



Teachingninja.in



Latest Govt Job updates



Private Job updates



Free Mock tests available



Visit - teachingninja.in

UKPSC Investigator & ASO

**Previous Year Paper
(Data Science)
08 Oct, 2024**



पेपर सील खोले बगैर इस तरफ से उत्तर शीट को बाहर निकालें ।
Without opening the Paper seal take out Answer Sheet
from this side.

ASD



परीक्षा का वर्ष : 2024

प्रश्न-पुस्तिका

प्रश्न-पत्र-II (वैकल्पिक विषय)

प्रश्न-पुस्तिका शृंखला

अपना अनुक्रमांक सामने अंकों में
बॉक्स के अन्दर लिखें
शब्दों में

A

डाटा विज्ञान

DATA SCIENCE

समय : 03:00 घंटे
पूर्णांक : 200

Time : 03:00 Hours
Maximum Marks : 200

प्रश्नों के उत्तर देने से पहले नीचे लिखे अनुदेशों को ध्यान से पढ़ लें ।

महत्वपूर्ण निर्देश

- प्रश्न-पुस्तिका के कवर पेज पर अनुक्रमांक के अतिरिक्त कुछ न लिखें ।
- यदि किसी प्रश्न में किसी प्रकार की कोई मुद्रण या तथ्यात्मक त्रुटि हो तो प्रश्न के अंग्रेजी तथा हिन्दी रूपान्तरों में से अंग्रेजी रूपान्तर को मानक माना जायेगा ।
- अभ्यर्थी अपने अनुक्रमांक, विषय-कोड एवं प्रश्न-पुस्तिका की सीरीज का अंकन OMR Answer Sheet में निर्दिष्ट कॉलम में सही-सही करें, अन्यथा उत्तर-पत्रक का मूल्यांकन नहीं किया जायेगा ।
- अभ्यर्थी रफ कार्य हेतु केवल प्रश्न-पुस्तिका (बुकलेट) के अन्त में दिये गये पृष्ठों का ही उपयोग करें । अलग से इस हेतु वर्किंग शीट उपलब्ध नहीं करायी जायेगी ।
- इस प्रश्न-पुस्तिका में 200 प्रश्न (वस्तुनिष्ठ प्रकार) हैं । प्रश्न पुस्तिका में अंकित सभी प्रश्न अनिवार्य हैं और प्रत्येक प्रश्न के समान अंक हैं । आपके जितने उत्तर सही होंगे उन्ही के अनुसार अंक दिये जायेंगे ।
- प्रत्येक प्रश्न के चार वैकल्पिक उत्तर, प्रश्न के नीचे (a), (b), (c) एवं (d) दिये गये हैं । इन चारों में से केवल एक ही सही उत्तर है । जिस उत्तर को आप सही या सबसे उचित समझते हैं, उत्तर-पत्रक (ओ.एम.आर. आंसर शीट) में उसके अक्षर वाले वृत्त को काले अथवा नीले बॉल प्वाइंट पेन से पूरा काला/नीला कर दें ।
- आयोग द्वारा आयोजित की जाने वाली वस्तुनिष्ठ प्रकृति की परीक्षाओं में ऋणात्मक मूल्यांकन (Negative Marking) पद्धति अपनायी जायेगी । अभ्यर्थी द्वारा प्रत्येक प्रश्न हेतु दिए गए गलत उत्तर के लिए या अभ्यर्थी द्वारा एक प्रश्न के एक से अधिक उत्तर देने के लिए (चाहे दिए गए उत्तर में से एक सही ही क्यों न हो), उस प्रश्न के लिए निर्धारित अंकों का एक-चौथाई अंक दण्ड के रूप में काटा जाएगा । दण्ड स्वरूप प्राप्त अंकों के योग को कुल प्राप्तांकों में से घटाया जाएगा ।
- आपके द्वारा सभी उत्तर केवल ओ.एम.आर. उत्तर-पत्रक (OMR Answer Sheet) पर ही दिया जाना अनिवार्य है । ओ.एम.आर. उत्तर-पत्रक के अतिरिक्त अन्य कहीं पर दिया गया उत्तर मान्य नहीं होगा ।
- ओ.एम.आर. उत्तर-पत्रक पर कुछ लिखने के पूर्व उसमें दिये गये सभी अनुदेशों को सावधानीपूर्वक पढ़ लें । ओ.एम.आर. उत्तर-पत्रक में वांछित सूचनाओं को अभ्यर्थी द्वारा परीक्षा प्रारम्भ होने से पूर्व भरा जाना अनिवार्य है ।
- ओ.एम.आर. उत्तर-पत्रक तीन प्रतियों (मूल प्रति, कार्यालय प्रति एवं अभ्यर्थी प्रति) में है । परीक्षा समाप्ति के उपरान्त अभ्यर्थी ओ.एम.आर. उत्तर-पत्रक की मूल प्रति एवं कार्यालय प्रति अन्तरीक्षक (Invigilator) को हस्तगत करने के उपरान्त ही कक्ष छोड़ें । अन्यथा की स्थिति में आयोग द्वारा नियमानुसार कार्यवाही की जाएगी । ओ.एम.आर. उत्तर-पत्रक की अभ्यर्थी प्रति, अभ्यर्थी अपने साथ ले जा सकते हैं ।
- यदि आपने इन अनुदेशों को पढ़ लिया है, इस पृष्ठ पर अपना अनुक्रमांक अंकित कर दिया है और ओ.एम.आर. उत्तर-पत्रक पर वांछित सूचनायें भर दी हैं, तो तब तक प्रतीक्षा करें, जब तक आपको प्रश्न-पुस्तिका खोलने को नहीं कहा जाता ।
- ओ.एम.आर. उत्तर-पत्रक (O.M.R. Answer Sheet) का मूल्यांकन ओ.एम.आर. आंसर शीट पर अभ्यर्थी द्वारा अंकित सीरीज कोड (A, B, C, D) के आधार पर ही किया जायेगा ।
- प्रश्न-पुस्तिका (Question Booklet) में से ओ.एम.आर. उत्तर-पत्रक (O.M.R. Answer Sheet) निकालने के पश्चात् ओ.एम.आर. उत्तर-पत्रक पर प्रश्न-पुस्तिका क्रमांक एवं प्रश्न-पुस्तिका के सीरीज कोड (A, B, C, D) की प्रविष्टि सावधानीपूर्वक करें । यदि उक्तानुसार कार्यवाही नहीं की जाती है, तो उसके लिए अभ्यर्थी स्वयं जिम्मेदार होगा ।

जब तक कहा न जाय इस प्रश्न-पुस्तिका को न खोलें ।

महत्वपूर्ण : प्रश्न-पुस्तिका खोलने पर तुरन्त जाँच कर देख लें कि प्रश्न-पुस्तिका के सभी पेज भली-भाँति छपे हुए हैं । यदि प्रश्न-पुस्तिका सीलबंद न हो अथवा कोई अन्य कमी हो, तो अन्तरीक्षक को दिखाकर उसी सीरीज की दूसरी प्रश्न-पुस्तिका प्राप्त कर लें ।

1. Which probability distribution is often used for modeling continuous variables in Bayesian algorithms ?
 (a) Bernoulli Distribution (b) Gaussian Distribution
 (c) Poisson Distribution (d) Exponential Distribution
2. Data Fishing is sometimes referred to as _____.
 (a) Data Bagging (b) Data Dredging
 (c) Data Merging (d) None of the above
3. What is the role of correlation matrix in data science ?
 (a) Identification of missing values
 (b) Finding relationship between variables
 (c) Visualisation of data distribution
 (d) Calculation of summary
4. In exploratory data analysis, the term 'Kurtosis' is meant for –
 (a) the presence of outlier
 (b) the spread of data distribution
 (c) the symmetry of data distribution
 (d) the shape of data distribution
5. To visualize the relationship between two categorical variables in a contingency table format, which EDA technique is suitable ?
 (a) Scatter plot (b) Mosaic plot and Cross tabulation
 (c) Line plot (d) Whisker's plot
6. How many types of exploratory data analysis are there ?
 (a) 2 (b) 3 (c) 6 (d) 4
7. What does the vertical line in Box-plot represent ?
 (a) Mode (b) Range (c) Standard Deviation (d) Median
8. Which of the following is not a descriptive statistics ?
 (a) Mean (b) Standard Deviation (c) T-test (d) Range
9. Among the following identify the one in which the dimensionality reduction reduces.
 (a) Performance (b) Collinearity (c) Entropy (d) Stochastics
10. To get insight from the data, data scientists prefer _____.
 (a) Data Warehouse (b) Data Mining
 (c) Data Visualisation (d) Data Analysis
11. The type of test (one tailed/two tailed) is defined by which of the following ?
 (a) Null Hypothesis (b) Simple Hypothesis
 (c) Alternate Hypothesis (d) Composite Hypothesis
12. Goodness of fitness of a distribution is tested by _____.
 (a) T-test (b) Chi-square test (c) F-test (d) P-test
13. _____ is not a step in exploratory data analysis.
 (a) Data visualization (b) Data cleaning
 (c) Data summarization (d) Model training

1. बेयसियन (बायेसियन) एल्गोरिद्म में कंटीन्यूअस वेरियेबल के मॉडलिंग के लिए प्रायः किस प्रोबेबिलिटी डिस्ट्रीब्यूशन (वितरण) का उपयोग किया जाता है ?
 (a) बर्नोली डिस्ट्रीब्यूशन (b) गॉसियन डिस्ट्रीब्यूशन
 (c) प्वासों डिस्ट्रीब्यूशन (d) एक्सपोनेन्शियल डिस्ट्रीब्यूशन
2. डाटा फिशिंग को कभी-कभी _____ भी कहा जाता है ।
 (a) डाटा बैगिंग (b) डाटा ड्रेजिंग (c) डाटा मर्जिंग (d) उपरोक्त में से कोई नहीं
3. डाटा साइन्स में सह-संबंध मैट्रिक्स की क्या भूमिका होती है ?
 (a) लुप्त वैल्यूज की पहचान (b) चरों के बीच में संबंध पता करना
 (c) डाटा वितरण की कल्पना करना (d) सारांश आँकड़ों की गणना
4. अन्वेषणात्मक डाटा विश्लेषण में, 'कुर्तोजिस' शब्द का मतलब होता है
 (a) आउटलायर की मौजूदगी (b) डाटा वितरण का फैलाव
 (c) डाटा वितरण की समरूपता (d) डाटा वितरण का आकार
5. कन्टिन्जेन्सी टेबल फॉर्मेट में, दो श्रेणीगत चरों के बीच संबंध की कल्पना करने के लिए ई डी ए की कौन सी तकनीक उपयुक्त है ?
 (a) स्कैटर प्लॉट (b) मोजेक प्लॉट और क्रॉस टेबुलेशन
 (c) लाइन प्लॉट (d) व्हिस्कर प्लॉट
6. अन्वेषणात्मक डाटा विश्लेषण कितने प्रकार के होते हैं ?
 (a) 2 (b) 3 (c) 6 (d) 4
7. बॉक्स प्लॉट में खड़ी लाइन क्या दर्शाती है ?
 (a) बहुलक (b) परास (c) मानक विचलन (d) माध्यिका
8. निम्नलिखित में से कौन सा एक वर्णनात्मक आँकड़ा नहीं है ?
 (a) माध्य (b) मानक विचलन (c) टी-टेस्ट (d) परास
9. निम्नलिखित में से उसे पहचानें जिसमें डायमैन्शनलिटी में कमी से कम हो जाती है ।
 (a) प्रदर्शन (b) समरैखिकता (c) एन्ट्रॉपी (d) स्टोकेस्टिक्स
10. डाटा से अंतर्दृष्टि प्राप्त करने के लिए, डाटा वैज्ञानिक _____ पर भरोसा करते हैं ।
 (a) डाटा बेअरहाऊस (b) डाटा खनन (c) डाटा विज्युलाइजेशन (d) डाटा एनालिसिस
11. परीक्षण का प्रकार (एक टेल्ड/दो टेल्ड) निम्नलिखित में से किसके द्वारा परिभाषित किया जाता है ?
 (a) शून्य परिकल्पना (b) सरल परिकल्पना (c) वैकल्पिक परिकल्पना (d) संयुक्त परिकल्पना
12. किसी वितरण की उपयुक्तता की अच्छाई का परीक्षण _____ द्वारा किया जाता है ।
 (a) टी-टेस्ट (b) कार्ई-स्क्वायर टेस्ट (c) एफ-टेस्ट (d) पी-टेस्ट
13. _____ अन्वेषणात्मक डाटा विश्लेषण में एक कदम (चरण) नहीं है ।
 (a) डाटा विज्युलाइजेशन (b) डाटा क्लीनिंग
 (c) डाटा सारांशीकरण (d) मॉडल प्रशिक्षण (ट्रेनिंग)

