MULTIPLE DISEASE PREDICTION SYSTEM

A project report submitted to

Rajiv Gandhi Proudyogiki Vishwavidhyalaya, Bhopal

in partial fulfillment for the award of

the degree of

Bachelor of Technology

in

Computer Science & Engineering

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

SUSHILA DEVI BANSAL COLLEGE OF TECHNOLOGY
INDORE- 453331

2022 - 2023

MULTIPLE DISEASE PREDICTION SYSTEM

A project report submitted to

Rajiv Gandhi Proudyogiki Vishwavidhyalaya, Bhopal

in partial fulfillment for the award of

the degree of

Bachelor of Technology

in

Computer Science & Engineering

PROJECT GUIDE
Prof. Shailendra Singh Bhalla

SUBMITTED BY
Ishita Agrawal 0829CS191060
Ankita Vishwas 0829CS191020

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

SUSHILA DEVI BANSAL COLLEGE OF TECHNOLOGY
INDORE- 453331

2022 - 2023

ACKNOWLEDGEMENT

With great pleasure and a sense of obligation, we express my heartfelt gratitude to **Prof. Shailendra Singh Bhalla** our project guide and project coordinator. We are highly indebted to him for his valuable guidance and ever-ready support in completing this project in the stipulated time. Also, we express our gratitude to him for his persisting encouragement, perpetual motivation, everlasting patience, and excellent expertise in discussions, during the progress of the project work, which have benefited us to an extent, which is beyond expressing. His contributions are beyond the preview of acknowledgment.

We are again grateful to **Prof. Arjun Singh Parihar**, Head of the Computer Science Department, SDBCT, Indore, for his encouragement throughout my project work.

We would like to give a warm expression of thanks to **Dr. Premanand Singh Chouhan** Director of SDBC, Indore for providing all facilities and academic environment for my project work.

We sincerely wish to express our gratefulness to all the members of the staff. We sincerely thank all our friends and well-wishers for directly or indirectly helping during work.

Ishita Agrawal 0829CS191060
Ankita Vishwas 0829CS191020

SUSHILA DEVI BANSAL COLLEGE OF TECHNOLOGY
INDORE, 453331

CERTIFICATE

This is to certify that **ISHITA AGRAWAL(0829CS191060) and ANKITA VISHWAS(0829CS191020)** have completed their project work, titled **"MULTIPLE DISEASE PREDICTION SYSTEM"** as per the syllabus and have submitted a satisfactory report on this project as a part of fulfillment towards the degree of **"BACHELOR OF TECHNOLOGY" (Computer Science & Engineering)** from **RAJIV GANDHI PROUDYOGIKI VISHWAVIDHYALAYA, BHOPAL.**

HEAD OF THE DEPARTMENT **PROJECT GUIDE**

DIRECTOR

SUSHILA DEVI BANSAL COLLEGE OF TECHNOLOGY
INDORE, 453331

CERTIFICATE

This is to certify that **ISHITA AGRAWAL(8029CS191060) AND ANKITA VISHWAS(0829CS191020)** have completed their project work, titled **"MULTIPLE DISEASE PREDICTION SYSTEM"** as per the syllabus and have submitted a satisfactory report on this project as a part of fulfillment towards the degree of **"BACHELOR OF TECHNOLOGY" (Computer Science & Engineering)** from **RAJIV GANDHI PROUDYOGIKI VISHWAVIDHYALAYA, BHOPAL.**

INTERNAL EXAMINER **EXTERNAL EXAMINER**

TABLE OF CONTENTS

ABSTRACT

Machine learning and Artificial Intelligence are playing a huge role in today's world. From self-driving cars to medical fields, we can find them everywhere. The medical industry generates a huge amount of patient data which can be processed in a lot of ways. So, with the help of machine learning, we have created a Prediction System that can detect more than one disease at a time. Many of the existing systems can predict only one disease at a time and that too with lower accuracy. Lower accuracy can seriously put a patient's health in danger. We have considered three diseases for now which are Heart, Parkinson's, and Diabetes and in the future, many more diseases can be added. The user has to enter various parameters of the disease and the system would display the output whether he/she has the disease or not. This project can help a lot of people as one can monitor the person's condition and take the necessary precautions thus increasing life expectancy.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

ECG	Electrocardiography
PPE	Pitch Period Entropy
DFA	Detrended Fluctuation Analysis
RPDE	Recurrence Period Density Entropy
SVM	Support Vector Machine
CSV	Comma Separated Values
API	Application Programming Interface

I. Software Requirement Specification

1.1 Introduction

1.1.1 Purpose

The main motivation for doing this research is to present a multiple disease prediction model for the prediction of the occurrence of heart, diabetes, and Parkinson's disease. Further, this research work is aimed at identifying the best classification algorithm for identifying the possibility of disease in a patient. This work is justified by performing a comparative study and analysis using the classification algorithms namely Logistic Regression and SV, are used at different levels of evaluation. Although these are commonly used machine learning algorithms, heart disease prediction is a vital task involving the highest possible accuracy. Hence, both the algorithms are evaluated at numerous levels and types of evaluation strategies. This will provide researchers and medical practitioners to establish a better.

1.1.2 Scope

As we know today life has become so fast and anyone can get a heart attack anytime and anywhere. People hardly get time to go to the hospital far away from their places and get a body(heart) checkup so we have attempted so that people can have fast and easy access to their heart-related health measures whenever and wherever required.

We have tried our level best to make software that provides everyone with this facility and ease.

1.1.3 Problem in the existing system

Many of the existing machine learning models for health care analysis are concentrating on one disease per analysis. For example first is for liver analysis, one for cancer analysis, and one for lung diseases like that. If a user wants to predict more than one disease, he/she has to go through different sites. There is no common system where one analysis can perform more than one disease prediction. Some of the models have lower accuracy which can seriously affect patients' health. When an organization wants to analyze their patient's health reports, they have to deploy many models which in turn increases the cost as well as time. Some of the existing systems consider very few parameters which can yield false results.

1.1.4 Statement of Problem

In multiple disease prediction, it is possible to predict more than one disease at a time. So the user doesn't need to traverse different sites to predict the diseases. We are taking three diseases that are Parkinson's, Diabetes, and Heart. As all three diseases are correlated with each other. To implement multiple disease analyses we are going to use machine learning algorithms and Streamlit. When the user is accessing this API, the user has to send the parameters of the disease along with the disease name. Streamlit will invoke the correspond-ding model and returns the status of the patient.

II. System Requirement Analysis

2.1 Overall Description

2.1.1 Product Perspective

The perspective of this project called "Multiple Disease Prediction System" is to predict the accurate disease of the patient using all their general information and also the symptoms. Using this information there we will compare with our datasets of the patients and predicts the disease which he/she is been through.

2.1.2 Product Functions

The main function of this product is that it can predict multiple diseases for a patient in one platform only. The user does not have to download new software to predict his/her multiple diseases.

2.1.3 User Classes and Characteristics

Diabetes Disease:

1. No. of pregnencies
2. Plasma glucose concentration in oral glucose tolerance test
3. Blood Pressure(mm/Hg)
4. Skin fold thickness(mm)
5. 2 hr serum insulin
6. Body mass index(Kg/m^2)
7. Diabetes pedigree function
8. Age in years

Heart Disease:

1 Age: displays the age of the individual.

2 Sex: displays the gender of the individual using the following format: 1 = male 0 = female 2 = trans.

3 Chest-pain type: 1 = typical angina, 2 = atypical angina, 3 = non - anginal pain, 4 = asymptotic

4 Blood Pressure mmHg (unit)

5 Serum Cholesterol mg/dl (unit)

6 Fasting Blood Sugar: If fasting blood sugar > 120mg/dl then : 1 (true) else : 0 (false)

7 Resting ECG : 0 = normal 1 = having ST-T wave abnormality 2 = left ventricular hypertrophy

8 Max heart rate achieved.

9 Exercise-induced angina : 1 = yes 0 = no

10 ST depression induced by exercise relative to rest

11 Peak exercise ST segment : 1 = upsloping 2 = flat 3 = down sloping

12 Number of major vessels (0-3) colored by fluoroscopy

13 Thal : displays the thalassemia

Parkinson's Disease:

1 Gender

2 PPE

3 DFA

4 RPDE

5 numPulses

6 numPeriodPulses

7 meanPeriodPulses

8 stdDevPeriodPulses

9 locPctJitter

10 locAbsJitter

11 rapJitter

12 locShimmer, etc

2.2 External Interface Requirements

2.2.1 Software Requirements

Name of Component	Specification
Operating System	Windows 9, Windows XP, Windows 7, Linux
Language	Python 3.7
Database	MS Excel
Browser	Any of Mozilla, Opera, Chrome, etc.
Web Server	Streamlit
Software Development Kit	Streamlit
Scripting Language Enable	Python Script

Table 2.1 – Software Requirements

2.2.2 Hardware Requirements

Name of Component	Specification
Processor	Pentium III 630 MHz
RAM	128 MB
Hard Disk	20 GB
Monitor	15'' color monitor
Keyboard	122 keys

Table 2.2 – Hardware Requirements

2.3 Functional Requirement

2.3.1 Dataset preparation and pre-processing

Data collection is defined as the procedure of collecting, measuring, and analyzing accurate insights for research using standard validated techniques. A researcher can evaluate their hypothesis based on collected data. In most cases, data collection is the primary and most important step for research, irrespective of the field of research. The approach to data collection is different for different fields of study, depending on the required information. The most critical objective of data collection is ensuring that information-rich and reliable data is collected for statistical analysis so that data-driven decisions can be made for research.

2.3.2 Data Visualization

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. Example –

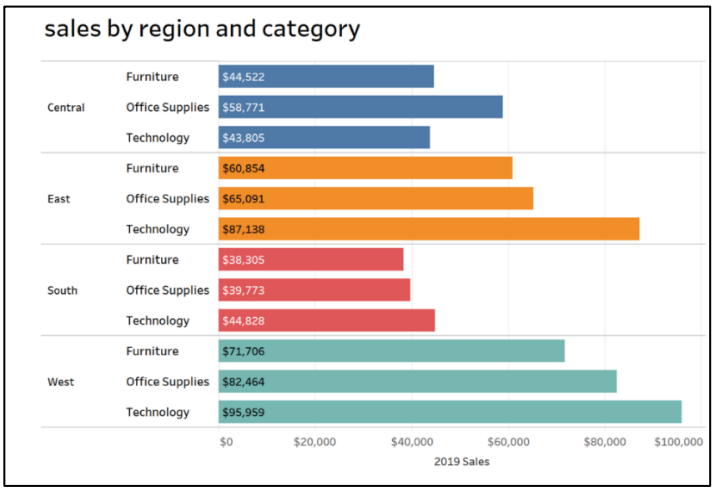


Fig. 2.1 – Data Visualization

2.3.3 Data Labelling

Supervised machine learning, which we'll talk about below, entails training a predictive model on historical data with predefined target answers. An algorithm must be shown which target answers or attributes to look for. Mapping these target attributes in a dataset is called labeling. Data labeling takes much time and effort as datasets sufficient for machine learning may require thousands of records to be labeled. For instance, if your image recognition algorithm must classify types of bicycles, these types should be clearly defined and labeled in a dataset.

2.4.4 Data Selection

Data selection is defined as the process of determining the appropriate data type and source, as well as suitable instruments to collect data. Data selection precedes the actual practice of data collection. This definition distinguishes data selection from selective data reporting (selectively excluding data that is not supportive of a research hypothesis) and interactive/active data selection (using collected data for monitoring activities/events, or conducting secondary data analyses). The process of selecting suitable data for a research project can impact data integrity. After having collected all information, a data analyst chooses a subgroup of data to solve the defined problem. For instance, if you save your customers' geographical location, you don't need to add their cell phones and bank card numbers to a

dataset. But purchase history would be necessary. The selected data includes attributes that need to be considered when building a predictive model.

