www.prepp.in

Prepp
Your Personal Exams Guide

Series : QR1SP

SET-4

प्रश्न-पत्र कोड
Q.P. Code 368

रोल नं.
Roll No.

परीक्षार्थी प्रश्न-पत्र कोड को उत्तर-पुस्तिका के मुख-पृष्ठ पर अवश्य लिखें ।
Candidates must write the Q.P. Code on the title page of the answer-book.

नोट :

(I) कृपया जाँच कर लें कि इस प्रश्न-पत्र में मुद्रित पृष्ठ **23** हैं ।

(II) प्रश्न-पत्र में दाहिने हाथ की ओर दिए गए प्रश्न-पत्र कोड को परीक्षार्थी उत्तर-पुस्तिका के मुख-पृष्ठ पर लिखें ।

(III) कृपया जाँच कर लें कि इस प्रश्न-पत्र में **21** प्रश्न हैं ।

(IV) **कृपया प्रश्न का उत्तर लिखना शुरू करने से पहले, उत्तर-पुस्तिका में यथा स्थान पर प्रश्न का क्रमांक अवश्य लिखें ।**

(V) इस प्रश्न-पत्र को पढ़ने के लिए 15 मिनट का समय दिया गया है । प्रश्न-पत्र का वितरण पूर्वाह्न में 10.15 बजे किया जाएगा । 10.15 बजे से 10.30 बजे तक परीक्षार्थी केवल प्रश्न-पत्र को पढ़ेंगे और इस अवधि के दौरान वे उत्तर-पुस्तिका पर कोई उत्तर नहीं लिखेंगे ।

NOTE :

(I) Please check that this question paper contains **23** printed pages.

(II) Q.P. Code given on the right hand side of the question paper should be written on the title page of the answer-book by the candidate.

(III) Please check that this question paper contains **21** questions.

(IV) **Please write down the serial number of the question in the answer-book at the given place before attempting it.**

(V) 15 minute time has been allotted to read this question paper. The question paper will be distributed at 10.15 a.m. From 10.15 a.m. to 10.30 a.m., the candidates will read the question paper only and will not write any answer on the answer-book during this period.

डाटा साइंस
DATA SCIENCE

निर्धारित समय : 2 घण्टे | *अधिकतम अंक : 50*
Time allowed : 2 hours | *Maximum Marks : 50*

Prepp
Your Personal Exams Guide

सामान्य निर्देश :

(i) *कृपया निर्देशों को ध्यान से पढ़ें ।*

(ii) *इस प्रश्न-पत्र में **दो** खण्डों में **21** प्रश्न हैं : **खण्ड – क** और **खण्ड – ख** ।*

(iii) ***खण्ड – क** में वस्तुनिष्ठ प्रकार के प्रश्न हैं, जबकि **खण्ड – ख** में विषयपरक प्रकार के प्रश्न हैं ।*

(iv) *दिए गए (5 + 16 =) 21 प्रश्नों में से, उम्मीदवार को **2** घंटे के आबंटित (अधिकतम) समय में (5 + 10 =) 15 प्रश्नों के उत्तर देने हैं ।*

(v) *किसी विशेष खण्ड के सभी प्रश्नों को सही क्रम में करने का प्रयास किया जाना चाहिए ।*

(vi) ***खण्ड – क :** वस्तुनिष्ठ प्रकार के प्रश्न (**24** अंक) :*

(a) *इस खण्ड में **5** प्रश्न हैं ।*

(b) *कोई नकारात्मक अंकन नहीं है ।*

(c) *दिए गए निर्देशों के अनुसार कीजिए ।*

(d) *प्रत्येक प्रश्न / भाग के सामने आबंटित अंकों का उल्लेख किया गया है ।*

(vii) ***खण्ड – ख :** विषयपरक प्रकार के प्रश्न (**26** अंक) :*

(a) *इस खण्ड में **16** प्रश्न हैं ।*

(b) *उम्मीदवार को **10** प्रश्न करने हैं ।*

(c) *दिए गए निर्देशों के अनुसार कीजिए ।*

(d) *प्रत्येक प्रश्न / भाग के सामने आबंटित अंकों का उल्लेख किया गया है ।*

www.prepp.in

Prepp
Your Personal Exams Guide

General Instructions :

(i) *Please read the instructions carefully.*

(ii) *This question paper consists of **21** questions in **two** Sections : **Section – A** & **Section – B**.*

(iii) ***Section – A** has Objective Type Questions whereas **Section – B** contains Subjective Type Questions.*

(iv) *Out of the given (5 + 16 =) 21 questions, a candidate has to answer (5 + 10 =) 15 questions in the allotted (maximum) time of **2** hours.*

(v) *All questions of a particular section must be attempted in the correct order.*

(vi) ***Section – A :** Objective Type Questions (**24** marks) :*

- (a) *This section has **5** questions.*
- (b) *There is no negative marking.*
- (c) *Do as per the instructions given.*
- (d) *Marks allotted are mentioned against each question/part.*

(vii) ***Section – B :** Subjective Type Questions (**26** marks) :*

- (a) *This section has **16** questions.*
- (b) *A candidate has to do **10** questions.*
- (c) *Do as per the instructions given.*
- (d) *Marks allotted are mentioned against each question/part.*

www.prepp.in

Prepp
Your Personal Exams Guide

खण्ड – क

(वस्तुनिष्ठ प्रकार के प्रश्न) **(24 अंक)**

1. रोजगार कौशल पर दिए गए **6** प्रश्नों में से किन्हीं **4** प्रश्नों के उत्तर दीजिए । **4 × 1 = 4**

