

How Severity the Accidents is ?

CRISTIAN ALEXANDER CASTAÑO MONTOYA WILSON ALBERTO VILLA MUÑETON
LUIS FELIPE CADAVID CHICA

Los accidentes de tránsito ocurren diariamente, vamos a analizar los accidentes ocurridos en los estados unidos en un periodo de 5 años (2016-2021), dadas unas características (Street, Temperature(F), Visibility(mi), Wind_Speed(mph), Weather_Condition, Crossing, Traffic_Signal, etc.) que nos ayudarán a comprender que tan grave fue el accidente ocurrido.

Para esto vamos a usar un dataset de [Kaggle](#), que contiene 1.048.575 accidentes (filas) que será modificado según la necesidad del proyecto y 42 variables (columnas) que se puede observar a continuación:

Severity,	Wind_Chill(F),	Roundabout,
Start_Time, End_Time,	Humidity(%),	Station, Stop,
Start_Lat, Start_Lng,	Pressure(in),	Traffic_Calming,
End_Lat, End_Lng,	Visibility(mi),	Traffic_Signal,
Distance(mi),	Wind_Direction,	Turning_Loop
Description, Number,	Wind_Speed(mph),	, Sunrise_Sunset,
Street Side, City County,	Precipitation(in),	Civil_Twilight,
State Zipcode, Country,	Weather_Condition,	Nautical_Twilight,
Timezone, Airport_Code,	Amenity, Bump, Crossing,	Astronomical_Twilight
Weather_Timestamp,	Give_Way, Junction,	
Temperature(F),	No_Exit, Railway,	

Nuestro problema de machine learning, es de clasificación, pues nuestra variable a predecir "Severity" está definida en 4 valores de 1 a 4 respectivamente, que definen que tan grave es el accidente. Teniendo en cuenta esto, nuestra métrica de machine learning va a ser el accuracy, el cual mide la proporción de predicciones correctas que realiza el modelo en comparación con el total de predicciones realizadas, expresada como porcentaje. Entre mayor sea el porcentaje de nuestro accuracy, mejor será el rendimiento del modelo. Por ejemplo, si el *severity* de un accidente es de 2 y nuestro modelo predice que es 3, y así sucesivamente con distintos datos, al final se sumará y obtendremos cuántos fueron las buenas predicciones y cuántas fueron las malas y obtendremos nuestro accuracy.

Debemos de tener un accuracy mínimo del 90%, solo así valdría la pena tener nuestro modelo funcionando. Que las predicciones que haga nuestro modelo después del entrenamiento, las buenas superen el 90%.