

Roll. No.

Question Booklet Number

601517

O.M.R. Serial No.

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

BCA (SEM.-VI) (NEP) EXAMINATION, 2024

COMPUTER APPLICATION

(Data Science and Machine Learning)

[BCA-6004]

Paper Code						
Z	2	0	0	0	9	2

Time : 1 : 30 Hours

Question Booklet

Series

A

Max. Marks : 75

Instructions to the Examinee :

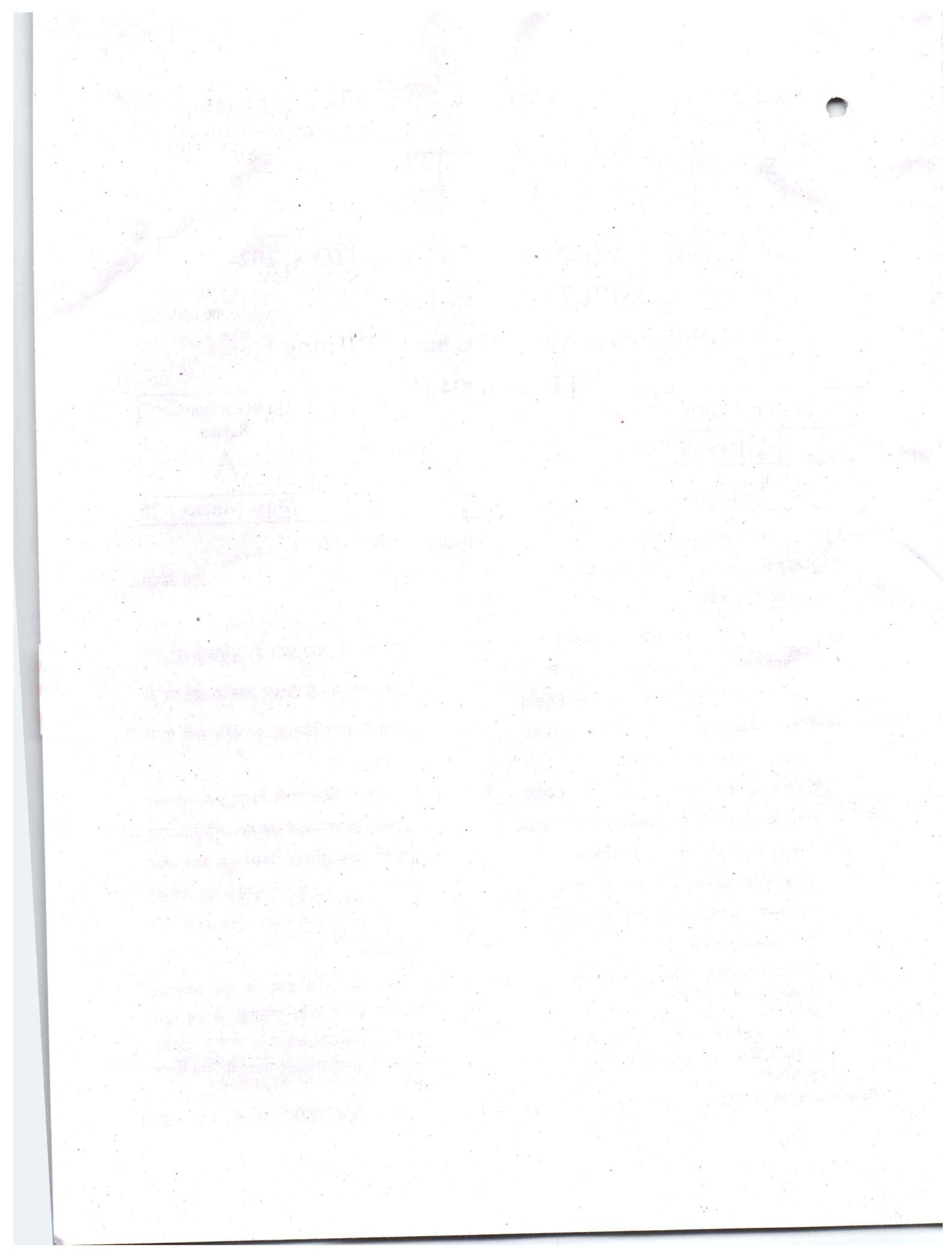
1. Do not open the booklet unless you are asked to do so.
2. The booklet contains 100 questions. Examinee is required to answer 75 questions in the OMR Answer-Sheet provided and not in the question booklet. All questions carry equal marks.
3. Examine the Booklet and the OMR Answer-Sheet very carefully 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 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 गए हों या प्रश्न एक से अधिक बार छप गए हों या उसमें किसी अन्य प्रकार की कमी हो, उसे तुरन्त बदल लें।
4. प्रश्न-पुस्तिका में प्रत्येक प्रश्न के चार सम्भावित उत्तर- A, B, C एवं 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)



