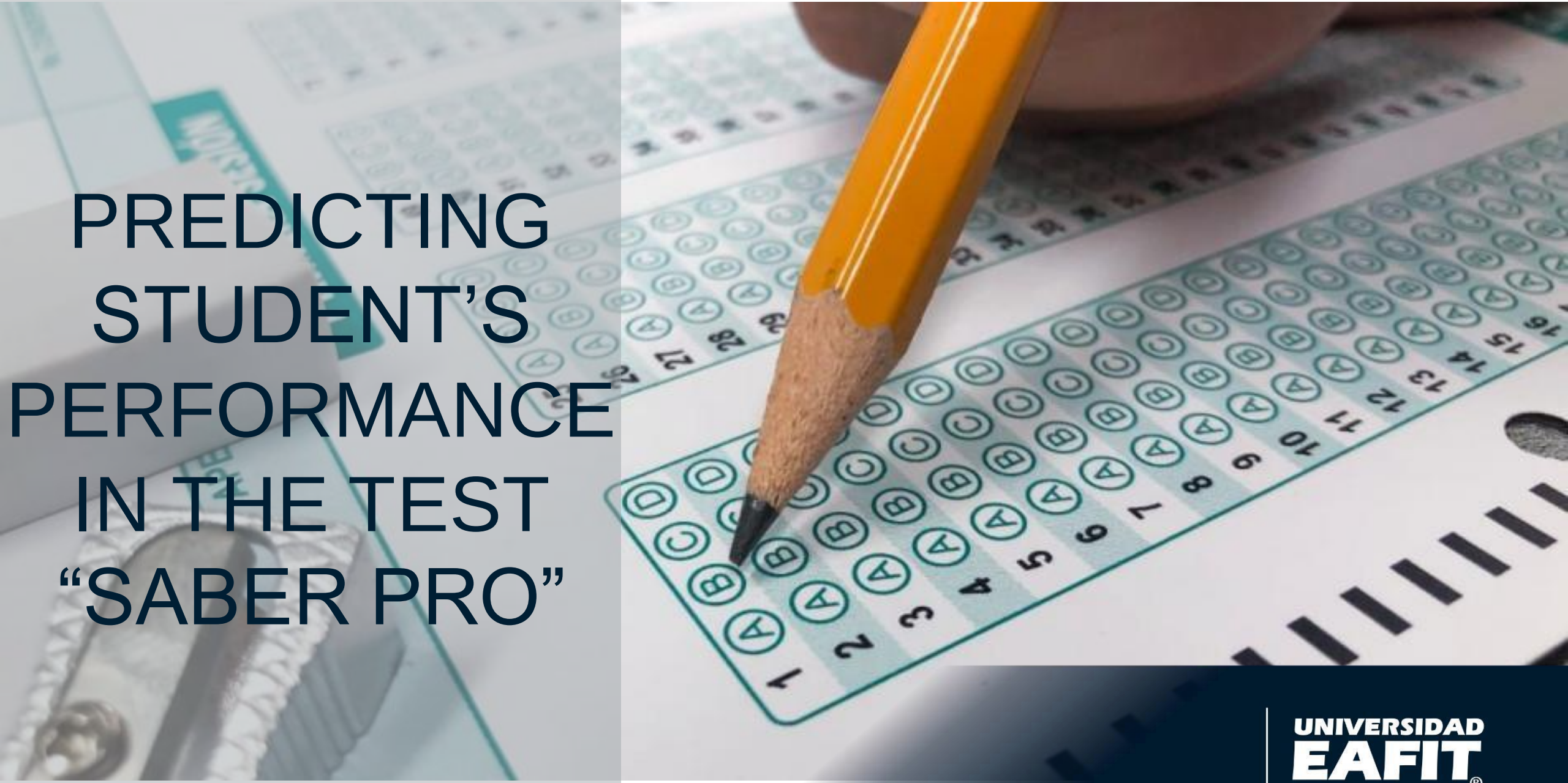
PREDICTING STUDENT'S PERFORMANCE IN THE TEST "SABER PRO"



