IQRA UNIVERSITY IU

Higher Education Commition of Pakistan

Heart Disease Prediction

Final Year Project II

Project Advisor/Supervisor : Dr Raza
Co Supervisor : Ms Sania Marium

https://iqra.edu.pk/

IQRA UNIVERSITY

Where Your Future Begins!

Project participation – 30%

37728

Yasir Ul Haq

Assistant Developer

Project participation – 40%

37888

Haris Bin Irfan

Leader/Developer

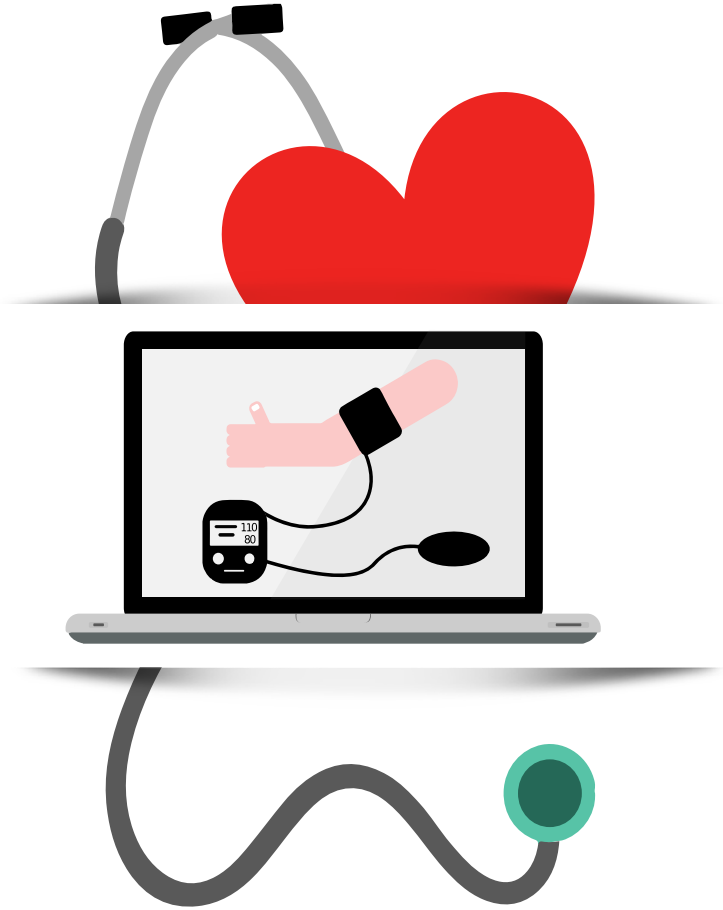
Project participation – 30%

38633

Syed Usama Ahmed

Designer/Support Staff

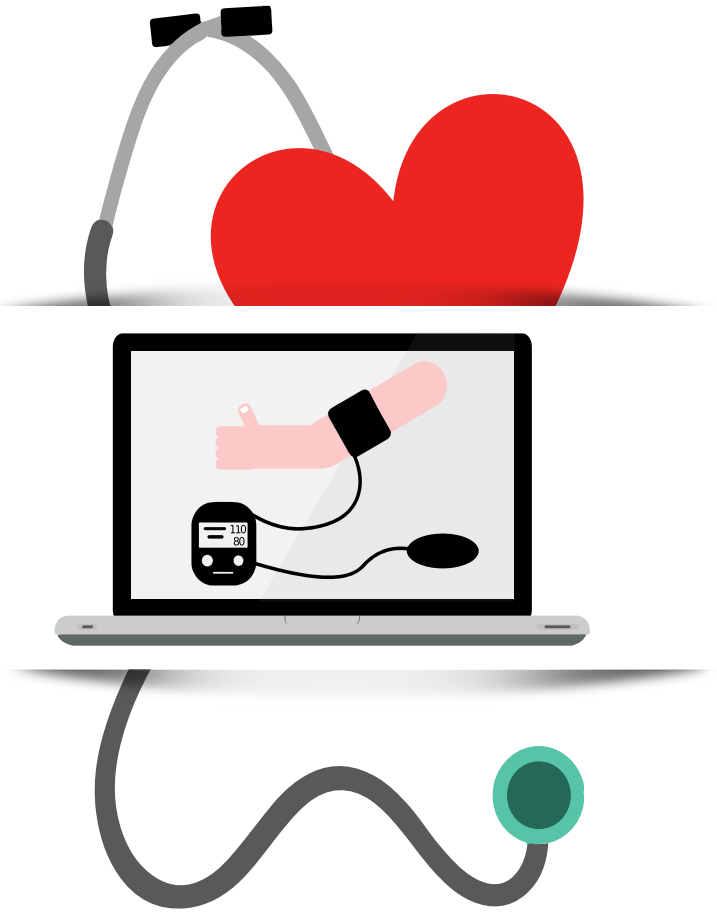
Agenda



INTRODUCTION

Data mining is the process of finding patterns and trends previously unknown in datasets, and using this information to build predictive models. To extract secret patterns and associations from large databases, data mining incorporates statistical analysis, machine learning and computer technologies. The World Health Statistics 2012 study highlights the fact that one in three people globally has elevated blood pressure-a condition that causes roughly half of all stroke and heart disease deaths. Heart disease, also known as cardiovascular disease (CVD), includes a number of heart attack conditions-not just heart attacks . Heart attack was the leading cause of deaths in the various countries like India. In the United States, heart disease affects a single adult every 34 seconds. Other types of heart disease include coronary heart disease, cardiomyopathy, and cardiovascular disease. The word "cardiovascular disease" involves a wide range of conditions that affect the heart and blood vessels, and how blood is pumped through the body and distributed. Diagnosis is a complicated and important task which requires accurate and efficient execution.

