BCA

(SEM. VI) MODEL PAPER – I
BCA – 6004 : DATA SCIENCE AND MACHINE LEARNING

Time : 1 :30 Hours *Maximum Marks : 75*

Q.1. Which one is NOT from Phase 1 of Data Science Life Cycle?
(1) Learning the target domain
(2) Developing initial hypothesis
(3) Visualize initial hypothesis
(4) Identifying key variables
Ans. (3) Visualize initial hypothesis

Q.2 Which of the following is the most important language for Data Science?
(1) Ruby (2) R (3) Java (4) None
Ans. (2) R

Q.3. A collection of information about a related topic is referred to as a_____
(1) Visualisation (2) Analysis
(3) Conclusion (4) Data
Ans. (1) Visualisation

Q.4. To find the _____ you add up all the numbers and then divide by how many numbers you have.
(1) Median (2) Mean
(3) Mode (4) Range
Ans. (2) Mean

Q.5. Which of the following is performed by Data Scientist?
(1) Create reproducible code (2) Challenge results
(3) Define the question (4) All of the above
Ans. (2) Challenge results

Q.6. Which is not a tool for Statistical Data Analysis?
(1) Logistic Regression (2) Linear & Non-linear Regression
(3) Histogram (4) ANOVA
Ans. (3) Histogram

Q.7. What is the mean of test scores?{70, 70, 80, 85, 85, 90, 95, 95, 100, 100}
(1) 85, 95, and 100 (2) 30
(3) 87 (4) None
Ans. (3) 87

Q.8. Choose the correct keyword for this definition: A graphical representation of a data set
(1) Data Set (2) Investigative Cycle
(3) Visualisation (4) None
Ans. (3) Visualisation

Q.9. To find the _____ you put all numbers in order from least to greatest and find the number that is in the middle.
(1) Median (2) Mode
(3) Mean (4) Range
Ans. (1) Median

Q.10 R is an interpreted language so it can access through _____?
(1) Command line interpreter
(2) Disk operating system
(3) Operating system
(4) User interface operating system
Ans. (1) Command line interpreter

Q.11. Data has been collected on visitors' viewing habits at a bank's website. Which technique is used to identify pages commonly viewed during the same visit to the website?
(1) Clustering (2) Classification
(3) Association Rules (4) Regression
Ans. (3) Association Rules

Q.12. A relationship between two or more variables is referred to as a _____
(1) Trend (2) Spike
(3) All of above (4) None of above
Ans. (1) Trend

Q.13. A graphical representation of a data set is referred to as a _____
(1) Visualization (2) Data Set
(3) Investigative Cycle (4) None
Ans. (1) Visualization

Q.14. Which of the following step is performed by data scientist AFTER acquiring the data?
(1) Data Integration (2) Data Replication
(3) Data Cleansing (4) All of the above
Ans. (3) Data Cleansing

Q.15. Data that sits outside the trend is referred to as a _____
(1) Outlier (2) Trend
(3) Spike (4) Both 1 & 2
Ans. (4) Both 1 & 2

Q.16. Which of the following approach should be used to ask a Data Analysis question?
(1) Find out the question which is to be answered
(2) Find only one solution for particular problem
(3) Find out answer from dataset without asking question
(4) None
Ans. (1) Find out the question which is to be answered

Q.17. Which of the following is NOT a machine learning algorithm?
(1) SVG (2) Random Forest
(3) SVM (4) None
Ans. (3) SVG

Q.18. What is Big Data?
(1) Data with the word 'big' in it
(2) Data about people who are big
(3) Data with a large size
(4) Data made with a big purpose
Ans. (3) Data with a large size

Q.19. What is R an implementation of?
(1) Logical Scoping
(2) S Programming Language
(3) Lexical Scoping
(4) Q Programming Language
Ans. (2) S Programming Language

Q.20. The 5 steps required to identify a problem and come up with a solution are referred to as the _____ Cycle
(1) Visualization (2) Investigative
(3) Conclusion (4) None
Ans. (1) Investigative

Q.21. Which of the following is characteristic of Processed Data?
(1) Hard to use for data analysis
(2) Data is not ready for analysis
(3) All steps should be noted
(4) None of the above
Ans. (4) None of the above

Q.22. Which was not mentioned as a latest trend tool_____
(1) Excel (2) Pentaho
(3) SPSS (4) Notepad
Ans. (4) Notepad

Q.23. Which of the following is one of the key data science skill ?
(1) Machine Learning (2) Statistics
(3) Data Visualization (4) All of the above
Ans. (4) All of the above

Q.24. Which of the following is not a stage in the Investigative Cycle?
(1) Investigate (2) Analysis
(3) Conclusion (4) None
Ans. (1) Investigate

Q.25. Vectors come in two parts_____ and _____
(1) Atomic vectors and list
(2) Atomic vectors and array
(3) Atomic vectors and matrix
(4) None
Ans. (1) Atomic vectors and list

Q.26 Choose the correct keyword for this definition. A collection of information about a related topic
(1) Trend (2) Spike
(3) Data Set (4) None
Ans. (3) Data Set

Q.27 The process of evaluating data through analytical and statistical tools.
(1) Data Mining (2) Data Exploration
(3) Data Analysis (4) Data Visualization
Ans. (3) Data Analysis

Q.28 Which of the following is key characteristic of hacker?
(1) Willing to find answers on their own
(2) Afraid to say they don't know the answer
(3) Not Willing to find answers on their own
(4) All of the mentioned
Ans. (1) Willing to find answers on their own

Q.29 Which of the following characteristic of big data is relatively more concerned to data science?
(1) Variety (2) Volume
(3) Velocity (4) None
Ans. (1) Variety

Q.30. R is an_____ programming language?
(1) GPL (2) Open source
(3) Closed source (4) Definite source
Ans. (2) Open source

Q.31 What is the primary goal of data science?
(1) Data storage
(2) Extracting actionable insights from data
(3) Creating data visualizations
(4) Software development
Ans. (2) Extracting actionable insights from data

Q.32 Which of the following is a characteristic of structured data?
(1) Unorganized and unstructured
(2) Easily fit into relational databases
(3) Text-based only
(4) Primarily used in deep learning
Ans. (2) Easily fit into relational databases

Q.33 What is the main purpose of exploratory data analysis (EDA)?
(1) Making predictions
(2) Confirming hypotheses
(3) Summarizing and understanding data
(4) Deploying machine learning models
Ans. (3) Summarizing and understanding data

Q.34 In machine learning, what does "supervised learning" involve?
(1) Learning without labeled data
(2) Learning from examples with labeled outcomes
(3) Learning without a predefined goal
(4) Learning from unstructured data
Ans. (2) Learning from examples with labeled outcomes

Q.35 What does the term "feature engineering" refer to in the context of machine learning?
(1) Creating new features from existing ones
(2) Designing graphical user interfaces
(3) Selecting the most important features
(4) Debugging machine learning models
Ans. (1) Creating new features from existing ones

Q.36 Which programming language is widely used for statistical analysis and data manipulation in data science?
(1) Java (2) Python
(3) C++ (4) Ruby
Ans. (2) Python

Q.36 What is the purpose of cross-validation in machine learning?
(1) Ensuring data security
(2) Evaluating model performance using multiple subsets of data
(3) Encrypting sensitive information
(4) Creating ensemble models
Ans. (2) Evaluating model performance using multiple subsets of data

Q.37 What does the term "Big Data" generally refer to?
(1) Any large dataset
(2) Datasets with high variability
(3) Extremely large and complex datasets
(4) Datasets with low dimensionality
Ans. (3) Extremely large and complex datasets

Q.38 What role does a data scientist play in a business context?
(1) Managing IT infrastructure
(2) Extracting insights from data to inform decision-making
(3) Developing marketing strategies
(4) Writing software code
Ans. (2) Extracting insights from data to inform decision-making

Q.39 When did the term "data science" first emerge?
(1) 1980s (2) 1990s
(3) 2000s (4) 2010s
Ans. (2) 1990s

Q.40 Which technological advancement played a significant role in the growth of data science by enabling the storage and processing of massive datasets?
(1) Mainframe computers (2) Cloud computing
(3) Quantum computing (4) Supercomputers
Ans. (2) Cloud computing

Q.41 In the context of the evolution of data science, what is the significance of the "Fourth Paradigm"?
(1) Emergence of big data technologies
(2) Integration of computational methods in scientific research
(3) Introduction of machine learning algorithms
(4) Rise of data visualization tools
Ans. (2) Integration of computational methods in scientific research

Q.43 Which era of data science is characterized by the widespread use of machine learning algorithms and advanced analytics techniques?
(1) Era of Statistical Analysis
(2) Era of Big Data
(3) Era of Machine Learning
(4) Era of Data Visualization
Ans. (3) Era of Machine Learning

Q.44 What role did open-source software play in the evolution of data science?
(1) Hindered progress by limiting access to proprietary tools
(2) Facilitated collaboration and innovation through freely available tools
(3) Increased reliance on commercial software
(4) Had no impact on the field
Ans. (2) Facilitated collaboration and innovation through freely available tools

