

PREDICTING STUDENT'S PERFORMANCE IN THE TEST “SABER PRO”



Team Presentation



Juan Pablo
Rincon



Julián
Gómez



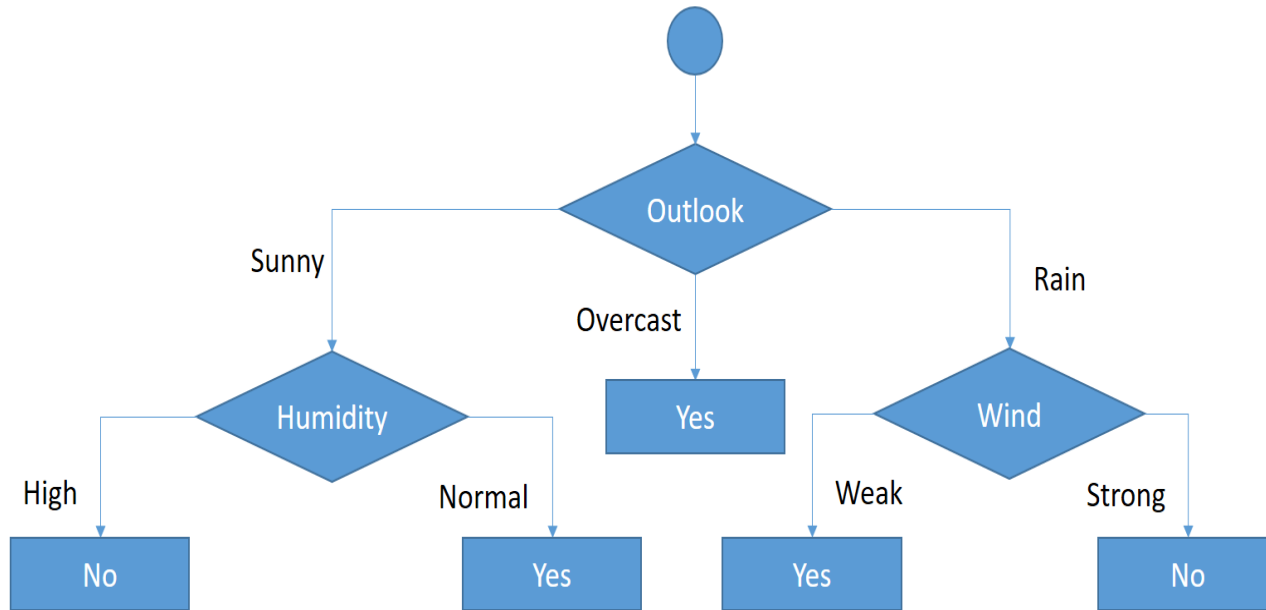
Miguel
Correa



Mauricio
Toro

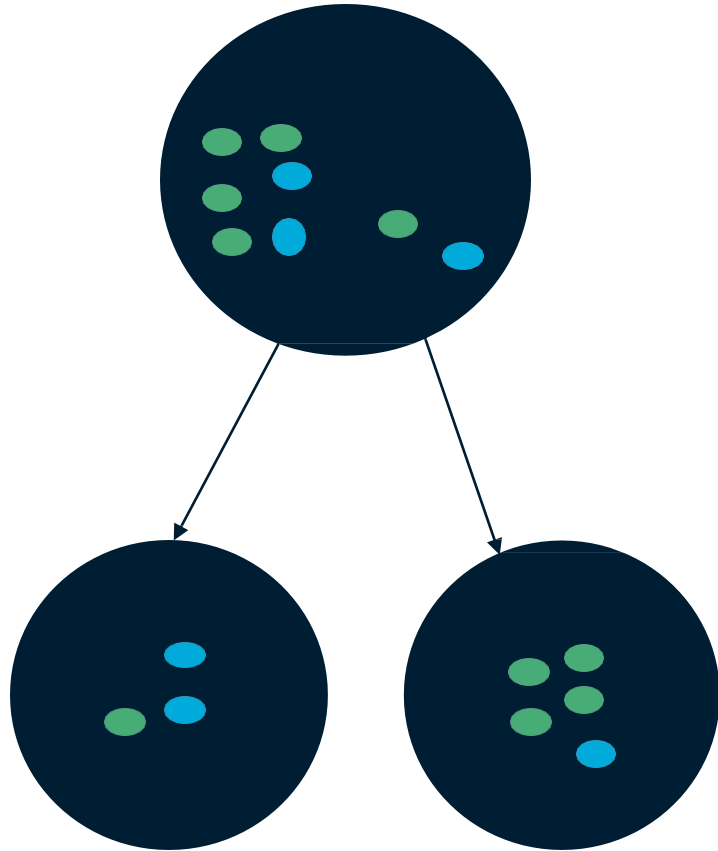


github.com/jgomezb11/ST0245-002/proyecto

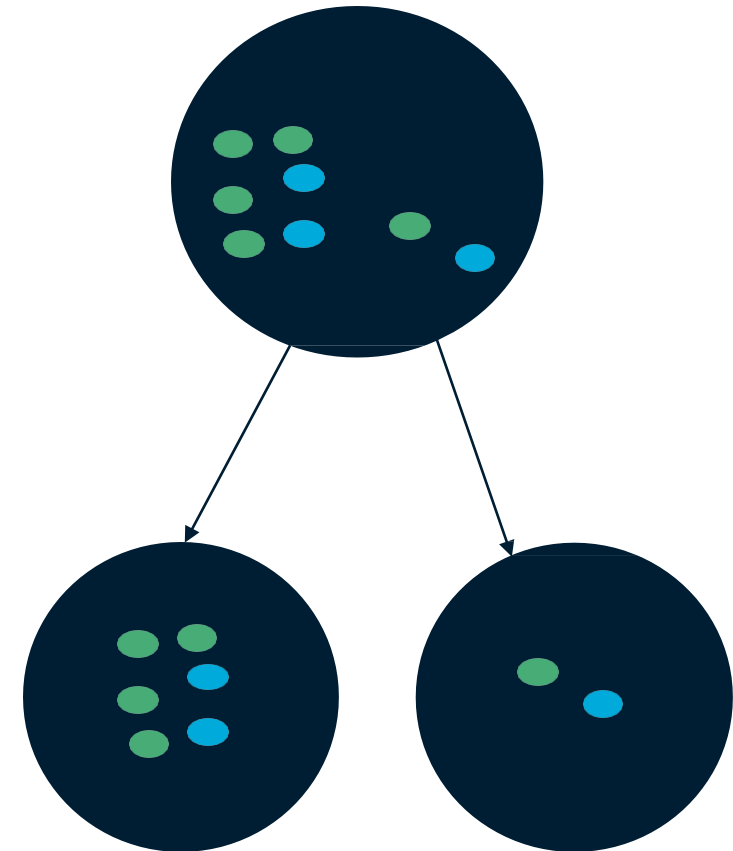


Algorithm to build a binary decision tree using *CART*. In this example, we show a model to predict whether or not to go shopping, according to weather.

Node Splitting



As an example, this split is based on the condition
“desemp_ingles == A-.”
For this case weighed Gini impurity is 0.38.



