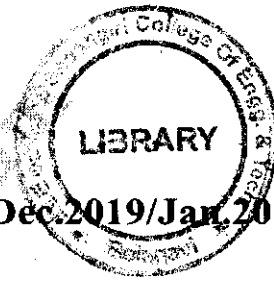


|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|



## Eighth Semester B.E. Degree Examination, Dec. 2019/Jan. 2020 Big Data Analytics

Time: 3 hrs.

Max. Marks: 80

Note: Answer FIVE full questions, choosing one full question from each module.

### Module-1

- 1 a. What is HDFS? List all the components of HDFS and explain any four components. (09 Marks)
- b. With example, explain the different general HDFS commands. (07 Marks)

OR

- 2 a. Explain with neat diagram, the parallel mapreduce data flow. (07 Marks)
- b. Write the java code for MAP and REDUCE of word count problem. Describe the steps of compiling and removing the mapreduce program. (09 Marks)

### Module-2

- 3 a. Describe with neat diagram, the two step sqoop data export and import method. (08 Marks)
- b. With block diagram discuss the various frameworks that run under YARN. (08 Marks)

OR

- 4 a. Discuss the different views supported by Apache Ambari. (06 Marks)
- b. Describe the various features of hadoop YARN administration. (04 Marks)
- c. Explain the different HDFS administration features. (06 Marks)

### Module-3

- 5 a. What is Business Intelligence (BI)? List the different BI applications and explain in detail any five applications. (10 Marks)
- b. With neat block diagram, explain data warehouse architecture. (06 Marks)

OR

- 6 a. Explain with diagram CRISP-DM data mining cycle. (08 Marks)
- b. Describe the common data mining mistakes. (04 Marks)
- c. List and describe the various charts used for data visualization. (04 Marks)

### Module-4

- 7 a. Explain the different steps for constructing the decision tree for the following example (Table 7(a)) (08 Marks)

| OUTLOOK  | TEMP | HUMIDITY | WINDY | PLAY |
|----------|------|----------|-------|------|
| SUNNY    | HOT  | HIGH     | FALSE | NO   |
| SUNNY    | HOT  | HIGH     | TRUE  | NO   |
| OVERCAST | HOT  | HIGH     | FALSE | YES  |
| RAINY    | MILD | HIGH     | FALSE | YES  |
| RAINY    | COOL | NORMAL   | FALSE | YES  |
| RAINY    | COOL | NORMAL   | TRUE  | NO   |
| OVERCAST | COOL | NORMAL   | TRUE  | YES  |
| SUNNY    | MILD | HIGH     | FALSE | NO   |
| SUNNY    | COOL | NORMAL   | FALSE | YES  |
| RAINY    | MILD | NORMAL   | FALSE | YES  |
| SUNNY    | MILD | NORMAL   | TRUE  | YES  |
| OVERCAST | MILD | HIGH     | TRUE  | YES  |
| OVERCAST | HOT  | NORMAL   | FALSE | YES  |
| RAINY    | MILD | HIGH     | TRUE  | NO   |

Table 7(a)

- b. Describe the advantages and disadvantages of using Regression models.

(08 Marks)

OR

- 8 a. Write the different steps involved in developing an artificial neural network. (05 Marks)  
 b. Describe the advantages of using ANN. (03 Marks)  
 c. For the following example (Table 8c) describe the different steps of forming association rules using Apriori algorithm. (08 Marks)

| S.No. | TRANSACTION LIST |         |         |         |
|-------|------------------|---------|---------|---------|
| 1     | MILK             | EGG     | BREAD   | BUTTER  |
| 2     | MILK             | BUTTER  | EGG     | KETCHUP |
| 3     | BREAD            | BUTTER  | KETCHUP |         |
| 4     | MILK             | BREAD   | BUTTER  |         |
| 5     | BREAD            | BUTTER  | COOKIES |         |
| 6     | MILK             | BREAD   | BUTTER  | COOKIES |
| 7     | MILK             | COOKIES |         |         |
| 8     | MILK             | BREAD   | BUTTER  |         |
| 9     | BREAD            | BUTTER  | EGG     | COOKIES |
| 10    | MILK             | BUTTER  | BREAD   |         |
| 11    | MILK             | BREAD   | BUTTER  |         |
| 12    | MILK             | BREAD   | COOKIES | KETCHUP |

Table 8(c)

**Module-5**

- 9 a. Describe the difference between text mining and data mining. (06 Marks)  
 b. Briefly describe SVM technique. (04 Marks)  
 c. Explain Naive Bayes model to classify the text data into right class using following dataset (Table Q9 (c)) (06 Marks)

| Training Set | Document ID | Keyword in the document   | Class = h (Healthy) |
|--------------|-------------|---------------------------|---------------------|
|              | 1           | Love Happy Joy Joy Love   | Yes                 |
|              | 2           | Happy Love KICK JOY Happy | Yes                 |
|              | 3           | Love Move Joy Good        | Yes                 |
|              | 4           | Love Happy Joy Pain Love  | Yes                 |
|              | 5           | Joy Love Pain Kick pain   | No                  |
|              | 6           | Pain Pain Love Kick       | No                  |
| Test data    | 7           | Love Pain Joy Love Kick   | ?                   |

Table Q9 (c)

OR

- 10 a. What is web mining? Explain the different types of web mining. (08 Marks)  
 b. Discuss the application and practical consideration of social network analysis. (08 Marks)

\*\*\*\*\*