

SEGUNDA 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

## Contenido

Introducción.....	3
Exploración descriptiva del Dataset.....	4
Métrica.....	6
Iteraciones de desarrollo. ....	7
Preprocesado de datos.....	7
Modelos supervisados.....	7
Modelos no supervisados.....	7
Resultados, métricas y curvas de aprendizaje.....	7
Retos y consideraciones de despliegue.....	8
Conclusiones.....	8

## Introducción.

Los accidentes de tránsito son un serio problema que afecta a muchas personas cada año. Estos accidentes, a menudo causados por imprudencias o condiciones de los conductores, se han convertido en una preocupación importante para los países. Además de enfrentar el reto de controlar el tráfico, también hay que lidiar con el impacto que estos accidentes tienen en la salud de los involucrados.

La gestión adecuada de los datos es clave en este escenario, sobre todo cuando se trata de riesgos relacionados con la seguridad vial, como los accidentes de coches. Con un conjunto de datos exhaustivo, se busca pronosticar la probabilidad de accidentes considerando sus causas y efectos, lo que permitirá enfocar los esfuerzos en la prevención y reducción de accidentes, así como proveer información vital a los interesados, como hospitales, para que puedan anticiparse a un posible incremento en la llegada de pacientes debido a accidentes de tráfico.

La inteligencia artificial (IA) surge como una solución prometedora en este ámbito, ya que puede procesar y analizar grandes cantidades de datos para extraer información valiosa. Con la IA, se busca obtener conocimientos que permitan acciones más efectivas contra las causas y efectos de los accidentes viales, y se espera que también facilite la toma de decisiones en la planificación y respuesta médica. En última instancia, la combinación de IA y análisis de datos tiene un gran potencial para mejorar la prevención de accidentes de tránsito y reducir su impacto negativo.

## Exploración descriptiva del Dataset

El conjunto de datos de accidentes de tráfico de EE. UU. contiene información detallada sobre 259.077 accidentes automovilísticos que ocurrieron en todos los estados de EE. UU. (excepto en 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, los factores contribuyentes y la información demográfica de los automovilistas involucrados en el accidente. El programa de muestreo se vio profundamente afectado por la pandemia COVID19, por lo que no hay nada más allá del año 2020 debido a una revisión importante del programa para acomodar el trabajo remoto. Los datos de 2020 pueden ser muy escasos para los meses de marzo a mayo, dado el hecho de que estos informes provienen de los departamentos de policía locales y muchos de ellos se vieron afectados en la primavera de 2020.

La base de datos proporciona una gran cantidad de información relevante sobre cada accidente registrado. Algunos de los datos incluidos son:

1. Ubicación: Latitud y longitud del lugar donde ocurrió el accidente.
2. Fecha y hora: Información sobre cuándo ocurrió el accidente.
3. Tipo: Clasificación del tipo de accidente, como colisión de vehículos, atropello, choque con un objeto fijo, etc.
4. Condiciones: Descripción de las condiciones climáticas en el momento del accidente.
5. Estado: Información sobre el estado de la carretera, como seca, mojada, helada, etc.
6. Contribuciones: Factores que se consideran que contribuyen al accidente, como la velocidad, el consumo de alcohol, el uso del cinturón de seguridad, etc.
7. Gravedad: Indicación de la gravedad del accidente en términos de personas fallecidas, heridas graves o heridas leves.

El análisis de estos datos proporcionará información valiosa sobre las causas y consecuencias de los accidentes de tráfico. Esto a su vez facilitará el desarrollo de estrategias de prevención y la toma de decisiones informadas para mejorar la seguridad vial y reducir la tasa de accidentes en el futuro.

STRATUM: Podría ser una clasificación o estratificación de los datos.
REGION: La región geográfica donde ocurrió el accidente.
URBANCITYNAME: Indica si el área es urbana o rural.
VE_TOTAL: Número total de vehículos involucrados en el accidente.

NUM_INJNAME: Número de personas heridas en el accidente.
MONTHNAME: El mes en que ocurrió el accidente.
YEARNAME: El año del accidente.
DAY_WEEKNAME: Día de la semana cuando ocurrió el accidente.
HOURLNAME: Hora del día cuando ocurrió el accidente.
MINUTENAME: Minuto exacto cuando ocurrió el accidente.
HARM_EV: Código o identificador del evento dañino principal en el accidente.
HARM_EVNAME: Descripción del evento dañino principal.
ALCOHOL: Indicador de si el alcohol estuvo involucrado en el accidente.
ALCOHOLNAME: Descripción de la implicación del alcohol.
MAX_SEV: Gravedad máxima del accidente.
MAX_SEVNAME: Descripción de la gravedad máxima del accidente.
MAN_COLL: Tipo de colisión (por ejemplo, frontal, trasera, lateral).
MAN_COLLNAME: Descripción del tipo de colisión.
RELJCT1NAME: Información sobre si el accidente ocurrió en una intersección.
RELJCT2NAME: Información más específica sobre la relación con la intersección.
TYP_INT: Tipo de intersección donde ocurrió el accidente.
WRK_ZONENAME: Indica si el accidente ocurrió en una zona de trabajo.
REL_ROADNAME: Relación del accidente con la carretera (en la carretera, adyacente, etc.).
LGT_CONDDNAME: Condiciones de luz en el momento del accidente.
SCH_BUSNAME: Indica si un autobús escolar estuvo involucrado en el accidente.
INT_HWYNAME: Indica si el accidente ocurrió en una carretera interestatal.
WEATHERNAME: Condiciones meteorológicas en el momento del accidente.
WKDY_IMNAME: Posiblemente una codificación del día de la semana.
HOURL_IMNAME: Posiblemente una codificación de la hora del accidente.
MINUTE_IMNAME: Posiblemente una codificación del minuto del accidente.
EVENT1_IMNAME: Primer evento importante registrado en el accidente.
MANCOL_IMNAME: Codificación del tipo de colisión.
RELJCT1_IMNAME: Codificación de la relación del accidente con la intersección.
RELJCT2_IMNAME: Codificación de información más específica sobre la intersección.
LGTCOON_IMNAME: Codificación de las condiciones de luz.
WEATHR_IMNAME: Codificación de las condiciones meteorológicas.
MAXSEV_IMNAME: Codificación de la gravedad máxima del accidente.
NO_INJ_IMNAME: Número de personas no heridas o codificación de este dato.
ALCHL_IMNAME: Codificación de la implicación del alcohol en el accidente.
WEIGHT: Peso o factor de ponderación para estadísticas o muestras.

