

Building the Classification Model

The train.csv file contains the training data. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. The datasets consists of several medical predictor variables and one target variable, Outcome. Predictor variables includes the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.

The following task has to be performed to train the model:

- a) Read the data
- b) Display first 5 rows of the dataset
- c) Separate the feature and target column in 2 different variable and name them as “X”(features), “y”(Target)
- d) Now divide the data into train and test (use train test split function of sklearn)
- e) Use Logistic regression and Naïve Bayes to train the model (use .fit)
- f) Find the test accuracy on both the models and compare the results.
- g) Plot the confusion matrix and precision recall curve.

Note : This assignment is made for practice only and is not graded.