

Marking Scheme
Strictly Confidential
(For Internal and Restricted use only)
Senior Secondary School Examination, 2026 (XIIth)
SUBJECT NAME : Data Science (Q.P. CODE 844/368)

General Instructions: -

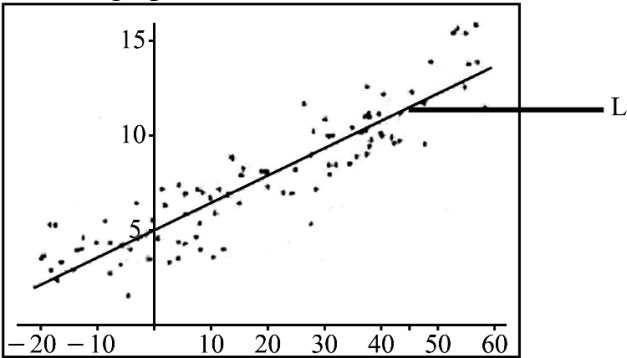
1	The CBSE has decided to introduce On Screen Marking (OSM) for the evaluation of Class XII answer Book with the 2026 Examination.
2	You are aware that evaluation is the most important process in the actual and correct assessment of the candidates. A small mistake in evaluation may lead to serious problems which may affect the future of the candidates, education system and teaching profession. To avoid mistakes, it is requested that before starting evaluation, you must read and understand the spot evaluation guidelines carefully.
3	“Evaluation policy is a confidential policy as it is related to the confidentiality of the examinations conducted, evaluation done and several other aspects. Its leakage to public in any manner could lead to derailment of the examination system and affect the life and future of millions of candidates. Sharing this policy/document to anyone, publishing in any magazine and printing in Newspaper/Website, etc. may invite action under various rules of the Board and IPC.”
4	Evaluation is to be done as per instructions provided in the Marking Scheme. It should not be done according to one’s own interpretation or any other consideration. Marking Scheme should be strictly adhered to and religiously followed. However, while evaluating, answers which are based on latest information or knowledge and/or are innovative, they may be assessed for their correctness otherwise and due marks be awarded to them. In Class-XII, while evaluating two competency-based questions, please try to understand given answer and even if reply is not from marking scheme but correct competency is enumerated by the candidate, due marks should be awarded.
5	The Marking scheme carries only suggested value points for the answers. These are in the nature of Guidelines only and do not constitute the complete answer. The students can have their own expression and if the expression is correct, the due marks should be awarded accordingly.
6	The Head-Examiner must go through the first five answer books evaluated by each evaluator on the first day, to ensure that evaluation has been carried out as per the instructions given in the Marking Scheme. If there is any variation, the same should be zero after deliberation and discussion. The remaining answer books meant for evaluation shall be given only after ensuring that there is no significant variation in the marking of individual evaluators.
7	Evaluators will mark (✓) wherever answer is correct. For wrong answer CROSS ‘X’ be marked. Evaluators will not put right (✓) while evaluating which gives an impression that answer is correct and no marks are awarded. This is most common mistake which evaluators are committing.
8	If a question has parts, please award marks on the right-hand side for each part in the OSM Portal. Marks awarded for different parts of the question will be totaled up by the OSM System.
9	If a question does not have any parts, marks must be awarded in the left-hand margin in the OSM Portal. This may also be followed strictly.

10	No marks to be deducted for the cumulative effect of an error. It should be penalized only once.
11	A full scale of marks _____ (example 0 to 80/70/60/50/40/30 marks as given in Question Paper) has to be used. Please do not hesitate to award full marks if the answer deserves it.
12	Every examiner has to necessarily do evaluation work for full working hours i.e., 8 hours every day and evaluate 20 answer books per day in main subjects and 25 answer books per day in other subjects (Details are given in Spot Guidelines). This is in view of the reduced syllabus and number of questions in question paper.
13	<p>Ensure that you do not make the following common types of errors committed by the Examiner in the past :-</p> <ul style="list-style-type: none"> • Answers marked as correct, but marks not awarded. (Ensure that the right tick mark is correctly and clearly indicated. It should merely be a line. Same is with the X for incorrect answer.) • Half or a part of answer marked correct and the rest as wrong, but no marks awarded.
14	While evaluating the answer books if the answer is found to be totally incorrect, it should be marked as cross (X) and awarded zero (0) Marks.
15	The Examiners should acquaint themselves with the guidelines given in the “Guidelines for Spot Evaluation” before starting the actual evaluation.
16	The candidates are entitled to obtain photocopy of the Answer Book on request on payment of the prescribed processing fee. All Examiners/Additional Head Examiners/Head Examiners are once again reminded that they must ensure that evaluation is carried out strictly as per value points for each answer as given in the Marking Scheme.
17	If a candidate attempts both alternatives/options in a question where only one option/ alternative is required to be attempted, the Evaluator shall award marks in both the options. The system will take the higher of two scores and disregard the other response.
18	In a question having two options/alternatives, if a candidate has attempted only one, then the evaluator shall mark “NA” (Not attempted) against the option that has not been attempted by the candidate.

