

**Indian Institute of Technology Kharagpur**  
**Department of Industrial and Systems Engineering**  
**Statistical Learning Lab (IM39202)**

**Topic:** Classification and regression trees, Bagging and random forest

**Objective:** In this lab the students will learn the basic techniques of fitting the decision tree for classification and regression dataset, also bagging and random forest technique.

**Tools required:** R and R-studio.

**Tasks:**

- 1) Import the designated data file
- 2) Data cleaning and pre-processing
- 3) Identify a response variable
- 4) Convert categorical inputs or consider it while fitting the data
- 5) Fit a classification and regression model
- 6) Plot the decision tree for fitted model
- 7) Prune the tree by changing the best value
- 8) Observe the results by calculating the misclassification rate or accuracy
- 9) Fit a bagging and random forest model
- 10) Change the value of number of predictors considered for each split and observe the results
- 11) Find the best model using parameter tuning and calculate the accuracy.

**Report Format:**

Create an R markdown file and submit the pdf with all the code snippets, plots, results and explanation.

**Dataset: (Perform above tasks on any one of the following dataset)**

- 1) Pulsar star (<https://www.kaggle.com/datasets/colearninglounge/predicting-pulsar-starintermediate>)
- 2) Drugs dataset (<https://www.kaggle.com/datasets/pablomgomez21/drugs-a-b-c-x-y-for-decision-trees>)