**SRR & CVR GOVT DEGREE COLLEGE(A):: VIJAYAWADA**

**I BSc (Data Science) Semester -II**

**Revised Syllabus 2021-22**

**INTRODUCTION TO DATA SCIENCE WITH R**

**MODEL PAPER**

**Time: 3hrs Max Marks:60**

**SECTION-A**

**Answer Any FIVE of the following Questions 5 X 4= 20 marks**

1. Define data science and applications of data science?
2. List the sources of data.
3. Write about RStudio default display panes?
4. Write about Help functions in R?
5. How to use apply() function on matrices in R. Explain with an example
6. Explain about Factors
7. What is the use of summary command in R
8. What is a recursive function? Explain
9. Explain about F-statistics
10. Explain about correlation matrix

**SECTION – B**

**Answer All the following questions 5 X 8=40M.**

1. a) Explain about the Data Science life cycle?

(OR)

b) i) Explain about types of data? ii) classification of digital data

1. a) Explain about data structures in the R language ?

(OR)

b) What is Vector and explain about common vector operations ?

1. a) Explain about various functions applied on matrix rows and columns?

(OR)

b) What is Data Frame ? Explain procedure to create Data Frame with example?

1. a) What Is Data Preprocessing & What Are The Steps Involved?

(OR)

b) How is R used in data visualisation?

1. a) Explain in detail about linear regression and multiple linear regression.

(OR)

b) i) Explain how to handle packages in R ii) Explain about types of Learning

**SRR & CVR GOVT DEGREE COLLEGE(A):: VIJAYAWADA**

**I BSc (Data Science) Semester -II**

**Revised Syllabus 2021-22**

**INTRODUCTION TO DATA SCIENCE WITH R**

**Question Bank**

**SECTION-A**

**SHORT ANSWER QUESTIONS - 4 MARKS**

Unit-I

1. Define data science and applications of data science?
2. write short notes on data collection ?
3. List the sources of data.

Unit II:

1. Write about RStudio default display panes?
2. Write about Help functions in R?
3. Explain about break and next statements in R
4. Explain about Vector recycling ?
5. How to declare and call a function in R?
6. How to declare Vector in R and vector indexing?
7. How to apply the ifelse() function on vectors? Explain with examples.
8. Explain about NA and NULL Values
9. Write short notes on Filtering ?

Unit III

1. How to use apply() function on matrices in R. Explain with an example
2. Explain procedure to add and delete rows and columns in matrix ?
3. Write short notes on recursive lists R?
4. Explain about merging of Data Frames in R?
5. Briefly explain the usage of table() function in R Language
6. Differentiate between split() and tapply() in R
7. Explain about Factors

Unit - IV

1. What is the use of summary command in R
2. Explain common measures of variability
3. How to Read and Write datasets in various formats in R
4. Explain about measures of central tendency
5. Explain about recursive functions
6. What is Exploratory Data analysis
7. Unit - V
8. Explain about F-statistics
9. Explain about correlation matrix

**LONG ANSWER QUESTIONS 8 MARKS**

Unit I

1. Explain about the Data Science life cycle?
2. Explain about different types of Databases in data Science?
3. i) Explain about types of data? ii) classification of digital data

Unit II

1. Explain about data structures in the R language ?
2. Explain about R operators ?
3. Explain about Control Structures in R ?
4. What is Vector and explain about common vector operations ?

Unit III

1. Explain about various functions applied on matrix rows and columns?
2. What is List? Explain about various operations on Lists?
3. What is Data Frame ? Explain procedure to create Data Frame with example?

Unit-IV

1. What Is Data Preprocessing & What Are The Steps Involved?
2. What are the steps involved in Data Cleaning. Explain in Detail
3. How is R used in data visualisation?

Unit-V

1. Explain in detail about linear regression and multiple linear regression.
2. Explain how to handle packages in R
3. Explain about types of Learning

**SRR & CVR GOVT DEGREE COLLEGE(A):: VIJAYAWADA**

**SEMESTER-III**

**COURSE 7: DATA MINING TECHNIQUES USING R**

Theory Credits: 3 3 hrs/week

**Aim and objectives of Course:**

* To understand Data mining techniques and algorithms.
* Comprehend the data mining environments and applications.

**Learning outcomes of Course:**

* Students who complete this course will be able to
* Compare various conceptions of data mining as evidenced in both research and application.
* Evaluate mathematical methods underlying the effective application of data mining.
* Should be able to apply the type of techniques based on the problems considered.
* Can find out the market patterns and association amongst different products.

**UNIT I:**

An idea on Data Warehouse, Data mining-KDD versus data mining, Stages of the Data MiningProcess-Task primitives., Data Mining Techniques – Data mining knowledge representation.

**UNIT II**

Data mining query languages- Integration of Data Mining System with a Data Warehouse- Issues, Data pre-processing – Data Cleaning, Data transformation – Feature selection – Dimensionality reduction

**UNIT III**

Concept Description: Characterization and comparison What is Concept Description, Data Generalization by Attribute-Oriented Induction(AOI), AOI for Data Characterization, Efficient Implementation of AOI.

