

Descripción del Problema Predictivo a Resolver:

El problema que se busca resolver se relaciona con la predicción de si un asegurado presentará un reclamo en los próximos 6 meses o no. Para ello, se utilizarán diversos atributos del asegurado y su vehículo, como la duración de la póliza, la edad del automóvil, la edad del propietario del automóvil, la densidad de población de la ciudad, la marca y modelo del automóvil, la potencia, el tipo de motor, entre otros. El objetivo es desarrollar un modelo de aprendizaje que pueda prever con precisión si un asegurado será propenso a presentar un reclamo en los próximos 6 meses.

Dataset Utilizado:

<https://www.kaggle.com/datasets/ifteshanajnin/carinsuranceclaimprediction-classification?select=train.csv>

El dataset contiene información detallada sobre los asegurados, incluyendo los atributos mencionados anteriormente, así como la variable objetivo que indica si el asegurado presentó un reclamo en los próximos 6 meses o no. Los datos se han recopilado previamente y se utilizarán para entrenar y evaluar el modelo predictivo.

Cuenta con 44 columnas de las cuales más del 10% son categóricas, algunos ejemplos son:

- **Segment:** A/B1/B2/C1/C2
- **fuel_type:** CNG/Petrol/Diesel
- **rear_brakes_type:** Drum/Disc
- **transmission_type:** Manual/Automatic
- **steering_type:** Power/Electric
- **area_cluster:**
C1/C2/C3/C4/C5/C6/C7/C8/C9/C10/C11/C12/C13/C14/C15/C16/C17/C18/C19/C20/C21/C22

Además, contiene 58592 registros (filas) de los cuales solamente usaremos 10000 para nuestro modelo en el que vamos a simular el requisito de 5% de datos faltantes en 3 columnas.

Nuestra columna objetivo se llama **is_claim**, representa los valores de 1 si el asegurado presento un reclamo del seguro y 0 si no lo hizo dentro de los últimos 6 meses.

Métricas de Desempeño Requeridas:

Para evaluar el desempeño del modelo, se utilizará el F1-score. Esta métrica ayudará a medir la capacidad del modelo para predecir correctamente los casos de reclamos y no reclamos.

Desde una perspectiva de negocio, se evaluará la métrica de "Ahorro Potencial". Esta métrica proporciona una estimación del dinero que la compañía de seguros podría ahorrar como resultado de la implementación del modelo de predicción.

El cálculo del ahorro potencial se basaría en la siguiente fórmula:

Ahorro Potencial = (Número de Reclamos Evitados) x (Costo Promedio de un Reclamo)

Donde:

"Número de Reclamos Evitados" es el número de reclamos que el modelo predice como probables y, por lo cual se evita asegurar al vehículo de ese cliente.

"Costo Promedio de un Reclamo" es el costo promedio que la compañía de seguros normalmente paga por un reclamo hecho por un asegurado.

Desempeño Deseable en Producción:

El modelo debe ser capaz de reducir la cantidad de reclamos en mínimo un 20%, esto para que el ahorro potencial cubra tanto los gastos del modelo como los gastos en las reclamaciones futuras.