**M.Sc. Final (Computer Science) Sem IV**

**Session 2021-2022**

**Paper I**

**DATA MINING**

**Multiple Choice Questions & Answers**

**Unit I**

Q1. Data Mining System Classification consists of?

A. Database Technology

B. Machine Learning

C. Information Science

**D. All of the above**

Answer: D. All of the above

Q2. How many categories of functions involved in Data Mining?

**A. 2**

B. 3

C. 4

D. 5

Answer: A. 2

Q3. "Efficiency and scalability of data mining algorithms" issues comes under? A. Mining Methodology and User Interaction Issues

**B. Performance Issues**

C. Diverse Data Types Issues

D. None of the above

Answer: B. Performance Issues

Q4. Which of the following is correct application of data mining?

A. Market Analysis and Management

B. Corporate Analysis & Risk Management

C. Fraud Detection

**D. All of the above**

Answer: D. All of the above

Q5. What is true about data mining?

A. Data Mining is defined as the procedure of extracting information from huge sets of data B. Data mining also involves other processes such as Data Cleaning, Data Integration, Data Transformation

C. Data mining is the procedure of mining knowledge from data.

**D. All of the above**

Answer: D. All of the above

Q6. \_\_\_\_\_\_\_\_\_\_ refers to the description and model regularities or trends for objects whose behavior changes over time.

A. Outlier Analysis

**B. Evolution Analysis**

C. Prediction

D. Classification

Answer: B. Evolution Analysis

Q7. Which of the following statements is correct about data mining?

A. It can be referred to as the procedure of mining knowledge from data

B. Data mining can be defined as the procedure of extracting information from a set of the data

C. The procedure of data mining also involves several other processes like data cleaning, data transformation, and data integration

**D. All of the above**

Answer: D. All of the above

Q8. Which of the following activities is NOT a data mining task?

A. Predicting the future stock price of a company using historical records B. Monitoring and predicting failures in a hydropower plant

**C. Extracting the frequencies of a sound wave**

D. Monitoring the heart rate of a patient for abnormalities

Answer: C. Extracting the frequencies of a sound wave

Q9. To detect fraudulent usage of credit cards, the following data mining task should be used **A. Outlier analysis**

B. prediction

C. association analysis

D. feature selection

Answer: A. Outlier analysis

Q10. Which of the following is NOT example of ordinal attributes?

**A. Zip codes**

B. Ordered numbers

C. Movie ratings

D. Military ranks

Answer: A. Zip codes

Q11. Identify the example of Nominal attribute

A. Temperature

B. Salary

C. Mass

**D. Gender**

Answer: D. Gender

Q12. Synonym for data mining is \_\_\_\_\_\_\_\_\_\_\_\_

A. Data Warehouse

**B. Knowledge discovery in database**

C. Business intelligence

D. OLAP

Answer: B. Knowledge discovery in database

Q13. Nominal and ordinal attributes can be collectively referred to as\_\_\_\_\_\_\_\_\_ attributes A. perfect

**B. qualitative**

C. consistent

D. optimized

Answer: B. qualitative

Q14. Incorrect or invalid data is known as \_\_\_\_\_\_\_\_\_

A. Missing data

B. Outlier

C. Changing data

**D. Noisy data**

Answer: D. Noisy data

Q15. "Handling of relational and complex types of data" issue comes under? A. Mining Methodology and User Interaction Issues

B. Performance Issues

**C. Diverse Data Types Issues**

D. None of the above

Answer: C. Diverse Data Types Issues

Q16. Data set {brown, black, blue, green , red} is example of \_\_\_\_ A. Continuous attribute

B. Ordinal attribute

C. Numeric attribute

**D. Nominal attribute**

Answer: D. Nominal attribute

Q17. The \_\_\_\_\_\_\_\_\_ refers to extracting knowledge from larger amount of data. A. data abstraction.

B. data warehouse.

C. database.

**D. data mining.**

Answer: D. data mining.

Q18. \_\_\_\_\_\_\_\_\_\_\_ is used for discrete target variable.

