

Predicción del
éxito de
Saber PRO
utilizando
algoritmo de
árbol de
decisiones



Presentación del Equipo



Simón Álvarez
Ospina



David Madrid
Restrepo



Miguel
Correa



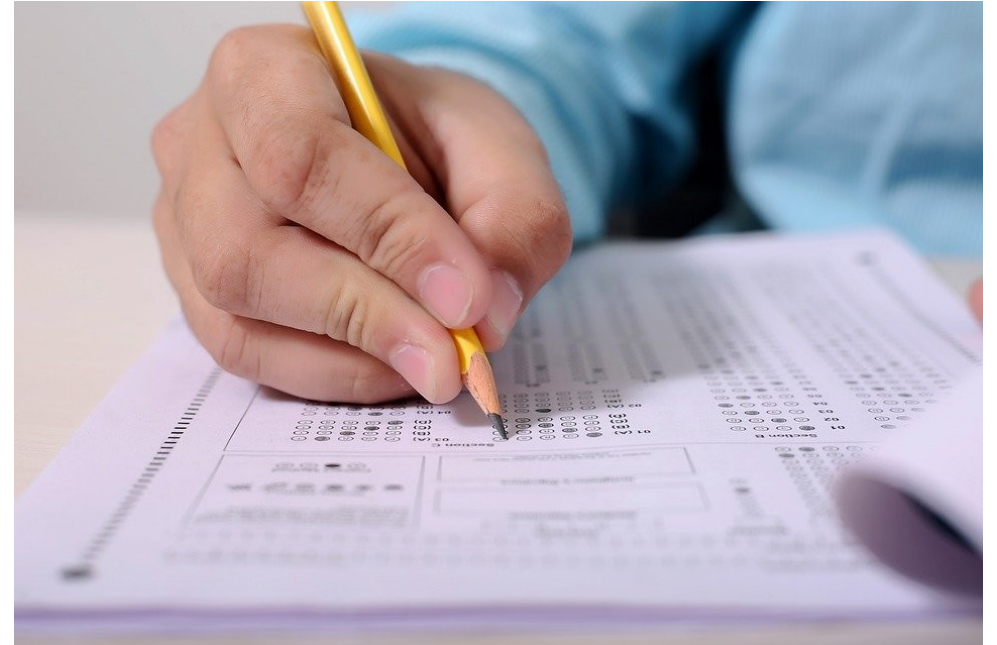
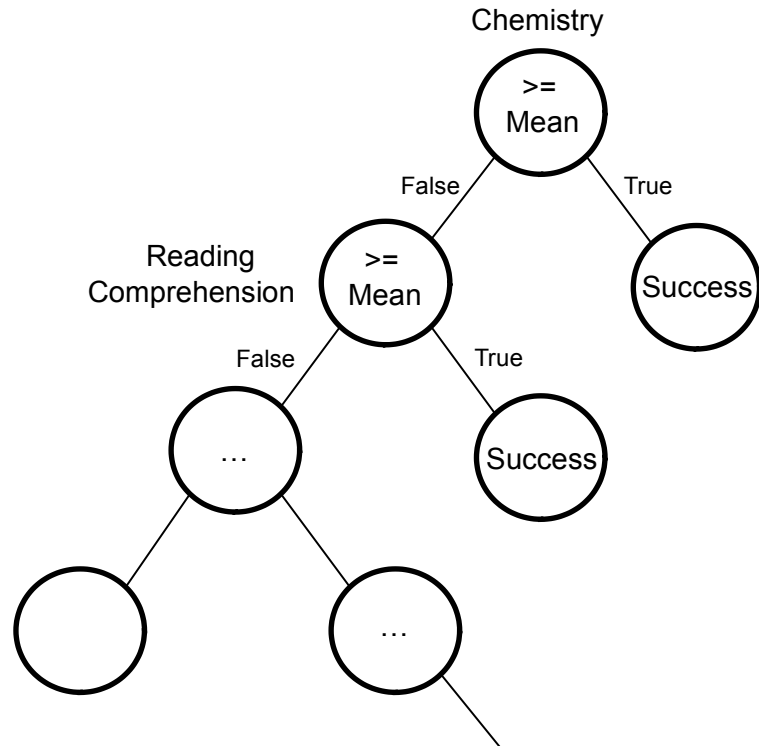
Mauricio
Toro



