Series : 4YWZX

SET – 4

रोल नं.

Roll No.

प्रश्न-पत्र कोड Q.P. Code	**368**

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.

डाटा साइंस

DATA SCIENCE

निर्धारित समय : 2 घण्टे | *अधिकतम अंक : 50*

*Time allowed : **2** hours* | *Maximum Marks : **50***

- कृपया जाँच कर लें कि इस प्रश्न-पत्र में मुद्रित पृष्ठ **23** हैं ।
- कृपया जाँच कर लें कि इस प्रश्न-पत्र में **21** प्रश्न हैं ।
- प्रश्न-पत्र में दाहिने हाथ की ओर दिए गए प्रश्न-पत्र कोड को परीक्षार्थी उत्तर-पुस्तिका के मुख-पृष्ठ पर लिखें ।
- ***कृपया प्रश्न का उत्तर लिखना शुरू करने से पहले, उत्तर-पुस्तिका में यथा स्थान पर प्रश्न का क्रमांक अवश्य लिखें ।***
- 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 ।

- Please check that this question paper contains **23** printed pages.
- Please check that this question paper contains **21** questions.
- 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.
- **Please write down the serial number of the question in the answer-book at the given place before attempting it.**
- 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.

~

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 ।*

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.*

खण्ड – क

(वस्तुनिष्ठ प्रकार के प्रश्न)

1. रोजगारपरक कौशल पर दिए गए **6** प्रश्नों में से किन्हीं **4** प्रश्नों के उत्तर दीजिए । **$4 \times 1 = 4$**

(i) एक व्यक्ति के व्यक्तित्व का वर्णन करने के संदर्भ में FFM का पूर्ण रूप है :

(A) Flexible Framework Model (B) Functional Flexibility Model

(C) Focused Feedback Method (D) Five Factor Model

(ii) निम्नलिखित में से किस बाधा का सामना उद्यमी अपने उद्यम का संचालन करते हुए कर सकते हैं ?

(A) कुशल श्रमिकों की उपलब्धता

(B) आत्म-विश्वास

(C) समय पर मौद्रिक संसाधनो (monetary resources) का उपलब्ध न होना

(D) जोखिम लेने की क्षमता

(iii) निम्नलिखित में से एक्टिव वॉइस (Active Voice) वाक्य की पहचान कीजिए ।

(A) होम टीम द्वारा गेम जीता गया । (B) रिपोर्ट मैनेजर द्वारा लिखी जा रही है ।

(C) मैनेजर रिपोर्ट लिख रहा है । (D) छात्र द्वारा होमवर्क किया गया ।

(iv) एक प्रेजेंटेशन (presentation) में नई स्लाइड जोड़ने के लिए शॉर्टकट की (key) है :

(A) Ctrl + N (B) Ctrl + Alt + N

(C) Ctrl + M (D) Ctrl + Alt + M

(v) उस अभिप्रेरणा (motivation) के लिये क्या शब्द प्रयोग किया जाता है जो बाहरी पारितोषिक (external rewards) के स्थान पर व्यक्ति के अंदर से आती है ?

(A) बाह्य (Extrinsic) अभिप्रेरणा

(B) वेतनवृद्धि (Increment) अभिप्रेरणा

(C) आंतरिक (Intrinsic) अभिप्रेरणा

(D) प्रोत्साहन-आधारित (Incentive-based) अभिप्रेरणा

(vi) ग्रीन जॉब्स के संदर्भ में उपयुक्त टेक्नोलॉजी (Appropriate Technology) _________ टेक्नोलॉजी है जो पर्यावरण अनुकूल है और स्थानीय आवश्यकताओं के अनुरूप है ।

(A) लार्ज-स्केल (large-scale) (B) स्मॉल-स्केल (small-scale)

(C) रिमोट-स्केल (remote-scale) (D) ओल्ड-स्केल (old-scale)

Section – A

(Objective Type Questions)

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

(i) With reference to describing an individual's personality, FFM stands for :

(A) Flexible Framework Model (B) Functional Flexibility Model

(C) Focused Feedback Method (D) Five Factor Model

(ii) Which of the following is a barrier that entrepreneurs may face while running their ventures ?

(A) Availability of skilled labours

(B) Self-confidence

(C) Unavailability of monetary resources on time

(D) Risk taking capability

(iii) Identify the active voice sentence from the following :

(A) The game was won by the home team.

(B) The report is being written by the manager.

