

NAALAIYA THIRAN – EXPERIENTIAL PROJECT BASED LEARNING INITIATIVE

**19CS532 - PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY
AND
ENTERPRENURSHIP**

A PROJECT REPORT ON

STATISTICAL MACHINE LEARNING APPROACHES TO LIVER DISEASE PREDICTION

Submitted by

TEAM ID : PNT2022TMID03236

AKASH R	212219040007
KARTHICK S	212219040058
CHANDRAKISHORE T	212219040164
DURGESH H	212219040027

Submitted in partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING



**SAVEETHA ENGINEERING COLLEGE
ANNA UNIVERSITY: CHENNAI- 600 025
NOVEMBER 2022**

TABLE OF CONTENTS

- 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
- 9. RESULTS**
 - 9.1 Performance Metrics
- 10. ADVANTAGES & DISADVANTAGES**
- 11. CONCLUSION**
- 12. FUTURE SCOPE**
- 13. APPENDIX**
 - Source Code
 - GitHub & Project Demo Link

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 help the doctors to diagnose the disease within a short duration of time. Discovering the 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-glutamyl transferase. The Liver Disease score can be used as part of health counseling, and for planning further liver investigations and follow-up.

CHAPTER 2

LITERATURE SURVEY

2.1 EXISTING PROBLEM

Using machine learning algorithms to predict disease is made possible by increasing access to hidden attributes in medical data sets. Various kinds of data sets, such as blood panels with liver function tests, histologically stained slide images, and the presence of specific molecular markers in blood or tissue samples, have been used to train classifier algorithms to predict liver disease with good accuracy. The ML methods described in previous studies have been evaluated for accuracy by a combination of confusion matrix, receiver operating characteristic under area under curve, and k-fold cross-validation.

Singh et al. designed software based on classification algorithms (including logistic regression, random forest, and naive Bayes) to predict the risk of liver disease from a data set with liver function test results.

Vijayarani and Dhavanand found that SVM performed better over naive Bayes to predict cirrhosis, acute hepatitis, chronic hepatitis, and liver cancers from patient liver function test results. SVM with particle swarm optimization (PSO) predicted the most important features for liver disease detection with the highest accuracy over SVM, random forest, Bayesian network, and an MLP-neural network. SVM more accurately predicted drug-induced hepatotoxicity with reduced molecular descriptors than Bayesian and other previously used models.

Phan and Chan et al demonstrated that a convolutional neural network (CNN) model predicted liver cancer in subjects with hepatitis with an accuracy of 0.980. The ANN model has been used to predict liver cancer in patients with type 2 diabetes. Neural network ML methods can help differentiate between types of liver cancers when applied to imaging data sets. Neural network algorithms have even been trained to predict a patient's survival after liver tumour removal using a data set containing images of processed and stained tissue from biopsies.

ML methods can facilitate the diagnosis of many diseases in clinical settings if trained and tested thoroughly. More widespread application of these methods to varying data sets can further improve accuracy in current deep learning methods. This study aimed to (i) impute missing data using the MICE algorithm; (ii) determine variable selection using eigen decomposition of a data matrix by PCA and to rank the important variables using the Gini index; (iii) compare among several statistical learning methods the ability to predict binary classifications of liver disease; (iv) use the synthetic minority oversampling technique (SMOTE) to oversample minority class to regulate overfitting; (v) obtain confusion matrices for comparing actual classes with predictive classes; (vi) compare several ML approaches to assess a better performance of liver disease diagnosis.(viii) evaluate receiver

operating characteristic (ROC) curves for determining the diagnostic ability of binary classification of liver disease.

2.2 REFERENCES

1. Wang, Y.; Li, Y.; Wang, X.; Gacesa, R.; Zhang, J.; Zhou, L.; Wang, B. Predicting Liver Disease Risk Using a Combination of Common Clinical Markers: A Screening Model from Routine Health Check-Up. *Dis. Markers* **2020**, 2020, 8460883.
2. Torkadi, P.P.; Apte, I.C.; Bhute, A.K. Biochemical evaluation of patients of alcoholic liver disease and non-alcoholic liver disease. *Indian J. Clin. Biochem.* **2014**, 29, 79–83. [
3. Ceriotti, F.; Henny, J.; Queraltó, J.; Ziyu, S.; Özarda, Y.; Chen, B.; Boyd, J.C.; Panteghini, M. Common reference intervals for aspartate aminotransferase (AST), alanine aminotransferase (ALT) and -glutamyl transferase (GGT) in serum: Results from an IFCC multicenter study. *Clin. Chem. Lab. Med.* **2010**, 48, 1593–1601. *Livers* **2021**, 1 311.
4. Chalasani, N.; Younossi, Z.; Lavine, J.E.; Charlton, M.; Cusi, K.; Rinella, M.; Harrison, S.A.; Brunt, E.M.; Sanyal, A.J. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. *Hepatology* **2018**, 67, 328–357.
5. Woreta, T.A.; Saleh, A.A. Evaluation of abnormal liver tests. *Med Clin.* **2014**, 98, 1–16.
6. Robles-Diaz, M.; Garcia-Cortes, M.; Medina-Caliz, I.; Gonzalez-Jimenez, A.; Gonzalez-Grande, R.; Navarro, J.M.; Castiella, A.; Zapata, E.M.; Romero-Gomez, M.; Blanco, S.; et al. The value of serum aspartate aminotransferase and gamma-glutamyl transpeptidase as biomarkers in hepatotoxicity. *Liver Int.* **2015**, 35, 2474–2482.
7. Borroni, G.; Ceriani, R.; Cazzaniga, M.; Tommasini, M.; Roncalli, M.; Maltempo, C.; Felling, C.; Salerno, F. Comparison of simple tests for the non-invasive diagnosis of clinically silent cirrhosis in chronic hepatitis C. *Aliment. Pharmacol. Ther.* **2006**, 24, 797–804.