(i) अंग्रेजी में __________ बुनियादी विरामचिह्न या संकेत उपयोग किए जाते हैं । **1**

(A) 10 (B) 12

(C) 15 (D) 20

(ii) बाह्य प्रेरणा __________ या ____________ के कारण उत्पन्न होती है । **1**

(iii) ऐसे व्यक्ति जो आस–पास के लोगों के साथ बातचीत करना पसंद करते हैं और आमतौर पर बातूनी होते हैं, उन्हें क्या कहा जा सकता है ? **1**

(A) बहिर्मुखी (B) अंतर्मुखी

(C) शर्मीला (D) डरा हुआ

(iv) स्प्रेडशीट में डेटा दर्ज करने के चरणों का सही क्रम क्या है ? **1**

(A) डेटा टाइप करें, सेल पर क्लिक करें और एंटर दबाएँ

(B) सेल पर क्लिक करें, डेटा टाइप करें और एंटर दबाएँ

(C) एंटर दबाएँ, सेल पर क्लिक करें और डेटा टाइप करें

(D) सेल पर क्लिक करें, एंटर दबाएँ और डेटा टाइप करें

(v) जो व्यक्ति समाज को लाभ पहुँचाने वाले समाधान विकसित करने पर ध्यान केंद्रित करते हैं, उन्हें __________ उद्यमी कहा जाता है । **1**

(A) पहली पीढ़ी (B) आई.टी.

(C) सामाजिक (D) पारिवारिक

(vi) भारत सरकार के शहरी विकास मंत्रालय ने ठोस कचरे को ___________ और _____________ के आधार पर 14 श्रेणियों में वर्गीकृत किया है । **1**

www.prepp.in

Prepp
Your Personal Exams Guide

Section – A

(Objective Type Questions) **(24 Marks)**

1. Answer any **4** questions out of given **6** questions on Employability Skills : **4 × 1 = 4**

(i) There are ________ basic punctuation marks or signs used in English. **1**

(A) 10 (B) 12

(C) 15 (D) 20

(ii) Extrinsic motivation arises because of ________ or __________. **1**

(iii) Individuals who love interacting with people around and are generally talkative can be termed as – **1**

(A) Extroverts (B) Introverts

(C) Shy (D) Scared

(iv) The correct order of steps for entering data in a spreadsheet is – **1**

(A) type the data, click the cell and press enter

(B) click the cell, type the data and press enter

(C) press enter, click the cell and type the data

(D) click the cell, press enter and type the data

(v) Individuals who focus on developing solutions that benefit the society are called ________ entrepreneurs. **1**

(A) First generation (B) I.T.

(C) Social (D) Family

(vi) The Ministry of Urban Development, Government of India has classified solid waste in 14 categories based on the _________ and __________. **1**

2. दिये गये **6** प्रश्नों में से किन्हीं **5** प्रश्नों के उत्तर दीजिए । **5 × 1 = 5**

(i) कौन सा डेटा गोपनीयता कानून (data privacy legislation) विशेष रूप से इस बात से संबंधित है कि वेबसाइटें और ऑनलाइन कंपनियाँ 13 वर्ष से कम उम्र के बच्चों से डेटा कैसे एकत्र करती हैं ? **1**

(A) GDPR (B) HIPAA

(C) CCPA (D) COPPA

(ii) बिजॉय एक डेटासेट (dataset) पर Exploratory Data Analysis (EDA) कर रहा है, जिसमें केवल एक चर (variable), **`Math_Score`**, है । वह यह जाँच करना चाहता है कि क्या इस चर में कोई आउटलायर (outlier) हैं । इस उद्देश्य के लिए कौन सा ग्राफ सबसे उपयुक्त है ? **1**

(A) हिस्टोग्राम (Histogram) (B) बॉक्स प्लॉट (Box Plot)

(C) बार चार्ट (Bar Chart) (D) पाई चार्ट (Pie Chart)

(iii) एक डिसीजन ट्री में प्रत्येक आंतरिक नोड (internal node) का प्राथमिक कार्य क्या है ? **1**

(A) यह अंतिम क्लास लेबल (class label) को रखता है ।

(B) यह एक परीक्षण के परिणाम (outcome) को दर्शाता है ।

(C) यह किसी विशेष क्लास (class) को चुनने के प्रश्न को दर्शाता है ।

(D) यह एक निर्णय की सफलता की संभावना (probability) को दर्शाता है ।

(iv) K-NN एल्गोरिथम (algorithm) में **k** का एक बड़ा मान (value) चुनने का सामान्य प्रभाव क्या होता है ? **1**

(A) यह आउटलायर (outliers) के प्रभाव को बढ़ाता है और निर्णय सतह (decision surface) को अधिक विशिष्ट (specific) बनाता है ।

(B) यह आउटलायर (outliers) के प्रभाव को कम करता है और निर्णय सतह को सामान्यीकृत (generalize) करने का कारण बनता है ।

(C) यह एल्गोरिथम की मेमोरी दक्षता (memory efficiency) को बढ़ाता है ।

(D) यह डेटासेट (dataset) की परवाह किए बिना हमेशा पूर्वानुमान (prediction) की सटीकता में सुधार करता है ।

www.prepp.in

2. Answer any **5** out of the given **6** questions. **5 × 1 = 5**

(i) Which data privacy legislation specifically deals with how websites and online companies collect data from children under 13 years of age ? **1**

(A) GDPR (B) HIPAA

(C) CCPA (D) COPPA

(ii) Bijoy is performing Exploratory Data Analysis (EDA) on a dataset, which contains only one variable, **`Math_Score`**. He wants to check if there are any outliers in this variable. Which graph is most suitable for this purpose ? **1**

(A) Histogram (B) Box Plot

(C) Bar Chart (D) Pie Chart

(iii) What is the primary function of each internal node in a Decision Tree ? **1**

(A) It holds the final class label.

