
Project Summary

Batch details	PGPDSE-FT Online DEC21
Team members	RAVI SHANKAR A V, TINTU B JOHN, FEBIN THOMAS, BASKAR D, ABDUL RAHMAN P J
Domain of Project	HEALTH CARE ANALYTICS
Proposed project title	CARDIOVASCULAR DISEASE PREDICTION
Group Number	6
Team Leader	RAVI SHANKAR A V
Mentor Name	Vibha Santhanam

Date: 28/06/2022

Signature of the Mentor

Signature of the Team Leader

Table of Contents

SI NO	Topic	Page No
1	Overview	3
2	Business problem goals	3
3	The topic survey in depth	4
4	Critical assessment of topic survey	4
5	Methodology to be followed	5
6	References	6

Project Details

OVERVIEW

Business problem statement (GOALS)

1. Business Problem Understanding

Heart disease is one of the leading causes of death for people in our country. Public health estimates indicate that India accounts for approximately 60% of the world's heart disease burden. About half of all Indians have at least 1 of 3 key risk factors for heart disease: high blood pressure, high cholesterol, and smoking. Other key indicator includes diabetic status, obesity (high BMI), not getting enough physical activity or drinking too much alcohol. We need a system to predict a patient's heart condition using these factors

2. Business Objective

To create a solution for predicting a patient's heart condition using the key indicators from the acquired data

3. Approach

Build a machine learning model to accurately detect "patterns" from the data that can predict a patient's heart condition using the key indicators.

4. Conclusion

This project will provide a deep insight into machine learning techniques for the classification of heart diseases. The existing techniques are studied and compared for finding efficient and accurate systems. Machine learning techniques significantly improve the accuracy of cardiovascular risk prediction through which patients can be identified. A model to predict a patient's heart condition is developed.

TOPIC SURVEY IN BRIEF

1. Problem understanding

Heart disease is the number one cause of mortality and a silent epidemic among Indians. The annual number of deaths from CVD in India is projected to rise from 2.26 million (1990) to 4.77 million (2020). Due to several contributing factors, such as age, chest pain, diabetes, cholesterol fluctuation, exhaustion, and many others. An early diagnosis of such disease has been sought for many years, and many data analytics tools have been applied to help health care providers identify some of the early signs of HD. Many tests can be performed on potential patients to take the extra precautions to reduce the effect of having such a disease

2. The current solution to the problem

Patients take the test to determine the condition of their heart but these tests just tell us if they currently have heart disease. These tests do not predict the likelihood of them getting the heart disease in the future based on their key indicators

3. The proposed solution to the problem

Predict the likelihood of getting the heart disease based on the key indicators which will give us a more accurate result and suggest a change to help them get into a healthy lifestyle

4. Reference to the problem

[Lifestyle changes to lower heart disease risk - Harvard Health](#)

[Know Your Risk for Heart Disease | cdc.gov](#)

[Lifestyle Changes for Heart Attack Prevention | American Heart Association](#)

CRITICAL ASSESSMENT OF TOPIC SURVEY (50-100 words)

1. Find the key area and gaps identified in the topic survey where the project can add value to the customers and business

