

SHRI GNANAMBICA DEGREE COLLEGE: MADANAPALLE



(AUTONOMOUS)
Course 7: Data Mining Techniques with R (MAJOR)
SEMESTER III
(W.E.F.2024-25)
Program: BSC (DS)



Hours per week: 4

Credits: 3

Course Objectives

To understand data mining techniques and algorithms, and to comprehend data mining environments and their applications.

Course Outcomes

Upon successful completion of the course:

- 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

Introduction to Data Mining: An idea on Data Warehouse, Data Mining – KDD versus Data Mining, Stages of the Data Mining Process, Task Primitives, Data Mining Techniques, Data Mining Knowledge Representation.

UNIT II

Data Preprocessing and Query Languages: 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 and Pattern Mining: 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, Frequent Itemset Mining Methods – Apriori Method, Generating Association Rules, Improving the Efficiency of Apriori, Pattern-Growth Approach for Mining Frequent Itemsets.

UNIT IV

Classification Techniques: Classification – Basic Concepts, Steps to Build a Classification Model, Decision Tree Induction, Decision Tree Induction Algorithm, Attribute Selection Measures, Tree Pruning, Bayes Classification Methods.



c. Mahesh Babu
CHAIRMAN
BOARD OF STUDIES
Shri Gnanambica Degree College (A),
MADANAPALLE - 517 325

UNIT V

Association and Clustering Methods: Association Rule Mining – Antecedent, Consequent, Multi-relational Association Rules, ECLAT, Case Study on Market Basket Analysis, Cluster Analysis, Impact of Data on Clustering Techniques, Partitioning Methods, Hierarchical Methods, Density-based Methods – DBSCAN.

References :

Online References

<https://www.geeksforgeeks.org/data-mining-in-r/>,

<https://www.tutorialspoint.com/exploring-data-mining-with-r>,

<https://www.rdatamining.com/>

Reference Books

1. Jiawei Han, Micheline Kamber, Jian Pei, *Data Mining: Concepts and Techniques*, 3rd Edition, Morgan Kaufmann Publishers, 2011.
2. Adelchi Azzalini, Bruno Scapa, *Data Analysis and Data Mining*, 2nd Edition, Oxford University Press, 2012.
3. Charu C. Aggarwal, *Data Mining: The Textbook*, 2015.
4. Alex Berson, Stephen J. Smith, *Data Warehousing, Data Mining & OLAP*, 10th Edition, Tata McGraw Hill, 2007.
5. G. K. Gupta, *Introduction to Data Mining with Case Studies*, 1st Edition, Eastern Economy Edition, PHI, 2006.



c Mahesh Bose
CHAIRMAN
BOARD OF STUDIES
Shri Gnanambica Degree College (A,
MADANAPALLE - 517 325

SHRI GNANAMBICA DEGREE COLLEGE: MADANAPALLE



(AUTONOMOUS)
Course 7: Data Mining Techniques with R (MAJOR)- Practicals
SEMESTER III
(W.E.F.2024-25)
Program: BSC (DS)



Hours per week: 2

Credits: 1

List of Experiments

1. Get and Clean data using dplyr exercises.
2. Visualize all Statistical measures (Mean, Mode, Median, Range, Inter Quartile Range etc., using Histograms, Boxplots and Scatter Plots).
3. Create a data frame with at least 10 entries of columns EMPID, EMPNAME, SALARY, STARTDATE
 - a. Extract two column names using column name.
 - b. Extract the first two rows and then all columns.
 - c. Extract 3rd and 5th row with 2nd and 4th column.
4. Create a data frame with 10 observations and 3 variables and add new rows and columns to it using 'rbind' and 'cbind' function.
5. Create a function to discretize a numeric variable into 3 quantiles and label them as low, medium, and high. Apply it on each attribute of any dataset to create a new data frame.
6. Create a simple scatter plot using any dataset using 'dplyr' library. Use the same data to indicate distribution densities using box whiskers.
7. Write R Programs to implement k-means clustering, k-medoids clustering and density-based clustering on any datasets.
8. Write a R Program to implement decision trees using 'reading Skills' dataset.
Management
9. Implement decision trees using any dataset using package party and 'rpart'
10. Generate top 5 association rules using apriori
11. Generate top 5 association rules using ECLAT
12. Write an R program to implement Naïve bayes Classification.



C. Mahesh Babu
CHAIRMAN
BOARD OF STUDIES
Shri Gnanambica Degree College (A)
MADANAPALLE - 517 325

SHRI GNANAMBICA DEGREE COLLEGE: MADANAPALLE

(AUTONOMOUS)

Course 7: Data Mining Techniques with R (MAJOR)

SEMESTER III

(W.E.F.2024-25)

Program: BSC (DS)

Question Paper – Blue Print

Time : 3 Hrs

Marks : 70

PART-A

Answer any 4 of the 8. Each Question Carries 5 marks.

(4 x 5 =20)

1. Question
2. Question
3. Question
4. Question
5. Question
6. Question
7. Question
8. Question

PART-B

Answer one from each unit. Each Question Carries 10 marks.

(5X10=50)

UNIT 1

9. Question

OR

10. Question

UNIT 2

11. Question

OR

12. Question

UNIT 3

13. Question

OR

14. Question

UNIT 4

15. Question

OR

16. Question

UNIT 5

17. Question

OR

18. Question



C Mahesh Babu
CHAIRMAN
BOARD OF STUDIES
Shri Gnanambica Degree College (A)
MADANAPALLE - 517 325