SAVEETHA ENGINEERING COLLEGE
AUTONOMOUS
Approved by AICTE | Affiliated to Anna University
TNEA CODE 1216

A Project Report

on

STATISTICAL MACHINE LEARNING APPROACHES TO LIVER DISEASE PREDICTION

Submitted on partial fulfilment for the award of the degree

of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

Under the guidance of

Dr. RAMYADEVI M.E., Ph.D

Associate Professor/CSE

Submitted by

Team ID: PNT2022TMID03302

212219040078 – MOHAMED KASIM RAJA M
212219040091 – NAVEEN PRABHU S
212219040018 – CHANDRAMOHAN A
212219040085 – MUHAMMED RASHID VP

NAALAIYATHIRAN – EXPERIENTIAL PROJECT BASED LEARNING INITIATIVE

19CS532 – PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTREPRENEURSHIP

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SAVEETHA ENGINEERING COLLEGE
(AUTONOMOUS)

CHENNAI - 602105

NOVEMBER 2022

Project Report Format

1. **INTRODUCTION**
   - 1.1 Project Overview
   - 1.2 Purpose
2. **LITERATURE SURVEY**
   - 2.1 Existing problem
   - 2.2 References
   - 2.3 Problem Statement Definition
3. **IDEATION & PROPOSED SOLUTION**
   - 3.1 Empathy Map Canvas
   - 3.2 Ideation & Brainstorming
   - 3.3 Proposed Solution
   - 3.4 Problem Solution fit
4. **REQUIREMENT ANALYSIS**
   - 4.1 Functional requirement
   - 4.2 Non-Functional requirements
5. **PROJECT DESIGN**
   - 5.1 Data Flow Diagrams
   - 5.2 Solution & Technical Architecture
   - 5.3 User Stories
6. **PROJECT PLANNING & SCHEDULING**
   - 6.1 Sprint Planning & Estimation
   - 6.2 Sprint Delivery Schedule
   - 6.3 Reports from JIRA
7. **CODING & SOLUTIONING (Explain the features added in the project along with code)**
   - 7.1 Feature 1
   - 7.2 Feature 2
   - 7.3 Database Schema (if Applicable)
8. **TESTING**
   - 8.1 Test Cases
   - 8.2 User Acceptance Testing

CHAPTER 1

INTRODUCTION

1.1 PROJECT OVERVIEW

Liver diseases avert the normal function of the liver. Mainly due to the large amount of alcohol consumption liver disease arises. Early prediction of liver disease using classification algorithms is an efficacious task that can helpthe doctors to diagnose the disease within a short duration of time. Discoveringthe existence of liver disease at an early stage is a complex task for the doctors.The main objective of this project is to analyze the parameters of various classification algorithms and compare their predictive accuracies so as to find out the best classifier for determining the liver disease. This Project examines data from liver patients concentrating on relationships between a key list of liver enzymes, proteins, age and gender using them to try and predict the likeliness of liver disease. Here we are building a model by applying various machine learning algorithms find the best accurate model. And integrate to flask based web application. User can predict the disease by entering parameters in the web application.

1.2 PURPOSE

Current screening strategies for liver disease focus on detection of subclinical advanced liver fibrosis but cannot identify those at high future risk of severe liver disease. Our aim was to develop and validate a risk prediction model for incident liver disease in the general population based on widely available factors. The diagnosis is often delayed until severe complications occur and prognosis becomes poor. In order to identify individuals in the general population who have a high risk of developing severe liver disease in the future, we developed and validated a Liver Disease risk prediction with or without measurement of the liver enzyme gamma-glutamyltransferase. The Liver Disease score can be used as part of health counseling, and for planning furtherliver investigations and follow-up.

CHAPTER 2

LITERATURE SURVEY

2.1 EXISTING PROBLEMS

- A number of network models based on neural have been developed in recent research assisting physicians with liver diagnosis in the medical field, such as the diagnostic support system, the expert system, the perceptive diagnostic model, and the hybrid recommendation framework.
- Christopher N suggested a system for the diagnosis of medical diseases, taking into account six benchmarks: liver, hepatitis heart, diabetes, breast, and lymph disorders. The researchers developed WSO and C4.5- based systems.
- Ramana has also conducted an acritical study on the diagnosis of liver diseases.

Drawbacks and Limitations

- From Christopher's Observation, the researchers developed WSO and C4.5-based systems, with a precision of 64.60 percent with 19 liver disorder dataset rules and 62.89 percent with 43 WSO and C4.5 rules, respectively. In the evaluation of identified categorization techniques.
- From Ramana's view, On the Naïve Bayes classifier, the authors gained51.59 percent accuracy, 55.94 percent on the C4.5 algorithm, 66.66 percentwith respect to BPNN, 62.6 percent with respect to Knowledge discovery, and sixty-two percent accuracy with respect to vector machine support algorithm.

Proposed Solutions

- SVM is a technique of supervised learning that pertained to classification as well as regression. It has effective performance in generalization. Moreover, when the algorithm requires input space with high dimensions, there is no requirement to add a previous understanding. This helps make ita very effective classifier for quality. The primary purpose of the SVM classifier is to classify between groups of various classifications by choosing the best classifier function in the training

data. A generalized linear method of classification is SVM. At the same time, the geometric margin is maximized and the classification error is minimized.

- This Project examines data from liver patients concentrating on relationships between a key list of liver enzymes, proteins, age, and gender using them to try and predict the likeliness of liver disease. Here we are building a model by applying various machine learning algorithms to find the best accurate model. And integrate into flask-based web applications. Users can predict the disease by entering parameters in the web application.

Conclusion

SVM, Logistic Regression, comprises two main machine learning techniques used. Using all the models, the prediction analysis has been implemented and their performance has been assessed. The probability of liver disease prediction was attained with an accuracy of 96%. In the future, the present scenario can be compared with other techniques such as naïve Bayes classification, Random forest, etc.

2.2 REFERENCES

- Yongjun Piao, Minghao Piao, Keun Ho Ryu, **MULTICLASS CANCER CLASSIFICATION USING A FEATURE SUBSET BASED ENSEMBLE FROM microRNA EXPRESSION PROFILES, COMPUTERS IN BIOLOGY AND MEDICINE 80** -2017.
- Sanjay Kumar, Sarthar Katyal, **EFFECTIVE ANALYSIS AND DIAGNOSIS OFLIVER DISORDER BY DATA MINING** – 2018.
- Ajay S Singh, MD Irfan, Abishek**, PREDICTION OF LIVER DISEASE USINGCLASSIFICATION ALGORITHMS** – 2018
- Md Fazle, SM Mahedy Hasan, Arifa , Md Asif, Md kamrul, **PREDICTION OFLIVER DISORDERS USING ML ALGORITHMS**-2020
- C Geetha, AR Arunachalam, **EVALUATION BASED APPROACHES FOR LIVER DISEASE PREDICTION USING MACHINE LEARNING ALGORITHMS** –Jan 2021.

2.3PROBLEM STATEMENT DEFINITION

To develop a Machine Learning Model for Liver Disease Prediction at an early stage by applying various machine learning algorithms and finding an accurate model. After that, it is integrated into the flask-based web application where users can predict the disease by entering parameters in the web application.

CHAPTER 3

IDEATION AND PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS

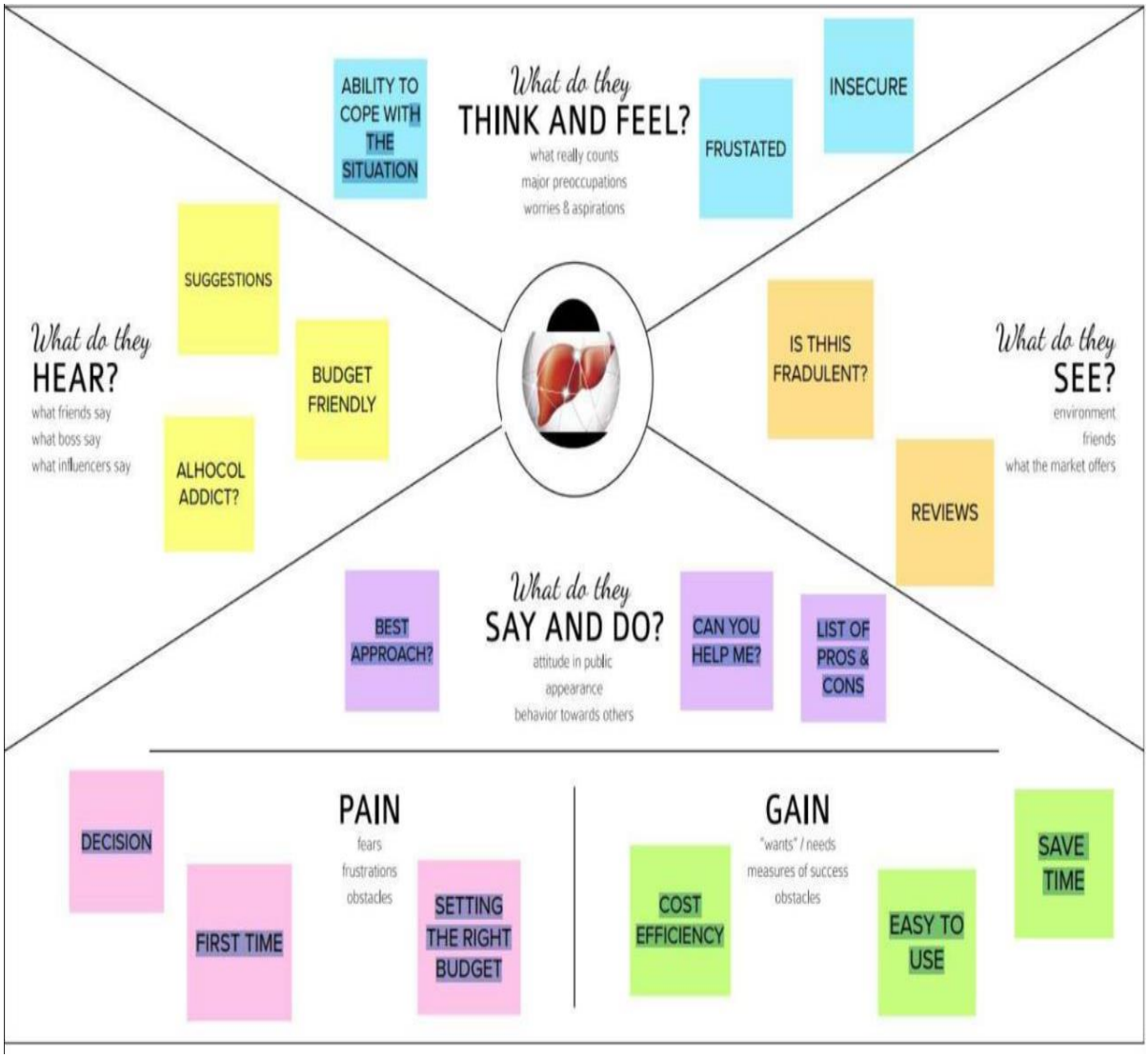

Figure 3.1 - Empathy Map

An empathy map is a widely-used visualization tool within the field. In relation to empathetic design, the primary purpose of an empathy map is to bridge the understanding of the end use.