<https://github.com/dmadridr/ST0245-002/tree/master/proyecto/>

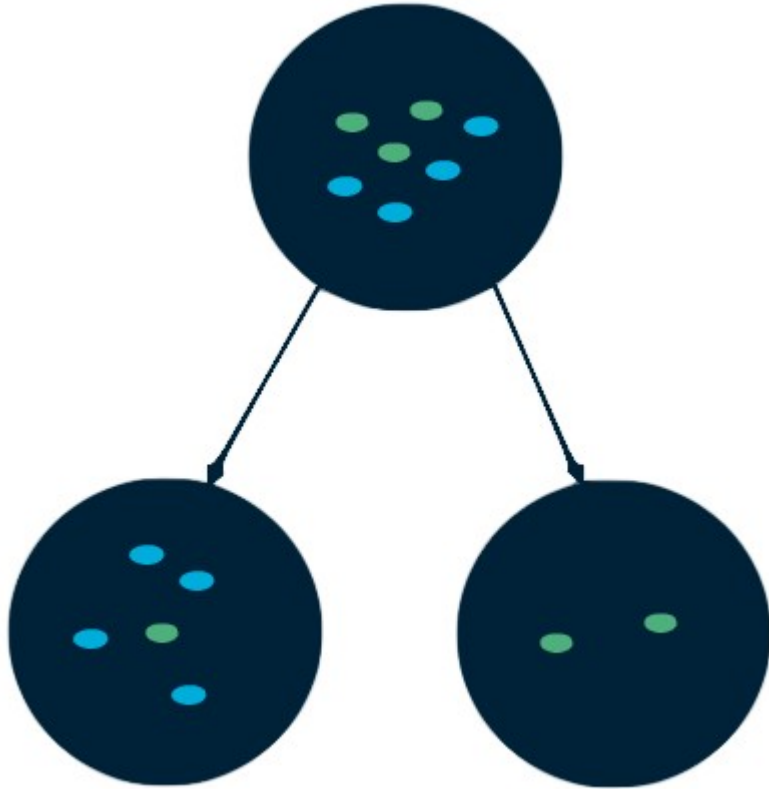


Diseño del Algoritmo

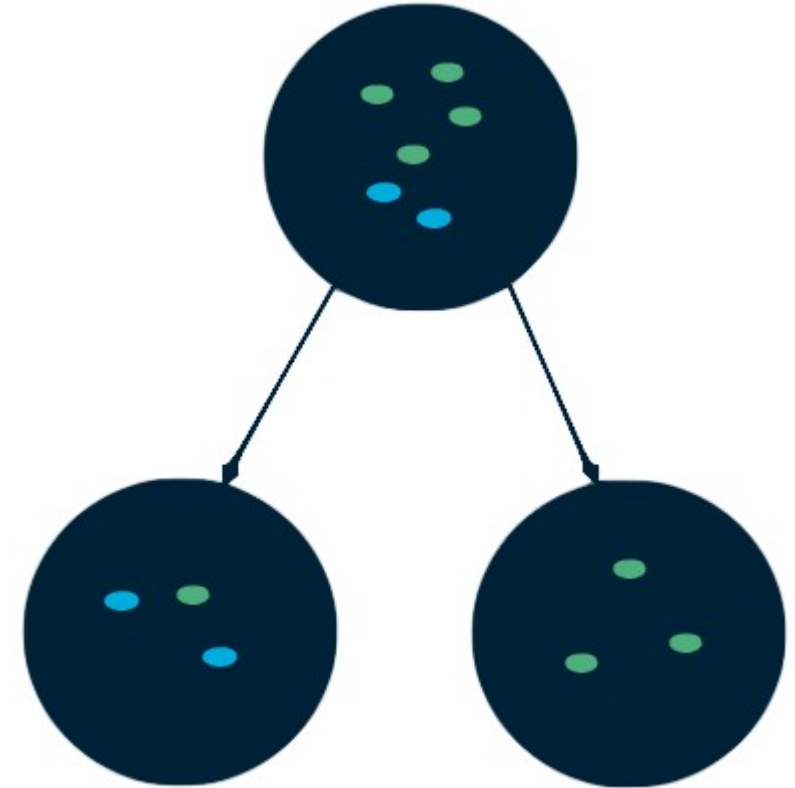


Algoritmo para construir un árbol binario de decisión usando CART. En este modelo se ilustra cómo va a ser el proceso de decisión para predecir si la persona estará o no sobre el promedio.

División de un nodo



Esta división está basada en la condición “desempeño inglés \geq A2.” Para este caso, la impureza Gini de la izquierda es 0.32, la impureza Gini de la derecha es 0 y la ganancia es de 0.26122.



