

<b>COURSE CODE</b>	<b>COURSE TITLE</b>	<b>L</b>	<b>T</b>	<b>P</b>	<b>C</b>
<b>1152IT149</b>	<b>DATA SCIENCE &amp; BIG DATA ANALYTICS</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>3</b>

### **UNIT I - INTRODUCTION TO BIG DATA ANALYTICS:**

Big Data overview, State of the practice in analytics role of data scientists, Big Data Analytics in industry verticals

### **UNIT II - END-TO-END DATA ANALYTICS LIFE CYCLE**

Key roles for successful analytic project, main phases of life cycle, Developing core deliverables for stakeholders

### **UNIT III - BASIC ANALYTIC METHODS:**

Introduction to “R”, analyzing and exploring data with “R”, statistics for model building and evaluation

### **UNIT IV - ADVANCED ANALYTICS AND STATISTICAL MODELING FOR BIG DATA**

Naïve Bayseian Classifier, K-means Clustering, Association Rules, Decision Trees, Linear and Logistic Regression, Time Series Analysis, Text Analytics;

### **UNIT V - MAPREDUCE/HADOOP**

Technology and Tools – MapReduce/Hadoop , In- database Analytics, MADlib and advanced SQL Tools

### **TEXT / REFERENCES BOOKS**

1. Noreen Burlingame ,The little book on Big Data, New Street publisher(eBook)  
<http://www.prlog.org/11800911-just-published-the-little-book-of-big-data-2012-edition.html>
2. Norman Matloff ,The Art of R Programming: A Tour of Statistical Software Design , ISBN-13: 978-1-59327-384-2; ISBN-10: 1-59327-384-3
3. [http://www.johndcook.com/R\\_language\\_for\\_programmers.html](http://www.johndcook.com/R_language_for_programmers.html)
4. <http://bigdatauniversity.com/>
5. <http://home.ubalt.edu/ntsbarsh/stat-data/topics.htm#rintroduction>