**Lab Assign-08- Decision Tree - Predict Tumor Log Volume**

* Read the dataset **"prostrate.csv"** that is provided to you.
* Build a suitable decision tree predictive model to predict **Tumor Log Volume** based on predictor information.
* Plot the decision tree and develop some metrics to determine the accuracy of your model. (Compute various evaluation parameters of the tree model built).
* Cross validate and optimize the model using hold back V-fold technique.
* Use method of pruning to avoid over-fitting and derive the best size of the decision tree.

**Notes**

1. Data-set contains biopsy results for *n = 97 men of various ages.*
2. Information includes
   1. Gleason Score (gleason): scores are assigned to the two most common tumor patterns, ranging from 2 to 10; in this data set, the range is from 6 to 9.
   2. Prostate-specific antigen (psa): laboratory results on protein production.
   3. Capsular penetration (cp): reach of cancer into gland lining.
   4. Benign prostatic hyperplasia amount (bph): size of prostate.
3. The goal is to predict tumor log volume (which measures the tumor’s size/spread). The predicted tumor size affects the patients’ treatment options, which include chemotherapy, radiation treatment, and surgical removal of the prostate.
4. The data of the 97 prostate cancer patients is given. The response is log volume (lcavol). We try to predict this variable from five covariates (age; logarithms of bph, cp, and psa; and the Gleason score).
5. Here the response is a continuous measurement variable, and we are dealing with a regression tree. We use the sum of squared residuals as the impurity (fitting) criterion.