

Decision
trees
to predict
a student's
performance
on Saber
Pro



*Complete this slide
For the first deliverable*



Kevin
Torres



Vladlen
Shatunov



Miguel
Correa



Mauricio
Toro

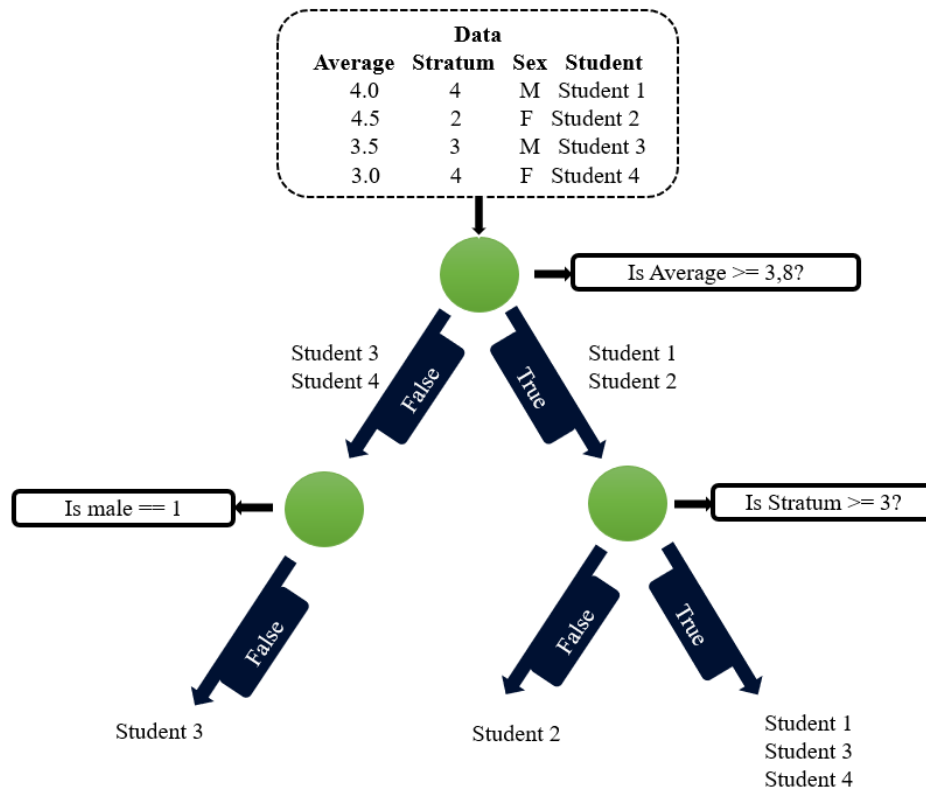


<http://github.com/>

vshatunovv

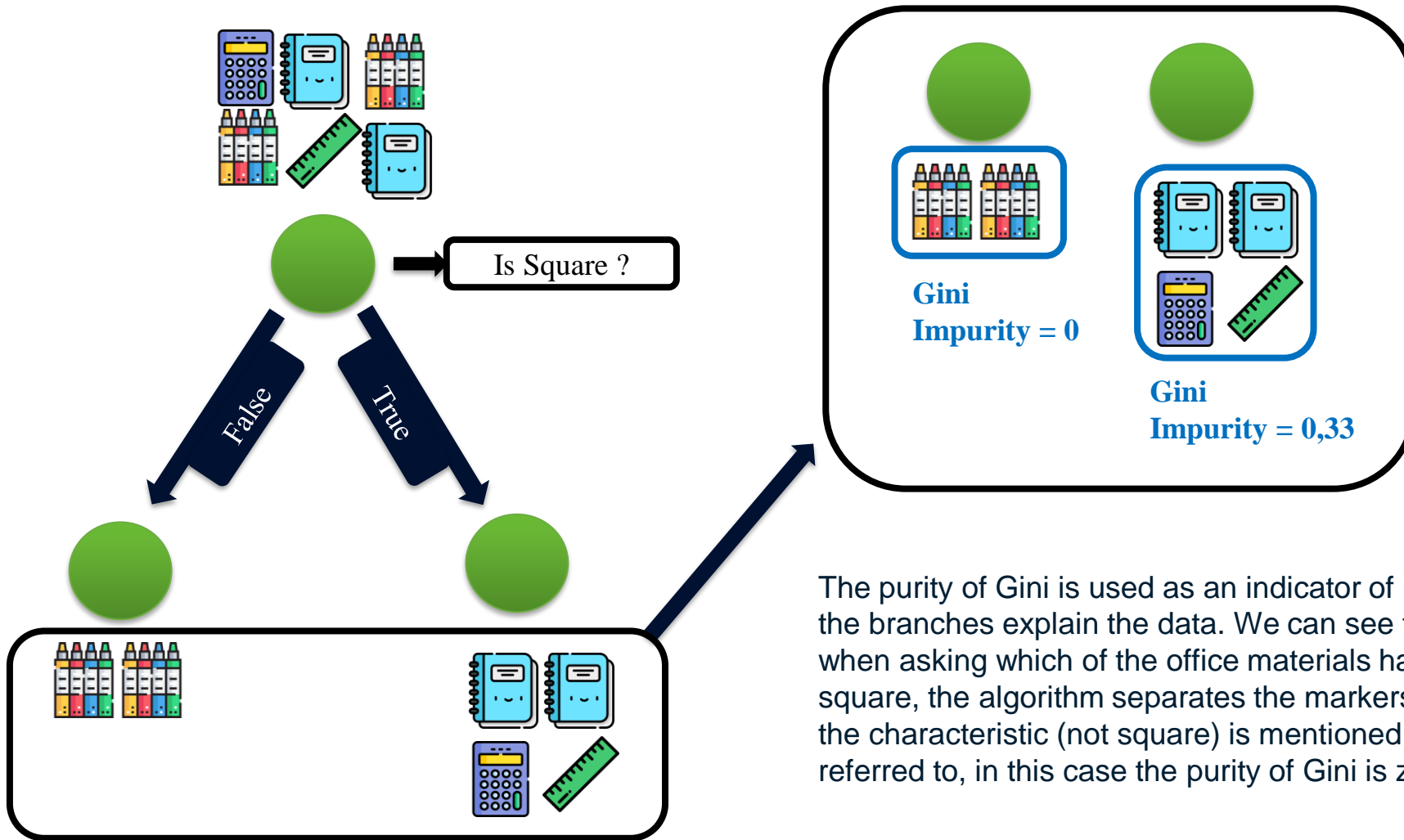
/proyecto/

Algorithm Design



The following image explains how to separate the parameters or variables that describe each copy in the CART decision tree. The following example shows a decision tree for 4 students where the parameters are socioeconomic stratum, average in high school and sex

Node Splitting



The purity of Gini is used as an indicator of how the conditions and depth of the branches explain the data. We can see that in the previous example, when asking which of the office materials have the characteristic of being square, the algorithm separates the markers from the rest of them, so when the characteristic (not square) is mentioned, the markers will always be referred to, in this case the purity of Gini is zero.