14. In a two-factor ANOVA, how many main effects are there ?
(a) 2 (b) 1 (c) 3 (d) 4
15. In a one-way ANOVA, the factor refers to _____.
(a) A categorical variable with 1 level.
(b) A categorical variable with 2 or more groups.
(c) A continuous variable.
(d) A variable with missing values.
16. Which of the following is not a commonly used sampling technique ?
(a) Parameter Sampling (b) Simple Random Sampling
(c) Systematic Sampling (d) Convenience Sampling
17. Which of the following is a commonly used method for cross-validation ?
(a) K-fold cross-validation (b) ANOVA
(c) Regression Analysis (d) Factorial Design
18. Karl Pearson correlation coefficient is suitable for which type of data ?
(a) Ordinal data (b) Nominal data
(c) Continuous data (d) Categorical data
19. The square root of the variance is called _____.
(a) Empirical data (b) Standard Deviation
(c) Mean (d) Continuous data
20. In a 2×2 factorial design, how many levels are there for each factor ?
(a) 2 levels for each factor (b) 4 levels for each factor
(c) 4 factors and 4 levels (d) 8 levels for each factor
21. Which of the following is a parameter of interest when estimating population proportion ?
(a) Mean (b) Median
(c) Standard Deviation (d) Proportion
22. The rejection probability of NULL Hypothesis when it is true is called as _____.
(a) Level of Confidence (b) Level of Significance
(c) Level of Margin (d) Level of Rejection
23. Which statistical measure is used to describe the central tendency of a numerical variable ?
(a) Mean (b) Standard Deviation
(c) Interquartile range (d) Range
24. A researcher wants to estimate the average height of a student in a University. Which of the following sampling method is most appropriate ?
(a) Stratified sampling (b) Convenience sampling
(c) Cluster sampling (d) Simple Random sampling

14. दो फैक्टर (कारक) एनोवा में, कितने मुख्य प्रभाव होते हैं ?
 (a) 2 (b) 1 (c) 3 (d) 4
15. एक-तरफा एनोवा में, फैक्टर (कारक) का तात्पर्य है _____।
 (a) एक स्तर वाला एक कैटेगोरिकल वेरिएबल
 (b) दो या अधिक समूह वाला एक कैटेगोरिकल (श्रेणीबद्ध) वेरिएबल
 (c) कन्टीन्यूअस वेरिएबल
 (d) मिसिंग वैल्यू वाला एक वेरिएबल
16. निम्नलिखित में से कौन सी तकनीक आमतौर पर इस्तेमाल की जाने वाली नमूनाकरण (सैम्पलिंग) तकनीक नहीं है ?
 (a) पैरामीटर सैम्पलिंग (b) सरल रैन्डम सैम्पलिंग (c) व्यवस्थित सैम्पलिंग (d) सुविधा सैम्पलिंग
17. निम्नलिखित में से कौन सी विधि आमतौर पर क्रॉस वैलिडेशन के लिए इस्तेमाल की जाती है ?
 (a) के-फोल्ड क्रॉस वैलिडेशन (b) एनोवा
 (c) रिग्रेशन एनालिसिस (d) फैक्टोरियल डिजाइन
18. कार्ल पीयरसन कोरिलेशन कोएफिशिएंट किस प्रकार के डाटा के लिए उपयुक्त है ?
 (a) ऑर्डिनल डाटा (b) नॉमिनल डाटा (c) कन्टीन्यूअस डाटा (d) कैटेगोरिकल डाटा
19. प्रसरण (वेरिएंस) के वर्गमूल को _____ कहा जाता है।
 (a) एम्पीरिकल डाटा (b) स्टैण्डर्ड (मानक) डिविएशन
 (c) मीन (d) कन्टीन्यूअस डाटा
20. 2×2 फैक्टोरियल डिजाइन में, प्रत्येक फैक्टर के लिए कितने स्तर होते हैं ?
 (a) प्रत्येक फैक्टर के लिए 2 लेवल (b) प्रत्येक फैक्टर के लिए 4 लेवल
 (c) 4 फैक्टर और 4 लेवल (स्तर) (d) प्रत्येक फैक्टर के लिए 8 लेवल
21. जनसंख्या अनुपात अनुमान लगाते समय कौन सा रुचि का पैरामीटर है ?
 (a) माध्य (b) माध्यिका (c) मानक विचलन (d) अनुपात
22. शून्य परिकल्पना (नल हाइपोथिसिस) के सत्य होने पर उसकी अस्वीकृति की संभावना को _____ कहा जाता है।
 (a) लेवल ऑफ कांफिडेन्स (b) लेवल ऑफ सिग्नीफिकेन्स
 (c) लेवल ऑफ मार्जिन (d) लेवल ऑफ रिजेक्शन
23. संख्यात्मक चर की केन्द्रीय प्रवृत्ति का वर्णन करने के लिए किस सांख्यिकीय माप का उपयोग किया जाता है ?
 (a) मीन (b) स्टैण्डर्ड डेविएशन (मानक विचलन)
 (c) इंटरक्वार्टाईल रेंज (d) रेंज
24. एक शोधकर्ता को एक विश्वविद्यालय में छात्रों की औसत ऊँचाई का अनुमान लगाना है। निम्नलिखित में से कौन सी सैपलिंग विधि सबसे उपयुक्त है ?
 (a) स्तरीकृत (स्ट्रेटीफाइड) सैपलिंग (b) सुविधा (कन्वीनीएन्स) सैपलिंग
 (c) क्लस्टर सैपलिंग (d) सरल रैन्डम सैपलिंग

25. Least Square Method is used in _____.
(a) Linear Regression (b) Linear Polynomial
(c) Linear Sequence (d) Square Polynomial
26. In context of hypothesis testing, identify the correct statement.
(a) Answering Yes/No questions about data
(b) Estimating numerical characteristics of data
(c) Describing associations within the data
(d) Modeling relationships within the data
27. _____ is/are the basic building block(s) of qualitative data.
(a) Data chunk (b) Categories
(c) Numeric figures (d) None of the above
28. _____ statistical process(es) are used for estimating relationship among variable.
(a) Causal only (b) Regression only
(c) Multivariate only (d) Both (a) & (c)
29. Data Discretionary in data pre-processing is primarily used for _____.
(a) Cleaning Data (b) Reducing Data
(c) Transforming Data (d) Controlling access to sensitive data
30. Residual _____ plots investigate normality of the errors.
(a) RR (b) PP (c) QQ (d) SS
31. Which of the following forms the basis for frequency interpretation of probabilities ?
(a) Asymptotics (b) Symptotics (c) Asymmetry (d) All of above
32. Which coefficient represents the slope and intercept of the regression line in linear regression ?
(a) Correlation Coefficient (b) Regression Coefficient
(c) Rank Coefficient (d) Karl-Pearson Coefficient
33. How many fact tables are there in a star schema ?
(a) One (b) Three (c) Two (d) Four
34. Which of the following is transaction driven ?
(a) Data store (b) Data mart
(c) Data warehouse (d) OLTP
35. Full form of ETL is _____.
(a) Execute, Transmit and Load
(b) Extract, Transform and Load
(c) Examine, Transmit and Load
(d) Examine, Transform and Load
36. The primary goal of data pre-processing is _____.
(a) Analysing raw data
(b) Storing data in databases
(c) Preparing data for machine learning algorithm
(d) Generating the data visualizations

25. लीस्ट स्क्वायर मेथड (विधि) _____ में प्रयुक्त होती है ।
 (a) लीनियर रिग्रेशन (b) लीनियर पॉलीनॉमियल (c) लीनियर सीक्वेंस (d) स्क्वायर पॉलीनॉमियल
26. परिकल्पना परीक्षण के संदर्भ में, सही वाक्य को चुनिए ।
 (a) डाटा के बारे में हाँ/ना प्रश्नों का उत्तर देना । (b) डाटा की संख्यात्मक विशेषताओं का अनुमान लगाना ।
 (c) डाटा के अंदर संगठनों का वर्णन करना । (d) डाटा के अंदर संबंधों के नमूने को बताना ।
27. _____ गुणात्मक डाटा के बुनियादी निर्माण खण्ड है/हैं ।
 (a) डाटा खण्ड (b) श्रेणियाँ (c) संख्यात्मक आँकड़े (d) उपरोक्त में से कोई नहीं
28. _____ सांख्यिकीय प्रक्रिया/प्रक्रियाओं के द्वारा वेरिएबल्स के बीच में सम्बन्धों का आकलन किया जाता है ।
 (a) केवल कॉजल (b) केवल रिग्रेशन (c) केवल मल्टीवेरिएट (d) दोनों (a) और (c)
29. डाटा प्री-प्रोसेसिंग में डाटा विविक्तिकरण का उपयोग मुख्य रूप से _____ के लिए किया जाता है ।
 (a) डाटा क्लीनिंग (b) डाटा कम करना
 (c) डाटा बदलना (d) संवेदनशील डाटा तक पहुँच नियंत्रित करना
30. रेसीडुअल (अवशिष्ट) _____ प्लॉट्स त्रुटियों की नोर्मैलिटी की जाँच करते हैं ।
 (a) आर आर (b) पी पी (c) क्यू क्यू (d) एस एस
31. निम्नलिखित में से कौन संभावनाओं (प्रोबेबिलिटी) की आवृत्ति व्याख्या (फ्रीक्वेंसी इन्टरप्रेटेशन) का आधार बनता है ?
 (a) एसिमेट्रिक्स (b) सिमेट्रिक्स (c) एसिमिट्री (विषमता) (d) उपरोक्त सभी
32. कौन सा कोएफीशिएंट लीनियर रिग्रेशन में रिग्रेशन लाइन के स्लोप और इण्टरसेप्ट को संदर्भित करता है ?
 (a) कोरिलेशन कोएफीशिएंट (b) रिग्रेशन कोएफीशिएंट
 (c) रैंक कोएफीशिएंट (d) कार्ल पीयरसन कोएफीशिएंट
33. एक स्टार स्कीमा में कितनी तथ्य तालिकाएँ होती हैं ?
 (a) एक (b) तीन (c) दो (d) चार
34. निम्नलिखित में से कौन सा ट्रान्ज़ैक्शन संचालित है ?
 (a) डाटा भंडार (b) डाटा मार्ट (c) डाटा वेअरहाऊस (d) ओ एल टी पी
35. ETL का पूर्ण रूप _____ है ।
 (a) एक्जिक्यूट, ट्रांसमिट एण्ड लोड (b) एक्सट्रैक्ट, ट्रांसफॉर्म एण्ड लोड
 (c) एक्जामिन, ट्रांसमिट एण्ड लोड (d) एक्जामिन, ट्रांसफॉर्म एण्ड लोड
36. डाटा प्री-प्रोसेसिंग का प्राथमिक लक्ष्य _____ है ।
 (a) कच्चे (रॉ) डाटा का विश्लेषण करना
 (b) डाटाबेस में डाटा संगृहीत करना
 (c) मशीन लर्निंग एल्गोरिदम के लिए डाटा तैयार करना
 (d) डाटा विज़्युलाइजेशन उत्पन्न करना

