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

UNIVERSIDAD EAFIT®

Santiago
Gonzalez
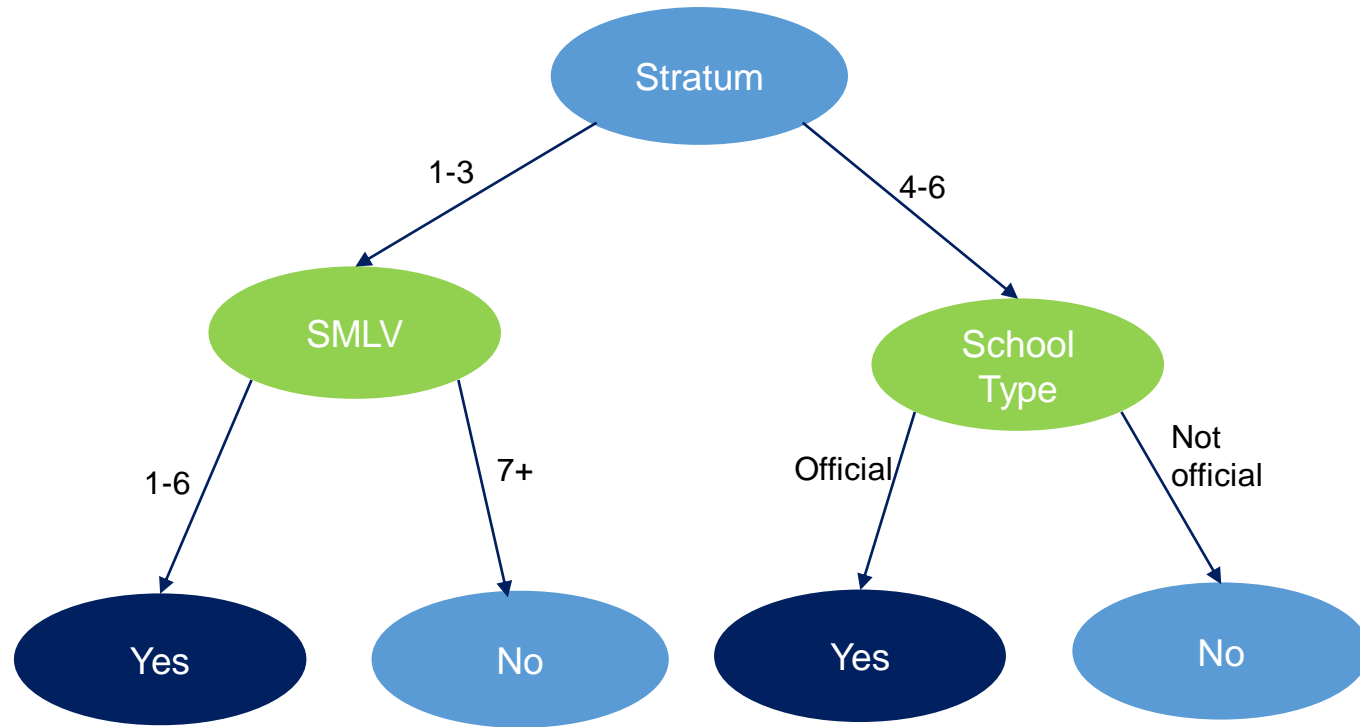
Mariana
Vasquez

Miguel
Correa
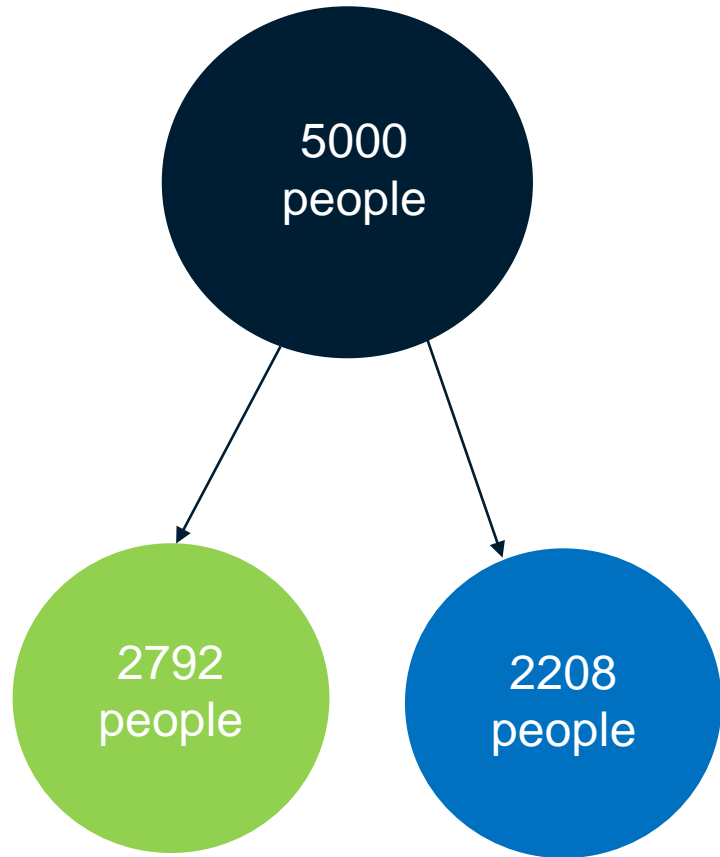
Mauricio
Toro

UNIVERSIDAD
EAFIT®

Algorithm Design

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.

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.
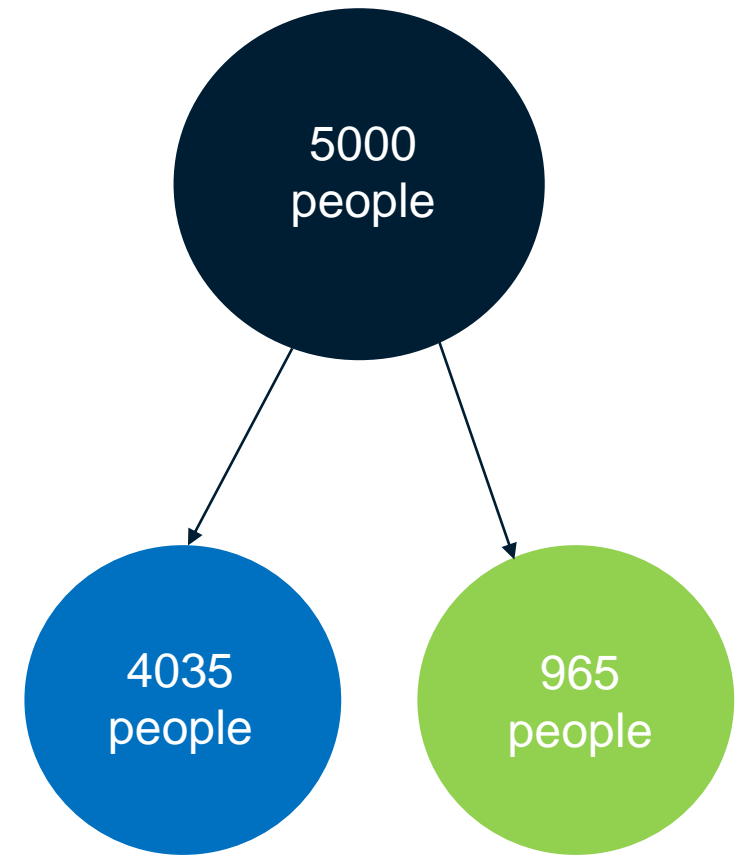
As an example, this split is based on the condition "stratum <= 3." For this case, left Gini impurity is 0.81, right Gini impurity is 0.19, and weigthed Gini impurity is 0.31.

Algorithm Complexity

	Time Complexity	Memory Complexity
Training the model	$O(N^2+M)$	$O(N^2*2M)$
Testing the Model	$O(N*M+4^{1-N})$	$O(2N^2+2^{1-M})$

Time and memory complexity of the CART algorithm. In this table, N means the number of students and due to the differences in the methods of the code, M in the training is the amount of data that is stored in the nodes, and the number of nodes in the testing phase.

