

JK

Roll No. _____

Question Booklet Number

3733957

O.M.R. Serial No. :

--	--	--	--	--	--	--

BCA VI Semester (NEP) Examination, 2025-26

Data Science and Machine Learning

Paper Code

B	C	A	6	0	0	4
---	---	---	---	---	---	---

Question Booklet
Series

A

Time : 1 : 30 Hours]

[Maximum Marks : 75

Instructions to the Examinee :

1. Do not open the booklet unless you are asked to do so.
2. The booklet contains 100 questions. Examinee is required to answer 75 questions in the OMR Answer-Sheet provided and not in the question booklet. **All questions carry equal marks.**
3. Examine the Booklet and the OMR Answer-Sheet very carefully before you proceed. Faulty question booklet due to missing or duplicate pages/questions or having any other discrepancy should be got immediately replaced.
4. Four alternative answers are mentioned for each question as – A, B, C & D in the booklet. The candidate has to choose the correct answer and mark the same in the OMR Answer-Sheet as per the direction :

(Remaining instructions on the last page)

परीक्षार्थियों के लिए निर्देश :

1. प्रश्न-पुस्तिका को तब तक न खोलें जब तक आपसे कहा न जाए।
2. प्रश्न-पुस्तिका में 100 प्रश्न हैं। परीक्षार्थी को 75 प्रश्नों को केवल दी गई OMR आन्सर-शीट पर ही हल करना है, प्रश्न-पुस्तिका पर नहीं। सभी प्रश्नों के अंक समान हैं।
3. प्रश्नों के उत्तर अंकित करने से पूर्व प्रश्न-पुस्तिका तथा OMR आन्सर-शीट को सावधानीपूर्वक देख लें। दोषपूर्ण प्रश्न-पुस्तिका जिसमें कुछ भाग छपने से छूट गये हों या प्रश्न एक से अधिक बार छप गए हों या उसमें किसी अन्य प्रकार की कमी हो, तो उसे तुरन्त बदल लें।
4. प्रश्न-पुस्तिका में प्रत्येक प्रश्न के चार सम्भावित उत्तर- A, B, C तथा D हैं। परीक्षार्थी को उन चारों विकल्पों में से सही उत्तर छाँटना है। उत्तर को OMR उत्तर-पत्रक में सम्बन्धित प्रश्न संख्या में निम्न प्रकार भरना है :

(शेष निर्देश अन्तिम पृष्ठ पर)

1. Data Science is a combination of which fields?
 - (A) Statistics
 - (B) Computer Science
 - (C) Domain Knowledge
 - (D) All of the above
2. The term Data Science became widely popular after:
 - (A) 1990
 - (B) 2000
 - (C) 2010
 - (D) 2020
3. Which role mainly builds predictive models?
 - (A) Data Analyst
 - (B) Data Scientist
 - (C) Data Engineer
 - (D) Business Analyst
4. First stage of a Data Science project:
 - (A) Data Modeling
 - (B) Problem Definition
 - (C) Data Visualization
 - (D) Deployment
5. Which field uses Data Science?
 - (A) Healthcare
 - (B) Finance
 - (C) Marketing
 - (D) All of the above
6. What is the main purpose of data preprocessing in data mining?
 - (A) To increase the volume of data
 - (B) To improve data quality and make it suitable for analysis
 - (C) To store data in a database
 - (D) To encrypt personal data
7. Removing incorrect or incomplete data is called:
 - (A) Data Integration
 - (B) Data Cleaning
 - (C) Data Reduction
 - (D) Data Collection
8. Which statement about outliers is true?
 - (A) Outliers should be part of the training dataset but should not be present in the test data
 - (B) Outliers should be identified and removed from a dataset.
 - (C) The nature of the problem determines how outliers are used
 - (D) Outliers should be part of the test dataset but should not be present in the training data

9. Mean is also called:
- (A) Average
 - (B) Median
 - (C) Mode
 - (D) Range
10. Standard deviation measures:
- (A) Central tendency
 - (B) Data spread
 - (C) Data storage
 - (D) Data entry
11. Correlation measures:
- (A) Relationship between variables
 - (B) Data deletion
 - (C) Data printing
 - (D) Data storage
12. ANOVA stands for:
- (A) Analysis of Variance
 - (B) Average of Numbers
 - (C) Analysis of Values
 - (D) Average Variance
13. Machine learning allows machines to:
- (A) Learn from data
 - (B) Store data
 - (C) Print data
 - (D) Delete data
14. Regression predicts:
- (A) Continuous values
 - (B) Categories
 - (C) Groups
 - (D) Images
15. Classification predicts:
- (A) Categories
 - (B) Continuous values
 - (C) Random numbers
 - (D) Images
16. Neural networks are inspired by:
- (A) Human brain
 - (B) CPU
 - (C) Database
 - (D) Hard disk
17. Basic unit of neural network:
- (A) Neuron
 - (B) Cell
 - (C) Processor
 - (D) Register