37. KDD stands for _____.
 (a) Knowledge Data Definition (b) Knowledge Definition Data
 (c) Knowledge Discovery in Database (d) Knowledge Data Discovery
38. These two techniques are used for descriptive analysis of historical data :
 (a) Data Warehousing & Data Science (b) Data Aggregation & Data Visualization
 (c) Data Integration & Data Mining (d) Database & Data Integration
39. Commonly used algorithm for association rule mining is _____.
 (a) Apriori (b) Decision Tree
 (c) Linear Regression (d) K-Nearest Neighbours
40. In data mining, how many categories of functions are included ?
 (a) 5 (b) 4 (c) 2 or 3 (d) 1
41. Which of the following sampling takes consecutive series of items ?
 (a) Block (b) Set (c) Random (d) Heuristic
42. _____ is used to obtain information from unstructured textual data.
 (a) Information access (b) Information retrieval
 (c) Information finding (d) Information storage
43. Predicting the number of newborns, the final number of total newborns can be termed as _____.
 (a) Features (b) Observations (c) Attributes (d) Outcome
44. DMQL stands for _____.
 (a) Data Mart Query Language
 (b) DB Miner Query Language
 (c) Data Mining Query Language
 (d) Datawarehouse Manipulation Query Language
45. Associate Rule Mining is a _____ learning technique.
 (a) Supervised (b) Unsupervised (c) Hybrid (d) Semi-supervised
46. Procedural Domain Knowledge in a rule-based system is in the form of _____.
 (a) Control Rules (b) Meta Rules
 (c) Production Rules (d) Refutation Rules
47. In data analytics-sampling is a process used in _____.
 (a) Network setting (b) Statistical analysis
 (c) Semantic analysis (d) Systematic analysis
48. Data collection is one of the steps in statistical data analytics. The step is performed after _____.
 (a) Model building (b) Model implementation
 (c) Evaluation (d) Business objective
49. Mapping of a class with the use of predefined class is known as _____.
 (a) Data set (b) Data characterization
 (c) Data discrimination (d) Data group
50. Which of the following is not a common data mining task ?
 (a) Classification (b) Regression (c) Clustering (d) Visualisation

37. के डी डी का मतलब _____ है ।
 (a) नॉलेज डाटा डेफिनेशन (b) नॉलेज डेफिनेशन डाटा
 (c) नॉलेज डिस्कवरी इन डाटाबेस (d) नॉलेज डाटा डिस्कवरी
38. निम्नलिखित दो तकनीकों का इस्तेमाल ऐतिहासिक डाटा के वर्णनात्मक विश्लेषण के लिए किया जाता है :
 (a) डाटा वेअरहाऊसिंग और डाटा विज्ञान (b) डाटा एक्त्रीकरण और डाटा विज्युलाइजेशन
 (c) डाटा एकीकरण और डाटा खनन (d) डाटाबेस और डाटा एकीकरण
39. आमतौर पर एसोसिएशन रूल माइनिंग (खनन) के लिए _____ एल्गोरिदम उपयोग किया जाता है ।
 (a) एप्रीऑरी (b) डिसिशन ट्री (c) लीनियर रिग्रेशन (d) के-नीयेरेस्ट नेबर्स
40. डाटा माइनिंग (खनन) में फंक्शन्स की कितनी श्रेणियाँ शामिल हैं ?
 (a) 5 (b) 4 (c) 2 अथवा 3 (d) 1
41. निम्नलिखित में से कौन सा नमूनाकरण वस्तुओं की लगातार श्रृंखला लेता है ?
 (a) ब्लॉक (b) सेट (c) रैन्डम (d) ह्यूरिस्टिक
42. _____ अनस्ट्रक्चर्ड टेक्स्टयूअल डाटा से सूचना प्राप्त करने के लिए इस्तेमाल किया जाता है ।
 (a) इन्फॉर्मेशन एक्सेस (b) इन्फॉर्मेशन रिट्रिवल (c) इन्फॉर्मेशन फाइन्डिंग (d) इन्फॉर्मेशन स्टोरेज
43. नवजात शिशुओं की संख्या की भविष्यवाणी करने वाले उदाहरणों में, नवजात शिशुओं की कुल संख्या को _____ कहते हैं ।
 (a) विशेषताएँ (b) अवलोकन (c) गुण (d) परिणाम
44. डी एम क्यू एल का मतलब _____ है ।
 (a) डाटा मार्ट क्वेरी लैंग्वेज (b) डी बी माइनर क्वेरी लैंग्वेज
 (c) डाटा माइनिंग क्वेरी लैंग्वेज (d) डाटा-वेअरहाऊस मैनिपुलेशन क्वेरी लैंग्वेज
45. एसोसिएट रूल माइनिंग एक _____ लर्निंग तकनीक है ।
 (a) सुपरवाइज्ड (b) अनसुपरवाइज्ड (c) हाइब्रिड (d) सेमी-सुपरवाइज्ड
46. नियम आधारित प्रणाली में प्रक्रियात्मक डोमेन ज्ञान _____ के रूप में होता है ।
 (a) नियंत्रण नियम (b) मेटा नियम (c) उत्पादन नियम (d) खण्डन नियम
47. डाटा एनालिटिक्स में नमूनाकरण प्रक्रिया _____ में इस्तेमाल की जाती है ।
 (a) नेटवर्क सेटिंग (b) सांख्यिकी एनालिसिस (c) शब्दार्थ एनालिसिस (d) व्यवस्थित एनालिसिस
48. डाटा संग्रह सांख्यिकीय डाटा विश्लेषणों के चरणों में से एक है । यह _____ चरण के बाद किया जाता है ।
 (a) मॉडल बनाना (b) मॉडल इम्प्लीमेंटेशन (c) मूल्यांकन (d) व्यावसायिक उद्देश्य
49. एक क्लास की मैपिंग एक प्री-डिफाइन्ड क्लास के उपयोग के साथ करने को _____ कहा जाता है ।
 (a) डाटा सेट (b) डाटा केरेक्ट्राइजेशन (c) डाटा डिसक्रिमिनेशन (d) डाटा ग्रुप
50. निम्नलिखित में से कौन सा सामान्य डाटा माइनिंग कार्य नहीं है ?
 (a) क्लासीफिकेशन (b) रिग्रेशन (c) क्लस्टरिंग (d) विज्युलाइजेशन

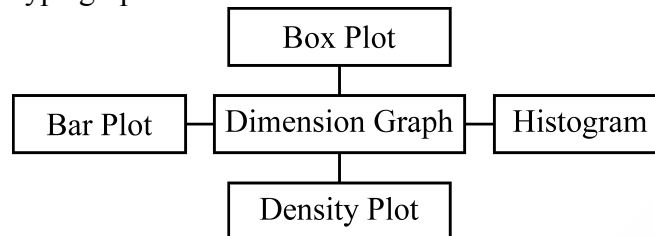
51. Which of the following is not a common data mining algorithm ?
 (a) K-Means clustering (b) Linear Regression
 (c) Apriori Algorithm (d) Support Vector Machine
52. Which of the clustering technique make use of merging approach ?
 (a) Partitioned (b) Naïve Bayes (c) Hierarchical (d) Both (a) & (c)
53. Which of the following is not used for data visualisation ?
 (a) Fever maps (b) Bubble charts (c) Bullet graphs (d) Heat maps
54. Which of the following is a data-visualization tool ?
 (a) Attensity (b) SAS (c) Clarabridge (d) Intensity
55. Which of the following provides state-of-art technology to store and process big data ?
 (a) Yahoo (b) Google (c) Twitter Map (d) Hadoop
56. Visual Analytics is the science of _____ supported by the interactive visual interface.
 (a) Optics (b) Computers
 (c) Analytical Reasoning (d) Analog Signal
57. Visualisation of data produces images that are filtered with the help of _____ technique.
 (a) Clutter Reduction (b) Predictive Analysis
 (c) Logical Reasoning (d) Open Source
58. Which of the following features of virtualization means that a virtual machine can be replaced by using a single file ?
 (a) Partitioning (b) Isolation
 (c) Encapsulation (d) Application Virtualization
59. A pie-chart is often referred to as _____.
 (a) Bar chart (b) Pair plot (c) Doughnut chart (d) Circle chart
60. Which of the following is used to plot the distribution of a continuous variable ?
 (a) Bar chart (b) Box chart (c) Scatter plot (d) Histogram
61. Visualization is a core-part for which of the data-mining technique ?
 (a) Link Analysis (b) Database Segmentation
 (c) Reinforcement (d) Link Segmentation
62. _____ graph displays information as a series of data points connected by a straight line.
 (a) Bar (b) Scatter (c) Histogram (d) Line
63. _____ are drawn with respect to the histogram created.
 (a) Frequency polygon (b) Box plot
 (c) Bar plot (d) Doughnut plot
64. Which method shows hierarchical data in a nested format ?
 (a) Tree maps (b) Scatter plots (c) Population pyramids (d) Area charts
65. _____ plot is also described as five-number summary plot.
 (a) Frequency polygon (b) Box plot
 (c) Histogram (d) Scatter plot

51. निम्नलिखित में से कौन सा सामान्य डाटा माइनिंग एल्गोरिद्म नहीं है ?
 (a) के-मीन्स क्लस्टरिंग (b) लीनियर रिग्रेशन (c) एप्रीऑरी एल्गोरिद्म (d) सपोर्ट वेक्टर मशीन
52. क्लस्टरिंग तकनीक में से किसको विलय (मरजिंग) अप्रोच की आवश्यकता है ?
 (a) विभाजन (b) नेव बेअस (बेयस) (c) हायरार्किकल (d) (a) और (c) दोनों
53. निम्नलिखित में से किसका उपयोग डाटा विज्युलाइजेशन में नहीं किया जाता है ?
 (a) फीवर मैप्स (b) बबल चार्ट्स (c) बुलेट ग्राफ्स (d) हीट मैप्स
54. निम्नलिखित में से कौन सा डाटा-विज्युलाइजेशन टूल है ?
 (a) अटेंसिटी (b) एस ए एस (SAS) (c) क्लारब्रिज (d) इंटेसिटी
55. निम्नलिखित में से कौन सी अत्याधुनिक तकनीक बिग डाटा को स्टोर और प्रोसेस करने के लिए उपयोग की जाती है ?
 (a) याहू (b) गूगल (c) ट्विटर मैप (d) हडूप
56. दृश्य विश्लेषण इंटरैक्टिव दृश्य इंटरफेस द्वारा समर्थित _____ का विज्ञान है ।
 (a) ऑप्टिक्स (b) कम्प्यूटर (c) विश्लेषणात्मक तर्क (d) एनालॉग सिग्नल
57. डाटा के विज्युलाइजेशन द्वारा उत्पादित छवि को _____ तकनीक से फिल्टर किया जाता है ।
 (a) क्लटर रिडक्सन (b) प्रीडिक्टव एनालिसिस (c) लॉजिकल रीजनिंग (d) ओपेन सोर्स
58. निम्नलिखित वर्चुअलाइजेशन विशेषता के द्वारा एक वर्चुअल मशीन को एकल फाइल से बदला जा सकता है :
 (a) विभाजन (b) आइसोलेशन
 (c) एनकैप्सुलेशन (d) एप्लीकेशन वर्चुअलाइजेशन
59. एक पाई चार्ट को अकसर _____ भी कहा जाता है ।
 (a) बार चार्ट (b) पेयर प्लॉट (c) डोनट चार्ट (d) सर्किल चार्ट
60. एक संतत चर के वितरण को आलेखित करने के लिए निम्नलिखित में से किसका उपयोग किया जाता है ?
 (a) बार चार्ट (b) बॉक्स चार्ट (c) स्कैटर प्लॉट (d) हिस्टोग्राम
61. विज्युलाइजेशन डाटा माइनिंग की किस तकनीक का मुख्य भाग है ?
 (a) लिंक विश्लेषण (b) डाटाबेस विभाजन (c) रि-इन्फोर्समेंट (d) लिंक विभाजन
62. _____ ग्राफ सीधी रेखा से जुड़े डाटा बिंदुओं की एक शृंखला को जानकारी के रूप में प्रदर्शित करता है ।
 (a) बार (b) स्कैटर (c) हिस्टोग्राम (d) लाइन
63. _____ बनाए गये हिस्टोग्राम के संबन्ध में तैयार किए जाते हैं ।
 (a) फ्रीक्वेंसी पॉलीगॉन (b) बॉक्स प्लॉट (c) बार प्लॉट (d) डोनट प्लॉट
64. कौन सी विधि नेस्टेड प्रारूप में हायरार्किकल डाटा दर्शाती है ?
 (a) ट्री मैप्स (b) स्कैटर प्लॉट्स (c) जनसंख्या पिरामिड्स (d) एरिया चार्ट
65. _____ प्लॉट को पाँच नंबर सारांश प्लॉट के रूप में भी वर्णित किया जाता है ।
 (a) फ्रीक्वेंसी पॉलीगॉन (b) बॉक्स प्लॉट (c) हिस्टोग्राम (d) स्कैटर प्लॉट

66. _____ plots are often used for checking randomness in time-series.