(C) The manager is writing the report.

(D) The homework was done by the student.

(iv) The shortcut key to add a new slide in a presentation is :

(A) Ctrl + N (B) Ctrl + Alt + N

(C) Ctrl + M (D) Ctrl + Alt + M

(v) Which term is used for motivation that comes from within an individual rather than external rewards ?

(A) Extrinsic motivation (B) Increment motivation

(C) Intrinsic motivation (D) Incentive-based motivation

(vi) With respect to green jobs, an Appropriate technology is the _________ technology that is environment friendly and suited to local needs.

(A) large-scale (B) small-scale

(C) remote-scale (D) old-scale

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

(i) बताइये कि निम्नलिखित कथन सत्य है या असत्य :

डाटा प्राइवेसी के सन्दर्भ में, कंपनियों को यूजर डाटा (user data) का एकमात्र स्वामी माना जाता है।

(ii) निम्नलिखित हिस्टोग्राम (histogram) क्या दर्शाता है ?

(A) यूनिवेरिएट एनालिसिस (Univariate Analysis)

(B) बाइवेरिएट एनालिसिस (Bivariate Analysis)

(C) मल्टीवेरिएट एनालिसिस (Multivariate Analysis)

(D) क्लस्टरिंग (Clustering)

(iii) उस एल्गोरिथ्म (algorithm) का नाम बताइए जो कई decision trees के परिणामों के आधार पर पूर्वानुमान (predictions) करता है।

(A) क्लासिफिकेशन एल्गोरिथ्म (Classification algorithm)

(B) के-मीन्स क्लस्टरिंग (K-means clustering)

(C) रेन्डम फॉरेस्ट एल्गोरिथ्म (Random Forest algorithm)

(D) के-नियरेस्ट नेबर एल्गोरिथ्म (K-nearest neighbour algorithm)

(iv) कौन सी विशेषता K-Nearest Neighbours (K-NN) algorithm का वर्णन करती है ?

(A) लेज़ी लर्निंग (Lazy learning) और पैरामेट्रिक लर्निंग (parametric learning)

(B) इगर लर्निंग (Eager learning) और नॉन-पैरामेट्रिक लर्निंग (non-parametric learning)

(C) लेज़ी लर्निंग (Lazy learning) और नॉन-पैरामेट्रिक लर्निंग (non-parametric learning)

(D) इगर लर्निंग (Eager learning) और पैरामेट्रिक लर्निंग (parametric learning)

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

(i) State whether the following statement is True or False :

In terms of data privacy, companies are considered the sole owners of user data.

(ii) The following histogram represents :

(A) Univariate Analysis
(B) Bivariate Analysis
(C) Multivariate Analysis
(D) Clustering

(iii) Name the algorithm which makes predictions based on the outcomes of several decision trees.

(A) Classification algorithm
(B) K-means clustering
(C) Random Forest algorithm
(D) K-nearest neighbour algorithm

(iv) Which characteristics describe K-Nearest Neighbours (K-NN) algorithm ?

(A) Lazy learning and parametric learning
(B) Eager learning and non-parametric learning
(C) Lazy learning and non-parametric learning
(D) Eager learning and parametric learning

(v) वर्षा और फसल वृद्धि (crop growth) के बीच संबंध और वर्षा के निश्चित स्तर के लिए फसल वृद्धि की मात्रा की जाँच करने के लिए किस प्रकार का विश्लेषण उपयुक्त है ?

(A) क्लासिफिकेशन (Classification)

(B) क्लस्टरिंग (Clustering)

(C) लीनियर रिग्रेशन (Linear Regression)

(D) डिसीजन ट्री (Decision Tree)

(vi) उत्पाद बेचने वाली वेबसाइट यह पूर्वानुमान लगाने के लिए कि किसी ग्राहक के क्या उत्पाद खरीदने की संभावना है, रिकमेंडेशन इंजन (recommendation engines) का प्रयोग करती है । यह कार्य किसके माध्यम से किया जाता है ?

(A) सुपरवाइज्ड लर्निंग (Supervised Learning)

(B) अनसुपरवाइज्ड लर्निंग (Unsupervised Learning)

(C) रीइनफोर्समेंट लर्निंग (Reinforcement Learning)

(D) नैचुरल लैंग्वेज प्रोसेसिंग (Natural Language Processing)

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

(i) CCPA का पूर्ण रूप है :

(A) California Consumer Privacy Act

(B) Children's Consumer Privacy Act

(C) California Consumer Public Act

(D) Children's Consumer Public Act

(ii) एक डाटासेट में मिसिंग डाटा (missing data) को हैंडल करने के लिए निम्नलिखित में से कौन सी सामान्य तकनीकें हैं ?

