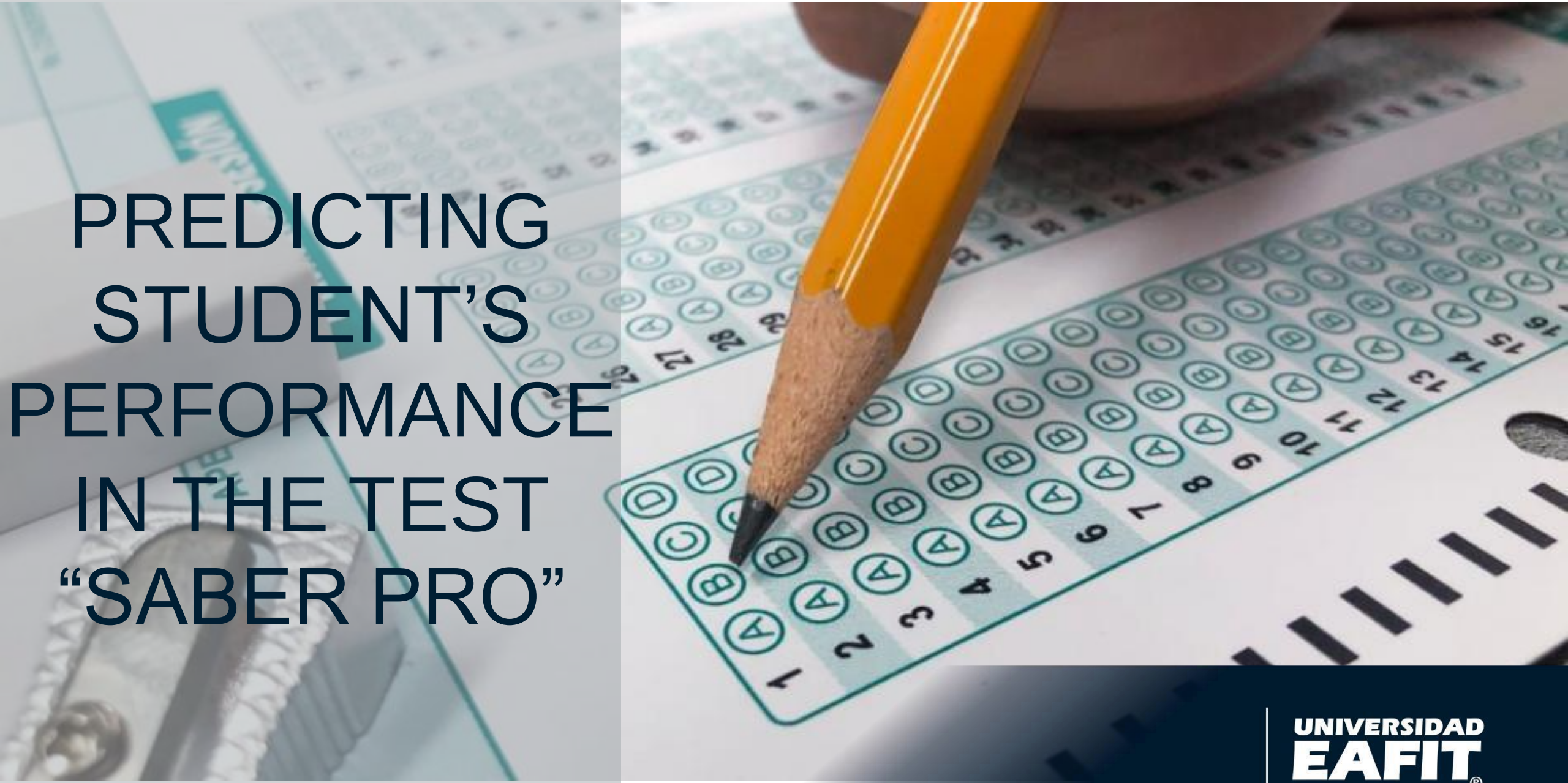
PREDICTING STUDENT'S PERFORMANCE IN THE TEST "SABER PRO"



