USN | | | | | | | | | | |

10IS74

Seventh Semester B.E. Degree Examination, April 2018
Data Warehousing and Data Mining

Time: 3 hrs. Max. Marks:100

Note: ***Answer any FIVE full questions, selecting atleast TWO questions from each part.***

PART – A

1 a. What is Operational Data Store? Explain the Operational Data Store (ODS) structure, with a neat diagram. **(10 Marks)**

b. Compare the ODS and Data warehouse. Explain the steps in data cleaning. **(10 Marks)**

2 a. List and explain Codd's OLAP characteristics. **(10 Marks)**

b. Describe Slice and Dice operation. Explain the guidelines for OLAP implementation. **(10 Marks)**

3 a. Discuss the tasks of data mining, with suitable examples. **(10 Marks)**

b. Explain the various challenges that motivated the development of data mining. **(10 Marks)**

4 a. Develop the Apriori algorithm for generating frequent item set. **(08 Marks)**

b. For the transactional database shown below with minimum support = 3, construct a FP tree and show for each transaction how the tree evolves. **(12 Marks)**

TID	Items Bought
1	f, a, c, d, g, i, m, p
2	a, b, c, f, ℓ, m, o
3	b, f, h, j, o
4	b, c, k, s, p
5	a, f, c, e, ℓ, p, m, n

PART – B

5 a. Construct a decision tree for the following database : **(12 Marks)**

Day	Outlook	Temperature	Humidity	Wind	Play ball
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Strong	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

b. Summarize the characteristics of Nearest – Neighbor classifiers. **(08 Marks)**

6 a. Write an algorithm to generate the topology of a Bayesian Network. Discuss characteristics of Bayesian Belief Network. **(10 Marks)**

b. Summarize the characteristics of Naive Bayes classifiers. With an example, explain Multicast problem. **(10 Marks)**

7 a. Discuss different types of clusters. **(10 Marks)**

b. Write Bisecting K – means Algorithm. **(10 Marks)**

8 Write short notes on the following :

a. Web structure mining.

b. Web usage mining.

c. Spatial databases.

d. Temporal databases. **(20 Marks)**

*****