The image is a composite. The left side shows a young man with dark hair, wearing a light-colored long-sleeved shirt, sitting on the ground and looking down at a book or paper. He is surrounded by various items: a plastic bottle of 'am' brand vitamin water, a metal dog bowl, a pair of sandals, and some papers. The right side shows a dog, possibly a pit bull mix, sitting upright and holding a small calculator in its mouth. The dog is wearing a dark red hoodie. The background is a blurred outdoor setting with a motorcycle visible on the left.

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

Team Presentation



Samuel David Ben Jacob
Arango Henao



Miguel
Correa

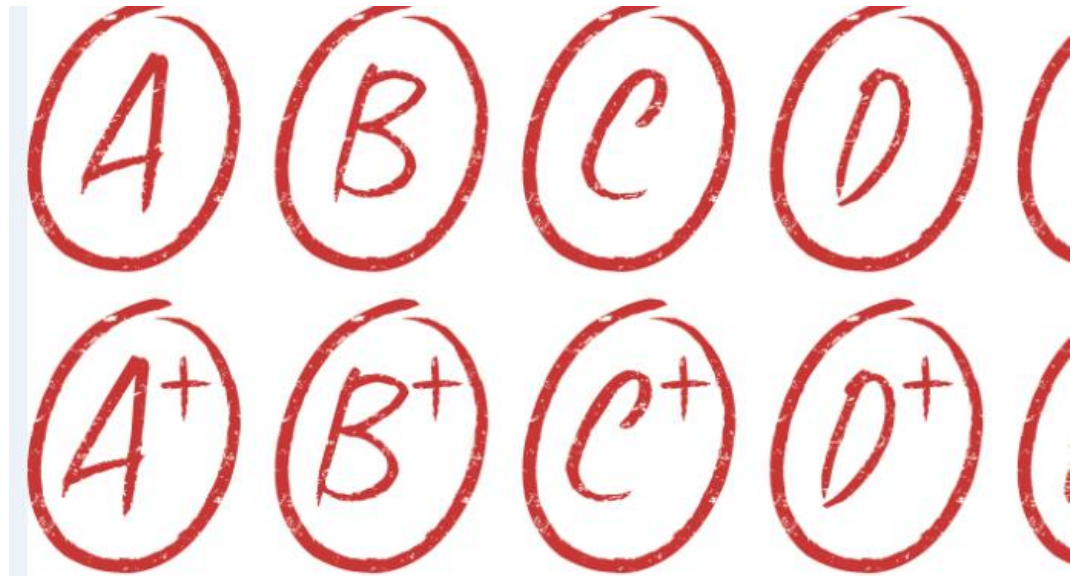
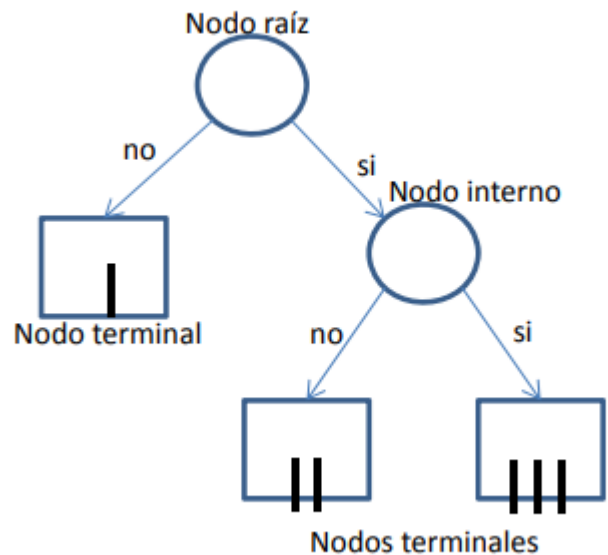