UNIVERSIDAD EAFIT

Team Presentation

Juan Pablo Rincon

Julián Gómez

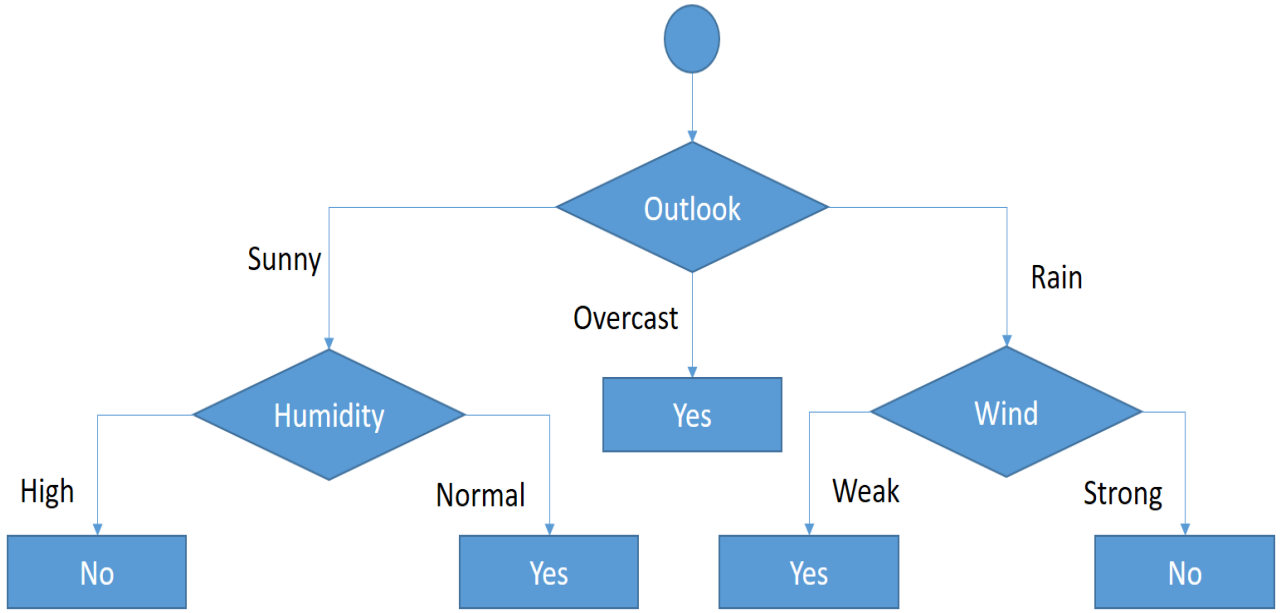
Miguel Correa

Mauricio Toro

github.com/jgomezb11/ST0245-002/proyecto

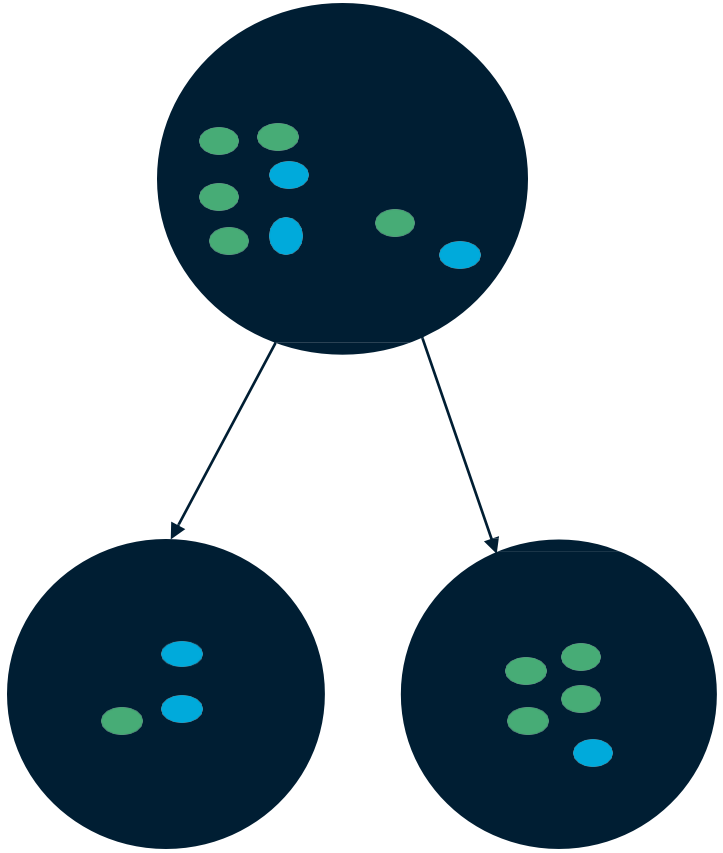
UNIVERSIDAD EAFIT

Algorithm Design

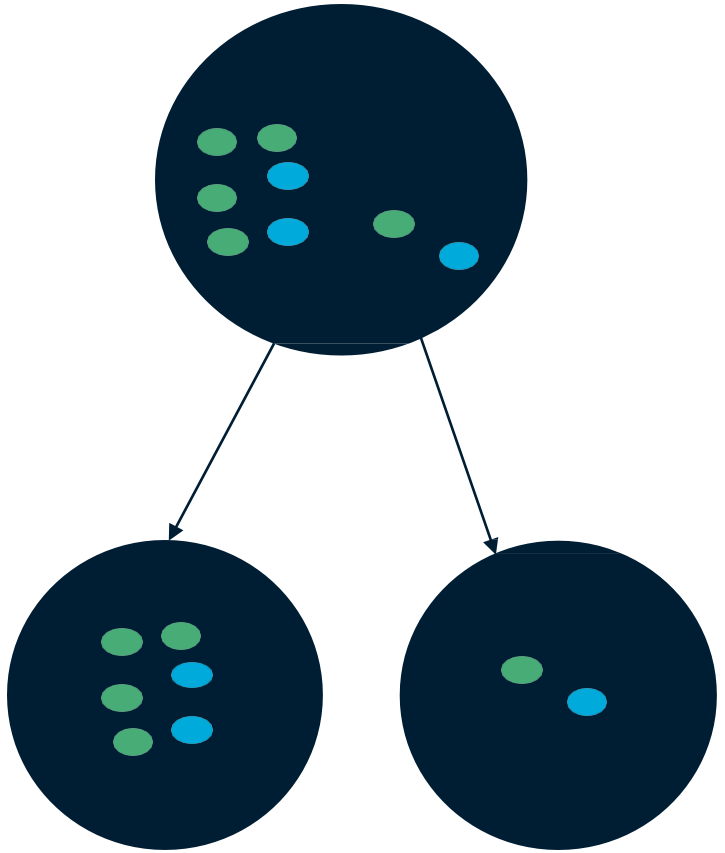


