Cluster Analysis: Types of data in Clustering Analysis, Categorization of Major Clustering methods, Hierarchical methods, Density-based methods, Grid-based methods, Model-based Clustering methods

Mining Complex Types of Data: Multidimensional analysis & Descriptive mining of Complex data objects, Mining Spatial Databases, Mining Multimedia Databases, Mining Time-series & Sequence data, Mining Text databases, Mining World -Wide Web

Data Mining Applications and Trends in Data Mining: Massive Datasets/Text mining, Agent-Based Mining

TEXT BOOKS, AND/OR REFERENCE MATERIAL

1. Jiawei Han and MichelineKamber, "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, 2000 (ISBN: 1-55860-489-8).
2. Ian H. Witten and Eibe Frank, "Data Mining: Practical Machine Learning Tools and Techniques with Java implementations", Morgan Kaufmann Publishers, San Fransisco, CA (2000).
3. Dorian Pyle, "Data Preparation for Data Mining", Morgan Kaufmann, (1999)
4. Korth, Silbertz, Sudarshan, "Database Concepts", McGraw Hill
5. Elmasri, Navathe, "Fundamentals Of Database Systems", Addision Wesley

DEPARTMENT: COMPUTER SCIENCE AND ENGINEERING
COURSE CODE: CSPC-322
COURSE TITLE: DATA ANALYTICS LABRATORY
COURSE DESIGNATION: REQUIRED
PRE-REQUISITES: NONE
CONTACT HOURS/CREDIT SCHEME: (L-T-P-C: 0-0-2-1)
COURSE ASSESSMENT METHODS: Assignments for each topic to be evaluated in the lab, and final evaluation at the end which includes Viva Voce, Conduct of experiment.

COURSE OUTCOMES
After the course completion, the student will be able to

1. Preparing for data summarization, query, and analysis.
2. Applying data modelling techniques to large data sets
3. Creating applications for Big Data analytics
4. Building a complete business data analytic solution

Course Outcomes	Program outcomes											
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CSPC-322												
CO 1	H		M		M							
CO 2	M	M	H	M	M				H			
CO 3					H	M						
CO 4		H		H		H		H		M		M

LIST OF PRACTICALS

1. Perform setting up and Installing Hadoop in its two operating modes:
2. Implement the following file management tasks in Hadoop:
3. Run a basic Word Count Map Reduce program to understand Map Reduce Paradigm.
4. Implement Stop word elimination problem Using suitable example.
5. Implement a Map Reduce program that mines weather data.
6. Install and Run Pig then write Pig Latin scripts to sort, group, join, project, and filter your data.
7. Write a Pig Latin scripts for finding TF-IDF value for book dataset (A corpus of eBooks available at: Project Gutenberg)
8. Install and Run Hive then use Hive to create, alter, and drop databases, tables, views, functions, and indexes.
9. Install, Deploy & configure Apache Spark Cluster. Run apache spark applications using Scala.
10. Apply suitable data analytics techniques using Apache Spark on Amazon food dataset, find all the pairs of items frequently reviewed together.