

Himanshu Mahajan
101803620
COE-28

classmate

Date _____

Page _____

Diabetes Prediction

Novelty of Project

1.) Optimization of Hyper parameters :-

After trying out various combinations of train-test split size & n-estimators for my random forest classifier, I found out that the best accuracy by taking n-estimators as 20 & train-test split size as 0.20 respectively. Any further increase in no. of n-estimators lead to increase in training loss & decrease in accuracy due to overfitting.

2.) Methodology :- I used statistical measures depending on the data distribution to fill the missing values in the features, then trained a random forest classifier in order to provide a reliable prediction for diabetes.