Often the diagnosis is made, based on the experience & expertise of the doctor. This adds to unnecessary patient outcomes & disproportionate medical cost of therapies. Hence an automatic system of medical diagnosis would be extremely beneficial. The healthcare industry collects huge amounts of healthcare data and that need to be mined to discover hidden information for effective decision making. Discover of hidden patterns and relationships often go unexploited [1]



PROBLEM STATEMENT

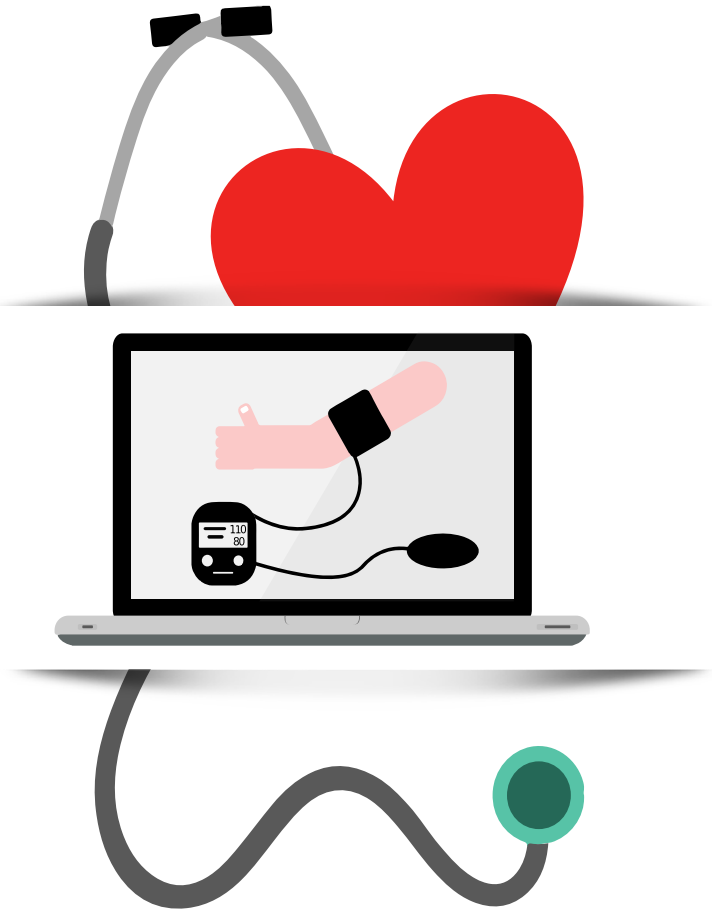
A mixture of lifestyle changes, medication and, in some cases, surgery, will effectively manage heart disease. The signs of heart disease can be reduced with the right treatment, and heart functioning increased. The projected outcomes can be used to avoid and thus reduce cost of surgical and other expensive treatments. The heart is a very critical part of the human body. It pumps blood into the whole body If the flow of blood in the body becomes insufficient, organs like the brain suffer, and if the heart stops working completely, death occurs within minutes. Some of the heart disease risk factors are:

1. Smoking: Smokers are twice as likely to have a heart attack as non smokers.
2. Cholesterol: A low cholesterol diet and Tran's saturated fat can help lower cholesterol levels and reduce heart disease risk.
3. Blood pressure: High BP leads to heart Attack.
4. Diabetes: If not controlled, diabetes can result in severe heart damage including heart attack and death.
5. Sedentary life style: Simple leisure activities, such as gardening and walking, can reduce our risk of heart disease.
6. Eating Habits: A healthy diet of the soul, low in salt, calories, saturated fat, trans fat, cholesterol and refined sugars can decrease our risk of heart disease.
7. Stress: Poorly controlled stress can result in heart attacks and strokes.

There are many methods related to prediction of disease. Yet heart-related disease in particular has been analyzed and the level of risk is produced. But there are usually no such tools that are used for specific disease prediction.

The main objective is to predict the Boolean class heart disease prediction, which represents whether a patient has heart disease or not:

- False does not represent heart disease.
- True represents heart disease present.



STATE OF THE ART

The pre-existing system operates on deep learning as well as data mining collections. By applying the powerful prediction algorithm, the existing system modules produce a comprehensive report. The main aim of the existing system was to compare and check the prepatient with disease outputs and new patient disease and then identify future possibilities of cardiac disease for a specific patient. By implementing following classification models:

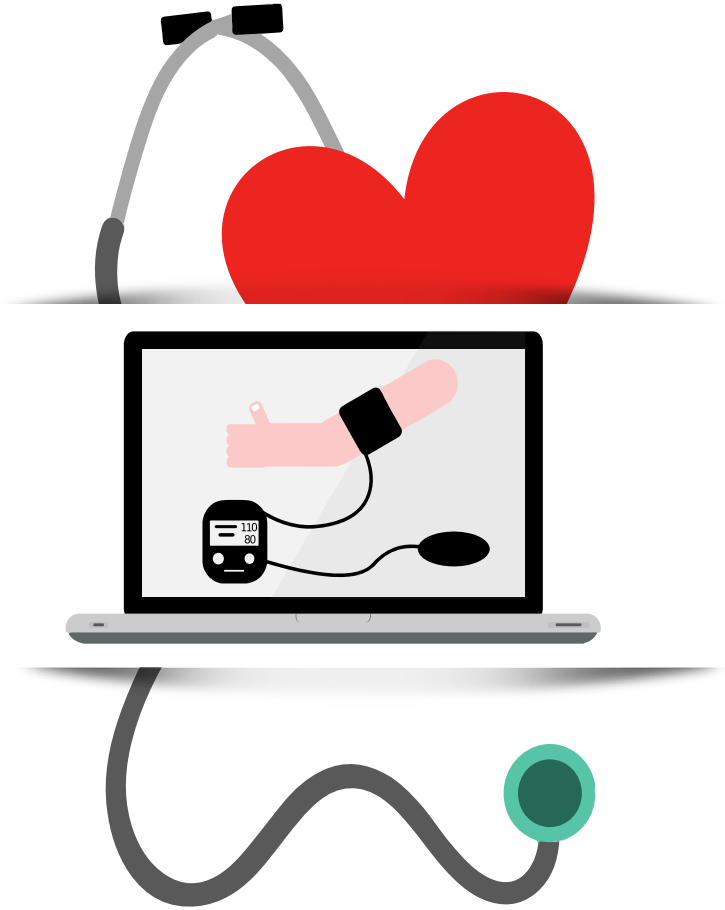
- Naive Bayes
- Logistic Regression
- Decision Tree
- Random Forest
- LightGBM
- XGboost

But the accuracy is much less by using all the existing systems.

