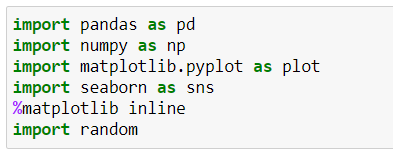
Random Forest:

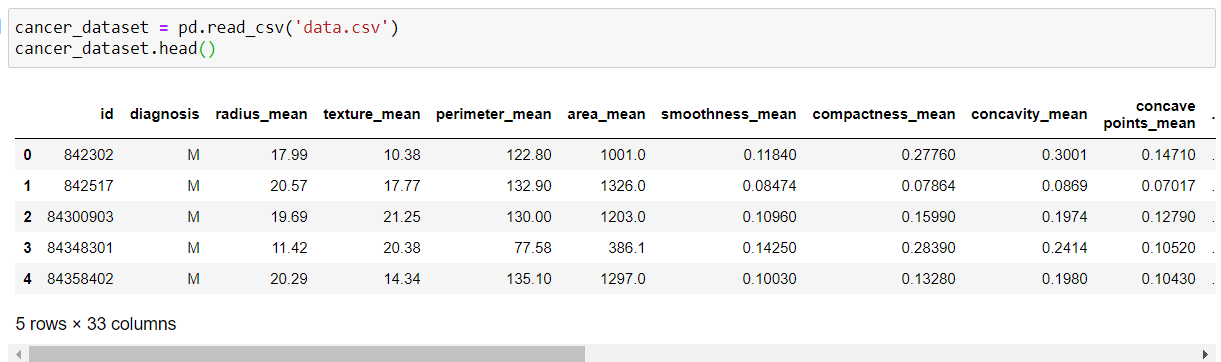
Random forest is known as Random Decision Forest. It is an ensemble method used for classification and regression problems. Mainly, ensemble methods use multiple learning models to gain better predictive results.

Implement a Random Forest Model for classification of the FNA images:

Importing the modules:

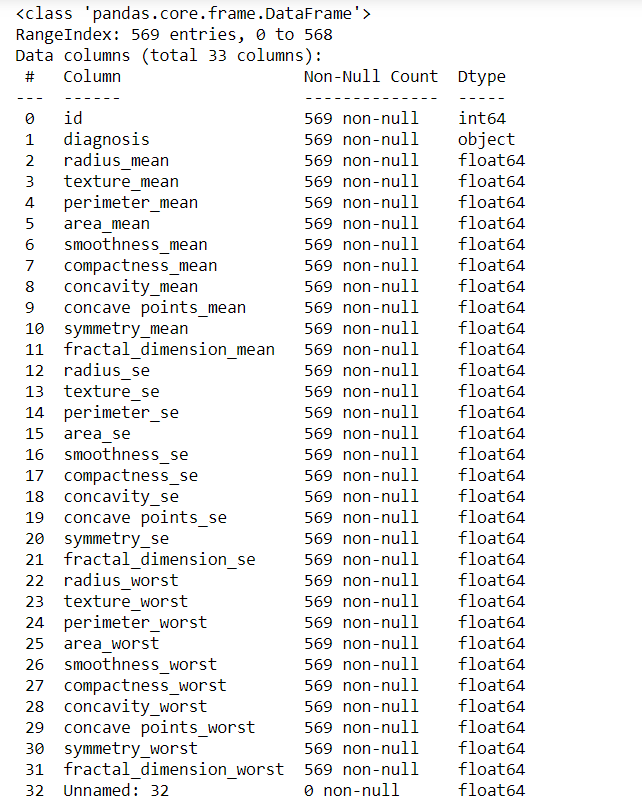


Loading the dataset into the data frame:

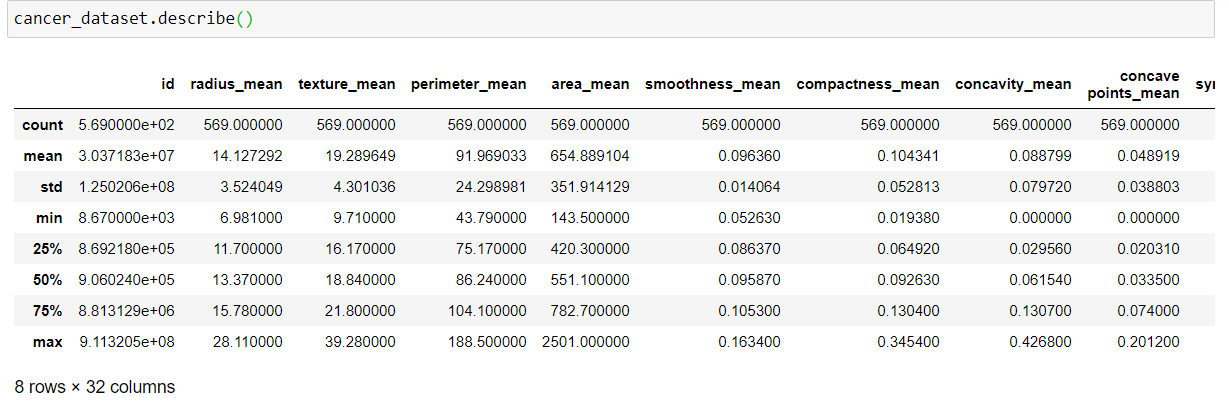


To print the dataset index datatypes and columns, non-null values and memory usage.

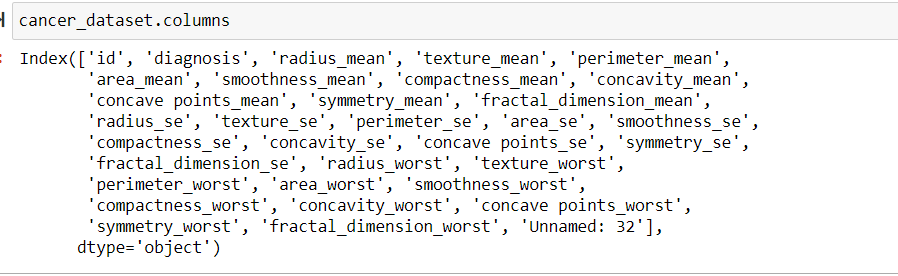




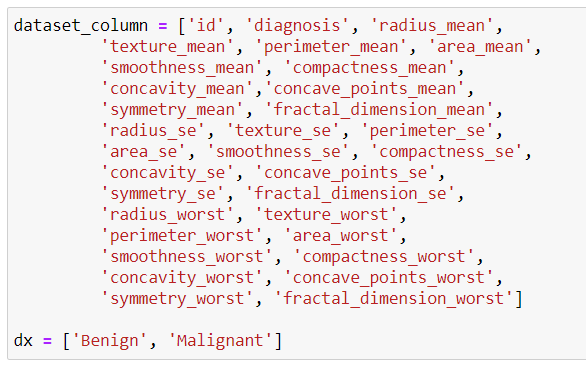
To get the statistical description:



To get all the columns in the dataset:

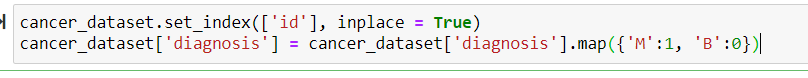


Loading the data into pandas data frame:

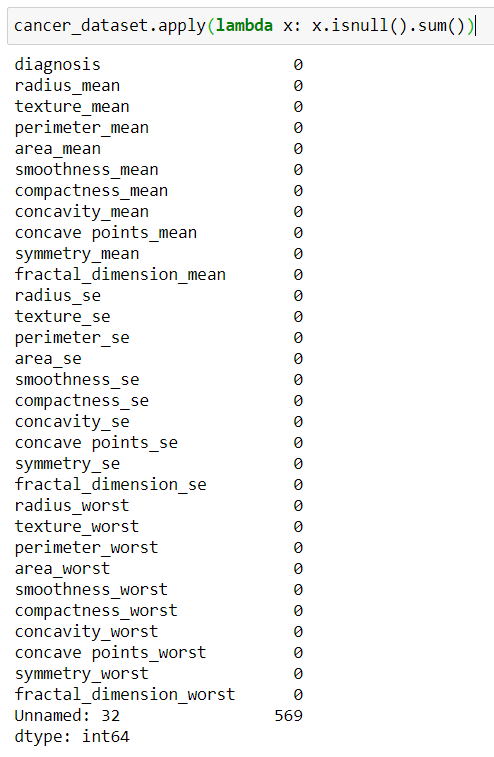


Data Cleaning:

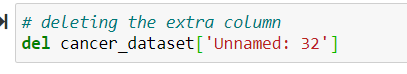
Here, we are setting id\_number to be the data frame index:



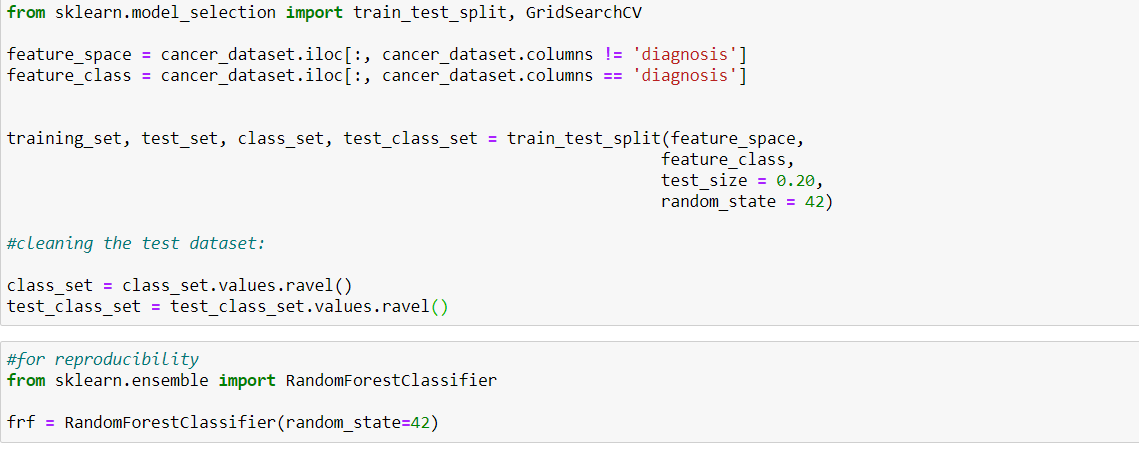
Verifying the column name and missing values for that column:



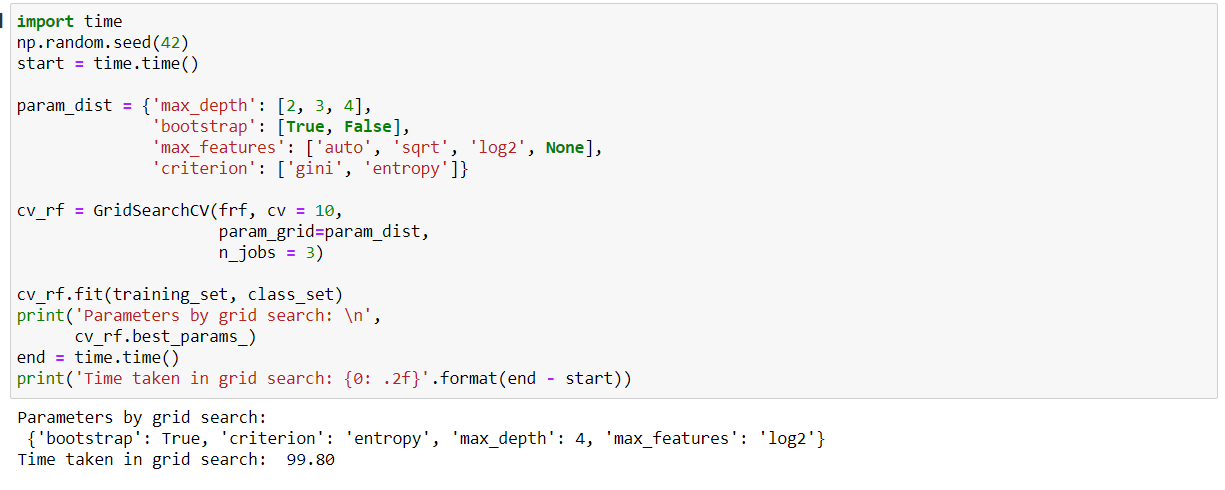
Deleting the extra column which we doesn’t use in the process:



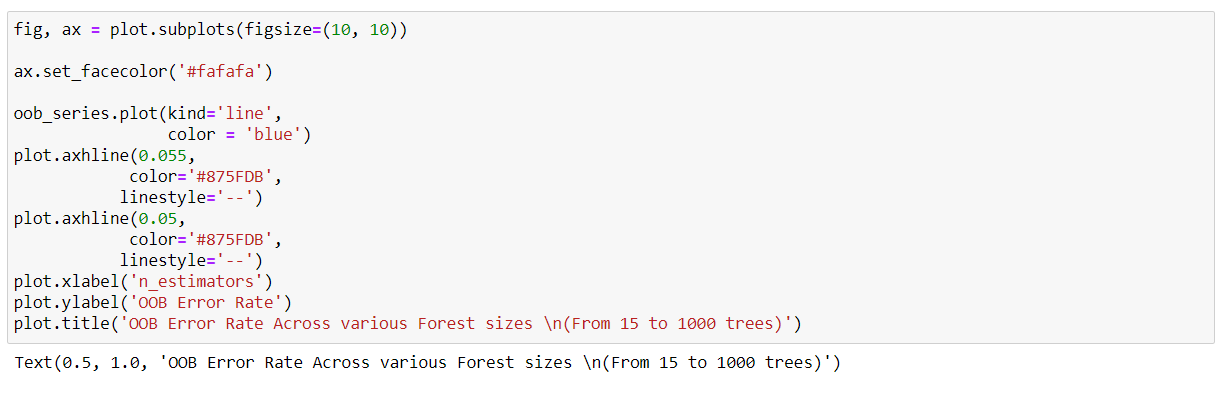
Now, splitting the dataset into training and test sets so that it creates 80-20% split.