Q.45 What is the primary goal of data science in a business context?
(1) Managing IT infrastructure
(2) Developing marketing strategies
(3) Extracting insights from data to inform decision-making
(4) Writing software code
Ans. (3) Extracting insights from data to inform decision-making

Q.46 Which term refers to the process of cleaning and preparing raw data for analysis in data science?
(1) Data visualization (2) Data engineering
(3) Data wrangling (4) Data mining
Ans. (3) Data wrangling

Q.47 In machine learning, what is the role of feature engineering?
(1) Selecting the most important features
(2) Creating new features from existing ones
(3) Debugging machine learning models
(4) Conducting statistical analyses
Ans. (2) Creating new features from existing ones

Q.48 What is the purpose of cross-validation in machine learning?
(1) Ensuring data security
(2) Evaluating model performance using multiple subsets of data
(3) Encrypting sensitive information
(4) Creating ensemble models
Ans. (2) Evaluating model performance using multiple subsets of data

Q.49 Which aspect of data science is concerned with making predictions based on historical data and patterns?
(1) Exploratory Data Analysis (EDA)
(2) Predictive Modeling
(3) Descriptive Statistics
(4) Data Visualization
Ans. (2) Predictive Modeling

Q.50 Which stage of the data science process involves defining the problem, understanding the goals, and formulating hypotheses?
(1) Data Collection (2) Data Cleaning
(3) Problem Definition (4) Data Analysis
Ans. (3) Problem Definition

Q.51 During which stage of data science do data scientists explore and manipulate the data to gain insights and identify patterns?
(1) Data Visualization
(2) Exploratory Data Analysis (EDA)
(3) Data Modeling
(4) Data Interpretation
Ans. (2) Exploratory Data Analysis (EDA)

Q.52 What is the primary purpose of data collection in the context of data science?
(1) To store data in databases
(2) To ensure data privacy
(3) To extract insights and knowledge from raw data
(4) To encrypt sensitive information
Ans. (3) To extract insights and knowledge from raw data

Q.53 Which of the following is an example of primary data collection?
(1) Analyzing existing research papers
(2) Conducting surveys and interviews
(3) Scraping data from websites
(4) Purchasing data from a third-party vendor
Ans. (2) Conducting surveys and interviews

Q.54 What is the role of sampling in data collection?
(1) Storing data efficiently
(2) Selecting a representative subset of the population
(3) Encrypting sensitive information
(4) Performing statistical analyses
Ans. (2) Selecting a representative subset of the population

Q.55 Which data collection method involves observing and recording behavior in a natural environment without interference?
(1) Experiments (2) Surveys
(3) Observational studies (4) Case studies
Ans. (3) Observational studies

Q.56 In social media analytics, what type of data is often collected for sentiment analysis?

(1) Numerical data (2) Textual data
(3) Image data (4) Audio data
Ans. (2) Textual data

Q.57. What is the purpose of data cleaning in the preprocessing stage?
(1) Encrypting data
(2) Transforming data into a usable format
(3) Selecting the most important features
(4) Creating new features
Ans. (2) Transforming data into a usable format

Q.58. Why is handling missing data important in data preprocessing?
(1) To reduce data storage costs
(2) To improve data visualization
(3) To avoid biased analysis and predictions
(4) To encrypt sensitive information
Ans. (3) To avoid biased analysis and predictions

Q.59. Which technique involves transforming categorical variables into numerical format for machine learning models?
(1) Data normalization (2) Feature scaling
(3) One-hot encoding (4) Data imputation
Ans. (3) One-hot encoding

Q.60. What is the purpose of outlier detection and treatment in data preprocessing?
(1) To create new features
(2) To remove irrelevant data
(3) To handle extreme values that may affect analysis
(4) To improve data visualization
Ans. (3) To handle extreme values that may affect analysis

Q.61. What does the term "feature scaling" refer to in data preprocessing?
(1) Creating new features from existing ones
(2) Standardizing numerical features to a common scale
(3) Selecting the most important features
(4) Handling missing data
Ans. (2) Standardizing numerical features to a common scale

Q.62. What is the primary goal of probability sampling in data collection?
(1) To simplify the data analysis process
(2) To ensure every data point is included
(3) To reduce bias in the sample
(4) To guarantee anonymity of respondents
Ans. (3) To reduce bias in the sample

Q.63. Which data collection method is suitable for studying the impact of a specific intervention or treatment?
(1) Surveys (2) Observational studies
(3) Experiments (4) Case studies
Ans. (3) Experiments

Q.64. In stratified sampling, how are subgroups determined?
(1) Randomly
(2) Based on pre-defined characteristics
(3) According to sample size
(4) Alphabetically
Ans. (2) Based on pre-defined characteristics

Q.65. What is the purpose of random sampling in data collection?
(1) To guarantee anonymity of respondents
(2) To select a representative subset of the population
(3) To simplify the data analysis process
(4) To ensure every data point is included
Ans. (2) To select a representative subset of the population

Q.66. In cross-sectional studies, how is data collected?
(1) Continuously over time
(2) From a single point in time
(3) Through experiments
(4) Using case studies
Ans. (2) From a single point in time

Q.67. Which sampling method is often used in market research to gather insights from a specific target audience?
(1) Convenience sampling (2) Stratified sampling
(3) Snowball sampling (4) Quota sampling
Ans. (4) Quota sampling

Q.68. What is the primary goal of Exploratory Data Analysis (EDA) in data science?
(1) To predict future trends
(2) To summarize data using statistical measures
(3) To visualize and understand the characteristics of the data
(4) To transform raw data into a usable format
Ans. (3) To visualize and understand the characteristics of the data

Q.69. Which visualization technique is suitable for understanding the distribution of a continuous numerical variable?
(1) Pie chart (2) Histogram
(3) Bar chart (4) Line chart
Ans. (2) Histogram

Q.70. In EDA, what does a boxplot primarily represent?
(1) Central tendency
(2) Outliers and distribution
(3) Correlation between variables
(4) Trend over time
Ans. (2) Outliers and distribution

Q.71. What is the purpose of a scatter plot in EDA?
(1) Showing the frequency of categorical variables
(2) Visualizing the distribution of a single variable
(3) Displaying the relationship between two numerical variables

(4) Highlighting outliers in the data

Ans. (3) Displaying the relationship between two numerical variables

Q.72 Which statistical measure is commonly used to assess the central tendency of a dataset in EDA?

(1) Variance (2) Median
(3) Range (4) Standard Deviation

Ans. (2) Median

Q.73 What does the term "correlation" represent in the context of EDA?

(1) The strength and direction of a linear relationship between two variables
(2) The spread of values in a dataset
(3) The average value of a dataset
(4) The occurrence of outliers

Ans. (1) The strength and direction of a linear relationship between two variables

Q.74 What is the purpose of a heat map in EDA?

(1) Displaying the distribution of a single variable
(2) Highlighting outliers in the data
(3) Showing the correlation between multiple variables
(4) Representing the central tendency of a dataset

Ans. (3) Showing the correlation between multiple variables

Q.75 Which EDA technique is useful for identifying missing values in a dataset?

(1) Boxplot (2) Scatter plot
(3) Bar chart (4) Missing data matrix

Ans. (4) Missing data matrix

Q.76 What does the term "skewness" indicate when analyzing a dataset during EDA?

(1) The presence of outliers
(2) The symmetry of the data distribution
(3) The spread of values in a dataset
(4) The strength of correlation between variables

Ans. (2) The symmetry of the data distribution

Q.77 In EDA, what is the purpose of the "pair plot" technique?

(1) Visualizing the distribution of a single variable
(2) Comparing the means of multiple groups
(3) Showing the relationship between every pair of numerical variables
(4) Highlighting extreme values in the data

Ans. (3) Showing the relationship between every pair of numerical variables

Q.78 What does the arithmetic mean represent in a dataset?

(1) The middle value
(2) The most frequent value
(3) The sum of all values divided by the number of values

(4) The range of values

Ans. (3) The sum of all values divided by the number of values

Q.79 How is the mean affected by extreme values or outliers in a dataset?

(1) It remains unchanged (2) It increases
(3) It decreases (4) It becomes undefined

Ans. (2) It increases

Q.80 In a positively skewed distribution, where is the mean located in relation to the median and mode?

(1) Mean = Median = Mode (2) Median < Median < Mode
(3) Mean > Median > Mode (4) Mean > Median < Mode

Ans. (3) Mean > Median > Mode

Q.81 What is the alternative name for the arithmetic mean?

(1) Central Tendency (2) Average
(3) Median (4) Mode

Ans. (2) Average

Q.82 If all values in a dataset are the same, what is the relationship between the mean, median, and mode?

(1) Mean = Median = Mode (2) Mean < Median < Mode
(3) Mean > Median > Mode (4) Mean > Median < Mode

Ans. (1) Mean = Median = Mode

Q.83 What is machine learning?