Mining Frequent Patterns, Associations and Correlations: Basic Concepts, FrequentItemset Mining Methods: Apriori method, generating Association Rules, Improvingthe Efficiency of Apriori,Pattern-Growth Approach for mining Frequent Item sets.

**UNIT-IV**

Classification Basic Concepts: Basic Concepts, Decision Tree Induction: Decision

TreeInduction Algorithm, Attribute Selection Measures, Tree Pruning. Bayes Classification Methods.

**UNIT-V**

Association rule mining: Antecedent, consequent , muti-relational association rules,

ECLAT.Case study on Market Basket Analysis.

Cluster Analysis: Cluster Analysis, Partitioning Methods, Hierarchal methods, Density based methods-DBSCAN.

**TEXT BOOKS:**

1. Jiawei Han, MichelineKamber, Jian Pei.“Data Mining: Concepts and Techniques”, 3

rd Edition,Morgan Kaufmann Publishers, 2011.

2. AdelchiAzzalini, Bruno Scapa, “Data Analysis and Data mining” , 2 ndEdiiton, Oxford Univeristy Press Inc., 2012.

3. Data Mining, The Textbook (2015) by Charu Aggarwal.

**REFERENCES BOOKS:**

1. Alex Berson and Stephen J. Smith, “Data Warehousing, Data Mining & OLAP”, 10th

Edition, TataMcGraw Hill Edition , 2007.

2. G.K. Gupta, “Introduction to Data Mining with Case Studies”, 1st Edition, EasterEconomy Edition, PHI, 2006.

**SRR & CVR GOVT DEGREE COLLEGE(A):: VIJAYAWADA**

**B.Sc (DATA SCIENCE) II YEAR III SEMESTER**

**Course 7 : DATA MINING TECHNIQUES USING R**

**MODEL QUESTION PAPER**

**Time: 3 hrs Max. Marks: 70**

**Section - A**

**Answer Any Five Questions: 5 x 4 = 20M**

1. Explain about Data Mining Knowledge Representation
2. What are Data Mining Task Primitives? Explain
3. Explain DMQL
4. What is dimensionality reduction? What is its use?
5. What is Concept Description?
6. Explain the basic principle of AOI
7. List attribute selection measure in Decision Tree and explain one in detail.
8. Explain about Tree Pruning
9. Explain steps in K-Means clustering algorithm
10. Write about Grid based clustering methods

**SECTION - B**

**Answer any ONE question from each unit**

**Each answer carries 10 marks 5 x 10 = 50M**

1. a) Explain the stages of KDD process

(OR)

b) Explain in detail about Data Warehouse

1. a) Explain about Data Cleaning in detail.

(OR)

b) Explain about Data Transformation in detail

1. a) Explain about the Apriori algorithm for finding frequent item sets with an example

(OR)

b) Explain FP Growth Algorithm

1. a) Explain about Decision Tree Induction Algorithm

(OR)

b) Explain about Bayesian classification

1. a) Explain about Hierarchical methods in cluster analysis

(OR)

b) Explain the ECLAT algorithm using a simple example of market basket analysis?

**SRR & CVR GOVT DEGREE COLLEGE(A):: VIJAYAWADA**

**B.Sc (DATA SCIENCE) II YEAR III SEMESTER**

**Course 7 : DATA MINING TECHNIQUES USING R**

**QUESTION BANK**

SHORT ANSWER QUESTIONS

1. Explain about Data Mining Knowledge Representation
2. What are Data Mining Task Primitives? Explain
3. Explain the concept of Data Warehouse
4. Explain DMQL
5. Explain about Feature Selection
6. What is dimensionality reduction? What is its use?
7. Explain about Antecedent, consequent in association rules
8. Explain about Support and confidence in association rules
9. What is Concept Description?
10. Explain the basic principle of AOI
11. List attribute selection measure in Decision Tree and explain one in detail.
12. Explain about Tree Pruning
13. Explain about partitioning scenarios in Decision Tree
14. Explain steps in K-Means clustering algorithm
15. Write about Grid based clustering methods
16. Write about Density Based Clustering Techniques
17. Explain about Partitioning based Clustering Techniques
18. Explain about DBSCAN Clustering algorithm

**Long Answer Questions**

Explain the stages of KDD process

Explain about Data mining Technique

Explain in detail about Data Warehouse

Explain about Data Cleaning in detail.

Explain about Data Transformation in detail

Explain about Dimensionality reduction in detail.

Explain about the Apriori algorithm for finding frequent item sets with an example

Explain FP Growth Algorithm

Explain about AOI algorithm

Explain about Decision Tree Induction Algorithm

Explain about Bayesian classification

Explain about Hierarchical methods in cluster analysis

Explain the ECLAT algorithm using a simple example of market basket analysis?