CHAROTAR UNIVERSITY OF SCIENCE AND TECHNOLOGY

DEVANG PATEL INSTITUTE OF ADVANCE TECHNOLOGY AND RESEARCH

Subject : Data Science and Analytics **Semester:** 7

Subject Code: CS442 **Academic Year :**2023-24(ODD)

Course Outcome (COs):

At the end of the course, the students will be able to:

CO1	Use an ethically responsible approach to evaluate and interpret data
CO2	Demonstrate expertise in statistical data processing
CO3	Use of various algorithms as well as mathematical and statistical models and optimization concepts to formulate and the use analyze data appropriately
CO4	Develop the ability to build and evaluate data-based models.
CO5	To learn difference between conventional SQL query language and NoSQL And MongoDB basic concepts
CO6	Utilizing data science principles and approaches to solve real-life situational problems and effectively communicate them.

	AIM	Hours	CO	PO	PEO
1.	To perform data pre-processing of IBM Churn dataset from https://www.kaggle.com/datasets/yeanzc/telco-customer-churn-ibm-dataset. • Load data • Find missing values • Clean data • Find co-relations between attributes. • Remove redundant attributes. • Normalize data • Visualize the data Use numpy, pandas, and matplotlib.	6	2	1,3	1
2.	To install Hadoop framework, configure it and setup a single node cluster. Use web based tools to monitor your Hadoop setup.	4	1,2	1,3, 4,5	1,2
3.	To implement file management tasks in Hadoop HDFS and perform Hadoop commands.	2	1,2	1,2, 3,5	1,2
4.	To implement Map, reduce, filter and lambda in python.	4	1,2	1,3, 4,5	1,2

5.	To implement a word count application using the MapReduce programming model.	4	1,2	1,3, 4,5	1,2
6.	To design and implement MapReduce algorithms to take a very large file of integers and produce as output: a) The largest integer b) The average of all the integers. c) The count of the number of distinct integers in the input.	2	1,2,3,5	1,2, 3,5	1,2
7.	To implement basic functions and commands in R Programming. Use R-Studio and build WordCloud and data visualization using R for easy to understand and better visualization than a data table.	4	1,2	1,2, 3,5	1,2
8.	To implement supervised learning algorithms (linear regression and logistic regression) using R.	4	1,2	1,2, 3,5	1,2
9.	To implement unsupervised learning algorithms (k-means) using R.	4	1,2	1,2, 3,5	1,2
10.	To install and implement basic database operations in MongoDB. To implement basic CRUD operations (create, read, update, delete) in MongoDB.	6	1,2	2,3	1,2
11.	To design a dashboard using Google data studio.	4	1,2	2,3	1,2
12.	To install, deploy & configure Apache Spark. To Select the fields from the dataset using Spark SQL.	2	1,2,5	1,2, 3,5	1,2
13.	To implement logistic regression using IBM churn dataset with Apache Spark.	4	1,2,5	1,2, 3,5	1,2
14.	To perform Graph Path and Connectivity analytics and implement basic queries after loading data using Neo4j	4	1,2,5	2, 5,11	1,2,5
15.	To perform case study of the following platforms for solving any big data analytic problem of your choice. (1) Amazon web services,(2) Microsoft Azure, (3)Google App engine	6	1,2,5	2, 5,11	1,2,5