

Code Challenge – Advanced Analytics

Objetivo

El principal objetivo es realizar un análisis exploratorio en el contexto de una empresa de seguros. Además, se debe realizar una clasificación binaria para poder predecir que pólizas harán un reclamo de siniestro y cuáles no.

Todo el desarrollo deberá estar hecho en Python.

Contexto

Se facilitan dos datasets. El primero con información de las pólizas y el segundo con información de los vehículos. Ambos datasets están vinculados por el VEHICLE_ID.

insurance_data.csv

- INSR_BEGIN – Fecha de inicio de la póliza
- INSR_END – Fecha de fin de la póliza
- CUSTOMER_SENIORITY – Antigüedad del cliente en la empresa
- SEX – Genero del cliente
- INSR_TYPE – Tipo de póliza
- INSURED_VALUE – Suma del valor asegurado
- PREMIUM – Cuota a pagar por la póliza
- VEHICLE_ID – Id del vehículo asegurado
- USAGE – Uso del vehículo asegurado
- CLAIM_PAID – Pago realizado al cliente por un reclamo de siniestro. Si es nulo significa que no hubo reclamo

vehicle_info.csv

- VEHICLE_ID – Id del vehículo
- PROD_YEAR – Año de fabricación
- SEATS_NUM – Cantidad de asientos
- CARRYING_CAPACITY – Capacidad de carga
- TYPE_VEHICLE – Tipo de vehículo
- CCM_TON – Potencia del motor
- MAKE – Marca del vehículo

Desafío 1

Realizar un análisis exploratorio con el objetivo de entender mejor la relación entre las pólizas y los vehículos. Al negocio le interesa especialmente conocer su cartera de clientes y conocer cuáles son los clientes que hicieron un reclamo de siniestro.

Se debe documentar paso a paso y las conclusiones obtenidas.

Desafío 2

Desarrollar un modelo predictivo que genere una clasificación binaria sobre las pólizas del archivo **test.csv**. El objetivo es poder identificar cuáles son las pólizas que realizaron un reclamo de siniestros. Se pueden utilizar todas las técnicas que consideres apropiadas para mejorar el modelo.

Se debe seleccionar que métrica se utilizara para evaluar el/los distintos modelos. Además, seleccionar con que algoritmo se creara el modelo justificando la elección.

Si bien es importante la predicción generada y la capacidad predictiva del modelo, nos interesa más el proceso y como llegaste a generar ese modelo. Sentite libre de agregar todos los comentarios que creas necesarios para que nos sea más fácil entender el proceso!

Entrega

La entrega se deberá realizar mediante un link a un repositorio de Github. El mismo deberá tener:

- Una notebook con el EDA, los comentarios y conclusiones
- Una notebook con el desarrollo del modelo, las métricas que se utilizaron para evaluar y los distintos modelos que se probaron
- Un archivo CSV con las predicciones del modelo

Todo el desarrollo deberá estar hecho en Python.

Recomendaciones

- No te trabes. Si algo no sale, dejalo y seguí adelante.
- Hace todos los comentarios que creas necesarios sobre el código. Nos sirve mucho para interpretar lo que fuiste desarrollando.