Mauricio
Toro

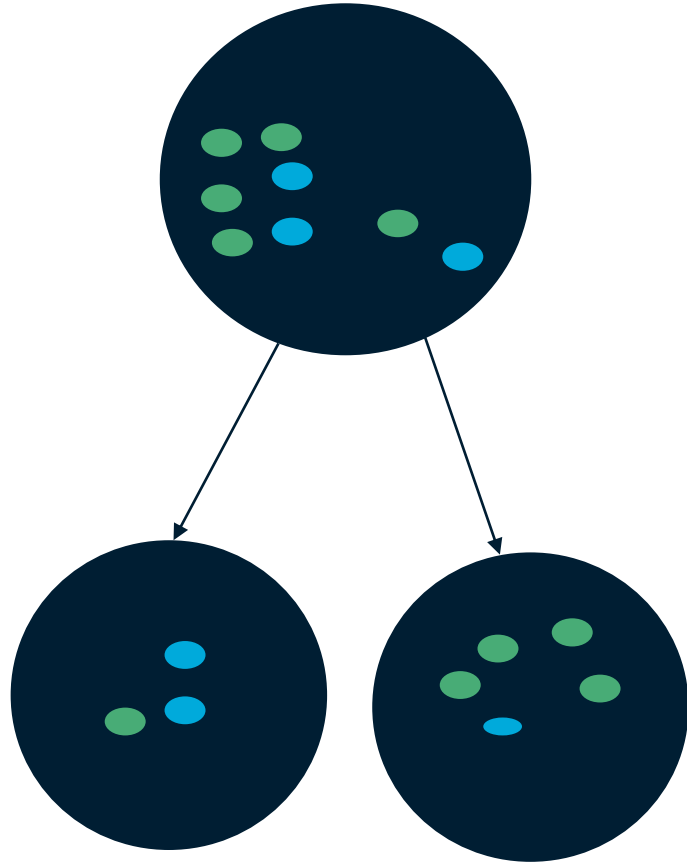


<http://github.com/Sdarangoh/proyecto/>

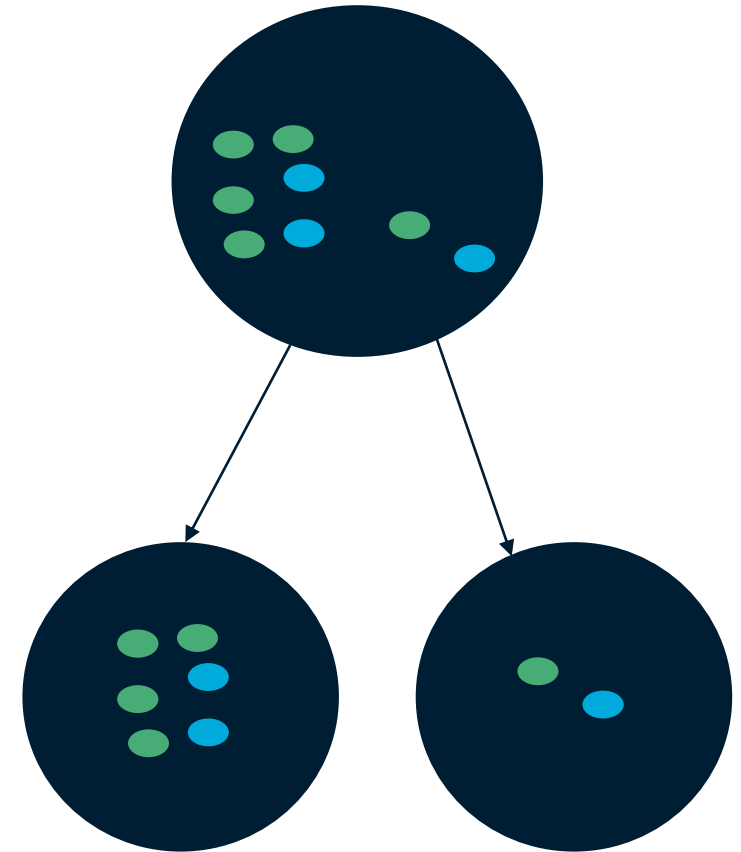
The algorithm chosen is a CART algorithm, which classifies the results obtained from the data searching for that variables more homogenous.



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.