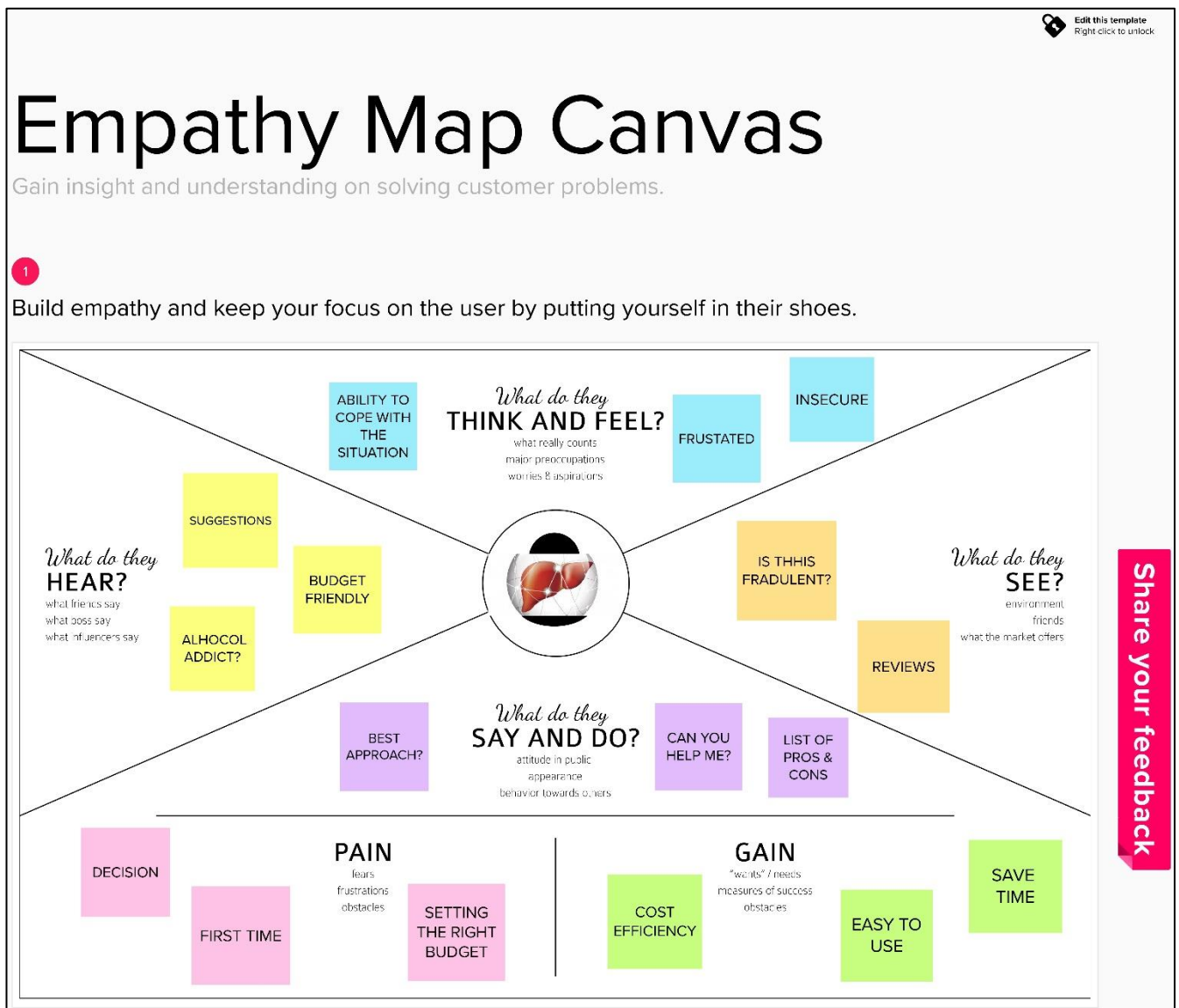
2.3 PROBLEM STATEMENT DEFINITION

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 help the doctors to diagnose the disease within a short duration of time. Discovering the existence of liver disease at an early stage is a complex task for the doctors. The main objective of this project is to analyse 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.

CHAPTER 3

IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS



3.2 IDEATION & BRAINSTORMING

Brainstorm & idea prioritization

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

- 10 minutes to prepare
- 1 hour for 10 iterations
- 3-5 people per session

[Share template feedback](#)

1 Problem Statement

To develop a machine learning model for Liver disease prediction in early stage

PROBLEM

To choose a classification algorithm which best fits for the given problem statement and able to map output with low cost of learning and with high accuracy

2 Liver disease prediction using machine learning

To analyze the patients data and to select a machine learning model which gives high accuracy for predicting the liver disease of the patients in the early stage

Akash R

High Accuracy	Security	Confidential
User Friendly	SMS/email Support	Forwarding result to nearest Hospital

Karthick S

Rapid Prediction	Instant Accurate results	Early stage Prediction
Survival Rate Increases	Easy Report Generation	Risk Free

Chandrakishore T

Enhanced UI	Reliability	Cost reduction
Early Diagnosis	Trustable	Availability

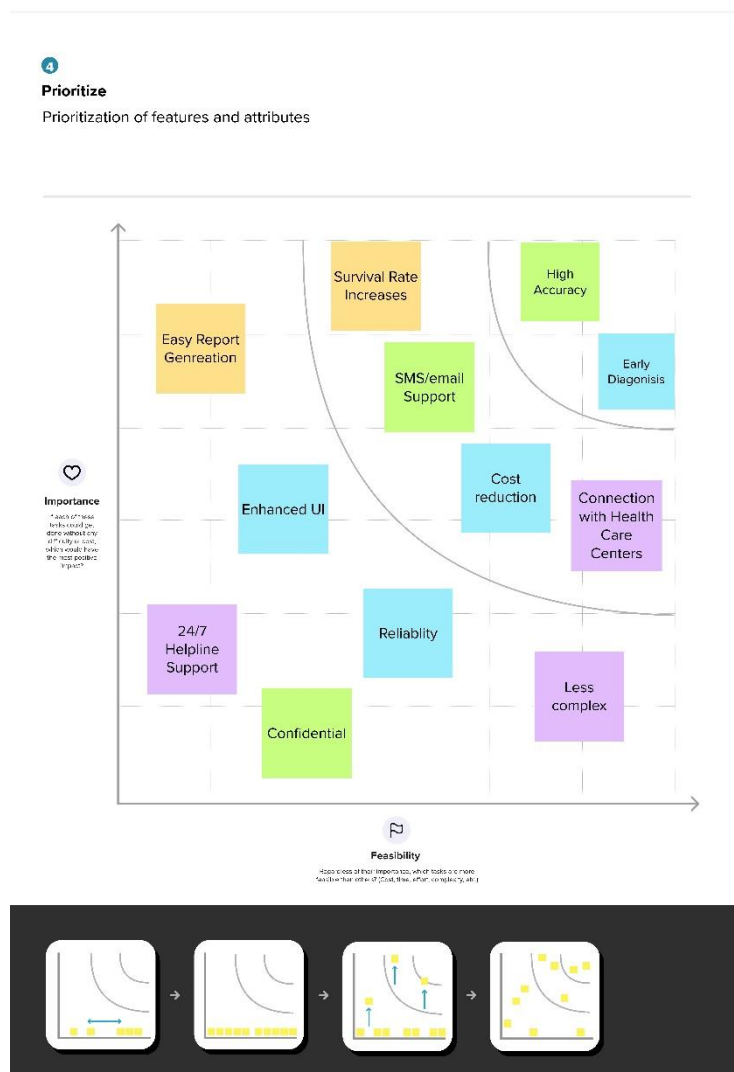
Durgesh H

Adoptability	Easy access	Integrity
Connection with Health Care Centers	24/7 Helpline Support	Less complex

Need some inspiration?

