PREDICCIÓN DE RESULTADOS EN PRUEBAS SABER PRO UTILIZANDO MACHINE LEARNING

UNIVERSIDAD EAFIT®

Presentación del Equipo

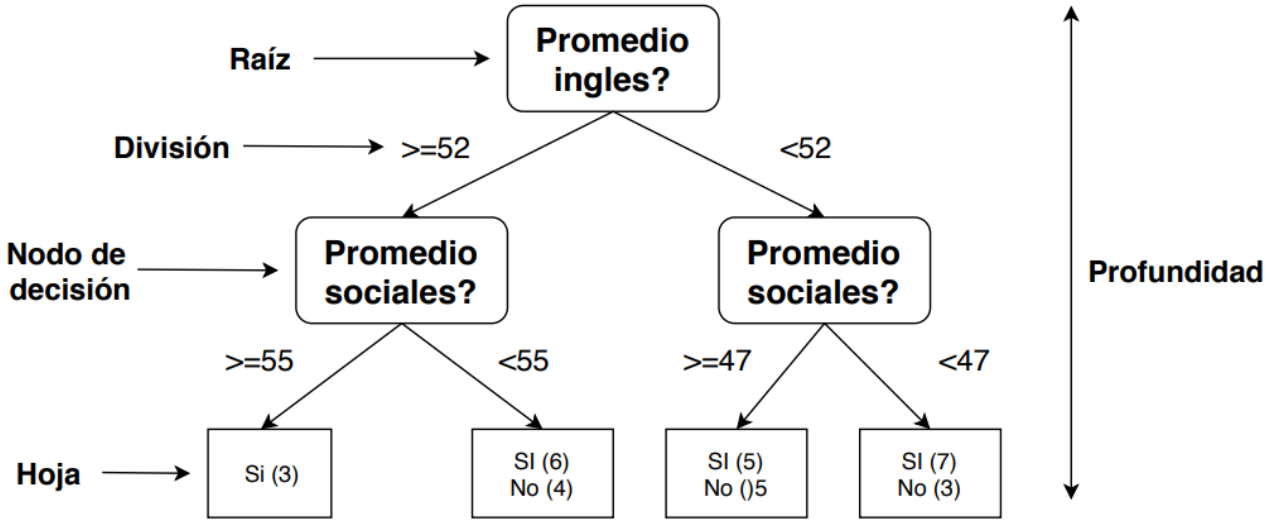
Simón Correa

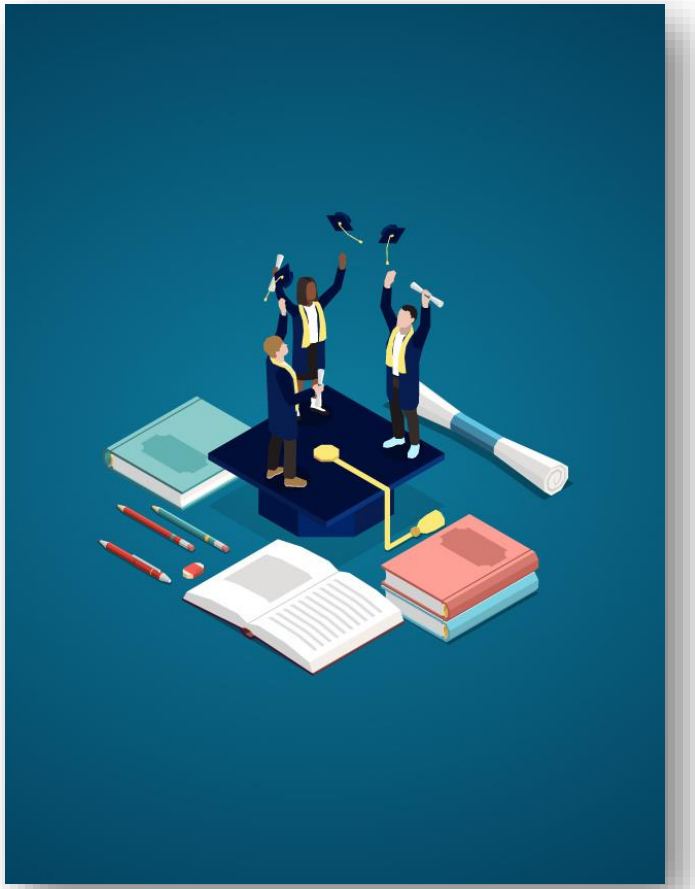
David Gomez

Miguel Correa

Mauricio Toro

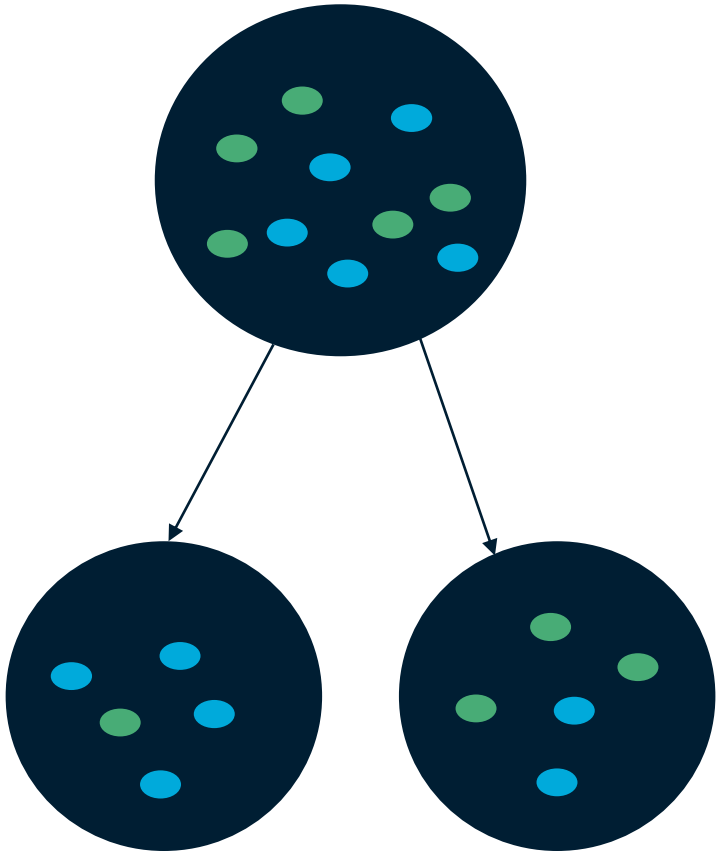
Diseño del Algoritmo



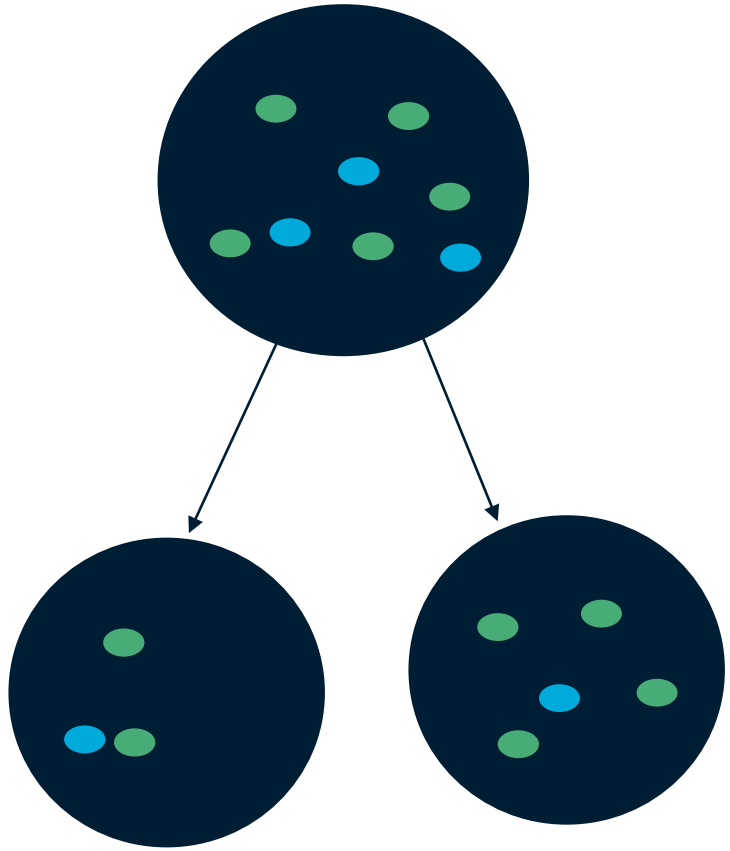


