2 COPY

QP CODE 34623

Sixth Semester B.C.A Degree Examinations

MAY/JUNE 2026

(CBCS NEP Scheme)

PAPER NBF 0230: DATA MINING AND DATA WAREHOUSE

Time: 2 hrs] [Max. Marks: 60

Instructions:

1. ***The Students should legibly write Section number along with question numbers within the margin ONLY.***
2. ***The answer without Section number and question numbers will not be valued.***
3. ***Draw neat and labelled diagram wherever necessary.***
4. ***Section – I multiple Choice Questions should be answered in the first page of main answer book and same questions shall not be answered repeatedly.***
5. ***Any identifications like marking, ticking, dots etc., in the options of MCQ (Section – I) on the question paper is strictly prohibited. Do not write anything (except register number) on the question paper.***

SECTION – I

I. Select the most appropriate answer from the options provided: $10 \times 1 = 10$

I - 1) Which step is not a part of the KD process
a) data cleaning b) data selection c) data duplication d) data mining

I - 2) Which of the following is a major issue in data mining
a) Noisy Data b) Scalability
c) Data privacy d) All of the above

I - 3) Which of the following best describes the Apriori Algorithm
a) It uses top down approach to generate frequent item sets
b) It uses a bottom up approach to generate frequent item sets
c) It compresses the data set uses an FP tree
d) It is used for clustering

I - 4) How does the FP growth algorithm initially process the data set
a) By generating candidate item sets of increasing length
b) By constructing an FP tree to represent the data set
c) By clustering similar transactions
d) By building decision tree

Contd..... 2

I - 5) Which of the following classifier uses Bayes theorem for classifications.

a) Decision tree b) K – nearest neighbour

c) Naïve Bayer d) Rule based classifier

I - 6) Which of the following is a common distance metric used in k – nearest neighbor algorithm.

a) Bayesian distance b) Decision distance

c) Euclidean distance d) Rule distance

I - 7) What is the advantage of DBSCAN algorithm

a) It require number of cluster to be specified in advance

b) It can find cluster of arbitrary shape

c) It assumes clusters are spherical

d) It is sensitive to initial centroid placement.

I - 8) Which of the following is a grid based clustering method.

a) DBSCAN b) Sting

c) K – means d) K – medoids

I - 9) Which of the following is not a characteristic of a data warehouse

a) Subject oriented b) Integrated

c) Time variant d) Volatile

I - 10) Which OLAP system stores data in a Multi dimensional cube?

a) ROLAP b) HOLAP c) MOLAP d) POLAP

SECTION – II

II. Answer any FIVE of the following: $5 \times 3 = 15$

II - 1) Explain KDD.

II - 2) Discuss the Application of Data Mining.

II - 3) Write the Mining Association Rules.

II - 4) Explain Decision tree induction.

II - 5) Explain lazy Learners.

II - 6) Briefly Explain Cluster Analysis.

II - 7) Discuss the basic concepts of data warehouse.

Contd..... 3

II - 8) Explain Data Cube Operations.

SECTION – III

$3 \times 5 = 15$

III. Answer any THREE questions from the following:

III - 1) Differentiate between DBMS and Data mining.

III - 2) What is Apriori Algorithm? Explain in detail.

III - 3) Discuss the K – Nearest Neighbour approach.

III - 4) Explain cluster partitioning method.

III - 5) Explain OLAP.

SECTION – IV

$2 \times 10 = 20$

IV. Answer the following:

IV - 1) a) Explain the Data Mining Techniques.

b) Explain data mining issues and challenges. (5 + 5)

OR

a) Write short note on mining frequent patterns.

b) Explain data cleaning process in data mining. (5 + 5)

IV - 2) a) Explain Density based clustering.

b) Discuss the evaluation of clustering method. (5 + 5)

OR

a) Explain data warehouse Modeling.

b) Explain the implementation of OLAP. (5 + 5)

* * * * * * *