**Heart Attack Possibility Prediction**

**Milestone: Project Report**

Group 18

Student 1 Karan Paresh

Student 2 Tianyu Yang

857-294-7660 (Tel of Karan Paresh)

617-368-0332 (Tel of Tianyu Yang)

[samani.k@northeastern.edu](mailto:samani.k@northeastern.edu)

[yang.tianyu@northeastern.edu](mailto:yang.tianyu@northeastern.edu)

**Percentage of Effort Contributed by Student 1: 50%**

**Percentage of Effort Contributed by Student2: 50%**

**Signature of Student 1: Karan Paresh**

**Signature of Student 2: Tianyu Yang**

**Submission Date: 12/13/2021**

**Problem setting:**

The dataset contains 14 attributes which are responsible for predicting whether the patient may suffer from a heart disease or not. The "target" attribute indicates the presence of heart disease in the patient. It is a discrete value 0 = no/less chance of heart attack and 1 = more chance of heart attack. When the prediction made is accurate, we can not only avoid false diagnosis but also save human resources. Making the correct predictions can solve a lot of trouble. Applying machine learning algorithms for the medical prediction, we will be able to extricate plenty of human resource because we do not need the complex diagnosis procedure in medical sector. The algorithms that will be put into test are Logistic Regression, SVM, Naïve Bayes, Random Forest, ANN to output a binary number 1 or 0.

**Problem definition: -**

Heart disease can be treated right and its symptoms can be condensed with an accurate predicted result and thereby reducing the surgery cost and other medication expenses. The main objective of the project is to predict the possibilities of heart disease explicitly with few records and attributes. Decisions are generally in the data sets and databases which therefore assists the practitioners in making better decisions and in recommending the required treatment.

Data mining holds great latent in medical sector which empowers the health care industry to use the data and analytics to pin down errors and reduce the overall medical lobby spending.

**Data source: -**

The dataset contains 14 attributes which contributes in predicting the target variable, the target variable refers to the presence of heart disease in the patient. It is a discrete valued 0=no/less chance of heart attack and 1=more chance of heart attack.

URL: <https://www.kaggle.com/nareshbhat/health-care-data-set-on-heart-attack-possibility>

**Data description: -**

1. Age:- The age of the people from the dataset.

2. Sex:- The sex of a person and it has has only 2 possible values in this dataset: 1 - Male and 0 - Female.

3. cp(Chest pain type):- This column defines the chest pain severity in scale of 0-4.

— Value 0: asymptomatic

— Value 1: atypical angina

— Value 2: non-anginal pain

— Value 3: typical angina

4. trestbps(Resting blood pressure):- An healthy has has a blood pressure of 80/120 mmHg , person is at high risk if it is above 180mmHg and lesser than 50mmHg.It has continuous values.

5. chol(Serum cholestrol):- The normal cholestrol level of a healthy person is between 125-200 mm/dl and it is considerable till 240 mm/dl and anything greater than this value will cause higher risk of heart attack and this feature has a continuous value.

6. fbs(Fasting blood sugar):- This feature has only two unique values - 1 if FBS is > 120 mg/dl otherwise 0.If the blood sugar is higher than the value mentioned the person is at high risk of getting heart attack.

7. restecg:- resting electrocardiographic results

— Value 0: normal

— Value 1: showing probable or definite left ventricular hypertrophy by Estes’ criteria

— Value 2: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)

8. thalach:- Maximum heart rate achieved and the normal heart rate of a person should be below 100 bpm anything above that is risky and person might tend to have a heart attack.

9. exang(exercise induced angina):- This feature has two values (1 = yes; 0 = no) means if a person faces angina due to exercise than the value is 1 else it is 0.

10. oldpeak(ST depression induced by exercise):- If the range is lesser than 1.5 its a high risk and if it is greter than 1.5 the person is at low risk of getting a heart attack.

11. slope:- the slope of the peak exercise ST segment

0: downsloping;

1: flat;

2: upsloping

when the value is 0 and 2 the possibility of heart attack is high and when it is 1 the chances of getting heart attack is less.

12. ca(number of major vessels colored by flouropsy):- This column has a discrete value from [0-3].

13. thal:- A blood disorder called thalassemia, it has discrete values

Value 1: normal blood flow

Value 2: fixed defect (no blood flow in some part of the heart)

Value 3: reversible defect (a blood flow is observed but it is not normal)

14. target:- This column has a discrete values

0 = less chance of heart attack, 1= more chance of heart attack.