Decision-Tree Model
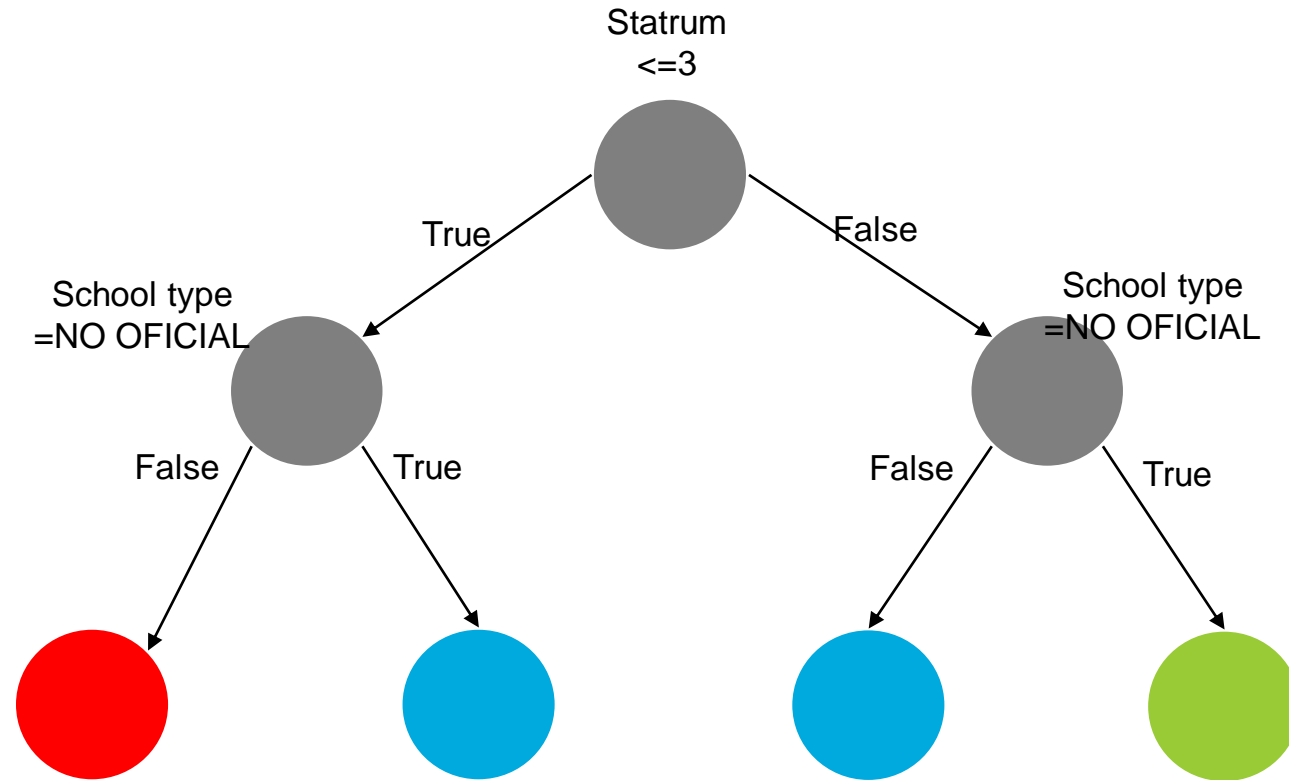
A binary decision tree to predict Saber Pro scores based on the results of Saber 11. Green nodes represent those with a high probability of success, blue medium probability and red a low probability of success.
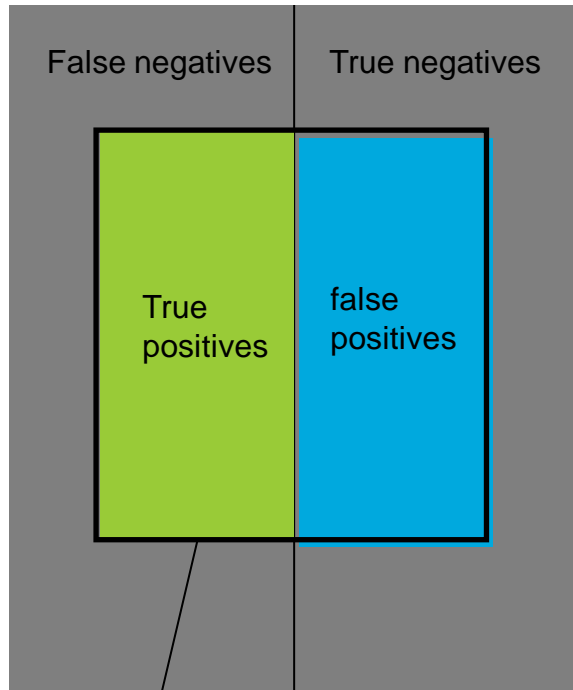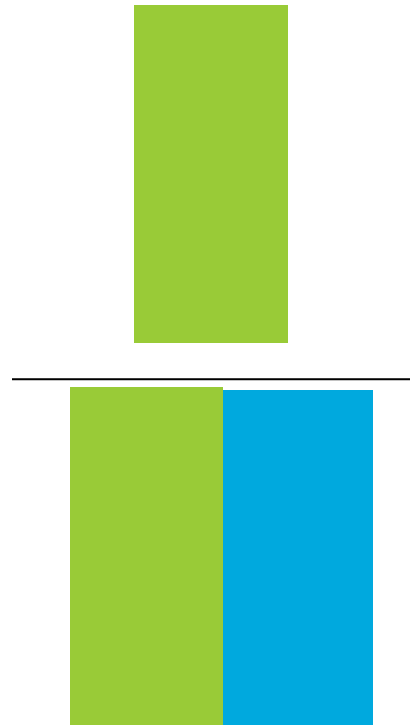
Most Relevant Features
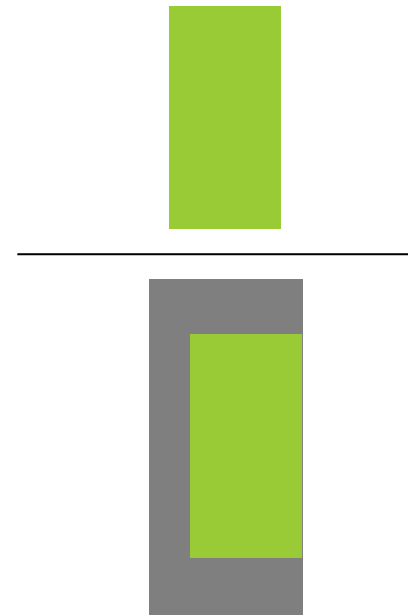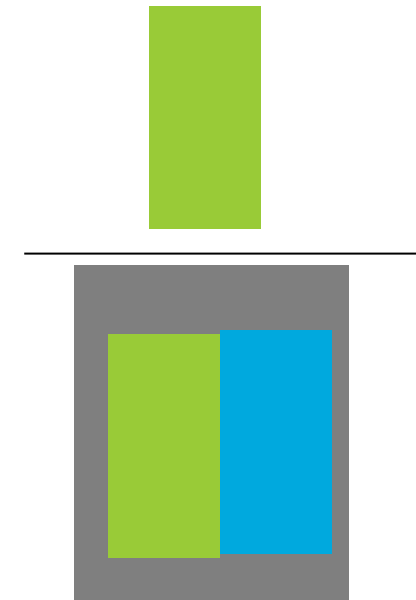
Previous scores