3.2 IDEATION AND BRAINSTORMING

Step-1:

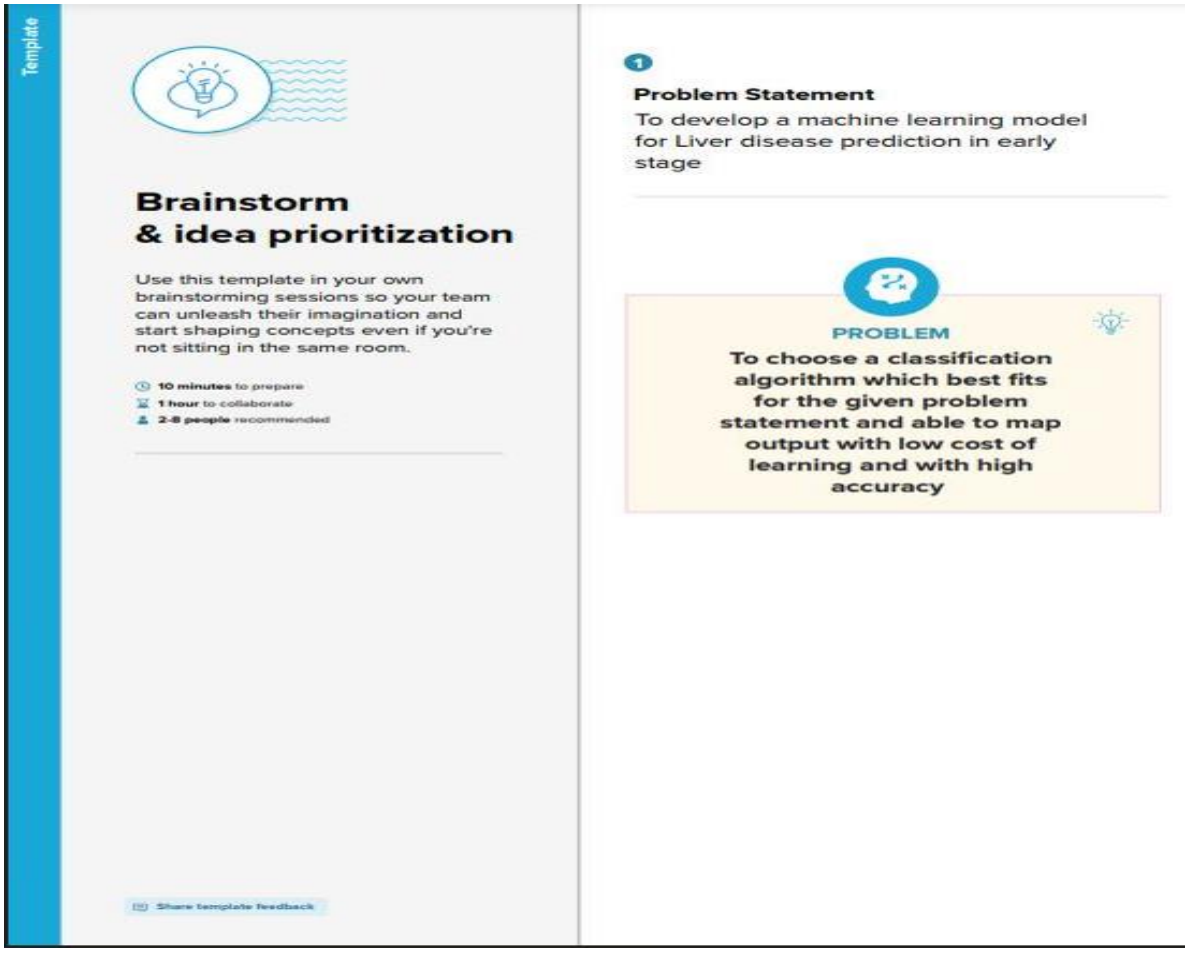

Figure 3.2 - Ideation and Brainstorming

A principal difference between ideation and brainstorming is that ideation iscommonly more thought of as being an individual pursuit, while brainstorming isalmost always a group activity.

Step-2:

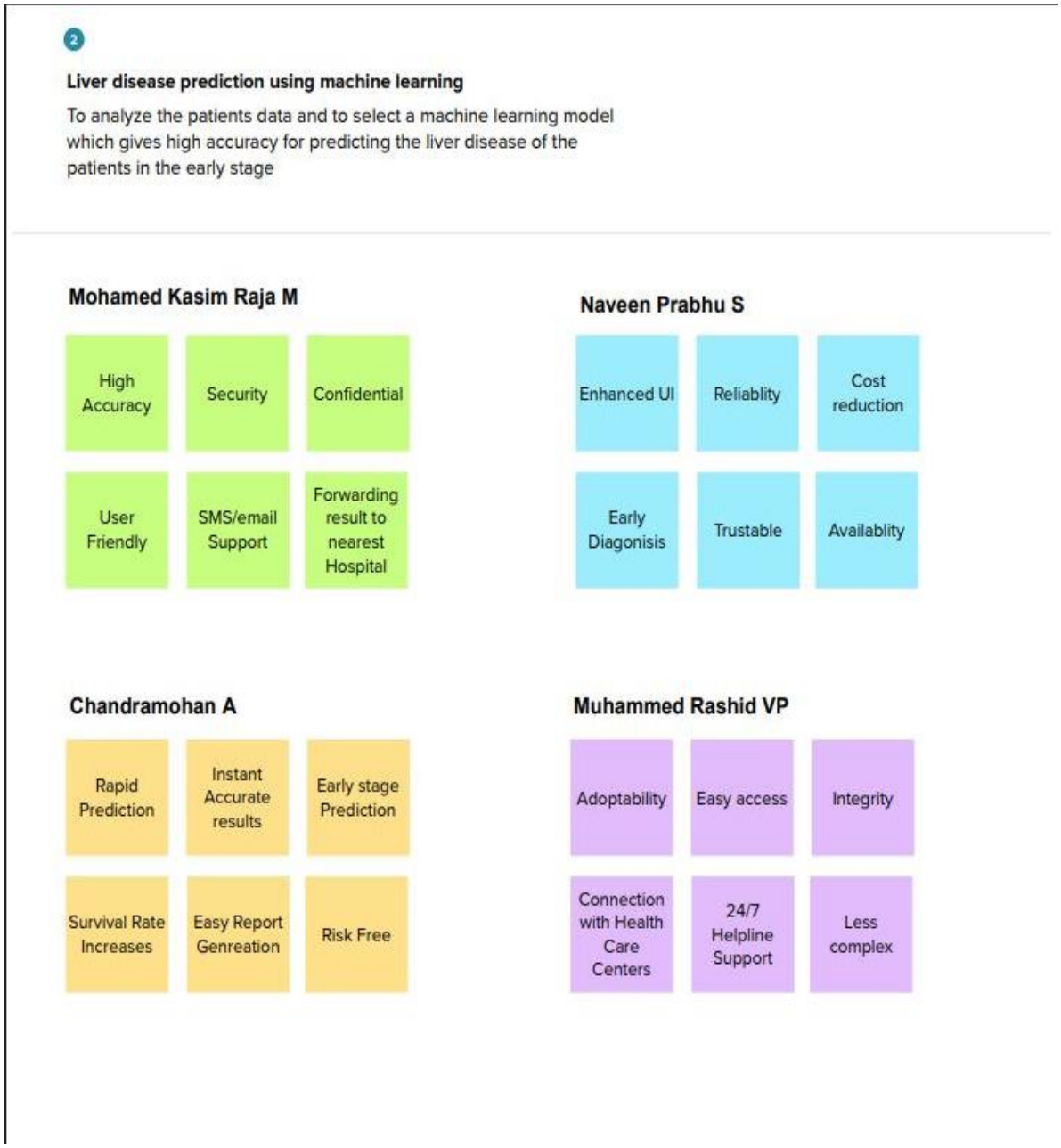

Figure 3.3 - Brainstorm, Idea Listing and Grouping

The idea listing and grouping is used to organize and analyse large numbers of ideas by categorising them. By organising and reorganising ideas, students gain a better appreciation of, and dialogue about, their ideas. As students create idea clusters, new contexts and connections among themes emerge.

Step-3:

Figure 3.4 - Idea Prioritization

Idea prioritization is just a part of the idea management process. Having a structured idea management process and a systematic way of gathering, evaluating and prioritizing new ideas takes time. To make it work, the entire idea management process should be integrated to the everyday ways of working.

