PREDICTING ACADEMIC SUCCESS USING DECISION TREES

UNIVERSIDAD EAFIT

Team Presentation

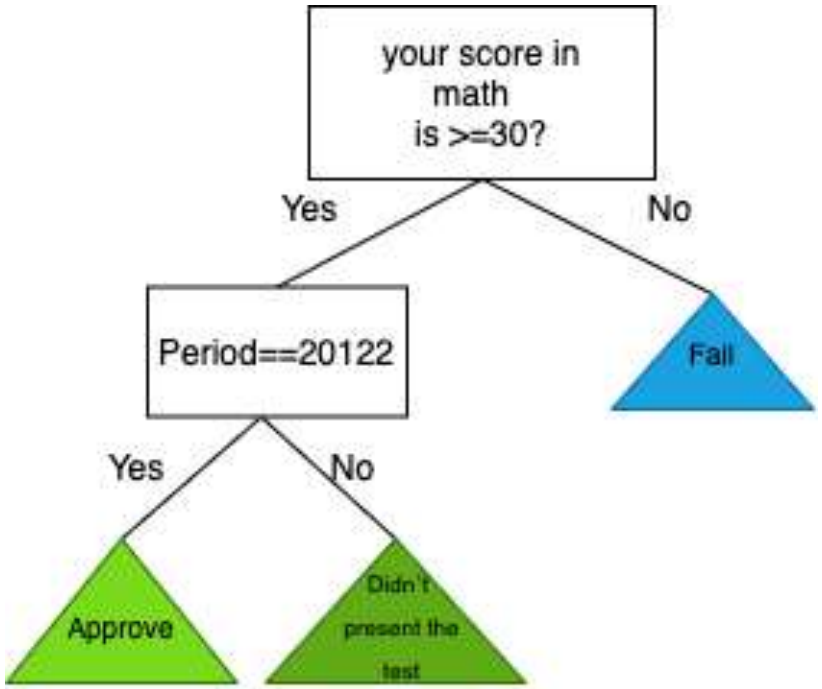
Paulina Ocampo

Maria José Gutiérrez

Miguel Correa

Mauricio Toro

Algorithm Design

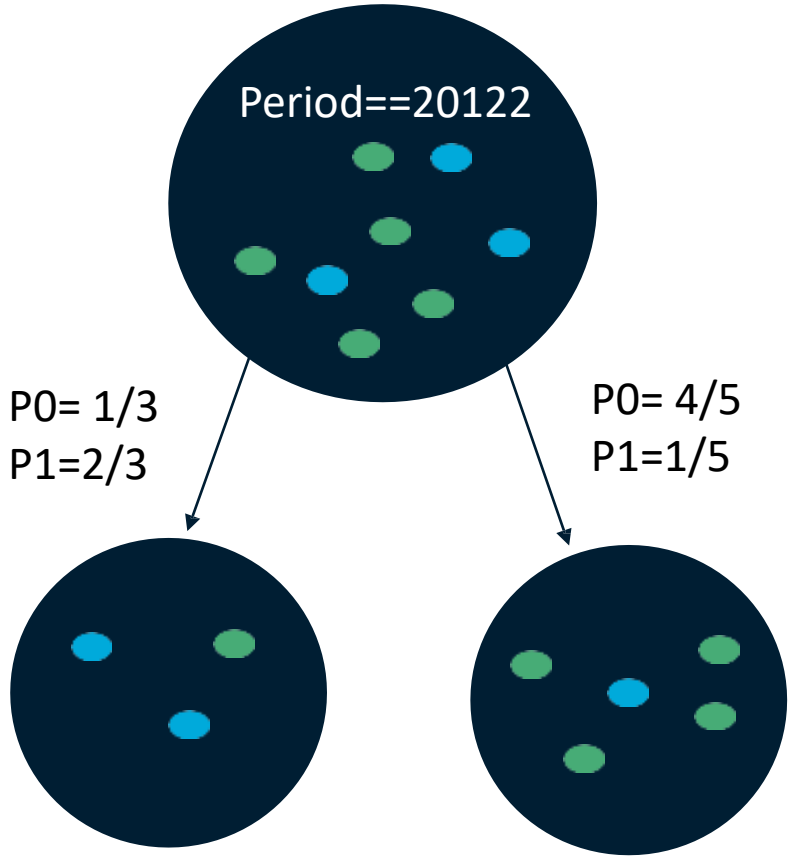


We focused on decision trees to give a solution about the prediction results in PRO knowledge test, we omit some variables to avoid discrimination. In this example, we show a CART algorithm model to predict if a student can approve an exam or not based on their score and the period in which the ICFES test were performed.

