PREDICTING ACADEMIC SUCCESS USING DECISION TREES

UNIVERSIDAD EAFIT

Team Presentation

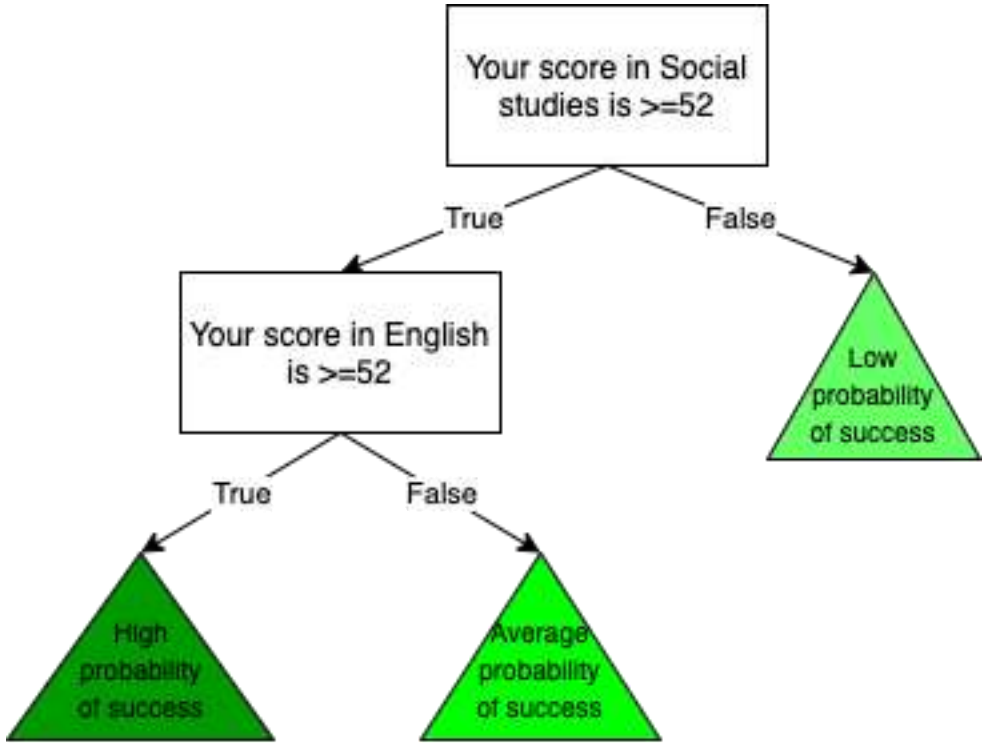
Paulina Ocampo

Maria José Gutiérrez

Miguel Correa

Mauricio Toro

Algorithm Design

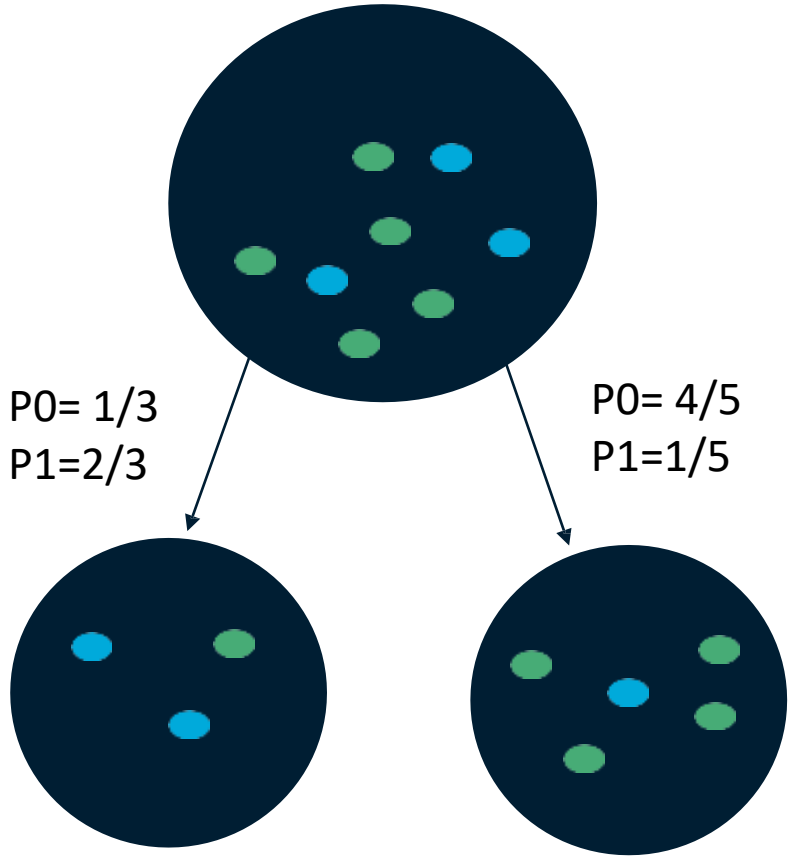


We focused on decision trees to give a solution about the prediction results in PRO knowledge test, we omit some variables to avoid discrimination. In this example, we show a CART algorithm model to predict if a student can approve an exam or not based on their score in Social studies and English in which the ICFES test were performed.

Took from: https://en.wikipedia.org/wiki/Test_(assessment)I