3.3 PROPOSED SOLUTION

S.No	Parameter	Description
1.	Problem Statement (Problem to be solved)	Discovering the existence of liver diseases at early stage is a complex task for doctors. The challenge is to predict the liver disease patient fast and accurate and to diagnose the patients in early stage .
2.	Idea / Solution description	Machine learning model which uses statistical data to predict the liver disease of the patients.
3.	Novelty / Uniqueness	Accurately classifies the intensity of the liver disease from the patients concentrating on relationship between a key list of enzymes, proteins, age and gender using them to predict the likeliness of the liver Disease.
4.	Social Impact / Customer Satisfaction	✓ Capable of predicting the liver disease in early stage ✓ Works accurately and precisely to predict the liver disease ✓ Doctors can be able to diagnose the live patients in early stage to save many lives.
5.	Business Model (RevenueModel)	This system can be integrated with any Health sector domain, It solves the complex process of predicting the liver disease of patients and makes ease to the doctors to diagnose the liver disease. ✓ The user can be able to get consulting with doctors.

6.	Scalability of the Solution	✓ Can be extended to predict many classification of diseases in early stage. This can be integrated to with any hospitals and health sectors to get patient records securely through APIs.

Table 3.1 - Proposed Solution

3.4 PROBLEM SOLUTION FIT

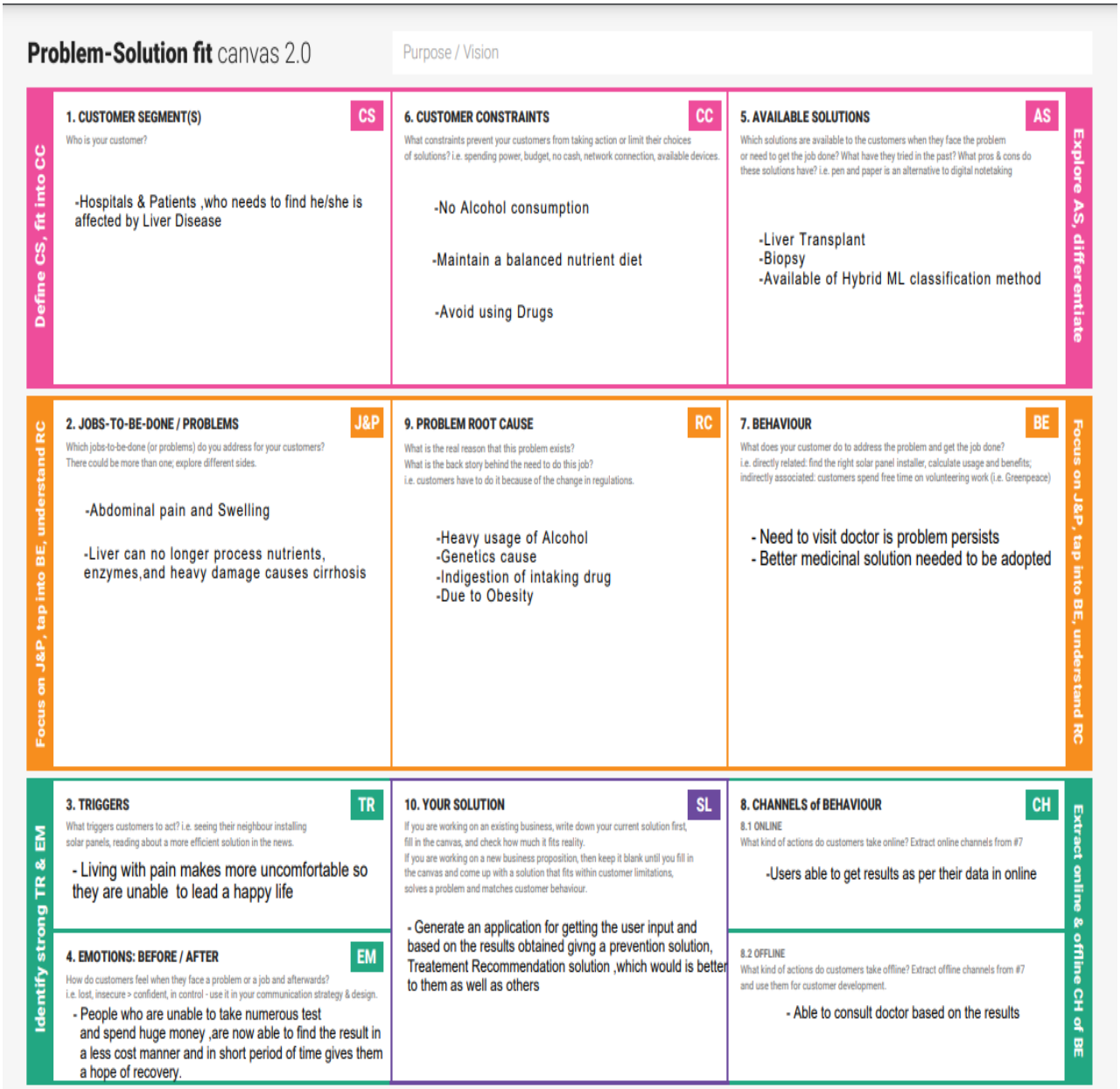

Figure 3.5 - Solution fit of design with user requirements

This occurs when the user have evidence that customers care about certain jobs, pains, and gains. At this stage the user proved the existence of a problem and have designed a value proposition that addresses customers' jobs, pains andgains.

CHAPTER 4

REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	**User Registration**	Registration through Form Registration through Gmail
FR-2	**User Confirmation**	Confirmation via Email Confirmation via OTP
FR-3	**Website Entry**	Collecting user's data and storing it in the Database
FR-4	**Permissions**	Location, Contacts, Storage

Table 4.1 - Functional Requirements

4.2 NON-FUNCTIONAL REQUIREMENT

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Defines how difficult it will be for a user to learn and operate the system. Usability can be accessed from different points
NFR-2	**Security**	Security requirements ensure that the software is protected from unauthorized access to the system and it's stored in data.
NFR-3	**Reliability**	Reliability defines how likely it is for the software to work without failure for a given period. Reliability decreases because of bugs in the code, hardware, failures and problems with other system component.
NFR-4	**Performance**	It is quality attribute that describer's responsiveness of system to the various user interactions with it.

NFR-5	**Availability**	It is gauged by period that system's functionality &services are available for use with all operations.
NFR-6	**Scalability**	Scalability describes how the system must grow without negative influence on its performance. This means serving more users, processing more data, doing more transactions. Website traffic limit must be scalable enough to support 2,00,000 users at a time.

Table 4.2 – Non-Functional Requirements

CHAPTER 5

PROJECT DESIGN

5.1 DATAFLOW DIAGRAM

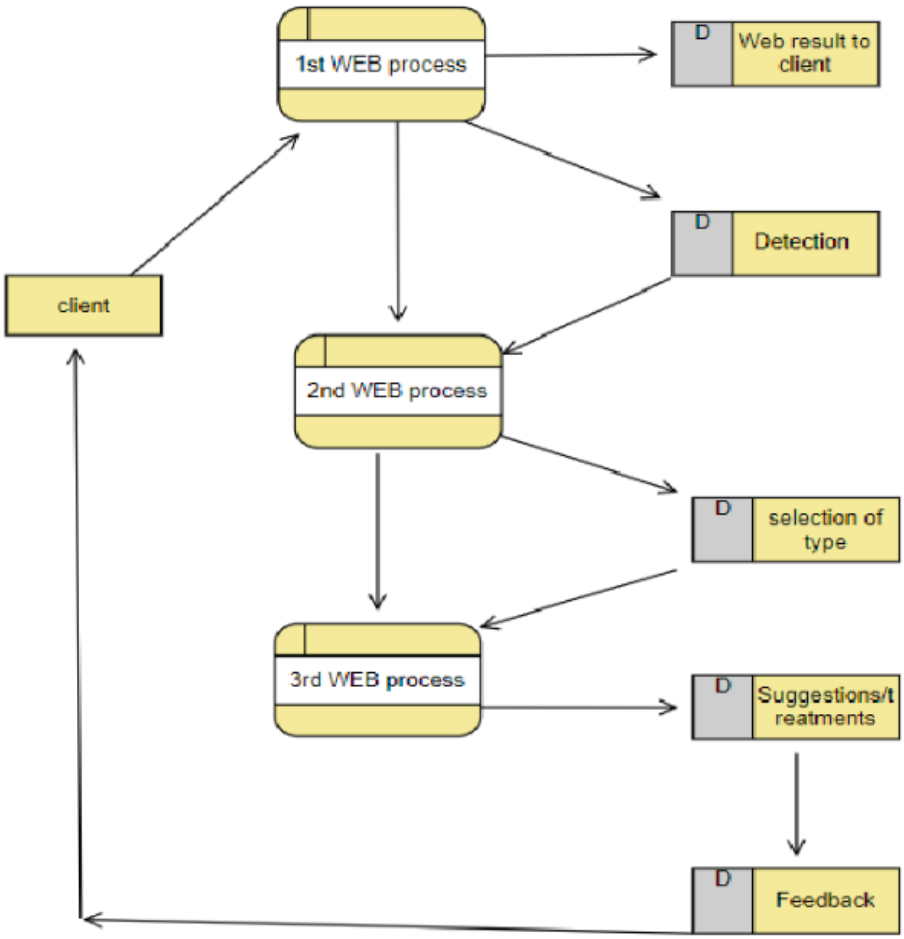

Figure 5.1 - Data flow Diagram

A data flow diagram is a graphical or visual representation using a standardized set of symbols and notations to describe a business's operations through data movement. They are often elements of a formal methodology such as Structured Systems Analysis and Design Method.

