

TILAK MAHARASHTRA VIDYAPEETH, PUNE
MASTER OF SCIENCE (M.Sc) in COMPUTER APPLICATIONS
EXAMINATION : MAY - 2026
SEMESTER – III

Sub : Data Mining & KDD (MSC24-301)

Date : 30/05/2026

Total Marks : 60

Time: 2.00 pm To 4.30 pm

Instruction:

1. All questions are compulsory unless and otherwise stated.
2. Bold figures to the right of every question are the maximum marks for that question.
3. Candidates are advised to attempt questions in order.
4. Answers written illegibly are likely to be marked zero.
5. Use of scientific calculators, Log tables, Mollier Charts is allowed.
6. Draw neat and labelled diagram wherever necessary.

Q. 1. Answer the following in 2 - 3 lines. (Any 5)

(10)

1. What is a fact table?
2. What is a data cube?
3. Define confidence for an association rule.
4. What is model usage?
5. Give one advantage of Naïve Bayes algorithm.
6. What is the Gini index?
7. What is the Gain Ratio?

Q. 2. Answer the following in short. (Any 4)

(20)

1. What is prediction in data mining?
2. Explain logistic regression in data mining.
3. How does machine learning relate to data mining?
4. Difference between OLAP and OLTP.
5. Describe hierarchical clustering?
6. Explain "subject-oriented" in data warehousing.

Q. 3. Answer the following in detail. (Any 3)

(30)

1. How do data mining tools support real-world problem solving?
2. Describe distance-based agglomerative clustering.
3. Why is visualization important in data mining tools? Explain with example.
4. Describe K-Means algorithm. What is the main limitation of K-Means?
5. Describe Bayes' theorem.