UNIVERSIDAD EAFIT

Team Presentation

Juan Pablo Rincon

Julián Gómez

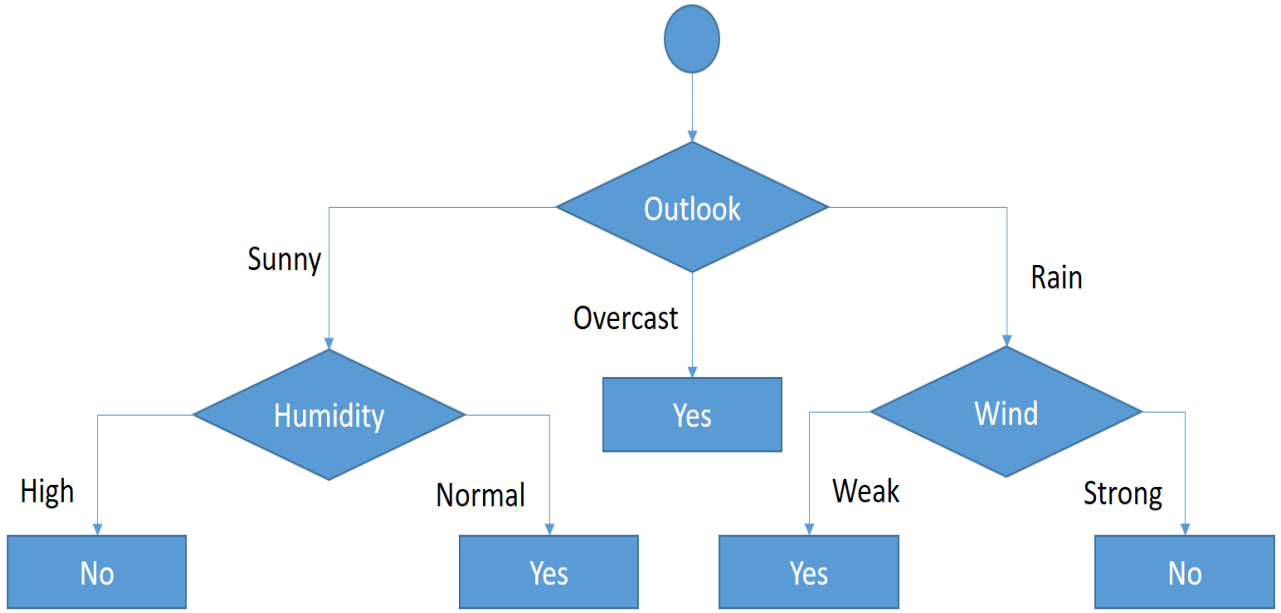
Miguel Correa

Mauricio Toro

github.com/jgomezb11/ST0245-002/proyecto

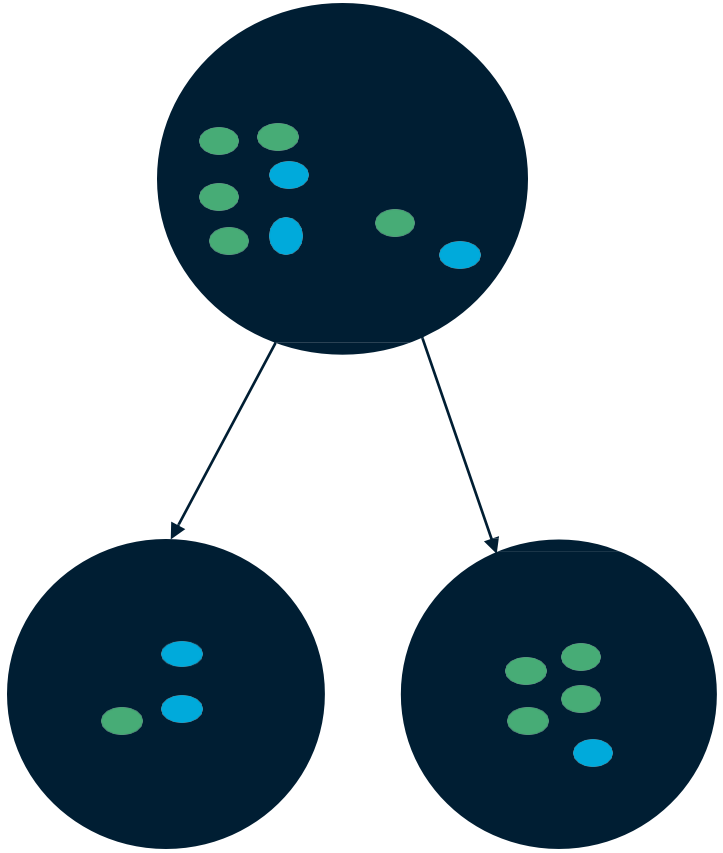
UNIVERSIDAD EAFIT

Algorithm Design

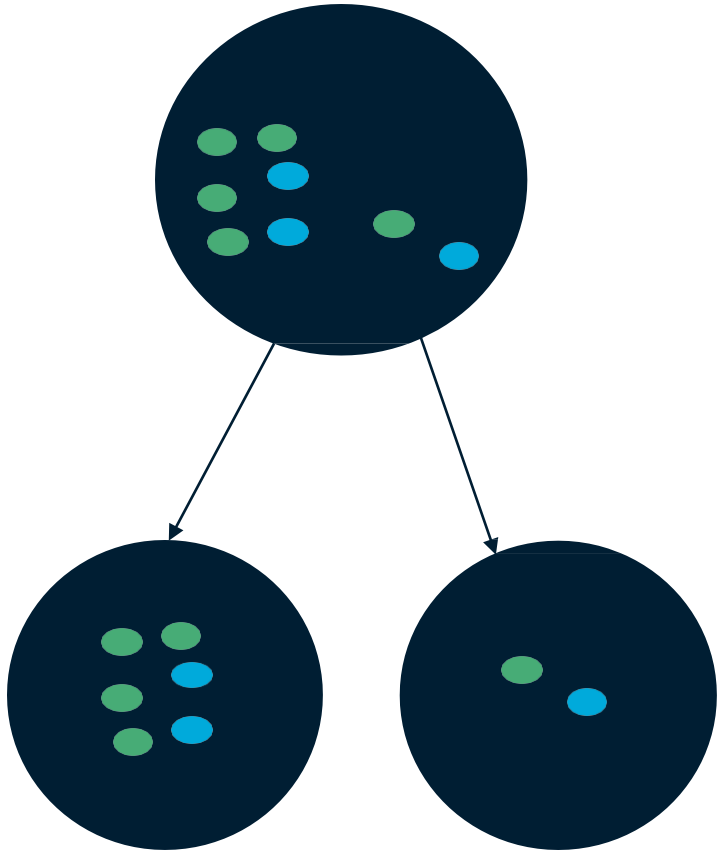


Algorithm to build a binary decision tree using *CART*. In this example, we show a