(1) A type of computer hardware
(2) A branch of artificial intelligence
(3) A programming language
(4) A type of robotic technology

Ans. (2) A branch of artificial intelligence

Q.84 Which of the following is a key goal of machine learning?

(1) Automating tasks without explicit programming
(2) Creating complex algorithms
(3) Building advanced user interfaces
(4) Developing virtual reality systems

Ans. (1) Automating tasks without explicit programming

Q.85 What is supervised learning?

(1) Learning from labeled data
(2) Learning without any guidance
(3) Learning in a team environment
(4) Learning from unstructured data

Ans. (1) Learning from labeled data

Q.86 In machine learning, what does "label" refer to?

(1) A brand name
(2) Output or outcome variable
(3) An algorithm
(4) A programming language

Ans. (2) Output or outcome variable

Q.87.Which algorithm is commonly used for classification problems in machine learning?
(1) K-means clustering (2) Decision trees
(3) Linear regression (4) Apriori algorithm
Ans. (2) Decision trees

Q.88.What is unsupervised learning?
(1) Learning from labeled data
(2) Learning without any guidance
(3) Learning in a team environment
(4) Learning from unstructured data
Ans. (2) Learning without any guidance

Q.89.What is the purpose of feature engineering in machine learning?
(1) To design user interfaces
(2) To create artificial features
(3) To extract relevant information from data
(4) To optimize hardware performance
Ans. (3) To extract relevant information from data

Q.90.Which of the following is an example of a regression problem?
(1) Image classification
(2) Spam detection
(3) Predicting house prices
(4) Customer segmentation
Ans. (3) Predicting house prices

Q.91.What is the basic building block of a neural network?
(1) Neuron (2) Feature
(3) Instance (4) Layer
Ans. (1) Neuron

Q.92.In a feed forward neural network, how are neurons organized?
(1) In a single layer
(2) In multiple layers
(3) In a random configuration
(4) In a circular arrangement
Ans. (2) In multiple layers

Q.93.What is the role of an activation function in a neural network?
(1) It defines the learning rate
(2) It connects input and output layers
(3) It introduces non-linearity
(4) It performs feature scaling
Ans. (3) It introduces non-linearity

Q.94.What does the term "backpropagation" refer to in neural networks?
(1) Forward pass of information
(2) Optimization algorithm for training
(3) Activation function
(4) Input layer of the network
Ans. (2) Optimization algorithm for training

Q.95.What is the purpose of the "bias" term in a neural network?
(1) To add randomness to the model
(2) To balance the data distribution
(3) To account for errors in predictions
(4) To ensure the model fits the training data perfectly
Ans. (3) To account for errors in predictions

Q.96.In a convolutional neural network (CNN), what is the primary operation performed by convolutional layers?
(1) Matrix multiplication
(2) Element-wise multiplication
(3) Convolution operation
(4) Activation function application
Ans. (3) Convolution operation

Q.97.What is the purpose of pooling layers in a CNN?
(1) To reduce the spatial dimensions of the input
(2) To increase the number of filters
(3) To add more hidden layers
(4) To apply the activation function
Ans. (1) To reduce the spatial dimensions of the input

Q.98.Which type of neural network is designed to learn from sequential data and time-series information?
(1) Convolutional Neural Network (CNN)
(2) Recurrent Neural Network (RNN)
(3) Multilayer Perceptron (MLP)
(4) Radial Basis Function Network (RBFN)
Ans. (2) To reduce the spatial dimensions of the input

Q.99.What is the vanishing gradient problem in neural networks?
(1) The model converges too quickly
(2) The model fails to converge
(3) The gradients become too large during training
(4) The gradients approach zero, hindering learning in deep networks
Ans. (4) The gradients approach zero, hindering learning in deep networks

Q.100.Which neural network architecture is inspired by the structure and function of the human brain?
(1) Convolutional Neural Network (CNN)
(2) Recurrent Neural Network (RNN)
(3) Radial Basis Function Network (RBFN)
(4) Neural Network with Backpropagation
Ans. (4) Neural Network with Backpropagation

BCA

(SEM. VI) MODEL PAPER – II

BCA – 6004 : DATA SCIENCE AND MACHINE LEARNING

Time : 1 30 Hours *Maximum Marks :*

Q.1. Data science is the process of diverse set of data through?
(1) Organizing data (2) Processing data
(3) Analysing data (4) All of the above
Ans. (4) All of the above

Q.2. Point out the correct statement.
(1) Raw data is original source of data
(2) Preprocessed data is original source of data
(3) Raw data is the data obtained after processing steps
(4) None of the above
Ans. (1) Raw data is original source of data

Q.3. How do we perform Bayesian classification when some features are missing?
(1) We integrate the posteriors probabilities over the missing features
(2) We ignore the missing features
(3) We assuming the missing values as the mean of all values
(4) Drop the features completely
Ans. (1) We integrate the posteriors probabilities over the missing features

Q.4. The modern conception of data science as an independent discipline is sometimes attributed to?
(1) John McCarthy (2) Arthur Samuel
(3) William S. (4) Dennis Ritchie
Ans. (3) William S.

Q.5. _____ graph displays information as a series of data points connected by straight line segments.
(1) Bar (2) Scatter
(3) Histogram (4) Line
Ans. (4) Line

Q.6. Data fishing is sometimes referred to as
(1) Data bagging (2) Data dredging
(3) Data merging (4) None of the mentioned
Ans. (2) Data dredging

Q.7. Which is one of the significant data science skills?
(1) Statistics (2) Data Visualization
(3) Machine Learning (4) All of the above
Ans. (4) All of the above

Q.8 A method used to make vector of repeated values?
(1) read() (2) data()
(3) rep() (4) view()
Ans. (3) rep()

Q.9. Which of the following step is performed by the data scientist after acquiring the data?
(1) Data Replication (2) Data Integration
(3) Data Cleansing (4) All of the Mentioned
Ans. (3) Data Cleansing

Q.10 K- nearest neighbors algorithm is based on _____ learning
(1) Unsupervised (2) Supervised
(3) Association (4) Correlation
Ans. (2) Supervised

Q.11 Which of the following statement is true?
(1) The nature of our business problem determines how outliers are used
(2) Outliers is a data point that is significantly close to other data points
(3) Outliers should be identified and removed always from a dataset
(4) Outliers can never be present in the testing dataset
Ans. (1) The nature of our business problem determines how outliers are used

Q.12 Which of the following function gives information about top-level data?
(1) tail (2) summary
(3) head (4) None of the above
Ans. (3) head

Q.13 Which of the following approach should be used if someone can't fix the variable?
(1) Non-stratify it (2) Randomize it
(3) Generalize it (4) None of the above
Ans. (4) None of the above

Q.14 Which of the following is good way of performing experiments in data science?
(1) Measure variability (2) Generalize to the problem
(3) Have Replication (4) All of the above
Ans. (4) All of the above

Q.15 Which of the following technique comes under practical machine learning?
(1) Bagging (2) Forecasting
(3) Boosting (4) None of the above
Ans. (3) Boosting

Q.16 CNN is mostly used for which type of data
(1) Structured Data
(2) Unstructured Data
(3) Both Structured and Unstructured
(4) None of the above
Ans. (2) Unstructured Data

Q.17 Which of the following transforms can be performed with data value?
(1) log10 (2) cos
(3) log2 (4) All of the above
Ans. (4) All of the above

Q.18 Which of the following SGD variants is based on both momentum and adaptive learning?
(1) Adagrad (2) RMSprop
(3) Adam (4) Nesterov
Ans. (3) Adam

Q.19 What is the work of a Data Architect?
(1) Utilize large data sets to gather information that meets the company's needs
(2) Work with businesses to determine the best usage of the information yielded from data
(3) are Build data solutions that optimized for performance and design applications
(4) All of the above
Ans. (3) are Build data solutions that optimized for performance and design applications

Q.20 A perfect negative correlation is signified by _____
(1) 1 (2) -1 (3) 2 (4) 0
Ans. (1) -1

Q.21 Which stage of the data science process involves defining the problem, understanding the goals, and formulating hypotheses?
(1) Data Collection (2) Data Cleaning
(3) Problem Definition (4) Data Analysis
Ans. (3) Problem Definition

Q.22 During which stage of data science do data scientists explore and manipulate the data to gain insights and identify patterns?
(1) Data Visualization
(2) Exploratory Data Analysis (EDA)
(3) Data Modeling
(4) Data Interpretation
Ans. (2) Exploratory Data Analysis (EDA)

Q.23 What is the primary goal of the data modeling stage in data science?
(1) Cleaning and preparing data
(2) Making predictions and building models
(3) Visualizing data

(4) Defining the business problem
Ans. (2) Making predictions and building models

Q.24 In which stage of the data science process do data scientists deploy models into a production environment?
(1) Data Interpretation
(2) Model Deployment
(3) Data Cleaning
(4) Exploratory Data Analysis (EDA)
Ans. (2) Model Deployment

