

Árboles
de decisión
para
pronosticar
el éxito
en las
pruebas
Saber Pro



Presentación del Equipo



Stiven Ossa
Sanchez



Juan David
Correa Duque



Miguel
Correa



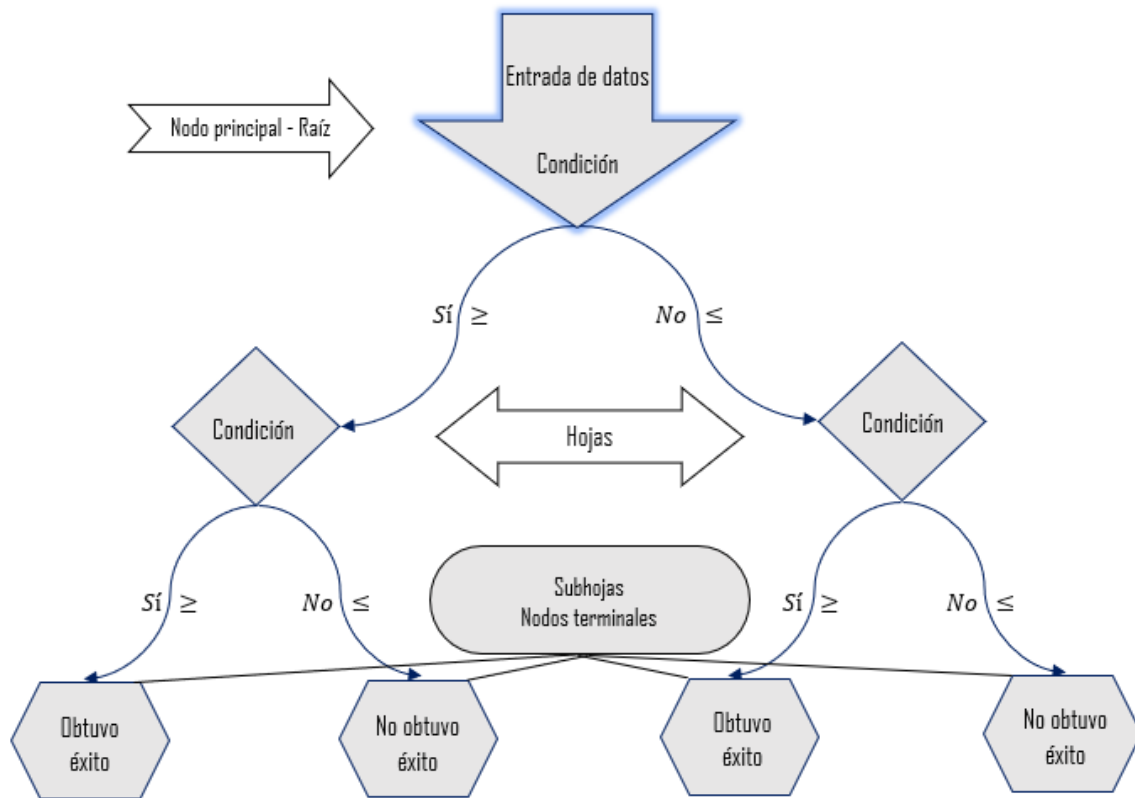
Mauricio
Toro



<http://github.com/juandacd/ST0245-002/proyecto/>



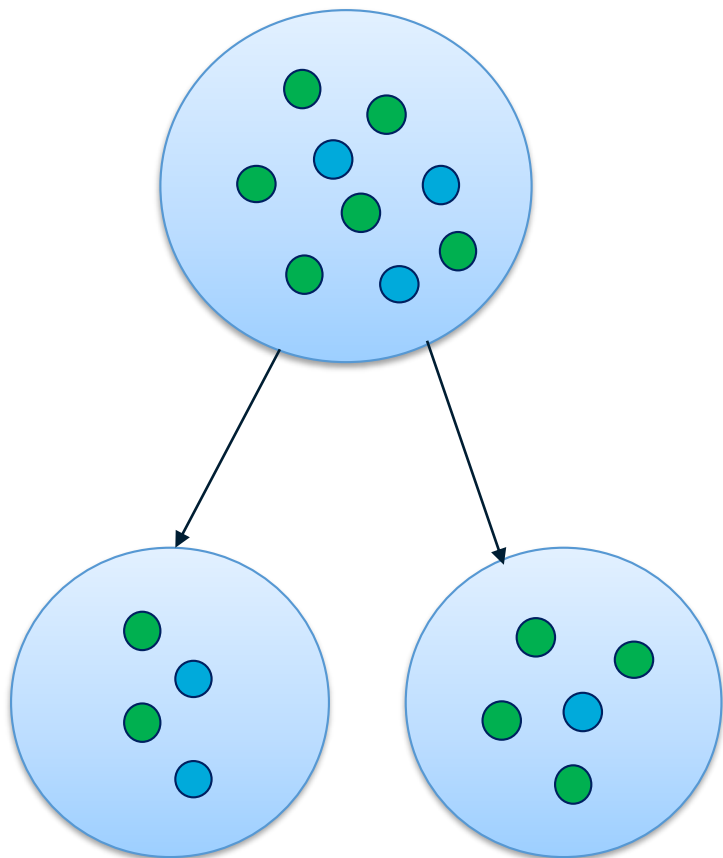
Diseño del Algoritmo



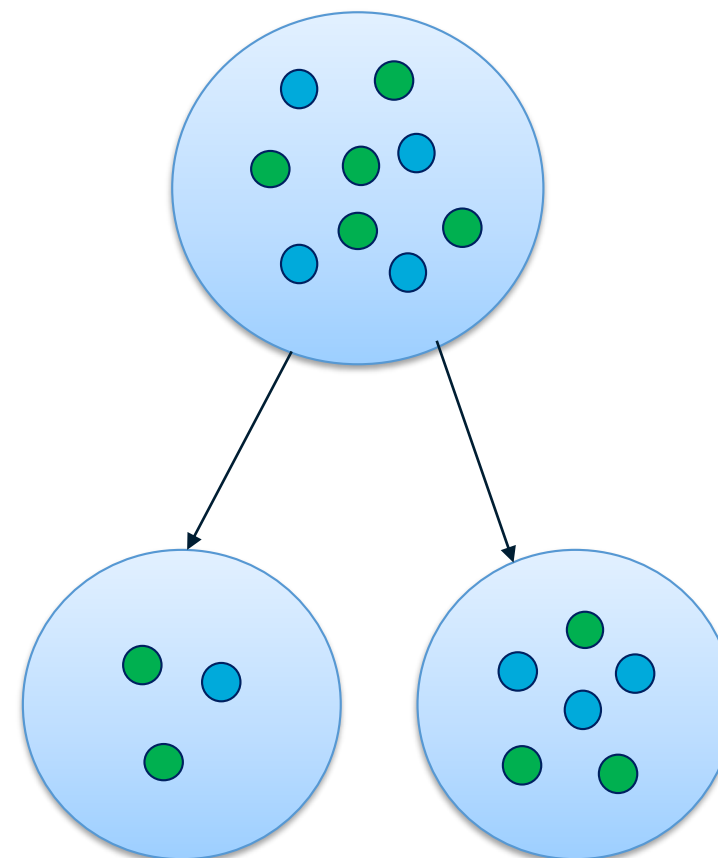
Algoritmo para construir un árbol binario de decisión usando el algoritmo CART. En este ejemplo, mostramos un diseño de un modelo para predecir si un estudiante tendrá éxito en su período académico.



División de un nodo



*Esta división está basada en la condición “puntaje en biología ≥ 59 .”
Para este caso, la impureza Gini de la izquierda es 0.45, la impureza Gini de la derecha es 0.38 y la impureza ponderada es de 0.41.*



*Esta división está basada en la condición “puntaje en matemáticas ≥ 50 ”
Para este caso, la impureza Gini de la izquierda es 0.37, la impureza Gini de la derecha es 0.5 y la impureza ponderada es 0.41.*

