

# Taller 3 - Making Money with Machine Learning

Harold Stiven Acuña

José David Cuervo  
José David Dávila  
César Augusto Alfaro

17 de abril de 2025

## Resumen

Texto del abstract

**Palabras clave:** precio de propiedades, aprendizaje automático

**Clasificación JEL:** J31, C53, J16

*Repositorio GitHub:* [https://github.com/alfarocesar/BDML\\_Making\\_Money\\_Equipo8](https://github.com/alfarocesar/BDML_Making_Money_Equipo8)

# 1. Introducción

Incluir la introducción

## 2. Datos

### 2.1. Adecuación de los datos

### 2.2. Construcción de la muestra

#### 2.2.1. Análisis y limpieza de las bases de datos

#### 2.2.2. Transformación de variables de personas

#### 2.2.3. Agregación de datos a nivel de hogar

### 2.3. Limpieza de datos y tratamiento de valores faltantes

### 2.4. Análisis descriptivo

#### 2.4.1. Distribución de variables numéricas

#### 2.4.2. Análisis de outliers

#### 2.4.3. Relación con la variable objetivo

### 2.5. Selección final de variables

## 3. Modelos y Resultados

### 3.1. Modelo de Selección y Entrenamiento

#### 3.1.1. Metodología

#### 3.1.2. Variables utilizadas

### 3.2. Resultados

### 3.3. Matrices de Confusión

## 4. Conclusión

## Anexos

## Referencias

- Anthropic (2023). Conversaciones con claude ai assistant. Utilizado para estructura de documentos LaTeX, y mejoramiento del código en R.
- Banco Mundial (2024). Trayectorias: Prosperidad y reducción de la pobreza en el territorio colombiano. Technical report, Banco Mundial, Washington, D. C., Estados Unidos de América. Publicado el 3 de diciembre de 2024.
- Castillo-Álvarez, G. and Sarmiento-Barbieri, I. (2024a). Data pre-processing: Visualizing and handling missing values (part 1). Notebook proporcionado como material del curso.
- Castillo-Álvarez, G. and Sarmiento-Barbieri, I. (2024b). Data pre-processing: Visualizing and handling missing values (part 2). Notebook proporcionado como material del curso.
- Huang, S., Shao, Y., Hu, L., Bi, J., and Ma, Z. (2025). Unveiling the spatio-temporal patterns and mechanisms of poverty incidence in southwest china through multi-sourced data and interpretable machine learning. *Applied Geography*, 176:103541.
- Huffman, C. and Nájera, H. (2024). What we talk about when we talk about measurement in poverty research. *Problemas del desarrollo*, 55(217):89–102.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R*. Springer, New York, NY, 2 edition.
- Karmaker, S. C., Rjbongshi, A., Pal, B., Sen, K. K., and Chapman, A. J. (2025). Machine learning-based prediction of energy poverty in bangladesh: Unveiling key socioeconomic drivers for targeted policy actions. *Socio-Economic Planning Sciences*, 99:102213.
- OpenAI (2023). Conversaciones con chatgpt. Utilizado para estructura de documentos LaTeX, y mejoramiento del código en R.
- Sarmiento-Barbieri, I. (2024). Limpieza de datos con tidyverse. Notebook proporcionado como material del curso.
- Sarmiento-Barbieri, I. and Castillo-Álvarez, G. (2024a). Caret para clasificación. Material del curso BDML, Universidad de los Andes. Notebook proporcionado como material del curso.
- Sarmiento-Barbieri, I. and Castillo-Álvarez, G. (2024b). Classification - cuaderno de clase. Notebook proporcionado como material del curso.
- Sarmiento-Barbieri, I. and Castillo-Álvarez, G. (2024c). Uniendo bases y calculando pobreza. Material del curso BDML, Universidad de los Andes. Notebook proporcionado como guía para el taller.
- Tal, E. (2016). How does measuring generate evidence? the problem of observational grounding. *Journal of Physics: Conference Series*, 772(1):012001.