Q.25 Which stage involves communicating the results, findings, and insights to stakeholders in a clear and understandable manner?
(1) Data Interpretation (2) Data Visualization
(3) Model Deployment (4) Problem Definition
Ans. (1) Data Interpretation

Q.26 What is the primary purpose of encryption in data security?
(1) Improving data processing speed
(2) Making data more readable
(3) Protecting data from unauthorized access
(4) Reducing storage costs
Ans. (3) Protecting data from unauthorized access

Q.27 Which type of attack involves tricking individuals into providing sensitive information, such as passwords or credit card numbers?
(1) Phishing (2) Denial-of-Service (DoS)
(3) Malware (4) Spoofing
Ans. (1) Phishing

Q.28 What is the purpose of access controls in data security?
(1) Encrypting data during transmission
(2) Managing user permissions and restricting access
(3) Detecting and removing malware
(4) Monitoring network traffic
Ans. (2) Managing user permissions and restricting access

Q.29 Which data security measure involves regularly updating software and systems to patch vulnerabilities?
(1) Firewall protection
(2) Intrusion Detection System (IDS)
(3) Penetration testing
(4) Software patching
Ans. (4) Software patching

Q.30 What does the term "data masking" refer to in data security?
(1) Hiding data from view
(2) Encrypting data at rest
(3) Creating backups of data
(4) Monitoring network traffic
Ans. (1) Hiding data from view

Q.16 CNN is mostly used for which type of data
(1) Structured Data
(2) Unstructured Data
(3) Both Structured and Unstructured
(4) None of the above
Ans. (2) Unstructured Data

Q.17 Which of the following transforms can be performed with data value
(1) log10 (2) cos
(3) log2 (4) All of the above
Ans. (4) All of the above

Q.18 Which of the following SGD variants is based on both momentum and adaptive learning?
(1) Adagrad (2) RMSprop
(3) Adam (4) Nesterov
Ans. (3) Adam

Q.19 What is the work of a Data Architect?
(1) Utilize large data sets to gather information that meets the company's needs
(2) Work with businesses to determine the best usage of the information yielded from data
(3) are Build data solutions that optimized for performance and design applications
(4) All of the above
Ans. (3) are Build data solutions that optimized for performance and design applications

Q.20 A perfect negative correlation is signified by _____
(1) 1 (2) -1 (3) 2 (4) 0
Ans. (1) -1

Q.21 Which stage of the data science process involves defining the problem, understanding the goals, and formulating hypotheses?
(1) Data Collection (2) Data Cleaning
(3) Problem Definition (4) Data Analysis
Ans. (3) Problem Definition

Q.22 During which stage of data science do data scientists explore and manipulate the data to gain insights and identify patterns?
(1) Data Visualization
(2) Exploratory Data Analysis (EDA)
(3) Data Modeling
(4) Data Interpretation
Ans. (2) Exploratory Data Analysis (EDA)

Q.23 What is the primary goal of the data modeling stage in data science?
(1) Cleaning and preparing data
(2) Making predictions and building models
(3) Visualizing data
(4) Defining the business problem
Ans. (2) Making predictions and building models

Q.24 In which stage of the data science process do data scientists deploy models into a production environment?
(1) Data Interpretation
(2) Model Deployment
(3) Data Cleaning
(4) Exploratory Data Analysis (EDA)
Ans. (2) Model Deployment

Q.25 Which stage involves communicating the results, findings, and insights to stakeholders in a clear and understandable manner?
(1) Data Interpretation (2) Data Visualization
(3) Model Deployment (4) Problem Definition
Ans. (1) Data Interpretation

Q.26 What is the primary purpose of encryption in data security?
(1) Improving data processing speed
(2) Making data more readable
(3) Protecting data from unauthorized access
(4) Reducing storage costs
Ans. (3) Protecting data from unauthorized access

Q.27 Which type of attack involves tricking individuals into providing sensitive information, such as passwords or credit card numbers?
(1) Phishing (2) Denial-of-Service (DoS)
(3) Malware (4) Spoofing
Ans. (1) Phishing

Q.28 What is the purpose of access controls in data security?
(1) Encrypting data during transmission
(2) Managing user permissions and restricting access
(3) Detecting and removing malware
(4) Monitoring network traffic
Ans. (2) Managing user permissions and restricting access

Q.29 Which data security measure involves regularly updating software and systems to patch vulnerabilities?
(1) Firewall protection
(2) Intrusion Detection System (IDS)
(3) Penetration testing
(4) Software patching
Ans. (4) Software patching

Q.30 What does the term "data masking" refer to in data security?
(1) Hiding data from view
(2) Encrypting data at rest
(3) Creating backups of data
(4) Monitoring network traffic
Ans. (1) Hiding data from view

Q.31 What is the primary consideration when choosing a sampling method in data collection?
(1) Ensuring anonymity of respondents
(2) Cost-effectiveness
(3) Reducing selection bias
(4) Establishing causation
Ans. (3) Reducing selection bias

Q.32 What is the purpose of data cleaning in the data preprocessing pipeline?
(1) To create new features
(2) To handle missing or erroneous data
(3) To improve data visualization
(4) To guarantee data privacy
Ans. (2) To handle missing or erroneous data

Q.33 Which step in data preprocessing involves transforming categorical variables into a numerical format for machine learning models?
(1) Feature scaling (2) Data imputation
(3) One-hot encoding (4) Data normalization
Ans. (3) One-hot encoding

Q.34 Why is handling imbalanced datasets important in data preprocessing?
(1) To increase data variety
(2) To avoid bias in model training
(3) To simplify the data analysis process
(4) To create new features
Ans. (2) To avoid bias in model training

Q.35 What does the term "data imputation" refer to in data preprocessing?
(1) Removing outliers from the dataset
(2) Handling missing data
(3) Encoding categorical variables
(4) Creating new features
Ans. (2) Handling missing data

Q.36 In feature scaling, what is the purpose of standardizing numerical features to a common scale?
(1) To handle missing data
(2) To create new features
(3) To improve data visualization
(4) To ensure fairness in model training
Ans. (4) To ensure fairness in model training

Q.37 What is the role of outlier detection and treatment in data preprocessing?
(1) To remove irrelevant data
(2) To improve data visualization
(3) To handle extreme values that may affect analysis

(4) To create new features
Ans. (3) To handle extreme values that may affect analysis

Q.38 Which technique involves converting text data into numerical representations suitable for machine learning models?
(1) Feature scaling (2) Tokenization
(3) Data normalization (4) Data imputation
Ans. (2) Tokenization

Q.39 What is the purpose of data normalization in data preprocessing?
(1) To handle missing data
(2) To create new features
(3) To standardize numerical features to a common scale
(4) To remove outliers
Ans. (3) To standardize numerical features to a common scale

Q.40 Which step in data preprocessing involves checking for and addressing duplicated records?
(1) Data normalization (2) Data deduplication
(3) Feature scaling (4) Data imputation
Ans. (2) Data deduplication

Q.41 What is the primary purpose of data preprocessing in machine learning?
(1) To increase data storage capacity
(2) To improve data visualization
(3) To enhance the performance of machine learning models
(4) To encrypt sensitive information
Ans. (3) To enhance the performance of machine learning models

Q.42 What is the primary goal of data cleaning in the context of data science?
(1) To increase data storage capacity
(2) To transform raw data into a usable format
(3) To guarantee data privacy
(4) To improve data visualization
Ans. (2) To transform raw data into a usable format

Q.43 Why is handling missing data important in data cleaning?
(1) To create new features
(2) To improve data visualization
(3) To avoid biased analysis and predictions
(4) To increase data variety
Ans. (3) To avoid biased analysis and predictions

Q.44 What technique involves removing or correcting inaccurate values in a dataset during data cleaning?
(1) Outlier detection (2) Data imputation
(3) Data deduplication (4) Error correction
Ans. (4) Error correction

Q.45 In the context of data cleaning, what is the purpose of handling outliers?
(1) To simplify the data analysis process
(2) To create new features
(3) To remove irrelevant data
(4) To handle extreme values that may affect analysis
Ans. (4) To handle extreme values that may affect analysis

Q.46 What is the role of deduplication in data cleaning?
(1) Handling missing data
(2) Removing irrelevant data
(3) Handling outliers
(4) Identifying and eliminating duplicate records
Ans. (4) Identifying and eliminating duplicate records

Q.47 What does the standard deviation measure in a dataset?
(1) Central tendency (2) Spread or dispersion
(3) Skewness (4) Median
Ans. (2) Spread or dispersion

Q.48 How is the standard deviation calculated?
(1) The range of values
(2) The square root of the variance
(3) The mean of the dataset
(4) The interquartile range (IQR)
Ans. (2) The square root of the variance

Q.49 If the standard deviation of a dataset is zero, what can be concluded about the data?
(1) The data is normally distributed
(2) All values in the dataset are the same
(3) The data is skewed
(4) The dataset has outliers
Ans. (2) All values in the dataset are the same

Q.50 How does the standard deviation react to outliers in a dataset?
(1) It increases (2) It decreases
(3) It remains unchanged (4) It becomes undefined
Ans. (1) It increases

Q.51 In a normal distribution, what percentage of data falls within one standard deviation of the mean?
(1) 25% (2) 50% (3) 68% (4) 95%
Ans. (3) 68%.