El algoritmo usado para construir un árbol de decisión binario, y para predecir el exito de un individuo fue el CART. En este ejemplo, mostramos un modelo para predecir si un estudiante va a tener éxito en las Pruebas Saber Pro, basándonos en sus resultados de las pruebas Icfes

División de un nodo



Esta división está basada en la condición “Puntaje Ingles >= 52” Para este caso, la impureza Gini de la izquierda es 0.3, la impureza Gini de la derecha es 0.43 y la impureza ponderada es de 0.39.



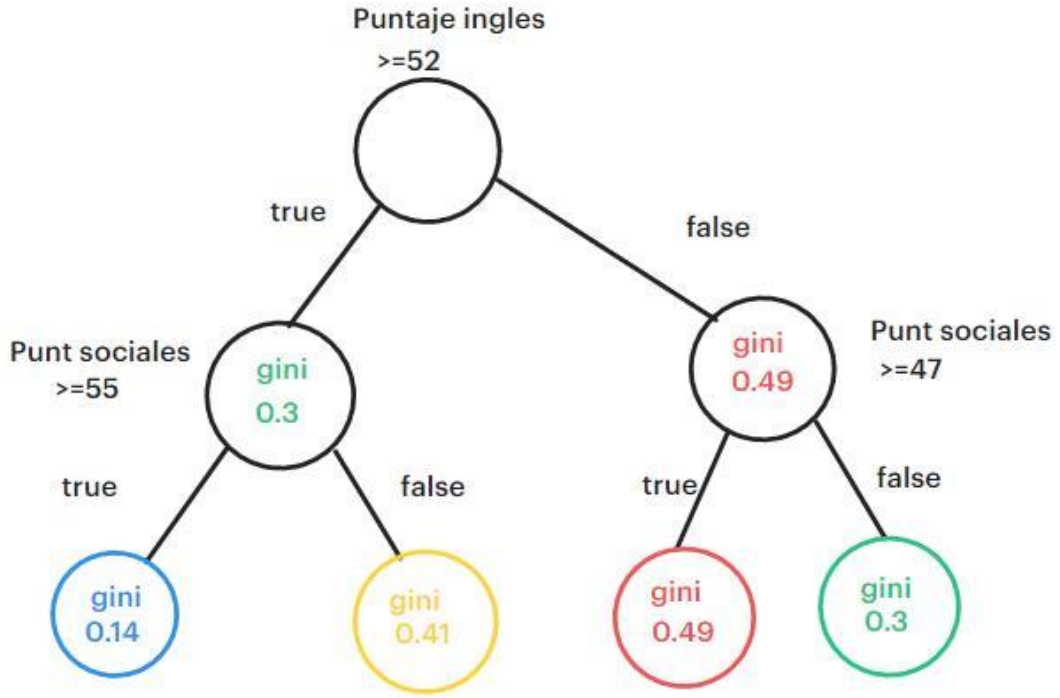
Esta división está basada en la condición “Puntaje Sociales >= 47.” Para este caso, la impureza Gini de la izquierda es 0.49, la impureza Gini de la derecha es 0.29 y la impureza ponderada es 0.39.

