

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





Simon
Cardenas
Villada



Juan Pablo
Yepes García



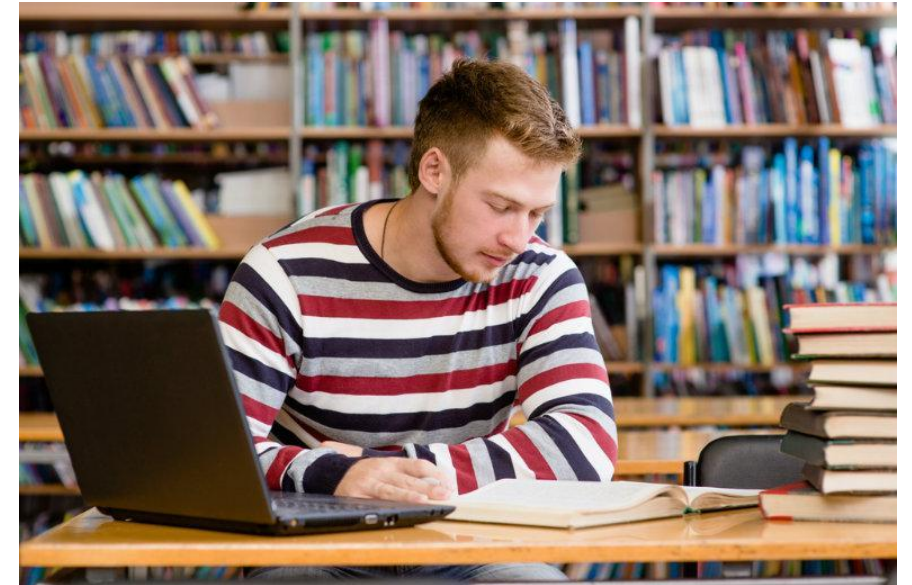
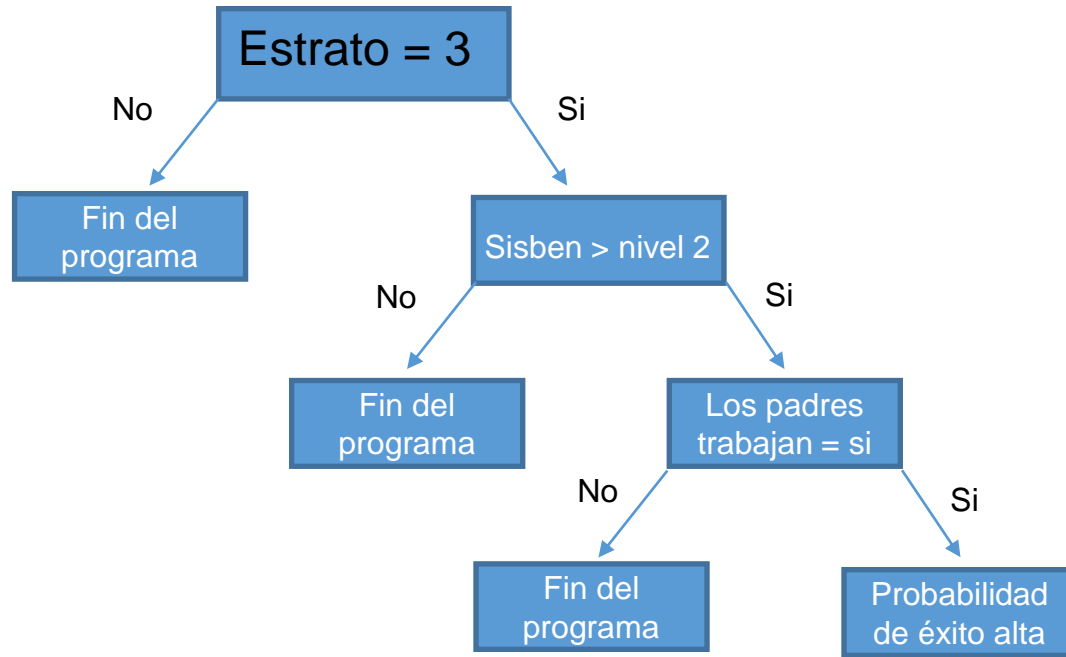
Miguel
Correa