Node Splitting

WI=(0,44·3)+(0,32·5)/3+5
WI= 0,365



LI=1-((1/3)^2 + (2/3)^2)=1-(5/9)
LI= 0,44

RI= 1-((4/5)^2 + (1/5)^2)=1-(17/25)
RI= 0,32

As an example, this split is based on the condition “Score in Social studies is >=52.” For this case, left Gini impurity is 0.44, right Gini impurity is 0.32 and weighted Gini impurity is 0.365. It help us to find the minimum percentage to know the success for each node.

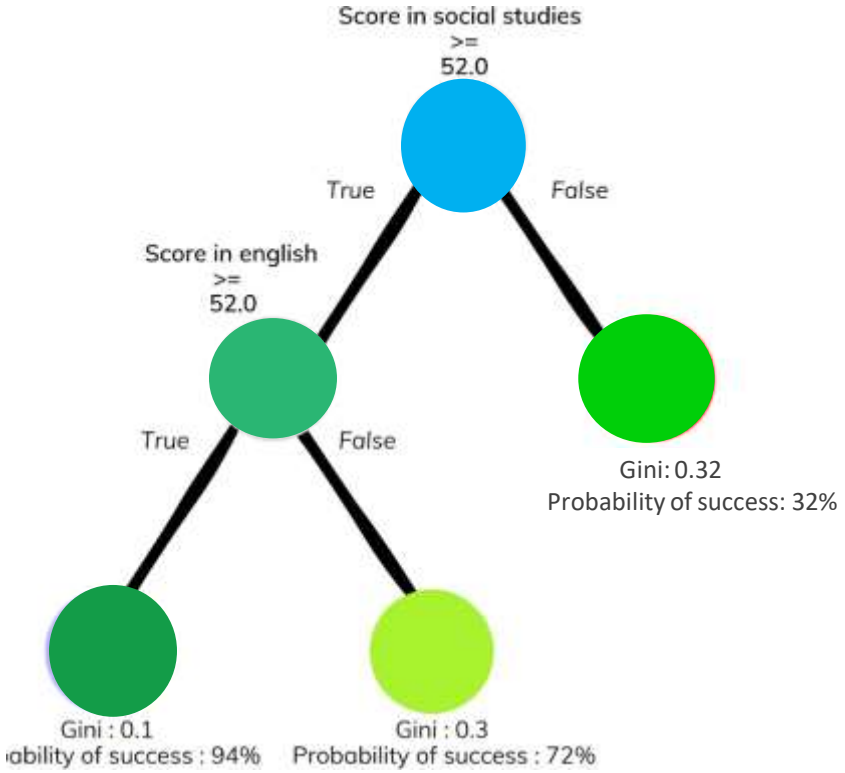
Algorithm Complexity

	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 CART algorithm.
Being M the amount of columns and N the amount of rows

