

STA 235 - JITT 9

Please answer before Sunday Apr 18th, 11:59 PM

* Required

1. Email *

2. EID *

Decision Trees

Imagine we are working with the simple example we have seen in class, where you are working for Disney+ and want to predict whether someone is going to cancel their subscription or not. To simplify the example, we have 200 customers and we only have two variables for each:

1. Mandalorian: Takes a value of 1 if the customer has finished the Mandalorian and 0 if not.
2. Log: Takes a value of 1 if the customer has logged into the platform in the past week and 0 if not.

C (in red) represents those that cancelled the service, and NC (in blue) those that have not.

We want to build a simple decision tree, so you will have to do some simple calculations!

3. Using the table below, which variable would you use for splitting first and why? Explain some of your calculations! (Note: You need to use the data below for this). *

		Log	
		Yes	No
Mandalorian	Yes	15 C 75 NC	50 C 10 NC
	No	5 C 25 NC	10 C 10 NC

4. What is the main disadvantage of having a tree with multiple splits? *

Feedback and questions

This section is optional. Please answer truthfully to help improve your experience!

5. Did you find the last lecture...

Mark only one oval.

☐ Too fast-paced

☐ Pace was ok

☐ Too slow-paced

6. Do you have any questions related to the material we covered in the previous classes?

This content is neither created nor endorsed by Google.

Google Forms