Summer Internship Project Report

on

CARDIOVASCULAR DISEASE PREDICTION

|  |
| --- |
| cid:image001.jpg@01D1FC90.7D09CEE0 |
| **LeadingIndia.ai** |

|  |  |
| --- | --- |
| Submitted To: | Submitted By: Team Number - 2 |
| DR. TANVEER AHMED,  Assistant Professor | V Sreejith, Amrita Vishwa Vidyapeetham |
|  | C Sudarshan , Amrita Vishwa Vidyapeetham |
|  | M Sukul , Amrita Vishwa Vidyapeetham |
|  |  |

Department of Computer Science Engineering

School of Engineering and Applied Sciences

Bennett University

Greater Noida

Uttar Pradesh

**Contents :**

Title Page Number

1. Abstract 3

2. Introduction 4

3. Why we chose this project 4

4. Study of Similar Projects 5

5. Flow Chart 5

6. Snapshots 6

7. Project Development Time Schedule 9

8. Learnings and Reflections from the Project 9

9. Future Enhancements 9

10. References 10

**Abstract:**

The main objective of our project is to predict if a person has a chance of getting a cardiovascular disease based on the person’s medical examination readings (Age ,BMI , Blood pressure , Cholesterol etc). This is a project for predicting potential cardiovascular disease.The significance of heart disease is increasing as the population ages. Existing modes of diagnosis is typically slow and may have undesirable side effects.

The data collected from patients during medical examination contain some hidden information , which is very useful for making effective decisions . For enhancing the results and making efficient decisions on this big data , some advanced ML techniques are used. We develop an efficient cardiovascular disease prediction using ML and DL models for predicting the risk level of cardiovascular disease . The model uses 11 medical parameters such as age, gender, blood pressure, cholesterol, for prediction. The model predicts the likelihood of patients getting heart disease. It derives significant patterns and relationships between the input features and learns the features . After the model is trained it is then used for predicting for the testing data . The obtained results have illustrated that the model can effectively predict the risk level of cardiovascular diseases.

**Introduction :**

The human cardiovascular system is made up of the heart and blood vessels. Cardiovascular disease (CVD) is a set of disorders related to the heart and blood vessels . Annually large number of people die from CVD . Globally it is the number one cause of death.

The most common cardiovascular disease is coronary artery disease, which can lead to heart attacks ,chest pain or stroke . Diagnosing the heart condition of the patient is a challenging task . Heart diseases are diagnosed from the signs , symptoms and physical examination of the patient. There are several factors that increase the risk of heart disease, such as smoking habit, body cholesterol level, obesity, high blood pressure, and lack of physical exercise.

A major challenge faced by health care organizations, such as hospitals and medical centres, is the provision of quality services at affordable costs. Our project aims to predict the presence of cardiovascular disease in a person using important features .

There are 3 types of input features:

* Objective: factual information;
* Examination: results of medical examination
* Subjective: information given by the patient.

The dataset that we used for our project is from Kaggle . It consists of 70,000 records. The dataset includes 11 features like age , gender , height , weight , blood pressure, cholesterol etc which were used to predict whether the person has a cardiovascular disease or not .

The dataset that we used consists of both numerical and categorical data. Before further processing, cleaning and filtering are applied on these records in order to filter the irrelevant data from the dataset.

The proposed system can determine an exact hidden knowledge, ie, patterns and relationships associated with heart disease from a heart disease dataset. It can also answer the complex queries for diagnosing heart disease; therefore, it can be helpful to health care practitioners to make intelligent clinical decisions. Results showed that the proposed system has its unique potency in realizing the objectives of the defined mining goals.

**Why we chose this Project:**

An estimated 17.5 million deaths occur due to cardiovascular diseases worldwide.

More than 75% deaths due to cardiovascular diseases occur in the middle-income and

low-income countries. Also, 80% of the deaths that occur due to CVDs are because of

stroke and heart attack . India too has a growing number of CVD patients added

every year. Currently, the number of heart disease patients in India is more than 30

million. Over two lakh open heart surgeries are performed in India each year. A

matter of growing concern is that the number of patients requiring coronary

interventions has been rising at 20% to 30% for the past few years .

As the risk of having cardiovascular disease is increasing day to day , Effective and efficient automated heart disease prediction models can be beneficial in medical sector for cardiovascular disease prediction. By deploying these automated heart disease prediction model in the medical sector , it will reduce the number of tests to be taken by a patient.

Hence, it will not only save cost but also the time of both, analysts and patients.

