



**VII SEMESTER B.TECH. (INFORMATION TECHNOLOGY)**  
**END SEMESTER EXAMINATIONS, FEB 2021**  
**SUBJECT: PROGRAM ELECTIVE VI-ADVANCED DATA SCIENCE**  
**[CRA 4012]**  
**REVISED CREDIT SYSTEM**  
**(03/02 /2021)**

Time: 3 Hours

MAX. MARKS: 50

### Instructions to Candidates:

- ❖ Answer **ALL** the questions.
- ❖ Missing data if any, may be suitably assumed.

- 1A.** What is the purpose of regularized regression? Write the form of penalized residual sum of squares (PRSS) used in Ridge and Lasso regression. Which one is better? Justify 5
- 1B.** Describe fluidPage layout in shiny framework. Also, write the function prototype for creating fluid page layouts along with arguments description. 3
- 1C.** Write the functionality of the following user interface components in Shiny framework:
- |                      |                   |   |
|----------------------|-------------------|---|
| (i) tabsetPanel()    | (iii) tabPanel()  |   |
| (ii) brushedPoints() | (iv) nearPoints() | 2 |
- 2A.** Write R code to perform the following tasks using caret package. Load the “words.train” and “words.test” data sets available in “ElemStatLearn” library. Set the variable x to be a factor variable in both the training and test set. Set the seed to 33855. Fit (1) a random forest predictor and (2) a boosted predictor using the “gbm” method, relating the factor variable y to the remaining variables with the train() command. Print the accuracy among the test set samples where the two methods agree. 5
- 2B.** Write R code for adding up two integers and returning their difference in a shiny application. 3
- 2C.** Explain with an example how the code written in R can be analyzed? 2
- 3A.** Explain with suitable example the functions supported by caret package in R to handle skewed data and missing values. 5
- 3B.** What is forecasting? Mention the specific patterns to be considered for time series data. Mention the outcome of the following commands
- |                     |                              |  |
|---------------------|------------------------------|--|
| (i) ma(ts, order=3) | (ii) ets(train, model="MMM") |  |
|---------------------|------------------------------|--|
- 3
- 3C.** Justify the statement “Unsupervised prediction is effectively an exploratory technique”. 2
- 4A.** What are covariates? Explain two levels of covariate creation. 5

**4B.** Mention the steps involved in Prediction study design. What are the implications, in K-fold cross validation if you choose a) large K value b) Small K value? What are the other approaches for cross validation? **3**

**4C** Given the prediction outcomes of two models ModelA and ModelB, comment on In sample and Out sample errors for these two models.

Model A	TRUE	FALSE
NOT SPAM	5	0
SPAM	0	5

Model B	TRUE	FALSE
NOT SPAM	5	1
SPAM	0	4

**2**

**5A.** List and explain with appropriate example, the essential elements of R package. **5**

**5B.** Define delayed reactivity in Shiny framework. Design a web page using Shiny framework as shown in Figure Q. 5B such that on submit of the button after the bar is glided, the plot changes.

**submitButton example**

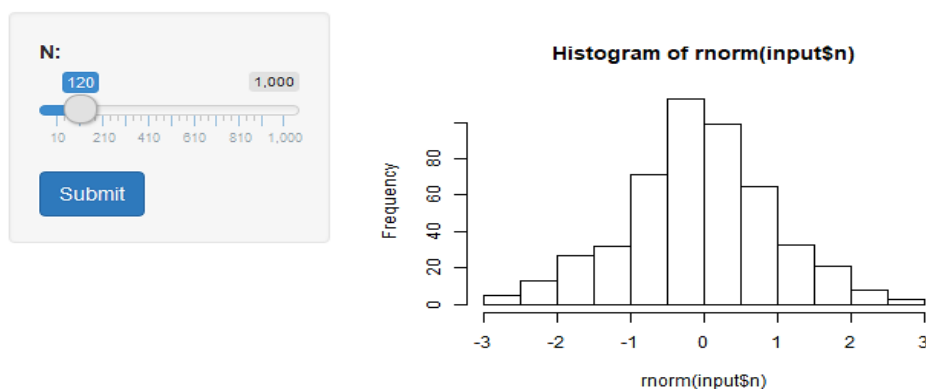


Figure Q.5B: UI for Bar Plot

**3**

**5C.** What are the problems solved by reproducible research? Give an example. **2**