

Primera entrega de proyecto

POR:

Jorge Luis Gonzalez Morelo

MATERIA:

Introducción a la inteligencia artificial

PROFESOR:

Raul Ramos Pollan



UNIVERSIDAD DE ANTIOQUIA

FACULTAD DE INGENIERÍA

MEDELLÍN 2021

1. Planteamiento del problema

En cualquier parte del mundo, muchas personas son víctimas cada año en accidentes de tránsito, ya sea, por imprudencia o confusiones. Lo cierto es que los datos estadísticos muestran que se han convertido en una gran problemática para los países del mundo, ya que no sólo deben lidiar con la circulación de tránsito, sino también con la salud de las personas afectadas.

La importancia del buen manejo de los datos puede ser vital en situaciones que impliquen riesgo en la seguridad de las personas como por ejemplo los accidentes automovilísticos, mediante este dataset se buscará predecir la accidentalidad teniendo en cuenta las causas y consecuencias, para poder trabajar desde la prevención y reducir la tasa de accidentalidad

2. Dataset

El dataset que se va a utilizar proviene de Kaggle, el cual está compuesto por un archivo CSV con datos de accidentes de tránsito ocurridos en Reino Unido desde el año 2005 al 2014, recolectados por el gobierno del Reino Unido.

El dataset contiene un registro de más de 1.8 millones de accidentes, sin embargo para el proyecto solo se usarán los datos más recientes, es decir los correspondientes al año 2014, los cuales son 146322 accidentes en total, y contiene la siguiente información:

- 1) Accident_Index,
- 2) Location_Easting_OSGR,
- 3) Location_Northing_OSGR,
- 4) Longitude,
- 5) Latitude,
- 6) Police_Force,
- 7) Accident_Severity,
- 8) Number_of_Vehicles,
- 9) Number_of_Casualties,
- 10) Date,
- 11) Day_of_Week,
- 12) Time,
- 13) Local_Authority_(District),
- 14) Local_Authority_(Highway),
- 15) 1st_Road_Class,
- 16) 1st_Road_Number,
- 17) Road_Type,
- 18) Speed_limit,
- 19) Junction_Control,
- 20) 2nd_Road_Class,
- 21) 2nd_Road_Number,
- 22) Pedestrian_Crossing-Human_Control,
- 23) Pedestrian_Crossing_Physical_Facilities,Light_Conditions,
- 24) Weather_Conditions,
- 25) Road_Surface_Conditions,
- 26) Special_Conditions_at_Site,

- 27) Carriageway_Hazards,
- 28) Urban_or_Rural_Area,
- 29) Did_Police_Officer_Attend_Scene_of_Accident,
- 30) LSOA_of_Accident_Location,
- 31) Year

3. Métricas

La principal métrica que se utilizará en el modelo de predicción de accidentes de tránsito es la Raíz del Error Cuadrático Medio o Root Mean Square Error (RMSE) y se calcula de la siguiente manera

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} = \sqrt{MSE}$$

Donde RMSE es la raíz cuadrada del promedio de la suma de diferencias cuadradas entre los valores observados en la serie y los esperados según el modelo de tendencia. Donde y_i corresponde a la serie observada e \hat{y}_i a la serie estimada, y N el número de datos totales. En los resultados se busca que mientras menor sea el error, más adecuado es el modelo.

4. Bibliografía

- <https://www.kaggle.com/datasets/devansodariya/road-accident-united-kingdom-uk-dataset?resource=download>