



EXAMEN PARCIAL

2024-1

CC57 Machine Learning

Secciones: Todas

I. PARTE PRÁCTICA

Caso tarjetas de crédito

El gerente de negocios de la entidad financiera Perú Bank está preocupado porque cada vez más clientes dan de baja sus servicios de tarjetas de crédito. Con el objetivo de reducir la tasa de abandono, el gerente está planificando realizar campañas de marketing, para que, de manera proactiva, se pueda gestionar y fidelizar a los clientes con mayor propensión de abandono y pueda cambiar su decisión de dejar el banco.

Para ello, se dispone de un conjunto de datos que consta de 3,254 clientes que mencionan su edad, salario, estado civil, límite de tarjeta de crédito, categoría de tarjeta de crédito, etc.

A continuación, se detalla la descripción de las variables:

- **CLIENTNUM:** Numero de cliente. Identificador único del cliente titular de la cuenta.
- **Target:** Hace referencia a si el cliente dio de baja su tarjeta de crédito (1: dio de baja, 0: no se dio de baja).
- **Customer_Age:** edad del cliente en años.
- **Gender:** género del cliente (M=Masculino, F=Femenino).
- **Dependent_count:** número de dependientes que tiene el cliente.
- **Education_Level:** nivel de educación del cliente.
- **Marital_Status:** estado civil del cliente.
- **Income_Category:** categoría del ingreso anual del cliente.
- **Card_Category:** tipo de tarjeta de crédito del cliente.
- **Months_on_book:** antigüedad del cliente en el banco en meses.
- **Total_Relationship_Count:** número de productos del cliente en el banco.
- **Months_Inactive_12_mon:** número de meses inactivos en los últimos 12 meses.
- **Contacts_Count_12_mon:** número de contactos que ha tenido el cliente con el banco en los últimos 12 meses.
- **Credit_Limit:** línea de tarjeta de crédito.
- **Total_Revolving_Bal:** saldo de tarjeta de crédito.
- **Avg_Open_To_Buy:** línea de crédito abierta para comprar en promedio en los últimos 12 meses.

- **Total_Trans_Amt:** monto total de las transacciones con la tarjeta de crédito en los últimos 12 meses.
- **Total_Trans_Ct:** número total de las transacciones con la tarjeta de crédito en los últimos 12 meses.
- **Avg_Utilization_Ratio:** índice de utilización promedio de la tarjeta de crédito.
- **Flag_Muestra:** partición de la base de datos en entrenamiento y validación (1=train, 0=test).

Con esta información y haciendo uso de una muestra de entrenamiento (Flag_Muestra = 1) / validación (Flag_Muestra = 0), desarrollar lo siguiente:

1. En base a la descripción del caso, plantear el objetivo de machine learning. (1 punto)
2. Realizar un análisis de identificación y tratamiento de valores faltantes. Presentar sus conclusiones. (1 punto)
3. Según corresponda, realizar un análisis descriptivo haciendo uso del gráfico de boxplot entre el target y las variables. Presentar sus conclusiones. (1 punto)
4. Según corresponda, realizar un análisis descriptivo haciendo uso de un gráfico de barras al 100% apilado entre el target y las variables. Presentar sus conclusiones. (1 punto)
5. Haciendo uso del modelo de Machine Learning KNN, programar y presentar una función para determinar el "K" más adecuado en función al AUC. La función debe presentar de manera gráfica en el eje vertical los distintos valores de AUC y en el eje horizontal lo distintos valores de "K" (entre 1 y 30), así como el valor de "K" más adecuado. Presentar sus conclusiones. (3 puntos)
6. Programar una función para presentar en una sola gráfica la curva ROC (Receiver Operating Characteristic) haciendo uso de los modelos de Machine Learning Naive Bayes, KNN y Redes Neuronales Artificiales (como mínimo Perceptron) . Presentar sus conclusiones. (3 puntos)

Usar el conjunto de datos "**peru_bank.csv**"

Indicaciones:

- No olvide completar sus datos personales (apellidos, nombres y código) en el Notebook antes de adjuntarlo con la solución.
- Renombre el archivo **EP_PRIMERPELLIDO_PRIMERNOMBRE_codigo.ipynb** con sus datos personales.