CSCSVAXX	ADVANCED DATA SCIENCE	L	T	P
		3	0	0

COURSE OBJECTIVES:
- To equip the students with some of the basic concepts and principles of data science.
- To make the students learn techniques and tools to deal with data collection and integration, exploratory data analysis, modeling, evaluation and effective communication.
- To develop skill sets needed to be a data scientist.
- To work effectively on data science projects.

Introduction: Big Data and Data Science hype, Datafication, Data Scientist, Current landscape of perspectives, Statistical Inference - Populations and Samples, Statistical modeling, Probability distributions, Modeling - Exploratory Data Analysis - Philosophy- Data Science Process - Case Study: RealDirect.

Algorithms: Linear Regression, k-NN, k-means, Spam Filters, Naive Bayes, Wrangling - Logistic Regression: Classifiers, Case Study: M6D Logistic Regression.

Feature Generation Brainstorming, Role of domain expertise, and Place for imagination – Feature Selection: Filters, Wrappers, Decision Trees, Random Forests.

Recommendation Engines: Nearest Neighbors - Dimensionality Problem-Singular Value Decomposition, Principal Component Analysis - Social Network Analysis.

Data Visualization: Basic principles, ideas and tools for data visualization, Sample projects – Data Engineering algorithms - Data Scientists and Ethics.

REFERENCES:
1. Rachel Schutt and Cathy O'Neil , "Doing Data Science, Straight Talk From The Frontline", O'Reilly Media, 2013.
2. Jure Leskovek, Anand Rajaraman and Jerey D. Ullman, "Mining of Massive Datasets", 2nd Edition, Cambridge University Press. 2014.
3. Kevin P. Murphy. Machine Learning: A Probabilistic Perspective, MIT Press, Cambridge, 2013.
4. Foster Provost and Tom Fawcett. Data Science for Business: What You Need to Know about Data Mining and Data-analytic Thinking, O'Reilly Media, 2013.

COURSE OUTCOMES:
After the completion of the course, the students will be able to
1. Describe the basic concepts of Data Science.
2. Apply basic machine learning algorithms for predictive modeling.
3. Apply Exploratory Data Analysis and Data Science process in a case study.
4. Identify approaches used for Feature Generation and Feature Selection and use in applications.
5. Create effective visualization of given data.