



Module 8: Random Forest Assignment

Problem Statement:

You work in XYZ Company as a Python Data Scientist. The company officials have collected some data on diabetes based on years of experience and wish for you to create a model from it.

Dataset: diabetes.csv

Tasks To Be Performed:

1. Load the dataset using pandas
2. Extract data from outcome column is a variable named Y
3. Extract data from every column except outcome column in a variable named X
4. Divide the dataset into two parts for training and testing in 70% and 30% proportion
5. Create and train Random Forest Model on training set
6. Make predictions based on the testing set using the trained model
7. Check the performance by calculating the confusion matrix and accuracy score of the model