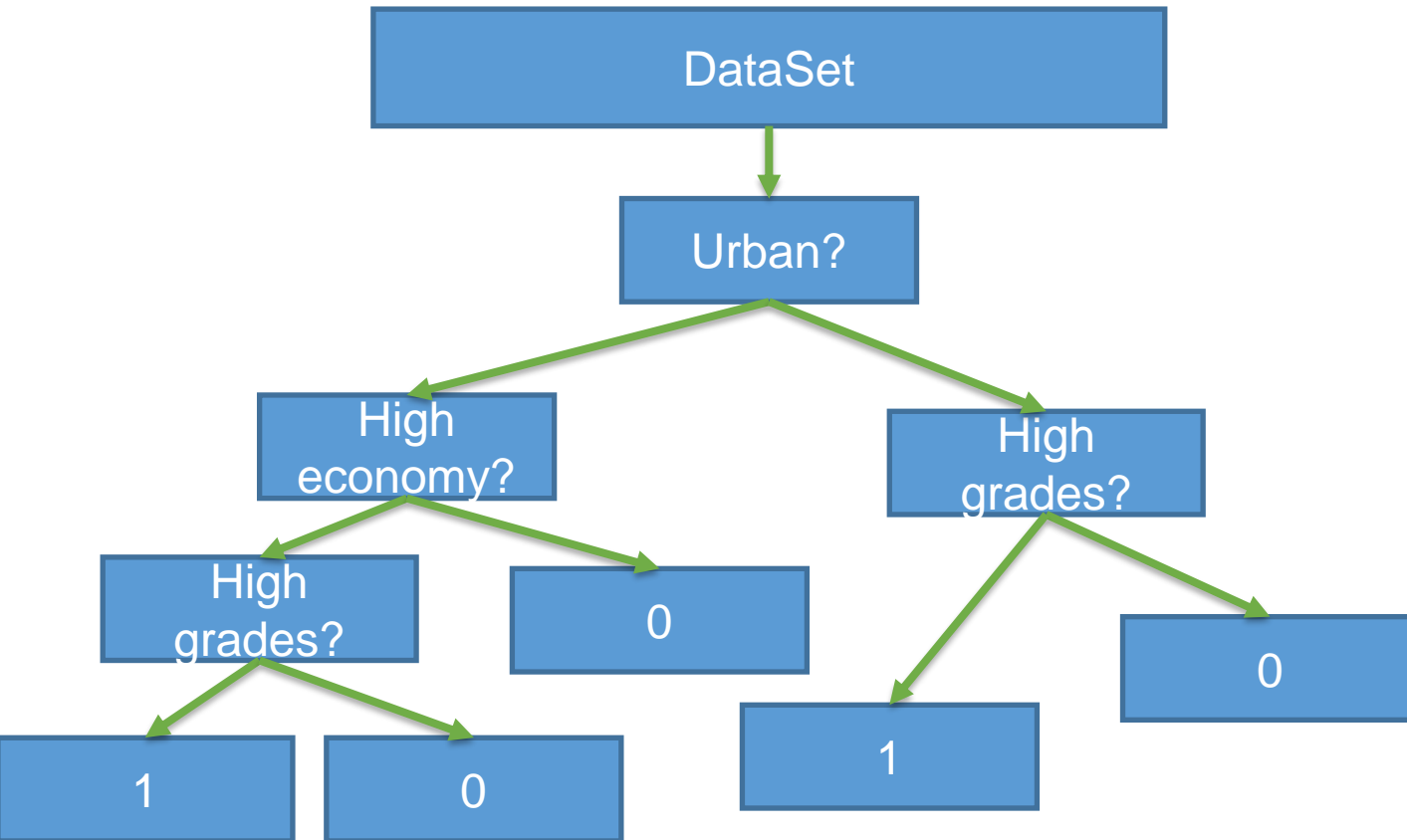
As an example, the área where is ubicated the educational stablshment seems to have an important relevance about the succes.



| | Time Complexity | Memory Complexity |
|--------------------|------------------|-------------------|
| Training the model | $O(N^m \cdot M)$ | $O(N \cdot M)$ |
| Testing the Model | $O(N \cdot M)$ | $O(1)$ |

