What are the limitations of the time series models. (CO4)

7-b. How can we describe the approach used in the Holt-Winters double exponential smoothing model? Explain three different types of Seasonal and Cyclic patterns? Explain with examples? (CO4) 10

8. Answer any one of the following:-

8-a. What sort of correlation would be expected between a company's expenditure on health and safety and the number of work related accidents. Explain in detail. (CO5) 10

8-b. What is the difference between Data Science and Machine Learning? What are feature selection and feature scaling in machine learning? What is Exploratory Data Analysis (EDA)? (CO5) 10

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 regression analysis is inappropriate when (CO1) 1

(a) you have two variables that are measured on an interval or ratio scale.

(b) you want to make predictions for one variable based on information about another variable.

(c) the pattern of data points forms a reasonably straight line

(d) there is heteroscedasticity in the scatter plot.

1-b. In the regression equation $y = bo + b1x$, bo is the (CO1) 1

(a) slope of the line

(b) independent variable

(c) y intercept

(d) coefficient of determination

1-c. If there is a very strong correlation between two variables then the correlation coefficient must be (CO2) 1

(a) any value larger than 1

(b) much smaller than 0, if the correlation is negative

(c) much larger than 0, regardless of whether the correlation is negative or positive

(d) None of these alternatives is correct.

1-d. In least squares regression, which of the following is not a required assumption about the error term ε? (CO2) 1

(a) The expected value of the error term is one.
(b) The variance of the error term is the same for all values of x.
(c) The values of the error term are independent.
(d) The error term is normally distributed.

1-e. In a logistic regression, if the predicted logit is 0, what's the transformed probability? (CO3) 1

(a) 0
(b) 1
(c) 0.5
(d) 0.05

1-f. _____ are defined as the ratio of the probability of an event occurring to the probability of the event not occurring (CO3) 1

(a) Simple
(b) Even
(c) Regex
(d) Odds

1-g. Which of the following is relatively easier to estimate in time series modeling? (CO4) 1

(a) Seasonality
(b) Cyclical
(c) Seasonality & Cyclical
(d) None of these

1-h. Which of the following is not a technique used in smoothing time series? (CO4) 1

(a) Tree based models like (CART)
(b) Smoothing Splines
(c) Locally weighted scatter plot smoothing
(d) None of These

1-i. Which of the following is an example of feature extraction? (CO5) 1

(a) construction bag of words from an email
(b) applying pca to project high dimensional data
(c) removing stop words
(d) forward selection

1-j. If the values of two variables move in the opposite direction, _____ (CO5) 1

(a) The correlation is said to be linear
(b) The correlation is said to be non-linear
(c) The correlation is said to be positive
(d) The correlation is said to be negative

2. Attempt all parts:-

2.a. What does coefficient of determination explain (in terms of variation) ? (CO1)

2.b. Explain the concept of overfitting and underfitting? (CO2)

2.c. What are R-squared and Adjusted R-squared? (CO3)

2.d. What are the two types of time series? (CO4)

2.e. Can you give me an example of where feature scaling would be required? (CO5)

SECTION-B

3. Answer any five of the following:-

3-a. What are the benefits of predictive analytics? (CO1)

3-b. Does causation imply correlation. Explain. Also state the difference between correlation and simple linear regression? (CO1)

3-c. Which factors play a role in the selection of a loss function for a mode? (CO2)

3-d. How do L1 and L2 regularization differ in improving the accuracy of machine learning models? (CO2)

3.e. What is the difference between the outputs of the Logistic model and the Logistic function? (CO3)

3.f. Describe what we use hypothesis testing for? What are p values and describe how p values are used in hypothesis testing? (CO4)

3.g. What is meant by Mean/Target Encoding? (CO5)

SECTION-C

4. Answer any one of the following:-

4-a. Differentiate between Approximation error and Estimation error? With Diagram? (CO1)

4-b. List out some of the best practices for Data Cleaning? Also, state the difference between multicollinearity and heteroskedasticity in regression analysis? (CO1)

5. Answer any one of the following:-

5-a. What do you understand by Regularization Techniques? What is Shrinkage method? Also, highlight the relationship between the target and available independent variables. (CO2)

5-b. You created a predictive model of a quantitative outcome variable using multiple regressions. How would you validate a model you created to generate a predictive analysis? (CO2)

6. Answer any one of the following:-

6-a. What are the advantages and disadvantages of conditional and unconditional methods of MLE? (CO3)

6-b. Why accuracy is not a good evaluation metric for a dataset of imbalanced classes? (CO3)

7. Answer any one of the following:-

7-a. Explain different trend trading strategies using a variety of technical indicators?