleansing terlebih dahulu, merubah tipe data supaya sesuai
- Data merge untuk menggabungkan semua data
- Membuat data baru untuk regression, yaitu groupby by date lalu yang di aggregasi adalah qty di sum
- Akan ada sekitar 365 rows
- Menggunakan metode time series ARIMA

Hints

Machine Learning Clustering

- Tujuan dari pembuatan model machine learning ini adalah untuk dapat membuat cluster customer-customer yang mirip
- Data cleansing terlebih dahulu, merubah tipe data supaya sesuai
- Data merge untuk menggabungkan semua data
- Membuat data baru untuk clustering, yaitu groupby by customerID lalu yang di agregasi adalah :
 - Transaction id count
 - Qty sum
 - Total amount sum
- Menggunakan metode clustering KMeans
- Untuk proses queri data, dapat memilih alternatif pemrograman yang sesuai preferensi kamu.
- Sedangkan untuk proses machine learning, Kamu diperbolehkan memilih dan menggunakan satu bahasa pemrograman di antara R dan Python untuk membantumu mengerjakan Final Project.
- (Untuk Proses Queri Database) Pengumpulan tugas dalam bentuk format file SQL script yang berisi perintah-perintah SQL secara berurutan yang nantinya dapat dieksekusi untuk memperoleh hasil yang diinginkan

Hints

- (Untuk Pemodelan Machine Learning) Terdapat dua skenario pengumpulan tugas berdasarkan bahasa pemrograman yang kamu gunakan untuk mengerjakan final project ini.
 - Apabila kamu menggunakan R untuk mengerjakan tugas ini, kamu perlu mengumpulkan:
 - File Code dengan format .R
 - File R Markdown dengan format .Rmd
 - File infografis sebagai media presentasi yang telah kamu buat dengan format .pdf
 - Apabila kamu menggunakan Python untuk mengerjakan tugas ini, kamu perlu mengumpulkan:
 - File Code dengan format .py
 - File Python Notebook dengan format .ipynb
 - File infografis/ppt sebagai media presentasi yang telah kamu buat dengan format .pdf
- Kerjakan end-to-end solution dan siapkan seluruh file yang perlu kamu kumpulkan dengan mengacu kepada No. 2 dan 3
- Gabung file yang telah kamu selesaikan dalam satu folder ZIP yang diberi judul "[Nama Lengkap]_Kalbe"
- Submit tugasmu dalam format template yang sudah disediakan pada bagian "Submission"

Submission

Setelah Anda menyelesaikan Challenge yang telah diberikan, Anda **DIWAJIBKAN** untuk membuat **File Final Submission** dalam format PPT dengan template <https://bit.ly/template-final-task-vix> yang berisi:

- Biodata Diri
- Hasil Pengerjaan
- Link Folder/File Hasil Pengerjaan (ikuti petunjuk pengerjaan)
- Link Github Hasil Pengerjaan (ikuti petunjuk pengerjaan)
- Link Video Presentasi Hasil Pekerjaan / Penjelasan Project yang Telah Anda Kerjakan

Video presentasi yang telah dibuat, harus Anda upload pada **YouTube** atau **Google Drive** untuk mendapatkan link yang akan dicantumkan pada file final submission nantinya.

Kemudian, **kumpulkan** *File Final Submission* tersebut dalam bentuk **PDF** dengan format **FinalTask_Kalbe_DS_Nama Lengkap**

File tersebut yang akan **diberikan kepada company** sebagai hasil akhir dari program VIX ini. Diharapkan Anda dapat membuat *File Final Submission* ini dengan sungguh-sungguh. Terima kasih.



Thank You!



KALBE
Nutritionals