**PROJECT PART - 2**

**FORECASTING HEART PROBLEMS BY IMPLEMENTING MACHINE LEARNING ALGORITHMS**

**PROJECT TEAM MATES**

700742905 - Mareedu Sai Sritha

700742428 - Nallakalva Rishitha Reddy

700734096 - Nandini Ramshetty

700741157 - Kolanupaka Sai Sumanth

**INDEX**

1. Motivation.................................................................3
2. Significance...............................................................4
3. Objectives..................................................................4
4. Detail design of features........................................5
5. Analysis......................................................................6
6. Implementation.......................................................7
   1. Import libraries..................................................7
   2. Import dataset....................................................8
   3. Understanding the data...................................11
   4. Data Processing.................................................14
   5. Machine Learning Methods............................15
      1. K Neighbors Classifier................................15
      2. Support Vector Classifier............................17
      3. Decision Tree Classifier..............................18
      4. Random ForestClassifier............................20
7. Implementation work status................................21
   1. Work completed.................................................21
   2. Work to be completed......................................22
   3. Concerns..............................................................22
8. References................................................................22

**MOTIVATION :**

In my life time I have seen ‘N’ no.of people suffering from Heart related issues in those my grandma was one among them last year my grandma wasn’t feeling good so we went to hospital for check up at that time i came to know how expensive is echocardiogram [ECG,Ultrasound] and angiogram was. On that day i had a question, can’t we diagnose the heart conditions without these tests in early stage?

That incident made me to pick this project and I would like to develop this with the help of supervised and unsupervised learning algorithms. And deploy a GUI Interface to predict the chances.

**SIGNIFICANCE** :

1. The significance of diagnosing heart disease in preliminary stage will minimize the risk of patients life.
2. Over here, machine learning will is used to diagnose,detect and forecast many problems in the medical industry.
3. The main purpose of this project is to give disorders analyst a tool to identify cardiac problems at an preliminary stage.

**OBJECTIVES** :

1 Machine learning is used to discover patterns from the user data and then make predictions based on trained data. ML helps in analyzing the data and identifying trends.

2.The objective of this project is to check whether the patient is likely to be diagnosed with any cardiovascular heart diseases based on their medical attributes such as gender, age, chest pain, sugar level, etc.

Over here, We collected the dataset from Kaggle ([https://www.kaggle.com/ronitf/heart-disease-uci](https://www.kaggle.com/ronitf/heart-disease-uci" \t "/Users/macbook/Downloads/x/_blank)) .

The datset which we are considering for our project consists of 303 samples in those 13 features with one target variable.

**DETAIL DESIGN OF FEATURES** :

1. Age
2. Sex
3. cp(chest pain)
4. trestbps(resting blood pressure)
5. chol(cholestoral)
6. fbs(Fasting blood sugar)
7. restecg(resting electrocardiographic )
8. thalach(peak heart rate )
9. exang(exercise induced angina)
10. oldpeak(depression induced by exercise relative to rest)
11. slope(slope of peak exercise ST)
12. ca(No.of major vessels)
13. thal
14. target(0:low risk of heart attack ; 1:high risk of heart attack)

**ANALYSIS** :

After analysing the data

1.chest pain = 0 : typical angina

1 : atypial angina

2 : non angina

3 : asymptomatic

2.blood pressure is in mm Hg.

3.cholestrol is in mg/dl

4.bloodsugar >120mg/dl [1=true; 0=false]

1. ecg = 0: normal

1: having ST-T wave abnormality

over here T = inversion

ST = depression >0.05mv

2:shows left ventricular

1. thalach is reaching max heart rate
2. exang is 1 = yes ; 0 = no

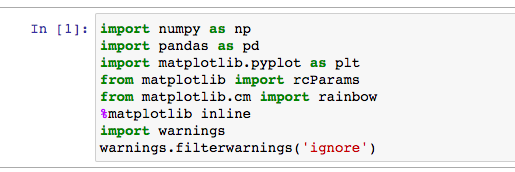
8.thal 3= normal;6 = fixed defect ; 7 = reversible defect

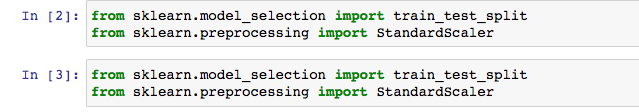
**IMPLEMENTATION** :

With the help of datset and by implementing the machine learning models we will be making predictions on heart disease.

