

**PRIMERA ENTREGA:**

**POR:**

Alejandro Álvarez López

**MATERIA:**

Introducción a la inteligencia artificial

**PROFESOR:**

Raul Ramos Pollan



**UNIVERSIDAD  
DE ANTIOQUIA**

1 8 0 3

**UNIVERSIDAD DE ANTIOQUIA**

**FACULTAD DE INGENIERÍA**

**MEDELLÍN 2022**

## 1. Planteamiento del problema

Cada día, el número de vehículos que son vendidos a lo largo del mundo va en aumento considerablemente, dicho esto, se sabe que la congestión vial es una problemática que se presenta diariamente en la mayoría de grandes ciudades del mundo, complicando una movilidad sencilla y efectiva, y presentando una variedad de problemas tanto a ciudadanos como al medio ambiente.

Teniendo en cuenta información agregada sobre vehículos detenidos, tiempos de espera en intersecciones, y distancias de frenado se va a predecir la congestión vial en 4 ciudades importantes de Estados Unidos (Atlanta, Boston, Chicago y Filadelfia).

## 2. Dataset

Se usará el dataset de Kaggle de la competición “BigQuery-Geotab Intersection congestion” encontrada en el enlace: <https://www.kaggle.com/competitions/bigquery-geotab-intersection-congestion/overview>. La base de datos a utilizar presenta aproximadamente 80000 observaciones (filas) y un total de 30 columnas o variables a utilizar, algunas de las más importantes son:

- Los tiempos totales que han permanecido los vehículos totales: Los tiempos totales se dividen en 5 columnas (20%, 40%, 50%, 60%, 80%), estos porcentajes representan el tiempo total que ha permanecido dicho porcentaje de autos detenido en la intersección correspondiente.
- Los tiempos que se han tardado los autos en volver a detenerse después de cruzar una intersección. Estos tiempos también se dividen en otras 5 columnas, siguiendo el patrón de los porcentajes mencionados anteriormente
- La distancia de la intersección a la cual se han detenido los autos. Como en los 2 ítems mencionados anteriormente, esta distancia presenta un comportamiento igual, 5 columnas divididas por porcentajes.
- Variable de si es fin de semana o no
- El mes
- La hora
- La ciudad
- Las calles

Adicional, el dataset cuenta con un archivo con menos observaciones llamado “train”, para los datos de entrenamiento, y un archivo llamado “sample\_submission”, que es el archivo que se sobrescribirá con los datos obtenidos por las predicciones.

### 3. Métricas

Como métrica de Machine learning se usará el RMSE (Error cuadrático medio), el cual es definido por la competencia. El RMSE es básicamente la desviación estándar de los residuales del modelo.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

Como métrica de negocio se podría usar la disminución del estrés de los pare y siga tanto de viajeros como de agentes de tránsito y el gobierno.

### 4. Desempeño

El objetivo de este proyecto será el identificar la congestión vehicular causada por las intersecciones (semáforos, señales de PARE, etc) en algunas ciudades de Estados Unidos, para de esta forma, entender mejor dicho fenómeno, sus causas y tratar de encontrar algunas posibles soluciones, para mejorar, de esta forma, la movilidad en distintas zonas.