Drawbacks of the existing system:

1. It's very difficult to maintain the system.
2. There is a high chance of obtaining inaccurate results.
3. GUI is much less user-friendly.
4. It takes longer to process the activities.

PROPOSED SOLUTION

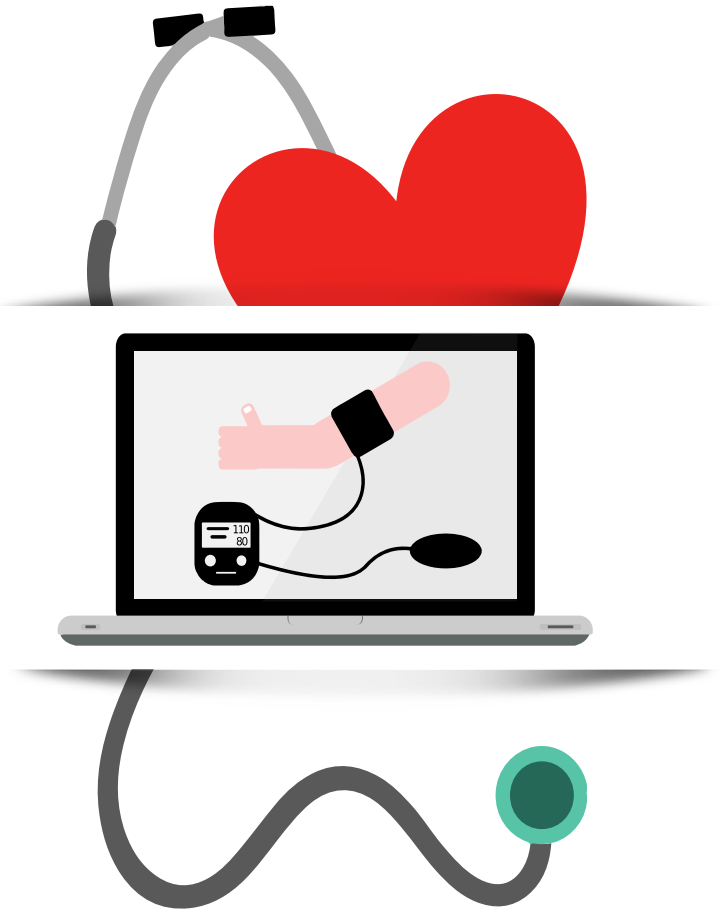


MOTIVATION AND NEED

Healthcare organizations face a major challenge in providing quality facilities at affordable rates. Quality service includes correct diagnosis of patients and effective treatments. Bad medical decisions can lead to catastrophic and therefore unacceptable consequences. Hospitals also need to reduce the expense of clinical testing. By using effective computer-based data and/or decision support systems, they can achieve these outcomes. Many hospitals today use some kind of hospital information systems to monitor their health care or patient information. Usually, these systems create huge amounts of data in the form of numbers, text, graphs and photographs. These data are, sadly, rarely used to assist medical decision-making. Such data contain a variety of hidden information that is largely unused. This raises a critical question: "How will we have a tendency to turn data into helpful knowledge that helps health care professionals to create smart medical decisions?" This is the main driving force behind this study.

CHALLENGES

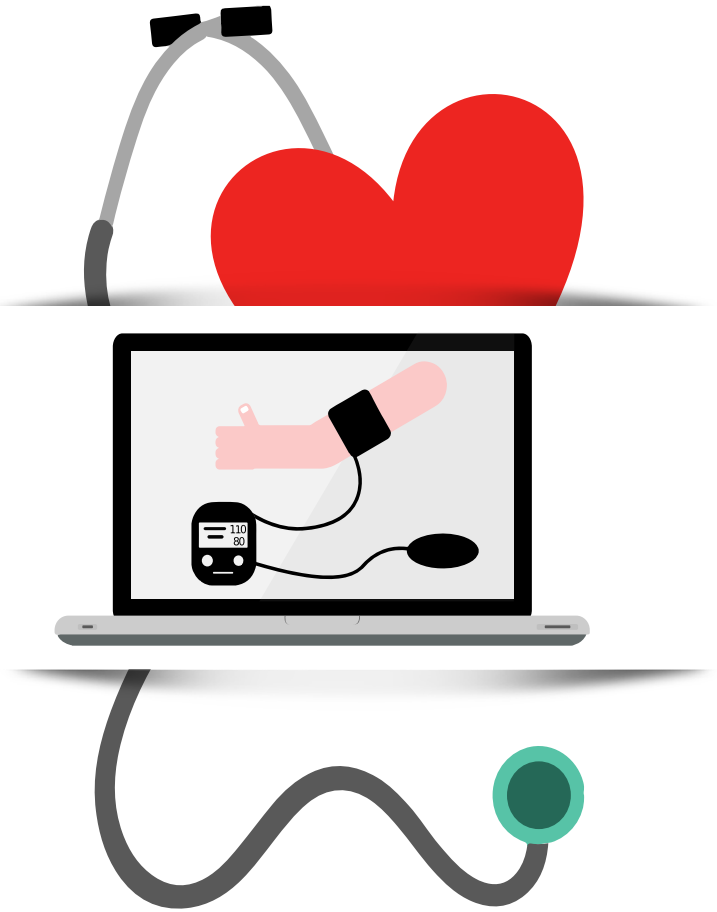
Medical diagnosis is considered to be an important yet complex task that needs to be accurately and efficiently carried out. It would be very useful to automate the same. Clinical decisions are often made based on the expertise and experience of the practitioner rather than the knowledge-rich information that is hidden in the server. This practice leads to unacceptable biases errors and high medical costs that impact patient's quality of service. Data mining has the potential to create a knowledge-rich environment that can significantly enhance the quality of clinical decision-making. Choosing the best algorithm for the purpose of training. Implementing the high Accuracy rate algorithm in system to generate accurate heart disease predictions.



