



## Propuesta - Proyecto final

**Maestrando:** Rolando Mora De León

**Cédula:** 9-756-1619

**Tutor:** Juan Marcos Castillo, PhD

**Curso:** Modelos Predictivos

**Título de la investigación:** Predicción del rendimiento académico de estudiantes de Educación Superior mediante Análisis de Datos

### 1. Descripción de la base de datos

La base de datos “Higher Education Students Performance Evaluation” (Evaluación del rendimiento de estudiantes de educación superior) proviene del repositorio UCI Machine Learning. Contiene información recopilada en el año 2019 de 145 estudiantes universitarios de dos facultades (Ingeniería y Ciencias de la Educación). En total incluye 31 variables predictoras sobre cada estudiante más la variable objetivo, sin contar un identificador de estudiante. Estas variables abarcan aspectos personales, familiares y hábitos académicos: por ejemplo edad, sexo, tipo de escuela secundaria de procedencia, si el estudiante tiene beca, situación laboral, actividades extracurriculares, estado civil, nivel educativo de los padres, número de hermanos, ingresos familiares, horas semanales de estudio, frecuencia de lectura, asistencia a seminarios, hábitos de estudio (tomar apuntes, atención en clase, etc.), promedio académico previo (GPA) y esperado, entre otros (todos codificados numéricamente en categorías). La variable objetivo es la calificación final del estudiante en determinada asignatura, registrada en una escala de 8 categorías (desde 0 = “Fail” o reprobado, hasta 7 = “AA”, que equivale a la nota máxima).

### 2. Motivo de la selección de los datos

El rendimiento académico en educación superior es un tema crítico, considerado el principal indicador de éxito o fracaso de un estudiante. Analizar y predecir dicho rendimiento tiene implicaciones prácticas importantes, ya que permitiría identificar estudiantes en riesgo y mejorar los resultados educativos. El conjunto de datos incluye una variedad de factores personales, socioeconómicos y académicos. Esto permite explorar cuáles de ellos son más determinantes en el éxito estudiantil, fomentando el autodescubrimiento de conocimiento al relacionar distintas dimensiones con las calificaciones. Además, los datos provienen de una fuente, están claramente documentados y no presentan valores faltantes, lo que reduce el esfuerzo de limpieza. Al ser un conjunto relativamente pequeño (145 casos), es manejable para efectos de un proyecto académico y computacionalmente eficiente de analizar en herramientas comunes.

El tema de la predicción del desempeño estudiantil resulta de interés personal y profesional, ya que combina el campo educativo con técnicas de análisis de datos. Esto ofrece la oportunidad de aplicar

