

PREDICCIÓN DE EXCELENCIA
SABER PRO



UNIVERSIDAD
EAFIT®

Presentación del Equipo



Juan Felipe
Ortiz Salgado



Tomas Duque
Giraldo



Miguel
Correa



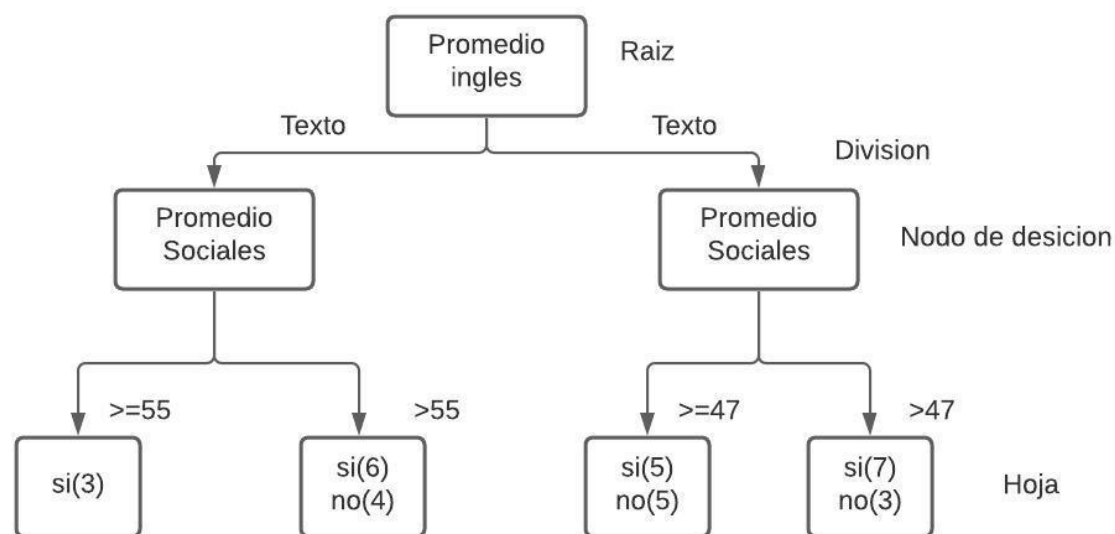
Mauricio
Toro



<http://github.com/>

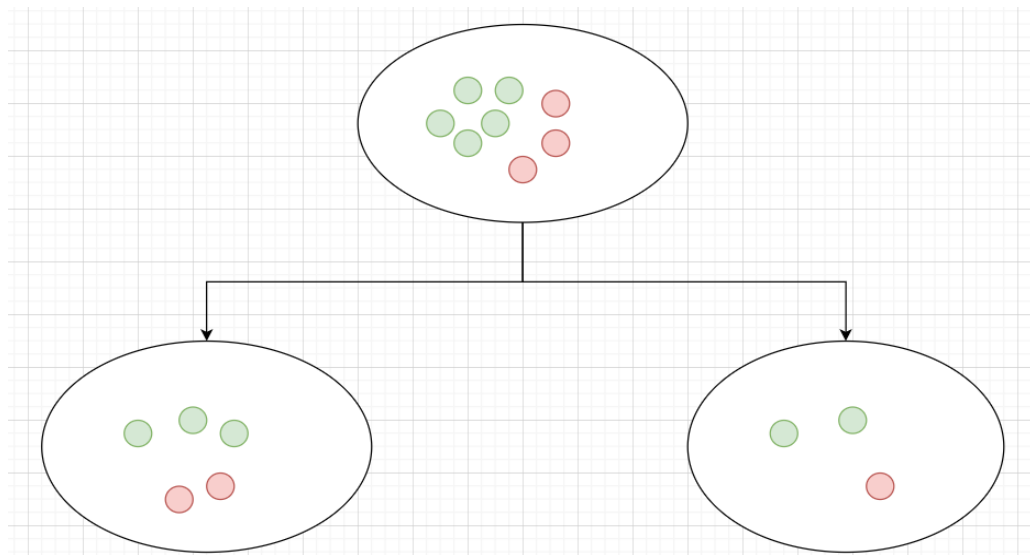
Tduqueg/proyecto/



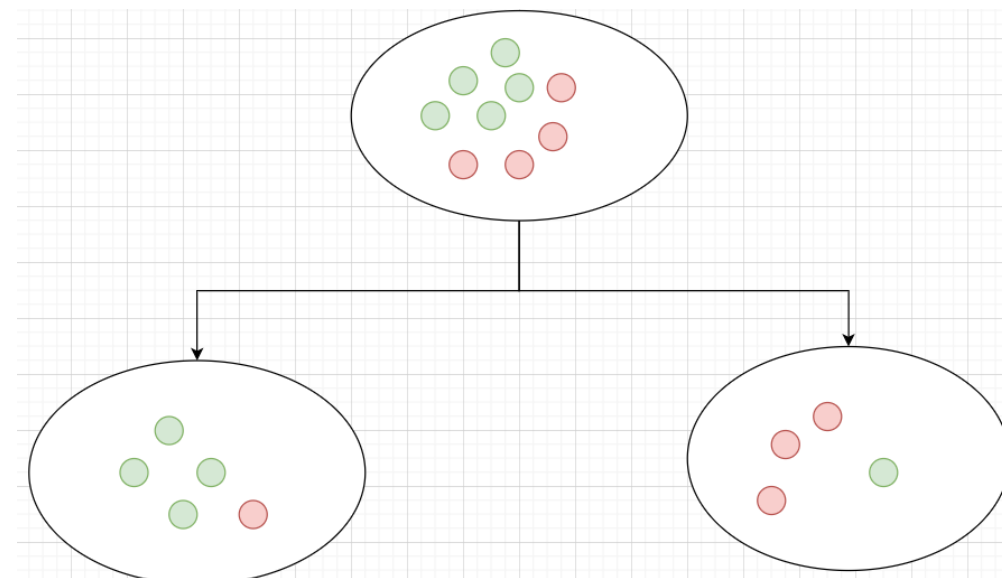


El entrenamiento de nuestro algoritmo CART, que fue el que usamos debe tomar decisiones a partir de los datos dados y con este