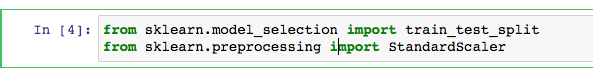
### **Import libraries**

Initially importing all the required libraries. I'll use numpy and pandas to start with. For resolution, We will use pyplot subpackage of matplotlib, using rcParams to add style to the plots and rainbow for colors. For performing Machine Learning models and to process the data, We use the sklearn library.



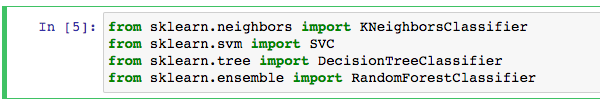


To process the data, We import few libraries. To divide the dataset for testing and training, We are going to use the train\_test\_split method. To scale the features, we use StandardScaler.



Next, We are importing all the ML Algorithms.

1. K - Neighbors Classifier(KNN)
2. Support Vector Classifiers(SVC)
3. Decision Tree Classifier(DTC)
4. Random Forest Classifier(RFC)

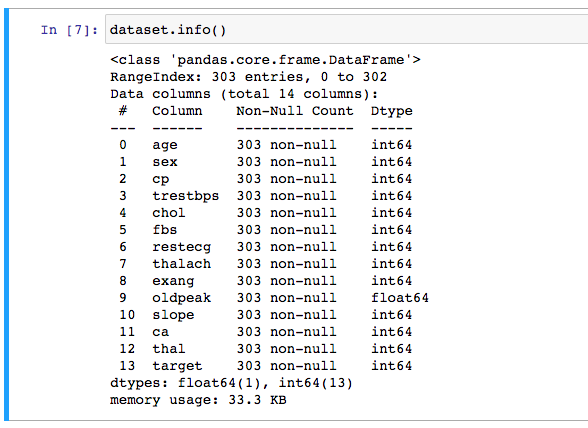


### **Import dataset**

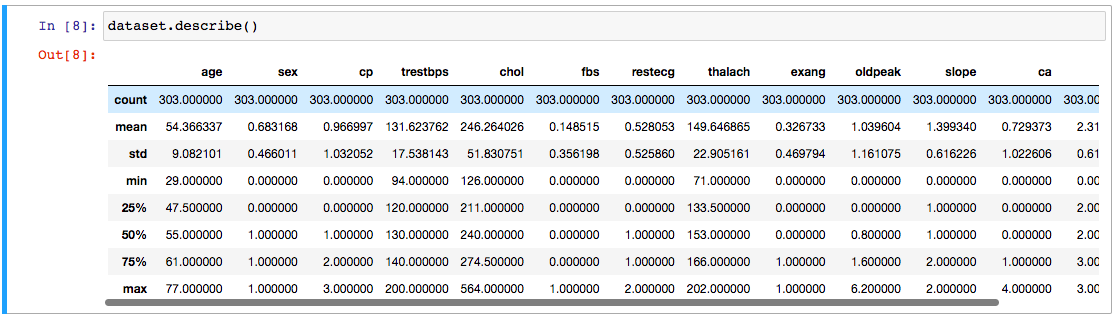
Now we have all the libraries that are needed, Next importing the dataset. The dataset is stored as file dataset.csv. We use pandas read\_csv method to read the dataset.

