Question Paper

Exam Date & Time: 28-Nov-2022 (02:00 PM - 05:00 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

SEVENTH SEMESTER B.TECH END SEMESTER EXAMINATIONS, NOV 2022

ADVANCED DATA SCIENCE - PART III [CRA 4062]

Marks: 50 **Duration: 180 mins.**

A

Answer all the questions.

Instructions to Candidates: Answer ALL questions Missing data may be suitably assumed

1) Write R code to upload the quantmod package and use the Google stock price starting from September 2021 until the latest you can get. (5)

A)

a. Produce time plot of the Google stock price.

b. Produce forecast using exponential smoothing function.

c. Report the accuracy of the forecast model.

B) The following confusion matrix summarizes the predictions made by the model: (3)

		Predicted	
		Drafted = Yes	**Drafted = No**
Actual	**Drafted = Yes**	**120** (True Positive)	**40** (False Negative)
	Drafted = No	**70** (False positive)	**170** (True Negative)

Compute Misclassification Rate, Gini Index and Information Gain for the above-mentioned model.

C) Consider spam dataset with outcome variable type=Spam/Not spam. Write R code for using functions present in the Caret package (2)

i. Create a data partition with 75% for training and 25% for testing data.

ii. Split data into k number of folds based on the outcome, returning data as list and should also return training set.

2) John enjoys posting pictures of his cat online. The past 6 pictures received the following number of likes in order. (5)

A)

10, 15, 15, 16, 17, 18, 21

Compute Median Absolute Division for the above scenario.

B) Find the Mean Squared Error for the Following Set of Values: (43,41),(44,45),(45,49),(46,47), (47,44). The equation of the line is given by (3)

y= 9.2 + 0.8x.

C) Write the form of penalized residual sum of squares (PRSS) used in Ridge and Lasso regression. (2)

3) Illustrate principal components analysis with the help of R function/code snippet to get information about standard deviation, and eigenvectors and plot the percentage variation. (5)

A)

B) With the help of R code snippet show how to slice the data into *k* folds for cross-validation and return *k* lists of indices. Explain in detail all the parameters associated with the function. (3)

C) Explain the functionality of ***caret*** package. (2)

4) A specific user would like to organize a shiny app into two separate files:ui.R,consisting of the user interface elements of the app, andserver .R, contains the logic of the app, including code for loading and handling data. The code snippet for each of these files is given below. With appropriate justification, briefly describe the flow of the interface. (5)

A)

App 1: ui.R

```
library(shiny)
fluidPage(
titlePanel("Data science FTW!"),
sidebarLayout(
sidebarPanel(
h3("Sidebar Text")
),
mainPanel(
h3("Main Panel Text")
)
)
)
```

App 1: server.R

```
library(shiny)
function(input, output){
}
```

B) Illustrate the process toinsert R code into an R Markdown document with appropriate examples. (3)

C) Illustrate the functionality of googleVis with an example. (2)

5) Illustrate the following functionalities in R Markdown with suitable examples. (5)

A)

i. Yaml front matter
ii. Loading the R markdown package
iii. Render your R markdown into a webpage
iv. Document view in HTML

v. PDF creation

B) Illustrate how to change the user interface of the Shiny app displayed depending on other UI inputs in your app. (3)

C) Write a function to add a base layer to your interactive map and to sequentially add features to your map with appropriate syntax. (2)

-----End-----