**ACADEMIC PLANNER**

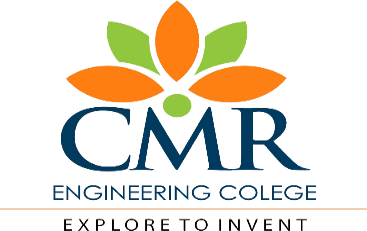
ON

***“*Data Analytics Using R*”***

**(AI504PC)**

Department of

**Computer Science & Engineering (AI&ML)**



**CMR ENGINEERING COLLEGE**

(Approved by AICTE-New Delhi, Affiliated to JNTU, Hyderabad)

Kandlakoya(v), Medchal Road,Hyderabad-501401,Telangana State,India. Website:[www.cmrec.ac.in](http://www.cmrec.ac.in/)

**(2025-26)**

**ACADEMIC PLANNER**

**Subject: Data Analytics Using R (III-B.Tech I Sem)**

**S.NO CONTENTS**

1. **- Preamble/Introduction**

# - Prerequisites

1. **- Objectives and Outcomes**

# - Syllabus

* 1. **R22-CMREC**

# GATE

* 1. **IES(Not Applicable)**

1. **- List of Expert Details**(Local/National/International with Contact details/Profile link/Blogs/their research

Contribution towards the subject)

# - Journals with min 5 ref paper for literature study

1. **- Subject -Lesson plan**
2. **- Suggested Books (**prescribed and References)

# - Websites for self learning Resources like

[*www.geeksforgeeks.org,*](http://www.geeksforgeeks.org/) [*www.schools.com,*](http://www.schools.com/) [*Coursera*](https://www.theeducationmagazine.com/word-art/best-educational-websites/#Coursera)*,*[*edX*](https://www.theeducationmagazine.com/word-art/best-educational-websites/#edX)*,* [*Udemy*](https://www.theeducationmagazine.com/word-art/best-educational-websites/#Udemy)*,* [*Khan Academy*](https://www.theeducationmagazine.com/word-art/best-educational-websites/#Khan%20Academy), NPTEL etc along Registration procedures*)*

# - Question Banks 1.CMREC/JNTUH/Model papers

**2.GATE**

# - Two case study presentations with Project / Product/ Model /prototypes/ Industrial applications.

1. **- Assignment Question/Innovative Assignments sets.**

# - List of topics for students Seminars with Guidelines

1. **- STEP/Course material in softcopy**

# - Expert Lectures with topics &Schedules (if any)

1. **Preamble/Introduction**

Data Analytics is the process of examining raw data to find the patterns, draw conclusions and make the predictions. It helps businesses to make better decisions, improve efficiency & gain deeper knowledge understanding their customer and operations.

# PREREQUISITES

* + Database Management Systems
  + Knowledge of probability and statistics.
  + Knowledge on Machine Learning concepts.

1. **OBJECTIVES AND OUTCOMES**

**COURSE OBJECTIVES**

* + To explore the fundamental concepts of data analytics
  + To learn the principles and methods of statistical analysis
  + Discover interesting patterns, analyze supervised and unsupervised models and estimate the accuracy of the algorithms.
  + To understand the various search methods and visualization techniques.

**COURSE OUTCOMES**

**CO1:** **Describe** the Concepts of Data Management

**CO2:** **Use** different techniques to Purify the raw Data.

**CO3:** **Analyze** the Various Machine Learning Models

**CO4:** **Categorize** the different Learning Techniques

**CO5:** **Apply** various Visualization Techniques to Generate Graphs.

# SCOPE

The Scope of this subject is to provide understanding of fundamental concepts of data analytics.

# (4.1) SYLLABUS – CMREC

**UNIT I**

**Data Management**: Design Data Architecture and manage the data for analysis, understand various sources of Data like Sensors/Signals/GPS etc. Data Management, Data Quality (noise, outliers, missing values, duplicate data) and Data Processing & Processing.

**UNIT II**

**Data Analytics**: Introduction to Analytics, Introduction to Tools and Environment, Application of Modeling in Business, Databases & Types of Data and variables, Data Modeling Techniques, Missing Imputations etc. Need for Business Modeling.

**UNIT III**

**Regression** — Concepts, Blue property assumptions, Least Square Estimation, Variable Rationalization, and Model Building etc.

Logistic Regression: Model Theory, Model fit Statistics, Model Construction, Analytics applications to various Business Domains etc.

