Student Exam No. _____

GANPAT UNIVERSITY

M. Tech SEMESTER II - INFORMATION TECHNOLOGY

REGULAR EXAMINATION MAY/JUNE - 2012

3IT203: Data Warehousing, OLAP & Data Mining

Time: 3 Hours] **[Total Marks: 70**

Instructions:

1. Figures to the right indicate full marks
2. Each section should be written in a separate answer book
3. Be precise and to the point in your answer

SECTION-I

Q.1

(A) Explain the architecture of data mining system? **4**

(B) Briefly describe the following advanced database systems and applications: object-oriented databases, spatial databases, text databases, multimedia databases, the world wide web. **5**

(C) Explain how data mining is used in health care analysis. **3**

OR

Q.1

(A) What is Association Rule Mining? How it differs from other data mining techniques? A dataset has six transactions. Let ***min_sup*** = 60% and ***min_conf*** = 60%. **6**

TID	Items_bought
T100	K, A, D, B, C
T200	D, A, C, E, B
T300	C, A, B, E
T400	B, A, D, E
T500	C, D, K, E, F, B
T600	A, B, C, E

- Find all large item-sets in database using ***Apriori*** and ***FP-growth***.
- Compare the efficiency of the two mining process.
- Strong association rules and Exact association rules for dataset.

(B) Explain data cleaning, data reduction, data integration & transformation in brief. **6**

Q.2

(A) Discuss the approaches for mining multi level association rules from the transactional databases. Give relevant example. **3**

SECTION-II

Q.4

(A) Describe the differences between the following architectures for the integration of a data mining system with a database or data warehouse system: *no coupling, loose coupling, semitight coupling, and tight coupling.* State which architecture you think is the most popular, and why? **6**

(B) What is web mining and how it is different from Data mining? How Web Content Mining and Web Structure Mining are different? What are the problems with web logs? Explain issues in both web taxonomies. **6**

OR

Q.4

(A) For the given training set in Table predict classification of the sample: **6**

a) {2, 1, 1} b) {0, 1, 1}

Using ***Simple Bayesian Classifier.***

Training data set for a classification using Naïve Bayesian Classifier

Sample	Attr1	Attr2	Attr3	Class C
1	1	0	1	1
2	0	1	1	2
3	2	1	2	1
4	1	1	2	2
5	0	2	2	1
6	2	2	1	1
7	2	0	2	2
8	1	0	1	2

(B) What is web mining and how it is different from Data mining? How Web Content Mining and Web Usage Mining are different? What are the problems with web logs? **6**

Q.5

(A) Draw a decision tree for following data using the concept buys_computer. Each internal node should represent a test on an attribute. Use ***InfoGain*** to find the best split. **6**

TID	Age	Income	Student	Credit Rating	Class: buys_computer
1	<=30	High	No	Fair	No
2	<=30	High	No	Excellent	No
3	31..40	High	No	Fair	Yes
4	>40	Medium	No	Fair	Yes
5	>40	Low	Yes	Fair	Yes
6	>40	Low	Yes	Excellent	No
7	31..40	Low	Yes	Excellent	Yes
8	<=30	Medium	No	Fair	No
9	<=30	Low	Yes	Fair	Yes
10	>40	Medium	Yes	Fair	Yes
11	<=30	Medium	Yes	Excellent	Yes
12	31..40	Medium	No	Excellent	Yes
13	31..40	High	Yes	Fair	Yes
14	>40	Medium	No	Excellent	No

(B) What are the drawbacks of ***Apriori*** rule mining algorithm and explain any ***TWO*** approaches to improve these drawbacks. **5**

OR

Q.5

(A) Explain the three-tier data warehouse architecture. **4**

(B) Differentiate Inter-transaction Association Rule Mining with standard Association Rule Mining? Why it is not applicable to market basket analysis? Explain steps to mine Inter-transaction Association Rule using Sliding window. Take suitable example for explanation. **4**

(C) Explain ***Gini Index*** an attribute selection measure used in classification. **3**

Q.6

(A) Which is popular in the data warehouse design, star schema model (or) snowflake schema model, Why? **5**

(B) Explain roll-up, drill-down, slice and dice operation with suitable example. **4**

(C) List out the differences between OLTP and OLAP. **3**

--- END OF PAPER ---