



TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

AN AUTONOMOUS INSTITUTION

Accredited by NAAC with 'A+' Grade.

(Sponsored by TKR Educational Society, Approved by AICTE, Affiliated to JNTU H)

Medbowli, Meerpet, Balapur, Hyderabad, Telangana – 500 097

Phone: 9100377790, email: info@tkrcet.ac.in, web site: www.tkrct.ac.in



B.TECH–COMPUTER SCIENCE & ENGINEERING (DATASCIENCE)

L/T/P/C

3/0/0/1

FUNDAMENTALS OF DATA SCIENCE (OPEN ELECTIVE)

Course Objectives:

- Learn concepts, techniques and tools they need to deal with various facets of data science

Course Outcomes:

1. Understand basic terms of statistical modeling and data science.L2
2. Apply the Data Types & Statistical Description in R.L3
3. Apply the Vectors, factors and list in R. L3
4. Apply the conditional statements and functions in R.L3
5. utilize R elements for data visualization and prediction.L3

UNIT- I

Introduction

Definition of Data Science- Big Data and Data Science hype – and getting past the hype - Datafication- Current landscape of perspectives - Statistical Inference - Populations and samples - Statistical modeling, probability distributions, fitting a model – Over fitting.

Basics of R: Introduction, R-Environment Setup, Programming with R, Basic Data Types.

UNIT- II Data Types & Statistical Description

Types of Data: Attributes and Measurement, Attribute, The Type of an Attribute, The Different Types of Attributes, Describing Attributes by the Number of Values, Asymmetric Attributes, Binary Attribute, Nominal Attributes, Ordinal Attributes, Numeric Attributes, Discrete versus Continuous Attributes. Basic Statistical Descriptions of Data: Measuring the Central Tendency: Mean, Median, and Mode, Measuring the Dispersion of Data: Range, Quartiles, Variance, Standard Deviation, and Interquartile Range, Graphic Displays of Basic Statistical Descriptions of Data.

UNIT- III

Vectors: Creating and Naming Vectors, Vector Arithmetic, Vector sub setting, Matrices: Creating and Naming Matrices, Matrix Sub setting, 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. Lists: Introduction, creating a List: Creating a Named List, Accessing List Elements, Manipulating List Elements, Merging Lists, Converting Lists to Vectors.

**TKR COLLEGE OF ENGINEERING AND TECHNOLOGY****AN AUTONOMOUS INSTITUTION**

Accredited by NAAC with 'A+' Grade.

(Sponsored by TKR Educational Society, Approved by AICTE, Affiliated to JNTU H)

Medbowli, Meerpet, Balapur, Hyderabad, Telangana – 500 097

Phone: 9100377790, email: info@tkrcet.ac.in, web site: www.tkrct.ac.in

**B.TECH–COMPUTER SCIENCE & ENGINEERING (DATASCIENCE)****UNIT- IV**

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.

UNIT- V

Charts and Graphs: Introduction, Pie Chart: Chart Legend, Bar Chart, Box Plot, Histogram, Line Graph: Multiple Lines in Line Graph, Scatter Plot.

Regression: Linear Regression Analysis, Multiple Linear regression

TEXT BOOKS:

1. Doing Data Science, Straight Talk from The Frontline. Cathy O’Neil and Rachel Schutt, O’Reilly, 2014.
2. K G Srinivas, G M Siddesh, “Statistical programming in R”, Oxford Publications.

REFERENCE BOOKS:

1. Jiawei Han, Micheline Kamber and Jian Pei. Data Mining: Concepts and Techniques, 3rd ed. The Morgan Kaufmann Series in Data Management Systems.
2. Introduction to Data Mining, Pang-Ning Tan, Vipin Kumar, Michael Steinbanch, Pearson Education.
3. Brain S. Everitt, “A Handbook of Statistical Analysis Using R”, Second Edition, 4 LLC, 2014.
4. Dalgaard, Peter, “Introductory statistics with R”, Springer Science & Business Media, 2008.
5. Paul Teetor, “R Cookbook”, O’Reilly, 2011.



TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

AN AUTONOMOUS INSTITUTION

Accredited by NAAC with 'A+' Grade.

(Sponsored by TKR Educational Society, Approved by AICTE, Affiliated to JNTU H)

Medbowli, Meerpet, Balapur, Hyderabad, Telangana – 500 097

Phone: 9100377790, email: info@tkrcet.ac.in, web site: www.tkracet.ac.in



B.TECH–COMPUTER SCIENCE & ENGINEERING (DATASCIENCE)

L/T/P/C

3/0/0/1

INTRODUCTION TO BIG DATA TECHNOLOGIES (OPEN ELECTIVE)

Course objectives:

- To teach the fundamental techniques and Principles in achieving big data analytics with scalability and streaming capability.

