

Question Paper

Exam Date & Time: 04-Jan-2023 (10:00 AM - 01:00 PM)



MANIPAL ACADEMY OF HIGHER EDUCATION

Manipal School of Information Sciences (MSIS), Manipal
Third Semester M.Sc. Information Science Degree Examination - January 2023

Data Mining and Warehousing [MIS 603]

Marks: 100

Duration: 180 mins.

Wednesday, January 4, 2023

Answer all the questions.

- 1) Clearly illustrate the KDD process with a diagram, and briefly describe each stage. (10)
- 2) Discuss choosing an attribute subset using the following ways given the initial attribute set: "A1, A2, A3, A4, A5, A6." (10)
 - a) Step-wise forward selection
 - b) Step-wise backward elimination
 - c) Decision-tree induction
- 3) Explain conceptual modelling of DWH (10)
- 4) What requirements are to be considered while designing GUIs for data mining query languages? Describe the architecture of data mining systems in brief? 10 marks (10)
- 5) Determine information Gain for each of the following Target and Contrasting classes: gender, major, birth country, age range, and gpa (10)

?

gender	major	birth_country	age_range	gpa	count
M	Science	Canada	20-25	Very_good	16
F	Science	Foreign	25-30	Excellent	22
M	Engineering	Foreign	25-30	Excellent	18
F	Science	Foreign	25-30	Excellent	25
M	Science	Canada	20-25	Excellent	21
F	Engineering	Canada	20-25	Excellent	18

Candidate relation for Target class: Graduate students ($\Sigma=120$)

gender	major	birth_country	age_range	gpa	count
M	Science	Foreign	<20	Very_good	18
F	Business	Canada	<20	Fair	20
M	Business	Canada	<20	Fair	22
F	Science	Canada	20-25	Fair	24
M	Engineering	Foreign	20-25	Very_good	22
F	Engineering	Canada	<20	Excellent	24

Candidate relation for Contrasting class: Undergraduate students ($\Sigma=130$)

- 6) Compare descriptive and predictive data mining? write Definition of concept description? Write briefly about data generalization and characterization based on summarization? 10 marks (10)
- 7) Give A Decision Tree using the following training data for "buys computer" ? Evaluate Gain based on age, income, students, and credit rating? (10)

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) Provide brief summaries of A) Attribute Selection Measures. B) Tree pruning and overfitting (10)
- 9) Using a precise diagram, describe a Neuron A Hidden/Output Layer Unit? (10)
- 10) Cluster the following 8 points into three clusters using (X, Y) as locations: A1(2,10), A2(2,5), A3(8,4), A4(5,8), A5(7,5), A6(6,4), A7(1,2), A8 (4,9). (10)
A1, A4, and A7 are the initial cluster centers.

-----End-----