2.3.5 Data Pre-processing

Data pre-processing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. The purpose of pre-processing is to convert raw data into a form that fits machine learning. Structured and clean data allows a data scientist to get more precise results from an applied machine learning model. The technique includes data formatting, cleaning, and sampling.

2.3.6 Data Splitting

A dataset used for machine learning should be partitioned into three subsets — training, test, and validation sets.

Training set - A data scientist uses a training set to train a model and define its optimal parameters it must learn from data.

Test set - A test set is needed for an evaluation of the trained model and its capability for generalization. The latter means a model's ability to identify patterns in new unseen data after having been trained over training data. It is crucial to use different subsets for training and testing to avoid model overfitting, which is the incapacity for generalization we mentioned above.

Validation set - The purpose of a validation set is to tweak a model's hyperparameters-higher-level structural settings that cannot be directly learned from data. These settings can express, for instance, how complex a model is and how fast it finds patterns in data.

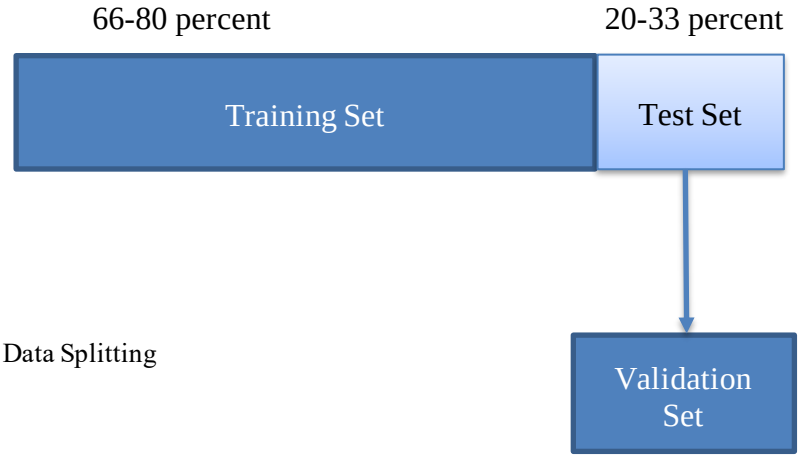


Fig 2.2 – Data Splitting

2.3.7 Modelling

After pre-processing the collected data and splitting it into three subsets, we can proceed with model training. This process entails "feeding" the algorithm with training data. An algorithm will process data and output a model that can find a target value (attribute) in new data - an answer you want to get with predictive analysis. The purpose of model training is to develop a model.

2.3.8 Model Deployment

Deployment is the method by which you integrate a machine-learning model into an existing production environment to make practical business decisions based on data. It is one of the last stages in the machine learning life cycle and can be one of the most cumbersome. Often, an organization's IT systems are incompatible with traditional model-building languages, forcing data scientists and programmers to spend valuable time and brainpower rewriting them.

2.4 Nonfunctional Requirements

2.4.1 Usability

The system should be easy to use. The system also should be user-friendly for users because anyone can use it instead of programmers.

2.4.2 Reliability

This software will be developed with machine learning, feature engineering, and deep learning techniques. So, in this step, there is no certain reliable percentage that is measurable. Also, user-provided data will be used to compare with results and measure reliability. With recent machine learning techniques, user-gained data should be enough for reliability if enough data is obtained.

2.4.3 Performance

Processing time and response time should be as little as possible providing the result at a faster rate when compared to other methods.

2.4.4 Supportability

The system should require Python knowledge to maintain. If any problem acquires on the user side and deep learning methods, it requires code knowledge and deep learning background to solve.

2.5 Project Plan

2.5.1 Team Members

- Ishita Agrawal
  Roll no. - 0829CS191060

- Ankita Vishwas
  Roll No. - 0829CS191020

2.5.2 Division of Work

Ankita Vishwas

- Dataset preparation and pre-processing
- Data Visualization
- Data Labelling
- Data Selection

Ishita Agrawal

- Data Pre-processing
- Data Splitting
- Modeling
- Model Deployment

2.5.3 Time Schedule

Sprint 1	**Sprint 2**	**Sprint 3**	**Sprint4**	**Sprint 5**
June 2022	July 2022 – August 2022	September 2022	October 2022	November 2022
Discussion of Project Selection	Literature survey, Synopsis, SDD, Research Notes	Freezing of algorithm, framework, and software	Coding and implementation	Testing and adding last-minute features

Table 2.3 – Time Schedule

III. Analysis

3.1 Methodology Used

Agile:

Agile is a process by which a team can manage a project by breaking it up into several stages and involving constant collaboration with stakeholders and continuous improvement and iteration at every stage. It promotes continuous iteration of development and testing throughout the software development life cycle of the project. Both development and testing activities are concurrent.

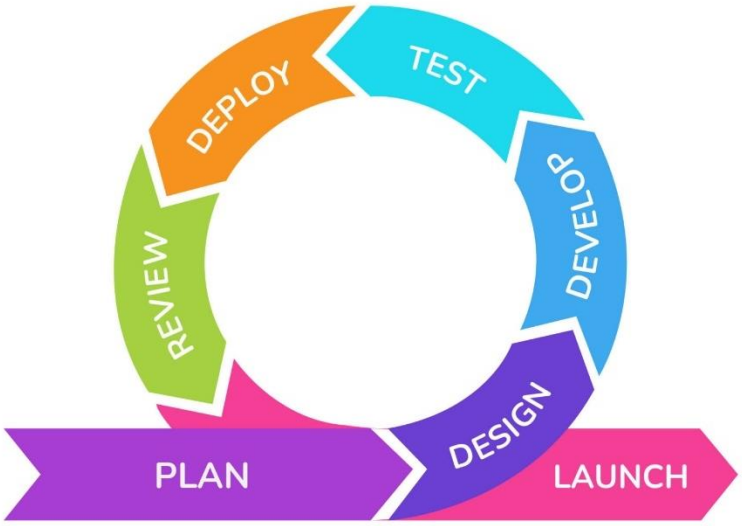


Fig 3.1 – Agile Methodology

Scrum:

SCRUM is an agile development method that concentrates specifically on how to manage tasks within a team-based development environment. Scrum encourages teams to learn through experiences, and self-organize while working on a problem.

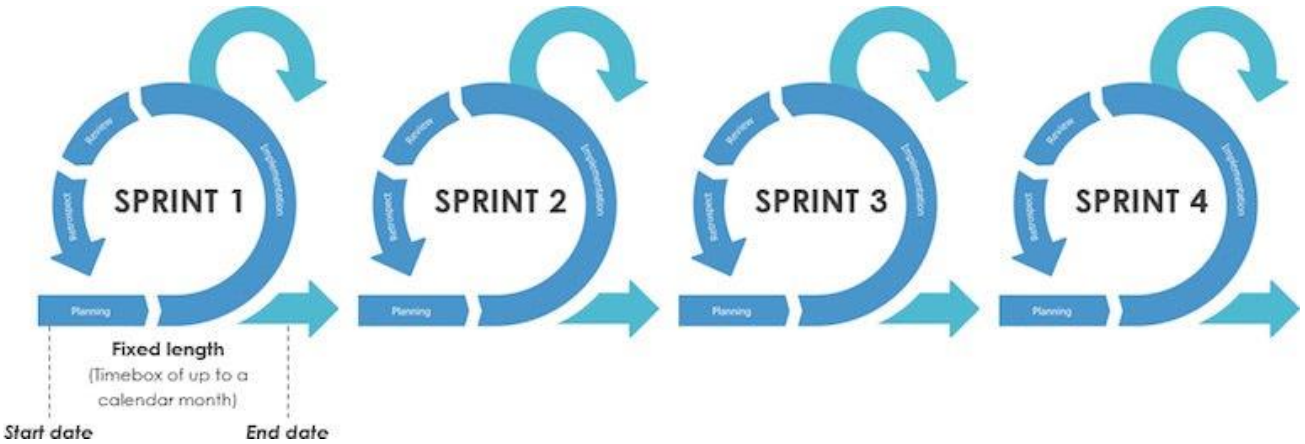


Fig 3.2 - Scrum

3.2 Use Case diagram

Though use cases are not recommended by agile development or feature-driven development, use cases were created to better understand the requirements of the system. The use cases are packaged according to domain areas. A detailed description of only some of the use cases is present in this section.

