

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



Presentación del Equipo



Simón Álvarez
Ospina



David Madrid
Restrepo



Miguel
Correa



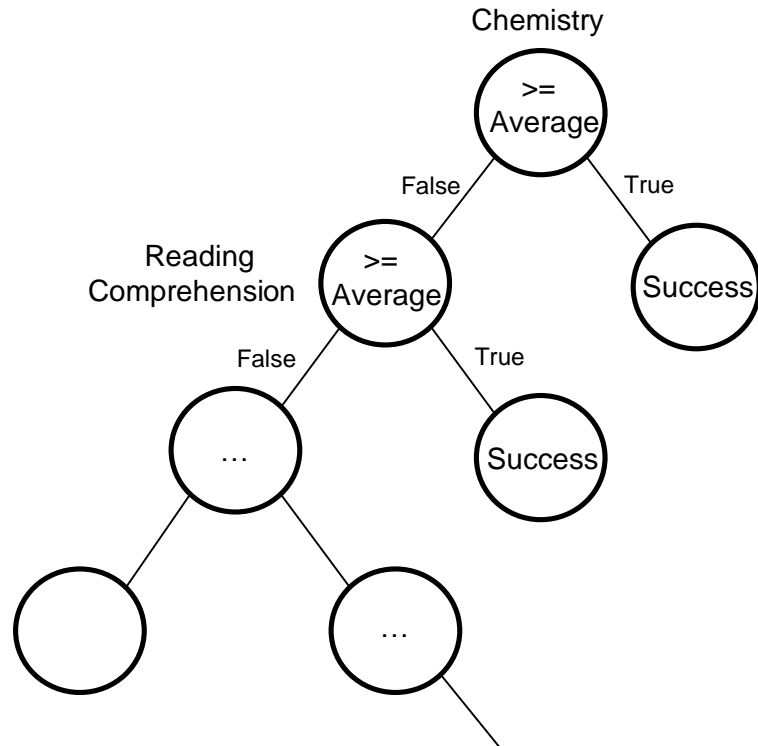
Mauricio
Toro



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

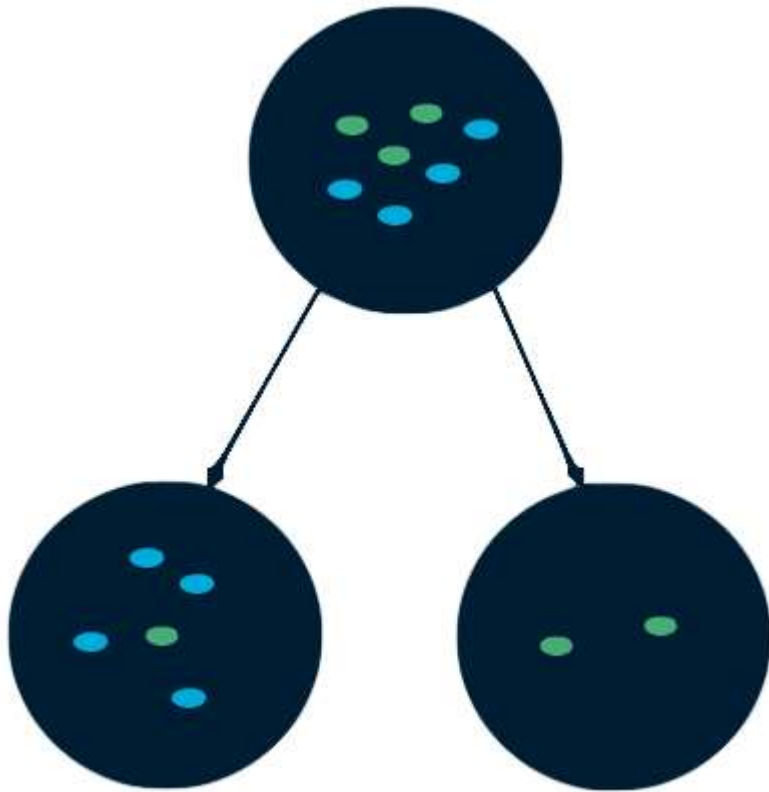


Diseño del Algoritmo

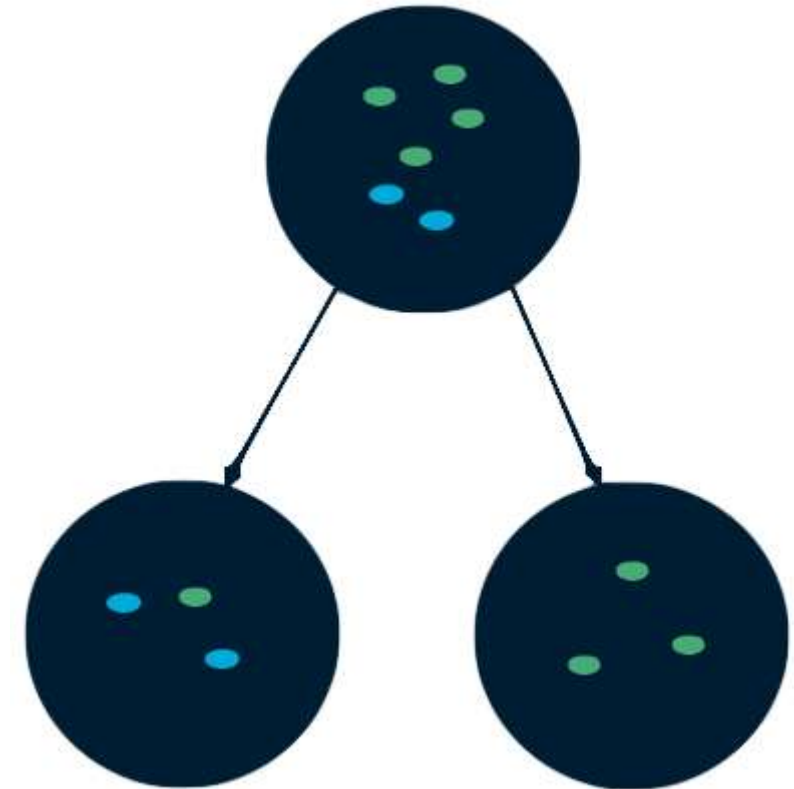


Algoritmo para construir un árbol binario de decisión usando CART. En este modelo se ilustra cómo va a ser el proceso de decisión para predecir si la persona estará o no sobre el promedio.

División de un nodo