As an example, this split is based on the condition “Fam_numlibros
== “0 A 10 LIBROS”.”
For this case weighed Gini impurity is 0.48.

Algorithm Complexity

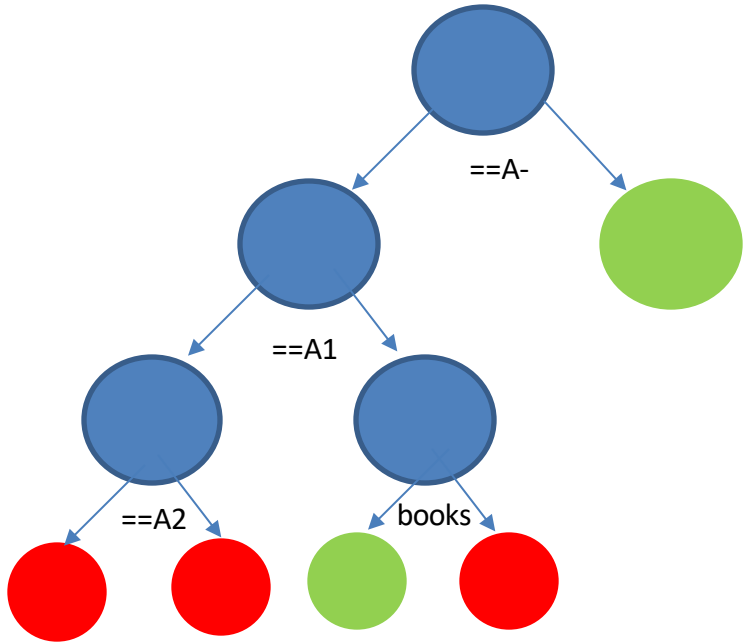


	Time Complexity	Memory Complexity
Training the model	$O(N^3 * M * \log(N))$	$O(N * M)$
Testing the Model	$O(N^3 * M * \log(N))$	$O(N * M)$

Time and memory complexity of the CART algorithm. Where N is the number of rows and M is the number of columns in the dataset



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 non-leafs nodes and red a low probability of success.

