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

UNIVERSIDAD EAFIT

Team Presentation

Complete this slide
For the first deliverable

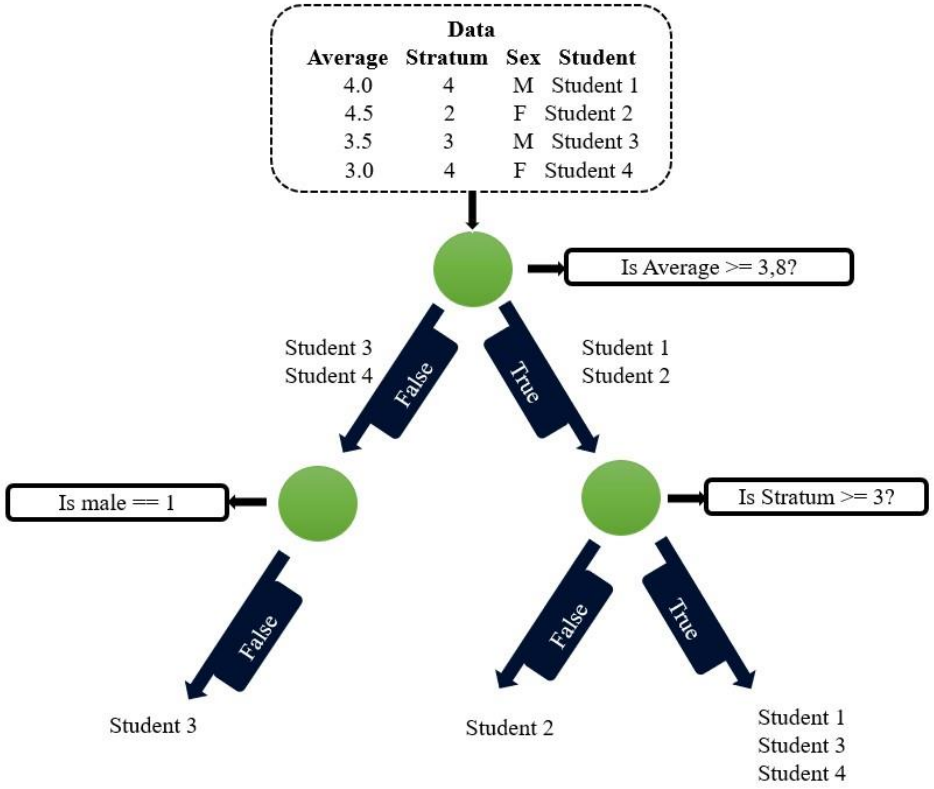
Kevin Torres

Vladlen Shatunov

Miguel Correa

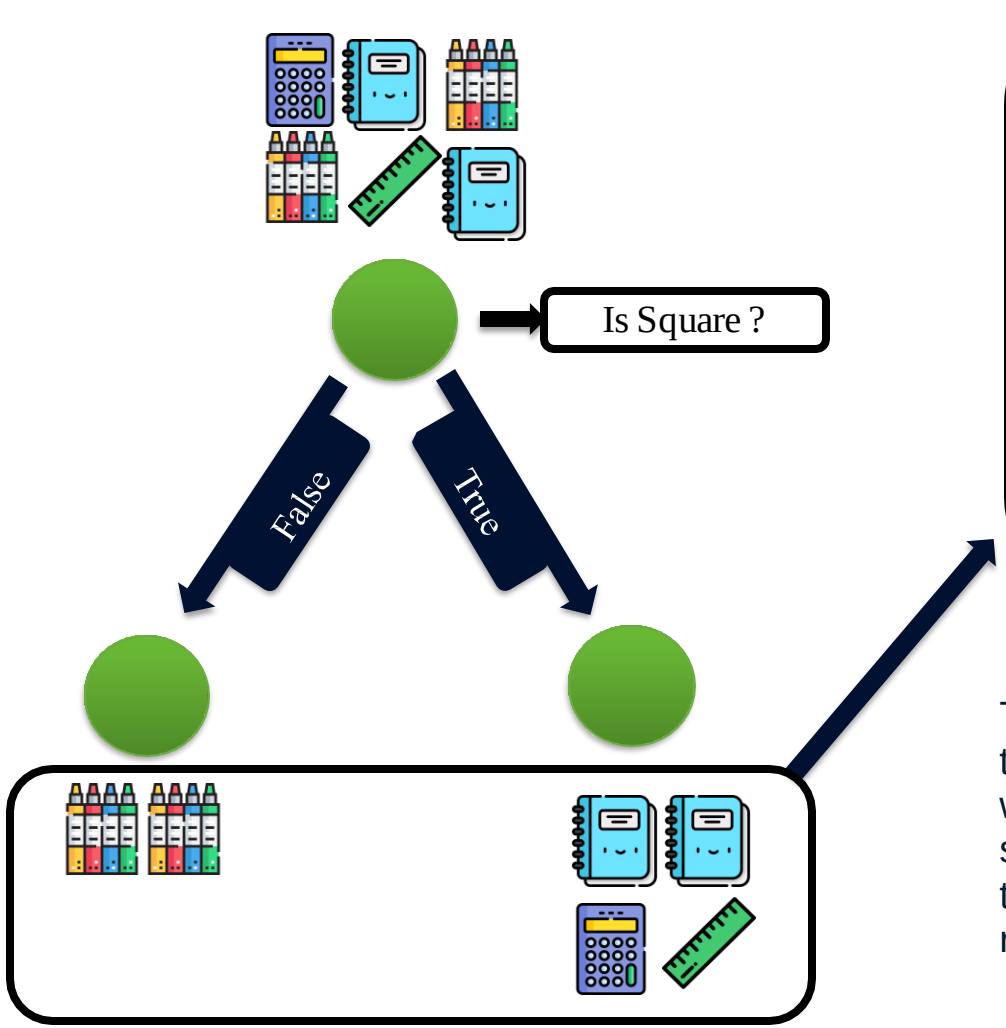
Mauricio Toro

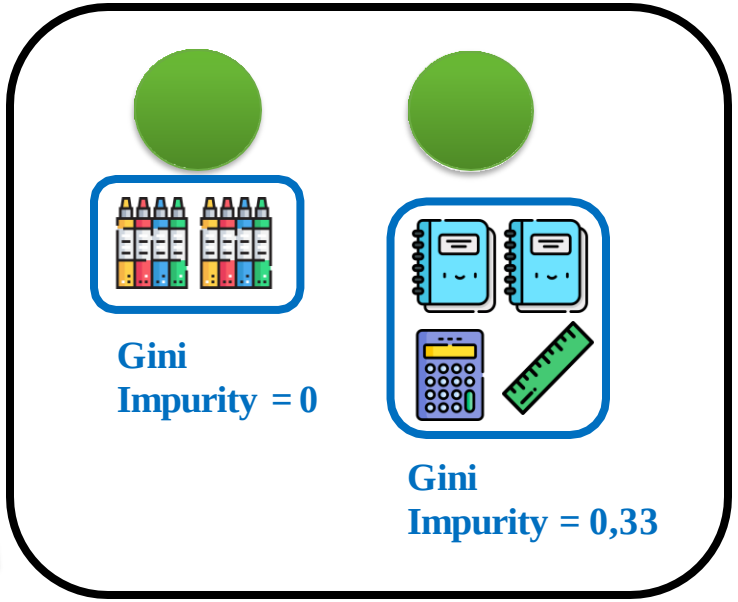
Algorithm Design



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





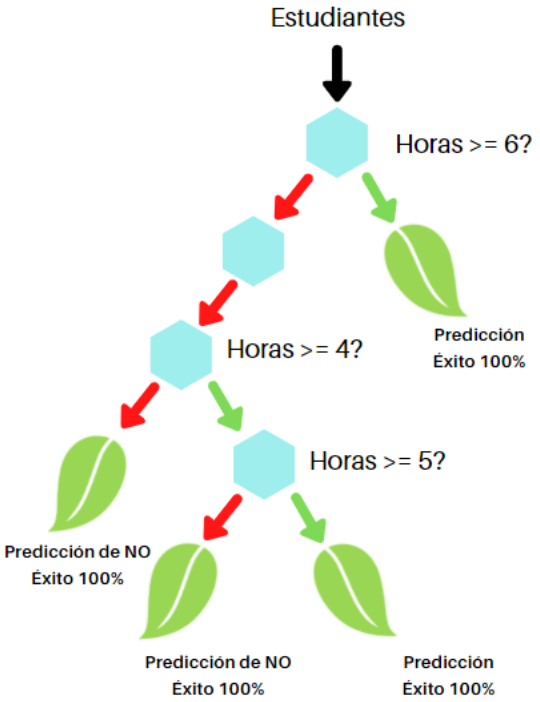
The purity of Gini is used as an indicator of how the conditions and depth of the branches explain the data. We can see that in the previous example, when asking which of the office materials have the characteristic of being square, the algorithm separates the markers from the rest of them, so when the 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

Decision-Tree Model



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.