Esta división está basada en la condición “`ingresos == 10`.” Para este caso, la impureza Gini de la izquierda es 0.44, la impureza Gini de la derecha es 0.32 y la impureza ponderada es de 0.37.



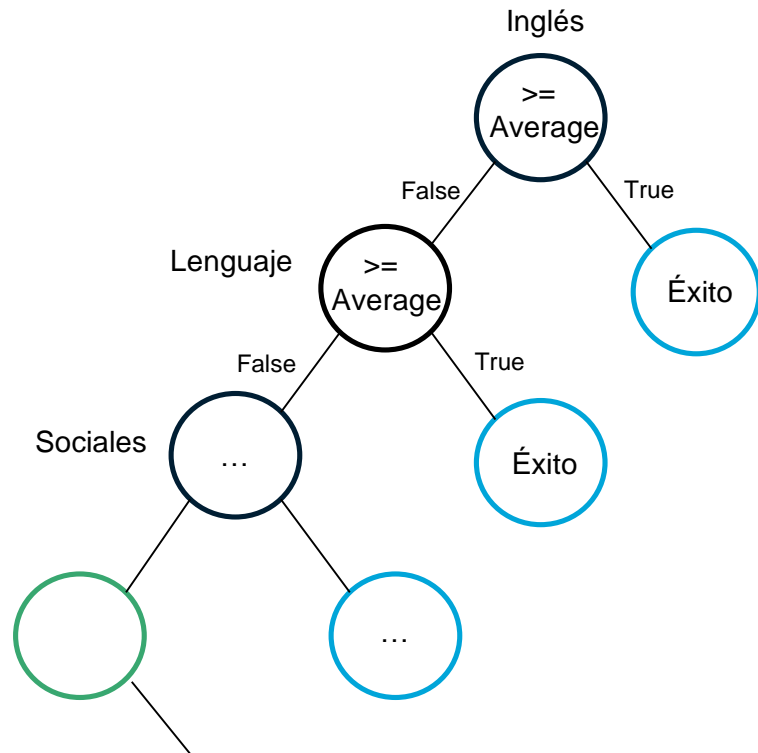
Esta división está basada en la condición “`estrato == 4`.” Para este caso, la impureza Gini de la izquierda es 0.44, la impureza Gini de la derecha es 0.32 y la impureza ponderada es 0.45.

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



Complejidad en tiempo y memoria del algoritmo. Donde N representa la altura del árbol y M representa el número de estudiantes.