Complejidad del Algoritmo

	Complejidad en tiempo	Complejidad en memoria
Entrenamiento del modelo	$O(N^2*M*2^M)$	$O(N*M*2^M)$
Validación del modelo	$O(N*M)$	$O(1)$

Complejidad en tiempo y memoria del algoritmo CART.
Siendo N la cantidad de filas y M la cantidad de columnas.

Modelo de Árbol de Decisión

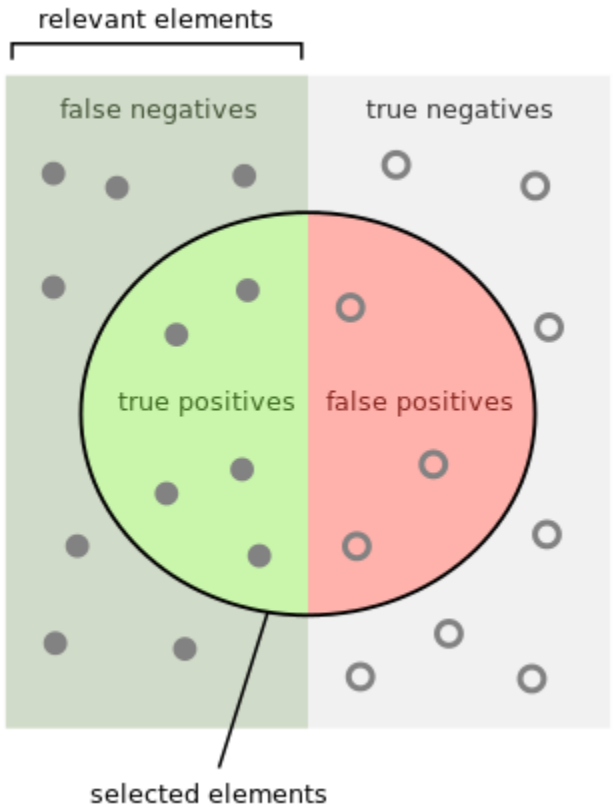


Un árbol de decisión binario para predecir Saber Pro basado en los resultados de Saber 11. Los nodos azules representan a aquellos con una alta probabilidad de éxito, los verdes con una probabilidad media, los amarillos una probabilidad media-baja, y los rojos con una baja probabilidad de éxito.

Características Más Relevantes

- Ciencias Sociales
- Inglés
- Biología

Métricas de Evaluación



Exactitud = $\frac{\text{true positives} + \text{true negatives}}{\text{relevant elements} + \text{true negatives}}$

Precisión = $\frac{\text{true positives}}{\text{true positives} + \text{false positives}}$

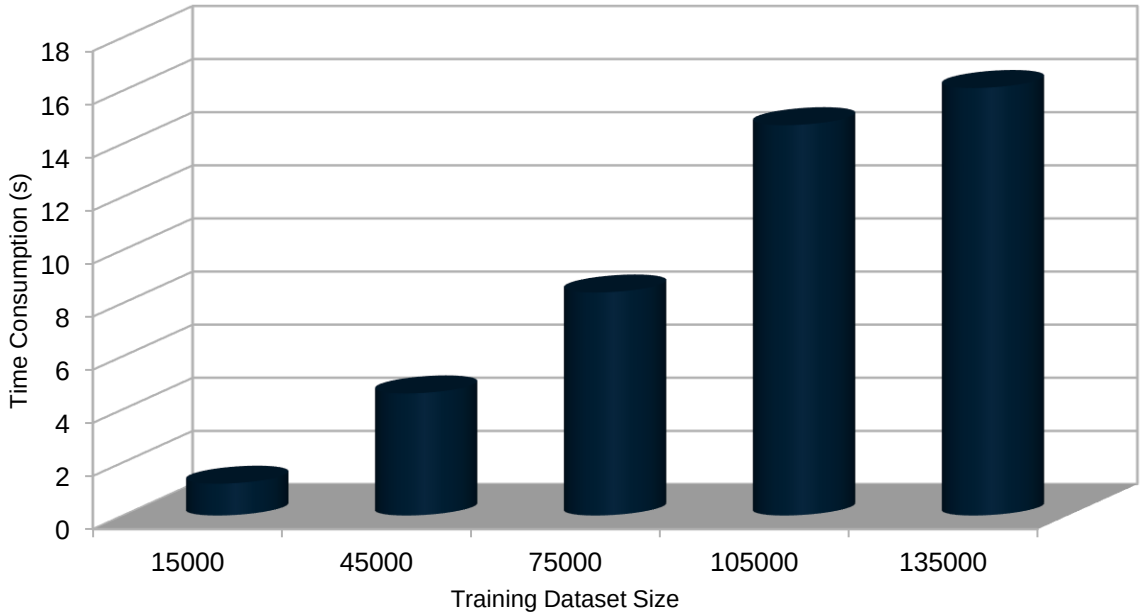
Sensibilidad = $\frac{\text{true positives}}{\text{relevant elements}}$

