Printed Page:-04	Subject Code:- ACSAI0516 Roll. No:				
NOIDA INSTITUTE OF ENGINEERING A (An Autonomous Institute Aft B.Te SEM: V - THEORY EXAM	IND TECHNOLOGY, GREATER NOIDA filiated to AKTU, Lucknow) ech IINATION (2024 - 2025)				
Subject: Predictive Analytics					
Time: 3 Hours	Max. Marks: 100				
General Instructions:					
IMP: Verify that you have received the question p	aper with the correct course, code, branch etc.				
1. This Question paper comprises of three Section	s -A, B, & C. It consists of Multiple Choice				
Questions (MCQ s) & Subjective type questions. 2 Maximum marks for each question are indicate	d on right hand side of each question				
3. Illustrate your answers with neat sketches when	ever necessary.				
<i>4. Assume suitable data if necessary.</i>					
5. Preferably, write the answers in sequential ord	er.				
6. No sheet should be left blank. Any written mate	rial after a blank sheet will not be				
evaluated/checked.					
<u>SECTION-A</u>					
1. Attempt all parts:-					
1-a. Identify below you are given a summary regression analysis from a sample of siz coefficient of determination is? (CO1, F	y of the output from a simple linear 1 te 15, SSR=100, SST = 152. The (X1)				
(a) 0.52	OV V				
(b) 0.6579 1					
(c) 0.8111					
(d) 1.52					
1-b. Identify If a regression coefficient is les K1)	s than 1 then the other should be? (CO1, 1				
(a) Equal to 1					
(b) Less than 1					
(c) 0					
(d) Greater than 1					
1-c. State when knowledge discovered can b future , it is known as analytics?	be used to suggest actions to be taken in 1 (CO2, K1)				
(a) Predictive					

- Cognitive (b)
- Diagnostic (c)
- Prescriptive (d)

•

1-d. State whether multiple linear regression (MLR) is a ______ type of statistical 1 analysis? (CO2, K1)

- (a) univariate
- (b) bivariate
- (c) multivariate
- 1-e.

1-f.

- State if the large values of the log-likelihood statistic indicate:? (CO3, K1)
- (a) That there are a greater number of explained vs. unexplained observations.
- (b) That the statistical model fits the data well.
- (c) That as the predictor variable increases, the likelihood of the outcome occurring decreases.
- (d) That the statistical model is a poor fit of the data.
- State whether logistic regression assumes a:? (CO3, K1)
 - (a) Linear relationship between continuous predictor variables and the outcome variable

1

1

1

(b) Linear relationship between continuous predictor variables and the logit of the outcome variable.

- (c) Linear relationship between continuous predictor variables.
- (d) Linear relationship between observations.
- 1-g. State if the method by which companies analyze customer data or other types of 1 information in an effort to identify patterns and discover relationships between different data elements is often referred to as? (CO4, K1)
 - (a) Data mining
 - (b) Data digging
 - (c) Customer data management
 - (d) Consumer engagement
- 1-h. State what Autocovariance measures _____? (CO4, K1)
 - (a) Linear dependence between multiple points on the different series observed at different times

(b) Quadratic dependence between two points on the same series observed at different times

- (c) Linear dependence between two points on different series observed at same time
- (d) Linear dependence between two points on the same series observed at different times
- 1-i. Identify what is the application of machine learning methods to a large database 1 called? (CO5, K1)
 - (a) Big Data Computing
 - (b) Internet of Things
 - (c) Data Mining
 - (d) Artificial Intelligence
- 1-j. Identify which of the following is true for the coefficient of correlation? (CO5,K1) 1

(a)	The coefficient	of correlation	is not dependent	on the change of scale
· · ·			1	0

(b) The coefficient of correlation is not dependent on the change of origin

(c) The coefficient of correlation is not dependent on both the change of scale and change of origin

