RayatShikshanSantha's

S M Joshi College,Hadapsar Pune 28

T .Y .B.C.A (Science) Sem-I

Internal Examination 2021-2022

Data Mining and Data Science (BCA-352)

Date: - 5-12-2020 **Marks-70**

Time:-3 Hrs

Q.1) A) Choose the correct option: **(1*5=5)**

1) ----------is not data mining functionality?

a. Clustering and Analysis b. Selection and Interpretation

c. Classification and Regression d. Characterization and Discrimination

2) To remove noise and inconsistent data ---------- is needed.

a. Data Cleaning b. Data Transformation

c. Data Reduction d. Data Integration

3) OLAP stands for----------

a. Online Analytical Processing b. Online Analysis Processing

c. Online Transaction Processing d. Online Aggregate Processing

4) Which classification scheme makes use of IF-THEN rules for class prediction?

a. Rule –based Classification b. Statistical based Classification

c. Distance based classification d. Decision tree based classification

5) Number of students of a class is an example of ----------

a. Discrete Data b. Ordinal Data

c. Nominal Data d. Summary Data

Q1.B) Define the Following terms. **(5)**

1) Star Schema

2) Information Gain

3) Regression

4) Classification

5) Data Cleaning

Q 2.Solve any **Five** Questions (3*5) =15

a) List the characteristics of OLAP systems.

b) Discuss about the major issues in data mining.

c) Differentiate between Query processing and Data Mining

d).List out steps of KDD process.

e) What are the advantages of using decision trees?

f) What is decision tree pruning?

Q 3.Solve any **Five** Questions (4*5) =20

a .Explain various applications of data science.

b. Describe the K-nearest neighbour classifier.

c. What is difference between BI and data Science?

d. What is major difference between the star schema and the snowflake schema?

e. What are the benefits of building an enterprise data warehouse?

f. Describe different challenges of data science technology.

g. Describe the features of a data warehouse.

Q 4.Solve any **Five** Questions (5*5) =25

a. What is Data Cleaning? Describe various methods of data cleaning.
b. Describe in detail the concept of Machine Learning.
c. Describe Bayesian classification.
d. What is the difference between Data Analysis and Data Analytics?
e. Explain the working of SVM classifier.
f. Explain a multidimensional view and a data cube.
g. Explain various steps in data pre-processing.

RayatShikshanSantha's

S M Joshi College,Hadapsar Pune 28

T .Y .B.C.A (Science) SEM-V

Internal Examination 2022-2023

Data Mining and Data Science (BCA-352)

Date: - 29-11-2021 **Marks-70**

Time:-3 Hrs

Q.1) A) Choose the correct option: **(1*5=5)**

1) ----------is the output of KDD.

a. Query b. Useful Information

c. Data d. Information

2) ----------describes the data contained in the data warehouse.

a. Relational data b. Operational data

c. Metadata d. Informational data

3) ----------is a good alternative to the star schema.

a. Star schema b. Snowflake schema

c. Fact constellation d. Star-snowflake schema

4) Decision Nodes are represented by ----------.

a. Disks b. Squares

c. Circles d. Triangles

5) A collection of one or more items is called as ----------.

a. Itemset b. Confidence

c. Support d. Summary Data

Q1.B) Define the Following terms. **(5)**

1) Clustering

2) Regression

3) Supervised learning

4) Classifier

5) Association Rule

Q 2.Solve any **Five** Questions (3*5) =15

a) Describe different challenges of data science technology?

b) What are different applications of clustering?

c) Describe about Prediction.

d) List the classification categorization algorithms.

e) Describe the features of a data warehouse.

f) List out steps of KDD process.

Q 3.Solve any **Five** Questions (4*5) =20

a .Explain various steps in data pre-processing.

b. Describe in detail the concept of Machine Learning.

c. Describe Bayesian classification?

d. Explain in detail different types of data.

e. Differentiate between Agglomerative and Divisive clustering method.

f. Explain the star scheme technique of modeling a data warehouse.

g. Discuss about the major issues in Data Mining?

Q 4.Solve any **Five** Questions (5*5) =25

a. What is Data Cleaning? Describe various methods of data cleaning.
b. What are the major differences between OLTP and a data warehouse system?.
c. Explain K-nearest classifier with an example.
d. Explain different types of data used in cluster analysis?
e. What is major difference between the star schema and the snowflake schema?
f. Explain decision tree classifier with an example.
g. What is data Reduction? Discuss briefly the techniques for data reduction.

RayatShikshanSantha's

S M Joshi College,Hadapsar Pune 28

T .Y .B.C.A (Science) Sem-I

Internal Examination 2023-2024

Data Mining and Data Science (BCA-352)

Date: - 21-10-2023 **Marks-70**

Time:-3 Hrs

Q.1) A) Choose the correct option: **(1*5=5)**

1) Which of the following is not a data pre-processing methods?

a. Data Visualization b. Data Discretization

c. Data Cleaning d. Data Reduction

2) The data Warehouse is ----------

a. read only b. write only

c. read write only d. None of the mentioned

3) ----------is used to map a data item to a real valued prediction variable.

a. Classification b. Regression

c. Clustering d.Summarization

4) Frequency of occurrence of an itemset is called as----------

a. Support b. Confidence

c. Support Count d. Itemset

5) ----------type of data contains only numbers.

a. Quantitative b. Unstructured

c. Structured d. Qualitative

Q1.B) Define the Following terms. **(5)**

1. Data Science

2. Clustering

3. Regression

4. Classifier

5. Data Mining

Q 2.Solve any **Five** Questions (3*5) =15

a. Discuss about the major issues in Data Mining.

b. Describe Bayesian classification.

c. Describe the features of a data warehouse.

d. What are advantages and disadvantages of K-means algorithm?

e. Explain continuous data with example.

f. compare descriptive and predictive data mining.

Q 3.Solve any **Five** Questions (4*5) =20

a .Explain various applications of data science.

b. Describe the K-nearest neighbor classifier.

c. Describe in detail the concept of Machine Learning.

d. Explain various steps in data pre-processing.

e. What are the benefits of building an enterprise data warehouse?

f. Explain the working of SVM classifier.

g. What are different types of clustering?

Q 4.Solve any **Five** Questions (5*5) =25

a. What is Data Cleaning? Describe various methods of data cleaning.
b. What are the major differences between an OLTP system and a data warehouse system?
c. Write a short note on linear and non-linear regression.
d. Consider the following transaction table and generate the candidate itemsets and frequent itemsets where the minimum support count is 2.

TID	List Of Items
1	Milk,Tea,Cake
2	Eggs,Tea,Cold Drink
3	Milk,Eggs,Tea,Cold Drink
4	Eggs,Cold Drink
5	Juice

e. Describe difference between Regression and Classification.
f. Explain a multidimensional view and a data cube.
g. Suppose that the data mining task is to cluster the following eights point (with(x;y)representing location into three cluster.
A1(2;10);A2(2;5);A3(8;4);A4(5;8);A5(7;5);A6(6;4);A7(1;2);A8(4;9).
The distance function is Euclidean distance.Suppose we assign A1, A4, A7 as the center of each cluster.Use K-means Algorithm.
a.The three cluster centers after the first round of execution.
b.The final three cluster.