

# Ejercicio de Inteligencia Artificial Aplicada a Analítica de Datos

---

## Título del ejercicio:

Clasificación de Declarantes del Impuesto de Renta utilizando Técnicas de Machine Learning

## Contexto del ejercicio

La Dirección de Impuestos y Aduanas Nacionales (DIAN), en conjunto con el DANE, ha puesto a disposición pública bases de datos de carácter tributario y socioeconómico. Uno de los conjuntos de datos disponibles en el portal <https://microdatos.dane.gov.co> el censo de edificaciones, que incluye variables de las construcciones en Colombia. Existe un dataset por cada trimestre.

A partir de este conjunto de datos, se propone crear un modelo que estime el valor por metro cuadrado de una construcción, con base en las variables disponibles más el trimestre en que se haga la estimación.

## Objetivos del ejercicio

1. Cargar y explorar el dataset de la ECV (último disponible en el portal de microdatos).  
<https://microdatos.dane.gov.co/index.php/catalog/852/data-dictionary>
2. Seleccionar variables relevantes para el modelo de clasificación (features).
3. Aplicar técnicas de preprocesamiento:
  - - Imputación y manejo de valores nulos
  - - Codificación de variables categóricas
  - - Estandarización y normalización de variables numéricas
4. Construir modelos de regresión utilizando al menos 4 técnicas diferentes:
5. Utilizar cross-validation para asegurar generalización del modelo.
6. Evaluar y comparar los modelos utilizando las métricas:
  - - Accuracy
  - - Precision
  - - Recall

- - F1-score
- - Curvas ROC y AUC

7. Visualizar los resultados mediante gráficos comparativos de métricas y curvas ROC.

### **Entrega esperada**

- Un notebook en Jupyter con todas las etapas del flujo de trabajo.
- Comparación tabular de las métricas por modelo.
- Curvas ROC de todos los modelos superpuestas.
- Una conclusión escrita de mínimo 500 palabras describiendo el proceso.
- Una conclusión escrita de mínimo 300 palabras sobre qué modelo ofrece mejor rendimiento y por qué, teniendo en cuenta la interpretabilidad, complejidad y precisión.