A. Nominal.

**B. Classification.**

C. Clustering.

D. Association

Answer: B. Classification.

Q19. Binary attribute are

**A. This takes only two values. In general, these values will be 0 and 1 and they can be coded as one bit**

B. The natural environment of a certain species.

C. Systems that can be used without knowledge of internal operations.

D. None of these

Answer: A. This takes only two values. In general, these values will be 0 and 1 and they can be coded as one bit

Q20. Which of the following is/are the Data mining tasks?

A. Regression

B. Classification

C. Clustering

D. inference of associative rules

**E. All A, B, C and D above.**

Answer: E. All A, B, C and D above.

Q21. Dimensionality reduction reduces the data set size by removing \_\_\_\_\_\_\_\_\_\_\_\_. A. relevant attributes.

**B. Irrelevant attributes.**

C. Support vector attributes.

D. Mining attributes

Answer: B. Irrelevant attributes.

Q22. Which indicator is used for similarity between two sets?

A. Rope Tree

**B. Jaccard Coefficient**

C. Tango Tree

D. MinHash Coefficient

Answer: B. Jaccard Coefficient

Q23. Height of a person, can be considered as an attribute of \_\_\_\_\_type? A. Nominal

B. Ordinal

C. Interval

**D. Ratio**

Answer: D. Ratio

Q24. Minkowski distance is a function used to find the distance between two A. Binary vectors

B. Boolean-valued vectors

**C. Real-valued vectors**

D. Categorical vectors

Answer: C. Real-valued vectors

Q25. Which of the following distance measure is similar to Simple Matching Coefficient (SMC)?

A. Euclidean distance

**B. Hamming distance**

C. Jaccard distance

D. Manhattan distance

Answer: B. Hamming distance

Q26. What is the range of the cosine similarity of the two documents?

**A : Zero to One**

B : Zero to infinity

C : Infinity to infinity

D : Zero to Zero

Answer: A : Zero to One

Q27. When are the members of two sets more common relatively? **A. Jaccard Index is Closer to 1**

B. Jaccard Index is Closer to 0

C. Jaccard Index is Closer to -1

D. Jaccard Index is Farther to 1

Answer: A. Jaccard Index is Closer to 1

Q28. Among these, which sampling is based on equal probability?

**A. Simple random sampling**

B. Stratified random sampling

C. Systematic sampling

D. Probability sampling

Answer: A. Simple random sampling

Q29. Which of the following is a predictive model?

A. clustering.

**B. regression.**

C. summarization.

D. association rules

Answer: B. regression.

Q30. Which of the following is a descriptive model?

A. classification.

B. regression.

**C. sequence discovery.**

D. association rules.

Answer: C. sequence discovery.

……………………………………………………………………………………………**Unit II**

Q1. Which of the following statement is true about the classification? A. It is a measure of accuracy

**B. It is a subdivision of a set**

C. It is the task of assigning a classification

D. None of the above

Answer: B. It is a subdivision of a set

Q2. The analysis performed to uncover interesting statistical correlations between associated-attribute-value pairs is called?

A. Mining of Association

B. Mining of Clusters

**C. Mining of Correlations**

D. None of the above

Answer: C. Mining of Correlations

Q3. The important characteristics of structured data are

A. Sparsity, Resolution, Distribution, Tuples

B. Sparsity, Centroid, Distribution , Dimensionality

C. Resolution, Distribution, Dimensionality ,Objects

**D. Dimensionality, Sparsity, Resolution, Distribution**

Answer: D. Dimensionality, Sparsity, Resolution, Distribution

Q4. Classification rules are extracted from\_\_\_\_\_\_\_\_\_\_.

A. root node.

**B. decision tree.**

C. siblings.

D. branches.

Answer: B. decision tree.

Q5. In Data Characterization, class under study is called as?

A. Study Class

B. Intial Class

**C. Target Class**

D. Final Class

Answer: C. Target Class

Q6. OLAP stands for \_\_\_\_\_\_\_\_.

