Heart Disease Prediction Using Machine Learning Algorithms

By

Jaswanth Sakamuri (Id:1152228)

Advised by

Prof. Ausif Mahmood

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIRMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN COMPUTER ENGINEERING

THE SCHOOL OF ENGINEERING UNIVERSITY OF BRIDGEPORT CONNECTICUT

December, 2023

ABSTRACT:

In recent times, Heart Disease prediction is one of the most complicated tasks in medical field. In the modern era, approximately one person dies per minute due to heart disease. Data science plays a crucial role in processing huge amount of data in the field of healthcare. As heart disease prediction is a complex task, there is a need to automate the prediction process to avoid risks associated with it and alert the patient well in advance. This paper makes use of heart disease dataset available in UCI machine learning repository. The proposed work predicts the chances of Heart Disease and classifies patient's risk level by implementing different data mining techniques such as Naive Bayes, Decision Tree, Logistic Regression and Random Forest. The trial results verify that KNN algorithm has achieved the highest accuracy of 90.16% compared to other ML algorithms implemented.

Chapter 1

INTRODUCTION

1.1 Overview

According to the World Health Organization, every year 12 million deaths occur worldwide due to Heart Disease. Heart disease is one of the biggest causes of morbidity and mortality among the population of the world. Prediction of cardiovascular disease is regarded as one of the most important subjects in the section of data analysis. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduces the complications. Machine learning proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the health care industry. This project aims to predict future Heart Disease by analyzing data of patients which classifies whether they have heart disease or not using machine-learning algorithm. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources, classifying them under suitable headings & finally analyzing to extract the desired data we can say that this technique can be very well adapted to do the prediction of heart disease.

The work proposed in this paper focus mainly on various data mining practices that are employed in heart disease prediction. Human heart is the principal part of the human body. Basically, it regulates blood flow throughout our body. Any irregularity to heart can cause distress in other parts of body. Any sort of disturbance to normal functioning of the heart can be classified as a Heart disease. In today's contemporary world, heart disease is one of the primary reasons for occurrence of most deaths. Heart disease may occur due to unhealthy lifestyle,

smoking, alcohol and high intake of fat which may cause hypertension. A healthy lifestyle and earliest detection are only ways to prevent the heart related diseases.

The main challenge in today's healthcare is provision of best quality services and effective accurate diagnosis. Even if heart diseases are found as the prime source of death in the world in recent years, they are also the ones that can be controlled and managed effectively. The whole accuracy in management of a disease lies on the proper time of detection of that disease. The proposed work makes an attempt to detect these heart diseases at early stage to avoid disastrous consequences.

Records of large set of medical data created by medical experts are available for analyzing and extracting valuable knowledge from it. Data mining techniques are the means of extracting valuable and hidden information from the large amount of data available. Mostly the medical database consists of discrete information. Hence, decision making using discrete data becomes complex and tough task. Machine Learning (ML) which is subfield of data mining handles large scale well-formatted dataset efficiently. In the medical field, machine learning can be used for diagnosis, detection and prediction of various diseases. The main goal of this paper is to provide a tool for doctors to detect heart disease as early stage. This in turn will help to provide effective treatment to patients and avoid severe consequences. ML plays a very important role to detect the hidden discrete patterns and thereby analyze the given data. After analysis of data ML techniques help in heart disease prediction and early diagnosis.

1.2 Relevant contemporary issues

Machine learning proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the health care industry. This project aims to predict future Heart Disease by analyzing data of patients which classifies whether they have heart disease or not using machine-learning algorithm. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the

data from various sources, classifying them under suitable headings & finally analysing to extract the desired data we can say that this technique can be very well adapted to do the prediction of heart disease.

Machine Learning is one of the most widely used concepts around the world. It will be essential in the healthcare sectors which will be useful for doctors to fasten the diagnosis. In this article, we will be dealing with the Heart disease dataset and will analyze, predict the result whether the patient has heart disease or normal, i.e. Heart disease prediction using Machine Learning. This prediction will make it faster and more efficient in healthcare sectors which will be a time-consuming process.

Chapter 2

LITERATURE SURVEY

[1] TITLE: Effective Heart Disease Prediction using Hybrid Machine Learning Techniques

YEAR: 2019

AUTHOR: Senthilkumar Mohan

ABSTRACT: The prediction model is introduced with different combinations of features and several known classification techniques. We produce an enhanced performance level with an accuracy level of 88.7% through the prediction model for heart disease with the hybrid random forest with a linear model (HRFLM)

ALGORITHMS: Random forest, HRFLM

CONCLUSION: Identifying the processing of raw healthcare data of heart information will help in the long term saving of human lives and early detection of abnormalities in heart conditions. Machine learning techniques were used in this work to process raw data and provide a new and novel discernment towards heart disease. Heart disease prediction is challenging and very important in the medical field. However, the mortality rate can be drastically controlled if the disease is detected at the early stages and preventative measures are adopted as soon as possible. Further extension of this study is highly desirable to direct the investigations to real-world datasets instead of just theoretical approaches and simulations. The proposed hybrid HRFLM approach is used combining the characteristics of Random Forest (RF) and Linear Method (LM). HRFLM proved to be quite accurate in the prediction of heart disease

[2] TITLE: Heart Disease Prediction using Machine Learning Techniques

YEAR: 2020

AUTHOR: Devansh Shah

ABSTRACT: This research paper presents various attributes related to heart disease, and the model on basis of supervised learning algorithms as Naïve Bayes, decision tree, K-nearest neighbor, and random forest algorithm. It uses the existing dataset from the Cleveland database of UCI repository of heart disease patients. The dataset comprises 303 instances and 76

attributes. Of these 76 attributes, only 14 attributes are considered for testing, important to substantiate the performance of different algorithms. This research paper aims to envision the probability of developing heart disease in the patients. The results portray that the highest accuracy score is achieved with K-nearest neighbor.

ALGORITHMS: Decision tree, Naïve Bayes, K-NN, Random forest

CONCLUSION: The overall aim is to define various data mining techniques useful in effective heart disease prediction. Efficient and accurate prediction with a lesser number of attributes and tests is our goal. In this study, I consider only 14 essential attributes. I applied four data mining classification techniques, K-nearest neighbor, Naive Bayes, decision tree, and random forest. The data were pre-processed and then used in the model. K-nearest neighbor, Naïve Bayes, and random forest are the algorithms showing the best results in this model. I found the accuracy after implementing four algorithms to be highest in K-nearest neighbors (k = 7). We can further expand this research incorporating other data mining techniques such as time series, clustering and association rules, support vector machine, and genetic algorithm. Considering the limitations of this study, there is a need to implement more complex and combination of models to get higher accuracy for early prediction of heart disease.

[3] TITLE: Latest trends on heart disease prediction using machine learning and image fusion

YEAR: 2020

AUTHOR: Neeraj kumar

ABSTRACT: Within this paper we include a review of the classification methods for machine learning and image fusion that have been demonstrated to help healthcare professionals identify heart disease.

