**“SPECT Heart Disease Prediction”**

A PROJECT REPORT

***Submitted by***

Souvik Roy Chowdhury (CSE 3G, 81)

Anirban Dey (CSE 3G, 82)

**Data Science & Data Analytics Laboratory (CS695A)**

Project Guide - Prof. Sankhadip Chatterjee & Prof. Somarpita Dutta

B. Tech 3rd year

2019



University of Engineering & Management (UEM)

University Area, Plot No. III - B/5, New Town, Action Area - III, Kolkata, West Bengal 700156

**CERTIFICATE**

Certified that this project report “**SPECT Heart Disease Prediction”** is the bonafide work of

1. Souvik Roy Chowdhury
2. Anirban Dey

of B.Tech, CSE, who carried out the project work under our supervision.

………………………………..

………………………………..

**SIGNATURE**

**Examiner:**

**ACKNOWLEDGEMENT**

The completion of this project could not have been accomplished

without the support of our teachers and guide Prof Sankhadip Chatterjee & Prof. Somarpita Dutta. We are thankful to you for allowing us your time to research and write.

We are also very thankful to our respected teachers for their co-operation and suggestion regarding the project work.

Last but not the least we are very thankful to our HOD, Prof. Sukalyan Goswami for giving us an opportunity of doing such an interesting project work.

* **Souvik Roy Chowdhury**
* **Anirban Dey**

**OVERVIEW**

At present, when one suffers from particular disease, then the person has to visit to doctor which is time consuming and costly too. Also if the user is out of reach of doctor and hospitals it may be difficult for the user as the disease can not be identified. So, if the above process can be completed using a automated program which can save time as well as money, it could be easier to the patient which can make the process easier. There are other Heart related Disease Prediction System using data mining techniques that analyses the risk level of the patient.

Disease Predictor is a web based application that predicts the disease of the user with respect to the symptoms given by the user. Disease Prediction system has data sets collected from different health related sites. With the help of Disease Predictor the user will be able to know the probability of the disease with the given symptoms.

As the use of internet is growing every day, people are always curious to know different new things. People always try to refer to the internet if any problem arises. People have access to internet than hospitals and doctors. People do not have immediate option when they suffer with particular disease. So, this system can be helpful to the people as they have access to internet 24 hours.

**INTRODUCTION**

PYTHON:

**Python is a widely used general-purpose, high level programming language. It was initially designed by Guido van Rossum in 1991 and developed by Python Software Foundation. It was mainly developed for emphasis on code readability, and its syntax allows programmers to express concepts in fewer lines of code.**

**Python is a programming language that lets you work quickly and integrate systems more efficiently.**

**There are two major Python versions- Python 2 and Python 3. Both are quite different.**

**Beginning with Python programming:**

**Finding an Interpreter:**

**Before we start Python programming, we needed to have an interpreter to interpret and run our programs. There are certain online interpreters like https://ide.geeksforgeeks.org/, http://ideone.com/ or http://codepad.org/ that can be used to start Python without installing an interpreter.**

JUPYTER NOTEBOOK:

**The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter Notebook is maintained by the people at Project Jupyter.**

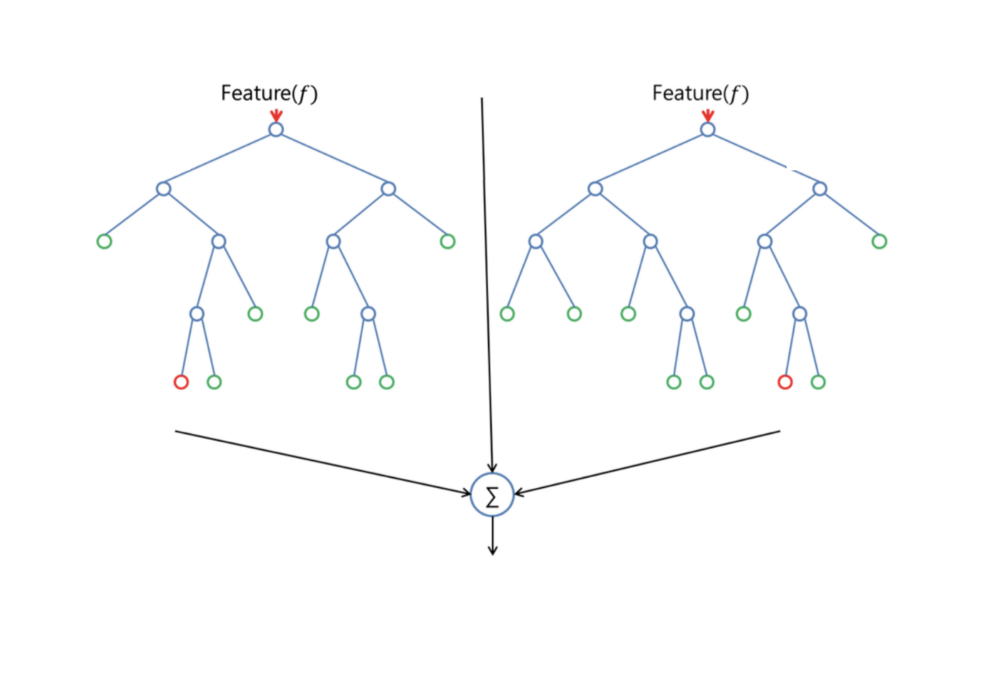
**Jupyter Notebooks are a spin-off project from the IPython project, which used to have an IPython Notebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python, but there are currently over 100 other kernels that you can also use.**

