Center for Technical Education

First Semester , 2018 - 2019

Course Handout

Course Title : Data Science and Machine Learning

Course Description : This course takes students from zero, to a position where they can do something with Data and Machine Learning Algorithms

Instructors : Fenil Suchak, Kaustav Tamuly, Nikhil Garg
Mentors : Atharva Joshi, Pranav Mahajan

Students will learn to play with Data. Cut it, clean it, prepare it, look into it, diagnose it. They will also learn state of the art Machine Learning Algorithms , which they will be able to use and apply to the Data at hand and get the best results. Machine learning being part science and part engineering, the students will be able to actually see how current research in this area is being applied to real world problem. This being a developing field, students interested in both science and engineering can substantially make the most out of it. As a by-product of this course students will also learn basic programming techniques, Python Language, Python Libraries for Machine learning,
Git Version Control System, Basic Image Processing techniques, Statistics, Linear Algebra, Kaggle Platform.

Prerequisites : None

Scope & Objective : It aims give students the tools and ideas of a very vast and dynamic field of Machine Learning so that they know when to use what to get the best out of their data.Expose students to various ongoing research in the field and how state of the art research can be applied in practice.

Week	Topic	What will you learn?	What you should refer?
1	What is ML? What is Data Science? Why is it there? Python and Dev Env Introduction	The birds eye view of the whole subject Supervised ML, Unsupervised ML. Appreciate the potential of Data	Class ,Additional Material and wherever you feel like
2	Python,R and Dev Env Introduction	Why use Python? Why use R?	Python,R Documentation
3	Mathematics of Machine Learning Part 1	Statistics, Probability and why it is the secret weapon to success	Class, Additional Material and wherever you feel like.
4	Mathematics of Machine Learning Part 2	Linear Algebra and why it is a swiss army knife	Class, Additional Material and wherever you feel like.
5	Bias-Variance Tradeoff the Real science Regression, KNN & Regularization	Data Cleaning, Pandas, Scikit-learn and basic but very powerful ML algorithms	Class, Additional Material and wherever you feel like. Github repo also
6	Competitive Data Science and Machine Learning	Validation techniques, Best Metric Selection, Loss function optimization, EDA. When should you give up on your data.	Class, Additional Material and wherever you feel like. Github repo also Kaggle

7	Naive Bayes, LDA, PCA, High Dimensional Data.	Learn to work with high dimensional data and visualize it better to get more insights.	Class, Additional Material and wherever you feel like. Github repo also Kaggle.
8	How bad can data be? Correlation? Missing Values? How to handle it.	To get the best out of the worst kind of data	Class, Additional Material and wherever you feel like. Github repo UCI ML repository
9	Trees, Boosting and Ensembling Why ensembling is so successful	Learn a new class of ML algorithms which are used widely for prediction	Class, Additional Material and wherever you feel like. Github repo
10	SVM , Neural Networks What makes NNets so good? And When you should use them	Learn state of the art algorithms and their application	Class, Additional Material and wherever you feel like.
11	SVM , Neural Networks What makes NNets so good? And When you should use them	Different types NNets, Convolutional Neural Networks, Recurrent Neural Networks Various Architectures	Class, Additional Material and wherever you feel like.
12	SVM , Neural Networks What makes NNets so good? And When you should use them	Different NNets, Convolutional Neural Networks, Recurrent Neural Networks. Various Architectures	Class, Additional Material and wherever you feel like.

13	Neural Networks. What algorithm makes most sense with which type of data. And when not to go with the hype.	How to smartly chose between algorithm given the data at hand.	Class, Additional Material and wherever you feel like.

Component	Tentative Weightage	Tentative Date	Remarks
Mini Assignment - I	20%	Week4	OB
Assignment - 2	30%	Week7	OB
Mini Assignment - 3	20%	Week9	OB
Final Project	30%	Week11	OB