Course code	Course Name	L-T-P - Credits	Year of Introduction
IT404	**Data Analytics**	**3-0-0-3**	**2016**

Prerequisite: CS208 Principles of database design

Course Objectives
- To understand the data analysis techniques
- To understand the concepts behind the descriptive analytics and predictive analytics of data
- To familiarize with Big Data and its sources
- To familiarize data analysis using R programming
- To understand the different visualization techniques in data analysis

Syllabus

Data Analysis, Analysis Vs Reporting, Different Statistical Techniques of Data Analysis, Descriptive Analytics, Regressive Models, Neural Networks. Descriptive Analytics- Association and Sequential Rules, Big Data and its characteristics, Data Analysis using R language, Data visualization techniques.

Expected outcome .
- The student will understand the techniques to analyze different types of data, characterize it and can apply them to make decision modeling process more intelligent

Text Book:
1. EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. John Wiley & Sons, 2015.
2. Jaiwei Han, Micheline Kamber, "Data Mining Concepts and Techniques", Elsevier, 2006.
3. Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer, 2007.

References:

Applications", John Wiley & Sons, 2014
1. Bart Baesens," Analytics in a Big Data World: The Essential Guide to Data Science and its Business Intelligence and Analytic Trends", John Wiley & Sons, 2013
   Challenges and Future Prospects, Springer, 2014.
2. Michael Minelli, Michele Chambers, Ambiga Dhiraj , "Big Data, Big Analytics: Emerging Min Chen, Shiwen Mao, Yin Zhang, Victor CM Leung ,Big Data: Related Technologies,

	Course Plan		
Module	**Contents**	**Hours**	**Sem. Exam Marks**
I	Introduction to Data Analysis - Evolution of Analytic scalability, analytic processes and tools, Analysis vs reporting - Modern data analytic tools. Statistical concepts: Sampling distributions, re-sampling, statistical inference, prediction error.	8	15%
II	Predictive Analytics – Regression, Decision Tree, Neural Networks. Dimensionality Reduction - Principal component analysis	6	15%
	FIRST INTERNAL EXAMINATION		

III	Descriptive Analytics - Mining Frequent itemsets - Market based model – Association and Sequential Rule Mining - Clustering Techniques – Hierarchical – K- Means	6	15%
IV	Introduction to Big data framework - Fundamental concepts of Big Data management and analytics - Current challenges and trends in Big Data Acquisition	7	15%
SECOND INTERNAL EXAMINATION			
V	Data Analysis Using R - Introduction to R, R Graphical User Interfaces, Data Import and Export, Attribute and Data Types, Descriptive Statistics, Exploratory Data Analysis, Visualization Before Analysis, Dirty Data, Visualizing a Single Variable, Examining Multiple Variables, Data Exploration Versus Presentation, Statistical Methods for Evaluation	8	20%
VI	Popular Big Data Techniques and tools- Map Reduce paradigm and the Hadoop system- Applications Social Media Analytics- Recommender Systems- Fraud Detection.	7	20%
END SEMESTER EXAM			

QUESTION PAPER PATTERN

Maximum Marks: 100   Exam Duration: 3 hours

The question paper shall consist of Part A, Part B and Part C.

Part A shall consist of three questions of 15 marks each uniformly covering Modules I and II. The student has to answer any two questions ($15\times2=30$ marks).

Part B shall consist of three questions of 15 marks each uniformly covering Modules III and IV. The student has to answer any two questions ($15\times2=30$ marks).

Part C shall consist of three questions of 20 marks each uniformly covering Modules V and VI. The student has to answer any two questions ($20\times2=40$ marks).

Note : Each question can have a maximum of 4 subparts, if needed