**ADA BOOST AND XGBOOST**

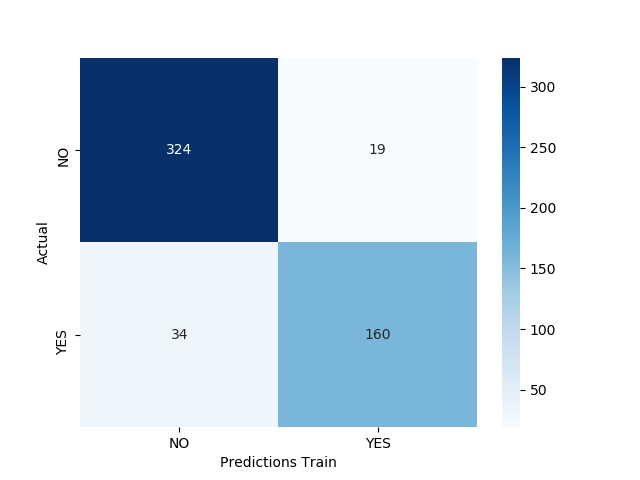
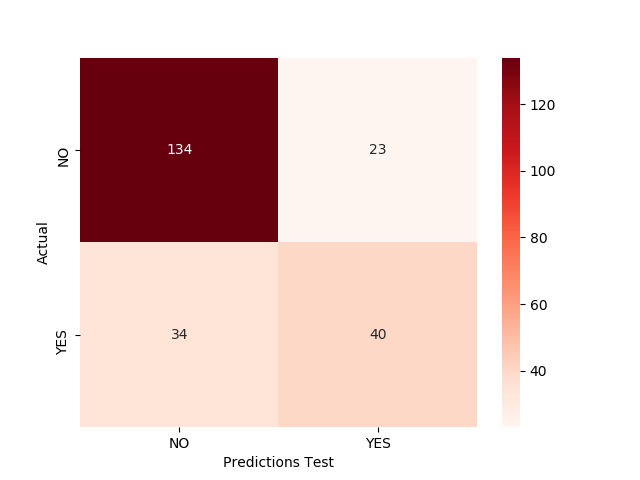
**Business Problem** = ﻿﻿﻿﻿﻿Perform AdaBoost and Extreme Gradient Boosting for the Diabetes\_RF dataset.

* **Name of the File: -** Diabetes\_RF.csv
* **Size of the File: -** 29 KB
* **Necessary Data : -** 768 Observations, 9 Features.

**Exploratory data Analysis** =

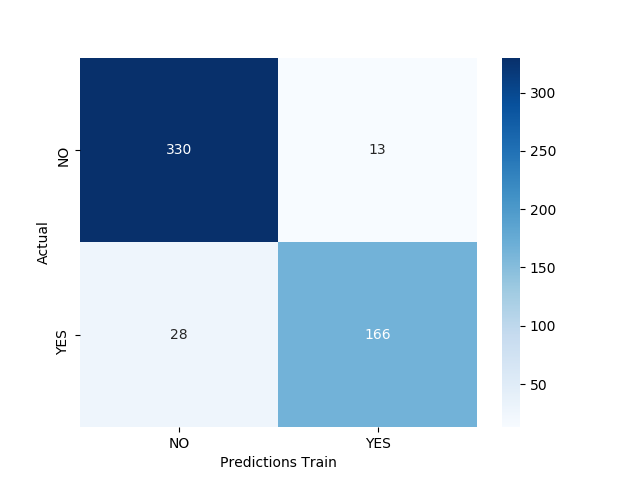
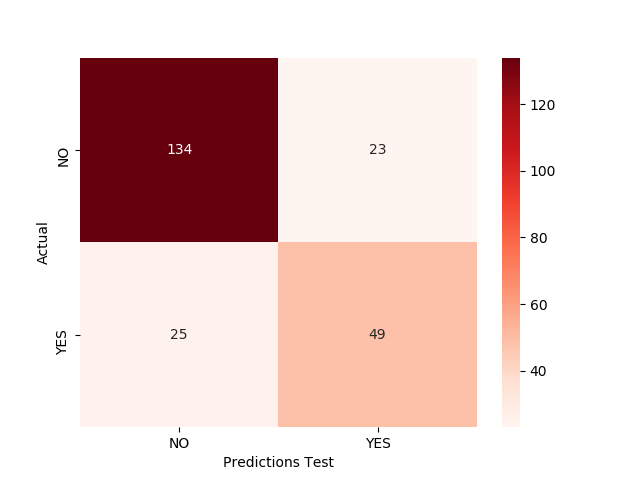
* **Outliers: -**  Outliers are not presents.
* **Missing Value: -** Data don’t have Missing Values
* **Output:** - Categorical
* **Sampling:**- Stratified Sampling (65% - 35%)