- (a) Auto correlation (b) Auto rank
- (c) Auto causation (d) Auto search

67. Which dimension type graph is shown below ?



- (a) 3 – Dimension (b) 1 – Dimension (c) 2 – Dimension (d) 0 – Dimension

68. Mongo DB is a _____ database.

- (a) SQL (b) DBMS (c) NO SQL (d) RDBMS

69. Which of the following is not a characteristic of Big Data ?

- (a) Volume (b) Velocity (c) Variety (d) Validation

70. Common Big Data Framework used in IT Industry are _____.

- (a) Python and Java (b) Hadoop and Spark
- (c) Linux and Windows (d) SQL and NO SQL

71. _____ is a component on top of spark core.

- (a) Spark streaming (b) RDD
- (c) Spark SQL (d) Spark Gift

72. _____ is used to process and analyze Big Data.

- (a) Spreadsheet Software (b) Relational Database
- (c) Apache-Hadoop (d) Microsoft Windows

73. Spark was initially started by _____ at UC Berkeley.

- (a) Mark Zuckerberg (b) Matei Zaharia
- (c) James Gosling (d) Doug Cutting

74. The size of a HDFS block is _____.

- (a) 16 MB (b) 32 MB (c) 1 GB (d) 64 MB

75. _____ manages data-traffic between virtual and physical machine.

- (a) Hypervisor (b) Map reduce (c) Hidesign (d) Mobile network

76. _____ is managed by Map Reduce.

- (a) Image (b) Unstructured data (c) Web logs (d) Users

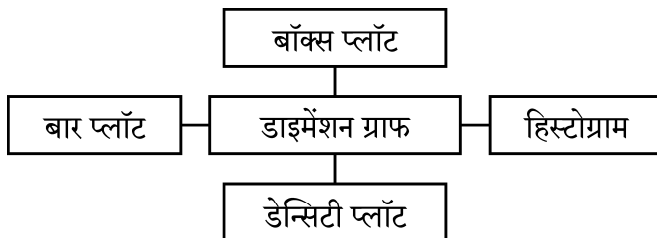
77. _____ command is used in HDFS for block directives.

- (a) FSCK (b) FCSK (c) FKSC (d) FKCS

78. HDFS stands for _____.

- (a) Hadoop's Division File System (b) Hadoop's Distributed File System
- (c) Hadoop's Data File System (d) Highly Dense File System

66. टाइम सीरीज़ में रैंडमनेस चेक करने के लिए _____ प्लॉट्स का प्रयोग प्रायः किया जाता है ।
 (a) ऑटो कोरिलेशन (b) ऑटो रैंक (c) ऑटो कॉजेशन (d) ऑटो सर्च
67. कौन से आयाम प्रकार का ग्राफ नीचे दर्शाया गया है ?



- (a) 3-डाइमेंशन (b) 1-डाइमेंशन (c) 2-डाइमेंशन (d) 0-डाइमेंशन
68. मोनो डी बी एक _____ डेटाबेस है ।
 (a) एस क्यू एल (b) डी बी एम एस (c) नो एस क्यू एल (d) आर डी बी एम एस
69. निम्नलिखित में से बिग डाटा की विशेषता नहीं है :
 (a) वॉल्यूम (आयतन) (b) वेलोसिटी (वेग) (c) वैरायटी (विविधता) (d) वैलिडेशन (सत्यापन)
70. आई टी इण्डस्ट्री में सामान्यतः उपयोग किए जाने वाले बिग डाटा फ्रेमवर्क हैं
 (a) पायथन और जावा (b) हडूप और स्पार्क
 (c) लाइनेक्स और विंडोज (d) एस क्यू एल और नो एस क्यू एल
71. _____ स्पार्क कोर के शीर्ष पर एक कम्पोजेंट (घटक) है ।
 (a) स्पार्क स्ट्रीमिंग (b) आर डी डी (c) स्पार्क एस क्यू एल (d) स्पार्क गिफ्ट
72. _____ का उपयोग बिग डाटा को प्रोसेस और विश्लेषण में किया जाता है ।
 (a) स्प्रेडशीट सॉफ्टवेयर (b) रिलेशनल डाटाबेस (c) अपाचे हडूप (d) माइक्रोसॉफ्ट विंडोज
73. स्पार्क की शुरुआत यू सी बरक्ले में _____ द्वारा की गई थी ।
 (a) मार्क जुकरबर्ग (b) माटेई जहरिया (c) जेम्स गोसलिंग (d) डोग कटिंग
74. HDFS (एच डी एफ एस) ब्लॉक का आकार _____ है ।
 (a) 16 MB (b) 32 MB (c) 1 GB (d) 64 MB
75. _____ वर्चुअल मशीन व फिजिकल मशीन के बीच डाटा ट्रैफिक का प्रबन्धन करता है ।
 (a) हाइपरवाइजर (b) मैप रेड्यूस (c) हाइडिज़ाइन (d) मोबाइल नेटवर्क
76. _____ मैप रेड्यूस द्वारा प्रबन्धन किया जाता है ।
 (a) इमेज (b) अनस्ट्रक्चर्ड डाटा (c) वेब लॉग्स (d) यूजर्स
77. _____ कमाण्ड एच डी एफ एस ब्लॉक को दिशा-निर्देश प्रदान करता है ।
 (a) FSCK (b) FCSK (c) FKSC (d) FKCS
78. HDFS का तात्पर्य है
 (a) हडूप डिविजन फाइल सिस्टम (b) हडूप डिस्ट्रीब्यूटेड फाइल सिस्टम
 (c) हडूप डाटा फाइल सिस्टम (d) हाईली डेंस फाइल सिस्टम

79. Procedural Domain Knowledge in a rule based system, is the form of _____.
 (a) Control Rules (b) Meta Rules (c) Production Rules (d) Beta Rules
80. Reporting does not involve _____.
 (a) Charts (b) Graphs
 (c) Tables (d) Predictive Model
81. The data node and name node in Hadoop are _____.
 (a) Worker node and Master node (b) Both are Worker nodes
 (c) Both are Master nodes (d) Master node and Worker node
82. _____ package is used to write programming code for Hadoop's Map Reduce Model directly.
 (a) HDFS (b) Map visor (c) rmr (d) RJSONIO
83. Following file system provides read-only access to HDFS over HTTPS :
 (a) HAR (b) HDFS
 (c) HXTP (d) None of the above
84. Which of the following is not a Big Data Technology ?
 (a) Apache Spark (b) Apache Pytorch (c) Apache Hadoop (d) Apache Kafka
85. Which computer technology uses the virtual resources on a dispersed network ?
 (a) Distributed computing (b) Cloud computing
 (c) Soft computing (d) Parallel computing
86. Following Command Line Interface is used to interact with HDFS :
 (a) HDFS shell (b) DFS shell (c) K shell (d) FS shell
87. The data which has already been collected by someone else is called _____ data.
 (a) Primary (b) Secondary (c) Quantitative (d) Qualitative
88. Missing value treatment to the data is _____.
 (a) necessary to get correct results (b) should never be practiced
 (c) to drop some missing values (d) not required in Big Data
89. The _____ distance is computed for those space where points are represented as the direction of vector components.
 (a) Euclidean (b) Hamiltonian (c) Union (d) Cosine
90. In HDFS cluster _____ manages the cluster metadata.
 (a) Name Node (b) Data Node (c) I Node (d) Cluster Node
91. Which of the following is a hypothesis testing method ?
 (a) Black Box Test (b) Chi-square Test (c) White Box Test (d) Mutation Test
92. The architecture of spark contains _____.
 (a) API and Data storage
 (b) API, Resource Manager and Data Storage
 (c) Resource Manager and API
 (d) Data storage and Resource Manager

79. रूल बेस्ड सिस्टम में एक प्रोसीजरल डोमेन नॉलेज _____ के रूप में है ।
 (a) कंट्रोल रूल्स (b) मेटा रूल्स (c) प्रोडक्शन रूल्स (d) बीटा रूल्स
80. _____ रिपोर्टिंग में निहित नहीं है ।
 (a) चार्टस (b) ग्राफ्स (c) टेबल्स (d) प्रीडिक्टिव मॉडल
81. हडूप में डाटा नोड और नाम नोड हैं
 (a) वर्कर नोड और मास्टर नोड (b) दोनों ही वर्कर नोड
 (c) दोनों ही मास्टर नोड (d) मास्टर नोड और वर्कर नोड
82. _____ पैकेज द्वारा हडूप के मैप रेड्यूस मॉडल में सीधे ही प्रोग्रामिंग कोड लिखे जाते हैं ।
 (a) एच डी एफ सी (b) मैप वाइजर
 (c) आर एम आर (d) आर जे एस ओ एन आई ओ
83. निम्नलिखित फाइल सिस्टम एच टी टी पी एस पर एच डी एफ एस को केवल पढ़ने के लिए पहुँच प्रदान करता है :
 (a) एच ए आर (b) एच डी एफ सी (c) एच एक्स टी पी (d) उपरोक्त में से कोई नहीं
84. निम्नलिखित में से कौन बिग डाटा टेक्नॉलॉजी नहीं है ?
 (a) अपाचे स्पार्क (b) अपाचे पाइटॉर्च (c) अपाचे हडूप (d) अपाचे काफका
85. एक डिस्पर्स नेटवर्क पर वर्चुअल रीसोर्सेज का उपयोग करने वाली कम्प्यूटर तकनीक कौन सी है ?
 (a) डिस्ट्रीब्यूटेड कम्प्यूटिंग (b) क्लाउड कम्प्यूटिंग
 (c) सॉफ्ट कम्प्यूटिंग (d) पैरेलल कम्प्यूटिंग
86. निम्नलिखित कमाण्ड लाइन इंटरफेस के द्वारा एच डी एफ एस के साथ इंटरएक्ट किया जाता है :
 (a) एच डी एफ एस शेल (b) डी एफ एस शेल
 (c) के शेल (d) एफ एस शेल
87. जो डाटा पहले से ही किसी के द्वारा एकत्रित किया गया हो _____ डाटा कहलाता है ।
 (a) प्राइमरी (b) सेकण्डरी (c) क्वांटिटेटिव (d) क्वालिटेटिव
88. डाटा में अनुपलब्ध (लापता) वैल्यू के उपचार _____ है ।
 (a) सही परिणाम प्राप्त करने के लिए आवश्यक (b) कभी प्रयोग नहीं करना चाहिए
 (c) कुछ लापता वैल्यू को छोड़ देना चाहिए (d) बिग डाटा में जरूरत नहीं
89. _____ दूरी की गणना उन स्थानों के लिए की जाती है जहाँ बिन्दुओं को वेक्टर घटकों की दिशा के रूप में दर्शाया जाता है ।
 (a) यूक्लीडियन (b) हैमिल्टोनियन (c) यूनियन (d) कोसाइन
90. एच डी एफ एस क्लस्टर में _____ क्लस्टर मेटाडाटा को व्यवस्थित करता है ।
 (a) नेम नोड (b) डेटा (डाटा) नोड (c) आई नोड (d) क्लस्टर नोड
91. निम्नलिखित में से कौन परिकल्पना परीक्षण की विधि है ?
 (a) ब्लैक बॉक्स टेस्ट (b) कार्ड-स्क्वायर टेस्ट (c) वाइट बॉक्स टेस्ट (d) म्यूटेशन टेस्ट
92. स्पार्क की संरचना का हिस्सा है
 (a) ए पी आई और डाटा स्टोरेज (b) ए पी आई, रीसोर्स मैनेजर और डाटा स्टोरेज
 (c) रीसोर्स मैनेजर और ए पी आई (d) डाटा स्टोरेज और रीसोर्स मैनेजर

