NAASCOM DATA SCIENCE MCQS

Module 1: Artificial Intelligence & Big Data Analytics – An Introduction

Module 2: Basic Statistical Concepts

Module 3: Advanced Statistical Concepts

Module 4: Statistical Tools and Usage

Module 5: Importing Data

Module 6: Pre-processing Data

Module 7: Exploring Data

Module 8: Data Structures and Algorithms

Module 9: Graph Algorithms

Module 10: String Algorithms

Module 11: Neural Networks

Module 12: Programming for Data Science

Module 13: Applications of pre-designed Algorithms

Module 1: Artificial Intelligence & Big Data Analytics – An Introduction

1. **What is the primary goal of Artificial Intelligence (AI)?**
   a) To replace human intelligence
   b) To create systems that can perform tasks requiring human intelligence
   c) To eliminate the need for data analysis
   d) To focus only on robotics
   Answer: b) To create systems that can perform tasks requiring human intelligence
2. **Which of the following is a characteristic of Big Data?**
   a) Low volume
   b) Structured data only
   c) High velocity, volume, and variety
   d) Limited to text data
   Answer: c) High velocity, volume, and variety
3. **What does the term "velocity" in Big Data refer to?**
   a) The size of the data
   b) The speed at which data is generated and processed
   c) The variety of data types
   d) The accuracy of the data
   Answer: b) The speed at which data is generated and processed
4. **Which of the following is NOT a type of AI?**
   a) Narrow AI
   b) General AI
   c) Super AI
   d) Structured AI
   Answer: d) Structured AI
5. **What is the main challenge of Big Data analytics?**
   a) Lack of data
   b) Managing and processing large volumes of data
   c) Limited storage options
   d) High cost of hardware
   Answer: b) Managing and processing large volumes of data

Module 2: Basic Statistical Concepts

6. **What is the mean of the dataset: [2, 4, 6, 8, 10]?**

   a) 5

   b) 6

   c) 7

   d) 8

   Answer: b) 6

7. **Which measure of central tendency is most affected by outliers?**

   a) Mean

   b) Median

   c) Mode

   d) Range

   Answer: a) Mean

8. **What is the median of the dataset: [3, 5, 7, 9, 11]?**

   a) 5

   b) 7

   c) 9

   d) 11

   Answer: b) 7

9. **What does standard deviation measure?**

   a) Central tendency

   b) Spread of data

   c) Skewness of data

   d) Correlation between variables

   Answer: b) Spread of data

10. **Which of the following is a measure of dispersion?**

    a) Mean

    b) Median

    c) Variance

    d) Mode

    Answer: c) Variance

Module 3: Advanced Statistical Concepts

11. **What does a p-value less than 0.05 indicate in hypothesis testing?**
    a) The null hypothesis is true
    b) The null hypothesis is rejected
    c) The alternative hypothesis is false
    d) The test is inconclusive
    Answer: b) The null hypothesis is rejected
12. **Which of the following is a non-parametric test?**
    a) t-test
    b) ANOVA
    c) Chi-square test
    d) Pearson correlation
    Answer: c) Chi-square test
13. **What is the purpose of regression analysis?**
    a) To classify data into categories
    b) To predict the relationship between variables
    c) To measure central tendency
    d) To visualize data
    Answer: b) To predict the relationship between variables
14. **What is the null hypothesis in a statistical test?**
    a) The hypothesis that there is no effect or no difference
    b) The hypothesis that there is a significant effect
    c) The hypothesis that the data is normally distributed
    d) The hypothesis that the sample size is large enough
    Answer: a) The hypothesis that there is no effect or no difference
15. **What is the confidence interval used for?**
    a) To estimate the range of a population parameter
    b) To measure the spread of data
    c) To test the normality of data
    d) To calculate the mean of a dataset
    Answer: a) To estimate the range of a population parameter

Module 4: Statistical Tools and Usage

16. **Which tool is commonly used for statistical analysis and data visualization?**
    a) Tableau
    b) R
    c) Excel
    d) All of the above
    Answer: d) All of the above
17. **What is the primary purpose of a box plot?**
    a) To show the distribution of data
    b) To display correlation between variables
    c) To predict future trends
    d) To calculate the mean of a dataset
    Answer: a) To show the distribution of data
18. **Which statistical tool is best suited for handling large datasets?**
    a) Excel
    b) SPSS
    c) R
    d) Python
    Answer: c) R
19. **What is the purpose of a histogram?**
    a) To show the frequency distribution of data
    b) To display relationships between variables
    c) To calculate the mean of a dataset
    d) To test hypotheses
    Answer: a) To show the frequency distribution of data
20. **Which of the following is NOT a statistical software?**
    a) SAS
    b) MATLAB
    c) PowerPoint
    d) Stata
    Answer: c) PowerPoint

