

# Prediction and analysis of state exams via the use of decision trees





Simon  
Cardenas  
Villada



Juan Pablo  
Yepes García



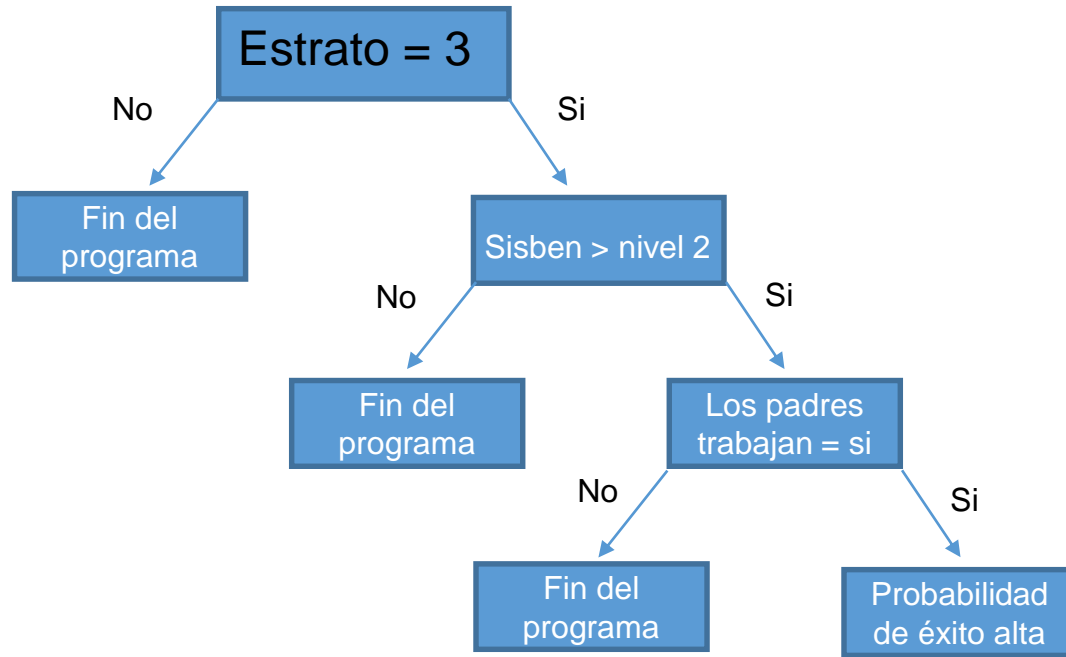
Miguel  
Correa



Mauricio  
Toro



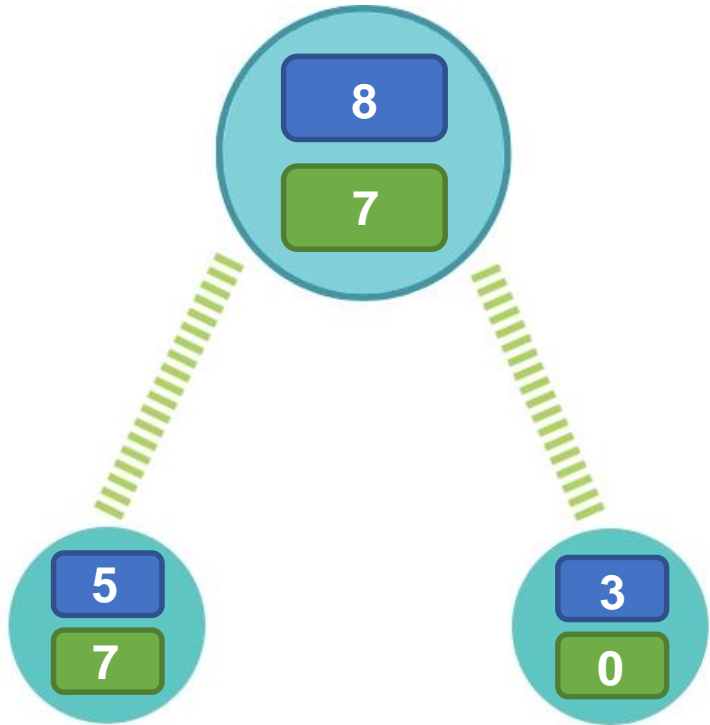
[http://github.com/](http://github.com/Jpyepes/proyecto/) Jpyepes /proyecto/



