

# ARBOLES DE DECISIÓN PARA PREDICCIÓN DEL ÉXITO EN PRUEBAS SABER PRO



# Presentación del Equipo



Tomas  
Atehortua



Sebastian  
Velez



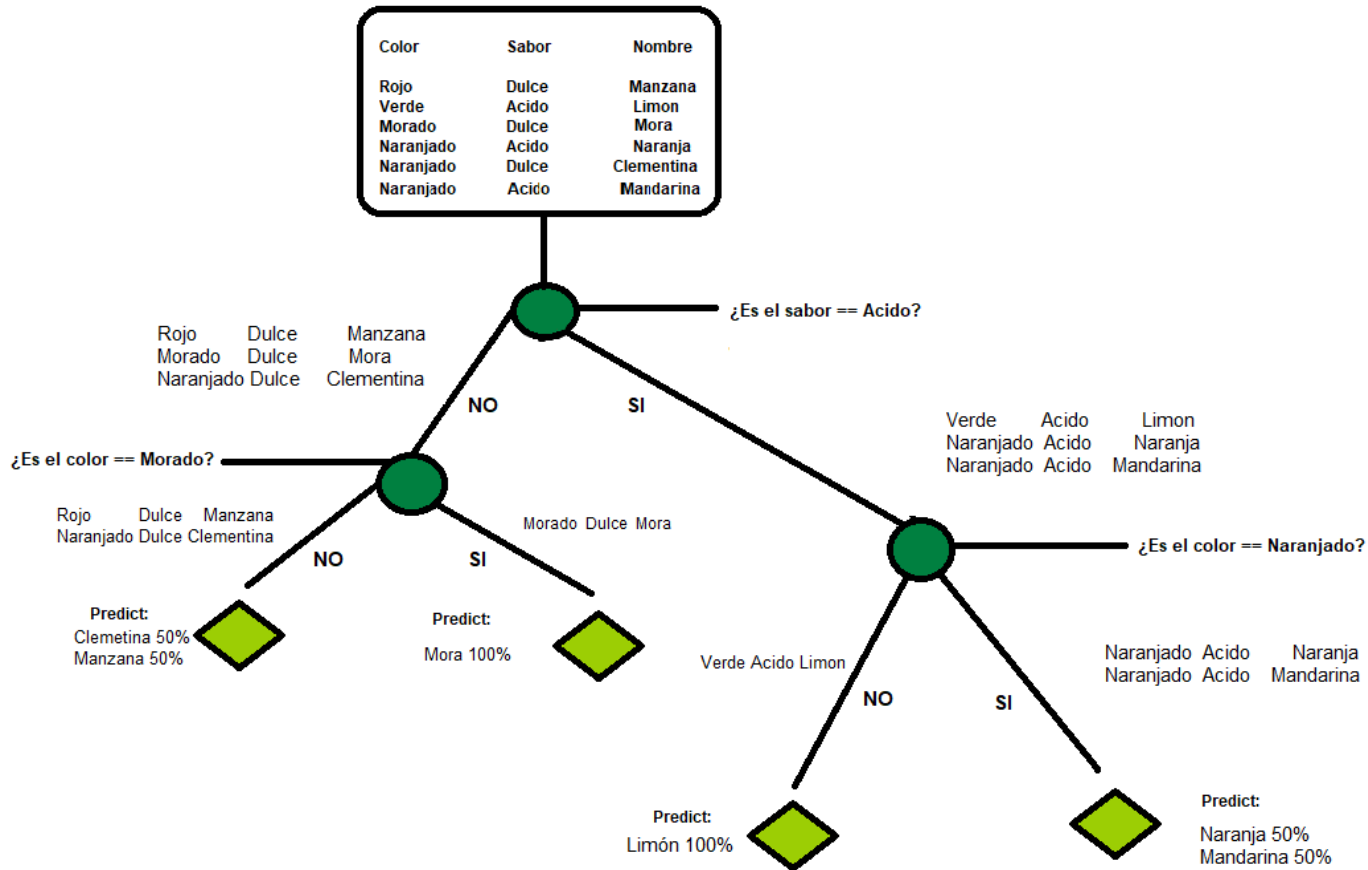
Miguel  
Correa



Mauricio  
Toro

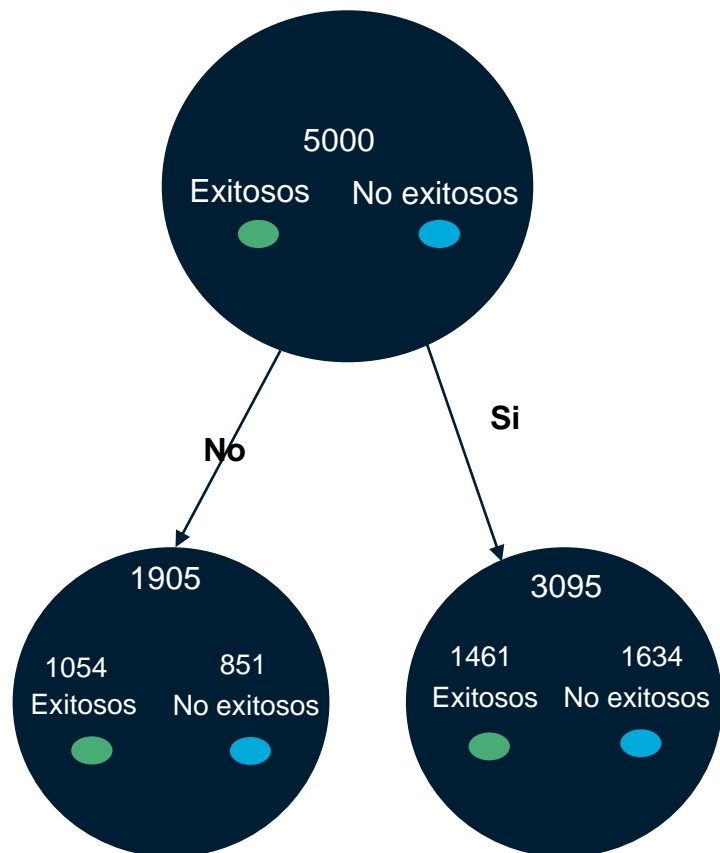


# Diseño del Algoritmo

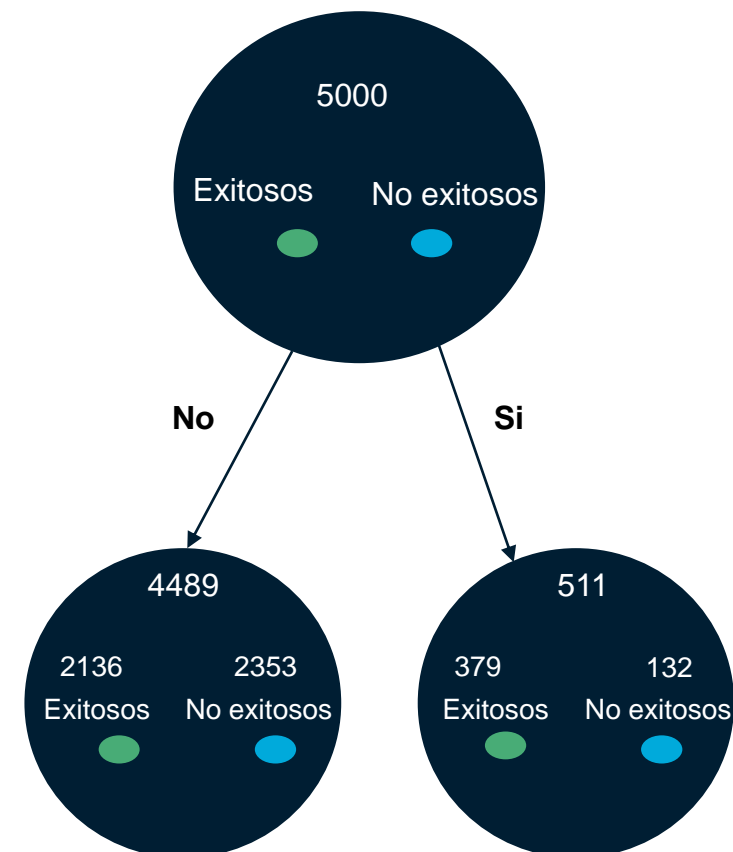


Algoritmo para construir un árbol binario de decision usando (CART).  
 En este ejemplo, mostramos un modelo para predecir que fruta comer en  
 base al sabor y al color distintivo de cada una de las seleccionadas.

# División de un nodo



Esta división está basada en la condición "Genero == F." Para este caso, la impureza Gini de la izquierda es 0.49, la impureza Gini de la derecha es 0.49 y la impureza ponderada es de 0.49.