Most Relevant Features



English

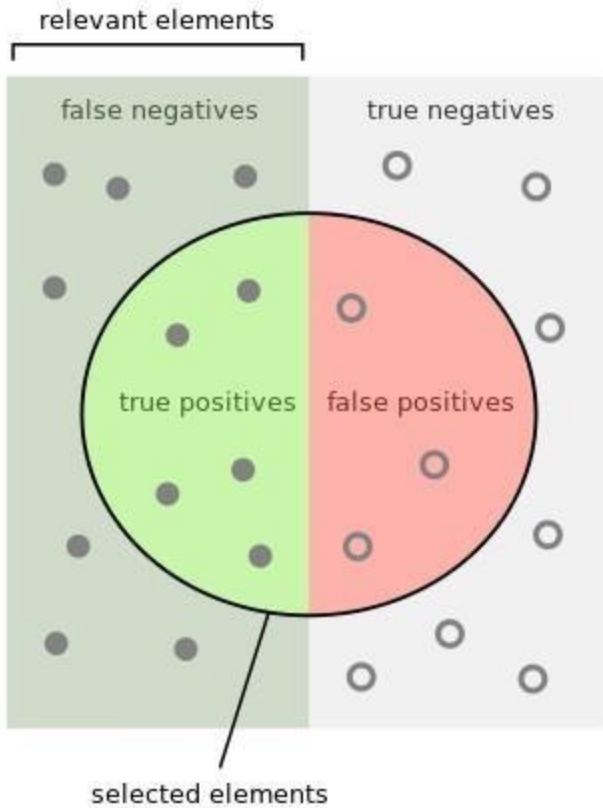


Number of
books

Evaluation Metrics

Keep this title

Complete this slide
For the third deliverable



Use vectorized figures to
explain the algorithm the evaluation
metrics, so they are not pixelated like mines

Use these
Colors for
Your figures

How many selected
items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant
items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

Explain Accuracy too...
In the same manner

If possible, avoid equations for
simple concepts that can be
explained through diagrams

Evaluation Metrics



Keep this title

*Complete this slide
For the third deliverable*



*Create the table in Powerpoint. Do not
copy pixelated screenshots from the
technical report please!*

	Training data set	Testing data set
Accuracy	0.8	0.62
Precision	0.6	0.55
Recall	0.76	0.61

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



*Explain the tables in your
own words*



*Include another HD picture related
to the example that you modeled
in the decision tree*