**ALGORITHMS**

RANDOM FOREST:

**Random Forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because it’s simplicity and the fact that it can be used for both classification and regression tasks. In this post, you are going to learn, how the random forest algorithm works and several other important things about it.**

**One big advantage of random forest is, that it can be used for both classification and regression problems, which form the majority of current machine learning systems. I will talk about random forest in classification, since classification is sometimes considered the building block of machine learning.**

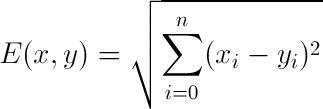
****

Random Forest, only a random subset of the features is taken into consideration by the algorithm for splitting a node. You can even make trees more random, by additionally using random thresholds for each feature rather than searching for the best possible thresholds (like a normal decision tree does).

K-Nearest Neighbors Algorithm:

The k-Nearest-Neighbors (kNN) method of classification is one of the simplest methods in machine learning, and is a great way to introduce yourself to machine learning and classification in general. At its most basic level, it is essentially classification by finding the most similar data points in the training data, and making an educated guess based on their classifications. Although very simple to understand and implement, this method has seen wide application in many domains, such as in recommendation systems, semantic searching, and anomaly detection.

As we would need to in any machine learning problem, we must first find a way to represent data points as feature vectors. A feature vector is our mathematical representation of data, and since the desired characteristics of our data may not be inherently numerical, preprocessing and feature-engineering may be required in order to create these vectors. Given data with N unique features, the feature vector would be a vector of length N, where entry I of the vector represents that data point’s value for feature I. Each feature vector can thus be thought of as a point in R^N.



Support Vector Machine:

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two dimentional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side.



Support Vector Machines are supervised learning models for classification and regression problems. They can solve linear and non-linear problems and work well for many practical problems. The idea of Support Vector Machines is simple: The algorithm creates a line which separates the classes in case e.g. in a classification problem. The goal of the line is to maximizing the margin between the points on either side of the so called decision line. The benefit of this process is, that after the separation, the model can easily guess the target classes (labels) for new cases.

**MAINTENANCE AND SUPPORT**

Corrective Maintenance:

In case of any bugs left in the system, the bugs and issues will be fixed for smooth running of the application. The accuracy of the system can be further improved with other algorithms if needed.

Adaptive Maintenance:

The features in the application can be added such as history of the disease can be kept in the log. The available list of symptoms can also be added for covering more number of diseases.