1. डेटा का पूरा कॉलम हटाना ।
2. मिसिंग वैल्यूज (missing values) वाले डेटा की पंक्ति हटाना ।
3. मिसिंग डेटा में वैरिएबल (variable) के मीन (mean) या मोड (mode) के पास की वैल्यू डालना ।
4. मिसिंग डेटा को वैसे छोड़ देना जैसा वह है ।
5. कोई रैंडम वैल्यू (random value) डालना ।

(A) 1 और 2 (B) 2 और 3

(C) 3 और 4 (D) 4 और 5

(v) Which type of analysis is suitable for examining the relationship between rainfall and crop growth and the amount of crop growth for a certain level of rainfall ?

(A) Classification (B) Clustering

(C) Linear Regression (D) Decision Tree

(vi) A website selling products uses recommendation engines to predict what products a customer is likely to purchase. This functionality is achieved through :

(A) Supervised Learning (B) Unsupervised Learning

(C) Reinforcement Learning (D) Natural Language Processing

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

(i) CCPA stands for :

(A) California Consumer Privacy Act

(B) Children's Consumer Privacy Act

(C) California Consumer Public Act

(D) Children's Consumer Public Act

(ii) Which of the following are common techniques for handling missing data in a dataset ?

1. Remove the entire column of data.
2. Remove the row of data with missing values.
3. Insert a value close to the mean or mode of the variable with missing data.
4. Leave the missing data as it is.
5. Insert a random value.

(A) 1 and 2 (B) 2 and 3

(C) 3 and 4 (D) 4 and 5

(iii) क्लासिफिकेशन ट्री (Classification tree) के मामले में, अदृष्ट डाटा (unseen data) पर पूर्वानुमान _______ का प्रयोग करके किए जाते हैं।

(A) माध्य (Mean)
(B) माध्यिका (Median)
(C) बहुलक (Mode)
(D) RMSE

(iv) आयामीता की समस्या (curse of dimensionality) K-Nearest Neighbours (K-NN) algorithm की परफार्मेंस को किस प्रकार प्रभावित करती है ?

(A) K-NN इनपुट वेरिएबल्स (input variables) की बढ़ती संख्या के साथ अधिक अच्छा परफोर्म करता है।

(B) K-NN इनपुट वेरिएबल्स की बढ़ती संख्या के साथ सटीक रूप से आउटपुट का पूर्वानुमान करने के लिए स्ट्रगल करता है।

(C) K-NN बढ़ी हुई कंप्यूटेशनल पावर के कारण अधिक इनपुट वेरिएबल से और अधिक तीव्र बन जाता है।

(D) K-NN इनपुट वेरिएबल्स की बढ़ती हुई संख्या के साथ आउटलायर्स (outliers) के प्रति कम संवेदनशील बन जाता है।

(v) **अभिकथन (A) :** वास्तविक-जीवन परिदृश्य में, एक छोटी आर एम एस ई वैल्यू (RMSE value) डेटा और अधिक परिशुद्धि के लिए उपयुक्त एक बेहतर मॉडल दर्शाती है।

कारण (R) : एक बड़ी RMSE value बताती है कि मॉडल डाटा के लिए पूरी तरह फिट है।

(A) (A) और (R) दोनों सही हैं, और (R), (A) का सही स्पष्टीकरण है।

(B) (A) और (R) दोनों सही हैं, परंतु (R), (A) का सही स्पष्टीकरण नहीं है।

(C) (A) सही है, परंतु (R) गलत है।

(D) (A) गलत है, परंतु (R) सही है।

(vi) बताइये कि निम्नलिखित कथन सत्य है या असत्य :

क्लस्टरिंग तकनीक (Clustering Techniques) तब लागू होती है जब उदाहरणों (instances) को प्राकृतिक समूहों में विभाजित किया जाना हो।

(iii) In case of a classification tree, predictions on unseen data are made using the ________.

(A) Mean (B) Median

(C) Mode (D) RMSE

(iv) How does the curse of dimensionality affect the performance of the K-Nearest Neighbours (K-NN) algorithm ?

(A) K-NN performs better with an increasing number of input variables.

