SEMESTER S6

DATA ANALYSIS

Course Code	OEEVT611	CIE Marks	40
Teaching Hours/Week (L: T:P: R)	3:0:0:0	ESE Marks	60
Credits	3	Exam Hours	2 Hrs. 30 Min.
Prerequisites (if any)	None	Course Type	Theory

Course Objectives:

1. To understand the basic concepts of data analytics.
2. To enable learners to perform data analysis on a real world scenario using appropriate tools.

SYLLABUS

Module No.	Syllabus Description	Contact Hours
1	Introduction to Data Analysis - Analytics, Analytics Process Model, Analytical Model Requirements. Data Analytics Life Cycle overview. Basics of data collection, sampling, preprocessing and dimensionality reduction	9
2	Descriptive statistics - Measures of central tendency and dispersion, Association of two variables - Discrete variables, Ordinal and Continuous variable, Probability calculus - probability distributions, Inductive statistics - Point estimation, Interval estimation, Hypothesis Testing - Basic definitions, t- test	9
3	Supervised Learning - Classification, Naive Bayes, KNN, Linear Regression. Unsupervised Learning - Clustering, Hierarchical algorithms – Agglomerative algorithm, Partitional algorithms - K- Means. Association Rule Mining - Apriori algorithm	9
4	Big Data Overview – State of the practice in analytics, Example Applications - Credit Risk Modeling, Business Process Analytics.Big Data Analytics using Map Reduce and Apache Hadoop, Developing and Executing a HadoopMapReduce Program.	9

	Overview of modern data analytic tools.Data Analysis Using R - Introduction to R - R Graphical User Interfaces, Data Import and Export, Attribute and Data Types, , Exploratory Data Analysis - Visualization Before Analysis, Dirty Data, Visualizing a Single Variable, Examining Multiple Variables.	

Course Assessment Method
(CIE: 40 marks, ESE: 60 marks)

Continuous Internal Evaluation Marks (CIE):

Attendance	Assignment/ Microproject	Internal Examination-1 (Written)	Internal Examination- 2 (Written)	Total
5	15	10	10	40

End Semester Examination Marks (ESE)

In Part A, all questions need to be answered and in Part B, each student can choose any one full question out of two questions

Part A	Part B	Total
• 2 Questions from each module. • Total of 8 Questions, each carrying 3 marks **(8x3 =24marks)**	• Each question carries 9 marks. • Two questions will be given from each module, out of which 1 question should be answered. • Each question can have a maximum of 3 sub divisions. **(4x9 = 36 marks)**	**60**

Course Outcomes (COs)

At the end of the course students should be able to:

Course Outcome		Bloom's Knowledge Level (KL)
CO1	Explain the basic and mathematical concepts for data analytics **(CognitiveKnowledge Level: Apply)**	**K2**
CO2	Illustratevariouspredictiveanddescriptiveanalyticsalgorithms(**Cognitive Knowledge Level: Apply)**	**K3**
CO3	DescribethekeyconceptsandapplicationsofBigDataAnalytics**(Cognitive Knowledge Level: Understand)**	**K2**
CO4	Use R programming tool to perform data analysis and visualization	**K3**

Note*: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create*

CO-PO Mapping Table (Mapping of Course Outcomes to Program Outcomes)

	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	**3**	**3**	**2**	**2**								**2**
CO2	**3**	**3**	**2**	**2**								**2**
CO3	**3**	**3**	**2**	**2**	**2**							**2**
CO4	**3**	**3**	**2**	**2**	**3**							**2**

Note: *1: Slight (Low), 2: Moderate (Medium), 3: Substantial (High), -: No Correlation*

Text Books				
Sl. No	**Title of the Book**	**Name of the Author/s**	**Name of the Publisher**	**Edition and Year**
1	Analytics in a Big Data World: The Essential Guide to Data Science and its Business Intelligence and Analytic Trends",	Bart Baesens	John Wiley & Sons,	2013.
2	EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data	David Dietrich	John Wiley & Sons,	2015.
3	Data Mining Concepts and Techniques'	Jaiwei Han, MichelineKamber	Elsevier,.	2006
4	Introduction to Statistics and DataAnalysis	Christian Heumann and Michael Schomaker	Springer,	2016

Reference Books				
Sl. No	**Title of the Book**	**Name of the Author/s**	**Name of the Publisher**	**Edition and Year**
1	Data Mining: Introductory and Advanced Topics.	Margaret H. Dunham,	Pearson,	2012.
2	Intelligent Data Analysis	Michael Berthold, David J. Hand	Springer	2007

Video Links (NPTEL, SWAYAM…)	
Module No.	**Link ID**
1	https://onlinecourses.nptel.ac.in/noc21_cs45/preview
2	https://archive.nptel.ac.in/courses/106/107/106107220/