93. Which of the following scheme provides best performance ?
 (a) Hashing by division (b) Hashing by multiplication
 (c) Universal hashing (d) Open hashing
94. HDFS operates in a _____ manner.
 (a) Master – Worker Architecture (b) Master – Master Architecture
 (c) Peer – Peer Architecture (d) Worker – Slave Architecture
95. Apache Hive is an example of _____ tool.
 (a) Mapping (b) Querying (c) Map Reduce (d) Input
96. Which of the following is not a component of the Hadoop Ecosystem ?
 (a) HDFS (b) Spark (c) Hive (d) SQL Server
97. _____ has the world's largest Hadoop cluster.
 (a) Apple (b) Datamatics (c) Facebook (d) Informatica
98. _____ is a supervised machine learning model used for text classification.
 (a) Naïve-Bayes Classifier (b) K-Means Classifier
 (c) Logistic Classifier (d) Simple Classifier
99. Identify a generative model used in machine learning.
 (a) Logistic Regression (b) Support Vector Machine
 (c) Linear Regression (d) Naïve Bayes
100. Which of the following is not a supervised learning algorithm ?
 (a) Naïve – Bayesian (b) Linear Regression
 (c) PCA (d) Decision Tree
101. Statistical technique for parameter estimation is commonly known as _____.
 (a) Minimum likelihood estimation (b) Maximum likelihood estimation
 (c) Bayesian estimation (d) Poisson estimation
102. In Bayesian networks, the directed edges represent _____.
 (a) dependency between variables (b) conditional probability
 (c) prior probability (d) independence probability
103. Following Bayesian algorithm is used for solving optimization problem :
 (a) Naïve Bayes (b) Bayesian Network
 (c) Bayesian-Linear Regression (d) Bayesian Optimization
104. Following Algorithm is used to estimate the performance of a machine-learning model :
 (a) Cross-validation (b) Clustering
 (c) Soft Computing (d) Swarm Intelligence
105. In information theory, the entropy of a random variable is the average level of _____.
 (a) Probability (b) Uncertainty
 (c) Certainty (d) Sum of its values
106. _____ does not fall under document classification.
 (a) Identifying text in a document (b) Categorization of document
 (c) Reading the document (d) Tagging

93. निम्नलिखित में कौन सी विधि श्रेष्ठ प्रदर्शन प्रदान करती है ?
 (a) विभाजन द्वारा हैशिंग (b) गुणन द्वारा हैशिंग
 (c) यूनिवर्सल हैशिंग (d) ओपन हैशिंग
94. एच डी एफ एस _____ तरीके से संचालित होती है ।
 (a) मास्टर-वर्कर आर्किटेक्चर (b) मास्टर-मास्टर आर्किटेक्चर
 (c) पीयर-पीयर आर्किटेक्चर (d) वर्कर-स्लेव आर्किटेक्चर
95. अपाचे हाइव _____ टूल का एक उदाहरण है ।
 (a) मैपिंग (b) क्वेयरिंग (c) मैप रेड्यूस (d) इनपुट
96. निम्नलिखित में से कौन हडूप इकोसिस्टम का घटक नहीं है ?
 (a) एच डी एफ एस (b) स्पार्क (c) हाइव (d) एस क्यू एल सर्वर
97. _____ के पास विश्व का सबसे बड़ा हडूप क्लस्टर है ।
 (a) एप्पल (b) डेटामैटिक्स (c) फेसबुक (d) इन्फॉमैटिका
98. _____ एक सुपरवाइज्ड मशीन लर्निंग मॉडल को टेक्स्ट वर्गीकरण के लिए प्रयोग किया जाता है ।
 (a) नेव-बेयस क्लासीफायर (b) के-मीन्स क्लासीफायर
 (c) लॉजिस्टिक क्लासीफायर (d) सरल (सिंपल) क्लासीफायर
99. मशीन लर्निंग में उपयोग किए जाने वाले जनरेटिव मॉडल की पहचान करें ।
 (a) लॉजिस्टिक रिग्रेशन (b) सपोर्ट वेक्टर मशीन (c) लीनियर रिग्रेशन (d) नेव-बेयस
100. निम्नलिखित में कौन सा एल्गोरिद्म सुपरवाइज्ड लर्निंग नहीं उपयोग करता है ?
 (a) नेव-बेयसियन (b) लीनियर रिग्रेशन (c) पी सी ए (d) डिसिशन (डिसिज़न) ट्री
101. पैरामीटर आकलन के लिए सांख्यिकीय तकनीक को सामान्यतया _____ कहा जाता है ।
 (a) न्यूनतम संभावना अनुमान (b) अधिकतम संभावना अनुमान
 (c) बेयसियन अनुमान (d) प्वासों अनुमान
102. बेयसियन नेटवर्क में निर्देशित किनारे (डायरेक्टेड एजेंस) _____ दर्शाते हैं ।
 (a) वेरियेबल्स के बीच निर्भरता (b) कंडीशनल प्रोबेबिलिटी
 (c) प्रायर प्रोबेबिलिटी (d) स्वतंत्र प्रोबेबिलिटी (आत्मनिर्भर)
103. निम्नलिखित बेयसियन एल्गोरिद्म का उपयोग ऑप्टिमाइजेशन समस्या हल के लिए किया जाता है :
 (a) नेव-बेयस (b) बेयसियन नेटवर्क
 (c) बेयसियन लीनियर रिग्रेशन (d) बेयसियन ऑप्टिमाइजेशन
104. निम्न एल्गोरिद्म के द्वारा मशीन लर्निंग मॉडल के कौशल का अनुमान लगाया जाता है :
 (a) क्रॉस वेलिडेशन (b) क्लस्टरिंग (c) सॉफ्ट कम्प्यूटिंग (d) स्वार्म इंटेलेजेंस
105. सूचना सिद्धान्त में रैंडम वेरिएबल की एन्ट्रॉपी _____ का औसत स्तर होता है ।
 (a) प्रोबेबिलिटी (b) असर्टेनिटी (c) सर्टेनिटी (d) इसकी संख्याओं का जोड़
106. _____ डॉक्यूमेन्ट वर्गीकरण के अंतर्गत नहीं आता है ।
 (a) डॉक्यूमेन्ट में टेक्स्ट की पहचान करना (b) डॉक्यूमेन्ट को कैटेग्राइज करना
 (c) डॉक्यूमेन्ट पढ़ना (d) टैगिंग

107. Multiclass Logistic Regression is also known as
(a) Mutual Information Method
(b) Maximum Entropy Method
(c) Minimum Entropy Method
(d) Average Entropy Method
108. The Bayes rule can be used for _____.
(a) increasing the complexity
(b) decreasing the complexity
(c) answering probabilistic query
(d) reducing the data
109. _____ is a multinomial model.
(a) Dirichlet Model
(b) Erlang Model
(c) Binomial Distribution Model
(d) Tree Model
110. _____ technique is used for classifying text documents.
(a) Bag of Alphabets
(b) Bag of Numbers
(c) Bag of Words
(d) Bag of Alphanumeric
111. The probabilities of the states before obtaining sample information are called _____.
(a) Prior probabilities
(b) Join probabilities
(c) Conditional probabilities
(d) Revised probabilities
112. Statistical technique for parameter estimation is commonly known as _____.
(a) Minimum Likelihood Estimation
(b) Maximum Likelihood Estimation
(c) Bayesian Estimation
(d) Bayesian Belief Network Estimation
113. What does K stands for in K-Means algorithm ?
(a) Number of Clusters
(b) Number of Attributes
(c) Number of Data
(d) Number of Iterations
114. Identify the clustering method which takes care of the variance in data.
(a) Gaussian Mixture Model
(b) Decision Tree
(c) K-Means
(d) Gaussian Blind Model
115. Clustering belongs to _____ data analysis.
(a) supervised
(b) unsupervised
(c) early
(d) late
116. The CPN provides practical approach for implementing
(a) Pattern Approximation
(b) Pattern Classification
(c) Pattern Mapping
(d) Pattern Clustering
117. For clustering we do not require _____.
(a) Unlabelled data
(b) Categorical data
(c) Numerical data
(d) Labelled data
118. Which of the following is not a distance metric used in hierarchical clustering ?
(a) Euclidean distance
(b) Manhattan distance
(c) Cosine similarity
(d) Pearson correlation coefficient

107. मल्टीक्लास लॉजिस्टिक रिग्रेशन को इस नाम से भी जाना जाता है
 (a) पारस्परिक सूचना विधि (b) अधिकतम एन्ट्रॉपी विधि
 (c) न्यूनतम एन्ट्रॉपी विधि (d) औसत एन्ट्रॉपी विधि
108. बेयस नियम का उपयोग _____ के लिए किया जा सकता है ।
 (a) जटिलता बढ़ाने (b) जटिलता घटाने
 (c) प्रोबेबिलिस्टिक क्वेरी का उत्तर देने (d) डाटा को कम करने
109. _____ एक मल्टीनॉमियल मॉडल है ।
 (a) डीरीख्लेट मॉडल (b) अलॉग मॉडल
 (c) बाइनॉमियल डिस्ट्रीब्यूशन मॉडल (d) ट्री मॉडल
110. _____ तकनीक द्वारा टेक्सट डॉक्यूमेंट्स का वर्गीकरण होता है ।
 (a) बैग ऑफ़ एल्फाबेट्स (b) बैग ऑफ़ नम्बर्स
 (c) बैग ऑफ़ वर्ड्स (d) बैग ऑफ़ एल्फान्यूमेरिक
111. सैम्पल जानकारी (इन्फॉर्मेशन) प्राप्त करने से पहले की प्रोबेबिलिटीस को _____ कहा जाता है ।
 (a) प्रायर प्रोबेबिलिटीस (b) ज्वाइन प्रोबेबिलिटीस
 (c) कंडीशनल प्रोबेबिलिटीस (d) रिवाइज्ड प्रोबेबिलिटीस
112. पैरामीटर आकलन के लिए सामान्यतया _____ सांख्यिकीय तकनीक का उपयोग किया जाता है ।
 (a) न्यूनतम संभावना अनुमान (b) अधिकतम संभावना अनुमान
 (c) बेयसियन अनुमान (d) बेयसियन बिलीफ नेटवर्क अनुमान
113. K-माध्य एल्गोरिथ्म (एल्गोरिद्म) में K किसको दर्शाता है ?
 (a) समूहों की संख्या (b) गुणों की संख्या (c) डाटा की संख्या (d) पुनरावृत्ति की संख्या
114. क्लस्टरिंग विधि जो डाटा में भिन्नता का ध्यान रखती है
 (a) गॉसियन मिश्रण मॉडल (b) डिसिशन ट्री
 (c) K-मीन्स (d) गॉसियन ब्लाइण्ड मॉडल
115. क्लस्टरिंग _____ डाटा विश्लेषण से संबंधित है ।
 (a) सुपरवाइज्ड (b) अनसुपरवाइज्ड (c) पहल (अर्ली) (d) देर (लेट)
116. सी.पी.एन. (CPN) _____ कार्यान्वयन के लिए व्यावहारिक (प्रेक्टिकल) दृष्टिकोण प्रदान करता है ।
 (a) पैटर्न एप्रोक्सिमेशन (b) पैटर्न क्लासीफिकेशन (c) पैटर्न मैपिंग (d) पैटर्न क्लस्टरिंग
117. क्लस्टरिंग के लिए हमें _____ की आवश्यकता नहीं है ।
 (a) लेबल रहित डाटा (b) कैटेगिरी (वर्गीकृत) डाटा
 (c) संख्यात्मक डाटा (d) लेबल किया गया डाटा
118. निम्नलिखित में से हायरारिकल क्लस्टरिंग में उपयोग की जाने वाली डिस्टेंस मैट्रिक्स नहीं है :
 (a) यूक्लीडियन डिस्टेंस (b) मैनहट्टन डिस्टेंस
 (c) कोसाइन समानता (d) पीयरसन कोरिलेशन कोएफिशिएंट

