

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

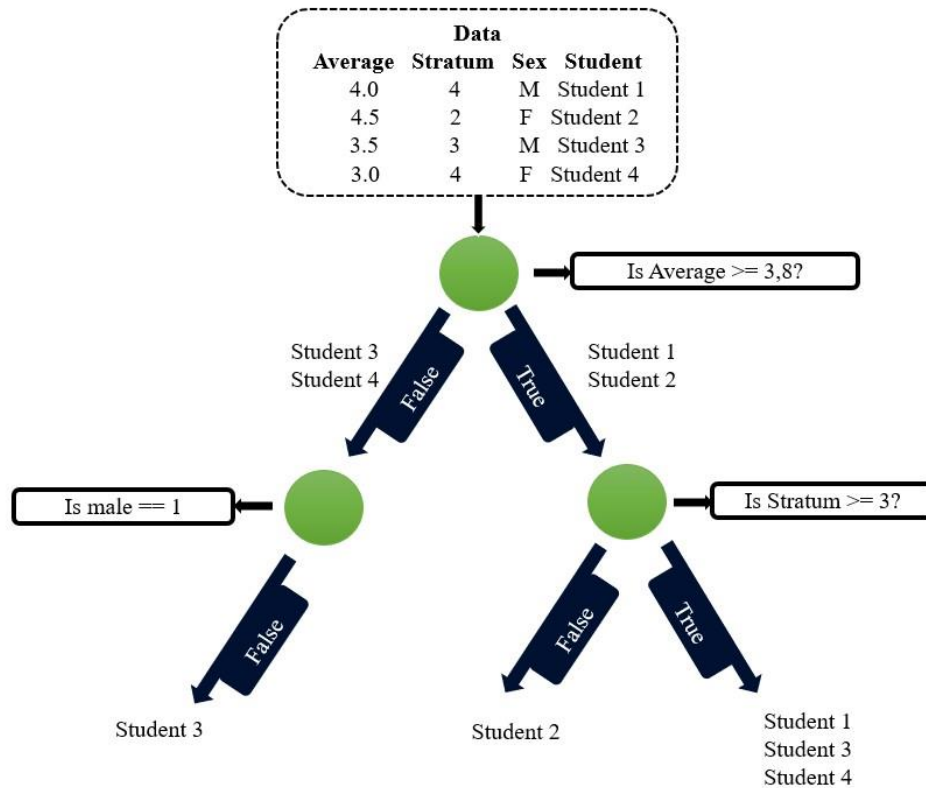


<http://github.com/>

vshatunovv

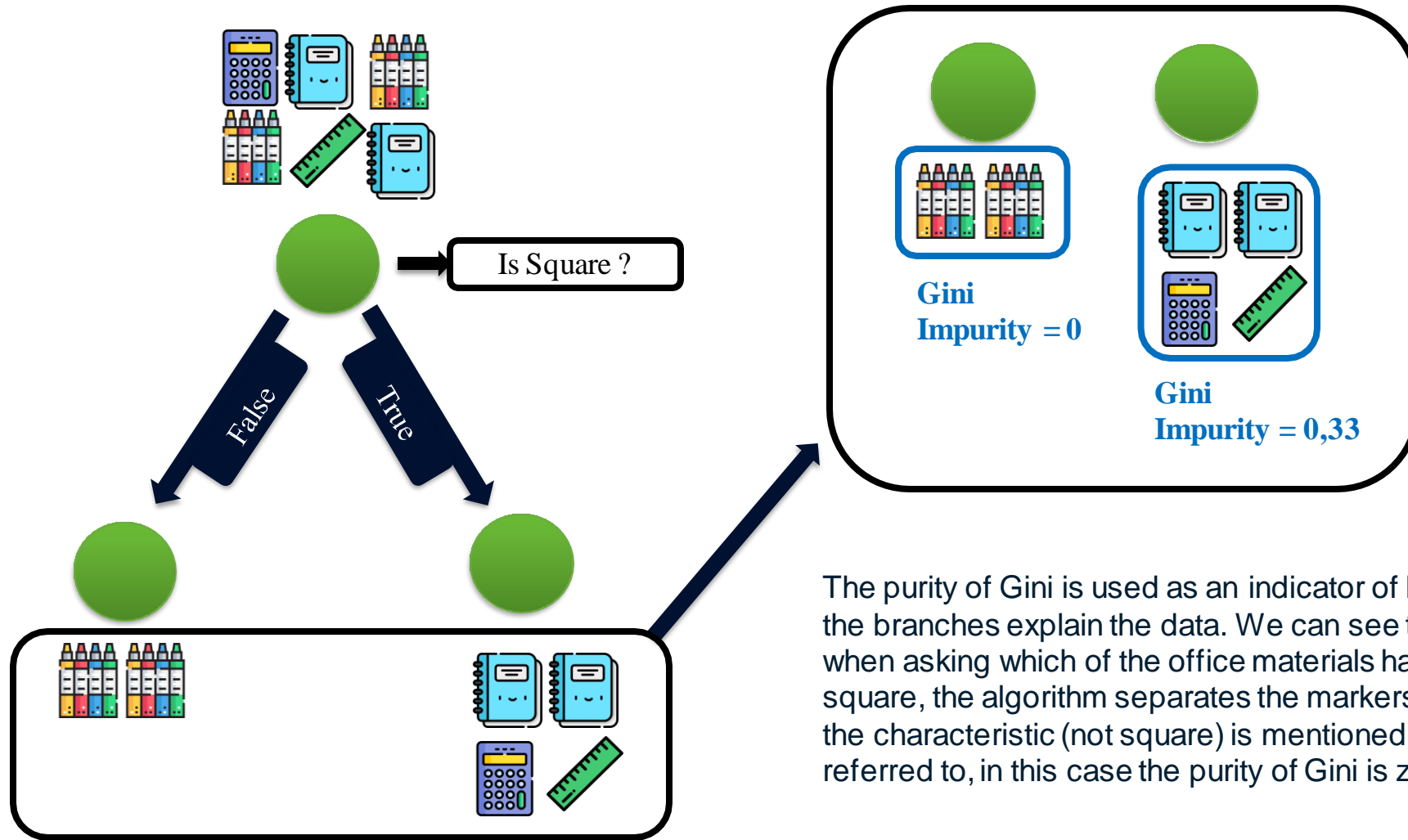
/proyecto/

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.