PROPOSED SOLUTION (Cont.)

AIM & OBJECTIVES

The proposed system has details that are categorized according to features in it if patients have heart disease or not. This proposed system will attempt to use this dataset to create a model that tries to predict whether or not a person has this disease. In this proposed system, we use 4 Classification algorithums BaggingClassifier(GaussianNB), MLPClassifier, SVC and AdaBoostClassifier(RandomForestClassifier). Calculating the score using the sklearn library. Implements VotingClassifier to get best accuracy results. Using the Comparing Models and Confusion Matrix to finally analyze the results. It should be grouped into separate structured data depending on the features of the patient's heart from the dataset we have. First, we have to import the dataset. Read the dataset, the data should contain different variables like age, gender, sex, cp(chest pain),slope, target etc. The data should be explored so that the information is verified. The records are divided into two datasets: dataset training 80% and dataset testing 20%. To avoid bias, the records are selected randomly for each set. Using Voting Classifier, we need to create a model that predicts the disease of the patient.



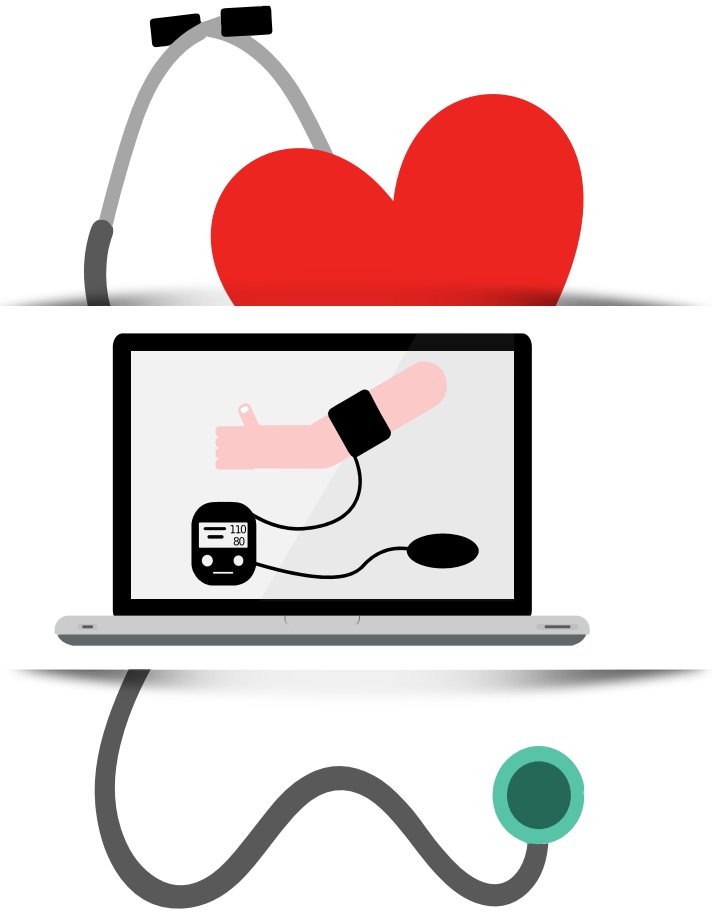
PROPOSED SOLUTION (Cont.)

METHODOLOGY

Methodology is a system that includes steps to convert raw data into recognized data patterns to extract user knowledge. The approach suggested involves measures, referred to as the preprocessing stage where the information are thoroughly analyzed. It will deal with missing values, balance information and normalize attributes depending on the algorithms used. Using classification models and EnsembleVoteClassifier, predictive analysis of the data is done after pre-processing of data. Eventually, prescriptive modeling is performed, where different performance metrics are used to test the predictive model in terms of performance and accuracy.

Data set for experiment: The data set for this research has been taken from the UCI data repository. Used data is freely available from the UCI Machine Learning Repository [10]. The Cleveland data were collected from the above-mentioned DM warehouse. This database includes 76 attributes and 14 attributes have been picked after neglecting redundant and obsolete attributes.

The list of 14 attributes and their brief description are shown in above table. In particular, several researchers used the Cleveland datasets and found that they were appropriate for the creation of a mining model due to lower missing values and outliers. Before they were submitted to the proposed algorithm for training and testing, the data were cleaned and preprocessed. Therefore, 303 are the appropriate instances for the development of supervised machine-learning model building. Attribute selection technique was used for further data reduction to make patterns easier and more comprehensible, but negligible effects were found on model performance observations undertaken in this study. All 13 attributes are considered in order to develop a model for the classifier and to achieve a predictive outcome for heart disease. GaussianNB, MLPClassifier, SVM, RandomForestClassifier algorithm are the classification techniques used in this research. The EnsembleVoteClassifier was used to evaluate the algorithms involved in the classification. The model was built using the Google Colab tool. In these experiments, 5-fold cross-validations were used to divide the data set into training and test sets that meets the model training and testing purpose requirement. As a result, the accuracy rate of this study was over 85%.



