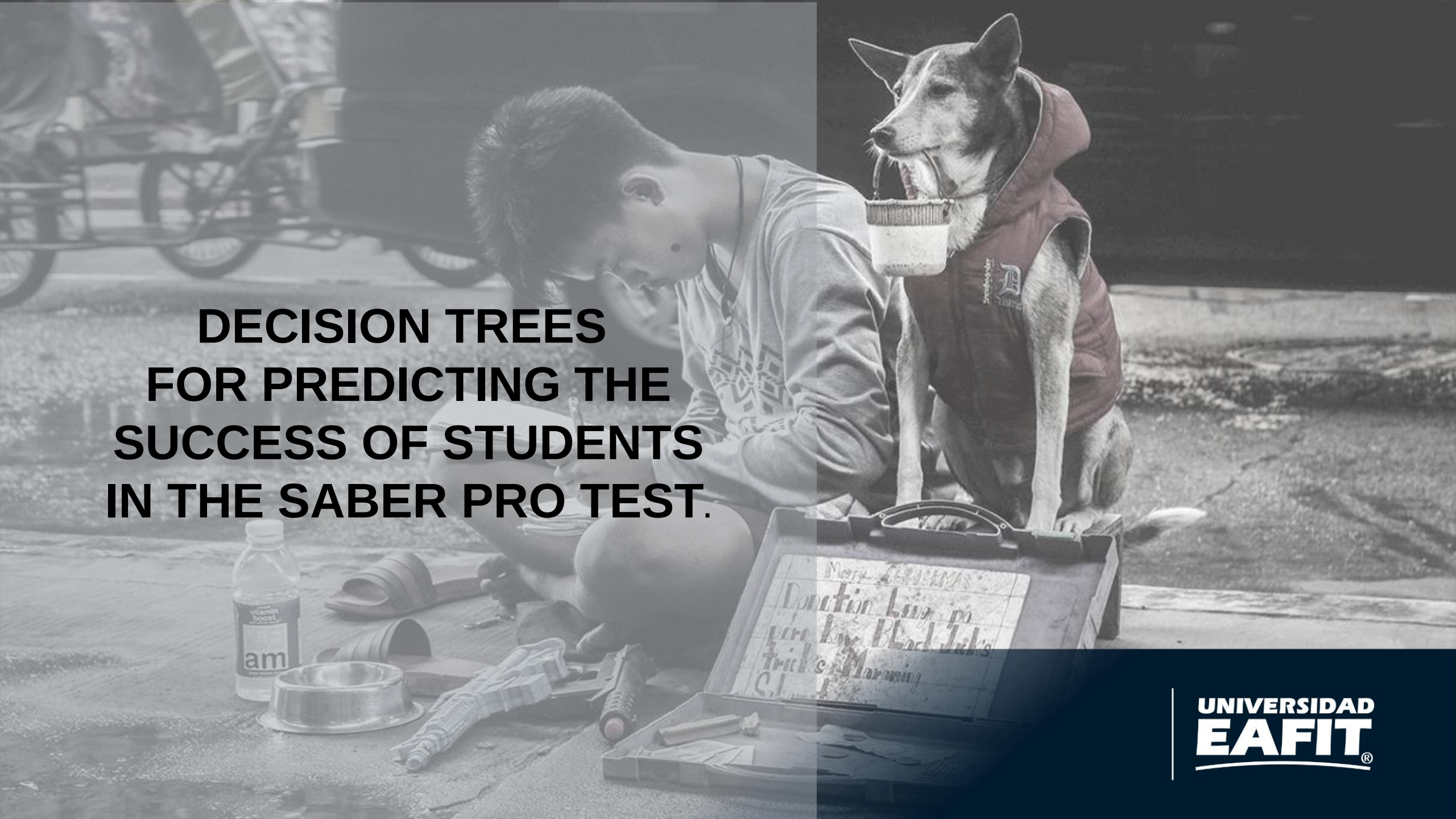


DECISION TREES FOR PREDICTING THE SUCCESS OF STUDENTS IN THE SABER PRO TEST.