model to predict whether or not to go shopping, according to weather.

Node Splitting



As an example, this split is based on the condition "desemp_ingles == A-."
For this case weigthed Gini impurity is 0.38.



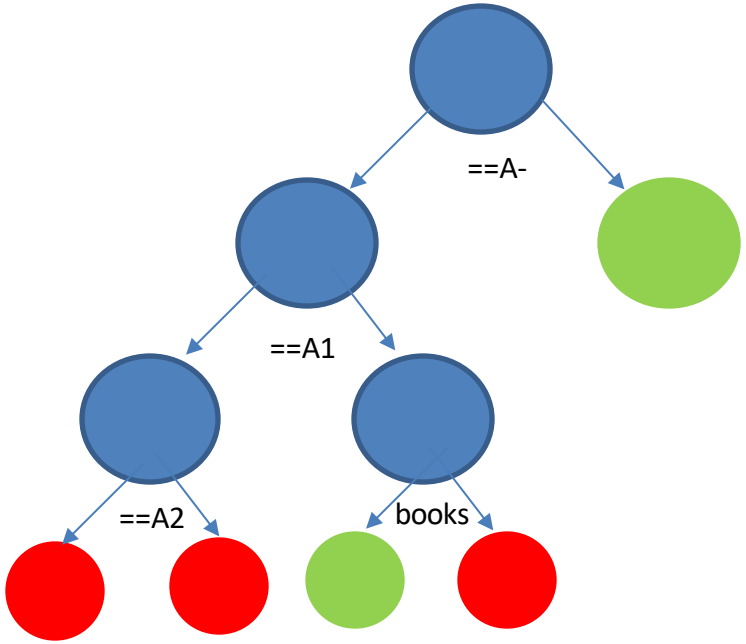
As an example, this split is based on the condition "Fam_numlibros == "0 A 10 LIBROS"."
For this case weigthed Gini impurity is 0.48.