MARKING SCHEME
Data Science (Subject Code-844)
(PAPER CODE : 368) (P3680844)

Q.No.	EXPECTED OUTCOMES/VALUE POINTS		Marks
	SECTION – A		
1.	Answer any 4 out of the given 6 questions on Employability Skills. रोजगार कौशल पर दिए गए 6 प्रश्नों में से किन्हीं 4 के उत्तर दीजिए।		
(i)	(C) 15	U-1 Pg-10	1
(ii)	Incentives or external rewards	U-2 Pg-24	1
(iii)	(A) Extroverts	U-2 Pg-33	1
(iv)	(B) Click the cell, type the data, and press enter.	U- 3 Pg-44	1
(v)	(C) Social	U-4 Pg-84	1
(vi)	Source of origin and type of waste	U-5 Pg-117	1
2.	Answer any 5 out of the given 6 questions. दिए गए 6 प्रश्नों में से किन्हीं 5 के उत्तर दीजिए।		(5x1=5)
(i)	Which data privacy legislation specifically deals with how websites and online companies collect data from children under 13 years of age ? (A) GDPR (B) HIPAA (C) CCPA (D) COPPA 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 ? (A) GDPR (B) HIPAA (C) CCPA (D) COPPA	Chapter-1 Page-3	1

(i)	Answer : (D) COPPA (1 mark for correct answer)		
(ii)	<p>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 ?</p> <p>(A) Histogram (B) Box Plot (C) Bar Chart (D) Pie Chart</p> <p>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 ?</p> <p>(A) हिस्टोग्राम (Histogram) (B) बॉक्स प्लॉट (Box Plot) (C) बार चार्ट (Bar Chart) (D) पाई चार्ट (Pie Chart)</p>	Chapter-2 Page-8	1
(ii)	Answer : (B) Box Plot (1 mark for correct answer)		
(iii)	<p>What is the primary function of each internal node in a Decision Tree ?</p> <p>(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.</p> <p>एक डिसीजन ट्री में प्रत्येक आंतरिक नोड (internal node) का प्राथमिक कार्य क्या है ?</p> <p>(A) यह अंतिम क्लास लेबल (class label) को hold करता 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</p>	Chapter-3 Page-13	1

	दर्शाता है।		
(iii)	<p>Answer : (C) It denotes a question on choosing a particular class.</p> <p>(1 mark for correct answer)</p>		
(iv)	<p>What is the common effect of choosing a big value of k in the k-NN algorithm ?</p> <p>(A) It increases the effect of outliers and makes the decision surface more specific.</p> <p>(B) It reduces the effect of outlier and causes the decision surface to generalize.</p> <p>(C) It increases the memory efficiency of the algorithm.</p> <p>(D) It always improves prediction accuracy regardless of the dataset.</p> <p>k- NN एल्गोरिथम (algorithm) में k का एक बड़ा मान (value) चुनने का सामान्य प्रभाव क्या होता है ?</p> <p>(A) यह आउटलायर (outliers) के प्रभाव को बढ़ाता है और निर्णय सतह (decision surface) को अधिक विशिष्ट (specific) बनाता है।</p> <p>(B) यह आउटलायर (outliers) के प्रभाव को कम करता है और निर्णय सतह को सामान्यीकृत (generalize) करने का कारण बनता है।</p> <p>(C) यह एल्गोरिथम की मेमोरी दक्षता (memory efficiency) को बढ़ाता है।</p> <p>(D) यह डेटासेट (dataset) की परवाह किए हमेशा prediction (पूर्वानुमान) सटीकता में सुधार करता है।</p>	Chapter-4 Page-23	1
(iv)	<p>Answer : (B) It reduces the effect of outlier and causes the decision surface to generalize.</p> <p>(1 mark for correct answer)</p>		
(v)	<p>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 _____.</p> 	Chapter-5 Page-34	1

	<p>(A) Line of Best Fit (B) Line of Worst Fit (C) Line of Median (D) Line of Mode</p> <p>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 के रूप में चिन्हीत रेखा (line) को _____ के रूप में जाना जाता है।</p> <p>(A) Line of Best Fit (B) Line of Worst Fit (C) Line of Median (D) Line of Mode</p>		
(v)	<p>Answer : (A) Line of Best Fit (1 mark for correct answer)</p>		
(vi)	<p>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 ?</p> <p>(A) Supervised Learning (B) Expectation-Maximization (C) Clustering (D) Anomaly detection</p> <p>संपूर्ण raw डेटा को कई समूहों में विभाजित करने की प्रक्रिया को क्या कहा जाता है ताकि 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 ?</p> <p>(A) सुपरवाइज्ड लर्निंग (Supervised Learning) (B) एक्सपेक्टेडेशन-मैक्सिमाइज़ेशन (Expectation-Maximization) (C) क्लस्टरिंग (Clustering) (D) विसंगति का पता लगाना (Anomaly detection)</p>	Chapter-7 Page-45	1
(vi)	<p>Answer : (C) Clustering (1 mark for correct answer)</p>		
3.	<p>Answer any 5 out of the given 6 questions. दिए गए 6 प्रश्नों में से किन्हीं 5 के उत्तर दीजिए।</p>		(5x1=5) (5x1=5)
(i)	<p>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</p>	Chapter-1 Page-2	1