(B) K-NN struggles to predict the output accurately with an increasing number of input variables.

(C) K-NN becomes faster with more input variables due to increased computational power.

(D) K-NN becomes less sensitive to outliers with an increasing number of input variables.

(v) **Assertion (A) :** In real-life scenarios, a small RMSE value indicates a better model fit to the data and higher accuracy.

Reason (R) : A large RMSE value suggests that the model fits the data well.

(A) Both (A) and (R) are true, and (R) is the correct explanation of (A).

(B) Both (A) and (R) are true, but (R) is not the correct explanation of (A).

(C) (A) is true, but (R) is false.

(D) (A) is false, but (R) is true.

(vi) State if the following statement is True or False :

"Clustering Techniques apply when the instances are to be divided into natural groups."

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

(i) निम्नलिखित में से कौन डेटा प्राइवेसी (data privacy) को परिभाषित करता है ?

(A) व्यक्तियों का यह नियंत्रित करने का अधिकार कि उनकी व्यक्तिगत सूचना कैसे एकत्रित और प्रयोग की जाती है।

(B) मार्केटिंग उद्देश्यों के लिए व्यक्तिगत सूचना एकत्र करने के लिए संगठनों का दायित्व।

(C) सभी प्लेटफार्मों पर व्यक्तिगत सूचना को अबाधित रूप से साझा करना।

(D) अनुसंधान के उद्देश्यों के लिए बिना सहमति व्यक्तिगत सूचना का प्रयोग।

(ii) किसी शहर में घर के मूल्यों का पूर्वानुमान बैडरूम की संख्या, घर का आकार (वर्ग फूटेज), अवस्थिति और सम्पत्ति की आयु जैसे कारकों पर निर्भर करता है। इस परिदृश्य के लिए कौन सी विश्लेषण तकनीक (analysis technique) उपयुक्त है ?

(A) यूनिवेरिएट (Univariate)

(B) ज़ीरोवेरिएट (Zerovariate)

(C) मल्टीवेरिएट (Multivariate)

(D) यूनिवेरिएट (Univariate) और मल्टीवेरिएट (Multivariate) दोनों

(iii) K-Nearest Neighbours (K-NN) एल्गोरिथ्म किस मूलभूत सिद्धान्त पर ऑपरेट करता है ?

(A) जो ऑब्जेक्ट (Objects) समान होते हैं, उनकी प्रवृत्ति अलग दूर स्थित होने की होती है।

(B) जो ऑब्जेक्ट (Objects) समान होते हैं, उनकी प्रवृत्ति बेतरतीब (random) दूरियों पर स्थित होने की होती है।

(C) जो ऑब्जेक्ट (Objects) समान होते हैं, उनकी प्रवृत्ति बिखरे (scattered) तरीके से स्थित होने की होती है।

(D) जो ऑब्जेक्ट (Objects) समान होते हैं, उनकी प्रवृत्ति एक दूसरे के निकट स्थित होने की होती है।

(iv) Linear regression किस प्रकार के variable का पूर्वानुमान करता है ?

(A) सुस्पष्ट (Categorical) (B) पृथक (Discrete)

(C) सतत (Continuous) (D) द्विगुण (Binary)

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

(i) Which of the following defines data privacy ?

(A) The right of individuals to control how their personal information is collected and used.

(B) The obligation of organisations to collect personal information for marketing purposes.

(C) The unrestricted sharing of personal information across platforms.

(D) The use of personal information without consent for research purposes.

(ii) Prediction of house prices in a city depends on factors such as the number of bedrooms, size of the house (square footage), location and age of the property. Which analysis technique is appropriate for this scenario ?

(A) Univariate

(B) Zerovariate

(C) Multivariate

(D) Both Univariate and Multivariate

(iii) Which fundamental principle does the K-Nearest Neighbours (K-NN) algorithm oparate on ?

(A) Objects that are similar tend to be located far apart.

(B) Objects that are similar tend to be located at random distances.

(C) Objects that are similar tend to be located in a scattered manner.

(D) Objects that are similar tend to be located close to each other.

(iv) Which type of variable does linear regression predict ?