(B) It represents the outcome of a test.

(C) It denotes a question on choosing a particular class.

(D) It represents the probability of success for a decision.

(iv) What is the common effect of choosing a big value of **k** in the K-NN algorithm ? **1**

(A) It increases the effect of outliers and makes the decision surface more specific.

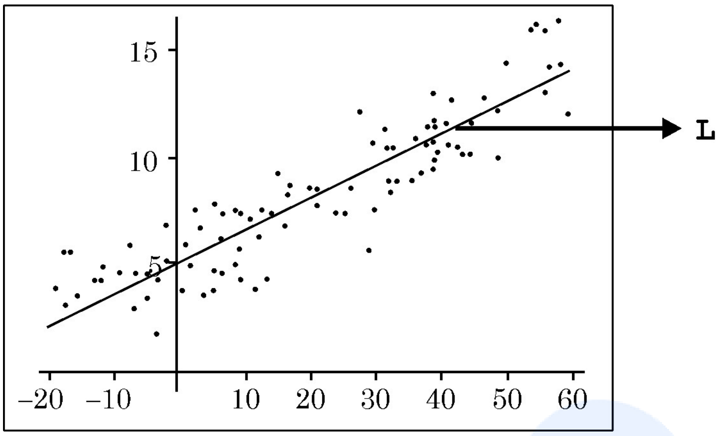
(B) It reduces the effect of outliers and causes the decision surface to generalize.

(C) It increases the memory efficiency of the algorithm.

(D) It always improves prediction accuracy regardless of the dataset.

www.prepp.in
Prepp
Your Personal Exams Guide

(v) नीचे दिखाए गए स्कैटर प्लॉट (scatter plot) पर विचार करें, जो दो चर (variables), X और Y के बीच संबंध दिखाता है । नीचे दिए गए ग्राफ में **L** से चिह्नित रेखा (line) को ________ कहते हैं । **1**



(A) Line of Best Fit (B) Line of Worst Fit
(C) Line of Median (D) Line of Mode

(vi) संपूर्ण raw डेटा को कई समूहों में विभाजित करने की प्रक्रिया को क्या कहा जाता है जिसमें एक समूह में डेटा पॉइंट (data points) उसी समूह के अन्य डेटा पॉइंट के समान हों लेकिन अन्य समूहों के डेटा पॉइंट से अलग हों ? **1**

(A) सुपरवाइज़्ड लर्निंग (Supervised Learning)
(B) एक्सपेक्टेशन–मैक्सिमाइज़ेशन (Expectation-Maximization)
(C) क्लस्टरिंग (Clustering)
(D) विसंगति का पता लगाना (Anomaly detection)

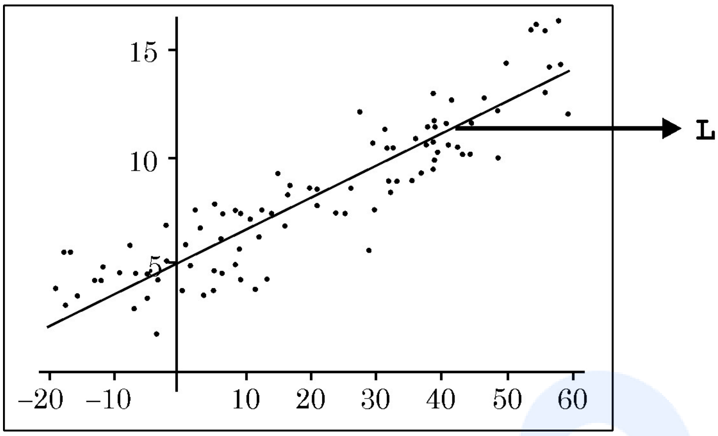
3. दिये गये **6** प्रश्नों में से किन्हीं **5** प्रश्नों के उत्तर दीजिए । **5 × 1 = 5**

(i) एक कॉलेज का प्रवेश कार्यालय (admissions office) आवेदनों को प्री–स्क्रीन (pre-screen) करने के लिए एक स्वचालित प्रणाली का उपयोग करता है । यह प्रक्रिया व्यक्तिगत छात्र के ग्रेड या टेस्ट स्कोर (test scores) की परवाह किए बिना, कुछ कम प्रदर्शन वाले स्कूलों के आवेदकों को स्वत: रूप से फ़िल्टर आउट (filter out) करने के लिए डिज़ाइन की गई है ।

यह स्वत: फ़िल्टरिंग प्रक्रिया (automated filtering process) किस नैतिक दिशानिर्देश (ethical guidelines) का उल्लंघन करती है ? **1**

(A) डेटा की सुरक्षा (Security of Data)
(B) डेटा की गोपनीयता (Privacy of Data)
(C) गैर भेदभाव (Non-discrimination)
(D) खुलापन (Openness)

(v) Consider the scatter plot shown below, that shows the relationship between two variables, X and Y. The line marked as **L** in the graph below is known as _______. **1**



(A) Line of Best Fit (B) Line of Worst Fit

(C) Line of Median (D) Line of Mode

(vi) What is the name of the process of dividing entire raw data into several groups such that data points in one group are similar to other data points in the same group but different from those in other groups ? **1**

(A) Supervised Learning (B) Expectation-Maximization

(C) Clustering (D) Anomaly detection

3. Answer any **5** out of the given **6** questions. **5 × 1 = 5**

(i) An admission office of a college uses an automated system to pre-screen applications. The process is designed to automatically filter out applicants who attended certain low-performing schools, regardless of the individual student's grades or test scores.

