BTS-VI(SS)-01.19-0022 Reg. No. | | | | | | | | |

C

B.Tech. Degree VI Semester Special Supplementary Examination January 2019

CS 15-1604 DATA MINING

(2015 Scheme)

Time: 3 Hours Maximum Marks: 60

PART A

(Answer *ALL* questions)

(10 × 2 = 20)

I. (a) Explain the various types of Binning in data preprocessing.
(b) Distinguish classification, prediction and clustering with examples.
(c) Explain OLAP and its differences from OLTP.
(d) Explain SVM's with an example.
(e) Explain the basic idea behind Apriori algorithm.
(f) What are different types of clustering algorithms?
(g) Write a note on Linear regression based classifiers.
(h) Explain the characteristics of big data.
(i) Explain the hadoop distributed architecture.
(j) Write notes on any two databases for the big data.

PART B

(4 × 10 = 40)

II. Explain the various stages and applications of data mining process in detail, with examples for each.

OR

III. Explain the various types of data preprocessing with examples.

IV. Explain the decision tree algorithm with an example dataset. Make your own assumptions.

OR

V. Compare and contrast neural networks with SVM. Explain the various mathematical notions for both models.

VI. (a) Distinguish DBSCAN with hierarchical clustering algorithm.
(b) Assume (2, 2), (2, 4), (2, 6), (12, 2), (12, 4), (12, 6). Assume two clusters are to be formed with these points. Explain how K-Means will proceed with sample calculations.

OR

VII. Distinguish K-Means and KNN with examples.

VIII. Explain the software architecture of Hadoop based distributed computing in detail with diagrams. How does HIVE and Hbase differ?

OR

IX. Explain the distributed processing using mapreduce. Show a sample implementation using JAVA.

***