UTILIZING DECISION TREES TO PREDICT STUDENT PERFORMANCE ON THE TEST "SABER PRO"

UNIVERSIDAD EAFIT®

Santiago Gonzalez
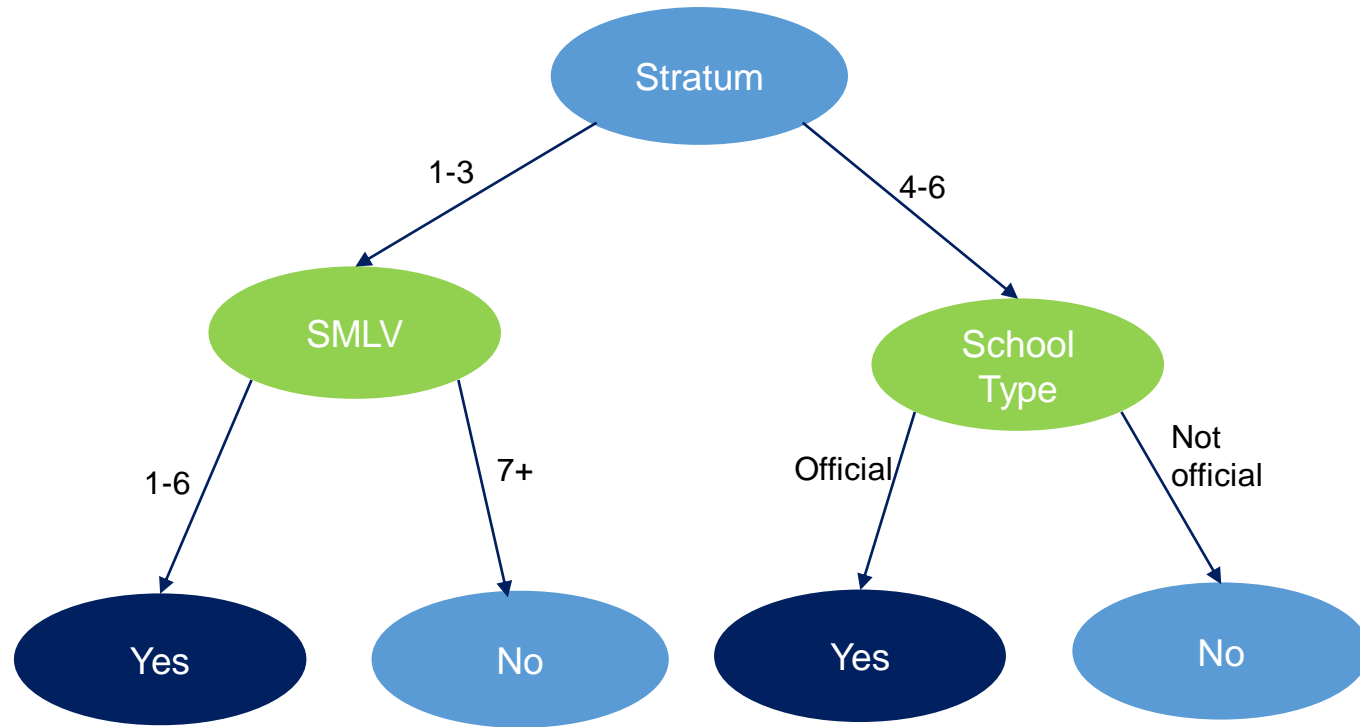
Mariana Vasquez

Miguel Correa

Mauricio Toro

UNIVERSIDAD EAFIT®
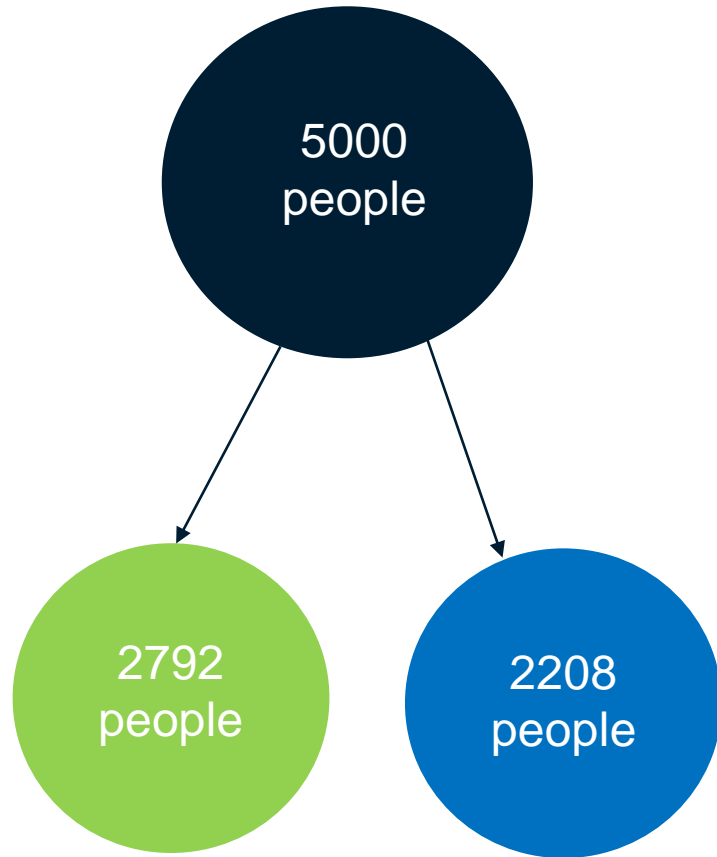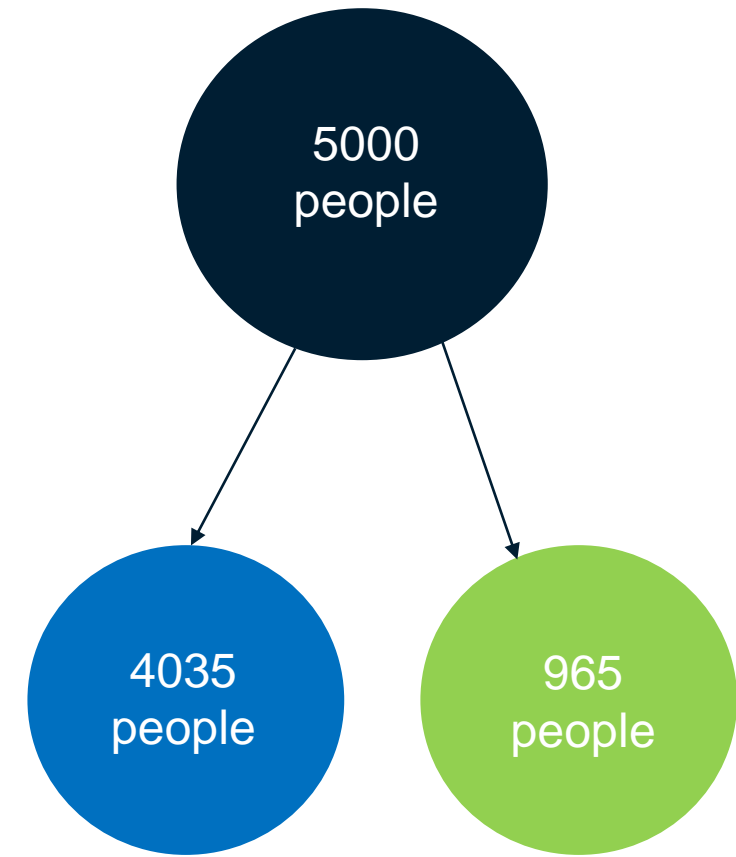
This is the training model for the CART tree that will be used in this project. In this example, we show a model to make decisions based on the stratum, the SMLV and the school type.

Node Splitting

As an example, this split is based on the condition "SchoolType==OFICIAL". For this case, left Gini impurity is 0.55, right Gini impurity is 0.44, and weigthed Gini impurity is 0.49.