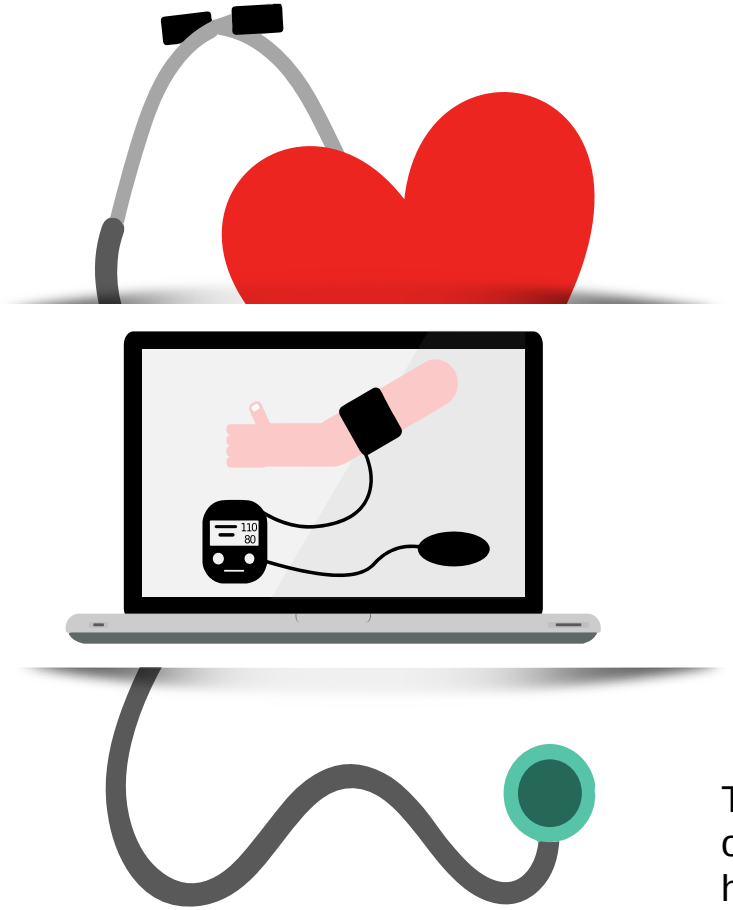
KEY BENEFITS & OUTCOMES

KEY BENEFITS AND BENEFICIARIE

Predictive modeling is useful as it offers accurate insight into any problem and allows users to create outcomes. In order to maintain a competitive advantage, insight into future events and results that challenge key assumptions is crucial.
An advertisement of the product will be placed on some social and high traffic websites through which user would be able to get more traffic on our heart disease prediction website. Banners are placed in various universities as well.

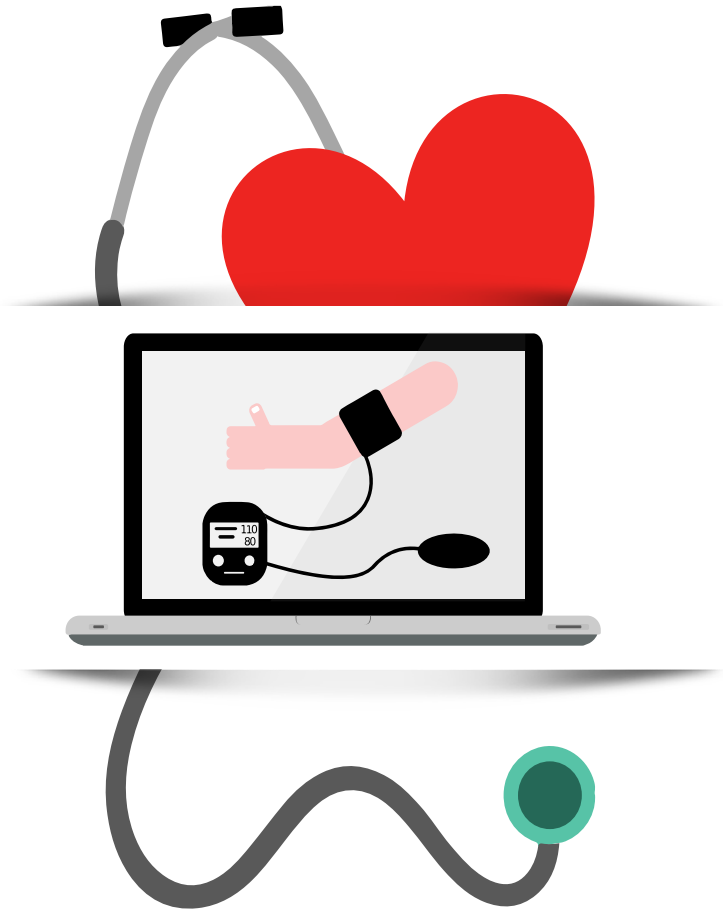
OUTCOMES

- Full Project Proposal Report with all the details of the project
- Complete Web Application with all feature mentioned above
- Server for Database
- Backup Database



TARGET MARKET

The proposed system will cover a particular field of hospital and healthcare center and the target users of this system are also specific like heart-related doctors who have directly used this system in hospitals and healthcare centers.



TECHNOLOGY USED

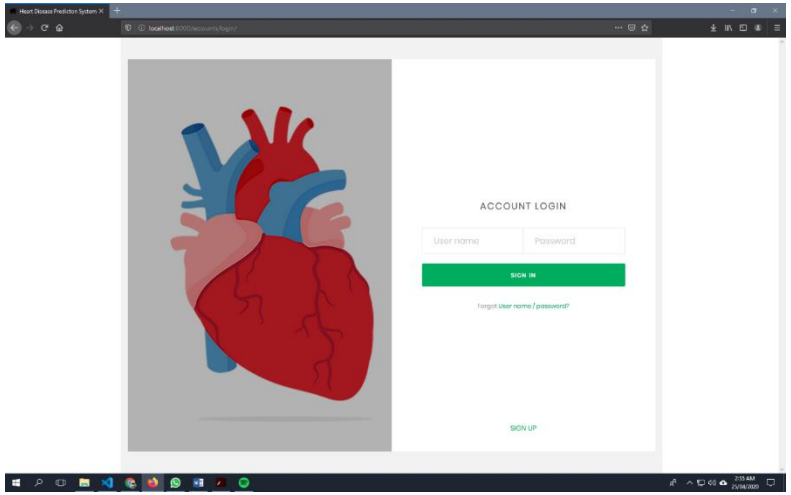
TOOLS & LANGUAGES

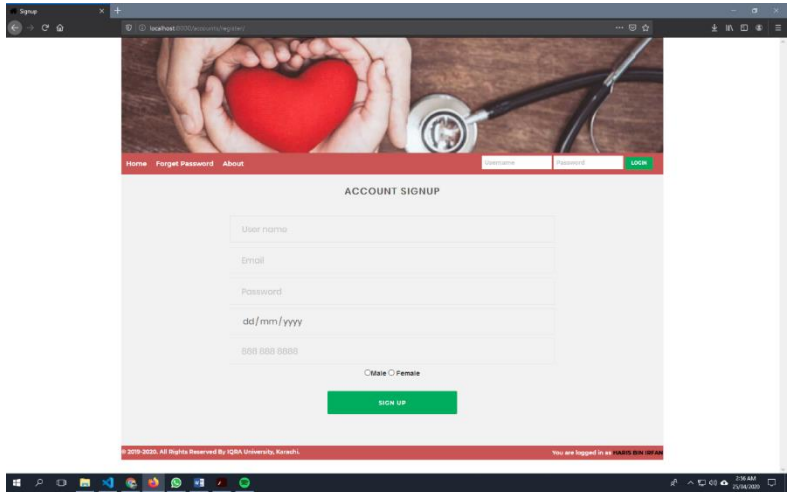
- ADOBE PHOTOSHOP
- ADOBE ILLUSTRATOR
- ADOBE XD
- BOOTSTRAP
- HTML/CSS/JS
- GOOGLE COLAB
- DJANGO FRAMEWORK
- POSTGRESS SQL
- VISUAL STUDIO CODE