**CONCLUSION AND RECOMMENDATION**

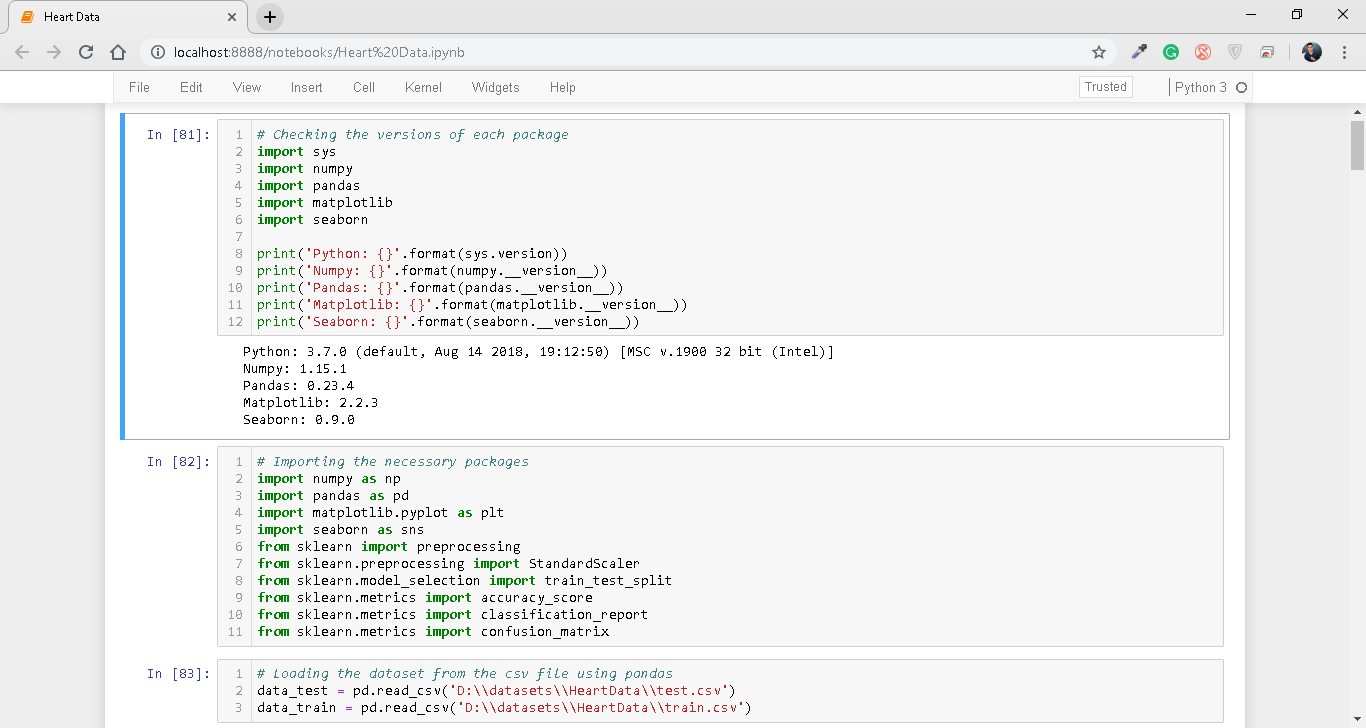
Conclusion:

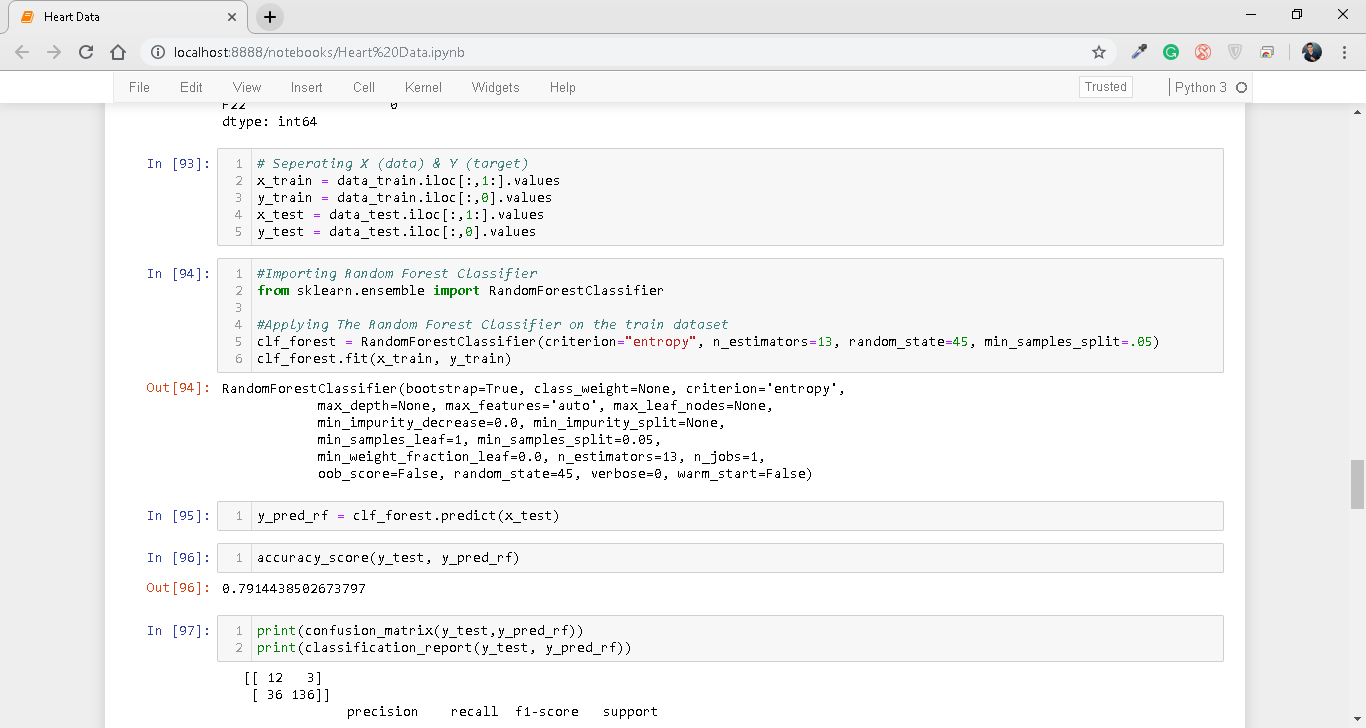
This project aims to predict the disease on the basis of the symptoms. The project is designed in such a way that the system takes symptoms from the user as input and produces output i.e. predict disease. Average prediction accuracy probability of 55% is obtained. Disease Predictor was successfully implemented using grails framework.