**UNIT IV**

**Object Segmentation**: Regression Vs Segmentation — Supervised and Unsupervised Learning, Tree Building — Regression, Classification, Overfilling, Pruning and Complexity, Multiple Decision Trees etc. Time Series Methods: Arima, Measures of Forecast Accuracy,

STL approach, Extract features from generated model as Height, Average Energy etc and Analyze for prediction

**UNIT V**

**Data Visualization:** Pixel-Oriented Visualization Techniques, Geometric Projection Visualization Techniques, Icon-Based Visualization Techniques, Hierarchical Visualization Techniques, Visualizing Complex Data and Relations.

# (4.2) SYLLABUS – GATE

**Not applicable**

# (4.3) SYLLABUS - IES

**Not applicable**

# LIST OF EXPERT DETAILS

**INTERNATIONAL**

1. “A Practical guide to Data Analysis Using R”John H. Maindonald, Statistics Research Associates, Wellington, New Zealand Cambridge University Press.
2. Gregory piatetsky-shapironational Michael Brodie, leading database researcher, industry leader, thinker. [SIGKDD](https://dblp.org/db/journals/sigkdd/sigkdd16.html#Piatetsky-Shapiro14), KDnuggets, Brookline, MA, USA.

**NATIONAL**

1. “Data Analytics Using R" Author: Seema Acharya Publisher: McGraw Hill Education (India).

2. Prof. D.Lakshmi, Professor, VIT, Bhopal Email: lakshmi.d@vitbhopal.ac.in Mobile No: 9945379089.

**REGIONAL**

1. Mr. Rajesh Prabhakar Kaila, Times Center for Learning Limited, Times of India Group, Visiting Faculty Member (Business Analytics & Finance) at Symbiosis International University, Hyderabad.
2. Dr. A.V.Krishna Prasad Faculty of CSE Dept, MVSR Engineering College, Hyderabad, MC member for CSI Hyderabad Chapter.
3. **JOURNALS**
4. <https://link.springer.com/article/10.1007/s10479-023-05390-7?utm_source=chatgpt.com>
5. <https://www.researchgate.net/publication/352258047_Data_Analytics_using_R_Programming>
6. <https://www.tandfonline.com/doi/epdf/10.1080/26939169.2022.2089410?needAccess=true>
7. <https://www.researchgate.net/publication/270272226_Data_Analysis_with_R_Selected_Topics_and_Examples>
8. <https://researchmethodscommunity.sagepub.com/blog/r-and-data-analysis-open-access-articles>

# SUBJECT (LESSON) PLAN

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S.NO** | **Topic (CMREC Syllabus)** | **Sub-Topic** | **No. Of lectures required** | | **Suggested Books** | **Teaching Methods** |
| **UNIT – I(11)** | | | | | | |
| **1** | **Data Management** | Introduction to Data Management. | **L1** | | **T1** | **M1** |
| **2** | Design Data Architecture and manage the data for analysis. | **L2-L3** | | **T1** | **M1** |
| **3** | understand various sources of Data like Sensors/Signals/GPS etc. | **L4-L5** | | **T1** | **M2(PPT)** |
| **4** | Data Management, Data Quality noise. | **L6-L7** | | **T1** | **M1** |
| **5** | outliers, missing values, duplicate data) | **L8-L9** | | **T1,R1** | **M2(PPT)** |
| **6** | Data Processing | **L10** | | **T1** | **M1** |
| **7** | Data Pre-Processing | **L11** | | **T1** | **M2(PPT)** |
| **UNIT – II(11)** | | | | | | |
| **8** | **Data Analytics** | Introduction to Analytics | **L12** | | **T1** | **M1** |
| **9** | Introduction to Tools and Environment | **L13-L14** | | **T1** | **M1** |
| **10** | Application of Modelling in Business | **L15-L16** | | **T1,R2** | **M2(PPT)** |
| **11** | Databases & Types of Data and variables | **L17-L18** | | **T1,R3** | **M1** |
| **12** | Data Modelling Techniques | **L19-20** | | **T1** | **M2(PPT)** |
| **13** | Missing Imputations etc | **L21** | | **T1** | **M2(PPT)** |
| **14** | Need for Business Modelling | **L22** | | **T1** | **M2(PPT)** |
|  | **Regression** | **UNIT – III(14)** | | | | |
| **15** | Regression Concepts | **L23-L24** | | **T1** | **M1** |
| **16** | Blue property assumptions | **L25-L26** | **T1,R3** | | **M1** |
| **17** | Least Square Estimation | **L27-L28** | | **T1** | **M1** |
| **18** | Variable Rationalization, Model Building etc. | **L29-L30** | | **T1,R2** | **M1(PPT)** |
| **19** | Logistic Regression: Model Theory | **L31-L32** | | **T1** | **M1(PPT)** |
| **20** |  | Model fit Statistics, Model Construction | **L33-L34** | | **T1,R2** | **M2(PPT)** |
| **21** |  | Analytics applications to various Business Domains etc. | **L35-L36** | | **T1,R2** | **M2(PPT)** |
|  | **Object Segmentation** | **UNIT-IV(13)** | | | | |
| **22** | Regression Vs Segmentation, Supervised and Unsupervised Learning | **L37-L38** | | **T1** | **M1** |
| **23** | Tree Building – Regression, Classification | **L39-40** | | **T1** | **M1** |
| **24** | Over fitting, Pruning and Complexity | **L41- L42** | | **T1** | **M1** |
| **25** | Multiple Decision Trees etc. | **L43** | | **T1** | **M1** |
| **26** | Time Series Methods: Arima, Measures of Forecast Accuracy | **L44-L45** | | **T1,R2** | **M1** |
| **27** | STL approach, Extract features from generated model as Height | **L46-L47** | | **T1,R2** | **M1** |
| **28** | Average Energy etc and Analyze for prediction | **L48-L49** | | **T1,R1** | **M1** |
| **UNIT- V(07)** | | | | | | |
| **29** | **Data Visualization** | Pixel-Oriented Visualization Techniques | **L50-L51** | | **TI** | **M1** |
| **30** | Geometric Projection Visualization Techniques | **L52-L53** | | **T1** | **M1** |
| **31** | Icon-Based Visualization Techniques | **L54** | | **T1** | **M1** |
| **32** | Hierarchical Visualization Techniques | **L55** | | **T1,R1** | **M1** |
| **33** | Visualizing Complex Data and Relations | **L56** | | **T1,R3** | **M1** |

