

GUJARAT TECHNOLOGICAL UNIVERSITY
Diploma Engineering – SEMESTER – 4 (NEW) – EXAMINATION – Summer-2024

Subject Code: 4341603

Date: 15-06-2024

Subject Name: Fundamentals Of Machine Learning

Time: 02:30 PM TO 05:00 PM

Total Marks: 70

Instructions:

1. Attempt all questions.
2. Make Suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.
4. Use of programmable & Communication aids are strictly prohibited.
5. Use of non-programmable scientific calculator is permitted.
6. English version is authentic.

		Marks
Q.1	(a) Define Machine Learning using suitable example?	03
પ્રશ્ન.1	(અ) યોગ્ય ઉદાહરણનો ઉપયોગ કરીને મશીન લર્નિંગને વ્યાખ્યાયિત કરો	૦૩
	(b) Explain the process of machine learning with the help of schematic representation	04
	(બ) મશીન લર્નિંગની પ્રક્રિયા સમજાવો સ્કેમાટીક રેખાકૃતિ ના ઉપયોગ કરીને	૦૪
	(c) Explain different types of machine learning with suitable application.	07
	(ક) યોગ્ય એપ્લિકેશન સાથે વિવિધ પ્રકારના મશીન લર્નિંગ સમજાવો.	૦૭
	OR	
	(c) What are various issues with machine learning? List three problems that are not to be solved using machine learning.	07
	(ક) મશીન લર્નિંગમાં વિવિધ સમસ્યાઓ શું છે? • ત્રણ સમસ્યાઓ કે જે મશીન લર્નિંગનો ઉપયોગ કરીને ઉકેલી શકાતી નથી.	૦૭
Q.2	(a) Give a summarized view of different types of data in a typical machine learning problem.	03
પ્રશ્ન.2	(અ) સામાન્ય મશીન લર્નિંગ સમસ્યામાં વિવિધ પ્રકારના ડેટાનો સારાંશ આપો.	૦૩
	(b) Given Attribute 1 values:32,37,47,50,59 Attribute 2 values:48,40,41,47,49 Calculate variance for both attributes. Determine which attribute is spread out around mean.	04
	(બ) આપેલ એટ્રિબ્યુટ ૧ વેલ્યુ:૩૨,૩૭,૪૭,૫૦,૫૯ એટ્રિબ્યુટ ૨ વેલ્યુ:૪૮,૪૦,૪૧,૪૭,૪૯ બંને એટ્રિબ્યુટ માટે વેરિયન્સ ગણતરી કરો. નક્કી કરો કે કઈ એટ્રિબ્યુટ મીન ને આસપાસ સ્પ્રેડ આઉટ છે.	૦૪
	(c) List Factors that lead to data quality issue. How to handle outliers and missing values.	07
	(ક) ડેટા ગુણવત્તા સમસ્યા તરફ દોરી જતા ફેક્ટર્સ નો લિસ્ટ બનાવો. આઉટલાયર્સ અને મિસિંગ વેલ્યુ કેવી રીતે હેન્ડલ કરવું.	૦૭

OR

Q.2	(a)	Give different machine learning activities.	03
પ્રશ્ન.2	(અ)	વિવિધ મશીન લર્નિંગ પ્રવૃત્તિઓ આપો	૦૩
	(b)	Calculate mean and median of the following numbers 12,15,18,20,22,24,28,30	04
	(બ)	નીચેની સંખ્યાઓના મીન અને મીડિયન ની ગણતરી કરો ૧૨,૧૫,૧૮,૨૦,૨૨,૨૪,૨૮,૩૦	૦૪
	(c)	Write a short note on dimensionality reduction and feature subset selection in context with data preprocessing.	07
	(ક)	ડેટા પ્રીપ્રોસેસિંગના સંદર્ભમાં ડાઇમેન્શનલિટી રીડક્શન અને ફીચર સબસેટ પસંદગી પર ટૂંકી નોંધ લખો.	૦૭