This automated filtering process violates which ethical guidelines ? **1**

(A) Security of Data (B) Privacy of Data

(C) Non-discrimination (D) Openness

(ii) एक डेटा विश्लेषक (data analyst) एक डेटासेट पर काम कर रही है । वह देखती है कि एक चर (variable) में कुछ प्रविष्टियाँ (entries) गायब हैं । डेटा क्लीनिंग (data cleaning) प्रक्रिया के दौरान इन गुम मानों (missing values) को handle करने का सबसे उपयुक्त तरीका क्या है/हैं ? 1

(i) गुम मानों को वैसे ही छोड़ दें ।

(ii) गुम चर की पंक्ति (row) को हटा दें ।

(iii) गुम चर के माध्य (mean)/बहुलक (mode) के करीब एक मान insert करें ।

(iv) गुम प्रविष्टियों की प्रतिलिपि (duplicate) बनाएँ

(v) गुम मानों को ऋणात्मक संख्याओं (negative numbers) से बदल दें ।

विकल्प :

(A) (i), (ii) और (iv)

(B) (ii) और (iii)

(C) (ii), (iii) और (v)

(D) (iii) और (iv)

(iii) सारा अपनी सप्ताहांत की गतिविधियों की योजना बनाने के लिए एक डिसीजन ट्री बना रही है । निर्णयों में से एक यह है कि हाइकिंग (hiking) के लिए जाना है या घर पर रहना है । अगर वह हाइकिंग के लिए जाती है, तो मौसम धूप वाला या बारिश वाला हो सकता है, लेकिन यह निश्चित नहीं है कि क्या होगा । डिसीजन ट्री में अनिश्चित मौसम की स्थिति (uncertain weather outcome) को दर्शाने के लिए, सारा को किस प्रकार के नोड (node) का उपयोग करना चाहिए ? 1

(A) वर्ग नोड (Square node)

(B) गोलाकार नोड (Circular node)

(C) त्रिकोणीय नोड (Triangular node)

(D) डायमंड नोड (Diamond node)

(iv) निम्नलिखित में से कौन उस समस्या का वर्णन करता है जहाँ इनपुट वैरिएबल (input variables) की संख्या बढ़ने पर K-NN एल्गोरिथम एक नए डेटा पॉइंट (data point) के आउटपुट का पूर्वानुमान (prediction) लगाने में संघर्ष करता है ? 1

(A) असंतुलित डेटा समस्या (Imbalanced Data Problem)

(B) प्रशिक्षण चरण का अभाव (No Training Step)

(C) आउटलायर्स के प्रति संवेदनशीलता (Sensitivity to Outliers)

(D) आयामी अभिशाप (Curse of Dimensionality)

Prepp
Your Personal Exams Guide

(ii) A data analyst is working on a dataset. She notices that a variable has some missing entries. What is the most appropriate way(s) to handle these missing values during the data cleaning process ? **1**

(i) Leave the missing values as they are

(ii) Remove the row of the variable that is missing

(iii) Insert a value close to mean/mode of the variable that is missing

(iv) Duplicate the missing entries

(v) Replace the missing values with negative numbers

Options :

(A) (i), (ii) and (iv)
(B) (ii) and (iii)
(C) (ii), (iii) and (v)
(D) (iii) and (iv)

(iii) Sara is creating a decision tree to plan her weekend activities. One of the decisions is whether to go hiking or stay home. If she goes hiking, the weather could be sunny or rainy, but she isn't sure what it will be. When representing the uncertain weather outcome in her decision tree, what type of node should Sara use ? **1**

(A) Square node
(B) Circular node
(C) Triangular node
(D) Diamond node

(iv) Which of the following describes the problem where the K-NN algorithm struggles to predict the output of a new data point as the number of input variables grows ? **1**

(A) Imbalanced Data Problem
(B) No Training Step
(C) Sensitivity to Outliers
(D) Curse of Dimensionality

www.prepp.in

Prepp
Your Personal Exams Guide

(v) रूट मीन स्क्वायर डेविएशन (Root Mean Square Deviation) किसका वर्गमूल (square root) है ? **1**

(A) रेसिडुअल का औसत (Mean of residuals)

(B) रेसिडुअल का विचरण (Variance of residuals)

(C) मानक विचलन (Standard deviation)

(D) मीन एब्सोल्यूट एरर (Mean absolute error)

(vi) वह एप्रोच (approach) जिसका K-मीन्स क्लस्टरिंग एल्गोरिथम (K-means Clustering Algorithm) एक दी गई क्लस्टरिंग समस्या को हल करने के लिए पालन करता है, उसे क्या कहा जाता है ? **1**

