# STA 235 - JITT 10

Please answer before Sunday Apr 25th, 11:59 PM

* Required

1. Email *

2. EID *

## Random Forests, Bagging, and Boosting!

Suppose you are a famous influencer with a substantial knowledge in data science and prediction, and you want to put your knowledge to use. You need to decide which post to upload, and you have a lot of data of how your past posts have been received (a lot of characteristics about the photo, whether it was a sponsored ad or not, length of the caption, time and day of the post, etc.). With that in mind, you decide to build a prediction model to predict the number of likes + shares... but really can't remember whether a single decision tree would be enough, or you could ran just a set of bagged trees, or just a full random forest. You are currently traveling (this fictitious scenario is post-COVID19), and only have an old laptop at hand.

3. What are the main differences between the three learners previously mentioned (i.e. single decision tree, bagged trees, and random forest)? You can basically talk about some of the advantages and disadvantages of each (Remember to base your answers based on the readings/videos!) *

4. You realize that you have data on over 5,000 posts and over 100 attributes/predictors. Would you need to use a model selection approach (e.g. stepwise, lasso) before using a decision tree or random forest to select which ones are more predictive? Why? *

## 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