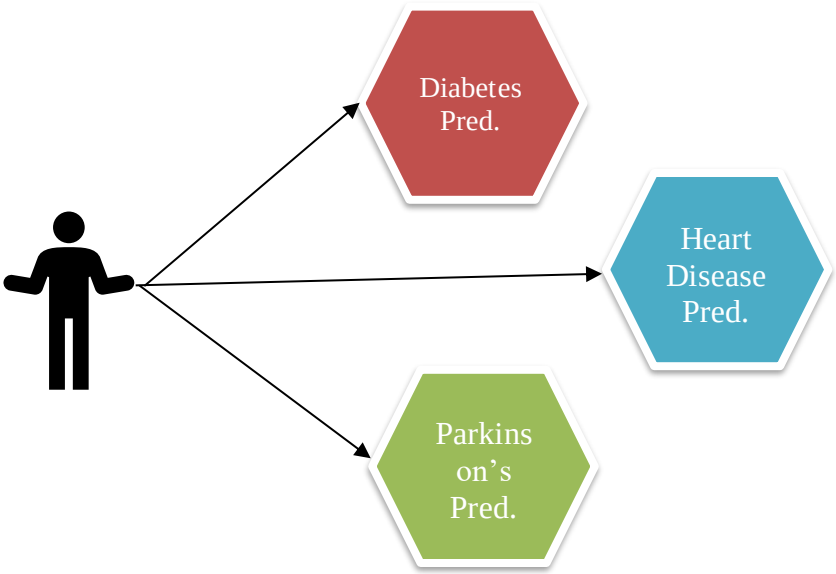


Fig 3.3 – Use Case Diagram

3.3 Data Flow Diagrams

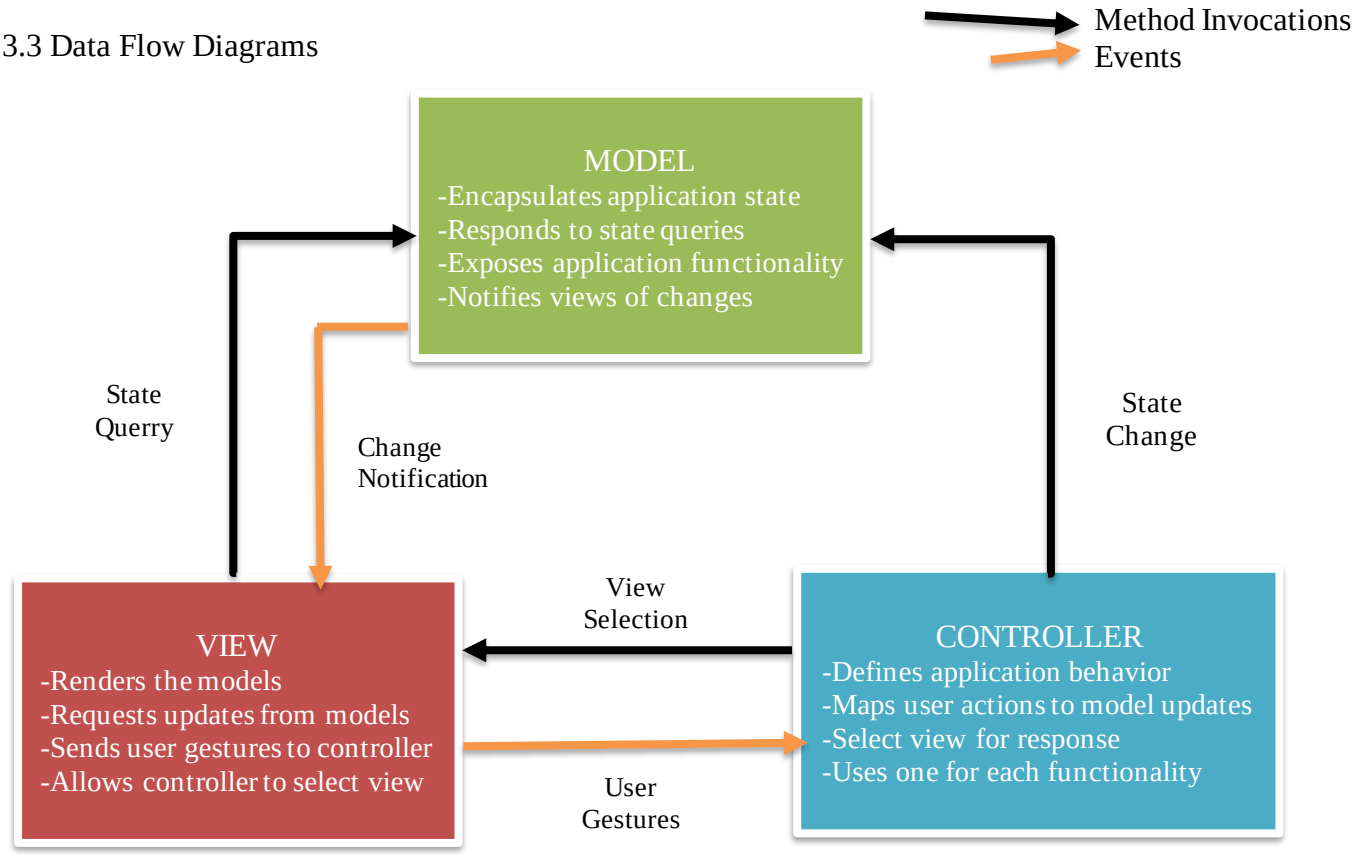


Fig 3.4 – Data Flow Diagram

3.4 ER Model

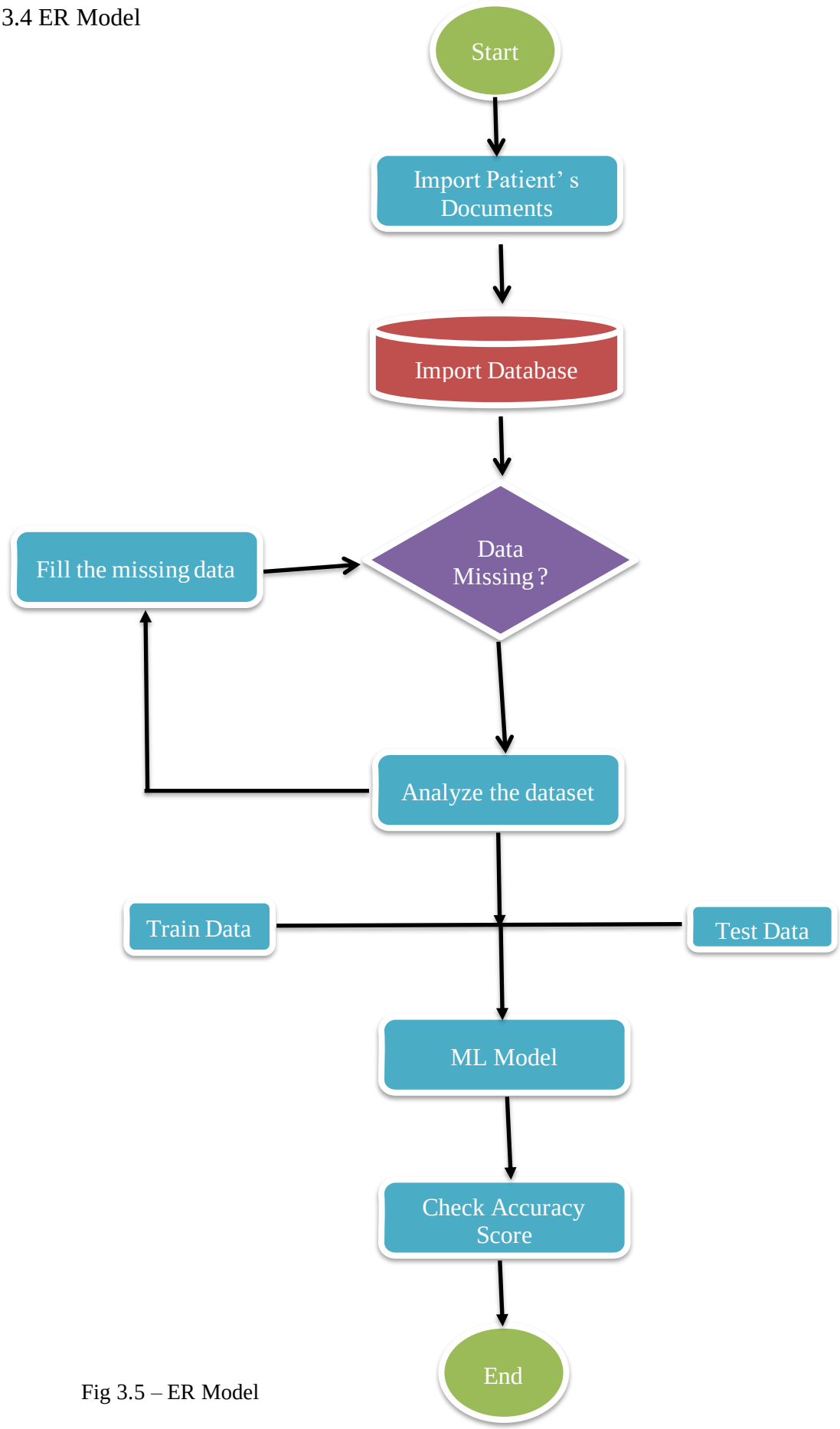


Fig 3.5 – ER Model

3.5 Process Specification

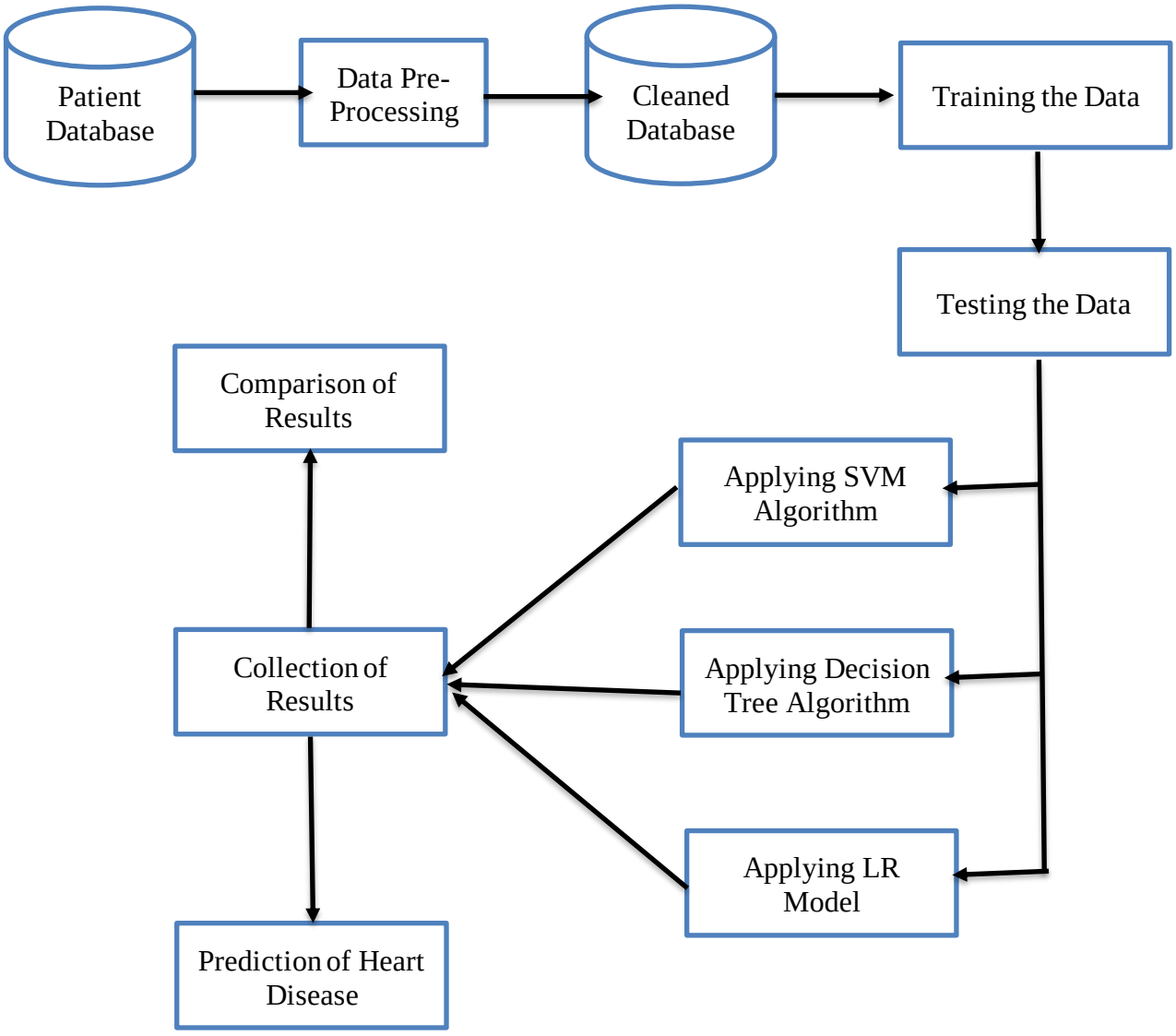


Fig 3.6 – Process Specification

3.6 CF Diagram

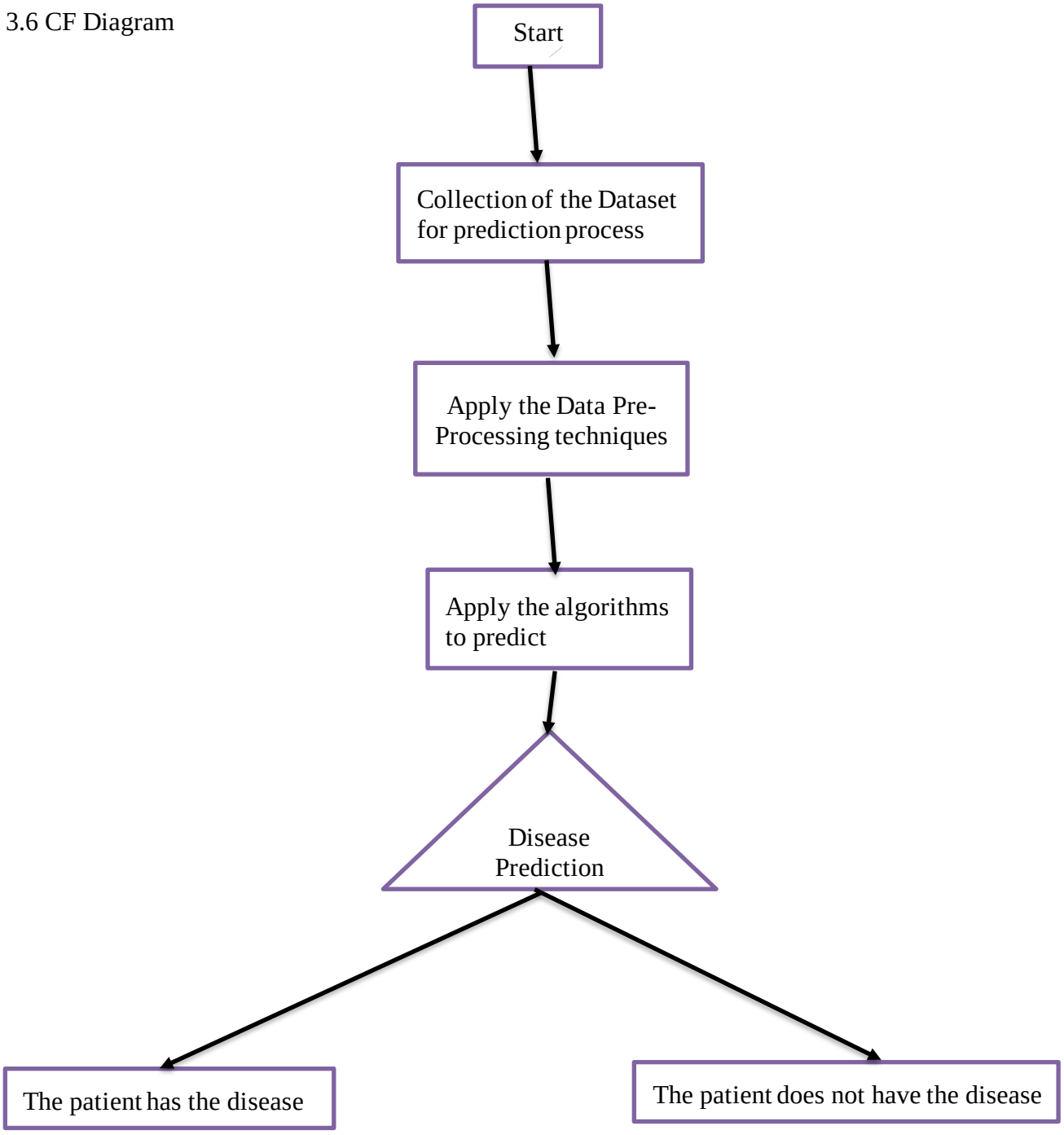


Fig 3.7 – CF Diagram

IV. Design

4.1 Architectural Design

4.1.1 System Architecture Diagram

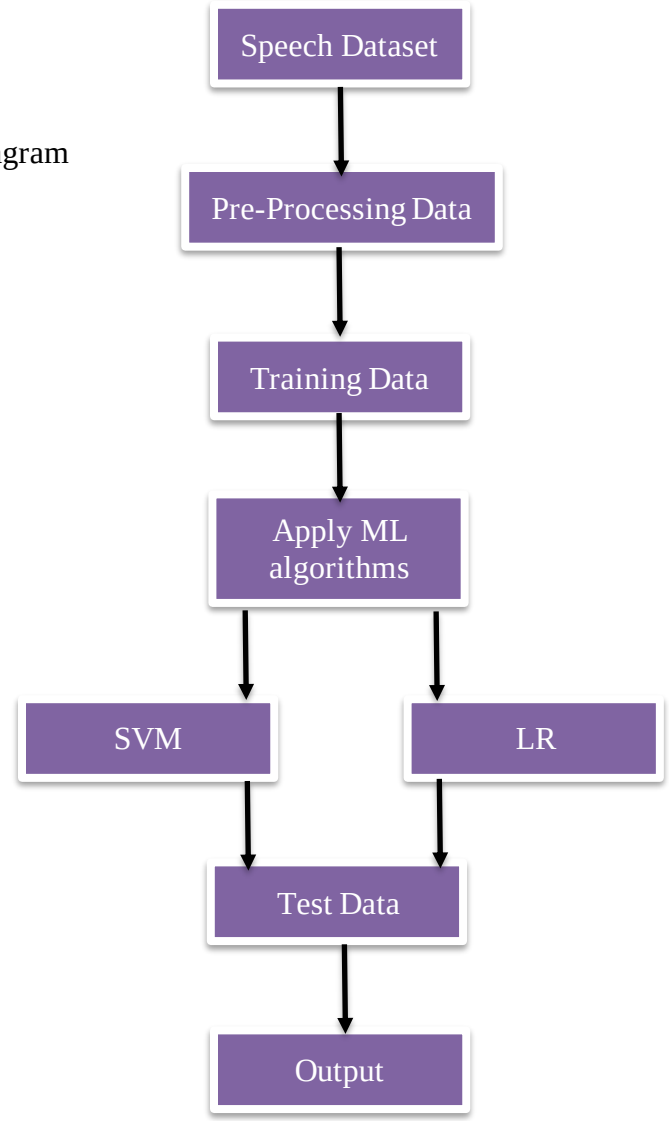


Fig 4.1 – System Architecture Diagram

4.1.2 Description of Architectural Design

Speech Dataset:

The main aim of this step is to spot and acquire all data-related problems. during this step, we'd like to spot the various data sources, as data are often collected from various speech dataset pre-processing data training data apply machine learning algorithms SVM and logistic regression test data output sources like files and databases. The number and quality of the collected data will determine the efficiency of the output. The more are going to be the info, the more accurate are going to be the prediction. We've collected our data from the Kaggle website.

Data Pre-Processing

The main aim of this step is to study and understand the nature of data that was acquired in the

previous step and also to know the quality of data. A real-world data generally contains noises, missing values, and may be in an unusable format that cannot be directly used for machine learning models. Data pre-processing is a required task for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model. Identifying duplicates in the dataset and removing them is also done in this step.

Training data:

Splitting the dataset into a Training set and testing set:

In machine learning data preprocessing, we have to break our dataset into both training set and test set. This is often one among the crucial steps of knowledge preprocessing as by doing this, we will enhance the performance of our machine learning model.

If we train our model alright and its training accuracy is additionally very high, but we offer a replacement dataset there to, then it'll decrease the performance. So we always attempt to make a machine learning model which performs well with the training set and also with the test dataset.

Apply Machine Learning Algorithms

