Heart Disease Prediction Using Machine Learning as a Problem Solving Innovation

Anupam Garain Kashyap[1], Prof.Ramesh Kumar[2]

[1,2] Department of Computer Science & Engineering, K.K. University, Berauti, Nepura, Biharsharif, Nalanda, Bihar, 803115, India

Abstract:

The main aim of our project is to provide a solution to people who are unaware of multiple cardiac issues faced by them. We have utilized the technical advantages of Machine Learning which has allowed us to analyze datasets and furthermore seek multiple algorithms that has allowed us to reach to most optimized and efficient results. We focus on providing significant advantages in health industry with the perfect amalgam of our technical expertise to ensure that each person gets a chance to be aware of cardiac issues faced by him.

Heart diseases contribute to a greater section of global deaths which is highly visible in the death index and it has been noted that majority of people were unable of their cardiac diseases. These cardiovascular diseases have been the cause of death for almost 17.9 million deaths as per 2019 census. This is highly believed that almost 1/3rd of deaths are caused due to health negligence and unable to comprehend the common causes of cardiovascular diseases. In low income or developing countries this issue is a major concern and one of the most prominent reasons for the same is their eating habits and lifestyle which has serious long term effects. Nowadays people are using smart watches that allow them to track their breath rate and heart rate but without a reliable portal that allows them to analyze the gathered information it just remains a data sets. Without proper management and medical care these diseases have serious concerns in the long term which could affect the lifespan of the person. Heart stroke and regular chest pains are more common issues then they are understood which is the major reason why a person must keep a track of his cardiac health.

Keywords: Machine Learning, Cardiovascular Diseases, heart disease prediction, Logistic Regression, Random Forest.

Brief Introduction To The Problem

Heart diseases are one of the major reason for rising global deaths and to cope up with the same the medical industry has been seeking for a reliable solution. The cardiovascular diseases are treatable but only when they are treated at the right time and a majority of people are unaware of their cardiovascular diseases. So we wish to solve the issue without innovative approach and touch of Machine Learning. We focus on developing a project which would predict the cardiovascular health issues based on the parameters provided by the users.

Methodology

Here we would be discussing about series of concepts utilized in making our project successful and efficient. Multiple stages of project implementation are discussed in detail with systematic management and functionality of each process.

Data Collection

In the initial process of data collection we believe in fetching data from the data sets which would be utilized by our application to generate results. These data sets are arranged and organized based on multiple variables provided by users which include health details like heartbeat, breath rate and lot more. This initial stage of data collection is the base step in which the users must ensure that the data entered is accurate to achieve the most efficient results.

Exploratory Data Analysis

Once the data is collection and starts processing in the manner of data sets, now a pipeline structure of data processing and management begins in which the data is furthermore bundled down for use. The primary data analysis starts working on the data sets which is streamlined using multiple algorithms for the same.

Data Cleaning

The process of data input is a cumbersome task and there are multiple instances when users can enter faulty information or the information entered does not match the default format. In such instances this process of methodology ensures that incorrect data is sorted and filtered out which makes it easier to increases program efficiency.

Feature Engineering

This process of the procedure involves converting the raw information provided by the user to the information which can be assessed by the system that would allow them to filter out the required information. This is an integral part of the process and use of multiple algorithms has been implemented in this step.

Model Selection

There are multiple predefined data models created by us in this program and based on the data sets the most suitable and significant model is selected by our program. This model has multiple criteria and parameters which are selected and sorted based on the user input.

Model Training

Once done with the model selection then the requisite details and code snippets are loaded in the active memory and the process of model training begins in which data is processed in the programs models to enhance accuracy. These training model sets provide most fluent byproducts to check the precision.

Evaluate Model

This part of the procedure involves thorough check to ensure that code works reliably and efficiently and for the same it has three simple parameters which are Precision. Recall and Accuracy.

Hyperparameter Tuning

This is a mode of working in which multiple trials run simultaneously with no effect on the latency and turnaround time experienced by the users which using the application.

Prediction and Model Deployment

This is final phase in which the final products are achieved and the final model is set into production so it can be utilized for vivid purposes.

Literature Review

1. Shah, Devansh, Samir Patel, and Santosh Kumar Bharti. "Heart disease prediction using machine learning techniques." SN Computer Science 1.6 (2020): 1-6. In this paper a thorough research was conducted that allowed a deep concept of understanding for multiple cardiac issues faced by issues. Set of multiple datasets were utilized in the process which made work much easier and streamlined.

2. Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." IEEE access 7 (2019: Artificial Intelligence is the most innovative tool which makes it easier for user to understand multiple results based on the data sets provided to the chatbot. With efficient working methodology it became easier to comprehend heart related concerns.

