



**UNIVERSIDAD AUTÓNOMA DE NUEVO
LEÓN**



FACULTAD DE CIENCIAS FÍSICO- MATEMÁTICAS

Unidad de Aprendizaje:

Minería de datos

AVANCE I PROYECTO INTEGRADOR

Maestro:

Berrones Reyes Mayra Cristina

Alumnos:

Sifuentes Soto Aranza 1887979

Contreras Garcia Gabriel Adrian 1752950

Solis Agundis Valeria 1815413

Equipo: 12

Grupo: 002

Frecuencia: miércoles 7-10 pm

Licenciatura en Actuaría

San Nicolás de los Garza, Nuevo León, 28 de octubre de 2020

1.- Título de la base de datos

Base de datos: US Accidents (3.5 million records)

Url: <https://www.kaggle.com/sobhanmoosavi/us-accidents>

2.- Descripción de los datos.

Nuestro conjunto de datos recopila información sobre accidentes automovilísticos en Estados Unidos, donde cubre 49 estados. Los datos de accidentes se recopilan desde febrero de 2016 hasta junio de 2020, utilizando dos API que proporcionan datos de incidentes (o eventos) de tráfico en tiempo real. Estas API transmiten datos de tráfico capturados por una variedad de entidades, como los departamentos de transporte de los Estados Unidos y el estado, agencias de aplicación de la ley, cámaras de tráfico y sensores de tráfico dentro de las redes de carreteras. Actualmente, hay alrededor de 3,5 millones de registros de accidentes en este conjunto de datos.

Los datos recopilados van desde las condiciones del clima en el que se desarrolló el accidente, hasta la hora, fecha en que se llevó a cabo, teniendo datos históricos en nuestra base de datos.

La base de datos cuenta con 49 columnas, a continuación se describirán qué representa cada columna

1. Se registra un identificador único del registro de accidentes.
2. Se indica la fuente del informe del accidente (es decir, la API que informó el accidente).
3. Un accidente de tráfico puede tener un código de canal de mensajes de tráfico (TMC) que proporciona una descripción más detallada del evento, muestra el código.
4. Muestra la gravedad del accidente, un número entre 1 y 4, donde 1 indica el menor impacto en el tráfico (es decir, un retraso corto como resultado del accidente) y 4 indica un impacto significativo en el tráfico (es decir, un retraso largo).
5. Muestra la hora de inicio del accidente en la zona horaria local.
6. Muestra la hora final del accidente en la zona horaria local. La hora final aquí se refiere a cuando el impacto del accidente en el flujo de tráfico fue descartado.
7. Muestra la latitud en coordenadas GPS del punto de inicio.
8. Muestra la longitud en coordenadas GPS del punto de inicio.
9. Muestra la latitud en coordenadas GPS del punto final.
10. Muestra la longitud en coordenadas GPS del punto final.
11. La longitud de la extensión de la carretera afectada por el accidente en millas.
12. Muestra una descripción del accidente.
13. Muestra el número de la calle en el registro de dirección.
14. Muestra el nombre de la calle en el registro de direcciones.
15. Muestra el lado relativo de la calle (derecha / izquierda) en el registro de direcciones. Registrando solo R(derecha), L(izquierda) así sea el caso.
16. Muestra la ciudad en el registro de direcciones.
17. Muestra el condado en el registro de direcciones.
18. Muestra el estado en el registro de direcciones.
19. Muestra el código postal en el registro de direcciones.
20. Muestra el país en el registro de direcciones.
21. Muestra la zona horaria según la ubicación del accidente (este, centro, etc.).

