Printed Page:-04 Subject Code:- ACSAI0516

Roll. No:

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech.

SEM: V - THEORY EXAMINATION (2023 - 2024)

Subject: Predictive Analytics

Time: 3 Hours Max. Marks: 100

General Instructions:

IMP: Verify that you have received the question paper with the correct course, code, branch etc.

1. This Question paper comprises of three Sections -A, B, & C. It consists of Multiple Choice Questions (MCQ's) & Subjective type questions.
2. Maximum marks for each question are indicated on right -hand side of each question.
3. Illustrate your answers with neat sketches wherever necessary.
4. Assume suitable data if necessary.
5. Preferably, write the answers in sequential order.
6. No sheet should be left blank. Any written material after a blank sheet will not be evaluated/checked.

SECTION A

20

1. Attempt all parts:-

1-a. A residual is defined as: (CO1) 1

(a) The difference between the actual Y values and the mean of Y

(b) The difference between the actual Y values and the predicted Y values.

(c) The predicted value of Y for the average X value

(d) The square root of the slope.

1-b. The correlation coefficient is used to determine: (CO1) 1

(a) A specific value of the y-variable given a specific value of the x-variable

(b) A specific value of the x-variable given a specific value of the y-variable

(c) The strength of the relationship between the x and y variables

(d) None of these

1-c. When knowledge discovered can be used to suggest actions to be taken in future , it is known as _____ analytics. (CO2) 1

(a) Predictive

(b) Cognitive

(c) Diagnostic

(d) Prescriptive

1-d. The main purpose of LR is : (CO2) 1

(a) Predicting one variable on the basis of another

(b) Explaining one variable in terms of another

(c) Describing the relationship between one variable and another

(d) Exploring the relationship between one variable and another

1-e. The disadvantages of Logistic Regression? (CO3) 1

(a) Sometimes a lot of Feature Engineering is required

(b) It is quite sensitive to noise and overfitting

(c) Both A and B

(d) None of the above

1-f. Can we solve the multiclass classification problems using Logistic Regression? (CO3) 1

(a) Yes

(b) No

(c) Can be yes or no

(d) Can not say

1-g. Which of the following is not an example of a time series model? (CO4) 1

(a) Naive approach

(b) Exponential smoothing

(c) Moving Average

(d) None of the above

1-h. Autocovariance measures _____ (CO4) 1

(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. Which of the following is a good test dataset characteristic? (CO5) 1

(a) Large enough to yield meaningful results

(b) Is representative of the dataset as a whole

(c) Both A and B

(d) None of the above

1-j. The independent variable is used to explain the dependent variable in _____ (CO5) 1

(a) Linear regression analysis

(b) Multiple regression analysis

(c) Non-linear regression analysis

(d) None of the above

2. Attempt all parts:-

2.a. Define adjusted R^2? (CO1) 2

2.b. Differentiate between classification and regression? (CO2) 2

2.c. Explain the use of ROC curve and the AUC of a ROC Curve. (CO3) 2

2.d. Name four types of forecasting model? (CO4) 2

2.e. Why is it important to understand your data before starting a project? (CO5) 2

SECTION B

30

3. Answer any <u>five</u> of the following:-

3-a. Can single outlier decrease or increase the correlation with a big magnitude? Is Pearson coefficient very sensitive to outliers? (CO1) 6

3-b. Which type of data is used in predictive analytics? How forecasting differs from predictive analytics. (CO1) 6

3-c. Explain polynomial regression and its application areas. (CO2) 6

3-d. Explain the regularization methods in Machine Learning? (CO2) 6

3.e. How do we handle categorical variables in Logistic Regression?(CO3) 6

3.f. Why would you use a log transformation of your data? How would you interpret the coefficients in a log-log transformed model?(CO4) 6

3.g. Explain Hash Encoding? (CO5) 6

SECTION C

50

4. Answer any <u>one</u> of the following:-

4-a. Explain the different types of Learning/ Training models in ML? (CO1) 10

4-b. How do you know if data is structured or unstructured? Support your answer with an example and also write their benefits and challenges. (CO1) 10

5. Answer any <u>one</u> of the following:-

5-a. Explain linear regression and how does it work? What are the assumptions of linear regression model? (CO2) 10

5-b. What is multicollinearity and how you can overcome it? Explain in detail? (CO2) 10

6. Answer any <u>one</u> of the following:-

6-a. Explain the following with the help of an example: (CO3) 10
a)Precision
b)Recall
c)F1 score
d)Accuracy
e)Specificity

6-b. How does logistic regression classification work? Explain its advantages, disadvantages and its application areas. (CO3). 10

7. Answer any <u>one</u> of the following:-

7-a. What are the components of time series model ? Explain each component involved in it. How do you estimate the seasonal component of time series? (CO4). 10

7-b. What is meant by the term stationary, as applied to a time series model? Explain how the notation I (0) and I (1) is related to the concept of stationarity. Give one example of a stationary model and one of a non-stationary model. (CO4) 10

8. Answer any <u>one</u> of the following:-

8-a. Why there is a need of Feature Selection. differentiate between feature selection and feature extraction methods.(CO5) 10

8-b. Explain the following: (CO5) 10
i)Data normalization in machine learning models
ii)Exploratory data analysis