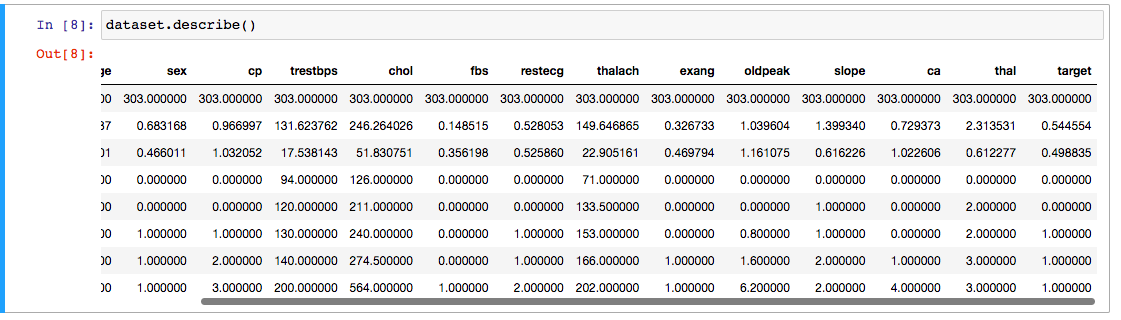
Screen Shot 2022-11-06 at 5.58.46 PM

The dataset is now packed into variable dataset. By using the methods desribe() and info() we can have a glimpse of the data and start processing and visualizing it.



As mentioned earlier, The datset is having 303 rows and there are no missing values. There are total 13 features with one target value which we need to find.

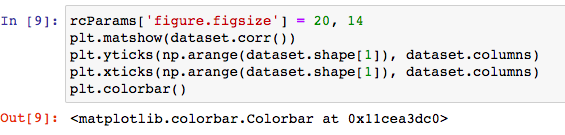


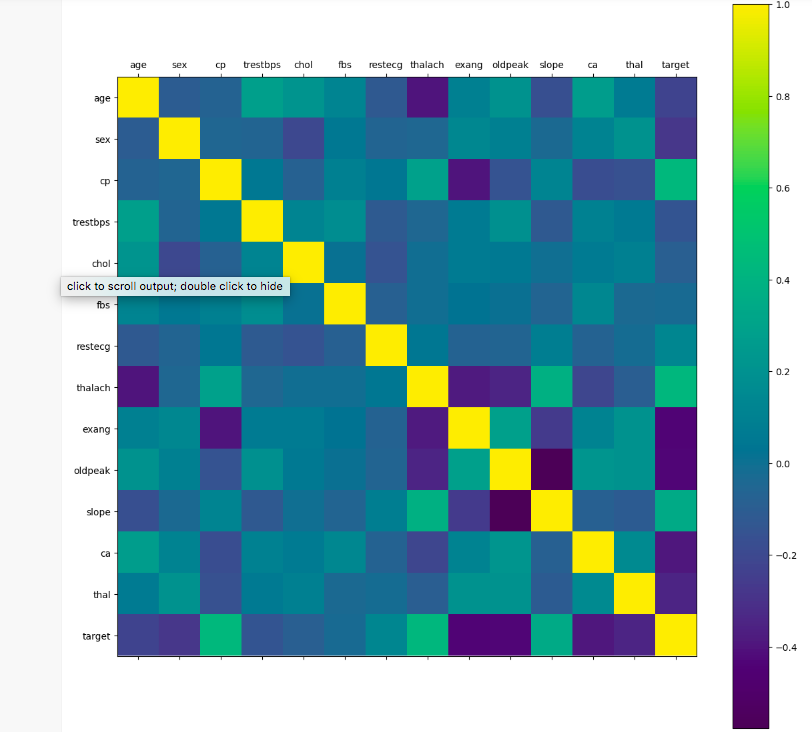


The scale of each featured column is different . Here the maximum for age is 77, the max of chol (serum cholestoral) is 564.