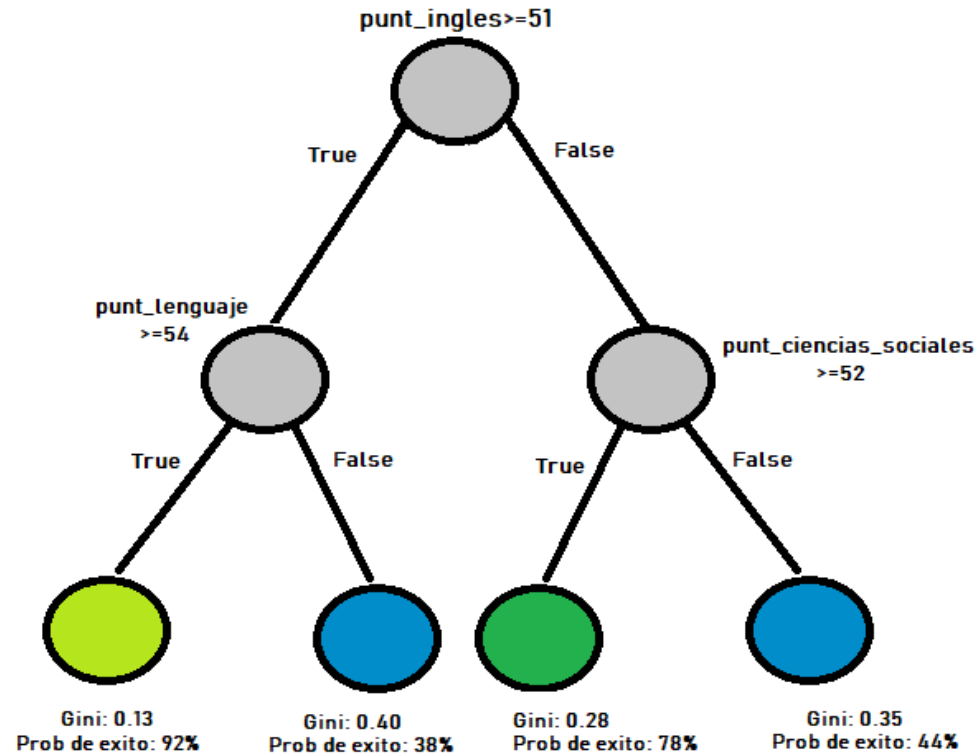
Esta división está basada en la condición "estrato == 4." Para este caso, la impureza Gini de la izquierda es 0.49, la impureza Gini de la derecha es 0.38 y la impureza ponderada es 0.47.

	Complejidad en tiempo	Complejidad en memoria
Entrenamiento del modelo	$O(N^2 * M * 2^M)$	$O(N * M * 2^M)$
Validación del modelo	$O(N * \text{LOG} M)$	$O(N * M)$

Complejidad en tiempo y memoria del algoritmo CART que diseñamos, donde  $n$  es la cantidad de estudiantes y  $m$  la cantidad de columnas



# Modelo de Árbol de Decisión



Este es un ejemplo de un árbol de decisión para predecir el resultado del Saber Pro usando los resultados del Saber 11. Verde lima representa nodos con alta probabilidad de éxito; verde oscuro media probabilidad; y azul baja probabilidad.

## Características Más Relevantes



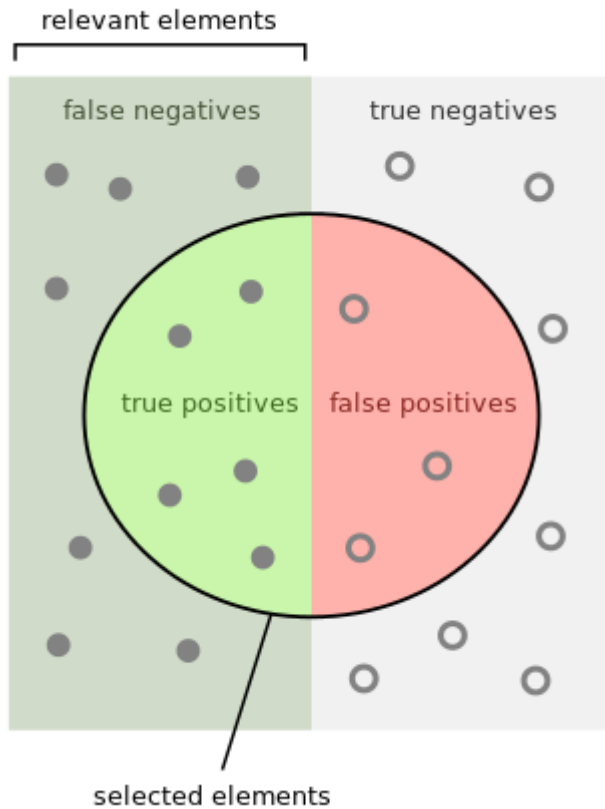
Ciencias Sociales



Inglés



Lenguaje (Lectura crítica)



How many selected items are relevant?

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

How many relevant items are selected?

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

How many items are positives?