**METHODS OF TEACHING:**

|  |  |  |
| --- | --- | --- |
| **M1 : Lecture Method** | **M4 : Presentation /PPT** | **M7 : Assignment** |
| **M2 : Demo Method** | **M5 : Lab/Practical** | **M8 : Industry Visit** |
| **M3 : Guest Lecture** | **M6 : Tutorial** | **M9 : Project Based** |

**NOTE:**

1. Any Subject in a Semester is supposed to be completed in 51 to 58 Periods.

1. Each Period is of 50 minutes.
2. Each unit duration & completion should be mentioned in the Remarks column.
3. List of Suggested books can be marked with Codes like T1, T2, R1, R2 etc.

# SUGGESTED BOOKS Text books

**T1:** Student's Handbook for Associate Analytics — II, III.

**T2:** Data Mining: Concepts and Techniques, 3rd ed. - Jiawei Han and Kamber Morgan Kauffmann .Publishers.

**T3:** Seema Acharya Mcgrawhill Publication Data Analytics Using R.

**REFERENCE BOOKS**

**R1.** Introduction to Data Mining Tan, Steinbach and Kumar Addison Welsy 2006 **R2.** Data Mining Analysis and Concepts M. Zaki and W. Meira

**R3.** Mining on Massive Datasets Jure Leskovec Stanford Univ, Anand Rajaram Milliway Labs, Jeffery D Ullman Stanford University.

# WEBSITES

* + <https://www.coursera.org/browse/data-science/data-analysis>
  + <https://onlinecourses.nptel.ac.in/noc21_cs45/preview>
  + <https://www.coursera.org/professional-certificates/google-data-analytics>
  + <https://www.udemy.com/topic/data-analysis/>
  + <https://www.geeksforgeeks.org/r-tutorial/?ref=dhm>
  + <https://www.analyticsvidhya.com/blog/2023/04/what-is-data-analytics/>
  + <https://datavizcatalogue.com/>
  + <https://online.hbs.edu/blog/post/data-visualization-techniques>
  + <https://online.stanford.edu/courses/soe-ycs0007-mining-massive-data-sets>

# QUESTION BANK

# PREVIOUS QUESTION PAPERS

# 

# 

**CMR ENGINEERING COLLEGE: : HYDERABAD**

**UGC AUTONOMOUS**

**III–B.TECH–I–Semester End Examinations (Regular) - December- 2022**

**DATA ANALYTICS USING R**

**(CSE)**

**[Time: 3 Hours] [Max. Marks: 70]**

**Note:** This question paper contains two parts A and B.

Part A is compulsory which carries 20 marks. Answer all questions in Part A.

