

MCQ On Predictive Modelling

1. Predictive modeling is a technique used to:

- a) Understand historical data
- b) Make projections about future outcomes
- c) Analyze data visually
- d) Create descriptive statistics

Correct Option: b) Make projections about future outcomes

2. Which of the following is NOT a common application of predictive modeling?

- a) Customer churn prediction
- b) Stock market analysis
- c) Data visualization
- d) Credit risk assessment

Correct Option: c) Data visualization

3. The primary goal of predictive modeling is to:

- a) Describe past events
- b) Forecast future events
- c) Summarize data
- d) Perform data cleaning

Correct Option: b) Forecast future events

4. Which algorithm is commonly used for binary classification tasks in predictive modeling?

- a) K-means
- b) Decision tree
- c) Apriori
- d) Principal Component Analysis (PCA)

Correct Option: b) Decision tree

5. In predictive modeling, the target variable is also known as the:

- a) Independent variable
- b) Dependent variable
- c) Control variable
- d) Confounding variable

Correct Option: b) Dependent variable

6. What is the main purpose of feature engineering in predictive modeling?

- a) To remove outliers from the dataset
- b) To select the best algorithm for modeling
- c) To transform raw data into informative features
- d) To perform data imputation

Correct Option: c) To transform raw data into informative features

7. Which evaluation metric is commonly used for regression tasks in predictive modeling?

- a) Accuracy
- b) F1 score

c) Root Mean Squared Error (RMSE)

d) Precision

Correct Option: c) Root Mean Squared Error (RMSE)

8. Ensemble methods in predictive modeling involve:

a) Using multiple models to make a single prediction

b) Combining categorical and numerical data

c) Using only one algorithm for all tasks

d) Eliminating outliers from the dataset

Correct Option: a) Using multiple models to make a single prediction

9. Overfitting in predictive modeling occurs when:

a) The model performs well on the training data but poorly on the test data

b) The model performs well on both training and test data

c) The model underestimates the target variable

d) The model has too few features

Correct Option: a) The model performs well on the training data but poorly on the test data

10. In predictive modeling, what is cross-validation used for?

a) Evaluating the model's performance on new data

b) Feature selection

c) Data visualization

d) Handling missing values

Correct Option: a) Evaluating the model's performance on new data

11. Which of the following algorithms is a type of decision tree-based ensemble method?

a) Random Forest

b) Naive Bayes

c) k-Nearest Neighbors (k-NN)

d) Linear Regression

Correct Option: a) Random Forest

12. Which technique is used to deal with imbalanced datasets in predictive modeling?

a) Data imputation

b) Principal Component Analysis (PCA)

c) Oversampling

d) Feature scaling

Correct Option: c) Oversampling

13. In predictive modeling, what is the purpose of the training dataset?

a) To validate the model's performance

b) To train the model and adjust its parameters

c) To test the model's generalization

d) To evaluate the model's accuracy

Correct Option: b) To train the model and adjust its parameters

14. Which of the following is NOT an example of a classification problem in predictive modeling?

- a) Spam email detection
- b) House price prediction
- c) Disease diagnosis
- d) Customer segmentation

Correct Option: b) House price prediction

15. Which of the following algorithms is unsupervised and used for clustering in predictive modeling?

- a) Linear Regression
- b) k-Nearest Neighbors (k-NN)
- c) k-Means
- d) Random Forest

Correct Option: c) k-Means

16. In predictive modeling, the process of splitting the data into training and testing sets is known as:

- a) Cross-validation
- b) Feature engineering
- c) Data imputation
- d) Train-test split

Correct Option: d) Train-test split

17. Which statistical technique is used to reduce the dimensionality of data in predictive modeling?

- a) Chi-squared test
- b) Principal Component Analysis (PCA)
- c) t-test
- d) Analysis of Variance (ANOVA)

Correct Option: b) Principal Component Analysis (PCA)

18. In predictive modeling, what does the term "outlier" refer to?

- a) A target variable with extremely high values
- b) A data point that deviates significantly from other data points
- c) A categorical variable with only two unique values
- d) A feature that has missing data

