Colin Pham

Professor Brooks

Foundations of AI

13 November 2024

Assignment 5: Working with Uncertainty

Problem 1: Decision Trees in scikit-learn

Run RandomForestClassifer on your dataset with 10, 25, and 50 estimators using both gini and entropy as separators

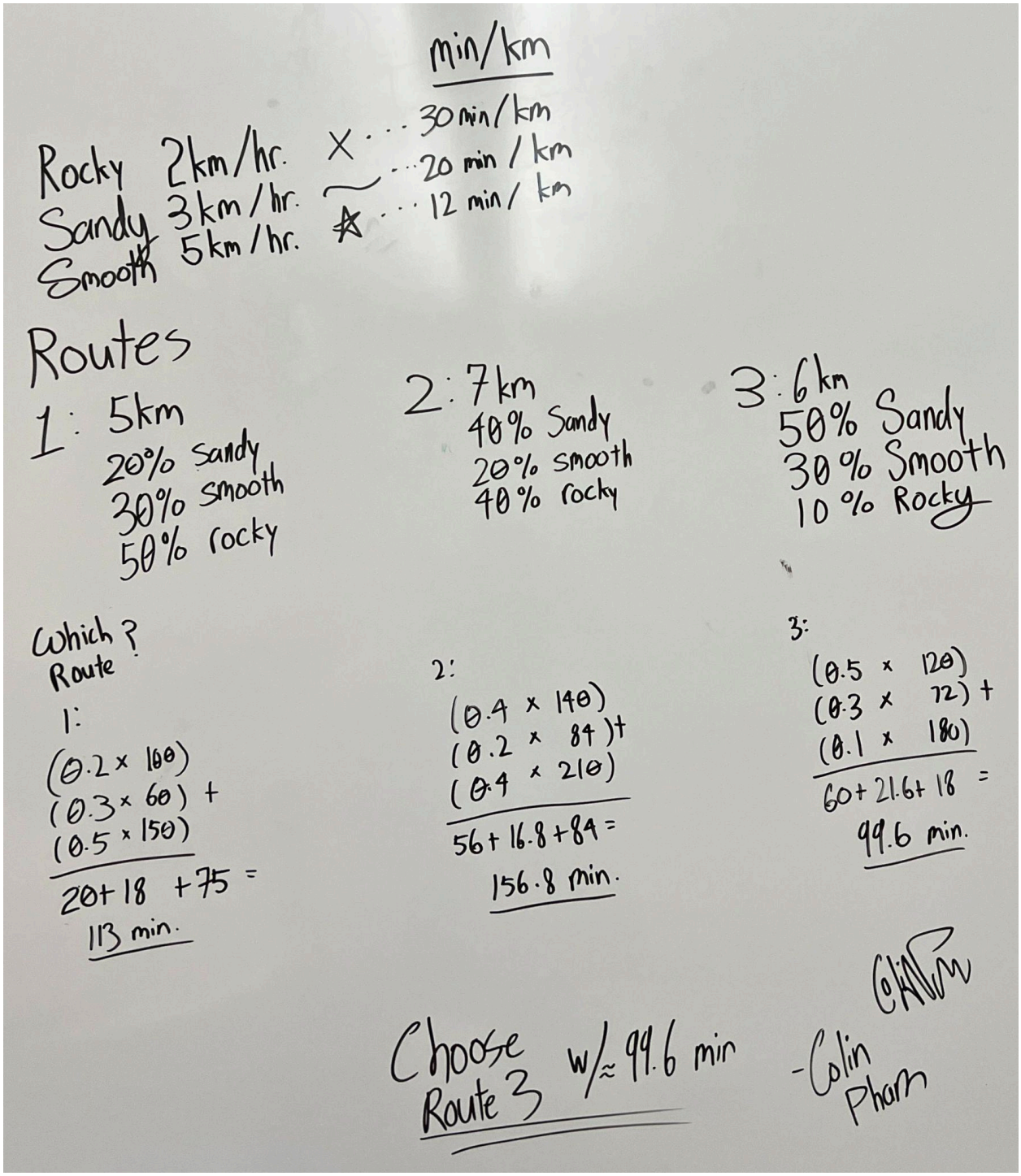
Legend: N = n_estimators. S = Separators. Can be G for Gino or E for Entropy

N:10 S:G	N:10 S:E	N:25 S:G	N:25 S:E	N:50 S:G	N:50 S:E
0.9083333333333333	0.8888888888888888	0.9333333333333333	0.9111111111111111	0.9305555555555556	0.95
0.8416666666666667	0.9083333333333333	0.9027777777777778	0.8833333333333333	0.9111111111111111	0.9083333333333333
0.9303621169916435	0.9303621169916435	0.9526462395543176	0.9610027855153204	0.9637883008356546	0.9526462395543176
0.9359331476323120	0.9331476323119777	0.9582172701949860	0.9693593314763231	0.9610027855153204	0.9693593314763231
0.8941504178272981	0.8913649025069638	0.9052924791086351	0.9220055710306406	0.9192200557103064	0.9359331476323120

Doing hyperparameter search by hand is annoying. Sklearn makes it easy for us with the GridSearchCV class. This lets us provide a list of models and hyperparameters, and it does cross-validation for each model and parameter combination and summarizes the result. Very handy!