UNIVERSIDAD EAFIT®

Team Presentation

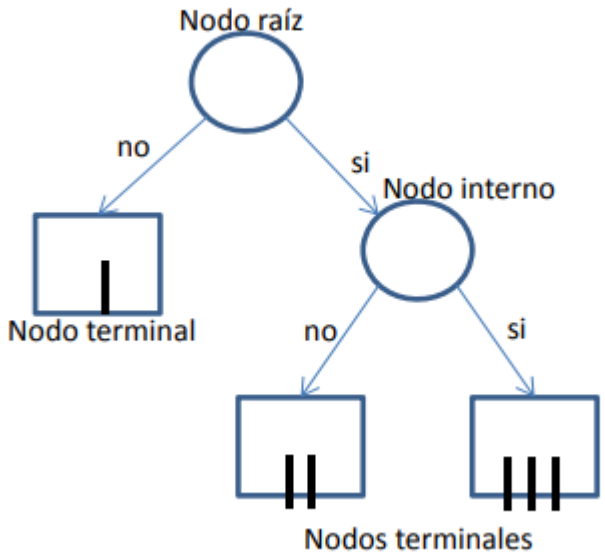
Samuel David Ben Jacob Arango Henao

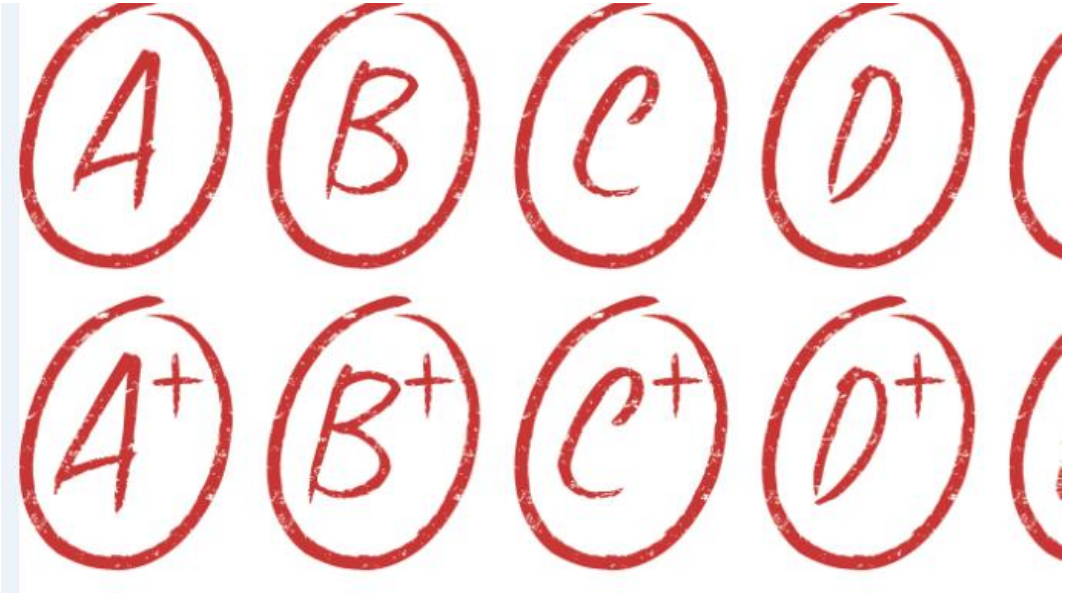
Miguel Correa

Mauricio Toro

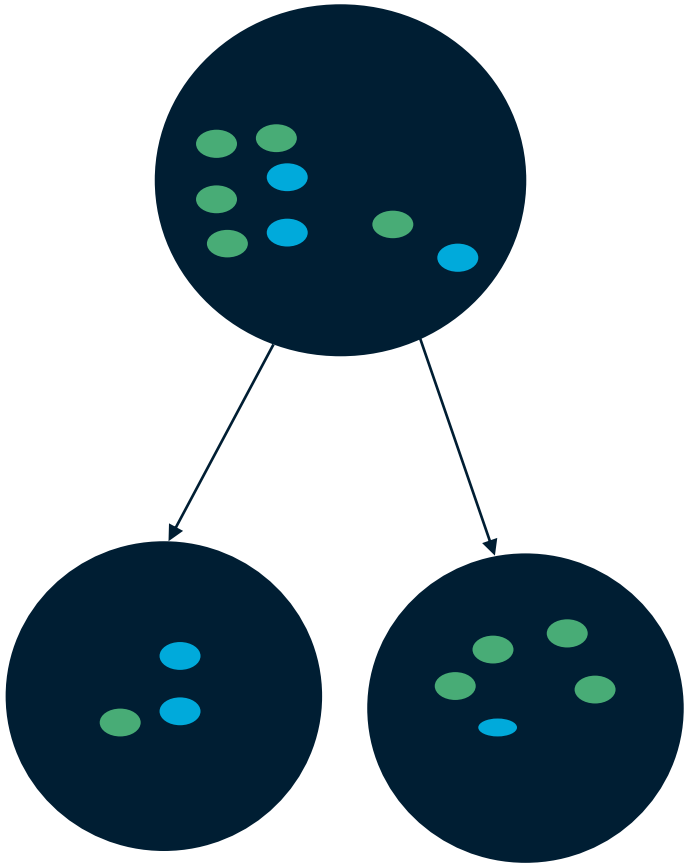
Algorithm Design

The algorithm chosen is a CART algorithm, which classifies