Course outcomes:

- Analyze the characteristic of Bigdata making use of data collection, preparation & visualization.
- Apply Analytical Theory and Methods in their Clustering and Associated Algorithms to solve problems encountered in the domain of big data analytics.L3
- Apply Analytical Theory and Methods in their classifications to solve problems of big data analytics.L3
- Interpret business models and scientific computing paradigms, and apply software tools for big data analytics.L3
- Achieve adequate perspectives of big data analytics in various applications like recommender systems, social media applications etc.L3

UNIT I

Introduction to Big Data, Characteristics of Data, and Big Data Evolution of Big Data, Definition of Big Data, Challenges with big data, Why Big data? Data Warehouse environment, Traditional Business Intelligence versus Big Data. State of Practice in Analytics, Key roles for New Big Data Ecosystems, Examples of big Data Analytics. Big Data Analytics, Introduction to big data analytics, Classification of Analytics, Challenges of Big Data, Importance of Big Data, Big Data Technologies, Data Science, Responsibilities, Soft state eventual consistency. Data Analytics Life Cycle

UNIT II

Analytical Theory and Methods I: Clustering and Associated Algorithms, Association Rules, Apriori Algorithm, Candidate Rules, Applications of Association Rules, Validation and Testing, Diagnostics, Regression, Linear Regression, Logistic Regression, Additional Regression Models..

UNIT III

Analytical Theory and Methods II: Classification, Decision Trees, Naïve Bayes, Diagnostics of Classifiers, Additional Classification Methods, Time Series Analysis, Box Jenkins methodology, ARIMA Model, Additional methods. Text Analysis, Steps, Text Analysis Example, Collecting Raw Text, Representing Text, Term Frequency-Inverse Document Frequency (TFIDF), Categorizing Documents by Topics, Determining Sentiments

UNIT IV

Data Product, Building Data Products at Scale with Hadoop, Data Science Pipeline and Hadoop Ecosystem, Operating System for Big Data, Concepts, Hadoop Architecture, Working with Distributed file system, Working with Distributed Computation, Framework for Python and Hadoop Streaming, Hadoop Streaming, MapReduce with Python, Advanced MapReduce. In-Memory Computing with



TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

AN AUTONOMOUS INSTITUTION

Accredited by NAAC with 'A+' Grade.

(Sponsored by TKR Educational Society, Approved by AICTE, Affiliated to JNTU H)

Medbowli, Meerpet, Balapur, Hyderabad, Telangana – 500 097

Phone: 9100377790, email: info@tkrcet.ac.in, web site: www.tkrct.ac.in



B.TECH–COMPUTER SCIENCE & ENGINEERING (DATASCIENCE)

Spark, Spark Basics, Interactive Spark with PySpark, Writing Spark Applications,

UNIT V

Distributed Analysis and Patterns, Computing with Keys, Design Patterns, Last-Mile Analytics, Data Mining and Warehousing, Structured Data Queries with Hive, HBase, Data Ingestion, Importing Relational data with Sqoop, Ingesting stream data with flume. Analytics with higher level APIs, Pig, Spark's higher level APIs..

TEXT BOOKS:

1. Big Data Science and Analytics A Hands-on Approach. By Arshdeep Bahga, Vijay Madiseti



TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

AN AUTONOMOUS INSTITUTION

Accredited by NAAC with 'A+' Grade.

(Sponsored by TKR Educational Society, Approved by AICTE, Affiliated to JNTU H)

Medbowli, Meerpet, Balapur, Hyderabad, Telangana – 500 097

Phone: 9100377790, email: info@tkrcet.ac.in, web site: www.tkrct.ac.in



B.TECH–COMPUTER SCIENCE & ENGINEERING (DATASCIENCE)

L/T/P/C

3/0/0/1

SOCIAL MEDIA ANALYTICS (OPEN ELECTIVE)

Course Objectives: Exposure to various web and social media analytic techniques.

Course Outcomes:

1. Knowledge on decision support systems.L2
2. Apply natural language processing concepts on text analytics.L3
3. Understand sentiment analysis.L2
4. Knowledge on Web Analytics, Web Mining.L2
5. Knowledge on search engine optimization and web analytics.L2

UNIT - I

An Overview of Business Intelligence, Analytics, and Decision Support: Analytics to Manage a Vaccine Supply Chain Effectively and Safely, Changing Business Environments and Computerized Decision Support, Information Systems Support for Decision Making, The Concept of Decision Support Systems (DSS),

UNIT - II

Text Analytics and Text Mining: Machine Versus Men on Jeopardy!: The Story of Watson, Text Analytics and Text Mining Concepts and Definitions, Natural Language Processing, Text Mining Applications, Text Mining Process, Text Mining Tools.

UNIT - III

Sentiment Analysis: Sentiment Analysis Overview, Sentiment Analysis Applications, Sentiment Analysis Process, Sentiment Analysis and Speech Analytics.

UNIT - IV

Social Analytics and Social Network Analysis: Social Analytics and Social Network Analysis, Social Media Definitions and Concepts, Social Media Analytics.

UNIT - V

Prescriptive Analytics - Optimization and Multi-Criteria Systems: Multiple Goals, Sensitivity Analysis, What-If Analysis, and Goal Seeking.