Algorithm Complexity

	Time Complexity	Memory Complexity
Training the model	O(N^3*M*log(N))	O(N*M)
Testing the Model	O(N^3*M*log(N))	O(1)

Time and memory complexity of the CART algorithm. Where N is the number of rows and M is the number of columns in the dataset

Decision-Tree Model



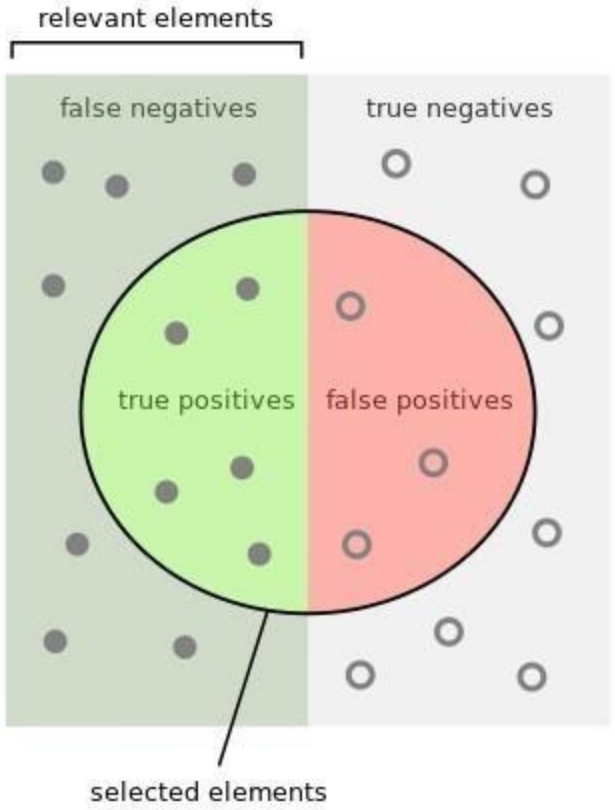
A binary decision tree to predict Saber Pro scores based on the results of Saber 11. Green nodes represent those with a high probability of success, blue non-leafs nodes and red a low probability of success.

Most Relevant Features

EN English

Number of books

Evaluation Metrics



How many selected items are relevant?

$$Precision = \frac{\text{true positives}}{\text{selected elements}}$$

How many relevant items are selected?

$$Recall = \frac{\text{true positives}}{\text{relevant elements}}$$

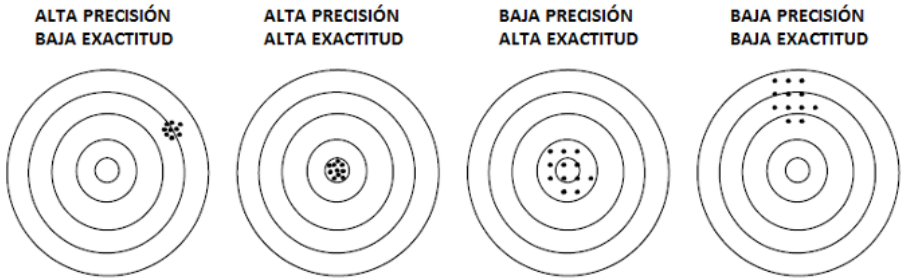
		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

$$Accuracy = \frac{TrueNegatives + TruePositive}{TruePositive + FalsePositive + TrueNegative + FalseNegative}$$