As an example, this split is based on the condition "stratum <= 3." For this case, left Gini impurity is 0.81, right Gini impurity is 0.19, and weigthed Gini impurity is 0.31.
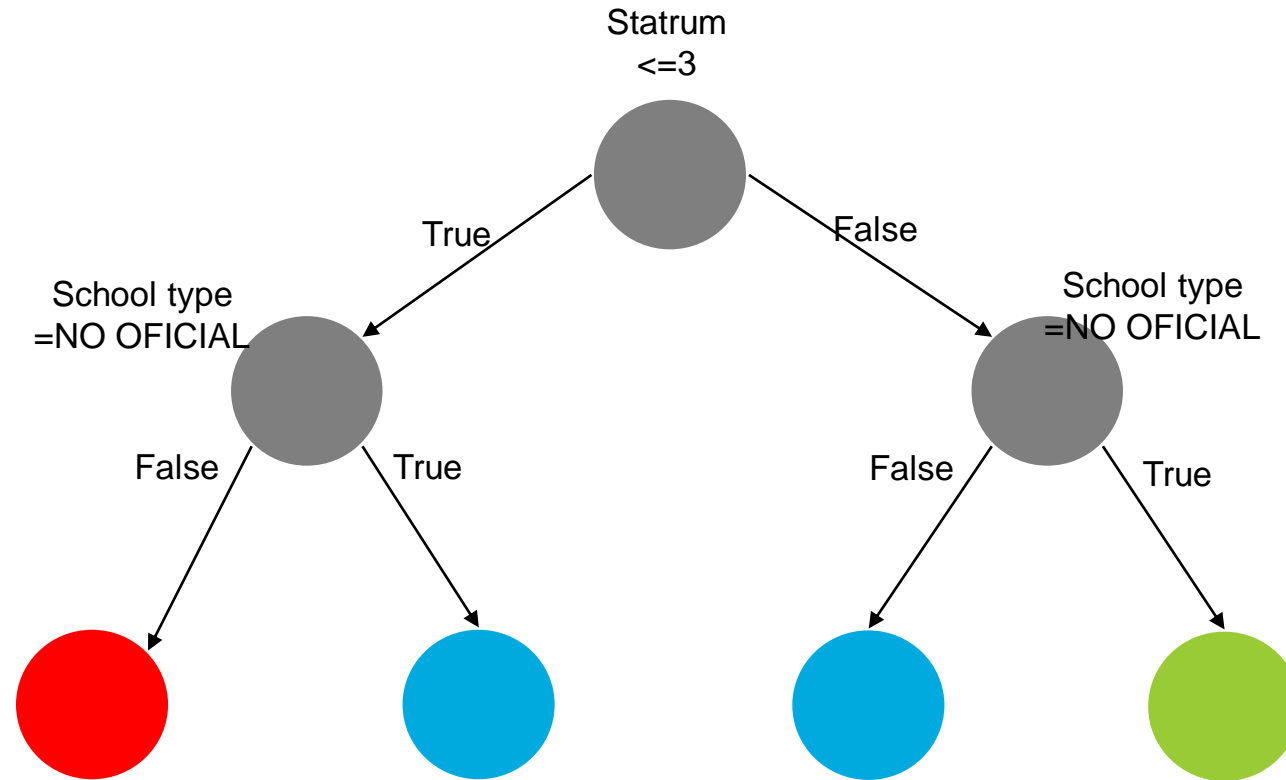
	Time Complexity	Memory Complexity
Training the model	$O(N^2{*}M)$	$O(N^2)$
Testing the Model	$O(N{*}M)$	$O(N)$

Time and memory complexity of the CART algorithm. In this table, N means the number of students and M represents the depth of the tree.