(A) Categorical (B) Discrete

(C) Continuous (D) Binary

(v) नीचे दिया गया फॉर्मूला देखें :

y ~ f(x, β)

जहाँ **x** इंडीपेंडेंट वेरिएबल्स (independent variables) का एक vector है और **y** डिपेंडेंट वेरिएबल (dependent variable) है।

निम्नलिखित में से उपयुक्त रिग्रेशन का चुनाव करें।

(A) लीनियर रिग्रेशन (Linear Regression)

(B) नॉन–लीनियर रिग्रेशन (Non-Linear Regression)

(C) कोरिलेशन रिग्रेशन (Correlation Regression)

(D) मल्टीपल लीनियर रिग्रेशन (Multiple Linear Regression)

(vi) विषमता का पता लगाने (anomaly detection) में अनसुपरवाइज्ड लर्निंग का महत्वपूर्ण एप्लीकेशन क्या है ?

(A) स्ट्रक्चर्ड डेटा (Structured data) में पैटर्न की पहचान करना।

(B) डेटा प्वाइंट्स को प्रीडिफाइन्ड श्रेणियों में वर्गीकृत करना।

(C) असामान्य डेटा प्वाइंट्स का पता लगाना।

(D) भावी परिणामों (future outcomes) का पूर्वानुमान लगाना

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

(i) COPPA का पूर्ण रूप है _________ Online Privacy Protection Act.

(A) California's (B) Children's

(C) Center's (D) Channel's

(ii) कौन सा कथन डिपेंडेंट वेरिएबल (dependent variable) की प्रकृति के आधार पर रिग्रेशन और क्लासिफिकेशन ट्रीज (regression and classification trees) के प्रयोग का सही वर्णन करता है ?

(A) Regression trees सतत (continuous) डिपेंडेंट वेरिएबल्स के लिए उपयुक्त हैं, जबकि classification trees सुस्पष्ट (categorical) डिपेंडेंट वेरिएबल्स के लिए उपयुक्त हैं।

(B) Classification trees सतत (continuous) डिपेंडेंट वेरिएबल्स के लिए उपयुक्त हैं, जबकि Regression trees सुस्पष्ट (categorical) डिपेंडेंट वेरिएबल्स के लिए उपयुक्त हैं।

(C) Regression और classification trees दोनों सतत (continuous) डिपेंडेंट वेरिएबल्स के लिए उपयुक्त हैं।

(D) Regression और classification trees दोनों सुस्पष्ट (categorical) डिपेंडेंट वेरिएबल्स के लिए उपयुक्त हैं।

(v) Consider the formula given below :

y ~ f(x, β)

where **x** is a vector of independent variables and **y** is the dependent variable.

Identify the type of regression from the following.

(A) Linear Regression (B) Non-Linear Regression

(C) Correlation Regression (D) Multiple Linear Regression

(vi) What is a significant application of unsupervised learning in anomaly detection ?

(A) Identifying patterns in structured data

(B) Classifying data points into predefined categories

(C) Detecting unusual data points

(D) Predicting future outcomes

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

(i) COPPA stands for _________ Online Privacy Protection Act.

(A) California's (B) Children's

(C) Center's (D) Channel's

(ii) Which statement correctly describes the use of regression and classification trees based on the nature of the dependent variable ?

(A) Regression trees are suitable for continuous dependent variables, while classification trees are suitable for categorical dependent variables.

(B) Classification trees are suitable for continuous dependent variables, while regression trees are suitable for categorical dependent variables.

(C) Both regression and classification trees are suitable for continuous dependent variables.

(D) Both regression and classification trees are suitable for categorical dependent variables.

(iii) निम्नलिखित चित्र (diagram) क्या दर्शाता है :

[redacted]

(A) क्रॉस वेलिडेशन (Cross validation) (B) रिग्रेशन (Regression)
(C) क्लासिफिकेशन (Classification) (D) न्यूरल नेटवर्क्स (Neural Networks)

(iv) लीनियर रिग्रेशन (Linear regression) में best fit की line का निर्धारण करते समय वास्तविक डेटा प्वाइंट्स से line का deviation न्यूनतम करने के लिए कौन सी विधि प्रयोग की जाती है ?

(A) Sum of Squared Errors (SSE) को अधिकतम करना ।
(B) Mean Absolute Error (MAE) को न्यूनतम करना ।
(C) Principal Component Analysis (PCA) का प्रयोग करना ।
(D) F-statistic को ऑप्टिमाइज (Optimise) करना ।

(v) निम्नलिखित ग्राफ दर्शाता है :

[redacted]

(A) लीनियर रिग्रेशन (Linear Regression)
(B) मल्टी–लीनियर रिग्रेशन (Multi-Linear Regression)
(C) क्लासिफिकेशन (Classification)
(D) कोरिलेशन (Correlation)

(vi) K-means algorithm में कौन सी समस्या समाधान की पद्धति अपनाई जाती है ?

