

**INTRODUCCIÓN ARTIFICIAL PARA LAS CIENCIAS E INGENIERÍAS
PRIMERA ENTREGA DE PROYECTO**

Docente:

Raúl Ramos Pollan

Presentado por:

Diego Andrés Flórez Ruano



Universidad de Antioquia

Facultad de Ingeniería

Medellín, Colombia

2022

1. PLANTEAMIENTO DEL PROBLEMA

Los accidentes de tránsito son la principal causa de muerte en niños de 5 a 14 años y jóvenes de 15 a 29 años. Además, son la octava causa de muerte para todos los grupos de edad. En muchos casos, estos accidentes son ocasionados por problemáticas como el estado de la vía, las condiciones de iluminación, entre otros, y gran parte de las víctimas fatales corresponden a motociclistas; una de las causas más subestimadas es la falta de carriles específicos para este tipo de vehículos automotores de 2 ruedas. Es por esto que es importante realizar un seguimiento a los accidentes de tránsito ocurridos en el último tiempo y establecer un modelo que permita conocer las causas más comunes de los distintos accidentes en los que se vieron involucrados motociclistas, con el fin de implementar estrategias específicas que permitan evitar tales sucesos como el mejoramiento de las carreteras, iluminación, etc.

2. DATASET

El dataset a utilizar proviene de Kaggle (<https://www.kaggle.com/code/sohamsave/accident-data-analysis/data>) en el cual se proporcionan datos de distintos accidentes ocurridos entre 2005 y 2014, y detalla las circunstancias de dichos accidentes tales como el clima, la ubicación latitud/longitud, el tipo de área, el tipo de carretera, la fecha, la hora, el día de la semana, entre otros. El dataset está compuesto por un conjunto de archivos .csv que proporcionan la información general de los accidentes, de las víctimas y de los vehículos implicados en estos mismos.

El archivo que contiene los datos del accidente se denomina “Accidents0514” y presenta las siguientes columnas:

- **Accident_Index:** Identificador para cada accidente.
- **Location_Easting_OSGR:** Coordenada cartesiana medida hacia el Este.
- **Location_Northing_OSGR:** Coordenada cartesiana medida hacia el Norte.
- **Accident_Severity:** Gravedad del accidente.
- **Number_of_Vehicles:** Número de vehículos implicados en el accidente.
- **Number_of_Casualties:** Número de víctimas.
- **Road_Surface_Conditions:** Condiciones de la carretera
- **Light_Conditions:** Condiciones de luz.

Por otro lado, el archivo que contiene información de los vehículos implicados en el accidente se denomina “Vehicles0514” y contiene las siguientes columnas:

- **Accident_Index:** Identificador para cada accidente.
- **Vehicle_Type:** Tipo de vehículo accidentado.
- **Vehicle_Manoeuvre:** Maniobra realizada por el vehículo.
- **Vehicle_Location-Restricted_Lane:** Ubicación del vehículo accidentado

- **1st_Point_of_Impact:** Primer punto de impacto en la colisión.

Finalmente, el archivo que contiene información de las víctimas del accidente se denomina “Casualties0514” y presenta las siguientes columnas:

- **Accident_Index:** Identificador para cada accidente.
- **Casualty_Class:** Tipo de víctimas del accidente.
- **Sex_of_Casualty:** Sexo de la víctima
- **Age_of_Casualty:** Edad de la víctima
- **Casualty_Severity:** Gravedad de la víctima.

Es importante anotar que se contará con un archivo .xls denominado como “Road-Accident-Safety-Data-Guide” que servirá como guía para mapear las variables codificadas en los distintos valores de los dataset anteriormente descritos.

3. MÉTRICAS

La métrica de evaluación principal para el modelo propuesto será el Root Mean Squared Logarithmic Error (RMSLE) que mide la desviación estándar de los residuos (errores de predicción). El RMSLE se calcula de la siguiente manera:

$$\epsilon = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(p_i + 1) - \log(a_i + 1))^2}$$

En donde:

$\epsilon = RMSLE$

$n =$ Total del número de observaciones en el dataset

$p_i =$ Predicción de la variable a evaluar

$a_i =$ Valor actual de la variable a evaluar

En cuanto a la métrica de negocio, se podrían evitar un gran número de accidentes de tránsito de motociclistas gracias a la utilización del modelo, lo que permitiría a las diferentes entidades gubernamentales evaluar la implementación de diferentes mecanismos como la inversión en infraestructura vial en conjunto con campañas de educación vial.

4. DESEMPEÑO

Con el modelo planteado se esperaría conocer con qué frecuencia se repiten accidentes de tránsito que involucran motociclistas, en qué lugares y qué factores externos influyen para que se ocasionaron dichas colisiones (estado de la vía, tipo de

vía, condiciones de iluminación, etc). Esto con el fin de establecer distintas estrategias por parte de las autoridades locales que permitan minimizar el riesgo de posibles accidentes por causas repetitivas, así como realizar un estudio financiero con el fin de invertir en proyectos que permitan mejorar la seguridad de los motociclistas para evitar más muertes ocasionadas por accidentes de tránsito. Se esperaría que el modelo propuesto tenga un porcentaje de acierto superior al 50%.

5. BIBLIOGRAFÍA:

- <https://www.kaggle.com/code/sohamsave/accident-data-analysis/data?select=Casualties0514.csv>
- <https://www.kaggle.com/competitions/nyc-taxi-trip-duration/overview/evaluation>