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

Propuesta de Investigación

**Título del Proyecto:**

Predicción de nivel de ingreso basado en características sociodemográficas usando machine learning.

**Descripción de la Base de Datos:**

La base de datos utilizada en este proyecto proviene del UCI Machine Learning Repository y se titula "Adult Income Dataset". Contiene información de 48,842 personas adultas de Estados Unidos, con 15 atributos que describen variables sociodemográficas como edad, educación, ocupación, horas trabajadas por semana, y país de origen. La variable objetivo es el nivel de ingreso, clasificado en dos categorías: menor igual que 50 mil dólares, mayor de 50 mil dólares anuales.  
  
Los atributos incluyen tanto variables numéricas como categóricas, lo cual permite aplicar diferentes técnicas de análisis estadístico y predictivo.

**Motivo de la Selección de los Datos:**

Esta base fue seleccionada por su amplio uso en tareas de clasificación y su relevancia práctica en áreas como recursos humanos, políticas públicas y análisis de desigualdad. Además, su estructura con atributos tanto numéricos como categóricos la hace ideal para poner en práctica diversas técnicas de minería de datos y algoritmos de predicción estudiados en el curso.

**Introducción del Contexto:**

En un contexto global donde las brechas económicas persisten, comprender qué factores influyen en los ingresos de una persona puede tener implicaciones valiosas en políticas públicas, planes de inclusión laboral y análisis económico. A través de este estudio, se busca identificar patrones en las características sociodemográficas que permitan predecir si una persona tiene una probabilidad alta o baja de generar ingresos superiores a los $50,000 anuales.  
  
Esta investigación también persigue un objetivo académico: fortalecer el dominio de herramientas de análisis predictivo mediante la aplicación práctica de algoritmos como Random Forest, Árboles de Decisión y Regresión Logística en Weka.

# Línea de Tiempo del Proyecto:

A screenshot of a graph

AI-generated content may be incorrect.