Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
(A constituent unit of MAHE, Manipal)

VII SEMESTER B.TECH.
MAKE UP EXAMINATIONS, DECEMBER 2019
SUBJECT: PROGRAM ELECTIVE VI: ADVANCED DATA SCIENCE [CRA 4012]
REVISED CREDIT SYSTEM
(/ 12 /2019)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:
- Answer **ALL** the questions.
- Missing data, if any, may be suitably assumed.

1A. What are the different data slicing techniques present in Caret package? Explain these techniques considering the dataset for credit card fraud, with the predicting variable status= {fraud, genuine}. **5**

1B. What is the significance of Shiny gadgets? Create a shiny gadget that takes two different vectors to populate two selectInputs. Choose two numbers within the gadget and return the product of those two numbers. **3**

1C. What is R markdown? List down three uses of R markdown. **2**

2A. Consider the dataset containing the attributes Crime_rate, area_size and house_price. The house_price is the target variable. Write R code to fit a linear model for the selected points on scatter plot and draw a line of best fit for the resulting model. **5**

2B. How is bagging performed? What are its advantages? **3**

2C. What are covariates? What covariates can be constructed for text files and images? **2**

3A. Define the structure of a shiny application. Write R code to create a web page containing the following fields:

i. Enter your name
ii. Enter your CGPA
iii. Select your semester

5

On submit, the details entered should appear on the main panel. Name the title of the web page as "Student Details".

3B. Write R command for creating 10 folds for cross validation and return the training set indices and display the structure of the training set indices. **3**

3C. Write R command to perform following operations using swirl:

i. Create a new lesson
ii. Get the current working lesson
iii. Write a multiple choice question
iv. Add lesson to manifest

2

4A. What is the significance of using Principal Component Analysis. Write R code to perform the following: **5**

i. Perform PCA on data
ii. Print the eigenvector/rotations first 5 rows and PCs
iii. Plot the first two principal components

4B. With a help of R code, explain how multiple views can be given to a shiny application? **3**

4C. What is meant by generic function in R? Explain with an example, how can you determine the class of a function? **2**

5A. Given testing, training and validation data, and outcome variable *y* perform the following operations using R **5**

i. Set the variable y to be a factor variable in both the training and test
ii. Set the seed to 33833.
iii. 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.
iv. Print the accuracy among the test set samples where the two methods agree

5B. Write R code to create the following types of charts using googleVis **3**

i. Line chart
ii. Geo chart
iii. Motion chart

5C. What are the important considerations while interpreting results of forecasting? **2**

****