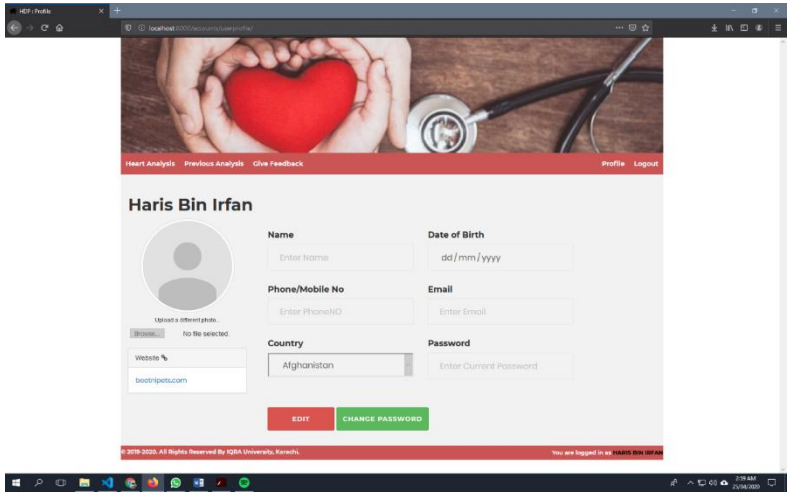
CLASSIFIERS

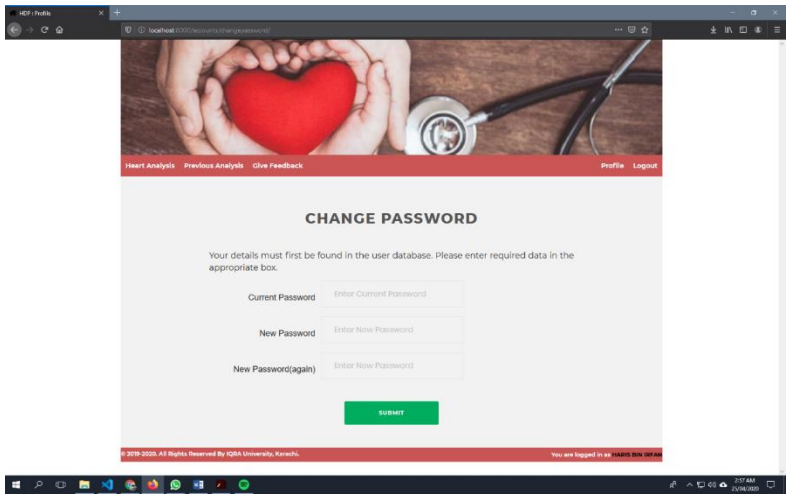
- A multilayer perceptron (MLP)
- A Support Vector Machine (SVM)
- Gaussian Bayes of Naive
- AdaBoostClassifier(RandomForestClassifier)
- EnsembleVoteClassifier