ALGORITHMS: ANN

CONCLUSION: The conclusion of algorithm has been determined; the proposed system may be used in the useful area. More specific feature selection approaches are used to improve algorithm precision, so that reliable results can be obtained. If the particular type of heart disease is diagnosed, care for that specific condition should be provided to the patient. Essentially, we would conclude that a dataset of appropriate samples and reliable data will be used to create a predictive model of heart disease. Accordingly, the dataset must be pre-

processed because pre-processing is the most critical part that prepares the dataset used by the machine learning algorithm and gets better results. An appropriate algorithm must be used to develop a predictive model that delivers accurate results. We observe that ANN has good effects for predicting heart disease in most models. Finally, the use of machine learning and image fusion to detect heart disease is an essential activity, and it can be of assistance to both healthcare authorities and patients. It is still an increasing area, and not much of it is reported because of the vast availability of patient data in hospitals or clinics. Most researchers obtained their datasets from the same source that is the repository of UCI. Given that the quality of the dataset is an significant factor in the accuracy of the forecast, more hospitals should be encouraged to publish high-quality datasets (while preserving patients' privacy) so that researchers will have a reliable source to help them improve their models and get good results that can help people profit from and cure heart disease in its initial stage.

[4] TITLE: Implementation of Machine Learning Model to Predict Heart Failure Disease

YEAR: 2019

AUTHOR: Fahd Saleh Alotaibi

ABSTRACT: This paper aims to improve the HF prediction accuracy using UCI heart disease dataset. For this, multiple machine learning approaches used to understand the data and predict the HF chances in a medical database. Furthermore, the results and comparative study showed that, the current work improved the previous accuracy score in predicting heart disease.

ALGORITHMS: Decision tree, Navie bayies

CONCLUSION: The ratio of heart failure patients has been increasing every day. To overcome this dangerous situation and deteriorate the chances of heart failure disease, there is a need of a system that can generate rules or classify the data using machine learning approaches. Therefore, this research discussed, proposed and implemented a machine learning model by combining five different algorithms. Rapid miner is the tool used in this research, which computed the high accuracy than Matlab and Weka tool. In comparison with the previous researches, this study has shown significant improvement and high accuracy than previous work. As far as UCI dataset concerns, the dataset needs to be amplified. As the main limitation in this work is the small size of the dataset. The dataset has limited number of patient's records; therefore, the dataset was augmented using appropriate techniques.

[5] TITLE: Design and Implementing Heart Disease Prediction Using Naives Bayesian

YEAR: 2019

AUTHOR: Anjan Nikhil Repaka

ABSTRACT: The research focuses on establishing SHDP (Smart Heart Disease Prediction that takes into consideration the approach of NB (Naive Bayesian) classification and AES (Advanced Encryption Standard) algorithm for resolving the issue of heart disease prediction. its revealed that in regard to accuracy, the prevailing technique surpasses the Naive Bayes by yielding an accuracy of 89.77%in spite of reducing the attributes. AES yields in high security performance evaluation in comparison to PHEA (Parallel Homomorphic Encryption Algorithm).

ALGORITHMS: Navies Bayesian

CONCLUSION: Data collection is carried out using numerous sources that are primary factors responsible for any sort of heart disease and thereby using a structure the database is constructed. The research focuses on establishing SHDP (Smart Heart Disease Prediction that takes into consideration the approach of NB (Naive Bayesian) classification and AES (Advanced Encryption Standard) algorithm for resolving the issue of heart disease prediction. its revealed that in regard to accuracy, the prevailing technique surpasses the Naive Bayes by yielding an accuracy of 89.77%in spite of reducing the attributes. AES yields in high security performance evaluation in comparison to PHEA (Parallel Homomorphic Encryption Algorithm).

Chapter 3

OVERVIEW OF THE PROJECT

3.1 Objective

The main objective of this research is to develop a heart prediction system. The system can discover and extract hidden knowledge associated with diseases from a historical heart data set Heart disease prediction system aims to exploit data mining techniques on medical data set to assist in the prediction of the heart diseases.

3.2 Existing Methods

- Various different ML algorithms that can be used for classification of heart disease. Research was carried out to study Decision Tree, KNN and K-Means algorithms that can be used for classification and their accuracy were compared. This research tells that accuracy obtained by Decision Tree was highest further it was inferred that it can be made efficient by combination of different techniques and parameter tuning.

- Rapid Miner tool is used which results in higher accuracy compared to Matlab and Weka tool. The accuracy of Decision Tree, Logistic Regression, Random forest, Naive Bayes and SVM classification algorithms were compared. Decision tree algorithm had the highest accuracy.

- NB (Naïve Bayesian) techniques are used for classification of dataset and AES (Advanced Encryption Standard) algorithm is used for secure data transfer for prediction of disease.

Drawbacks

- As the main limitation in this work is the small size of the dataset. The dataset has limited number of patient's records; therefore, the dataset was augmented using appropriate techniques.
- It will take more time to train the datasets

3.3 Proposed system

- The proposed work predicts heart disease by exploring the above mentioned four classification algorithms and does performance analysis. The objective

of this study is to effectively predict if the patient suffers from heart disease. The health professional enters the input values from the patient's health report. The data is fed into model which predicts the probability of having heart disease.

- Data Collection and Preprocessing: The dataset used here is the Heart disease Dataset which is a combination of 4 different databases, but only the UCI Cleveland dataset was used. This database consists of a total of 76 attributes but all published experiments refer to using a subset of only 14 features. Therefore, we have used the already processed UCI Cleveland dataset available in the Kaggle website for our analysis.

Chapter 4:

SOFTWARE REQUIREMENT ANALYSIS

4.1 INTRODUCTION TO SRS

The introduction of the Software Requirements Specification (SRS) provides an overview of the entire SRS with purpose, scope, definitions, acronyms, abbreviations, references and overview of the SRS. The aim of this document is to gather, analyse, and give an in-depth insight of the complete "Heart disease prediction" by defining the problem statement in detail. The detailed requirements of the Fish disease Classification related functions are provided in this document.

4.2 PURPOSE

The Purpose of the Software Requirements Specification is to provide the technical, Functional and non-functional features, required to develop a web application App. The entire application designed to provide user flexibility for finding the shortest and/or time saving path. In short, the purpose of this SRS document is to provide a detailed overview of our software product, its parameters and goals. This document describes the project's target audience and its user interface, hardware and software requirements. It defines how our client, team and audience see the product and its functionality.

4.3 TECHNOLOGY:

Python:

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-

oriented approach aim to help programmers write clear, logical code for small and large-scale projects.[27]

Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library

Scikit Learn

Scikit-learn (formerly scikits.learn and also known as sklearn) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, *k*-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy. Scikit-learn is largely written in Python, and uses numpy extensively for high-performance linear algebra and array operations. Furthermore, some core algorithms are written in Cython to improve performance. Support vector machines are implemented by a Cython wrapper around LIBSVM; logistic regression and linear support vector machines by a similar wrapper around LIBLINEAR. In such cases, extending these methods with Python may not be possible.

Tensor flow:

TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks. It is used for both research and production at Google. TensorFlow is Google Brain's second-generation system. Version 1.0.0 was released on February 11, 2017. While the reference implementation runs on single devices, TensorFlow can run on multiple CPUs and GPUs (with optional CUDA and SYCL extensions for general-purpose computing on graphics processing units). TensorFlow is available on 64-bit Linux, macOS, Windows, and mobile computing platforms including Android and iOS.

Theano

Theano is a Python library and optimizing compiler for manipulating and evaluating mathematical expressions, especially matrix-valued ones. In Theano, computations are expressed using a NumPy-esque syntax and compiled to run efficiently on either CPU or GPU architectures.

Theano is an open source project primarily developed by a Montreal Institute for Learning Algorithms (MILA) at the Université de Montréal.