diagrama Podemos dar un pequeño ejemplo en donde mostramos un modelo para predecir si un estudiante va a tener exito en las pruebas Saber Pro, basandonos en resultados de las pruebas Icfes.

División de un nodo

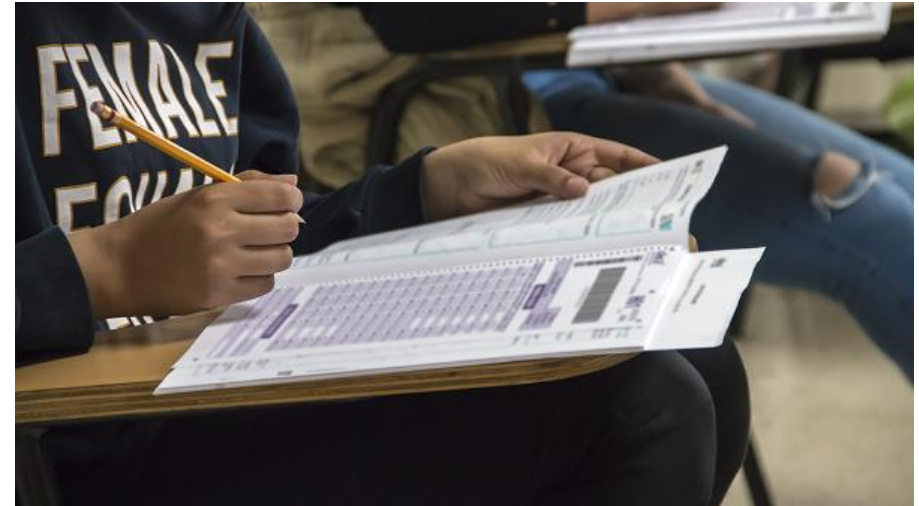


Es este diagrama empezamos tratando la impureza de gini con “Puntaje Ingles ≥ 52 ” Para este caso, la impureza Gini de la izquierda es 0.3, la impureza Gini de la derecha es 0.43 y la impureza ponderada es de 0.39.



