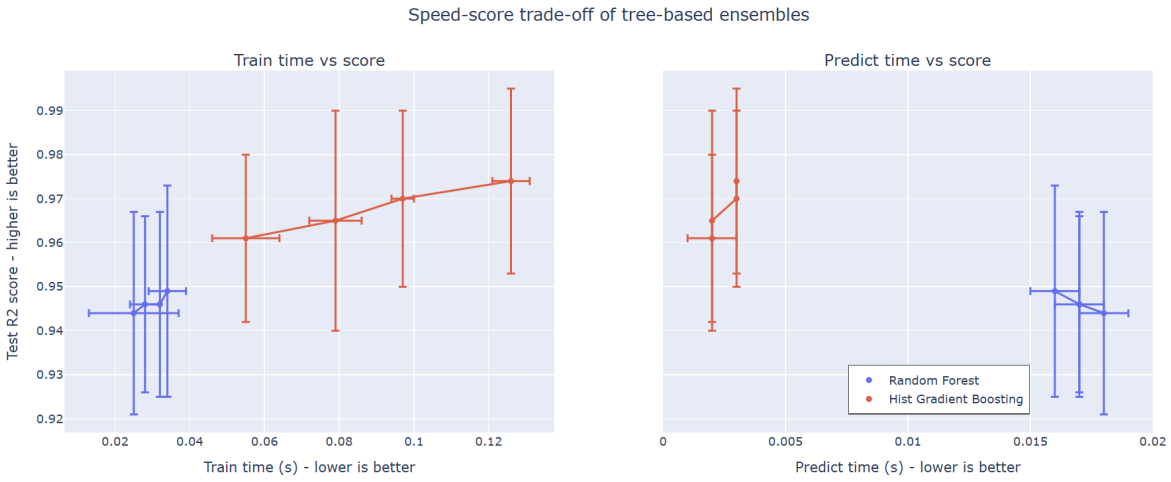
1. **Decision Trees in scikit-learn.**
    b. **Recall that the Random Forest is an ensemble-based approach uses multiple decision trees. Replace the Decision Tree with the Random Forest ClassifierLinks to an external site.. Run it on your dataset with 10, 25, and 50 estimators using both gini and entropy as separators. Create a table showing the results and add it to the PDF with your written answers.**

	10 estimators	25 estimators	50 estimators
Gini	[0.944, 0.861, 0.944, 0.885, 0.714]	[0.917, 0.972, 0.944, 1.0, 0.885]	[0.888, 0.916, 0.944, 1.0, 0.914]
Entropy	[0.972, 0.888, 0.944, 0.942, 0.685]	[0.916, 0.861, 0.944, 1.0, 0.885]	[0.916, 0.888, 0.888, 1.0, 0.857]

    d. **The last part shows how to use plotly to generate a scatterplot showing your results. Generate a plot and add it to the PDF with your answers.**



4. **Utility.**

   This is a pen-and-paper (or typed) exercise. Please include a PDF with your answers in your repository.

   Our Mars rover has been out collecting samples, when it detects that a sandstorm is coming. It needs to return to the safety of the charging station as quickly as possible.

   It knows that over rocky terrain it can go 3 km/h. Over sandy terrain it can go 4 km/h, and over smooth terrain it can go 5 km/h.

   There are three routes it might choose from. Unfortunately, our terrain data for the three routes is incomplete, so we only have estimates.

   Route 1 is 5 km long. There is a 25% chance it is rocky, 25% chance it is sandy, and a 50% chance it is smooth.

   Route 2 is 6 km long. There is a 50% chance it is rocky, a 30% chance it is sandy, and a 20 % chance it is smooth.

   Route 3 is 4 km long. There is a 70% chance it is rocky, a 10% chance it is sandy, and a 20% chance it is smooth.

   a. **Which route should we pick? Show your work. You may find it easier to convert km/hr to mins/km (how many minutes does it take to go 1km on each surface?)**

Rocky:

$$\frac{3\ km}{1\ hr} \leftrightarrow \frac{1\ hr}{3\ km}$$

$$\frac{1\ hr}{3\ km} \times \frac{60\ min}{1\ hr} = 20\ min/km$$

Sandy:

$$\frac{4\ km}{1\ hr} \leftrightarrow \frac{1\ hr}{4\ km}$$

$$\frac{1\ hr}{4\ km} \times \frac{60\ min}{1\ hr} = 15\ min/km$$

Smooth:

$$\frac{5\ km}{1\ hr} \leftrightarrow \frac{1\ hr}{5\ km}$$

$$\frac{1\ hr}{5\ km} \times \frac{60\ min}{1\ hr} = 12\ min/km$$

Route 1:
$5\ km \times ((20\ min/km \times 0.25) + (15\ min/km \times 0.25) + (12\ min/km \times 0.50)) = 73.75\ min$

Route 2:
$6\ km \times ((20\ min/km \times 0.50) + (15\ min/km \times 0.30) + (12\ min/km \times 0.20)) = 101.4\ min$

Route 3:
$4\ km \times ((20\ min/km \times 0.70) + (15\ min/km \times 0.10) + (12\ min/km \times 0.20)) = 71.6\ min$

$71.6\ min < 73.75\ min < 101.4\ min$

She should pick Route 3.

Route 1 contains a crater. If the wall of the crater is intact, we can take a shortcut through the crater, which will save 10 minutes. If the wall has been damaged, we will need to go around, which will add 15 minutes to our journey. There is a 30% chance that the wall is damaged.

Route 3 contains a bridge. If that bridge is damaged, we will need to repair it, which will add 40 minutes to our time. There is a 60% chance that the bridge is damaged.

b. Update your estimates for the travel time for each route. Now which route seems best?

Route 1:
$73.75\ min - 10\ min = 63.75\ min$
$73.75\ min + 15\ min = 88.75\ min$
$(63.75\ min \times 0.70) + (88.75 \times 0.3) = 71.25\ min$

Route 3:
$(71.6\ min \times 0.40) + (111.6\ min \times 0.60) = 95.6\ min$

$71.25\ min < 95.6\ min < 101.4\ min$

Route 1 seems best.

Now we have an additional piece of potential information. There is an orbiting satellite that can tell us whether route 2 is rocky or not. If not, that would be great news, and would make it much more appealing! The only problem is that the satellite is not yet in position. How long should we wait for the satellite?

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

Route 2:

$6\,km \times (15\,min/km \times 0.60) + (12\,min/km \times 0.40) = 82.8\,min$

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

$P(rocky2) = 0.5$

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

We take Route 1.

Route 1: $71.25\,min$

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

We should not wait for the satellite because even if the satellite says route 2 is not rocky, it would still take longer than route 1.