Numpy

NumPy or sometimes /ˈnʌmpi/ (*NUM-pee*)) is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. The ancestor of NumPy, Numeric, was originally created by Jim Hugunin with contributions from several other developers. In 2005, Travis Oliphant created NumPy by incorporating features of the competing Numarray into Numeric, with extensive modifications. NumPy is open-source software and has many contributors.

Pandas

In computer programming, pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license.[2] The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals.

OPEN CV

OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products.

Being a BSD-licensed product, OpenCV makes it easy for businesses to utilize and modify the code.

The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc. OpenCV has more than 47 thousand people of user community and estimated number of downloads exceeding 18 million. The library is used extensively in companies, research groups and by governmental bodies.

4.2 HARDWARE REQUIREMENTS

System	:	Pentium IV 2.4 GHz.
Hard Disk	:	250 GB.
Floppy Drive	:	1.44 Mb.
Monitor	:	15 VGA Colour
Mouse	:	Logitech.
Ram	:	1 GB.

4.3 SOFTWARE REQUIREMENTS

Operating system	:	Windows 10
Coding Language	:	Python
Software used	:	Anaconda, Jupiter Notebook

Chapter 5

SYSTEM ANALYSIS

Analysis is the process of finding the best solution to the problem. System analysis is the process by which we learn about the existing problems, define objects and requirements and evaluates the solutions. It is the way of thinking about the organization and the problem it involves, a set of technologies that helps in solving these problems. Feasibility study plays an important role in system analysis which gives the target for design and development.

5.1 Feasibility Study

All systems are feasible when provided with unlimited resource and infinite time. But unfortunately this condition does not prevail in practical world. So it is both necessary and prudent to evaluate the feasibility of the system at the earliest possible time. Months or years of effort, thousands of rupees and untold professional embarrassment can be averted if an ill- conceived system is recognized early in the definition phase. Feasibility & risk analysis are related in many ways. If project risk is great, the feasibility of producing quality software is reduced. In this case three key considerations involved in the feasibility analysis are

- Economical Feasibility
- Technical Feasibility
- Social Feasibility

5.1.1 Economical Feasibility

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Since the project is Machine learning based, the cost spent in executing this project would not demand cost for software and related products, as most of

the products are open source and free to use. Hence the project would consumed minimal cost and is economically feasible.

5.1.2 Technical Feasibility

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Since machine learning algorithms is based on pure math there is very less requirement for any professional software. Also, most of the tools are open source. The best part is that we can run this software in any system without any software requirements which makes them highly portable. Most of the documentation and tutorials make easy to learn the technology.

5.1.3 Social Feasibility

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The main purpose of this project which is based on creating an early prediction system of Heart disease. Thus, this is a noble cause for the sake of the society, a small step taken to achieve a secure and healthy future.

5.2 Analysis

5.2.1 Performance Analysis

Most of the software we use is open source and free. The models which we use in this software only are trained once in the beginning and there is no need to fed again in for the training phase. One can directly predict for values. Hence time-complexity is very less. Therefore this model is temporally sound.

5.2.2 Economical Analysis

Economic analysis is performed to evaluate the development cost weighed against the ultimate income or benefits derived from the developed system. The

completion of this project can be considered free of cost in its entirety. As the software used in building the model is free of cost and all the data sets used are being downloaded from Kaggle.

5.2.3 Technical Analysis

As mentioned earlier, the tools used in building this software is open source. Each tool contains simple methods and the required methods are overridden to tackle the problem.

Chapter 6

SYSTEM DESIGN

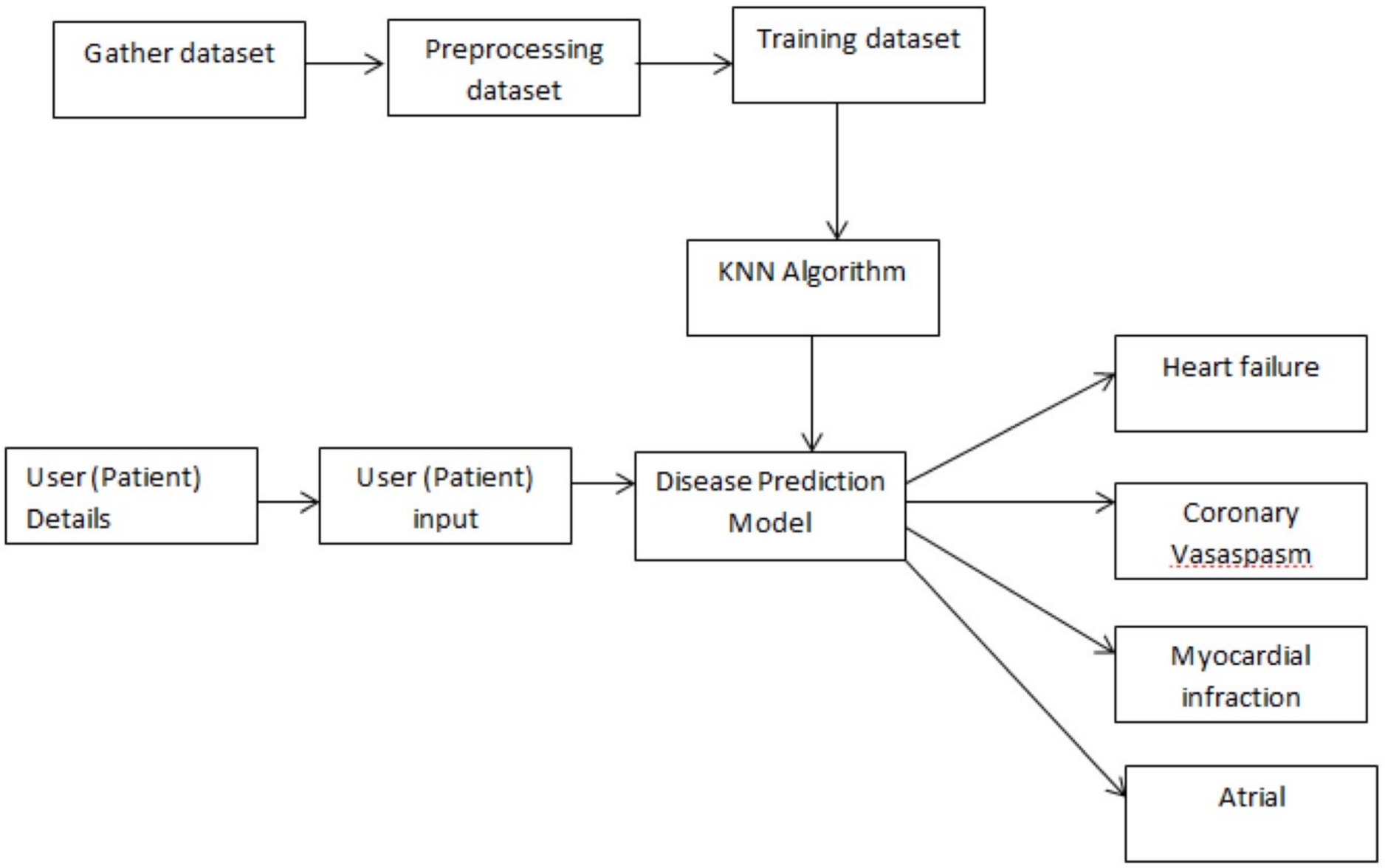
6.1 Introduction to Design document

The Software Design will be used to aid in software development for android application by providing the details for how the application should built. Within the Software Design, specifications are narrative and graphical documentation of the software design for the project includes use case models, sequence diagrams and other supporting requirement information.