Complejidad del Algoritmo

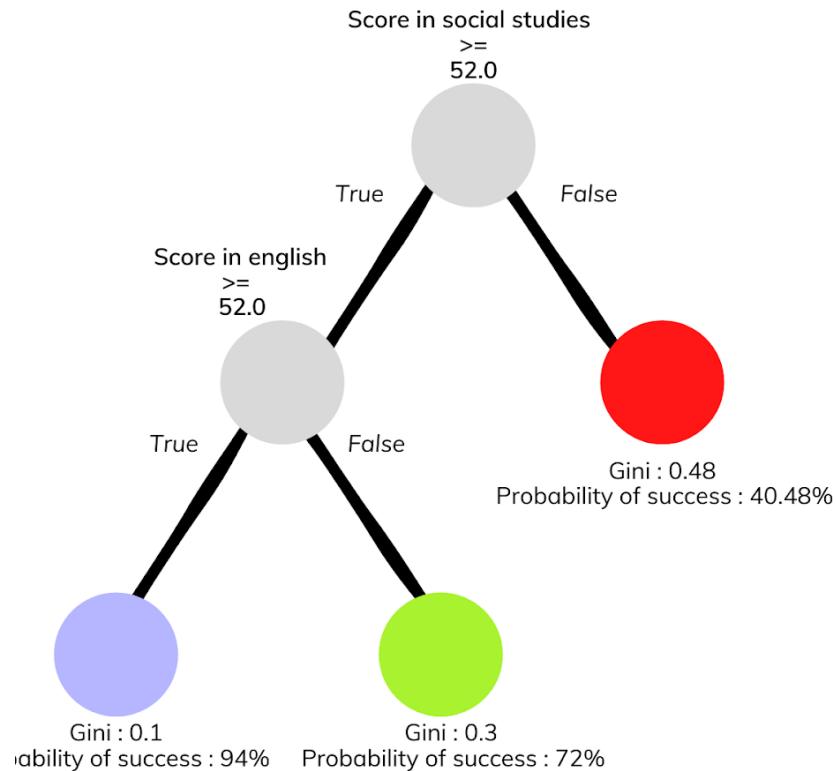


	Complejidad en tiempo	Complejidad en memoria
Entrenamiento del modelo	$O(2^M MN^2)$	$O(NM2^M)$
Validación del modelo	$O(NM)$	$O(N)$

Complejidad en tiempo y memoria del algoritmo (se empleó el algoritmo CART). Donde N son el número de filas (cantidad de estudiantes) y M son el número de columnas (cantidad de aspectos).



Modelo de Árbol de Decisión



Un árbol de decisión CART para predecir el resultado del Saber Pro usando los resultados del Saber 11. Violeta representa nodos con alta probabilidad de éxito; verde media probabilidad; y rojo baja probabilidad.

Características más relevantes:



Puntaje en matemáticas



Puntaje en lenguaje



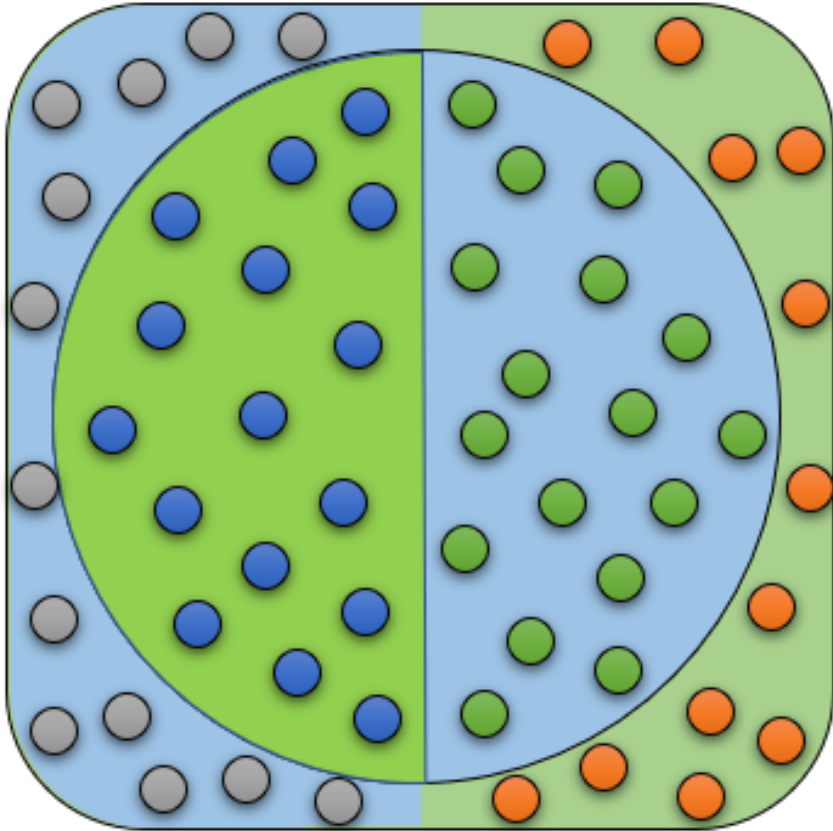
Puntaje en sociales



Puntaje en filosofía







Puntaje en inglés



$$\textit{Exactitud} = \frac{\textit{Positivos verdaderos} + \textit{Negativos verdaderos}}{\textit{Total}}$$

$$\textit{Precisión} = \frac{\textit{Positivos verdaderos}}{\textit{Total de positivos predichos}}$$

