

PGD Mid Exam (Machine Learning) 2024

Task 1:

Every year many students give the GRE exam to get admission in foreign Universities. The data set contains GRE Scores (out of 340), TOEFL Scores (out of 120), University Rating (out of 5), Statement of Purpose strength (out of 5), Letter of Recommendation strength (out of 5), Undergraduate GPA (out of 10), Research Experience (0=no, 1=yes), Admitted (0=no, 1=yes). Admitted is the target variable.

Data Set Available on kaggle (The last column of the dataset needs to be changed to 0 or 1)

```
[1 if each > 0.75 else 0 for each in df['Chance of Admit ']]
```

Dataset: Available on google classroom

The counselor of the firm is supposed check whether the student will get an admission or not based on his/her GRE score and Academic Score. So to help the counselor to take appropriate decisions build a machine learning model classifier using Decision tree to predict whether a student will get admission or not.

- Apply Data pre-processing (Label Encoding, Data Transformation....) techniques if necessary.
- Perform EDA
- Perform data-preparation (Train-Test Split)
- Apply Machine Learning Algorithm
- Evaluate Model.

Task 2:

Download heart dataset from following link.

<https://www.kaggle.com/datasets/zhaoyingzhu/heartcsv>

Perform following operation on given dataset.

- a) Find Shape of Data
- b) Find Missing Values
- c) Find data type of each column
- d) Finding out Zero's
- e) Find Mean age of patients
- f) Now extract only Age, Sex, ChestPain, RestBP, Chol. Randomly divide dataset in training (75%) and testing (25%).
- g- Create confusion matrix and classification report

Create confusion matrix based on given data mathematically and find

- I. Accuracy
- II. Precision
- III. Recall
- IV. F-1 score

Data: Through the diagnosis test I predicted 100 report as COVID positive, but only 45 of those were actually positive. Total 50 people in my sample were actually COVID positive. I have total 500 samples.

Task 3:

Discuss the following (Any 3)

- Overfitting
- Under fitting
- Best fit
- Bias
- Variance

Task 4:

House Price Prediction

The Ames Housing dataset is taken from kaggle competition. The aim of the project is to predict house price for houses in Boston Housing Dataset. You have to download dataset [train.csv](#) from a google classroom.

Perform Data pre-processing if required.

- 1-Perform EDA and show visualization among different features
- 2- Select a model for your data and train your model on 80%-20%
- 3- Calculate MAE , MSE and RMSE .