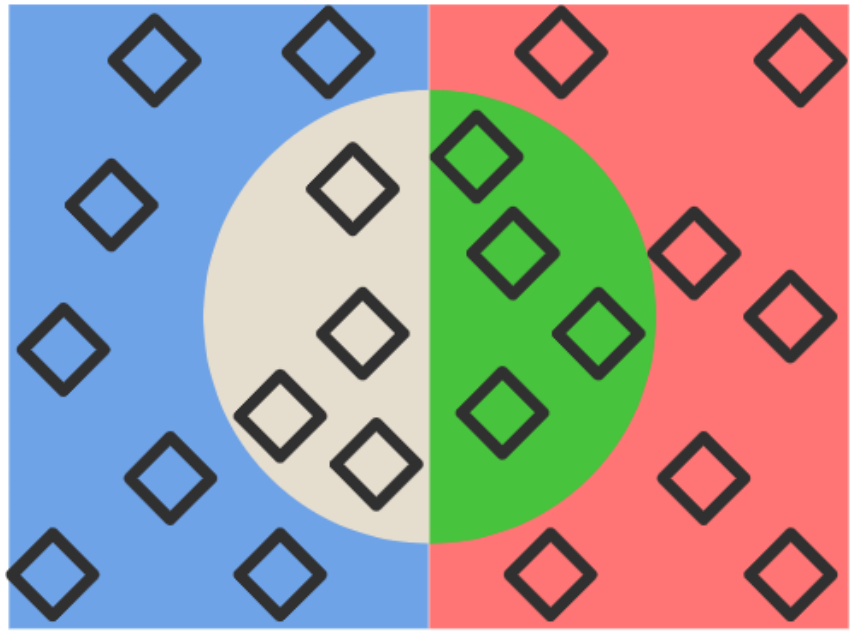
Most Relevant Features

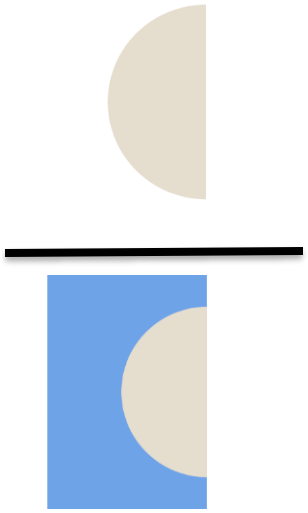
Social Studies

English

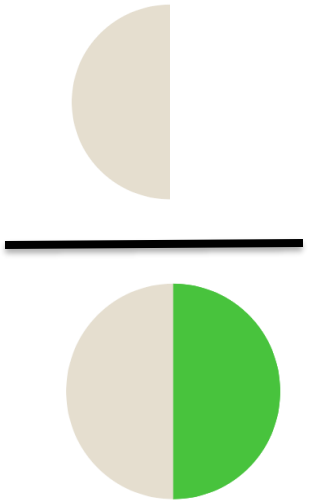
