**Lab Exercises-7**

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 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 Random Forest classifier and evaluate the performance of your classifier. Also you need to do the feature analysis and find out the important feature in iris dataset.

3. Use the [make\_classification() function](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.make_classification.html) to create a synthetic binary classification problem with 1,000 examples and 20 input features. Use this synthetic dataset to build a classification model using Random forest classifier. Evaluate your model using stratified cross fold validation.

4. Use the [make\_regression() function](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.make_classification.html) to create a synthetic binary classification problem with 1,000 examples and 20 input features. Use this synthetic dataset to build a classification model using Random forest regressor. Evaluate your model using stratified cross fold validation.

5. 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 random forest ensemble method to build the classification model. Evaluate your model performance. Use “gridsearchcv( )” to find the best value of decision trees for creating forest.

6. Use breast cancer dataset. The dataset contains a total number of 10 features labeled in either *benign* or *malignant* classes. The features have 699 instances out of which 16 feature values are missing. The dataset only contains numeric values. Implement the ensemble method using Decision tree classifiers.