Explorando el Impacto de las Estrategias de Enseñanza en el Rendimiento Académico

Estamos investigando cómo las diferentes formas de enseñar de los profesores influyen en las notas de los estudiantes de cuarto grado. No solo nos centramos en las calificaciones, sino que también queremos comprender qué hace que algunas clases sean más efectivas que otras.

Lo que Buscamos Descubrir:

1. Métodos de Enseñanza Efectivos: Nos interesa ver si hay estilos de enseñanza específicos que realmente ayuden a los estudiantes a mejorar sus calificaciones en materias clave.

2. El Rol del Entorno del Estudiante: Queremos explorar cómo factores como el entorno familiar y la educación de los padres pueden influir en el rendimiento académico.

3. Recomendaciones para Mejorar la Enseñanza: Con lo que descubramos, esperamos ofrecer a la escuela sugerencias prácticas para optimizar sus métodos de enseñanza.

4. Desarrollar un Modelo Predictivo: Nuestro objetivo es crear un modelo que pueda predecir cómo diferentes factores afectarán las calificaciones de los estudiantes, lo cual sería muy útil para la planificación y toma de decisiones de la escuela.

Tipo de Aprendizaje Elegido: Aprendizaje Supervisado

¿Por qué?

- Datos con Etiquetas: Cada estudiante tiene una "etiqueta" de rendimiento (aprobado o desaprobado en distintas materias), lo cual es perfecto para el aprendizaje supervisado.

- Relaciones Causa-Efecto: Queremos entender cómo ciertas variables, como las técnicas de enseñanza, afectan los resultados académicos.

- Predicción y Clasificación: Este enfoque nos permite analizar los datos actuales y hacer predicciones para el futuro.

- Modelos Concretos: Con el aprendizaje supervisado, podemos crear modelos precisos y medibles, algo esencial en educación.

Modelo y Algoritmo Elegidos

- Modelo de Clasificación: Elegimos este modelo porque las categorías a predecir son 'aprobado' o 'desaprobado'. Nos parece el más adecuado para categorizar a los estudiantes según su rendimiento académico, basándonos en variables como los métodos de enseñanza y el entorno del estudiante.

- Árboles de Decisión: Optamos por este algoritmo porque es fácil de entender y maneja bien tanto datos numéricos como categóricos. Además, nos ayuda a identificar qué variables son más influyentes en el rendimiento de los estudiantes.

Otros algoritmos que podríamos usar:

Regresión Logística: Adecuada para problemas de clasificación binaria. Es útil para entender la importancia de cada variable independiente.

Máquinas de Soporte Vectorial (SVM): Efectivas en espacios de alta dimensión y cuando la relación entre las clases no es lineal.

Redes Neuronales y Deep Learning: Apropiadas para capturar relaciones complejas en los datos, aunque pueden requerir una mayor cantidad de datos para un entrenamiento efectivo.

Métodos de Ensamble: Técnicas como Random Forest pueden ofrecer un rendimiento superior al combinar múltiples modelos débiles para formar uno más fuerte.

Implementación y Consideraciones Adicionales

- Usaremos scikit-learn en Python para implementar un árbol de decisión. Este framework nos permite dividir los datos en conjuntos de entrenamiento y prueba, entrenar el modelo y evaluar su precisión.

- Consideraremos la validación cruzada para asegurarnos de que nuestro modelo es robusto y ajustaremos los parámetros para evitar el sobreajuste y mejorar la precisión del modelo.

Análisis Exploratorio de Datos (EDA) e Ingeniería de Características

EDA:

- Revisión Inicial: Empezaremos con una revisión general para entender la estructura y distribución de los datos.

- Análisis Estadístico Descriptivo: Calcularemos estadísticas básicas para tener una visión general de las tendencias y variaciones.

- Visualización de Datos: Usaremos gráficos para visualizar distribuciones y posibles correlaciones.

- Valores Atípicos y Variables Faltantes: Identificaremos y trataremos valores atípicos y decidiremos cómo manejar las variables con datos faltantes.

Ingeniería de Características:

- Codificación de Variables Categóricas: Transformaremos variables categóricas en un formato numérico adecuado para el modelado.

- Normalización/Estandarización: Aseguraremos que todas las variables contribuyan equitativamente al modelo.

- Creación de Nuevas Características: Podríamos crear nuevas variables, como la edad del estudiante a partir de su fecha de nacimiento.