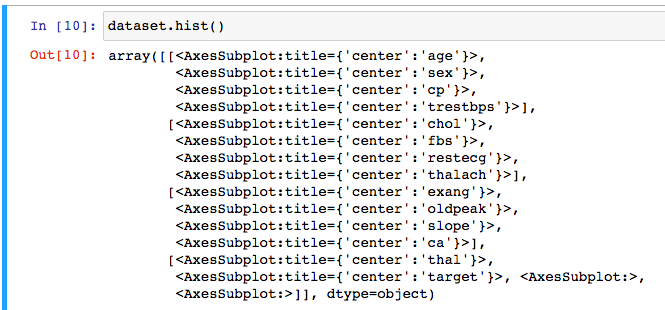
### **Understanding the data**

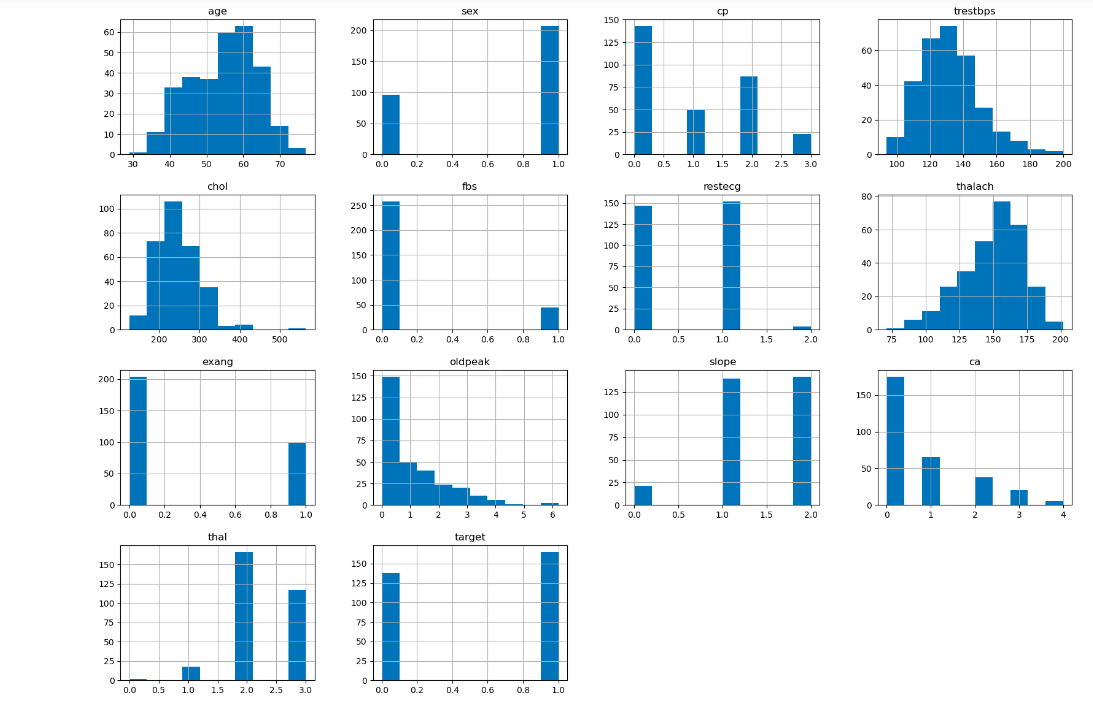
### Now, With the help of visualizations we can understand the data and implement the processings.