Q. 3	(a)	Does bias affect the performance of the ML model? Explain briefly.	03
પ્રશ્ન.3	(અ)	શું બાયસ એમ. એલ. મોડેલના પ્રદર્શનને અસર કરે છે? ટૂંકમાં સમજાવો.	૦૩
	(b)	Compare cross-validation and bootstrap sampling	04
	(બ)	ક્રોસ-વેલિડેશન અને બુટસ્ટ્રેપ સેમ્પલિંગની સરખામણી કરો	૦૪
	(c)	A model predicts the sale of a sports shop. 83 times model predicts that customer will buy cricket Bat and shop actually make a sale of cricket Bat. 7 times model predicts for sale but customer don't buy any bat. 5 times model predicts that customer is keen to know the prices just and won't buy bat but customer actually buys the bat after knowing prices. 5-time model predicts that customer is not a buyer and customer don't buy anything. Draw confusion matrix and Calculate	07
		a. Error rate	
		b. Precision	
		c. Recall	
		d. F-measure	
	(ક)	એક મોડેલ, સ્પોર્ટ્સ શોપના વેચાણની આગાહી કરે છે. 83 વખત મોડેલ અનુમાન કરે છે કે ગ્રાહક ક્રિકેટ બેટ ખરીદશે અને વાસ્તવમાં ગ્રાહક ક્રિકેટ બેટનું ખરીદી પણ કરે છે. 7 વખત મોડેલ વેચાણ માટે આગાહી કરે છે પરંતુ ગ્રાહક કોઈ બેટ ખરીદતો નથી. 5 વખત મોડેલ અનુમાન કરે છે કે ગ્રાહક માત્ર કિંમતો જાણવા ઉત્સુક છે અને બેટ ખરીદશે નહીં પરંતુ ગ્રાહક ખરેખર કિંમતો જાણ્યા પછી બેટ ખરીદે છે. 5-ટાઇમ મોડેલ અનુમાન કરે છે કે ગ્રાહક ખરીદનાર નથી અને ગ્રાહક કંઈપણ ખરીદતો નથી. કન્ફ્યુઝન મેટ્રિક્સ દોરો અને ગણતરી કરો અ એરર રેટ બ પ્રીસીઝન ક રેકાલ ડી. એફ-મેઝર	૦૭

OR

Q. 3	(a)	Define in brief	03
		a) Target function b) Cost function c) Loss Function	
પ્રશ્ન.3	(અ)	સંક્ષિપ્તમાં વ્યાખ્યાયિત કરો	૦૩
		અ) ટાર્ગેટ ફંક્શન બ) કોસ્ટ ફંક્શન ક) લોસ ફંક્શન	
	(b)	Explain balanced fit, underfit and overfit	04
	(બ)	બેલેન્સ્ડ ફિટ, અંડરફિટ અને ઓવરફિટ સમજાવો	૦૪
	(c)	A model predicts the sale of a sports shop. 83 times model predicts that customer will buy cricket Bat and shop actually make a sale of cricket Bat. 7 times model predicts for sale but customer don't buy any bat. 5 times model predicts that customer is keen to know the prices just and won't buy bat but customer actually buys the bat after knowing prices. 5-time model predicts that customer is not a buyer and customer don't buy anything. Draw confusion matrix and Calculate	07

- a. Accuracy
- b. Kappa Value

- (ક) એક મોડેલ, સ્પોર્ટ્સ શોપના વેચાણની આગાહી કરે છે. 83 વખત મોડેલ અનુમાન કરે છે કે ગ્રાહક ક્રિકેટ બેટ ખરીદશે અને વાસ્તવમાં ગ્રાહક ક્રિકેટ બેટનું ખરીદી પણ કરે છે. 7 વખત મોડેલ વેચાણ માટે આગાહી કરે છે પરંતુ ગ્રાહક કોઈ બેટ ખરીદતો નથી. 5 વખત મોડેલ અનુમાન કરે છે કે ગ્રાહક માત્ર કિંમતો જાણવા ઉત્સુક છે અને બેટ ખરીદશે નહીં પરંતુ ગ્રાહક ખરેખર કિંમતો જાણ્યા પછી બેટ ખરીદે છે. 5-ટાઇમ મોડેલ અનુમાન કરે છે કે ગ્રાહક ખરીદનાર નથી અને ગ્રાહક કંઈપણ ખરીદતો નથી. કન્ફ્યુશન મેટ્રિક્સ દોરો અને ગણતરી કરો
- અ) એક્યુરેસી
બ) કપ્પા વેલ્યુ

- Q. 4** (a) Give classification learning steps. **03**
પ્રશ્ન.4 (અ) ક્લાસીફિકેશન લર્નિંગ સ્ટેપ્સ આપો **03**
(b) Given **04**

Hours used	smartphones	Exam Scores
2		85
3		80
4		75
5		70
6		60

find a linear relationship between the hours smartphone used (input feature) and the exam score(output target)