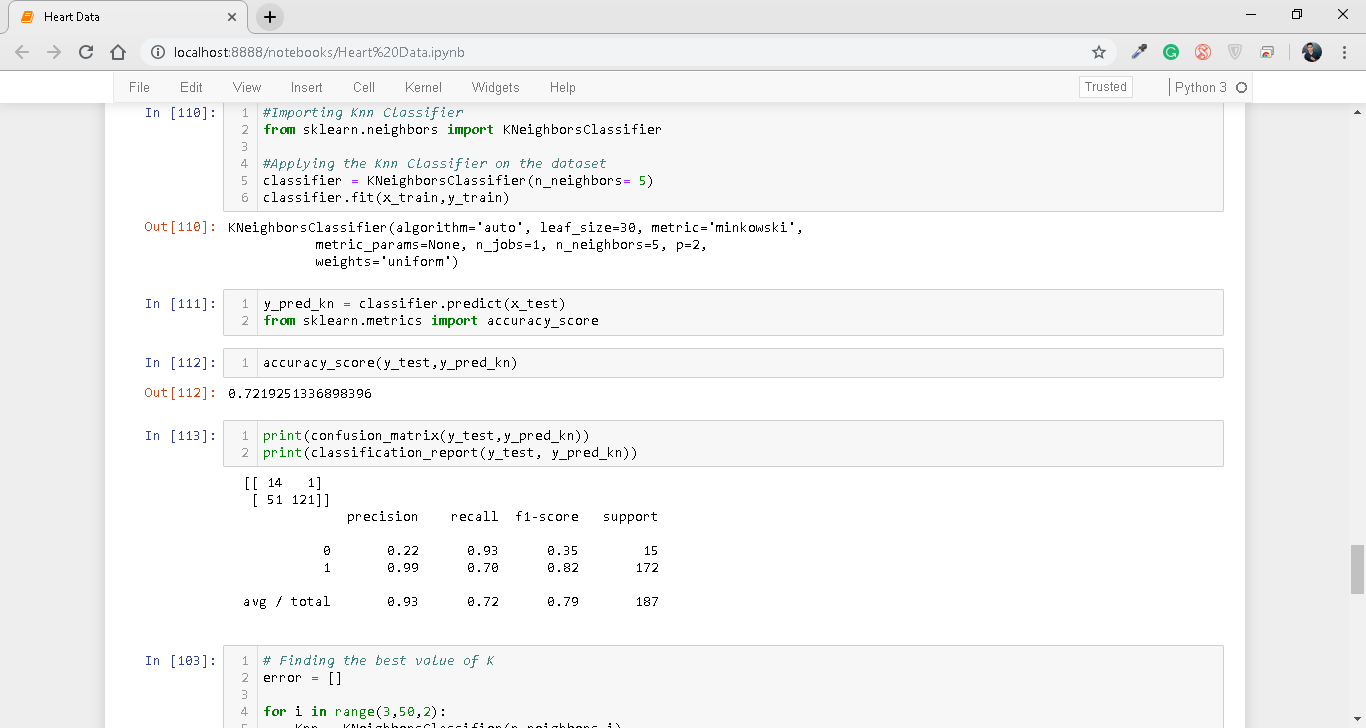
Recommendations:

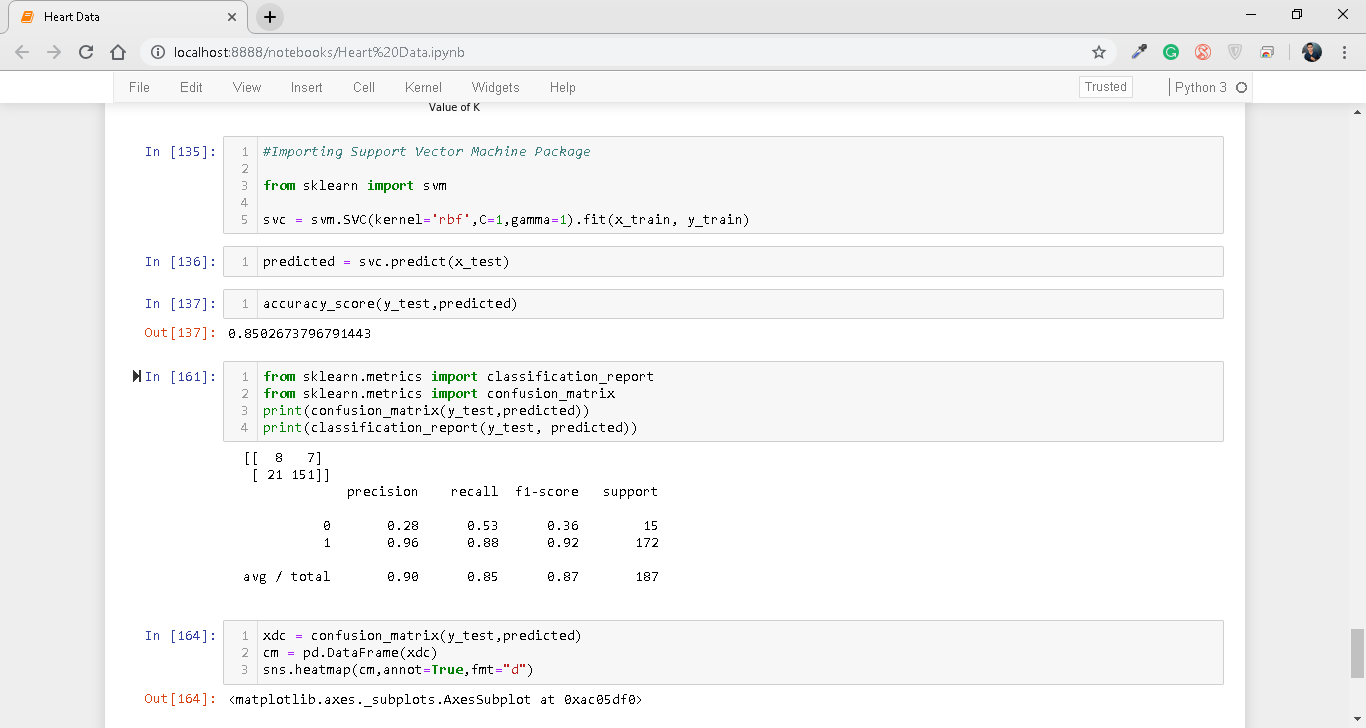
This project has not implemented recommendation of medications to the user. So, medication recommendation can be implemented in the project. History about the disease for a user can be kept as a log and recommendation can be implemented for medications.

**APPENDIX**









**BIBLIOGRAPHY**

The SPECT Heart Data Set was taken from:

<https://archive.ics.uci.edu/ml/datasets.html>

A.Davis, D., V.Chawla, N., Blumm, N., Christakis, N., & Barbasi, A. L. (2008). Predicting Individual Disease Risk Based On Medical History.

Adam, S., & Parveen, A. (2012). Prediction System For Heart Disease Using Random Forest.

Al-Aidaroos, K., Bakar, A., & Othman, Z. (2012). Medical Data Classification With KNN Approach. *Information Technology Journal* .

Darcy A. Davis, N. V.-L. (2008). Predicting Individual Disease Risk Based On Medical History.

JyotiSoni, Ansari, U., Sharma, D., & Soni, S. (2011). Predictive Data Mining for Medical Diagnosis: An Overview Of Heart Disease Prediction with SVM Approach.

K.M. Al-Aidaroos, A. B. (n.d.).

K.M. Al-Aidaroos, A. B. (n.d.). 2012. *Medical Data Classification With Random Forest*.

NisharBanu, MA; Gomathy, B;. (2013). Disease Predicting System Using Data Mining Techniques.