the results obtained from the data searching for that variables more homogenous.

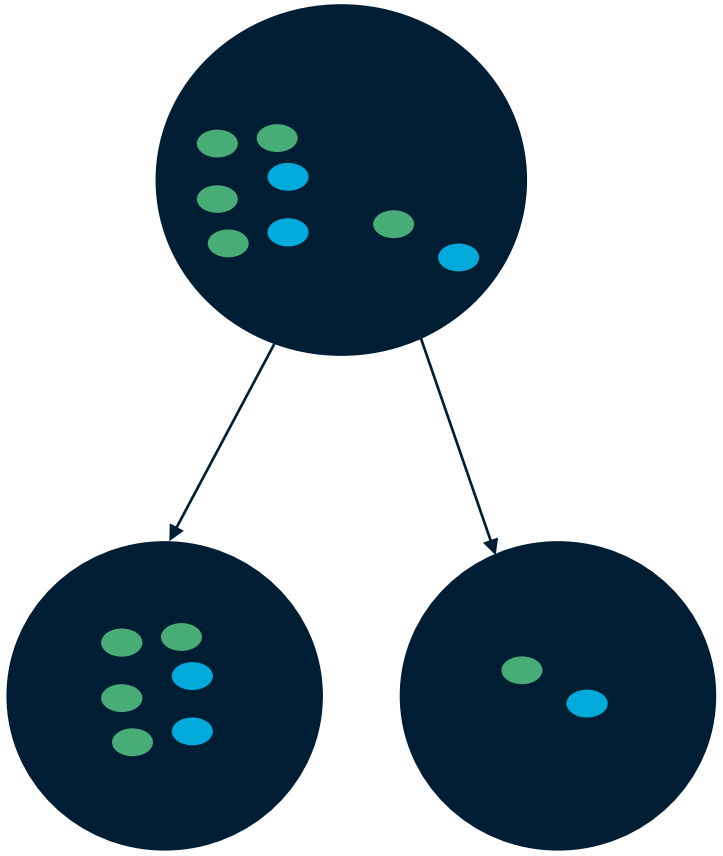




Node Splitting



The Algorithm chooses the variable which gives most homogeneity to the problem and put it on the top of the tree, and so on with the second most common.



