

## Propuesta de Proyecto Aprendizaje Automático

Tec. Ciencia de datos e Inteligencia Artificial.  
Aprendizaje Automático  
2 año, 1 cuatrimestre  
Nombre: Albornoz Elias



## Análisis de delitos en la ciudad de Buenos Aires

### Contexto y relevancia del problema:

En los últimos años la Ciudad de Buenos Aires implementó políticas que incluyen la publicación periódica de datos detallados sobre delitos registrados en sus localidades. Estos datos están disponibles en el Portal de Datos Abiertos ([data.buenosaires.gob.ar](http://data.buenosaires.gob.ar)), contiene información sobre tipo de delito, barrio, franja, uso arma, entre otras variables contextuales, separados en datasets por años.

El dataset elegido "2022 Delitos (XLSX) ", representa una oportunidad para aplicar técnicas de Aprendizaje Automático a una problemática de alto interés social como entender los factores asociados a la gravedad de los delitos. En particular, identificar si un delito involucra el uso de un arma permite distinguir entre incidentes de mayor y menor riesgo, lo que puede ser útil para el análisis de tendencias en la seguridad urbana.

### Objetivo General - Pregunta interesante:

Se plantea un problema de clasificación supervisada, con el objetivo de predecir una variable categórica binaria respondiendo a la pregunta: **¿Se utilizó un arma durante la comisión del delito?**

Esta variable está disponible directamente en el dataset con el nombre **uso\_arma (SI/NO)**.

### Objetivos Específicos:

- 1- Explorar y comprender la estructura del dataset "2022 Delitos (XLSX) "
- 2- Realizar el preprocesamiento necesario: limpieza de datos, codificación de variables categóricas, normalización, creación de nuevas variables en caso de ser necesario.
- 3- Evaluar las relaciones entre las variables predictoras y la variable objetivo, con el objetivo de encontrar el algoritmo más adecuado.
- 4- Implementar y comparar al menos un modelo de clasificación, justificando su elección en función de las propiedades observadas en los datos.

## Posibles modelos y por qué usarlos:

**Regresión Logística:** Si hay tendencias lineales, muestra cuanto influye cada variable con coeficientes, es ideal para clasificaciones binarias.

**Para K-NN y Árbol de decisión:** Los delitos con arma se agrupan en zonas/horas específicas?

**K-NN:** útil si la relación entre variables es no lineal o si los patrones dependen, por ejemplo: ciertos barrios + ciertos horarios, lo cual se clasificaría por proximidad.

**Árbol de Decisión:** permite modelar interacciones complejas entre las variables mediante reglas lógicas, no requiere escalado de variables.