Algorithm to build a binary decision tree using *CART*. In this example, we show a model to predict whether or not to go shopping, according to weather.

Node Splitting



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



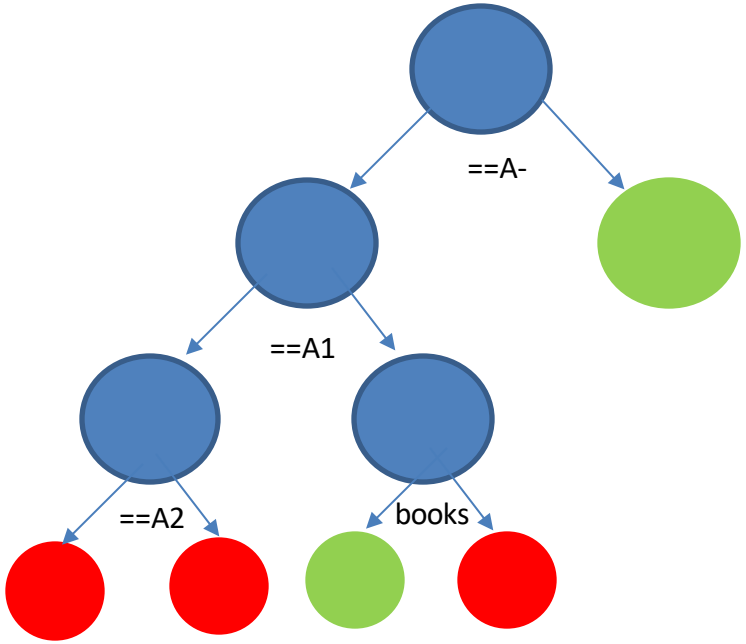
As an example, this split is based on the condition "Fam_numlibros == "0 A 10 LIBROS"."
For this case weigthed Gini impurity is 0.48.

