



### #1. What is Data Science?

☐

1. The study of computer algorithms

☐

2. The process of extracting insights from complex and unstructured data

☐

3. The analysis of data using statistical methods

☐

4. The study of computer networks

☐

5. None of the above

### #2. Which programming languages are commonly used in Data Science?

☐

1. Java, C++, Python

☐

2. R, Python, SQL

☐

3. HTML, CSS, JavaScript

☐

4. Ruby, Swift, Kotlin

☐

5. None of the above

### #3. What is the purpose of exploratory data analysis (EDA) in Data Science?

☐

1. To predict future trends

☐

2. To summarize data and gain insights

☐

3. To develop machine learning models

☐

4. To create data visualizations

☐

5. None of the above

#4. What is the main goal of data preprocessing in the context of machine learning?

☐

1. To make the data easier to understand and interpret

☐

2. To remove all the data entries with missing values

☐

3. To increase the complexity of the data

☐

4. To reduce the size of the dataset by removing columns and rows

☐

5. None of the above

#5. Which of the following is used for feature selection in machine learning?

☐

1. Principal Component Analysis (PCA)

☐

2. Linear Regression

☐

3. Random Forest

☐

4. K-Nearest Neighbors

☐

5. None of the above

#6. What does the term “overfitting” mean in the context of machine learning?

☐

1. The model fits the training data too well but performs poorly on new data

☐

2. The model has too few features and lacks complexity

☐

3. The model is trained on a very small dataset

☐

4. The model is not fitting the training data well enough

☐

5. None of the above

#7. What is a confusion matrix used for in the evaluation of classification models?

☐

1. To visualize the relationship between variables

☐

2. To summarize the performance of a classification algorithm

☐

3. To calculate correlation coefficients

☐

4. To identify outliers in the data

☐

5. None of the above

#8. What is the purpose of cross-validation in machine learning?

☐

1. To split the dataset into training and testing sets

☐

2. To evaluate a model's performance on an independent dataset

☐

3. To visualize the data using cross-shaped plots

☐

4. To transform categorical variables into numerical values

☐

5. None of the above

#9. Which algorithm is commonly used for both classification and regression tasks in machine learning?

☐

1. Support Vector Machine (SVM)

☐

2. K-Means Clustering

☐

3. Decision Tree

☐

4. Neural Network

☐

5. None of the above

#10. What is the primary purpose of regularization techniques in machine learning?

☐

1. To add more features to the model

☐

2. To reduce the complexity of the model

☐

3. To increase the accuracy of the model

☐

4. To increase the training time of the model

☐

5. None of the above

#11. What is the difference between supervised and unsupervised learning in

machine learning?

- ☐
- 1. In supervised learning, the model is trained with labeled data; in unsupervised learning, the model is trained with unlabeled data
- ☐
- 2. Supervised learning requires a human supervisor; unsupervised learning does not
- ☐
- 3. Supervised learning is used for regression tasks; unsupervised learning is used for classification tasks
- ☐
- 4. In unsupervised learning, the model predicts outcomes; in supervised learning, the model does not predict outcomes
- ☐
- 5. None of the above

#12. What does the term “feature engineering” refer to in the context of machine learning?

- ☐
- 1. Creating new features from existing data
- ☐
- 2. Transforming features into labels
- ☐
- 3. Removing features with missing values
- ☐
- 4. Engineering physical devices based on machine learning algorithms
- ☐
- 5. None of the above

#13. What is the main goal of clustering algorithms in unsupervised learning?

- ☐
- 1. To predict an output value based on input features

☐

2. To group similar data points together

☐

3. To classify data points into predefined classes

☐

4. To draw decision boundaries between classes

☐

5. None of the above

#14. What is the purpose of dimensionality reduction techniques like PCA (Principal Component Analysis) in machine learning?

☐

1. To increase the number of features in the dataset

☐

2. To reduce the number of features while retaining essential information

☐

3. To add noise to the data and increase variability

☐

4. To remove outliers from the dataset

☐

5. None of the above

#15. In data preprocessing, what is imputation used for?

☐

1. To increase the number of features in the dataset

☐

2. To remove outliers from the dataset

☐

3. To fill missing values in the dataset using various techniques

☐

4. To transform categorical data into numerical values

☐

5. None of the above

#16. What is the primary purpose of the term frequency-inverse document frequency (TF-IDF) in text mining and natural language processing?

- ☐
- 1. To calculate the frequency of words in a document
- ☐
- 2. To measure the importance of words in a document based on their frequency and rarity in the entire corpus
- ☐
- 3. To summarize the content of a document
- ☐
- 4. To translate text from one language to another
- ☐
- 5. None of the above

#17. What does the term “precision” represent in the context of classification models?

- ☐
- 1. The ability of the model to find all the relevant cases
- ☐
- 2. The ability of the model to correctly identify positive cases
- ☐
- 3. The ability of the model to avoid classifying negative cases as positive
- ☐
- 4. The ability of the model to generalize well to new, unseen data
- ☐
- 5. None of the above

#18. Which algorithm is commonly used for anomaly detection in data science?

- ☐
- 1. K-Means Clustering
- ☐

2. Decision Tree

☐

3. Isolation Forest

☐

4. Support Vector Machine (SVM)

☐

5. None of the above

#19. What is the purpose of ensemble methods in machine learning?

☐

1. To increase the accuracy of the model

☐

2. To decrease the complexity of the model

☐

3. To decrease the training time of the model

☐

4. To increase the number of features in the dataset

☐

5. None of the above

#20. Which metric is commonly used for evaluating regression models in data science?

☐

1. Accuracy

☐

2. F1-Score

☐

3. Mean Squared Error (MSE)

☐

4. Precision-Recall Curve

☐

5. None of the above



Next  
Results