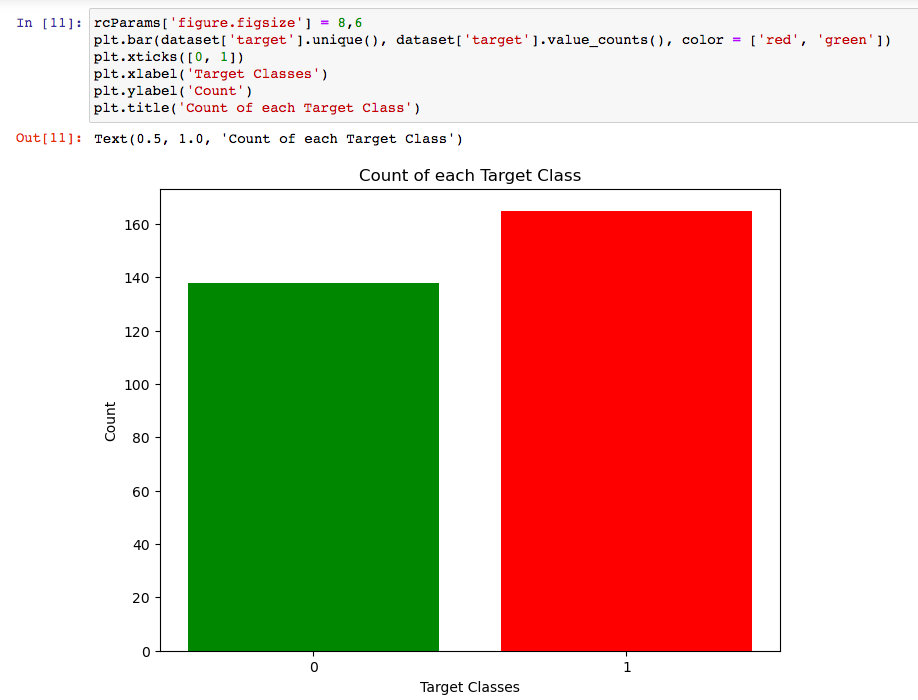




By having a sigh at the correlation matrix above, it's easy to see that few have negative correlation with the target value while some have positive. Next, We can have a look into the histograms for each variable.



By having a sigh at the histograms above, We can come to know that each feature has varied range of distribution. Thus, using scaling before our predictions will be good use. Even, the categorical features ares unique. It's always a great practice to work with dataset where the target classes are of equal size. Thus, let's check for the similar one.



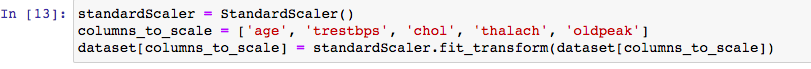
The above two classes are not exactly 50% each but the ratio is better to continue without dropping the data.

### **Data Processing**

After examining the dataset, Here some categorical variables need to be converted into dummy variables and we need to scale the values and train the ML models. Initially, We will use the get\_dummies method to create dummy columns for each categorical variables.

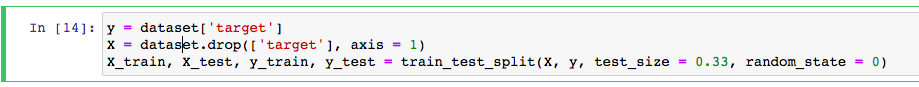
Screen Shot 2022-11-06 at 6.16.45 PM

Now, We will be using the StandardScaler from sklearn to scale the dataset.



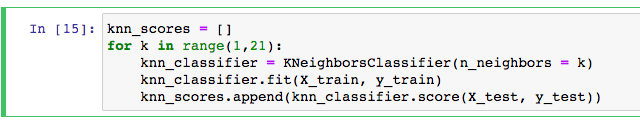
### **Machine Learning Methods :**

Now, By importing  train\_test\_split  we will divide the train and test datasets. Afterthat, We import all Machine Learning models we will be using that to train and test the data.