Use a variety of online resources to generate ideas.

[Open source](#)




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	<ul style="list-style-type: none">• 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 (Revenue Model)	<ul style="list-style-type: none">• 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	<ul style="list-style-type: none">• Can be extended to predict many classifications of diseases in early stage• This can be integrated to with any hospitals and health sectors to get patient records securely through APIs

3.4 PROBLEM SOLUTION FIT

Problem-Solution fit canvas 2.0		Purpose / Vision		
Define CS, fit into CC	1. CUSTOMER SEGMENT(S) CS Who is your customer? -Hospitals & Patients ,who needs to find he/she is affected by Liver Disease	6. CUSTOMER CONSTRAINTS CC What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices. -No Alcohol consumption -Maintain a balanced nutrient diet -Avoid using Drugs	5. AVAILABLE SOLUTIONS AS Which solutions are available to the customers when they face the problem or need to get the job done? What have they tried in the past? What pros & cons do these solutions have? i.e. pen and paper is an alternative to digital notetaking -Liver Transplant -Biopsy -Available of Hybrid ML classification method	Explore AS, differentiate
	2. JOBS-TO-BE-DONE / PROBLEMS J&P Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one; explore different sides. -Abdominal pain and Swelling -Liver can no longer process nutrients, enzymes, and heavy damage causes cirrhosis	9. PROBLEM ROOT CAUSE RC What is the real reason that this problem exists? What is the back story behind the need to do this job? i.e. customers have to do it because of the change in regulations. -Heavy usage of Alcohol -Genetics cause -Indigestion of intaking drug -Due to Obesity	7. BEHAVIOUR BE What does your customer do to address the problem and get the job done? i.e. directly related: find the right solar panel installer, calculate usage and benefits; indirectly associated: customers spend free time on volunteering work (i.e. Greenpeace) - Need to visit doctor is problem persists - Better medicinal solution needed to be adopted	
Identify strong TR & EM	3. TRIGGERS TR What triggers customers to act? i.e. seeing their neighbour installing solar panels, reading about a more efficient solution in the news. - Living with pain makes more uncomfortable so they are unable to lead a happy life	10. YOUR SOLUTION SL If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much it fits reality. If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behaviour. - Generate an application for getting the user input and based on the results obtained giving a prevention solution, Treatment Recommendation solution ,which would is better to them as well as others	8. CHANNELS of BEHAVIOUR CH 8.1 ONLINE What kind of actions do customers take online? Extract online channels from #7 -Users able to get results as per their data in online 8.2 OFFLINE What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development. - Able to consult doctor based on the results	Extract online & offline CH of BE
	4. EMOTIONS: BEFORE / AFTER EM How do customers feel when they face a problem or a job and afterwards? i.e. lost, insecure > confident, in control - use it in your communication strategy & design. - People who are unable to take numerous test and spend huge money ,are now able to find the result in a less cost manner and in short period of time gives them a hope of recovery.			



Problem-Solution fit canvas is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 license
 Created by Daria Nepriakhina / Amaltama.com



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

4.2 NON-FUNCTIONAL REQUIREMENT

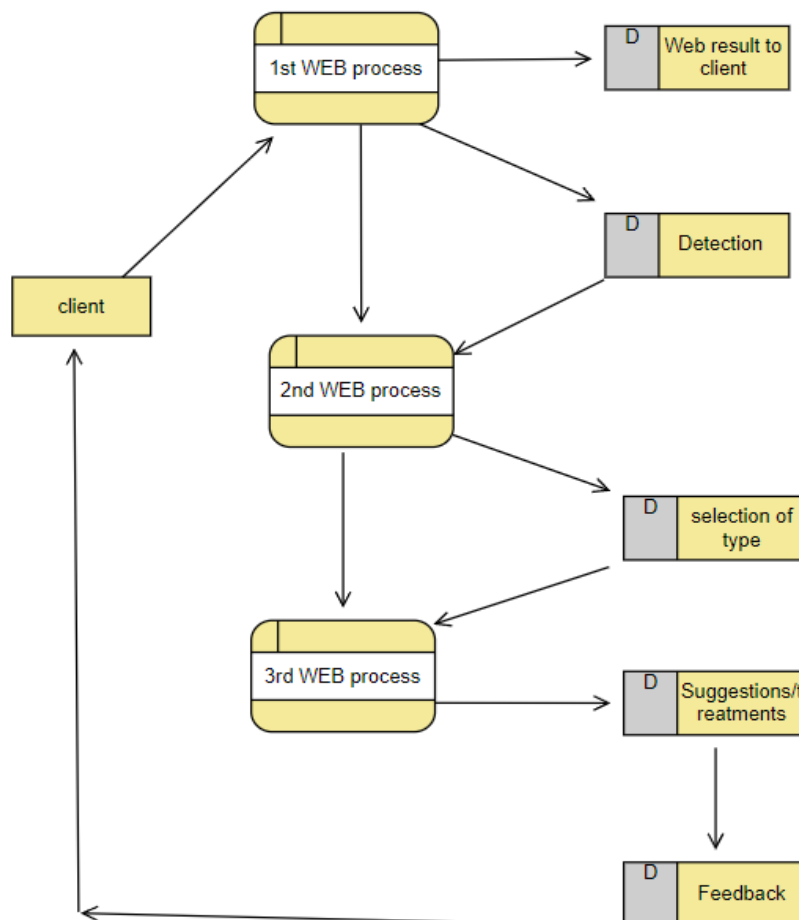
FR No.	Non-Functional Requirement	Description
NFR-1	Usability	Defines how difficult it will be for a user to learn and operate the system. Usability can be accessed from different points
NFR-2	Security	Security requirements ensure that the software is protected from unauthorized access to the system 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.

