

Course Type	Course Code	Name of Course	L	T	P	Credit
DC	NCSC516	Data Analytics	3	0	0	3

#### Course Objective

The main goal of this course is to help students learn, understand, and practice big data analytics and machine learning approaches, which include the study of modern computing big data technologies and scaling up machine learning techniques focusing on industry applications. Mainly the course objectives are: conceptualization and summarization of big data and machine learning, trivial data versus big data, big data computing technologies, machine learning techniques, and scaling up machine learning approaches.

#### Learning Outcomes

- Ability to identify the characteristics of datasets and compare the trivial data and big data for various applications.
- Ability to select and implement machine learning techniques and computing environment that are suitable for the applications under consideration.
- Ability to recognize and implement various ways of selecting suitable model parameters for different machine learning techniques.
- Ability to integrate machine learning libraries and mathematical and statistical tools with modern technologies like Hadoop and map reduce.

Unit No.	Topics to be Covered	Lecture Hours	Learning Outcome
1	Introduction, Big Data; Descriptive Statistics; Descriptive Statistics and Probability Distributions.	5	Understanding what is Big data and use of statistics for various data analysis.
2	Inferential Statistics: Inferential Statistics through hypothesis tests Permutation & Randomization Test.	4	To get knowledge of inferential statistics.
3	Regression & ANOVA: Regression and ANOVA (Analysis of Variance).	4	To know how to predict next by the use of regression technique and using ANOVA related to given data set.
4	Machine Learning: Introduction and Concepts: Differentiating algorithmic and model based Frameworks.	5	To know the application of machine learning in data analytics.
5	Regression: Ordinary Least Squares, Ridge Regression, Lasso Regression, K Nearest Neighbours Regression & Classification.	5	To have the knowledge of prediction and classification techniques for a given data set.
6	Supervised Learning with Regression and Classification techniques -1: Bias-Variance Dichotomy; Model Validation Approaches; Logistic Regression; Linear Discriminant Analysis; Quadratic Discriminant Analysis.	7	To know various techniques related to different learning techniques.

7	Regression and Classification Trees Support Vector Machines; Ensemble Methods: Random Forest; Neural Networks; Deep learning; Unsupervised Learning and Challenges for Big Data Analytics: Clustering, Associative Rule Mining Challenges for big data analytics.	9	Further to know various approaches related to different learning techniques.
8	Scalable Computing: Hadoop and Map Reduce.	3	To have the knowledge of using Hadoop and Map Reduce.
<b>Total</b>		<b>42</b>	

#### **Text Books:**

1. Data Analytics Made Accessible, by A. Maheshwari
2. Big Data Analytics, By Seema Acharya, Wiley

#### **Reference Books:**

1. Machine Learning and Big Data by Kareem Alkaseer
2. Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die by E. Siegel
3. Too Big to Ignore: The Business Case for Big Data, by award-winning author P. Simon,