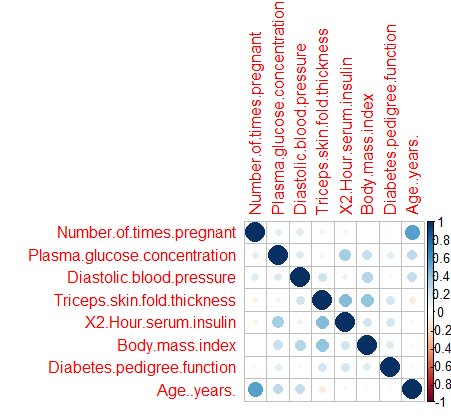
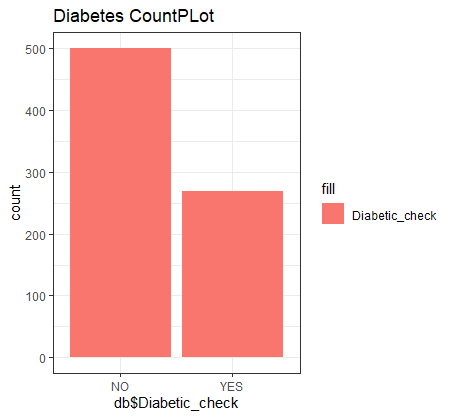
Ensemble Techniques

1. **Diabetes\_RF:**

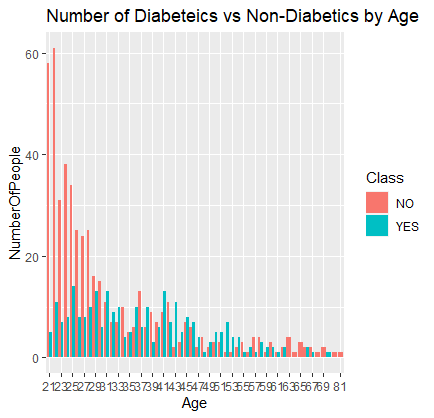


We can good amount of correlation between Age.Years and Number of Times Pregnant.

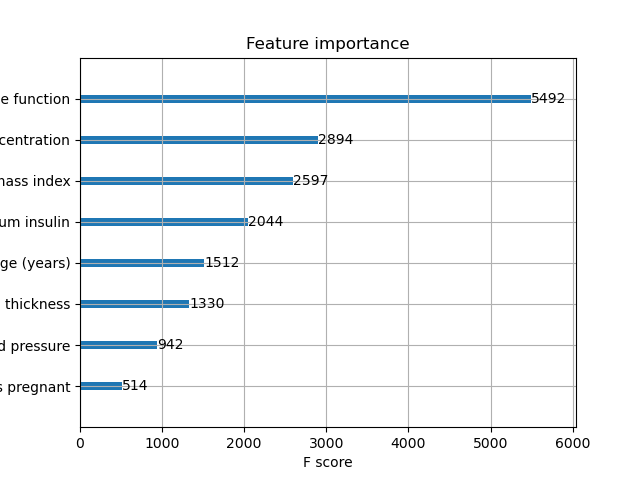
And between Body mass index and Triceps skin fold thickness.



In the above barplot the number of patients with diabetes are far less than diabetes patients. (NO – above 500 , YES- 260)



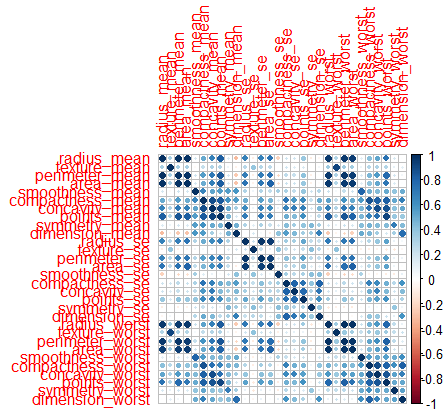
Majority of people are in the range of 21-53 that is 60 for 21 in which there are somewhere around 5 people from (21 group) having YES -diabetes



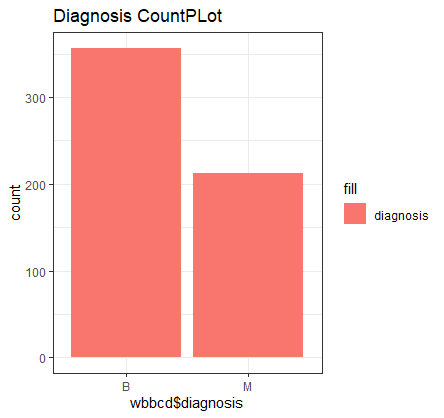
Most important feature according the XGboost is Diabetes pedigree function.