Took from: https://www.epicentrochile.com/2015/08/02/la-pobreza-ralentiza-el-desarrollo-de-materia-gris-en-el-cerebro-de-los-ninos/

Decision-Tree Model

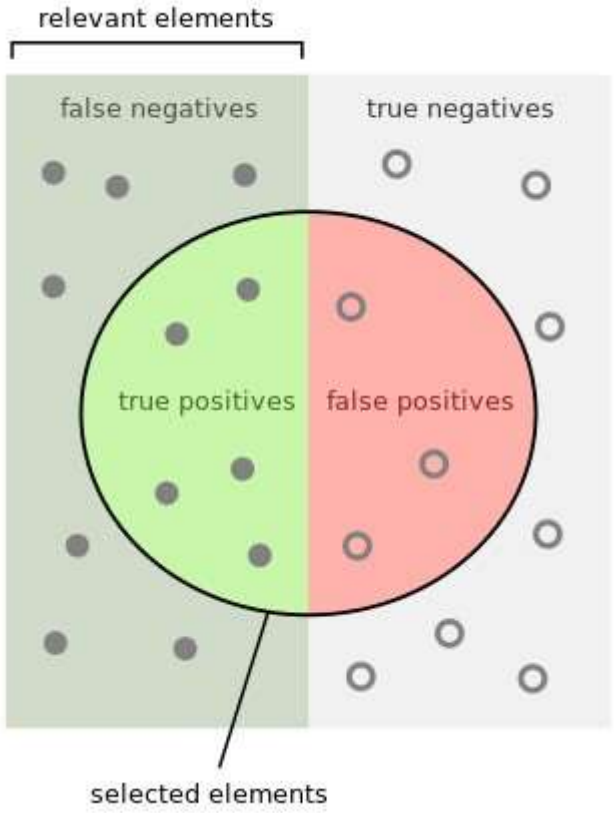


A binary decision tree to predict Saber Pro scores based on the results of Saber 11. Dark green node represent those with a high probability of success, light green an average probability and medium green a low probability of success.