**A. Online Analytical Processing.**

B. Online Linear Analytical Processing.

C. Online Animated Process.

D. Online Analytical Problem.

Answer: A. Online Analytical Processing.

Q7. The information on two attributes is displayed in \_\_\_\_\_\_\_\_\_\_\_\_ in scatter diagram.

A. visualization space.

B. scatter space.

**C. cartesian space.**

D. interactive space.

Answer: C. cartesian space.

Q8. OLAP is used to explore the \_\_\_\_\_\_\_\_\_\_\_ knowledge.

A. shallow.

B. deep.

**C. multidimensional.**

D. hidden

Answer: C. multidimensional.

Q9. A natural way to visualize the process of training a self-organizing map is called \_\_\_\_\_\_\_\_\_\_.

**A. kohonen movie.**

B. kohonen map.

C. frame.

D. scatter diagram.

Answer: A. kohonen movie.

Q10. The decision support system is used only for \_\_\_\_\_\_\_.

A. cleaning.

B. coding.

C. selecting.

**D. queries.**

Answer: D. queries.

Q11. \_\_\_\_\_\_\_\_ is the technique which is used for discovering patterns in dataset at the beginning of data mining process.

A. Kohenon map.

**B. Visualization.**

C. OLAP.

D. SQL.

Answer: B. Visualization.

Q12. Which one of the following is not true about OLAP?

A. They create no new knowledge.

**B. OLAP is powerful that data mining tool.**

C. They cannot search for new solution.

D. OLAP tool store their data in special multidimensional format.

Answer: B. OLAP is powerful that data mining tool.

Q13. Classification rules are extracted from\_\_\_\_\_\_\_\_\_\_.

A. root node.

**B. decision tree.**

C. siblings.

D. branches

Answer: B. decision tree.

Q14. Which of the following statement is true about the classification? A. It is a measure of accuracy

**B. It is a subdivision of a set**

C. It is the task of assigning a classification

D. None of the above

Answer: B. It is a subdivision of a set

Q15. Which of the following can be considered as the classification or mapping of a set or class with some predefined group or classes?

A. Data set

B. Data Characterization

C. Data Sub Structure

**D. Data Discrimination**

Answer: D. Data Discrimination

Q16. The analysis performed to uncover the interesting statistical correlation between associated -attributes value pairs are known as the \_\_\_\_\_\_\_.

A. Mining of association

**B. Mining of correlation**

C. Mining of clusters

D. All of the above

Answer: B. Mining of correlation

Q17. The issues of “Scalability and efficiency of the data mining algorithms” come under:

A. User Interaction and Mining Methodology Issues

B. Diverse Data Types Issues

**C. Performance Issues**

D. None of the above

**Answer:** C. Performance Issues

Q18. Decision tree is the most powerful for \_\_\_\_\_\_\_\_

A. classification

B. prediction

**C. both a and b**

D. none of these

Answer: C. both a and b

Q19. Decision-tree algorithm falls under the category of \_\_\_\_\_\_\_\_

A. unsupervised learning algorithms

B. reinforcement learning algorithm

**C. supervised learning algorithms**

D. prone to errors in classification problems with many class

Answer: C. supervised learning algorithms

Q20. \_\_\_\_\_\_\_is the measure of uncertainty of a random variable, it characterizes the impurity of an arbitrary collection of examples.

A. Information Gain

B. Gini Index

**C. Entropy**

D. none of these

Answer: C. Entropy

Q21. \_\_\_\_\_\_\_is a metric to measure how often a randomly chosen element would be incorrectly identified.

A. Information Gain

**B. Gini Index**

C. Entropy

D. none of these

Answer: B. Gini Index

………………………………………………………………………………………

**Unit III**

Q1. \_\_\_\_\_\_\_\_\_ is an example for case based-learning.

A. Decision trees.

B. Neural networks.

C. Genetic algorithm.

**D. K-nearest neighbor.**

Answer: D. K-nearest neighbor.

Q2. In K-nearest neighbor algorithm K stands for \_\_\_\_\_\_\_\_.

