Practical Data Science
(Assignment-I)

Solve the following problems.

Problem1: First attempts with Predictive Analytics Problems

Go through the following problem of kaggle:
https://www.kaggle.com/c/ghouls-goblins-and-ghosts-boo/data

Do the following tasks:

a. Apply random predictions/majority based prediction to each observation and find out how much accurate your predictions are by submitting to kaggle?
b. Find out the pattern in the data manually with visual EDA and then hard-code the logic. Find out the accuracy by submitting to kaggle.
c. Find out the pattern in the data with machine learning approach using decision tree algorithm. Find out the accuracy after submitting to kaggle and also locally using cross-validation technique.

Problem 2: Applying DecisionTree Algorithms

Given the following training data with 4 categorical variables and 1 target variable.

RID	age	income	student	credit_rating	Class: buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31 . . . 40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31 . . . 40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31 . . . 40	medium	no	excellent	yes
13	31 . . . 40	high	yes	fair	yes
14	>40	medium	no	excellent	no

Do the following:

Practical Data Science
(Assignment-I)

Build a tree model using decision tree algorithm with following parameter combinations and predict the class of following test observation using the model you constructed:

age<=30, income=medium, student=yes, credit_rating=fair

a. Criterion = information-gain and depth = 3
b. Criterion = gini-gain and depth = 3