

**DISEÑO DE UN
ALGORITMO
PARA PREDECIR
EL ÉXITO EN LAS
PRUEBAS
SABER PRO**



Presentación del equipo



Dennis
Castrillon



Sebastian
Castaño



Miguel
Correa



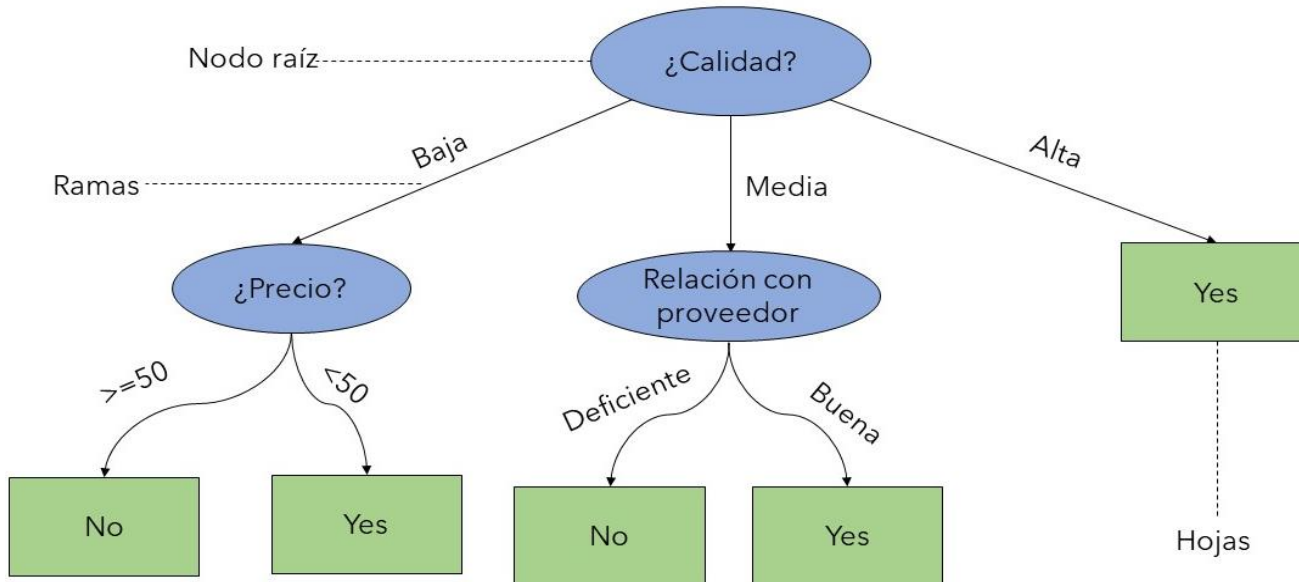
Mauricio
Toro



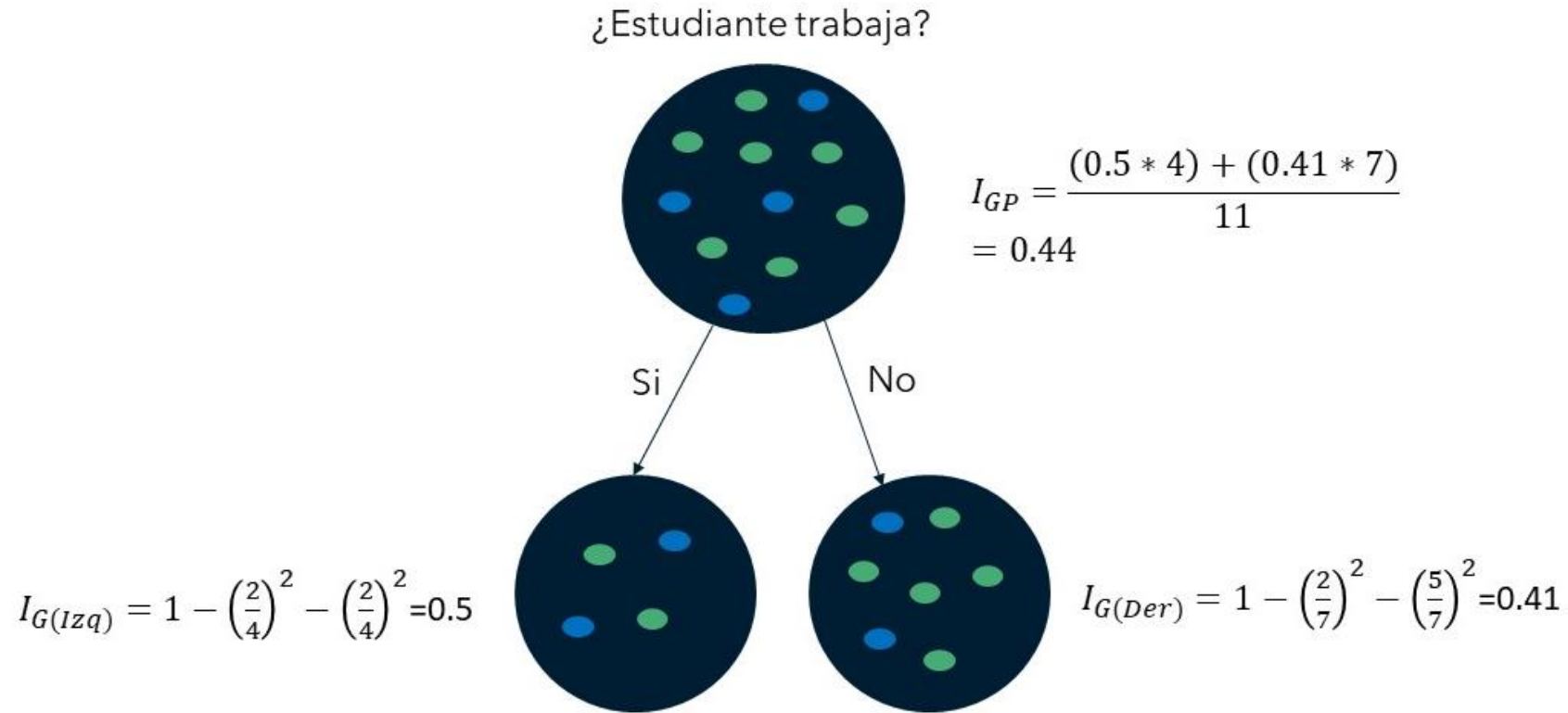
<http://github.com/scasta31/ST0245-003/proyecto/>



Diseño del Algoritmo



Algoritmo para construir un árbol binario de decisión usando C4.5. En este ejemplo, mostramos un modelo para predecir si uno debe o no adquirir un material específico en una compañía, dependiendo de la calidad del material, precio y relación con el proveedor.



Esta división está basada en la condición “Estudiante trabaja?”. Para este caso la impureza de Gini del nodo de la izquierda es de 0.5 y para el nodo de la derecha es 0.41. Finalmente, la impureza de Gini ponderada es de 0.44

