**HEART DISEASE PREDICTION**

**PROJECT REPORT**

**Submitted by**

**ULTRON**

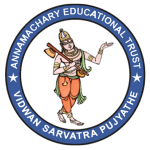
**➢ K.NAVEEN KUMAR**

**➢ T.DILEEP**

***In partial fulfilment for the award of the Certificate***

**of**

**SUMMER INTERNSHIP PROGRAM**

**Department of Computer Science and Engineering**

**Annamacharya Institute of Technology and Sciences**

**Venkatapuram Village , Renigunta Mandal , Tirupati , Andhra Pradesh 517520**

**July 2019.**

### **BONAFIDE CERTIFICATE**

This is to certify that the project entitled ”**HEART DISEASE PREDICTION**” submitted by **K.Naveen Kumar & T.Dileep** in partial fulfilment for the requirements for the award of internship certification in technologies of Machine learning and Deep learning is an authentic work carried out by them under my supervision and guidance.

To the best of my knowledge, the matter embodied in the project report has not been submitted to any other University/Institute for the award of any Degree or Diploma.

### **Signature of Supervisor                                       Signature of Head of the Department**

 Akshay Kumar Kothuri, M.Tech       Mrs.B.Rupa Devi M.Tech,Ph.D.,

AI & IoT Developer Assistant Professor & HOD,

SmartBridge Educational Services PVT LTD Department of CSE,

Hyderabad – 500096 AITS,Tirupati.

Index

1. Title of the project

1.1 Introduction

1.2 Objective of Research

1.3 Problem Statement

1.4 Industry profile

2) Review of literature

3) Data Collection

4) Methodology

4.1 Exploratory data analysis

4.2 Figures and Tables

4.3 Statistical techniques and visualization

4.4 Data modelling and visualization

5) Findings and suggestions

6) Conclusion

**Introduction**

Cardiovascular is a term covering any disorder of the heart. Heart diseases have become a major concern to deal with as studies show that the number of deaths due to heart diseases have increased significantly over the past few decades in India, in fact it has become the leading cause of death in India.

A study shows that from 1990 to 2016 the death rate due to heart diseases have increased around 34 per cent from 155.7 to 209.1 deaths per one lakh population in India.

Thus preventing Heart diseases has become more than necessary. Good data-driven systems for predicting heart diseases can improve the entire research and prevention process, making sure that more people can live healthy lives. This is where Machine Learning comes into play. Machine Learning helps in predicting the Heart diseases, and the predictions made are quite accurate.

Problem Description :

A dataset is formed by taking into consideration some of the information of 920 individuals. The problem is : based on the given information about each individual we have to calculate that whether that individual will suffer from heart disease.

# **Objective of Research :**

The goal of this project is to build a model that can predict the probability of heart disease occurrence, based on a combination of features that describes the disease.

The two inputs of the component are the model and the prediction set. The future result shows the predicted data, actual data, and the probability of different results in each group.

## **Software and libraries used in this project**

# **Python language was used in this project, with many of its libraries listed below:**

# **Pandas: data analysis library Numpy: scientific computing library Sklearn: machine learning library Itertools: library contains functions for creating iterators to use for efficient looping Matplotlib: 2D plotting library Seaborn: statistical data visualization**

# 

# **Problem statement :**

The heart diseases are very common in now a days every 3 people out of 10 are getting cardiovascular diseases so we are taking some random dummies and we will able to find whether the person is suffering from heart disease or not

# **Dataset Collection:**

The dataset consists of 920 individuals data. There are 15 columns in the dataset, however the first column name is not a good parameter as far as machine learning is considered so, there are effectively 14 columns.

# **(age)**

# **(sex) (0 = female, 1 = male)**

# **(cp) cp: chest pain type -- 1: typical angina -- 2: atypical angina -- 3: non-anginal pain -- 4: asymptomatic**

# **(trestbps) trestbps: resting blood pressure (in mm Hg on admission to the hospital)**

# **(chol) chol: serum cholesterol in mg/dl**

# **(fbs) fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)**

# **(restecg) restecg: resting electrocardiographic results -- Value 0: normal -- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) -- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria**

# **(thalach) halach: maximum heart rate achieved**

# **(exang) exercise induced angina (1 = yes; 0 = no)**

# **(oldpeak) oldpeak = ST depression induced by exercise relative to rest**

# **(slope) slope: the slope of the peak exercise ST segment -- Value 1: upsloping -- Value 2: flat -- Value 3: downsloping**

# **(ca) ca: number of major vessels (0-3) colored by flourosopy**

# **(thal) thaldur: duration of exercise test in minutes**

# **(num) (the predicted attribute) num: diagnosis of heart disease (angiographic disease status) -- Value 0: < 50% diameter narrowing -- Value 1: > 50% diameter narrowing (in any major vessel: attributes 59 through 68 are vessels)**

# 

# **Why these parameters:**

Age: Age is the most important risk factor in developing cardiovascular or heart diseases, with approximately a tripling of risk with each decade of life.[24] Coronary fatty streaks can begin to form in adolescence.[25] It is estimated that 82 percent of people who die of coronary heart disease are 65 and older.[26] Simultaneously, the risk of stroke doubles every decade after age 55.