(A) ग्रेडिएंट डिसेंट (Gradient Descent)
(B) एक्स्पेक्टेशन – मेक्सिमाइज़ेशन (Expectation-Maximisation)
(C) डिसिज़न ट्रीज़ (Decision Trees)
(D) सपोर्ट वेक्टर मशीन्स (Support Vector Machines)

(iii) The following diagram represents :

[redacted]

(A) Cross validation
(B) Regression
(C) Classification
(D) Neural Networks

(iv) When determining the line of best fit in linear regression, which method is used to minimize the deviation of the line from the actual data points ?

(A) Maximising the Sum of Squared Errors (SSE).
(B) Minimising the Mean Absolute Error (MAE).
(C) Using Principal Component Analysis (PCA).
(D) Optimising the F-statistic.

(v) The following graph represents :

[redacted]

(A) Linear Regression
(B) Multi-Linear Regression
(C) Classification
(D) Correlation

(vi) Which problem-solving approach does the K-means algorithm employ ?

(A) Gradient Descent
(B) Expectation-Maximisation
(C) Decision Trees
(D) Support Vector Machines

खण्ड – ख

(विषयपरक प्रकार के प्रश्न)

रोजगारपरक कौशल पर दिये गये **5** प्रश्नों में से किन्हीं **3** प्रश्नों का प्रत्येक का उत्तर **20-30** शब्दों में दीजिए । **$3 \times 2 = 6$**

6. एक वाक्य (sentence) में articles क्या हैं ? ऐसे वाक्य का उदाहरण भी दीजिए जिसमें article का प्रयोग दिखाया गया हो ।

7. नीचे दी गई वर्कशीट पर विचार कीजिए :

	A	B	C	D	E
1	**S. No.**	**Item**	**Quantity**	**Price**	**Amount**
2	**1**	Pen	10	60	
3	**2**	Glue stick	11	45	
4	**3**	Notebook	11	75	
5	**4**	Register	19	90	
6	**5**	Scale	5	5	

   (a) स्प्रेडशीट में, प्रत्येक फॉर्मूला एक _________ sign से आरंभ होता है ।

   (b) ग्लू-स्टिक के लिए राशि (amount) कैलकुलेट करने के लिए फॉर्मूला लिखिये । (राशि मूल्य (Price) × मात्रा (Quantity) के रूप में कैलकुलेट की जाती है ।)

8. निम्नलिखित व्यक्तित्व विकारों (personality disorders) की पहचान कीजिए और उनका नाम दीजिए :

   (a) एक विकार जिसके लक्षण हैं – अत्यधिक आत्म-संदेह, आश्वासन की निरंतर आवश्यकता, निर्णय लेने के लिये दूसरों पर निर्भरता और एकाकीपन से बचना ।

   (b) एक विकार जिसमें व्यक्ति यह विश्वास करते हैं कि वे दूसरों से श्रेष्ठ हैं, उनमें संवेदना (empathy) की कमी होती है और वे अपनी उपलब्धियों को बढ़ा-चढ़ाकर बखान करते हैं ।

9. परिवहन (transportation) में ग्रीन जॉब्स का वर्णन कीजिए ।

10. एक सफल उद्यमी के किन्हीं दो गुणों का वर्णन कीजिए ।

Section – B

(Subjective Type Questions)

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

6. What are articles in a sentence ? Also give an example of a sentence exhibiting the usage of article.

7. Consider the following worksheet :

	A	B	C	D	E
1	**S. No.**	**Item**	**Quantity**	**Price**	**Amount**
2	**1**	Pen	10	60	
3	**2**	Glue stick	11	45	
4	**3**	Notebook	11	75	
5	**4**	Register	19	90	
6	**5**	Scale	5	5	

(a) In a spreadsheet, every formula starts with a/an _________ sign.

(b) Write the formula to calculate the amount for Glue-stick. (Amount is calculated as Price × Quantity)

8. Identify and name the following personality disorders :

(a) A disorder characterised by intense self-doubt, a constant need for reassurance, dependency on others for decision-making, and avoidance of solitude.

(b) A disorder where individuals believe they are superior to others, lack empathy, and exaggerate their own achievements.