UNIVERSIDAD EAFIT®

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 medium probability and red a low probability of success.
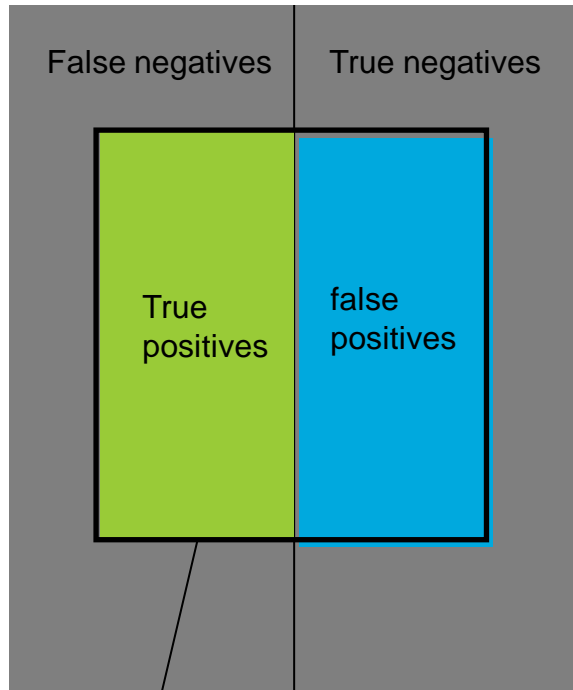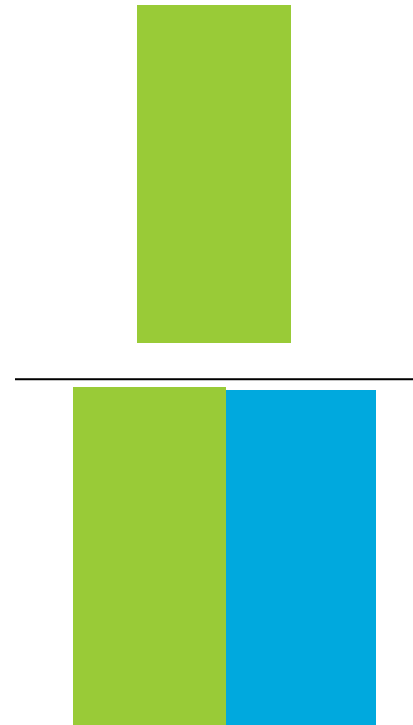
Most Relevant Features
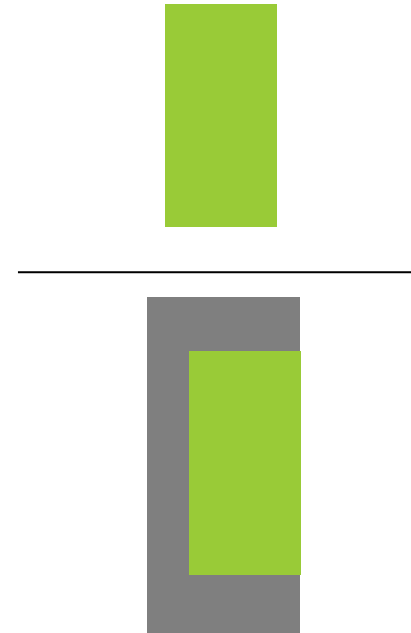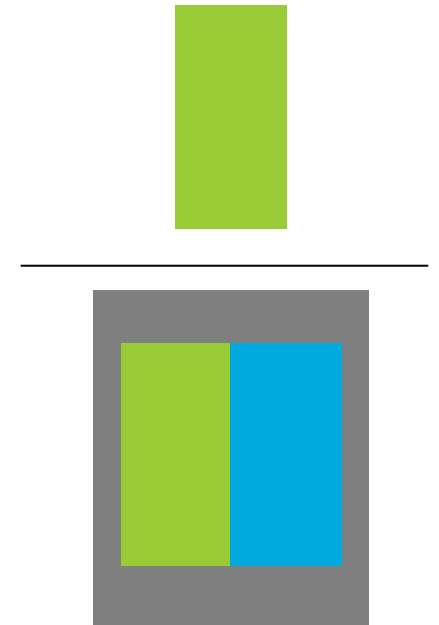
Previous scores

School type

Stratum

False negatives | True negatives

True positives | false positives

Selected elements

Precision

Recall

Accuracy

Evaluation Metrics

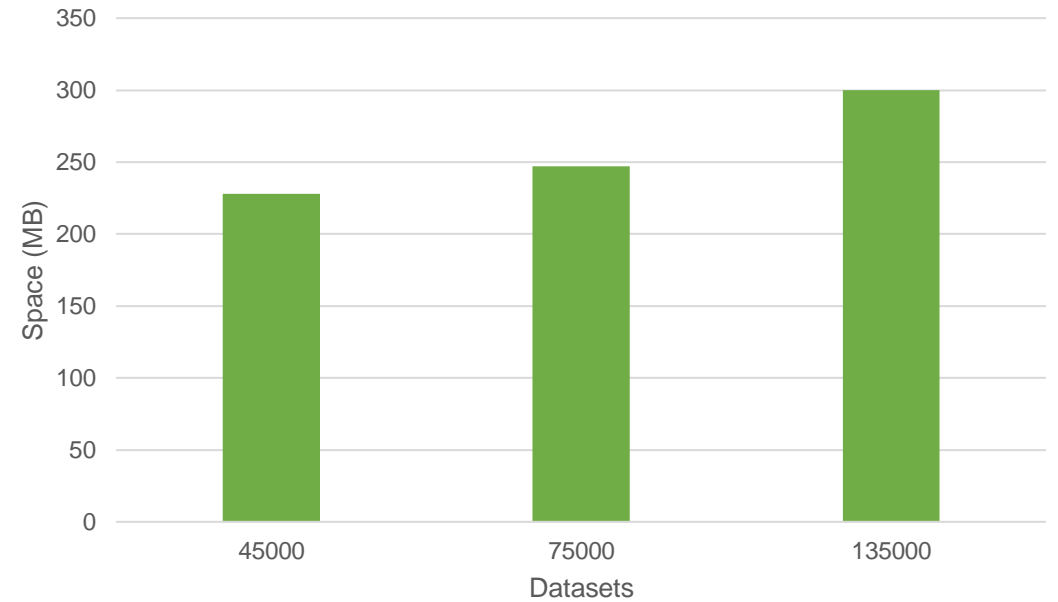|	Training data set	Testing data set
Accuracy	0.80	0.96
Precision	0.74	0.90
Recall	0.78	0.98

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

Time and Memory Consumption

THANK YOU!