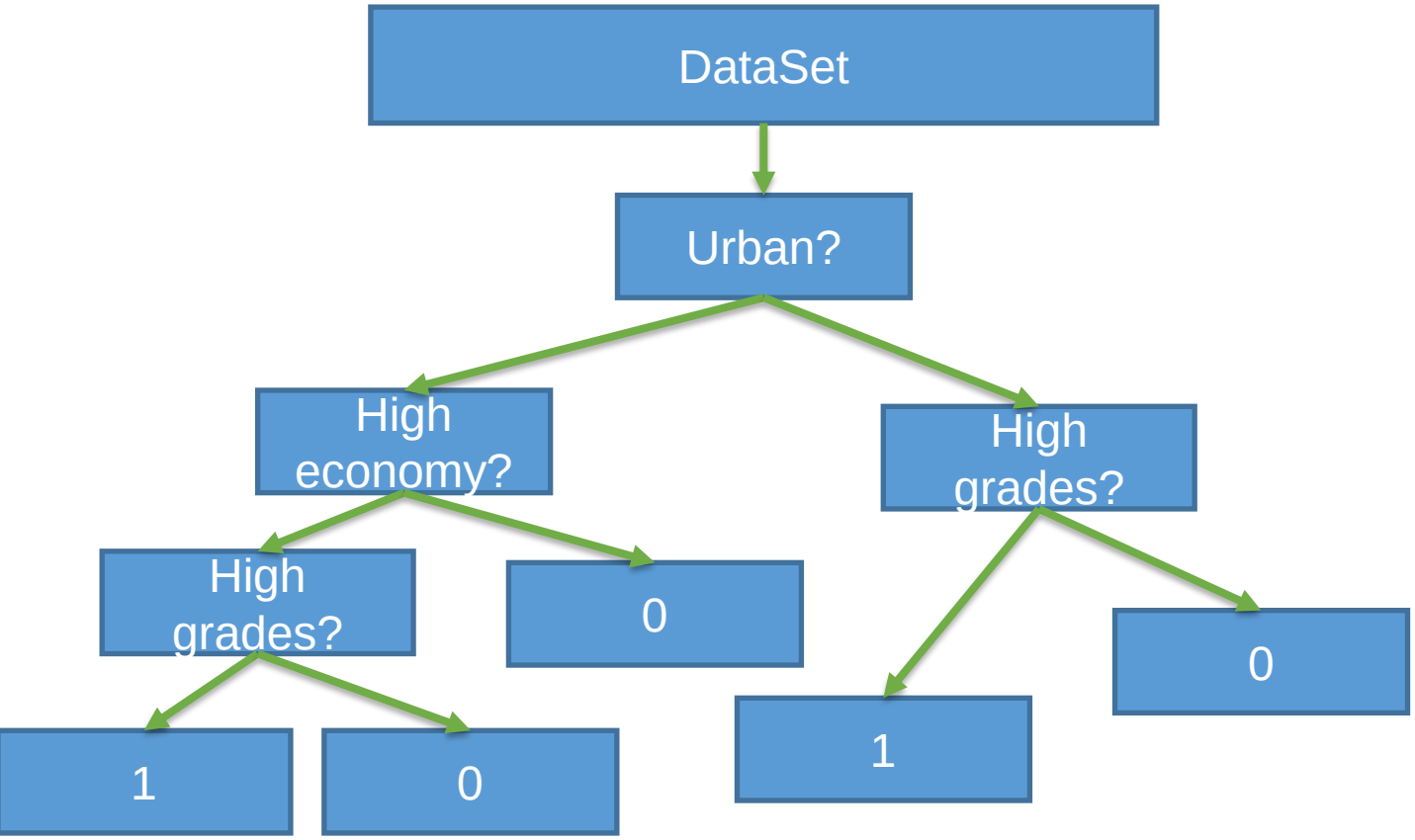
As an example, the área where is ubicated the educational stablishment seems to have an important relevance about the succes.

Algorithm Complexity

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

I’ve choose a CART algorithm, which is simple and gives a high accuracy.

Decision-Tree Model



Most Relevant Features

- Economy
- Results
- Residence

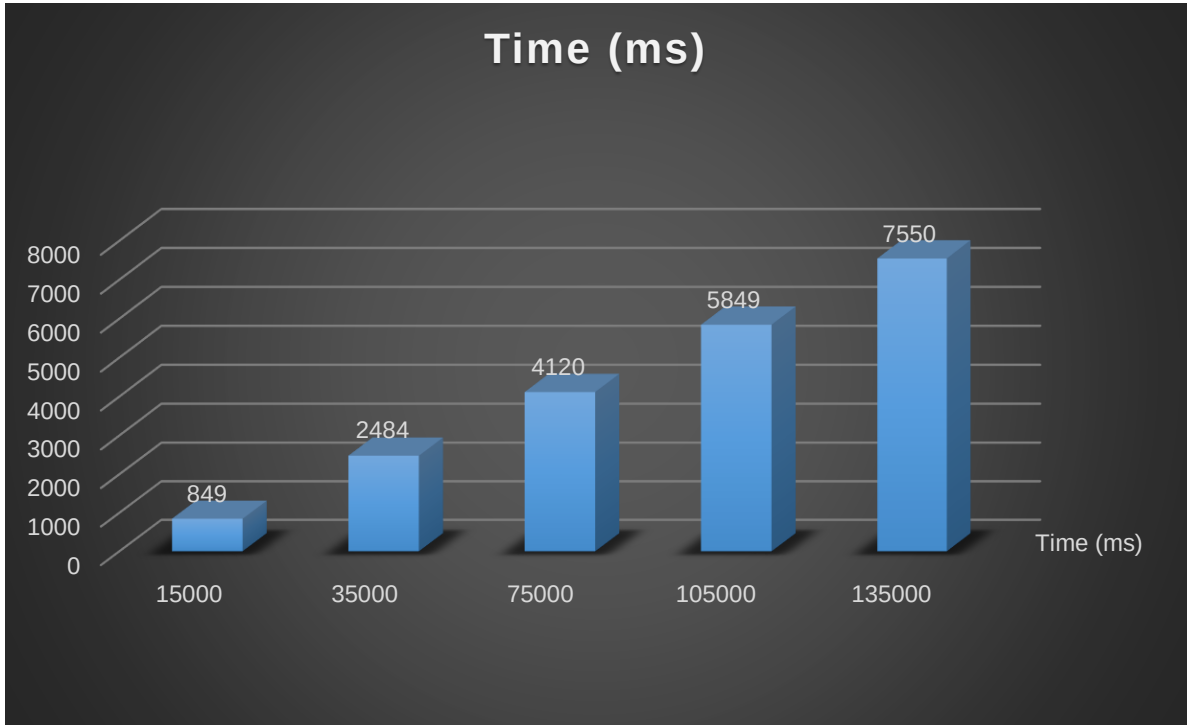
The present model takes the grades of the student ass a determinant variable, but considers first other possibilities.