	<p>guidelines ?</p> <p>(A) Security of Data</p> <p>(B) Privacy of Data</p> <p>(C) Non discrimination</p> <p>(D) Openness</p> <p>एक कॉलेज 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 स्वतः रूप से फ़िल्टर आउट (filter out) करने के लिए डिज़ाइन की गई है। यह स्वतः फ़िल्टरिंग प्रक्रिया किस नैतिक दिशानिर्देश (ethical guidelines) का उल्लंघन करती है ?</p> <p>(A) Security of Data</p> <p>(B) Privacy of Data</p> <p>(C) Non discrimination</p> <p>(D) Openness</p>		
(i)	<p>Answer : (C) Non discrimination</p> <p>(1 mark for correct answer)</p>		
(ii)	<p>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 ?</p> <p>(i) Leave the missing values as they are</p> <p>(ii) Remove the row of the variable that is missing</p> <p>(iii) Insert a value close to mean/mode of the variable that is missing</p> <p>(iv) Duplicate the missing entries</p> <p>(v) Replace the missing values with negative numbers</p> <p>एक डेटा विश्लेषण (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 ?</p> <p>(i) गुम मानों को वैसे ही छोड़ दें</p> <p>(ii) गुम चर की पंक्ति (row) को हटा दें</p> <p>(iii) गुम चर के माध्य (mean)/ बहुलक (mode) के करीब एक मान</p>	<p>Chapter-2</p> <p>Page-11</p>	<p>1</p>

	<p>insert करें</p> <p>(iv) गुम प्रविष्टियों की प्रतिलिपि (duplicate) बनाएँ</p> <p>(v) गुम मानों को ऋणात्मक संख्याओं (negative numbers) से बदल दें</p> <p>Options :</p> <p>(A) (i), (ii) and (iv)</p> <p>(B) (ii) and (iii)</p> <p>(C) (ii), (iii) and (v)</p> <p>(D) (iii) and (iv)</p>		
(ii)	<p>Answer : (B) (ii) and (iii)</p> <p>(1 mark for correct answer)</p>		
(iii)	<p>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 which it will be. When representing the uncertain weather outcome in her decision tree, what type of node should Sara use ?</p> <p>(A) Square node</p> <p>(B) Circular node</p> <p>(C) Triangular node</p> <p>(D) Diamond node</p> <p>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 वह निश्चित नहीं है कि क्या होगा। डिसीजन ट्री में uncertain weather outcome को दर्शाने के लिए, सारा को किस प्रकार के नोड (node) का उपयोग करना चाहिए ?</p> <p>(A) चौकोर नोड (Square node)</p> <p>(B) गोलाकार नोड (Circular node)</p> <p>(C) त्रिकोणीय नोड (Triangular node)</p> <p>(D) डायमंड नोड (Diamond node)</p>	<p>Chapter-3</p> <p>Page-16,17</p>	<p>1</p>
(iii)	<p>Answer : (B) Circular node</p> <p>(1 mark for correct answer)</p>		
(iv)	Which of the following describes the problem where the K-NN	Chapter-4	1

	<p>algorithm struggles to predict the output of a new data point as the number of input variables grows ?</p> <p>(A) Imbalanced Data Problem (B) No Training Step (C) Sensitivity to Outliers (D) Curse of Dimensionality</p> <p>निम्नलिखित में से कौन सी 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) करने में संघर्ष करता है ?</p> <p>(A) Imbalanced Data Problem (B) No Training Step (C) Sensitivity to Outliers (D) Curse of Dimensionality</p>	Page-25	
(iv)	<p>Answer : (D) Curse of Dimensionality (1 mark for correct answer)</p>		
(v)	<p>The Root Mean Square Deviation is the square root of :</p> <p>(A) Mean of residuals (B) Variance of residuals (C) Standard deviation (D) Mean absolute error</p> <p>रूट मीन स्क्वायर डेविएशन (Root Mean Square Deviation) किसका वर्गमूल (square root) है ?</p> <p>(A) Mean of residuals (B) Variance of residuals (C) Standard deviation (D) Mean absolute error</p>	Chapter-5 Page-35	1
(v)	<p>Answer : (B) Variance of residuals (1 mark for correct answer)</p>		
(vi)	<p>The approach that the K-means Clustering Algorithm follows to solve a given clustering problem is called :</p> <p>(A) Anomaly Detection (B) Expectation-Maximization (C) Decision Tree (D) Linear Regression</p> <p>वह approach जिसका K-मीन्स क्लस्ट्रिंग एल्गोरिथम (K-means Clustering Algorithm) एक दी गई क्लस्ट्रिंग समस्या को हल करने के लिए पालन करता है, उसे क्या कहा जाता है ?</p> <p>(A) विसंगति का पता लगाना (Anomaly Detection)</p>	Chapter-7 Page-46	1

