

--	--	--	--	--	--	--	--	--	--



INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal, Hyderabad - 500 043

MODEL QUESTION PAPER

M.Tech I Semester End Examinations (Regular), January - 2020

Regulation: IARE-R18

DATA SCIENCE

(COMPUTER SCIENCE AND ENGINEERING)

Time: 3Hours

Max Marks: 70

Answer ONE Question from each Unit

All parts of the question must be answered in one place only

UNIT - I

1. (a) Summarize the stages of data science project? Discuss the fundamental differences of R with other programming languages? [7M]
- (b) Elaborate the process of working with different data files. Write about file choose() function. Compute the given mathematical formula and display on console in R. $X = e^8 \log 18 + \sqrt{296}$ [7M]
2. (a) List the different R functions to read and write the data from disk and R object. Write R script to choose the character data dynamically from user. [7M]
- (b) Compare the different forms of data types and create a list object. Write R script to create and display list object of stores items having:
{Fruits: {orange, mango, apple, watermelon, banana} Juices: {appy, fruty, slice}
Milkshakes: {Mango, papaya, sapota, pineapple}}. [7M]

UNIT - II

3. (a) State the different ways to access different types of data files? Discuss the relevant packages and methods to access .csv, .exl files. [7M]
- (b) Differentiate SQL and No SQL databases in detail? Give example of XML data extraction and operations using R. [7M]
4. (a) Compute the covariance matrix and correlation matrix for the four numerical attributes. Interpret the statistical findings to know more about hidden nature in data. [7M]
- (b) Distinguish simple and multiple regression analysis and its applications working with numerical and categorical data? [7M]

UNIT - III

5. (a) List out the applications of machine learning with example. Discuss about a consistent learner and what it means for a set of training examples to be linearly separable. [7M]
- (b) Describe the process to create and evaluate the data model for the given data. To predict whether an email is a spam and should be delivered to the Junk folder. Suggest suitable data model. [7M]
6. (a) Discuss about clustering in detail? Elaborate k means algorithm. Write a R script to implement. [7M]
- (b) State about conditional probability in the bayes theorem for different type of data classification problem with suitable example. Calculate the Jaccard coefficient for the given data $p = 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0$, $q = 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 1$; [7M]

UNIT - IV

7. (a) Give the basic structure of neural network and different types of ANN with real time examples. [7M]
- (b) How to evaluate hypothesis of different types with examples? Explain the basics of sampling theory. [7M]
8. (a) Discuss the difference of error in two hypotheses. Differentiate the MAP (maximum a posteriori) and ML (maximum likelihood) hypothesis. Give an example of a scenario in which a MAP hypothesis is preferable to an ML hypothesis. [7M]
- (b) Compare the learning algorithms with example in terms of problem nature, accuracy and error rate. [7M]

UNIT - V

9. (a) Generalize the graphical analysis in data analysis? List out the various plots in R and explain in detail. [7M]
- (b) How to plot the word (text) data based on frequency of words. Write R script to plot a data frame having: {df1: {sea, river, pond, lake, pool} df2: {3,5,8,10,34}} using relevant plot. [7M]
10. (a) How would you get the multiple plots in single window? Plot the regression model along with residuals. Write a R script for creating a box plot of iris sepal length attribute. [7M]
- (b) Elaborate how to export a graph using graphics parameters. How to export the text data to plot with example. [7M]



INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

COURSE OBJECTIVES:

The course should enable the students to:

I	Summarize the fundamental knowledge on basics of data science and R programming.
II	Develop programs in R language for understanding and visualization of data using statistical functions and plots.
III	Learn to apply hypotheses and data into actionable predictions.
IV	Understand a range of machine learning algorithms along with their strengths and weaknesses.
V	Able to document and transfer the results and effectively communicate the findings using visualization techniques.

COURSE OUTCOMES:

Students, who complete the course, will have demonstrated the ability to do the following:

S No	Course Outcomes
CO 1	Understand the process and different stages of data science and relevant data descriptions in R language
CO 2	Illustrate various SQL, NOSQL databases connecting with R and perform correlation and regression analysis systems.
CO 3	Evaluate different data models and perform clustering analysis
CO 4	Solve various real time problems using artificial neural networks techniques and comparing different learning algorithms.
CO 5	Explore on various ways to deliver results through documentation and plots of multivariate data and matrix data.

COURSE LEARNING OUTCOMES (CLOs):

1. Understand and develop relevant programming abilities.
2. Understand and intuition of the whole process line of extracting knowledge from data.
3. Equip with the fundamental knowledge on basics of data science and R programming
4. Critically analyze and evaluate variety of NoSQL databases.
5. Develop the ability to build and assess Data-based models.
6. Analyze data analysis and make models using regression analysis.
7. Familiarize with variety of machine learning tasks: clustering, dimensionality reduction, regression and classification.
8. Understand how to formalize practical problems using methods of machine learning.
9. Understand neural networks techniques solve real time problems.
10. Understand the different learning algorithms.
11. Chose an appropriate learning Algorithms to solve particular problems.
12. Based on delivering results make documentation for various results sets.
13. Understand how to plot graphs for multivariate and matrix data.

Mapping of Semester End Examination to Course Learning Outcomes:

SEE Question No.		Course Learning Outcomes		Course Outcomes	Blooms Taxonomy Level
1	a	BCSB06.03	Equip with the fundamental knowledge on basics of data science and R programming.	CO 1	Understand
	b	BCSB06.01	Understand and develop relevant programming abilities.	CO 1	Remember
2	a	BCSB06.02	Understand and intuition of the whole process line of extracting knowledge from data.	CO 1	Understand
	b	BCSB06.02	Understand and intuition of the whole process line of extracting knowledge from data.	CO 1	Understand
3	a	BCSB06.03	Equip with the fundamental knowledge on basics of data science and R programming.	CO 2	Understand
	b	BCSB06.03	Equip with the fundamental knowledge on basics of data science and R programming.	CO 2	Understand
4	a	BCSB06.05	Develop the ability to build and assess Data-based models.	CO 2	Understand
	b	BCSB06.06	Analyze data analysis and make models using regression analysis.	CO 2	Understand
5	a	BCSB06.08	Understand how to formalize practical problems using methods of machine learning.	CO 3	Understand
	b	BCSB06.07	Familiarize with variety of machine learning tasks: clustering, dimensionality reduction, regression and classification.	CO 3	Remember
6	a	BCSB06.07	Familiarize with variety of machine learning tasks: clustering, dimensionality reduction, regression and classification.	CO 3	Understand
	b	BCSB06.08	Understand how to formalize practical problems using methods of machine learning.	CO 3	Understand
7	a	BCSB06.09	Understand neural networks techniques solve real time problems.	CO 4	Remember
	b	BCSB06.10	Understand the different learning algorithms.	CO 4	Remember
8	a	BCSB06.10	Understand the different learning algorithms.	CO 4	Understand
	b	BCSB06.11	Chose an appropriate learning Algorithms to solve particular problems.	CO 4	Understand
9	a	BCSB06.13	Understand how to plot graphs for multivariate and matrix data.	CO 5	Remember
	b	BCSB06.13	Understand how to plot graphs for multivariate and matrix data.	CO 5	Understand
10	a	BCSB06.12	Based on delivering results make documentation for various results sets.	CO 5	Understand
	b	BCSB06.13	Understand how to plot graphs for multivariate and matrix data.	CO 5	Understand

HOD, CSE