School of Computer Science Engineering and Technology

Course- BTech
Course Code- CSET301
Year- 2022
Date- 12-09-2022

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

Lab Assignment 4.2_2

Exp. No.	Name	CO-1	CO-2	CO-3
4.2_2	Naïve bayes Classifier	✓	✓	--

Objective: Implement Naïve bayes Classifier to model wine quality based on physicochemical test

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)

1. Load the dataset from the link https://archive.ics.uci.edu/ml/datasets/wine+quality (5)
2. Read the dataset (5)
3. Check the presence of missing values (5)
4. Impute the missing values and remove any undesirable feature from the dataset. (10)
5. Check for the outliers in the columns and treat the outliers if present. (10) (Optional Part)

6. Handle the Target columns. Map the quality column to numeric form such as: " >8 " to 3 (Best Quality) and " Greater than 6 and less than 8" to 2 (medium quality), "less than 5" to 1 (Worst)
7. Split the dataset into train and test. (Ratio: 70:30, 80:20) (10)
8. Construct Naïve Bayes model (Hint: use GaussianNB model) (10)
9. Evaluate the performance of model on train and test subsets using accuracy, recall. Also check the values in confusion matrix. (20)
10. Explore the different parameters while creating naïve bayes classifier model.

Suggested Platform: Python: Jupyter Notebook/Azure Notebook/Google Colab.