Now, we've both the train and test data. The subsequent step is to spot the possible training methods and train our models. As this is often a classification problem, we've used two different classification methods SVM and Logistic Regression. Both algorithms has been run over the Training dataset and their performance in terms of accuracy is evaluated alongside the prediction wiped out the testing data set.

1. Logistic Regression

Logistic regression is additionally one among the foremost popular Machine Learning algorithms, which comes under the Supervised Learning technique. it's used for predicting the specific variable employing a given set of independent variables. It becomes a classification technique only a choice threshold is brought into the image. The setting of the edge value may be a vital aspect of Logistic regression and depends on the classification problem itself.

In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.

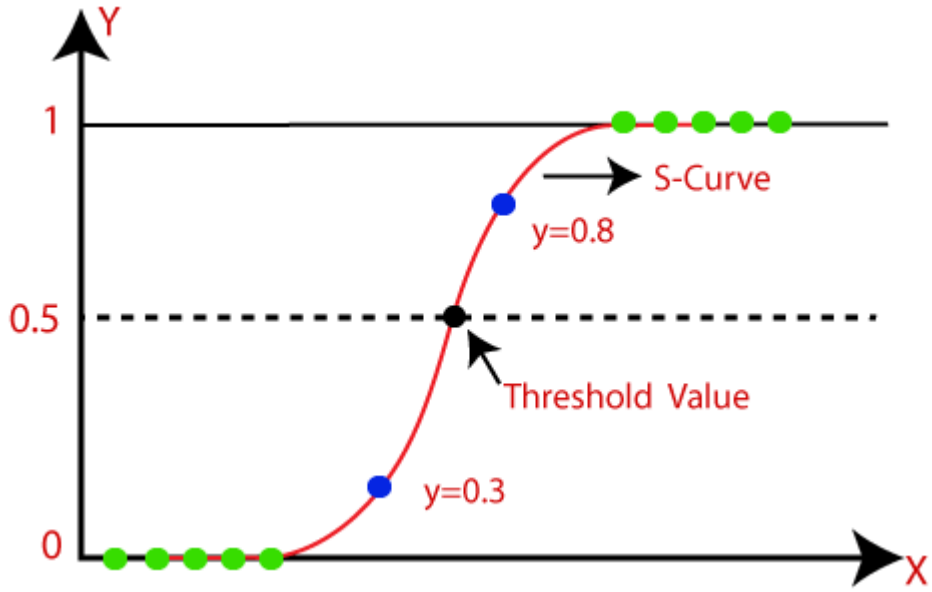


Fig 4.2 – S-Shaped Curve

Support Vector Machine

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:

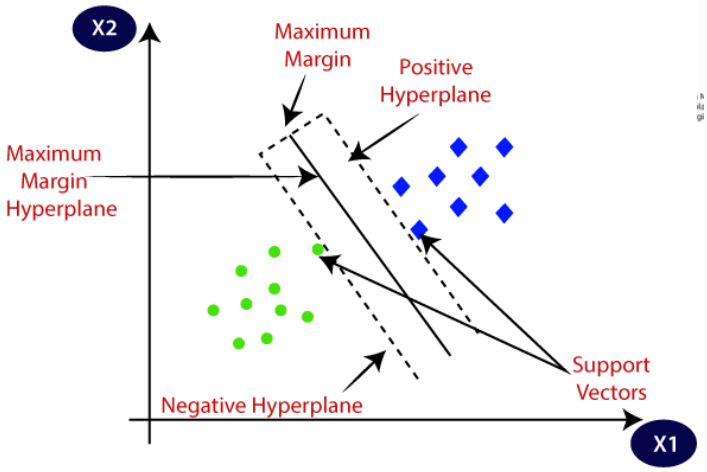


Fig 4.3 - SVM

Testing Data

Once the disease's Prediction model has been trained on the pre-processed dataset, then the model is tested using different data points. In this testing step, the model is checked for correctness and accuracy by providing a test dataset to it. All the training methods need to be verified for finding out the best model to be used. After fitting our model with training data, we used this model to predict values for the test dataset. These predicted values on testing data are used for model comparison and accurate calculation.

User Interface

Our Front-End implementation is completed using Python and Streamlit Framework in Scientific Python Development Environment (Spyder). The user interface is extremely essential for any project because everyone who tries to utilize the system for a purpose will attempt to access it using an interface. Indeed, our system also features a user interface built to facilitate users to utilize the services we provide. Where we have used Streamlit a Python library, utilized for creating web sites. Streamlit gives the sorts of choice when developing web applications, it provides you with tools, methods, and mechanics that allow you to build, create a various application but it will not enforce any dependencies or tell you the way how the project should be look like.