Algorithm Complexity



Keep this title

Complete this slide
For the third deliverable



Create the table in Powerpoint. Do not
copy pixelated screenshots from the
technical report please!

	Time Complexity	Memory Complexity
Training the model	$O(N^2 * M * 2^M)$	$O(N * M * 2^M)$
Testing the Model	$O(N * M)$	$O(1)$

Time and memory complexity of the (In this semester, one could be CART, ID3, C4.5... please choose) algorithm. (Please explain what do N and M mean in this problem. PLEASE DO IT!)



Explain the tables in your
own words



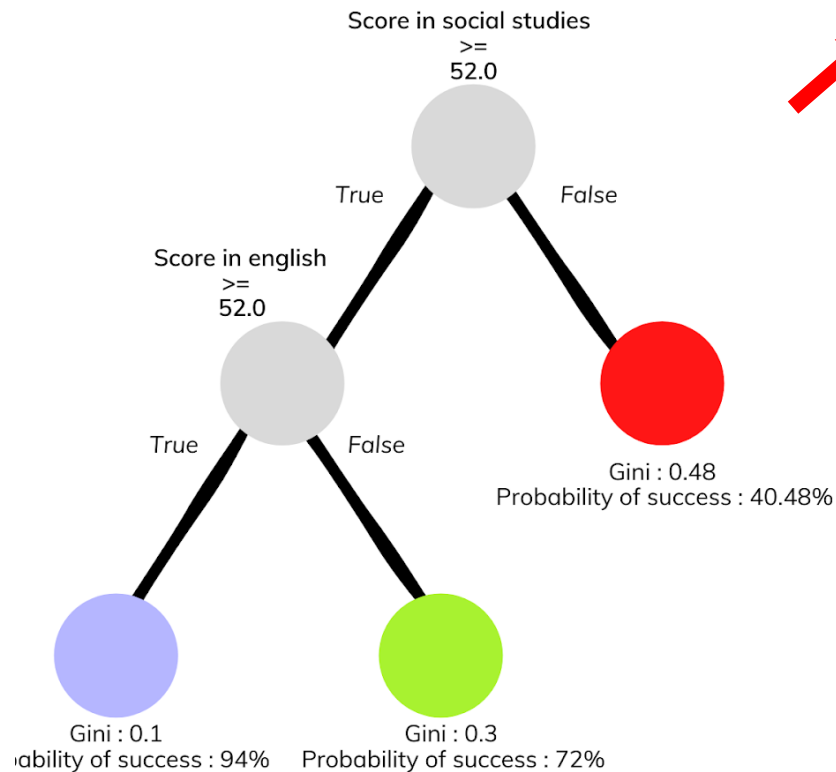
Include another HD picture related
to the example that you modeled
in the decision tree