(A) विसंगति का पता लगाना (Anomaly Detection)

(B) एक्सपेक्टेशन–मैक्सिमाइज़ेशन (Expectation-Maximization)

(C) डिसीजन ट्री (Decision Tree)

(D) लीनियर रिग्रेशन (Linear Regression)

4. दिये गये **6** प्रश्नों में से किन्हीं **5** प्रश्नों के उत्तर दीजिए । **5 × 1 = 5**

(i) 11 दिसंबर, 2019 को, किस मंत्रालय (ministry) ने भारतीय संसद में Personal Data Protection Bill, 2019 पेश किया ? **1**

(A) स्वास्थ्य और परिवार कल्याण मंत्रालय

(B) कानून और न्याय मंत्रालय

(C) वित्त मंत्रालय

(D) इलेक्ट्रॉनिक्स और सूचना प्रौद्योगिकी मंत्रालय

(ii) डेटा दर्ज (entering) करते समय, रेहान ने गलती से `Distance` (दूरी) चर (variable) को संख्यात्मक `(numeric)` के बजाय टेक्स्ट `(text)` के रूप में दर्ज कर दिया है । इस परिदृश्य के लिए निम्नलिखित में से कौन सा कथन सत्य है ? **1**

(A) एक बार डेटा दर्ज हो जाने के बाद, उसे डेटा विश्लेषण के किसी भी चरण के दौरान क्लीन (clean) नहीं किया जा सकता है ।

(B) मॉडल बनाने के लिए डेटा का उपयोग करते समय यह समस्याएँ उत्पन्न करेगा ।

(C) यह बिना किसी समस्या के सभी गणितीय कार्यों (mathematical operations) को होने देगा ।

(D) इसका किसी भी प्रकार के विश्लेषण पर कोई प्रभाव नहीं पड़ेगा ।

Prepp
Your Personal Exams Guide

(v) The Root Mean Square Deviation is the square root of : **1**

(A) Mean of residuals (B) Variance of residuals

(C) Standard deviation (D) Mean absolute error

(vi) The approach that the K-means Clustering Algorithm follows to solve a given clustering problem is called : **1**

(A) Anomaly Detection (B) Expectation-Maximization

(C) Decision Tree (D) Linear Regression

4. Answer any **5** out of the given **6** questions. **5 × 1 = 5**

(i) On 11th December, 2019, which ministry tabled the Personal Data Protection Bill, 2019 in the Indian Parliament ? **1**

(A) Ministry of Health and Family Welfare

(B) Ministry of Law and Justice

(C) Ministry of Finance

(D) Ministry of Electronics and Information Technology

(ii) While entering data, Rehan has mistakenly recorded the variable **`Distance`** as **`text`** instead of **`numeric`**. Which of the following statements is true for this scenario ? **1**

(A) Once the data is recorded, it cannot be cleaned during any phase of data analysis.

(B) It will cause problems while using the data to build a model.

(C) It will allow all mathematical operations without any issues.

(D) It has no effect on any type of analysis.

www.prepp.in

(iii) K-NN को नॉन–पैरामीट्रिक एल्गोरिथम (non-parametric algorithm) क्यों माना जाता है ? **1**

(A) यह वर्गीकरण (classification) से पहले डेटा के लिए एक विशिष्ट वितरण (specific distribution) मानता है ।

(B) इसे डेटा वितरण के लिए लर्निंग पैरामीटर (learning parameters) सीखने की आवश्यकता होती है ।

(C) यह डेटा के वितरण के बारे में कुछ भी नहीं मानता है ।

(D) यह डेटा को मॉडल करने के लिए पैरामीटर (parameters) की एक निश्चित संख्या का उपयोग करता है ।

(iv) स्नेहा लीनियर रिग्रेशन (linear regression) कर रही है । यह निर्धारित करने के लिए कि अवलोकित बिंदु (observed points), मॉडल के अनुमानित मानों (predicted values) के कितने करीब हैं, वह रूट मीन स्क्वायर डेविएशन (Root Mean Square Deviation) विधि का उपयोग कर रही है । यदि एक बड़ा विचलन मान (deviation value) प्राप्त होता है तो अनुशंसित कार्रवाई (recommended action) क्या है ? **1**

(A) मॉडल से आश्रित चर (dependent variable) को हटा दें ।

(B) स्वतंत्र चर (independent variable) को आश्रित चर (dependent variable) में बदल दें ।

(C) मॉडल को पुन: प्रशिक्षित करें (Retrain the model)

(D) लीनियर रिग्रेशन से वर्गीकरण एल्गोरिथम (classification algorithm) पर स्विच करें ।

(v) कथन-1 : लोगारिद्मिक फंक्शन्स (Logarithmic functions) रैखिक फंक्शन्स (linear functions) के उदाहरण हैं ।

कथन-2 : नॉन–लीनियर रिग्रेशन (non-linear regression) का ग्राफ एक वक्र (curve) के समीकरण का अनुसरण (follow) करता है । **1**

(A) कथन-1 और कथन-2 दोनों सही हैं ।

(B) कथन-1 और कथन-2 दोनों गलत हैं ।

(C) कथन-1 सही है, लेकिन कथन-2 गलत है ।

(D) कथन-2 सही है, लेकिन कथन-1 गलत है ।

(vi) K-मीन्स क्लस्टरिंग (K-means clustering) एक ऐसी तकनीक है जिसका उपयोग एक डेटासेट में डेटा क्लासों (data classes) के समूह (cluster) का पता लगाने के लिए किया जाता है और इसे ______ के रूप में वर्गीकृत किया जाता है । **1**

(A) एक सुपरवाइज़्ड लर्निंग तकनीक (A supervised learning technique)

(B) एक वर्गीकरण तकनीक (A classification technique)

(C) एक अनसुपरवाइज़्ड मशीन लर्निंग तकनीक (An unsupervised machine learning technique)

(D) एक रिग्रेशन तकनीक (A regression technique)

(iii) Why is K-NN considered as non-parametric algorithm ? **1**

(A) It assumes a specific distribution for the data before classification.

