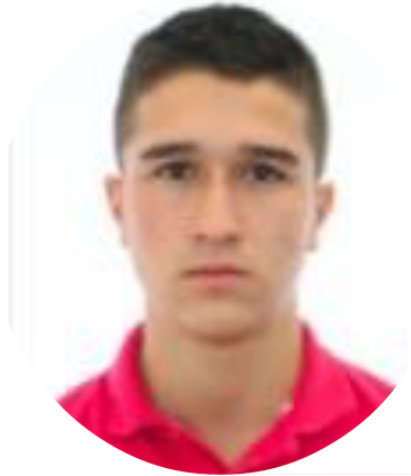


Predicting  
pruebas saber  
pro using  
decisión trees



# Team Presentation



Vicente  
Aristizabal



Miguel  
Correa



Mauricio  
Toro



<http://github.com/>

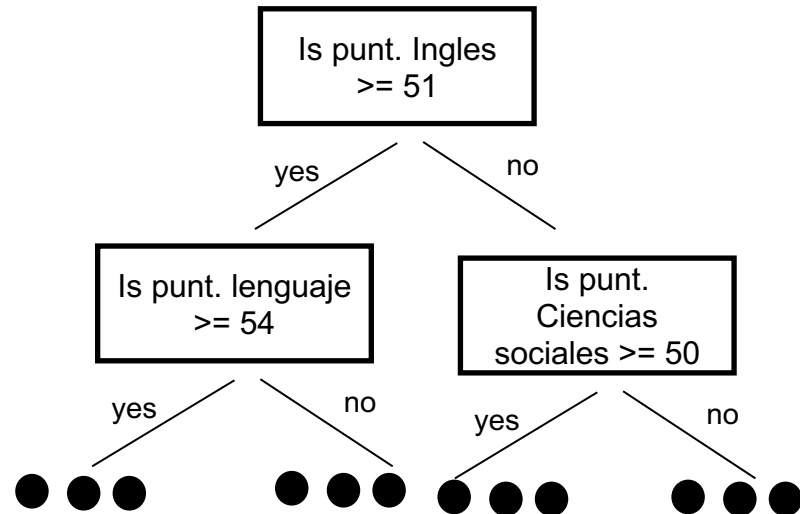
varisti7

/proyecto/



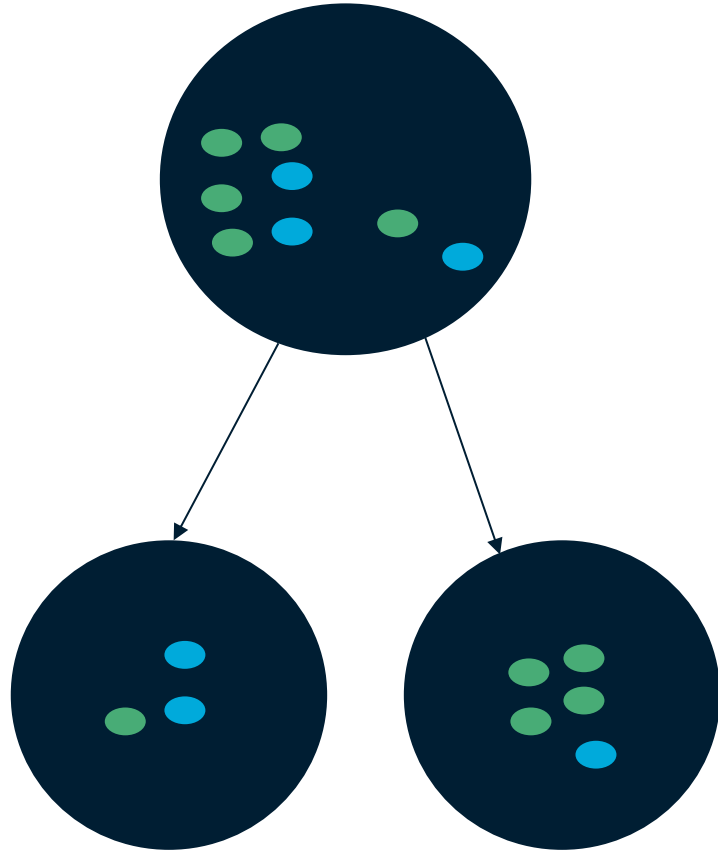


# Algorithm Design

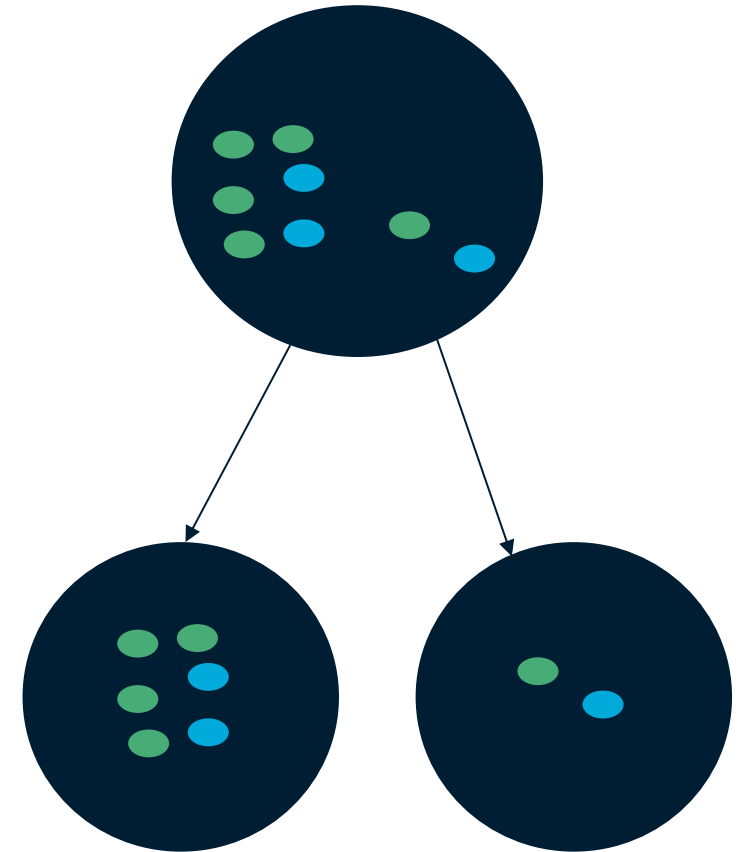


The algorithm used in the project to decide if the student will have success in the pruebas saber pro was a CART decision tree algorithm.

# Node Splitting



As an example, this split is based on the condition punt. Ingles  $\geq$  51.