Time and Memory Consumption

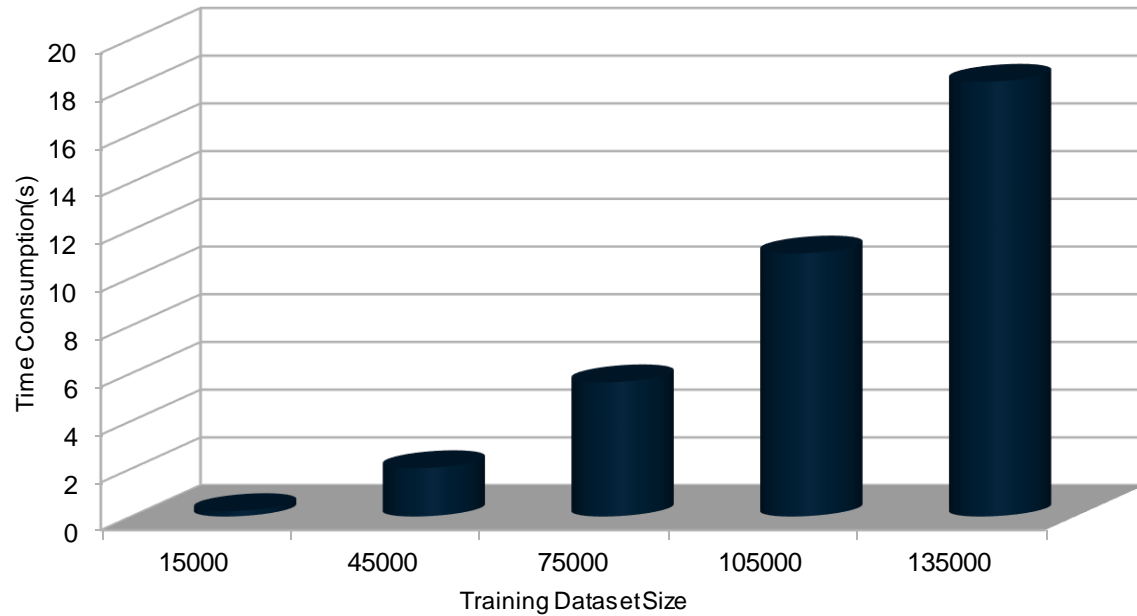


Keep this title

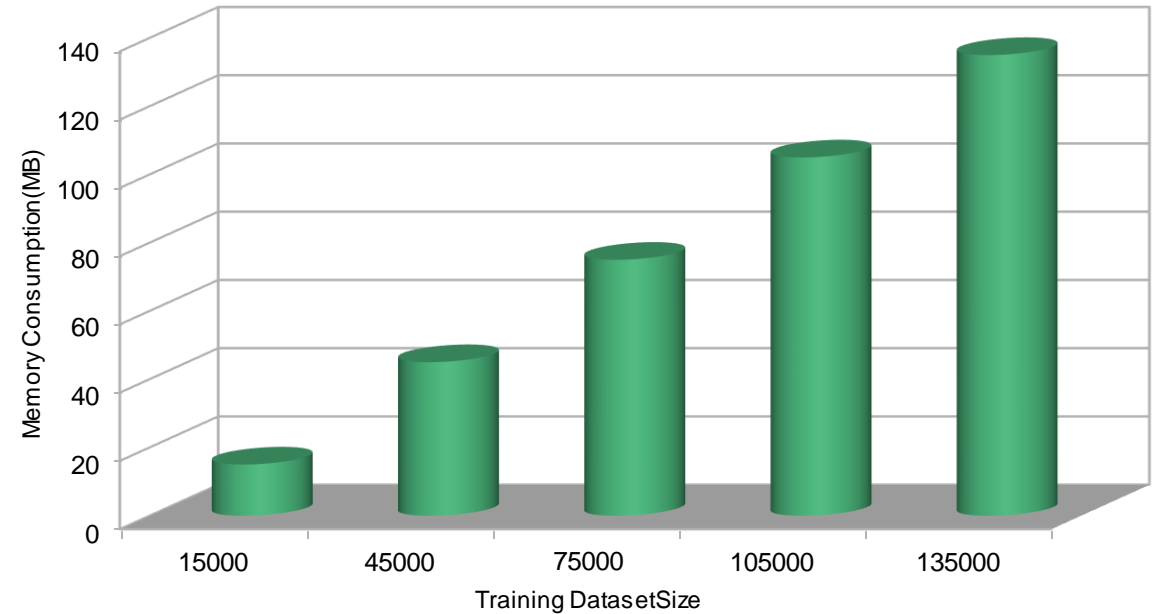
Complete this slide
For the third deliverable



Create the plots in Excel. Do not copy
pixelated screenshots from the technical
report please!



Time Consumption



Memory Consumption



Include the citation of the report
in arXiv and link

C. Patiño-Forero, M. Agudelo-Toro, and M. Toro. Planning system for deliveries in Medellín. ArXiv e-prints, Nov. 2016. Available at: <https://arxiv.org/abs/1611.04156>

Include a
screenshot

Cornell University

arXiv.org > cs > arXiv:1611.04156

Computer Science > Data Structures and Algorithms

[Submitted on 13 Nov 2016]

Planning system for deliveries in Medellín

Catalina Patiño-Forero, Mateo Agudelo-Toro, Mauricio Toro

Here we present the implementation of an application capable of planning the shortest delivery route in the city of Medellín, Colombia. We discuss the different approaches to this problem which is similar to the famous Traveling Salesman Problem (TSP), but differs in the fact that, in our problem, we can visit each place (or vertex) more than once. Solving this problem is important since it would help people, especially stores with delivering services, to save time and money spent in fuel, because they can plan any route in an efficient way.

Comments: 5 pages, 9 figures

Subjects: **Data Structures and Algorithms (cs.DS)**

ACM classes: F.2.0; G.2.2

Cite as: [arXiv:1611.04156](https://arxiv.org/abs/1611.04156) [cs.DS]
(or [arXiv:1611.04156v1](https://arxiv.org/abs/1611.04156v1) [cs.DS] for this version)



THANK YOU!