CHAPTER 5

PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



5.2 SOLUTION & TECHNICAL ARCHITECTURE

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2

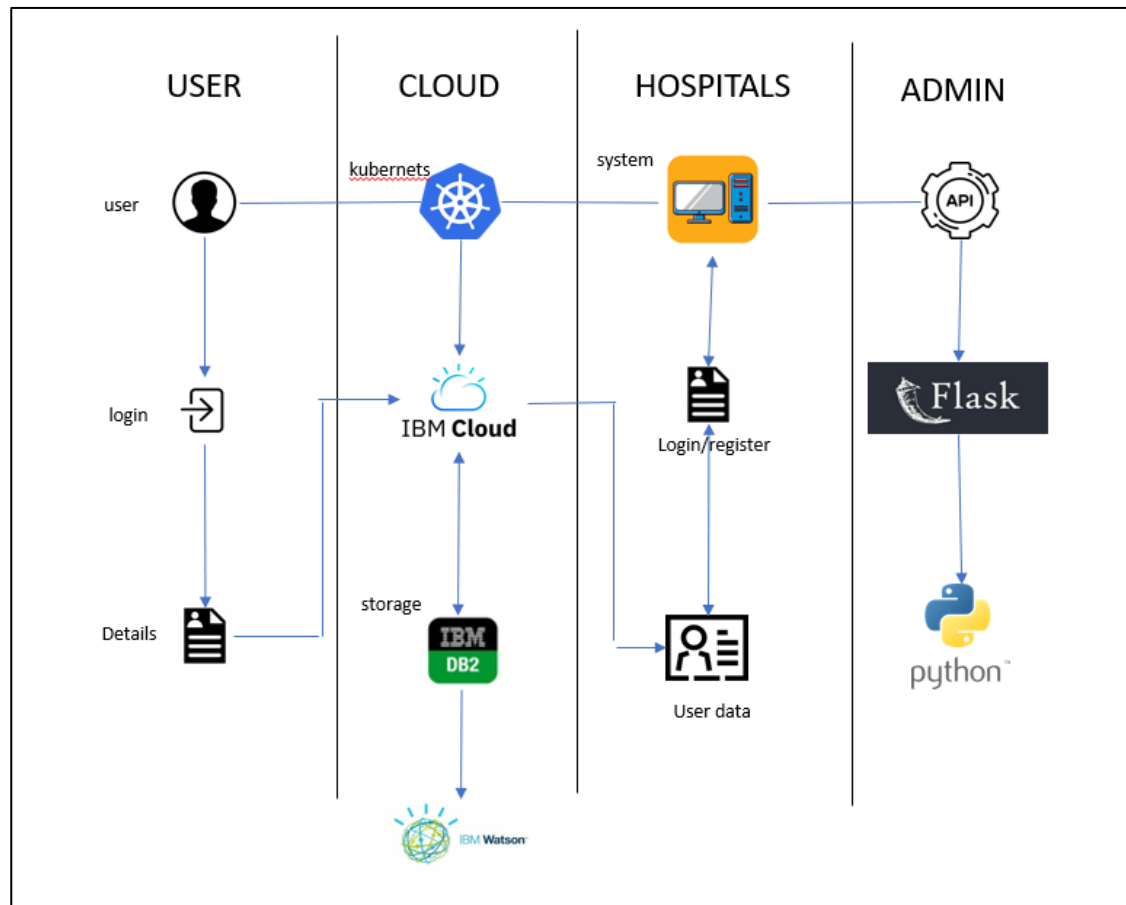


TABLE-1: COMPONENTS & TECHNOLOGIES:

S.NO	COMPONENT	DESCRIPTION	TECHNOLOGY
1.	User Interface	User access to the application through the mobile application.	HTML
2.	Application Logic-1	Creating an application interface	Python & Flask
3.	Application Logic-2	Creating an AI assistant that gives medical services to the user.	IBM Watson Assistance
4.	File Storage	Files are stored in the local storage and stored in the cloud.	IBM Block Storage or Other Storage Service or Local Filesystem
5.	External API-1	Use this REST API to manage locations. Get all locations. URI, /admin/resources/locations. Method, GET.	IBM Location REST API

6.	Deep Learning Model	Creating an algorithm to calculate case information provides by the hospitals.	Object Recognition Model, etc.
7.	Infrastructure (Cloud)	IBM Cloud App Configuration is a centralized feature-management and configuration service on IBM Cloud.	IBM Cloud Foundry & Kubernetes

TABLE-2: APPLICATION CHARACTERISTICS:

S.NO	CHARACTERISTICS	DESCRIPTION	TECHNOLOGY
1.	Open-Source Frameworks	There are no open-source frameworks in this application.	Python
2.	Security Implementations	Blockchain technology is used for Security implementation its private framework protects all data.	Blockchain
3.	Scalable Architecture	Users are provided with medical services online and giving awareness to people by giving therapeutic medicines and monitoring user movements in pandemic zones and alerts before they are affected.	IBM Cloud
4.	Availability	Medicinal Recommendations, Test kits, Doctor suggestions, and Updated Contaminated zones are available in applications.	IBM Watson Assistant
5.	Performance	The geo-fencing algorithm is updated daily and shows the day-to-day updates of the contaminated zones.	Geofence

5.3 USER STORIES

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 confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once 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 application through 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 access details 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