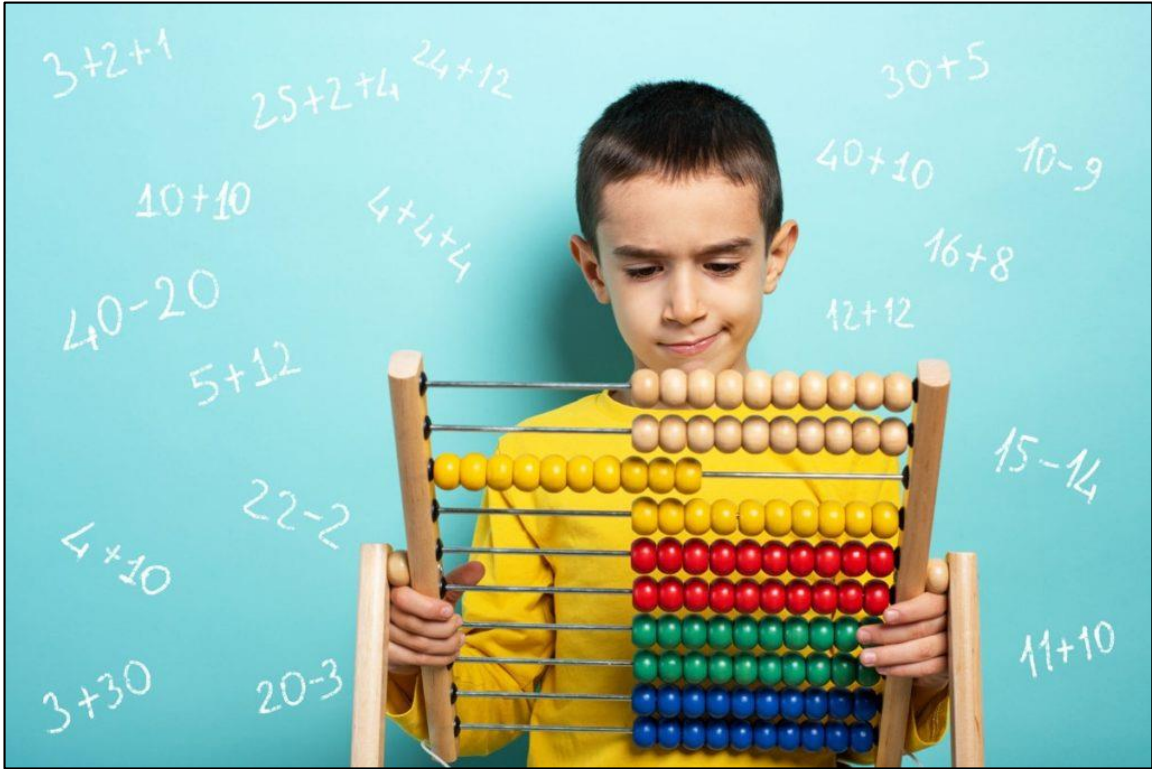
$$\textit{Sensibilidad} = \frac{\textit{Positivos verdaderos}}{\textit{Positivos reales}}$$

- | | |
|---|--|
|  <i>Positivos verdaderos</i> |  <i>Falsos positivos</i> |
|  <i>Negativos verdaderos</i> |  <i>Falsos negativos</i> |