	(B) एक्सपेक्टेडेशन-मैक्सिमाइज़ेशन (Expectation-Maximization) (C) डिसीजन ट्री (Decision Tree) (D) लीनियर रिग्रेशन (Linear Regression)		
(vi)	Answer : (B) Expectation-Maximization (1 mark for correct answer)		
4.	Answer any 5 out of the given 6 questions. दिए गए 6 प्रश्नों में से किन्हीं 5 के उत्तर दीजिए।		(5x1=5) (5x1=5)
(i)	On 11 th December, 2019, which ministry tabled the personal Data protection Bill, 2019 , in the Indian parliament ? (A) Ministry of Health and Family Welfare (B) Ministry of Law and Justice (C) Ministry of Finance (D) Ministry of Electronics and Information Technology 11 दिसंबर, 2019, को किस मंत्रालय (ministry) ने भारतीय संसद में personal Data protection Bill, 2019 , पेश किया ? (A) स्वास्थ्य और परिवार कल्याण मंत्रालय (B) कानून और न्याय मंत्रालय (C) वित्त मंत्रालय (D) इलेक्ट्रॉनिक्स और सूचना प्रौद्योगिकी मंत्रालय	Chapter-1 Page-4	1
(i)	Answer : (D) Ministry of Electronics and Information Technology (1 mark for correct answer)		
(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 ? (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. 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 ? (A) एक बार डेटा दर्ज हो जाने के बाद, इसे डेटा विश्लेषण के किसी भी चरण के दौरान साफ (clean) नहीं किया जा सकता है।	Chapter-2 Page-10	1

	<p>(B) मॉडल बनाने के लिए डेटा का उपयोग करते समय यह समस्याएँ उत्पन्न करेगा।</p> <p>(C) यह बिना किसी समस्या के सभी गणितीय कार्यों (mathematical operations) को होने देगा।</p> <p>(D) इसका किसी भी प्रकार के विश्लेषण पर कोई प्रभाव नहीं पड़ेगा।</p>		
(ii)	<p>Answer : (B) It will cause problems while using the data to build a model.</p> <p>(1 mark for correct answer)</p>		
(iii)	<p>Why is K-NN considered as non-parametric algorithm ?</p> <p>(A) It assumes a specific distribution for the data before classification.</p> <p>(B) It requires learning parameters for the data distribution.</p> <p>(C) It does not assume anything about the distribution of the data.</p> <p>(D) It uses a fixed number of parameters to model the data.</p> <p>K-NN को नॉन-पैरामीट्रिक एल्गोरिथम (non-parametric algorithm) क्यों माना जाता है ?</p> <p>(A) यह वर्गीकरण (classification) से पहले डेटा के लिए एक विशिष्ट वितरण (specific distribution) मानता है।</p> <p>(B) इसे डेटा वितरण के लिए लर्निंग पैरामीटर (learning parameters) सीखने की आवश्यकता होती है।</p> <p>(C) यह डेटा के वितरण के बारे में कुछ भी नहीं मानता है।</p> <p>(D) यह डेटा को मॉडल करने के लिए पैरामीटर (parameters) की एक निश्चित संख्या का उपयोग करता है।</p>	Chapter-4 Page-23	1
(iii)	<p>Answer : (C) It does not assume anything about the distribution of the data.</p> <p>(1 mark for correct answer)</p>		
(iv)	<p>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 ?</p> <p>(A) Remove the dependent variable from the model.</p> <p>(B) Change the independent variable to the dependent variable (y).</p> <p>(C) Retrain the model.</p> <p>(D) Switch from linear regression to a classification algorithm.</p> <p>स्नेहा लीनियर रिग्रेशन (linear regression) कर रही है और यह निर्धारित करने के लिए रूट मीन स्क्वायर डेविएशन (Root Mean</p>	Chapter-5 Page-35	1

	<p>Square Deviation) विधि का उपयोग कर रही है कि observed points मॉडल के अनुमानित मानों (predicted values) के कितने करीब हैं। यदि एक बड़ा विचलन मान (deviation value) प्राप्त होता है तो recommended action क्या है ?</p> <p>(A) मॉडल से dependent variable (y) को हटा दें।</p> <p>(B) independent variable (x) को dependent variable (y) में बदल दें।</p> <p>(C) मॉडल को पुनः प्रशिक्षित करें (Retrain the model)।</p> <p>(D) लीनियर रिग्रेशन से वर्गीकरण एल्गोरिथम (classification algorithm) पर स्विच करें।</p>		
(iv)	<p>Answer : (C) Retrain the model.</p> <p>(1 mark for correct answer)</p>		
(v)	<p>Statement (1) : Logarithmic functions are examples of linear function.</p> <p>Statement (2) : The graph of non-linear regression follows the equation of a curve.</p> <p>(A) Both Statement (1) and Statement (2) are correct</p> <p>(B) Both Statement (1) and Statement (2) are incorrect</p> <p>(C) Statement (1) is correct but Statement (2) is incorrect</p> <p>(D) Statement (1) is correct but Statement (2) is incorrect</p> <p>कथन (1) : Logarithmic functions linear function के उदाहरण हैं।</p> <p>कथन (2) : नॉन-लीनियर रिग्रेशन (non-linear regression) का ग्राफ एक वक्र (curve) के समीकरण को follow करता है।</p> <p>(A) कथन (1) और कथन (2) दोनों सही हैं</p> <p>(B) कथन (1) और कथन (2) दोनों गलत हैं</p> <p>(C) कथन (1) सही है लेकिन कथन (2) गलत है</p> <p>(D) कथन (2) सही है लेकिन कथन (1) गलत है</p>	Chapter-6 Page-41	1
(v)	<p>Answer :(D) Statement (2) is correct but Statement (1) is incorrect</p> <p>(1 mark for correct answer)</p>		
(vi)	<p>K-means clustering is a technique used to spot clusters of data classes in a dataset and is classified as_____.</p> <p>(A) A supervised learning technique</p> <p>(B) A classification technique</p>	Chapter-7 Page-45	1

