Assignment 3 and Resources

Your third assignment will be to create a decision tree model to predict the survival of passengers using the Titanic dataset (titanic.csv) present on the GitHub repo in 'Assignment 3' folder .

- Create new features, remove irrelevant features, go crazy and build a reliable model!
- Split your dataset into 75-25 parts and use the 75% to train the model.
- Use the 25% to test your model using model.score()
- Test it on three different train-test splits and take the average and print it at the end of your notebook (remember, you don't need to train thrice for this, just train once, and use the train-test split function to obtain two more different test datasets in addition to the original one obtained).
- I'll be putting up the best accuracy as and when I receive a submission from any of you on the whatsapp group. So let's see who does best :P
- This way, you'll work on improving your models incrementally which is a very important thing in machine learning.

Here are some more resources from the previous two weeks for reference, also, get you hands dirty, google, and figure stuff out on your own, it'll help you learn much better.

1. Session 2
    a. https://www.youtube.com/watch?v=u73PU6Qwl1I
    b. https://machinelearningmastery.com/linear-regression-for-machine-learning/
    c. https://machinelearningmastery.com/logistic-regression-for-machine-learning/
    d. https://towardsdatascience.com/machine-learning-part-3-logistics-regression-9d890928680f

2. Session 3 (Decision trees)
    a. https://towardsdatascience.com/entropy-how-decision-trees-make-decisions-2946b9c18c8
    b. https://medium.com/greyatom/decision-trees-a-simple-way-to-visualize-a-decision-dc506a403aeb
    c. https://www.hackerearth.com/fr/practice/machine-learning/machine-learning-algorithms/ml-decision-tree/tutorial/
    d. https://stackabuse.com/decision-trees-in-python-with-scikit-learn/
    e. https://scikit-learn.org/stable/modules/tree.html