A green and black logo

AI-generated content may be incorrect.A logo of a university

AI-generated content may be incorrect.Universidad Tecnológica de Panamá

A logo of a university

AI-generated content may be incorrect.

Maestría en Analítica de Datos

Materia: Modelos Predictivos

Profesor: Juan Marcos Castillo, PhD

Estudiante: Irma Martínez

Título: Propuesta de Proyecto - Predicción de Nivel de Ingreso

Fecha: 09/07/2025

# Introducción

Este trabajo tiene como objetivo aplicar las técnicas de modelado predictivo a la clasificación de personas que ganan más de $50,000 anuales a partir de datos de carácter socioeconómico. La base que se empleará para darle soporte a esta problemática es el Adult Income Dataset, ubicada en la biblioteca UCI (UC Irvine Machine Learning Repository)

# 2. Justificación

El estudio de los ingresos personales es una actividad clave en diversas disciplinas como la economía, la sociología y la planificación del gobierno. Determinar cuáles son los factores que inciden en el nivel de ingresos puede contribuir a la estructuración e implementación de mejores políticas sociales, detectar desigualdades sociales, y aportar al diseño de procesos de reclutamiento y evaluación de empresas.

# 3. Antecedentes

El Adult Income Dataset ha sido utilizado en múltiples estudios como referencia para comparar algoritmos de clasificación. Estudios anteriores han aplicado modelos como regresión logística, árboles de decisión, Naives Bayes, redes neuronales y random forest para predecir la variable objetivo: ingreso mayor a 50 mil dólares anuales o menor igual que esta cantidad.

# 4. Definición del problema

Problema: Tomando en cuenta ciertas variables, ¿se podría predecir si una persona gana más de $50,000 anuales en función de sus características demográficas y laborales?

Variable de salida: ingreso (binaria: >50K o <=50K)

Variables predictoras:

- edad  
- educacion  
- ocupación  
- estado civil  
- raza  
- genéro  
- horas por semana trabajadas  
- País de nacimiento

# 5. Avance del análisis predictivo

Hasta la fecha, se han realizado las siguientes tareas:  
- Carga y exploración preliminar del archivo determinando las dimensiones del dataset, variables disponibles y examinando una muestra de los datos que contiene este.   
- Limpieza básica: eliminación de caracteres especiales, valores faltantes y agregar encabezados y categorizar los rangos de edad y de horas trabajadas a la semana, además de eliminar registros con valores faltantes y eliminar caracteres de más dentro del campo ingreso, al igual que la transformación de ingreso a valor binario.  
- Elaboración de gráficas de visualización.

# 6. Gráficas de visualización

A continuación, se presentan algunas visualizaciones clave del análisis exploratorio:  
a) Distribución por ingresos  
b) Edad vs Ingresos  
c) Educación más frecuente por ingreso  
d) Horas de trabajo semanales según ingreso

A screenshot of a computer

AI-generated content may be incorrect.