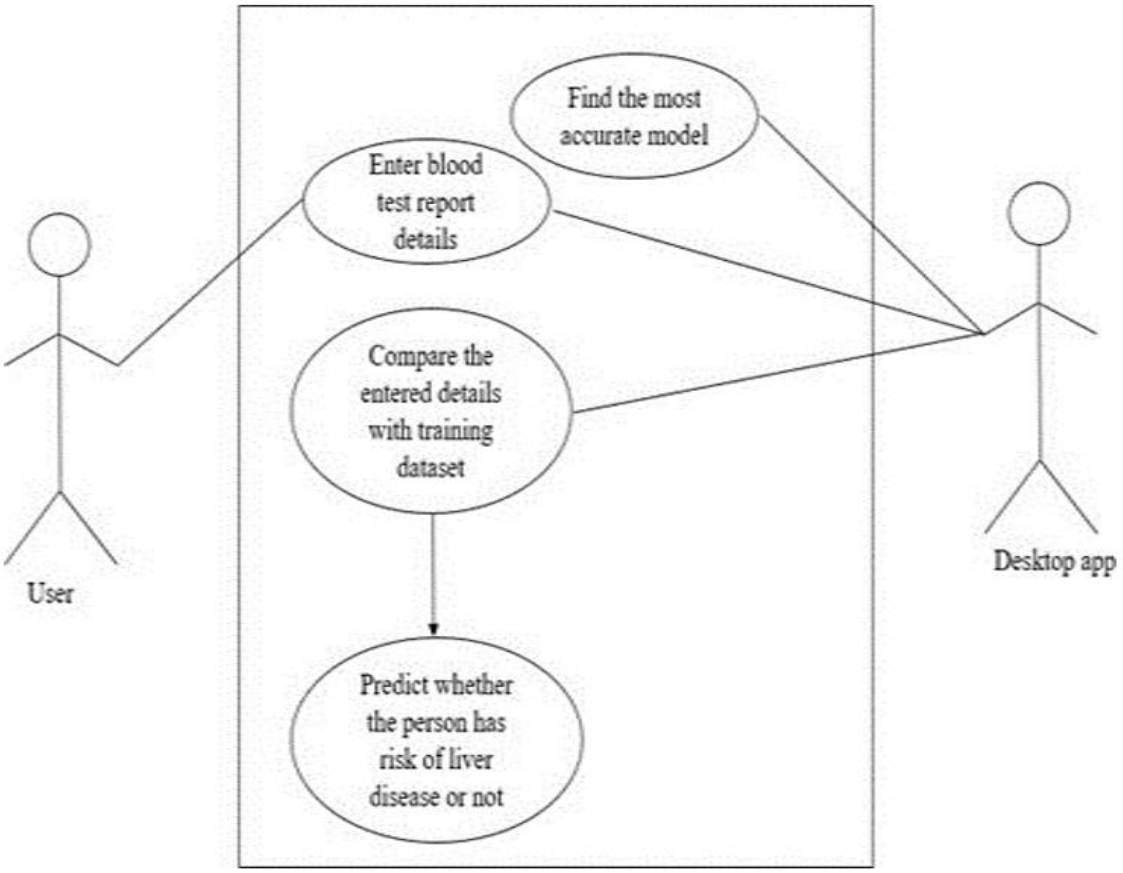

Figure 5.2 - Describes the flow of the project deployment

The flow through which applications, modules, updates, and patches are delivered from developers to users. The methods used by developers to build, test and deploy new code will impact how fast a product can respond to changes in customer preferences or requirements and the quality of each change.

5.2SOLUTION AND TECHNICAL ARCHITECTURE

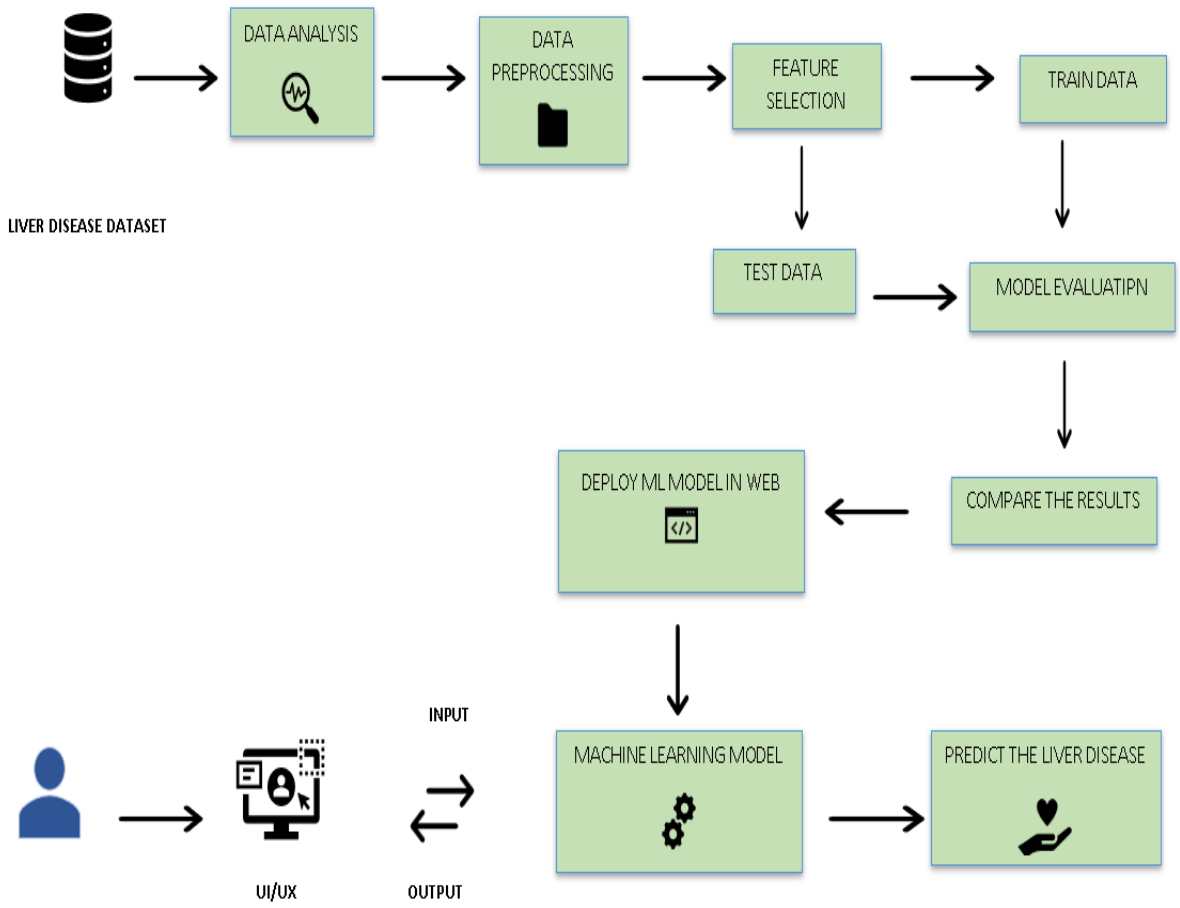

Figure 5.3 - The process of architecttural description

Solution Architects are most similar to project managers, ensuring that all parties, including stakeholders, are on the same page and moving in the right direction at all stages. Technical architects manage all activities leading to the successful implementation of a new application. A solution architect must have a technical background with at least eight years of work experience in one or more IT areas including but not limited to: IT architecture, infrastructure, and cloud development.

5.3USER STORY

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Web user)	Login	USN-1	As a user, I can register for the application by entering my email, password, and confirmingmy password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation emailonce I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the applicationthrough Mobile number	I can register & access the application	Medium	Sprint-2
	Dashboard	USN-4	As a user, I need to enter my Details	I can get information per details	High	Sprint-1
	Dashboard	USN-5	As a user, I need to enter my Test Details	I can get results based on test results	High	Sprint-1
Administrator	Services	USN-6	As an admin I need to provide valid result	I can get a result	High	Sprint-1
		USN-7	As an admin I need to provide valid/useful Suggestions	I can get suggestions	Medium	Sprint-1
	Mass Data Process	USN-8	As an admin I need to collect all the details and information	I can use it for later period	High	Sprint-1
		USN-9	As an admin I need to store all the details and information.	I can use it for later period	High	Sprint-1
Hospital Administrator	Login	USN-10	As an admin I need to login and accessdetails of customers	I can use it for further next step process	High	Sprint-1
	Dashboard	USN-11	As an admin I need to proceed the details with case head	I can use it for further next step process	High	Sprint-1

Table 5.1 - User Story

CHAPTER 6

PROJECT PLANNING AND SCHEDULING

6.1 SPRINT PLANNING AND ESTIMATION

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Login	USN-1	As a user, I can login for the application byentering my mail	1	High	Mohamed Kasim Raja M
Sprint-1		USN-2	As a user, I will login and get confirmation mailonce I have registered	2	High	Mohamed Kasim Raja M
Sprint-2		USN-3	As a user, I can login for the application through mobile number	2	Medium	Naveen Prabhu S
Sprint-3	Dashboard	USN-4	As a user, I need to enter my details	1	High	Naveen Prabhu S
Sprint-3	Dashboard	USN-5	As a user, I need to provide my Test Details	2	High	Mohamed Kasim Raja M
Sprint-3	services	USN-6	As a admin I need to provide valid result	3	High	Muhammed Rashid VP
Sprint-3		USN-7	As a admin I need to provide valid /useful suggestions	6	Medium	Chandramohan A
Sprint-4	Data Process	USN-8	As a admin need to collect all the details and information.	2	High	Chandramohan A

Sprint-4		USN-9	As a admin I need to store all the details and information	3	High	Naveen Prabhu S
Sprint-4	Login	USN-10	As a admin I need to login and access details of customers	5	High	Chandramohan A
Sprint-4	Dashboard	USN-11	As a admin I need to proceed the details with case head	12	High	Muhammed Rashid VP

Table 6.1 - Sprint planning and estimation

6.2 SPRINT DELIVERY SCHEDULE

Sprint	**Total Story Points**	**Duration**	**Sprint Start Date**	**Sprint End Date(Planned)**	**Story Points Completed (as on Planned End Date)**	**Sprint Release Date (Actual)**
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022		
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022		
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022		

