USN | | | | | | | | | | **15CS82**

Eighth Semester B.E. Degree Examination, June/July 2019
Big Data Analytics

Time: 3 hrs.   Max. Marks: 80

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

Module-1

1 a. How does the Hadoop MapReduce Data flow work for a word count program? Give an example. **(08 Marks)**
   b. Briefly explain HDFS Name Node Federation, NFS Gateway, Snapshots, Checkpoint and Backups. **(08 Marks)**

OR

2 a. What do you understand by HDFS? Explain its components with a neat diagram. **(10 Marks)**
   b. Bring out the concepts of HDFS block replication, with an example. **(06 Marks)**

Module-2

3 a. Explain Apache Squoop Import and Export method with neat diagrams. **(10 Marks)**
   b. Explain with a neat diagram, the Apache Oozie work flow for Hadoop architecture. **(06 Marks)**

OR

4 a. How do you run Map Reduce and Message Passing Interface (MPI) on YARN architecture? Discuss. **(10 Marks)**
   b. What do you understand by YARN Distributed-Shell? **(06 Marks)**

Module-3

5 a. Write any four Business Intelligence Application for various sectors. **(08 Marks)**
   b. Explain the star schema design of Data Warehousing with an example. **(06 Marks)**
   c. What is Confusion Matrix? **(02 Marks)**

OR

6 a. Explain CRISP-DM cycle with a neat diagram. **(08 Marks)**
   b. What do you understand by the term Data Visualization? How is it important in Big data Analytics? **(05 Marks)**
   c. Differentiate between Data Mining and Data Warehousing. **(03 Marks)**

Module-4

7 a. What is a splitting variable? Describe three criteria for choosing a splitting variable. **(04 Marks)**
   b. List some of the advantages and disadvantages of Regression Model. **(04 Marks)**
   c. Create a decision tree for the following data set.

Age	Job	House	Credit	Loan Approved
Young	False	No	Fair	No
Young	False	No	Good	No
Young	True	No	Good	Yes
Young	True	Yes	Fair	Yes
Young	False	No	Fair	No

Age	Job	House	Credit	Loan Approved
Middle	False	No	Fair	No
Middle	False	No	Good	No
Middle	True	Yes	Good	Yes
Middle	False	Yes	Excellent	Yes
Middle	False	Yes	Excellent	Yes
Old	False	Yes	Excellent	Yes
Old	False	Yes	Good	Yes
Old	True	No	Good	Yes
Old	True	No	Excellent	Yes
Old	False	No	Fair	No

Then solve the following problem using the model:

Age	Job	House	Credit	Loan Approved
Young	False	False	Good	???

(08 Marks)

OR

8　a. Explain the design principles of an Artificial Neural Network. (08 Marks)

　　b. How does the Apriori Algorithm work? Apply the same for the following example.

T_{ID}	List of Item-IDs
T_{100}	I_1, I_2, I_5
T_{200}	I_2, I_4
T_{300}	I_2, I_3
T_{400}	I_1, I_2, I_4
T_{500}	I_1, I_3
T_{600}	I_2, I_3
T_{700}	I_1, I_3
T_{800}	I_1, I_2, I_3, I_5
T_{900}	I_1, I_2, I_3

Assume the support count = 2. (08 Marks)

Module-5

9　a. What is Naïve Bayes Technique? Explain its model. (05 Marks)

　　b. What is a Support Vector Machine? Explain its model. (08 Marks)

　　c. Mention the 3-step process of Text Mining. (03 Marks)

OR

10　a. Explain briefly the three different types of web mining. (06 Marks)

　　b. Compute the rank values for the Nodes for the following network shown in Fig.Q10(b), which is the Highest ranked node. Solve the same with eight iterations.
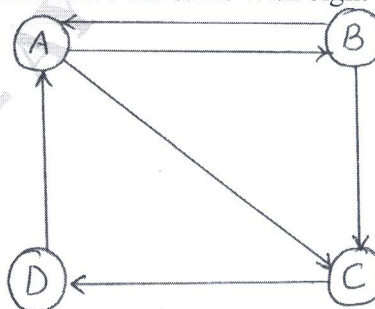

Fig.Q10(b) (10 Marks)

* * * * *