**Project Charter**

**[CARDIOVASCULAR DISEASES PREDICTOR]**

Intel® AI for Youth Bootcamp Batch 2

# Information:

I took the help of my AI instructors and the notebooks they provided in the beginning of the session. I also thank the other fellow AI students who helped me in case of any error. Firstly, I downloaded the dataset from a trusted website or source after extensive research on the problem (cardiovascular diseases a.k.a heart diseases). I tried various models and used the one which gave me the most accuracy after a many different levels of testing. After choosing the right model, I filled in the guidelines on how the person has to use the model and how he/she should expect the output to be.

# Problem scoping

**Background of the problem:**

Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year. CVDs are a group of disorders of the heart and blood vessels and include coronary heart disease, cerebrovascular disease, rheumatic heart disease and other conditions. Four out of 5 CVD deaths are due to heart attacks and strokes, and one third of these deaths occur prematurely in people under 70 years of age. After receiving this data,

I built the **4W** canvas.

Who is getting affected = ALL HUMAN BEINGS ON EARTH

What is getting affected = THEIR HEALTH

Why is the heart of the people getting affected = UNAWARENESS OF GOOD CARDIC HEALTH

Where are they getting affected = HEART

**Proposed solution:**

As **Cardiovascular diseases (CVDs)** is one of the leading causes for deaths all over the world, I thought AI can be brought in to save lives. The model helps a person by giving them info about the issues he/she is facing after the person fills in the information about the symptoms. This is not a final review as it is a computer’s suggestion, all the computer does is it tells you whether you are prone to heart diseases or not. My proposed solution is the:

**CARDIOVASCULAR DISEASES PREDICTOR**

**Reference links:** [Add all the links that you have referred to, to find the problem, and the solution]

|  |  |  |
| --- | --- | --- |
| Sr. No. | Topic | Link |
| 1 | Heart diseases | https://www.who.int/health-topics/cardiovasc ular-diseases#tab=tab\_1 |
| 2 | DATASET | https://archive.ics.uci.edu/ml/datasets/Heart+ Disease |
| 3 | Steps for construction of the model | [https://www.analyticsvidhya.com/blog/2020/0 4/confusion-matrix-machine-learning/](https://www.analyticsvidhya.com/blog/2020/0%204/confusion-matrix-machine-learning/) |
| 4 | Various models | https://towardsdatascience.com/machine-lear ning-basics-with-the-k-nearest-neighbors-algori thm-6a6e71d01761 |

# Data acquisition

**Data features**:

To achieve the solution, we need to extract data from the trusted source such as a government source. A .csv type file with data is the most widely used file extension (comma separated values). After this step it is important to organize, analyze and sort the data. We can also learn information about the data through functions like .info (), describe () which will help in finding out the mean, mode, median, which is known as standard deviation. It is also important to clean the wrong data to give **accurate models**.

**Reference links:**

|  |  |  |
| --- | --- | --- |
| Sr. No. | Topic | Link |
| 1 | Symptoms for the presence of a particular disease | https://www.hindawi.com/journals/cmmm/20 17/8272091/tab1/ |
| 2 | Cardiovascular diseases | https://www.who.int/health-topics/cardiovasc ular-diseases#tab=tab\_1 |
| 3 | Exercise induced angina | https://www.mayoclinic.org/diseases-condition s/angina/symptoms-causes/syc-20369373#:~:te xt=But%20when%20you%20increase%20the,ar teries%20slow%20down%20blood%20flow |
| 4 | Serum Cholesterol | https://www.healthline.com/health/serum-cho lesterol |
| 5 | Blood Sugar levels | https://www.medicalnewstoday.com/articles/3 17536 |
| 6 | ST depression | https://www.ncbi.nlm.nih.gov/pmc/articles/PM C4958709/ |

# Data Exploration

**Information about your data**:

I had collected a dataset from KAGGLE website and verified the source by going to a government website. The dataset provided me with information of over 350 patients with heart problems. I visited other websites to learn more about the symptoms of these diseases so that I get good and accurate information about them.

**Data cleaning:**

I removed the wrong data or the data which did not make sense from the dataset. I also used various functions to handle missing data like the .info (), describe (), dropna () etc. I also arranged the data in ranges so that it gives more accurate results. I replaced the words with numbers to make the code easier to understand.

**Data visualization:**

I plotted many graphs like bar graphs, histograms, line graphs, boxplots etc. to learn, organize, analyze, sort and clean the data. It is important to plot graphs as it gives us a lot of important data on the model we have to construct.

**Important features:**

Data Models ensure consistency in naming conventions, default values, semantics, security while ensuring quality of the data. Data Model structure helps to define the relational tables, primary and foreign keys and stored procedures. There are three types; conceptual, logical, and physical.

# Modeling

**Model Selection:**

The type of model chosen by me for this project was the K-nearest neighbors’ model as this model provided good accuracy. It is an easy model to train, test, implement and also for prediction. It is a model which uses its neighbor’s data to predict the end result.