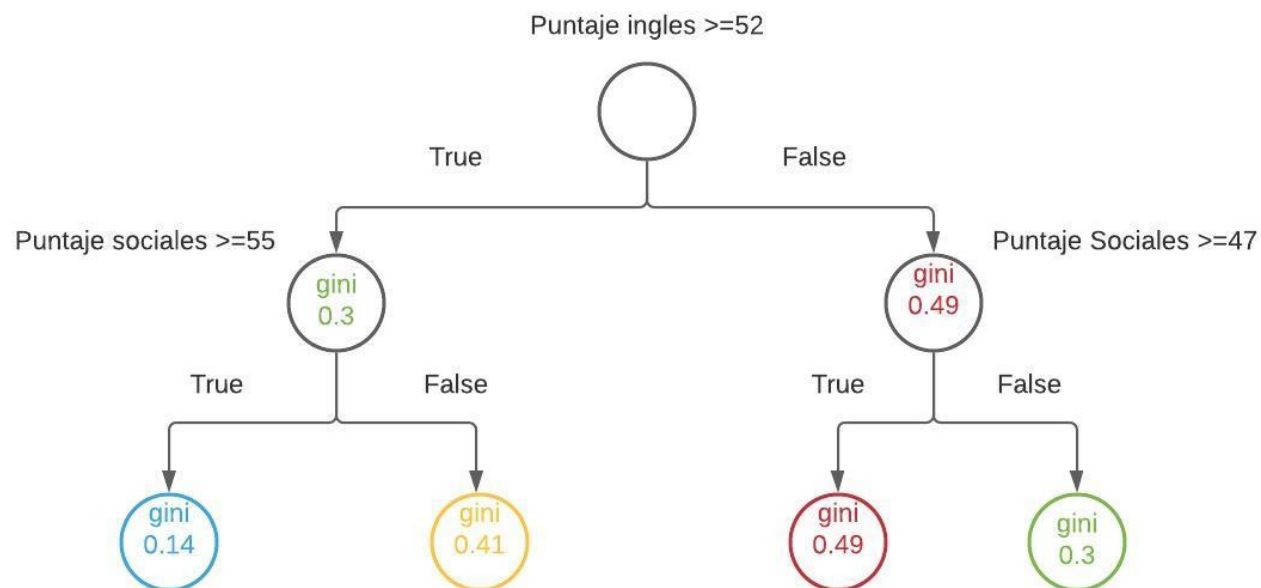
Esta división está basada en la condición “Puntaje Sociales ≥ 47 .” Para este caso, la impureza Gini de la izquierda es 0.49, la impureza Gini de la derecha es 0.29 y la impureza ponderada es 0.39.

	Complejidad 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 M la cantidad de columnas y N de filas.

Modelo del arbol



Un árbol de decisión para predecir el resultado del Saber Pro usando los resultados del Saber 11. los nodos azules representan aquellos con una alta probabilidad de éxito, los verdes con una probabilidad media, los amarillos una media-baja, y los rojos con una baja probabilidad de éxito.

