





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







métodos de machine learning en un contexto real y entender cómo la analítica de datos puede apoyar la toma de decisiones en educación.

3. Introducción y definición del problema

En el ámbito de la educación superior, mejorar el rendimiento académico de los estudiantes y reducir las tasas de reprobación o deserción son desafíos prioritarios. A pesar de contar con datos variados sobre los estudiantes, las universidades a menudo carecen de herramientas para sintetizar esa información y predecir qué estudiantes podrían tener dificultades al final del semestre. Se busca construir un modelo predictivo que, dados los atributos de un estudiante al inicio o durante el semestre, pueda estimar su categoría de desempeño. El objetivo final de la investigación es doble: por un lado, describir estadísticamente el conjunto de datos para obtener insights, y por otro, desarrollar un modelo predictivo que anticipe el rendimiento.

En síntesis, abordamos la pregunta: ¿Es posible predecir con precisión el rendimiento final de un estudiante universitario utilizando sus características personales, familiares y académicas, y cuáles de estos factores contribuyen más a dicha predicción? Resolver este problema aportaría valor tanto a nivel académico (demostrando la aplicación de técnicas de IA en educación) como práctico (informando políticas de apoyo estudiantil basadas en datos).

Línea de tiempo de la investigación

Actividad														Jı	ılio													
	1 2	2 3	4	5 6	3 7	8 9	9 1	0 11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Selección y análisis de base de datos																												
Formulación de propuesta																												
Entrega de propuesta																												
Pre-procesamiento y limpieza																												
Análisis descriptivo exploratorio																												
Aportes al reporte de avances																												
Selección de variables relevantes																												
Selección y entrenamiento de modelos																												
Aportes al reporte de avances																												
Evaluación de resultados y conclusiones																												
Entrega reporte final																												