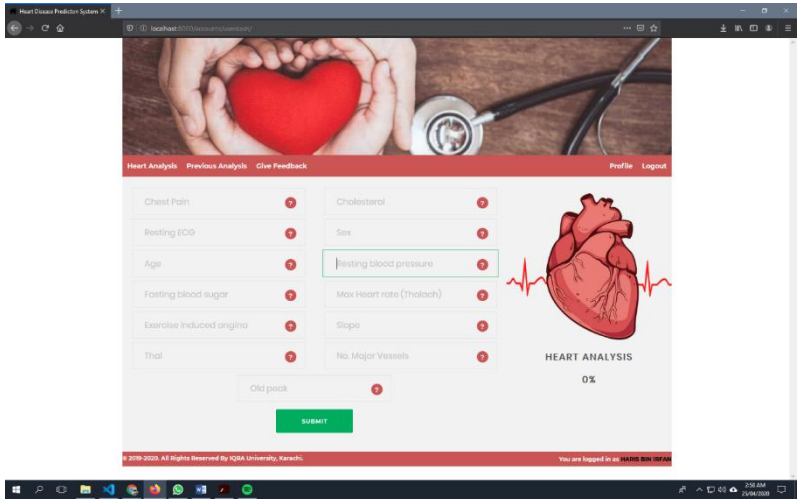
SNAPSHOT OF THE PROJECT INTERFACE

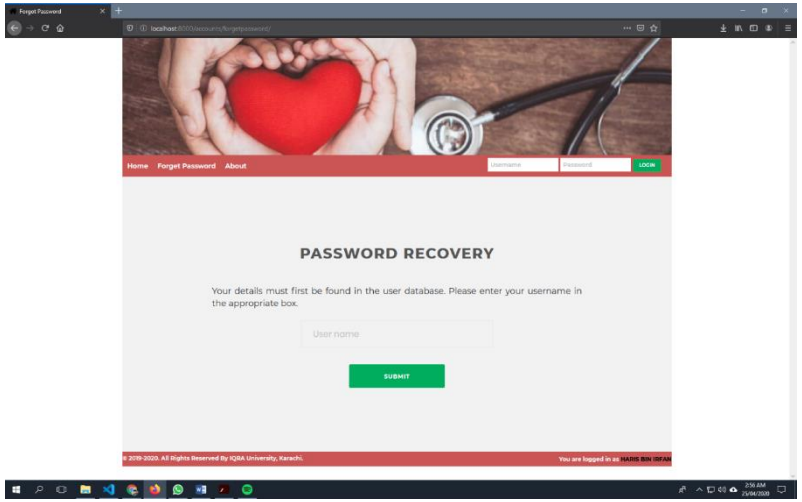




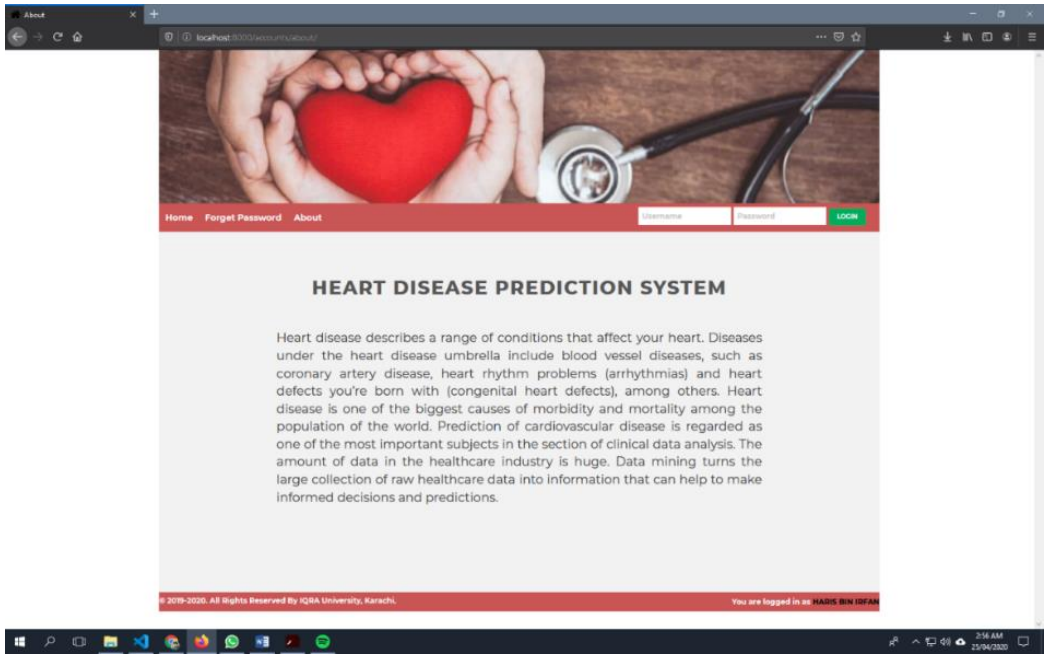


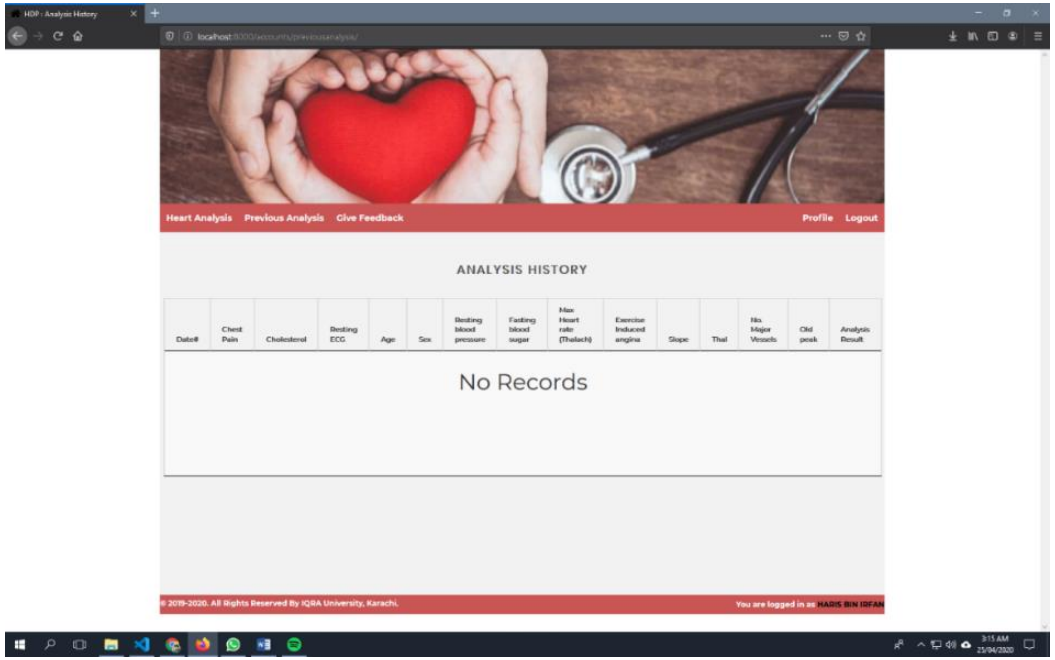






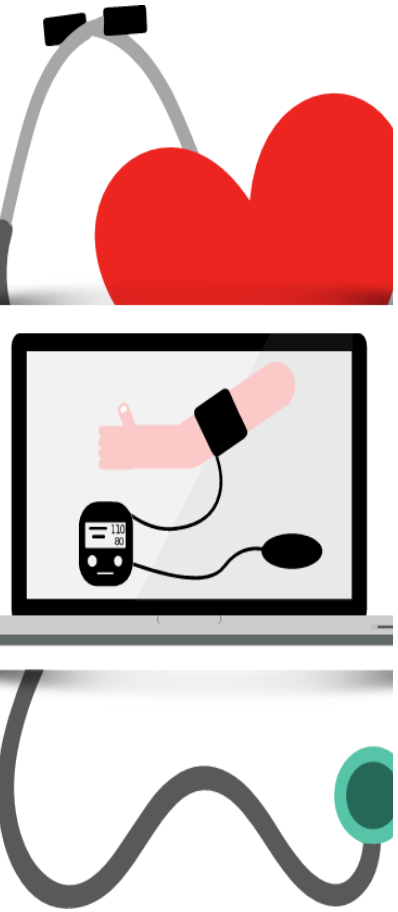
SNAPSHOT OF THE PROJECT INTERFACE





Lean Business Model Canvas

Model: Heart Disease Prediction



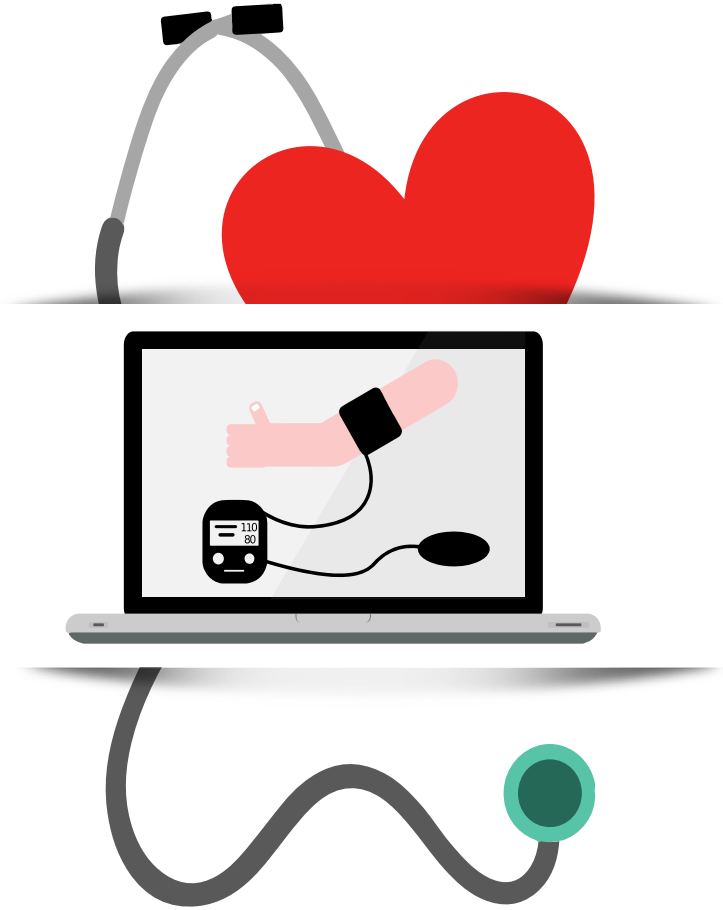
Problem	Solution	Unique Value Proposition	Unfair Advantages	Customer Segments
The main Problem which we have detected about the prediction of heart disease that's the consultants take much time to diagnose a cancer, Survivability and Reoccurrence. Cost is too high and ime waste in different tests.There are many methods related to prediction of disease. Yet heart-related disease in particular has been analyzed and the level of risk is produced. But there are usually no such tools that are used for specific disease prediction. The main objective is to predict the Boolean class heart disease prediction, which represents whether a patient has heart disease or not.	Heart Disease Prediction application is an end user support and online consultation project.a web application that allows users to get instant guidance on their heart disease through an intelligent system online. We use Machine learning algorithms to solve the mentioned problems faced by the patients and also for doctors. **Key Metrics** Our aim is to complete the project with the abilities of application which can easily: 1.Predict the heart disease. 2.Predict the survivability of patients.	From the proposed solution we have selected the machine learning algorithms which 1. MLPClassifier 2. SVM 3. GaussianNaïve Bayes 4. AdaBoostClassifier (RandomForestClassifier) 5. EnsembleVoteClassifier	As we are using the combination of machine learning algorithms in our project and it is very obvious that combination can never be the same in any backend development it may have different. **Channels** Channels for reaching to customers are: 1.Web Advertisement. 2.Facebook page. 3.Instagram page. 4.Hospitals.	Customers are: 1. Patients 2. Doctors

