



NATIONAL SCHOOL OF BUSINESS MANAGEMENT

**BSc in Management Information Systems (Special) & BSc (Hons) Software
Engineering**

4th Year 2nd Semester Examination

31- Oct- 2020

CS405.3 - Data Warehousing and Data Mining

Instructions to Candidates

- 1) Answer all questions.**
- 2) Time allocated for the examination is three (03) hours.**
- 3) This paper contains 5 pages. If a page or a part of this question paper is not printed, please inform the Supervisor immediately.**
- 4) Write your index number in all pages of answer script.**
- 5) Staple all answer sheets at the end of the examination.**

Question 1 -20 Marks

1. Briefly explain the difference between Online Transaction Processing (OLTP) and Online Analytical Processing(OLAP) in terms of **users, type of data and storage size** (3 marks)
2. What is Star Schema in multidimensional data modeling? (3 marks)
3. Explain what is a Concept Hierarchy by using an example (3 marks)
4. Briefly explain Roll-up and Roll-down operations in Online Analytical Processing (3marks)
5. By using an illustration explain Base cuboid and Apex cuboid (3 marks)
6. What are the differences between Full Cube Materialization and No Materialization? (2 marks)
7. Briefly explain the tasks involve in data preprocessing (3 marks)

Question 2 -20 Marks

1. Briefly describe Market Basket Analysis technique (2 marks)
2. Briefly describe what is frequent pattern mining (2 marks)
3. What is Apriori property? (1 marks)
4. Consider the following database table of sales which extracted from a super market. Records with “1” represents that item was bought in the particular transaction and “0” represents that item was not bought in the transaction. By using Apriori algorithm answer the questions accordingly. Minimum Support Count is 2.

TransactionId	Milk (M)	Bread (B)	Cheese (C)	Potato (P)	Onion (O)
T1	1	1	1	0	0
T2	0	0	0	1	1
T3	0	1	1	1	0
T4	1	1	1	0	0
T5	1	1	1	0	1
T6	1	1	1	1	0

- I. Find the candidate 1-Itemset for the database (3 marks)
- II. Find the candidate 2-itemset and frequent 2-item set for the database (8 marks)
- III. Mine the most frequent itemset with respective support counts (4 marks)

Question 3 -20 Marks

- Following table extracted from a transaction processing database. By using FP-tree algorithm answer below questions. Minimum Support Count is 2.

TransactionId	Items
T1	B, C, D
T2	A, B
T3	A, C, D
T4	A, D, E
T5	A, B, C
T6	B, C, E
T7	C
T8	D, E
T9	C, B
T10	B, C, D
T11	A, B, C, D, E

- Find the candidate 1-itemset for the dataset and the List denoted by "L" with ordered set of items in descending order. (4 marks)
- Draw the FP-Tree for given data set. (6 marks)
- Find the Conditional pattern bases for each item in the dataset. (5 marks)
- Mine the frequent patterns of each item using FP-Tree for each. (5 marks)

Question 4 -20 Marks

- Briefly explain Classification task in data mining. (2 marks)
- Bellow table shows the result of a classifier which used to predict Cancer patients. Horizontally (Rows) represents the actual classes and Vertically (Columns) represents the predictions of the classifier. By considering Accuracy, Sensitivity and Specificity measures evaluate the performance of this classifier. (3 marks)

Classes	Cancer = Yes	Cancer = No	Total
Cancer = Yes	70	180	250
Cancer = No	120	9630	9750
Total	190	9810	10000

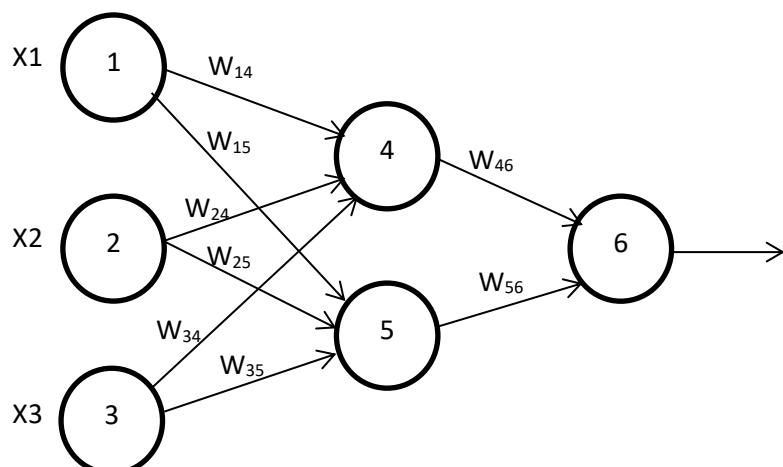
- Consider the following database table which was extracted from a fraud detection system used by a bank. By using Naïve Bayes classification algorithm find out that below query would classifies as fraud = Yes or No. (15 marks)

$X = (\text{Age} = \text{Youth}, \text{AssertDeclared} = \text{No}, \text{Income} = \text{High}, \text{EmploymentType} = \text{Contract})$

RecordID	Age	AssertDeclared	Income	EmploymentType	Fraud
T1	Youth	Yes	High	Permanent	No
T2	Middle	Yes	Medium	Permanent	No
T3	Middle	Yes	High	Permanent	No
T4	Middle	No	Medium	Permanent	No
T5	Youth	No	Low	Self	Yes
T6	Senior	Yes	Medium	Self	No
T7	Senior	No	High	Permanent	Yes
T8	Youth	Yes	Low	Contract	Yes
T9	Senior	Yes	Low	Contract	Yes
T10	Youth	No	High	Self	No
T11	Youth	Yes	Medium	Contract	No
T12	Middle	No	Low	Contract	No
T13	Senior	Yes	Medium	Permanent	No
T14	Middle	No	Low	Self	Yes
T15	Youth	No	High	Permanent	No

Question 5 -20 Marks

- Below diagram shows a multilayer feed forward neural network with one hidden layer. X_1, X_2, X_3 are the input tuples for the network and W_{14}, W_{15}, \dots represents the respective weights of the links between nodes. Initial weights and biases (represented as θ_4, θ_5 and θ_6 for the bias of unit4, unit5 and unit6 respectively) are given in the table below the network diagram. By using back propagation algorithm answer the questions accordingly.



X1	X2	X3	W ₁₄	W ₁₅	W ₂₄	W ₂₅	W ₃₄	W ₃₅	W ₄₆	W ₅₆	θ4	θ5	θ6
1	1	0	0.3	-0.4	-0.2	0.5	0.1	-0.5	-0.2	0.4	-0.2	0.3	0.1

- I. If X1 = 1, X2 = 1 and X3 = 0 inputs to the network via unit1, unit2 and unit3 respectively compute the Net Input and Output of the each unit using below two equations. (6 marks)

$$I_j = \sum W_{ij} O_i + \theta_j$$

$$O_j = \frac{1}{1 + e^{-I_j}}$$

- II. If Output of the unit 6 is given as 1 compute the error of output layer using below equation (2 marks)

$$Err_j = O_j(1 - O_j)(T_j - O_j)$$

- III. By using below equation compute the errors of units 5 and 4. (4 marks)

$$Err_j = O_j(1 - O_j) \sum_K Err_K W_{jk}$$

- IV. If learning rate (L) is given as 0.9 adjust the weights of W₅₆ and W₄₆ and bias of the unit6. (8 marks)

$$W_{ij} = W_{ij} + (L)Err_j O_i$$

$$\theta_j = \theta_j + (L)Err_j$$

.....**END OF THE PAPER.....**