TEXT BOOK:

1. Ramesh Sharda, Dursun Delen, Efraim Turban, BUSINESS INTELLIGENCE AND ANALYTICS: SYSTEMS FOR DECISION SUPPORT, Pearson Education.



TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

AN AUTONOMOUS INSTITUTION

Accredited by NAAC with 'A+' Grade.

(Sponsored by TKR Educational Society, Approved by AICTE, Affiliated to JNTU H)

Medbowli, Meerpet, Balapur, Hyderabad, Telangana – 500 097

Phone: 9100377790, email: info@tkrcet.ac.in, web site: www.tkrct.ac.in



B.TECH–COMPUTER SCIENCE & ENGINEERING (DATASCIENCE)

REFERENCE BOOKS:

1. Rajiv Sabherwal, Irma Becerra-Fernandez,” Business Intelligence – Practice, Technologies and Management”, John Wiley 2011.
2. Lariss T. Moss, ShakuAtre, “Business Intelligence Roadmap”, Addison-Wesley It Service.
3. Yuli Vasiliev, “Oracle Business Intelligence: The Condensed Guide to Analysis and Reporting”, SPD Shroff, 2012.



TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

AN AUTONOMOUS INSTITUTION

Accredited by NAAC with 'A+' Grade.

(Sponsored by TKR Educational Society, Approved by AICTE, Affiliated to JNTU H)

Medbowli, Meerpet, Balapur, Hyderabad, Telangana – 500 097

Phone: 9100377790, email: info@tkrcet.ac.in, web site: www.tkrct.ac.in



B.TECH–COMPUTER SCIENCE & ENGINEERING (DATASCIENCE)

L/T/P/C

3/0/0/1

DATA VISUALIZATION(OPEN ELECTIVE)

Course Objectives:

To understand the visual representation of structured and unstructured data.

Course Outcomes:

After completion of course, the students will be able to

1. Understand the visualization and Data basics L2
2. Understand the Visualization process and know the representation of Spatial & Geo spatial data L2
3. Analyze various Visualization techniques for Multivariate data and other structures of data .L4
4. Interacting the different operators and different data spaces.L2
5. Design effective visualization of modern toolkits.L3

UNIT I

Introduction

What is Visualization, History, Relationship visualization with other fields, The visualization Process, Pseudocode Conventions, The Scatter plot **Data Foundations:** Types of Data, Structure within and between the records, Data Processing.

UNIT II

Visualization Foundations

The Visual Process, Semiology of Graphical Symbols, The Eight Visual Variables, Historical Perspective, Taxonomies. Visualization Techniques for Spatial Data:One-Dimensional Data, Two-Dimensional Data, Three- Dimensional Data, Dynamic Data, Combining Techniques.

Visualization Techniques for Geospatial Data: Visualizing Spatial Data, Visualization of Point Data, Visualization of Line Data, Visualization of Area Data.

UNIT III

Visualization Techniques for Multivariate Data

Point-Based Techniques, Line-Based Techniques, Region-Based Techniques, Combinations of Techniques, Visualization Techniques for Trees, Graphs, and Networks: Displaying Hierarchical Structures, Displaying Arbitrary Graphs/Networks.

UNIT IV

Text and Document Visualization

Levels of Text Representation, The Vector Space Model, Single Document Visualizations, Document Collection Visualizations. **Interaction Concepts:** Interaction Operators, Interaction Operands and Spaces, A Unified Framework. **Interaction Techniques:** Object Space, Data Space, Attribute Space, Data Structure Space, Visualization Structure Space, Animating Transformations, Interaction Control.

**TKR COLLEGE OF ENGINEERING AND TECHNOLOGY****AN AUTONOMOUS INSTITUTION**

Accredited by NAAC with 'A+' Grade.

(Sponsored by TKR Educational Society, Approved by AICTE, Affiliated to JNTU H)

Medbowli, Meerpet, Balapur, Hyderabad, Telangana – 500 097

Phone: 9100377790, email: info@tkrcet.ac.in, web site: www.tkrct.ac.in**B.TECH–COMPUTER SCIENCE & ENGINEERING (DATASCIENCE)****UNIT V****Designing Effective Visualizations**

Steps in Designing Visualizations, Problems in Designing Effective Visualizations **Comparing and Evaluating Visualization Techniques:** User Tasks, User Characteristics, Data Characteristics, Visualization Characteristics, Structures for Evaluating Visualizations. **Visualization Systems:** Systems Based on Data Type, Systems Based on Analysis Type, Text Analysis and Visualization, Modern Integrated Visualization Systems, Toolkits

TEXT BOOKS:

1. Interactive Data Visualization Foundations, Techniques, and Applications by Grinstein, Georges Keim, Daniel Ward, Matthew O , CRC Press Taylor & Francis Group.
2. Digital Image. Processing. Third Edition. Rafael C. Gonzalez. University of Tennessee. Richard E. Woods. NledData Interactive. Pearson International Edition