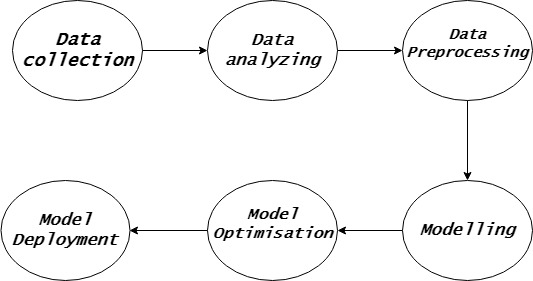
**Study of similar projects :**

We came after some similar cardiovascular disease prediction models :

1. <https://towardsdatascience.com/predicting-presence-of-heart-diseases-using-machine-learning-36f00f3edb2c>
2. <https://www.kaggle.com/neisha/heart-disease-prediction-using-logistic-regression>
3. <https://www.kaggle.com/fabijanbajo/heart-disease-prediction>
4. <https://www.kaggle.com/ronitf/predicting-heart-disease>
5. <https://www.kaggle.com/sonumj/heart-disease-dataset-from-uci>

The above projects are related to cardiovascular disease prediction but are for different dataset.

**Flow Chart :**



**Snapshots:**

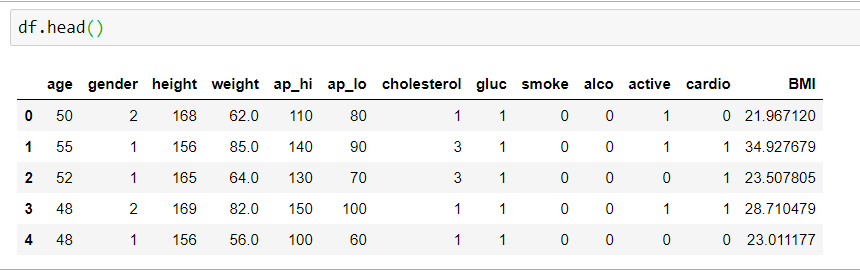


Fig 1: dataset

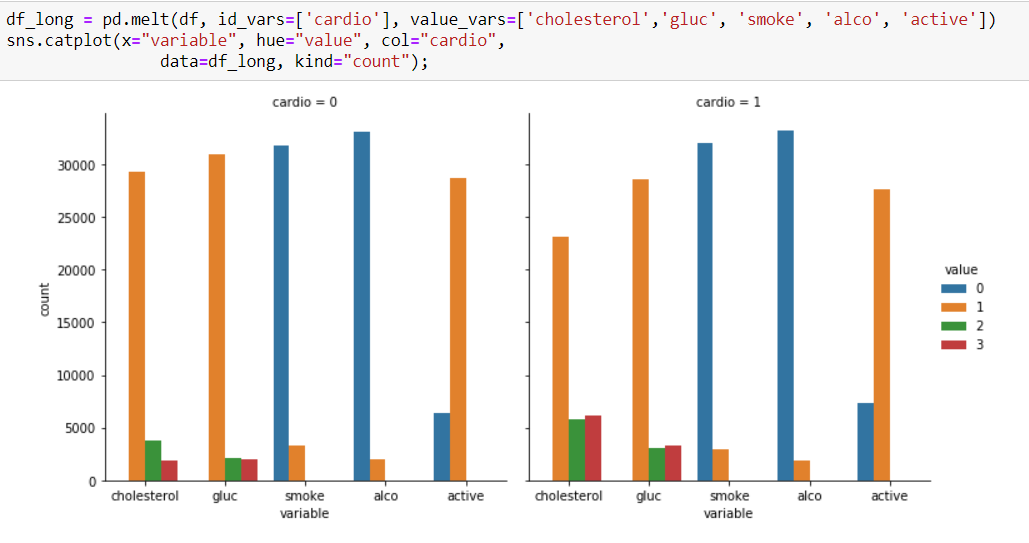


Fig 2 : Statistics based on target variable

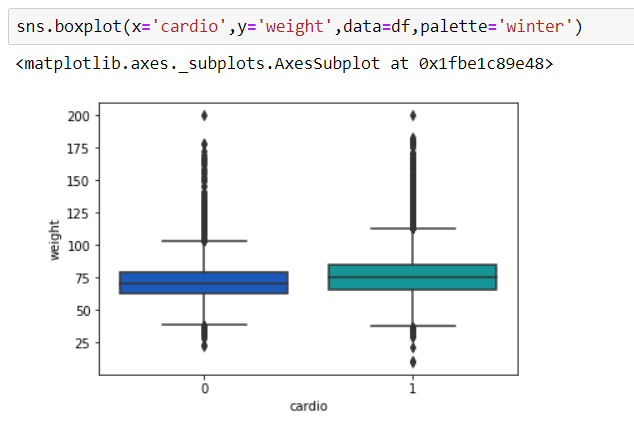


Fig 3 : Box plot

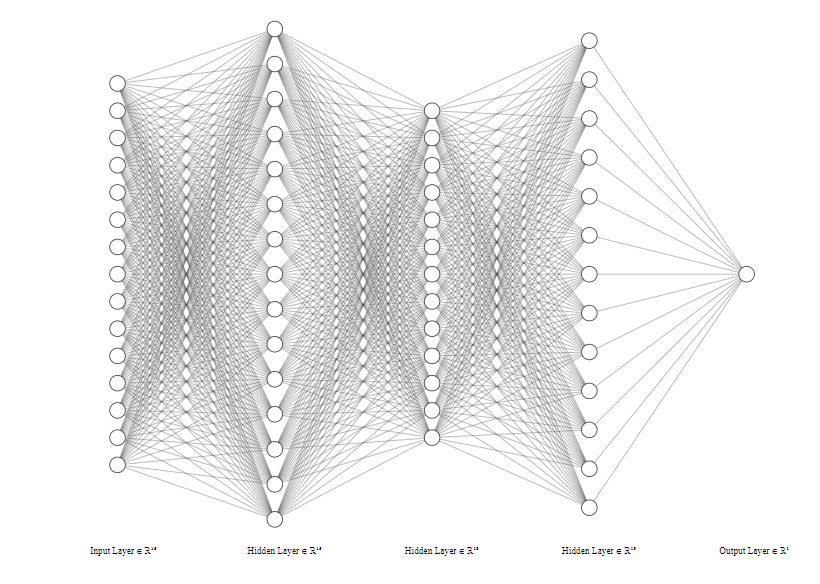


Fig 4 : Neural Network Model

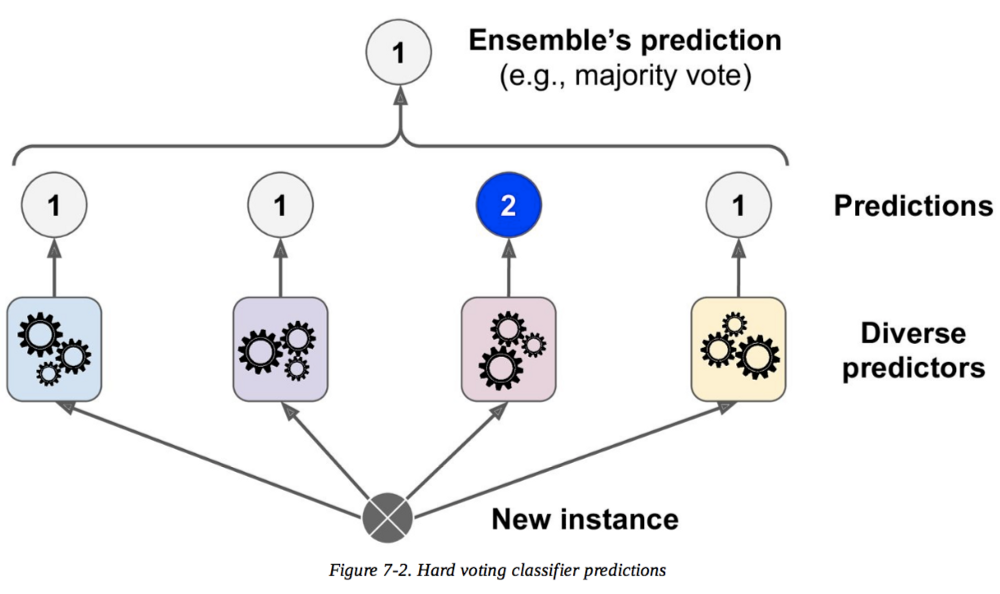


Fig 5 : Max Voting

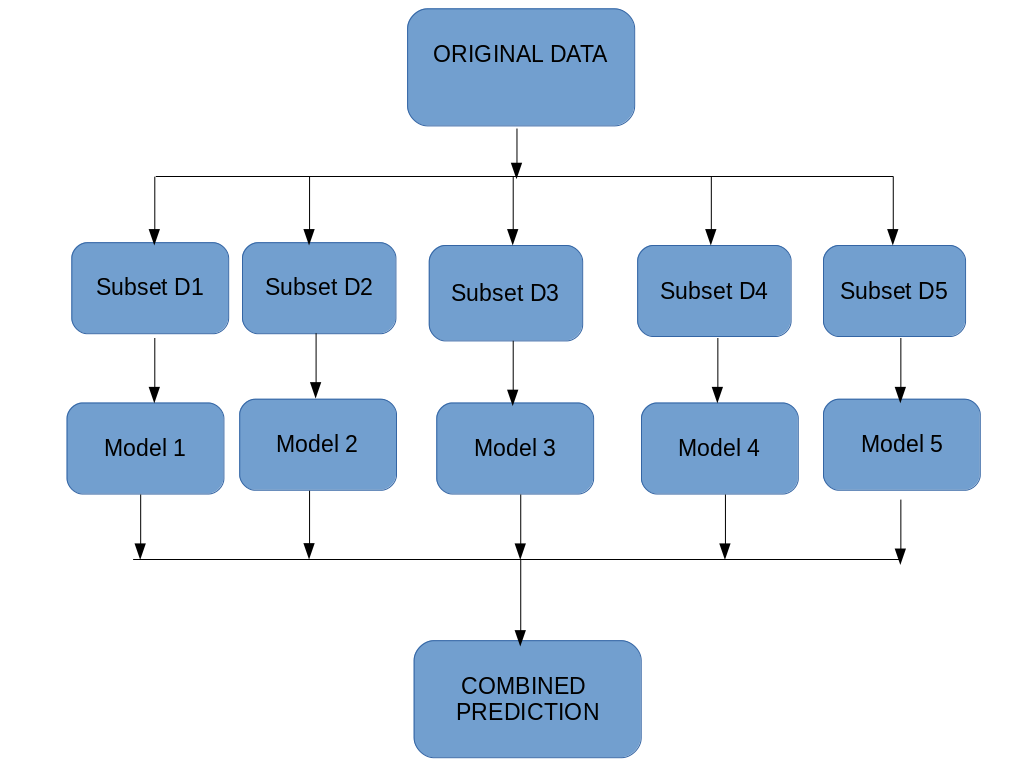


Fig 6 : Bagging

