
Welcome to Data Science Online Bootcamp

Know more about Decision Trees



Democratizing Data Science Learning

Learning Objectives

Decision Trees



Some Important Updates

- Some updates have been made to the code from yesterday's session, please **download the updated code** below:

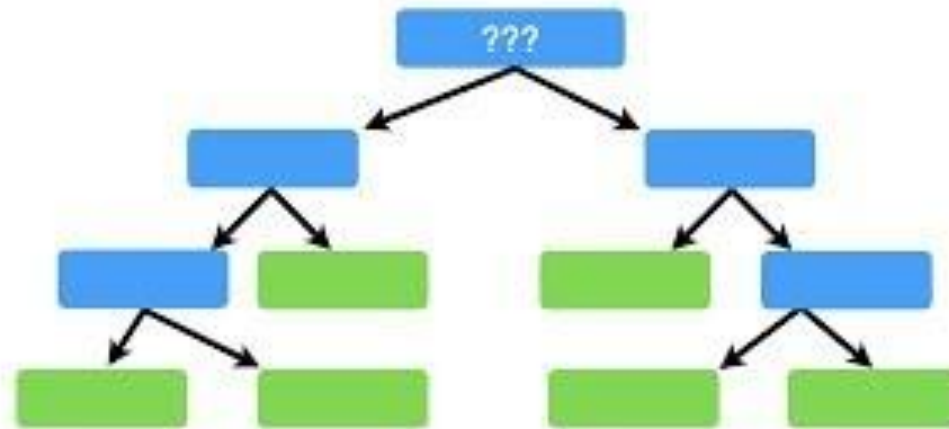
https://github.com/dphi-official/First_ML_Model

Understanding Decision Trees

- The following video explains clearly what a decision is and how it works with bunch of live examples.
- Alongside it will talk about some mathematical aspects such as gini impurity. It would be good to understand them in long run, though you may not want to overwhelm yourself with too many jargons at the moment.

Understanding Decision Trees

Decision Trees...



...clearly explained!

Building Decision Tree from Scratch



The diagram illustrates a decision tree for classifying fruits based on color and diameter. The root node asks "Is diameter ≥ 3 ?".

Training Data:

Color	Diam	Label
Green	3	Apple
Yellow	3	Apple
Red	1	Grape
Red	1	Grape
Yellow	3	Lemon

Decision Tree Structure:

- Root Node: Is diameter ≥ 3 ?
 - False branch: R 1 Grape, R 1 Grape
 - True branch: G 3 Apple, Y 3 Apple, Y 3 Lemon

Results:

- False branch: Apple 100%
- True branch: Predict Apple 50%, Lemon 50%

{ML} Let's Write a Decision Tree from Scratch

Decision Tree Resources

- A simple non-technical explanation of decision tree:
<https://towardsdatascience.com/a-beginners-guide-to-decision-tree-classification-6d3209353ea>
- A comprehensive article on decision trees that talks about optimisation criteria, pros and cons etc:
<https://www.datacamp.com/community/tutorials/decision-tree-classification-python>

Decision Tree Resources

- **Optional reading**, to understanding several aspects of decision trees:

<https://medium.com/greyatom/decision-trees-a-simple-way-to-visualize-a-decision-dc506a403aeb>

- [https://medium.com/@chiragsehra42/decision-trees-explained-easily-28f23241248#:~:text=Decision%20Trees%20\(DTs\)%20are%20a,and%20the%20fitter%20the%20model](https://medium.com/@chiragsehra42/decision-trees-explained-easily-28f23241248#:~:text=Decision%20Trees%20(DTs)%20are%20a,and%20the%20fitter%20the%20model)

Let's do some practice

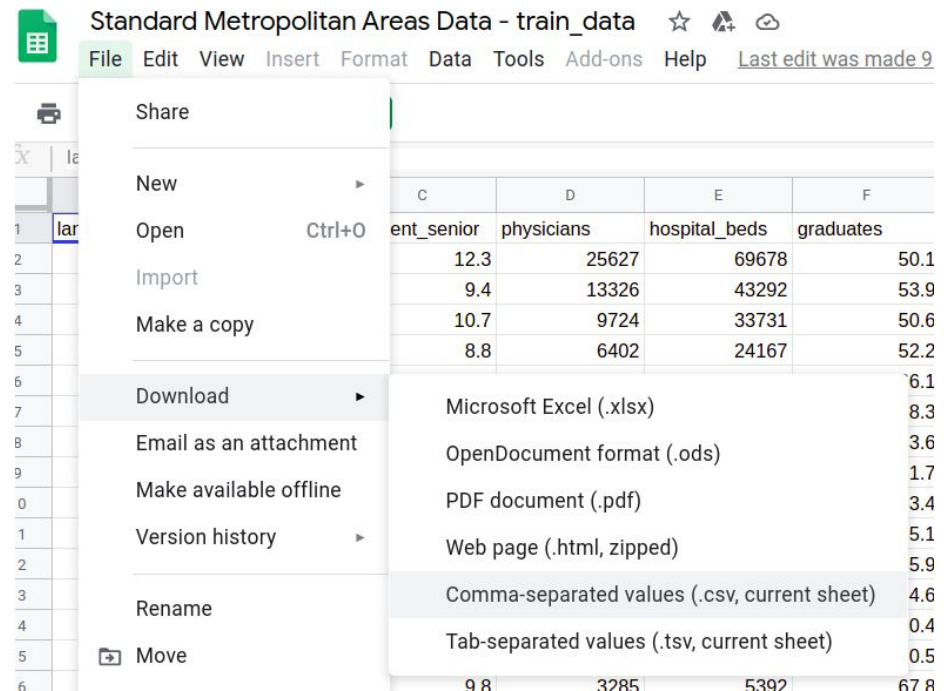
- Now that we know how to build a simple decision tree model on one of the datasets that we used previously **(dataset details are given in the next slide)**
- **Objective:** Imagine you were hired as a Data Scientist/Analyst by the Central Police Office of Metropolitan areas and your job is now to create a predictive model for `crime_rate`. For now, let's focus on building a simple decision tree.

Dataset

In this exercise, we'll be exploring the Standard Metropolitan Areas Dataset in depth.

Link to the Dataset: https://bit.ly/SMA_Dataset

Download the dataset as a CSV file.



Dataset Description

It contains data of 99 standard metropolitan areas in the US. The data set provides information on 10 variables for each area for the period 1976-1977. The areas have been divided into 4 geographic regions: 1=North-East, 2=North-Central, 3=South, 4=West. The variables provided are listed in the table below:

Variable name	Description	
land_area	size in square miles	
total_population	estimated population in thousands	
percent_city	percent of population in central city/cities	
percent_senior	percent of population ≤ 65 years	
physicians	number of professionally active physicians	
hospital_beds	total number of hospital beds	
graduates	percent of adults that finished high school	
work_force	number of persons in work force in thousands	
income	total income in 1976 in millions of dollars	
crime_rate	Ratio of number of serious crimes by total population	
region	geographic region according to US Census	



What next?

If you were able to build a model then don't forget to share it here along with your inferences with fellow learners:

<https://dphibootcamp.slack.com/archives/C014J5W1KUG>

Download link to slides

Download here:

<https://docs.google.com/presentation/d/1McLDUQ-t3eL0Y7zvpFPRLzseH5Q3DV9N2F99Cd9boe4/edit#slide=id.p1>

That's it for the day. Thank you!

Feel free to post any queries in the #help channel on Slack



Democratizing Data Science Learning