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

EN English

Number of books

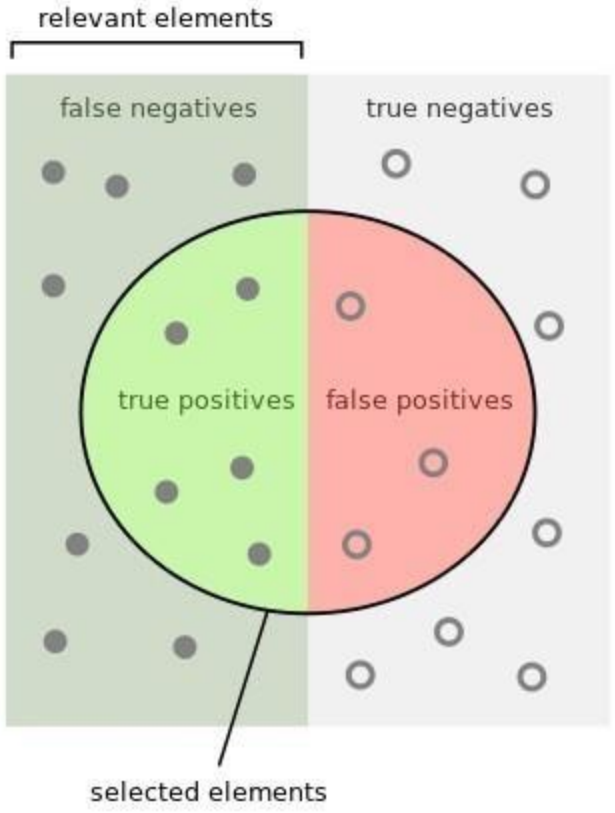
Evaluation Metrics

Keep this title

Complete this slide
For the third deliverable

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

Use these Colors for Your figures



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{selected elements}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{relevant elements}}$$

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

Explain Accuracy too…
In the same manner

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.