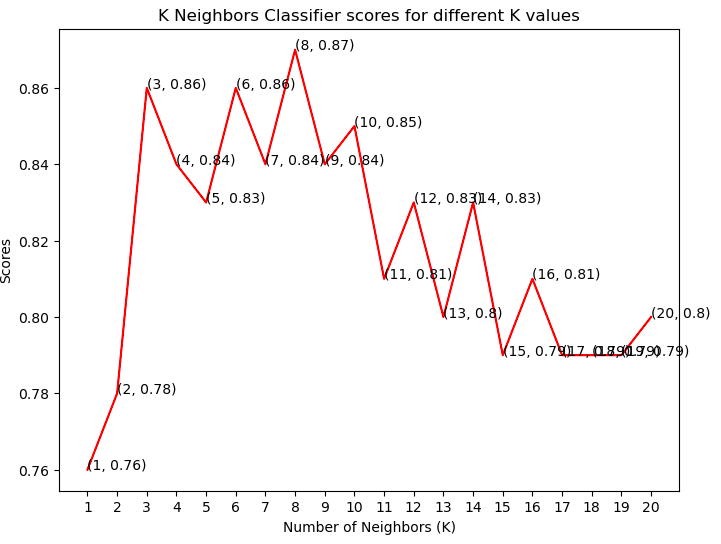
****

#### **K Neighbors Classifier**

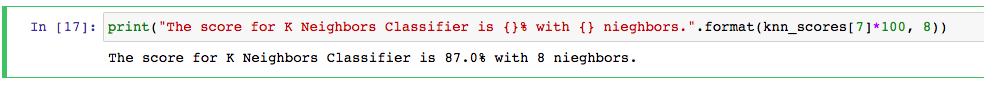
The classified score differs based on varied values of neighbors that we choose. Thus, We will plot a score graph for different K values and continuously check when we can get the best score.



We have the scores for different neighbor values in array knn\_scores. We now plot it and check for which value of K we get the best scores.

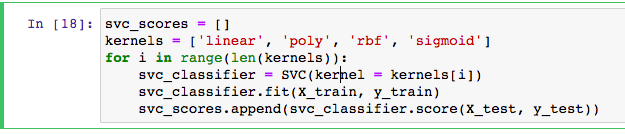


Byt he plot above, it is clear that the maximum score achieved was 0.87 for 8 neighbors.

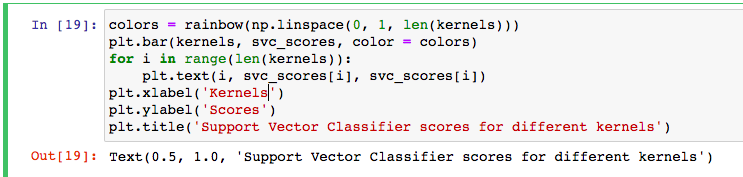
****

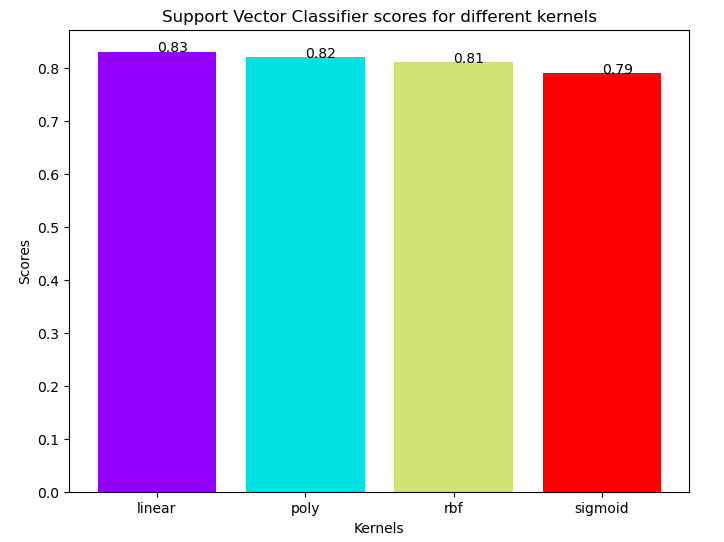
#### **Support Vector Classifier**

There are No.of kernels for Support Vector Classifier. We test some of them and check which has the best score.

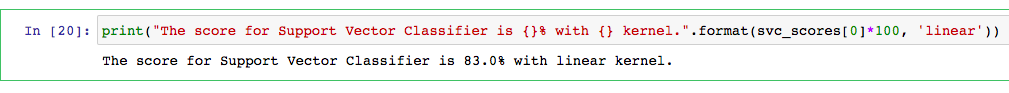


We now plot a bar plot for each kernel and see which got the best.



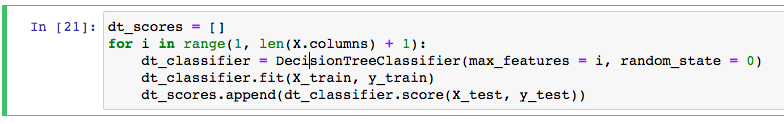


The linear kernel performed the best, which is slightly better than rbf kernel.

