Reg. No.										

MANIPAL ACADEMY OF HIGHER EDUCATION, MANIPAL
MANIPAL SCHOOL OF INFORMATION SCIENCES, MANIPAL

THIRD SEMESTER MASTER OF SCIENCE – M.Sc. (INFORMATION SCIENCE)

DEGREE EXAMINATION – NOVEMBER 2021

SUBJECT: MIS 603 – DATA MINING AND WAREHOUSING

Wednesday, November 24, 2021

Time: 10.00 – 13.00 Hrs. Max.Marks:100

All questions carry 10 marks each

1. Give comparison of OLTP & OLAP (10 Marks)

2. Apply the Apriory algorithm for the following TDB generate all the possible candidates? minimum is confidence 50%

$Sup_{min} = 2$

Database TDB

Tid	Items
10	A, C, D
20	B, C, E
30	A, B, C, E
40	B, E

3. List and describe the five *primitives* for specifying a data mining task. (10 Marks)

4. Write short notes on 'Architecture of data mining systems'? Discuss Syntax for DMQL for a. Specification of Interestingness Measures b. Specification of Pattern Presentation (10 Marks)

5. Demonstrate Mining Frequent Patterns without Candidate Generation using the following transaction DB by Construct FP-tree? Mine Frequent Patterns by Creating Conditional Pattern-Bases (10 Marks)

Table 6.1 Transactional Data for an *AllElectronics* Branch

TID	List of item_IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

6. How does ARCS work? Briefly explain with a neat diagram? List Techniques for Mining MD Associations and explain them briefly. (10 Marks)

7. For the class buys_computer(D) in following dataset calculate Entropy (D), Information Gain(age) and Information Gain(income). (10 Marks)

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

8. List the strengths and weakness of Neural Network as a Classifier give small explanation for each of them (10 Marks)

9. For the following dataset using the Naïve bayes classifier calculate (10 Marks)
   a. P(Ci):P(buys_computer = "yes") , P(buys_computer = "no")
   b. Given with X = (age <= 30 , income = medium, student = yes, credit_rating = fair), Compute P(X|Ci) for each class
   c. P(X|Ci)*P(Ci), mention X belongs to which class

10. Give an example of how specific clustering methods may be integrated, for example, where one clustering algorithm is used as a preprocessing step for another. In addition, provide reasoning on why the integration of two methods may sometimes lead to improved clustering quality and efficiency (10 Marks)

******