4.2 Database Design

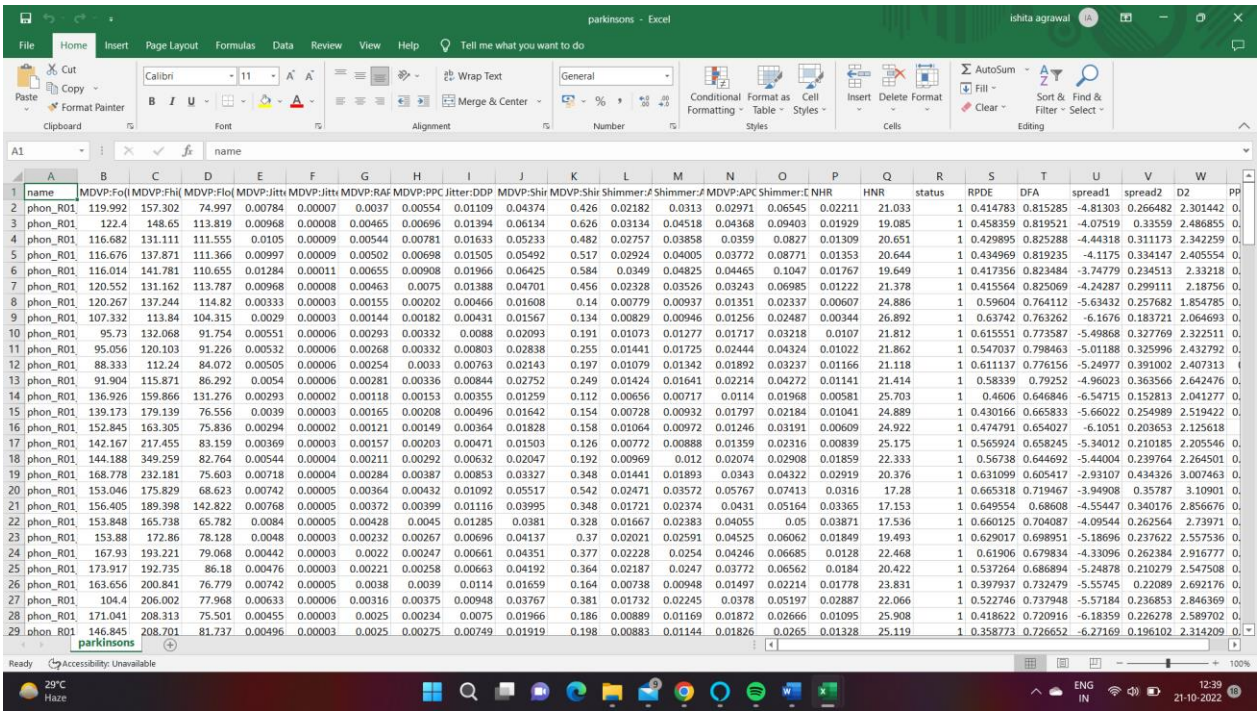
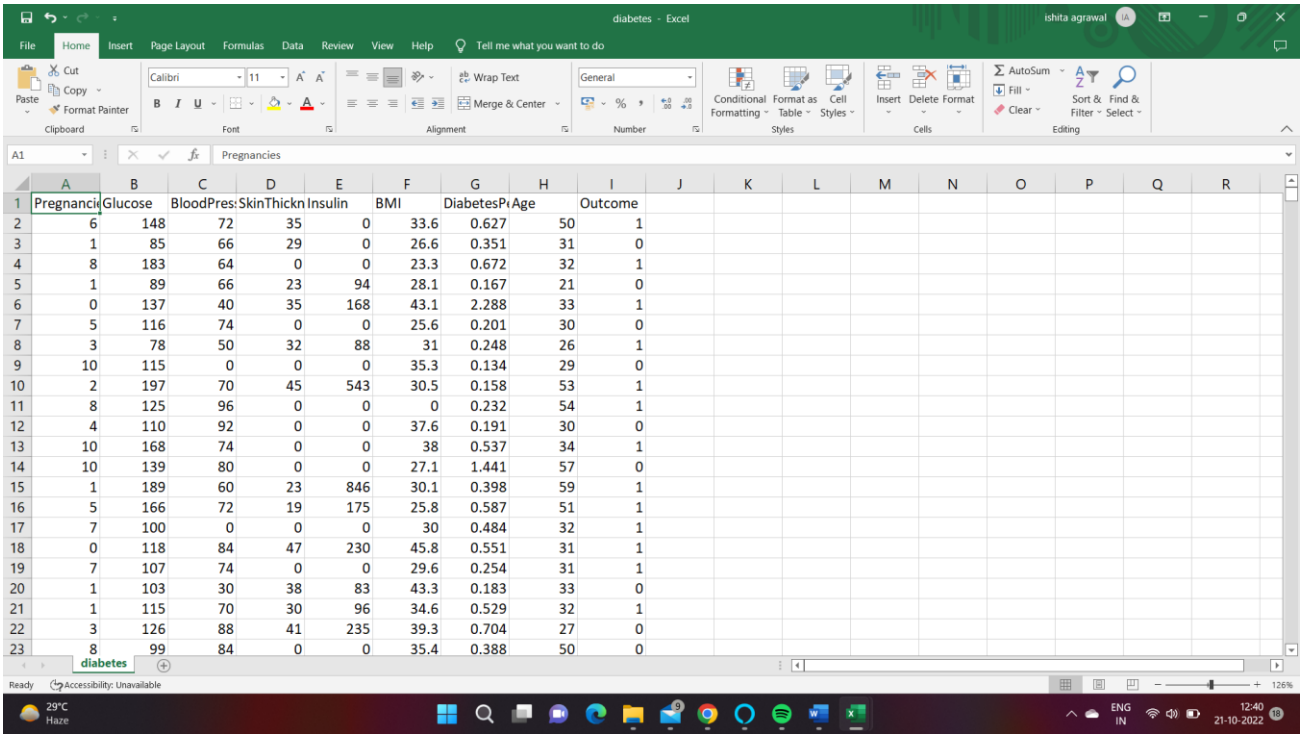
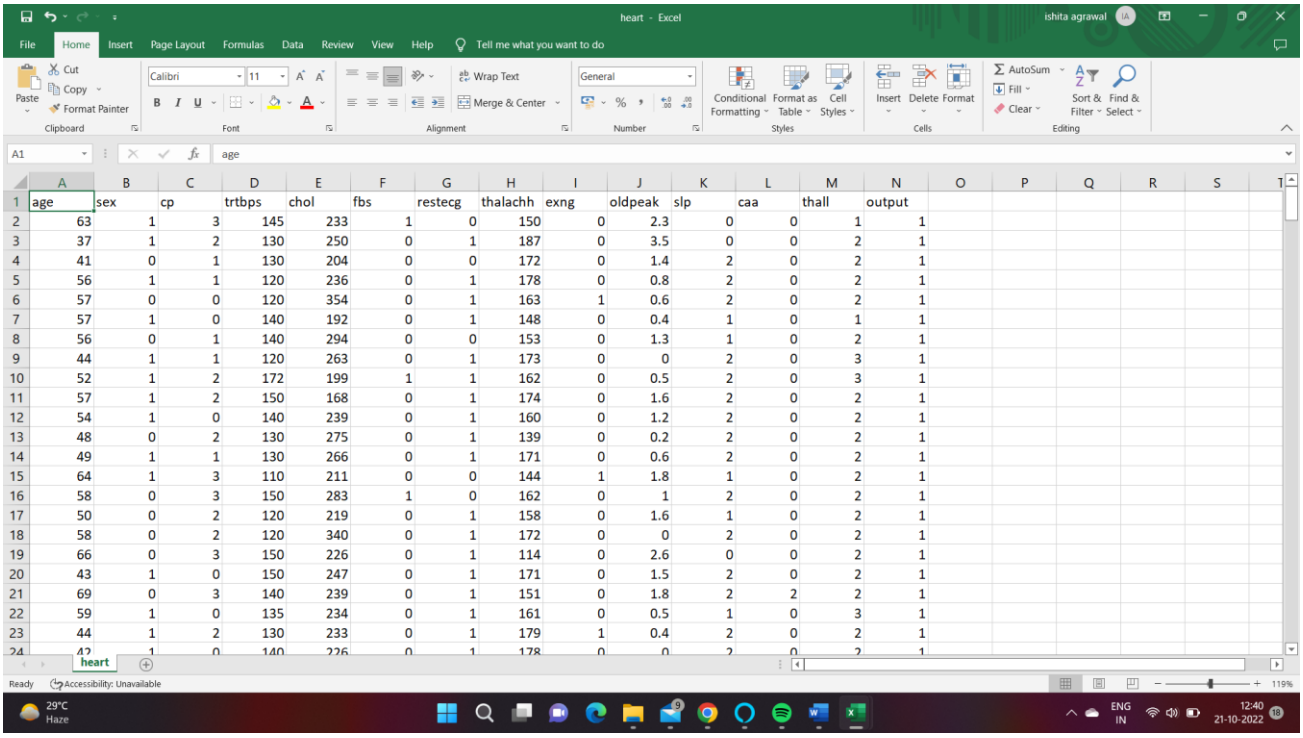


Fig 4.4(a) – Parkinsons Disease Database



	A	B	C	D	E	F	G	H	I
1	Pregnancie	Glucose	BloodPres	SkinThickn	Insulin	BMI	DiabetesP	Age	Outcome
2	6	148	72	35	0	33.6	0.627	50	1
3	1	85	66	29	0	26.6	0.351	31	0
4	8	183	64	0	0	23.3	0.672	32	1
5	1	89	66	23	94	28.1	0.167	21	0
6	0	137	40	35	168	43.1	2.288	33	1
7	5	116	74	0	0	25.6	0.201	30	0
8	3	78	50	32	88	31	0.248	26	1
9	10	115	0	0	0	35.3	0.134	29	0
10	2	197	70	45	543	30.5	0.158	53	1
11	8	125	96	0	0	0	0.232	54	1
12	4	110	92	0	0	37.6	0.191	30	0
13	10	168	74	0	0	38	0.537	34	1
14	10	139	80	0	0	27.1	1.441	57	0
15	1	189	60	23	846	30.1	0.398	59	1
16	5	166	72	19	175	25.8	0.587	51	1
17	7	100	0	0	0	30	0.484	32	1
18	0	118	84	47	230	45.8	0.551	31	1
19	7	107	74	0	0	29.6	0.254	31	1
20	1	103	30	38	83	43.3	0.183	33	0
21	1	115	70	30	96	34.6	0.529	32	1
22	3	126	88	41	235	39.3	0.704	27	0
23	8	99	84	0	0	35.4	0.388	50	0

Fig 4.4(b) – Diabetes Disease Database



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
2	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
3	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
4	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
5	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
6	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
7	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
8	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
9	44	1	1	120	263	0	1	173	0	0	2	0	3	1
10	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
11	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
12	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
13	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
14	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
15	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
16	58	0	3	150	283	1	0	162	0	1	2	0	2	1
17	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
18	58	0	2	120	340	0	1	172	0	0	2	0	2	1
19	66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
20	43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
21	69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
22	59	1	0	135	234	0	1	161	0	0.5	1	0	3	1
23	44	1	2	130	233	0	1	179	1	0.4	2	0	2	1
24	42	1	0	140	226	0	1	178	0	0	2	0	2	1