**A. number of neighbors that are investigated.**

B. number of iterations.

C. number of total records.

D. random number.

Answer: A. number of neighbors that are investigated.

Q3. The complexity of data mining algorithm is represented by \_\_\_\_\_\_\_\_.

A. log n.

B. 2n log n.

**C. n log n.**

D. 2 log n

Answer: C. n log n.

Q4. Data mining algorithms require \_\_\_\_\_\_\_\_\_\_\_

A. efficient sampling method.

B. storage of intermediate results.

C. capacity to handle large amounts of data.

**D. All of the above.**

Answer: D. All of the above.

Q5. In K- nearest neighbor the input is translated to \_\_\_\_\_\_\_\_\_\_.

A. values

**B. points in multidimensional space**

C. strings of characters

D. nodes

Answer: B. points in multidimensional space

Q6. Association rules are always defined on\_\_\_\_\_\_\_\_.

**A. binary attribute.**

B. single attribute.

C. relational database.

D. multidimensional attribute

Answer: A. binary attribute.

Q7. The input unit of perceptron is called as\_\_\_\_\_\_\_.

A. associators.

B. responders.

C. neuron.

**D. photo receptors.**

Answer: D. photo receptors.

Q8. The intermediate layers in a back-propagation network consists of\_\_\_\_\_\_\_\_\_\_.

A. photo receptors.

B. responders.

**C. hidden nodes.**

D. associators.

Answer: C. hidden nodes.

Q9. The left hand side of an association rule is called\_\_\_\_\_\_\_\_.

A. Consequent

B. Onset

**C. Antecedent**

D. Precedent

Answer: C. Antecedent

Q10. is the most well-known association rule algorithm and is used in most commercial products

**A. Apriori algorithm**

B. Partition algorithm

C. Distributed algorithm

D. Pincer-search algorithm

Answer: A. Apriori algorithm

Q11. The basic idea of the Apriori algorithm is to generate\_\_\_\_\_item sets of a particular size & scans the database

**A. Candidate**

B. Primary

C. Secondary

D. Superkey

Answer: A. Candidate

Q12. Rule based classification algorithms generate\_\_\_\_\_\_\_\_\_rule to perform the classification

**A. If-then**

B. While

C. Do while

D. Switch

Answer: A. If-then

Q13. If a set is a frequent set and no superset of this set is a frequent set, then it is called \_\_\_\_\_\_\_\_.

**A. maximal frequent set.**

B. border set.

C. lattice.

D. infrequent sets

Answer: A. maximal frequent set.

Q14. A priori algorithm is otherwise called as \_\_\_\_\_\_\_\_\_\_.

A. width-wise algorithm.

**B. level-wise algorithm.**

C. pincer-search algorithm.

D. fp growth algorithm

Answer: B. level-wise algorithm.

Q15**.** The FP-growth algorithm has \_\_\_\_\_\_\_\_ phases.

A. one.

**B. two.**

C. three.

D. four.

Answer: B. two.

Q16. A frequent pattern tree is a tree structure consisting of \_\_\_\_\_\_\_\_.

A. an item-prefix-tree.

B. a frequent-item-header table.

C. a frequent-item-node.

**D. both a & b.**

Answer: D. both a & b.

Q17. SVM is classified into how many types?

A. One

**B. Two**

C. Three

D. Four

Answer**:** B. Two

Q18. SVM is a supervised Machine Learning can be used for

Options :

A. Regression

B. Classification

**C. Either a or b**

D. None of These

Answer: C. Either a or b

Q19. Linear separator, Hyper plane

Options :

A. f(x)=sign(w/x+b)

B. f(x)=sign(w+x+b)

**C. f{x)=sign(w.x+b)**

D. f(x)=sign(w-x+b)

Answer: C. f{x)=sign(w.x+b)

Q20. Closest Point to the hyper plane are support vectors

**A. True**

B. False

C. Unpredictable

D. None of these

Answer: A. True

………………………………………………………………………………………………

**Unit IV**

Q1. Hierarchical clustering should be primarily used for exploration.

