COURSE TITLE	**DATA MINING FOR BUSINESS ANALYTICS USING R**
COURSE CODE	**04MB1134**
COURSE CREDITS	**3**

Course Outcomes: After completion of this course, student will be able to:

1. Develop control structures and functions in R, such as if-else loops, for and while loops, and loop functions.
2. Execute data manipulation and exploration in R, including handling packages, working with directories, data types, and performing common analytical tasks.
3. Perform descriptive statistical analysis and data visualization in R, using functions for basic arithmetic operations, statistical measures, and graphical presentations like bar charts, histograms, and scatterplots.
4. Apply simulation techniques and hypothesis testing in R, including generating random numbers, simulating random sampling, and solving problems with binomial, Poisson, normal, and exponential distributions.
5. Implement data mining and pre-processing techniques, such as classification, clustering, association rule generation, data normalization, cleaning, and feature selection, along with understanding their applications in real-world scenarios.

Pre-requisite of course: Understanding of mean, median, standard deviation, correlation, and probability distributions.

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
3	0	0	50	30	20	0	0

Contents : Unit	Topics	Contact Hours
1	**Introduction to R** Introduction, Downloading and installing R, IDE and Text Editors, handling packages in R. Getting Started with R – Working with Directory, Data Types in R, Commands for Data Exploration. Loading and Handling Data – Assigning Objects, Expression, Variables, Functions, Missing Values, Vectors, Matrices, Factors, Lists, subsetting a vector, subsetting a matrix, Common analytical tasks, grouping variables, Methods of Reading and Writing data in R. Exploring Data in R – Data Frames, Functions on Data Frame, Load Data Frames Managing Data Frames with the dplyr package	6

DR. MEETA JOSHI
Digitally signed by (Name of HOD)

DR. SANJEET SINGH
Digitally signed by (Name of Dean/ Principal)

Contents : Unit	Topics	Contact Hours
2	**Control Structures:** if-else loops, for Loops, while Loops, next, break Functions and Loop Functions: Functions: Functions in R, Your First Function, , Argument Matching, Lazy Evaluation, The Argument, Arguments Coming After the Argument, Functions using control statements , Loop Functions: Looping on the Command Line, lapply(), sapply(), split(), Splitting a Data Frame, tapply, apply(), Col/Row Sums and Means, Other Ways to Apply, mapply().	9
3	**Descriptive Statistics:** Basic Arithmetic Operations, Standard Functions like abs(), sqrt(), round(), sum(), product(), log(), log10(), Statistical Functions like min(), max(), range(), mean(), quantile(), summary(), var(), sd(), scale(), boxplot(), cov(), cor(). , Frequency Measures and Graphical Presentation frequency distribution and cumulative frequency distribution tables,, Bar Chart, Pie Chart, Histogram, Box-Whisker Plot, Scatterplots. Simulation: Generating Random Numbers, Setting the random number seed, Simulating Random Sampling, R function for solution of Binomial, Poisson, Normal and Exponential distribution problems. Hypothesis Testing: Testing Means (Single mean and Two Means)	10
4	**Data and Pre-processing** Concept of data mining, Reason of data mining, Data Mining Process, Relation to Business Intelligence Techniques, Introduction to Data Mining Task (Classification, Clustering, Association Analysis, Anomaly Detection). , Overview of model, Basic terminologies, Predictive modelling, Real-world Data mining Applications. Data and Pre-processing: Understanding of Data, what is Data?, Types of Attributes, Properties of attribute values, Sampling, Data Normalization, Data Cleaning, Similarity Measures, and Feature Selection	10
5	**Classification:** Decision-Tree Based Approach, Rule-based Approach, Instance-based classifiers (Nearest Neighbour), , Classification Model Selection and Evaluation. Association Rules: Discovering and Generating Association Rules in Transaction Databases, Data Format, Frequent Item set Generation, Frequency-plot, Rule Generation, The Process of Rule Selection and Interpreting the Results, FP-Growth Algorithm and Evaluation of Association Patterns. Overview of Cluster Analysis. Overview of Data Visualization.	10
Total Hours		**45**

Textbook :

1 R Programming for Data Science , Roger D. Peng , Lean Publishing , 2015

2 Data Mining: Concepts and Techniques , Jiawei Han , Morgan Kaufmann Publishers , 2012

DR. MEETA JOSHI
Digitally signed by (Name of HOD)

DR. SANJEET SINGH
Digitally signed by (Name of Dean/ Principal)

References:

1 Data Mining, Data Mining, Charu C. Aggarwal , Springer International Publishing , 2015

Suggested Theory Distribution:

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery and evaluation

Remember / Knowledge	Understand	Apply	Analyze	Evaluate	Higher order Thinking / Creative
0.00	10.00	25.00	25.00	20.00	20.00

Instructional Method:

1 Use a blended learning approach with theory, R-based hands-on labs, real business case studies, and project-based assessment

Supplementary Resources:

1 https://www.w3schools.com/r/r_intro.asp
2 https://posit.co/download/rstudio-desktop/
3 https://posit.co/products/open-source/rstudio/?sid=1

DR. MEETA JOSHI
Digitally signed by (Name of HOD)

DR. SANJEET SINGH
Digitally signed by (Name of Dean/ Principal)