Mauricio
Toro

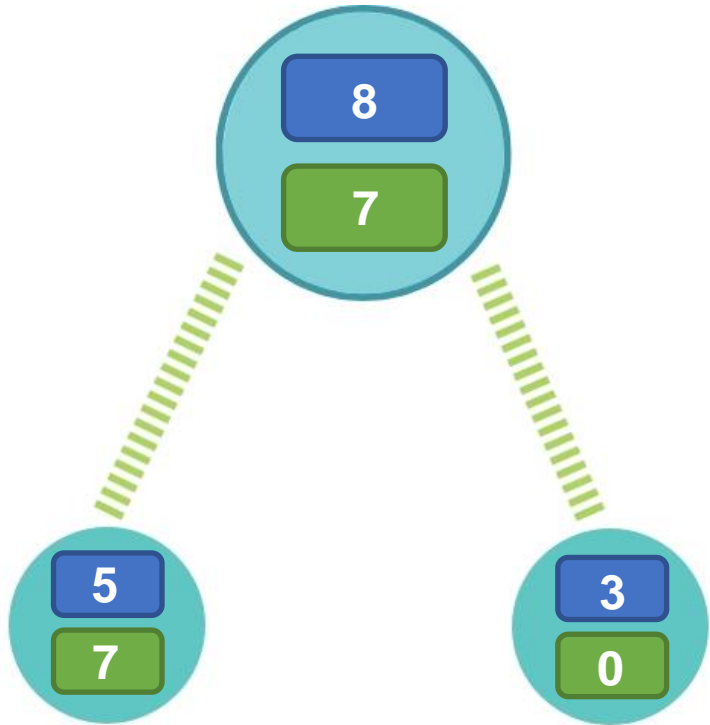


[http://github.com/](http://github.com/Jpyepes/proyecto/) Jpyepes /proyecto/

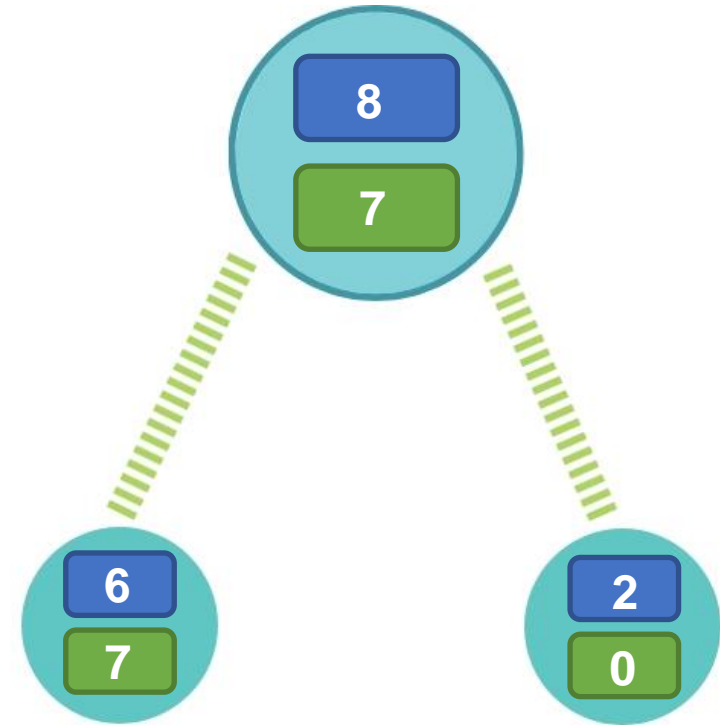


This model represents a hypothetical situation in which 3 conditions approximate the probability of success of the student.

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_estratovivienda.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