## Métrica

La métrica principal empleada en el modelo de predicción de accidentes de tránsito es el Error Cuadrático Medio (RMSE, por sus siglas en inglés), que se calcula de la siguiente manera:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum (y_i - \hat{y}_i)^2}$$

El Error Cuadrático Medio (RMSE) es la métrica principal utilizada en el modelo de predicción de accidentes de tránsito. Se calcula como la raíz cuadrada del promedio de la suma de las diferencias al cuadrado entre los valores observados en la serie y los valores esperados según el modelo de tendencia.

Donde  $y_i$  corresponde a los valores observados en la serie y  $\hat{y}_i$  representa los valores estimados por el modelo.  $N$  representa el número total de datos en la serie.

El RMSE se utiliza como una medida de la discrepancia entre los valores observados y los valores estimados. Cuanto menor sea el valor del RMSE, más adecuado será el modelo de predicción, ya que indicará que las predicciones se acercan más a los valores reales.

Al utilizar el RMSE como métrica de evaluación, se busca obtener un modelo de predicción que minimice el error y se ajuste de manera precisa a los datos de accidentes de tránsito. Esto permitirá realizar predicciones más precisas y confiables, lo que a su vez facilitará la toma de decisiones informadas y la implementación de estrategias efectivas para prevenir accidentes de tránsito en el futuro.

## Iteraciones de desarrollo.

Preprocesado de datos

Modelos supervisados

Modelos no supervisados

Resultados, métricas y curvas de aprendizaje.

## Retos y consideraciones de despliegue

El modelo de predicción de víctimas en accidentes automovilísticos tiene dificultad para predecir con precisión el número de víctimas. Esto se debe a que el dataset actual no incluye toda la información necesaria, como las condiciones climáticas, el estado de la vía o la presencia de señales de tránsito. Para mejorar el modelo, es necesario recopilar más datos y evaluarlo con profesionales de la salud y servicios de emergencia.

El modelo actual no incluye información sobre factores que pueden afectar el número de víctimas, como las condiciones climáticas o el estado de la vía. Para mejorar el modelo, es necesario recopilar datos sobre estos factores. Esto podría implicar costos adicionales, ya que se necesitaría realizar encuestas o recopilar información de fuentes secundarias.

Evaluación con los profesionales de la salud y servicios de emergencia: Es importante evaluar el modelo con profesionales de la salud y servicios de emergencia para determinar si es preciso y útil. Esta evaluación podría realizarse mediante simulaciones o pruebas piloto.

## Conclusiones

Para mejorar el rendimiento y reducir el sesgo del modelo de predicción de accidentes de tránsito, es necesario obtener más datos representativos. Estos datos deben ser más completos y diversos para que los modelos puedan capturar mejor la variabilidad de los accidentes. También es importante considerar otros modelos disponibles en el campo de la predicción de accidentes de tránsito, ya que cada modelo tiene sus propias fortalezas y debilidades. Por último, es posible que sea necesario aplicar técnicas de preprocesamiento y manejo de datos para abordar el sesgo presente en los datos.

1. Obtención de más datos representativos: El dataset actual no es representativo de la variabilidad de los accidentes de tránsito. Esto se debe a que los datos actuales están sesgados hacia accidentes con pocas víctimas. Para mejorar el rendimiento del modelo, es necesario obtener más datos que reflejen la realidad de los accidentes de tránsito. Estos datos deben incluir accidentes con muchas víctimas, así como accidentes con pocas víctimas.



2. Selección del modelo: Los tres modelos evaluados inicialmente tienen resultados similares. Sin embargo, es importante considerar otros modelos disponibles en el campo de la predicción de accidentes de tránsito. Cada modelo tiene sus propias fortalezas y debilidades, y probar diferentes enfoques permitirá tener una visión más completa y robusta de las posibles soluciones.
3. Tratamiento del sesgo: El sesgo en los datos puede estar influenciado por la propia naturaleza de los accidentes de tránsito. Si existe una acumulación significativa de valores cercanos a 1 en los datos, esto puede plantear desafíos para que los modelos "aprendan" de manera efectiva. Es posible que los modelos tengan dificultades para capturar y generalizar patrones en los datos debido a esta falta de variabilidad. Por lo tanto, es fundamental considerar técnicas de preprocesamiento y manejo de datos que aborden este sesgo.

La mejora del modelo de predicción de accidentes de tránsito requiere superar los desafíos de la obtención de datos representativos, la selección del modelo y el tratamiento del sesgo. Sin embargo, un modelo preciso podría ayudar a mejorar la seguridad vial.