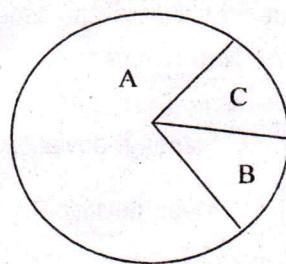
1. A collection of information about a related topic is referred to as a _____
- (A) Visualization
(B) Conclusion
(C) Analysis
(D) Data
2. Machine learning is a subset of :
- (A) Analytics
(B) Deep learning
(C) Artificial intelligence
(D) None of the above
3. Identify the key data science skills among the following :
- (A) Data visualization
(B) Machine learning
(C) Statistics
(D) All of the above
4. Which is not a tool for Statistical Data Analysis?
- (A) Logistic Regression
(B) Histogram
(C) Linear and Non-linear regression
(D) ANOVA
5. What is the mean of test scores?
 $\{70, 70, 80, 85, 85, 90, 95, 95, 100, 100\}$
- (A) 87
(B) 90
(C) 85
(D) 70, 85, 95 and 100
6. Which of the following steps are performed by data scientists after acquiring the data?
- (A) Data integration
(B) Data cleansing
(C) Data mining
(D) None of these
7. Data that lies near/outside the boundary of the cluster is called as :
- (A) Outliers
(B) Key Points
(C) Spike
(D) Cluster Head
8. Which of the following is characteristic of Processed Data?
- (A) Hard to be used for data analysis
(B) Data is not yet ready for analysis
(C) Data has been replicated.
(D) None of these
9. Which of the following is one of the key data science skill?
- (A) Machine learning
(B) Data visualization
(C) Statistics
(D) All of these
10. Audio data is :
- (A) Structured
(B) Semi-structured
(C) Unstructured
(D) Cannot be determined

11. Among the following options identify the one which is false regarding regression.
- (A) It is used for prediction
(B) It discovers causal relationships.
(C) It is used for interpretation.
(D) It relates inputs to outputs.
12. What does K stand for in K mean algorithm?
- (A) Number of data points
(B) Number of clusters
(C) Number of attributes
(D) Number of iterations
13. Among the following options choose which one of the following focuses on the discovery of unknown properties on the data.
- (A) Big data
(B) Machine learning
(C) Data-mining
(D) Deep learning
14. Among the following, choose the correct application of data science in Healthcare :
- (A) Data science for genomics
(B) Drug discovery
(C) Data science for medical imaging
(D) All of these
15. The different types of machine learning are:
- (A) Supervised learning
(B) Reinforced learning
(C) Unsupervised learning
(D) All of these.
16. Point out the correct statement :
- (A) Raw data is original source of data
(B) Pre-processed data is original source of data
(C) Raw data is the data obtained after processing steps
(D) None of the mentioned
17. Which of the following are the data sources in data science?
- (A) Unstructured data
(B) Both (A) and (C)
(C) Structured data
(D) None of these
18. Why Machine Learning in Data Science?
- (A) For Prediction
(B) For Visualization
(C) For Cleaning
(D) All of the above
19. Which of the following is definition of Raw Data?
- (A) Easy to use for data analysis
(B) Set of Measurement on Recorded Values
(C) Processed Data
(D) None of the above