**Ada Boosting =**

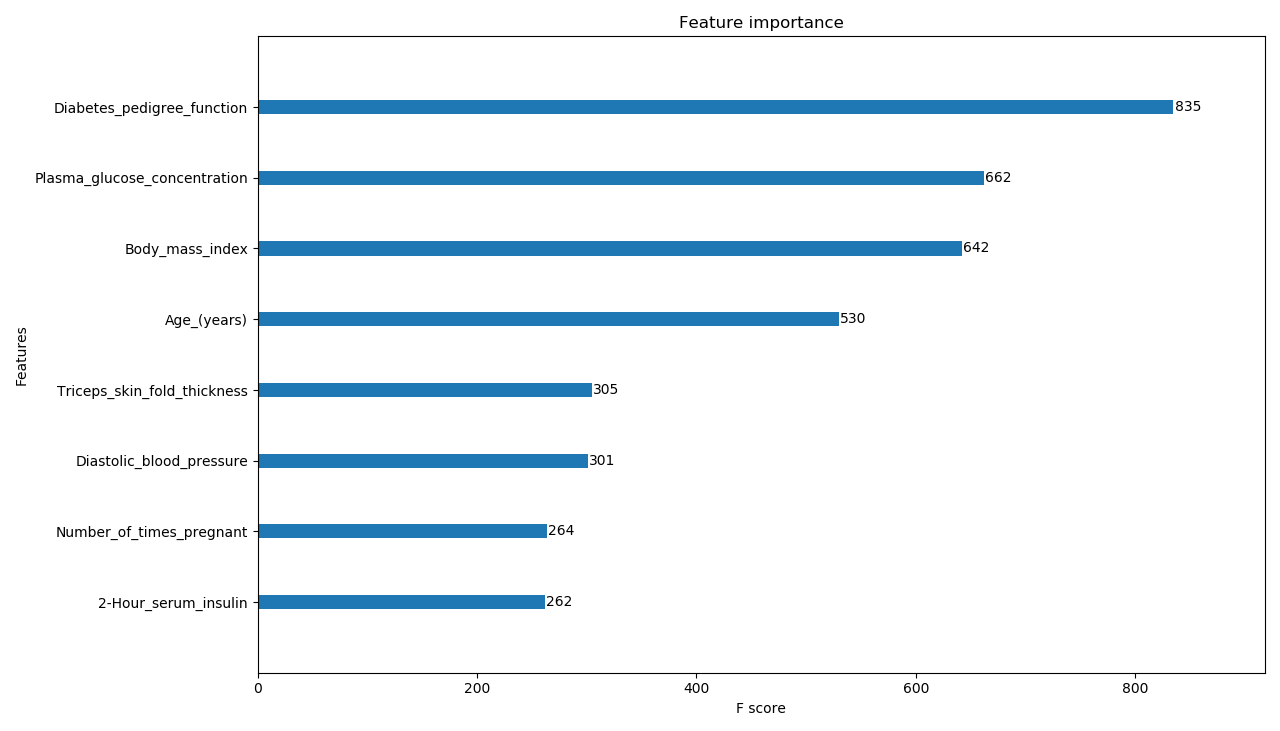
* **﻿Criterion:-** Entropy
* **Base of ﻿Estimators: -** Decision Tree
* **No. of ﻿Estimators:-** 200
* **﻿Learning Rate:-** 1
* **Accuracy Score Train:-** 90 %
* **Accuracy Score Test:-** 75%
* **Confusion Matrix : -**

Accuracy of model on Train data is high but at test data it is very low, It is over fit issue. Let’s try ﻿Extreme Gradient Boosting.

**XG Boosting =**

* **﻿﻿Max Depth:-** 4
* **No. of ﻿Estimators:-** 365
* **﻿Learning Rate:-** 0.02
* **﻿Sub Sample:-** 1
* **Accuracy Score Train:-** 90 %
* **Accuracy Score Test:-** 79%
* **Confusion Matrix : -**

Accuracy of model on test data is higher as compare to Ada Boost. We can use this model for predication

* **Important Feature Plot:-**

**Python code file**: - [Diabetes Analysis.py](https://github.com/nilaydeshmukh0/AdaBoost-XGBoost/blob/master/Diabetes%20Analysis/Diabetes%20Analysis.py)