Took from: http://clipart-library.com/maths-examination-cliparts.html

Node Splitting

WI=(0,44·3)+(0,32·5)/3+5
WI= 0,365



LI=1-((1/3)^2 + (2/3)^2)=1-(5/9)
LI= 0,44

RI= 1-((4/5)^2 + (1/5)^2)=1-(17/25)
RI= 0,32

As an example, this split is based on the condition “period == 20122.” For this case, left Gini impurity is 0.44, right Gini impurity is 0.32 and weighted Gini impurity is 0.365. It help us to find the minimum percentage to know the success for each node.

Algorithm Complexity

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!

	Time Complexity	Memory Complexity
Training the model	$O(N^2*M*2^M)$	$O(N*M*2^M)$
Testing the Model	O(N*M)	O(1)

Time and memory complexity of the (In this semester, one could be CART, ID3, C4.5… please choose) algorithm. (Please explain what do N and M mean in this problem. PLEASE DO IT!

Explain the tables in your own words

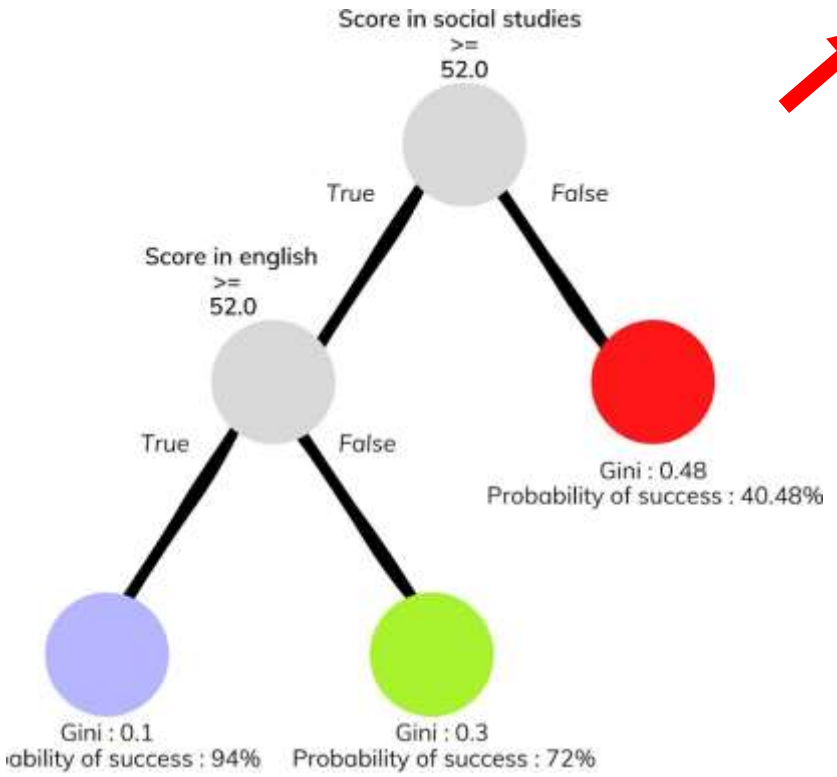
Include another HD picture related to the example that you modeled in the decision tree

UNIVERSIDAD EAFIT

Decision-Tree Model

Keep this title

Complete this slide
For the third deliverable



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

A binary decision tree to predict Saber Pro scores based on the results of Saber 11. Violet nodes represent those with a high probability of success, green medium probability and red a low probability of success.

Explain the Figures in your own words

Most Relevant Features

- Social Studies
- English
- Gender

Use an icon for each feature!

Is it ethical to make a model that predicts academic success based on gender?