18. Activation function is used to:
- (A) Introduce nonlinearity
 - (B) Store data
 - (C) Delete data
 - (D) Compress files
19. Example of activation function:
- (A) Sigmoid
 - (B) ReLU
 - (C) Tanh
 - (D) All of the above
20. Deep learning uses:
- (A) Multi-layer neural networks
 - (B) Single neuron
 - (C) No neuron
 - (D) Data tables
21. The process of gathering data is called:
- (A) Data Mining
 - (B) Data Collection
 - (C) Data Analysis
 - (D) Data Reduction
22. Removing incorrect or incomplete data is called:
- (A) Data Integration
 - (B) Data Cleaning
 - (C) Data Reduction
 - (D) Data Collection
23. Combining data from multiple sources is:
- (A) Data Integration
 - (B) Data Cleaning
 - (C) Data Transformation
 - (D) Data Reduction
24. If the standard deviation of a dataset is 0, what does this indicate?
- (A) The dataset has high variance
 - (B) All data points are equal
 - (C) The mean is 0
 - (D) There are missing values
25. What is the primary goal of Exploratory Data Analysis (EDA) in data science?
- (A) To build predictive machine learning models
 - (B) To understand data, uncover patterns and find anomalies
 - (C) To replace all missing values with mean
 - (D) To collect raw data from databases

26. Which technique is used to quantify the linear relationship between two quantitative variables?
- (A) Scatter plot
 - (B) Correlation coefficient
 - (C) Box plot
 - (D) Median
27. Which plot is best suited to check for outliers in a numerical variable?
- (A) Pie chart
 - (B) Bar chart
 - (C) Box plot
 - (D) Line graph
28. Which of the following is an example of a bivariate exploratory graph?
- (A) Histogram of a single feature
 - (B) Box plot of a single variable
 - (C) Scatter plot of feature X versus feature Y
 - (D) Density plot
29. Which statistical measure is most robust (less affected) to extreme outliers?
- (A) Mean
 - (B) Standard Deviation
 - (C) Median
 - (D) Variance
30. Which of the following is characteristic of Processed Data?
- (A) Data is not ready for analysis
 - (B) All steps should be noted
 - (C) Hard to use for data analysis
 - (D) None of the mentioned
31. Which of the following technique comes under practical machine learning?
- (A) Bagging
 - (B) Boosting
 - (C) Forecasting
 - (D) None of the mentioned
32. Which of the following is commonly referred to as 'data fishing'?
- (A) Data dredging
 - (B) Data bagging
 - (C) Data merging
 - (D) Data booting
33. Which of the following is characteristic of Raw Data?
- (A) Data is ready for analysis
 - (B) Original version of data
 - (C) Easy to use for data analysis
 - (D) None of the mentioned
34. Which of the following input can be accepted by DataFrame?
- (A) DataFrame
 - (B) Series
 - (C) Structured ndarray
 - (D) All of the mentioned

35. Which of the following is the most common problem with messy data?
- (A) Variables are stored in both rows and columns
 - (B) Column headers are values
 - (C) A single observational unit is stored in multiple tables
 - (D) All of the mentioned
36. Which of the following is used to extract data from HTML code of websites?
- (A) Webscraping
 - (B) Webcleaning
 - (C) Webdredging
 - (D) All of the mentioned
37. Which of the following gave rise to the need for graphs in data analysis?
- (A) Decision making
 - (B) Communicating results
 - (C) Data visualization
 - (D) All of the mentioned
38. Which of the following testing is concerned with making decisions using data?
- (A) Hypothesis
 - (B) Probability
 - (C) Causal
 - (D) None of the mentioned
39. Which of the following function can create the indices for time series type of splitting?
- (A) newTimeSlices
 - (B) createTimeSlices
 - (C) binTimeSlices
 - (D) none of the mentioned
40. Which of the following package tools are present in caret?
- (A) pre-processing
 - (B) feature selection
 - (C) model tuning
 - (D) all of the mentioned
41. Which Python library is used for creating basic visualizations such as line and bar charts?
- (A) NumPy
 - (B) Pandas
 - (C) Matplotlib
 - (D) Seaborn
42. A dataset has duplicate rows causing issues in analysis. Which Pandas method will you use to fix this?
- (A) drop_duplicates()
 - (B) dropna()
 - (C) fillna()
 - (D) groupby()