Sex: Men are at greater risk of heart disease than pre-menopausal women.[24][30] Once past menopause, it has been argued that a woman's risk is similar to a man's[30] although more recent data from the WHO and UN disputes this.[24] If a female has diabetes, she is more likely to develop heart disease than a male with diabetes.

Angina (Chest Pain): Angina is chest pain or discomfort caused when your heart muscle doesn't get enough oxygen-rich blood. It may feel like pressure or squeezing in your chest. The discomfort also can occur in your shoulders, arms, neck, jaw, or back. Angina pain may even feel like indigestion.

Resting Blood Pressure: Over time, high blood pressure can damage arteries that feed your heart. High blood pressure that occurs with other conditions, such as obesity, high cholesterol or diabetes, increases your risk even more.

Serum Cholestrol: A high level of low-density lipoprotein (LDL) cholesterol (the "bad" cholesterol) is most likely to narrow arteries. A high level of triglycerides, a type of blood fat related to your diet, also ups your risk of heart attack. However, a high level of high-density lipoprotein (HDL) cholesterol (the "good" cholesterol) lowers your risk of heart attack.

Fasting Blood Sugar: Not producing enough of a hormone secreted by your pancreas (insulin) or not responding to insulin properly causes your body's blood sugar levels to rise, increasing your risk of heart attack.

Resting ECG: For people at low risk of cardiovascular disease, the USPSTF concludes with moderate certainty that the potential harms of screening with resting or exercise ECG equal or exceed the potential benefits. For people at intermediate to high risk, current evidence is insufficient to assess the balance of benefits and harms of screening.

Max heart rate achieved: The increase in the cardiovascular risk, associated with the acceleration of heart rate, was comparable to the increase in risk observed with high blood pressure. It has been shown that an increase in heart rate by 10 beats per minute was associated with an increase in the risk of cardiac death by at least 20%, and this increase in the risk is similar to the one observed with an increase in systolic blood pressure by 10 mm Hg.

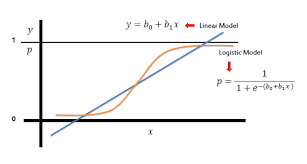
Exercise induced angina: The pain or discomfort associated with angina usually feels tight, gripping or squeezing, and can vary from mild to severe. Angina is usually felt in the centre of your chest, but may spread to either or both of your shoulders, or your back, neck, jaw or arm. It can even be felt in your hands. o Types of Angina a. Stable Angina / Angina Pectoris b. Unstable Angina c. Variant (Prinzmetal) Angina d. Microvascular Angina

ST depression induced by exercise relative to rest :

Peak exercise ST segment: A treadmill ECG stress test is considered abnormal when there is a horizontal or down-sloping ST-segment depression ≥ 1 mm at 60–80 ms after the J point. Exercise ECGs with up-sloping ST-segment depressions are typically reported as an ‘equivocal’ test. In general, the occurrence of horizontal or down-sloping ST-segment depression at a lower workload (calculated in METs) or heart rate indicates a worse prognosis and higher likelihood of multi-vessel disease. The duration of ST-segment depression is also important, as prolonged recovery after peak stress is consistent with a positive treadmill ECG stress test. Another finding that is highly indicative of significant CAD is the occurrence of ST-segment elevation > 1 mm (often suggesting transmural ischaemia); these patients are frequently referred urgently for coronary angiography.

**Logistic Regression:**

Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.



\* This is the graph of the Linear Regression

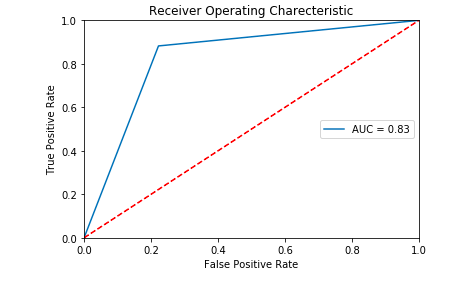
**Methodology :**

We can train our prediction model by analyzing existing data because we already know whether each patient has heart disease. This process is also known as supervision and learning. The trained model is then used to predict if users suffer from heart disease. The training and prediction process is described as follows:

**Statistical Techniques and Visualization**

# 

.



# 

**Exploratory data analysis**

The figures mentioned above is the First data divided into two parts using component splitting. In this experiment, data is split based on a ratio of 80:20 for the training set and the prediction set. The training set data is used in the logistic regression component for model training, while the prediction set data is used in the prediction component

**Data Modelling and Visualization**

**\*Data Modelling with Nodered**

Nodered is a ibm cloud web service which is used to create data flows of which we give like dataset of our project and based on the program code we deploy and then it will be generate a ui. Based on that Graphical user interface we can conclude as our project result

\*Like this we will get the flow connections then we need to deploy it and change the url at the last as ‘red’ to ‘ui’

# 

\* Data Visualization

# \* This is the actual GUI’s of the Nodered flow of our project

**Findings and Suggestions :**

Python regression model has lower error than WEKA model.

It gives good results for simple model.

**In order to improve the model:**

dataset cleaning could be more efficient. According to histogram graph of attributes, they have lots of inconsistent values.

**Conclusion**

**Therefore we used the machine learning techniques based on the python programming language and we predicted the heart disease will effect a particular person or not**