Gender

Evaluation Metrics

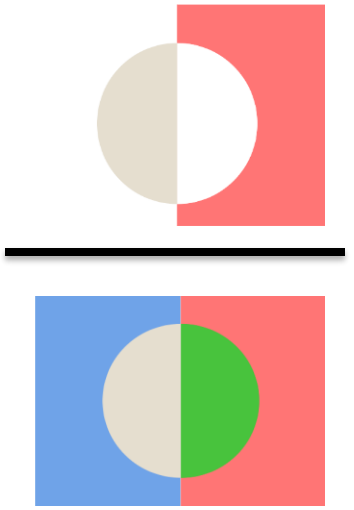




Recall



Presicion

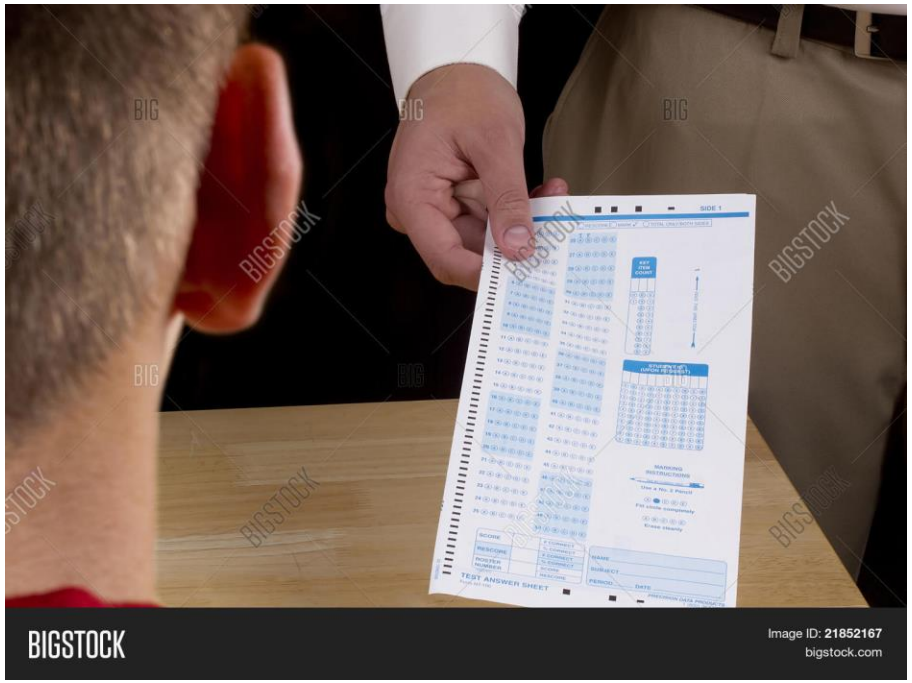


Accuracy

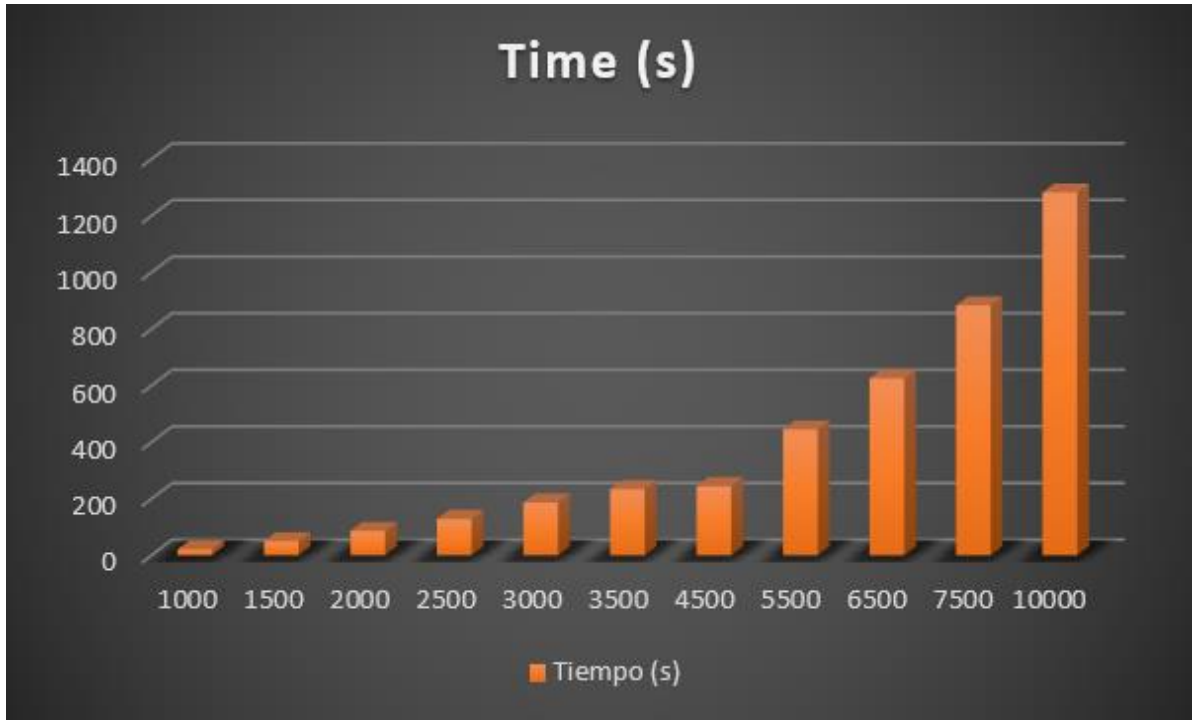