(B) It requires learning parameters for the data distribution.

(C) It does not assume anything about the distribution of the data.

(D) It uses a fixed number of parameters to model the data.

(iv) Sneha is performing linear regression. She is using the Root Mean Square Deviation method to determine how close the observed points are to the model's predicted values. What is the recommended action if a large deviation value is obtained ? **1**

(A) Remove the dependent variable from the model.

(B) Change the independent variable to the dependent variable.

(C) Retrain the model.

(D) Switch from linear regression to a classification algorithm.

(v) Statement-1 : Logarithmic functions are examples of linear functions.

Statement-2 : The graph of non-linear regression follows the equation of a curve. **1**

(A) Both Statement-1 and Statement-2 are correct.

(B) Both Statement-1 and Statement-2 are incorrect.

(C) Statement-1 is correct but Statement-2 is incorrect.

(D) Statement-2 is correct but Statement-1 is incorrect

(vi) K-means clustering is a technique used to spot clusters of data classes in a dataset and is classified as ________. **1**

(A) A supervised learning technique

(B) A classification technique

(C) An unsupervised machine learning technique

(D) A regression technique

www.prepp.in

Prepp
Your Personal Exams Guide

5. दिये गये **6** प्रश्नों में से किन्हीं **5** प्रश्नों के उत्तर दीजिए । **5 × 1 = 5**

(i) निम्नलिखित में से कौन GDPR कानून का एक महत्वपूर्ण पहलू (aspect) है ? **1**

(A) डेटा गुणन (Data multiplication)

(B) भुला दिए जाने का अधिकार (Right to be forgotten)

(C) अनिवार्य डेटा साझाकरण (Mandatory data sharing)

(D) असीमित डेटा भंडारण (Unlimited data storage)

(ii) कथन-1 : रिग्रेशन ट्रीज़ (Regression trees) का उपयोग तब किया जाता है जब आश्रित चर (dependent variable) श्रेणीबद्ध (categorical) होता है ।

कथन-2 : एक डिसीजन ट्री में एक नोड की दो या दो से अधिक शाखाएँ (branches) हो सकती हैं । **1**

(A) कथन-1 और कथन-2 दोनों सही हैं ।

(B) कथन-1 और कथन-2 दोनों गलत हैं ।

(C) कथन-1 सही है, लेकिन कथन-2 गलत है ।

(D) कथन-2 सही है, लेकिन कथन-1 गलत है ।

(iii) K-NN एल्गोरिथम के लिए निम्नलिखित में से कौन सा सही है ? **1**

(A) यह असंतुलित डेटा (imbalanced data) के साथ अत्यधिक सटीक (accurate) है ।

(B) यह आयामी अभिशाप (curse of dimensionality) से प्रतिरक्षित (immune) है ।

(C) यह धीमा और मेमोरी-अकुशल (memory-inefficient) है ।

(D) यह आउटलायर्स (outliers) के लिए रोबस्ट (robust) है ।

(iv) एक सरल लीनियर रिग्रेशन (simple linear regression) का समीकरण `Y = m * X + b` के रूप का होता है । समीकरण में **m** और **b** क्या दर्शाते हैं ? **1**

(A) ढाल (Slope), स्वतंत्र चर (independent variable)

(B) स्वतंत्र चर (Independent variable), इंटरसेप्ट (Intercept)

(C) ढाल (Slope), इंटरसेप्ट (Intercept)

(D) आश्रित चर (Dependent Variable), ढाल (Slope)

(v) एक सरल रिग्रेशन समीकरण के दाहिने हाथ की ओर एक इंटरसेप्ट (intercept) और एक्स्प्लेनेटरी वेरिएबल (explanatory variable) के साथ एक _______ होता है । **1**

(A) कोफीसिएंट (Co-efficient)

(B) आश्रित चर (Dependent variable)

(C) इंटरसेप्ट (Intercept)

(D) लीनियर रिग्रेशन (Linear regression)

Prepp
Your Personal Exams Guide

5. Answer any **5** out of the given **6** questions. **5 × 1 = 5**

(i) Which of the following is an important aspect of GDPR legislation ? **1**

(A) Data Multiplication (B) Right to be forgotten

(C) Mandatory Data Sharing (D) Unlimited Data Storage

(ii) Statement-1 : Regression trees are used when the dependent variable is categorical.

Statement-2 : A node can have two or more branches in a decision tree. **1**

(A) Both Statement-1 and Statement-2 are correct.

(B) Both Statement-1 and Statement-2 are incorrect.

(C) Statement-1 is correct but Statement-2 is incorrect.

(D) Statement-2 is correct but Statement-1 is incorrect.

(iii) Which of the following is correct for the K-NN algorithm ? **1**

(A) It is highly accurate with imbalanced data.

(B) It is immune to the curse of dimensionality.