119. Which clustering algorithm is sensitive to the order of the data points ?
 (a) GMM (b) K-Means (c) Agglomerative (d) DBSCAN
120. The algorithm most sensitive to outliers is _____.
 (a) K-Means (b) K-Modes (c) K-Medians (d) K-Values
121. Algorithm is used to handle clusters of varying shapes and sizes.
 (a) DBSCAN (b) K-Means (c) Mean – Shift (d) Agglomerative
122. _____ clustering algorithm is based on the concept of centroids.
 (a) Mean – Shift (b) DBSCAN (c) K-Means (d) Agglomerative
123. How is a hypothesis represented in concept learning ?
 (a) Scalar only (b) Vector only
 (c) Polynomial (d) Either Scalar or Vector
124. _____ is a generative model used in machine learning.
 (a) Support Vector Machine (b) Naïve Bayes
 (c) Logistic Regression (d) Linear Regression
125. What is the Manhattan Distance between a data point (9, 7) and a new query instance (3, 4) ?
 (a) 7 (b) 4 (c) 3 (d) 9
126. Which of the following is not based on Decision Tree Algorithm ?
 (a) ID3 (b) C 4.5 (c) DBSCAN (d) CART
127. What is the objective of regression analysis ?
 (a) To predict categorical outcomes. (b) To predict continuous outcomes.
 (c) To classify data into different classes. (d) To cluster data points.
128. In binary classification, what does F1 score represent ?
 (a) Harmonic mean of precision and recall
 (b) Arithmetic mean of precision and recall
 (c) Geometric mean of precision and recall
 (d) Product of precision and recall
129. SVM stands for _____.
 (a) Super Versatile Machine (b) Support Vector Machine
 (c) Simple Vector Machine (d) Share Variables Machine
130. What is the disadvantage of decision tree ?
 (a) Prone to overfit
 (b) Robust to outliers
 (c) Cannot perform classification
 (d) Cannot perform regression
131. Singular value decomposition is used in _____.
 (a) Lower Simple Indexing (b) Lateral Simple Indexing
 (c) Latent Semantic Indexing (d) Latent Syntax Indexing
132. In pattern recognition system, the most popular distance metric is _____.
 (a) Manhattan (b) Mean (c) Minkowski (d) Euclidean

119. कौन सी क्लस्टरिंग एल्गोरिद्म डाटा बिन्दुओं के क्रम के प्रति संवेदनशील है ?
 (a) जी.एम.एम. (GMM) (b) के-मीन्स
 (c) एग्लोमेराटिव (d) डी.बी.एस.सी.ए.एन.
120. आउटलायर्स से प्रभावित अत्यंत संवेदनशील _____ एल्गोरिद्म है ।
 (a) के-मीन्स (b) के-मोड्स (c) के-मीडियन (d) के-वैल्यूस
121. _____ एल्गोरिद्म का उपयोग क्लस्टरों के विभिन्न आकार एवं आकृति में किया जाता है ।
 (a) डी.बी.एस.सी.ए.एन. (b) के-मीन्स
 (c) मीन-शिफ्ट (d) एग्लोमेराटिव
122. _____ क्लस्टरिंग एल्गोरिद्म कॉन्सेप्ट ऑफ सेंटरायड (केन्द्रक) पर आधारित है ।
 (a) मीन-शिफ्ट (b) डी.बी.एस.सी.ए.एन. (c) के-मीन्स (d) एग्लोमेराटिव
123. कॉन्सेप्ट (अवधारणा) लर्निंग में एक हाइपोथिसिस (परिकल्पना) को कैसे दर्शाया जाता है ?
 (a) सिर्फ स्कैलर (b) सिर्फ वेक्टर (c) पॉलिनामियल (d) या तो स्कैलर या वेक्टर
124. _____ मशीन लर्निंग में उपयोग किया जाने वाला एक जेनरेटिव मॉडल है ।
 (a) सपोर्ट वेक्टर मशीन (b) नेव बेयस
 (c) लॉजिस्टिक रिग्रेशन (d) लीनियर रिग्रेशन
125. डाटा बिन्दु (9, 7) और नई क्वेरी इंस्टांस (3, 4) के बीच की मैनहट्टन दूरी क्या है ?
 (a) 7 (b) 4 (c) 3 (d) 9
126. निम्नलिखित में से कौन डिजिटल ट्री एल्गोरिद्म पर आधारित नहीं है ?
 (a) आई.डी. 3 (b) सी 4.5 (c) डी.बी.एस.सी.ए.एन. (d) सी.ए.आर.टी.
127. रिग्रेशन विश्लेषण का उद्देश्य क्या है ?
 (a) श्रेणीबद्ध परिणामों की भविष्यवाणी करना । (b) निरंतर परिणामों की भविष्यवाणी करना ।
 (c) डाटा को विभिन्न वर्गों में वर्गीकृत करना । (d) डाटा बिन्दुओं को क्लस्टर करने के लिए ।
128. बाइनरी वर्गीकरण में एफ-1 स्कोर क्या दर्शाता है ?
 (a) प्रिसिशन और रिकॉल का हार्मोनिक मीन (b) प्रिसिशन और रिकॉल का अर्थमैटिक मीन
 (c) प्रिसिशन और रिकॉल का जियोमैट्रिक मीन (d) प्रिसिशन और रिकॉल का गुणनफल
129. एस.वी.एम. का तात्पर्य है
 (a) सुपर वर्सेटाइल मशीन (b) सपोर्ट वेक्टर मशीन
 (c) सिंपल वेक्टर मशीन (d) शेयर वेरिएबल्स मशीन
130. डिजिटल ट्री के प्रति क्या प्रतिकूल (अनुपयोगी) है ?
 (a) ओवरफिट के प्रति प्रोन है । (b) आउटलायर्स के लिए रॉबस्ट है ।
 (c) क्लासीफिकेशन (वर्गीकृत) नहीं कर सकते हैं । (d) रिग्रेशन नहीं कर सकते हैं ।
131. सिंगुलर वैल्यू डिक्वॉम्पोजिशन _____ में प्रयुक्त होता है ।
 (a) लोअर सिंपल इंडेक्सिंग (b) लैटरल सिंपल इंडेक्सिंग
 (c) लैटेंट सीमेंटिक इंडेक्सिंग (d) लैटेंट सिन्टेक्स इंडेक्सिंग
132. नमूना पहचान प्रणाली में, सबसे लोकप्रिय मापीय दूरी _____ है ।
 (a) मैनहट्टन (b) मीन (c) मिन्कोव्स्की (d) यूक्लीडियन

133. Consider the following data :
 True Positive (TP) = 9
 False Positive (FP) = 6
 False Negative (FN) = 26
 True Negative (TN) = 70
 then value of Error Rate = _____ ?
 (a) 45.5% (b) 28.8% (c) 20.6% (d) 89.9%
134. Which of the following is an example of supervised learning concept ?
 (a) Classification of mails (b) Bar graph representation
 (c) K-means algorithm (d) Recommendation system
135. KNN stands for _____.
 (a) K-Non-Neighbour Network (b) K-Normal Neighbour
 (c) K-Nearest Neighbour (d) K-Neural Network
136. From the following, Pattern Recognition is used for _____.
 (a) Computer Vision (b) Speech Recognition
 (c) Both (a) and (b) (d) Neither (a) nor (b)
137. Which of the following is true about classification ?
 (a) It is a measure of accuracy. (b) It is a sub-division process.
 (c) It is a process of multiplication. (d) It is a process of addition.
138. Which of the following is not related to decision trees ?
 (a) Uses divide and conquer method (b) Leaf
 (c) Pruning (d) Node
139. A 2D filtered data is also called _____.
 (a) Political Map (b) Social Map (c) Google Map (d) Feature Map
140. What can be explained by PAC learning ?
 (a) Sample complexity (b) Overfitting
 (c) Label function (d) Underfitting
141. Which of the following is not a machine learning algorithm ?
 (a) SVG (b) Random Forest (c) SVM (d) KNN
142. Machine learning is a subset of which of the following ?
 (a) Random Tree Learning (b) Deep Learning
 (c) Artificial Intelligence (d) Data Learning
143. Error correction is a type of _____ learning.
 (a) Supervised (b) Unsupervised (c) Micro (d) Macro
144. Logistic Regression algorithm is _____ type of algorithm.
 (a) Cost-function minimization (b) Ranking
 (c) Regression (d) Classification

133. निम्नलिखित डाटा पर विचारण कीजिए :

टी पी = 9

एफ पी = 6

एफ एन = 26

टी एन = 70

तो त्रुटि दर (एर रेट) होगा _____

- (a) 45.5% (b) 28.8% (c) 20.6% (d) 89.9%

134. निम्नलिखित में से सुपरवाइज्ड लर्निंग सिद्धान्त का उदाहरण है :

- (a) मेल का वर्गीकरण (b) बार ग्राफ निरूपण (c) के-मीन्स एल्गोरिद्म (d) रिकमेन्डेशन सिस्टम

135. के एन एन का तात्पर्य _____ है ।

- (a) के-नॉन नेबर नेटवर्क (b) के-नॉर्मल नेबर
(c) के-नीयरेस्ट नेबर (d) के-न्यूरल नेटवर्क

136. निम्नलिखित में से पैटर्न रिकॉग्निशन _____ में प्रयुक्त होता है ।

- (a) कम्प्यूटर विज्ञान (b) स्पीच रिकॉग्निशन
(c) दोनों (a) और (b) (d) (a) और (b) में से कोई नहीं

137. निम्नलिखित में से वर्गीकरण के बारे में कौन सही है ?

- (a) यह सटीकता का माप है । (b) यह एक उपविभाजन प्रक्रिया है ।
(c) यह एक गुणा का कार्य है । (d) यह एक जमा का कार्य है ।

138. निम्नलिखित में से डिसिशन ट्री से संबन्धित नहीं है :

- (a) यह डिवाइड और काँकर पद्धति का प्रयोग करता है । (b) लीफ
(c) प्रूनिंग (d) नोड

139. 2D फिल्टर्ड डाटा को _____ भी कहा जाता है ।

- (a) पॉलिटिकल मैप (b) सोशल मैप (c) गूगल मैप (d) फीचर मैप

140. PAC लर्निंग द्वारा क्या समझाया जा सकता है ?

- (a) सैम्पल (नमूना) कॉम्प्लेक्सिटी (b) ओवरफिटिंग
(c) लेबल फंक्शन (d) अन्डरफिटिंग

141. निम्नलिखित में से कौन मशीन लर्निंग एल्गोरिद्म नहीं है ?

- (a) एस वी जी (b) रैण्डम फॉरेस्ट (c) एस वी एम (d) के एन एन

142. निम्नलिखित में से मशीन लर्निंग किसका उपसमूह है ?