Evaluation Metrics

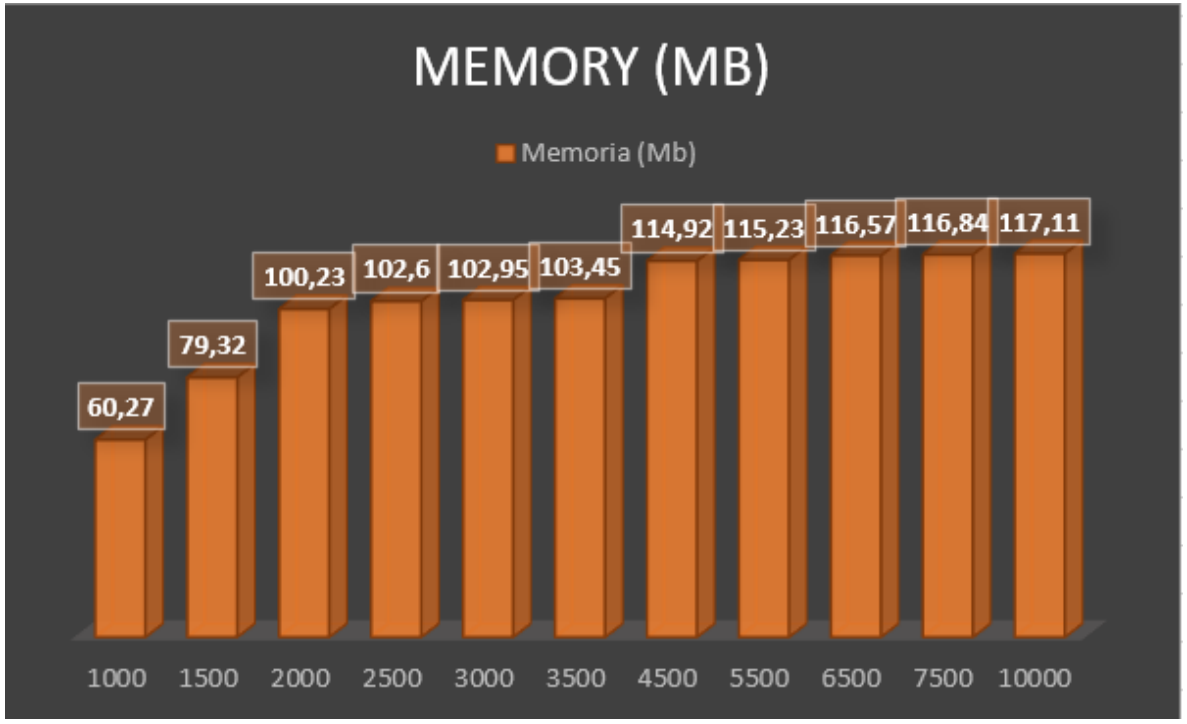
	Training data set	Testing data set
Accuracy	96,29%	97,56%
Precision	92,6%	95,13%
Recall	100%	100%