43. Which type of neural network is best suited for image processing tasks?
- (A) Recurrent Neural Network (RNN)
 - (B) Convolutional Neural Network (CNN)
 - (C) Perceptron
 - (D) Linear Regression
44. Which of the following techniques is commonly used to prevent overfitting in deep neural networks?
- (A) Increasing the number of hidden layers
 - (B) Dropout
 - (C) Reducing the training data
 - (D) Increasing the learning rate
45. What is the purpose of the back-propagation algorithm?
- (A) To perform forwarding propagation
 - (B) To calculate the gradient of the loss function with respect to weights
 - (C) To normalize input data
 - (D) To initialize the weights
46. In a feedforward ANN, how does information flow?
- (A) In one direction (input to output)
 - (B) In both directions
 - (C) In a circular manner
 - (D) From output to input
47. Which of the following activation functions is most commonly used in the hidden layers of a Convolutional Neural Network to avoid the vanishing gradient problem?
- (A) Sigmoid
 - (B) ReLu (Rectified Linear Unit)
 - (C) Tanh
 - (D) Softmax
48. What does the Softmax activation function in the output layer of a classifier produce?
- (A) A single binary value
 - (B) A linear value
 - (C) A probability distribution over classes
 - (D) A random output
49. The mapping or classification of a class with some predefined group or class is known as?
- (A) Data Characterization
 - (B) Data Discrimination
 - (C) Data Set
 - (D) Data Sub Structure
50. Deep learning uses:
- (A) Multi-layer neural networks
 - (B) Single neuron
 - (C) No neuron
 - (D) Data tables

51. In linear regression, the method used to minimize error is:
- (A) Gradient Boosting
 - (B) Least Squares Method
 - (C) K-Means
 - (D) PCA
52. What does R-squared represent?
- (A) Error rate
 - (B) Variance explained by the model
 - (C) Mean value
 - (D) Probability score
53. Which assumption is NOT part of linear regression?
- (A) Linearity
 - (B) Independence
 - (C) Normality of errors
 - (D) Non-linear decision boundary
54. Logistic regression is mainly used for:
- (A) Regression problems
 - (B) Classification problems
 - (C) Clustering
 - (D) Forecasting time series
55. The output of logistic regression is:
- (A) Continuous value
 - (B) Probability between 0 and 1
 - (C) Integer value
 - (D) Negative value only
56. Which function is used in logistic regression?
- (A) Linear function
 - (B) Sigmoid function
 - (C) ReLU function
 - (D) Softmax only
57. What is the decision boundary in logistic regression?
- (A) Curve
 - (B) Straight line or hyperplane
 - (C) Random boundary
 - (D) Circular boundary
58. Machine learning allows machines to:
- (A) Learn from data
 - (B) Store data
 - (C) Print data
 - (D) Delete data
59. What is the output at each node in a neural network called?
- (A) Synapse
 - (B) Activation or node value
 - (C) Weight
 - (D) Bias

60. Classification predicts:

- (A) Categories
- (B) Continuous values
- (C) Random numbers
- (D) Images

61. What type of algorithm is a Decision Tree?

- (A) Unsupervised learning
- (B) Supervised learning
- (C) Reinforcement learning
- (D) Deep learning

62. Decision Trees can be used for:

- (A) Only classification
- (B) Only regression
- (C) Both classification and regression
- (D) Clustering only

63. Activation function is used to:

- (A) Introduce nonlinearity
- (B) Store data
- (C) Delete data
- (D) Compress files

64. Which algorithm is commonly used to build decision trees?

- (A) K-Means
- (B) ID3
- (C) Apriori
- (D) DBSCAN

65. Which measure is used in ID3 to select the best feature?

- (A) Gini Index
- (B) Entropy
- (C) Accuracy
- (D) Variance

66. Which algorithm uses Gini Index as a splitting criterion?

- (A) ID3
- (B) C4.5
- (C) CART
- (D) KNN

67. What does entropy measure in a dataset?

- (A) Mean value
- (B) Impurity or randomness
- (C) Distance
- (D) Correlation

68. Overfitting in decision trees can be reduced by:
- (A) Increasing depth
 - (B) Pruning
 - (C) Adding more features
 - (D) Using more nodes
69. What is pruning in decision trees?
- (A) Adding nodes
 - (B) Removing unnecessary branches
 - (C) Increasing tree height
 - (D) Splitting nodes
70. Who proposed the first artificial neuron model?
- (A) Alan Turing
 - (B) McCulloch and Pitts
 - (C) Geoffrey Hinton
 - (D) John McCarthy
71. Which field combines biology and computing to design neural networks?
- (A) Data Mining
 - (B) Bioinformatics
 - (C) Artificial Intelligence
 - (D) Computer Networks
72. Which part of a biological neuron receives signals?
- (A) Axon
 - (B) Dendrites
 - (C) Synapse
 - (D) Nucleus
73. The axon of a neuron is responsible for:
- (A) Receiving signals
 - (B) Processing data
 - (C) Sending signals
 - (D) Storing data
74. Artificial Neural Networks are mainly used in:
- (A) Only hardware design
 - (B) Only networking
 - (C) Pattern recognition and learning
 - (D) File storage
75. Which of the following is NOT a component of an artificial neuron?
- (A) Weights
 - (B) Activation function
 - (C) Bias
 - (D) Compiler