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



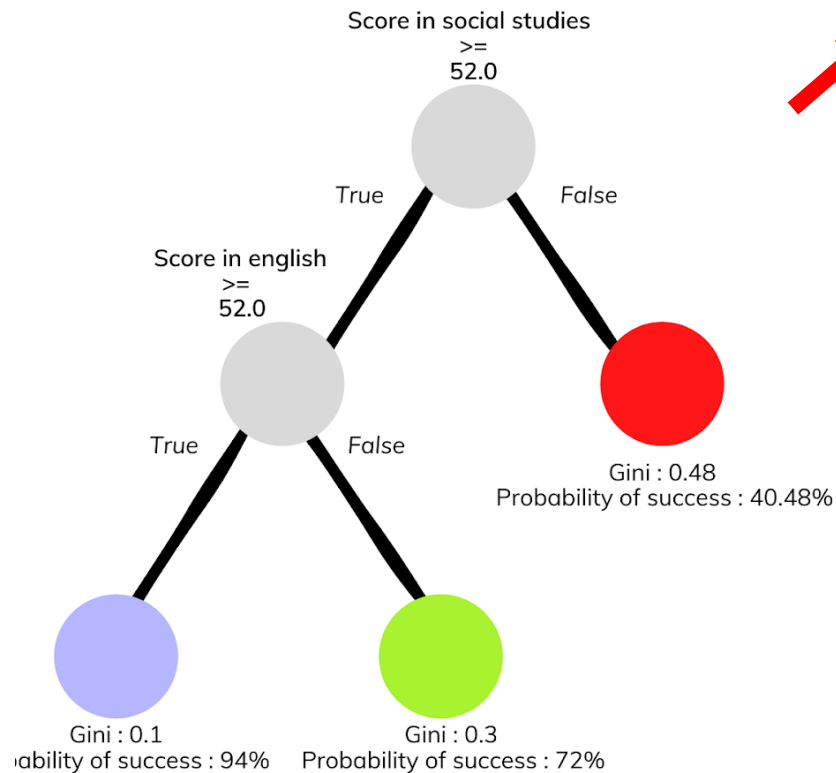
Include another HD picture related to the example that you modeled in the decision tree

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?