That obtains the most F1 score .

1. **Wbcd \_Dataset:**

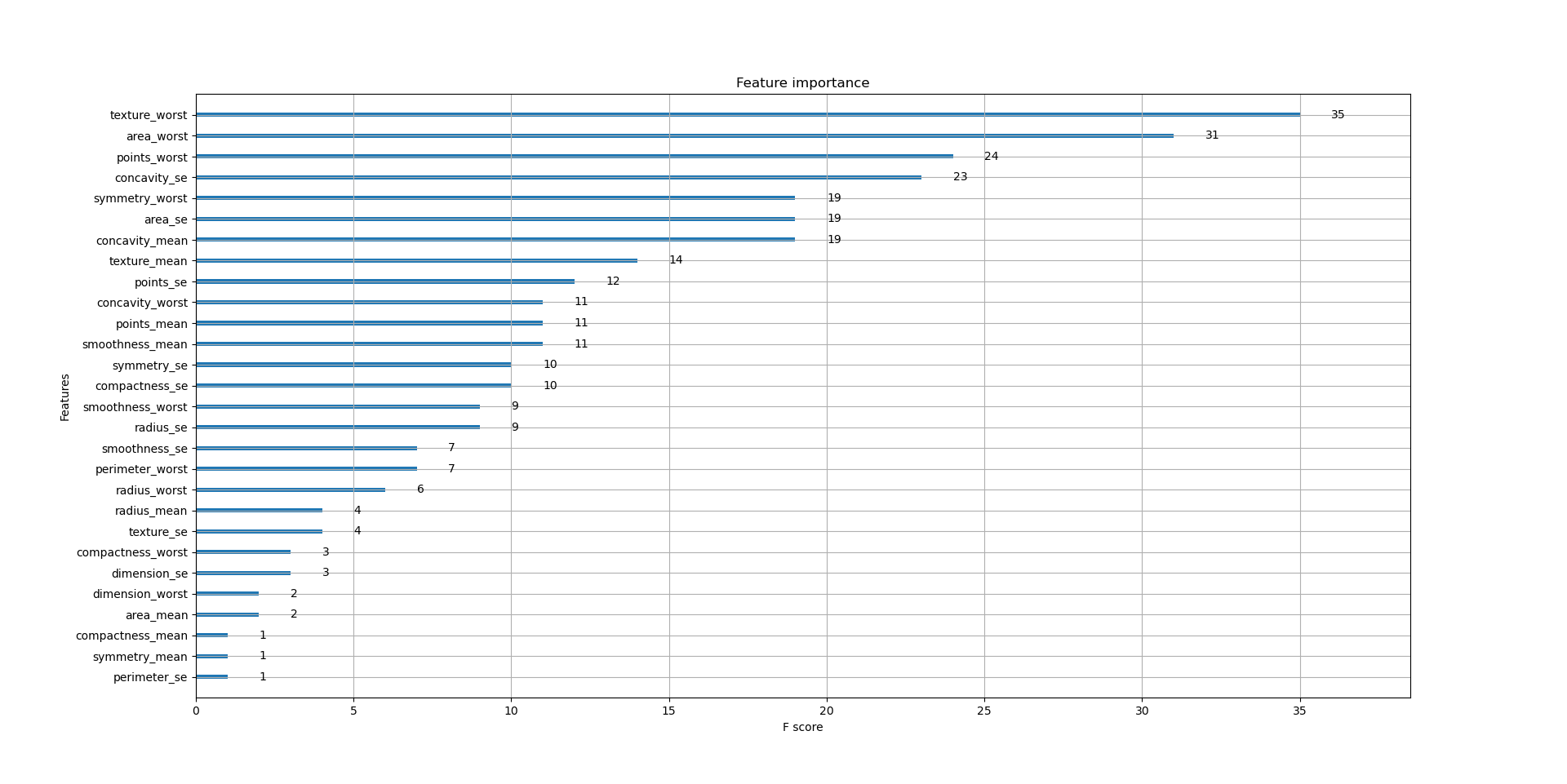


Too many parameters , might PCa to find the important features .Of that Texture worst seems to have the best impact on accuracy.(important feature)



There are more people with Benign than malignant .

About 320 count in for Benign and 200 for Malignant.



The XGB tree at creating unique trees after many iteration over the gamma and using gridSerachCV best params we get.