Cost Structure

Technical HR Deployment Cost	Rs.60,000
Support Staff	Rs.150,000
Equipment	Rs.452,606
Traveling	Rs.90,000
Boarding & Lodging	Rs.180,000
Miscellaneous	Rs.102,000
Sub Total:	**Rs.1,034,606**
Audit Charges	Rs.25,000
Contingency	Rs.50,000
Institutional/Organizational Overheads	Rs.200,000
Total Budget:	**Rs.1,309,606**

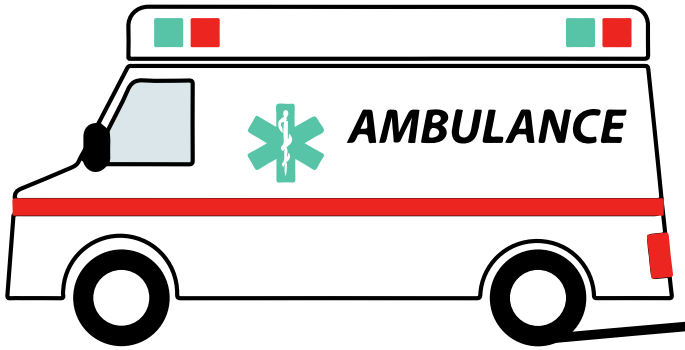
Revenues Streams

We will generate our revenue form CPM display advertising on web application. (e.g. banners ads and skyscrapers).
CPM stands for "cost per thousand" where M denotes "Mille". The site owner charges advertisers a rate card price (for example 50 GBP CPM) according to the number of its ads shown to site visitors. Ads may be served by the site owners own ad server or more commonly through a third-party ad network service such as Google AdSense as is the case with my site.



CONCLUSION

The proposed model is a web application system based on GUI, user-friendly, scalable, responsive and expandable. The proposed working model will also help to reduce the cost of care by offering prompt initial diagnostics. The model will also serve the function of the medical student training program, and will be a soft diagnostic method available to doctors and cardiologists. General practitioners can use this method for initial cardio-patient diagnosis. There are several potential enhancements which could be explored to enhance this prediction system's scalability and accuracy. Since we've built a simplified method, we can use this method to analyze various data sets in future. The dimensionality of the heart database is typically high in DM warehouse so defining and selecting significant attributes for better diagnosis of heart disease are very daunting tasks for future study.



Thank You