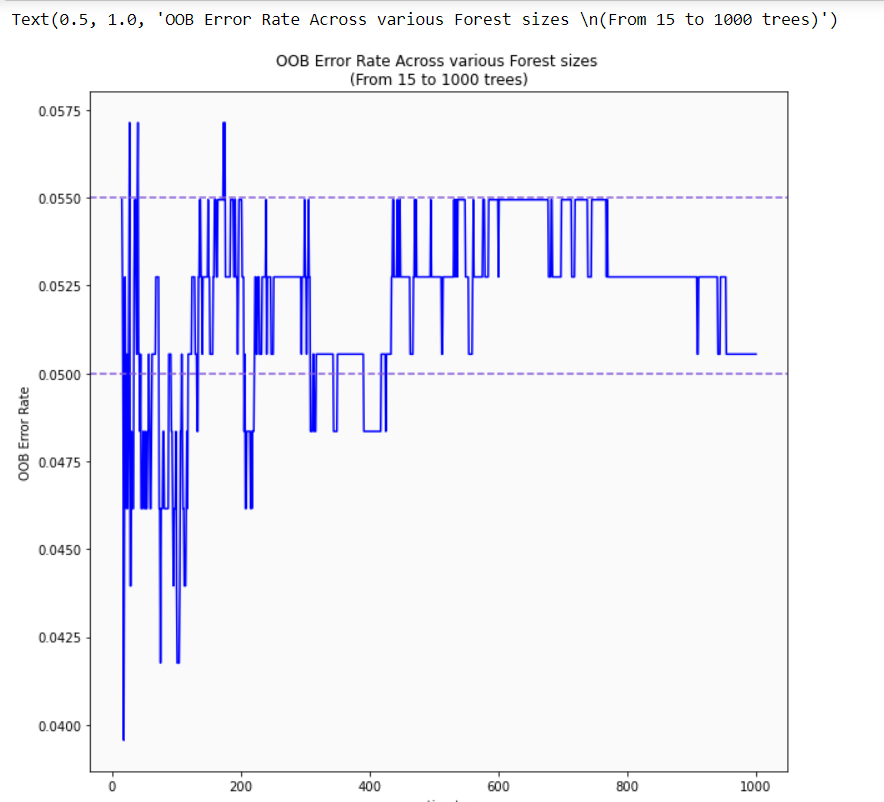


By using the GridSearchCV functionaly, we have created a dictionary with parameter in which we optimize to create the best model for the dataset.

