REPORTE DE AVANCE DEL PROYECTO

Carrera: Maestría en Analítica de Datos – UTP

Profesor: Juan Marcos Castillo, PhD

Estudiante: Irma Martínez

Cédula: 8-886-1536

Fecha: 07/10/2025

# 1. 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.

# 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 concepción 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, SVM, redes neuronales y random forest para predecir la variable objetivo: income (>50K o <=50K).

# 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: income (binaria: >50K o <=50K)

Variables predictoras:

- age  
- education  
- occupation  
- marital-status  
- race  
- sex  
- hours-per-week  
- native-country

# 5. Avance del análisis predictivo

Hasta la fecha, se han realizado las siguientes tareas:  
- Carga y exploración preliminar del archivo  
- Limpieza básica: eliminación de caracteres especiales, valores faltantes y agregar encabezados y categorizar los rangos de edad y de horas trabajadas a la semana.  
- 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.