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

Decision-Tree Model



Most Relevant Features



Economy



Results



Residence

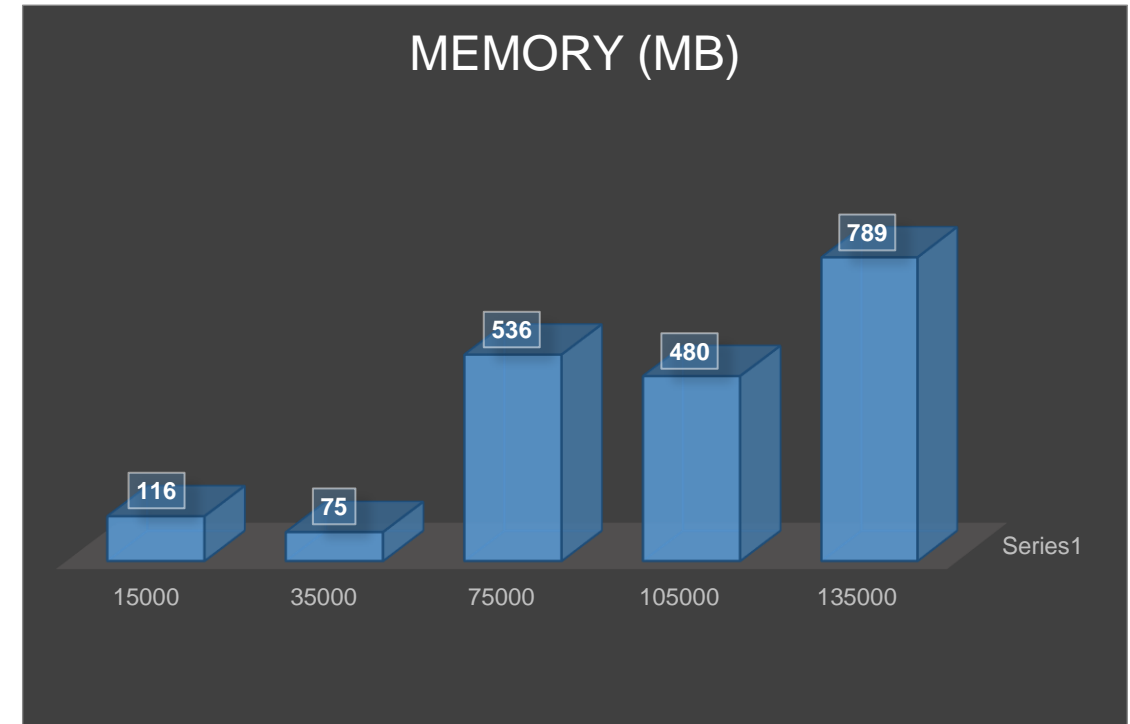
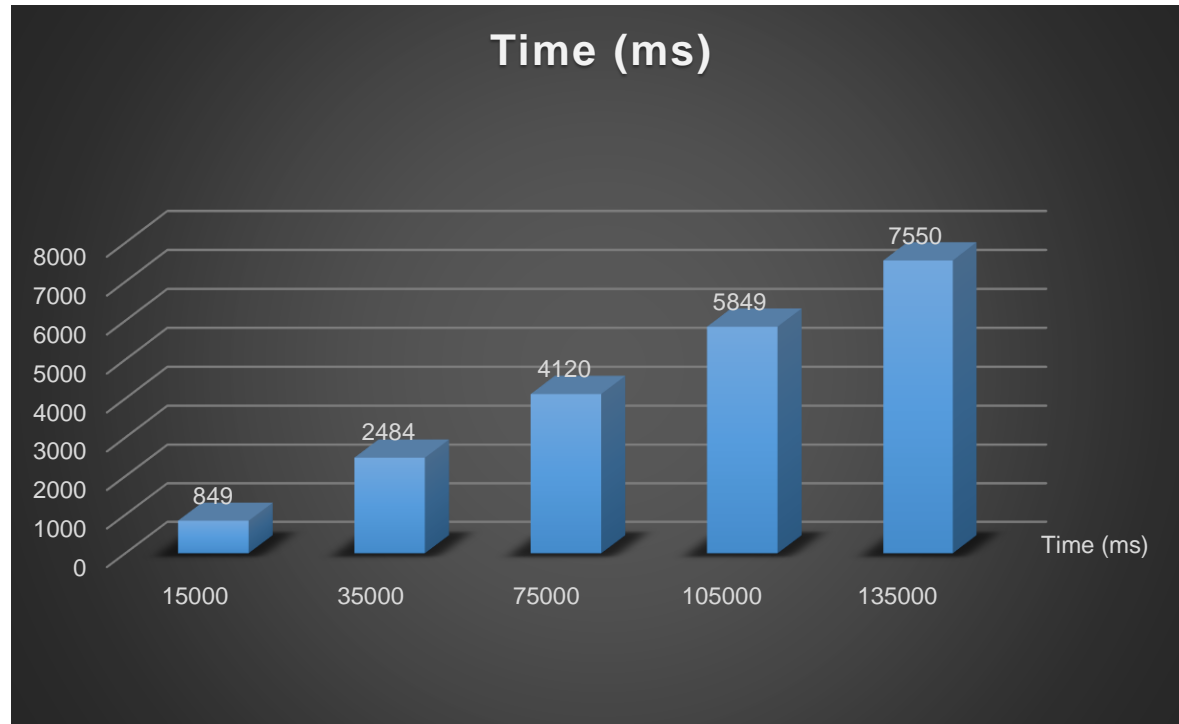
The present model takes the grades of the student as a determinant variable, but considers first other possibilities.

| | 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



**Haciendo el trabajo de EDA a
última hora**

**Y empieza a hablarme la
escopeta.**

THANK YOU!

[https://www.uv.es/mlejarza/actuariales/tam/
arbolesdecision.pdf](https://www.uv.es/mlejarza/actuariales/tam/arbolesdecision.pdf)