Question Paper Code : X 20774

B.E./B.Tech. DEGREE EXAMINATIONS, NOVEMBER/DECEMBER 2020

Sixth/Seventh Semester

Information Technology

IT 6702 – DATA WAREHOUSING AND DATA MINING

(Common to Computer Science and Engineering)

(Regulations 2013)

(Also Common to PTIT 6702 – Data Warehousing and Data Mining for B.E.

Part-time – Sixth Semester – Computer Science and Engineering – Regulations – 2014)

Time : Three Hours   Maximum : 100 Marks

Answer ALL questions

PART – A   **(10×2=20 Marks)**

1. List the characteristics of a data warehouse.
2. Differentiate fact table and dimension table.
3. Define OLAP.
4. State the purpose of OLAP servers in data warehouse architecture.
5. Define data mining.
6. Outline the need for data pre-processing in data mining.
7. Define tree pruning.
8. What are lazy learners ?
9. Outline the difference between agglomerative clustering and divisive clustering.
10. What is outlier analysis ?

PART – B   **(5×13=65 Marks)**

11. a) What is a data warehouse ? With a neat sketch, explain the various components in a data warehouse. **(13)**

(OR)

b) Explain in the following DBMS schemas with an example :

    i) Star schema **(5)**

    ii) Snowflake schema **(4)**

    iii) Fact constellation schema. **(4)**

12. a) What is Cognos Impromptu ? Explain its features. **(13)**

　　　　　(OR)

　　b) Outline the following OLAP operations with an example and relevant diagrams.

　　　i) Roll-up **(5)**

　　　ii) Drill-down **(4)**

　　　iii) Slice and dice. **(4)**

13. a) i) Explain the steps of knowledge discovery in databases with neat sketch. **(7)**

　　　ii) Outline the primitives for specifying a data mining task. **(6)**

　　　　　(OR)

　　b) Explain min-max normalization and Z-score normalization with an example. **(13)**

14. a) Discuss the Apriori algorithm for discovering frequent item sets. Apply the Apriori algorithm to the following data set and find the frequent 3-item set with the minimum support value to be 2. **(13)**

TID	Items
1	Milk, Tea, Cake
2	Eggs, Tea, Cold drink
3	Milk, Eggs, Tea, Cold drink
4	Eggs, Cold drink
5	Juice

　　　　　OR)

　　b) Outline the steps in decision tree induction with an example. **(13)**

15. a) Present an outline of the different categories of clustering algorithms. **(13)**

(OR)

b) Explain density based clustering approach with an example. **(13)**

PART – C **(1×15=15 Marks)**

16. a) Describe K-means clustering with an example. **(15)**

(OR)

b) Apply Bayesian classification on the following data set and predict the class for the sample {Age = Youth, Income = medium, Student = yes, Credit = Fair}. **(15)**

Age	Income	Student	Credit	Target Buys Computer
Youth	High	No	Fair	No
Youth	High	No	Excellent	No
Middle	High	No	Fair	Yes
Senior	Medium	No	Fair	Yes
Senior	Low	Yes	Fair	Yes
Senior	Low	Yes	Excellent	No
Middle	Low	Yes	Excellent	Yes
Youth	Medium	No	Fair	No
Youth	Low	Yes	Fair	Yes
Senior	Medium	Yes	Fair	Yes
Youth	Medium	Yes	Excellent	Yes
Middle	Medium	No	Excellent	Yes
Middle	High	Yes	Fair	Yes
Senior	Medium	No	Excellent	No

—————————