9. Discuss green jobs in transportation.

10. Discuss any two qualities of a successful entrepreneur.

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

11. सामाजिक लाभ और टेक्नोलॉजी का उत्तरदायी प्रयोग सुनिश्चित करने में उनके महत्त्व पर जोर देते हुए, सॉफ्टवेयर विकास और डेटा प्रबंधन से संबंधित दो नीतिपरक दिशा-निर्देशों (ethical guidelines) को स्पष्ट कीजिए ।

12. सिंगल वेरिएबल का विश्लेषण (univariate analysis) करने के लिए उपयुक्त एक प्रकार के चार्ट / ग्राफ और दो वेरिएबल्स के बीच संबंध का विश्लेषण (bivariate analysis) करने के लिए एक प्रकार के चार्ट / ग्राफ का नाम बताइए ।

13. डेटा एनालिसिस (data analysis) में क्रॉस वेलिडेशन (cross-validation) को स्पष्ट कीजिए ।

14. रिग्रेशन ट्रीज़ (Regression trees) और क्लासिफिकेशन ट्रीज़ (Classification trees) के बीच अंतर के कोई दो बिन्दु बताइए ।

15. रूट मीन स्क्वेयर डिविएशन (Root Mean Square Deviation) क्या है ? डेटा विश्लेषण में इसका प्रयोग क्यों होता है ?

16. टेक्स्ट डॉक्यूमेंट (Text documents) को सूचना पुनःप्राप्ति प्रणालियों और रिकमेंडेशन इंजन (recommendation engines) के लिए उन्हें कुशलतापूर्वक व्यवस्थित करने और वर्गीकृत करने के लिए प्रायः उनकी विषय-वस्तु समानता (content similarity) के आधार पर समूहीकृत किया जाता है ।

    इस संदर्भ में उपयोग में लाई गई अनसुपरवाइज़्ड लर्निंग (unsupervised learning) पद्धति की पहचान कीजिए और परिभाषित कीजिए ।

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

17. PDP को विस्तार से लिखिए और स्पष्ट कीजिए ।

18. मल्टीवेरिएट अनेलिसिस (Multivariate Analysis) को विस्तार से समझाइए । अपने उत्तर की पुष्टि में कम से कम एक उदाहरण दीजिए ।

19. डिसिज़न ट्री मेथोडोलोजी (Decision tree methodology) का प्रयोग करने के कोई चार लाभ बताइए ।

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

11. Explain two ethical guidelines relevant to software development and data management, emphasizing their importance in ensuring societal benefit and responsible use of technology.

12. Name one type of chart/graph suitable for analysing a single variable (univariate analysis) and one type for analyzing the relationship between two variables (bivariate analysis).

13. Explain cross-validation in data analysis.

14. Mention any two points of differentiation between Regression trees and Classification trees.

15. What is Root Mean Square Deviation ? Why it is used in data analysis ?

16. Text documents are often grouped based on their content similarity to efficiently organize and categorise them for information retrieval systems and recommendation engines.

    Identify and define the unsupervised learning method utilized in this context.

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

17. Expand and explain PDP.

18. Explain Multivariate Analysis in detail. Give at least one example to support your answer.

19. List any four advantages of using a decision tree methodology.

20. आरुष ने 6 महीने पहले एक कारोबार आरंभ किया । यह देखा गया है कि उसकी आय प्रत्येक महीने बहुत अधिक तेजी से (exponentially) बढ़ती है ।

(a) आरुष के आय की प्रवृत्ति को समझाने के लिए उपयुक्त regression model की पहचान कीजिए ।

(b) Regression model का फॉर्मूला/समीकरण लिखिए ।

(c) उपरोक्त विवेचित Regression model को दर्शाने वाला काल्पनिक डेटा के साथ सैंपल चार्ट तैयार कीजिए ।

21. (a) Clustering शब्द की परिभाषा दीजिए । वास्तविक जीवन के उदाहरण के साथ अपने उत्तर का समर्थन करें ।

(b) किन्हीं दो clustering पद्धतियों के नाम बताइए ।

20. Aarush started a business 6 months back. It has been noted that his income increases exponentially every month.

   (a) Identify the appropriate regression model to illustrate Aarush's score trend.

   (b) Write the formula/equation for the regression model.

   (c) Draw a sample chart with dummy data depicting the above discussed regression model.

21. (a) Define the term clustering. Support your answer with a real life example.

   (b) Name any two clustering methods.