6.1.1 Scope

This software Design Document is for a base level system, which will work as a proof of concept for the use of building a system that provides a base level of functionality to show feasibility for large-scale production use. The software Design Document, the focus placed on generation of the documents and modification of the documents. The system will used in conjunction with other pre-existing systems and will consist largely of a document interaction faced that abstracts document interactions and handling of the document objects. This Document provides the Design specifications of “Heart disease prediction”.

6.2 Architecture Diagram



- The dataset was taken from machine learning data repository.
- The data is cleaned and preprocessed before it is submitted to the proposed algorithm for training and testing.
- Feature selection aims to select and remove irrelevant or less important features.
- Dataset is divided into two parts, namely training data and testing data, where training set to used to train the machine learning model and testing set is used to test the model.
- Using Decision tree, K-Nearest Neighbour, Logistic Regression.

Data Collection and Pre-processing

The data set for this research was taken from UCI data repository.14 Data accessed from the UCI Machine Learning Repository is freely available. In particular, the Cleveland and Hungarian databases have been used by many researchers and found to be suitable for developing a mining model, because of lesser missing values and outliers. The data is cleaned and pre-processed before it is submitted to the proposed algorithm for training and testing.

The UCI Machine Learning Repository is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms.

The overall objective of our work is to predict more accurately the presence of heart disease. In this paper, UCI repository dataset are used to get more accurate results. Two data mining classification techniques were applied contains 76 attributes, but all published experiments refer to using a subset of 14 of them.

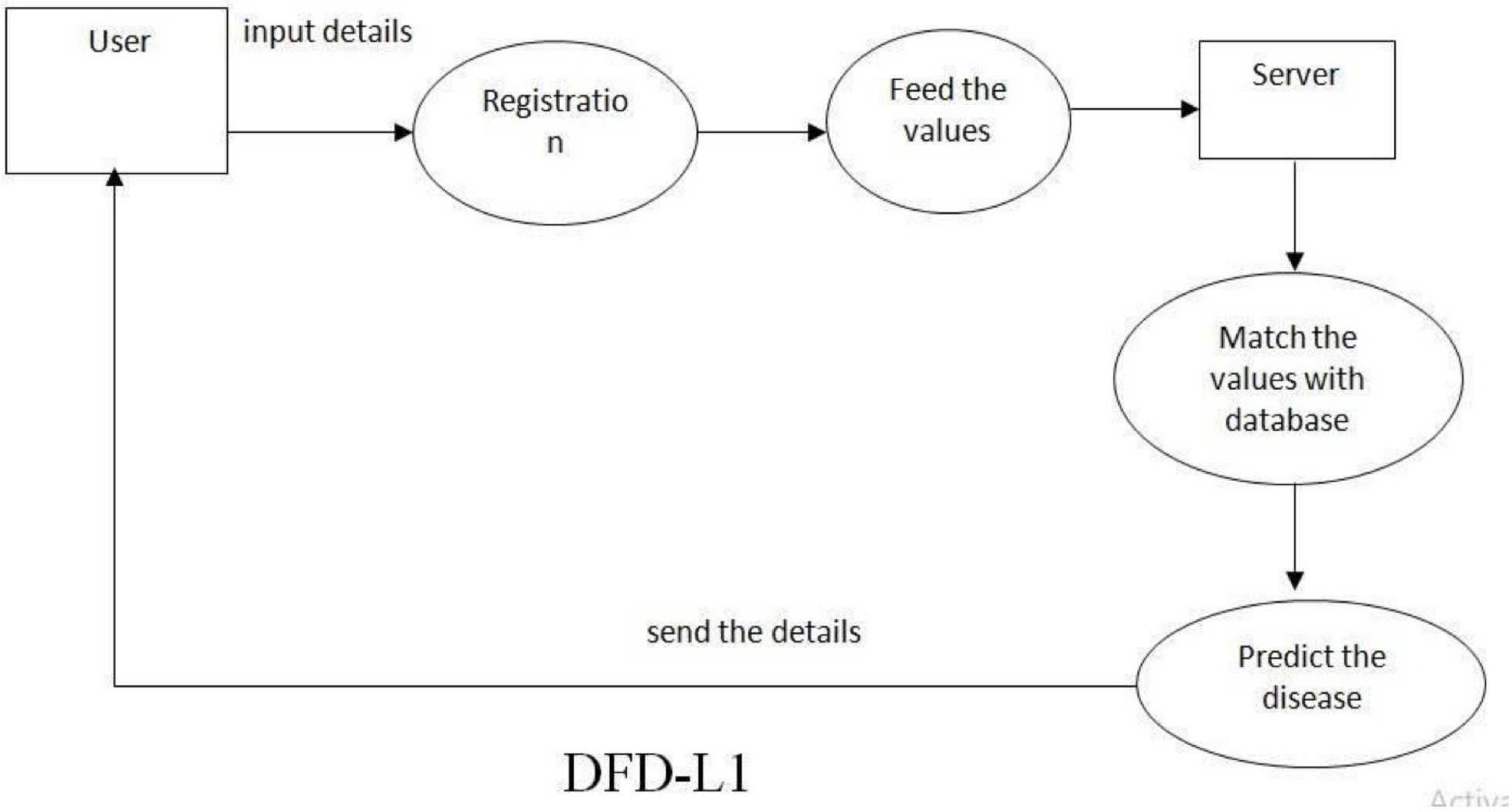
Classifiers Used for Experiments.

#	Attributes	Description	Values
1	Age	Patient's age in years	Continuous Value
2	Sex	Sex of Patient	1 = Male 0 = Female
3	Cp	Chest pain	Value 1: typical angina Value 2: atypical angina Value 3: non-angina pain Value 4: asymptomatic
4	Trestbps	Resting blood pressure	Continuous value in mm/Hg
5	Chol	Serum cholesterol in mg/dl	Continuous value in mg/dl
6	Fbs	Fasting blood sugar	1 ≥120 mg/dl 0 ≤ 120 mg/dl
7	Restcg	Resting electrocardiographic results	0 = normal 1 = having_ST_T wave abnormal 2 = left ventricular hypertrophy
8	Thalach	Maximum heart rate achieved	Continuous value
9	Exang	Exercise induced angina	1: yes 0: no
10	Oldpeak	ST depression induced by exercise relative to rest	Continuous value
11	Slope	the slope of the peak exercise ST segment	1: upsloping 2: flat 3: down sloping
12	Ca	number of major vessels colored by fluoroscopy	0-3 value
13	Thal	defect type	3 = normal 6 = fixed defect 7 = reversible defect
14	num	diagnosis of heart disease	no_heart_disease have_heart_disease

6.3 Data Flow Diagram

DFD-L0: This is the initial idea for the flow of the data. The data has to be flown from user to server and from server to the user for the prediction of the disease by entering details and sending the data.