Conserven ese título

Completen esta lámina
en la tercera entrega



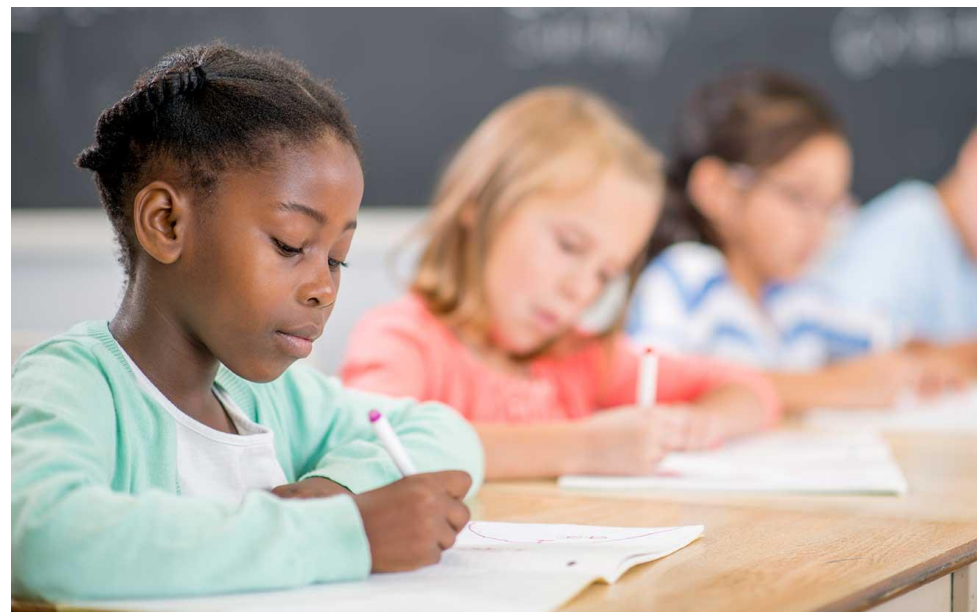
Creen esta tabla en Powerpoint. ¡No
copien pantallazos pixelados del porte
aquí!

	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 (En este semestre, una opción puede ser CART, ID3, C4.5, elijan uno). (Por favor, expliquen qué es N y qué es M en este problem. ¡POR FAVOR, HÁGANLO!)



Expliquen las tablas con
sus propias palabras



Incluyan una foto de alta definición
relacionada con el problema que
están modelando

Modelo de Árbol de Decisión

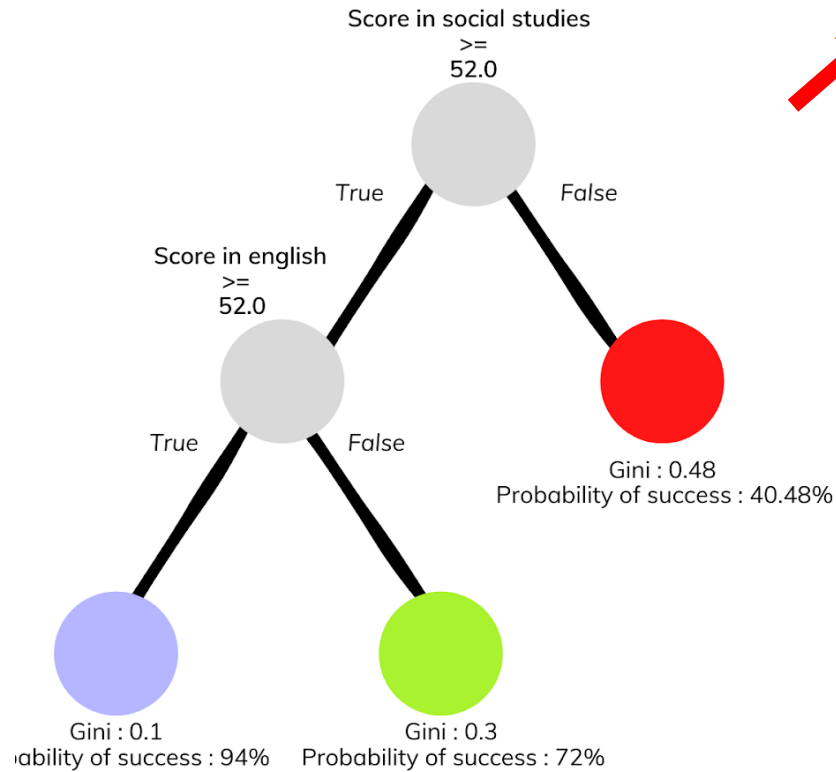
Conserven ese título

Completen esta lámina
en la tercera entrega



Usen estos colores
en sus gráficas

Creen una gráfica, en español, en
Powerpoint. ¡No copien pantallazos
pixelados del reporte técnico, por favor!



Un árbol de decisión 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.

