c. Discovering natural groupings or clusters within the data

d. Analysing relationships between variables

6. **Which data mining technique is used for predicting a target variable based on input features?**
   a. Clustering
   b. Classification
   c. Association rule mining
   d. Regression

7. **What is the measure of the accuracy of a classification model in data mining?**
   a. Support
   b. Confidence
   c. Lift
   d. Accuracy

8. **What does the term 'overfitting' mean in the context of data mining models?**
   a. The model is too simple and lacks complexity
   b. The model performs well on the training data but poorly on new, unseen data
   c. The model is not able to learn from the data
   d. The model is too flexible and fits the noise in the data

9. **In which phase of the data mining process are outliers and anomalies detected?**
   a. Data preprocessing
   b. Data cleaning
   c. Data integration
   d. Data exploration

10. **What is the primary purpose of the 'lift' metric in association rule mining?**
    a. Measures the degree of confidence in the rules
    b. Identifies the most frequent item sets
    c. Evaluates the performance of the model compared to a random model
    d. Measures the level of support for a rule

11. **What is the primary purpose of feature selection in data mining?**
    a. To reduce dimensionality and improve model performance
    b. To increase the complexity of the model
    c. To overfit the training data
    d. To improve data preprocessing efficiency

12. **Which algorithm is commonly used for anomaly detection in data mining?**
    a. K-Means
    b. Apriori
    c. Isolation Forest
    d. Decision Trees

Unit-III

1. **What is the primary goal of data mining in the context of knowledge discovery?**
   a. Data storage
   b. Data integration
   c. Pattern discovery and knowledge extraction
   d. Data visualisation

2. **Which phase of the data mining process involves selecting and transforming the data for analysis?**
   a. Data cleaning
   b. Data integration
   c. Data selection
   d. Data preprocessing

3. **What is the term for the process of finding hidden patterns, trends or relationships in large datasets?**
   a. Data cleansing
   b. Data exploration
   c. Data mining
   d. Data modeling

4. **Which algorithm is commonly used for association rule mining in data mining?**
   a. K-Means
   b. Apriori
   c. Decision Trees
   d. Support Vector Machines (SVM)

5. **What is the primary objective of clustering in data mining?**
   a. Predicting future outcomes
   b. Classifying data into predefined categories

13. In data mining, what does the term 'ensemble learning' refer to?
   a. Learning from labelled training data
   b. Combining predictions from multiple models to improve accuracy
   c. Extracting patterns from structured data
   d. Clustering data into natural groups

14. What is the primary goal of time series analysis in data mining?
   a. Identifying patterns and relationships between variables
   b. Detecting outliers and anomalies
   c. Analysing data over time to make predictions
   d. Clustering data into temporal groups

15. Which technique is used in data mining for imputing missing values in a dataset?
   a. Principal Component Analysis (PCA)
   b. Singular Value Decomposition (SVD)
   c. k-Nearest Neighbours (k-NN)
   d. Association Rule Mining

16. What is the role of the lift chart in evaluating a predictive model's performance in data mining?
   a. Measures the degree of confidence in the rules
   b. Identifies the most frequent item sets
   c. Evaluates the performance compared to a random model
   d. Measures the level of support for a rule

17. Which algorithm is commonly used for text mining and natural language processing in data mining?
   a. Decision Trees
   b. Naive Bayes
   c. Support Vector Machines (SVM)
   d. Apriori

18. What is the purpose of dimensionality reduction techniques in data mining?
   a. To increase the complexity of the model
   b. To identify frequent itemsets in the data
   c. To reduce the number of features while preserving important information
   d. To visualise high-dimensional data

19. Which data mining technique is suitable for identifying patterns and relationships in a large dataset with high dimensionality?
   a. Association Rule Mining
   b. Clustering

   c. Principal Component Analysis (PCA)
   d. Classification

20. What is the purpose of cross-validation in the evaluation of a predictive model in data mining?
   a. To train the model on a large dataset
   b. To test the model on the training data
   c. To assess the model's performance on new, unseen data
   d. To increase the complexity of the model

21. What is the primary goal of a Knowledge Management System (KMS)?
   a. Data storage and retrieval
   b. Information sharing and collaboration
   c. Automation of routine tasks
   d. Real-time transaction processing

22. What is data mining?
   a. Extracting minerals from the earth
   b. Extracting useful patterns and information from large datasets
   c. Data encryption technique
   d. Database optimisation process