	<p>(C) An unsupervised machine learning technique</p> <p>(D) A regression technique</p> <p>K-मीन्स क्लस्टरिंग (K-means clustering) एक ऐसी तकनीक है जिसका उपयोग एक डेटासेट में डेटा क्लास (data classes) के क्लस्टर (समूह) का पता लगाने के लिए किया जाता है और इसे इस रूप में वर्गीकृत किया जाता है _____।</p> <p>(A) एक सुपरवाइज्ड लर्निंग तकनीक (A supervised learning technique)</p> <p>(B) एक वर्गीकरण तकनीक (A classification technique)</p> <p>(C) एक अनसुपरवाइज्ड मशीन लर्निंग तकनीक (An unsupervised machine learning technique)</p> <p>(D) एक रिग्रेशन तकनीक (A regression technique)</p>		
(vi)	<p>Answer : (C) An unsupervised machine learning technique</p> <p>(1 mark for correct answer)</p>		
5.	<p>Answer any 5 out of the given 6 questions.</p> <p>दिए गए 6 प्रश्नों में से किन्हीं 5 के उत्तर दीजिए।</p>		<p>(5x1=5)</p> <p>(5x1=5)</p>
(i)	<p>Which of the following is an important aspect of GDPR legislation ?</p> <p>(A) Data multiplication</p> <p>(B) Right to be forgotten</p> <p>(C) Mandatory data sharing</p> <p>(D) Unlimited data storage</p> <p>निम्नलिखित में से कौन सा GDPR कानून का एक महत्वपूर्ण पहलू (aspect) हैं ?</p> <p>(A) Data multiplication</p> <p>(B) Right to be forgotten</p> <p>(C) Mandatory data sharing</p> <p>(D) Unlimited data storage</p>	Chapter-1 Page-3	1
(i)	<p>Answer : (B) Right to be forgotten</p> <p>(1 mark for correct answer)</p>		
(ii)	<p>Statement-1 : Regression trees are used when the dependent variable is categorical.</p> <p>Statement-2 : A node can have two or more branches in a decision tree.</p> <p>(A) Both Statement (1) and Statement (2) are correct</p> <p>(B) Both Statement (1) and Statement (2) are incorrect</p> <p>(C) Statement (1) is correct but Statement (2) is incorrect</p> <p>(D) Statement (2) is correct but Statement (1) is incorrect</p>	Chapter-3 Page-16	1

	<p>कथन (1) : रिग्रेशन ट्री (Regression trees) का उपयोग तब किया जाता है जब डिपेंडेंट वेरिएबल (dependent variable) श्रेणीबद्ध (categorical) होता है।</p> <p>कथन (2) : एक डिसीजन ट्री में एक नोड में दो या दो से अधिक शाखाएँ (branches) हो सकती हैं।</p> <p>(A) कथन (1) और कथन (2) दोनों सही हैं</p> <p>(B) कथन (1) और कथन (2) दोनों गलत हैं</p> <p>(C) कथन (1) सही है लेकिन कथन (2) गलत है</p> <p>(D) कथन (2) सही है लेकिन कथन (1) गलत है</p>		
(ii)	<p>Answer : (D) Statement (2) is correct but Statement (1) is incorrect</p> <p>(1 mark for correct answer)</p>		
(iii)	<p>Which of the following is correct for the K-NN algorithm ?</p> <p>(A) It is highly accurate with imbalanced data.</p> <p>(B) It is immune to the curse of dimensionality.</p> <p>(C) It is slow and memory-inefficient.</p> <p>(D) It is robust to outliers.</p> <p>K-NN एल्गोरिथम के लिए निम्नलिखित में से कौन सा सही है ?</p> <p>(A) यह असंतुलित डेटा (imbalanced data) के साथ अत्यधिक सटीक (accurate) है।</p> <p>(B) यह curse of dimensionality से प्रतिरक्षित (immune) है।</p> <p>(C) यह धीमा और मेमोरी-अकुशल (memory-inefficient) है।</p> <p>(D) यह आउटलायर के लिए robust है।</p>	Chapter-4 Page-24	1
(iii)	<p>Answer : (C) It is slow and memory-inefficient.</p> <p>(1 mark for correct answer)</p>		
(iv)	<p>The equation for a simple linear regression is of the form $y = m \cdot x + b$. What do 'm' and 'b' represent in the equation.</p> <p>(A) Slope, Independent Variable</p> <p>(B) Independent Variable, Intercept</p> <p>(C) Slope, Intercept</p> <p>(D) Dependent Variable, Slope</p> <p>एक सरल लीनियर रिग्रेशन (simple linear regression) का समीकरण $y = m \cdot x + b$ के रूप का होता है। समीकरण में 'm' और 'b' क्या दर्शाते हैं ?</p> <p>(A) Slope इंडिपेंडेंट वेरिएबल (Independent Variable)</p> <p>(B) इंडिपेंडेंट वेरिएबल (Independent Variable) इंटरसेप्ट</p>	Chapter-5 Page-34	1