76. The output of an artificial neuron is determined by:
- (A) Only input values
 - (B) Weighted sum and activation function
 - (C) Only bias
 - (D) Only weights
77. Which neural network model uses multiple layers?
- (A) Single-layer perceptron
 - (B) Multilayer perceptron (MLP)
 - (C) Linear regression
 - (D) K-Means
78. Backpropagation is used for:
- (A) Clustering
 - (B) Training neural networks
 - (C) Data cleaning
 - (D) Data visualization
79. Which of the following is a type of neural network?
- (A) Decision Tree
 - (B) Convolutional Neural Network
 - (C) KNN
 - (D) Naive Bayes
80. Recurrent Neural Networks (RNN) are mainly used for:
- (A) Static data
 - (B) Sequential data
 - (C) Clustering
 - (D) Sorting
81. Which of the following is NOT a neural network model?
- (A) Hopfield Network
 - (B) Boltzmann Machine
 - (C) Support Vector Machine
 - (D) Feedforward Network
82. Standard deviation measures:
- (A) Central tendency
 - (B) Spread or dispersion of data
 - (C) Skewness
 - (D) Median
83. Which of the following is a subset of machine learning?
- (A) Neural Network
 - (B) Perceptron
 - (C) Deep Learning
 - (D) All of the above

84. Positive skewness indicates:
- (A) Left tail is longer
 - (B) Right tail is longer
 - (C) Symmetrical distribution
 - (D) No variation
85. Kurtosis measures:
- (A) Central tendency
 - (B) Spread
 - (C) Shape/peakedness of distribution
 - (D) Correlation
86. In a box plot, the line inside the box represents:
- (A) Mean
 - (B) Mode
 - (C) Median
 - (D) Range
87. Outliers in a box plot are typically:
- (A) Inside the box
 - (B) Equal to median
 - (C) Points far from whiskers
 - (D) Always zero
88. A pivot table is mainly used for:
- (A) Data cleaning
 - (B) Data summarization and aggregation
 - (C) Data encryption
 - (D) Model training
89. In which of the following applications can we use deep learning to solve the problem?
- (A) Protein structure prediction
 - (B) Prediction of chemical reactions
 - (C) Detection of exotic particles
 - (D) All of the above
90. Which of the following would have a constant input in each epoch of training a Deep Learning model?
- (A) Weight between input and hidden layer
 - (B) Weight between hidden and output layer
 - (C) Biases of all hidden layer neurons
 - (D) Activation function of output layer
91. Which method can be used to handle missing data?
- (A) Ignoring records
 - (B) Mean/median imputation
 - (C) Prediction methods
 - (D) All of the above

92. Data transformation includes:
- (A) Data cleaning only
 - (B) Normalization and scaling
 - (C) Data deletion
 - (D) Data duplication
93. Which of the following is a data reduction technique?
- (A) Sampling
 - (B) Feature selection
 - (C) Dimensionality reduction
 - (D) All of the above
94. In data preprocessing, noise refers to:
- (A) Useful data
 - (B) Random errors or irrelevant data
 - (C) Structured data
 - (D) Clean data
95. Which technique is used to smooth noisy data?
- (A) Binning
 - (B) Sorting
 - (C) Classification
 - (D) Clustering
96. Which of the following is NOT a data reduction method?
- (A) Sampling
 - (B) Aggregation
 - (C) Feature selection
 - (D) Data duplication
97. Combining multiple attributes into one is called:
- (A) Data cleaning
 - (B) Data integration
 - (C) Data aggregation
 - (D) Data sampling
98. Data discretization is the process of:
- (A) Converting continuous data into intervals
 - (B) Removing missing values
 - (C) Combining datasets
 - (D) Sorting data
99. Which preprocessing step ensures consistency in data?
- (A) Data integration
 - (B) Data cleaning
 - (C) Data reduction
 - (D) Data visualization
100. Which of the following is an example of transformation?
- (A) Removing duplicates
 - (B) Normalizing values between 0 and 1
 - (C) Dropping rows
 - (D) Collecting data