Roll No:-_____

(Sem - IV) Diploma Exam 2024 (Even)

[Time: 3 Hours] **(ARTIFICIAL INTELLIGENCE(AI) AND MACHINE LEARNING) (Theory)** **[Full. Marks: 70]**

Introduction to Machine Learning (2044403)

- 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)

		Marks	CO	BL
Q.1	**Answer all questions as directed. (निर्देशानुसार सभी प्रश्नों के उत्तर दें)** **(2x10=20)**			
a)	Match the type of dataset with its example:	2	1	1

1.Text Dataset	a. Financial stock prices
2. Image Dataset	b. Articles and news
3. Time Series Dataset	c. Photos and video

डेटा सेट के प्रकार को उसके उदाहरण से मिलाएं:

1. टेक्स्ट डेटा सेट	a. वित्तीय स्टॉक की कीमतें
2. इमेज डेटा सेट	b. लेख और समाचार
3. समय श्रृंखला डेटा सेट	c. तस्वीरें और वीडियो

b)	Match the preprocessing technique with its description:	2	1	2

1. Data Cleaning	a. Converting categorical data into numerical form
2. Feature Encoding	b. Removing noise and inconsistencies from data
3. Dimensionality Reduction	c. Reducing the number of features while preserving information

प्रीप्रोसेसिंग तकनीक को उसके विवरण से मिलाएं:

1. डेटा सफाई	a. श्रेणीय डेटा को संख्यात्मक रूप में बदलना
2. फीचर एन्कोडिंग	b. डेटा से शोर और असंगतियों को निकालना
3. आयाम न्यूनीकरण	c. जानकारी को संरक्षित करते हुए विशेषताओं की संख्या को कम करना

c)	Supervised learning requires labelled data for training. (True/False) पर्यवेक्षित लर्निंग को प्रशिक्षण के लिए लेबल डेटा की आवश्यकता होती है।(सही /गलत)	2	2	2

d) PCA (Principal Component Analysis) is a type of clustering technique. (True/False) PCA (मुख्य घटक विश्लेषण) एक प्रकार की क्लस्टरिंग तकनीक है।(सही /गलत)	2	5	2
e) Cross-validation helps in assessing the performance of a machine learning model. (True/False) क्रॉस-वैलिडेशन मशीन लर्निंग मॉडल के प्रदर्शन का आकलन करने में मदद करता है।(सही /गलत)	2	1	2
f) In supervised learning, the main goal is to __ the relationship between input and output variables. (discover/predict/ignore) पर्यवेक्षित लर्निंग में, मुख्य लक्ष्य __ के बीच के संबंध को पहचानना है। (खोजना/भविष्यवाणी करना/अनदेखा करना)	2	2	2
g) Data __ involves reducing the number of features in a dataset. (cleaning/normalization/dimensionality reduction) डेटा __ में डेटासेट में विशेषताओं की संख्या को कम करना शामिल है। (सफाई/सामान्यीकरण/आयाम न्यूनीकरण)	2	5	3
h) The __ dataset is used to evaluate the performance of a machine learning model after training. (training/validation/test) __ डेटा सेट का उपयोग प्रशिक्षण के बाद मशीन लर्निंग मॉडल के प्रदर्शन का मूल्यांकन करने के लिए किया जाता है। (प्रशिक्षण/मान्यता/परीक्षण)	2	2	3
i) .....clustering is a method where clusters are formed from individual data points. (Divisive/Agglomerative/Hierarchical) ……. क्लस्टरिंग एक विधि है जहां क्लस्टर व्यक्तिगत डेटा बिंदुओं से बनते हैं। (विभाजित/संकेद्रण/पदानुक्रमित)	2	4	1
j) Which of the following is not a type of dataset? a) Text Dataset b) Image Dataset c) String Dataset d) Time Series Dataset निम्नलिखित में से कौन सा डेटा सेट का प्रकार नहीं है? a) टेक्स्ट डेटा सेट b) इमेज डेटा सेट c) स्ट्रिंग डेटा सेट d) समय श्रृंखला डेटा सेट	2	1	2

Group (B) (ग्रुप -बी)

Answer all five questions. (सभी पाँच प्रश्नों के उत्तर दें|) **4x5=20**

Q.2 Define Machine Learning and explain its relationship with Artificial Intelligence. मशीन लर्निंग को परिभाषित करें और इसका आर्टिफिशियल इंटेलिजेंस के साथ संबंध समझाएं।	4	1	1
OR (अथवा) Discuss the differences between Supervised, Unsupervised, and Reinforcement Learning. पर्यवेक्षित, अप्रशिक्षित, और पुनर्बलन लर्निंग के बीच अंतर पर चर्चा करें।	4	2	2

