

# **Proyecto Claims Severity Prediction**

## **ENTREGA 1**

### **INTEGRANTES:**

YOHEL OSVALDO PEREZ GARCIA  
TATIANA ELIZABETH SÁNCHEZ SANIN  
DANIELA ANDREA PAVAS BEDOYA

### **MATERIA:**

Introducción a la Inteligencia Artificial para las Ciencias e Ingenierías

### **Profesor:**

RAUL RAMOS POLLAN

UNIVERSIDAD DE ANTIOQUIA  
FACULTAD DE INGENIERÍA

## 1. Planteamiento del Problema

Para abordar este problema, una compañía de seguros puede recopilar y analizar datos sobre reclamos pasados para identificar patrones o tendencias que puedan ayudar a predecir la gravedad de reclamos futuros. Esto puede implicar analizar una variedad de variables, que incluyen datos demográficos, información de la póliza y detalles sobre la naturaleza del reclamo.

Al usar algoritmos de aprendizaje automático para analizar estas variables, las compañías de seguros pueden construir modelos que puedan predecir con precisión la gravedad de los reclamos de seguros. Esto puede ayudar a las compañías de seguros a evaluar mejor el riesgo asociado con diferentes pólizas y tomar decisiones más informadas sobre cómo gestionar los reclamos. En última instancia, esto puede ayudar a las compañías de seguros a proporcionar servicios más eficaces y eficientes a sus clientes.

## 2. Datos

Los datos provienen de una competencia de Kaggle, (<https://www.kaggle.com/competitions/allstate-claims-severity/data>), que contiene datos simulados de reclamos de seguros y detalles de los clientes. El conjunto de datos contiene 188,318 registros y 132 características, incluidas las características categóricas y numéricas.

El conjunto de datos de entrenamiento consta de 188,318 observaciones y 131 variables que incluyen 72 variables categóricas binarias, 43 variables no binarias con 3-326 niveles, 14 variables continuas y la variable de resultado, "pérdida". Dado que todas las variables predictoras se anonimizan, no se divulga información específica sobre ellas. No hay valores perdidos. También se observa un conjunto de datos de prueba con 125,546 filas.

## 3. Métricas

La métrica de desempeño de machine learning utilizada será el Mean Absolute Error (MAE), ya que es la métrica utilizada en la competencia.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Donde  $n$  es la cantidad de datos,  $y_i$  los datos observados y  $\hat{y}_i$  los valores pronosticados.

Como métrica de negocio, se podría utilizar la reducción del costo de los reclamos para la compañía de seguros.

#### **4. Desempeño**

Un desempeño deseable en producción podría ser tener un modelo que pueda predecir el monto de la reclamación con un MAE de menos de 1200 unidades monetarias (la métrica utilizada en la competencia). De esta forma poder hacer un mejor análisis de la gravedad y costo de los reclamos realizados por los clientes y permitir la toma de decisiones orientada a la reducción de estos en al menos un 5% en comparación con el modelo existente o sin modelo alguno.

#### **5. Bibliografía**

Allstate Insurance, "Allstate Claims Severity," Kaggle, 2016. [Online]. Available:

<https://www.kaggle.com/competitions/allstate-claims-severity/overview>.

[Accessed: 11-Mar-2023].

Prof, C. (2022, 3 enero). 3 Ways to Calculate the Mean Absolute Error (MAE) in R

[Examples]. CodingProf.com. <https://www.codingprof.com/3-ways-to-calculate-the-mean-absolute-error-mae-in-r-examples/>