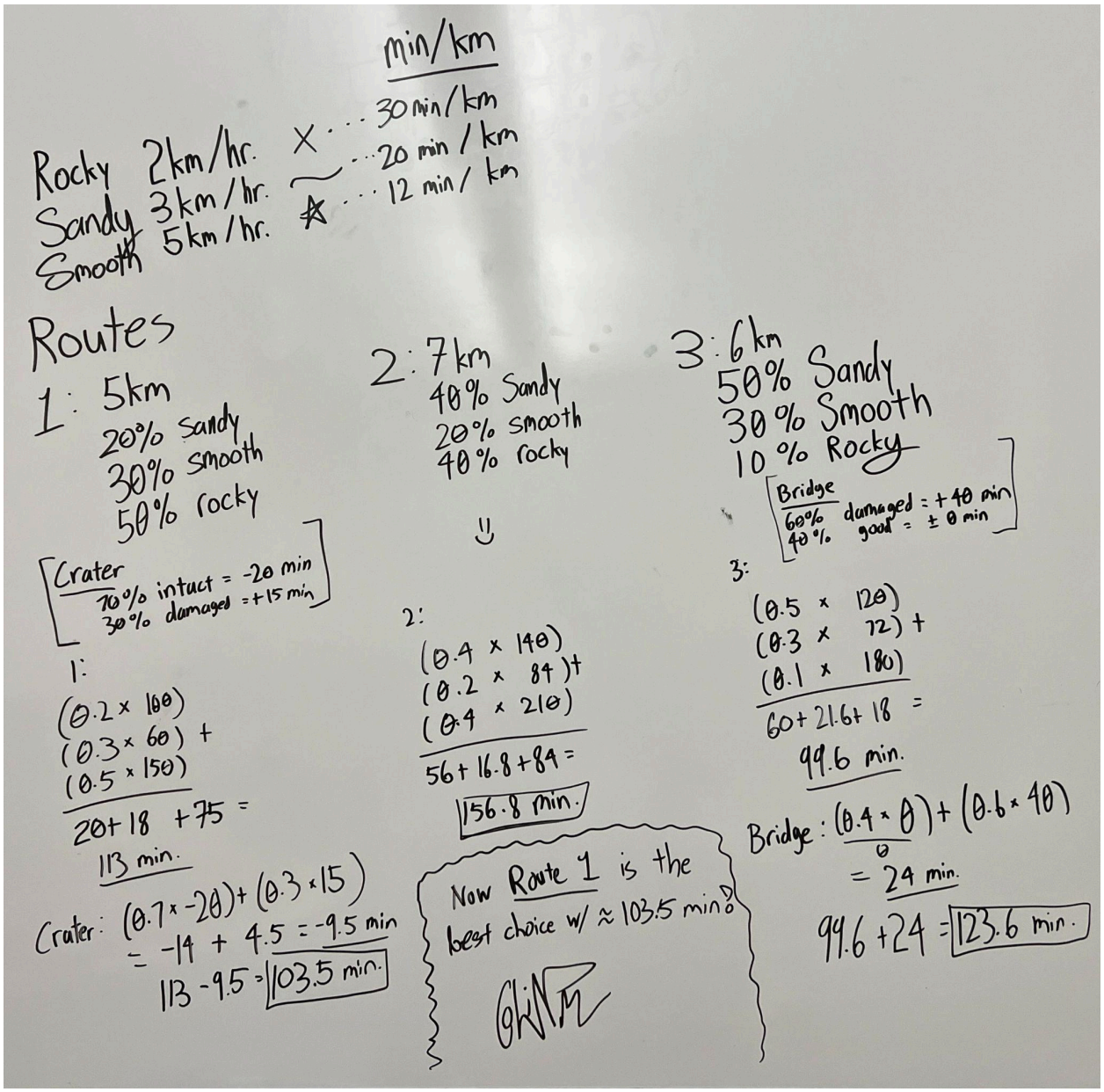
Woah! Look at that forest go (you can't because it's so fast)!

