#1. 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

#2. 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

#3. 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

#4. What does the term “logistic regression” refer to in machine learning?

☐
1. A regression algorithm used for predicting continuous values
☐
2. A classification algorithm used for binary and multiclass classification
☐
3. A technique for finding outliers in a dataset
☐
4. A method for transforming categorical data into numerical values
☐
5. None of the above

#5. What is the primary purpose of the “bag-of-words” model in natural language processing?

☐
1. To represent text as numerical vectors
☐
2. To summarize the content of a document
☐
3. To identify the sentiment expressed in a piece of text
☐
4. To translate text from one language to another

☐
5. None of the above

#6. What is the primary 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

#7. What is the purpose of the “dropout” technique in neural networks?

☐
1. To reduce the learning rate
☐
2. To add noise to the input data to increase variability
☐
3. To randomly deactivate some neurons during training
☐
4. To increase the number of layers in the network
☐
5. None of the above

#8. Which technique is commonly used for text classification tasks in natural language processing?

☐
1. K-Means Clustering

☐
2. Naive Bayes Classifier
☐
3. Random Forest
☐
4. Support Vector Machine (SVM)
☐
5. None of the above

#9. What is the purpose of the “bagging” technique in ensemble learning?

☐
1. Combining multiple weak learners into a strong learner
☐
2. Training multiple models independently and combining their predictions
☐
3. Assigning weights to features based on their importance
☐
4. Adjusting hyperparameters of the model
☐
5. None of the above

#10. What is the primary goal of feature scaling in machine learning?

☐
1. To increase the complexity of the model
☐
2. To make different features comparable by bringing them to a similar scale
☐
3. To reduce the dimensionality of the dataset
☐
4. To add noise to the data and increase variability
☐
5. None of the above

#11. 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

#12. What is the primary purpose of the “SVD” (Singular Value Decomposition) technique in recommendation systems?

☐
1. To classify items into predefined categories
☐
2. To reduce the dimensionality of the user-item interaction matrix
☐
3. To calculate the frequency of items in a dataset
☐
4. To generate random numbers for recommendations
☐
5. None of the above

#13. 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

#14. 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 create multiple train-test splits and average the performance
☐
4. To transform categorical variables into numerical values
☐
5. None of the above

#15. Which metric is commonly used for evaluating classification models in data science?

☐
1. Mean Absolute Error (MAE)
☐
2. Root Mean Square Error (RMSE)
☐
3. Area Under the Receiver Operating Characteristic Curve (AUC-ROC)
☐
5. F1-Score

#16. What is the 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 “ensemble learning” refer to in machine learning?

☐
1. Using multiple algorithms to solve a problem
☐
2. Combining predictions from multiple models to make a more accurate prediction
☐
3. Training a single model on large amounts of data
☐
4. Feature selection techniques
☐
5. None of the above

#18. What is the primary purpose of the “K-Means” clustering algorithm?

☐
1. To classify data points into predefined classes
☐
2. To find patterns and relationships in data points
☐
3. To group similar data points together
☐
4. To predict continuous values
☐

5. None of the above

#19. In machine learning, what does the term “bias” refer to?

☐
1. The difference between predicted values and actual values
☐
2. The error introduced by approximating a real-world problem as a simplified model
☐
3. The features used to make predictions
☐
4. The variance in the model's predictions
☐
5. None of the above

#20. What is the primary objective of reinforcement learning in the context of artificial intelligence?

☐
1. To process and understand natural language
☐
2. To optimize decisions over time using trial and error
☐
3. To recognize patterns and images in data
☐
4. To perform calculations and computations automatically
☐
5. None of the above

Next

Results