

**PROYECTO FINAL**

**DATA SCIENCE**

Autor:

Jean Piero Marcon

Institución:

Coderhouse

Fecha de presentación:

10/2022

Un letrero de color negro

Descripción generada automáticamente con confianza baja

**Contenido**

1. Descripción del negocio y de la finalidad del estudio
2. Diccionario de los datos
3. Herramientas tecnológicas utilizadas
4. Dataset a utilizar
5. Exploración de los datos
6. Procesamiento de los datos
7. Análisis exploratorio de los datos
8. Entrenamiento de modelos supervisados

9.1 Decision Tree

9.2 Random Forest

9.3 K-Nearest Neighborg

9.4 Logistic Regression

9.5 Support Vector Machine

9.6 Evaluación de modelos

1. Optimización de parámetros

10.1 GridSearchCV

10.2 RandomSearchCV

10.3 Comparación de métricas

1. Evaluación del modelo seleccionado
2. Conclusión
3. **Descripción del negocio y de la finalidad del estudio**

El análisis de los clientes ayuda a modificar un producto en función a los diferentes segmentos de ellos. Por ejemplo, desarrollar estrategias de ventas enfocadas a los clientes de mayor potencial, y con productos específicos.

En la siguiente sección, iremos a través de un proyecto de ciencia de datos sobre el análisis de los clientes con python, buscando predecir aquellos que sean potencialmente aceptadores de las campañas de marketing, todo esto con datos recopilados de datos personales y desde las ventas para tener un amplio panorama del comportamiento.

1. **Diccionario de los datos**

**People**

1. ID: Customer's unique identifier
2. Year\_Birth: Customer's birth year
3. Education: Customer's education level
4. Marital\_Status: Customer's marital status
5. Income: Customer's yearly household income
6. Kidhome: Number of children in customer's household
7. Teenhome: Number of teenagers in customer's household
8. Dt\_Customer: Date of customer's enrollment with the company
9. Recency: Number of days since customer's last purchase
10. Complain: 1 if the customer complained in the last 2 years, 0 otherwise

**Products**

1. MntWines: Amount spent on wine in last 2 years
2. MntFruits: Amount spent on fruits in last 2 years
3. MntMeatProducts: Amount spent on meat in last 2 years
4. MntFishProducts: Amount spent on fish in last 2 years
5. MntSweetProducts: Amount spent on sweets in last 2 years
6. MntGoldProds: Amount spent on gold in last 2 years

**Promotion**

1. NumDealsPurchases: Number of purchases made with a discount
2. AcceptedCmp1: 1 if customer accepted the offer in the 1st campaign, 0 otherwise
3. AcceptedCmp2: 1 if customer accepted the offer in the 2nd campaign, 0 otherwise
4. AcceptedCmp3: 1 if customer accepted the offer in the 3rd campaign, 0 otherwise
5. AcceptedCmp4: 1 if customer accepted the offer in the 4th campaign, 0 otherwise
6. AcceptedCmp5: 1 if customer accepted the offer in the 5th campaign, 0 otherwise
7. Response: 1 if customer accepted the offer in the last campaign, 0 otherwise

**Place**

1. NumWebPurchases: Number of purchases made through the company’s website
2. NumCatalogPurchases: Number of purchases made using a catalogue
3. NumStorePurchases: Number of purchases made directly in stores
4. NumWebVisitsMonth: Number of visits to company’s website in the last month
5. **Herramientas tecnológicas utilizadas**

* Anaconda y Google Colaboratory: Para la creación y manipulación de archivos Jupyter Notebook con Python. A la vez, utilizando librerías dentro de python, como Numpy, Pandas, Matplotlib, Seaborn y Sklearn para el entrenamiento y evaluación de los modelos de Machine Learning.
* Ilustrator y Canva: Para la creación y edición de imágenes.
* Microsoft Word: Para la creación de informes.
* Microsoft Power BI: Para la creación de un Dashboard informativo y explicativo donde se integren los datos y sea de fácil manipulación.
* Kaggle: Base de datos de datasets funcional para ejemplos de investigación.

1. **Dataset a utilizar**

El conjunto de datos a estudiar fue sacado de la página web <https://raw.githubusercontent.com/amankharwal/Website-data/master/marketing_campaign.csv>, como un archivo sin formato .csv para luego ser codificado en python y trabajado como un dataframe.