

Question Paper

Exam Date & Time: 13-May-2023 (09:30 AM - 12:30 PM)



MANIPAL ACADEMY OF HIGHER EDUCATION

INTERNATIONAL CENTRE FOR APPLIED SCIENCES END SEMESTER THEORY EXAMINATION-MAY 2023 IV SEMESTER B.Sc.(APPLIED SCIENCES) IN ENGG.

DATA WAREHOUSING AND DATA MINING [ICS 247]

Marks: 50

Duration: 180 mins.

Answer all the questions.

Missing data, if any, may be suitably assumed.

- 1) What is data mining? What are the benefits of using data mining techniques in business and other industries? Discuss by taking one example. (5)
 - a)
 - b)
 - c)
- 2) What is data preprocessing, and why is it essential for data mining? (3)
 - a)
 - b)
 - c)
- 3) What are some examples of real-world applications of data mining? Justify. (2)
- 4) Use the Apriori algorithm to generate frequent item sets and association rules from the table given below with a $\text{mini_support} \geq 2$ and $\text{confidence} \geq 50\%$. (5)
 - a)
 - b)
 - c)

Trans-ID	Items Purchased
T001	Bread, Milk, Eggs
T002	Bread, Diapers, Beer, Eggs
T003	Milk, Diapers, Beer, Chips
T004	Bread, Milk, Diapers, Beer
T005	Bread, Milk, Diapers, Chips

- 5) How does the Apriori algorithm compare with other frequent itemset mining techniques, such as the FP-growth algorithm? (3)
 - a)
 - b)
 - c)
- 6) What is OLAP, and how does it differ from OLTP (Online Transaction Processing)? (2)
 - a)
 - b)
 - c)
- 7) How does feature selection help to improve model performance and reduce overfitting? Discuss one technique. (5)
 - a)
 - b)
 - c)
- 8) What is the difference between feature selection and feature extraction? (3)
 - a)
 - b)
 - c)
- 9) Give the advantages and disadvantages of using decision trees. (2)
- 10) How does the value of k affect the overall fit of the model in the k -NN (5)

- algorithm?
- a) Give an example and show the results obtained as k varies. What happens when $k=N$?
 - b) How do you construct a Bayesian belief network and what factors are involved in the process? Discuss with one example. (3)
 - c) What is dimensionality reduction, and why is it necessary in data preprocessing? (2)
- 5) What is Clustering? What are the basic steps involved in the clustering process? Explain one method with an example. (5)
- a)
 - b) Can you discuss some applications of artificial neural networks in real-world scenarios? (3)
 - c) What is the kernel trick in Support Vector Machine? (2)

-----End-----