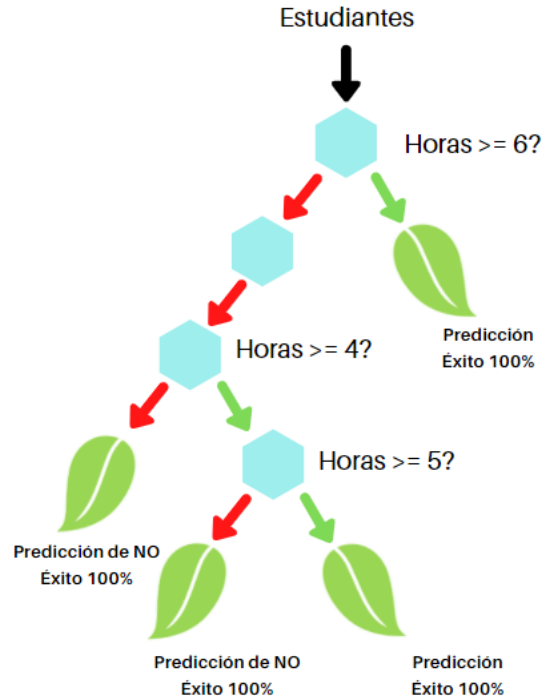
Algorithm Complexity



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

Time and memory complexity of the CART algorithm.
N means the number of operations.
M means the number of variables