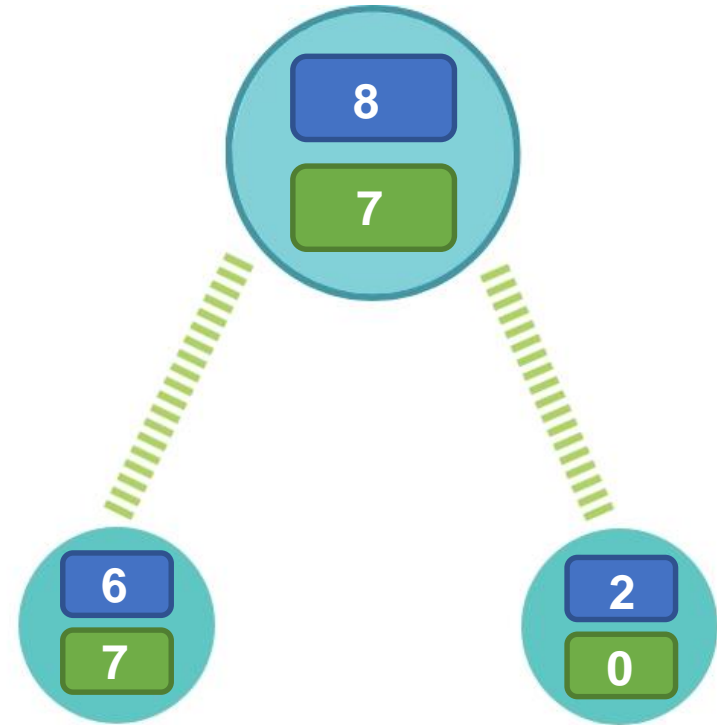
This model represents a hypothetical situation in which 3 conditions approximate the probability of success of the student.



# Node Splitting



This node split is based on the condition “fami\_nivelsisben == Nivel 2.” For this case, left Gini impurity is 0.474, right Gini impurity is 0.0, and weighted Gini impurity is 0.385. Elements in blue boxes were marked as having success “0”, and elements in green boxes with success “1”.



This node split is based on the condition “fami\_estrato vivienda.1 == Estrato 1” For this case, left Gini impurity is 0.90, right Gini impurity is 0, and weighted Gini impurity is 0.429. Elements in blue boxes were marked as having success “0”, and elements in green boxes with success “1”.

# Complejidad del Algoritmo

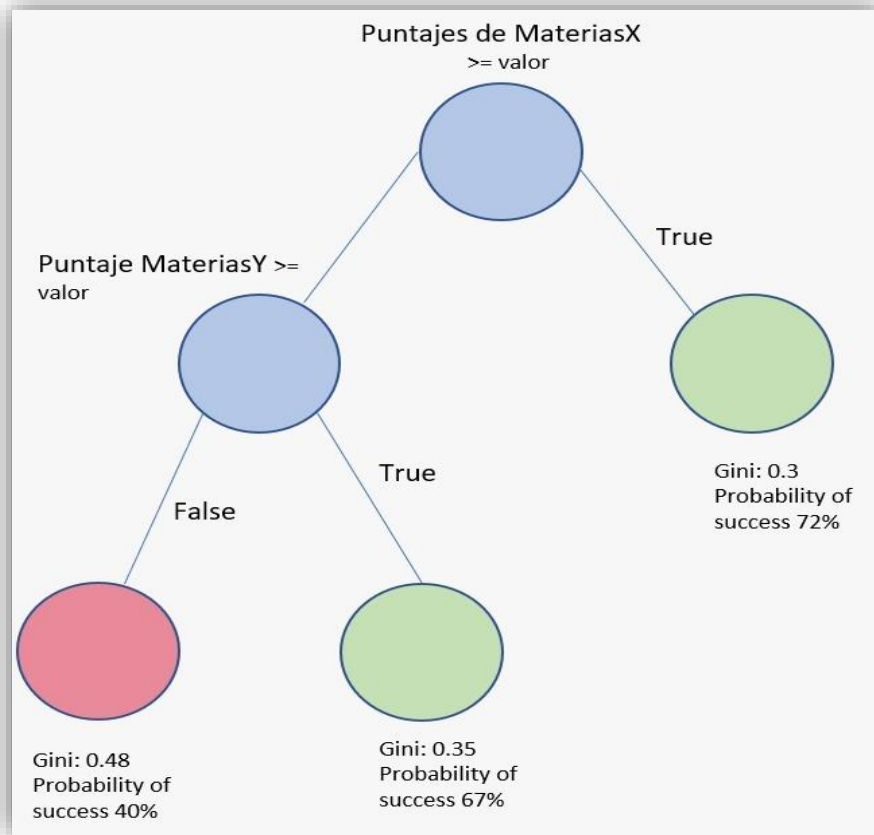


	Complejidad en tiempo	Complejidad en memoria
Entrenamiento del modelo	$O(m*n*2^m)$	$O(m*n)$
Validación del modelo	$O(m*n*\log n)$	$O(m*n)$

Donde  $n$  es el número de filas y  $m$  el de columnas de la matriz que almacena los datasets.



# Modelo de Árbol de Decisión



Gráfica de un árbol con posibles valores

## Determinantes de la predicción

*Resultados en las distintas áreas académicas que evalúa el ICFES*



El árbol solo toma en cuenta las columnas que contienen los resultados de los módulos académicos (Lenguaje, matemáticas, etc.). Se tomo esta decisión por que consideramos que es la forma mas objetiva y la que menos errores presenta en un su implementación.



	Conjunto de entrenamiento	Conjunto de validación
Exactitud	69%	
Precisión	83.55%	
Sensibilidad	52.5%	

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.