(d) None of the above

2. Attempt all parts:-

•

2.a.	Discuss what happens when p value for t test is lower than alpha i.e. what do you conclude? (CO1, K1)	2
2.b.	Explain the concept of overfitting and underfitting? (CO2, K1)	2
2.c.	State and explain which algorithm is better at handling outliers logistic regression or SVM? (CO3, K1)	2
2.d.	State what is residual analysis and its types? (CO4, K1)	2
2.e.	State whether do all features need to be scaled when using machine learning algorithms in detail? (CO5, K1)	2
SECTIO	<u>DN-B</u>	30
3. Answ	er any <u>five</u> of the following:-	
3-a.	Find the correlation coefficient for the following data and intrepret the result? (CO1,K1) X: 23 27 28 28 29 30 31 33 35 36 Y: 18 20 22 27 21 29 27 29 28 29	6
3-b.	Discuss how would you explain the difference between correlation and covariance? (CO1, K2)	6
3-с.	Discuss what do you understand by K-fold Cross-Validation? Explain using diagram about 5 fold cross validation? (CO2, K2)	6
3-d.	Discuss what do you understand by L1 and L2 Regularization, with an example? (CO2, K2)	6
3.e.	Express in detail "Why can't we use Linear Regression in place of Logistic Regression for Binary classification"? (CO3, K2)	6
3.f.	State what is white's test, when can we use this test and how, explain with an example? (CO4,K1)	6
3.g.	Explain what is the primary distinction between LDA and PCA also state how does LDA calculate the projection direction to maximize class separability? (CO5, K1)	6
SECTIO	<u>DN-C</u>	50
4. Answ	er any <u>one</u> of the following:-	
4-a.	Obtain the equations of line of regression for the following data : ? (CO1, K2) X: 1 2 3 4 5 6 7 8 9 Y: 9 8 10 12 11 13 14 16 15 Also obtain an estimate of x for $x = 6.2$	10
4-b.	Discuss what sort of correlation would be expected between a company's	10

expenditure on health and safety and the number of work related accidents and how can we determine whether the correlation is positive or negative and also its degree or extent and state what Is Required To Become A Data Analyst? Explain in detail (In points)? (CO1, K2)

- 5. Answer any one of the following:-
- 5-a. Explain Null Hypothesis in terms of Whites test also define the following terms, 10 with an example ? (CO2, K1)
 - 1. Homoscedasticity
 - 2. Independent Test
- 5-b. Explain what is the difference between bias and variance, and how do they impact 10 model performance, Can the (training) R-squared of a multiple linear regression model improve after removing/adding features? (CO2, K1)

10

10

- 6. Answer any one of the following:-
- 6-a. Explain the following with the help of an example? (CO3 K1)
 - 1. Precision
 - 2. Recall
 - 3. F1 Score
 - 4. Accuracy
- 6-b. Explain how to interpret the results of a Logistic regression model, Discuss how 10 "BIC is some what similar to AIC but with slight change", Discuss how do we measure the quality of statistical model in detail? (CO3, K1)
- 7. Answer any one of the following:-
- 7-a. Explain the following in detail? (CO4,K2)
 - a) Information Gain
 - b) Chi-square test
 - c) Fisher's Score
 - d) Correlation Coefficient
 - e) Variance Threshold
- 7-b. Applying the equation of two regression lines obtained in a correlation analysis of 10 60 observations are : 5x=6y+24 and 1000y=768x-3608, What is the correlation coefficient, Show that the ratio of coefficient of variability of x to that of y is 5/24, and what is the ration of variance of x and y? (CO4, K3)
- 8. Answer any one of the following:-
- 8-a. Explain how does feature scaling impact the performance of PCA and LDA? 10 (CO5, K1)
- 8-b. Explain what is the primary distinction between LDA and PCA also state how 10 does LDA calculate the projection direction to maximize class separability, Describe how in a facial recognition system, which technique would be more appropriate for feature extraction, and why? (CO5, K2)