Q.3 Explain the significance of preprocessing in Machine Learning. मशीन लर्निंग में पूर्वप्रसंस्करण का महत्व समझाएं। **OR (अथवा)** How does cross-validation improve the performance of Machine Learning models? क्रॉस–वैधीकरण मशीन लर्निंग मॉडल के प्रदर्शन को कैसे सुधारता है**?**	4 4	2 2	3 3
Q.4 What are the common types of datasets used in Machine Learning? मशीन लर्निंग में उपयोग किए जाने वाले सामान्य डेटा सेट प्रकार क्या हैं**?** **OR (अथवा)** Discuss the significance of the confusion matrix in classification problems? वर्गीकरण समस्याओं में भ्रम मैट्रिक्स का महत्व क्या है**?**	4 4	1 3	2 2
Q.5 Explain the concept of dimensionality reduction and its benefits. आयामी न्यूनीकरण की अवधारणा और इसके लाभों को समझाएं। **OR (अथवा)** Explain Linear Discriminant Analysis (LDA) and its applications. रेखीय भेदक विश्लेषण (LDA**)** और इसके अनुप्रयोगों को समझाएं।	4 4	5 5	2 3
Q.6 Explain the K-Nearest Neighbor algorithm and its use in classification. K-निकटतम पड़ोसी एल्गोरिदम और इसके वर्गीकरण में उपयोग को समझाएं। **OR (अथवा)** Explain the differences between linear regression and logistic regression. रेखीय प्रतिगमन और लॉजिस्टिक प्रतिगमन के बीच अंतर को समझाएं।	4 4	3 3	3 2
Group (C) (ग्रुप - सी) **Answer all five questions. (सभी पाँच प्रश्नों के उत्तर दें।) 6x5=30**			
Q.7 Explain the importance and applications of Independent Component Analysis (ICA) in Machine Learning, highlighting its advantages and limitations. मशीन लर्निंग में स्वतंत्र कंपोनेंट विश्लेषण (**ICA)** के महत्व और अनुप्रयोगों को समझाएं**,** इसके फायदे और सीमाएँ उजागर करें। **OR (अथवा)** Provide a comprehensive overview of hierarchical clustering, including its working mechanism, types, and applications. पदानुक्रमित क्लस्टरिंग का व्यापक अवलोकन प्रदान करें**,** जिसमें इसका कार्य तंत्र**,** प्रकार और अनुप्रयोग शामिल हैं।	6 6	5 4	2 3
Q.8 Describe the concept of dataset division, including the importance of train, test, and validation sets, and explain the process and benefits of cross-validation. डेटा सेट विभाजन की अवधारणा का वर्णन करें**,** जिसमें प्रशिक्षण**,** परीक्षण और सत्यापन सेटों का महत्व**,** और क्रॉस–वैधीकरण की प्रक्रिया और लाभों को समझाएं। **OR (अथवा)** Explain the importance of data normalization and how it is achieved. डेटा सामान्यीकरण का महत्व और इसे कैसे प्राप्त किया जाता है, समझाएँ।	6 6	1 1	2 2

Q.9	Discuss the evaluation measures for clustering, distinguishing between extrinsic and intrinsic measures and their importance. क्लस्टरिंग के लिए मूल्यांकन उपायों पर चर्चा करें, बाह्य और आंतरिक उपायों के बीच अंतर और उनका महत्व समझाएं। **OR (अथवा)** Describe the differences between training, validation, and test datasets with examples. प्रशिक्षण, मान्यता, और परीक्षण डेटा सेट के बीच अंतर को उदाहरण सहित समझाएँ।	6 6	4 1	2 3
Q.10	Explain the evaluation measures used in Machine Learning for model performance, including Sum of Squares Error (SSE), Mean Squares Error (MSE),and R-square. मशीन लर्निंग में मॉडल प्रदर्शन के लिए उपयोग किए जाने वाले मूल्यांकन उपायों का विवरण, जिसमें योगफल त्रुटि (**SSE),** माध्य त्रुटि वर्ग (**MSE),** और **R**-वर्ग शामिल हैं। **OR (अथवा)** Provide detailed real-life examples of machine learning and their impacts. मशीन लर्निंग के वास्तविक जीवन के उदाहरण और उनके प्रभाव को विस्तृत करें।	6 6	3 2	3 2
Q.11	Explain the process and importance of feature selection in Machine Learning, and discuss various techniques used for this purpose. मशीन लर्निंग में विशेषता चयन की प्रक्रिया और महत्व पर चर्चा करें और इस उद्देश्य के लिए उपयोग की जाने वाली विभिन्न तकनीकों को समझाएं। **OR (अथवा)** Describe the ROC curve, how it is plotted, and its importance in model evaluation. ROC वक्र, इसे कैसे प्लॉट किया जाता है, और मॉडल मूल्यांकन में इसकी महत्ता का वर्णन करें।	6 6	2 3	3 2

----------*****---------