	<p>(Intercept)</p> <p>(C) Slope इंटरसेप्ट (Intercept)</p> <p>(D) डिपेंडेंट वेरिएबल (Dependent Variable), Slope</p>		
(iv)	<p>Answer : (C) Slope, Intercept</p> <p>(1 mark for correct answer)</p>		
(v)	<p>A simple regression equation has an intercept on the right-hand side and an explanatory variable with a/an _____.</p> <p>(A) Coefficient</p> <p>(B) Dependent variable</p> <p>(C) intercept</p> <p>(D) Linear regression</p> <p>एक सरल रिग्रेशन समीकरण के दाहिने हाथ की ओर एक इंटरसेप्ट (intercept) और explanatory variable के साथ एक _____ होता है।</p> <p>(A) Coefficient</p> <p>(B) डिपेंडेंट वेरिएबल (Dependent variable)</p> <p>(C) इंटरसेप्ट (intercept)</p> <p>(D) लीनियर रिग्रेशन (Linear regression)</p>	Chapter-6 Page-40	1
(v)	<p>Answer : (A) Coefficient</p> <p>(1 mark for correct answer)</p>		
(vi)	<p>The K-means Clustering algorithm it first specifies the number of clusters, K depending on the input. Thereafter how does it initialize the centroids ?</p> <p>(A) By Assigning all data points to any one cluster.</p> <p>(B) By Computing the sum of the squared distance and assigning to the K clusters.</p> <p>(C) By Shuffling the data points and then selecting K data points randomly.</p> <p>(D) By Using only the labelled data for clusters, K.</p> <p>K-मीन्स क्लस्टरिंग एल्गोरिथम के मामले में, यह सबसे पहले इनपुट के आधार पर क्लस्टर्स (clusters) की संख्या K को निर्दिष्ट (specifies) करता है। इसके बाद यह सेंट्रोइड्स (centroids) को कैसे (initialize) करता है ?</p> <p>(A) सभी डेटा पॉइंट (data points) को किसी एक क्लस्टर को assign करना।</p> <p>(B) वर्ग दूरी (squared distance) के योग की गणना करना और K क्लस्टर्स को assign करना।</p> <p>(C) डेटा पॉइंट को शफल (Shuffling) करना और फिर randomly K डेटा पॉइंट का चयन करना।</p>	Chapter-7 Page-46	1

	(D) क्लस्टर, K के लिए केवल लेबल किए गए डेटा (labelled data) का उपयोग करना।		
(vi)	Answer : (C) By Shuffling the data points and then selecting k data points randomly. (1 mark for correct answer)		
	SECTION – B खण्ड - ख		30 marks
	Answer any 3 out of the given 5 questions (Question No 6 to 10) on Employability Skills in 20-30 words each. रोजगार कौशल पर दिए गए 5 प्रश्नों (प्रश्न संख्या 6 से 10) में से किन्हीं 3 प्रश्नों के उत्तर 20-30 शब्दों में दीजिए।		(3x2=6) (3x2=6)
6.	The four skills are :- Listening Speaking Reading Writing	U-1 Pg-1	2 (0.5x4)
7.	Motivation is derived from the word 'motive'. Thus directing behaviour towards certain motive or goal is the essence of motivation. An individual's motivation may come from within or be inspired by others or events.	U-2 Pg-24	2 (1x2=2)
8.	A digital presentation can be saved as a file on the computer. This can be opened later, viewed, edited shared with friends and colleagues and printed.	U-3 Pg-66	2 (1+1)
9.	Because (i) It requires specific progression and procedures to be followed (ii) It also requires skills to digress when required and yet make the entire activity profitable and growth oriented	U-4 Pg-80	2
10.	Appropriate technology is a small scale technology that is environment friendly and suited to local needs. Examples – (i) Bike powered or hand powered water pumps (ii) Solar lamps in streetlight (iii) Solar buildings etc	U-5 Pg-117	2 (1+1)
	Answer any 4 out of the given 6 questions (Question No 11 to 16) in 20 – 30 words each. दिए गए 6 प्रश्नों (प्रश्न संख्या 11 से 16) में से किन्हीं 4 प्रश्नों के उत्तर		(4x2=8) (4x2=8)

	20-30 शब्दों में दीजिए।		
11.	<p>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 ?</p> <p>डेटा गोपनीयता (data privacy) क्या है ? समझाएँ कि यदि कोई कंपनी व्यक्तिगत पहचान योग्य जानकारी (personal identifiable information) को एन्क्रिप्टेड (encrypted) फॉर्मेट में सुरक्षित रूप से संग्रहीत करती है, तब भी डेटा गोपनीयता नियमों का उल्लंघन क्यों हो सकता है ?</p> <p>Answer :</p> <p>Data privacy is the right of any individual to have control over how his/her personal information is collected and used.</p> <p>A violation can still occur even if data is stored securely in an encrypted format because data privacy is not just about secure data storage; if there is no agreement from the users regarding the collection of the data itself, it is violation of data privacy rules.</p> <p>(1 mark for each for each correct part)</p>	Chapter-1 Page-2	2
12.	<p>Raunak owns a small cafe 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.</p> <p>(a) What type of analysis should he use to study this correlation ?</p> <p>(b) Name any one graphical method he can use to visualize this data.</p> <p>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।</p> <p>(a) इस सहसंबंध का अध्ययन करने के लिए उसे किस प्रकार के विश्लेषण (analysis) का उपयोग करना चाहिए ?</p> <p>(b) इस डेटा को विजुअलाइज़ (visualize) करने के लिए वह जिस किसी एक ग्राफिकल विधि (graphical method) का उपयोग कर सकता है, उसका नाम बताएं।</p> <p>Answer :</p> <p>(a) Raunak should use bivariate analysis to study the</p>	Chapter-2 Page-9	2