Part B consists of 5 Units. Answer any one full question from each unit. Each question   
 carries 10 marks.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. a) | How can the default path to package library be changed in R? | | | [2M] |
| b) | List out two IDEs for R. | | | [2M] |
| c) | What are mean and median with a neat example? | | | [2M] |
| d) | What are the advantages of using data visualization? | | | [2M] |
| e) | Give the general equation for computing linear regression? | | | [2M] |
| f) | What is the syntax of lm() function? | | | [2M] |
| g) | What is a three-way contingency table? | | | [2M] |
| h) | What are the major diagnostic functions of the ‘LogisticDx’ package? | | | [2M] |
| i) | Name the packages used to build decision trees in R? | | | [2M] |
| j) | List out the names of learning algorithms that create a decision tree. | | | [2M] |
| **PART-B** **(50 Marks)** | | | | |
| 2. | a) Explain RSQLite package.  b) Explain the commands using R: summary (), str (), head (), tail (), view (), edit () | | | [5M]  [5M] |
| **OR** | | | | |
| 3. | Create a dataset, ‘Watch’ and store the information about watches of four different companies. Explain all the steps of simple analytical data processing from input to output on this dataset. | | | [10M] |
|  |  |  |  |  |
| 4. | a) What are the data frames? Write its significance in R-Language?  b) Explain the graphical techniques used by Exploratory Data Analysis using R. | | | [5M]  [5M] |
| **OR** | | | | |
| 5. | What is bar chart? Discuss the various types of bar charts using R? | | | [10M] |
|  |  |  |  |  |
| 6. | Compare and Contrast Multiple R-squared and Adjusted R-squared. | | | [10M] |
| **OR** | | | | |
| 7. | What is model Fitting? Explain various models and their commands in R. | | | [10M] |
|  |  |  |  |  |
| 8. | Create a table with a ‘pizza’ column that stores the information that is necessary to implement multinomial logistics regression. After placing the information, implement multinomial logistics regression on this table. | | | [10M] |
| **OR** | | | | |
| 9. | a) Explain binary logistic regression with a single categorical variable.  b) Explain about likelihood function. | | | [5M]  [5M] |
|  |  |  |  |  |
| 10. | Create a dataset that contains the features of apples. Now find out the “entropy” and “information gain” for this dataset. Also, find out the best feature of the apple dataset. | | | [10M] |
| **OR** | | | | |
| 11. | Write and explain ID3 decision tree construction algorithm. | | | [10M] |
| \*\*\*\*\*\*\*\*\*\*\*\* | | | | |

# 

|  |
| --- |
| **UNIT-1**  **Short Questions**   1. What is Data Management? 2. Discuss different Types of Data? 3. List out Enterprise Requirements. 4. What is Data Quality? 5. What did you understand by Data preprocessing?   **Long Questions**   1. Explain the sources of primary Data. 2. Explain Data Architecture in detail. 3. Write about data preprocessing needs. 4. Explain in detail for generating primary data. 5. Explain data architecture in detail. 6. Explain elements of data architecture. 7. Explain Survey methods and experimental method. 8. Explain the sources of secondary data. 9. Explain GPS and signal data. 10.What is data quality? Explain. |
| **UNIT-2**  **Short Questions**   1. What is Data Analytics. Explain? 2. Discuss different Types of Data? 3. Discuss different types of data. 4. What is Missing Imputation? 5. List the need for business modelling   **Long Questions**   1. Explain about Data modelling techniques? 2. Explain the multiple analytical methodologies 3.Explain supervised and unsupervised learning?   4.List out Machine learning tasks and explain.  5.Explain about Applications of Data modelling.   1. Distinguish between Machine learning and Data mining. 2. What are the different types of data. Explain? 3. Explain in detail about tools and environment. 4. Explain about Missing Imputations in detail. 5. Discuss about Data & Variable in detail.   **UNIT-3**  **Short Questions**   * 1. What is Linear regression?   2. Discuss about Logical regression?   3. What is Supervised learning. Explain with Example?   4. What is Unsupervised learning. Explain with Example?   5. Differentiate regression &classification?   **Long Questions**  1. Explain the different steps involved in Model Construction. 2. Demonstrate Tree Building & Pruning process in detail?  3. List the various applications of Data Analytics?  4. Discuss in detail about Least Square estimation.  5. Explain BLUE Properties in detail? |

**UNIT-4**

**Short Questions**

1. What is STL Approach?
2. Define Pruning with an Example?
3. Elaborate pruning and complexity
4. What is Multiple Decision tree.
5. How do you extract features from STL Approach. Discuss.

**Long Questions**

1. Define Time Series Analysis? Describe the process of ARIMA model?

2. Explain binary logistic regression with a covariate variable.

3. Write and explain ID3 decision tree construction algorithm.

4. Explain binary logistic regression with a single categorical variable.

5. Create a table with a ‘person’ column that stores the information like name, age, gender, annual

income and other. Implement the binary logistic regression with single categorical and three-way

contingency table after placing the required information on the table.