Fig 4.4(c) – Heart Disease Database

4.3 Component Design

4.3.1 Flow Chart

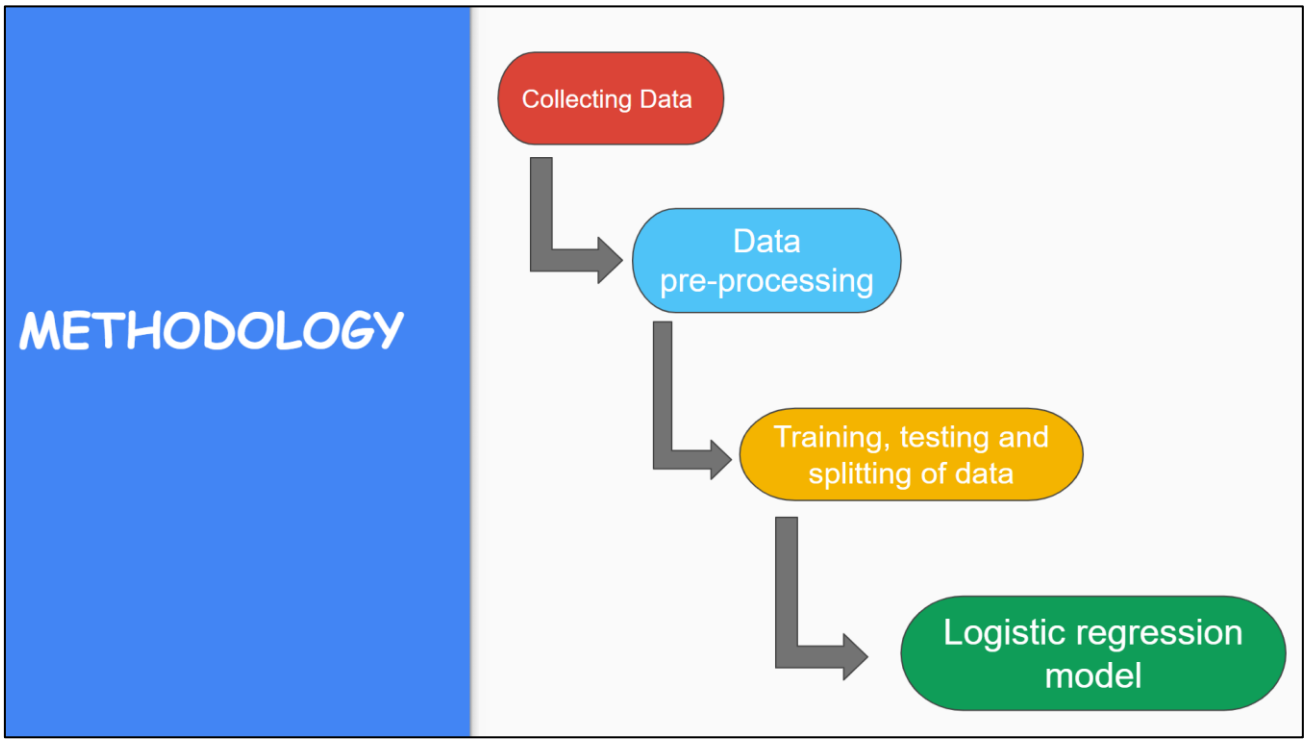


Fig 4.5 – Methodology

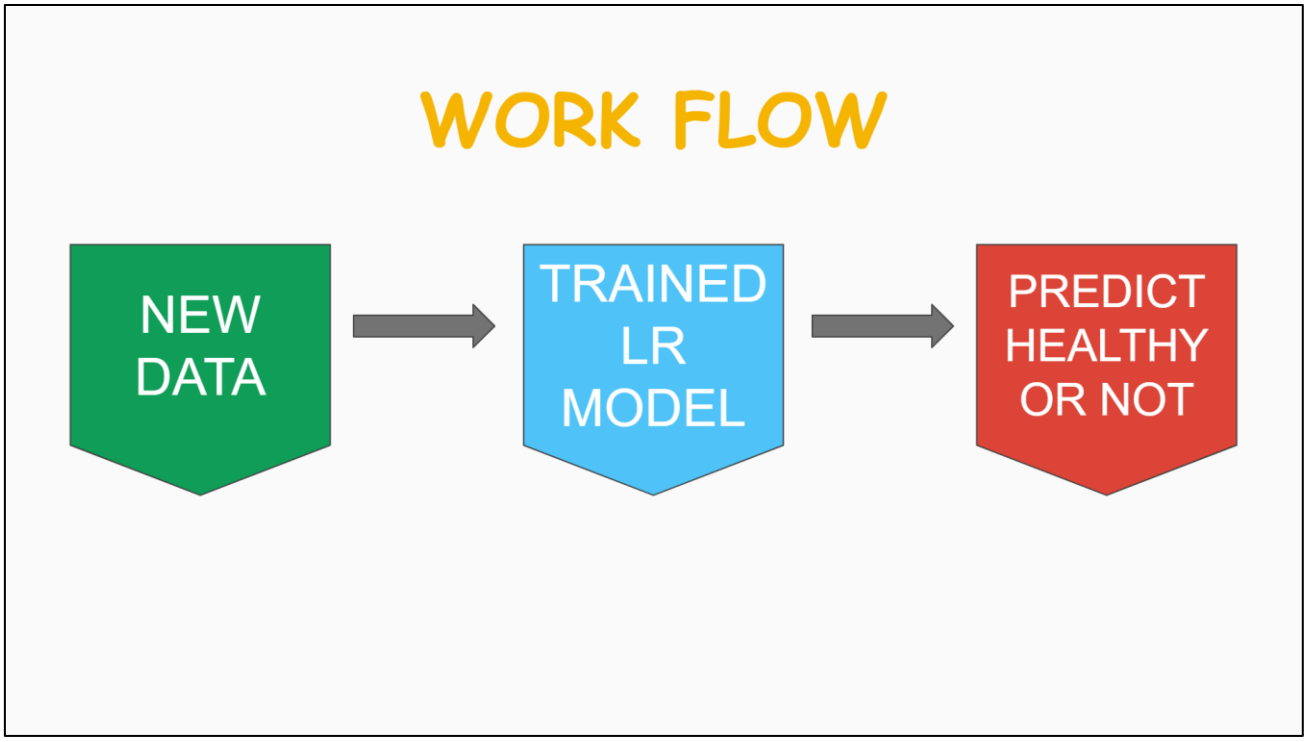
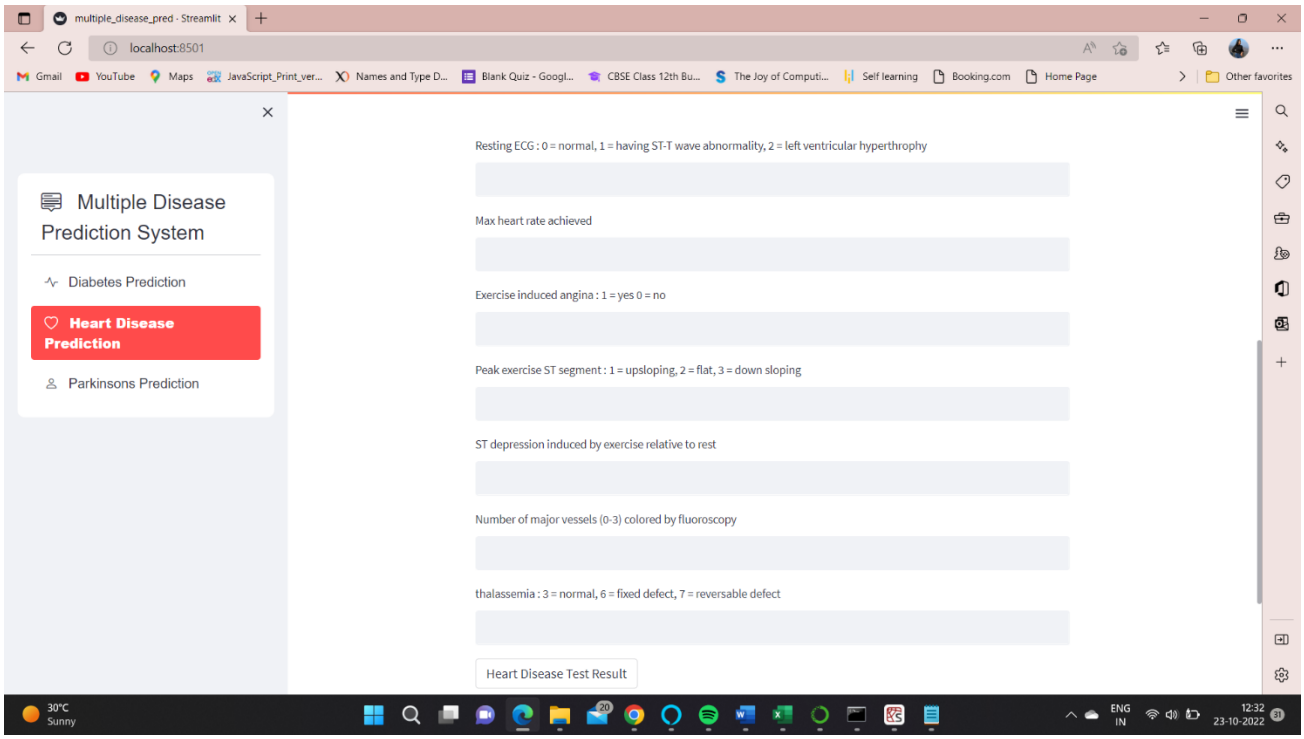
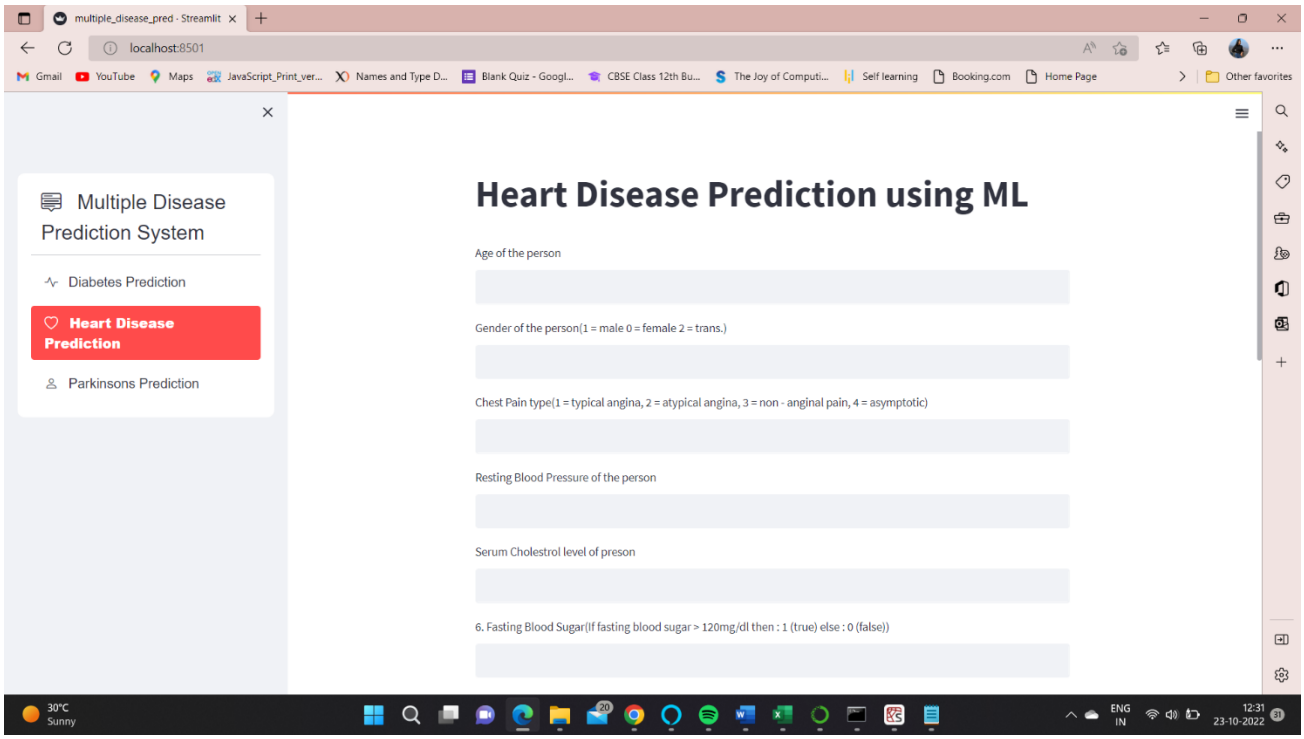