- (a) रैण्डम ट्री लर्निंग (b) डीप लर्निंग
(c) आर्टिफीशियल इंटेलीजेंस (d) डाटा लर्निंग

143. त्रुटि सुधार एक _____ लर्निंग है ।

- (a) सुपरवाइज्ड (b) अनसुपरवाइज्ड (c) माइक्रो (d) मैक्रो

144. लॉजिस्टिक रिग्रेशन एल्गोरिद्म एक _____ प्रकार का एल्गोरिद्म है ।

- (a) कॉस्ट फंक्शन मिनिमाइजेशन (b) रैंकिंग
(c) रिग्रेशन (d) क्लासीफिकेशन (वर्गीकरण)

145. Which evaluation metric is commonly used for regression tasks ?
 (a) Accuracy (b) Mean Absolute Error
 (c) F1-score (d) Precision
146. Logistic regression algorithm was proposed in year _____ by _____.
 (a) 1944, Joseph Berkson (b) 1968, Mark
 (c) 1948, Bernoulli (d) 1988, Biden
147. _____ is an example of unsupervised learning.
 (a) E-mail Spam Detection (b) Customer segmentation
 (c) Image classification (d) Stock-price prediction
148. By splitting the data into training data and test data we overcome _____.
 (a) Overfitting (b) Overlapping (c) Overlay(s) (d) One test
149. From the following _____ is not a pruning technique.
 (a) Cost-Based Pruning (b) Cost Complexity Pruning
 (c) Minimum Error Pruning (d) Maximum Error Pruning
150. _____ process is not a part of machine learning.
 (a) Feature selection (b) Model evaluation
 (c) Model deployment (d) Processing of data
151. Which test is appropriate for comparing means of two-independent variables ?
 (a) Z-test (b) T-test (c) Chi-square test (d) F-test
152. _____ Namenode is used when the primary Namenode goes down.
 (a) Rack (b) Secondary
 (c) Data (d) None of the above
153. Which of the following technique comes under practical machine learning ?
 (a) Semaphores (b) Big Oh Notation (c) Boosting (d) Formal proofs
154. Which of the following SGD variant is based on both momentum and adaptive learning ?
 (a) Adagrad (b) RMS prop (c) Adam (d) Nesterov
155. Which classifier performance metric is particularly useful when dealing with imbalanced datasets ?
 (a) Accuracy (b) Precision (c) Recall (d) F1-score
156. In machine learning, which of the following is not a common resampling method ?
 (a) Bootstrap
 (b) Grid Search Cross Validation
 (c) Leave One Out Cross-Validation (LOOCV)
 (d) Random Subspace Method
157. In binary classification, what does the F1-score represent ?
 (a) Harmonic mean of precision and recall (b) Arithmetic mean of precision and recall
 (c) Geometric mean of precision and recall (d) Product of precision and recall
158. How is classifier performance typically measured ?
 (a) Mean squared error (b) R-squared
 (c) Correlation and Coefficient (d) Precision and Recall

145. रिग्रेशन कार्यों के लिए आमतौर पर किस मूल्यांकन मेट्रिक का उपयोग किया जाता है ?
 (a) एक्यूरेसी (b) मीन एब्सोल्यूट एर (c) एफ-1 स्कोर (d) प्रिसिशन
146. लॉजिस्टिक रिग्रेशन एल्गोरिद्म सन् _____ में _____ द्वारा प्रस्तावित किया गया था ।
 (a) 1944, जोसेफ बर्कसन (b) 1968, मार्क
 (c) 1948, बर्नोली (d) 1988, बाइडन
147. _____ अनसुपरवाइज्ड लर्निंग का एक उदाहरण है ।
 (a) ई-मेल स्पैम का पता लगाना (b) ग्राहक सेगमेंटेशन (विभाजन)
 (c) ईमेज वर्गीकरण (d) स्टॉक-मूल्य भविष्यवाणी
148. डाटा को प्रशिक्षण और परीक्षण में विभाजित करने से _____ नहीं होता ।
 (a) ओवरफिटिंग (b) ओवरलैपिंग (c) ओवरले(स) (d) वन टेस्ट
149. निम्नलिखित में से एक प्रूनिंग तकनीक नहीं है :
 (a) कॉस्ट-बेस्ड प्रूनिंग (b) कॉस्ट कॉम्प्लेक्सिटी प्रूनिंग
 (c) मिनिमम एर प्रूनिंग (d) मैक्सीमम एर प्रूनिंग
150. _____ प्रक्रिया मशीन लर्निंग का हिस्सा नहीं है ।
 (a) फीचर सिलेक्शन (b) मॉडल इवैलुएशन (c) मॉडल डिप्लॉयमेंट (d) डाटा प्रोसेसिंग
151. दो स्वतंत्र चरों के माध्यों की तुलना के लिए कौन सा परीक्षण उपयुक्त है ?
 (a) जेड-टेस्ट (b) टी-टेस्ट (c) कार्ई-स्क्वायर टेस्ट (d) एफ-टेस्ट
152. _____ नेमनोड का उपयोग तब किया जाता है जब प्राथमिक नेमनोड नीचे चला (बंद हो) जाता है ।
 (a) रैक (b) द्वितीयक (सेकण्डरी) (c) डाटा (d) उपरोक्त में से कोई नहीं
153. निम्नलिखित में से कौन सी तकनीक व्यावहारिक मशीन लर्निंग के अन्तर्गत आती है ?
 (a) सेमाफोर (b) बिग ओह संकेतन (c) बूस्टिंग (d) औपचारिक प्रमाण
154. निम्नलिखित में से कौन सा एसजीडी संस्करण गति और अनुकूली शिक्षा दोनों पर आधारित है ?
 (a) एडाग्राड (b) आर एम एस प्राप (c) एडम (d) नेस्टरोव
155. असंतुलित डेटासेट से निपटने के दौरान कौन सा क्लासीफायर प्रदर्शन (परफॉर्मेंस) मेट्रिक विशेष रूप से उपयोगी है ?
 (a) एक्यूरेसी (b) प्रिसिशन (c) रिकॉल (d) एफ-1 स्कोर
156. मशीन लर्निंग में निम्नलिखित में से कौन सी सामान्य पुनःनमूनाकरण विधि नहीं है ?
 (a) बूटस्ट्रेप (b) ग्रिड सर्च क्रॉस वैलिडेशन
 (c) लीव-वन-आऊट क्रॉस वैलिडेशन (एलओओसीवी) (d) रैंडम सबस्पेस विधि
157. बाइनरी वर्गीकरण में, एफ-1 स्कोर क्या दर्शाता है ?
 (a) प्रिसिशन और रिकॉल का हार्मोनिक मीन (b) प्रिसिशन और रिकॉल का अर्थमेटिक मीन
 (c) प्रिसिशन और रिकॉल का जियोमेट्रिक मीन (d) प्रिसिशन और रिकॉल का गुणा (प्रोडक्ट)
158. क्लासीफायर का प्रदर्शन कैसे मापा जाता है ?
 (a) मीन स्क्वायर्ड एर (b) आर-स्क्वायर्ड
 (c) कोरिलेशन एण्ड कोएफिशिएंट (d) प्रिसिशन और रिकॉल

159. Which of the following is commonly used to assess the performance of a classification algorithm when dealing with binary outcomes ?
 (a) T-test (b) Binomial test (c) ANOVA (d) Chi-square test
160. When comparing the performance of multiple classification algorithms across multiple data sets, which of the following test is appropriate ?
 (a) ANOVA (b) MANOVA (c) Tukey's HSD test (d) Chi-square test
161. In a K-fold cross-validation, data set is divided as _____.
 (a) (K) equal subsets (b) (K-1) training subsets
 (c) Randomly into disjoint subsets (d) Some ratio of subsets
162. _____ is a hyper parameter in machine learning.
 (a) Training Data (b) Test Data (c) Learning Rate (d) Machine Data
163. A network with dynamics is known as _____.
 (a) Computer Network (b) Parallel Network
 (c) Stochastic Network (d) Neighbour Network
164. In K-fold cross validation, the value of K is typically chosen as _____.
 (a) 1 (b) same as the size of data set
 (c) a smaller integer such as 5 or 10 (d) a large integer such as 1500 or 2000
165. Which of the following is not a cross-validation ?
 (a) Holdout validation (b) K-Mean cross-validation
 (c) K-Fold cross-validation (d) Leave-one-out cross validation
166. F1 score is calculated using _____ and _____.
 (a) Accuracy, Precision (b) Precision, Recall
 (c) Sensitivity, Specificity (d) Accuracy, Recall
167. Which of the following is useful for imbalanced datasets ?
 (a) Accuracy (b) Precision (c) Recall (d) F1-score
168. In the following ANN, the connections between the units do not form a directed cycle :
 (a) Delta Neural Network (b) Genetic Neural Network
 (c) Feed forward Neural Network (d) Fuzzy-Gamma Neural Network
169. In a Neural network, each pass is called _____.
 (a) Epoch (b) Neuron (c) Bit (d) Weight
170. Which of the following is a decision making model used in discrete, stochastic, sequential environments ?
 (a) Back Propagation Model (b) Markov Decision Process Model
 (c) Feed Forward Model (d) Feed Backward Model
171. Which of the following component makes a neural network non-linear in nature ?
 (a) Activation function (b) Bias
 (c) Weight (d) Input layer
172. Gradient Descent algorithm was proposed by _____.
 (a) Ross Quinlan (b) Leslie Valient
 (c) Thomas Bayes (d) Augustin-Louis Cauchy

159. आऊटकम वेरिफेबल बाइनरी हो, तब वर्गीकरण एल्गोरिद्म के प्रदर्शन के मूल्यांकन के लिए निम्नलिखित का इस्तेमाल किया जाता है :
- (a) टी-टेस्ट (b) बाइनोमियल टेस्ट (c) एनोवा (d) कार्ई-स्क्वायर टेस्ट
160. मल्टीपल डाटासेट में मल्टीपल क्लासीफिकेशन एल्गोरिद्म के प्रदर्शन की तुलना के लिए निम्नलिखित में से कौन सा टेस्ट उपयुक्त है ?
- (a) एनोवा (b) मनोवा
(c) तुकी का एच एस डी टेस्ट (d) कार्ई-स्क्वायर टेस्ट
161. के-फोल्ड क्रॉस वेलिडेशन में डाटा सेट को _____ में विभाजित किया जाता है ।
- (a) (K) सामान सबसेट्स (b) (K-1) ट्रेनिंग सबसेट्स
(c) डिसज्वाइंट सबसेट में रैण्डमली विभाजन (d) किसी अनुपात पर सबसेट
162. मशीन लर्निंग में _____ एक हाइपर पैरामीटर है ।
- (a) ट्रेनिंग डाटा (b) टेस्ट डाटा (c) लर्निंग रेट (d) मशीन डाटा
163. ऐसा नेटवर्क जिसमें डायनामिक्स की भूमिका होती है
- (a) कम्प्यूटर नेटवर्क (b) पैरेलल नेटवर्क (c) स्टॉकास्टिक नेटवर्क (d) नेबर (पड़ोसी) नेटवर्क
164. के-फोल्ड क्रॉस-वेलिडेशन में “के” को आमतौर पर _____ चुना जाता है ।
- (a) 1 (b) डाटा सेट के बराबर
(c) एक छोटा पूर्णांक 5 या 10 (d) एक बड़ा पूर्णांक 1500 या 2000
165. निम्नलिखित में से क्रॉस वेलिडेशन नहीं है :
- (a) होल्डआऊट वेलिडेशन (b) के-मीन क्रॉस वेलिडेशन
(c) के-फोल्ड क्रॉस वेलिडेशन (d) लीव-वन-आऊट क्रॉस वेलिडेशन
166. एफ-1 स्कोर मापने के लिए _____ और _____ का उपयोग किया जाता है ।
- (a) एक्ज्युरैसी, प्रिसिशन (b) प्रिसिशन, रिकॉल
(c) सेन्सिटिविटी, स्पेसिफिसिटी (d) एक्ज्युरैसी, रिकॉल
167. निम्नलिखित में से कौन (क्या) असंतुलित डेटा-सेट के लिए उपयुक्त है ?
- (a) एक्ज्युरैसी (b) प्रिसिशन (c) रिकॉल (d) एफ-1 स्कोर
168. निम्नलिखित ए.एन.एन. में, इकाइयों के बीच में संयोजन से डायरेक्टेड साइकिल नहीं बनता है :
- (a) डेल्टा न्यूरल नेटवर्क (b) जेनेटिक न्यूरल नेटवर्क
(c) फीड फॉरवर्ड न्यूरल नेटवर्क (d) फज्ज़ी-गामा न्यूरल नेटवर्क
169. न्यूरल नेटवर्क में प्रत्येक पास _____ कहलाता है ।
- (a) इपोक (b) न्यूरॉन (c) बिट (d) वेट (वजन)
170. निम्न में कौन डिस्क्रीट, स्टॉकास्टिक, सीक्वेंसियल वातावरण में प्रयुक्त डिसिशन मेकिंग मॉडल है ?
- (a) बैक प्रोपेगेशन मॉडल (b) मार्कोव डिसिशन प्रोसेस मॉडल
(c) फीड फॉरवर्ड मॉडल (d) फीड बैकवर्ड मॉडल
171. निम्नलिखित में से कौन सा घटक न्यूरल नेटवर्क को नॉन-लीनियर बनाता है ?
- (a) एक्टिवेशन फंक्शन (b) बायस (c) वेट (वजन) (d) इनपुट लेआउट
172. ग्रेडिएण्ट डीसेंट एल्गोरिद्म _____ द्वारा प्रस्तावित किया गया ।
- (a) रॉस क्विनलॉन (b) लेसली वेलीयंट (c) थॉमस बेयस (d) ऑगस्टिन-लुई कॉशी