20. Which of the following are "Measures of Central Tendency"?
- (A) Mode, Mean, Median
 - (B) Mean, Standard Deviation, Range
 - (C) Mean, Range, Mode
 - (D) Range, Standard Deviation, Variance
21. What is it called when the data is sourced from the place of origin?
- (A) Secondary
 - (B) Secondary and primary
 - (C) Primary
 - (D) All of the above
22. The method of collecting primary data is called :
- (A) Questionnaire and schedule method
 - (B) Observation and Interview method
 - (C) None of the above
 - (D) Both (A) and (B)
23. Among the following option identify the one which is not a type of learning :
- (A) Supervised learning
 - (B) Reinforcement learning
 - (C) Semi- unsupervised learning
 - (D) Un-supervised learning
24. Identify the kind of learning algorithm for "facial identities for facial expressions".
- (A) Prediction
 - (B) Recognizing anomalies
 - (C) Recognition patterns
 - (D) Generating patterns
25. What is the application of machine learning methods to a large database called?
- (A) Big data computing
 - (B) Data mining
 - (C) Internet of things
 - (D) Artificial intelligence
26. What is Machine learning?
- (A) The selective acquisition of knowledge through the use of computer programs.
 - (B) The selective acquisition of knowledge through the use of manual programs.
 - (C) The autonomous acquisition of knowledge through the use of computer programs.
 - (D) The autonomous acquisition of knowledge through the use of manual programs
27. What is the key difference between supervised and unsupervised learning?
- (A) Supervised learning requires labeled data, while unsupervised learning does not.
 - (B) Supervised learning predicts labels, while unsupervised learning discovers patterns.
 - (C) Supervised learning is used for classification, while unsupervised learning is used for regression.
 - (D) Supervised learning is always more accurate than unsupervised learning.

28. Which type of machine learning algorithm falls under the category of "unsupervised learning"?
- (A) Linear Regression
 - (B) K-means Clustering
 - (C) Decision Trees
 - (D) Random Forest
29. Which of the following statements is not true about neural networks?
- (A) They are class of very powerful machine learning classifiers
 - (B) Neural networks are a class of computationally inexpensive algorithms
 - (C) These are capable of fitting almost any hypotheses
 - (D) NN has lots of interconnected nodes which are organized in layers
30. Which of the following statements is not true about the structure of a neural network?
- (A) It consists of a set of nodes connected by links
 - (B) Not every link in a neural network is associated with a weight
 - (C) The processing elements in the neural network are known as neurons
 - (D) A set of thresholds or activation is used
31. Which of the following is false?
- (A) Neural networks are artificial copies of the human brain.
 - (B) Neural networks have high computational rates than conventional computers.
 - (C) Neural networks learn by examples.
 - (D) None of these
32. Which algorithm is best suited for a binary classification problem?
- (A) K-nearest Neighbors
 - (B) Random Forest
 - (C) Decision Trees
 - (D) Linear Regression
33. To which input does the learner has access to?
- (A) Testing Data
 - (B) Label Data
 - (C) Training Data
 - (D) Cross-Validation Data
34. What is the learner's output also called?
- (A) Predictor, or Hypothesis, or Classifier
 - (B) Predictor, or Hypothesis, or Trainer
 - (C) Predictor, or Trainer, or Classifier
 - (D) Trainer, or Hypothesis, or Classifier
35. What is not accessible to the learner?
- (A) Training Set
 - (B) Label Set
 - (C) Labeling Function
 - (D) Domain Set

36. What are the possible values of A, B, and C in the following diagram?



- (A) A – Training Set, B – Domain Set, C – Cross-Validation Set
- (B) A – Training Set, B – Test Set, C – Cross-Validation Set
- (C) A – Training Set, B – Test Set, C – Domain Set
- (D) A – Test Set, B – Domain Set, C – Training Set

37. The shape of the Normal Curve is _____.

- (A) Bell Shaped
- (B) Flat
- (C) Circular
- (D) Spiked

38. Normal Distribution is applied for _____.

- (A) Continuous Random Distribution
- (B) Discrete Random Variable
- (C) Irregular Random Variable
- (D) All of these

39. Normal Distribution is symmetric is about _____.

- (A) Variance
- (B) Mean
- (C) Standard deviation
- (D) Covariance

40. The area under a standard normal curve is :

- (A) 0
- (B) 1
- (C) ∞
- (D) not defined

41. Normal Distribution is also known as _____.

- (A) Cauchy's Distribution
- (B) Laplacian Distribution
- (C) Gaussian Distribution
- (D) Lagrangian Distribution

42. Skewness of Normal distribution is _____.

- (A) Negative
- (B) Positive
- (C) Undefined
- (D) 0

43. For a normal distribution its mean, median, mode are equal.

- (A) True
- (B) False
- (C) Sometimes True sometimes False
- (D) Depending upon the data