Esta división está basada en la condición “puntaje lenguaje \geq promedio.” Para este caso, la impureza Gini de la izquierda es 0.44, la impureza Gini de la derecha es 0 y la ganancia es 0.166.

Complejidad del Algoritmo

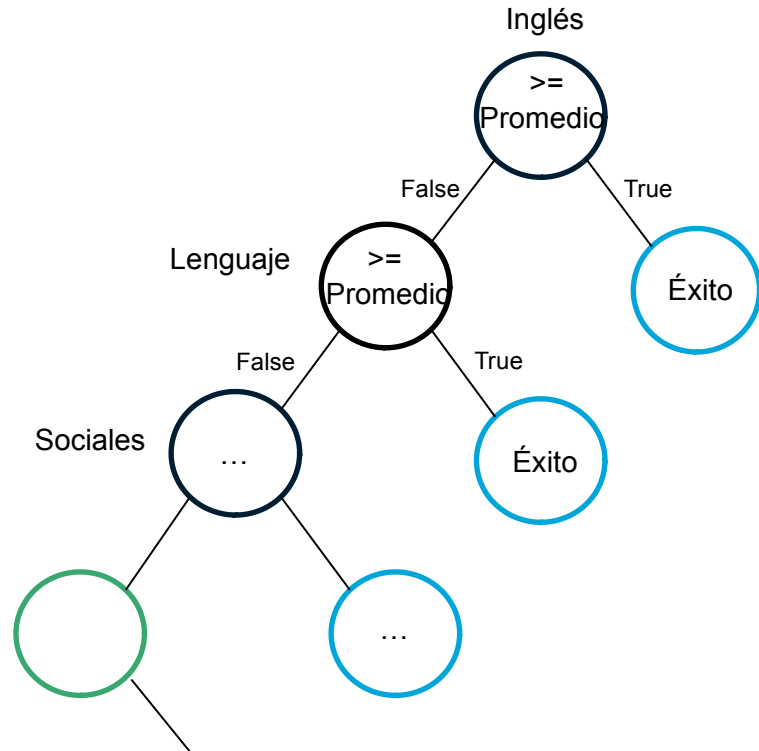


	Complejidad en tiempo	Complejidad en memoria
Entrenamiento del modelo	$O(M \cdot N^2 \cdot 2^N)$	$O(M \cdot 2^N)$
Validación del modelo	$O(N \cdot M)$	$O(M)$



Complejidad en tiempo y memoria del algoritmo. Donde M representa el número de estudiantes (filas) y N representa el número de variables (columnas).

Modelo de Árbol de Decisión



Un árbol de decisión para predecir el resultado del Saber Pro usando los resultados del Saber 11 y algunos aspectos socioeconómicos.