(C) It is slow and memory-inefficient.

(D) It is robust to outliers.

(iv) The equation for a simple linear regression is of the form `Y = m * X + b`. What do `m` and `b` in the equation ? **1**

(A) Slope, Independent variable

(B) Independent variable, Intercept

(C) Slope, Intercept

(D) Dependent Variable, Slope

(v) A simple regression equation has an intercept on the right-hand side and an explanatory variable with a/an _______. **1**

(A) Coefficient (B) Dependent variable

(C) Intercept (D) Linear regression

Prepp
Your Personal Exams Guide

(vi) K-मीन्स क्लस्टरिंग (K-means clustering) एल्गोरिथम, यह सबसे पहले इनपुट के आधार पर क्लस्टर्स (clusters) की संख्या, K को निर्दिष्ट (specifies) करता है । इसके बाद यह सेंट्रोइड्स (centroids) को कैसे इनीशियलाइज (initialize) करता है ? 1

(A) सभी डेटा पॉइंट (data points) को किसी एक क्लस्टर को एसाइन (assign) करना ।

(B) वर्ग दूरी (squared distance) के योग की गणना करना और K क्लस्टर्स को एसाइन (assign) करना ।

(C) डेटा पॉइंट को शफल (shuffling) करना और फिर रैंडमली (randomly) K डेटा पॉइंट का चयन करना ।

(D) क्लस्टर, K के लिए केवल लेबल किए गए डेटा (labelled data) का उपयोग करना ।

खण्ड – ख

(विषयपरक प्रकार के प्रश्न) **(26 अंक)**

रोज़गार कौशल पर दिये गये **5** प्रश्नों में से किन्हीं **3** प्रश्नों के उत्तर **20-30** शब्दों में दीजिए । **3 × 2 = 6**

6. "भाषा सीखने के लिए व्यक्ति को चार प्रमुख कौशल विकसित करने की आवश्यकता है ।" कौशल का नाम बताइए । **2**

7. प्रेरणा शब्द से आप क्या समझते हैं ? **2**

8. डिजिटल प्रेजेंटेशन को कम्प्यूटर पर कैसे सेव किया जा सकता है ? इसका कोई एक फ़ायदा बताइए । **2**

9. "उद्यमिता कला और विज्ञान का उत्तम संयोजन है", दिए गए कथन को समझाइए । **2**

10. हरित–नौकरियों से संबंधित उपयुक्त प्रौद्योगिकी से आपका क्या तात्पर्य है ? कोई दो उदाहरण भी दीजिए । **2**

Prepp
Your Personal Exams Guide

(vi) The K-means Clustering algorithm, first specifies the number of clusters, K depending on the input. Thereafter how does it initialize the centroids ? **1**

(A) By assigning all data points to any one cluster.

(B) By computing the sum of the squared distance and assigning to the K clusters.

(C) By shuffling the data points and then selecting K data points randomly.

(D) By using only the labelled data for clusters, K.

Section – B

(Subjective Type Questions) (26 Marks)

Answer any **3** out of the given **5** questions on Employability Skills in **20-30** words each. **3 × 2 = 6**

6. "To learn the language, one needs to develop four key skills." Name the skills. **2**

7. What do you understand by the term motivation ? **2**

8. How digital presentation can be saved on the computer ? Give any one advantage of it. **2**

9. "Entrepreneurship is the perfect combination of art and science." Explain the given statement. **2**

10. What do you mean by appropriate technology related to green-jobs ? Give any two examples also. **2**

Prepp
Your Personal Exams Guide

दिए गए **6** प्रश्नों में से किन्हीं **4** प्रश्नों के उत्तर **20-30** शब्दों में दीजिए । **4 × 2 = 8**

11. डेटा गोपनीयता (Data privacy) क्या है ? समझाइए कि यदि कोई कंपनी व्यक्तिगत पहचान योग्य जानकारी (personal identifiable information) को एन्क्रिप्टेड (encrypted) फॉर्मेट में सुरक्षित रूप से संगृहीत करती है, तब भी डेटा गोपनीयता नियमों का उल्लंघन क्यों हो सकता है ? **2**

12. रौनक का एक छोटा सा कैफे (cafe) है और वह यह समझना चाहता है कि बाहर के तापमान और प्रतिदिन बेचे जाने वाले कोल्ड ड्रिंक्स की संख्या के बीच कोई सहसंबंध (correlation) है या नहीं । उसने 30 दिनों का डेटा एकत्र किया है जिसमें दैनिक तापमान और कोल्ड ड्रिंक की बिक्री दिखाई गई है । **2**
    (a) इस सहसंबंध का अध्ययन करने के लिए उसे किस प्रकार के विश्लेषण (analysis) का उपयोग करना चाहिए ?
    (b) इस डेटा को विज़ुअलाइज़ (visualize) करने के लिए वह जिस किसी एक ग्राफिकल विधि (graphical method) का उपयोग कर सकता है, उसका नाम बताएँ ।

13. क्रॉस-वैलिडेशन (Cross-Validation) की किसी एक लोकप्रिय विधि का नाम बताइए और उसकी व्याख्या करें । **2**

14. डिसीजन ट्री (Decision Trees) को वर्सटाइल (versatile) क्यों माना जाता है ? किन्हीं दो बिंदुओं का उल्लेख करें । **2**