44. In Normal distribution, the highest value of ordinate occurs at _____.

- (A) Mean
- (B) Variance
- (C) Extremes
- (D) Same value occurs at all points

45. The shape of the normal curve depends on its _____.
 (A) Mean deviation
 (B) Standard deviation
 (C) Quartile deviation
 (D) Correlation
46. In Standard normal distribution, the value of mode is _____.
 (A) 2
 (B) 1
 (C) 0
 (D) Not fixed
47. What is the mean and variance for standard normal distribution?
 (A) Mean is 0 and variance is 1
 (B) Mean is 1 and variance is 0
 (C) Mean is 0 and variance is ∞
 (D) Mean is ∞ and variance is 0
48. Standard deviation is the square root of variance.
 (A) True
 (B) False
 (C) Depending upon the standard deviation.
 (D) Depending upon the data
49. Find the mean and variance of X :
- | | | | | | |
|------|-----|-----|-----|-----|-----|
| x | 0 | 1 | 2 | 3 | 4 |
| f(x) | 1/9 | 2/9 | 3/9 | 2/9 | 1/9 |
- (A) 2, 4/3
 (B) 3, 4/3
 (C) 2, 2/3
 (D) 3, 2/3
50. _____ is the simplest possible measure of dispersion and equal to the difference of extreme values of the series.
 (A) Range
 (B) Standard deviation
 (C) Mean deviation
 (D) Skewness
51. Quartile deviation is the _____ of/between the third and first quartile values.
 (A) Average
 (B) maximum
 (C) difference
 (D) minimum
52. When the values in the series do not have equal importance, we calculate _____.
 (A) Mode
 (B) Arithmetic mean
 (C) Weighted mean
 (D) None of these
53. Which of the following is not a disadvantage of using mean?
 (A) It is affected by the extreme values.
 (B) It doesn't possess the desired algebraic property.
 (C) It cannot be computed in grouped data with open ended class intervals.
 (D) None of these.

54. For calculating the median, all the items of the series must be arranged in a/an ____.
- Descending order
 - Ascending order
 - Both (A) and (B)
 - None of these
55. Mode refers to the value in the series, that occurs ____ number of times.
- Minimum
 - $n-1$, where n is the items in the list
 - at least 3
 - maximum
56. The value of extreme items don't influence the average for ____.
- mean
 - median
 - mode
 - none of these
57. ____ is not a measure of central tendency.
- mean
 - mode
 - median
 - range
58. ____ divides the data into 100 equal parts.
- Quartile
 - Percentile
 - Decile
 - None of these
59. The sum of deviation from _____ is always 0.
- mean
 - mode
 - median
 - None of these
60. Two unbiased coins are tossed. What's the probability of having access to at most one head?
- $1/5$
 - $3/4$
 - $1/2$
 - $1/3$
61. K-fold cross-validation is :
- Linear in K
 - Quadratic in K
 - Cubic in K
 - Exponential in K
62. The following is a weakness of the tree based classifiers :
- They are liable to over-fit or under-fit models
 - They always under-fit models
 - They always overfit models
 - They take a lot of time to fit models
63. The trade off between over fitting and under fitting training data is called :
- The bias-variance tradeoff
 - The tradeoff curve
 - The residual sum of squares
 - The null deviance

64. If the first quartile and third quartile are as 32 and 35 respectively with the median of 20, then distribution is skewed to :
- (A) Lower tail
 - (B) Open end tail
 - (C) Upper tail
 - (D) Closed end tail
65. If for a distribution the difference of first quartile and median is greater than difference of median and third quartile, then the distribution is classified as :
- (A) Positively skewed
 - (B) Negatively skewed
 - (C) Not skewed at all
 - (D) Absolute open ended
66. If the $\beta_1 = 9$, $\beta_2 = 11$, then coefficient of skewness is :
- (A) 0.589
 - (B) 0.789
 - (C) 0.689
 - (D) 0.489
67. The measurement techniques used to measure the extent of skewness in data set values are called :
- (A) Measure of tail distribution
 - (B) Measure of median tail
 - (C) Measure of distribution width
 - (D) Measure of skewness
68. The statistical measures such as average deviation, standard deviation and mean are classified as part of :
- (A) Moment system
 - (B) Decile system
 - (C) Quartile system
 - (D) Percentile system
69. The method of calculating skewness which is based on the positions of quartiles and median in a distribution is called :
- (A) Gary's coefficient of skewness
 - (B) Bowley's coefficient of skewness
 - (C) Sharma's coefficient of skewness
 - (D) Jack Karl's coefficient of skewness
70. The median of a skewed distribution is 8, third quartile is 12, first quartile is 8 and inter-quartile range is 4, then relative coefficient of skewness is :
- (A) $+/- 11$
 - (B) $+/- 1$
 - (C) $+/- 8$
 - (D) $+/- 9$
71. The moment about mean which is indication whether distribution is symmetrical or asymmetrical is considered as :
- (A) Third moment
 - (B) Second moment
 - (C) First moment
 - (D) Fourth moment