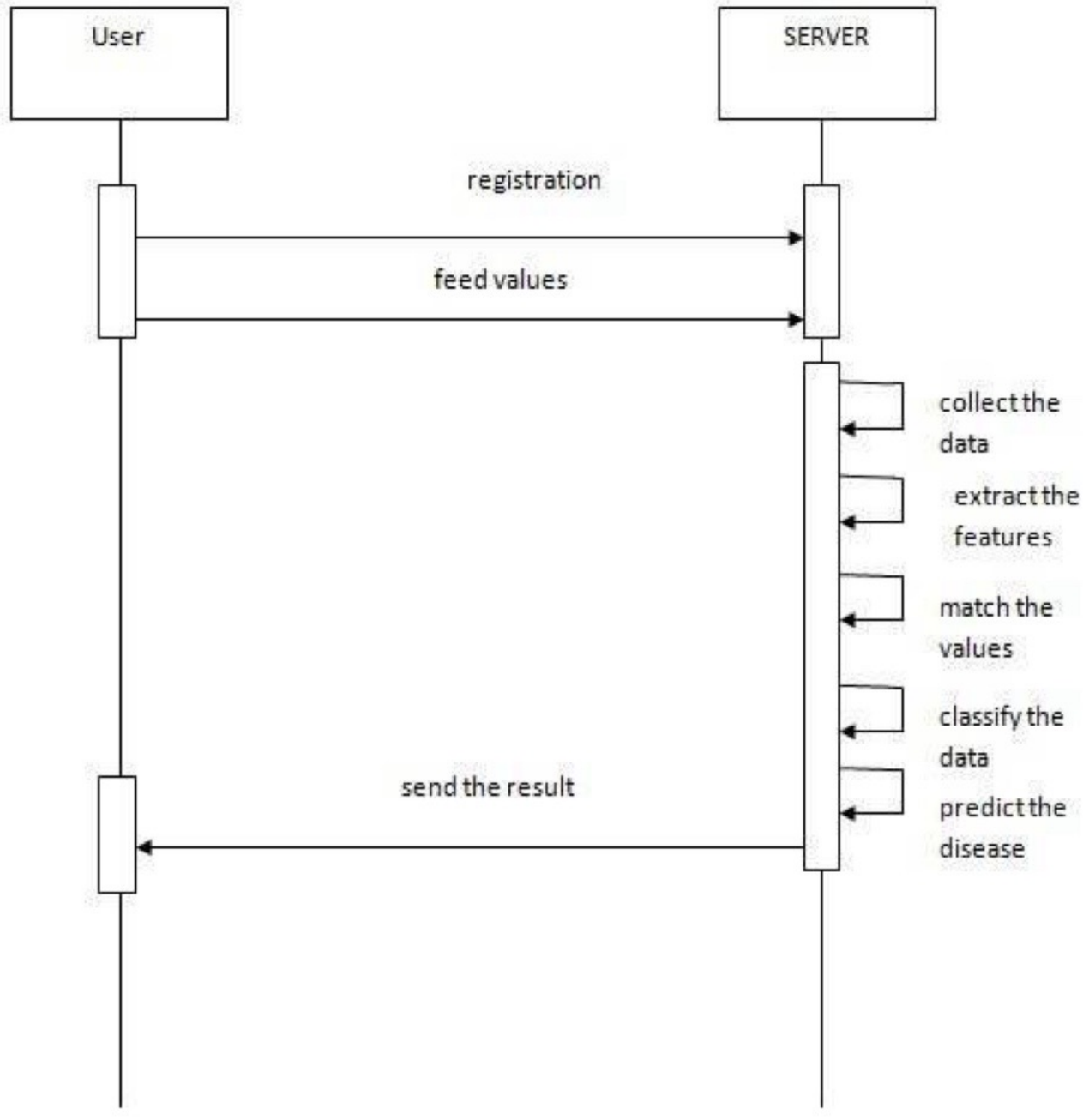
DFD-L1: This is the process or the idea where the data has been used to predict the disease by following several steps like registration (for new users), Feed the values (entering values), match the values and finally predict the disease (Final result). The registered users can login to their account and can enter the values that is data and then predict the result and then generate the report as similar to the newly registered users.



DFD-L1

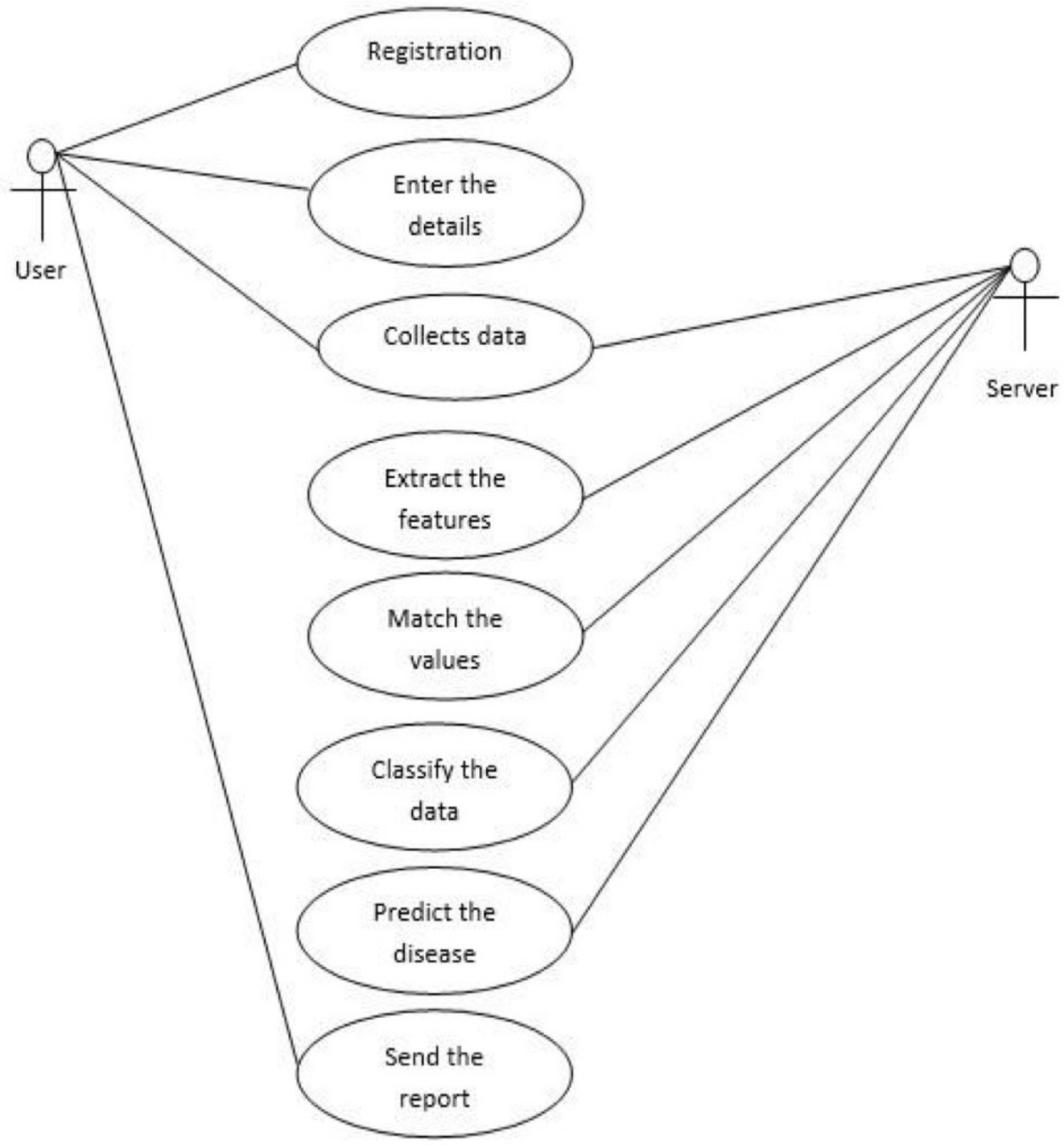
6.4 Sequence Diagram

The data which is flown from user to the system there it undergoes matching for data from the user input and the data which we have i.e datasets (trained data) Finding probability between them by comparing the values and then generating the report.