Q.52 Which measure of spread is resistant to extreme values and outliers?
(1) Range
(2) Interquartile Range (IQR)
(3) Mean Absolute Deviation (MAD)
(4) Standard Deviation
Ans. (2) Interquartile Range (IQR)

Q.53 What does a larger standard deviation indicate about a dataset?
(1) The data is more spread out
(2) The data is more concentrated
(3) The dataset is smaller
(4) The data is perfectly symmetric
Ans. (1) The data is more spread out

Q.54 If each value in a dataset is multiplied by a constant, how is the standard deviation affected?
(1) The standard deviation remains unchanged
(2) The standard deviation is divided by the constant
(3) The standard deviation is multiplied by the constant
(4) The standard deviation becomes undefined
Ans. (3) The standard deviation is multiplied by the constant

Q.55 What is the relationship between the variance and the standard deviation?
(1) They are equal
(2) Variance is the square root of the standard deviation
(3) Standard deviation is the square root of the variance
(4) They are unrelated
Ans. (3) Standard deviation is the square root of the variance

Q.56 In a positively skewed distribution, how is the standard deviation located in relation to the mean?
(1) Standard Deviation > Mean
(2) Standard Deviation < Mean
(3) Standard Deviation = Mean
(4) Standard Deviation is undefined
Ans. (1) Standard Deviation > Mean

Q.57 What does skewness measure in a probability distribution?
(1) Spread or dispersion (2) Symmetry
(3) Central tendency (4) Kurtosis
Ans. (2) Symmetry

Q.58 In a positively skewed distribution, where is the majority of the data located?
(1) Left side (2) Right side
(3) Center (4) Upper tail
Ans. (2) Right side

Q.59 How is skewness calculated for a dataset?
(1) The difference between the mean and median
(2) The sum of squared deviations from the mean
(3) The third standardized moment
(4) The range of values
Ans. (3) The third standardized moment

Q.60 What does a negative skewness value indicate about a dataset?
(1) Right-skewed (2) Left-skewed
(3) Symmetric (4) Normally distributed
Ans. (2) Left-skewed

Q.61 What is the range of possible values for skewness?
(1) -1 to 0 (2) -2 to 2
(3) -∞ to ∞ (4) 0 to 1
Ans. (3) -∞ to ∞

Q.62 How is kurtosis defined in statistics?
(1) Measure of spread
(2) Measure of symmetry
(3) Measure of the tailedness of a distribution
(4) Measure of central tendency
Ans. (3) Measure of the tailedness of a distribution

Q.63 In a leptokurtic distribution, how are the tails compared to a normal distribution?
(1) Thinner tails (2) Fatter tails
(3) Symmetric tails (4) Undefined tails
Ans. (2) Fatter tails

Q.64 What does a positive kurtosis value indicate about the tails of a distribution?
(1) Thinner tails (2) Fatter tails
(3) Symmetric tails (4) Undefined tails
Ans. (2) Fatter tails

Q.65 Which distribution has a kurtosis of 3 in the standard normal distribution?
(1) Mesokurtic (2) Platykurtic
(3) Leptokurtic (4) Undefined
Ans. (1) Mesokurtic

Q.66 What is the range of possible values for kurtosis in a normal distribution?
(1) -1 to 1 (2) -2 to 2
(3) 0 to 3 (4) -∞ to ∞
Ans. (2) -2 to 2

Q.67 Which type of machine learning algorithm is inspired by the structure and function of the human brain?
(1) Supervised learning
(2) Unsupervised learning
(3) Reinforcement learning
(4) Neural networks
Ans. (4) Neural networks

Q.68 What is the role of a bias term in linear regression?
(1) It measures the variability of the data
(2) It accounts for errors in predictions
(3) It represents the slope of the line
(4) It ensures the model passes through the origin
Ans. (4) It ensures the model passes through the origin

Q.69 Which method is used for feature scaling in machine learning?
(1) Standardization (2) Normalization
(3) Both 1 and 2 (4) None of the above
Ans. (3) Both a and b

Q.70 In the context of machine learning, what is a hyperplane?
(1) A plane in a 3D space
(2) A line that separates two classes in a feature space
(3) A type of neural network layer
(4) A method for data preprocessing
Ans. (2) A line that separates two classes in a feature space

Q.71 What is the purpose of the activation function in a neural network?
(1) To normalize the input data
(2) To calculate the loss function
(3) To introduce non-linearity
(4) To define the learning rate
Ans. (3) To introduce non-linearity

Q.72 What is ensemble learning?
(1) Training multiple models independently
(2) Combining the predictions of multiple models
(3) Learning from unstructured data
(4) Using neural networks for feature extraction
Ans. (2) Combining the predictions of multiple models

Q.73 Which algorithm is suitable for anomaly detection?
(1) K-means clustering (2) Isolation Forest
(3) Linear regression (4) Decision trees
Ans. (2) Isolation Forest

Q.74 What is the role of dropout in neural networks?
(1) It prevents overfitting
(2) It increases the learning rate
(3) It speeds up training time
(4) It adds noise to the input data
Ans. (1) It prevents overfitting

Q.75 What does the term "bias-variance tradeoff" refer to in machine learning?
(1) Balancing the model complexity and interpretability
(2) Balancing the bias and variance of a model for optimal performance
(3) Choosing between supervised and unsupervised learning
(4) Balancing the precision and recall in classification problems
Ans. (2) Balancing the bias and variance of a model for optimal performance

Q.76 What is the difference between precision and recall?
(1) Precision measures the ability to correctly predict positive instances, while recall measures the ability to correctly predict negative instances

(2) Precision measures the ability to correctly predict negative instances, while recall measures the ability to correctly predict positive instances.
(3) Precision and recall are the same thing.
(4) Precision and recall are not relevant in machine learning.

Ans. (1) Precision measures the ability to correctly predict positive instances, while recall measures the ability to correctly predict negative instances.

Q.77 What is the purpose of regularization in machine learning?
(1) To increase model complexity
(2) To decrease model complexity
(3) To eliminate bias in the model
(4) To speed up the training process

Ans. (2) To decrease model complexity

Q.78 Which of the following is a clustering algorithm?
(1) Random Forest (2) K-means
(3) Gradient Boosting (4) Decision Trees

Ans. (2) K-means

Q.79 What is the role of a learning rate in gradient descent optimization?
(1) To control the convergence speed
(2) To determine the number of iterations
(3) To define the size of the training dataset
(4) To set the threshold for model accuracy

Ans. (1) To control the convergence speed

Q.80 What is the purpose of the term "one-hot encoding" in machine learning?
(1) To encode numerical values
(2) To represent categorical variables as binary vectors
(3) To standardize data
(4) To normalize feature values

Ans. (2) To represent categorical variables as binary vectors

Q.81 What is the purpose of dropout layers in a neural network?
(1) To increase the learning rate
(2) To add noise to the input data
(3) To prevent overfitting
(4) To increase the number of neurons

Ans. (3) To prevent overfitting

Q.82 Which loss function is commonly used for binary classification problems in neural networks?
(1) Mean Squared Error (MSE)
(2) Cross-Entropy Loss
(3) Hinge Loss
(4) Huber Loss

Ans. (2) Cross-Entropy Loss

Q.83 What is the role of the learning rate in training a neural network?
(1) To control the convergence speed
(2) To determine the number of neurons
(3) To define the size of the training dataset
(4) To set the threshold for model accuracy

Ans. (1) To control the convergence speed

Q.84 What is the purpose of batch normalization in neural networks?
(1) To accelerate the training process
(2) To normalize the input features
(3) To prevent overfitting
(4) To apply the activation function

Ans. (2) To normalize the input features

Q.85 Which type of neural network is suitable for tasks like image recognition and object detection?
(1) Multilayer Perceptron (MLP)
(2) Recurrent Neural Network (RNN)
(3) Radial Basis Function Network (RBFN)
(4) Convolutional Neural Network (CNN)

Ans. (4) Convolutional Neural Network (CNN)

Q.86 What is the role of the output layer activation function in a neural network used for binary classification?
(1) Sigmoid function (2) Tanh function
(3) ReLU function (4) Softmax function

Ans. (1) Sigmoid function

Q.87 What is the purpose of the term "weight" in a neural network?
(1) It represents the spatial dimensions of the input
(2) It is the learning rate of the model
(3) It defines the strength of connections between neurons
(4) It represents the number of layers in the network