School type

Stratum

False negatives | True negatives

True positives | false positives

Selected elements

Precision

Recall

Accuracy

Evaluation Metrics

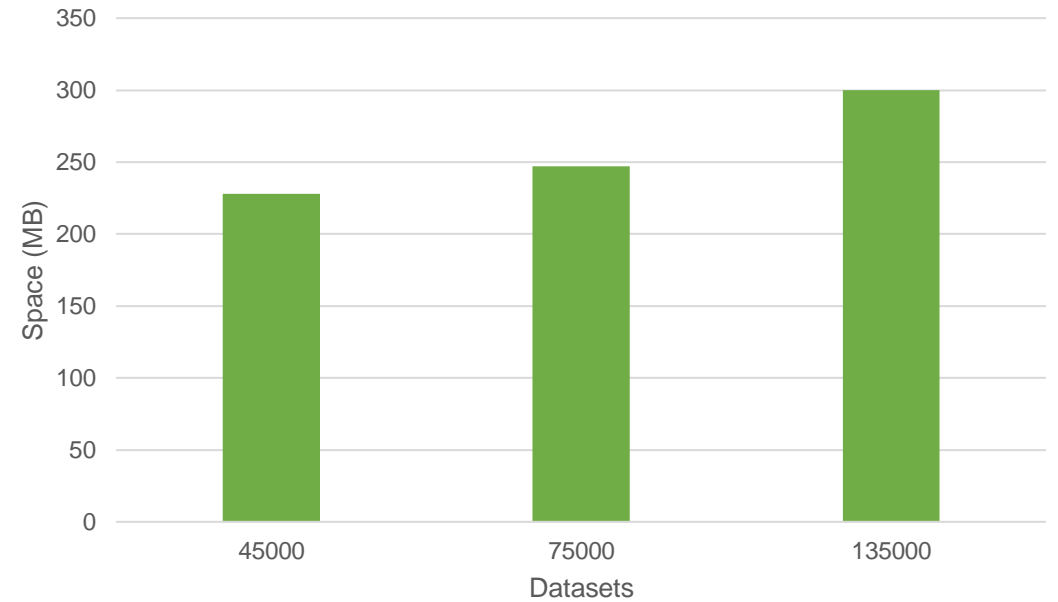|	Training data set	Testing data set
Accuracy	0.48	0.48
Precision	0.50	0.52
Recall	0.49	0.52

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

Time Consumption

Memory Consumption

THANK YOU!