22. Denota el aeropuerto que está más cerca de donde ocurrió el accidente.
23. Muestra la marca de tiempo del registro de observación meteorológica (en hora local).
24. Muestra la temperatura (en Fahrenheit).
25. Muestra la sensación térmica (en grados Fahrenheit).
26. Muestra la humedad (en porcentaje).
27. Muestra la presión del aire (en pulgadas).
28. Muestra visibilidad (en millas).
29. Muestra la dirección del viento.
30. Muestra la velocidad del viento (en millas por hora).
31. Muestra la cantidad de precipitación en pulgadas, si hay alguna.
32. Muestra las condiciones meteorológicas (lluvia, nieve, tormenta, niebla, etc.)
33. Una anotación de POI que indica la presencia de comodidades en un lugar cercano, donde se registra verdadero o falso según corresponda.
34. Una anotación de POI que indica la presencia de un tope o una joroba en una ubicación cercana, donde se registra verdadero o falso según corresponda.
35. Una anotación de POI que indica la presencia de cruce en una ubicación cercana, donde se registra verdadero o falso según corresponda.
36. Una anotación de POI que indica la presencia de una señal de dar el paso en una ubicación cercana, donde se registra verdadero o falso según corresponda.
37. Una anotación de POI que indica la presencia de una intersección en una ubicación cercana, donde se registra verdadero o falso según corresponda.
38. Una anotación de POI que indica la presencia de una señal de no salida en una ubicación cercana, donde se registra verdadero o falso según corresponda.
39. Una anotación de POI que indica la presencia de vías férreas en una ubicación cercana, donde se registra verdadero o falso según corresponda.
40. Una anotación de POI que indica la presencia de una rotonda en una ubicación cercana, donde se registra verdadero o falso según corresponda.
41. Una anotación de POI que indica la presencia de una estación en una ubicación cercana, donde se registra verdadero o falso según corresponda.
42. Una anotación de POI que indica la presencia de una parada en una ubicación cercana, donde se registra verdadero o falso según corresponda.
43. Una anotación de POI que indica la presencia de traffic_calming en una ubicación cercana, donde se registra verdadero o falso según corresponda.
44. Una anotación de POI que indica la presencia de una señal de tráfico en una ubicación cercana, donde se registra verdadero o falso según corresponda.
45. Una anotación de POI que indica la presencia de Turn_loop (Un circuito de giro es un segmento de carretera a menudo circular al final de un camino que se utiliza para dar la vuelta) en una ubicación cercana, donde se registra verdadero o falso según corresponda.
46. Muestra el período del día (es decir, día o noche) basado en el amanecer / atardecer.
47. Muestra el período del día (es decir, día o noche) basado en el crepúsculo civil.
48. Muestra el período del día (es decir, día o noche) según el crepúsculo náutico.
49. Muestra el período del día (es decir, día o noche) según el crepúsculo astronómico.

3.- Justificación del uso de datos.

Usamos esta base de datos ya que, creemos que es una base de datos completa, la cual contiene diversas variables que pueden resultar útiles para resolver diferentes problemáticas

que son muy comunes en la actualidad; Además que contiene un registro de tiempo muy completo, por lo que será más sencillo tener resultados favorecedores según la problemática que se plantee.

4.- Planteamiento del problema.

Como gobierno de los Estados Unidos, encontramos la problemática de que los accidentes automovilísticos a veces llegan a causar muchos estragos no solo en los involucrados si no también en la vialidad, ya que dependiendo de la magnitud de este puede llegar a cerrarse la vialidad por horas en lo que llegan policías y paramédicos al lugar, nosotros necesitamos encontrar una solución para poder atender más rápido al llamado de estos accidentes, así se podrían salvar más vidas y la vialidad no se vería muy afectada, ya que en caso que el choque impida el paso de automóviles, los policías podrían actuar más rápido y abrir de nuevo la vialidad, esto traería un gran beneficio al país y sobre todo a los ciudadanos.

5.- Objetivo final

El objetivo principal de este trabajo es realizar un modelo de predicción, necesitamos poder predecir cuándo sucederá un accidente y en donde, para poder ubicar a más policías dentro de estas zonas, tomaremos en cuenta diversos factores que son relevantes a la hora de que sucede un choque como lo pueden ser: ciudad, estado, temperatura, humedad, condición del clima (nublado, soleado, lluvioso, etc.) y periodo del día (día o noche). Estos factores serán analizados en conjunto para predecir una probabilidad de accidente, de esta forma se podría monitorear todos los estados de EU las 24 horas, y dependiendo de los datos que nos vaya arrojando el modelo se podría enviar más patrullas a los lugares con mayor probabilidad de que suceda un choque, así podría atenderse más rápido un accidente automovilístico y se podrían salvar más vidas.

Como objetivo secundario queremos analizar en donde se producen más accidentes automovilísticos y cuál es su severidad (el impacto que tienen estos choques en el tráfico), al contar con coordenadas exactas de longitud y latitud podemos realizar una segmentación y ver donde es más frecuente que ocurran los choques y saber cuánto impactan estos choques en el tráfico, se toma a medida de 1 – 4 donde 1 es un ligero retraso en el tráfico y 4 un gran retraso en el tráfico (en cuestión de horas). Esto nos podría ayudar a los cálculos del modelo predictivo, pero además podría usarse para otros ámbitos como en el área de los seguros para ver en qué zonas suceden con mayor frecuencia accidentes y con esto realizar mejores cálculos para las primas de los seguros de autos.

6.- Planeación de herramienta a utilizar

Se utilizará principalmente predicción, ya que tenemos un registro de datos históricos, los cuales nos servirán para predecir eventos futuros, por ejemplo: en este caso podemos predecir el lugar de futuros accidentes, el grado del accidente, etcétera.

Nuestra herramienta secundaria sería clasificación, ya que tenemos un registro de diferentes eventos con características diferentes, por lo que podemos utilizar esta herramienta para clasificar cada accidente según su conjunto de características, lo que nos facilitará el uso de los datos.

Y por último, utilizaremos visualización, ya que será útil ver nuestros datos de una manera gráfica, y de igual manera que la herramienta anterior nos facilitará el manejo de nuestros datos.