USN | | | | | | | | | |

13MCA442

Fourth Semester MCA Degree Examination, June/July 2019
Data Warehousing and Mining

Time: 3 hrs. Max. Marks:100

Note: ***Answer any FIVE full questions.***

1 a. Define a data warehouse. Explain the key features of a data warehouse. **(05 Marks)**
b. With a neat diagram, explain the three tier data warehouse architecture. **(10 Marks)**
c. Give the differences between :
i) Roll up and drill down
ii) Slice and dice. **(05 Marks)**

2 a. What is data mining? With neat diagram explain the KDD process in detail. **(05 Marks)**
b. Describe the different characteristics of data sets. **(05 Marks)**
c. Explain any four data mining techniques as detail. **(10 Marks)**

3 a. What is sampling? Disucss the various sampling techniques in detail. **(05 Marks)**
b. With a neat diagram, explain the feature subset selection process. **(05 Marks)**
c. Define :
i) Simple matching coefficient
ii) Jaccard coefficient
Calculate SMC and Jaccard values for the given binary vectors.
X = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)
Y = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1) **(10 Marks)**

4 a. Define Association rule discovery. Illustrate the Apriori principle with an example by considering the itemset {a, b, c, d, e} where {c, d, e} is the frequent itemset. **(10 Marks)**
b. Explain the two different rule generation techniques. **(10 Marks)**

5 a. What is a decision tree? Explain Hunts algorithm with an example. **(10 Marks)**
b. What is model over filtering? How do we handle model overfiltering in decision tree induction. **(05 Marks)**
c. Explain the different methods to evaluate the performance of a classifier. **(05 Marks)**

6 a. How do we build rule based classifier with an algorithm explain how do we extract rules using sequential covering method. **(10 Marks)**
b. Explain K-nearest algorithm in detail. **(05 Marks)**
c. What are the characteristics of Nearest neighbour classifiers? **(05 Marks)**

7 a. What is Cluster analysis? Discuss the different types of clustering. **(05 Marks)**
b. Explain K-means clustering algorithm with an example. **(10 Marks)**
c. What is hierarchical clustering? Discuss the key issues related to it. **(05 Marks)**

8 a. What is an outlier? Discuss the causes of anomalies. **(05 Marks)**
b. Discuss the different cluster based technique for outlier detection. **(10 Marks)**
c. Explain the proximity based outlier detection. **(05 Marks)**

* * * * *