Expliquen sus gráficos con
sus propias palabras

Características Más Relevantes



Ciencias Sociales



Inglés



Género

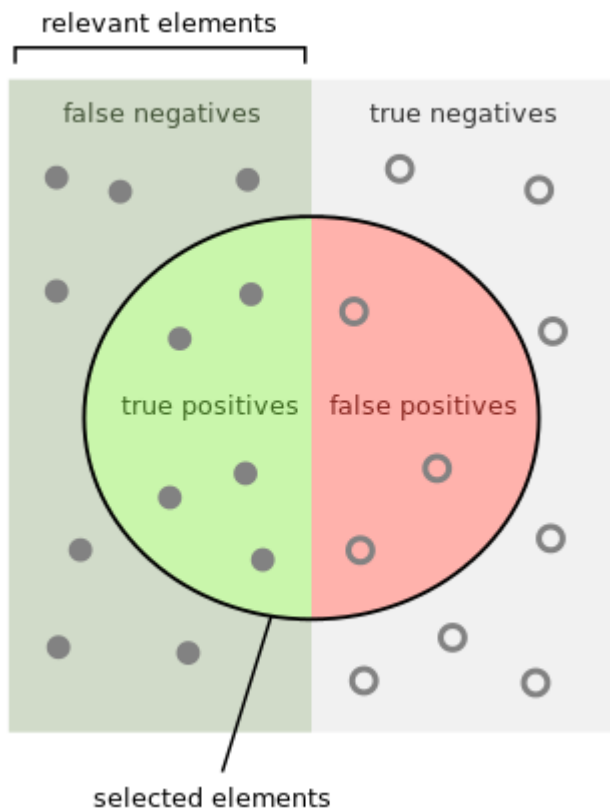
¡Usen un ícono para
representar cada
característica!

¿Es ético usar el género en
un modelo que sirve para
predecir el éxito académico?

Métricas de Evaluación

Conserven ese título

Completen esta lámina
en la tercera entrega



Usen gráficas vectorizadas, en español,
para explicar las métricas de evaluación,
de esa forma no les quedará pixelado
como las mías

How many selected
items are relevant?

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

How many relevant
items are selected?

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

Expliquen la exactitud
también....
De la misma manera

Usen estos colores
para sus gráficas

Si es posible, eviten usar
ecuaciones para explicar simples
conceptos que se pueden explicar
con diagramas coloridos

Métricas de Evaluación

Conserven ese título

Completen esta lámina
en la tercera entrega



Crean la tabla en Powerpoint. ¡No copien
pantallazos pixelados del reporte, por
favor!

	Conjunto de entrenamiento	Conjunto de validación
Exactitud	0.8	0.62
Precisión	0.6	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.



Incluyan otra gráfica en alta
definición relacionada con el
problema que están resolviendo.