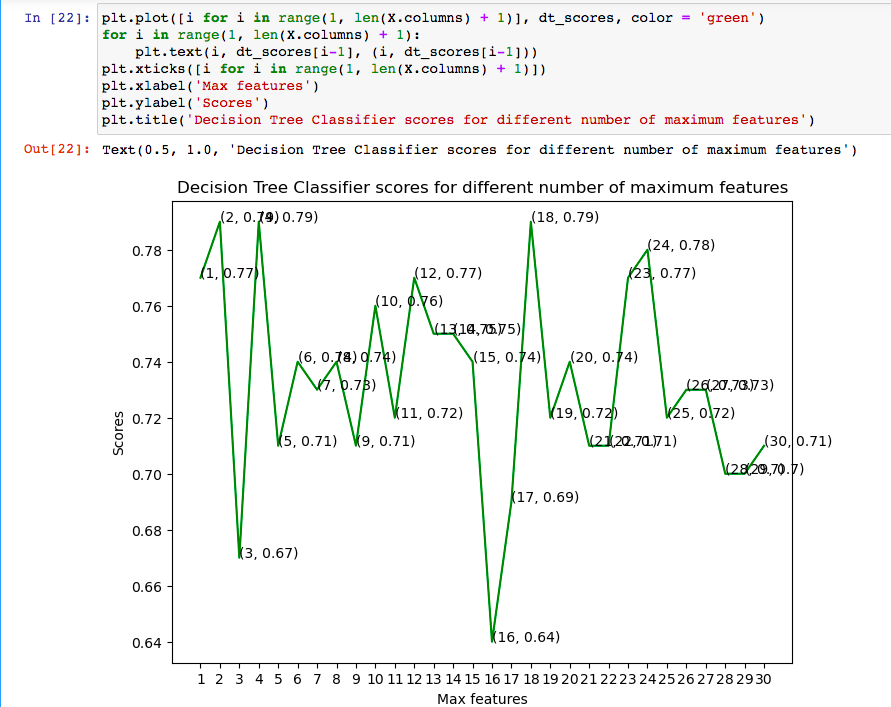


#### **Decision Tree Classifier**

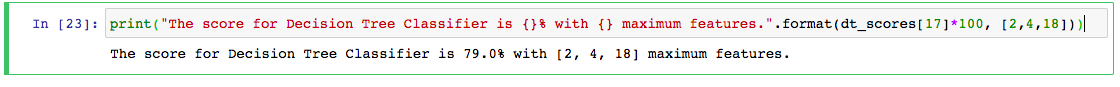
#### Here, We use the Decision Tree Classifier to model the problem. We vary between a set of max\_features and observes which returns the best accuracy.



We picked the utmost number of features from 1 to 30 to split. Here we can observe the scores for each case.

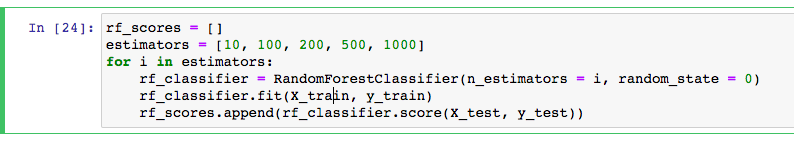


This ML model have attained the good accuracy at three values with maximum features, 2, 4 and 18

