1. Which function is used to split the dataset into training and testing sets in scikit-learn?
   **a) train_test_split()**
   b) split_data()
   c) data_split()
   d) train_split()
2. Which library in Python is primarily used for data manipulation and analysis, often used before applying machine learning algorithms?
   a) NumPy
   b) Matplotlib
   c) Pandas
   d) Seaborn
3. What method would you use to fit a linear regression model in scikit-learn?
   **a) fit()**
   b) train()
   c) model()
   d) learn()
4. Which function in scikit-learn can be used for performing K-Means clustering?
   **a) KMeans()**
   b) KMeansClustering()
   c) cluster_KMeans()
   d) k_means()
5. Which evaluation metric is commonly used for classification problems?
   a) Mean Squared Error
   **b) Accuracy**
   c) Euclidean Distance
   d) R-squared
6. What method in scikit-learn is used to predict the class labels for given data points?
   **a) predict()**
   b) classify()
   c) forecast()
   d) label()
7. Which algorithm is an example of a supervised learning algorithm?
   a) K-Means
   **b) Linear Regression**
   c) PCA
   d) t-SN
8. Which Python library is used to implement Support Vector Machines?
   a) scipy
   **b) sklearn**
   c) statsmodels
   d) tensorflow
9. Which method in scikit-learn is used to standardize features by removing the mean and scaling to unit variance?
   a) normalize()
   b) scale()
   c) standardize()
   **d) StandardScaler()**
10. Which of the following is not an unsupervised learning algorithm?
    a) K-Means
    b) DBSCAN
    **c) Decision Tree**
    d) Hierarchical Clustering
11. Which of the following functions in scikit-learn is used for performing Principal Component Analysis (PC?
    **a) PCA()**
    b) PrincipalComponent()
    c) ComponentAnalysis()
    d) PCA_Analysis()
12. Which algorithm is typically used for outlier detection in unsupervised learning?
    a) Linear Regression
    **b) DBSCAN**
    c) Logistic Regression
    d) K-Nearest Neighbors
13. Which library provides a comprehensive set of tools for machine learning in Python, including both supervised and unsupervised algorithms?
    a) NumPy
    **b) scikit-learn**
    c) Matplotlib
    d) Pandas

14. In scikit-learn, which method is used to find the optimal number of clusters in K-Means?
    a) fit()
    **b) elbow_method()**
    c) optimal_clusters()
    d) find_clusters()
15. Which technique is used to reduce the dimensionality of data while preserving as much variance as possible?
    a) Linear Regression
    **b) PCA**
    c) SVM
    d) Decision Tree
16. What function is used to create a confusion matrix in scikit-learn?
    **a) confusion_matrix()**
    b) matrix_confusion()
    c) confusion()
    d) create_matrix()
17. Which method is used in scikit-learn to perform feature selection?
    a) feature_selection()
    **b) SelectKBest()**
    c) choose_features()
    d) select_features()
18. Which parameter of the KMeans class in scikit-learn specifies the number of clusters?
    a) clusters
    **b) n_clusters**
    c) num_clusters
    d) k_clusters
19. Which type of machine learning algorithm groups data into clusters based on similarity?
    a) Supervised
    **b) Unsupervised**
    c) Reinforcement
    d) Semi-supervised
20. Which method is used to encode categorical variables in scikit-learn?
    a) LabelEncoder
    b) CategoryEncoder
    **c) OneHotEncoder**
    d) CategoryTransformer
21. Which scikit-learn module includes implementation for various machine learning algorithms?
    a) sklearn.metrics
    b) sklearn.datasets
    **c) sklearn.model_selection**
    d) sklearn.linear_model
22. Which algorithm is used for unsupervised learning?
    a) Decision Tree
    b) Random Forest
    **c) K-Means**
    d) SVM
23. Which function in scikit-learn is used to load the Iris dataset?
    **a) load_iris**
    b) get_iris
    c) iris_data
    d) fetch_iris
24. What is machine learning?
    a) A type of artificial intelligence that mimics human intelligence.
    **b) A field of study that gives computers the abilityto learn without being explicitly programmed.**
    c) The process of automating repetitive tasks using algorithms.
    d) An approach to problem-solving that relies on human input for decision-making.
25. Which of the following is an example of supervised learning?
    a) Predicting customer churn based on historical data.
    b) Grouping news articles into topics.
    c) Anomaly detection in network traffic.
    **d) Identifying spam emails.**
26. What is the primary goal of unsupervised learning?
    a) Predicting outcomes based on input data.
    **b) Finding patterns or structures in data.**
    c) Maximizing rewards through trial and error.
    d) Learning from feedback to improve performance.
27. Which of the following evaluation metrics is commonly used for regression problems?
    a) Accuracy
    b) Precision
    **c) Mean Squared Error (MSE)**
    d) F1 Score

28. What is the purpose of the validation set in machine learning?
    a) To train the model.
    b) To test the model's performance on unseen data.
    **c) To fine-tune hyperparameters.**
    d) To evaluate the model's performance during training.
29. Which of the following is NOT a step in the machine learning pipeline?
    a) Data preprocessing
    **b) Feature selection**
    c) Model evaluation
    d) Model training
30. What is overfitting in the context of machine learning?
    **a) When a model performs well on training databut poorly on unseen data.**
    b) When a model is too simple to capture the underlying patterns in the data.
    c) When a model generalizes well to unseen data.
    d) When a model's complexity matches the complexity of the data.
31. Which of the following techniques is used to address overfitting in machine learning?
    a) Feature scaling
    **b) Regularization**
    c) Dimensionality reduction
    d) Data augmentation
32. Which of the following is an ensemble learning technique?
    a) Linear Regression
    **b) Decision Trees**
    c) K-Means Clustering
    d) Logistic Regression
33. What does the term "bias-variance tradeoff" refer to in machine learning?
    a) Balancing computational resources with model accuracy.
    b) Balancing the complexity of a model with its ability to generalize to new data.
    c) Balancing the number of features in a model with its computational efficiency.
    **d) Balancing model bias with model variance to optimize performance.**
34. Which of the following is a hyperparameter for the K-Nearest Neighbors algorithm?
    a) Number of clusters (K)
    b) Learning rate
    **c) Number of neighbors (K)**
    d) Regularization parameter
35. What is the main difference between classification and regression problems in machine learning?
    **a) Classification predicts discrete categories, whileregression predicts continuous values.**
    b) Classification uses labeled data, while regression uses unlabeled data.
    c) Classification involves clustering data points, while regression involves fitting a curve to data.
    d) Classification aims to minimize error, while regression aims to maximize accuracy.