	<p>relationship between temperature and cold drink sales.</p> <p>(b) He can use a scatter plot (or line chart/pair plot) to visualize the relationship between the two variables.</p> <p>(1 mark for each for each correct part)</p>		
13.	<p>Name and explain any one popular method of cross-validation.</p> <p>क्रॉस वैलिडेशन (cross-validation) की किसी एक लोकप्रिय विधि का नाम बताइए और उसकी व्याख्या करें।</p> <p>Answer :</p> <p>One popular method to do cross validation is to use k-fold cross validation technique.</p> <p>In this method, we split the data into k different but similar folds or sets. We then perform k iterations and in each iteration, we choose one fold as the validation set or test set, and the rest as training sets.</p> <p>(1 mark for correctly naming the technique 1 mark for brief explanation)</p>	Chapter-4 Page-25	2
14.	<p>Why are decision trees considered versatile ? Mention any two points.</p> <p>डिसीजन ट्री (decision Trees) को versatile क्यों माना जाता है? किन्हीं दो बिंदुओं का उल्लेख करें।</p> <p>Answer :</p> <ol style="list-style-type: none"> 1. They can be used for any kind of problem, whether it is classification or regression. 2. Unlike linear models, decision trees can map both linear and non-linear relationships quite well. <p>(1 mark for mentioning each point)</p>	Chapter-3 Page-15	2
15.	<p>What is Mean Absolute Error (MAE) ? Also explain the process of 'Fitting the line to the data' that MAE.</p> <p>Mean Absolute Error (MAE) क्या है ? MAE का उपयोग करके 'डेटा पर लाइन को फिट करने' (Fitting the line to the data) की प्रक्रिया की भी व्याख्या करें।</p> <p>Answer :</p> <p>Mean Absolute Error (MAE) measures the average magnitude of the errors in predictions without considering their direction.</p> <p>The basic objective of linear regression is to try to reduce the vertical distance between the line of best fit and the data points to make it minimum. This process is called Fitting the line to the data.</p> <p>(1 mark for each for each correct part)</p>	Chapter-5 Page-34	2
16.	<p>Mention any two differences between Supervised Learning and Unsupervised Learning.</p> <p>सुपरवाइज्ड लर्निंग (Supervised Learning) और अनसुपरवाइज्ड लर्निंग (Unsupervised Learning) के बीच किन्हीं दो अंतरों का</p>	Chapter-7 Page-43,44	2

	<p>उल्लेख करें।</p> <p>Answer :</p> <table><tr><th>Supervised Learning</th><th>Unsupervised Learning</th></tr><tr><td>Algorithms are trained using data that is well-labelled or classified</td><td>Algorithms are not trained using data that is classified or labelled.</td></tr><tr><td>Uses the trained algorithm to make predictions on new sets of data for classification or regression problems</td><td>Algorithm discover hidden patterns or data groupings, thereby acting on data without human intervention.</td></tr><tr><td>Ideal for classification and regression problems.</td><td>Ideal for solving real-life problems such as exploratory data analysis, customer segmentation, and image recognition</td></tr></table> <p>(1 mark for each correct difference) (Any two)</p>	Supervised Learning	Unsupervised Learning	Algorithms are trained using data that is well-labelled or classified	Algorithms are not trained using data that is classified or labelled.	Uses the trained algorithm to make predictions on new sets of data for classification or regression problems	Algorithm discover hidden patterns or data groupings, thereby acting on data without human intervention.	Ideal for classification and regression problems.	Ideal for solving real-life problems such as exploratory data analysis, customer segmentation, and image recognition		
Supervised Learning	Unsupervised Learning										
Algorithms are trained using data that is well-labelled or classified	Algorithms are not trained using data that is classified or labelled.										
Uses the trained algorithm to make predictions on new sets of data for classification or regression problems	Algorithm discover hidden patterns or data groupings, thereby acting on data without human intervention.										
Ideal for classification and regression problems.	Ideal for solving real-life problems such as exploratory data analysis, customer segmentation, and image recognition										
	<p>Answer any 3 out of the given 5 questions (Question No 17 to 21) in 50 – 80 words each.</p> <p>दिए गए 5 प्रश्नों (प्रश्न संख्या 17 से 21) में से किन्हीं 3 प्रश्नों के उत्तर 50-80 शब्दों में दीजिए।</p>		<p>(3x4=12)</p> <p>(3x4=12)</p>								
17.	<p>Answer the following questions about HIPAA legislation.</p> <p>i. Expand the abbreviation HIPAA.</p> <p>ii. State any two objectives of HIPAA.</p> <p>iii. Give any two examples personal identifiers that HIPAA protects.</p> <p>HIPAA कानून (legislation) के बारे में निम्नलिखित प्रश्नों के उत्तर दें।</p> <p>i. HIPAA संक्षिप्त नाम (abbreviation) को expand करें।</p> <p>ii. HIPAA के किन्हीं दो उद्देश्यों (objectives) को बताएं।</p> <p>iii. personal identifiers के दो उदाहरण दें जिनकी HIPAA सुरक्षा करता है।</p> <p>Answer :</p> <p>i. Health Insurance Portability and Accountability Act (1 mark for correct expansion)</p> <p>ii.</p> <ul style="list-style-type: none">• To protect healthcare information from fraud and theft• Helps to manage personally identifiable information stored by healthcare and insurance companies.• Returns control of data to the individuals (1 mark for each correct objective) (Any two) <p>iii. Name, Phone numbers, email addresses, geographical identifiers, fingerprints or retinal prints, social security</p>	Chapter-1 Page-3	1+2+1=4								