UNIVERSIDAD EAFIT

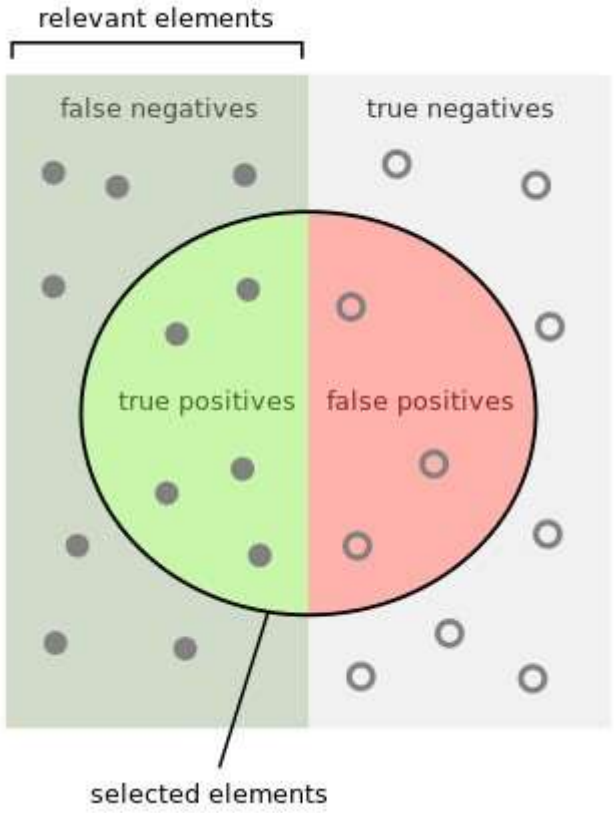
Evaluation Metrics

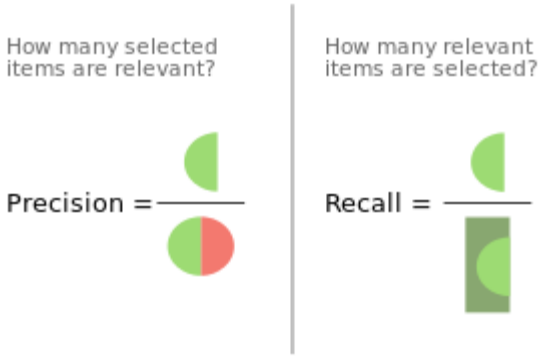
Keep this title

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





Explain Accuracy too…
In the same manner

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

UNIVERSIDAD EAFIT

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

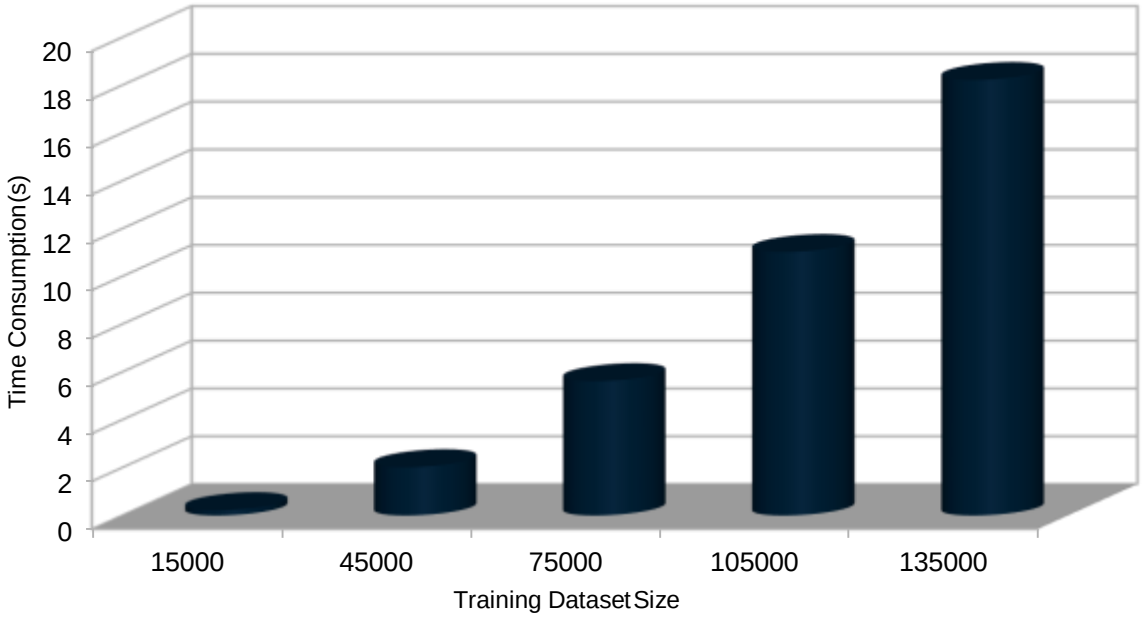
UNIVERSIDAD EAFIT

Time and Memory Consumption

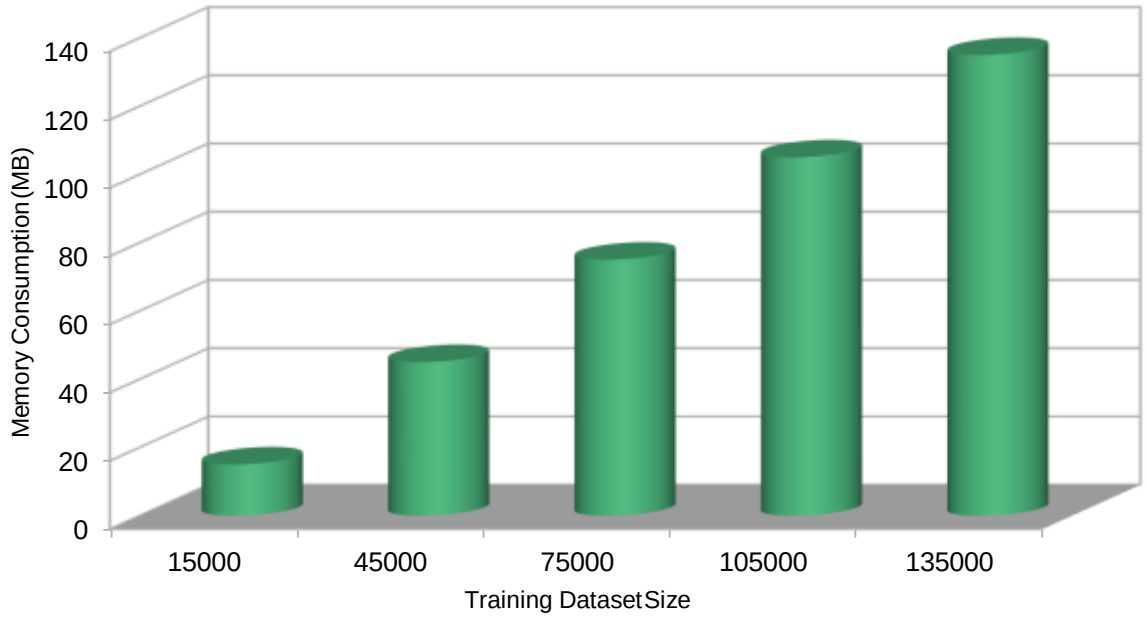