Ans. (3) It defines the strength of connections between neurons

Q.88 Which technique is used to address the exploding gradient problem in neural networks?
(1) Weight regularization
(2) Gradient clipping
(3) Learning rate annealing
(4) Dropout

Ans. (2) Gradient clipping

Q.89 What is the role of an optimizer in training a neural network?
(1) To define the architecture of the network
(2) To minimize the loss function and update weights
(3) To apply the activation function
(4) To initialize the weights randomly

Ans. (2) To minimize the loss function and update weights

Q.90 In a long short-term memory (LSTM) network, what is the purpose of the memory cell?
(1) To store the output of the network
(2) To store information for long periods
(3) To apply the activation function
(4) To control the learning rate
Ans (2) To store information for long periods

Q.91 What is the purpose of the term "epoch" in training a neural network?
(1) It represents the number of layers in the network
(2) It defines the learning rate
(3) It refers to one complete pass through the entire training dataset
(4) It is a measure of model accuracy
Ans. (3) It refers to one complete pass through the entire training dataset

Q.92 Which of the following is a type of unsupervised learning technique often used for dimensionality reduction in neural networks?
(1) K-means clustering
(2) Principal Component Analysis (PCA)
(3) Decision trees
(4) Support Vector Machines (SVM)
Ans (2) Principal Component Analysis (PCA)

Q.93 What is the purpose of the term "transfer learning" in neural networks?
(1) It refers to moving weights from one layer to another
(2) It involves transferring knowledge from one task to another
(3) It is the process of transferring data between neural networks
(4) It is a type of weight regularization technique
Ans. (2) It involves transferring knowledge from one task to another

Q.94 In a neural network, what is the role of the term "bias"?
(1) To add randomness to the model
(2) To balance the data distribution
(3) To account for errors in predictions
(4) To ensure the model fits the training data perfectly
Ans (3) It involves transferring knowledge from one task to another

Q.95 Which type of neural network architecture is designed for storing and recalling information over extended time periods?
(1) Convolutional Neural Network (CNN)
(2) Multilayer Perceptron (MLP)
(3) Radial Basis Function Network (RBFN)
(4) Recurrent Neural Network (RNN)
Ans (4) Recurrent Neural Network (RNN)

Q.96 What is the primary advantage of using a deep neural network compared to a shallow one?
(1) Reduced computational complexity
(2) Improved interpretability
(3) Ability to automatically extract hierarchical features
(4) Faster convergence speed
Ans (3) Ability to automatically extract hierarchical features

Q.97 In a neural network, what is the purpose of the term "activation"?
(1) To define the learning rate
(2) To apply the loss function
(3) To determine the number of neurons
(4) To introduce non-linearity to the model
Ans. (4) To introduce non-linearity to the model

Q.98 What is the role of the term "filter" in a convolutional neural network (CNN)?
(1) It is a type of activation function
(2) It represents the learning rate
(3) It defines the spatial dimensions of the input
(4) It is used to extract features from the input data
Ans (4) It is used to extract features from the input data

Q.99 What is the purpose of the term "hyperparameter" in a neural network?
(1) It is a parameter automatically learned by the model
(2) It is a parameter set prior to training
(3) It is a measure of model performance
(4) It is a type of regularization parameter
Ans. (2) It is a parameter set prior to training

Q.100 Which type of neural network is often used for solving regression problems?
(1) Convolutional Neural Network (CNN)
(2) Recurrent Neural Network (RNN)
(3) Radial Basis Function Network (RBFN)
(4) Multilayer Perceptron (MLP)
Ans. (4) Multilayer Perceptron (MLP)

❑❑

KPH for BCA

BCA

(SEM. VI) MODEL PAPER – III

BCA – 6004 : DATA SCIENCE AND MACHINE LEARNING

Time : 1 :30 Hours *Maximum M... 75*

Q.1. The applications of Data Science are
(1) Airline Route Planning
(2) Healthcare
(3) Fraud and Risk Detection
(4) All of the above
Ans. (4) All of the above

Q.2. A correct way to preprocess the data When performing regression or classification is
(1) Normalize the data → PCA → training
(2) PCA → normalize PCA output → training
(3) Normalize the data → PCA → normalize PCA output → training
(4) None of the above
Ans. (1) Normalize the data → PCA → training

Q.3. Which of the following uses data on some object to predict values for other object?
(1) Inferential (2) Predictive
(3) Exploratory (4) None of the above
Ans. (2) Predictive

Q.4. Point out the correct statement.
(1) Data has only qualitative value
(2) Data has both qualitative and quantitative value
(3) Data has only quantitative value
(4) None of the above
Ans. (2) Data has both qualitative and quantitative value

Q.5. Which of the following is the common goal of statistical modeling?
(1) Inference (2) Subsetting
(3) Summarizing (4) None of the above
Ans. (1) Inference

Q.6. Which of the following is a branch of statistics?
(1) Descriptive statistics (2) Industry statistics
(3) Inferential statistics (4) Both 1 and 3
Ans. (4) Both 1 and 3

Q.7. _____ is the basic data structure of pandas that can be thought of SQL table or a spreadsheet data representation.
(1) Series (2) List
(3) Dataframe (4) ndarray
Ans. (3) Dataframe

Q.8. Advantages of Data Science are
(1) Data Science is Versatile
(2) Abundance of Positions
(3) A Highly Paid Career
(4) All of the above
Ans. (4) All of the above

Q.9. Which of the following is another name for raw data?
(1) Destination data (2) Eggy data
(3) Machine learning (4) Secondary data
Ans. (2) Eggy data

Q.10. Disadvantages of Data Science are
(1) Arbitrary Data May Yield Unexpected Results
(2) Data Science is Blurry Term
(3) Large Amount of Domain Knowledge Required
(4) All of the above
Ans. (4) All of the above

Q.11. What is the primary goal of the data interpretation stage in data science?
(1) Preparing data for analysis
(2) Building machine learning models
(3) Communicating results and insights
(4) Defining the business problem
Ans. (3) Communicating results and insights

Q.12. In which stage of the data science process do data scientists use statistical techniques to extract patterns and insights from the data?
(1) Data Modeling
(2) Exploratory Data Analysis (EDA)
(3) Data Visualization
(4) Problem Definition
Ans. (2) Exploratory Data Analysis (EDA)

Q.13. Which stage involves deploying machine learning models into a production environment for practical use?
(1) Data Interpretation
(2) Model Deployment
(3) Data Cleaning
(4) Exploratory Data Analysis (EDA)
Ans. (2) Model Deployment

Q.14. What is a common challenge related to data security when implementing cloud-based solutions?
(1) Lack of data
(2) Increased control over data access
(3) Data encryption challenges
(4) Improved physical security
Ans. (3) Data encryption challenges

Q.15 Which type of cyber attack involves overwhelming a system with traffic, making it unavailable to users?
(1) Phishing
(2) Denial-of-Service (DoS)
(3) Malware
(4) Spoofing
Ans. (2) Denial-of-Service (DoS)

Q.16 What is the primary concern with data breaches in the context of data security?
(1) Loss of data integrity
(2) Unauthorized access and disclosure of sensitive information
(3) Slow data processing
(4) Lack of data variety
Ans. (2) Unauthorized access and disclosure of sensitive information

Q.17 What does the term "social engineering" refer to in the context of data security?
(1) Encrypting data during transmission
(2) Gaining unauthorized access to a system
(3) Tricking individuals into divulging confidential information
(4) Protecting data from physical damage
Ans. (3) Tricking individuals into divulging confidential information

Q.18 Which security measure involves restricting access to data based on user roles and responsibilities?
(1) Encryption (2) Access controls
(3) Firewalls (4) Data masking
Ans. (2) Access controls

Q.19 In healthcare, how can data science be applied to improve patient outcomes?
(1) Predicting disease outbreaks
(2) Enhancing medical imaging analysis
(3) Optimizing supply chain management
(4) Improving weather forecasting
Ans. (2) Enhancing medical imaging analysis

Q.20 What role does data science play in finance and banking?
(1) Predicting earthquakes
(2) Detecting fraudulent transactions
(3) Analyzing plant growth patterns
(4) Optimizing shipping routes
Ans. (2) Detecting fraudulent transactions

Q.21 How can data science be applied in e-commerce?
(1) Analyzing climate data
(2) Recommending products to users
(3) Studying ancient civilizations
(4) Controlling traffic signals
Ans. (2) Recommending products to users

Q.22 What is a common application of data science in marketing?
(1) Studying ocean currents
(2) Predicting volcanic eruptions
(3) Personalizing advertising campaigns
(4) Analyzing outer space phenomena
Ans. (3) Personalizing advertising campaigns

Q.23 How can data science contribute to urban planning and development?
(1) Analyzing DNA sequences
(2) Predicting stock market trends
(3) Optimizing public transportation routes
(4) Studying marine life
Ans. (3) Optimizing public transportation routes

Q.24 In education, how can data science be used to enhance learning experiences?
(1) Analyzing geological formations
(2) Personalizing learning materials
(3) Forecasting tsunamis
(4) Monitoring bird migrations
Ans. (2) Personalizing learning materials

