SHRI GNANAMBICA DEGREE COLLEGE, MADANAPALLE

(AUTONOMOUS)

PROGRAMME: B. Sc. (Data Science)

II YEAR-III-SEM

COURSE: Data Mining Techniques Using R

QUESTION BANK

UNIT I

Long Answer Questions

1. Explain the concept of a Data Warehouse?
2. Discuss in detail the stages of KDD. Differentiate between KDD (Knowledge Discovery in Databases) and Data Mining.
3. Explain the task primitives of data mining with suitable examples.
4. Discuss various data mining technique.
5. Explain the concept of data mining knowledge representation. Why is it important?

Short Answer Questions

1. What is a Data Warehouse?
2. Define Data Mining.
3. What is KDD?
4. Differentiate KDD and Data Mining.
5. List the stages of the data mining process.
6. What are task primitives?
7. Define knowledge representation in data mining.
8. Write any two techniques in data mining task.
9. What is OLAP?
10. What is data warehouse schema?

UNIT II

Long Answer Questions

1. Explain data mining query languages with suitable examples.
2. Discuss in detail the integration of data mining systems with data warehouses.
3. Discuss the different methods of data cleaning with examples.
4. Explain data transformation and its role in data preprocessing.
5. Write a detailed note on feature selection methods.

Short Answer Questions

1. What is DMQL?
2. Define integration of data mining with a data warehouse.
3. Mention issue in integrating data mining systems.
4. What is data preprocessing?
5. Define dimensionality reduction.
6. What is PCA (Principal Component Analysis)?
7. Mention advantages of dimensionality reduction.
8. Write issues in data preprocessing.
9. Differentiate dimensionality reduction and feature selection
10. Define redundancy and curse of dimensionality in data preprocessing.

UNIT III

Long Answer Questions

1. Explain the concept of “Concept Description” in data mining.
2. What is Attribute-Oriented Induction (AOI)? Explain with an example.
3. Discuss the implementation of AOI for data characterization and explain efficient implementation methods of AOI?
4. Discuss in detail the Apriori algorithm for frequent itemset mining.
5. How are association rules generated from frequent itemsets? Explain with an example.

Short Answer Questions

1. What is concept description?
2. Define data characterization.
3. Define data comparison.
4. Explain how Apriori efficiency can be improved.
5. Differentiate frequent patterns, associations, and correlations.
6. Write a short note on the Pattern Growth approach for frequent itemset mining.
7. Write the main idea of Apriori.
8. Define frequent itemset and FP-growth.
9. What is support, confidence and lift in association rule mining?
10. Explain steps in FP-growth?

UNIT IV

Long Answer Questions

1. Explain the basic concepts of classification in data mining and discuss the steps involved in building a classification model.
2. Explain Decision Tree Induction with an example.
3. Write a detailed note on attribute selection measures used in decision trees
4. Explain tree pruning and its importance in classification.
5. Discuss Bayes classification methods in detail and Naïve Bayes classification with an example.

Short Answer Questions

1. What is classification? And explain role of classifier in classification?
2. What is decision tree induction?
3. Define attribute selection.
4. Define pruning in decision trees.
5. Define overfitting and underfitting?
6. Define Bayes theorem and What is Naïve Bayes

7. Explain components of classification model.
8. Explain accuracy, precision, recall and confusion matrix?
9. Write applications of classification.
10. Compare Decision Tree and Naïve Bayes classifiers.

UNIT V

Long Answer Questions

1. Explain association rule mining. Define antecedent and consequent in association rules with examples.
2. Discuss multi-relational association rules in detail.
3. Explain the ECLAT algorithm with an example.
4. Write a detailed case study on Market Basket Analysis.
5. Explain partitioning methods for clustering with examples.

Short Answer Questions

1. What is an association rule?
2. Define antecedent in association rules.
3. Define consequent in association rules.
4. What is ECLAT?
5. Describe hierarchical clustering methods with advantages and disadvantages.
6. Discuss about clustering method.
7. Discuss DBSCAN as a density-based clustering method.
8. Compare partitioning, hierarchical, and density-based clustering methods.
9. Define noise in clustering.
10. Mention real-life applications of clustering.