**Lab Exercises-6**

1. Use petrol\_consumption dataset. Your task is to predict the gas consumption (in millions of gallons) in 48 of the US states based on petrol tax (in cents), per capita income (dollars), paved highways (in miles) and the proportion of population with the driving license. Build the regression model using Random Forest Regressor. Analyze the prediction ability of your model.

2. Use the wine dataset, which is a very famous multi-class classification problem. This data is the result of a chemical analysis of wines grown in the same region in Italy using three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines. The dataset comprises 13 features ('alcohol', 'malic\_acid', 'ash', 'alcalinity\_of\_ash', 'magnesium', 'total\_phenols', 'flavanoids', 'nonflavanoid\_phenols', 'proanthocyanins', 'color\_intensity', 'hue', 'od280/od315\_of\_diluted\_wines', 'proline') and a target (type of cultivars). This data has three types of cultivar classes: 'class\_0', 'class\_1', and 'class\_2'.

Your task is to build a model using k-nn classifier to classify the type of cultivar. Access the performance of your model using different values of ‘***k*’.**

3. Use iris flower dataset to create classification model. Your task is to predict the class to which these plants belong. There are three classes in the dataset: Iris-setosa, Iris-versicolor and Iris-virginica. Create the classification model using k-nn classifier and evaluate the performance of your classifier. Find the best value of ‘***k***’ for iris dataset.

4. Use breast cancer dataset, build classification model using k-nn classifier for predicting whether a person suffers from cancer or not. Find the best value of ‘***k***’ for breast cancer dataset. Evaluate the performance of your model for different ‘***k***’ values.

5. Use iris dataset of Q3 and build the k-nn model using k-fold cross validation, stratified cross fold validation. Compare the results of train\_test\_split, k-fold and skfold validation techniques.

6. Use diabetes dataset. The datasets consist of several medical predictor (independent) variables and one target (dependent) variable, Outcome. Independent variables include the number of pregnancies the patient has had, their BMI, insulin level, age, and so on. Use k-nn to build the classification model. Evaluate your model performance. Use “gridsearchcv( )” to find the best value of ‘***k***’.

7. Use titanic dataset. Handle the null values and convert the categorical values into numerical values. Make a classification model using k-nn classifier to predict the survival of a passenger on the ship. Use “gridsearchCV( )” to find the best value of **‘*k*’.**