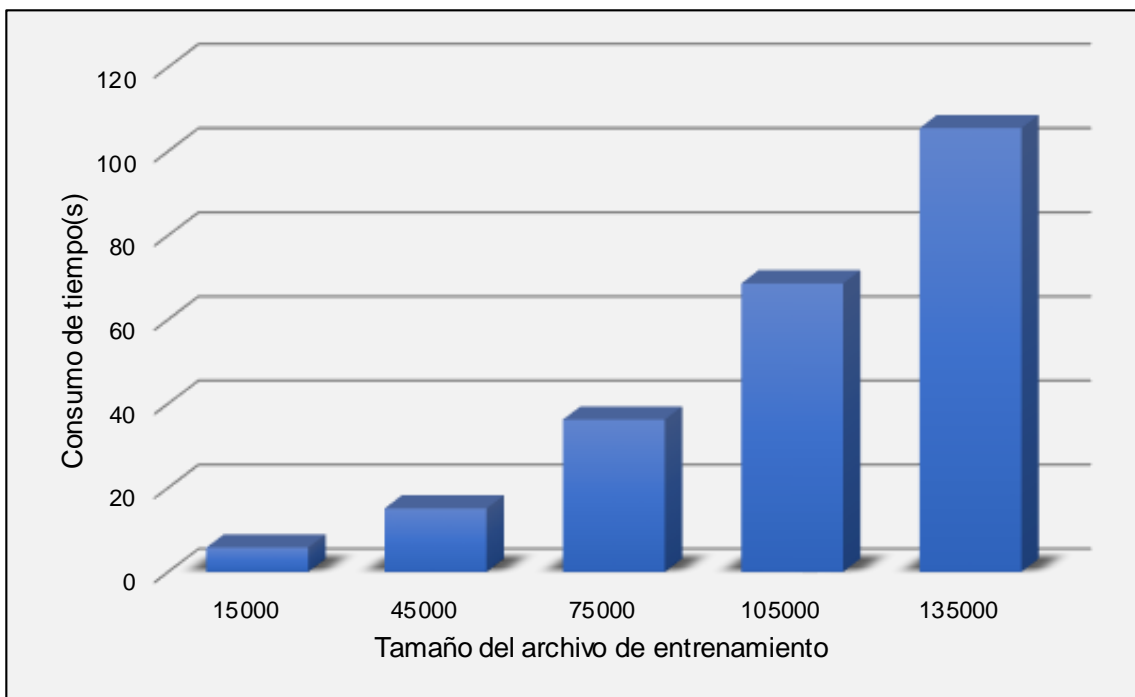


	Conjunto de entrenamiento	Conjunto de validación
Exactitud	0.78	0.7
Precisión	0.79	0.71
Sensibilidad	0.75	0.68

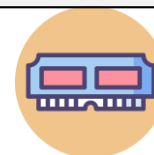
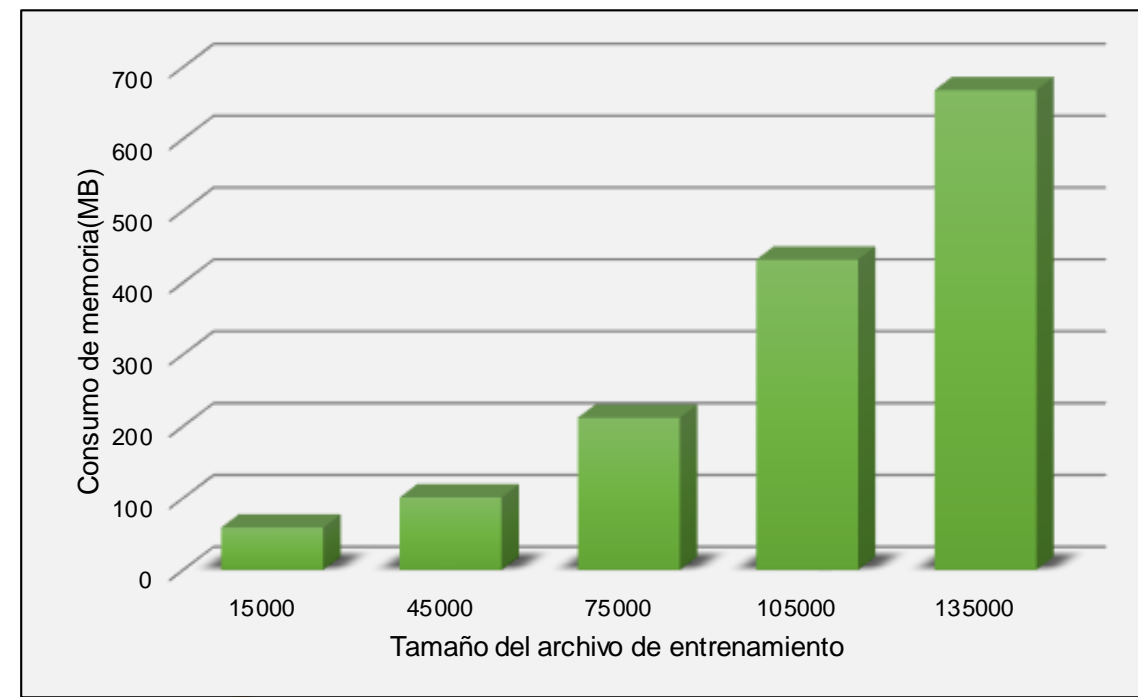
Métricas de evaluación obtenidas con el conjunto de datos de entrenamiento de 135,000 estudiantes y el conjunto de datos de validación de 45,000 estudiantes.



Consumo de tiempo y memoria



Consumo de tiempo



Consumo de memoria



¡GRACIAS!