



IIT ROORKEE



NPTEL ONLINE  
CERTIFICATION COURSE

# CLASSIFICATION & REGRESSION TREES PART-6

## LECTURE 41

DR. GAURAV DIXIT

DEPARTMENT OF MANAGEMENT STUDIES



# CLASSIFICATION & REGRESSION TREES

- Steps to classify a new observation using tree based models
  - At leaf node, majority class is assigned to the new observation
    - For class of interest scenario, proportion of records belonging to the class of interest is compared with the user specified cut off value for the same
- Open RStudio



# CLASSIFICATION & REGRESSION TREES

- Pruning
  - Avoid overfitting
    - Full grown tree leads to complete overfitting of data
    - Poor performance on new data
  - Overall error of tree models
    - Expected to decrease until the point where relationships between outcome variable and predictors are fitted
    - Then tree models start fitting to the noise and overall error starts increasing
      - Due to splits involving small number of observations



# Key References

- Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data by EMC Education Services (2015)
- Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner by Shmueli, G., Patel, N. R., & Bruce, P. C. (2010)



# Thanks...