Module 5: Importing Data

21. **Which file format is commonly used for importing structured data?**
    a) CSV
    b) JSON
    c) XML
    d) All of the above
    Answer: d) All of the above
22. **What is the first step in importing data into a data analysis tool?**
    a) Cleaning the data
    b) Loading the data
    c) Visualizing the data
    d) Analyzing the data
    Answer: b) Loading the data
23. **Which Python library is used to import CSV files?**
    a) NumPy
    b) Pandas
    c) Matplotlib
    d) Scikit-learn
    Answer: b) Pandas
24. **What is the purpose of a delimiter in a CSV file?**
    a) To separate rows
    b) To separate columns
    c) To compress the file
    d) To encrypt the file
    Answer: b) To separate columns
25. **Which of the following is NOT a data import method?**
    a) Reading from a database
    b) Reading from a CSV file
    c) Reading from a PDF file
    d) Reading from an API
    Answer: c) Reading from a PDF file

Module 6: Pre-processing Data

26. **What is the purpose of data normalization?**
    a) To remove missing values
    b) To scale data to a specific range
    c) To remove outliers
    d) To sort data in ascending order
    Answer: b) To scale data to a specific range
27. **Which technique is used to handle missing data?**
    a) Imputation
    b) Normalization
    c) Aggregation
    d) Visualization
    Answer: a) Imputation
28. **What is the purpose of data binning?**
    a) To group data into intervals
    b) To remove outliers
    c) To normalize data
    d) To sort data
    Answer: a) To group data into intervals
29. **Which of the following is NOT a data pre-processing step?**
    a) Data cleaning
    b) Data visualization
    c) Data transformation
    d) Data reduction
    Answer: b) Data visualization
30. **What is the purpose of outlier detection?**
    a) To identify errors in data
    b) To normalize data
    c) To group data into categories
    d) To sort data
    Answer: a) To identify errors in data

Module 7: Exploring Data

31. **What is the purpose of exploratory data analysis (EDA)?**
    a) To build predictive models
    b) To summarize the main characteristics of a dataset
    c) To clean the data
    d) To import data from external sources
    Answer: b) To summarize the main characteristics of a dataset
32. **Which plot is used to visualize the relationship between two numerical variables?**
    a) Bar chart
    b) Scatter plot
    c) Pie chart
    d) Histogram
    Answer: b) Scatter plot
33. **What is the purpose of a correlation matrix?**
    a) To show relationships between variables
    b) To calculate the mean of a dataset
    c) To visualize missing data
    d) To normalize data
    Answer: a) To show relationships between variables
34. **Which of the following is NOT a step in EDA?**
    a) Data cleaning
    b) Data visualization
    c) Hypothesis testing
    d) Data modeling
    Answer: d) Data modeling
35. **What is the purpose of a pair plot?**
    a) To visualize relationships between multiple variables
    b) To calculate the mean of a dataset
    c) To detect missing values
    d) To normalize data
    Answer: a) To visualize relationships between multiple variables

Module 8: Data Structures and Algorithms

36. **Which data structure follows the Last-In-First-Out (LIFO) principle?**
    a) Queue
    b) Stack
    c) Linked List
    d) Tree
    Answer: b) Stack
37. **What is the time complexity of a binary search algorithm?**
    a) O(1)
    b) O(log n)
    c) O(n)
    d) O(n^2)
    Answer: b) O(log n)
38. **Which data structure is used to implement a priority queue?**
    a) Stack
    b) Queue
    c) Heap
    d) Linked List
    Answer: c) Heap
39. **What is the time complexity of a linear search algorithm?**
    a) O(1)
    b) O(log n)
    c) O(n)
    d) O(n^2)
    Answer: c) O(n)
40. **Which of the following is NOT a linear data structure?**
    a) Array
    b) Linked List
    c) Stack
    d) Tree
    Answer: d) Tree

Module 9: Graph Algorithms

41. **Which algorithm is used to find the shortest path in a weighted graph?**
    a) Breadth-First Search (BFS)
    b) Depth-First Search (DFS)
    c) Dijkstra's algorithm
    d) Kruskal's algorithm
    Answer: c) Dijkstra's algorithm
42. **What is the purpose of a spanning tree in graph theory?**
    a) To connect all nodes with the shortest path
    b) To connect all nodes without cycles
    c) To find the longest path in a graph
    d) To remove all edges from a graph
    Answer: b) To connect all nodes without cycles
43. **Which algorithm is used to find the minimum spanning tree?**
    a) Dijkstra's algorithm
    b) Prim's algorithm
    c) Bellman-Ford algorithm
    d) Floyd-Warshall algorithm
    Answer: b) Prim's algorithm
44. **What is the time complexity of Dijkstra's algorithm?**
    a) O(V^2)
    b) O(V log V)
    c) O(E log V)
    d) O(V + E)
    Answer: c) O(E log V)
45. **Which of the following is NOT a graph traversal algorithm?**
    a) Breadth-First Search (BFS)
    b) Depth-First Search (DFS)
    c) Dijkstra's algorithm
    d) A* algorithm
    *Answer: d) A algorithm**

