Practice_Test1

1. Which of the following best describes Artificial Intelligence?

(A) Programming machines to perform repetitive tasks

(B) Enabling machines to mimic human-like perception, reasoning, and learning

(C) Using large datasets for statistical computation

(D) Developing complex mathematical algorithms only

2. The rational agent approach in AI focuses on:

(A) Following human emotions

(B) Acting in a way that maximizes expected performance measure

(C) Minimizing energy consumption

(D) Performing all possible actions simultaneously

3. Strong AI refers to systems that:

(A) Simulate human reasoning but lack consciousness

(B) Are pre-programmed with expert knowledge

(C) Exhibit genuine intelligence and self-awareness

(D) Are limited to rule-based systems

4. Which component of AI is responsible for drawing conclusions from data?

(A) Learning

(B) Reasoning

(C) Perception

(D) Knowledge representation

5. Expert systems fall under which category of AI techniques?

(A) Reactive

(B) Cognitive

(C) Knowledge-based

(D) Neural-based

6. Which of the following is not a type of learning in Machine Learning?

(A) Supervised Learning

(B) Reinforcement Learning

(C) Associative Learning

(D) Unsupervised Learning

7. A model that continuously improves its performance through feedback is using:

(A) Supervised Learning

(B) Reinforcement Learning

(C) Unsupervised Learning

(D) Semi-supervised Learning

8. The bias-variance tradeoff in ML refers to:

(A) Balancing training and test data

(B) Balancing underfitting and overfitting

(C) Balancing precision and recall

(D) Balancing classification and regression

9. The first stage in machine learning model development is:

(A) Model tuning

(B) Data preprocessing

(C) Feature engineering

(D) Evaluation

10. Which of the following is a real-world application of supervised learning?

(A) Market basket analysis

(B) Clustering customers

(C) Spam detection

(D) Topic modeling

11. The 5 Vs of Big Data include all except:

(A) Volume

(B) Velocity

(C) Variety

(D) Visualization

12. Which V of Big Data refers to data consistency and reliability?

(A) Variety

(B) Veracity

(C) Velocity

(D) Volume

13. The major influence of Big Data on AI is:

(A) Reducing model complexity

(B) Enabling better feature selection through vast datasets

(C) Making algorithms less interpretable

(D) Increasing bias in models

14. Data Science is best defined as:

(A) A subset of AI focusing on computer vision

(B) Extracting knowledge and insights from data using scientific methods

(C) Creating dashboards from data

(D) Building neural networks only

15. In a Data Science project pipeline, model evaluation comes after:

(A) Model deployment

(B) Data collection

(C) Model training

(D) Feature selection

16. A major application of Data Science in business is:

(A) Predictive analytics

(B) Static website generation

(C) File compression

(D) Encryption

17. The aggregate() function in Pandas

(A) Applies a single function to all columns

(B) Allows multiple aggregations per column

(C) Works only on numeric data

(D) Replaces missing values

18. The pivot() function in Pandas is mainly used to

(A) Transform rows into columns

(B) Combine multiple DataFrames

(C) Remove duplicates

(D) Group by values

19. The melt() function in Pandas performs the reverse of

(A) concat()

(B) join()

(C) pivot()

(D) groupby()

20. In Pandas, map() and applymap() differ in that

(A) map() works on DataFrames, applymap() on Series

(B) map() works on Series, applymap() on DataFrames

(C) both are identical

(D) applymap() cannot use lambda functions

21. A reduce operation on a DataFrame aims to

(A) Filter rows

(B) Aggregate multiple elements into a single value

(C) Create pivot tables

(D) Transform categorical data

22. To select all data from June month every year in a time series index

(A) df['June']

(B) df[df.index.month == 6]

(C) df['month'==6]

(D) df.query("month==6")