Decision-Tree Model



Keep this title

Complete this slide
For the third deliverable



A binary decision tree to predict Saber Pro scores based on the results of Saber 11. Violet nodes represent those with a high probability of success, green medium probability and red a low probability of success.



Explain the Figures in your own words

Create the Figure in Powerpoint. Do not copy pixelated screenshots from the technical report please!



Most Relevant Features



Social Studies



English



Gender



Use an icon for each feature!



Is it ethical to make a model that predicts academic success based on gender?

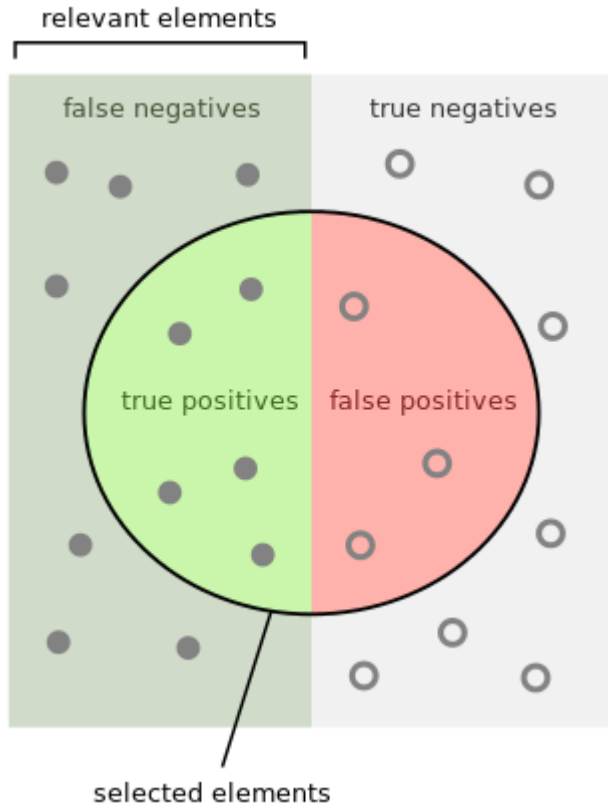
Evaluation Metrics

Keep this title

Complete this slide
For the third deliverable



Use these
Colors for
Your figures



How many selected
items are relevant?

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

How many relevant
items are selected?

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

Explain Accuracy too...
In the same manner

If possible, avoid equations for
simple concepts that can be
explained through diagrams

Evaluation Metrics



Keep this title

*Complete this slide
For the third deliverable*



*Create the table in Powerpoint. Do not
copy pixelated screenshots from the
technical report please!*

	Training data set	Testing data set
Accuracy	0.8	0.62
Precision	0.6	0.55
Recall	0.76	0.61

Evaluation metrics using a training dataset of 135,000 students and test dataset of 45,000 students.



*Explain the tables in your
own words*



*Include another HD picture related
to the example that you modeled
in the decision tree*