3. Rajdhan, Apurb, et al. "Heart disease prediction using machine learning." International Journal of Research and Technology. In this paper the final research was concluded based on the user input and multiple data models stored in the system which were trained to ensure ease of output. With well established algorithms and working it became easier to predict heart related diseases
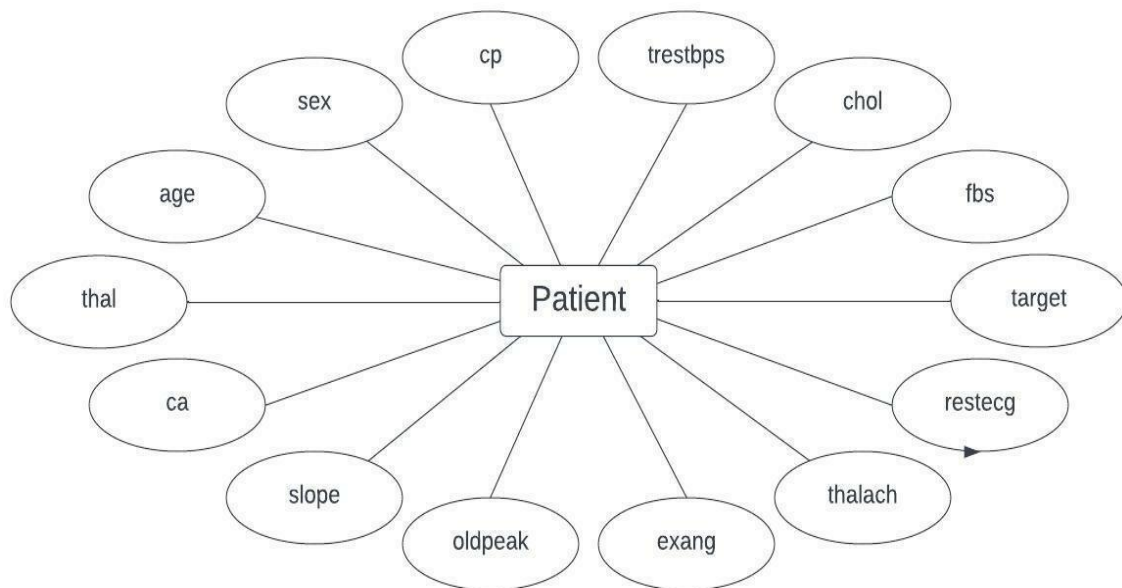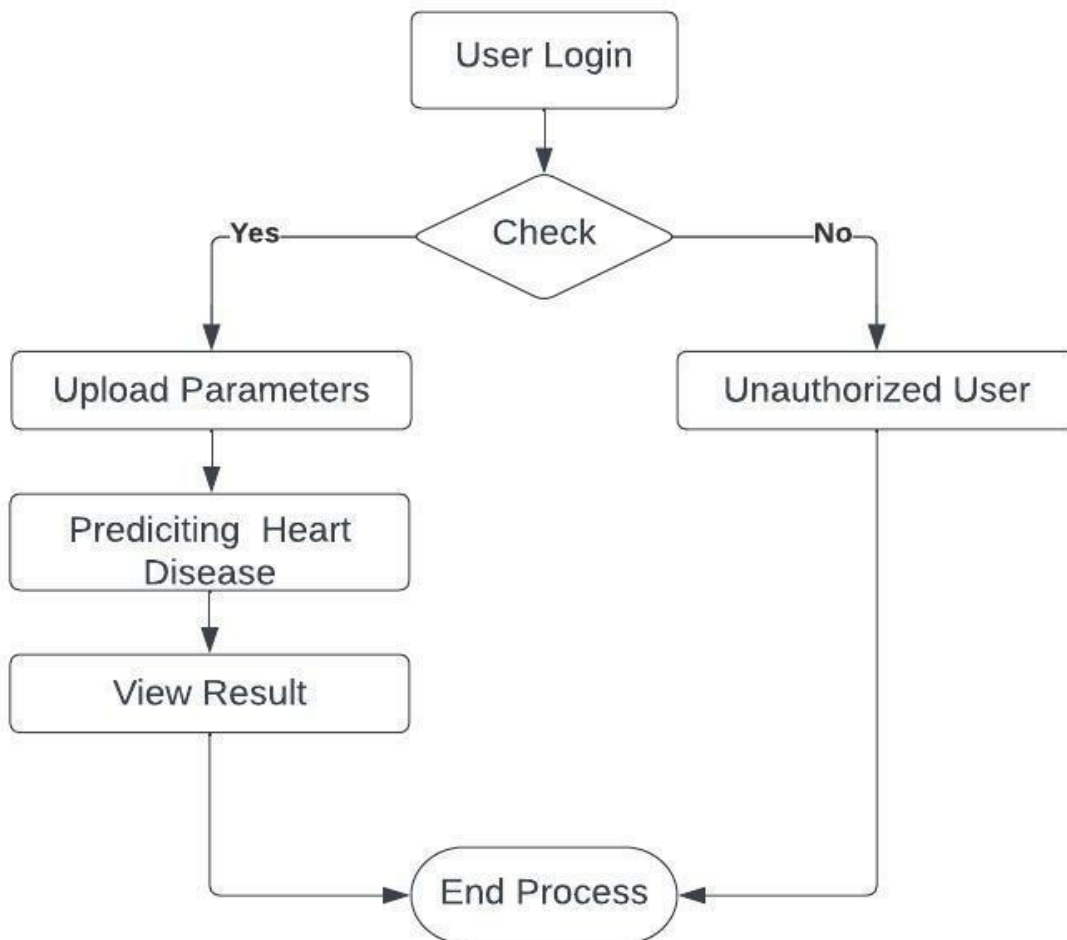
Working Procedure

