Problem 1 Thanh Cong Nguyen

Part 2

n_estimators	criterion	mean_accuracy
10	gini	0.870000
25	gini	0.943810
50	gini	0.932857
10	entropy	0.886825
25	entropy	0.921587
50	entropy	0.910317

Part 3

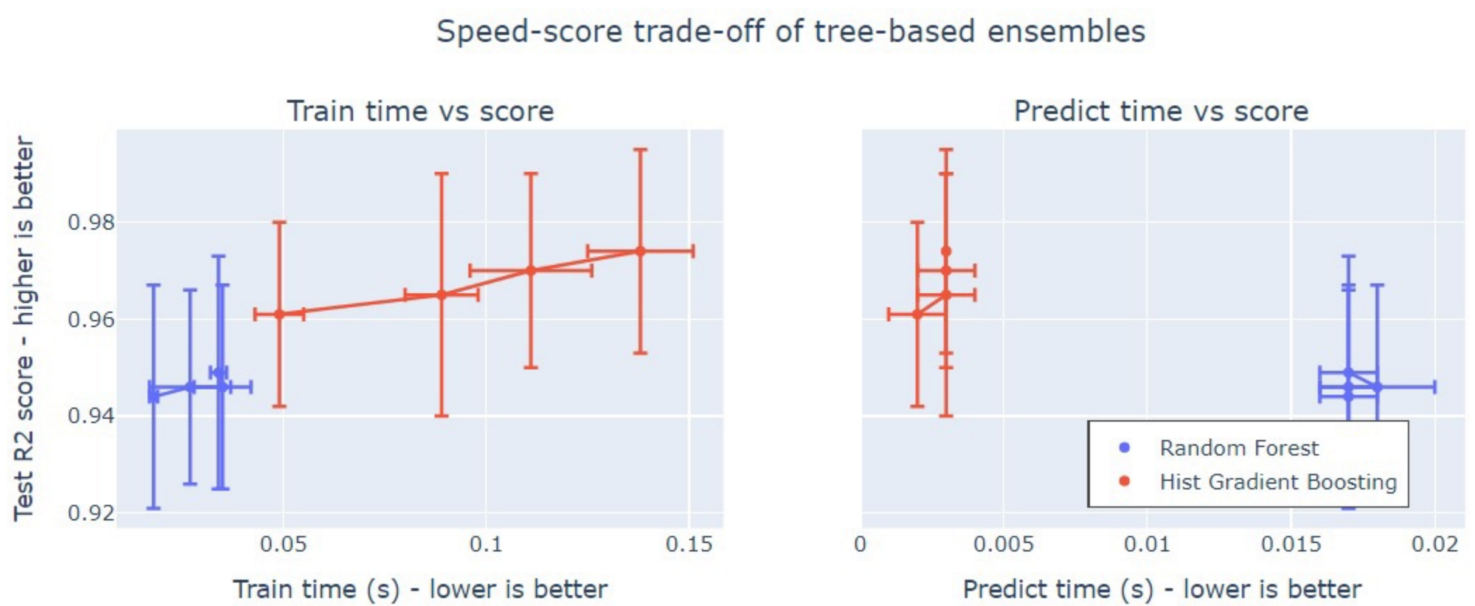
Model: Random Forest

	mean_fit_time	std_fit_time	...	mean_train_score	std_train_score
0	0.019624	0.002558	...	0.971002	0.005630
1	0.034013	0.008302	...	0.980226	0.005208
2	0.037480	0.007028	...	0.981986	0.001647
3	0.034993	0.005786	...	0.982866	0.002143

Model: Hist Gradient Boosting

	mean_fit_time	std_fit_time	...	mean_train_score	std_train_score
0	0.049333	0.002617	...	0.989456	0.002557
1	0.088954	0.005264	...	1.000000	0.000000
2	0.105162	0.012207	...	1.000000	0.000000
3	0.147699	0.014340	...	1.000000	0.000000

Plots



Speed-score trade-off of tree-based ensembles
Train time vs score
Predict time vs score
Test R2 score - higher is better
0.98
0.96
0.94
0.92
0.05
0.1
0.15
Train time (s) - lower is better
0
0.005
0.01
0.015
0.02
Predict time (s) - lower is better
Random Forest
Hist Gradient Boosting

Thanh Cong Nguyen

Problem 4:

Rocky terrain: 2 Km/h = 60 / 2 = 30 mins/Km
Sandy terrain: 3 Km/h = 60 / 3 = 20 mins/Km
Smooth terrain: 5 Km/h = 60 / 5 = 12 mins/Km

Part 1: Route 1 = $(0.2 * 20 + 0.3 * 12 + 0.5 * 30) * 5$
$= 22.6 * 5 = 113$ mins

Route 2 = $(0.4 * 20 + 0.2 * 12 + 0.4 * 30) * 7$
$= 22.4 * 7 = 156.8$ mins

Route 3 = $(0.5 * 20 + 0.4 * 12 + 0.1 * 30) * 6$
$= 17.8 * 6 = 106.8$ mins

=> We should pick route 3

Part 2

Route 1: If wall intact (70% chance): $113 - 20 = 93$ mins

If wall damaged (30% chance): $113 + 15 = 128$ mins

$\rightarrow 0.7 * 93 + 0.3 * 128 = 103.5$ mins

Route 3: If bridge damaged (60% chance): $106.8 + 40 = 146.8$ mins

If bridge intact (40% chance): 106.8 mins (no additional time)

$\rightarrow 0.6 * 146.8 + 0.4 * 106.8 = 130.8$ mins

$\Rightarrow$ Now, route 1 seems best.

Part 3:

First: If route 2 is not rocky, it means that would either be sandy or smooth.

$$P(\text{Sandy}) = \frac{40\%}{60\%} = \frac{2}{3} \text{ chance}$$

$$P(\text{Smooth}) = \frac{20\%}{60\%} = \frac{1}{3} \text{ chance}$$

$$\text{Route 2} : \left(\frac{2}{3} * 20 + \frac{1}{3} * 12\right) * 7 = (13.34 + 4) * 7$$

$$= 121.38 \text{ mins}$$

- Even route 2 is not rocky, it will take 121.38 mins
  => Route 1 is still the best route (103.5 mins)

Second: The probability that route 2 is not rocky is simply the probability that it is either sandy or smooth:

$$P(\text{not rocky}) = 1 - P(\text{rocky}) = 1 - 0.4 = 0.6 \ (60\%)$$

So, there's a 60% chance the satellite will say route 2 is not rocky.

Third : If route 2 is confirmed to be rocky :

$$30 \text{ min/Km} * 7 \text{ Km} = 210 \text{ mins}$$

=> We would stick with our previous best route, route 1, which has an expected utility of 103.5 minutes (after counting for the crater)

Last: * If the satellite says route 2 is not rocky (60% chance)

- Best expected utility = 103.5 mins (route 1)

* If the satellite says route 2 is rocky (40% chance)

- Best expected utility = 103.5 mins (route 1)

The expected utility if we ask for the satellite is:

$$0.6 * 103.5 + 0.4 * 103.5 = 62.1 + 41.4 = 103.5 \text{ mins}$$

=> Since the expected utility if we ask would be the same as route 1's time of 103.5 mins => It adds no benefit and only increase our travel time if we wait for the satellite

=> Therefore, we should proceed with route 1 without waiting.