**Syllabus for B. Tech. III Year I semester**

**Computer Science and Engineering**

**INTRODUCTION TO DATA SCIENCE**

**(Professional Elective-II)**

|  |  |  |  |
| --- | --- | --- | --- |
| **L** | **T** | **P/D** | **C** |
| **3** | **-** | **-** | **3** |

**Code: 6EC10**

**Course Objective:**

To know the fundamental concepts of Data Science. To explore tools and practices for working with Data Science. To learn about Principle component analysis and understand about Predictive Analytics.

**Course Outcomes:**

**At the end of this course, the student will be able to**

1. Learn about various data types, types of data sets a data quality
2. Implementation of R fundamentals and perform factors and data frames.
3. Implementation of data structures iterative programming & function concepts using R
4. Learn about data visualization techniques and apply suitable visualization techniques
5. Learn about dimensionality reduction based on examples illustrations
6. Perform predictive data analysis on variety of data along with appropriate statistical tests using R.

**UNIT-I**

**DATA TYPES & COLLECTION**

Types of Data: Attributes and Measurement, What is an Attribute?, The Type of an Attribute, The Different Types of Attributes, Describing Attributes by the Number of Values, Asymmetric Attributes, Binary Attribute (Pg.No:22-29, Text Book-1), Nominal Attributes, Binary Attributes, Ordinal Attributes, Numeric Attributes, Discrete versus Continuous Attributes (Pg. No. 39-44, Text-2), Types of Data Sets, General Characteristics of Data Sets, Record Data, Transaction or Market Basket Data, The Data Matrix, The Sparse Data Matrix, Graph Based Data, Graph- Based Data, Ordered Data. Handling Non-Record Data, Data Quality, Measurement and Data Collection Issues, Precision, Bias and Accuracy. (Pg. No. 29-39, Text-1)

**UNIT-II**

**Basics of R:** Introduction, R-Environment Setup, Programming with R, Basic Data Types, Vectors: Creating and Naming Vectors, Vector Arithmetic, Vector Subsetting, Matrices: Creating and Naming Matrices, Matrix Subsetting, Arrays, Class.

**Factors and Data Frames:** Introduction to Factors: Factor Levels, Summarizing a Factor, Ordered Factors, Comparing Ordered Factors, Introduction to Data Frame, Subsetting of Data Frames, Extending Data Frames, Sorting Data Frames. (Text Book-3)

**UNIT-III**

**Lists:** Introduction, Creating a List: Creating a Named List, Accessing List Elements, Manipulating List Elements, Merging Lists, Converting Lists to Vectors, Conditionals and

**Control Flow:** Relational Operators, Relational Operators and Vectors, Logical Operators, Logical Operators and Vectors, Conditional Statements.

**Iterative Programming in R:** Introduction, While Loop, For Loop, Looping Over List.

**Functions in R:** Introduction, Writing a Function in R, Nested Functions, Function Scoping, Recursion, Loading an R Package, Mathematical Functions in R.(Text Book -4)

**UNIT-IV:**

**DATA VISUALIZATION**

**Data Visualization**

Pixel-Oriented Visualization Techniques, Geometric Projection Visualization Techniques, Icon-Based Visualization Techniques, Hierarchical Visualization Techniques, Visualizing Complex Data and Relations. (Pg. No. 56-64, Text-2)

**Charts and Graphs :** Introduction, Pie Chart: Chart Legend, Bar Chart, Box Plot, Histogram, Line Graph: Multiple Lines in Line Graph, Scatter Plot.(Text Book-4)

**UNIT-V:**

**DIMENSIONALITY REDUCTION**

Eigen values and Eigenvectors of Symmetric Matrices, Definitions, Computing Eigen values and Eigenvectors, The Matrix of Eigenvectors, Principal-Component Analysis, An Illustrative Example, Using Eigenvectors for Dimensionality Reduction, Singular-Value Decomposition, Definition of SVD, Interpretation of SVD, Dimensionality Reduction Using SVD (Pg. No.405-422, Text Book-3)

**UNIT VI**

**PREDICTIVE ANALYTICS**

**Data Interfaces:** Introduction, CSV Files: Syntax, Importing a CSV File

**Statistical Applications:** Introduction, Basic Statistical Operations, Linear Regression Analysis, Chi-Squared Goodness of Fit Test, Chi-Squared Test of Independence, Multiple Regression. (Text Book-4)

**TEXT BOOKS:**

1.Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson Education Inc.

2. Han, Jiawei, Jian Pei, and Micheline Kamber, “Data mining: concepts and techniques”, 3 rd Edition, Elsevier, 2011.

3. Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman, Mining of Massive Datasets, Cambridge University Press

4. K G Srinivas ,G M Siddesh “Statistical programming in R”, Oxford Publications.

**REFERENCE BOOKS:**

1. Brain S. Everitt, “A Handbook of Statistical Analysis Using R”, Second Edition, 4 LLC, 2014.

2. Dalgaard, Peter, “Introductory statistics with R”, Springer Science & Business Media, 2008.

3. Samir Madhavan, “Mastering Python for Data Science”, Packt, 2015.

4. Paul Teetor, “R Cookbook, O’Reilly, 2011.