6.3REPORTS FROM JIRA
BURNDOWN CHART

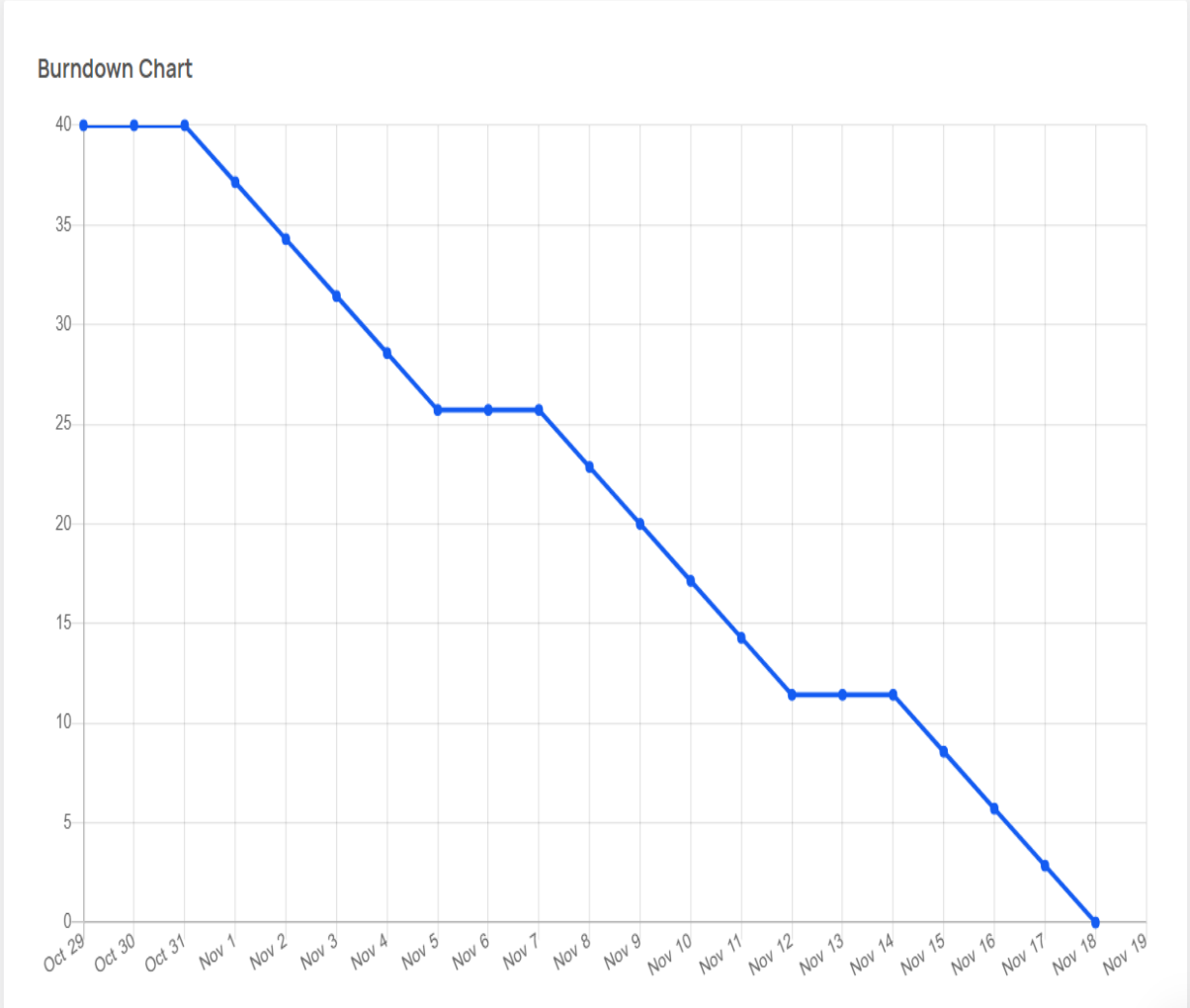

Figure 6.1 - Burndown Chart

A burndown chart shows the amount of work that has been completed in anepic or sprint, and the total work remaining. Burndown charts are used to predict your team's likelihood of completing their work in the time available.

BURNUP CHART

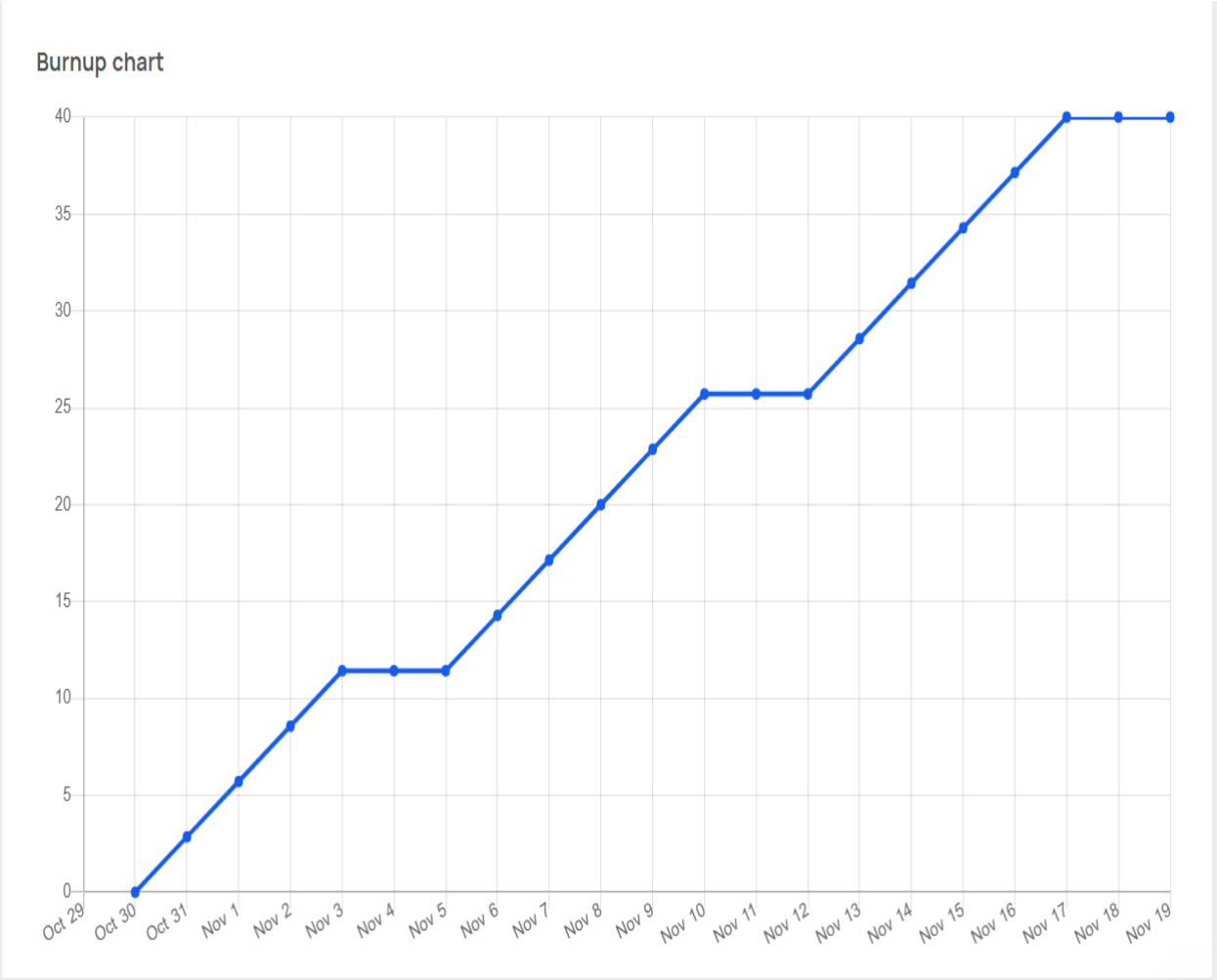

Figure 6.2 - Burnup Chart

A burnup chart highlights the work you've completed against your total project scope while a burn down chart highlights the amount of work remaining ina project. A burnup chart contains a work completed line and a project scope line.It displays the scope of a project and the work completed

CHAPTER 7

CODING AND SOLUTION

7.1 FEATURE 1

The first feature of the deployement is the process of Random Forest

Classifier is used to train and test the model for detecting the Liver Disease withthe help of collected and pre-processed dataset collections.

```
# Train Test Split:
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X_smote,y_smote,
test_size=0.3, random_state=33)
```

Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average incase of regression.

```
# RandomForestClassifier:
from sklearn.ensemble import
RandomForestClassifierRandomForest =
RandomForestClassifier() RandomForest =
RandomForest.fit(X_train,y_train)

# Predictions:
y_pred = RandomForest.predict(X_test)

# Performance:
print('Accuracy:', accuracy_score(y_test,y_pred))
print(confusion_matrix(y_test,y_pred))
rint(classification_report(y_test,y_pred))
```

Gradient boosting classifiers are a group of machine learning algorithms that combine many weak learning models together to create a strong predictivemodel. Decision trees are usually used when doing gradient boosting.

```
# GradientBoostingClassifier:
from sklearn.ensemble import
GradientBoostingClassifierGradientBoost =
GradientBoostingClassifier() GradientBoost =
GradientBoost.fit(X_train,y_train)

# Predictions:
y_pred = GradientBoost.predict(X_test)

# Performance:
print('Accuracy:', accuracy_score(y_test,y_pred))
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
```

AdaBoost can be used to boost the performance of any machine learningalgorithm. It is best used with weak learners. These are models that achieve accuracy just above random chance on a classification problem. The most suited and therefore most common algorithm used with AdaBoost are decisiontrees with one level.

```
# AdaBoostClassifier:
from sklearn.ensemble import
AdaBoostClassifierAdaBoost =
AdaBoostClassifier()
AdaBoost = AdaBoost.fit(X_train,y_train)
# Predictions:
y_pred = AdaBoost.predict(X_test)
```

```
# Performance:
print('Accuracy:', accuracy_score(y_test,y_pred))
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
```

7.2 FEATURE 2

Python flask is the first feature that helps to complete this project. Itallows the user to create local server and host the website in a local machine.

```
from flask import Flask, render_template, request
import numpy as np
import pickle
import requests
import json
```

Here we import all the necessary features of this project involving in Pythonflask.

```
header = {'Content-
Type': 'application/json', 'Authorization': 'Bearer '
+mltoken}
app = Flask(_name_)
model = pickle.load(open('liver2.pkl', 'rb'))

@app.route('/',methods
=['GET']) def Home():
    return render_template('index.html')
```

Here we created a local client's own server which serves the .html pages to the
sers.

```
@app.route("/predict",
methods=['POST'])def
```

```
predict():
    if request.method ==
        'POST': Age =
        int(request.form['
        Age'])
        Gender = int(request.form['Gender'])
        Total_Bilirubin = float(request.form['Total_Bilirubin'])
        Alkaline_Phosphotase = int(request.form['Alkaline_Phosphotase'])
        Alamine_Aminotransferase =
  int(request.form['Alamine_Aminotransferase'])Aspartate_Aminotransferase =
  int(request.form['Aspartate_Aminotransferase'])
        Total_Protiens = float(request.form['Total_Protiens'])
           Albumin = float(request.form['Albumin'])
                                                    Albumin_and_Glob
ulin_Ratio =
float(request.form['Albumin_and_Globulin_Ratio']                    v
alues                                               =
np.array([[Age,Gender,Total_Bilirubin,Alkaline_Phosphotase,Alamine_A
minotran
sferase,Aspartate_Aminotransferase,Total_Protiens,Albumin,Albumin_and
_Globulin_Ratio]])
        prediction = model.predict(values)
        return render_template('result.html',
prediction=predictionif _name_ == "_main_":
    app.run(debug=True)
```

Here we use the inputs from the html pages which has to be get by using request method in Python Flask. By validating the values from the database, we allow the user to access the home page. render_template: Used for rendering html pages on browser. url_for: Passing the control of the program to another function. session: Creates a separate session for the individual use.