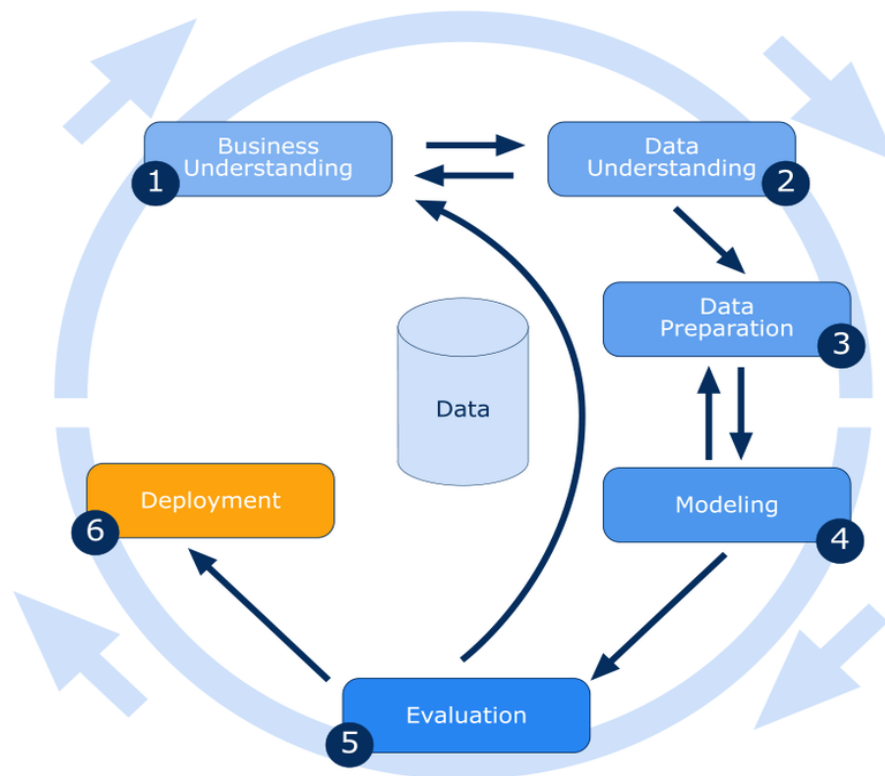
Heart disease is one of the most significant causes of mortality in the world today. Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis. Machine learning (ML) is effective in assisting in making decisions and predictions from the large quantity of data

produced by the healthcare industry. We have also seen ML techniques being used in recent developments in different areas of the Internet of Things (IoT).

2. What key gaps are you trying to solve?

In this project, we propose a novel method that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. The prediction model is introduced with different combinations of features and several known classification techniques.

METHODOLOGY TO BE FOLLOWED (Explain each step from 1-5)



1. Business Understanding:

India accounts for approximately 60% of the world's heart disease burden, despite having less than 20% of the world's population. Heart disease is the number one cause of mortality and a silent epidemic among Indians. A system is needed to predict the heart disease based on the key indicators

2. Data Understanding:

- The dataset has 319796 rows and 18 columns
- The target variable is the **HeartDisease** column which has two unique values:
 - 1 represents **YES**
 - 0 represents **NO**
- The dataset provides us with the various factors that lead to Heart Disease such as BMI, Smoking, Alcohol Drinking, Stroke, Physical Health, Mental Health, Diff Walking, Sex, Age Category, Race, Diabetic, Physical Activity, Gen Health, Sleep Time, Asthma, Kidney Disease, Skin Cancer

3. Data Preparation:

- Understanding the dataset and determining the target variable
- To check for null values and replace them with appropriate
- To check for outliers and treat them with appropriate methods
- To perform various EDA techniques like feature engineering, categorical encoding and ordinal encoding to get the ideal dataset for the machine learning technique used.

4. Modelling:

- After pre-processing of data, implement various machine learning models on the train data set to find which model fits the best to predict if a patient will have a Heart Disease or not.
- Out of the tested models, the one with the best accuracy is chosen and implemented on the test data set.

5. Evaluation:

An algorithm can be trained using the data and be used to generate more accurate predictions to identify the **risk of heart disease** in a patient based on the probabilistic label assigned to each feature variable by the algorithm.

Patients can be assigned a class "**Risk_of_heartdisease**" based on the predicted label such that:

- o **Low-risk** for patients with low chances of getting heart disease
- o **High-risk** for patients with High chances of getting heart disease

REFERENCES

https://www.cdc.gov/heartdisease/risk_factors.htm

https://www.cdc.gov/heartdisease/risk_factors.htm

Declaration:

This is to declare that the dataset that we are using for our capstone project does not have any relevant legality associated with it and can be used to showcase the work we do on it as a presentation in Great Learning.