Module 10: String Algorithms

46. **Which algorithm is used to find the longest common subsequence between two strings?**
    a) Knuth-Morris-Pratt algorithm
    b) Rabin-Karp algorithm
    c) Dynamic Programming
    d) Boyer-Moore algorithm
    Answer: c) Dynamic Programming
47. **What is the time complexity of the Knuth-Morris-Pratt algorithm for pattern matching?**
    a) O(n)
    b) O(n log n)
    c) O(n^2)
    d) O(log n)
    Answer: a) O(n)
48. **Which algorithm is used for string matching with hashing?**
    a) Knuth-Morris-Pratt algorithm
    b) Rabin-Karp algorithm
    c) Boyer-Moore algorithm
    d) Aho-Corasick algorithm
    Answer: b) Rabin-Karp algorithm
49. **What is the purpose of the Aho-Corasick algorithm?**
    a) To find the longest common subsequence
    b) To match multiple patterns in a text
    c) To perform string hashing
    d) To sort strings
    Answer: b) To match multiple patterns in a text
50. **Which of the following is NOT a string matching algorithm?**
    a) Knuth-Morris-Pratt algorithm
    b) Rabin-Karp algorithm
    c) Dijkstra's algorithm

d) Boyer-Moore algorithm

Answer: c) Dijkstra's algorithm

Module 11: Neural Networks

51. **What is the role of an activation function in a neural network?**
    a) To initialize weights
    b) To introduce non-linearity
    c) To reduce overfitting
    d) To normalize input data
    Answer: b) To introduce non-linearity
52. **Which of the following is a type of neural network architecture?**
    a) Convolutional Neural Network (CNN)
    b) Recurrent Neural Network (RNN)
    c) Generative Adversarial Network (GAN)
    d) All of the above
    Answer: d) All of the above
53. **What is the purpose of backpropagation in a neural network?**
    a) To initialize weights
    b) To update weights based on errors
    c) To normalize input data
    d) To reduce overfitting
    Answer: b) To update weights based on errors
54. **Which activation function is commonly used in the output layer of a binary classification neural network?**
    a) ReLU
    b) Sigmoid
    c) Tanh
    d) Softmax
    Answer: b) Sigmoid
55. **What is the purpose of dropout in a neural network?**
    a) To reduce overfitting

b) To increase accuracy

c) To initialize weights

d) To normalize input data

Answer: a) To reduce overfitting

Module 12: Programming for Data Science

56. **Which programming language is most commonly used in data science?**

    a) Java

    b) Python

    c) C++

    d) Ruby

    Answer: b) Python

57. **What is the purpose of the Pandas library in Python?**

    a) Data visualization

    b) Data manipulation and analysis

    c) Machine learning

    d) Web development

    Answer: b) Data manipulation and analysis

58. **Which Python library is used for data visualization?**

    a) NumPy

    b) Pandas

    c) Matplotlib

    d) Scikit-learn

    Answer: c) Matplotlib

59. **What is the purpose of the NumPy library in Python?**

    a) Data visualization

    b) Numerical computations

    c) Machine learning

    d) Web development

    Answer: b) Numerical computations

60. **Which of the following is NOT a Python library used in data science?**
    a) TensorFlow
    b) Keras
    c) Flask
    d) Scikit-learn
    Answer: c) Flask

Module 13: Applications of Pre-designed Algorithms

61. **Which algorithm is commonly used for recommendation systems?**
    a) Linear Regression
    b) K-Means Clustering
    c) Collaborative Filtering
    d) Decision Trees
    Answer: c) Collaborative Filtering
62. **What is the primary use of the Apriori algorithm?**
    a) Classification
    b) Clustering
    c) Association rule mining
    d) Regression
    Answer: c) Association rule mining
63. **Which algorithm is used for clustering data?**
    a) Linear Regression
    b) K-Means Clustering
    c) Decision Trees
    d) Support Vector Machines
    Answer: b) K-Means Clustering
64. **What is the purpose of the k-Nearest Neighbors (k-NN) algorithm?**
    a) Classification and regression
    b) Clustering
    c) Dimensionality reduction

d) Association rule mining

Answer: a) Classification and regression

65. **Which algorithm is used for dimensionality reduction?**

a) Principal Component Analysis (PCA)

b) K-Means Clustering

c) Decision TreESd) Support Vector Machines

Answer: a) Principal Component Analysis (PCA)