Expliquen las tablas con sus
propias palabras

Consumo de tiempo y memoria

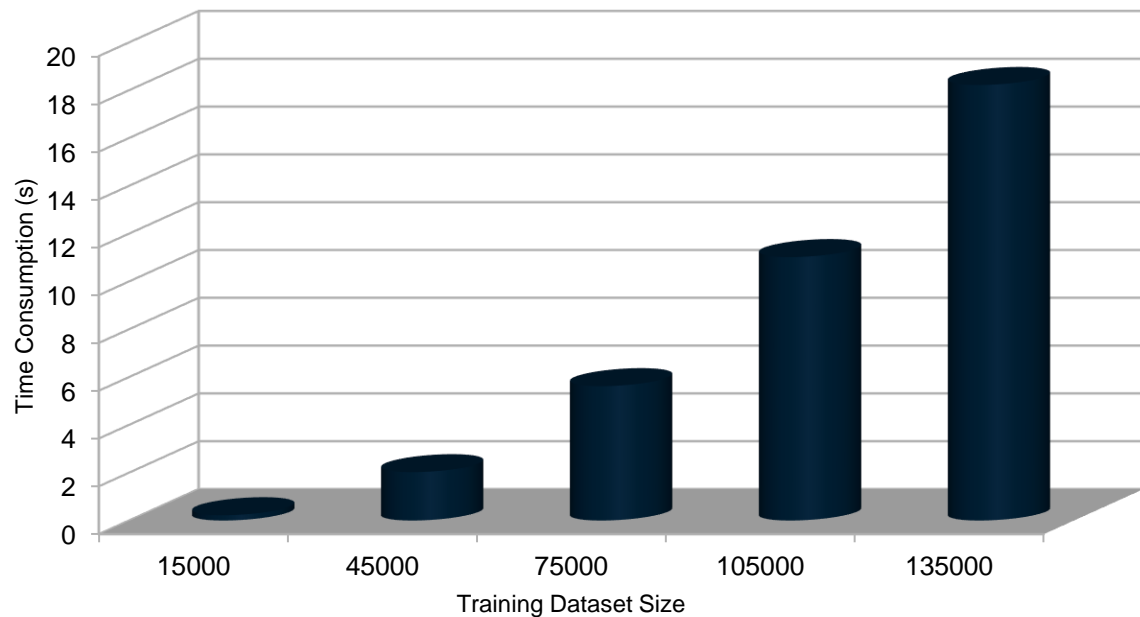
Conserven ese título

Completen esta lámina
en la tercera entrega

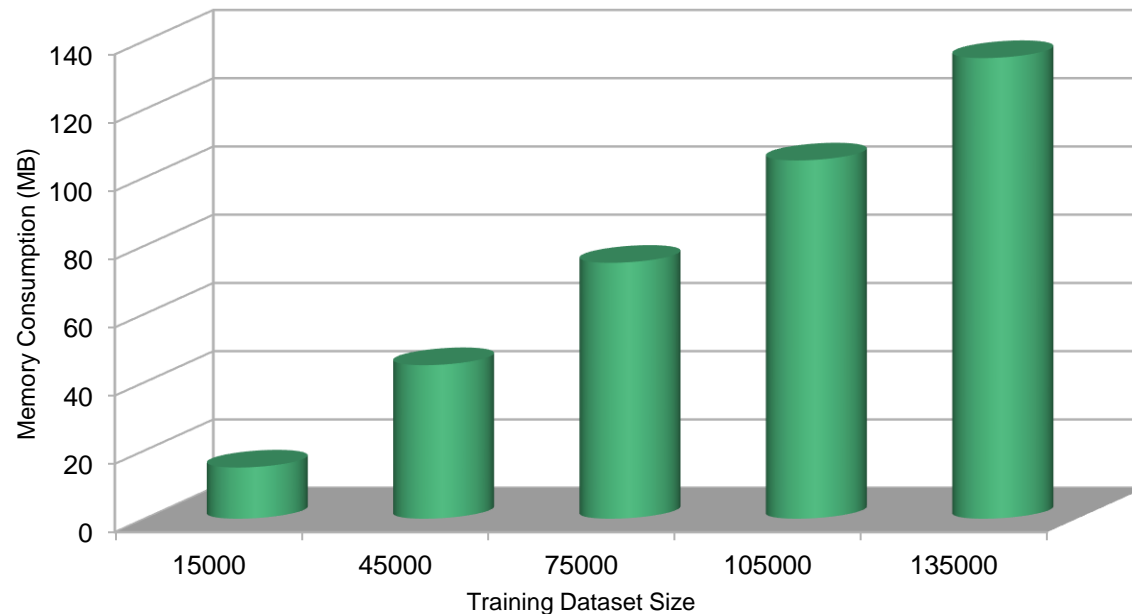


Creen las gráficas en Excel en español. ¡No
tomen pantallazos pixelados del reporte!

Usen estos colores
para sus gráficas



Consumo de tiempo



Consumo de memoria



Incluyan la citación del reporte en arXiv y
su vínculo comose muestra abajo

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>

Incluyan un
pantallazo

The screenshot shows the top of an arXiv page. At the top is the Cornell University logo and name. Below that is a red banner with the text 'arXiv.org > cs > arXiv:1611.04156'. Underneath is a grey banner with 'Computer Science > Data Structures and Algorithms'. The submission date '[Submitted on 13 Nov 2016]' is shown. The title 'Planning system for deliveries in Medellín' is in large bold black font. Below the title are the authors 'Catalina Patiño-Forero, Mateo Agudelo-Toro, Mauricio Toro' in blue. The abstract follows, starting with '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.' At the bottom, there are fields for 'Comments: 5 pages, 9 figures', 'Subjects: Data Structures and Algorithms (cs.DS)', 'ACM classes: F.2.0; G.2.2', and 'Cite as: arXiv:1611.04156 [cs.DS] (or arXiv:1611.04156v1 [cs.DS] for this version)'.

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 [cs.DS]**
(or **arXiv:1611.04156v1 [cs.DS]** for this version)

Completen esta lámina
en la tercera entrega

Digan gracias por
escucharnos

¡GRACIAS!