23. The need for data visualization is primarily to

(A) Reduce data redundancy

(B) Simplify data cleaning

(C) Communicate insights effectively

(D) Increase data storage

24. In Matplotlib, the anatomy of a figure includes all except

(A) Axes

(B) Title

(C) Legend

(D) Layout manager

25. Which Seaborn plot is most suitable for visualizing correlations

(A) Bar plot

(B) Heatmap

(C) Countplot

(D) Violin plot

26. Cross-sectional data refers to

(A) Data collected over time

(B) Data collected at a single point in time across entities

(C) Data with missing values

(D) Aggregated temporal data

27. Time series data differs from cross-sectional data mainly by

(A) Variable type

(B) Dependency over time

(C) Missing attributes

(D) Data format

28. The type of data representing categories with intrinsic order is

(A) Nominal

(B) Ordinal

(C) Interval

(D) Ratio

29. Ratio data differs from interval data by having

(A) Equal intervals

(B) True zero point

(C) Categorical values

(D) Constant variance

30. A major vulnerability of mean as a measure of central tendency is

(A) It ignores sample size

(B) It is affected by extreme values

(C) It is non-unique

(D) It cannot be computed on ratio data

31. The variance is more robust than the mean for detecting

(A) Skewness

(B) Outliers

(C) Central tendency

(D) Frequency

32. Which NumPy function computes the mean along a specified axis

(A) np.aggregate()

(B) np.mean()

(C) np.axis_mean()

(D) np.vectorize_mean()

33. The vectorized operations in NumPy are efficient because

(A) They use loops written in Python

(B) They are implemented in C and use broadcasting

(C) They store intermediate results in memory

(D) They use recursion

34. Which function is used to apply custom lambda logic element-wise to arrays

(A) np.vectorize()

(B) np.map()

(C) np.apply()

(D) np.lambda_apply()

35. A Type I error occurs when:

(A) A true null hypothesis is rejected

(B) A false null hypothesis is accepted

(C) Both hypotheses are rejected

(D) None of the above

36. The significance level (α) represents:

(A) Probability of Type II error

(B) Probability of rejecting true null hypothesis

(C) Power of the test

(D) Confidence interval

37. The p-value smaller than α indicates:

(A) Fail to reject null

(B) Accept null

(C) Reject null

(D) Insufficient evidence

38. The primary goal of EDA is to:

(A) Build predictive models

(B) Explore patterns, anomalies, and relationships in data

(C) Automate hypothesis testing

(D) Conduct feature scaling

39. Which plot is most effective for detecting outliers?

(A) Bar plot

(B) Box plot

(C) Line plot

(D) Histogram

40. Correlation matrix is used in EDA to:

(A) Identify missing values

(B) Measure linear relationships between features

(C) Detect non-linear patterns

(D) Normalize features

41. A right-skewed distribution indicates:

(A) Mean > Median

(B) Mean < Median

(C) Symmetrical shape

(D) No outliers

42. Feature engineering typically follows EDA because:

(A) EDA defines target variables

(B) EDA identifies feature relevance and relationships

(C) EDA eliminates class imbalance

(D) EDA generates labels automatically

43. In EDA, handling missing values using median instead of mean is preferred when:

(A) Data is normally distributed

(B) Data is skewed

(C) Data is categorical

(D) Data is binary

44. If your AI model performance improves only when data volume increases significantly, the issue likely lies in:

(A) Algorithm design

(B) Underfitting due to small dataset

(C) Data leakage

(D) Improper validation strategy

45. When using Pandas melt, the id_vars parameter is used to:

(A) Specify columns to unpivot

(B) Keep certain columns fixed

(C) Define index levels

(D) Remove duplicates

46. A low p-value with high R^2 indicates:

(A) Strong correlation but poor model

(B) Statistically significant and good fit

(C) Overfitted model

(D) Model bias

47. In Big Data pipelines, data veracity challenges are addressed using:

(A) Data duplication

(B) Data cleaning and validation

(C) Increasing sample size

(D) Compression techniques

48. The reduce() function in Python can be best described as:

(A) Sequentially combining elements using a function

(B) Removing duplicates

(C) Mapping elements to keys

(D) Filtering arrays

49. In hypothesis testing, increasing the sample size generally:

(A) Increases Type I error

(B) Decreases power

(C) Increases test sensitivity

(D) Makes test unreliable

50. The most appropriate EDA tool to identify feature correlation before ML model training is:

(A) Box plot

(B) Heatmap

(C) Pie chart

(D) Scatter matrix

51. A hospital implements an AI system that predicts disease risk using past patient data. The system improves with every new patient record.
Which AI technique is this system most likely using?

(A) Rule-based reasoning

(B) Machine Learning

(C) Natural Language Processing

(D) Expert Systems

52. A chatbot in an insurance company understands customer questions, retrieves information, and replies in natural language.
Which AI components are primarily at work?

(A) Vision and Motion

(B) Learning and Perception

(C) NLP and Knowledge Representation

(D) Planning and Robotics

53. A data scientist builds a regression model that performs well on training data but poorly on test data.
Which problem does this indicate?

(A) Underfitting

(B) Overfitting

(C) Regularization

(D) Cross-validation error

54. A retail company stores customer transactions from multiple stores across countries, generating terabytes of data daily.
Which Vs of Big Data are most relevant?

(A) Variety and Veracity

(B) Volume and Velocity

(C) Velocity and Veracity

(D) Volume and Visualization

55. An AI model predicting loan default improved after integrating massive transaction datasets.
What aspect of Big Data improved the AI performance?

(A) Data Veracity

(B) Data Variety

(C) Data Volume

(D) Data Velocity

56. A recommendation system suggests movies based on users with similar preferences.
Which ML method is being applied?

(A) Classification

(B) Clustering

(C) Reinforcement Learning

(D) Collaborative Filtering

57. An autonomous drone learns to land smoothly by adjusting based on success rewards.
Which type of learning is this?

(A) Supervised

(B) Reinforcement

(C) Unsupervised

(D) Semi-supervised

58. An e-commerce platform uses analytics to segment customers based on spending habits and predict purchase behavior.
This scenario represents:

(A) Data Visualization

(B) Predictive Analytics

(C) Data Warehousing

(D) Prescriptive Modeling

59. You have a NumPy array of monthly sales. You need the total and average sales per quarter.
Which function combination is most appropriate?

(A) np.mean(), np.sum()

(B) np.average(), np.min()

(C) np.median(), np.percentile()

(D) np.std(), np.var()

60. You want to calculate profit = revenue - cost for arrays of 1M elements each.
Which approach ensures fastest computation?

(A) For loop subtraction

(B) Map() with lambda

(C) Vectorized subtraction using NumPy

(D) List comprehension

61. A dataset contains columns: *Region, Month, Sales, Profit*.

You want the **average profit per region**.
Which operation should you apply?

(A) df.aggregate('mean')

(B) df.groupby('Region')['Profit'].mean()

(C) df.pivot_table(index='Month', values='Profit')

(D) df.melt('Profit')

62. A DataFrame stores monthly sales per region. You want to display **regions as columns** and **months as rows**.
Which Pandas function helps restructure data this way?

(A) groupby()

(B) melt()

(C) pivot()

(D) transpose()

63. You have a Series of product reviews. You want to apply sentiment scoring to each review.

Which is the most efficient approach?

(A) Use map() with a lambda function returning sentiment score

(B) Use reduce() to combine all reviews

(C) Use groupby()

(D) Use melt()

64. A company tracks daily sales data for 10 years. You want to compute **average sales in June across all years**.

Which Pandas filter works best?

(A) a) df[df.index.month == 6].mean()

(B) b) df.query('month == June')

(C) c) df[df.Month == 'June']

(D) d) df.groupby('Month').mean()

65. You are analyzing product sales distribution to check for outliers.

Which visualization is most suitable?

(A) a) Line chart

(B) b) Box plot

(C) c) Heatmap

(D) d) Scatter plot

66. In a heatmap of correlation between numerical features, the cell between *Price* and *Discount* shows **−0.9**.

What can be inferred?

(A) a) High price increases discount

(B) b) Higher discount corresponds to lower price

(C) c) Price and discount are independent

(D) d) Data is not correlated

67. A survey records satisfaction levels as *Poor, Average, Good, Excellent*.

Which data type is appropriate?

(A) a) Nominal

(B) b) Ordinal

(C) c) Interval

(D) d) Ratio

68. A dataset of salaries contains one very high executive salary that distorts the average.

Which measure should be used instead of mean?

(A) a) Variance

(B) b) Median

(C) c) Standard deviation

(D) d) Mode

69. A/B testing compares two webpage designs. The p-value is **0.03** and significance level α = **0.05**.

What should you conclude?

(A) a) Fail to reject null hypothesis

(B) b) Reject null hypothesis — difference is significant

(C) c) Increase sample size

(D) d) Accept null hypothesis

70. If a researcher concludes there is a difference when in fact there isn't, it is:

(A) a) Type I error

(B) b) Type II error

(C) c) Sampling bias

(D) d) Regression error