# TEERTHANKER MAHAVEER UNIVERSITY

**(Established under Govt. of U. P. Act No. 30, 2008)**

**Delhi Road, Moradabad (U.P.)**

## PhD PROGRAMME

## SYLLABUS FOR DISCIPLINE-SPECIFIC COURSE COMPUTER APPLICATIONS/ COMPUTER SCIENCE & ENGINEERING

| Course Code: PDS240137 | DATA SCIENCE AND DATA ANALYTICS | L | T | P | C |
|---|---|---|---|---|---|
| | | 0 | 0 | 0 | 4 |
| **Objective:** | The objective of this course is to develop proficiency in analyzing and interpreting large datasets using statistical, machine learning, and computational techniques to extract actionable insights. | | | | |
| **Course Outcomes:** | | | | | |
| **CO 1:** | Understanding Sources and nature of data and Data Analytics Lifecycle and analysis of different analytic tools. | | | | |
| **CO 2:** | Understanding different data analysis techniques for different applications. | | | | |
| **CO 3:** | Understanding data mining streams Analysis of different data models Real-time Analytics Platform (RTAP) applications. | | | | |
| **CO 4:** | Analyzing frequent item sets and clustering techniques. | | | | |
| **CO 5:** | Understanding MapReduce, Hadoop, Pig, Hive, HBase and Apply visual data analysis techniques. | | | | |
| **Course Content:** | | | | | |
| **Unit 1:** | **Introduction to Data Analytics:** Sources and nature of data, classification of data (structured, semi-structured, unstructured), characteristics of data, introduction to Big Data platform, need of data analytics, evolution of analytic scalability, analytic process and tools, analysis vs reporting, modern data analytic tools, applications of data analytics. Data Analytics Lifecycle: Need, key roles for successful analytic projects, various phases of the data analytics lifecycle – discovery, data preparation, model planning, model building, communicating results, operationalization. | | | | |
| **Unit 2:** | **Data Analysis:** Regression modelling, multivariate analysis, Bayesian modelling, inference and Bayesian networks, support vector and kernel methods, analysis of time series: linear systems analysis & nonlinear dynamics, rule induction, neural networks: learning and generalization, competitive learning, principal component analysis and neural networks, fuzzy logic: extracting fuzzy models from data, fuzzy decision trees, stochastic search methods. | | | | |
| **Unit 3:** | **Mining Data Streams:** Introduction to streams concepts, stream data model and architecture, stream computing, sampling data in a stream, filtering streams, counting distinct elements in a stream, estimating moments, counting oneness in a window, decaying window, Real-time Analytics Platform (RTAP) applications, Case studies – real-time sentiment analysis, stock market predictions. | | | | |

| | |
|---|---|
| **Unit 4:** | **Frequent Item sets and Clustering:** Mining frequent item sets, market-based modelling, Apriori algorithm, handling large data sets in main memory, limited pass algorithm, counting frequent item sets in a stream, clustering techniques: hierarchical, K-means, clustering high dimensional data, CLIQUE and ProCLUS, frequent pattern-based clustering methods, clustering in non-euclidean space, clustering for streams and parallelism. |
| **Unit 5:** | **Frame Works and Visualization:** MapReduce, Hadoop, Pig, Hive, HBase, MapR, Sharding, NoSQL Databases, S3, Hadoop Distributed File Systems, Visualization: visual data analysis techniques, interaction techniques, systems and applications. Introduction to R - R graphical user interfaces, data import and export, attribute and data types, descriptive statistics, exploratory data analysis, visualization before analysis, and analytics for unstructured data. |
| **Text Books:** | 1. Michael Berthold, David J. Hand, Intelligent Data Analysis, Springer<br>2. Anand Rajaraman and Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press.<br>3. Bill Franks, Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, John Wiley & Sons.<br>4. Michael Minelli, Michelle Chambers, and Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley. |
| **Reference Books:** | 1. David Dietrich, Barry Heller, Beibei Yang, “Data Science and Big Data Analytics”, EMC Education Series, John Wiley.<br>2. Frank J Ohlhorst, “Big Data Analytics: Turning Big Data into Big Money”, Wiley and SAS Business Series.<br>3. Colleen Mccue, “Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis”, Elsevier.<br>4. Anil Maheshwari, “Data Analytics”, McGraw Hill Education. |