**UNIT-5**

**Short Questions**

1. Define Dendrogram?

2. Why do we use Data Visualization?

3.Which packages build decision trees in R?

4. What is Visualizing complex data?

5. What do you mean by discrete value?

**Long Questions**

1. What are the benefits of Data Visualization. Explain?

2. What are the various Data Visualization Techniques? Explain?

3. Q10.What are the advantages &amp; disadvantages of Data Visualization?

4. Create a dataset that contains discrete values. Generate the decision tree for it.

5. Explain Icon-based Visualization and Hierarchal visualization techniques in detail.

1. **CASE STUDIES**

**1. Case Study: Leveraging Data Analytics in the E-commerce Industry**

The e-commerce industry has rapidly embraced data analytics to drive business decisions, enhance customer experience, and optimize operations. A leading online retail company implemented a data-driven approach to improve customer retention and increase conversion rates. Using R and Python, the company analyzed user behavior from web logs, click streams, and past purchase data.

Through customer segmentation using K-means clustering, the firm identified key buyer personas. Predictive models such as logistic regression and decision trees were used to estimate the likelihood of cart abandonment. Real-time recommendation systems powered by collaborative filtering were developed to personalize product suggestions.

As a result, the company achieved a 25% increase in conversion rate and a 15% reduction in cart abandonment. Moreover, inventory management was improved by using time series forecasting models to predict product demand, which reduced overstocking and stock outs.

This case highlights how the integration of data analytics into e-commerce operations empowers businesses to make strategic decisions, deliver personalized experiences, and optimize supply chains. The ability to process and analyze vast datasets in real-time has become a key differentiator in the highly competitive e-commerce landscape.

**2. Case Study: Enhancing Patient Health Diagnosis through Data Analytics**

A multispecialty hospital implemented data analytics to improve the accuracy and speed of patient health diagnosis. The hospital collected large volumes of structured data (lab results, vitals, EHRs) and unstructured data (doctor notes, radiology reports). Using R and Python, data scientists applied machine learning models to predict the likelihood of chronic diseases such as diabetes and heart disease.

The process began with data preprocessing, where missing values, duplicates, and inconsistencies were addressed. Feature engineering was then applied to extract meaningful variables such as BMI, glucose levels, and lifestyle indicators. Logistic regression, random forests, and support vector machines were used to build predictive models.

For example, using a dataset of over 20,000 patients, the hospital’s model predicted Type 2 diabetes with over 88% accuracy. Clinicians were provided with risk scores and diagnostic insights through an R Shiny dashboard, enabling early intervention for high-risk patients.

As a result, the hospital reduced diagnostic time by 30% and improved patient outcomes through timely care. This case demonstrates how integrating analytics into clinical workflows can significantly enhance decision-making, personalize treatment, and reduce healthcare costs.

1. **ASSIGNMENT QUESTIONS**

**Assignment-I Questions**

1a. What is Data Management? What are the benefits of good Data Management?

b. Explain the various sources of Date in detail?

2. a. How to design Data Architecture? What are the factors influence the data Architecture?

b. What is Pre-Processing? What are the different phases in Data Pre-Processing?

3. What are the various imputations techniques in Data Analytics?

4. Explain different types of data analytics?

5. a. Define Regression? How to draw a best-fit regression line and explain least square estimation?

b. Find the line of best fit for the following data of heights and weights of students of a school using the Least Square method:

Height (in centimeters): [160, 162, 164, 166, 168]

Weight (in kilograms) : [52, 55, 57, 60, 61]

**Assignment-II Questions**

1. What is a Decision Tree? Illustrate the process of building a Decision Tree?

2. What is Supervised Learning? Explain the various types of Supervised Learning techniques?

3. What is a Pixel based Visualization Technique? Explain?

4. List the various types of visualization methods with neat diagrams?

5. What is a Time Series Analysis? Explain ARIMA model to forecast the data?

# INNOVATIVE ASSIGNMENT QUESTIONS

* 1. Analyze any dataset of your interest available on Kaggle dataset.

# 2. Build logistic regression model for prediction of heart disease.

# TOPICS FOR STUDENT’S SEMINARS:

1. Different Sources of Data
2. Data Analytics Techniques
3. Types of Data Analytics
4. Linear and Logistic Regression Model
5. Decision trees
6. Analysis of real time dataset
7. ARMIA Model & STL Approach
8. Data Visualization Techniques

**14-STEP/Course material in softcopy**

****

**15-Expert Lectures with topics &Schedules (if any)**

A Topic on “Explorative Data Analytics” by Professor. Lakshmi, Professor, VIT BHOPAL by 20th of August 2025.