Características Más Relevantes



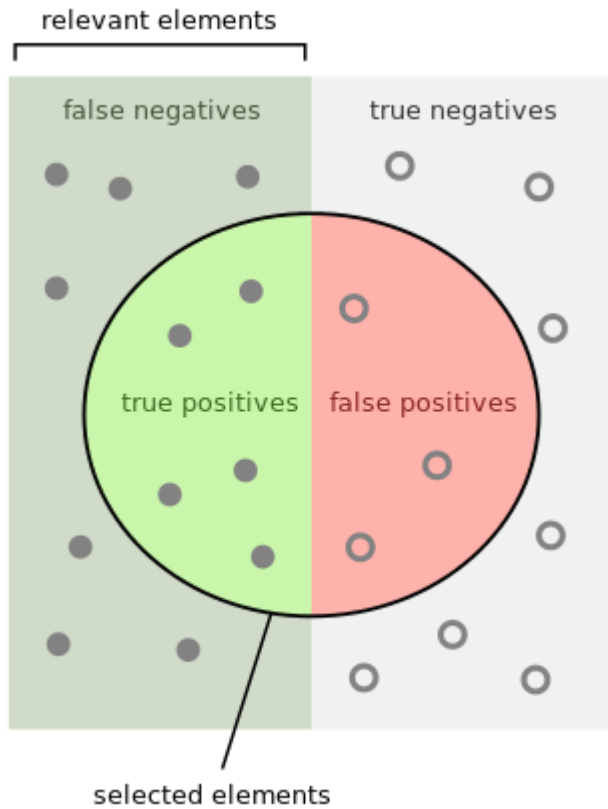
Ciencias Sociales



Inglés



Ciencias



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

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

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

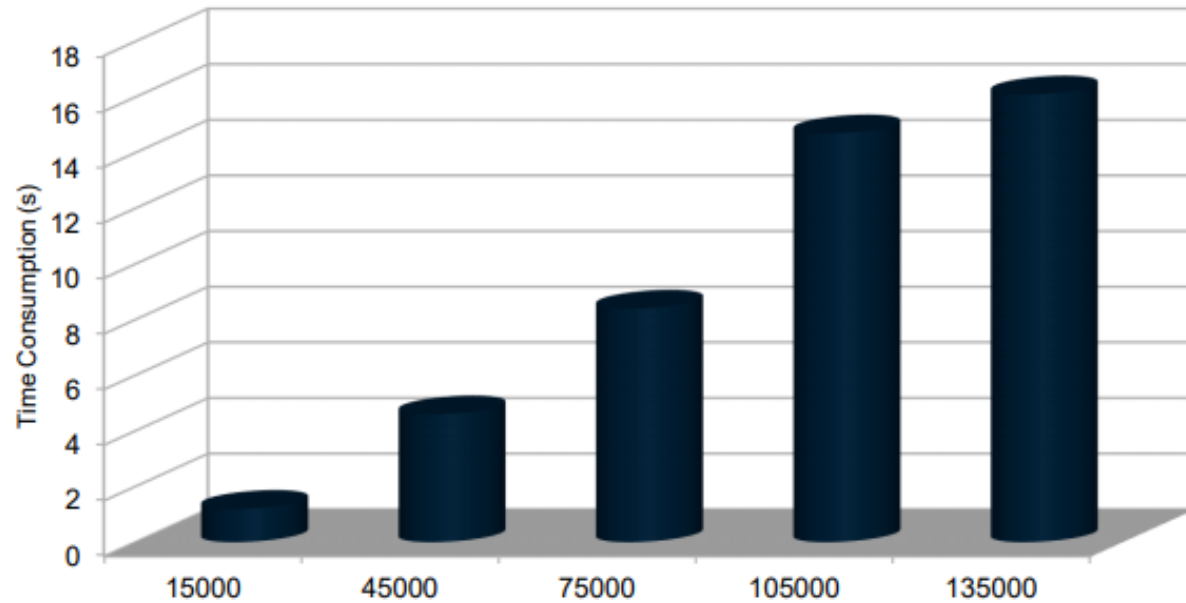


	Conjunto de entrenamiento	Conjunto de validación
Exactitud	0.8	0.67
Precisión	0.7	0.55
Sensibilidad	0.76	0.61

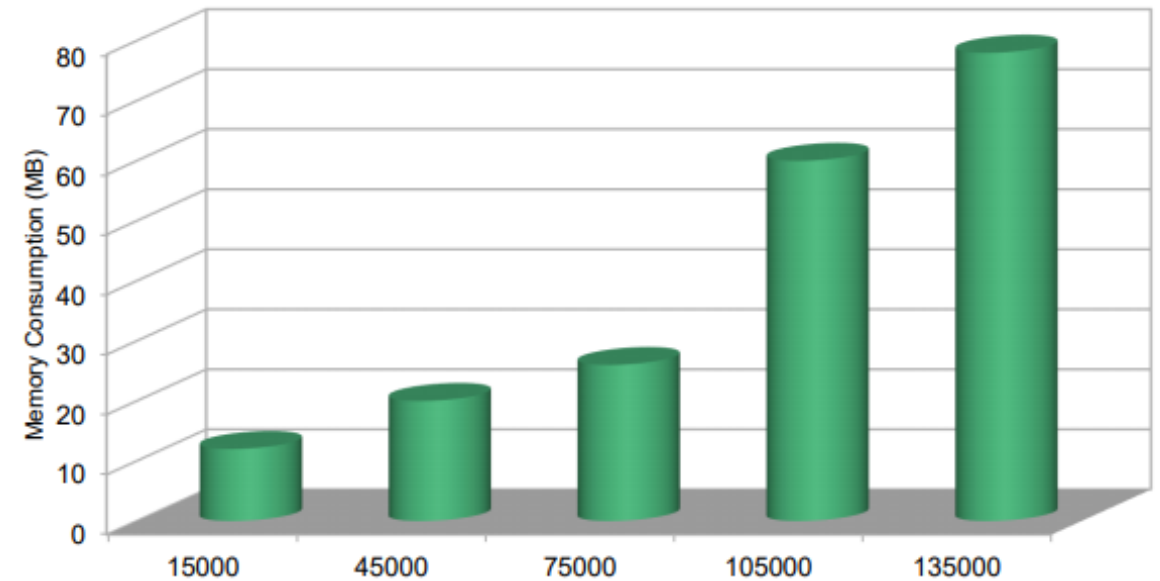
Métricas de evaluación obtenidas con el conjunto de datos de entrenamiento de 135,000 estudiantes y el conjunto de datos de validación de 45,000 estudiantes.



Consumo de tiempo y memoria



Consumo de tiempo



Consumo de memoria

A young man with dark hair is sitting on the ground, leaning forward and writing in a notebook. He is wearing a light-colored long-sleeved shirt. To his right, a medium-sized dog with white and brown fur is sitting upright, holding a small white cup in its mouth. The dog is wearing a dark red hooded jacket. In front of them, an open black toolbox sits on the ground. Inside the toolbox, there is a piece of paper with handwritten text: "Nota: CERRAMOS", "Donation Line po", "para Kay Black Jack's", "Trick & Marooning", and "C.I.". There are also some small items and papers inside the toolbox. To the left of the toolbox, there is a clear plastic bottle with a label that says "am", a metal bowl, and some other small items. The background is slightly out of focus, showing what appears to be a motorcycle and some other objects. The overall scene suggests a homeless person or a person in need, with the dog acting as a companion.

¡GRACIAS POR ESCUCHARNOS!