**Project Development Time Schedule:**

1st May 2019 : Project Initial talks and Project finalization

2nd May 2019 - 4th May 2019 : Dataset finalized (Data Collection)

5th May 2019 - 9th May 2019 : Data Analysing and Pre-processing

10th May 2019 - 11th May 2019 : Implementation of Traditional ML classifiers

12th May 2019 - 15th May 2019 : Implementation of Neural Networks and Ensemble

techniques

16th May 2019 - 17th May 2019 : Hyperparameter optimization

**Learnings and reflections from the project :**

It was really a great learning experience working in this project . Our team members were working on Deep learning for the first time so it was hard for us but we were successful to learn about Deep learning and working on it in a project based manner . We were also able to learn about its possibilities in future . Coming from the same university , we knew each other’s strength and were moving forward together as a team . Team work was our key to successfully completing our project .

**Future Enhancement:**

Our current project implemented on static dataset . It can be enhanced with dynamic dataset with real time inputs . Better datasets , more training will lead to much better results and eliminate all possible limitations .

**References :**

[1] <https://archive.ics.uci.edu/ml/datasets/heart+Disease>

[2] <https://www.kaggle.com/ronitf/heart-disease-uci>

[3] <https://physionet.org/>

[4] <https://www.analyticsvidhya.com/>

[5] <https://medium.com/>

[6] [https://www.datacamp.com](https://www.datacamp.com/)

[7] <https://machinelearningmastery.com/>

[8] <https://www.kaggle.com/neisha/heart-disease-prediction-using-logistic-regression>

[9] <https://www.kaggle.com/fabijanbajo/heart-disease-prediction>

[10] <https://www.kaggle.com/ronitf/predicting-heart-disease>

[11] <https://www.kaggle.com/imnikhilanand/heart-attack-prediction>

[12] <https://www.kaggle.com/sonumj/heart-disease-prediction-with-ann>

[13] <https://www.kaggle.com/gauravsinghania/heart-disease-prediction>

[14] <https://www.kaggle.com/sulianova/eda-cardiovascular-data>

[15] <https://www.kaggle.com/pratiq/initial-cardiovascular-disease-predition>

[16] <https://www.kaggle.com/manoj1993007/cardiovascular-disease-analysis-and-prediction>