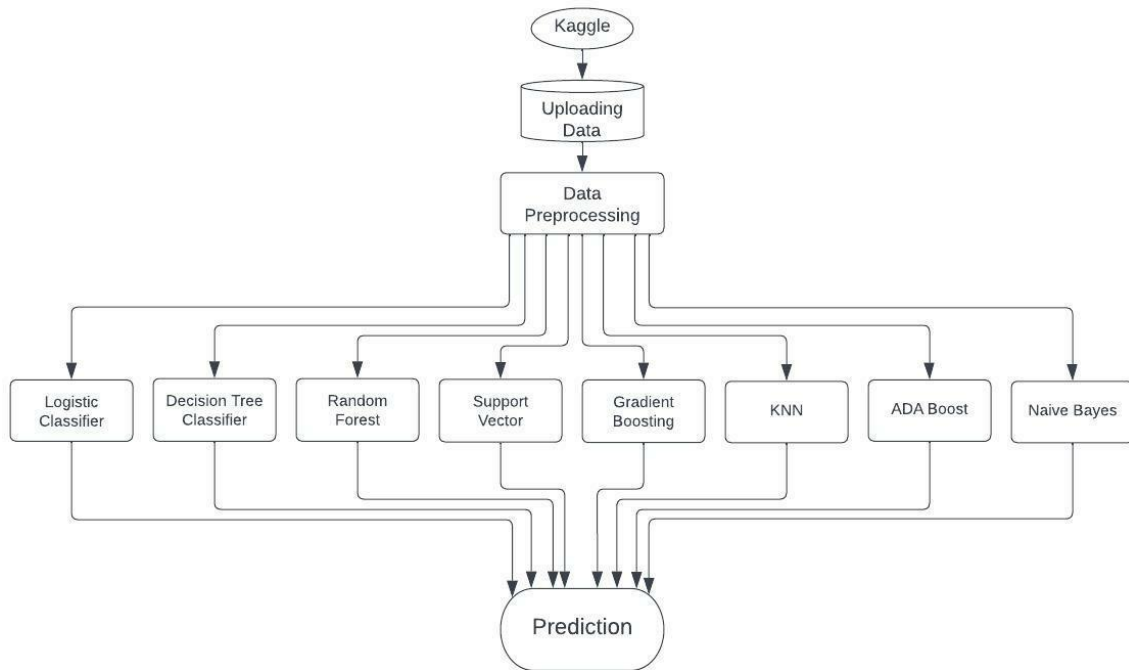Fig. Dataset Attributes

Fig: Algorithm flow chart

Fig: Working and Output

Multiple Algorithms Used In The Project

Logistic Regression

This is similar to a linear regression in which data is sorted and simplified to ensure the final outcome, an equation is formulated and utilized based on the weightage of coefficients and their significance in the equation.

Example: $y = e^{(b0 + b1*x)} / (1 + e^{(b0 + b1*x)})$

Decision Tree

This is well devised selection technique in which data is arranged and sorted based on the tree structures in which datasets are placed at the internal nodes, branches represent the decision structure provided and the leaf nodes indicate the outcome. With a graphical outcome the data becomes comprehensive and efficient. The **CART** (Classification and Regression Tree Algorithm) is created to ensure that decision efficiency of the project increases exponentially.

Support Vector Classified and Random Forest

This algorithm is a part of ensemble learning in which multiple classifiers are put together to find a solution to a complex problem. This algorithm is a vital part of machine learning and it utilizes multiple tree models to ensure that the most suitable result is reached with ease.

AdaBoost Classifier

This algorithm of machine learning allows the previous leaning to carry on to the next parameter allowing the principle boost in their memory, the AdaBoost algorithm has its core functionality in boosting the working with a slight visible difference.