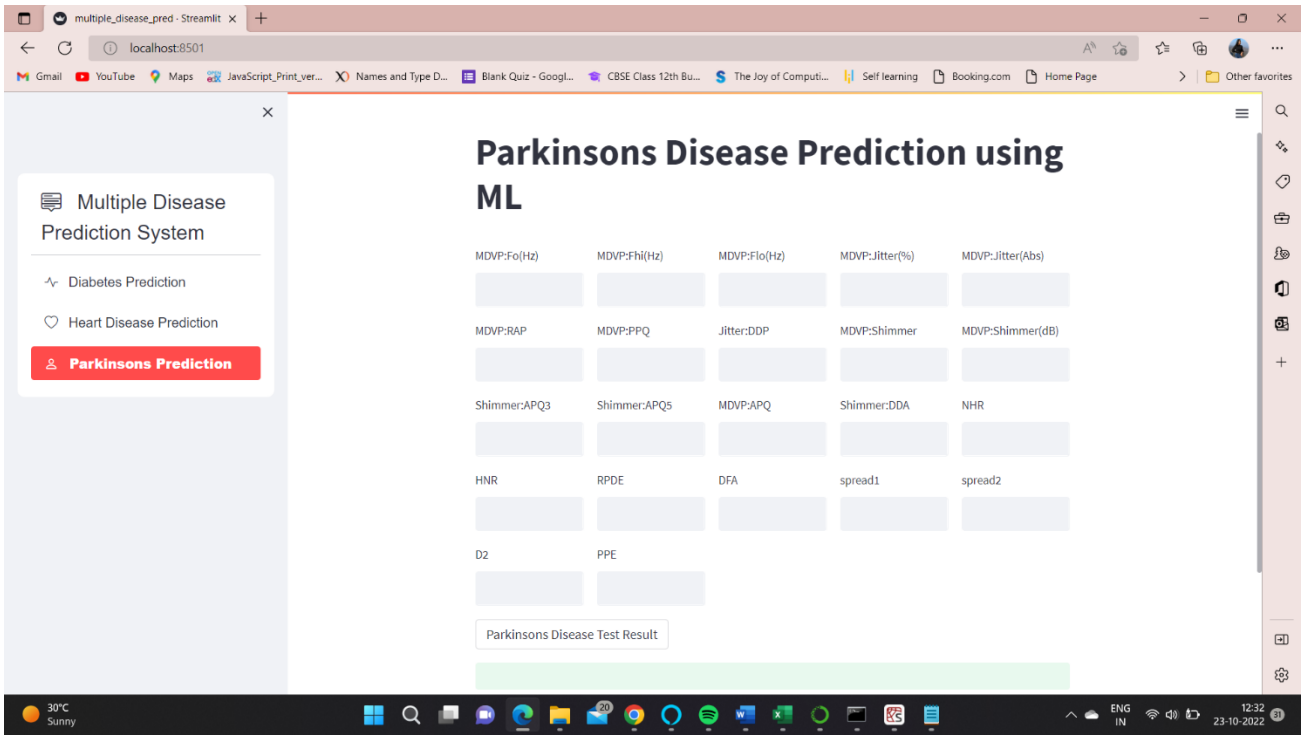
Fig 4.6 – Work Flow

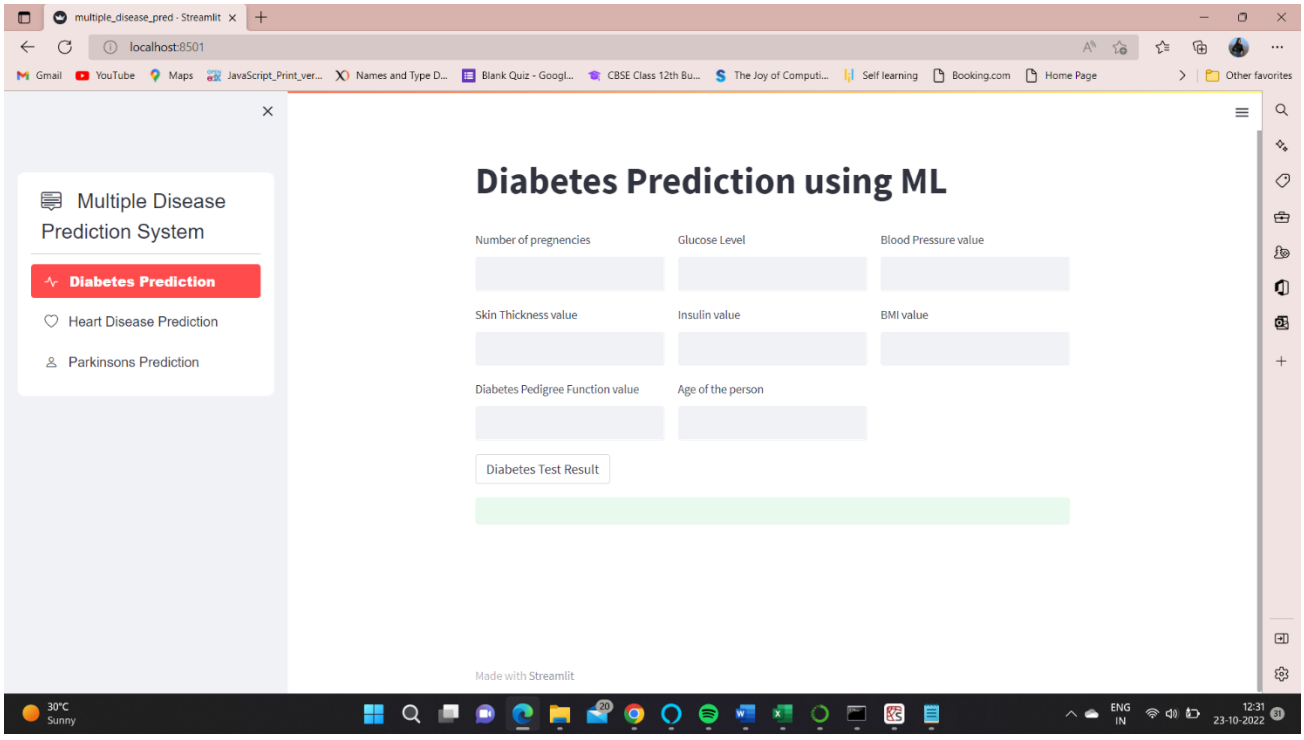
4.4 Interface Design

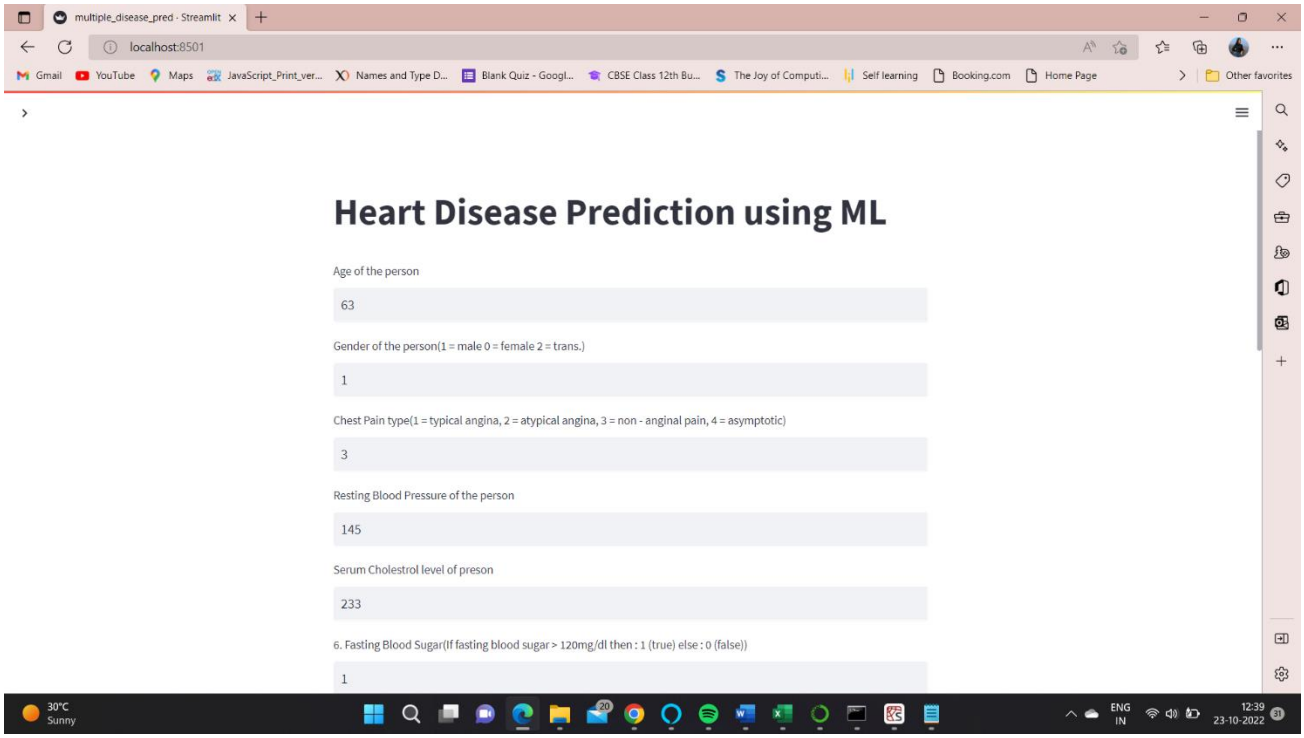
4.4.1 Screenshots

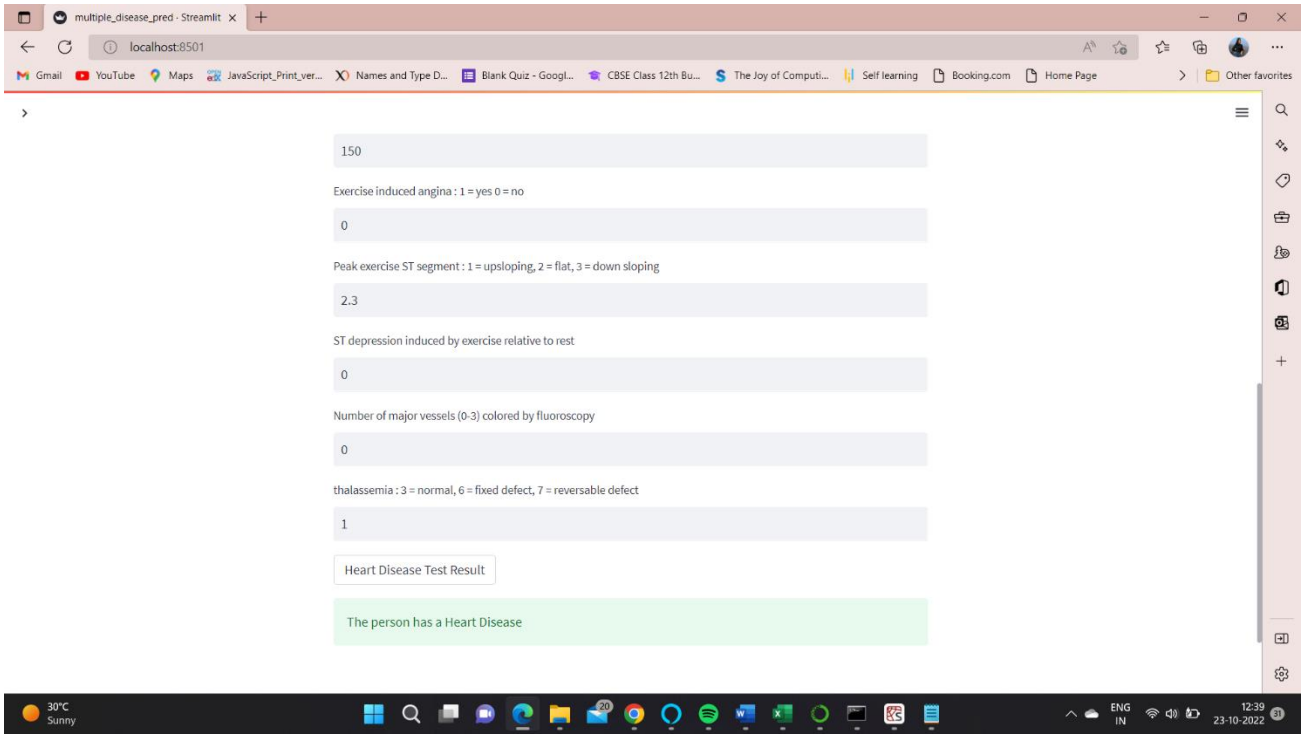


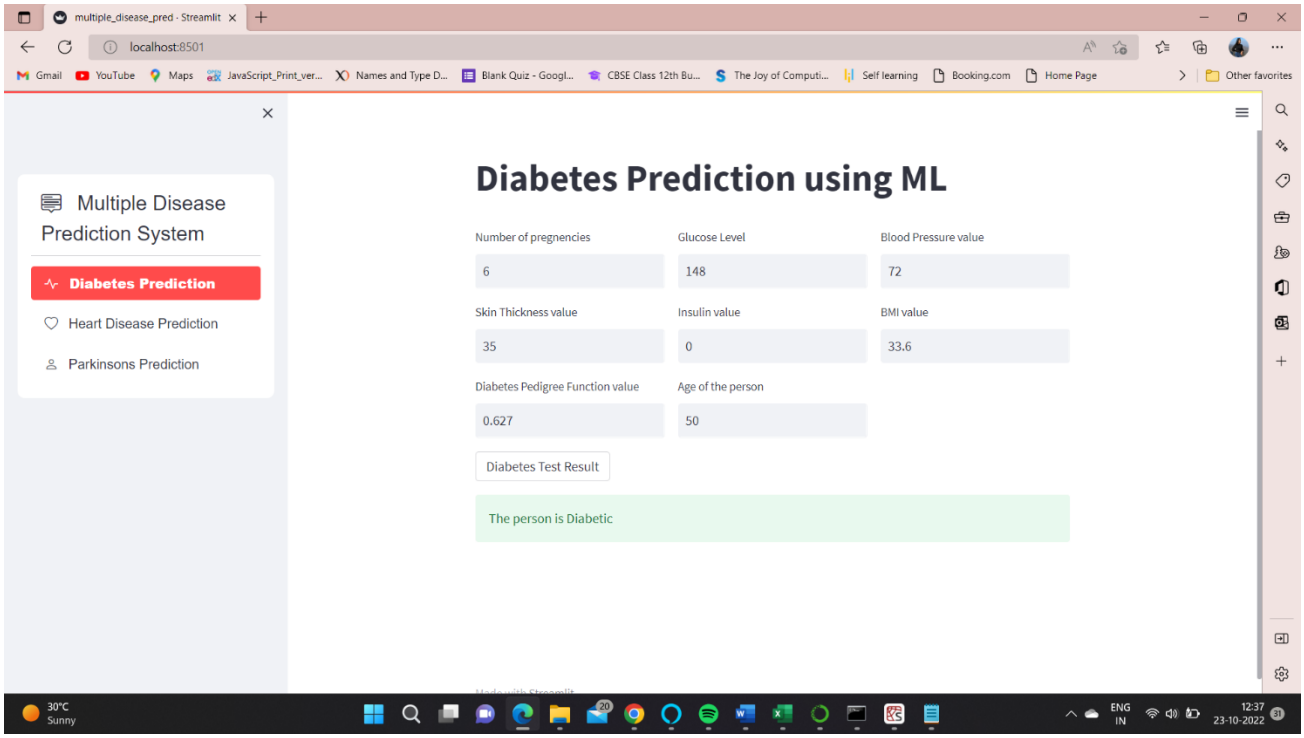


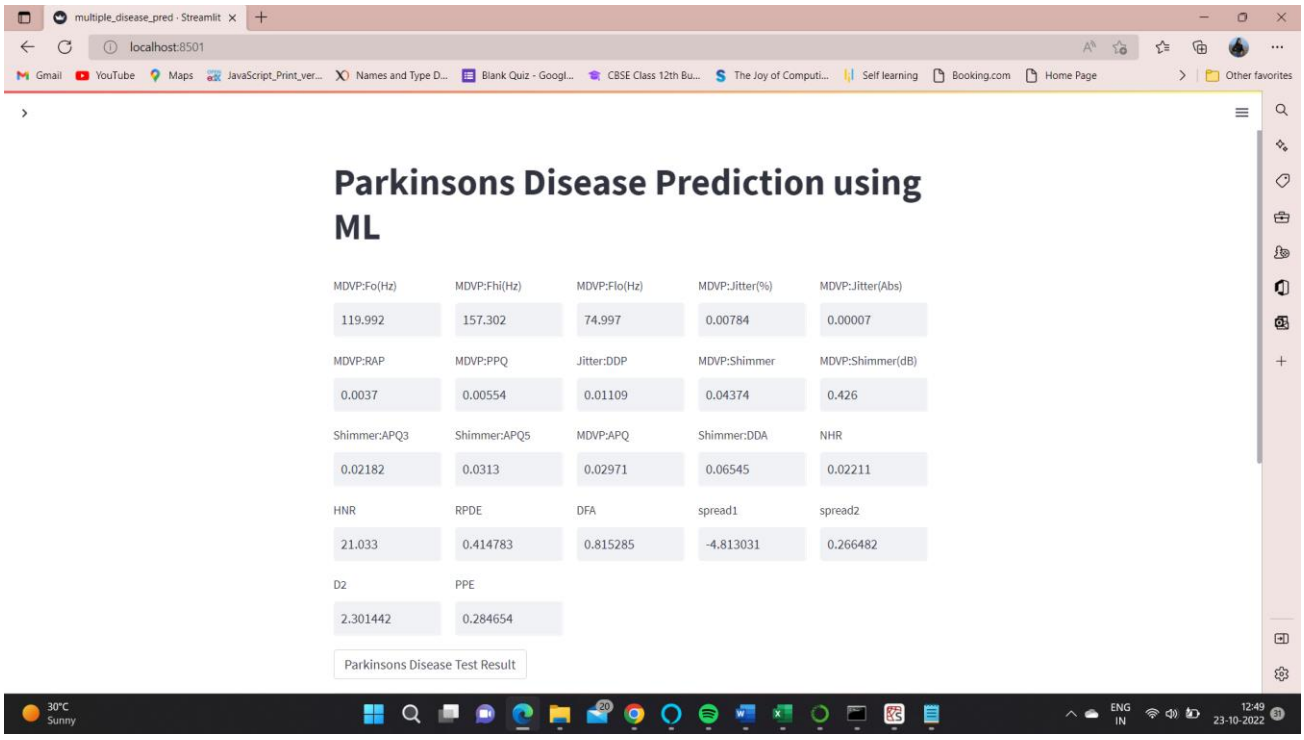

multiple_disease_pred - Streamlit
localhost:8501
Multiple Disease Prediction System
Diabetes Prediction
Heart Disease Prediction
Parkinsons Prediction
Parkinsons Disease Prediction using ML
MDVP:Fo(Hz)
MDVP:Fhi(Hz)
MDVP:Flo(Hz)
MDVP:Jitter(%)
MDVP:Jitter(Abs)
MDVP:RAP
MDVP:PPQ
Jitter:DDP
MDVP:Shimmer
MDVP:Shimmer(dB)
Shimmer:APQ3
Shimmer:APQ5
MDVP:APQ
Shimmer:DDA
NHR
HNR
RPDE
DFA
spread1
spread2
D2
PPE
Parkinsons Disease Test Result


multiple_disease_pred - Streamlit
localhost:8501
Multiple Disease Prediction System
Diabetes Prediction
Heart Disease Prediction
Parkinsons Prediction
Diabetes Prediction using ML
Number of pregnencies
Glucose Level
Blood Pressure value
Skin Thickness value
Insulin value
BMI value
Diabetes Pedigree Function value
Age of the person
Diabetes Test Result
Made with Streamlit


multiple_disease_pred · Streamlit
localhost:8501
Heart Disease Prediction using ML
Age of the person
63
Gender of the person(1 = male 0 = female 2 = trans.)
1
Chest Pain type(1 = typical angina, 2 = atypical angina, 3 = non - anginal pain, 4 = asymptotic)
3
Resting Blood Pressure of the person
145
Serum Cholestrol level of preson
233
6. Fasting Blood Sugar(If fasting blood sugar > 120mg/dl then : 1 (true) else : 0 (false))
1


multiple_disease_pred · Streamlit
localhost:8501
150
Exercise induced angina : 1 = yes 0 = no
0
Peak exercise ST segment : 1 = upsloping, 2 = flat, 3 = down sloping
2.3
ST depression induced by exercise relative to rest
0
Number of major vessels (0-3) colored by fluoroscopy
0
thalassemia : 3 = normal, 6 = fixed defect, 7 = reversable defect
1
Heart Disease Test Result
The person has a Heart Disease


multiple_disease_pred · Streamlit
localhost:8501
Multiple Disease Prediction System
Diabetes Prediction
Heart Disease Prediction
Parkinsons Prediction
Diabetes Prediction using ML
Number of pregnencies
6
Glucose Level
148
Blood Pressure value
72
Skin Thickness value
35
Insulin value
0
BMI value
33.6
Diabetes Pedigree Function value
0.627
Age of the person
50
Diabetes Test Result
The person is Diabetic


multiple_disease_pred · Streamlit
localhost:8501
Parkinsons Disease Prediction using ML
MDVP:Fo(Hz)
119.992
MDVP:Fhi(Hz)
157.302
MDVP:Flo(Hz)
74.997
MDVP:Jitter(%)
0.00784
MDVP:Jitter(Abs)
0.00007
MDVP:RAP
0.0037
MDVP:PPQ
0.00554
Jitter:DDP
