20AM3609- DATA SCIENCE

Syllabus

COURSE CONTENT:	
MODULE 1: Introduction to Data Science & Programming Tools for Data Science	**8 Hrs**
Concept of Data Science, Traits of Big data, Analysis vs Reporting, Toolkits using Python:, NumPy, Pandas, Scikit-learn, Matplotlib, Visualizing Data: Bar Charts, Line Charts, Scatterplot. Working with data: Reading Files, Scraping the Web, Using APIs (Example: Using the Twitter APIs), Cleaning and Munging, Manipulating Data, Rescaling, Dimensionality Reduction, Principal Component Analysis, Feature extraction	
MODULE 2: Mathematical Foundations	**8 Hrs**
Review of Probability theory, Correlation, Dependence and Independence, Conditional Probability, Baye's Theorem, The Normal Distribution, The Central Limit Theorem, Hypothesis and Inference: Statistical Hypothesis Testing, Confidence Intervals, P-hacking, Bayesian Inference	
MODULE 3 : Machine Learning	**8 Hrs**
Overview of Machine learning concepts – Over fitting and under fitting, feature selection, train/test splits, Types of Machine learning – Supervised, Unsupervised, Reinforced learning, Linear Regression- regularization (lasso, ridge, elastic net), Clustering algorithms, K-Means Clustering, Classification versus Regression	
MODULE 4 : Popular Machine Learning algorithms	**8 Hrs**
Naive Bayes, K-Nearest Neighbors, logistic regression, support vector machines (SVM), decision trees, and random forest, Classification performance metrics, Analysis of Time Series, Neural Networks- Learning and Generalization, Overview of Deep Learning.	
MODULE 5 : Case Studies of Data Science Application	**7 Hrs**
Weather forecasting, Stock market prediction, Object recognition, Real Time Sentiment Analysis.	

TEXT BOOK:

1. Joel Grus, "Data Science from Scratch: First Principles with Python", O'Reilly Media
2. Aurélien Géron, "Hands-On Machine Learning with Scikit-Learn and Tensor Flow: Concepts, Tools, and Techniques to Build Intelligent Systems", 1st Edition, O'Reilly Media

REFERENCES:

1. Ian Goodfellow, Yoshua Bengio and Aaron Courville, "Deep Learning", MIT Press http://www.deeplearningbook.org
2. Jiawei Han and Jian Pei, "Data Mining Concepts and Techniques", Third Edition, Morgan Kaufmann Publishers
3. Jain V.K., "Data Sciences", Khanna Publishing House, Delhi.
4. Jain V.K., "Big Data and Hadoop", Khanna Publishing House, Delhi.
5. Jeeva Jose, "Machine Learning", Khanna Publishing House, Delhi.
6. Chopra Rajiv, "Machine Learning", Khanna Publishing House, Delhi