CHAPTER 8

TESTING

8.1 TEST CASES

TEST CASE ID	15358	**TEST CASE DESCRIPTION**	STATISTICAL MACHINE LEARNING APPROACHES TO LIVER DISEASE PREDICTION

S.No.	PREREQUISITES	TEST DATA
1	Access to Chrome Browser	By clicking the website link
2	Entering the details required	Details should be in a integer format
3	Check for correct values	Data sholud be filled
4	Application to train the model	Provide the datasets for model training

Table 8.1 - Test Details

Test Scenario :

Step	Step Details	Expected Results	Actual Results	Pass/Fail/Not/ Executed/ Suspended
1	Navigate to website link	Site should open	As Expected	Pass
2	Enter the details	Details should be entered	As Expected	Pass
3	Click Submit	Check the result	As Expected	Pass
4	Output results	Result are generated	As Expected	Pass

Table 8.2 - Test Cases

8.2 USER ACCEPTANCE TESTING

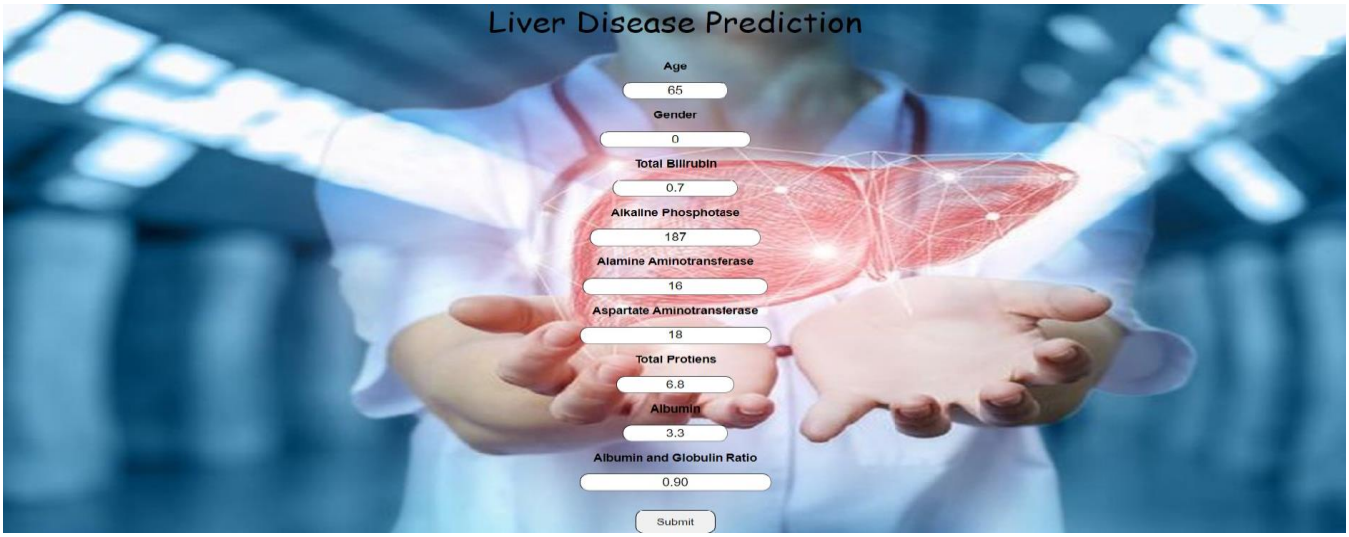

Figure 8.1 - User Acceptance Test 1

In this the user will be entering tested values which is moved for prediction.

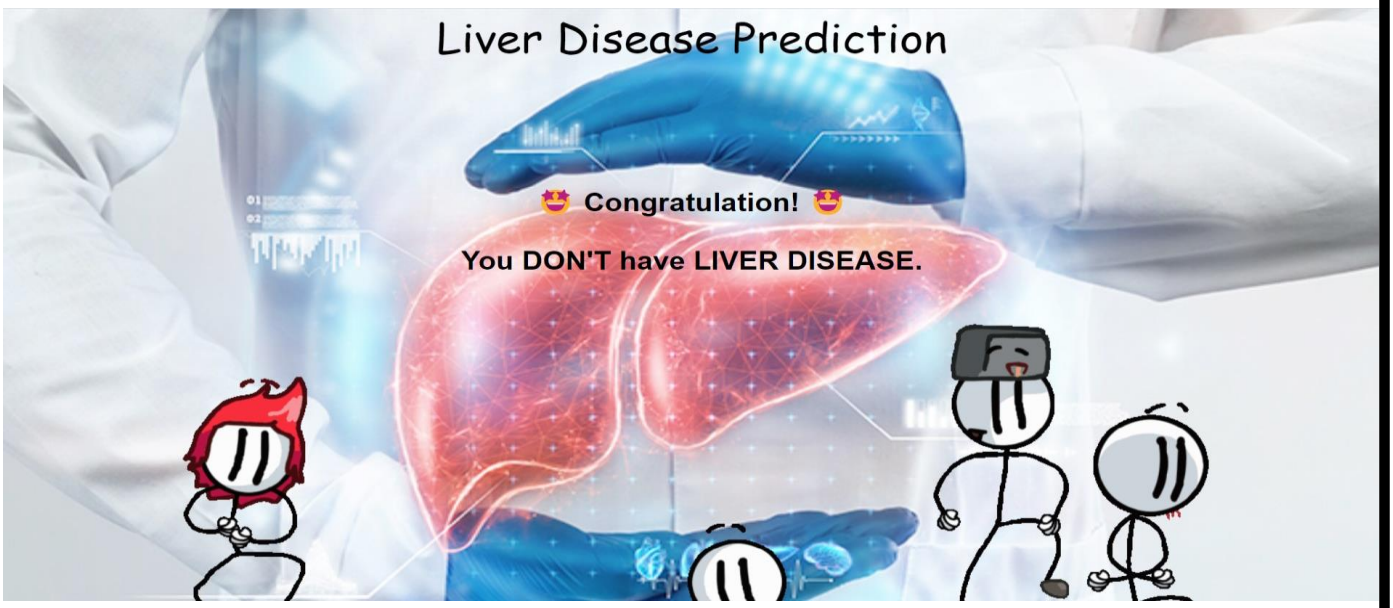

Figure 8.2 - User Acceptance Test 2

In this the data which was entered by the user will be analyzed. This figureshows the person does not have liver disease.

.

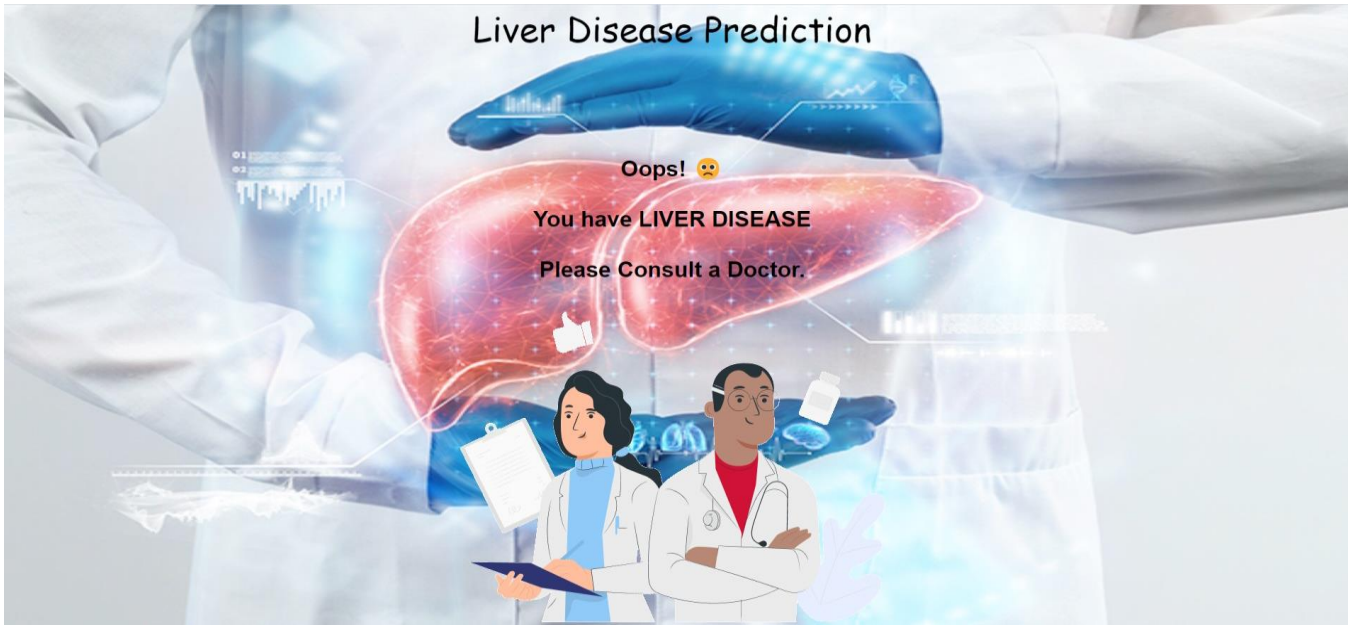

Figure 8.3 - User Acceptance Test 3

In this the data which was entered by the user will be analyzed. This figureshows the person have a liver disease.

