Roll No:-_____

(Sem-5_OBE) Diploma Exam 2025 (Odd)

[Time: 3 Hours] **(Computer Science) (Theory)** **[Full. Marks: 70]**

Data Science: Data Warehousing and Data Mining

(Computer Science)

(T2418503A)

- All questions are compulsory. (सभी प्रश्न अनिवार्य है।)

- Marks are mentioned on the right side of each question. (अंक सभी प्रश्न के दाई ओर अंकित किये है।)

Group (A) (ग्रुप -ए)

Q.1 Answer all questions as directed. (2x10=20)
(निर्देशानुसार सभी प्रश्नों के उत्तर दें)

	Marks	CO	BL
a) Fill in the Blanks: _____ is used to study relationships between dependent and independent variables. (Linear Regression / Logistic Regression / Delphi / Histogram) _____ का उपयोग आश्रित और स्वतंत्र चरों के बीच संबंधों का अध्ययन करने के लिए किया जाता है। (रैखिक प्रतिगमन / लॉजिस्टिक प्रतिगमन / डेल्फी / हिस्टोग्राम)	2	-	1
b) Identify the odd one out: Descriptive Analysis, Predictive Analysis, Exploratory Analysis, Operating System असंगत चुनिए: वर्णनात्मक विश्लेषण, भविष्यसूचक विश्लेषण, खोजपरक विश्लेषण, ऑपरेटिंग सिस्टम	2	-	1
c) Fill in the Blanks: _____ operation in OLAP is used to view data at increasing levels of aggregation. (Roll-up / Slice / Dice / Drill-down) OLAP में _____ ऑपरेशन का उपयोग बढ़ते स्तरों पर डेटा देखने के लिए किया जाता है। (रोल-अप / स्लाइस / डाइस / ड्रिल-डाउन)	2	-	1
d) OLTP systems are mainly used for analytical processing. (True/False) OLTP प्रणालियाँ मुख्य रूप से विश्लेषणात्मक प्रोसेसिंग के लिए उपयोग की जाती हैं। (सही / गलत)	2	-	1
e) Which of the following is a supervised learning algorithm? a) K-Means b) Decision Tree c) DBSCAN d) Apriori निम्नलिखित में से कौन-सा सुपरवाइज्ड लर्निंग एल्गोरिद्म है? a) K-मीन्स b) निर्णय वृक्ष c) DBSCAN d) एप्रियोरी	2	-	1
f) Match the Following: Column A — Column B 1. Bayesian Classification — a. Clustering algorithm 2. Decision Tree — b. Uses tree-like structure	2	-	2

	3. SVM c. Separates data using hyperplane 4. K-Means d. Uses probability theory मिलान कीजिए: स्तंभ A स्तंभ B 1. बेयसियन वर्गीकरण a. क्लस्टरिंग एल्गोरिद्म है 2. निर्णय वृक्ष b. पेड़ जैसी संरचना का उपयोग करता है 3. SVM c. हाइपरप्लेन का उपयोग करके डेटा को अलग करता है 4. K-मीन्स d. प्रायिकता सिद्धांत का उपयोग करता है			
g)	_____ algorithm is based on density of data points. (DBSCAN / K-Means / PageRank / SVM) _____ एल्गोरिद्म डेटा बिंदुओं की घनत्व पर आधारित है। (DBSCAN / K-मीन्स / पेजरैंक / SVM)	2	-	1
h)	Match the Following: Column A Column B 1. K-Means a. Centroid-based clustering 2. DBSCAN b. Bottom-up approach 3. Agglomerative c. Density-based clustering 4. Divisive d. Top-down approach मिलान कीजिए: स्तंभ A स्तंभ B 1. K-मीन्स a. सेंट्रॉइड-आधारित क्लस्टरिंग 2. DBSCAN b. बॉटम-अप दृष्टिकोण 3. एग्लोमरेटिव c. घनत्व-आधारित क्लस्टरिंग 4. डिवाइसिव d. टॉप-डाउन दृष्टिकोण	2	-	2
i)	Which of the following is a partition-based clustering technique? a) K-Means b) DBSCAN c) Agglomerative d) Divisive निम्नलिखित में से कौन-सी विभाजन-आधारित क्लस्टरिंग तकनीक है? a) K-मीन्स b) DBSCAN c) एग्लोमरेटिव d) डिवाइसिव	2	-	1
j)	Overfitting occurs when the model performs well on training data but poorly on unseen data.(True / False) ओवरफिटिंग तब होती है जब मॉडल प्रशिक्षण डेटा पर अच्छा प्रदर्शन करता है लेकिन नए डेटा पर खराब करता है।(सही / गलत)	2	-	1
	Group (B) (ग्रुप -बी) **Answer all five questions. (सभी पाँच प्रश्नों के उत्तर दें|)** **4x5=20**			
Q.2	Compare descriptive and inferential analysis in Data Science. डेटा विज्ञान में वर्णनात्मक और अनुमेय विश्लेषण की तुलना कीजिए।	4	1	2
	OR (अथवा)			
	Distinguish between linear regression and logistic regression. रैखिक प्रतिगमन और लॉजिस्टिक प्रतिगमन में अंतर बताइए।	4	1	2
Q.3	Compare the different data warehouse models. विभिन्न डेटा वेयरहाउस मॉडलों की तुलना कीजिए।	4	2	2
	OR (अथवा)			