6.5 Use Case Diagram

The steps from registering the user i.e., beginning step to the final generating of the report can all be explained using use case diagram where actors are used as users. Users register by using certain parameters and then they login to their accounts and enter their health conditions and values i.e., data collection is taken into consideration i.e., the data from the users need to be collected and check for the disease and predict the disease and finally a report need to be generated.


Registration
Enter the details
Collects data
Extract the features
Match the values
Classify the data
Predict the disease
Send the report
User
Server

Chapter 7

IMPLEMENTATION

Introduction

The project is implemented using Python which is an object oriented programming language and procedure oriented programming language. Object oriented programming is an approach that provides a way of modularizing program by creating partitioned memory area of both data and function that can be used as a template for creating copies of such module on demand.

This project is implemented using python programming language. Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library. The machine Learning techniques are used in this project.

Machine Learning overview

Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of Computer Programs that can change when exposed to new data. In this article, we'll see basics of Machine Learning, and implementation of a simple machine learning algorithm using python.

Machine learning involves a computer to be trained using a given data set, and use this training to predict the properties of a given new data. For example, we can train a computer by feeding it 1000 images of cats and 1000 more images which are not of a cat, and tell each time to the computer whether a picture is cat or not. Then if we show the computer a new image, then from the above training, the computer should be able to tell whether this new image is a cat or not. The process of training and prediction involves the use of specialized algorithms. We feed the training data to an algorithm, and the algorithm uses this training data to give predictions on a new test data. One such algorithm is K-Nearest-Neighbour classification (KNN classification). It takes a test data, and finds k nearest data values to this data from test data set. Then it selects the neighbour of maximum frequency and gives its properties as the prediction result.

CHALLENGES IN IMPLEMENTING MACHINE LEARNING:

Most insurers recognize the value of machine learning in driving better decision-making and streamlining business processes. Research for the Accenture Technology Vision 2018 shows that more than 90 percent of insurers are using, plan to use or considering using machine learning or AI in the claims or underwriting process.

Some of the challenges insurers typically encounter when adopting machine learning are.

Training requirements AI-powered intellectual systems must be trained in a domain, e.g., claims or billing for an insurer. This requires a separate training system, which insurers find hard to provide for training the AI model. Models need to be trained with huge volumes of documents/transactions to cover all possible scenarios.

Right data source The quality of data used to train predictive models is equally important as the quantity, in the case of machine learning. The datasets need to be representative and balanced so that they can give a better picture and avoid bias. This is important to train predictive models. Generally, insurers struggle to provide relevant data for training AI models

Difficulty in predicting returns It's not very easy to predict improvements that machine learning can bring to a project. For example, it's not easy to plan or budget a project using machine learning, as the funding needs may vary during the project, based on the findings. Therefore, it is almost impossible to predict the return on investment. This makes it hard to get everyone on board the concept and invest in it.

Data security The huge amount of data used for machine learning algorithms has created an additional security risk for insurance companies. With such an increase in collected data and connectivity among applications, there is a risk of data leaks and security breaches. A security incident could lead to personal information falling into the wrong hands. This creates fear in the minds of insurers.

Source Code

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score
```

```
heart_data = pd.read_csv('heart.csv')
```

```
heart_data.head()
```

```
heart_data.tail()
```

```
heart_data.shape
```

```
heart_data.info()
```

```
heart_data.isnull().sum()
```

```
heart_data.describe()
print("The Target Value")
```

```
print("",heart_data['target'].value_counts())
```

```
X = heart_data.drop(columns='target', axis=1)
```

```
Y = heart_data['target']

print(X)

print(Y)

X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, stratify=Y,
random_state=2)

print(X.shape, X_train.shape, X_test.shape)

model = KNeighborsClassifier()

model.fit(X_train, Y_train)

X_train_prediction = model.predict(X_train)
training_data_accuracy = accuracy_score(X_train_prediction, Y_train)

print('Accuracy on Training data : ', training_data_accuracy)

X_test_prediction = model.predict(X_test)
test_data_accuracy = accuracy_score(X_test_prediction, Y_test)

print('Accuracy on Test data : ', test_data_accuracy)

input_data = (25,3,0,200,190,1,2,180,1,0.3,2,0,7)
```

```
input_data_as_numpy_array= np.asarray(input_data)
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)
prediction = model.predict(input_data_reshaped)
print(prediction)
if (prediction[0]== 0):
    print('The Person has Heart failure')
elif (prediction[0] == 1):
    print('The Person has Myocardial infraction')
elif (prediction[0] == 2):
    print('The Person has Dilated cardiomyopathy')
elif (prediction[0] == 3):
    print('The Person has Coronary vasaspasm')
elif (prediction[0] == 4):
    print('The Person has Atrial fibrillation')
else:

    print('The Person has Arrhythmia- abnormal heart rythmn')
```

Chapter 8

TESTING

INTRODUCTION

Testing is the way toward running a framework with the expectation of discovering blunders. Testing upgrades the uprightness of the framework by distinguishing the deviations in plans and blunders in the framework. Testing targets distinguishing blunders – prom zones. This aides in the avoidance of mistakes in the framework. Testing additionally adds esteems to the item by affirming the client's necessity.

The primary intention is to distinguish blunders and mistake get-prom zones in a framework. Testing must be intensive and all around arranged. A somewhat tried framework is as terrible as an untested framework. Furthermore, the cost of an untested and under-tried framework is high. The execution is the last and significant stage. It includes client preparation, framework testing so as to guarantee the effective running of the proposed framework. The client tests the framework and changes are made by their requirements. The testing includes the testing of the created framework utilizing different sorts of information. While testing, blunders are noted and rightness is the mode.

OBJECTIVES OF TESTING

•Testing in a cycle of executing a program with the expectation of discovering mistakes.

•A effective experiment is one that reveals an up 'til now unfamiliar blunder.

Framework testing is a phase of usage, which is pointed toward guaranteeing that the framework works accurately and productively according to the client's need before the live activity initiates. As expressed previously, testing is indispensable to the achievement of a framework. Framework testing makes the coherent presumption that if all the framework is right, the objective will be

effectively accomplished. A progression of tests are performed before the framework is prepared for the client acknowledgment test.

TESTING METHODS

System testing is a stage of implementation. This helps the weather system works accurately and efficiently before live operation commences. Testing is vital to the success of the system. The candidate system is subject to a variety of tests: online response, volume, stress, recovery, security, and usability tests series of tests are performed for the proposed system are ready for user acceptance testing.

White Box Testing

The test is conducted during the code generation phase itself. All the errors were rectified at the moment of its discovery. During this testing, it is ensured that

- All independent module have been exercised at least one
- Exercise all logical decisions on their true or false side.
- Execute all loops at their boundaries.

Black Box Testing

It is focused around the practical necessities of the product. It's anything but a choice to white box testing; rather, it is a reciprocal methodology that is probably going to reveal an alternate class of blunders than White Box strategies. It is endeavored to discover mistakes in the accompanying classes.

- Incorrect or missing capacities

- Interface blunders
- Errors in an information structure or outside information base access

Unit Testing

Unit testing chiefly centers around the littlest unit of programming plan. This is known as module testing. The modules are tried independently. The test is done during the programming stage itself. In this progression, every module is discovered to be working acceptably as respects the normal yield from the module.

Integration Testing

Mix testing is an efficient methodology for developing the program structure, while simultaneously leading tests to reveal blunders related with the interface. The goal is to take unit tried modules and manufacture a program structure. All the modules are joined and tried in general.

Output Testing

Subsequent to performing approval testing, the following stage is yield trying of the proposed framework, since no framework could be valuable on the off chance that it doesn't create the necessary yield in a particular configuration. The yield design on the screen is discovered to be right. The organization was planned in the framework configuration time as indicated by the client needs. For the printed copy likewise, the yield comes according to the predefined prerequisites by the client. Subsequently yield testing didn't bring about any amendment for the framework.

User Acceptance Testing

Client acknowledgment of a framework is the vital factor for the achievement of any framework. The framework viable is tried for client acknowledgment by continually staying in contact with the imminent framework clients at the hour of creating and making changes at whatever point required.

VALIDATION

Toward the consummation of the reconciliation testing, the product is totally amassed as bundle interfacing blunders have been revealed and adjusted and a last arrangement of programming tests starts in approval testing. Approval testing can be characterized from multiple points of view, however a straightforward definition is that the approval succeeds when the product work in a way that is normal by the client. After approval test has been directed as follows:

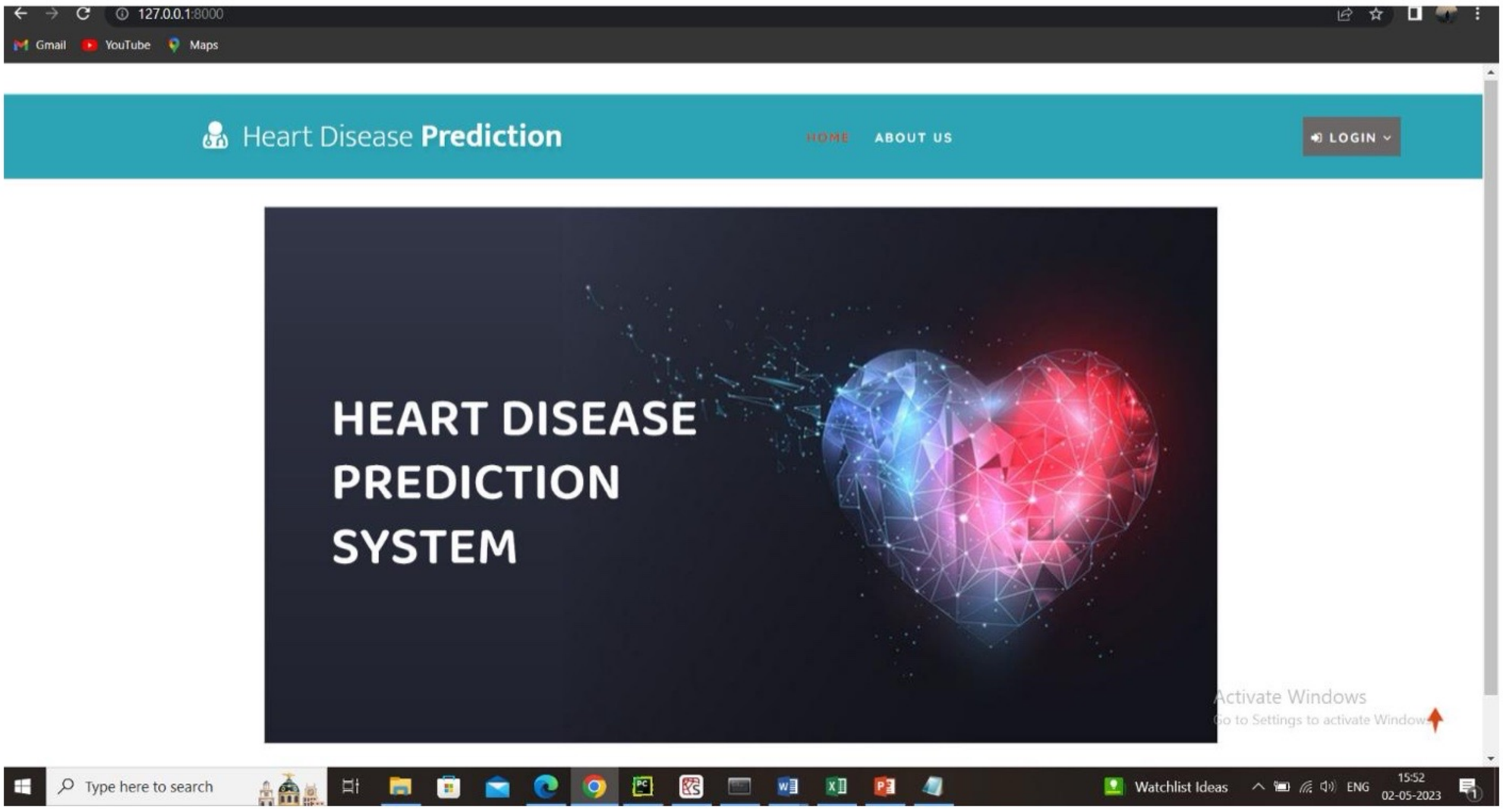
- The capacity or execution qualities adjust to detail and are acknowledged.
- A deviation from the particular is revealed and a lack list is made.
- Proposed framework viable has been tried by utilizing an approval test and discovered to be working acceptably.

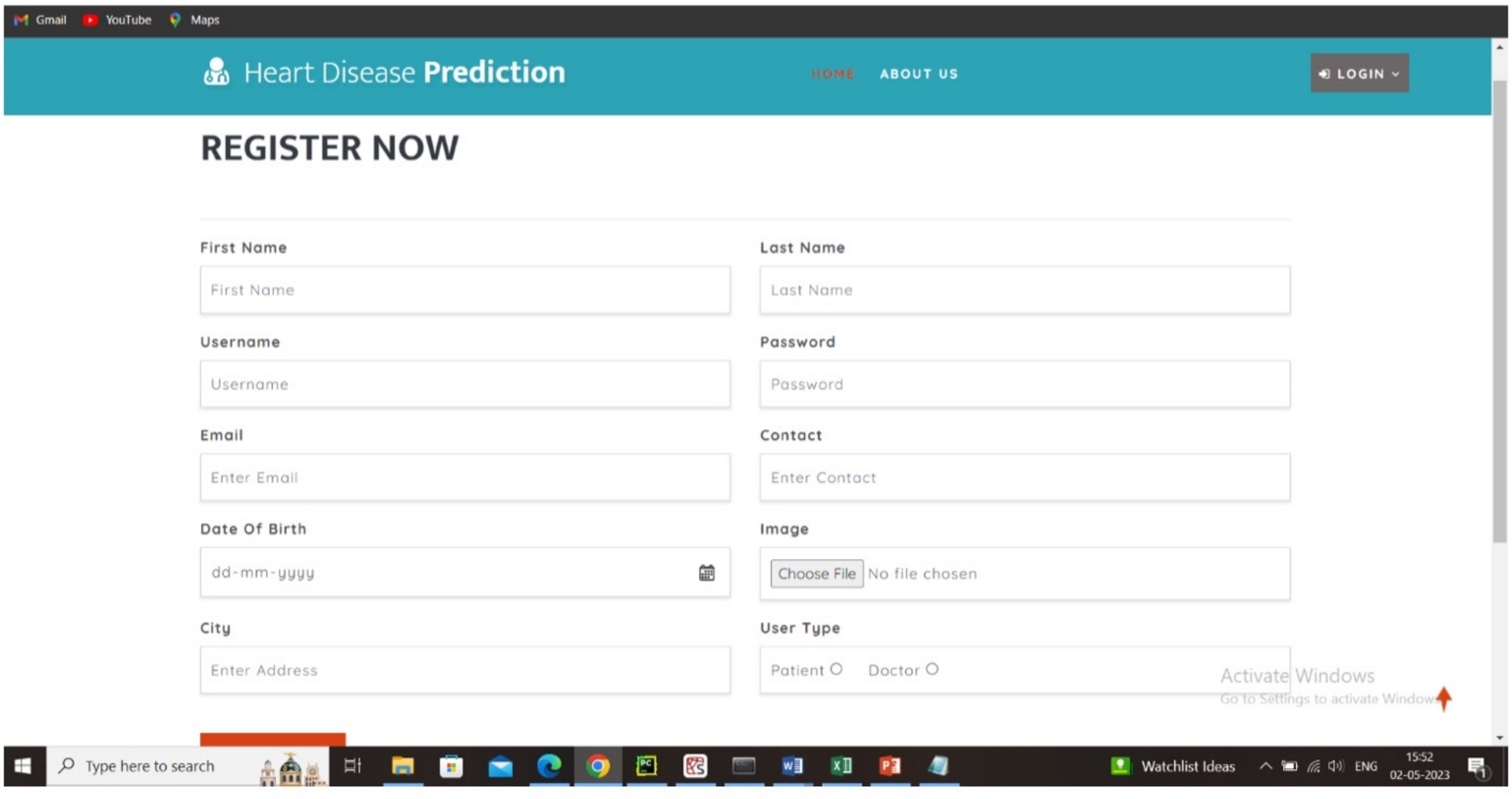