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

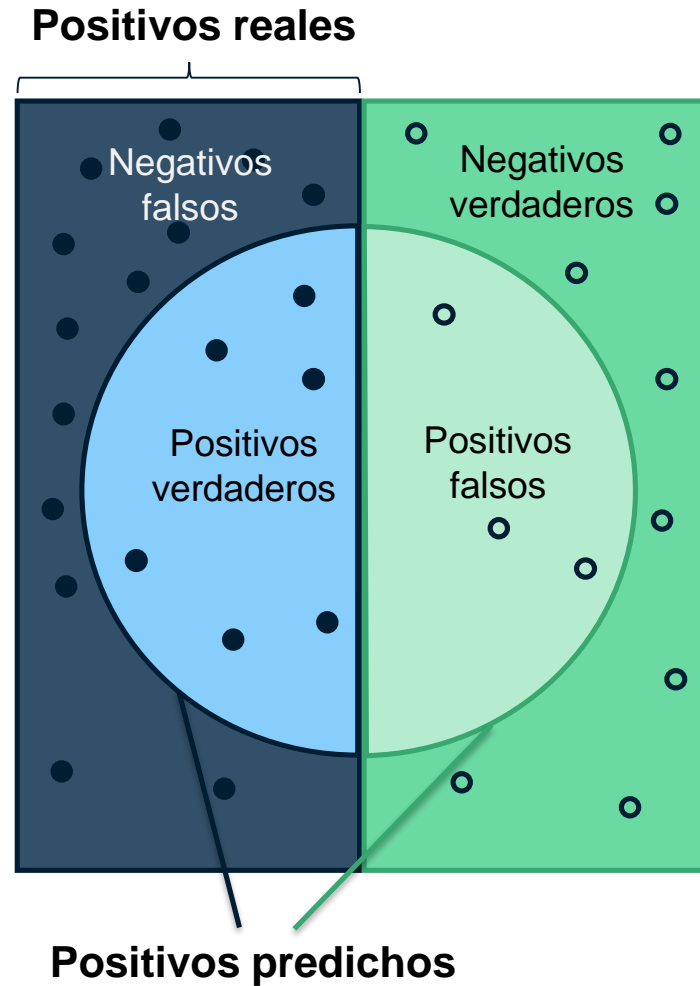


Inglés



Lenguaje

Métricas de Evaluación



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

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

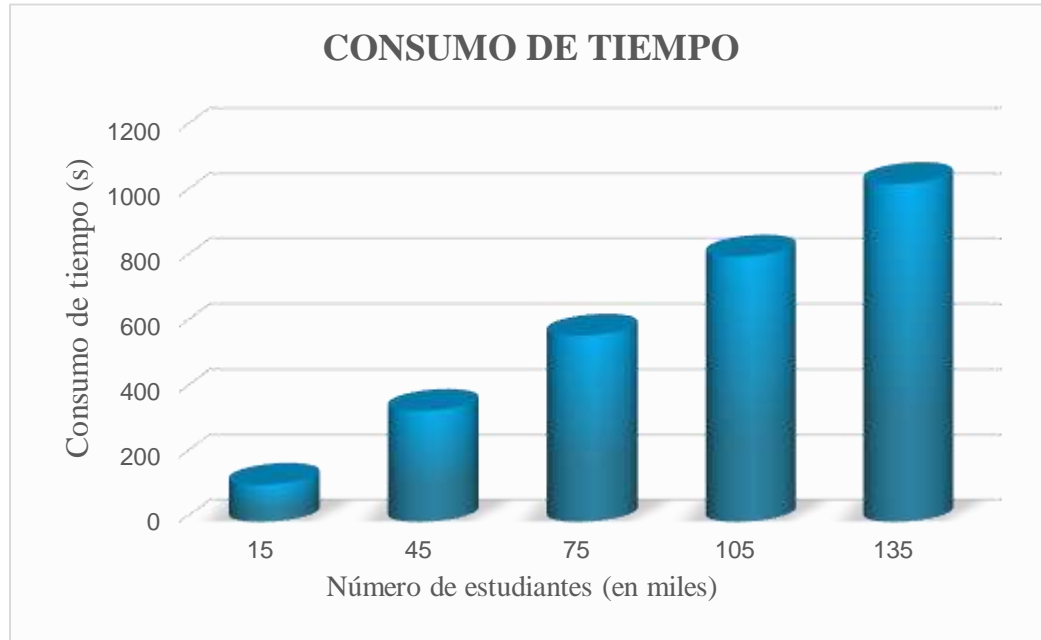
$$\text{Exactitud} = \frac{\text{Positivo verdadero} + \text{Negativo verdadero}}{\text{Total}}$$

	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



C. Patiño-Forero, M. Agudelo-Toro, and M. Toro. Planning system for deliveries in Medellín. ArXiv e-prints, Nov. 2016. Available at: <https://arxiv.org/abs/1611.04156>

 Cornell University

arXiv.org > cs > arXiv:1611.04156

Computer Science > Data Structures and Algorithms

[Submitted on 13 Nov 2016]

Planning system for deliveries in Medellín

Catalina Patiño-Forero, Mateo Agudelo-Toro, Mauricio Toro

Here we present the implementation of an application capable of planning the shortest delivery route in the city of Medellín, Colombia. We discuss the different approaches to this problem which is similar to the famous Traveling Salesman Problem (TSP), but differs in the fact that, in our problem, we can visit each place (or vertex) more than once. Solving this problem is important since it would help people, especially stores with delivering services, to save time and money spent in fuel, because they can plan any route in an efficient way.

Comments: 5 pages, 9 figures

Subjects: **Data Structures and Algorithms (cs.DS)**

ACM classes: F.2.0; G.2.2

Cite as: [arXiv:1611.04156](https://arxiv.org/abs/1611.04156) [cs.DS]
(or [arXiv:1611.04156v1](https://arxiv.org/abs/1611.04156v1) [cs.DS] for this version)



¡GRACIAS!