Most Relevant Features

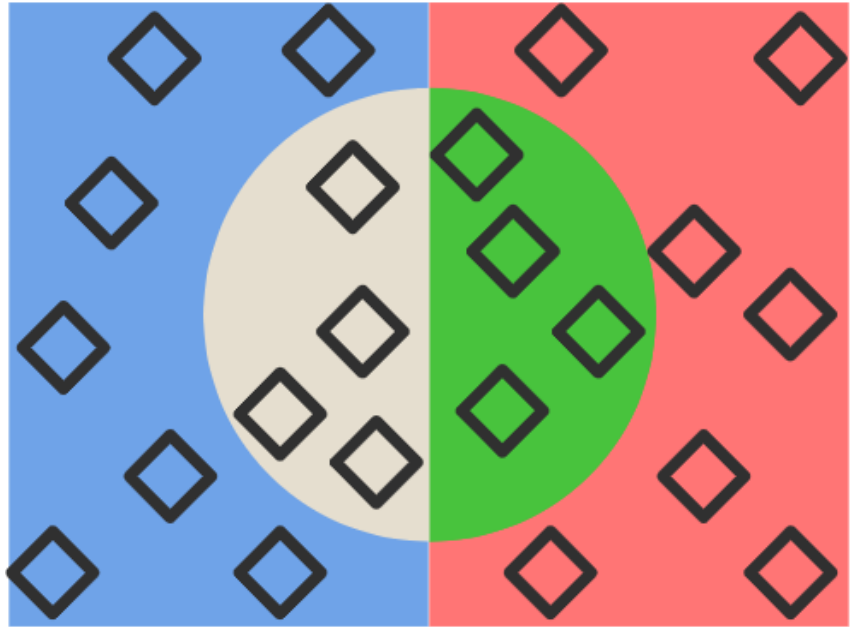
Social Studies

English

Gender

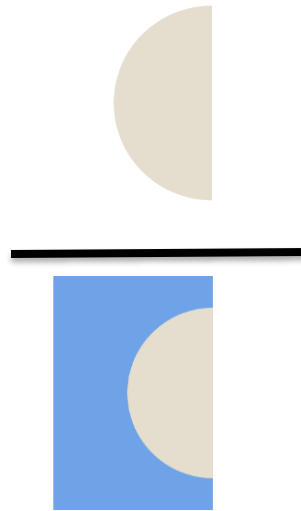
A binary decision tree to predict Saber Pro scores based on the results of Saber 11. Violet nodes represent those with a high probability of success, green medium probability and red a low probability of success.