**A. True**

B. False

Answer: A. True

Q2. Which one of the clustering technique needs the merging approach? A. Partitioned

B. Naïve Bayes

**C. Hierarchical**

D. Both A and C

Answer: C. Hierarchical

Q3. Which of the following forms of data mining assigns records to one of a predefined set of classes?

A. Classification

**B. Clustering**

C. Both A and B

D. None

Answer: B. Clustering

Q4. Cluster is

**A. Group of similar objects that differ significantly from other objects**

B. Operations on a database to transform or simplify data in order to prepare it for a machine-learning algorithm

C. Symbolic representation of facts or ideas from which information can potentially be extracted

D. None of these

Answer: A. Group of similar objects that differ significantly from other objects

Q5. Which Of the following sampling methods, which is a probability method? A. Judgement

B. Quota

**C. Simple random**

D. Convenience

Answer: C. Simple random

Q6. \_\_\_\_\_\_\_\_analysis divides data into groups that are meaningful, useful, or both. **A. Cluster.**

B. Association.

C. Classification.

D. Relation.

Answer: A. Cluster.

Q7. Which of the following is the not a types of clustering?

A. K-means.

B. Hierarchical.

C. Partitional.

**D. Splitting**

Answer: D. Splitting

Q8. A sequence of patterns that occur frequently is known as?

A. Frequent Item Set

**B. Frequent Subsequence**

C. Frequent Sub Structure

D. All of the above

Answer: B. Frequent Subsequence

Q9. Which of the following statements is incorrect about the hierarchal clustering?

**A. The hierarchal type of clustering is also known as the HCA**

B. The choice of an appropriate metric can influence the shape of the cluster

C. In general, the splits and merges both are determined in a greedy manner D. All of the above

Answer: A. The hierarchal type of clustering is also known as the HCA

Q10. Which one of the clustering technique needs the merging approach?

A. Partitioned

B. Naïve Bayes

**C. Hierarchical**

D. Both A and C

Answer: C. Hierarchical

Q11. Identify the method for incremental conceptual clustering. A. STING

**B. COBWEB**

C.COBRA

D.OLAP

Answer: B. COBWEB

**Q12.** The goal of \_\_\_\_\_ is to discover both the dense and sparse regions of a data set.

A. association rule.

B. classification.

**C. clustering.**

D. genetic algorithm.

Answer: C. clustering.

**Q13.** Which of the following is a clustering algorithm?

A. a priori.

**B. clara.**

C. pincer-search.

D. fp-growth.

Answer: B. clara.

**Q14.** In \_\_\_\_\_\_\_\_ algorithm each cluster is represented by the center of gravity of the cluster.

A. k-medoid.

**B. k-means.**

C. stirr.

D. rock.

Answer: B. k-means.

**Q15.** In \_\_\_\_\_\_\_\_\_\_\_ each cluster is represented by one of the objects of the cluster located near the center.

**A. k-medoid.**

B. k-means.

C. stirr.

D. rock.

Answer: A. k-medoid.

Q16. K-means clustering consists of a number of iterations and not deterministic.

**A. True**

B. False

Answer: A. True

Q17. Which is conclusively produced by Hierarchical Clustering?

A. final estimation of cluster centroids

**B. tree showing how nearby things are to each other**

C. assignment of each point to clusters

D. all of these

Answer: B. tree showing how nearby things are to each other

Q18. Which clustering technique requires a merging approach?

A. Partitional

**B. Hierarchical**

C. Naive Bayes

D. None of the mentioned

Answer: B. Hierarchical

Q19.Find odd man out

**A. DBSCAN**

B. K-mean

C. PAM

D. None of these

Answer: A. DBSCAN

Q20.Chameleon is \_\_\_\_\_\_\_\_\_\_

A. Density based cc

B. Partitioning based Algorithm

C. Model based Algorithm

**D. Hierarchical clustering Algorithm**

Answer: D. Hierarchical clustering Algorithm

………………………………………………………………………………………………