Time and Memory Consumption

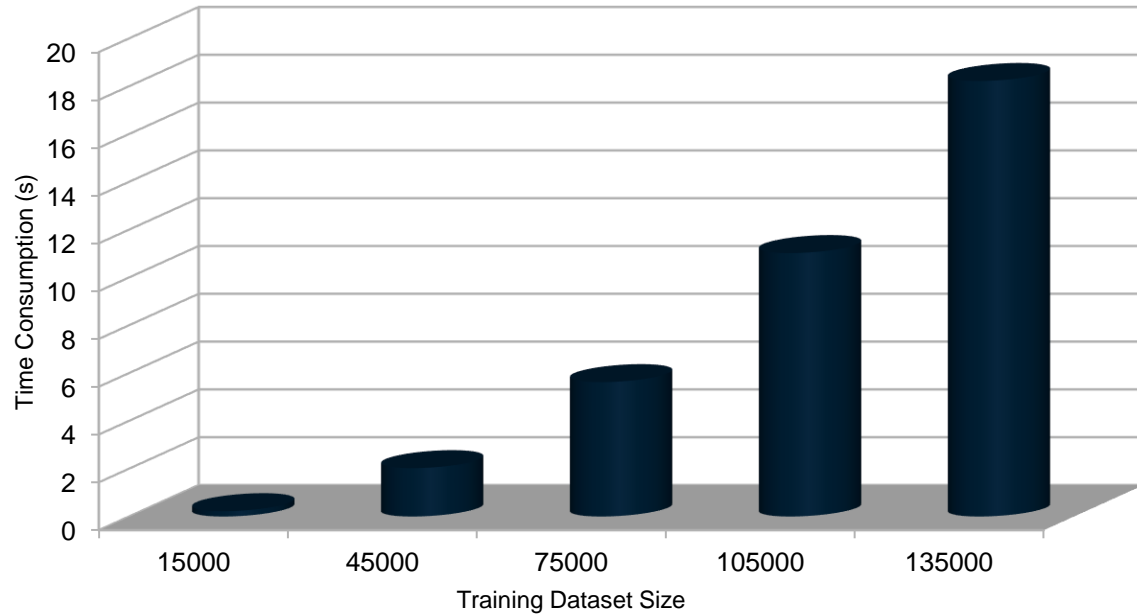


Keep this title

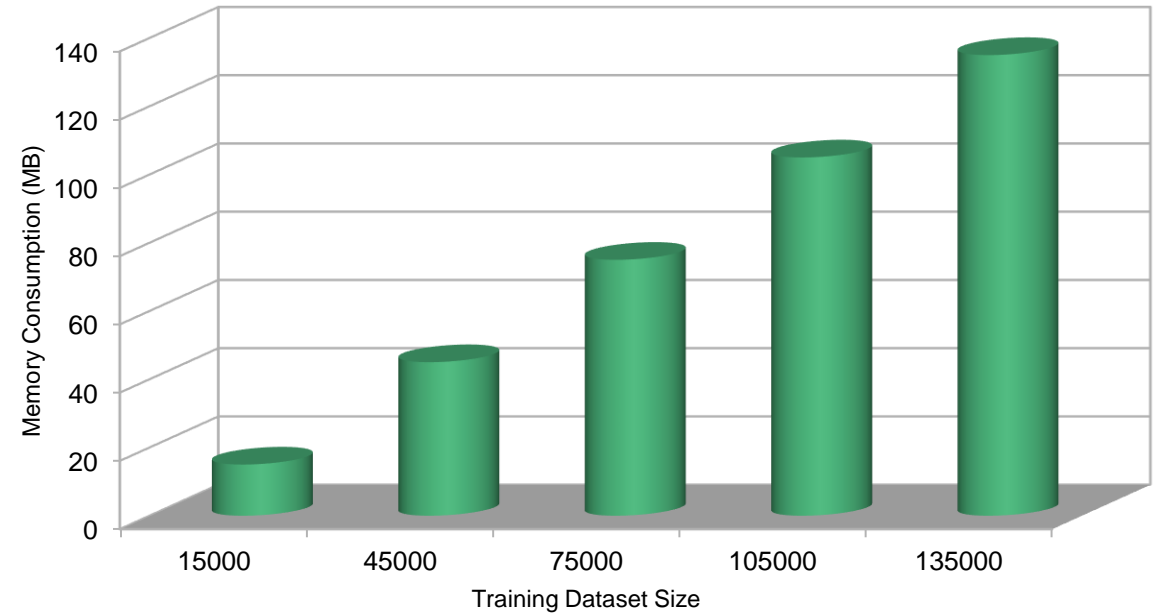
Complete this slide
For the third deliverable



Create the plots in Excel. Do not copy
pixelated screenshots from the technical
report please!



Time Consumption



Memory Consumption



Include the citation of the report
in arXiv and link

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>

Include a
screenshot

The screenshot shows the arXiv.org page for the paper 'Planning system for deliveries in Medellín'. The header includes the Cornell University logo and the text 'Cornell University'. Below this, the breadcrumb path 'arXiv.org > cs > arXiv:1611.04156' is displayed. The category 'Computer Science > Data Structures and Algorithms' is shown. The submission date '[Submitted on 13 Nov 2016]' is noted. The title 'Planning system for deliveries in Medellín' is prominently displayed, followed by the authors 'Catalina Patiño-Forero, Mateo Agudelo-Toro, Mauricio Toro'. The abstract text describes the implementation of an application for planning the shortest delivery route in Medellín, Colombia, comparing it to the Traveling Salesman Problem (TSP). The abstract mentions that the problem allows visiting each place more than once and is important for saving time and money in fuel. At the bottom, metadata is provided: '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)'.



Complete this slide
For the third deliverable

Say thank you for
listening!

THANK YOU!