CHAPTER 6

PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING & ESTIMATION

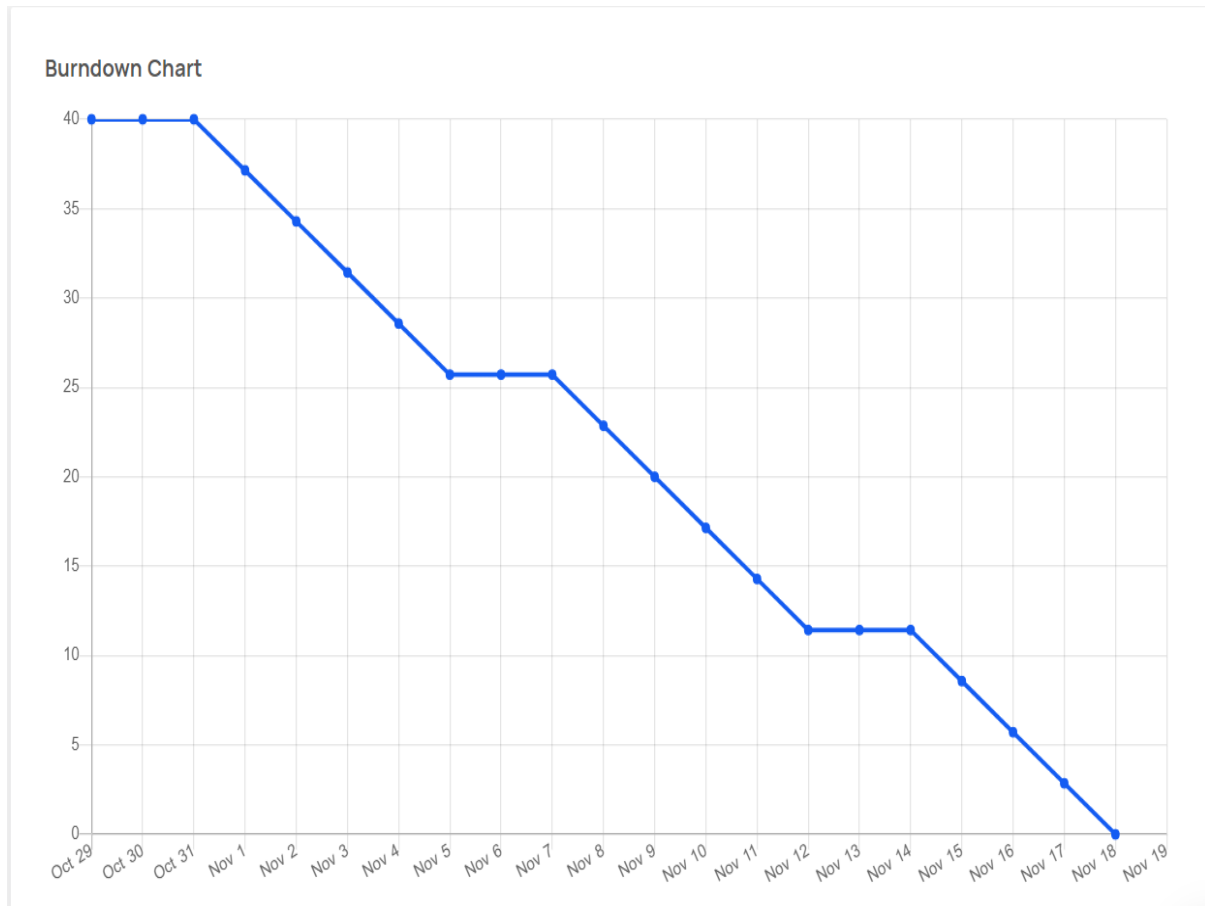
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Login	USN-1	As a user, I can login for the application by entering my mail	1	High	Akash R
Sprint-1		USN-2	As a user, I will login and get confirmation mail once I have registered	2	High	Akash R
Sprint-2		USN-3	As a user, I can login for the application through mobile number	2	Medium	Karthick S
Sprint-3	Dashboard	USN-4	As a user, I need to enter my details	1	High	Karthick S
Sprint-3	Dashboard	USN-5	As a user, I need to provide my Test Details	2	High	Akash R
Sprint-3	services	USN-6	As a admin I need to provide valid result	3	High	Durgesh H
Sprint-3		USN-7	As a admin I need to provide valid /useful suggestions	6	Medium	Chandrakis hore T
Sprint-4	Data Process	USN-8	As a admin need to collect all the details and information.	2	High	Chandrakis hore T
Sprint-4		USN-9	As a admin I need to store all the details and information	3	High	Karthick S
Sprint-4	Login	USN-10	As a admin I need to login and access details of customers	5	High	Chandrakis hore T
Sprint-4	Dashboard	USN-11	As a admin I need to proceed the details with case head	12	High	Durgesh H

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		

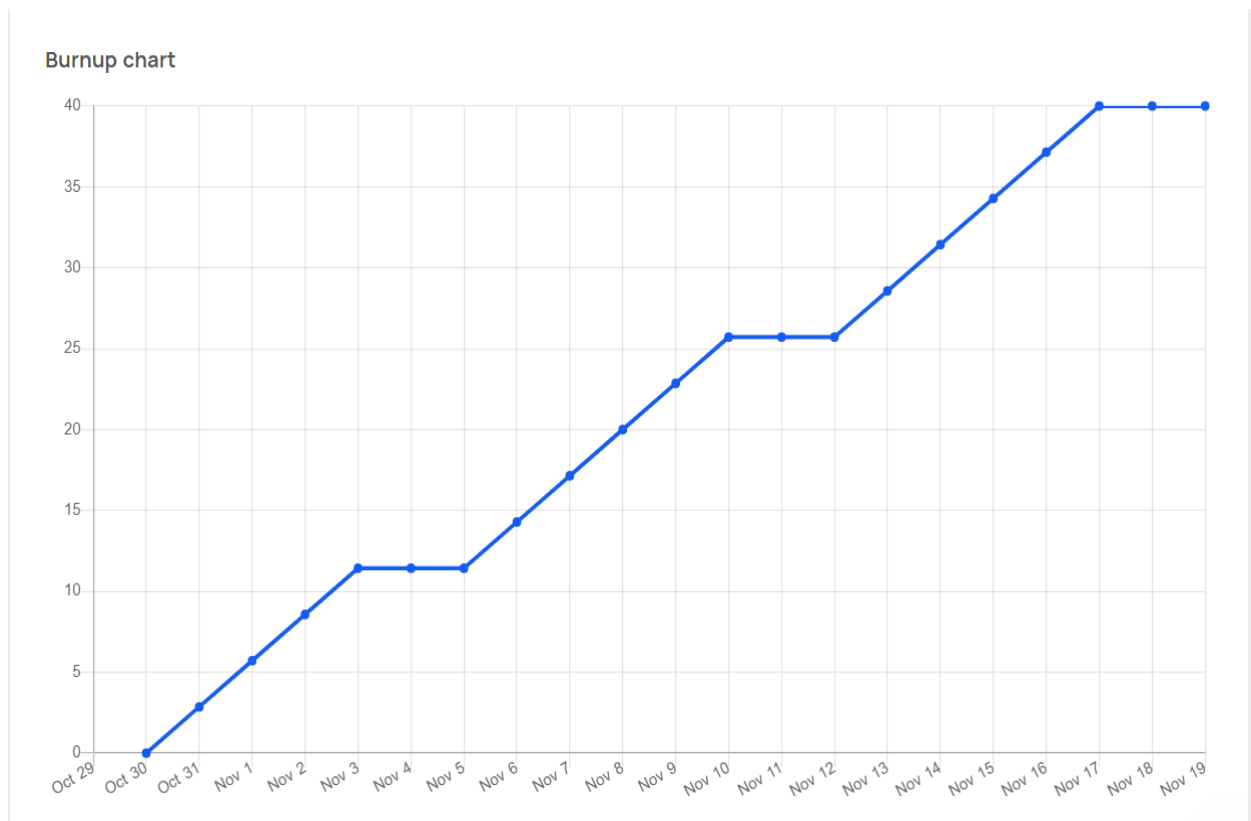
6.3 REPORTS FROM JIRA

Burndown Chart



A burndown chart shows the amount of work that has been completed in an epic 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



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 in a 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 deployment is the process of Random Forest

Classifier is used to train and test the model for detecting the Liver Disease with the 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 in case of regression.

