**Lab Exercises-4**

1. **Use pima\_Indian\_Diabetes dataset and perform the following:**
2. Use preprocessing methods to clean the dataset.
3. Spit the dataset into Train and Test dataset, respectively.
4. Visualize your training data to know whether the relationship between dependent and independent variable is linear.
5. Use train dataset to create a decision tree model.
6. Use test dataset for predictions.
7. Access the accuracy of your model.
8. **Use balance dataset and perform the following:**
9. Use preprocessing methods to clean the dataset.
10. Spit the dataset into Train and Test dataset, respectively.
11. Use train dataset to create a Decision tree model.
12. Use test set for making predictions
13. Access the quality of your classification model.

**3.** Use the **red\_wine** dataset. This dataset is related to red variants of the Portuguese “Vinho Verde” wine. Due to privacy and logistic issues, only physicochemical (inputs) and sensory (the output) variables are available (e.g. there is no data about grape types, wine brand, wine selling price, etc.).

You will take into account various input features like fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, alcohol. Based on these features predict the quality of the wine.

4. Consider the following problem:

In AB Company, there is a salary distribution table based on Year of Experience as given in “salary\_dataset.csv”. **“The scenario is you are a HR officer, you got a candidate with 5 years of experience. Then what is the best salary you should offer to him?”**

5. Use diabetes dataset of Q1. And classify the people who are at the risk of getting diabetes. Make the use of Logistic Regressor classifier to build classification model. Compare your results

**6. Use titanic dataset. Handle the null values and convert the categorical values into numerical values. Make a classification model using Logistic Regression to predict the survival of a passenger on the ship.**

**7. Use “bill\_authentication.csv” dataset. Your task is to predict whether a bank note is authentic or fake depending upon the four different attributes of the image of the note. The attributes are Variance of wavelet transformed image, curtosis of the image, entropy, and skewness of the image.**