** ACS COLLEGE OF ENGINEERINGC:\Documents and Settings\Administrator\Desktop\update kk\New Folder\naac-a-new.png**

**#207, Kambipura, Mysore road, Banglore-74**

**Department of Computer Science & Engineering**

**QUESTION BANK**

Sem/Subject/Code: **8th/BIG DATA ANALYTICS/15CS82**

Prepared By**: SUNITA CHALAGERI**

Year : **2019-20(EVEN)**

**Module-1**

1. Define HDFS? Explain the various components of HDFS with neat diagram.
2. What are the important aspects of HDFS.
3. Explain Block replication concepts in HDFS with diagram.
4. Write a note on HDFS user commands.
5. Practice Map reduce word count problem
6. Write a note on HDFS Safenode, Namenode, HDFS checkpoints and backups, HDFS snapshot and Gateway.
7. What is Map reduce? Explain Map reduce dataflow with neat diagram.
8. What are the steps involved in map reduce concept?
9. What are the steps involved to compile &run the program (word count program) from the command line.

**Module-2**

1. Write a note on Apache Pig and Hive.
2. What is Apache scoop & explain the import and export methods of scoop.
3. Compare apache scoop version1 with version2
4. How do acquire data streams using apache flume explain with neat diagram
5. Installation steps of Hadoop,pig,hive..etc
6. Write a note on Apache Hbase.
7. Explain the structure of Yarn and its applications.
8. Explain the different views of apache Ambari
9. How do you change the Hadoop properties using apache Ambari
10. Write a note on basic Hadoop administration procedure.

**Module-3**

1.Describe business intelligence and data mining (BIDM) cycle.

2. Describe data processing chain.

3. Explain different data mining techniques.

4.what are the applications of BI.

5.Write a note on BI tools.

6. what is the purpose of data warehouse and differentiate between DataMart and data warehouse.

7.What are the different key elements of data warehouse. Describe each one of them.

8. what are the sources and types of data for data warehouse.

9.write a note on data warehouse architecture.

10.explain star schema architecture for data warehouse.

11. what are the design considerations for data warehouse.

12.what is data mining? What are supervised and unsupervised learning technique.

13.what is confusion matrix.

14.what are the most popular data mining technique.

15.what are the major mistakes to be avoided while doing data mining.

16.Explain CRISP datamining cycle.

17.write a note on tools and platform for datamining

18.what is data visualization.

19.what are the key requirement for good visualization.

20. what are data visualization techniques and when would you use tables and graphs.

21.Describe some key steps in data visualization.

**Module-4**

1.What is a Decision tree? why are decision trees the most popular classification technique.

2.what is a splitting variable? Describe three criteria for choosing a splitting variable.

3. what is pruning? What are pre pruning and post pruning techniques? Why choose one over the other.

4.what are gini’s coefficient and information gain.

5.Solve problems on decision tree.

6.what is a regression model? Compare decision trees with regression models.

7.whta is a neural network? How does it work.

8. Explain design principles of an ANN.

9.Applications of ANN.

10.Steps required to build an ANN

11. advantages and disadvantages of ANN

12.Define cluster and what are the applications of cluster analysis.

13.Pseudocode for clustering.

14.write a note on clustering techniques.

14.Explain k-means algorithm for clustering.

15.solve problems using k-means

16.Advantages and Disadvantages of k-means algorithm.

17.Define association rule mining. what are the applications of association rules.

18.explain the representation of association rules.

19.Solve problems using apriori algorithm.

20.How does the apriori algorithm work.

21.What are frequent item sets.

**Module-5**

1. What is text mining? What are the applications of text mining.
2. Explain the process of text mining (text mining architecture)
3. Write a note on term document matrix and how to mine TDM explain with example.
4. Compare text mining and datamining.
5. Explain text mining best practices.
6. Define naïve-Bayes Analysis and explain the model of naïve Bayes with example.
7. Explain the classification of text with example.
8. What are the advantages and disadvantages of naïve-Bayes.
9. What is support vector machine and what are support vectors explain with example.
10. Explain the model of support vector machine (SVM)
11. Explain the kernel method.
12. What are the advantages & disadvantages of SVM.
13. Define web mining and what are the characteristics of optimized websites.
14. Explain the structure of web mining.
15. What are different web mining algorithms.
16. What is social network analysis and how it is different from other data mining techniques. (Clustering or Decision tree)
17. What are the applications of SNA.
18. Problems on finding the rank values of the node and compute the highest rank node
19. Comparison between social network analysis and data mining.
20. What are the different considerations that should be considered while designing SNA.