

PRIMERA ENTREGA DEL PROYECTO

ALEJANDRO ARIAS ORTIZ

LUIS MATEO OCHOA AGUDELO

DUVAN ESNEIDER GALLEGUO JIMENEZ

RAUL RAMOS POLLAN



MODELOS DE SISTEMAS 1 / INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL

UNIVERSIDAD DE ANTIOQUIA

FACULTAD DE INGENIERÍA

MEDELLÍN 2023

1. PLANTEAMIENTO DEL PROBLEMA:

En Estados Unidos y en todo el mundo, muchas personas se convierten en víctimas de accidentes de tráfico cada año, ya sea debido a acciones imprudentes o confusiones. Es innegable que las estadísticas demuestran que este problema se ha convertido en un desafío significativo para el país, ya que no solo debe abordar la gestión del tráfico, sino también atender a la salud de las personas afectadas por estos incidentes. La importancia de la adecuada gestión de los datos se vuelve crucial en situaciones que implican riesgos para la seguridad de las personas, como los accidentes de tráfico en Estados Unidos. A través de este conjunto de datos, se busca predecir el porcentaje de accidentalidad teniendo en cuenta sus causas y consecuencias, con el fin de centrarse en la prevención y reducir la tasa de accidentes en el país.

2. DATASET:

El dataset que se va a utilizar proviene de Kaggle, el cual contiene El conjunto de datos de accidentes de tráfico de EE. UU. Además contiene información detallada sobre 259,077 accidentes automovilísticos que ocurrieron en todos los estados de EE. UU. (excepto áreas remotas de Alaska y Hawái) entre 2016 y 2020. Los conjuntos de datos tienen información muy detallada sobre cada accidente, incluida la ubicación, la hora, el clima y los factores que contribuyeron. e información demográfica de los automovilistas involucrados en el accidente.

3. METRICAS:

La mejor métrica de desempeño para predecir un índice de accidentabilidad basado en un dataset que contiene accidentes reales es la raíz del error cuadrático medio (RMSE). La cual mide la diferencia promedio entre los valores predichos y los valores reales, en este caso, los valores predichos serían el índice de accidentabilidad, y los valores reales serían los índices de accidentabilidad observados en el dataset.

$$RMSE = \sqrt{(\sum (y_i - \hat{y}_i)^2 / n)}$$

donde:

y_i es el valor real del índice de accidentabilidad en el i-ésimo punto de datos

\hat{y}_i es el valor predito del índice de accidentabilidad en el i-ésimo punto de datos

n es el número de puntos de datos en el dataset

El dataset fue sacado de:

<https://www.kaggle.com/datasets/jonbown/us-2020-traffic-accidents>

4. DESEMPEÑO DESEABLE:

Un primer criterio sobre cuál sería el desempeño deseable en producción para un modelo de predicción de índices de accidentabilidad sería que la RMSE sea inferior a 1. Esto significaría que, en promedio, los valores predichos están a una distancia de 1 de los valores reales.

Un valor de RMSE inferior a 1 es deseable porque indica que el modelo está prediciendo los índices de accidentabilidad con un alto grado de precisión. Esto significa que el modelo es capaz de predecir los índices de accidentabilidad con un error relativamente pequeño.