

DATA SCIENCE PROJECTS FOR BUSINESS

MCQ 1

1. What is the main difference between supervised and unsupervised learning?
 - A) Supervised learning has labeled data, unsupervised learning does not.
 - B) Unsupervised learning requires more data than supervised learning.
 - C) Supervised learning is only used for classification, unsupervised learning is only used for regression.
 - D) Unsupervised learning always gives better accuracy than supervised learning.

2. Which of the following is an example of an unsupervised learning task?
 - A) Spam email classification
 - B) Sentiment analysis on customer reviews
 - C) Customer segmentation for marketing
 - D) Predicting house prices based on past sales

3. How can you measure the success or value of a machine learning project?
 - A) By achieving high accuracy or performance on test data
 - B) By evaluating improvements in key business metrics
 - C) By minimizing model training time and computational cost
 - D) By using the most advanced machine learning techniques available

4. How can you compute the Return on Investment (ROI) of a machine learning project?
 - A) $(\text{Total Revenue Generated} - \text{Model Development Cost}) / \text{Model Development Cost}$
 - B) $(\text{Model Accuracy} - \text{Baseline Accuracy}) / \text{Total Cost}$
 - C) $(\text{Total Number of Predictions} / \text{Model Training Time}) * \text{Cost per Prediction}$
 - D) $(\text{Computational Cost} - \text{Data Collection Cost}) / \text{Model Complexity}$

5. In a company with low data maturity, which type of project is more beneficial to start with?
 - A) AI project, to leverage advanced predictive models immediately
 - B) BI project, to improve data organization, reporting, and governance
 - C) AI project, because it requires less structured data
 - D) Either AI or BI, as data maturity has no impact on project success

6. What is a common source of bias in open-source generative AI models?
- A) The model architecture inherently favors certain outputs
 - B) Bias in the training data used to develop the model
 - C) Open-source models are designed to be completely unbiased
 - D) Bias only occurs in proprietary AI models, not in open-source ones
7. When should a company internalize a data project instead of outsourcing it?
- A) When the company has the necessary in-house expertise and wants to build long-term capabilities
 - B) When external vendors offer a faster and cheaper solution
 - C) When the project requires minimal domain knowledge and customization
 - D) When the company wants to experiment with AI without committing resources
8. What is the main objective of clustering in machine learning?
- A) To group similar data points together
 - B) To predict the future behavior of data points
 - C) To reduce the dimensionality of the dataset
 - D) To increase the accuracy of classification models
9. Which of the following is not a clustering algorithm?
- A) K-Means
 - B) DBSCAN
 - C) Decision Trees
 - D) Hierarchical Clustering
10. In K-Means clustering, what does the "K" represent?
- A) The number of clusters
 - B) The number of iterations
 - C) The number of features in the dataset
 - D) The number of centroids at the beginning
11. Which of the following metrics is commonly used to evaluate clustering quality?
- A) Accuracy
 - B) Silhouette Score
 - C) Mean Squared Error (MSE)
 - D) Precision and Recall

12. What is the purpose of Principal Component Analysis (PCA)?

- A) To cluster data into meaningful groups
- B) To transform high-dimensional data into a lower-dimensional space
- C) To improve the accuracy of supervised models
- D) To increase the number of features in a dataset

13. How does PCA achieve dimensionality reduction?

- A) By selecting the most important features
- B) By creating new uncorrelated variables (principal components)
- C) By removing missing values from the dataset
- D) By clustering data into groups

14. What is a principal component in PCA?

- A) A cluster of similar data points
- B) A new axis that captures the maximum variance in the data
- C) A new class label assigned to each data point
- D) A random feature generated from existing data

15. If the first two principal components explain 95% of the variance in the dataset, what does that mean?

- A) The remaining components contain very little useful information
- B) The dataset has only two features
- C) PCA has removed all noise from the dataset
- D) The first two components are the only important ones for clustering