Gradient Boost Classifier

These algorithms combines all the weaker models to create a complex and reliable model which reduces the chances of errors and increases the program efficiency.

K-Nearest Neighbors

This algorithm utilizes the stored data sets for its further learning which is the major reason why it is known as lazy learner algorithm as it is able to grasp the entire information in a longer run with increased efficiency.

Confusion Matrix

This matrix is a N x N matrix in which the target classes are stored and these classes compare the target values to check the implementation of program which makes things easier in sorting and implementation.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.00
mean	54.366337	0.683168	0.966997	131.623762	246.264026	0.148515	0.528053	149.646865	0.326733	1.039604	1.399340	0.729373	2.31
std	9.082101	0.466011	1.032052	17.538143	51.830751	0.356198	0.525860	22.905161	0.469794	1.161075	0.616226	1.022606	0.61
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000	0.00
25%	47.500000	0.000000	0.000000	120.000000				133.500000	0.000000	0.000000	1.000000	0.000000	2.00
50%	55.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000	153.000000	0.000000	0.800000	1.000000	0.000000	2.00
75%	61.000000	1.000000	2.000000	140.000000	274.500000	0.000000	1.000000	166.000000	1.000000	1.600000	2.000000	1.000000	3.00
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000	3.00

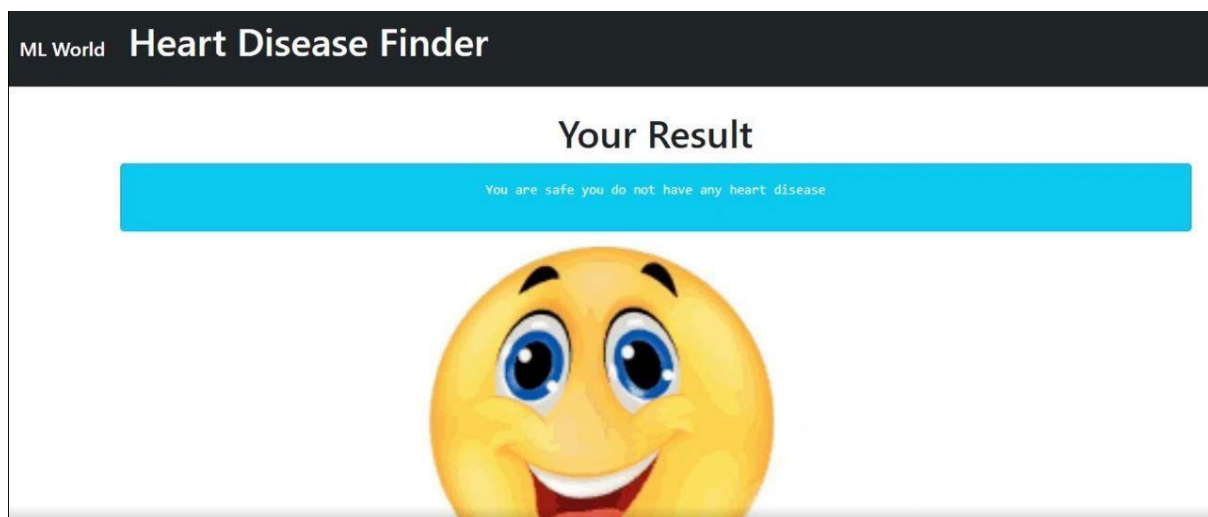Fig. Data set presentation

Fig. Chatbot

Fig: Output after processing information

Scope And Conclusion

The greatest regret posed by a human is unable to save a loved one and with this project we seek to put forward an innovation that would eliminate this regret. We have merged technology with patient records as data sets to provide a thorough analysis of heart diseases. The cardiovascular diseases is a greater segment and it continues to be a great contributor in global deaths. So we wish to put forward a project that would make it easier to predict heart disease and seek for reliable medical guidance.

References

- Shah, Devansh, Samir Patel, and Santosh Kumar Bharti. "Heart disease prediction using machine learning techniques." SN Computer Science 1.6 (2020): 1-6.
- Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." IEEE access 7 (2019): 81542-81554.
- Rajdhan, Apurb, et al. "Heart disease prediction using machine learning." International Journal of Research and Technology 9.04 (2020): 659-662.
- Ali, Md Mamun, et al. "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison." Computers in Biology and Medicine 136 (2021): 104672.
- Jindal, Harshit, et al. "Heart disease prediction using machine learning algorithms." IOP conference series: materials science and engineering. Vol. 1022. No. 1. IOP Publishing, 2021.
- Ali, Md Mamun, et al. "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison." Computers in Biology and Medicine 136 (2021): 104672.