Azul representa nodos con alta probabilidad de éxito y Verde representa nodos con baja probabilidad.

Características Más Relevantes



Ciencias Sociales

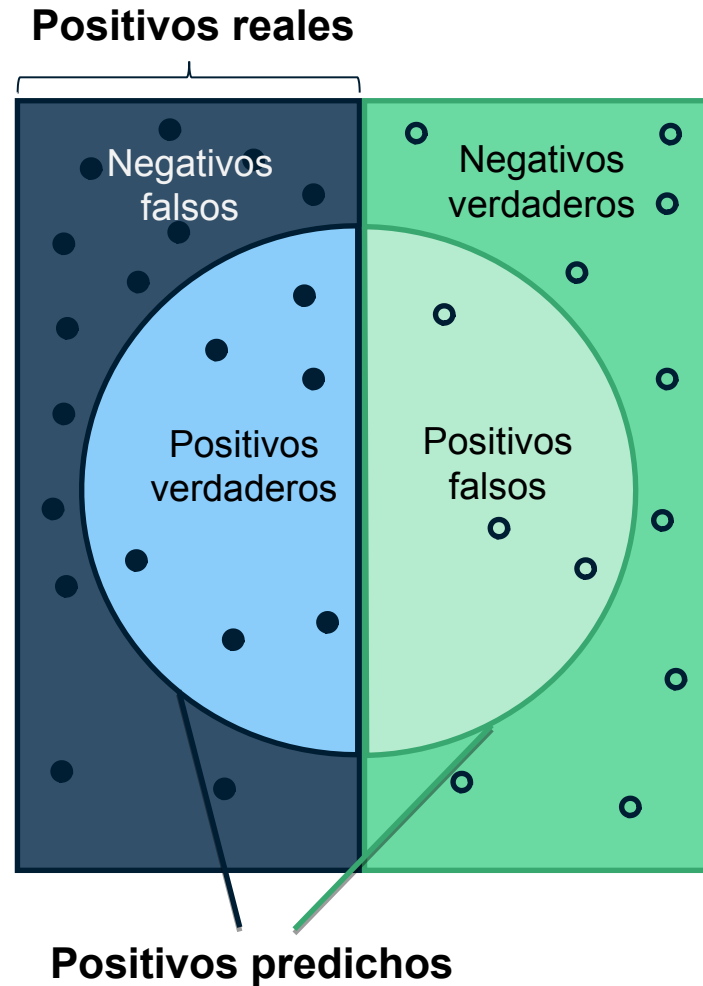


Inglés



Lenguaje

Métricas de Evaluación



$$\text{Precisión} = \frac{\text{Positivo verdadero}}{\text{Positivos predichos}}$$

$$\text{Sensibilidad} = \frac{\text{Positivo verdadero}}{\text{Positivos reales}}$$

