Sr. No. 003506

December 2023

B. Tech (CE/CE (Hindi Medium)) 5th Semester

Machine Learning (PEC-CS-D-501)

Time: 3 Hours | Max. Marks:75

Instructions:
1. *It is compulsory to answer all the questions (1.5 marks each) of Part -A in short.*
2. *Answer any four questions from Part -B in detail.*
3. *Different sub-parts of a question are to be attempted adjacent to each other.*
4. *Use of calculator is allowed*

PART -A

Q1 (a) What are different types of machine learning paradigm? (1.5)

(b) What is overfitting? How you can avoid it? (1.5)

(c) What is training set and test set in a machine learning model? How much data will you allocate for training, validation and test sets? (1.5)

(d) Explain confusion matrix with respect to machine learning algorithm. (1.5)

(e) Explain the difference between machine learning, artificial intelligence and deep leaning. (1.5)

(f) What is 'naïve' in naïve bayes classifier? (1.5)

(g) Define precision and recall. (1.5)

(h) What is pruning in decision trees? How it is done? (1.5)

(i) Why there is a need for regularizations? (1.5)

(j) Is a high variance in data good or bad? (1.5)

PART -B

Q2 (a) Consider the following data (effects of hours of mixing on temperature of wood pulp), draw the best fit line using linear regression. (7)

HOURS OF MIXING (X)	TEMPERATURE OF WOOD PULP (Y)
2	21
4	27
6	29
8	64
10	86
12	92

(b) Consider the above data calculate (4+4)

I. the value of Y for X=16

II. the total error i.e. sum of the residuals in the predicted value of Y

Q3 (a) How Random Forest works? (5)

(b) How prediction is done by time series regression. What are the components (10)

00356/2

of time series? Can time series algorithms estimate the total sales in next 3 years of an insurance company?

Q4 What is machine learning? Explain steps of implementation of machine learning algorithm? Why data preprocessing step is important? Discuss recent trends in various learning techniques of machine learning. (15)

Q5 (a) What is bias and variance in machine modeling? (5)

(b) Explain distance measure formulas used in clustering. Using distance-based measure cluster the following eight points (with (x, y) representing locations) into three clusters: (10)

A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)

Initial cluster centers are: A1(2, 10), A4(5, 8) and A7(1, 2).

Q6 (a) Differentiate between classification and clustering? How classification model is made to learn by machine learning algorithm? How do you design email spam filter? (10)

(b) How is Amazon able to recommend other things to buy? (5)

Q7 Write short notes on the following: (15)

I. Principal Component Analysis
II. Artificial Neural Network
III. Baye's theorem

****