	numbers, medical records (any two) (½ mark for each correct identifier)		
18.	<p>Explain the concept of Univariate Analysis. Describe its main purpose and list two different statistical and two different graphical methods associated with it.</p> <p>यूनिवैरिएट 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।</p> <p>Answer :</p> <p>Concept : only one variable analysed, no worry about causes or relationships</p> <p>Purpose : Describe the data and find patterns that exist within it</p> <p>Statistical Methods : (Mean, mode, median, range, variance, maximum, minimum, quartiles, standard deviation – any two)</p> <p>Graphical Methods : Frequency distribution tables, bar charts, histograms, frequency polygons, pie charts - any two)</p> <p>(1 mark for correct concept) (1 mark for correct purpose) (2 mark for correct methods)</p>	Chapter-2 Page-8	4
19.	<p>What is a decision tree ? Explain the steps involved in creating a decision tree.</p> <p>डिसीजन ट्री (decision tree) क्या है ? डिसीजन ट्री बनाने में शामिल चरणों की व्याख्या करें।</p> <p>Answer :</p> <p>A decision tree is a diagrammatic representation of a decision-making process with a tree-like structure. Each internal node represents a question, branches show possible outcomes, and leaf nodes indicate class labels or results.</p> <p>Steps for creating decision tree</p> <ul style="list-style-type: none"> • Define the main objective and place it at the root of the tree • Draw branches for each possible decision and add leaf nodes to show results. • Calculate the probability of success for each decision using data or past projects. <p>(1 mark for correct definition) (1 mark for each correct step)</p>	Chapter-3 Page-13,16,17	4
20.	(a) A health researcher wants to predict a patient's blood	Chapter-6	4

	<p>pressure based on four factors : age, weight, cholesterol level, and daily exercise duration .</p> <p>Identify and define the most suitable regression model to analyse the relationship. Explain why this model is appropriate for this scenario.</p> <p>(b) Rudhra 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 ?</p> <p>(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 ?</p> <p>(b) रुद्र ने छह महीने पहले अपना व्यवसाय शुरू किया था। उसकी आय में हर महीने exponential (5 गुना) वृद्धि हुई है। सातवें महीने में उसकी आय को (predict) करने के लिए किस प्रकार के रिग्रेशन (regression) का उपयोग किया जा सकता है ?</p> <p>Answer :</p> <p>(a) Multiple Linear Regression Model. It is a statistical method that models the relationship between one dependent variable and two or more independent variables by fitting a linear equation to the observed data. It is appropriate for this scenario as there is one dependent variable (blood pressure) and multiple independent variables (age, weight, cholesterol level, and daily exercise duration)</p> <p>(1 mark for identification and 1 mark for valid definition.)</p> <p>(1 mark for suitable explanation)</p> <p>(b) Non-Linear Regression</p> <p>(1 mark for correct answer)</p>	Page-40-42	
21.	<p>Explain the role of Unsupervised Learning in the following real-world applications :</p> <p>i. Recommendation engines</p> <p>ii. Anomaly detection</p> <p>iii. Medical images</p>	Chapter-7 Page-44	4

	<p>iv. News categorization निम्नलिखित वास्तविक-दुनिया के अनुप्रयोगों (real-world applications) में अनसुपरवाइज्ड लर्निंग (Unsupervised Learning) की भूमिका (role) की व्याख्या करें।</p> <p>i. रेकमेंडेशन इंजन (Recommendation engines)</p> <p>ii. अनोमली डिटेक्शन (Anomaly detection)</p> <p>iii. मेडिकल इमेजिंग (Medical imaging)</p> <p>iv. न्यूज़ वर्गीकरण (News categorization)</p> <p>Answer :</p> <p>i. Recommendation Engines : Helps predict products a customer is likely to buy by analysing past purchase behaviour and discovering trends.</p> <p>ii. Medical imaging : Assists in image detection, classification, and segmentation for faster and accurate diagnosis.</p> <p>iii. Anomaly detection : Identifies unusual data points in large datasets to detect fraud or security breaches.</p> <p>iv. News categorization : Categorizes news articles into sections like sports, entertainment, and international news.</p> <p>(1 mark for each correct part)</p>		
	- o O o -		