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



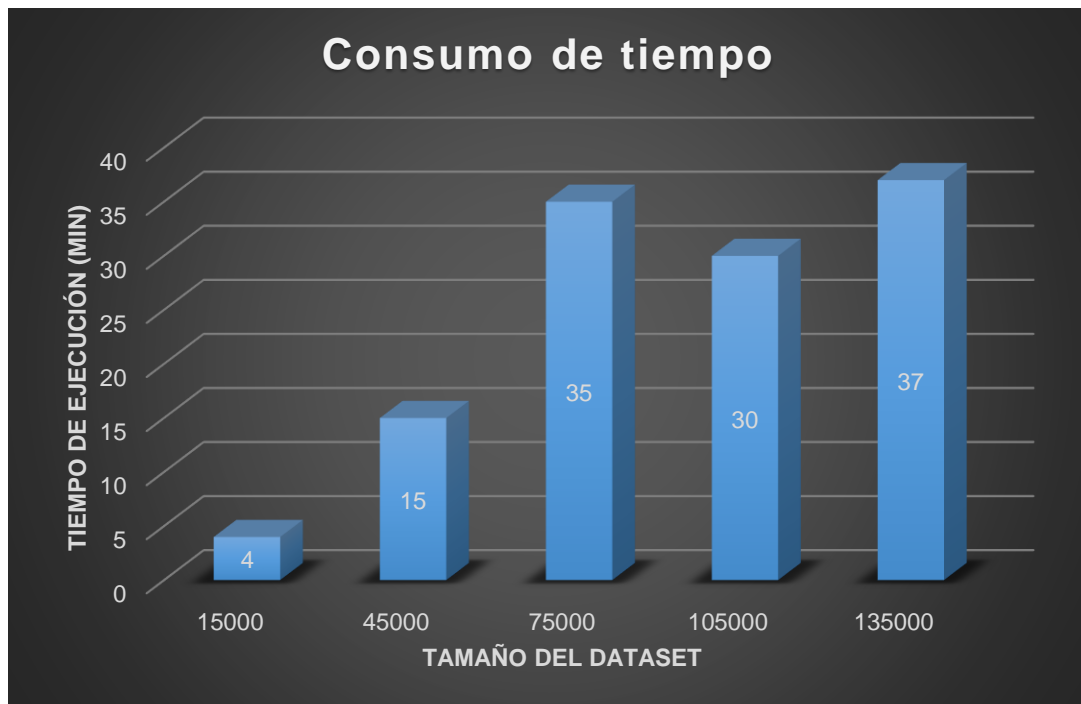
	Conjunto de entrenamiento	Conjunto de validación
Exactitud	0.7381	0.7380
Precisión	0.445	0.452
Sensibilidad	0.781	0.777

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

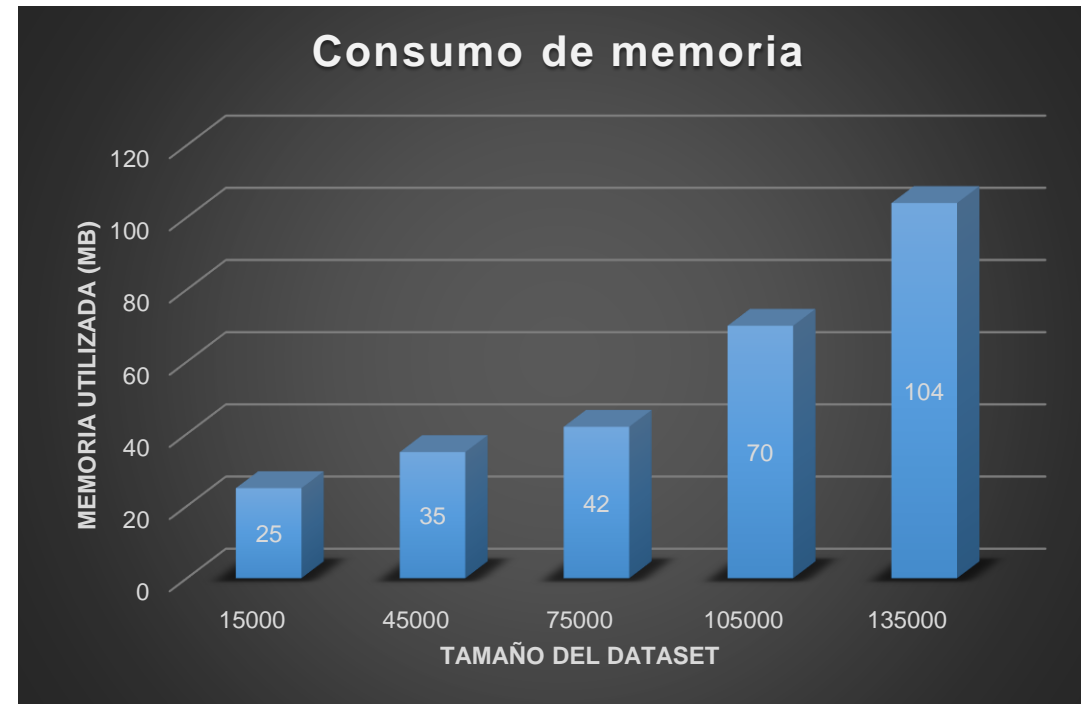




# Consumo de tiempo y memoria



Consumo de tiempo



Consumo de memoria



¡GRACIAS!