Test Cases

Test Case	**Test Purpose**	**Test condition**	**Expected outcome**	**Actual result**	**Pass or Fail**
Load Construction cost data	Upload data	Check for csv data	Uploaded successfully	Csv file is loaded Successfully.	Pass
Pre processing Of data	CSV data	Pre-processing of input data 1.filling missing values 2.data analysis 3.outliers	Pre Processing is done successfully	Csv file is pre pre-processed successfully.	Pass
Split data into train and test	Pre processed Of data	80% data for training and 20% data for testing	Data is splitted into x train and y train	As expected	Pass
Apply the regression algorithms	Machine learning algorithms	Pass train data to regression algorithms	Data is predicted successfully	As expected	Pass
Comparison result	Choose the best algorithms for prediction	Check for accuracy for each algorithms	Comparison result	Result is shown.	Pass

Chapter 9

RESULTS AND SNAPSHOTS





Gmail YouTube Maps

Heart Disease **Prediction** HOME PREDICT MY DETAIL FEEDBACK HISTORY HELLO,SUSHMA

HEART PREDICTION OUTPUT

Prediction output

Accuracy (%) is : 74.8971193415638

Result: You Have no Heart Failure Chances

© Facebook Twitter Google Plus

Activate Windows
Go to Settings to activate Windows.

Type here to search Watchlist Ideas ENG 15:51 02-05-2023

(2) WhatsApp | Heart Disease Prediction System

127.0.0.1:8000/predict_desease/3/51.028806584362144/

Gmail YouTube Maps

Heart Disease **Prediction** HOME PREDICT MY DETAIL FEEDBACK HISTORY HELLO,KEERTHANA

HEART PREDICTION OUTPUT

Prediction output

Accuracy (%) is : 51.028806584362144

Result: You are Coronary vasaspasm , Need to Checkup.

CONTACT OUR DOCTORS

#	Image	Full Name	Email	Contact	Address
No Record Found.					

Type here to search 29°C Rain ENG 15:09 06-05-2023

(2) WhatsApp | Heart Disease Prediction System

127.0.0.1:8000/predict_desease/4/51.028806584362144/

Gmail YouTube Maps

Heart Disease **Prediction**

HOME PREDICT MY DETAIL FEEDBACK HISTORY HELLO,KEERTHANA

HEART PREDICTION OUTPUT

Prediction output

Accuracy (%) is : 51.028806584362144

Result: You are Atrial fibrillation, Need to Checkup.

CONTACT OUR DOCTORS

#	Image	Full Name	Email	Contact	Address
No Record Found.					

Type here to search 29°C Rain ENG 15:09 06-05-2023

(2) WhatsApp | Heart Disease Prediction System

127.0.0.1:8000/predict_desease/1/51.028806584362144/

Gmail YouTube Maps

Heart Disease **Prediction**

HOME PREDICT MY DETAIL FEEDBACK HISTORY HELLO,KEERTHANA

HEART PREDICTION OUTPUT

Prediction output

Accuracy (%) is : 51.028806584362144

Result: You are Myocardial infraction , Need to Checkup.

CONTACT OUR DOCTORS

#	Image	Full Name	Email	Contact	Address
No Record Found.					

Type here to search 29°C Rain ENG 15:10 06-05-2023

CONCLUSION

Heart disease detection model has been developed using ML classification modelling techniques. This project predicts people with cardiovascular disease by extracting the patient medical history that leads to a fatal heart disease from a dataset that includes patients medical history such as chest pain, sugar level, blood pressure, etc. This Heart Disease detection system assists a patient based on his/her clinical information of them been diagnosed with a previous heart disease. The algorithms used in building the given model are Logistic regression, and KNN. Use of more training data ensures the higher chances of the model to accurately predict whether the given person has a heart failure chances, Myocardial infraction, Atrial fibrillation, coronary vasaspasm.

REFERENCES

[1] Avinash Golande, Pavan Kumar T, ”Heart Disease Prediction Using Effective Machine Learning Techniques”, International Journal of Recent Technology and Engineering, Vol 8, pp.944-950,2019.

[2] T.Nagamani, S.Logeswari, B.Gomathy,” Heart Disease Prediction using Data Mining with Mapreduce Algorithm”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019.

[3] Fahd Saleh Alotaibi,” Implementation of Machine Learning Model to Predict Heart Failure Disease”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.

[4] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, ”Design And Implementation Heart Disease Prediction Using Naives Bayesian”, International Conference on Trends in Electronics and Information(ICOEI 2019).

[5] Theresa Princy R,J. Thomas,’Human heart Disease Prediction System using Data Mining Techniques’, International Conference on Circuit Power and Computing Technologies,Bangalore,2016.

[6] Nagaraj M Lutimath,Chethan C,Basavaraj S Pol.,’Prediction Of Heart Disease using Machine Learning’, International journal Of Recent Technology and Engineering,8,(2S10), pp 474-477, 2019.

[7] UCI, —Heart Disease Data Set.[Online]. Available (Accessed on May 1 2020): https://www.kaggle.com/ronitf/heart-disease-uci.

[8] Sayali Ambekar, Rashmi Phalnikar,“Disease Risk Prediction by Using Convolutional Neural Network”,2018 Fourth International Conference on Computing Communication Control and Automation.

[9] C. B. Rjeily, G. Badr, E. Hassani, A. H., and E. Andres, —Medical Data Mining for Heart Diseases and the Future of Sequential Mining in Medical Field,‖ in Machine Learning Paradigms, 2019, pp. 71–99.

[10] Jafar Alzubi, Anand Nayyar, Akshi Kumar. "Machine Learning from Theory to Algorithms: An Overview", Journal of Physics: Conference Series, 2018

[11] Fajr Ibrahem Alarsan., and Mamoon Younes ‘Analysis and classification of heart diseases using heartbeat features and machine learning algorithms’,Journal Of Big Data,2019;6:81.