

Primera entrega del proyecto

POR:

Juan Sebastián Ortiz Tangarife

MATERIA:

Introducción a la Inteligencia Artificial

PROFESOR:

Raúl Ramos Pollan



Universidad de Antioquia

Facultad de Ingeniería

Medellín 2023

1. Planteamiento del problema

Dadas ciertas características anónimas evaluadas a los clientes de una empresa de pólizas de seguros, se quiere predecir la probabilidad con la que un cliente va a realizar un reclamo sobre esta póliza.

2. Dataset

Voy a utilizar el dataset de Kaggle

<https://www.kaggle.com/competitions/tabular-playground-series-sep-2021/overview>

que consta de 120 columnas anonimizadas pero que tienen propiedades relacionadas con características del mundo real.

3. Métricas

La métrica de machine learning que se va a utilizar es el área bajo la curva ROC que sirve para evaluar el rendimiento de un modelo de clasificación binaria. Por ejemplo si el AUC-ROC me da muy cercano a 1 significa que el modelo es capaz de distinguir correctamente y sin errores las dos clases (Si realizó o no un reclamo). Como métrica de negocio se va a usar el decremento de estos reclamos a lo largo del tiempo.

4. Desempeño

Como la empresa se preocupa mucho por la satisfacción del cliente y quiere predecir con antelación si va a realizar un reclamo, La AUC-ROC deberá ser mayor a 0.75 para ser aceptada ya que de lo contrario no tendría mucho sentido hacer estas predicciones porque sería casi que lo mismo si se realizara una clasificación de forma aleatoria.

5. Bibliografía

<https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>

<https://www.kaggle.com/competitions/tabular-playground-series-sep-2021/overview>