Universidad Galileo

Postgrado en Análisis y Predicción de Datos

Product Development, Sección L

Catedrático: Ing Preng Biba.

Auxiliar: Milton Godinez.

**Nombre: Kimberly Alejandra Rivera González**

**Carné: 20001676**

**Dataset a utilizar en proyecto**

**El dataset que utilizaré en el proyecto se llama:** Customer Personality Analysis marketing\_campaign.csv

**El contexto descrito en el dataset es:**

Planteamiento del problema: El análisis de la personalidad del cliente es un análisis detallado de los clientes ideales de una empresa. Ayuda a una empresa a comprender mejor a sus clientes y les facilita la modificación de productos de acuerdo con las necesidades, los comportamientos y las preocupaciones específicas de los diferentes tipos de clientes.

El análisis de la personalidad del cliente ayuda a una empresa a modificar su producto en función de sus clientes objetivo de diferentes tipos de segmentos de clientes. Por ejemplo, en lugar de gastar dinero para comercializar un nuevo producto para cada cliente en la base de datos de la empresa, una empresa puede analizar qué segmento de clientes es más probable que compre el producto y luego comercializar el producto solo en ese segmento en particular.

**Objetivo:** Necesita realizar agrupaciones para resumir los segmentos de clientes.

**Cantidad de registros:** 2240 filas.

**Cuenta con una cantidad de columnas de:** 29 columnas.

**Descripción de cada una de las columnas:** Contenido de Atributos

***Persona***

ID: identificador único del cliente

Year\_Birth: año de nacimiento del cliente

Education: nivel de educación del cliente

Marital\_Status: estado civil del cliente

Income: ingresos familiares anuales del cliente

Kidhome: número de niños en el hogar del cliente

Teenhome: Número de adolescentes en el hogar del cliente

Dt\_Customer: fecha de inscripción del cliente en la empresa

Recency: número de días desde la última compra del cliente

Queja: 1 si el cliente se quejó en los últimos 2 años, 0 en caso contrario

***Productos***

MntWines: cantidad gastada en vino en los últimos 2 años

MntFruits: Cantidad gastada en frutas en los últimos 2 años

MntMeatProducts: Cantidad gastada en carne en los últimos 2 años

MntFishProducts: Cantidad gastada en pescado en los últimos 2 años

MntSweetProducts: Cantidad gastada en dulces en los últimos 2 años

MntGoldProds: cantidad gastada en oro en los últimos 2 años

***Promoción***

NumDealsPurchases: Número de compras realizadas con descuento

AcceptedCmp1: 1 si el cliente aceptó la oferta en la primera campaña, 0 en caso contrario

AcceptedCmp2: 1 si el cliente aceptó la oferta en la segunda campaña, 0 en caso contrario

AcceptedCmp3: 1 si el cliente aceptó la oferta en la tercera campaña, 0 en caso contrario

AcceptedCmp4: 1 si el cliente aceptó la oferta en la cuarta campaña, 0 en caso contrario

AcceptedCmp5: 1 si el cliente aceptó la oferta en la quinta campaña, 0 en caso contrario

Response: 1 si el cliente aceptó la oferta en la última campaña, 0 en caso contrario

***Lugar***

NumWebPurchases: número de compras realizadas a través del sitio web de la empresa

NumCatalogPurchases: número de compras realizadas mediante un catálogo

NumStorePurchases: número de compras realizadas directamente en las tiendas

NumWebVisitsMonth: número de visitas al sitio web de la empresa en el último mes

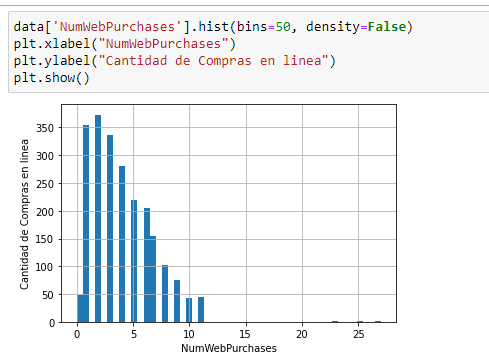
**La fuente de donde se tomó es de Kaggle:**

<https://www.kaggle.com/imakash3011/customer-personality-analysis>

**Visualización de los datos en Jupyter:**



**Target:** Se exploró y analizó la data y el target que elegí para el proyecto es NumWebPurchases: número de compras realizadas a través del sitio web de la empresa, esa será la variable que se va a predecir.



**Adjunto a este documento el dataset para su visualización:**

marketing\_campaign.csv

**Pasos realizados:**

Fueron tres notebooks.

* El primer notebook es el de Proyecto 1 Kimberly Rivera 20001676.ipynb, el cual contiene todo el análisis y toda la limpieza de datos y el llenado de las variables con data faltante.
* Luego el siguiente notebook que se encontrará en la carpeta es llamado Variable\_Selection\_And\_Model\_Training P1 - Kimberly Rivera 20001676, el cual contiene todos los modelos que se realizaron y que se entrenaron para validar cual era el mejor modelo que se tomaría para el pipeline, yo utilicé neuronas y elegí la neurona que me dio mejor resultado.
* El tercer notebook es llamado Machine\_Learning\_Pipeline P1 - Kimberly Rivera 20001676, donde se encuentra toda el esquema y la estructura del pipeline, así como generar el archivo PKL a utilizar.
* En cada uno de los notebooks se encuentran anotaciones y títulos de lo realizado.

**API desarrollada.**

El API se encuentra en la carpeta **apidemo**

En esta carpeta se encuentran los archivos de:

* App.py = archivo donde está el llamado desde postman.
* Config.py = archivo donde está toda la estructura del pipeline.
* My\_preprocessors.py = archivo que creamos al inicio para las variables de tiempo.
* Pipeline\_predict2.py = es el archivo donde se encuentra el llamado al archivo anteriormente guardado PKL que lleva por nombre WebPurchase\_pipeline.pkl
* marketing\_campaign.csv = archivo de prueba
* Pasos en el Prompt.txt = Un archivo que yo cree para realizar las pruebas en el postman y en el prompt de anaconda y contiene esta información:

C:\Users\kimbe\

cd OneDrive - Universidad Galileo

cd Octavo Trimestre

cd Product Development

cd Proyecto 1 Kimberly Rivera

cd apidemo

python pipeline\_predict2.py

OneDrive - Universidad Galileo\Octavo Trimestre\Product Development\Proyecto 1 Kimberly Rivera\apidemo

set FLASK\_APP=app

flask run

Prueba en postman:

{"ID":"2174","Year\_Birth":"1954","Education":"Graduation","Marital\_Status":"Single","Income":"46344","Kidhome":"1","Teenhome":"1","dia":"8","mes":"3","Year":"2014","Recency":"38","MntWines":"11","MntFruits":"1","MntMeatProducts":"6","MntFishProducts":"2","MntSweetProducts":"1","MntGoldProds":"6","NumDealsPurchases":"2","NumCatalogPurchases":"1","NumStorePurchases":"2","NumWebVisitsMonth":"5","AcceptedCmp3":"0","AcceptedCmp4":"0","AcceptedCmp5":"0","AcceptedCmp1":"0","AcceptedCmp2":"0","Complain":"0","Z\_CostContact":"3","Z\_Revenue":"11","Response":"0"}

* bbu5l-oriqg.json = el archivo .json que se utiliza en postman.
* Imágenes que confirman la conexión entre el postman y el api creada:
  + API Y POSTMAN.PNG
  + API.PNG
  + POSTMAN.PNG

Text

Description automatically generated

Text

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated