Savitribai Phule Pune University

T.Y. B.C.A (Science)

Semester – V

C.B.C.S 2019 Pattern

BCA357

DSE II Lab

(Data Mining)

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1.Write a R program to add, multiply and divide two vectors of integertype. (Vector length should be minimum 4) [10 Marks]

Q2.Consider the student data set. It can be downloaded from: https://drive.google.com/open?id=1oakZCv7g3mlmCSdv9J8kdSaqO 5_6dIOw . Write a programme in python to apply simple linear regression and find out mean absolute error, mean squared error and root mean squared error. [20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write an R program to calculate the multiplication table using afunction. [10 Marks]

Q2. Write a python program to implement k-means algorithms on asynthetic dataset. [20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write a R program to reverse a number and also calculate the sum ofdigits of that number. [10 Marks]

Q2. Consider the following observations/data. And apply simple linear regression and find out estimated coefficients b0 and b1.(use numpypackage)
x=[0,1,2,3,4,5,6,7,8,9,11,13]
y = ([1, 3, 2, 5, 7, 8, 8, 9, 10, 12,16, 18] [20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write a R program to calculate the sum of two matrices of given size. [10 Marks]

Q2. Consider following dataset

weather=['Sunny','Sunny','Overcast','Rainy','Rainy','Rainy','Overcast','Sunny','Sunny','Rainy','Sunn
y','Overcast','Overcast','Rainy']

temp=['Hot','Hot','Hot','Mild','Cool','Cool','Cool','Mild','Cool','Mild','Mild','Mild','Hot','Mild']

play=['No','No','Yes','Yes','Yes','No','Yes','No','Yes','Yes','Yes','Yes','Yes','No'].

Use Naïve Bayes algorithm to predict [0: Overcast, 2: Mild]tuple belongs to which class whether to play the sports or not.

[20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write a R program to concatenate two given factors. [10 Marks]

Q2. Write a Python program build Decision Tree Classifier using Scikit- learn package for diabetes data set (download database from https://www.kaggle.com/uciml/pima-indians-diabetes-database)

[20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write a R program to create a data frame using two given vectors and displaythe duplicate elements. [10 Marks]

Q2. Write a python program to implement hierarchical Agglomerative clusteringalgorithm. (Download Customer.csv dataset from github.com). [20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write a R program to create a sequence of numbers from 20 to 50 and findthe mean of numbers from 20 to 60 and sum of numbers from 51 to 91.

[10 Marks]

Q2. Consider the following observations/data. And apply simple linear regression and find out estimated coefficients b1 and b1 Also analyse theperformance of the model

(Use sklearn package)

x = np.array([1,2,3,4,5,6,7,8])

y = np.array([7,14,15,18,19,21,26,23]) [20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write a R program to get the first 10 Fibonacci numbers. [10 Marks]

Q2. Write a python program to implement k-means algorithm to build prediction model (Use Credit Card Dataset CC GENERAL.csv Download from kaggle.com) [20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write an R program to create a Data frames which contain details of 5 employees and display summary of the data. [10 Marks]

Q2. Write a Python program to build an SVM model to Cancer dataset. The dataset is available in the scikit-learn library. Check the accuracyof model with precision and recall. [20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write a R program to find the maximum and the minimum value of a givenvector [10 Marks]

Q2. Write a Python Programme to read the dataset ("Iris.csv"). dataset download from (https://archive.ics.uci.edu/ml/datasets/iris) and apply Apriori algorithm. [20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write a R program to find all elements of a given list that are not inanother given list.

```
AB st("x", "y", "z")
   st("X", "Y", "Z", "x", "y", "z")
```

[10 Marks]

Q2. Write a python program to implement hierarchical clustering algorithm.(Download Wholesale customers data dataset from github.com).

[20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write a R program to create a Dataframes which contain details of 5employees and display the details.

Employee contain (empno,empname,gender,age,designation)

[10 Marks]

Q2. Write a python program to implement multiple Linear Regression modelfor a car dataset. Dataset can be downloaded from:
https://www.w3schools.com/python/python_ml_multiple_regression.asp

[20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Draw a pie chart using R programming for the following datadistribution:

Digits on Dice	1	2	3	4	5	6
Frequency of getting each number	7	2	6	3	4	8

[10 Marks]

Q2. Write a Python program to read "StudentsPerformance.csv" file. Solvefollowing:

- To display the shape of dataset.

- To display the top rows of the dataset with their columns.Note: Download dataset from following link :

(https://www.kaggle.com/spscientist/students-performance-inexams?select=StudentsPerformance.csv) [20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write a script in R to create a list of employees (name) and perform thefollowing:

a. Display names of employees in the list.

b. Add an employee at the end of the list

c. Remove the third element of the list. [10 Marks]

Q2. Write a Python Programme to apply Apriori algorithm on Groceries dataset. Dataset can be downloaded from (https://github.com/amankharwal/Websitedata/blob/master/Groceries_dataset.csv).

Also display support and confidence for each rule.

[20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1.Write a R program to add, multiply and divide two vectors of integer type.(vector length should be minimum 4) [10 Marks]

Q2. Write a Python program build Decision Tree Classifier for shows.csvfrom pandas and predict class label for show starring a 40 years old American comedian, with 10 years of experience, and a comedy ranking of 7? Create a csv file as shown in https://www.w3schools.com/python/python_ml_decision_tree.asp

[20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write a R program to create a simple bar plot of given data

Year	Export	Import
2001	26	35
2002	32	40
2003	35	50

[10 Marks]

Q2. Write a Python program build Decision Tree Classifier using Scikit-learnpackage for diabetes data set (download database from https://www.kaggle.com/uciml/pima-indians-diabetes-database)

[20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write a R program to get the first 20 Fibonacci numbers.

[10 Marks]

Q2. Write a python programme to implement multiple linear regression modelfor stock market data frame as follows:

```
Stock_Market = {'Year':
[2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2017,2016,2
016,20,16,2016,2016,2016,2016,2016,2016,2016,2016,2016],
'Month': [12, 11,10,9,8,7,6,5,4,3,2,1,12,11,10,9,8,7,6,5,4,3,2,1],
'Interest_Rate': [2.75,2.5,2.5,2.5,2.5,2.5,2.5,2.25,2.25,2.25,2,2,2,1.75,1.75,1.75,1.75,1.75,1
.75,1.75,1.75,1.75,1.75,1.75],
'Unemployment_Rate':
[5.3,5.3,5.3,5.3,5.4,5.6,5.5,5.5,5.5,5.6,5.7,5.9,6,5.9,5.8,6.1,6.2,6.1,6.1,6.1,5
.9,6.2,6.2,6.1],
'Stock_Index_Price': [1464,1394,1357,1293,1256,1254,1234,1195,1159,1167,1130,1075,1047,
965,943,958,971,949,884,866,876,822,704,719] }
```

And draw a graph of stock market price verses interest rate.

[20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write a R program to find the maximum and the minimum value of a givenvector [10 Marks]

Q2. Consider the following observations/data. And apply simple linear regression and find out estimated coefficients b1 and b1 Also analyse theperformance of the model

(Use sklearn package)

x = np.array([1,2,3,4,5,6,7,8])

y = np.array([7,14,15,18,19,21,26,23]) [20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write a R program to create a Dataframes which contain details of 5 Studentsand display the details.

Students contain (Rollno,Studname,Address,Marks) [10 Marks]

Q2. Write a python program to implement multiple Linear Regression modelfor a car dataset. Dataset can be downloaded from:
https://www.w3schools.com/python/python_ml_multiple_regression.asp
[20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]

Savitribai Phule Pune University

T.Y. B.C.A. (Science) (Semester-V) Practical Examination

BCA 357: DSE II Lab (Data Mining)

Duration: 3Hrs. **Max Marks: 35+15=50**

Q1. Write a R program to create a data frame from four given vectors. [10 Marks]

Q2. Write a python program to implement hierarchical Agglomerativeclustering algorithm. (Download Customer.csv dataset from github.com). [20 Marks]

Q3. Viva [5 Marks]

Q4. Internal Assessment [15 Marks]