Problem 4: Utility



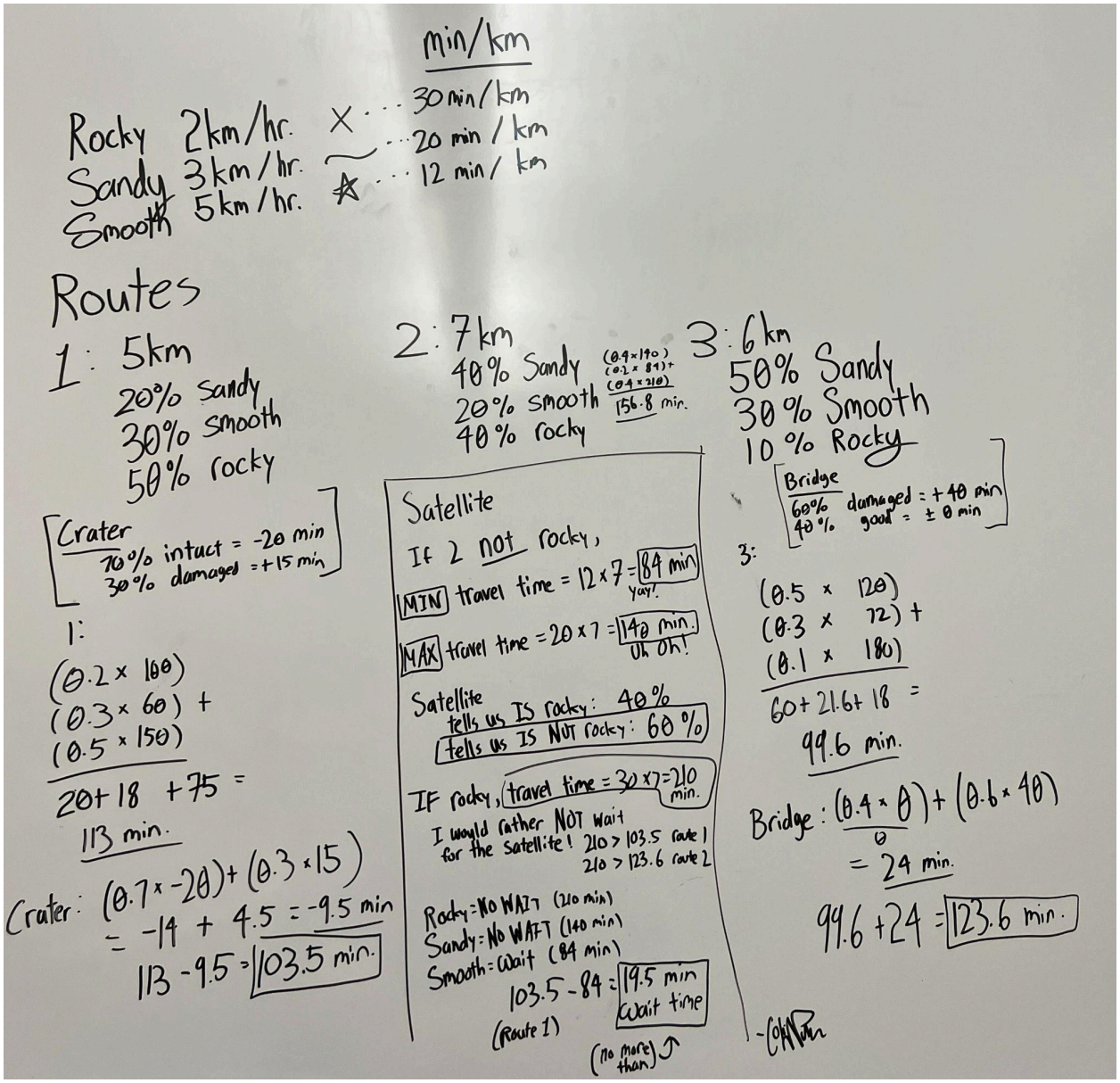
Which route should we pick?

- Route 3's odds outputted a calculated time of 99.6 minutes, which is less than the other routes of time 113 minutes and 156.8 minutes. **Route 3 is the best**



Now which route seems best?

- With the crater and the bridge cases for routes 1 and 3 respectively, Route 1's time decreased by 9.5 minutes and route 3's time increased by 24 minutes. **Route 1 became the more favorable route with a time of 103.5 minutes**



First: If the satellite said that route 2 was not rocky, how long would we expect it to take?

- If route 2 is NOT rocky, it can either be smooth or sandy. **A smooth path gives us an 84 minutes of travel while a sandy path gives us 140 minutes of travel**

Second: What's the probability that the satellite will tell us this?

- A satellite telling us the path is NOT rocky would be the probability of smooth (20) plus the probability of sandy (40) equalling to **60%**

Third: If the satellite tells us route 2 is in fact rocky, what do we do? How long will that take?

- If route 2 IS rocky, then the travel time will be **210 minutes**. How awful! Literally any other route would be a better choice at this point. **I would take route 1 at that point**

Last: given all of this, how long should we wait for the satellite?

- The satellite telling us it is NOT rocky and taking our chances to get a smooth path, our travel time would be 84 minutes. Taking that time and subtracting it from the next best path, route 1, of time 103.5, we would have to wait **no more than 19.5 minutes for a satellite response**. If the time exceeds, take route 1.

- However, there is STILL a possibility that the path can be sandy, making it still slower than route 1. Some more math has to be done here. After the 19.5 minute wait time: (0.66666 x 140) + (0.3333333 x 84) = 121.332 minutes. (⅔ chance for sandy and ⅓ chance for smooth) Wow! **Route 1 is still technically better**. Man. I would take route 1.