**Model training:**

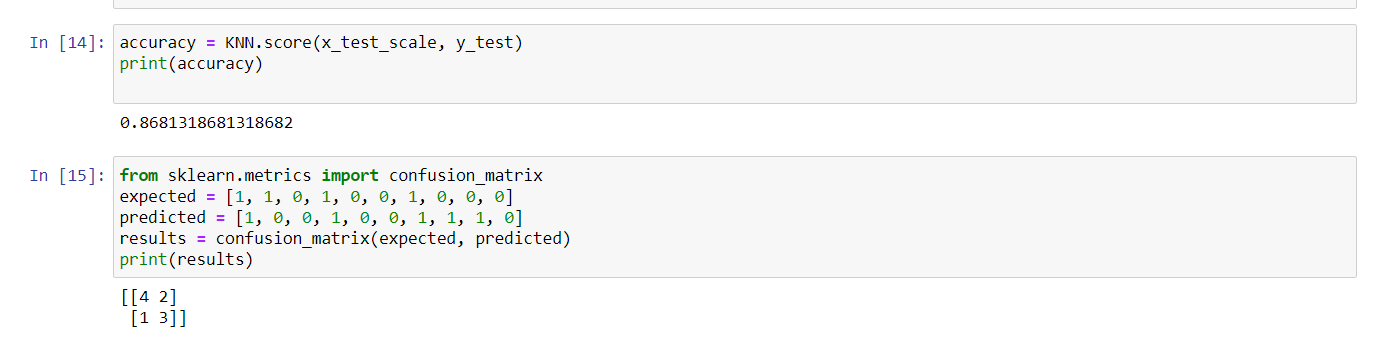
My model is extremely comprehendible. It just predicts whether the person is prone to heart problems or not. It uses an easy type of model (K - nearest neighbors) to predict the end result. The ratio of the test data is **0.3**, the ratio of the training data is **0.7.**

# Evaluation

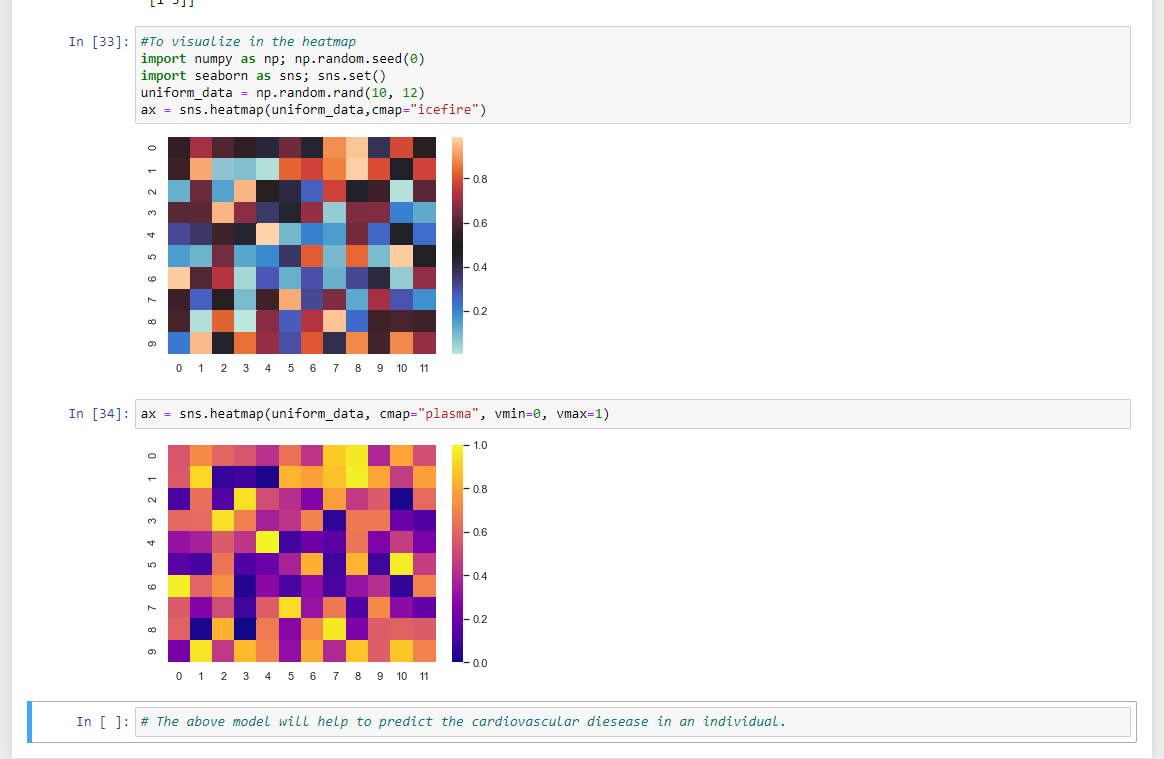
**Model accuracy:**

0.8681318681318682

**CONFUSION MATRIX:**



**HEATMAP:**



# Deployment

**Deployment method:**

The model is built in such a way that the user **MUST** follow the instructions given in the output section first and later enter the data below the “please give the information here” section. The user has to remember that he/she should understand the short forms by reading it from the output section. The user should then fill out information of each asked question inside the bracket carefully. The output would come either 1 or 0. If the output is 1, then symptoms are detected. If the output is 0, then there are no symptoms. The user has to remember to fill in accurate information. Else errors will be detected.