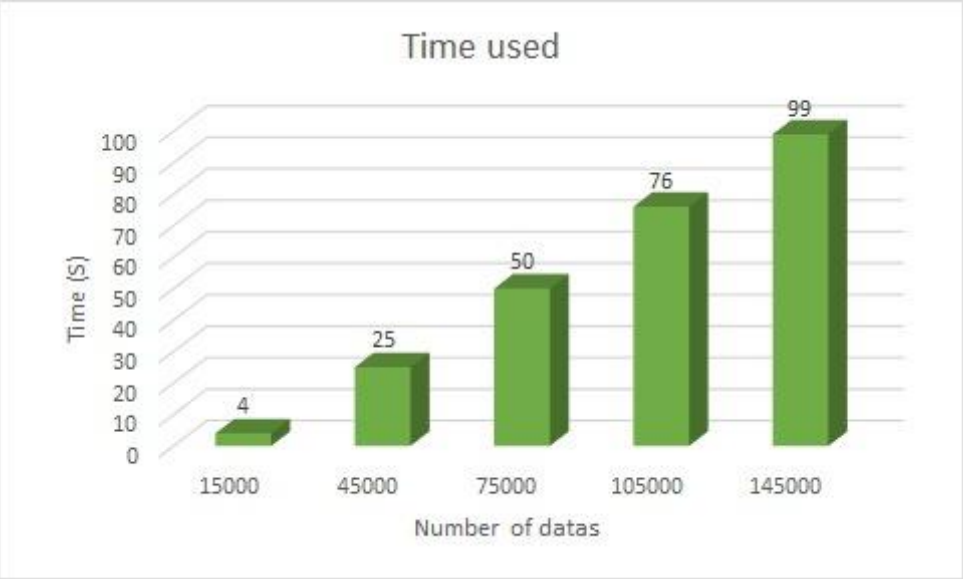
Evaluation Metrics

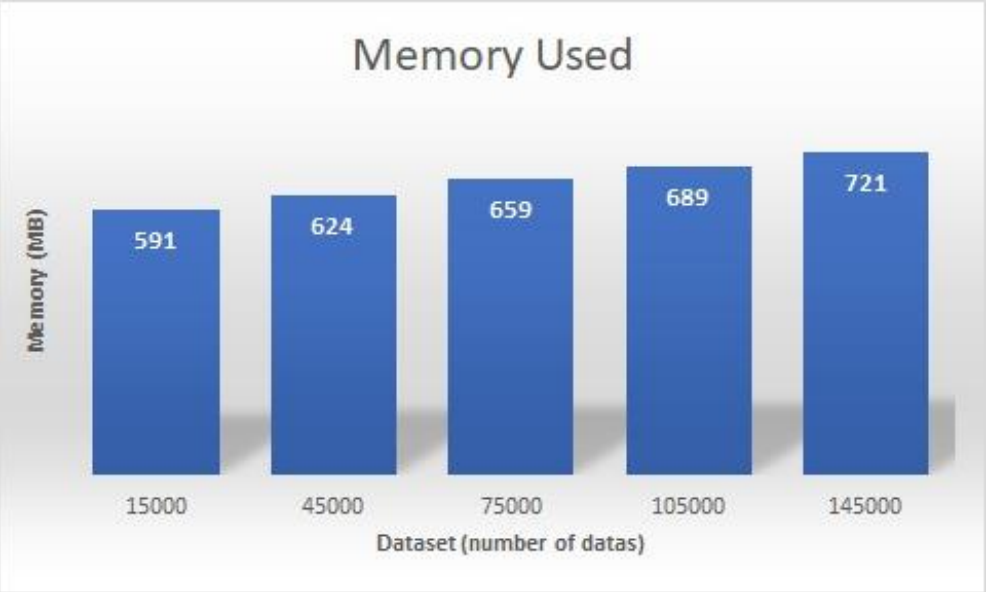
	Training data set	Testing data set
Accuracy	0.71	0.70
Precision	0.8	0.8
Recall	0.46	0.46