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





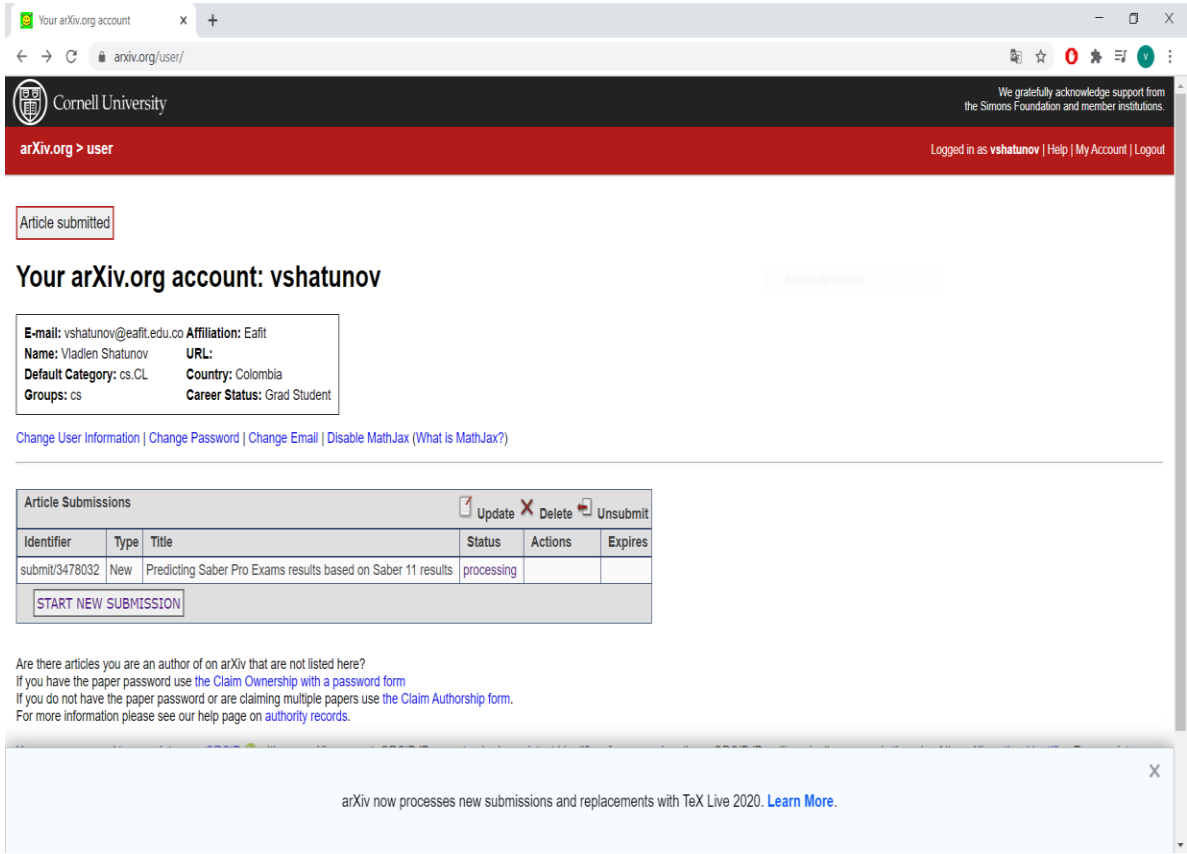
Time Consumption

Memory Consumption

UNIVERSIDAD EAFIT

Report Accepted on arXiv

Status: Scheduled to be announced



UNIVERSIDAD EAFIT


THANK YOU!