RandomForestClassifier:

```
from sklearn.ensemble import RandomForestClassifier RandomForest =
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))
print(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 predictive model. Decision trees are usually used when doing gradient boosting.

```
# GradientBoostingClassifier:

from sklearn.ensemble import GradientBoostingClassifier
GradientBoost = 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 learning algorithm. 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 decision trees with one level.

```
# AdaBoostClassifier:

from sklearn.ensemble import AdaBoostClassifier
AdaBoost = 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. It allows 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 Python flask.

```
header = {'Content-Type': 'application/json', 'Authorization': 'Bearer '  
+ mtoken}  
  
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 users.

```
@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_Aminotransferas  
e = 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_Aminotransferase,Aspartate_Aminotransferase>Total_Protiens,Albumin,Albumin_and_Globulin_Ratio]])

prediction = model.predict(values)

return render_template('result.html', prediction=prediction if
__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

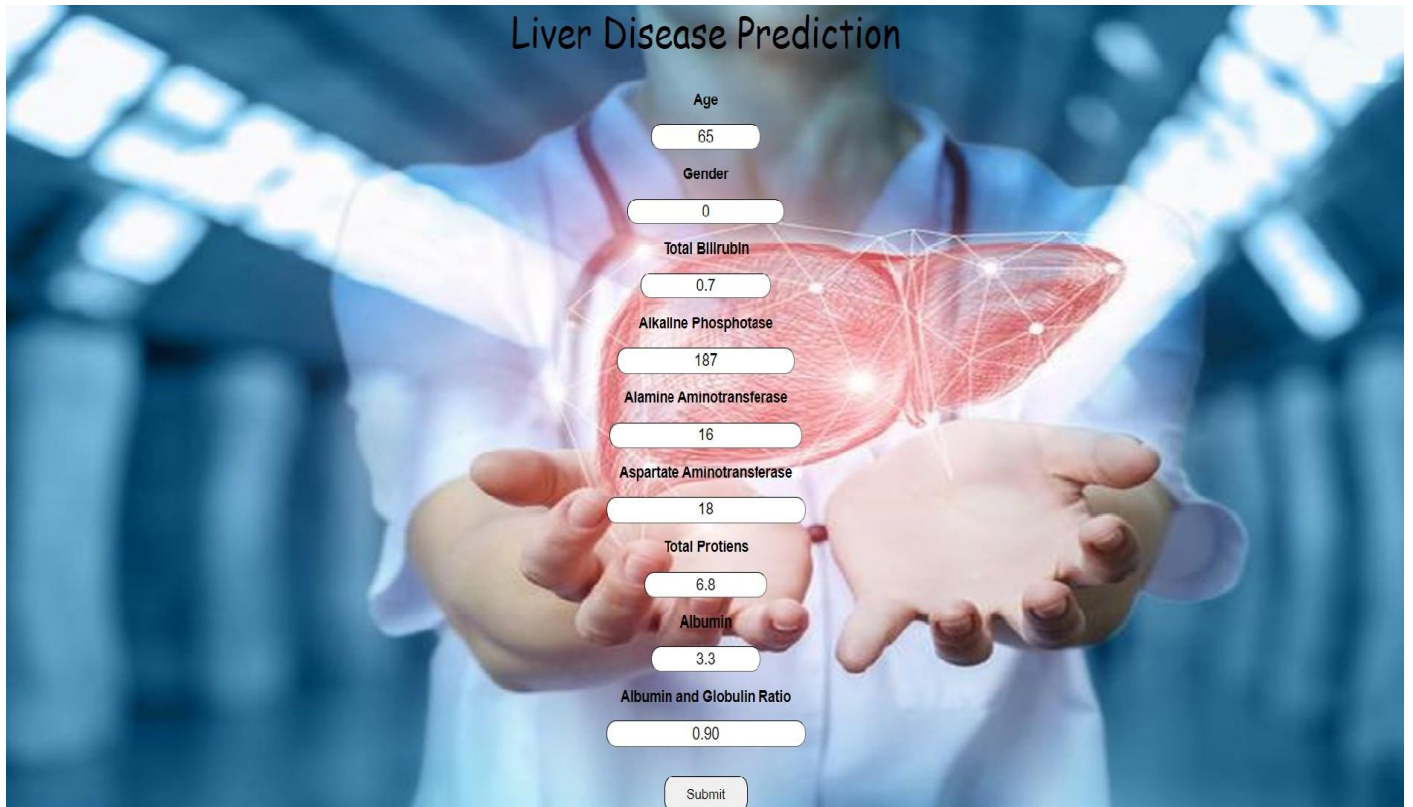
TEST CASE ID	15358	TEST CASE DESCRIPTION	STATISTICAL MACHINE LEARNING APPROACHES TO LIVER DISEASE PREDICTION