Keep this title

Complete this slide
For the third deliverable

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

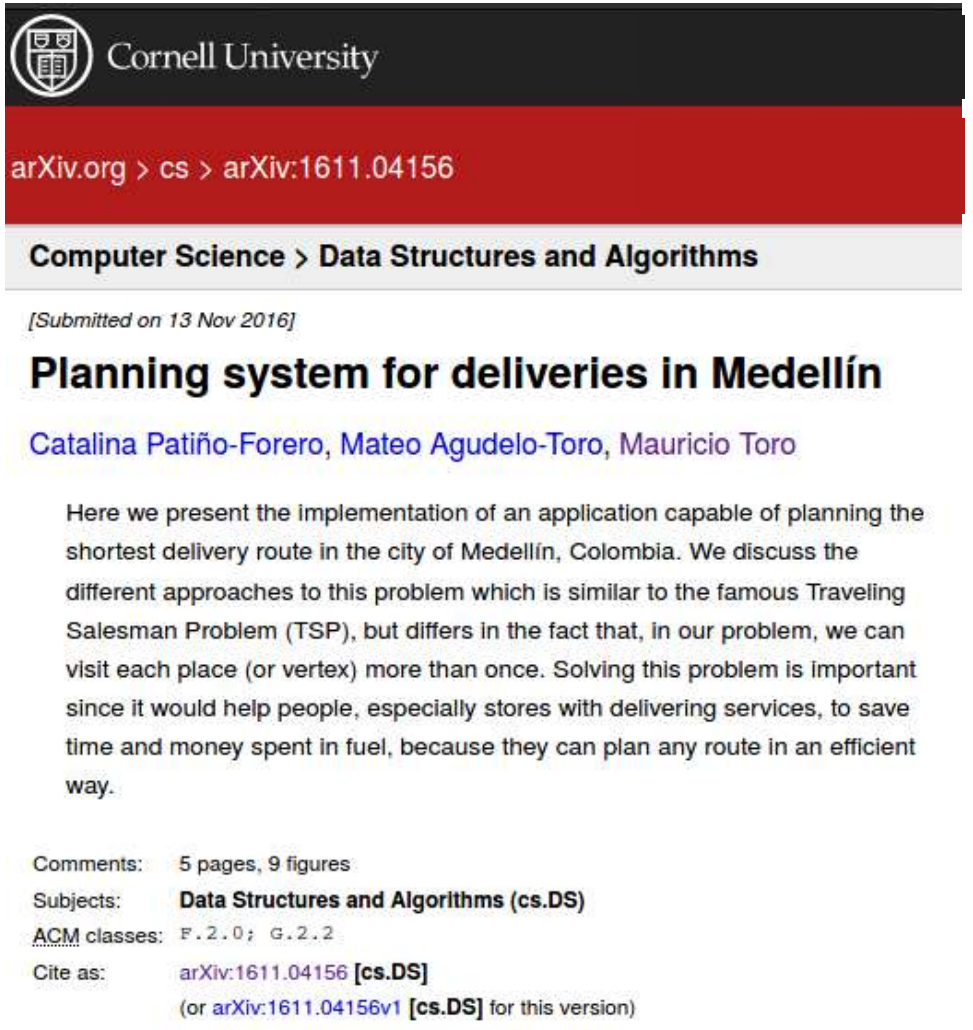
Keep this title

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



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 **[cs.DS]**
(or arXiv:1611.04156v1 **[cs.DS]** for this version)

UNIVERSIDAD EAFIT

Complete this slide
For the third deliverable

THANK YOU!

Say thank you for listening!