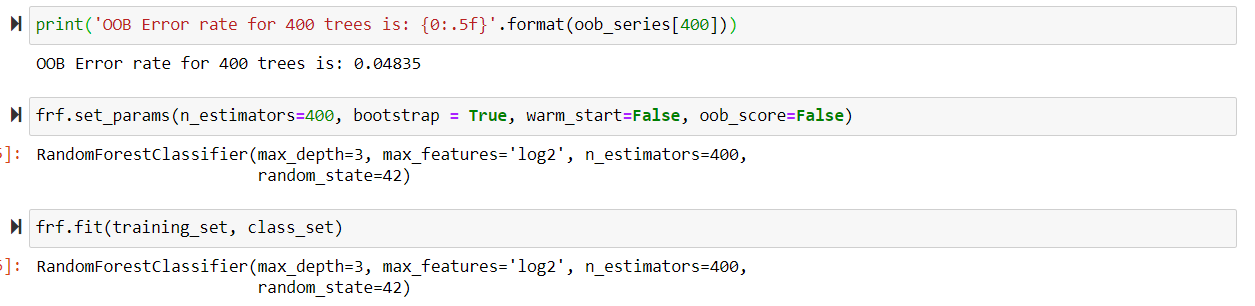




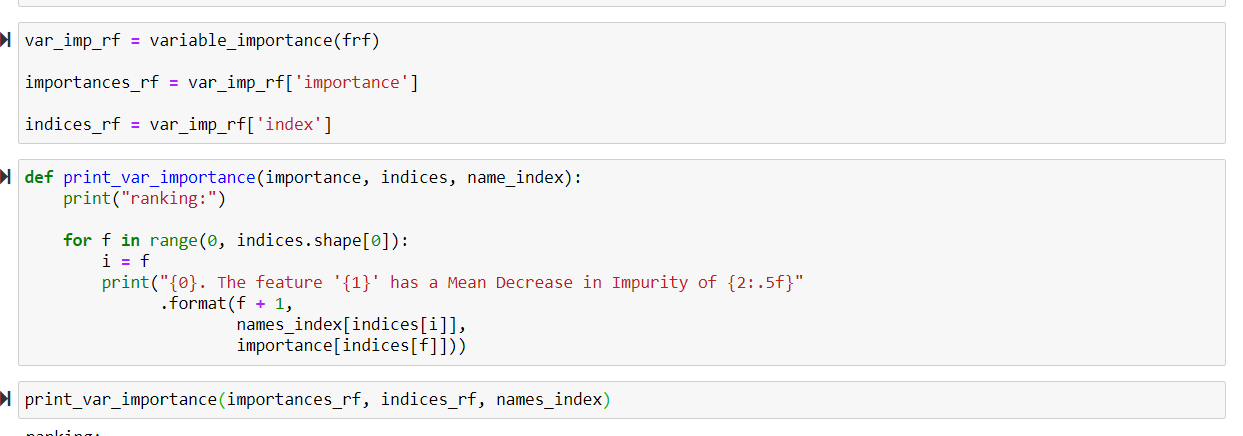


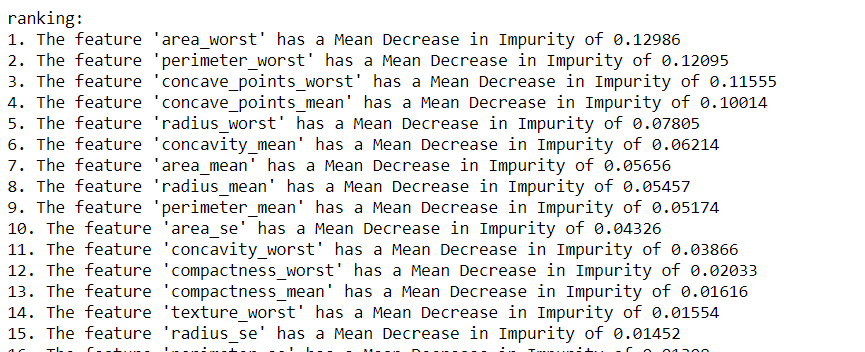


By the plot we can easily observe that the OOB error ratee starts to oscilate at 400 tress , so, we use 400 tress in the random forest.

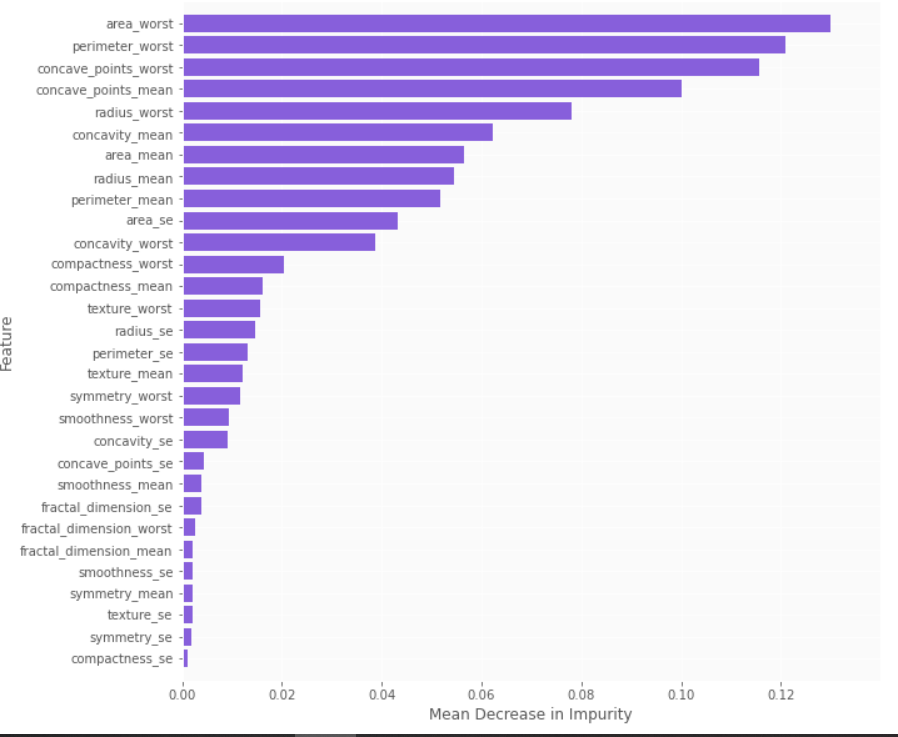






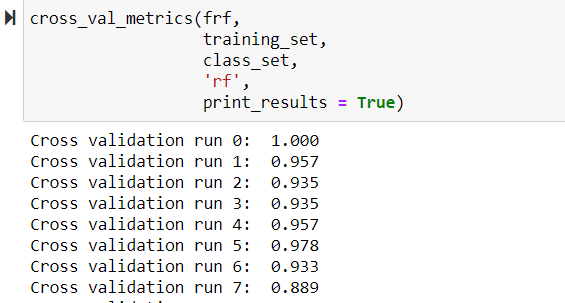






Cross Validation:

Here, I have used k-fold cross validation , and created 10 subsets to our data on which the training and test set methodology.



Conclusion Matrix:

