Unit 4: Data and Analysis

72 MCQs. Correct option in bold. 1 mark each.

1. Data science is an interdisciplinary field of mathematics, statistics and
A. physics **B. computer science**
C. history D. art

2. Data science can work with
A. only unstructured data B. only structured data
C. structured as well as unstructured data D. no data

3. Predicting a disease from X-rays is a use of data science in
A. healthcare B. sports
C. education D. entertainment

4. Fraud detection in banks is a use of data science in
A. airline B. education C. sports **D. finance**

5. A recommender system is a tool that
A. repairs hardware
B. draws pictures
C. recommends products, services or information based on past behaviour and interests
D. stores passwords

6. Sentiment analysis identifies
A. positive, negative or neutral sentiments in reviews
B. viruses
C. faulty hardware
D. network speed

7. In 1950, the Turing Test was proposed by
A. Tim Berners-Lee B. Charles Babbage
C. Bill Gates **D. Alan Turing**

8. The Turing Test measures
A. a CPU's speed
B. a machine's ability to exhibit intelligent behaviour
C. memory size
D. network bandwidth

9. Artificial Intelligence is the ability of a machine to
A. exhibit human-like behaviour such as problem solving and understanding language
B. connect to the internet only
C. store data
D. print documents

10. NLP stands for
A. Network Link Protocol B. Natural Logic Processing
C. New Language Program **D. Natural Language Processing**

11. ChatGPT and chatbots are examples of
A. databases B. computer vision
C. NLP D. robotics only

12. The face recognition feature in smartphones uses
A. spreadsheets **B. computer vision**
C. NLP D. reinforcement only

13. Which AI field enables smartphones to recognise faces and unlock devices?
A. Neural networks only B. Deep learning only
C. NLP **D. Computer vision**

14. Automation industry uses computers and robots to perform tasks
A. manually B. only with humans
C. automatically without human intervention D. slowly

15. The entire scope of data science and AI is based on algorithms developed in
A. machine learning B. graphics
C. word processing D. networking

16. Automated fraud detection is an example of
A. manual auditing **B. machine learning**
C. photo editing D. word processing

17. Machine learning is a subset of
A. web design B. databases
C. artificial intelligence D. networking

18. Which framework is important in machine learning?
A. WordPad B. Excel
C. Photoshop **D. TensorFlow**

19. Deep learning uses layers of
A. artificial neural networks B. databases
C. web pages D. spreadsheets

20. A neural network is a system designed to work like
A. a calculator **B. the human brain**
C. a router D. a printer

21. The three models of machine learning are supervised, unsupervised and
A. manual B. static
C. deep **D. reinforcement**

22. In supervised learning, the machine is given
A. no data B. random noise
C. labelled data D. unlabelled data

23. Labelled data means
A. questions provided with their correct answers
B. questions without answers
C. random numbers
D. blank files

24. In unsupervised learning, the machine is given
A. rewards only **B. unlabelled data**
C. labelled data D. penalties only

25. Unsupervised learning identifies patterns and divides data into
A. groups based on similarities and differences
B. penalties
C. labelled answers
D. rewards

26. In reinforcement learning, machines are trained using
A. no feedback B. unlabelled data only
C. labelled data only **D. positive or negative feedback**

27. The key terms of reinforcement learning are agent, environment, actions, rewards and
A. labels **B. penalties** C. clusters D. tables

28. The primary function of rewards and penalties in reinforcement learning is
A. classifying outcomes B. labelling data
C. improving action choices D. evaluating performance only

29. Learning to play chess by winning and losing moves is an example of
A. none of these **B. reinforcement learning**
C. unsupervised learning D. supervised learning

30. Training a toy robot to find its way out of a maze is
A. unsupervised learning B. supervised learning
C. none of these **D. reinforcement learning**

31. Sorting a basket of apples and bananas using given examples is
A. supervised learning B. unsupervised learning
C. reinforcement learning D. none of these

32. Grouping Lego blocks by colour without labels is
A. none of these B. supervised learning
C. unsupervised learning D. reinforcement learning

33. Categorising customer feedback into positive, negative or neutral (with labelled examples) suits
A. reinforcement learning B. unsupervised learning
C. deep learning only **D. supervised learning**

34. If an agent must achieve a goal, we use
A. unsupervised learning **B. reinforcement learning**
C. none D. supervised learning

35. In the machine learning process, cleaned data is first split into
A. training data and testing data B. text and images
C. files and folders D. rows and columns

36. The model learns patterns in the
A. testing set only B. hardware
C. training set D. output

37. The output of artificial intelligence systems is to
A. produce reports only **B. mimic human behaviour**
C. produce charts only D. store data

38. The output of machine learning is
A. prediction and classification B. only charts
C. mimicking humans D. insight and reports

39. The output of data science is
A. prediction only B. voice only
C. robots **D. insight, reports and visualization**

40. Churn prediction finds the possibility that
A. a server will fail
B. a product is cheap
C. a customer will stop using a company's services or products
D. a customer will buy more

41. Keeping existing customers loyal is usually
A. impossible B. less beneficial than getting new ones
C. useless **D. more beneficial than getting new ones**

42. Behavioral segmentation divides customers into groups based on their
A. behaviour B. name C. age only D. city only

43. Data visualization is
A. deleting data
B. the graphical representation of data and information
C. storage of data
D. encrypting data

44. A bar chart is used to
A. show correlation B. show proportions of a whole
C. compare discrete values D. show trends over time

45. A line chart is used to display
A. distributions of one value B. matrix colours
C. proportions of a whole **D. trends of data over time**

46. A pie chart represents
A. proportions of categories in a whole B. trends over time
C. distribution D. correlation

47. A scatter plot displays
A. trends over years
B. the relationship between numerical variables
C. matrix intensity
D. proportions

48. A histogram shows
A. trends over time
B. proportions
C. the distribution of a single numerical value
D. relationship between two variables

49. A box plot summarises data using
A. minimum, first quartile, median, third quartile and maximum
B. mode only
C. colours
D. mean only

50. A box plot is also called
A. bubble plot B. bar plot
C. line plot **D. box-and-whisker plot**

51. A heatmap represents data in matrix form using
A. pie slices B. bar height
C. colour intensity D. line slopes

52. A bubble chart is similar to a scatter plot but adds a third variable shown by
A. chart title **B. the size of the bubble**
C. colour of the axis D. line thickness

53. Quantitative visualization represents
A. numerical data B. categories only
C. locations only D. text only

54. Categorical visualization represents data that falls into
A. locations B. numbers only
C. distinct categories D. time

55. Temporal visualization displays data that
A. is geographic **B. changes over time**
C. is categorical D. is unstructured

56. Spatial visualization represents data related to
A. numbers only B. categories
C. time **D. physical locations or spaces**

57. Multivariate visualization represents data with
A. only time **B. more than two variables**
C. one variable D. no variable

58. Interactive visualization allows users to interact with data through
A. paper printouts B. books
C. dashboards and filters D. posters

59. Network diagrams, tree maps and word clouds are tools of
A. information visualization B. quantitative visualization
C. temporal visualization D. spatial visualization

60. A database is a collection of related data stored in the form of
A. sounds B. images only
C. folders only **D. tables**

61. Data in a database is retrieved using
A. JPEG files **B. SQL queries**
C. CSS styles D. HTML tags

62. The data science life cycle is a(n)
A. single step B. random process
C. iterative process D. one-time process

63. The first stage of the data science life cycle is
A. Problem definition B. Data analysis
C. Data collection D. Model deployment

64. In problem definition, KPI stands for
A. Key Process Input B. Known Problem Index
C. Key Programming Interface **D. Key Performance Indicator**

65. A KPI is
A. a database
B. a measurable target to track progress and success
C. a programming language
D. a type of chart

66. In data collection, APIs are
A. operating systems
B. charts
C. small programs that act like messengers between software
D. databases

67. API stands for
A. Application Programming Interface B. Application Process Interface
C. Applied Program Interface D. Automatic Program Input

68. Data cleaning is also called
A. data collection B. data deployment
C. data modelling **D. data preprocessing**

69. Which stage ensures a model's accuracy, reliability and compliance with privacy rules?
A. Maintenance and monitoring B. Data analysis
C. Model evaluation D. Model deployment

70. Which stage sets up the working model in a website, app or database?
A. Data collection B. Data cleaning
C. Problem definition **D. Model deployment**

71. Partial deployment means the model is
A. deleted **B. deployed in small steps to test it first**
C. never deployed D. deployed twice

72. Which is the key characteristic of the data cleaning stage?
A. Error removal and data organisation B. Data collection
C. Model deployment D. Pattern identification