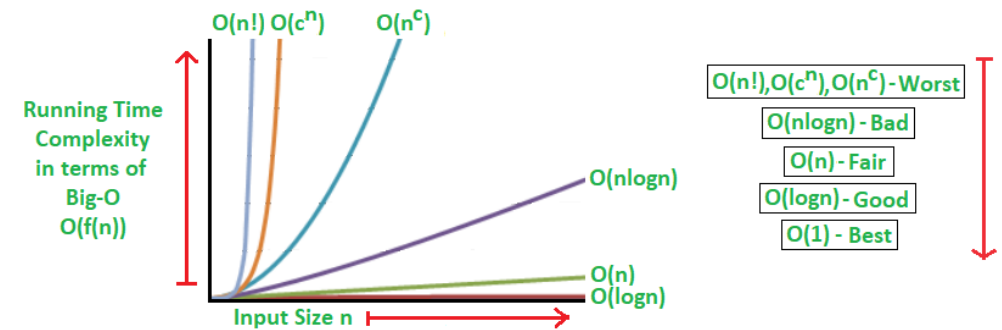
As an example, this split is based on the condition punt. lenguaje  $\geq$  54.

# Algorithm Complexity

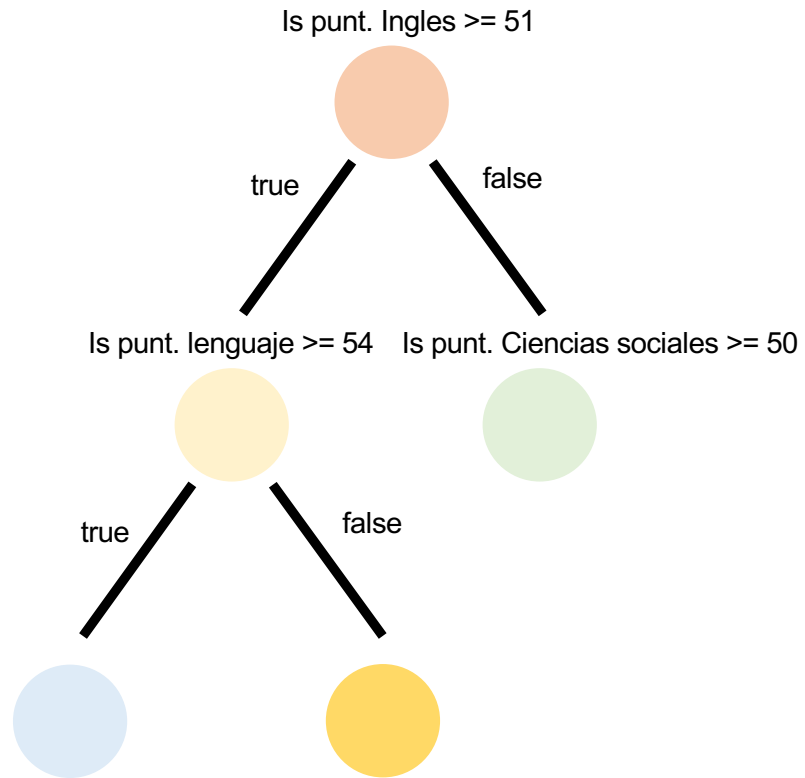


	Time Complexity	Memory Complexity
Training the model	$O(M \cdot N \cdot \log(N))$	$O(\log(N))$
Testing the Model	$O(N \cdot \log(N))$	$O(1)$

This complexities are understanding  $N$  as the number of students to evaluate and  $M$  as the number of possible questions (for example punt. Ingles times the amount of unique values present in all students, this for each possible question).