23. Which phase of the CRISP-DM model involves selecting the appropriate data for analysis?
   a. Data understanding
   b. Data preparation
   c. Data mining
   d. Data evaluation

24. Association rule mining is commonly used for:
   a. Classification
   b. Clustering
   c. Predictive modeling
   d. Discovering relationships in data

25. Which algorithm is commonly used for clustering in data mining?
   a. Apriori
   b. K-means
   c. Decision Trees
   d. Support Vector Machines

26. What is the goal of outlier detection in data mining?
   a. Identify unusual patterns that do not conform to expected behaviour
   b. Maximise the accuracy of classification models
   c. Optimise database performance
   d. Enhance data visualisation techniques

27. **Which of the following is an example of a classification algorithm?**
    a. K-means
    b. Apriori
    c. Decision Trees
    d. Hierarchical Clustering

28. **What is the purpose of the 'apriori' algorithm in data mining?**
    a. Association rule mining
    b. Clustering
    c. Classification
    d. Regression analysis

29. **Which data mining task involves assigning predefined classes to instances?**
    a. Clustering
    b. Regression
    c. Classification
    d. Association

30. **Which of the following is not a supervised learning algorithm?**
    a. Support Vector Machines
    b. K-means
    c. Decision Trees
    d. Neural Networks

31. **What is the primary goal of dimensionality reduction techniques in data mining?**
    a. Increase computational complexity
    b. Improve model interpretability
    c. Reduce the number of features in the dataset
    d. Enhance data visualisation

32. **Which of the following is an advantage of using decision trees in classification?**
    a. Prone to overfitting
    b. Limited interpretability
    c. Require normalisation of data
    d. Easy to understand and interpret

33. **What is the purpose of cross-validation in machine learning?**
    a. Optimise hyperparameters
    b. Assess model performance on new data
    c. Increase model complexity
    d. Reduce training time

34. **Which of the following is a common technique for handling missing data in a dataset?**
    a. Imputation
    b. Outlier detection
    c. Principal component analysis
    d. Support vector machines

35. **What is the difference between classification and regression in data mining?**
    a. Classification deals with categorical outcomes, while regression deals with numerical outcomes
    b. Regression deals with categorical outcomes, while classification deals with numerical outcomes
    c. Classification and regression are synonymous terms
    d. Both involve clustering techniques

36. **Which of the following is a data preprocessing technique to scale features in a similar range?**
    a. Normalisation
    b. Imputation
    c. Standardisation
    d. Encoding

37. **What is the main purpose of data warehousing in the context of data mining?**
    a. Real-time data analysis
    b. Storing and managing large volumes of data for analysis
    c. Data encryption
    d. Indexing database tables

38. **Which of the following is a technique for reducing the curse of dimensionality in data mining?**
    a. Feature engineering
    b. Principal Component Analysis (PCA)
    c. Cross-validation
    d. Association rule mining

39. **What is the role of the 'lift' metric in association rule mining?**
    a. Measures the predictive accuracy of a model
    b. Evaluates the performance of a clustering algorithm
    c. Indicates the ratio of the observed support to the expected support
    d. Measures the performance improvement of a marketing strategy compared to random chance

40. **What is the primary objective of the K-Nearest Neighbours (KNN) algorithm?**
    a. Association rule mining
    b. Clustering
    c. Classification
    d. Regression analysis

41. In the context of data mining, what is the purpose of the term 'overfitting'?
    a. Model fits the training data but fails to generalise to new data
    b. Model is too simple and lacks complexity
    c. Model has high bias and low variance
    d. Model perfectly fits both training and testing data

42. Which of the following is a benefit of ensemble learning techniques in data mining?
    a. Increased risk of overfitting
    b. Reduced computational efficiency
    c. Improved predictive performance
    d. Limited diversity in model predictions

43. What is the role of the 'confusion matrix' in evaluating classification models?
    a. Measures the spread of data points in a cluster
    b. Evaluates the performance of regression models
    c. Summarises the performance of a classification model
    d. Measures the correlation between features

44. Which of the following is a characteristic of a good Data Mining Model?
    a. High bias and low variance
    b. Complex and difficult to interpret
    c. Overfitting to the training data
    d. Generalises well to new, unseen data

45. What does the term 'apex data' refer to in the context of data mining?
    a. The most significant data points in a dataset
    b. Data with the highest level of noise
    c. Data stored in a data warehouse
    d. The first data point in a time series