Q.25 What is a key application of data science in the energy sector?
(1) Analyzing lunar landscapes
(2) Optimizing oil drilling operations
(3) Studying ancient civilizations
(4) Predicting forest fires
Ans. (2) Optimizing oil drilling operations

Q.26 How can data science be applied in the field of transportation?
(1) Analyzing weather patterns
(2) Predicting earthquake aftershocks
(3) Optimizing traffic flow and reducing congestion
(4) Studying the human genome
Ans. (3) Optimizing traffic flow and reducing congestion

Q.27 In agriculture, what is a common application of data science?
(1) Analyzing social media trends
(2) Predicting solar flares
(3) Monitoring crop health with remote sensing
(4) Analyzing human migration patterns
Ans. (3) Monitoring crop health with remote sensing

Q.28 How can data science contribute to environmental conservation?
(1) Analyzing stock market trends
(2) Predicting volcanic eruptions
(3) Monitoring wildlife populations and habitats
(4) Studying ancient civilizations
Ans. (3) Monitoring wildlife populations and habitats

Q.29 Which data security measure involves regularly updating software and systems to patch vulnerabilities?
(1) Firewall protection
(2) Intrusion Detection System (IDS)
(3) Penetration testing
(4) Software patching
Ans. (4) Software patching

Q.30 To tell python that we want to activate the mean function from the numpy package, we write _____ it front of the mean.
(1) npm (2) np
(3) ng (4) ngm
Ans. (2) np

Q.31 What is the primary purpose of handling inconsistent data in data cleaning?
(1) To simplify the data analysis process
(2) To create new features
(3) To improve data visualization
(4) To ensure data accuracy and reliability
Ans. (4) To ensure data accuracy and reliability

Q.32 Which data cleaning step involves standardizing units of measurement to ensure consistency?
(1) Error correction (2) Data imputation
(3) Unit conversion (4) Outlier detection
Ans. (3) Unit conversion

Q.33 Why is addressing duplicate data important in data cleaning?
(1) To remove irrelevant data
(2) To create new features
(3) To simplify the data analysis process
(4) To ensure accurate and unbiased analysis
Ans. (4) To ensure accurate and unbiased analysis

Q.34 In data cleaning, what does the term "inconsistency" refer to?
(1) Duplicated records (2) Errors in data entry
(3) Outliers in the dataset (4) Missing values
Ans. (2) Errors in data entry

Q.35 What is the primary goal of data reduction in the context of data science?
(1) To increase data storage capacity
(2) To transform raw data into a usable format
(3) To simplify the data analysis process
(4) To guarantee data privacy
Ans. (3) To simplify the data analysis process

Q.36 Which technique involves selecting a subset of relevant features to reduce the dimensionality of the dataset?
(1) Feature scaling
(2) Principal Component Analysis (PCA)

(3) Data imputation
(4) Outlier detection
Ans. (2) Principal Component Analysis (PCA)

Q.37 What is the purpose of aggregation in data reduction?
(1) To handle missing data
(2) To simplify the data analysis process
(3) To remove irrelevant data
(4) To create new features
Ans. (2) To simplify the data analysis process

Q.38 Which data reduction method involves combining similar data points into representative clusters?
(1) Principal Component Analysis (PCA)
(2) Clustering
(3) Feature scaling
(4) Outlier detection
Ans. (2) Clustering

Q.39 What is the role of binning in data reduction?
(1) To handle missing data
(2) To remove irrelevant data
(3) To simplify the data analysis process
(4) To group continuous numerical data into discrete intervals
Ans. (4) To group continuous numerical data into discrete intervals

Q.40 In feature selection for data reduction, what criterion is often used to evaluate the importance of features?
(1) Mean imputation (2) Entropy
(3) Data imputation (4) Outlier detection
Ans. (2) Entropy

Q.41 What is the primary advantage of using dimensionality reduction techniques like t-SNE or UMAP?
(1) Creating new features
(2) Simplifying the data analysis process
(3) Handling missing data
(4) Guaranteeing data privacy
Ans. (2) Simplifying the data analysis process

Q.42 What does the term "binning" refer to in the context of data reduction?
(1) Grouping similar data points into clusters
(2) Combining similar features into representative bins
(3) Handling missing data
(4) Removing irrelevant data
Ans. (2) Combining similar features into representative bins

Q.43 Which method involves transforming numerical features into a lower-dimensional representation while preserving the variance in the data?
(1) Clustering

(2) Binning
(3) Principal Component Analysis (PCA)
(4) Feature scaling

Ans. (3) Principal Component Analysis (PCA)

Q.44. What is the primary advantage of data reduction techniques in the context of machine learning?
(1) Increasing data storage capacity
(2) Improving data visualization
(3) Enhancing model interpretability and efficiency
(4) Guaranteeing data privacy

Ans. (3) Enhancing model interpretability and efficiency

Q.45. What does the box in a box plot represent?
(1) Interquartile Range (IQR)
(2) Median
(3) Range
(4) Mean

Ans. (1) Interquartile Range (IQR)

Q.46. In a box plot, what do "whiskers" represent?
(1) Range of values
(2) Standard Deviation
(3) Extremes or outliers
(4) Interquartile Range (IQR)

Ans. (3) Extremes or outliers

Q.47. How is the median represented in a box plot?
(1) A dot
(2) The central line inside the box
(3) The ends of the whiskers
(4) The center of the box

Ans. (2) The central line inside the box

Q.48. What information does a box plot provide about the spread of data?
(1) Range
(2) Variance
(3) Interquartile Range (IQR)
(4) Standard Deviation

Ans. (3) Interquartile Range (IQR)

Q.49. In a box plot, how are outliers typically identified?
(1) Values outside the whiskers
(2) Values inside the box
(3) Values at the median
(4) Values at the ends of the box

Ans. (1) Values outside the whiskers

Q.50. What does a longer box in a box plot indicate about the data?
(1) Greater spread (2) Smaller spread
(3) Symmetric distribution (4) Normal distribution

Ans. (1) Greater spread

Q.51. What is the purpose of the "notches" in some box plots?
(1) Indicate the mean
(2) Represent the range of values
(3) Identify outliers
(4) Provide an estimate of the median confidence interval

Ans. (4) Provide an estimate of the median confidence interval

Q.52. In a box plot, what does a skew to the right indicate?
(1) Positive skewness
(2) Negative skewness
(3) Symmetric distribution
(4) Mesokurtic distribution

Ans. (1) Positive skewness

Q.53. What is the key advantage of using a box plot for data visualization?
(1) It provides a detailed view of individual data points.
(2) It shows the full distribution of data
(3) It focuses on central tendency
(4) It is suitable only for small datasets

Ans. (2) It shows the full distribution of data

Q.54. How are multiple box plots typically used for comparison?
(1) Stacked vertically
(2) Stacked horizontally
(3) Overlaid in the same plot
(4) Only one box plot can be displayed at a time

Ans. (3) Overlaid in the same plot

Q.55. What is the primary purpose of a pivot table in data analysis?
(1) Summarizing and analyzing data
(2) Creating charts and graphs
(3) Sorting data alphabetically
(4) Importing external data

Ans. (1) Summarizing and analyzing data

Q.56. In a pivot table, what is the "values" field used for?
(1) Row labels (2) Column labels
(3) Data aggregation (4) Filter criteria

Ans. (3) Data aggregation

Q.57. What does the "row labels" area represent in a pivot table?
(1) The vertical axis of a chart
(2) The summarized data
(3) Categories or groups for rows
(4) The calculated fields

Ans. (3) Categories or groups for rows

Q.58. In Excel, how can you change the summary function for a field in a pivot table?
(1) It cannot be changed

(2) Right-click on the field and select "Change Function"
(3) Use a formula in the cell next to the pivot table
(4) Drag the field to a different area in the pivot table

Ans. (2) Right-click on the field and select "Change Function"

Q.59 What is the purpose of the "columns" area in a pivot table?
(1) Displaying row labels
(2) Sorting data alphabetically
(3) Aggregating data vertically
(4) Creating a multi-level hierarchy

Ans. (4) Creating a multi-level hierarchy

Q.60 How can you filter data in a pivot table to show only specific information?
(1) Use a slicer
(2) Change the font color
(3) Manually delete unwanted rows
(4) Copy and paste the desired data

Ans. (1) Use a slicer

Q.61 What is the purpose of the "report filter" in a pivot table?
(1) Sorting data
(2) Filtering data by criteria
(3) Changing the font style
(4) Calculating percentages

Ans. (2) Filtering data by criteria

Q.62 In a pivot table, what does the term "drill down" mean?
(1) Creating a new pivot table
(2) Expanding detailed data from a summary
(3) Reducing the size of the pivot table
(4) Renaming field labels