	Compare OLAP and OLTP systems. OLAP और OLTP प्रणालियों की तुलना कीजिए।	4	2	2
Q.4	Explain the concept and importance of Data Mining. डेटा माइनिंग की अवधारणा और महत्व को समझाइए। **OR (अथवा)**	4	3	1
	Explain the various data preprocessing methods. विभिन्न डेटा पूर्व-संसाधन विधियों को समझाइए।	4	3	1
Q.5	Discuss different approaches to solve classification problems. वर्गीकरण समस्याओं को हल करने के विभिन्न दृष्टिकोणों पर चर्चा कीजिए। **OR (अथवा)**	4	4	1
	Explain the working of Bayesian classification. बेयसियन वर्गीकरण के कार्य को समझाइए।	4	4	1
Q.6	Discuss the basic requirements for performing cluster analysis. क्लस्टर विश्लेषण करने की मूल आवश्यकताओं पर चर्चा कीजिए। **OR (अथवा)**	4	5	1
	Discuss distance measures used in clustering algorithms. क्लस्टरिंग एल्गोरिद्म में प्रयुक्त दूरी मापदंडों पर चर्चा कीजिए।	4	5	1
	Group (C) (ग्रुप - सी) **Answer all five questions. (सभी पाँच प्रश्नों के उत्तर दें।)** **6x5=30**	6	1	1
Q.7	Explain various forecasting techniques used in Data Science. डेटा विज्ञान में उपयोग की जाने वाली विभिन्न पूर्वानुमान तकनीकों को समझाइए। **OR (अथवा)**	6	1	1
	Discuss the need and process of hypothesis testing परिकल्पना परीक्षण की आवश्यकता और प्रक्रिया पर चर्चा	6	1	1
Q.8	Explain the concept and importance of a data warehouse. Discuss the need for a separate data warehouse system. डेटा वेयरहाउस की अवधारणा और महत्व को समझाइए।एक अलग डेटा वेयरहाउस सिस्टम की आवश्यकता पर चर्चा कीजिए। **OR (अथवा)**	6	2	1
	Discuss the relationship between data mining and OLAP. डेटा माइनिंग और OLAP के बीच संबंध पर चर्चा कीजिए।	6	2	1
Q.9	Explain the integration of data mining systems with databases. डेटाबेस के साथ डेटा माइनिंग प्रणालियों के एकीकरण को समझाइए। **OR (अथवा)**	6	3	1
	Discuss predictive modeling and its applications. भविष्यसूचक मॉडलिंग और इसके अनुप्रयोगों पर चर्चा कीजिए।	6	3	1
Q.10	Compare dimensionality reduction and pattern-based classification. विमीयता न्यूनीकरण और पैटर्न-आधारित वर्गीकरण की तुलना कीजिए। **OR (अथवा)**	6	4	2
	Distinguish between binary and multi-class classification. द्विआधारी और बहु-वर्ग वर्गीकरण में अंतर बताइए।	6	4	2
Q.11	Write short notes on SVM and Decision Tree. SVM और निर्णय वृक्ष पर संक्षिप्त नोट लिखिए। **OR (अथवा)**	6	4	3
	Write short notes on PageRank and HITS algorithms. पेजरैंक और HITS एल्गोरिद्म पर संक्षिप्त नोट लिखिए।	6	5	3