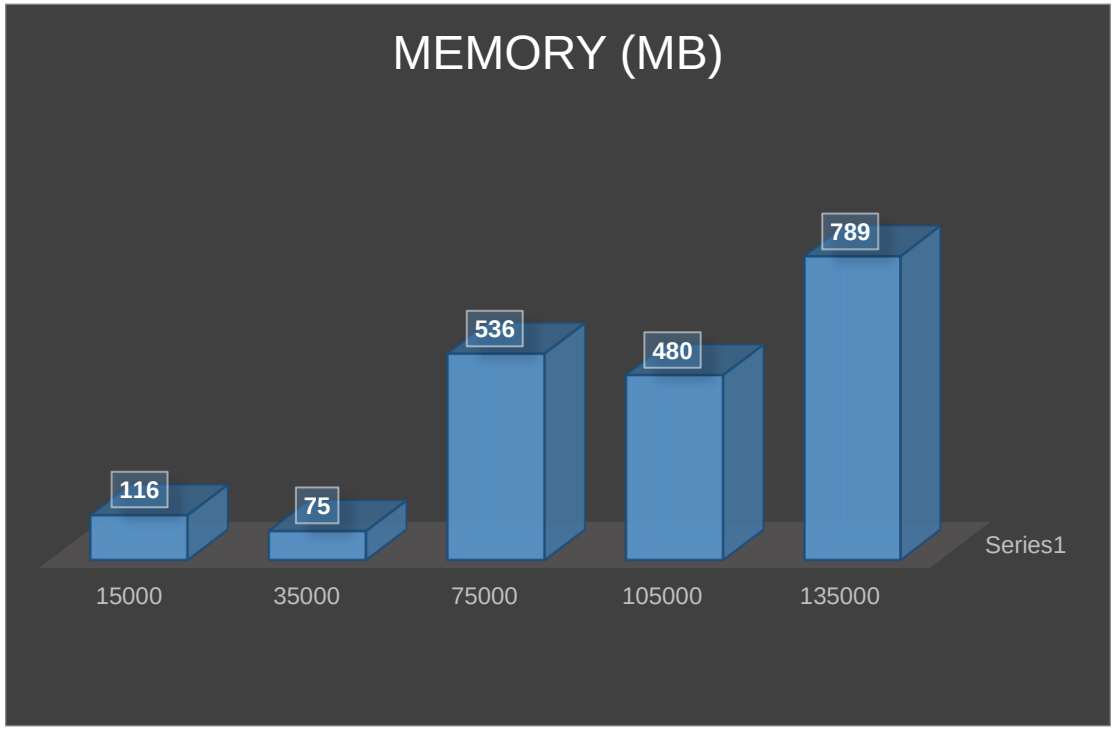
Evaluation Metrics

	Training data set	Testing data set
Accuracy	0.72	0.67
Precision	0.65	0.58
Recall	0.69	0.60

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

UNIVERSIDAD EAFIT

THANK YOU!

Is just a meme, don’t worry, i don’t have one.

https://www.uv.es/mlejarza/actuariales/tam/arbolesdecision.pdf