72. The frequency distribution is considered as negatively skewed if all the values of distribution moves to :
(A) Variance tail
(B) Upper tail
(C) Median tail
(D) Lower tail
73. In measures of skewness, the absolute skewness is equal to :
(A) mean - mode
(B) mean - median
(C) mean + mode
(D) mean + median
74. How many dependent variables does a two-way ANOVA have?
(A) 3 (B) 1
(C) 2 (D) 0
75. What type of data are best analysed in ANOVA?
(A) Co-relational
(B) Random
(C) Experimental
(D) Simple
76. Which of the following are data security consideration?
(A) Backups
(B) Disposal of Data
(C) Archival Storage
(D) All of the above
77. What is data privacy?
(A) Data privacy is the protection of personal data
(B) Users who should not have access to it
(C) The ability of individuals to determine who can access their personal information
(D) All of the above mentioned
78. What do you mean by data science?
(A) Dealing with huge amounts of data to find marketing patterns is known as data science
(B) Extracting a meaningful insight from the data is what data science is
(C) It is a study that deals with a huge amount of all types of data structured, unstructured, or semi-structured
(D) All of the above
79. For the Karl Pearson's skewness coefficient the value of skewness must be in limits :
(A) $+/- 3$
(B) $+/- 1$
(C) $+/- 2$
(D) $+/- 5$
80. ANOVAs cannot be used when testing data collected in educational research as it cannot be applied to social science.
(A) True
(B) False
(C) Sometimes True and sometimes False
(D) Cannot be sure

81. What is structured data?
- (A) Structured data is a type of data that is huge in number and has many inaccurate values
- (B) Structured data is a type of data that is comparatively less in number and can be stored in proper rows and columns
- (C) Structured data is a type of data that has inaccurate values but can be stored in rows and columns
- (D) None of these.
82. Data used to build a data mining model are :
- (A) Training data
- (B) Validation data
- (C) Test data
- (D) Hidden data
83. The distribution is considered leptokurtic if :
- (A) If $\beta_3 < 3$
- (B) If $\beta_3 > 3$
- (C) If $\beta_2 > 3$
- (D) If $\beta_2 < 2$
84. Among the following examples, which would you address using an supervised learning algorithm?
- (A) Given email labeled as spam or not spam, learn a spam filter.
- (B) Given a set of news articles found on the web, group them into set of articles about the same story.
- (C) Given a database of customer data, automatically discover market segments and group customers into different market segments.
- (D) Find the patterns in market basket analysis.
85. In kurtosis, the beta is greater than three and quartile range is preferred for :
- (A) Mesokurtic distribution
- (B) Leptokurtic distribution
- (C) Mega curve distribution
- (D) Platykurtic distribution
86. Which of the following is a good test dataset characteristic?
- (A) Large enough to yield meaningful results.
- (B) Representative of the dataset as a whole.
- (C) Both (A) and (B)
- (D) None of the above

87.

Following are the types of supervised learning :

- (A) Classification
- (B) Regression
- (C) Subgroup discovery
- (D) All of the above

88.

The distance metrics used in clustering :

- (A) Euclidean distance
- (B) Manhattan distance
- (C) Square distance
- (D) Both (A) and (B)

89.

The output of training process in machine learning is :

- (A) Machine learning model
- (B) Machine learning algorithm
- (C) Null
- (D) Accuracy

90.

A feature F1 can take certain value: A, B, C, D, E, & F and represents grade of students from a college. Here feature type is :

- (A) Nominal
- (B) Ordinal
- (C) Categorical
- (D) Boolean

91.

Prediction is :

- (A) the result of application of specific theory or rule in a specific case
- (B) discipline in statistics used to find projections in multidimensional data
- (C) value entered in database by expert
- (D) independent of data

92.

The kurtosis defines the peakness of the curve in the region which is :

- (A) Around the variance
- (B) Around the median
- (C) Around the mean
- (D) Around the mode

93.

Some companies wants to segment their customers into distinct groups, this is an example of :

- (A) Supervised learning
- (B) Reinforcement learning
- (C) Unsupervised learning
- (D) Data extraction

94.