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



Time and Memory Consumption





Time Consumption

Memory Consumption

Report Accepted on arXiv

Preview

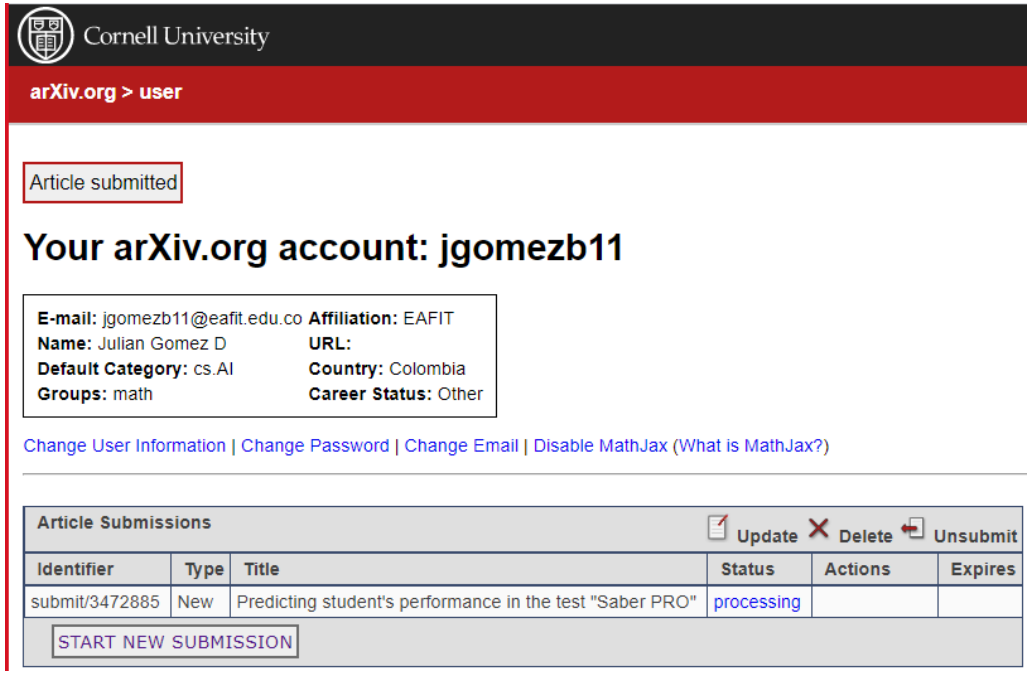
Predicting student's performance in the test "Saber PRO"

Juan Pablo Rincon Usma and Julian Gomez Benitez

Today technology gives us a proportion of data that we can use to predict results about everyday things that happen around us, we want to focus on predicting academic success through academic and sociodemographic variables, the idea is to analyze these data and predict whether students will be able to do well on the "Saber pro" tests. Similar studies have been carried out which allow us to determine how the students will do, and they have given good results, taking this into account, help can be offered to students who have difficulties and thus avoid that their grades decrease considerably.

Comments: 6 pages, 6figures

J. Rincon Usma and J. Gomez Benitez. Predicting student's performance in the test "Saber Pro" . ArXiv e-prints, Nov. 2016. Available at: To be defined.



Cornell University

arXiv.org > user

Article submitted

Your arXiv.org account: jgomezb11

E-mail: jgomezb11@eafit.edu.co **Affiliation:** EAFIT
Name: Julian Gomez D **URL:**
Default Category: cs.AI **Country:** Colombia
Groups: math **Career Status:** Other

Change User Information | Change Password | Change Email | Disable MathJax (What is MathJax?)

Article Submissions Update Delete Unsubmit

Identifier	Type	Title	Status	Actions	Expires
submit/3472885	New	Predicting student's performance in the test "Saber PRO"	processing		

START NEW SUBMISSION

UNIVERSIDAD EAFIT

THANK YOU!