Métricas de Evaluación

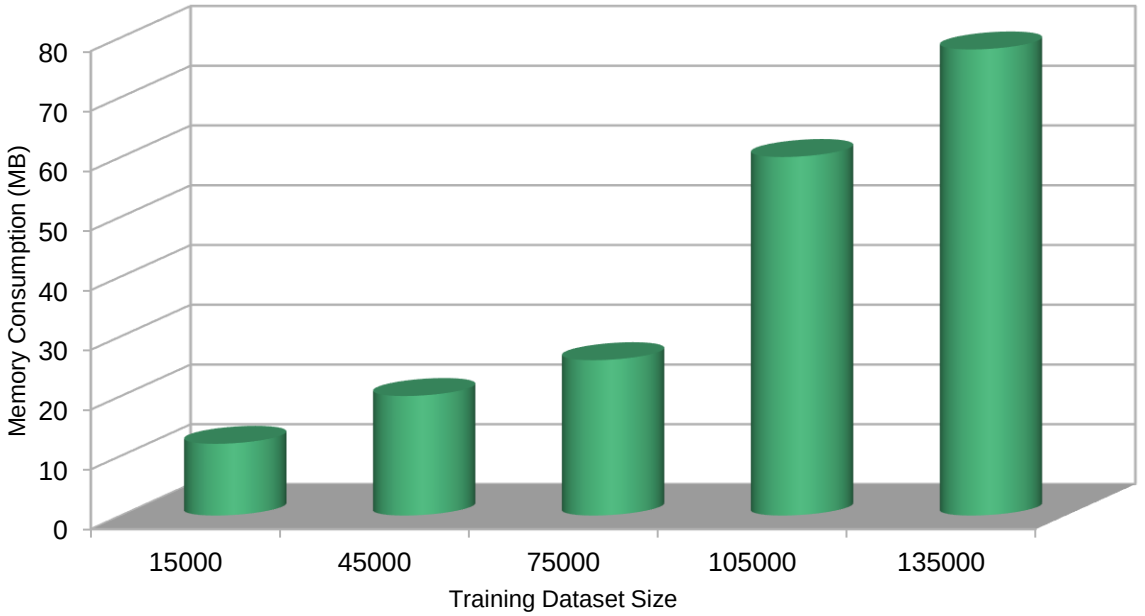
	Conjunto de entrenamiento	Conjunto de validación
Exactitud	0.78	0.78
Precisión	0.72	0.72
Sensibilidad	0.82	0.82

Métricas de evaluación obtenidas con el conjunto de datos de entrenamiento de 135,000 estudiantes y el conjunto de datos de validación de 45,000 estudiantes.

Consumo de tiempo y memoria



Consumo de tiempo



Consumo de memoria


¡GRACIAS!