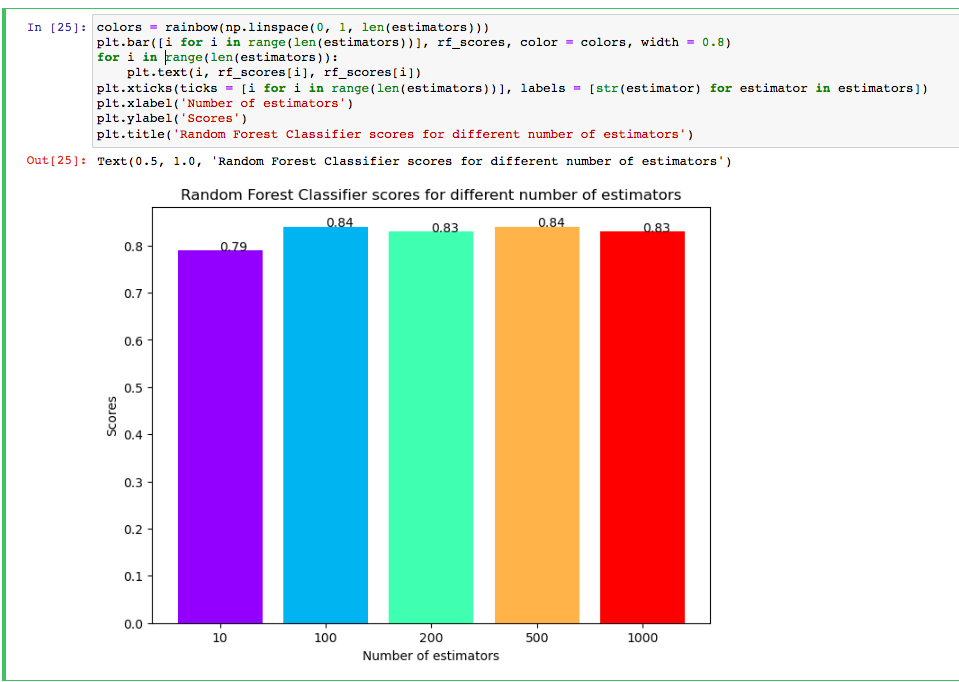


#### **Random Forest Classifiers(rfc):**

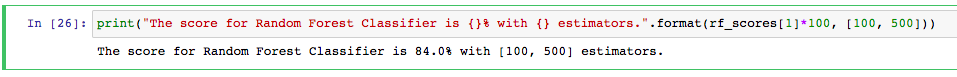
Now, We use the ensemble method, Random Forest Classifier, to create the model and vary the No.of estimators to see the effect.



The model is trained and the scores are noted. Now plotting a bar plot to compare the scores.



The highest score is achieved when the total estimators are 100 or 500.



With the help of ‘dataset.csv’ here are the results which we got by using machine learning algorithms.

1. **IMPLEMENTATION WORK STATUS :**

a) WORK COMPLETED :

\* Implemented **KNN Classifier**, **Suport vector classifier** and **Random forest classifier** to predict the accuracy.

\* By differentiating the above three classifiers we came to know which model is having the best accuracy for heart disease prediction.

|  |  |  |
| --- | --- | --- |
| TASKS COMPLETED | | |
| Action Plan | Team Member | Contribution |
| Collected and analyzed the data. | Sai Sumanth | 23% |
| Prioritized the problems and identified solutions. | Nallakalava Rishitha | 23% |
| Developed the code and done the documentation. | Mareedu Sai Sritha | 31% |
| Tested the classified model and collected reference papers. | Nandini Ramshetty | 23% |

b) WORK TO BE COMPLETED :

\* We would like to extend this project with html page which takes the input values from users and displays the predicted value.

c) CONCERNS :

The accuracy from the current work is around 87% which is not good enough for real time applications. Hence, to improve the accuracy we would like to try few more ML algorithms and also work on processing the data.

1. **REFERENCES**

1 A. Singh and R. Kumar, "Heart Disease Prediction Using Machine Learning Algorithms," 2020 International Conference on Electrical and Electronics Engineering (ICE3), 2020, pp. 452-457, doi: 10.1109/ICE348803.2020.9122958.

2 V. Sharma, S. Yadav and M. Gupta, "Heart Disease Prediction using Machine Learning Techniques," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2020, pp. 177-181, doi: 10.1109/ICACCCN51052.2020.9362842.

3.H. Singh, T. Gupta and J. Sidhu, "Prediction of Heart Disease using Machine Learning Techniques," 2021 Sixth International Conference on Image Information Processing (ICIIP), 2021, pp. 164-169, doi: 10.1109/ICIIP53038.2021.9702625.

4. A. Gavhane, G. Kokkula, I. Pandya and K. Devadkar, "Prediction of Heart Disease Using Machine Learning," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018, pp. 1275-1278, doi: 10.1109/ICECA.2018.8474922.