Correct Option: b) A data point that deviates significantly from other data points

19. Which method is used to handle missing data in predictive modeling?

- a) Removing the entire row with missing data
- b) Ignoring the missing values during model training
- c) Imputing the missing data with estimated values
- d) Replacing missing data with random values

Correct Option: c) Imputing the missing data with estimated values

20. The ROC curve in predictive modeling is used to visualize the trade-off between:

- a) Sensitivity and Specificity
- b) Precision and Recall
- c) True Positive Rate and False Positive Rate
- d) Accuracy and F1 score

Correct Option: a) Sensitivity and Specificity

21. In predictive modeling, what does the term "feature selection" mean?

- a) Selecting the best machine learning algorithm
- b) Choosing the most important features for modeling
- c) Removing outliers from the dataset
- d) Transforming raw data into informative features

Correct Option: b) Choosing the most important features for modeling

22. Which of the following algorithms is commonly used for time series forecasting in predictive modeling?

- a) Decision tree
- b) Support Vector Machine (SVM)
- c) Autoregressive Integrated Moving Average (ARIMA)
- d) Naive Bayes

Correct Option: c) Autoregressive Integrated Moving Average (ARIMA)

23. Which technique is used to prevent data leakage in predictive modeling?

- a) Data imputation
- b) Feature scaling
- c) Cross-validation
- d) Model training

Correct Option: c) Cross-validation

24. In predictive modeling, what is the purpose of the validation dataset?

- a) To train the model
- b) To tune hyperparameters
- c) To evaluate the model's performance on unseen data
- d) To visualize the data distribution

Correct Option: b) To tune hyperparameters

25. In predictive modeling, the process of

transforming categorical variables into numerical ones is called:

- a) Data normalization
- b) Data visualization
- c) One-hot encoding
- d) Data imputation

Correct Option: c) One-hot encoding

26. Which of the following algorithms is a type of support vector machine used in predictive modeling?

- a) Decision tree
- b) k-Nearest Neighbors (k-NN)
- c) Linear Regression
- d) Support Vector Regression (SVR)

Correct Option: d) Support Vector Regression (SVR)

27. What is the purpose of a confusion matrix in predictive modeling?

- a) To visualize data distribution
- b) To handle missing data
- c) To evaluate the model's performance on binary classification
- d) To perform data imputation

Correct Option: c) To evaluate the model's performance on binary classification

28. Which technique can be used to address multicollinearity in predictive modeling?

- a) Principal Component Analysis (PCA)
- b) Decision tree
- c) Data normalization
- d) Data imputation

Correct Option: a) Principal Component Analysis (PCA)

29. Which algorithm is used for collaborative filtering in recommendation systems?

- a) Decision tree
- b) k-Means
- c) Singular Value Decomposition (SVD)
- d) Naive Bayes

Correct Option: c) Singular Value Decomposition (SVD)

30. Which of the following techniques is used for feature scaling in predictive modeling?

- a) Min-Max scaling
- b) Decision tree
- c) One-hot encoding
- d) K-nearest neighbor

Correct Option: a) Min-Max scaling

31. In predictive modeling, what is the purpose of the test dataset?

- a) To train the model
- b) To evaluate the model's performance on unseen data
- c) To visualize the data distribution
- d) To tune hyperparameters

Correct Option: b) To evaluate the model's performance on unseen data

32. Which of the following evaluation metrics is used for binary classification tasks in predictive modeling?

- a) Mean Absolute Error (MAE)
- b) Root Mean Squared Error (RMSE)
- c) Area Under the Curve (AUC-ROC)
- d) R-squared (R2)

Correct Option: c) Area Under the Curve (AUC-ROC)

33. In predictive modeling, what is the "curse of dimensionality" referring to?

- a) The difficulty of visualizing high-dimensional data
- b) The presence of outliers in the dataset
- c) The impact of multicollinearity on the model's performance
- d) The increase in computational complexity as the number of features grows

Correct Option: a) The difficulty of visualizing high-dimensional data

34. Which algorithm is used for text classification tasks in predictive modeling?

- a) Decision tree
- b) Support Vector Machine (SVM)
- c) k-Means
- d) Naive Bayes

Correct Option: d) Naive Bayes

35. In predictive modeling, which technique is used to reduce overfitting in decision trees?

- a) Bagging
- b) Data normalization
- c) Principal Component Analysis (PCA)
- d) Feature scaling

Correct Option: a) Bagging

36. Which of the following is NOT a step in the predictive modeling process?

- a) Data visualization
- b) Data preprocessing
- c) Data imputation
- d) Data encryption

Correct Option: d) Data encryption

37. What is the primary purpose of hyperparameter tuning in predictive modeling?

- a) To make the model faster
- b) To improve the model's performance on the training data
- c) To reduce the number of features
- d) To handle missing data

Correct Option: b) To improve the model's performance on the training data

38. In predictive modeling, which of the following is an example of an imbalanced dataset?

- a) A dataset with an equal number of positive and negative samples
- b) A dataset with a significantly larger number of positive samples than negative samples
- c) A dataset with a large number of missing values
- d) A dataset with a high number of features

Correct Option: b) A dataset with a significantly larger number of positive samples than negative samples

39. Which of the following algorithms is NOT used for regression tasks in predictive modeling?

- a) Decision tree
- b) Random Forest
- c) k-Means
- d) Support Vector Regression (SVR)

Correct Option: c) k-Means

40. In predictive modeling, what does the term "precision" refer to?

- a) The percentage of true positive predictions
- b) The percentage of true negative predictions
- c) The percentage of correctly predicted positive cases
- d) The percentage of correctly predicted negative cases

Correct Option: c) The percentage of correctly predicted positive cases

41. Which of the following techniques is used for dimensionality reduction in predictive modeling?

- a) One-hot encoding
- b) Decision tree
- c) Singular Value Decomposition (SVD)
- d) k-Means

Correct Option: c) Singular Value Decomposition (SVD)

42. Which algorithm is used for time series forecasting in predictive modeling, particularly when dealing with seasonality?

- a) Random Forest
- b) Autoregressive Integrated Moving Average (ARIMA)
- c) Naive Bayes
- d) Decision tree

Correct Option: b) Autoregressive Integrated Moving Average (ARIMA)

43. In predictive modeling, which evaluation metric is used to assess the performance of a multi-class classification problem?

- a) Precision
- b) F1 score
- c) Area Under the Curve (AUC-ROC)
- d) Mean Absolute Error (MAE)

Correct Option: b) F1 score

44. Which of the following is a supervised learning algorithm used in predictive modeling?

- a) k-Means
- b) Principal Component Analysis (PCA)
- c) Decision tree
- d) Apriori

Correct Option: c) Decision tree

45. In predictive modeling, what is the purpose of a learning curve?

- a) To visualize data distribution
- b) To measure the model's accuracy
- c) To evaluate the impact of different hyperparameters
- d) To analyze feature importance

Correct Option: c) To evaluate the impact of different hyperparameters

46. Which of the following algorithms is used for data clustering in predictive modeling?

- a) Linear Regression
- b) k-Means
- c) Decision tree
- d) Naive Bayes

Correct Option: b) k-Means

47. In predictive modeling, what does the term "recall" refer to?

- a) The percentage of true positive predictions
- b) The percentage of true negative predictions
- c) The percentage of correctly predicted positive cases
- d) The percentage of correctly predicted negative cases

Correct Option: a) The percentage of true positive predictions

48. Which method can be used to address class imbalance in predictive modeling?

- a) Data normalization
- b) Oversampling
- c) Feature scaling
- d)

Data imputation

Correct Option: b) Oversampling

49. In predictive modeling, which technique is used to avoid overfitting in neural networks?

- a) Gradient boosting
- b) Feature engineering
- c) Regularization
- d) k-fold cross-validation

Correct Option: c) Regularization

50. Which of the following algorithms is used for clustering and density estimation in predictive modeling?

- a) Support Vector Machine (SVM)
- b) Principal Component Analysis (PCA)
- c) k-Means
- d) Gaussian Mixture Model (GMM)

Correct Option: d) Gaussian Mixture Model (GMM)

SATHAJ GYAN