

## **Entrega 1: Descripción y Formulación del Objetivo**

**Proyecto:** Clasificación de Riesgo Ambiental y Territorial en Barrios Populares del Gran Buenos Aires mediante Aprendizaje Automático

**Autora:** Miriam Velazque

### **Contexto del Problema**

En el Gran Buenos Aires, miles de familias habitan en barrios populares expuestos a múltiples factores de riesgo ambiental y territorial. Estos incluyen la cercanía a cuerpos y cursos de agua, la presencia de microbasurales, y la vulnerabilidad frente a inundaciones. La falta de infraestructura y planificación urbana agrava estas condiciones, afectando la salud, seguridad y calidad de vida de las comunidades.

Este proyecto se enmarca en el análisis de datos relevados por *DU-Techo* y *RENABAP*, con el objetivo de generar conocimiento útil para la toma de decisiones públicas, la planificación territorial y el diseño de políticas sociales. La combinación de capas geográficas y variables categóricas permite abordar el problema desde una perspectiva técnica y social.

### **Objetivos del Proyecto**

**Objetivo General:** Desarrollar un modelo de Aprendizaje Automático supervisado que clasifique zonas de barrios populares del Gran Buenos Aires según su nivel de riesgo ambiental (bajo, medio, alto), utilizando datos geospaciales y categóricos.

#### **Objetivos Específicos:**

- Integrar y procesar datasets geográficos de cuerpos de agua, cursos de agua, microbasurales y zonas inundables.
- Construir una variable objetivo que represente el nivel de riesgo ambiental por zona.
- Aplicar y comparar modelos de clasificación supervisada (Árbol de Decisión, KNN, Regresión Logística).
- Evaluar el desempeño de los modelos mediante métricas como accuracy, F1-score y matriz de confusión.
- Visualizar los resultados mediante mapas de calor y gráficos interpretativos para facilitar la comunicación de hallazgos.

### **Tipo de Problema**

Este proyecto aborda un problema de **clasificación supervisada**, ya que el objetivo es asignar una etiqueta de riesgo (bajo, medio, alto) a cada zona analizada, en función de sus características ambientales y territoriales.

### **Dataset y Variables**

El conjunto de datos está compuesto por más de **5.000 instancias** distribuidas en distintas zonas del conurbano bonaerense. Los archivos provienen de fuentes públicas y contienen variables espaciales y categóricas:

- **Cuerpos de Agua** (MULTIPOLYGON): tipo (Laguna, Humedal, etc.), ubicación geográfica.
- **Cursos de Agua** (MULTILINESTRING): trazado hidrográfico, localidad.
- **Microbasurales** (MULTIPOINT): ubicación puntual, provincia, localidad.
- **Zonas Inundables** (MULTIPOLYGON): frecuencia de inundación, extensión, severidad.

Las variables clave incluyen:

- Coordenadas geográficas (EPSG:4326)
- Tipo de riesgo ambiental
- Frecuencia de ocurrencia
- Ubicación administrativa (provincia, departamento, localidad)

Se analizó el archivo `du_techo_zonas_inundables.csv`, y tenemos dos variables claves que pueden funcionar como variable objetivo:

#### ◆ **se\_inunda\_**

- Valores como:
  - SÍ, TODO EL BARRIO
  - SÓLO EN UN SECTOR
  - NO SE INUNDA
- Esta variable indica el **grado de afectación por inundaciones**.

#### ◆ **con\_que\_fr**

- Valores como:
  - CADA VEZ QUE LLUEVE FUERTE
  - OCASIONALMENTE
  - NO SE INUNDA
- Esta variable representa la **frecuencia de inundación**.

### **Modelos Candidatos**

Se consideran los siguientes algoritmos de clasificación supervisada:

- **Árbol de Decisión**: para interpretar qué variables influyen más en el riesgo.
- **K-Nearest Neighbors (KNN)**: para clasificar zonas según similitud con otras.
- **Regresión Logística**: útil si se plantea una clasificación binaria (riesgoso/no riesgoso).
- **Clustering (K-Means)**: como técnica exploratoria para agrupar zonas por perfil de riesgo (opcional).

Todos los modelos serán implementados en **Python** utilizando la librería **scikit-learn**, y organizados en un repositorio Git bajo la estructura **Cookiecutter Data Science**, según las pautas del parcial.

### **Relevancia del Proyecto**

Este trabajo combina análisis técnico con una narrativa social potente. Permite visibilizar condiciones de vulnerabilidad en territorios marginados, y ofrece herramientas para mejorar la planificación urbana y la equidad territorial. Además, representa una oportunidad para aplicar Aprendizaje Automático en un contexto real, con impacto social directo.