

Prueba Técnica - Analítica

El Product Owner [PO] de una célula de trabajo está interesado en fortalecer la alianza comercial con una institución dedicada a la reparación de deudas de personas naturales. Estos productos se utilizan en usuarios con dificultades para manejar sus obligaciones financieras permitiéndoles ajustar sus obligaciones a su capacidad de pago, renegociando el plazo e incluso el monto y tipo de acreedores.

El PO, le ha sido asignado una Historia de Usuario [HU] en Jira que consta de tres partes:

- Parte 1: exploración de datos y entrenamiento de modelos
- Parte 2: optimización de pipelines de entrada y salida de datos del modelo
- Parte 3: aplicaciones del modelo en estrategias comerciales

Para efectos de la prueba, cada parte tiene un peso que se especifica a continuación.

Parte 1 [50%]:

En el equipo de analítica, se preparó una base de datos ["Usuarios_Reparadora.csv"], con el objetivo de identificar las características que tienen en común los usuarios que han sido elegidos y rechazados por por nuestro aliado de reparadoras de deudas en el pasado, con el objetivo de promover el producto a otros usuarios que les aplique.

El objetivo de esta primera parte es construir un modelo analítico supervisado de tipo clasificación para explicar la variable objetivo [Y], la cual evalúa la propensión de un usuario de ser aceptado por el aliado para adquirir el producto de reparadora de deudas.

En la solución planteada, construya un Jupyter Notebook en Python donde evidencie los siguientes análisis:

1. Análisis exploratorio de los datos
2. Proceso de ingeniería de características [puede dejar de utilizar algunas variables o generar nuevas a partir de las existentes]
3. Proceso de selección de características y balanceo de datos
4. Entrenar un modelo de regresión logística para la clasificación
5. Seleccionar un modelo diferente a la regresión logística y entrenarlo
6. Definición y justificación de la mejor métrica para medir la calidad de los modelos y comparar los resultados de los dos modelos

Nota: dentro de los acuerdos comerciales, se estableció que el saldo actual de la persona debe ser superior a COP 4.500.000 para que se le pueda generar el plan de liquidación.

Parte 2 [30%]:

Una vez tenga los modelos definitivos de la Parte 1, el objetivo de esta segunda parte es optimizar los procesos de entrada y de salida de los modelos cuando se ingresan datos nuevos o “*out-of-sample*”.

En el Jupyter Notebook anterior, le corresponde almacenar los procedimientos de preparación de datos y los objetos con los modelos entrenados en archivos tipo “pkl” o “pmmi”, con el objetivo que estos artefactos analíticos puedan ser reutilizados.

Así las cosas, construya un nuevo Jupyter Notebook en Python que contenga únicamente los procedimientos expresados a continuación:

1. Carga de la base de datos masiva de usuarios sin variable Y [utilice el archivo “Perfilador_Reparadora.csv”]
2. Inyección del pipeline de procesamiento de variables al *dataset* del paso anterior [variables X]
3. Inyección de los modelos pre-entrenados al *dataset* resultante del paso anterior y almacene la predicción del modelo como nuevas columnas en el *dataset*
4. Genere una nueva columna que sirva para particionar los usuarios a los que se les va a realizar una campaña comercial . Para esto puede utilizar diferentes reglas duras, mecanismos de votación de modelos, entre otros
5. Guarde los resultados y expórtelos en un archivo con extensión “.csv”.

Parte 3 [20%]:

La persona Líder de Mercadeo Digital, le pide algunas ideas para personalizar las campañas comerciales. Para esto, usted tiene la iniciativa de segmentar a los usuarios utilizando modelos de tipo no-supervisados sobre el *dataset* generado en la Parte 2.

Desarrolle un nuevo Jupyter Notebook en Python donde:

1. Cargue el archivo “.csv” de la Parte 2
2. Aplique una técnica de segmentación estadística e identifique en cuantos grupos deberían lanzarse las campañas de mercadeo
3. Resuma los principales mensajes para enfatizar con el equipo de mercado a partir de los *insights* encontrados en el punto anterior [Utilice los campos de Markdown del Notebook]