Evaluation Metrics

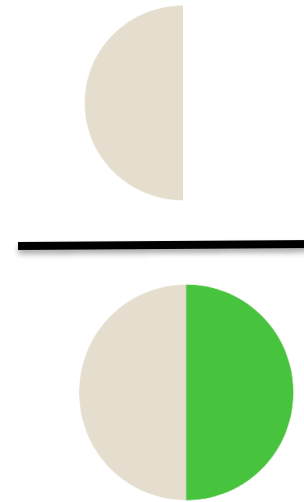


● True Positives ● False Positives

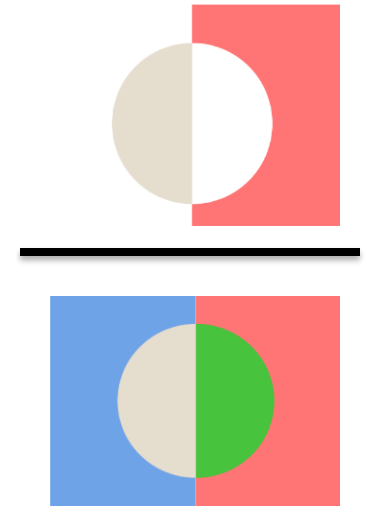
■ False Negatives ■ True Negatives



Recall



Precision



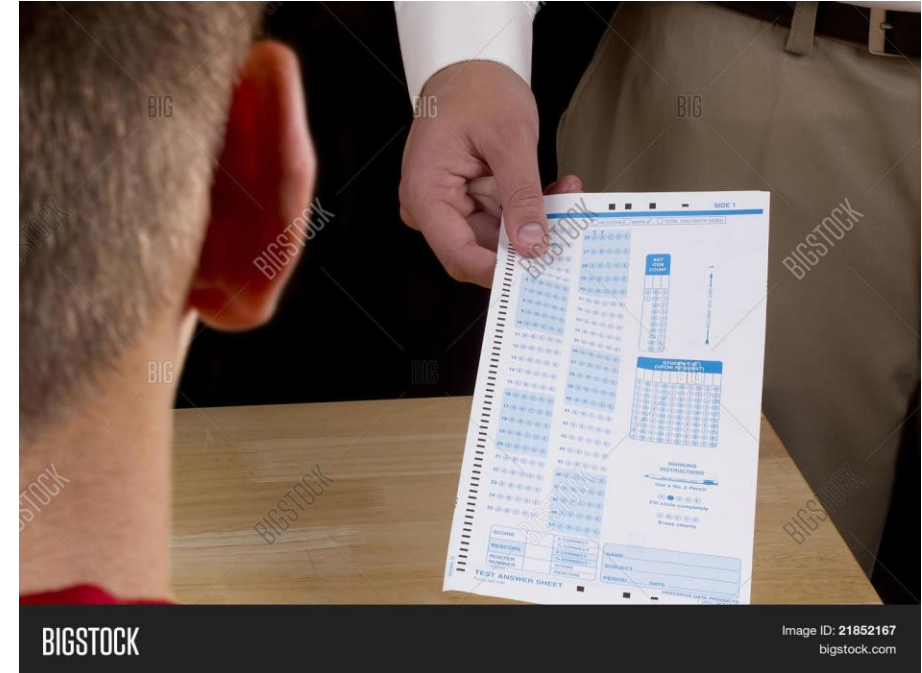
Accuracy

Evaluation Metrics

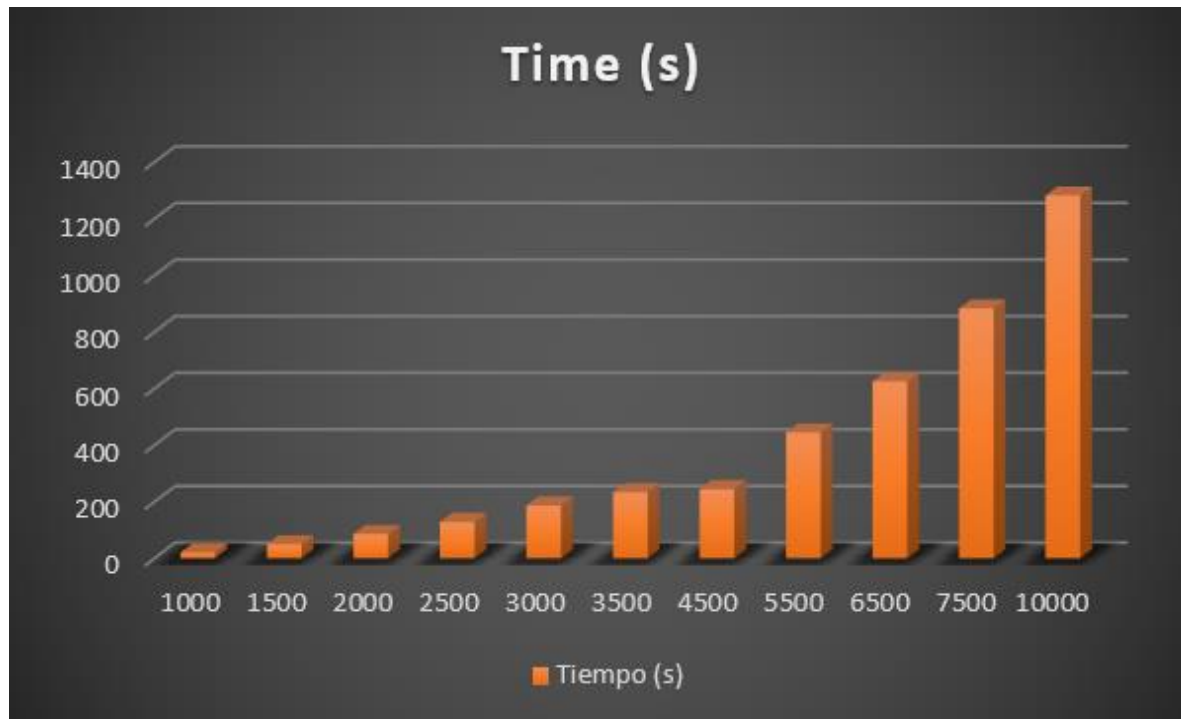


	Training data set	Testing data set
Accuracy	96,29%	97,56%
Precision	92,6%	95,13%
Recall	100%	100%