CHAPTER 9

RESULTS

9.1 PERFORMANCE METRICS

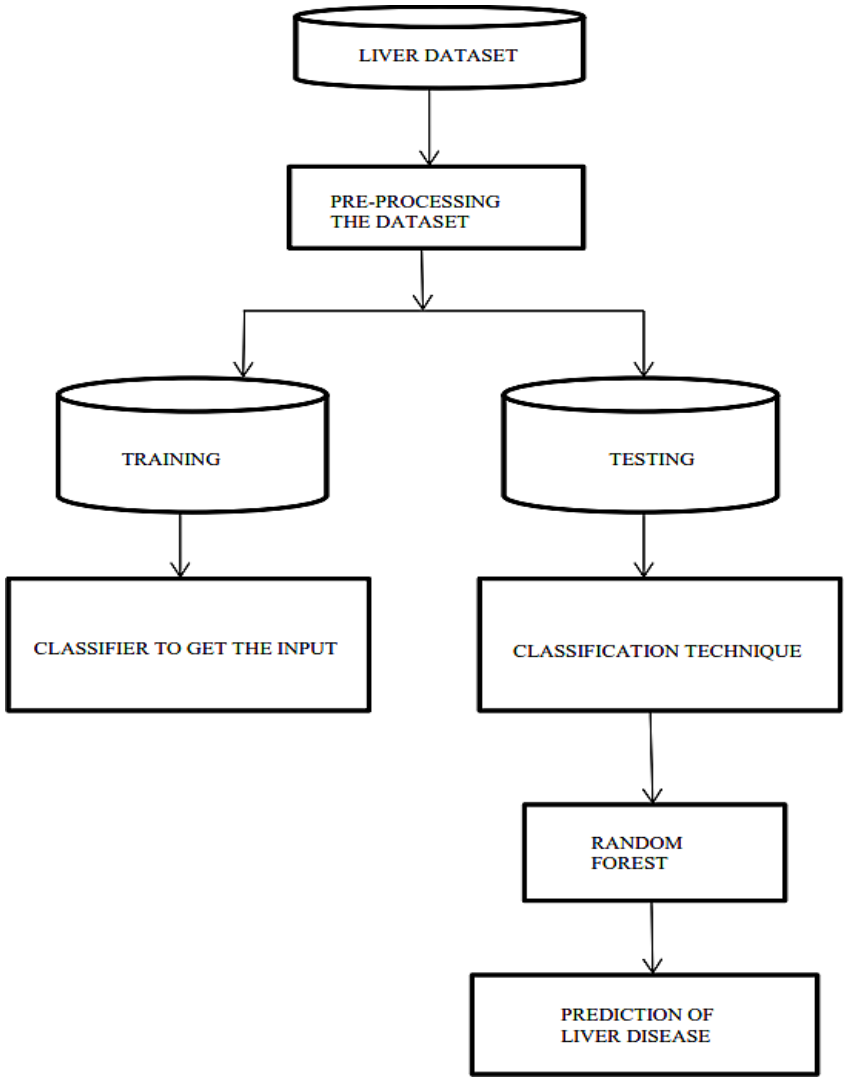

Figure 9.1 - Flow of the Performance

Performance Metrics is performed along for all the above assigned tasks.In each and every tasks various metrics is performed inorder to provide the optimum outcome.Results gives the best outcome as expected in project.

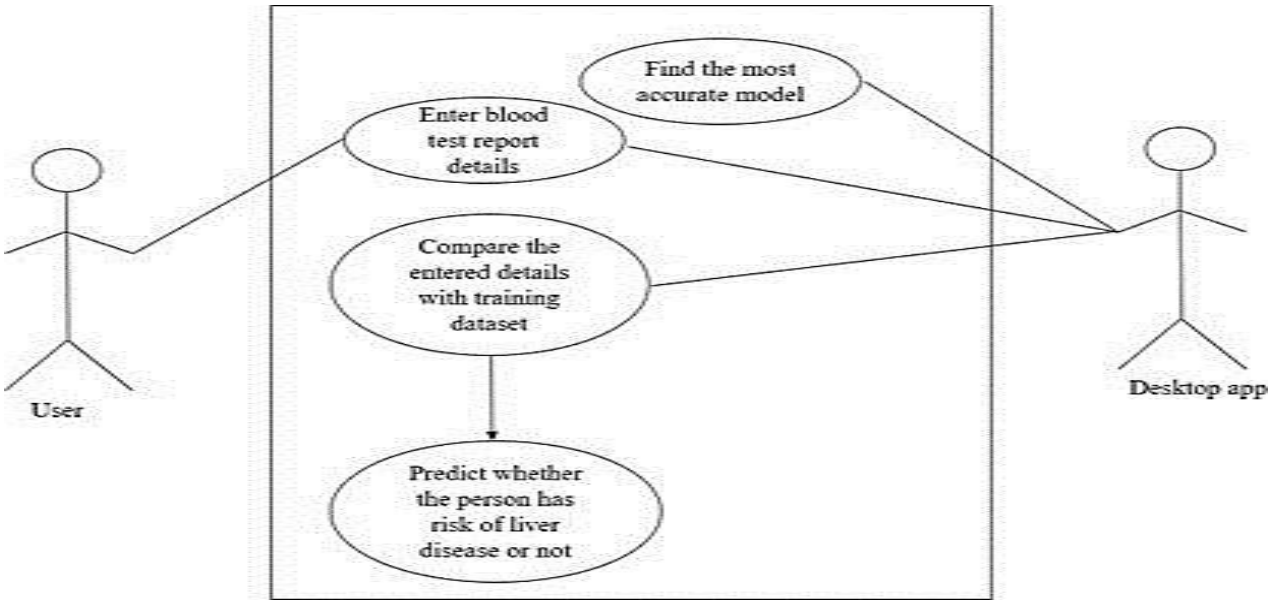

Figure 9.2 - Performance Matrix

One can use following execution measures for the request and figure of imperfection slanted module as shown by his/her own need.

Confusion Matrix: The confusion matrix is used to measure the introduction of two class issue for the given instructive record. The right cornerto corner parts TP (True positive) and TN (True Negative) adequately describe instances similarly as FP (false positive) and FN (false negative) wrongly request instances. Confusion Matrix correctly classify instance TP+TN incorrectly classify instances.

- True positives imply the positive liver tuples that were precisely named bythe classifier,
- True negatives are the negative liver tuples that were precisely set apart bythe classifier.
- False positives are the negative liver tuples that were erroneously set apartas positive tuples.
- False negatives are the positive liver tuples that were incorrectly stampednegative tuples.

CHAPTER 10

ADVANTAGES AND DISADVANTAGES

ADVANTAGES

In pharmaceutical industries, random forest can be used to identify the potential of a certain medicine or the composition of chemicals required formedicines. Random forest can also be used in hospitals to identify the diseasessuffered by a patient ,risk of liver disease and many other diseases where early analysis and research play a crucial role.Our study has successfully developed and validated the first risk prediction model and subsequent user-friendly scoring tool, the algorithm for Liver Function Investigations, for liver condition diagnosis in patients with no obvious liver condition at the time of incident liverfunction testing in primary care.

DISADVANTAGES

The idea for the approach of this project arises from the current situation regarding the increase in the confirmatory diagnosis of liver, and lack of treatment or the user's ignorance of its pathologies leads to irreversible kidney failure in the final stages of disease, such as dialysis for life, financially affecting the health system, as it is a costly treatment that generates the most significant amount of absorption of the resources available for health. This could be reduced by using tools such as machine learning to classify from the initial stages. Although the application of machine learning in healthcare and other areas is favorable, the field of liver disease has not yet exploited its full potential.

CHAPTER 11

CONCLUSION

The principal part of this work is to make an effective diagnosis system for liver infection patients. The application will have the option to predict liver infection prior and advise the wellbeing condition. This application can be surprisingly gainful in low-salary nations where our absence of medicinal foundations and just as particular specialists. In our study, there are a few bearings for future work in this field. We just explored some popular supervised machine learning algorithms, more algorithms can be picked to assemble an increasingly precise model of liver disease prediction and performance can be progressively improved. Additionally, this work likewise ready to assume a significant role in health care research and just as restorative focuses to anticipate liver infection.

CHAPTER 12

FUTURE SCOPE

Diseases related to liver is becoming more and more common with time. With continuous technological advancements, these are only going to increase in the future. Although people are becoming more conscious of healthnowadays and are joining yoga classes, dance classes; still the sedentary lifestyle and luxuries that are continuously being introduced and enhanced; theproblem is going to last long. So, in such a scenario, our project will be extremely helpful to the society. With the dataset that we used for this project, we got 81% accuracy for Random forest model, and though it might be difficult to get such accuracies with very large datasets, from this projects results, one can clearly conclude that we can predict the risk of liver diseases with accuracyof 90 % or more. Also it can be incorporated into a website and these app and website will be highly beneficial for a large section of society.