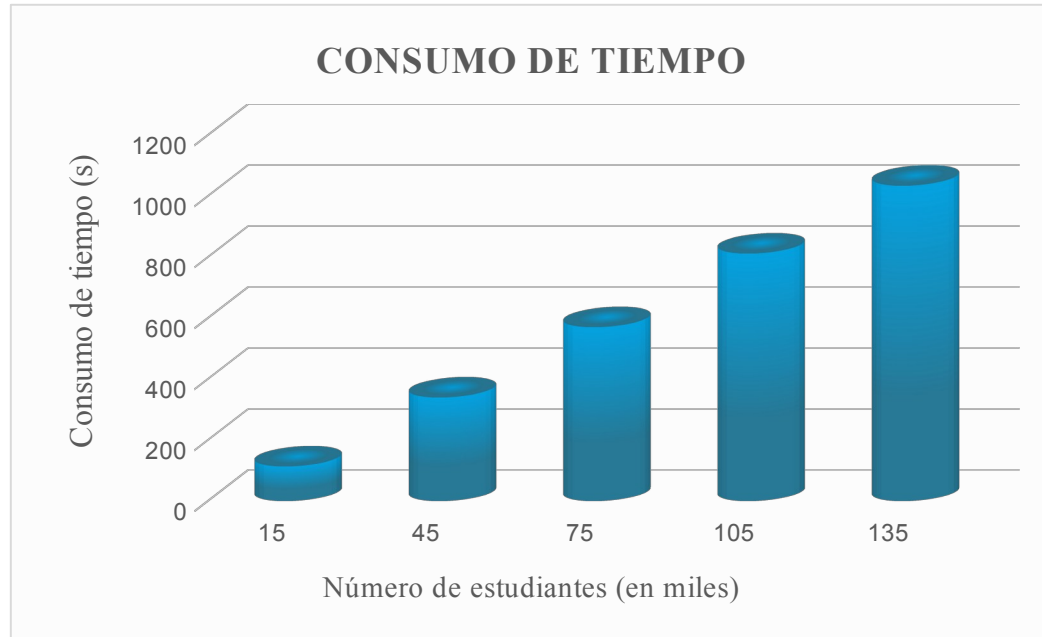
$$\text{Exactitud} = \frac{\text{Positivo verdadero} + \text{Negativo verdadero}}{\text{Total}}$$

	Conjunto de entrenamiento	Conjunto de validación
Exactitud	0.76	0.67
Precisión	0.76	0.67
Sensibilidad	0.77	0.67

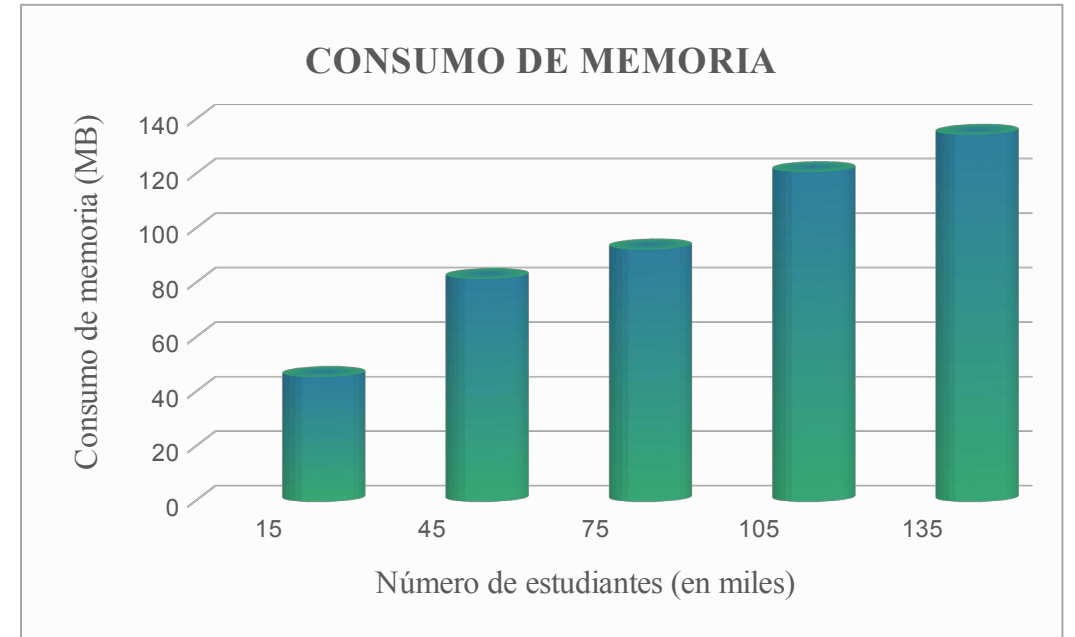
Métricas de evaluación obtenidas con el conjunto de datos de entrenamiento de 135,000 estudiantes y el conjunto de datos de validación de 45,000 estudiantes.



Consumo de tiempo y memoria



Consumo de tiempo



Consumo de memoria

¡GRACIAS!