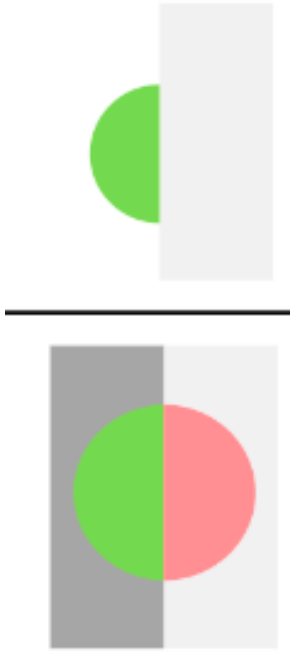
Most Relevant Features

- Social Studies
- English
- Chemistry

Evaluation Metrics



Accuracy



Precision

Recall

Evaluation Metrics

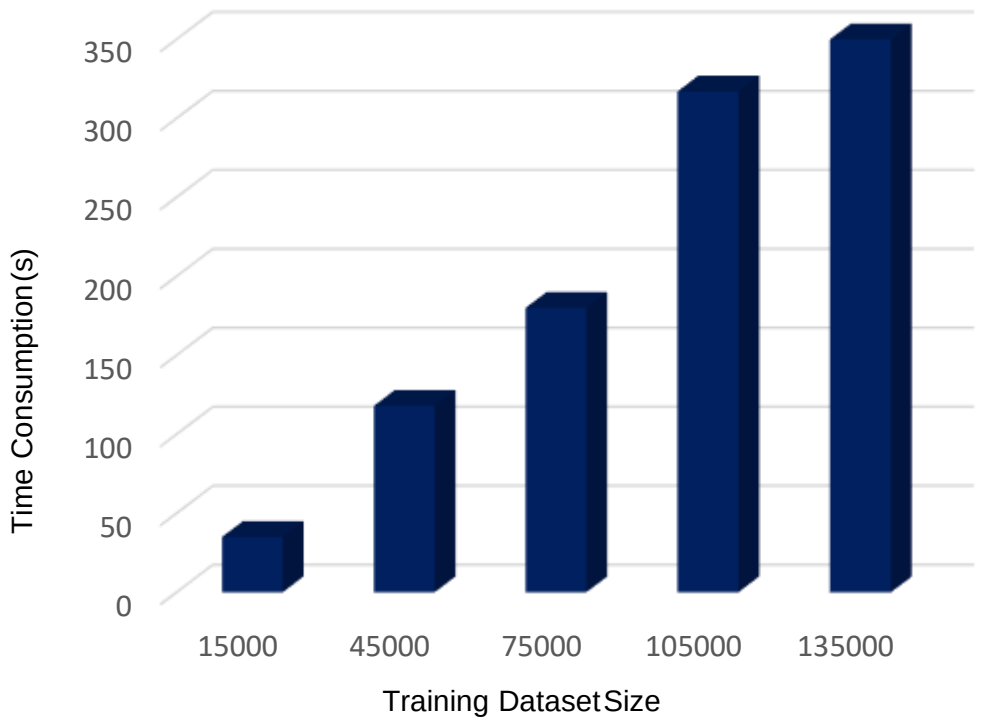
	Training data set	Testing data set
Accuracy	0.78	0.77
Precision	0.76	0.76
Recall	0.79	0.78

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

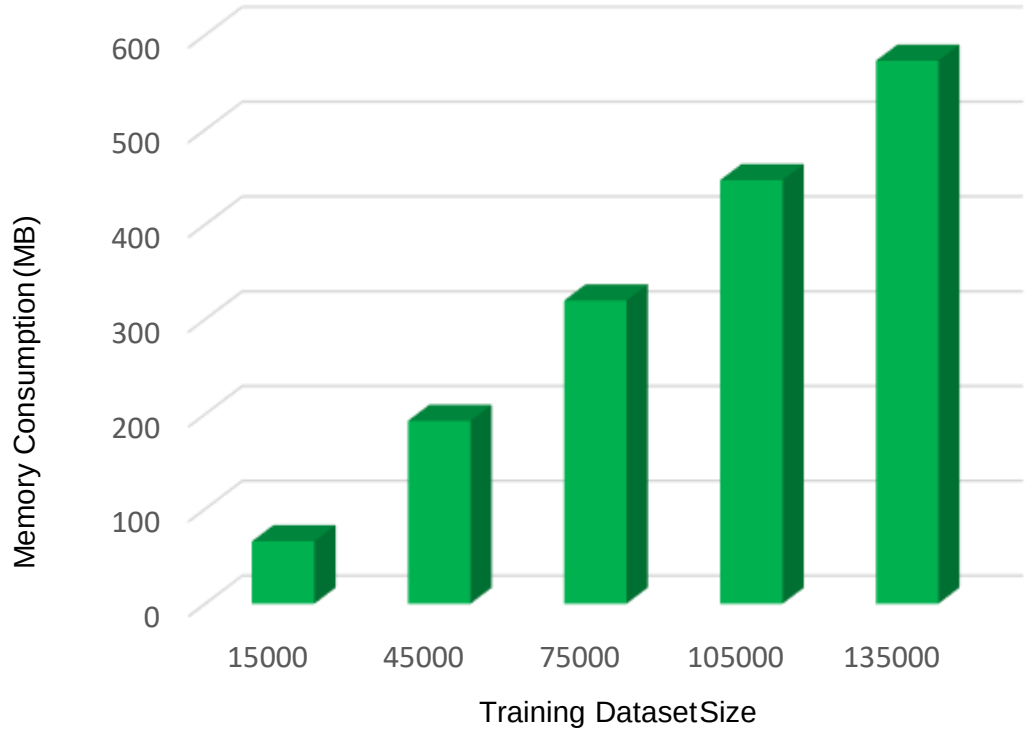
Took from: http://diariote.mx/?p=12695

UNIVERSIDAD EAFIT

Time and Memory Consumption



Time Consumption



Memory Consumption

THANK YOU!