# Decision-Tree Model



In the decision tree model there was 25 variables that were evaluated such as scores in subjects, internet, cellphones, between others. The most relevant ones were English, social science and language.

## Most Relevant Features



Social Science  
Studies



English



Language

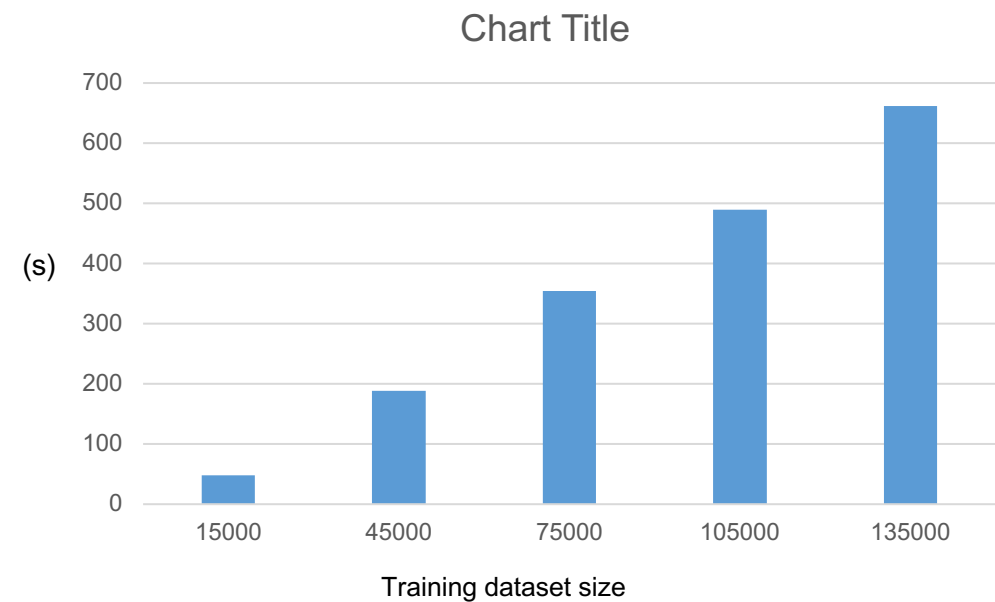
```
{'1': {'true': 16333, 'false': 6185}, '0': {'true': 16313, 'false': 6169}, 'total': {'1': 22518, '0': 22482}}
```

When the CART algorithm was tested with the 45.000 file and trained with the 135.000 file, the above were the results. From this info the precision that is true positive ('1': true) divided all positive ('1': true + '0': false) was of 72.58%.

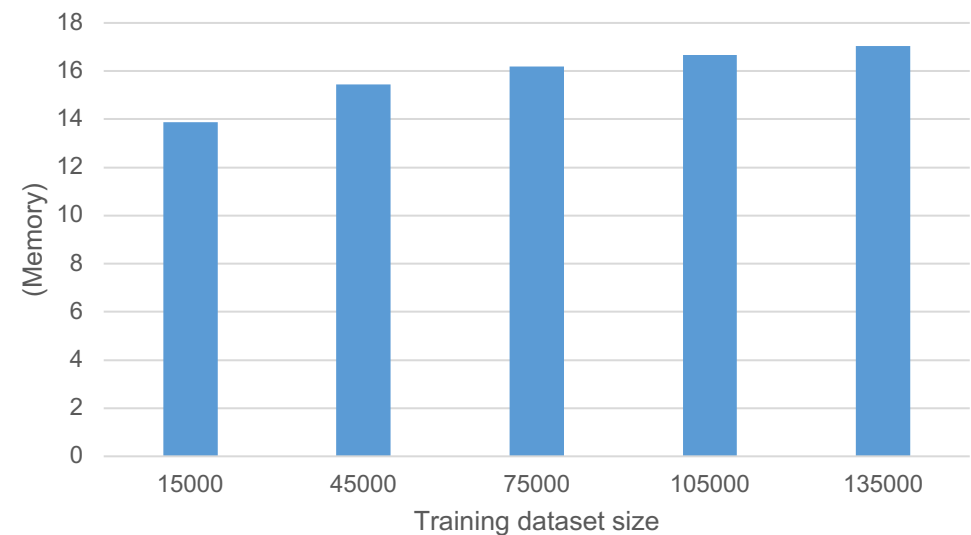
While the recall, that is the true positives ('1': true) divided the relevant elements ('total': 1) was of 72.53%.

And last but not least the accuracy that is the sum of true positives ('1': true) and true negatives ('0': true) divided the total of elements (45.000) was of 72.54%

	Testing data set
<b>Accuracy</b>	0.7254
<b>Precision</b>	0.7258
<b>Recall</b>	0.7253



Time Consumption



Memory Consumption





THANK YOU!