CHAPTER 13

APPENDIX

SOURCE CODE:

Algorithm :

```
#Importing The Libraries
 import pandas as pd
 import numpy as np
 import matplotlib.pyplot as plt
 import seaborn as sns
 import pickle

#Reading the Dataset data=pd.read_csv(path)

#Exploratory  Data  Analysis
data.head()
data.tail() data.info()
data.describe()

#Checking For Null Values And Handling Null Values
 data.isnull().any()
 data.isnull().sum()
data['Albumin_and_Globulin_Ratio']=data.fillna(data['Albumin_and_Globulin_Ratio']
.mode()[0])
data.isnull().sum()
 #Data Visualization
 sns.countplot(data=data,x='Gender',lable='Count')
 m,f=data['Gender'].value_counts()
 print("No of Males:",m)
 print("No of Females:",f)
```

```
sns.countplot(data=data, x='Dataset')
LD,NLD=data['Dataset'].value_counts()
print("liver disease patients:",LD)
print("Non-liver disease patients:",NLD)

#Splitting The Dataset Into Dependent And Independent Variable
x=data.iloc[:,0:-1]
y=data.iloc[:,-1]

#Split The Dependent And Independent Features Into Train Set And TestSetfrom
sklearn.model_selction import train_test_split
xtrain.xtest,ytrain,ytest=train_test_split(x,y,test_size=0.2)

#Check the shape of both xtrain and xtest.
xtrain.shape
xtest.shape
#importing the machine learning model
from sklearn.svm import svc
from sklearn.ensemble import RandomForestClassifierfrom
sklearn.neighbors import KNeighborsClassifier #initiailizing
the machine learning models
svm=SVC()
RFmodel=RandomForestClassifier()
KNNmodel=KNeighborsClassifiers()
svm=SVC()
#Train the data with SVM model
svm.fit(xtrain,ytrain)
#Random forest classifiers model
from sklearn.ensemble import RandomForestClassifier
RFmodel=RandomForestClassifier()
#Train the data with random forest model
RFmodel.fit(xtrain,ytrain)
SVMaccuracy=accuracy_score(SVMpred,ytest)SVMaccuracy
```

```
from sklearn.ensemble import RandomForestClassifier
RFmodel=RandomForestClassifier()
RFmodel.fit(xtrain,ytrain) RFpred=RFmodel.predict(xtest)
RFaccuracy=accuracy_score(RFpred,ytest)
RFaccuracy
RFcm=confusion_matrix(RFpred,ytest)
RFcm
from sklearn.neighbors import KNeighborsClassifier
KNN=KNeighborsClassifier()
KNN.fit(xtrain,ytrain)
KNNpred=KNN.preduct(xtest)
KNNaccuracy=accuracy_score(KNNpred,ytest)
KNNaccuracy
KNNcm=confusion_matrix(KNNpred,ytest)
KNNcm
# saving the model
import pickle
pickle.dump(svm,open('liver_analysis.pkl','wb'))
```

Flask Connection

```
from flask import Flask, render_template, request
import numpy as np
import pickle
app = Flask(__name__)
model = pickle.load(open('Liver2.pkl', 'rb'))
@app.route('/',methods=['GET'])
def Home():
    return render_template('index.html')
@app.route("/predict", methods=['POST'])def
predict():
    if request.method == 'POST': Age
        = int(request.form['Age'])
```

```
        Gender = int(request.form['Gender'])
        Total_Bilirubin = float(request.form['Total_Bilirubin']) Alkaline_Phosphotase =
        int(request.form['Alkaline_Phosphotase'])Alamine_Aminotransferase =
int(request.form['Alamine_Aminotransferase'])
        Aspartate_Aminotransferase =
int(request.form['Aspartate_Aminotransferase']) Total_Protiens =
        float(request.form['Total_Protiens'])Albumin =
        float(request.form['Albumin'])
        Albumin_and_Globulin_Ratio =
float(request.form['Albumin_and_Globulin_Ratio'])
        values =
np.array([[Age,Gender,Total_Bilirubin,Alkaline_Phosphotase,Alamine_Amino
transferase,Aspartate_Aminotransferase,Total_Protiens,Albumin,Albumin_a
nd_Globulin_Ratio]])
        prediction = model.predict(values)
        return    render_template('result.html',    prediction=prediction)    if
name_== "_main_":
   app.run(debug=True)
```

HTML Templates

index.htm

```
<!DOCTYPE html>
<html lang="en">
<head>
   <meta charset="UTF-8">
   <title>Liver Prediction Model</title>


</head>
<body>
   <div class="container">
      <h2 class='container-heading'><span class="heading_font">LiverDisease
Prediction</span></h2>
```

```
    </div>

    <div class="ml-container">
        <form action="{{ url_for('predict') }}" method="POST">
            <br>
            <h3>Age</h3>
            <input id="first" name="Age" placeholder="in Year"
required="required">
            <br>
            <h3>Gender</h3>
            <input id="second" name="Gender" placeholder="Male = 1,Female=0"
required="required">
            <br>
            <h3>Total Bilirubin</h3>
            <input id="third" name="Total_Bilirubin" placeholder="Total Bilirubin"
required="required">
            <br>
            <h3>Alkaline Phosphotase</h3>
            <input id="fourth" name="Alkaline_Phosphotase"
placeholder="Alkaline Phosphotase" required="required">
            <br>
            <h3>Alamine Aminotransferase</h3>
            <input id="fifth" name="Alamine_Aminotransferase" placeholder="Alamine
Aminotransferase" required="required">
            <br>
            <h3>Aspartate Aminotransferase</h3>
            <input id="sixth" name="Aspartate_Aminotransferase" placeholder="Aspartate
Aminotransferase" required="required">
            <br>
            <h3>Total Protiens</h3>
```

```
        <input id="seventh" name="Total_Protiens" placeholder="TotalProtiens"
required="required">
        <br>
        <h3>Albumin</h3>
        <input id="eight" name="Albumin" placeholder="Albumin"
required="required">
        <br>
        <h3>Albumin and Globulin Ratio</h3>
        <input id="ninth" name="Albumin_and_Globulin_Ratio" placeholder="Albumin
and Globulin Ratio" required="required">
        <br>
        <br>
        <br>
        <button id="sub" type="submit ">Submit</button>
        <br>
        <br>
        <br>
        <br>
      </form>
   </div>
<style>
/* Background Image */
body
{
background-
image:url("https:/     akcdn.detik.net.id/community/media/visual/2020/12/17
/6-makanan-untuk-mencegah-penyakit-liver-salah-satunya-kopi- 5_43.jpeg?w=250&q=");
height: 100%;
/* Center and scale the image nicely */
```

```
background-position: center;
background-repeat: no-repeat;
background-size: 100% 100%;
}
/* Color */
body{
    font-family: Arial, Helvetica,sans-serif;
    text-align: center;
    margin: 0;
    padding: 0;
    width: 100%;
    height: 100%;
    display: flex;
    flex-direction: column;
}
/* Heading Font */
.container-heading{
    margin: 0;
}
.heading_font{
    color: #black;
    font-family: 'Pacifico', cursive;
    font-size: 50px;
    font-weight: normal;
}
/* Box */
        #first {
            border-radius: 14px;height:
            25px;
            width: 150px;
```

```
    font-size: 20px;
    text-align: center;
}
#second {
    border-radius: 14px;height:
    25px;
    width: 220px;
    font-size: 20px;
    text-align: center;
}
#third {
    border-radius: 14px;height:
    25px;
    width: 180px;
    font-size: 20px;
    text-align: center;
}
#fourth {
    border-radius: 14px;height:
    25px;
    width: 250px;
    font-size: 20px;
    text-align: center;
}
#fifth {
    border-radius: 14px;height:
    25px;
    width: 270px;
    font-size: 20px;
    text-align: center;
```

```
        }
        #sixth {
            border-radius: 14px;height:
            25px;
            width: 280px;
            font-size: 20px;
            text-align: center;
        }
        #seventh {
            border-radius: 14px;height:
            25px;
            width: 170px;
            font-size: 20px;
            text-align: center;
        }
        #eight {
            border-radius: 14px;height:
            25px;
            width: 150px;
            font-size: 20px;
            text-align: center;
        }
        #ninth {
            border-radius: 14px;height:
            25px;
            width: 280px;
            font-size: 20px;
            text-align: center;
        }
    /* Submit Button */
```

```
#sub {
        width: 120px;height:
        43px;
        text-align: center;
        border-radius: 14px;
        font-size: 18px;
      }

 </style>
 </body>
 </html>
```

result.html

```
<!DOCTYPE html>
<html lang="en">
<head>
   <meta charset="UTF-8">
   <meta name="viewport" content="width=device-width, initial-scale=1.0">
   <title>Liver Disease Result</title>
</head>
<body>
   <div class="container">
      <form action="{{ url_for('predict')}}" method="post">
         <h2 class='container-heading'><span class="heading_font">LiverDisease
Prediction</span></h2>
      <br><br><br><br><br><br><br>
      <!-- Result -->
      <div class="results">
         {% if prediction==2 %}
            <h1><span class='danger'>Oops! □<br><br>You have LIVERDISEASE
<br><br>Please Consult a Doctor.</span></h1>
            <img class="gif" src="{{ url_for('static', filename='dr.gif')}}"
```

```
alt="LIVER Image">
        {% elif prediction==1 %}
            <h1><span class='safe'> Congratulation! <br><br>You DON'T have
LIVER DISEASE.</span></h1>
            <img class="gif1" src="{{ url_for('static', filename='yes.webp')}}"alt="Not
LIVER Image">
        {% endif %}
      </div>
      </form>
   </div>
<style>
/* Background Image */
body
{
background-image:url("https:/ d1vbn70lmn1nqe.cloudfront.net/prod/wp-
content/uploads/2022/05/11050117/hepatitis-kronis-halodoc.jpg"); height:
100%;
/* Center and scale the image nicely */
background-position: center; background-
repeat: no-repeat; background-size: 100%
100%;
}
/* Color */
body{
   font-family: Arial, Helvetica,sans-serif;
   text-align: center;
   margin: 0;
   padding: 0;
   width: 100%;
   height: 100%;
```

```
    display: flex;
    flex-direction: column;
}
/* Heading Font */
.container-heading{
    margin: 0;
}
.heading_font{
    color: #black;
    font-family: 'Pacifico', cursive;
    font-size: 50px;
    font-weight: normal;
}
</style>
</body>
</html>
```

Github and Project Video Demo Link

Github Link:

https://github.com/IBM-EPBL/IBM-Project-5486-1658767689

Project Video Demo Link:

https://drive.google.com/file/d/1zj4ueDfX36pbSc5uojRrfPHC8NTGUamG/view?usp=sharing