- (બ) આપેલ **08**

કલાકો સ્માર્ટફોન વપરાય છે	પરીક્ષાના સ્કોર્સ
૨	૮૫
૩	૮૦
૪	૭૫
૫	૭૦
૬	૬૦

સ્માર્ટફોન વપરાયેલ કલાકો (ઇનપુટ ફીચર) અને પરીક્ષાના સ્કોર (આઉટપુટ ફીચર) વચ્ચે લિનિયર સંબંધ શોધો

- (c) Explain classification steps in detail **07**
(ક) વર્ગીકરણના પગલાંને વિગતવાર સમજાવો **09**

OR

- Q. 4** (a) Does the choice of the k value influence the performance of the KNN algorithm? Explain briefly **03**
પ્રશ્ન.4 (અ) શું કે મૂલ્યની પસંદગી કે અન અન અલ્ગોરિધમના પ્રભાવને પ્રભાવિત કરે છે? ટૂંકમાં સમજાવો **03**
(b) Define Support Vectors in the SVM model. **04**
(બ) SVM મોડેલમાં સપોર્ટ વેક્ટર્સને વ્યાખ્યાયિત કરો. **08**
(c) Explain logistic regression in detail. **07**
(ક) લોજિસ્ટિક રીગ્રેશનને વિગતવાર સમજાવો. **09**

- Q.5** (a) Write a short note on Matplotlib python library. **03**
પ્રશ્ન.5 (અ) Matplotlib python library પર ટૂંકી નોંધ લખો. **03**

- (b) Suppose you have a dataset containing the following points in a two-dimensional space: **04**
 $\{(2,3),(3,3),(4,3),(5,3),(6,3),(7,3),(8,3),(25,20),(26,20),(27,20),(28,20),(29,20),(30,20)\}$

- Perform k-means clustering to partition these points into two clusters.
 (બ) ધારો કે તમારી પાસે ટ્વિ-પરિમાણીય જગ્યામાં નીચેના બિંદુઓ ધરાવતો ડેટાસેટ છે: **૦૪**
 $\{(૨,૩),(૩,૩),(૪,૩),(૫,૩),(૬,૩),(૭,૩),(૮,૩),(૨૫,૨૦),(૨૬,૨૦),(૨૭,૨૦),(૨૮,૨૦),(૨૯,૨૦),(૩૦,૨૦)\}$

- આ બિંદુઓને બે ક્લસ્ટરોમાં વિભાજિત કરવા માટે k-મીન્સ ક્લસ્ટરિંગ કરો.
 (c) Give functions and its use of Scikit-learn for **07**
 a. Data Preprocessing
 b. Model Selection
 c. Model Evaluation and Metrics
 (ક) સાઈકિટ-લર્ન ના ફંક્શન્સ અને તેનો ઉપયોગ આપો ફોર **૦૭**
 અ) ડેટા પ્રીપ્રોસેસિંગ
 બ) મોડલ પસંદગી
 ક) મોડલ મૂલ્યાંકન અને મેટ્રિક્સ

OR

- Q.5** (a) List out the major features of Numpy. **03**
પ્રશ્ન.5 (અ) Numpy ના મુખ્ય લક્ષણોની યાદી બનાવો. **૦૩**
 (b) Suppose we have the following one-dimensional dataset: **04**
 $\{1,2,4,5,7,8,10,11,12,14,15,17\}$

- Perform k-means clustering to partition these points into three clusters.
 (બ) ધારો કે તમારી પાસે નીચેનો એક-પરિમાણીય ડેટાસેટ છે: **૦૪**
 $\{૧,૨,૪,૫,૭,૮,૧૦,૧૧,૧૨,૧૪,૧૫,૧૭\}$

- આ બિંદુઓને ત્રણ ક્લસ્ટરોમાં વિભાજિત કરવા માટે k-મીન્સ ક્લસ્ટરિંગ કરો.
 (c) Give function and its use of Pandas library for **07**
 a. Data Preprocessing
 b. Data Inspection
 c. Data Cleaning and Transformation
 (ક) નમપાઇ ના ફંક્શન્સ અને તેનો ઉપયોગ આપો ફોર **૦૭**
 અ) ડેટા પ્રીપ્રોસેસિંગ
 બ) ડેટા ઇન્સ્પેક્શન
 ક) ડેટા ક્લીનિંગ અન્ડ ટ્રાન્સફોર્મએસન