You are given seismic data and you want to predict next earthquake, this is an example of :

- (A) Supervised learning
- (B) Reinforcement learning
- (C) Unsupervised learning
- (D) Dimensionality reduction

95. Imagine a Newly-Born starts to learn walking. It will try to find a suitable policy to learn walking after repeated falling and getting up. Specify what type of machine learning is best suited?
- (A) Classification
(B) Regression
(C) k-means algorithm
(D) Reinforcement learning
96. In multiclass classification number of classes must be :
- (A) less than two
(B) equals to two
(C) greater than two
(D) Both (A) and (B)
97. A perceptron adds up all the weighted inputs it receives, and if it exceeds a certain value, it outputs a 1, otherwise it just outputs a 0 :
- (A) true
(B) false
(C) sometimes – it can also output intermediate values as well.
(D) can't say
98. Analysis of variance performs which type of test?
- (A) F-test
(B) t-test
(C) z-test
(D) xy-test
99. You trained a binary classifier model which gives very high accuracy on the training data, but much lower accuracy on validation data. Which is FALSE?
- (A) This is an instance of over-fitting.
(B) This is an instance of under-fitting.
(C) The training was not well regularized.
(D) The training and testing examples may be sampled from different distributions.
100. 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

Rough Work / रफ कार्य

Example :

Question :

Q.1 A B C D

Q.2 A B C D

Q.3 A B C D

5. Each question carries equal marks. Marks will be awarded according to the number of correct answers you have.
6. All answers are to be given on OMR Answer Sheet only. Answers given anywhere other than the place specified in the answer sheet will not be considered valid.
7. Before writing anything on the OMR Answer Sheet, all the instructions given in it should be read carefully.
8. After the completion of the examination, candidates should leave the examination hall only after providing their OMR Answer Sheet to the invigilator. Candidate can carry their Question Booklet.
9. There will be no negative marking.
10. Rough work, if any, should be done on the blank pages provided for the purpose in the booklet.
11. To bring and use of log-book, calculator, pager & cellular phone in examination hall is prohibited.
12. In case of any difference found in English and Hindi version of the question, the English version of the question will be held authentic.

Impt. On opening the question booklet, first check that all the pages of the question booklet are printed properly. If there is any discrepancy in the question Booklet, then after showing it to the invigilator, get another question Booklet of the same series.

उदाहरण :

प्रश्न :

प्रश्न 1 A B C D

प्रश्न 2 A B C D

प्रश्न 3 A B C D

प्रत्येक प्रश्न के अंक समान हैं। आपके जितने उत्तर सही होंगे, उन्हीं के अनुसार अंक प्रदान किये जायेंगे।

सभी उत्तर केवल ओ०एम०आर० उत्तर-पत्रक (OMR Answer Sheet) पर ही दिये जाने हैं। उत्तर-पत्रक में निर्धारित स्थान के अलावा अन्यत्र कहीं पर दिया गया उत्तर मान्य नहीं होगा।

ओ०एम०आर० उत्तर-पत्रक (OMR Answer Sheet) पर कुछ भी लिखने से पूर्व उसमें दिये गये सभी अनुदेशों को सावधानीपूर्वक पढ़ लिया जाये।

परीक्षा समाप्ति के उपरान्त परीक्षार्थी कक्ष निरीक्षक को अपनी OMR Answer Sheet उपलब्ध कराने के बाद ही परीक्षा कक्ष से प्रस्थान करें। परीक्षार्थी अपने साथ प्रश्न-पुस्तिका ले जा सकते हैं।

निगेटिव मार्किंग नहीं है।

कोई भी रफ कार्य, प्रश्न-पुस्तिका में, रफ-कार्य के लिए दिए खाली पेज पर ही किया जाना चाहिए।

परीक्षा-कक्ष में लॉग-बुक, कैल्कुलेटर, पेजर तथा सेल्फ्युलर फोन ले जाना तथा उसका उपयोग करना वर्जित है।

प्रश्न के हिन्दी एवं अंग्रेजी रूपान्तरण में भिन्नता होने की दशा में प्रश्न का अंग्रेजी रूपान्तरण ही मान्य होगा।

महत्वपूर्ण: प्रश्नपुस्तिका खोलने पर प्रथमतः जाँच कर देख लें कि प्रश्नपुस्तिका के सभी पृष्ठ भलीभाँति छपे हुए हैं। यदि प्रश्नपुस्तिका में कोई कमी हो, तो कक्षनिरीक्षक को दिखाकर उसी सिरीज की दूसरी प्रश्नपुस्तिका प्राप्त कर लें।