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

Explain the tables in your own words

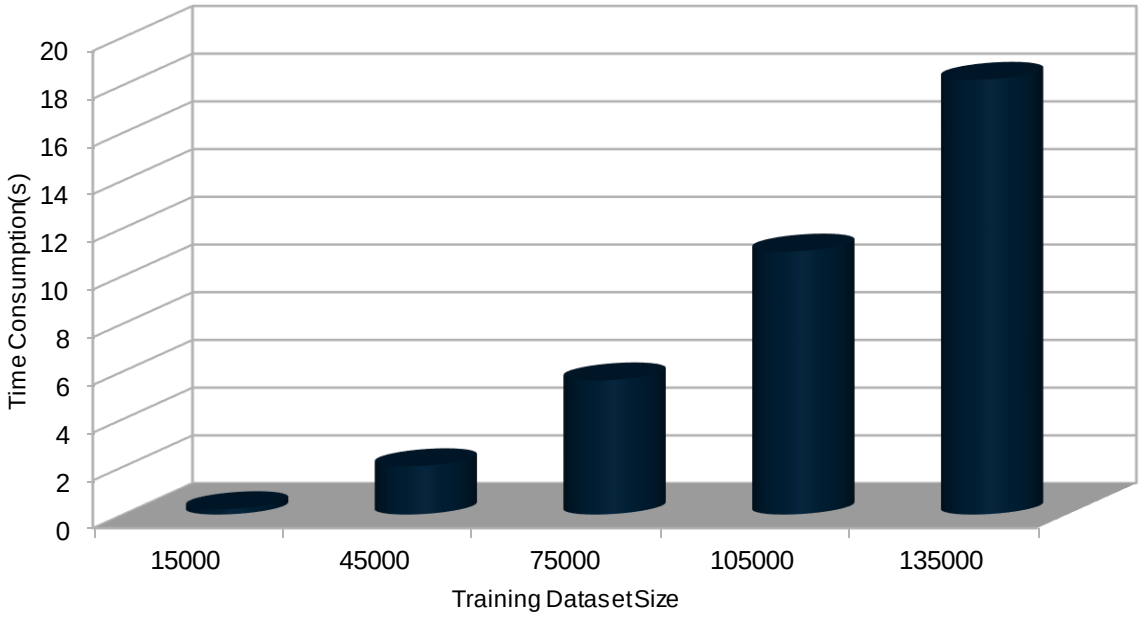
UNIVERSIDAD EAFIT

Time and Memory Consumption

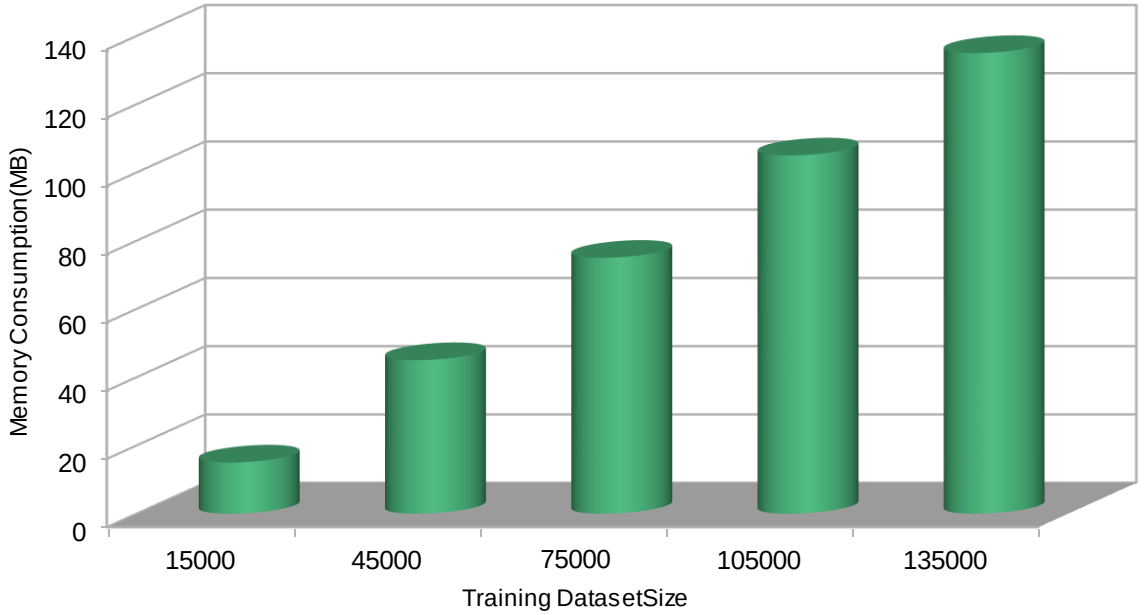
Keep this title

Complete this slide
For the thirddeliverable

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



Time Consumption



Memory Consumption

UNIVERSIDAD EAFIT

Report Accepted on arXiv

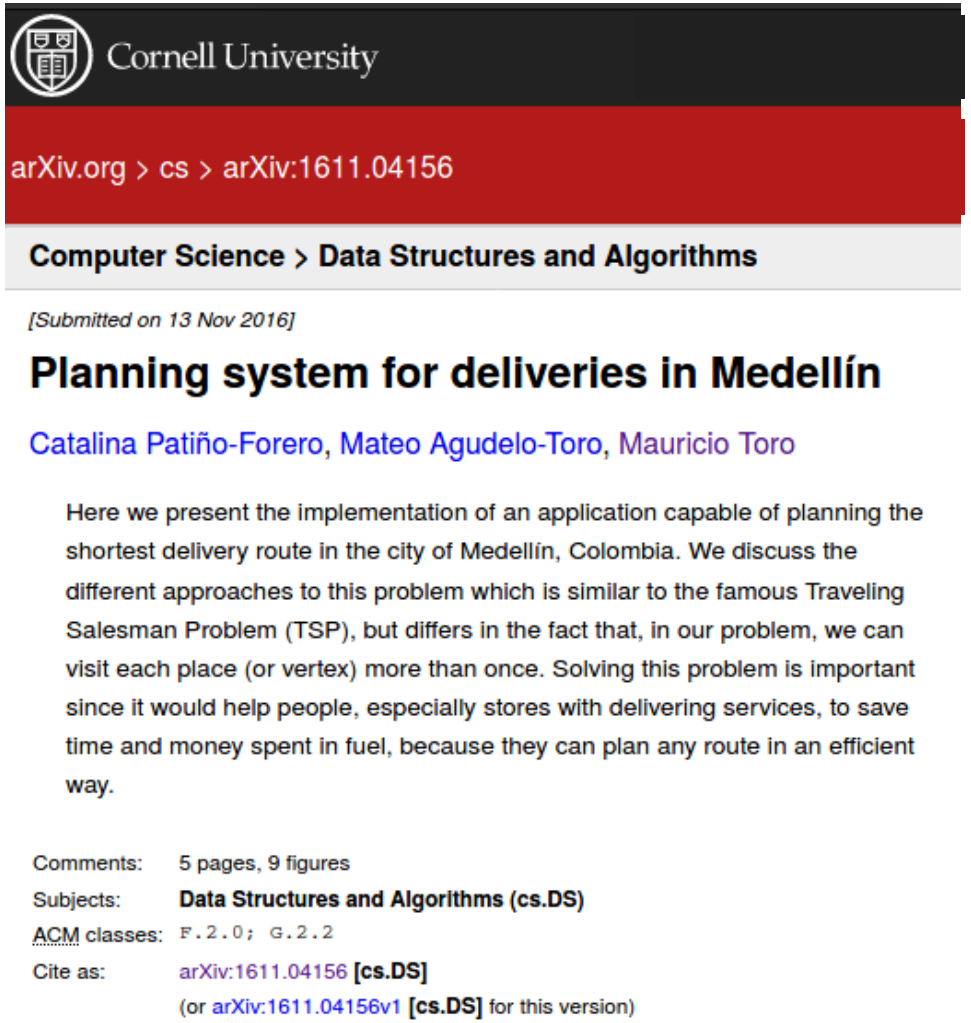
Keep this title

Complete this slide
For the third deliverable

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



UNIVERSIDAD EAFIT

THANK YOU!