Summer of science

CS18 : Data Science

Mentee – Pravesh Khaparde

Mentor – Ank kumar gupta

Plan of Action

WEEK 1: Python and some of its libraries for ML and Excel for Data Visualization, pivot tables, regression, graphs. Reference from “Introduction to Machine Learning” from NPTEL - IIT Madras.

WEEK 2: Linear Regression, loss functions, Logistic Regression, few problems practicing from “great learning” YouTube channel and W3schools, Machine Learning Specialization (coursera course by Andrew NG).

WEEK 3: Statistical concepts like confusion matrix, correlation, significance of R^2 value, evaluation metrics etc.

WEEK 4: Decision Trees and Ensemble Methods. Video lectures from NPTEL - IIT Madras and reading from scikit-learn.org. exercises from “An introduction to statistical learning” by Springer.

Midterm Report

WEEK 5: SVM, Basic Neural Networks, Forward Propagation and Back Propagation. exercises from “An introduction to statistical learning” by Springer.

WEEK 6 -7: Exploratory Data Analysis including Imputation techniques, KNN, Data Visualization, IQR, VIF, IQR, encoding etc.

WEEK 8: Unsupervised Learning and Clustering. Practice and coding.

End term Report

Reference:

- An introduction to statistical learning with applications in Python, , Vol. 112, Springer, New York Author: James, G., Witten, D., Hastie, T., Tibshirani, R. and Taylor, J.
- Introduction to Machine Learning (NPTEL - IIT Madras) : https://nptel.ac.in/courses/106106139
- Machine Learning Specialization (Coursera - Andrew NG) https://www.coursera.org/specializations/machine-learning-introduction
- Neural networks and deep learning, web-book http://neuralnetworksanddeeplearning.com/index.html