173. 2-D filtered data are called as _____.
 (a) Social Maps (b) Google Maps (c) Political Maps (d) Feature Maps
174. From the following _____ receives messages from other cells in Neurons.
 (a) Dendrites (b) Axon (c) Layer (d) Dexon
175. Of the following _____ optimization algorithm is used to train a machine-learning model.
 (a) Naïve Bayes (b) Gradient Descent
 (c) Markov Decision (d) Prim's Algorithm
176. Total types of layers in radial basis function neural network is _____.
 (a) 1 (b) 2 (c) 3 (d) 4
177. When both inputs are 1, the output of NAND gates based on Pitts model is _____.
 (a) 0 (b) 1 (c) Z (d) Either 0 or 1
178. What type of activation function is commonly used in LSTM gates ?
 (a) ReLU (b) Sigmoid (c) Tan h (d) Leaky ReLU
179. Which of the following is not a deep learning framework ?
 (a) Tensor Flow (b) Pytorch (c) Keras (d) Scikit-learn
180. The primary goal of deep learning is _____.
 (a) Data Extraction (b) Data Compression
 (c) Feature Engineering (d) Learning pattern from data
181. Fundamental unit of neural network is _____.
 (a) Nucleus (b) Neuron (c) Brain (d) Axon
182. From the following RNN stands for _____.
 (a) Receiver Neural Network (b) Reporting Neural Network
 (c) Recurrent Neural Network (d) Recording Neural Network
183. LSTM stands for _____.
 (a) Long Short Term Memory (b) Long Sequential Term Memory
 (c) Limited Short Term Memory (d) Linear Sequential Term Memory
184. In neural network context, Negative sign of a weight indicates _____.
 (a) Excitatory Input (b) Inhibitory Input
 (c) Excitatory Output (d) Inhibitory Output
185. CNN stands for _____.
 (a) Complex Neural Network (b) Costly Neural Network
 (c) Convolution Neural Network (d) Come-back Neural Network
186. _____ activation function is commonly used in hidden layers of deep neural network.
 (a) ReLU (b) Sigmoid (c) Tan h (d) Softmax
187. Markov Networks are _____ models.
 (a) Directed (b) Simple
 (c) Undirected (d) Complex-directed

173. 2-डी फिल्टर्ड डाटा को _____ कहा जाता है ।
 (a) सोशल मैप्स (b) गूगल मैप्स (c) पॉलिटिकल मैप्स (d) फीचर मैप्स
174. निम्नलिखित में से संदेशों को अन्य सेल से न्यूरोन में प्राप्त करता है :
 (a) डेन्ड्राइट्स (b) एक्सोन (c) लेयर (d) डेक्सोन
175. निम्नलिखित में से _____ ऑप्टिमाइजेशन एल्गोरिदम सामान्यतः मशीन लर्निंग मॉडल को ट्रेन करने के लिए प्रयुक्त है ।
 (a) नेव बेयस (बायस) (b) ग्रेडियेन्ट डीसेन्ट (c) मार्कोव डिजिजन (d) ग्रिम्स (ग्रिम) एल्गोरिदम
176. रेडिएल आधार फंक्शन न्यूरल नेटवर्क में लेयर के कुल प्रकार है
 (a) 1 (b) 2 (c) 3 (d) 4
177. जब दोनों इनपुट 1 हैं, तो पिट्स मॉडल आधारित NAND (नेन्ड) गेट की आऊटपुट _____ होगी ।
 (a) 0 (b) 1 (c) Z (d) या तो 0 या 1
178. एल एस टी एम गेट में आमतौर पर किस प्रकार का एक्टिवेशन फंक्शन उपयोग होता है ?
 (a) ReLU (रे एलयू) (b) सिग्माइड (c) Tan h (टैन एच) (d) लीकी रे एलयू
179. निम्न में से एक डीप लर्निंग ढाँचा नहीं है :
 (a) टेंसर फ्लो (b) पाईटार्च (c) केरास (d) साइकित लर्न
180. डीप-लर्निंग का प्राथमिक लक्ष्य _____ है ।
 (a) डाटा एक्सट्रैक्शन (b) डाटा कम्प्रेसन (c) फीचर इंजीनियरिंग (d) डाटा से पैटर्न सीखना
181. न्यूरल नेटवर्क की मूलभूत इकाई _____ है ।
 (a) न्यूक्लियस (b) न्यूरॉन (c) ब्रेन (d) एक्सॉन
182. निम्नलिखित में से आर एन एन का मतलब :
 (a) रिसीवर न्यूरल नेटवर्क (b) रिपोर्टिंग न्यूरल नेटवर्क
 (c) रिकॉर्ड न्यूरल नेटवर्क (d) रिकॉर्डिंग न्यूरल नेटवर्क
183. एल एस टी एम _____ का संक्षिप्त रूप है ।
 (a) लॉग शार्ट टर्म मेमोरी (b) लॉग सीक्वेंशियल टर्म मेमोरी
 (c) लिमिटेड शार्ट टर्म मेमोरी (d) लीनियर सीक्वेंशियल टर्म मेमोरी
184. न्यूरल नेटवर्क के संदर्भ में, वेट का नेगेटिव साइन _____ दर्शाता है ।
 (a) एक्साइटेटरी इनपुट (b) इनहिबिटरी इनपुट (c) एक्साइटेटरी आऊटपुट (d) इनहिबिटरी आऊटपुट
185. सी एन एन _____ का संक्षिप्त रूप है ।
 (a) कॉम्प्लेक्स न्यूरल नेटवर्क (b) कॉस्टली न्यूरल नेटवर्क
 (c) कॉन्वोल्यूशन न्यूरल नेटवर्क (d) कम-बैक न्यूरल नेटवर्क
186. _____ एक्टिवेशन फंक्शन आमतौर पर डीप न्यूरल नेटवर्क की हिडन (छिपी) परतों में उपयोग किया जाता है ।
 (a) रे एलयू (b) सिग्माइड (c) टैन एच (d) सॉफ्टमैक्स
187. मार्कोव नेटवर्क्स _____ मॉडल्स हैं ।
 (a) डायरेक्टेड (b) सिंपल (c) अनडायरेक्टेड (d) कॉम्प्लेक्स डायरेक्टेड

188. Which of the following language(s) are commonly used in Data Science ?
(a) Python (b) HTML (c) R (d) Both (a) & (c)
189. The process of Quantifying data is referred to as _____.
(a) Decoding (b) Structure (c) Enumeration (d) Coding
190. _____ is used when we want to visually examine the relationship between two quantitative variables.
(a) Bar Graph (b) Scatter Plot (c) Line Graph (d) Pie Chart
191. Amongst the following, which is/are not a major data analysis approach ?
(a) Text Analytics (b) Data Mining
(c) Topology (d) Predictive Analytics
192. Discovering rules that describe large partitions of data is called _____.
(a) Association (b) Clustering (c) Classifier (d) Cleansing
193. By the year 2025, the volume of data will increase at-least to the scale of _____.
(a) TB (b) YB (c) EB (d) ZB
194. Data that sits outside the trend is referred to as a/an _____.
(a) Trend (b) Spike (c) Outlier (d) Cluster
195. Which of the following is another name for raw data ?
(a) Destination Data (b) Eggy Data
(c) Secondary Data (d) Machine Learning Data
196. Which of the following is not a data science application ?
(a) Generating Random Number (b) Predicting Stock Price
(c) Image Recognition (d) Fraud Detection
197. What does DSP stand for in the context of Data Science ?
(a) Data Storage Platform (b) Data Science Processing
(c) Data Science Prediction (d) Digital Signal Processing
198. The information that is displayed in its most unprocessed form is referred to as _____.
(a) Formatted data (b) Unstructured data
(c) Structured data (d) Semi-structured data
199. A _____ is a subject-oriented, non-volatile, integrated and time-variant collection of data generated for the aim of making decision(s) by the management.
(a) Data warehouse (b) Mobile data
(c) Big data (d) Structured data
200. Which of the following steps are not included in a data-science process ?
(a) Quantum Computing (b) Data Visualisation
(c) Data Analysis (d) Data Collection

188. निम्नलिखित में से कौन सी भाषा(यें) आमतौर पर डाटा साइन्स में उपयोग की जाती हैं ?
 (a) पायथन (b) एच टी एम एल (c) आर (d) दोनों (a) और (c)
189. आँकड़ों की मात्रा निर्धारित करने की प्रक्रिया को _____ कहा जाता है ।
 (a) डिकोडिंग (b) स्ट्रक्चर (c) एन्यूमरेशन (d) कोडिंग
190. _____ का उपयोग तब किया जाता है जब हम दो मात्रात्मक वेरियेबल के बीच संबंधों को दृष्टि से जाँच करना चाहते हैं ।
 (a) बार ग्राफ (b) स्कैटर प्लॉट (c) लाइन ग्राफ (d) पाई चार्ट
191. निम्नलिखित में से कौन सा एक प्रमुख डाटा विश्लेषण दृष्टिकोण नहीं है ?
 (a) टेक्स्ट एनालिटिक्स (b) डाटा माइनिंग (c) टोपोलॉजी (d) प्रीडिक्टिव एनालिटिक्स
192. डाटा के बड़े हिस्से का वर्णन करने वाले नियमों की खोज को _____ कहा जाता है ।
 (a) एसोसिएशन (b) क्लस्टरिंग (c) क्लासिफायर (d) क्लीनजिंग
193. सन् 2025 तक डाटा की मात्रा कम से कम _____ स्केल तक बढ़ जाएगी ।
 (a) TB (b) YB (c) EB (d) ZB
194. प्रवृत्ति से बाहर रहने वाले डाटा को _____ कहा जाता है ।
 (a) प्रवृत्ति (b) स्पाइक (c) आउटलायर (d) क्लस्टर
195. निम्नलिखित में से कौन सा कच्चे (रॉ) डाटा का दूसरा नाम है ?
 (a) गंतव्य डाटा (b) एग्जी डाटा
 (c) द्वितीयक (सेकण्डरी) डाटा (d) मशीन लर्निंग डाटा
196. निम्नलिखित में से कौन सा डेटा विज्ञान का अनुप्रयोग नहीं है ?
 (a) जेनरेटिंग रैंडम नंबर (b) प्रीडिक्टिंग स्टॉक प्राइस
 (c) इमेज रिकॉग्निशन (d) फ्रॉड डिटेक्शन
197. डाटा साइन्स के संदर्भ में डी एस पी का अर्थ है
 (a) डाटा स्टोरेज प्लेटफॉर्म (b) डाटा साइन्स प्रोसेसिंग
 (c) डाटा साइन्स प्रीडिक्शन (d) डिजिटल सिग्नल प्रोसेसिंग
198. वह सूचना जो अपने सबसे मूल रूप (अनप्रोसेस्ड) में प्रदर्शित होती है, _____ कहलाती है ।
 (a) संरूपित डाटा (b) असंरचित डाटा (c) संरचित डाटा (d) अर्ध-संरचित डाटा
199. एक _____ डाटा का एक विषय-उन्मुख, नॉन-वोलेटाइल, एकीकृत और समय-भिन्न संग्रह है जो प्रबंधन को निर्णय लेने में मदद करता है ।
 (a) डाटा-वेअरहाउस (b) मोबाइल डाटा (c) बिग डाटा (d) संरचित डाटा
200. डाटा विज्ञान प्रक्रिया में निम्न में से कौन सा चरण शामिल नहीं है ?
 (a) क्वांटम कम्प्यूटिंग (b) डाटा विज्युलाइजेशन (c) डाटा विश्लेषण (d) डाटा संग्रह

Space For Rough Work / रफ कार्य के लिए जगह



Teachingninja.in