Ans. (2) Expanding detailed data from a summary

Q.63 What Excel feature is often used in combination with pivot tables for enhanced data visualization?
(1) Conditional formatting (2) Spell check
(3) Hyperlinks (4) WordArt

Ans. (1) Conditional formatting

Q.64 What is the curse of dimensionality in machine learning?
(1) The high cost of computing power
(2) The difficulty of handling large datasets
(3) The increased complexity with higher dimensions
(4) The limitation of feature selection algorithms

Ans. (3) The increased complexity with higher dimensions

Q.65 Which evaluation metric is commonly used for classification problems?
(1) Mean Squared Error (MSE)
(2) F1 score
(3) R-squared
(4) Root Mean Squared Error (RMSE)

Ans. (2) F1 score

Q.66 What is over fitting in machine learning?
(1) When a model generalizes well to new data
(2) When a model performs poorly on the training data
(3) When a model is too complex and fits noise in the data
(4) When a model is too simple and under fits the data

Ans. (3) When a model is too complex and fits noise in the data

Q.67 Which algorithm is used for reinforcement learning?
(1) Naive Bayes
(2) Support Vector Machines (SVM)
(3) Q-learning
(4) Random Forest

Ans. (3) Q-learning

Q.68 What is the role of a hyper parameter in machine learning?
(1) It is a type of algorithm
(2) It is a parameter automatically learned by the model
(3) It is a parameter set prior to training
(4) It is a measure of model performance

Ans. (3) It is a parameter set prior to training

Q.69 What is a confusion matrix used for in machine learning?
(1) Visualizing data distributions
(2) Evaluating classification model performance
(3) Generating random numbers
(4) Clustering data points

Ans. (2) Evaluating classification model performance

Q.70 Which of the following is an example of a semi-supervised learning algorithm?
(1) K-nearest neighbors
(2) Principal Component Analysis (PCA)
(3) Support Vector Machines (SVM)
(4) Label Propagation

Ans. (4) Label Propagation

Q.71 What is the primary difference between classification and regression tasks in machine learning?
(1) Classification involves predicting categories, while regression involves predicting numerical values.
(2) Classification involves predicting numerical values, while regression involves predicting categories.

(3) Classification and regression are interchangeable terms
(4) Classification and regression are unrelated concepts

Ans. (1) Classification involves predicting categories, while regression involves predicting numerical values.

Q.72 Which type of machine learning algorithm is well-suited for handling sequential data, such as time series?
(1) Decision trees
(2) Support Vector Machines (SVM)
(3) Recurrent Neural Networks (RNN)
(4) K-means clustering

Ans. (3) Recurrent Neural Networks (RNN)

Q.73 In reinforcement learning, what does the term "reward" represent?
(1) A penalty for incorrect actions
(2) The cost of the training dataset
(3) The output of the neural network
(4) Positive feedback for desirable actions

Ans. (4) Positive feedback for desirable actions

Q.74 What is the role of the term "kernel" in Support Vector Machines (SVM)?
(1) It is a data preprocessing technique.
(2) It is a hyper parameter that controls model complexity.
(3) It defines the decision boundary between classes.
(4) It is an activation function in neural networks.

Ans. (3) It defines the decision boundary between classes.

Q.75 What does the acronym "PCA" stand for in the context of machine learning?
(1) Principal Component Analysis
(2) Predictive Classification Algorithm
(3) Primary Cluster Aggregation
(4) Parametric Curve Adjustment

Ans. (1) Principal Component Analysis

Q.76 Which method is commonly used for handling missing data in machine learning?
(1) Deleting rows with missing values
(2) Replacing missing values with the mean or median
(3) Ignoring missing values during training
(4) All of the above

Ans. (4) All of the above

Q.77 What is the purpose of a learning curve in machine learning?
(1) To visualize the relationship between features
(2) To assess model performance over training iterations
(3) To represent the decision boundary in a classification problem
(4) To create a roadmap for feature engineering

Ans. (2) To assess model performance over training iterations

Q.78 What is the role of the term "batch size" in training neural networks?
(1) It represents the size of the input features.
(2) It determines the number of iterations in training.
(3) It controls the number of samples processed in each update.
(4) It is a regularization parameter.

Ans. (3) It controls the number of samples processed in each update.

Q.79 Which of the following is a type of ensemble learning algorithm that builds multiple models sequentially, where each model corrects the errors of the previous one?
(1) Bagging (2) Boosting
(3) Stacking (4) Random Forest

Ans. (2) Boosting

Q.80 What is the main advantage of using a deep neural network?
(1) Reduced computational complexity
(2) Increased interpretability
(3) Ability to automatically extract features
(4) Faster training time

Ans. (3) Ability to automatically extract features

Q.81 What is the primary function of the softmax activation function in the output layer of a neural network?
(1) To introduce non-linearity
(2) To calculate the loss function
(3) To normalize the output into probability distributions
(4) To speed up the training process

Ans. (3) To normalize the output into probability distributions

Q.82 Which technique is commonly used to address the vanishing gradient problem in recurrent neural networks (RNNs)?
(1) Weight regularization
(2) Gradient clipping
(3) Learning rate annealing
(4) Dropout

Ans. (2) Gradient clipping

Q.83 What is the purpose of the term "early stopping" in training a neural network?
(1) To stop training when the model is overfitting
(2) To stop training when the learning rate is too high
(3) To stop training when the loss function is too low
(4) To stop training when the model is underfitting

Ans. (1) To stop training when the model is overfitting

Q.84 Which type of neural network is designed to approximate complex non-linear functions?
(1) Multilayer Perceptron (MLP)
(2) Radial Basis Function Network (RBFN)
(3) Convolutional Neural Network (CNN)
(4) Recurrent Neural Network (RNN)
Ans. (1) Multilayer Perceptron (MLP)

Q.85 What is the primary purpose of the term "dropout" in a neural network?
(1) To add noise to the input data
(2) To prevent over fitting by randomly dropping neurons during training
(3) To increase the learning rate
(4) To normalize feature values
Ans. (2) To prevent over fitting by randomly dropping neurons during training

Q.86 Which activation function is commonly used in the hidden layers of a neural network?
(1) Sigmoid function (2) Tanh function
(3) ReLU function (4) Softmax function
Ans. (3) ReLU function

Q.87 What is the purpose of the term "momentum" in the context of neural network optimization algorithms?
(1) To slow down the convergence speed
(2) To add randomness to the model
(3) To accelerate the convergence by adding a fraction of the previous update
(4) To define the size of the training dataset
Ans. (3) To accelerate the convergence by adding a fraction of the previous update

Q.88 Which type of neural network is designed for handling and processing memory over long sequences?
(1) Convolutional Neural Network (CNN)
(2) Recurrent Neural Network (RNN)
(3) Radial Basis Function Network (RBFN)
(4) Multilayer Perceptron (MLP)
Ans. (2) Recurrent Neural Network (RNN)

Q.89 What is the purpose of the term "kernel" in a convolutional neural network (CNN)?
(1) It is a data preprocessing technique.
(2) It is a hyper parameter that controls model complexity.
(3) It defines the decision boundary between classes.

(4) It is used for convolution operations on the input.
Ans. (4) It is used for convolution operations on the input.

Q.90 In a neural network, what is the primary role of the loss function?
(1) To define the learning rate
(2) To introduce non-linearity
(3) To measure the difference between predicted and actual values
(4) To determine the number of layers in the network
Ans. (3) To measure the difference between predicted and actual values

Q.91 Which of the following is one of the key data science skills?
(1) Statistics (2) Machine Learning
(3) Data Visualization (4) All of the above
Ans. (4) All of the above

Q.92 Which of the following is a key characteristic of a hacker?
(1) Afraid to say they don't know the answer
(2) Willing to find answers on their own
(3) Not Willing to find answers on their own
(4) All of the above
Ans. (2) Willing to find answers on their own

Q.93 Raw data should be processed only one time.
(1) True (2) False
(3) Can be true or false (4) Can not say
Ans. (2) False

Q.94 Unstructured data is not organized.
(1) True (2) False
(3) Can be true or false (4) Can not say
Ans. (1) True

Q.95 Which of the following are correct component for data science?
(1) Data Engineering (2) Advanced computing
(3) Domain expertise (4) All of the above
Ans. (4) All of the above

Q.96 Causal analysis is commonly applied to census data.
(1) True (2) False
(3) Can be true or false (4) Can not say
Ans. (2) False

Q.97 Which of the following model is usually a gold standard for data analysis?
(1) Inferential (2) Descriptive
(3) Causal (4) All of the above
Ans. (3) Causal

Q.98.Which of the following is a revision control system?
(1) Git (2) Numpy
(3) Scipy (4) Slidify
Ans. (1) Git

Q.99.Which of the following focuses on the discovery of (previou
unknown properties on the data?
(1) Data mining (2) BigData
(3) Data wrangling (4) Machine Learning
Ans. (1) Data mining

Q.100.Data can be categorized into _____ groups.
(1) 1 (2) 2 (3) 3 (4) 4
Ans. (2) 2