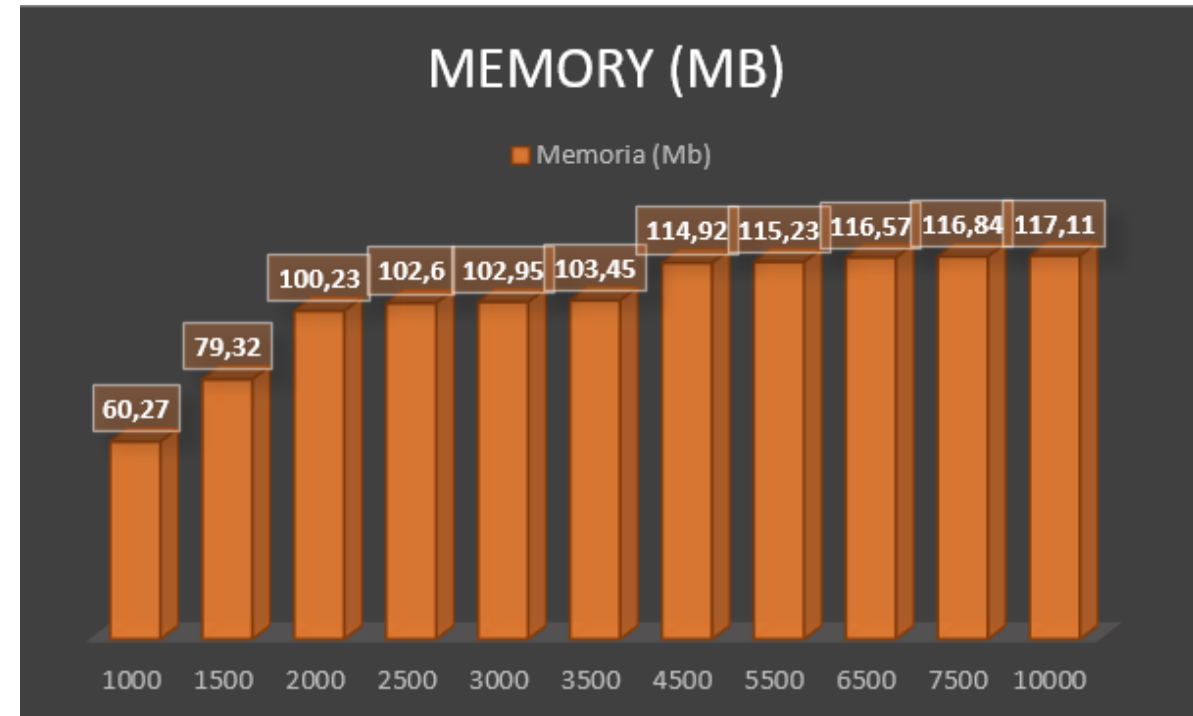
Evaluation metrics using a training dataset of 135,000 students and test dataset of 45,000 students. We can see that this model of our is very precise and has little to no errors when were talking of big amounts of data



Time and Memory Consumption



Time Consumption



Memory Consumption

Report Accepted on arXiv

Status: Scheduled to be announced

Your arXiv.org account

arxiv.org/user/

Cornell University

arXiv.org > user

Logged in as vshatunov | Help | My Account | Logout

Article submitted

Your arXiv.org account: vshatunov

E-mail: vshatunov@eafit.edu.co

Name: Vladlen Shatunov

Default Category: cs.CL

Groups: cs

Affiliation: Eafit

URL:

Country: Colombia

Career Status: Grad Student

Change User Information

Change Password

Change Email

Disable MathJax (What is MathJax?)

Article Submissions

Update

Delete

Unsubmit

Identifier	Type	Title	Status	Actions	Expires
submit/3478032	New	Predicting Saber Pro Exams results based on Saber 11 results	processing		

START NEW SUBMISSION

Are there articles you are an author of on arXiv that are not listed here?

If you have the paper password use the Claim Ownership with a password form

If you do not have the paper password or are claiming multiple papers use the Claim Authorship form.

For more information please see our help page on authority records.

arXiv now processes new submissions and replacements with TeX Live 2020. Learn More.

UNIVERSIDAD
EAFIT®



THANK YOU!