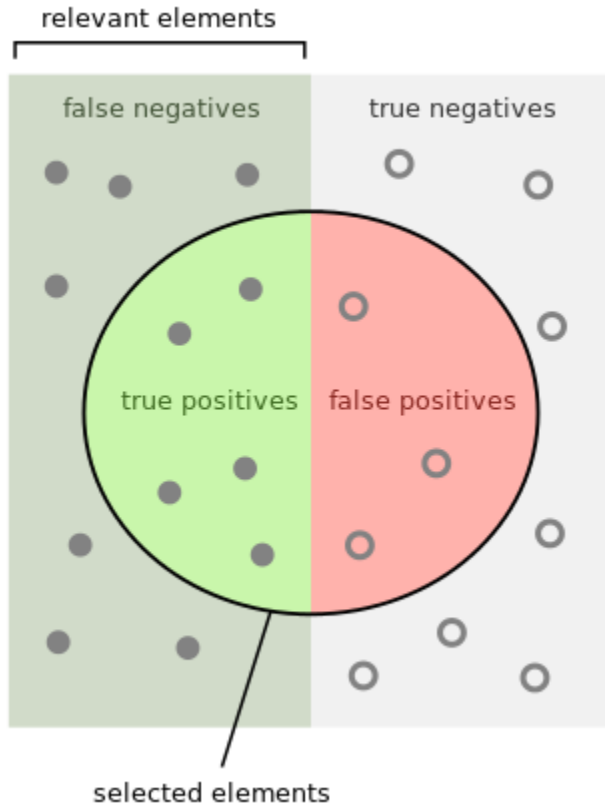
Evaluation Metrics

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

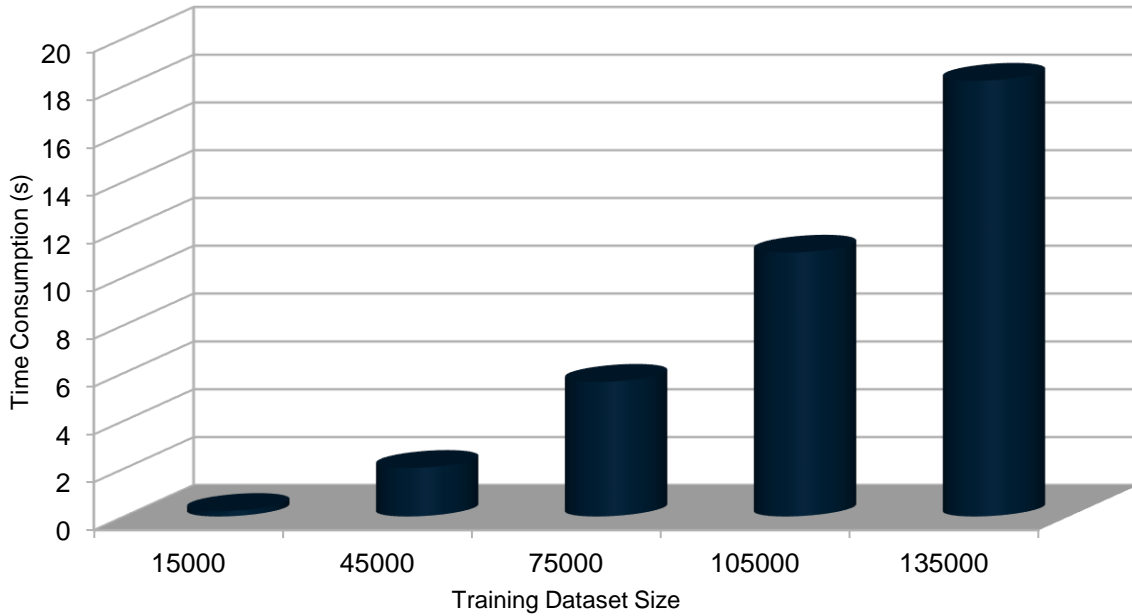


Keep this title

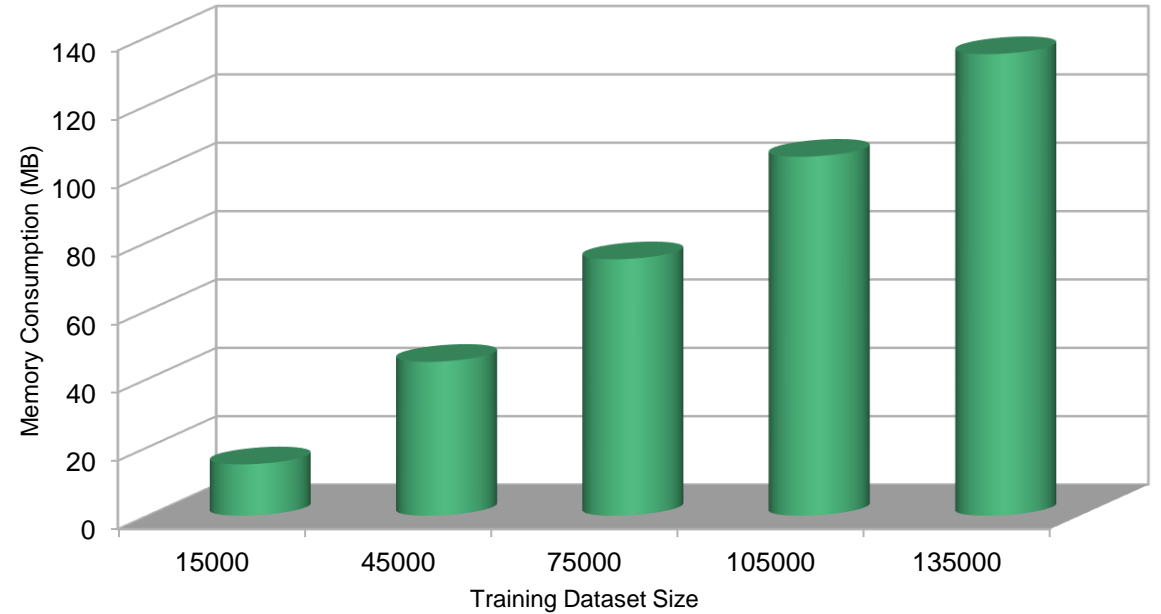
Complete this slide
For the third deliverable



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

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)



Complete this slide
For the third deliverable

Say thank you for
listening!

THANK YOU!