School of Computer Science Engineering and Technology

Course- BTech
Course Code- CSET301
Year- 2022
Date- 3-10-2022

Type- Core
Course Name-AIML
Semester- Even
Batch- V Sem

Lab Assignment No. 7 . 1 . 1

Exp. No.	Name	CO-1	CO-2	CO-3
7.1.1	Random Forest classifier	✓	✓	--

Lab Assignment No. 7.1.1

Objective: To Implement Random Forest classifier

About Dataset:

Data Set Characteristics:	Multivariate	**Number of Instances:**	4898	**Area:**	Business
Attribute Characteristics:	Real	**Number of Attributes:**	12	**Date Donated**	2009-10-07
Associated Tasks:	Classification, Regression	**Missing Values?**	N/A	**Number of Web Hits:**	1942447

The list of attributes with description is given below:

Input variables (based on physicochemical tests):

1 - fixed acidity
2 - volatile acidity
3 - citric acid
4 - residual sugar
5 - chlorides
6 - free sulfur dioxide
7 - total sulfur dioxide
8 - density
9 - pH
10 - sulphates
11 - alcohol

Output variable (based on sensory data):

12 - quality (score between 0 and 10)

Steps

1. **Dataset:** Download the dataset from the link (5)
   https://archive.ics.uci.edu/ml/datasets/wine+quality
2. Read the dataset (5)
3. Extract the Independent and Dependent Variable (5)
4. Convert the Output column quality (score between 0 and 10) into the three categories i.e., best, average, and poor. (10)

5. Split the dataset into training and testing using 75-25 division (10)
6. Perform normalization on numerical features (10)
7. Build a Random Forest classification model using Sklearn with default parameters. Predict the target values in the testing set. (10)
8. Check the performance of model using confusion matrix (10)
9. Calculate the accuracy, PR, RR and F1-score (10)
10. Playing with Random Forest: Change the following parameters of the random forest and analyze their performance for training and testing using the evaluation measures. (20)
    a. n_estimators
    b. criterion{“mse”, “mae”}
    c. max_depth
    d. min_samples_split
    e. min_samples_leaf
    f. max_features
    g. random_state
11. Compare the performance of the Random Forest model with Decision Tree Model. (10)

Suggested Platform: Python: Azure Notebook/Google Colab Notebook, packages such as numpy, nltk, regular expression package re.

data for each setup.