0.01109
MDVP:Shimmer
0.04374
MDVP:Shimmer(dB)
0.426
Shimmer:APQ3
0.02182
Shimmer:APQ5
0.0313
MDVP:APQ
0.02971
Shimmer:DDA
0.06545
NHR
0.02211
HNR
21.033
RPDE
0.414783
DFA
0.815285
spread1
-4.813031
spread2
0.266482
D2
2.301442
PPE
0.284654
Parkinsons Disease Test Result

CONCLUSION

We developed a Prediction Engine that enables the user to check whether he/she has diabetes/ heart disease/ Parkinsons Disease. The user interacts with the Prediction Engine by filling out a form that holds the parameter set provided as input to the trained models. The Prediction engine provides an optimal performance compared to other state of art approaches. The Prediction Engine makes use of two algorithm to predict the presence of a disease namely: Logistic Regression and Support Vector Machine (SVM). The reasons to choose these algorithms is:

- It is effective, if the training data is large.
- A single dataset can be provided as an input to this algorithms with minimal or no modification.
- A common scalar can be used to normalize the input provided to this algorithm.

FUTURE SCOPE

- To enhance the functionality of the prediction engine providing the details of 5 nearest hospitals or medical facilities to the user input location.
- Provide a user account that allows the user to keep track of their medical test data and get suggestions or support to meet the right specialists or the tests to be taken.
- Provide admin controls to upload, and delete the dataset which will be used to train the model.
- Automate the process of training the model and extracting pickle files of the trained models which will be consumed by the API's to predict the disease.
- Mail the detailed report of the prediction engine results along with the information of 5 nearest medical facilities details having location and contact information.

REFERENCES

1 Priyanka Sonar, Prof. K. JayaMalini, "Diabetes Prediction Using Different Machine Learning Approaches", Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019) IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4.

2 Samrat Kumar Dey, Ashraf Hossain, Md. Mahbubur Rahman, "Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm", 2018 21st International Conference of Computer and Information Technology (ICCIT).

3 Deeraj Shetty, Kishor Rit, Sohail Shaikh, Nikita Patil, "Diabetes Disease Prediction Using Data Mining", 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS).

4 Zhilbert Tafa, Nerxhivane Pervetica, Bertran Karahoda, "An Intelligent System for Diabetes Prediction", 4thMediterranean Conference on Embedded Computing MECO – 2015 Budva, Montenegro.