--------------	-------	-----------------------	---

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 should be filled
4	Application to train the model	Provide the datasets for model training

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

8.2 USER ACCEPTANCE TESTING



Liver Disease Prediction

Age: 65

Gender: 0

Total Bilirubin: 0.7

Alkaline Phosphatase: 187

Alamine Aminotransferase: 16

Aspartate Aminotransferase: 18

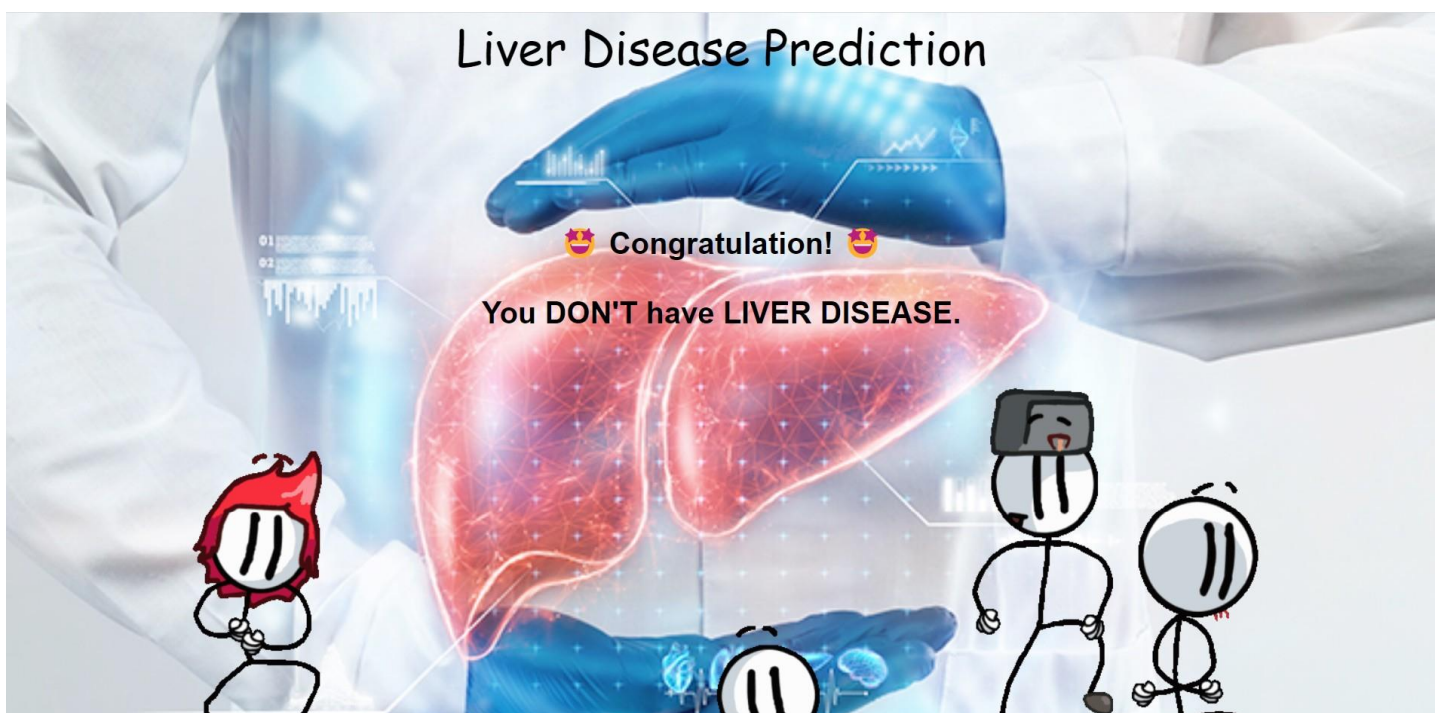
Total Protiens: 6.8

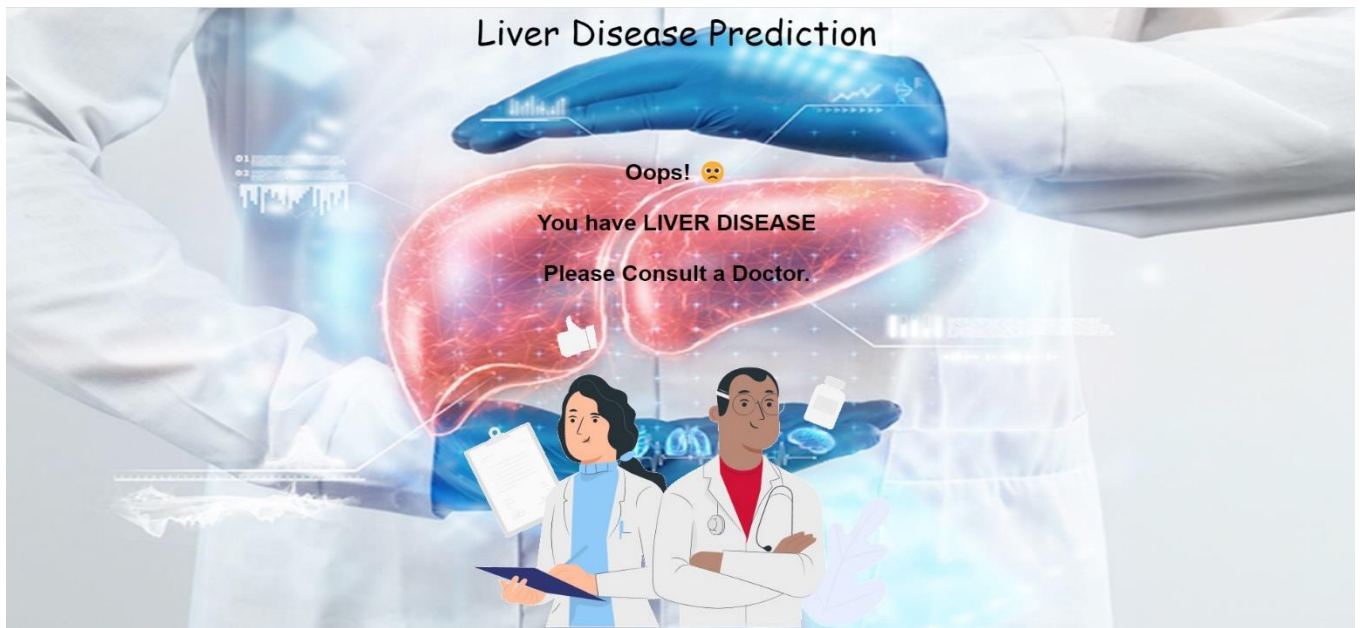
Albumin: 3.3

Albumin and Globulin Ratio: 0.90

Submit

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



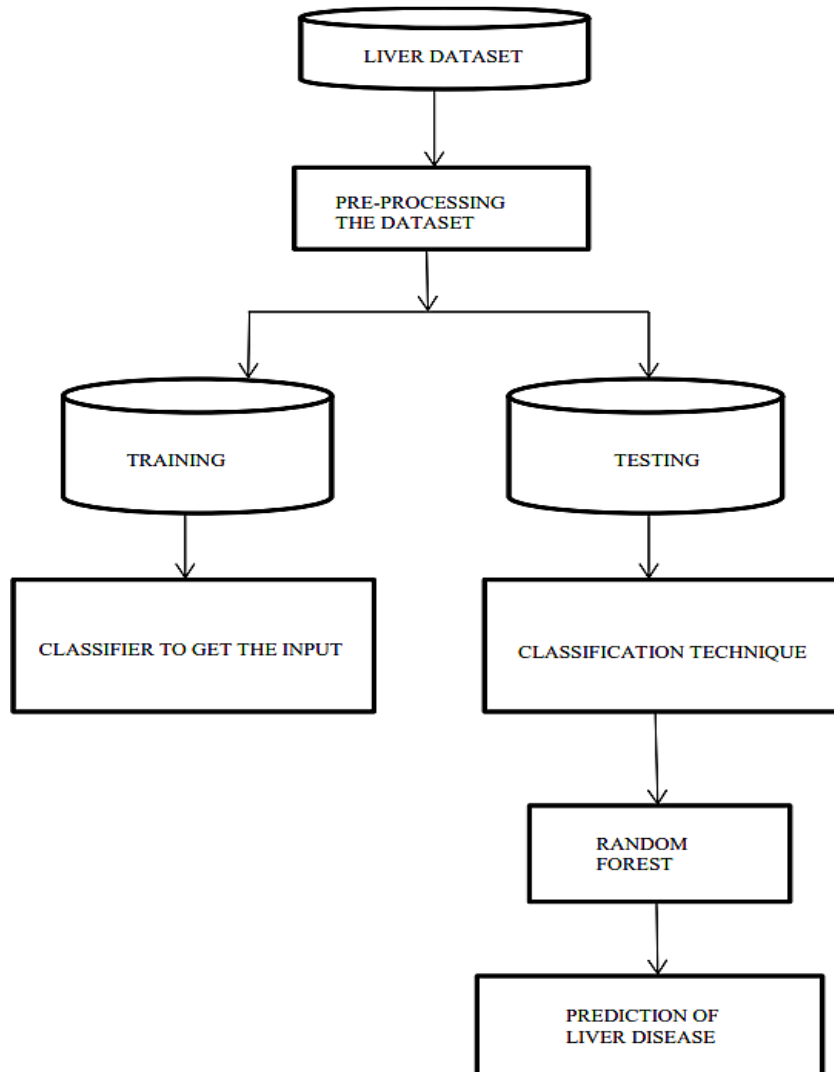


**In this the data which was entered by the user will be analyzed.
This figureshows the person have a liver disease.**

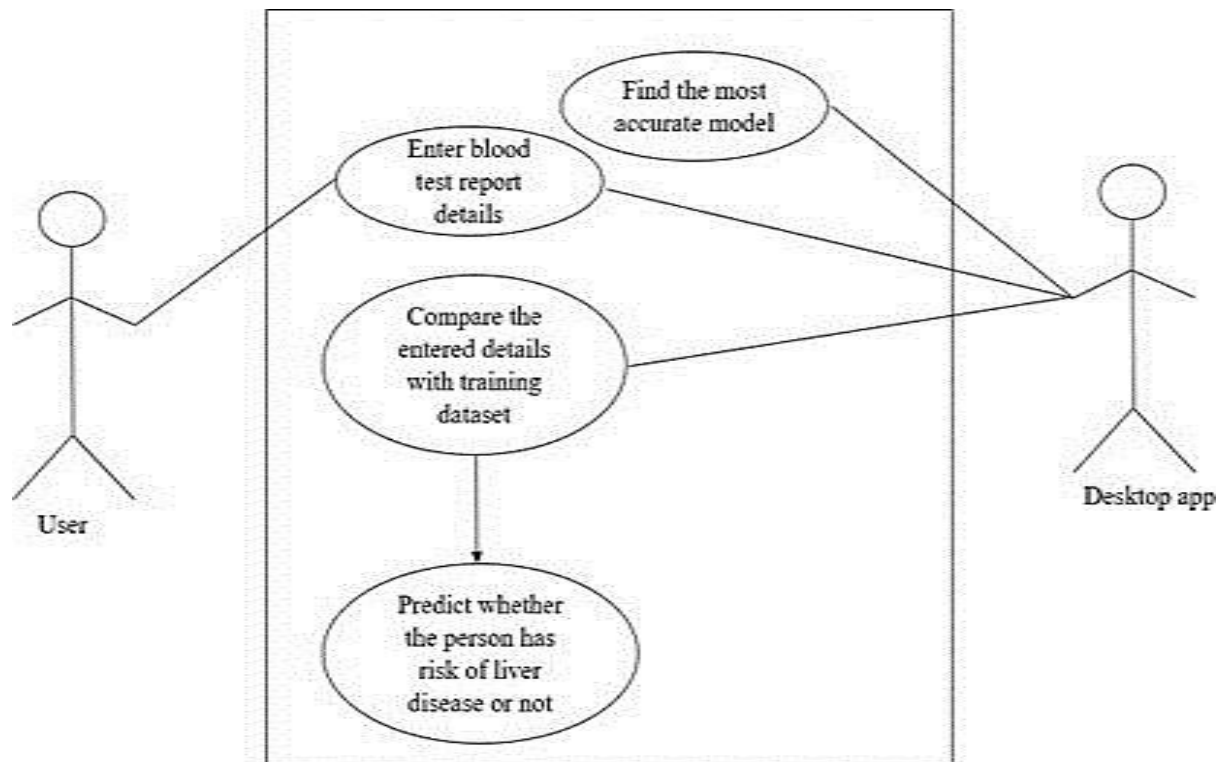
CHAPTER 9

RESULTS

9.1 PERFORMANCE METRICS



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



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 & DISADVANTAGES

ADVANTAGES

In pharmaceutical industries, random forest can be used to identify the potential of a certain medicine or the composition of chemicals required for medicines. Random forest can also be used in hospitals to identify the diseases suffered 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 liver function 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 health nowadays and are joining yoga classes, dance classes; still the sedentary lifestyle and luxuries that are continuously being introduced and enhanced; the problem 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 project's results, one can clearly conclude that we can predict the risk of liver diseases with accuracy of 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)
```

```
ns.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 TestSet

```
from sklearn.model_selection 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 RandomForestClassifier
```

```
from sklearn.neighbors import KNeighborsClassifier
```

#initializing 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.predict(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'])

```

```

    = 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.HTML

```

<!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">Liver Disease
        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="Total
Protiens" 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">Liver
Disease 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 LIVER DISEASE <br><br>Please Consult a Doctor.</span></h1>
```

```

```

```
{% elif prediction==1 %}
```

```
<h1><span class='safe'> 🎉 Congratulation!
```

```
🎉 <br><br>You DON'T have LIVER DISEASE.</span></h1>
```

```

```

```
{% 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 & PROJECT DEMO LINK

GITHUB LINK :

<https://github.com/IBM-EPBL/IBM-Project-17965-1659677477>

PROJECT DEMO LINK :

https://drive.google.com/file/d/1Ul3tnAC13Lpy0Ml4GVVTgMwTrDdxJU9X/view?usp=share_link