

Primera Entrega

POR

Jhon Alexander Botero Gómez

Giovani Cardona Marín

Sebastian Mora Piedrahita

CURSO

Introducción a la Inteligencia Artificial

PROFESOR

Raul Ramos Pollan



**UNIVERSIDAD
DE ANTIOQUIA**

1 8 0 3

UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE INGENIERÍA
MEDELLÍN 2023

1. Planteamiento del problema.

El problema planteado se centra en la necesidad de mejorar el proceso de reclamaciones de seguros en el contexto de accidentes automovilísticos graves. Cuando una persona se enfrenta a las consecuencias emocionales y físicas de un accidente, su prioridad es su bienestar y el apoyo de su familia y amigos. El proceso de presentar reclamaciones de seguros se convierte en una tarea incómoda y mentalmente agotadora que nadie desea enfrentar en ese momento. Es por esto que se desea desarrollar un algoritmo preciso para predecir la gravedad de los siniestros, lo que no solo beneficiará a la aseguradora, sino que también brindará a los asegurados la tranquilidad de saber que recibirán el apoyo adecuado en momentos difíciles.

2. Dataset.

El dataset a utilizar proviene de una competencia de kaggle la cual tiene 188318 muestras y 132 columnas. El dataset está compuesto por 3 archivos .csv que proporcionan la información requerida.

id Identificador de cada muestra que representa un siniestro

cat Variables categoricas.

cont Variables continuas.

Descripción de archivos

- **train.csv** - el conjunto de entrenamiento
- **test.csv** - el conjunto de prueba. Debe predecir el valor de pérdida para los identificadores de este archivo.
- **sample_submission.csv** - un archivo de envío de muestra en el formato correcto

3. Métricas.

La métrica de evaluación del modelo es el error absoluto medio (EAM) o por sus siglas en inglés (MAE)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n \underbrace{|y_i - \hat{y}_i|}_{\text{predicted value} - \text{actual value}}$$

test set

Donde y_i es el valor predicho y \hat{y}_i el valor verdadero y por lo tanto el valor absoluto de la diferencia $y_i - \hat{y}_i$ es igual al error absoluto $|e_i| = |y_i - \hat{y}_i|$.

Con respecto a la métrica del negocio, se considerará que el valor de pérdida estimado se encuentre en un rango entre 5% menos y 5% más del valor real. La precisión debe ser la suficiente para que la compañía pueda empezar a ofrecer este servicio al público con total seguridad de que aportará el valor suficiente.

4. Desempeño.

El modelo de predicción desarrollado en este proyecto tiene como objetivo principal estimar la gravedad de las reclamaciones de seguros para Allstate. Se basa en métodos avanzados de análisis de datos y algoritmos de aprendizaje automático para proporcionar estimaciones precisas de pérdida en un siniestro grave. Este modelo se va a destacar por su capacidad para reducir de manera efectiva el error absoluto medio (MAE), lo que resulta en una mejora significativa en la precisión de las predicciones de reclamaciones, contribuyendo así a la eficiencia y calidad del proceso de reclamaciones de la compañía.

5. Bibliografía.

- DanaFerguson, Meg Risdal, NoTrick, Sara R, Sillah, Tim Emmerling, Will Cukierski. (2016). Allstate Claims Severity. Kaggle.
<https://kaggle.com/competitions/allstate-claims-severity>.