15. मीन एब्सोल्यूट एरर (Mean Absolute Error-MAE) क्या है ? 'डेटा पर लाइन को फिट करने' (Fitting the line to the data) की प्रक्रिया की भी व्याख्या करें, जिसके लिए MAE का उपयोग होता है । **2**

16. सुपरवाइज़्ड लर्निंग (Supervised Learning) और अनसुपरवाइज़्ड लर्निंग (Unsupervised Learning) के बीच किन्हीं दो अंतरों का उल्लेख करें । **2**

दिये गये **5** प्रश्नों में से किन्हीं **3** प्रश्नों के उत्तर **50-80** शब्दों में दीजिए । **3 × 4 = 12**

17. HIPAA कानून (legislation) के बारे में निम्नलिखित प्रश्नों के उत्तर दें : **1+2+1=4**
    (a) HIPAA संक्षिप्त नाम (abbreviation) का विस्तार (expand) करें ।
    (b) HIPAA के किन्हीं दो उद्देश्यों (objectives) को बताएँ ।
    (c) पर्सनल आइडेंटीफायर्स (Personal identifiers) के कोई दो उदाहरण दें, जिनकी HIPAA सुरक्षा करता है ।

www.prepp.in

Prepp
Your Personal Exams Guide

Answer any **4** out of the given **6** questions in **20-30** words each. **4 × 2 = 8**

11. What is Data privacy ? Explain why data privacy rules can be violated even if a company stores personal identifiable information securely in an encrypted format ? **2**

12. Raunak owns a small café and wants to understand if there is a correlation between the temperature outside and the number of cold drinks sold daily. He has collected data for 30 days showing daily temperature and cold drink sales. **2**

    (a) What type of analysis should he use to study this correlation ?

    (b) Name any one graphical method he can use to visualize this data.

13. Name and explain any one popular method of cross-validation. **2**

14. Why are decision trees considered versatile ? Mention any two points. **2**

15. What is Mean Absolute Error (MAE) ? Also explain the process of 'Fitting the line to the data' that uses MAE. **2**

16. Mention any two differences between Supervised Learning and Unsupervised Learning. **2**

Answer any **3** out of the given **5** questions in **50-80** words each. **3 × 4 = 12**

17. Answer the following questions about HIPAA legislation : **1+2+1=4**

    (a) Expand the abbreviation HIPAA.

    (b) State any two objectives of HIPAA.

    (c) Give any two examples of personal identifiers that HIPAA protects.

www.prepp.in

Prepp
Your Personal Exams Guide

18. यूनिवेरिएट एनालिसिस (Univariate Analysis) की अवधारणा (concept) की व्याख्या करें। इसके मुख्य उद्देश्य (main purpose) का वर्णन करें और इससे जुड़ी दो अलग-अलग सांख्यिकीय (Statistical) और दो अलग-अलग ग्राफिकल (graphical) विधियों को सूचीबद्ध करें। 4

19. डिसीजन ट्री (Decision Tree) क्या है ? डिसीजन ट्री बनाने में शामिल चरणों की व्याख्या करें। 4

20. (a) एक स्वास्थ्य शोधकर्ता (health researcher) चार कारकों : आयु **(age)**, वजन **(weight)**, कोलेस्ट्रॉल स्तर **(cholesterol level)** और दैनिक व्यायाम की अवधि **(daily exercise duration)** के आधार पर एक मरीज के रक्तचाप (blood pressure) को predict करना चाहता है।

इस संबंध का विश्लेषण करने के लिए सबसे उपयुक्त रिग्रेशन मॉडल (regression model) की पहचान करें और उसे परिभाषित करें। समझाएँ कि यह मॉडल इस परिदृश्य (scenario) के लिए क्यों उपयुक्त है।

(b) रुद्र ने छह महीने पहले अपना व्यवसाय शुरू किया था। उसकी आय में हर महीने तेजी से (exponential) (5 गुना) वृद्धि हुई है। सातवें महीने में उसकी आय का अनुमान लगाने के लिए किस प्रकार के रिग्रेशन (regression) का उपयोग किया जा सकता है ? **3+1=4**

21. निम्नलिखित वास्तविक-दुनिया के अनुप्रयोगों (real-world applications) में अनसुपरवाइज़्ड लर्निंग (Unsupervised Learning) की भूमिका (role) की व्याख्या करें : 4

(i) रेकमेंडेशन इंजन (Recommendation engines)

(ii) अनोमली डिटेक्शन (Anomaly detection)

(iii) मेडिकल इमेजिंग (Medical imaging)

(iv) न्यूज़ वर्गीकरण (News categorization)

www.prepp.in

Prepp
Your Personal Exams Guide

18. Explain the concept of Univariate Analysis. Describe its main purpose and list two different statistical and two different graphical methods associated with it. **4**

19. What is a Decision tree ? Explain the steps involved in creating a decision tree. **4**

20. (a) A health researcher wants to predict a patient's blood pressure based on four factors : **`age`**, **`weight`**, **`cholesterol level`** and **`daily exercise duration`**.

    Identify and define the most suitable regression model to analyze the relationship. Explain why this model is appropriate for this scenario.

    (b) Rudra started her business six months back. Her income has increased exponentially (5 times) every month. Which type of regression can be used to predict her income in the seventh month ? **3+1=4**

21. Explain the role of Unsupervised Learning in the following real-world applications : **4**

    (i) Recommendation engines

    (ii) Anomaly detection

    (iii) Medical imaging

    (iv) News categorization

Your Personal Exams Guide