A REVIEW OF LIVER PATIENT ANALYSIS METHODS USING MACHINE LEARNING

Submitted in partial fulfillment of requirement for the award of the Degree

Bachelor of Computer Science

In the faculty of Computer Science of Bharathiar University, Coimbatore

Submitted by

TEAM ID: NM2023TMID1760

JUFRIYA HASHLIN. H

JERLIN JEMELA. A

HASINA. B

SASIKALA. A

Under the guidance of

Dr. V.PUNITHAVATHI MCA.,M.Phil., DOA., Ph.D.

Lecturer , Department of Computer Science

DEPARTMENT OF COMPUTER SCIENCE

LRG GOVERNMENT ARTS COLLEGE FOR WOMEN

(Affiliated To Bharathiar University)

TIRUPUR-4

APRIL-2023

LRG GOVERNMENT ARTS COLLEGE

NAAN MUDHALVAN PROJECT WORK

(AFFILIATED TO BHARATHIAR UNIVERSITY)

TIRUPUR-641602

TITLE :A Review of Liver Patient Analysis Methods Using ML

This is to certify that this is a bonafide record of work done by the above students of III B.Sc (CS) Degree **NAAN MUDHALVAN PROJECT** during the year 2022-2023

Submitted for the Naan Mudhalvan project work held on…….20

CLASS MENTOR **HEAD OF DEPARTMENT**

[Dr.V.Punithavathi] **[Dr.R.Parimala]**

TABLE OF CONTENTS

1. INTRODUCTION

A REVIEW OF LIVER PATIENT ANALYSIS

METHODS USING ML

Liver Disease is the leading cause of global death that impacts the massive quantity of humans around the world. This disease is caused by an assortment of elements that harm the liver. For example, obesity, an undiagnosed hepatitis infection, alcohol misuse which is responsible for abnormal nerve function, coughing up or vomiting blood, kidney failure, liver failure, jaundice, liver encephalopathy and there are many more. Diagnosis of liver infection at preliminary stage is important for better treatment. In today's scenario devices like sensors are used for detection of infections. Accurate classification techniques are required for automatic identification of disease samples.This disease diagnosis is very costly and complicated. Therefore, the goal of this work is to evaluate the performance of different Machine Learning algorithms in order to reduce the high cost of chronic liver disease diagnosis by prediction. In this work, we used five algorithms Logistic Regression, Decision Tree,

Support Vector Machine, Naïve Bayes, and Random Forest. The performance of different classification techniques was evaluated on different measurement techniques such as accuracy, precision, recall, and specificity. We found the accuracy 74%, 72%, 72%, 71%, and 57% for SVM,DT,RF,LR and NB. The analysis result shown the SVM achieved the highest accuracy. Moreover, our present study mainly focused on the use of clinical data for liver disease prediction and explores different ways of representing such data through our analysis.

2. PROBLEM SELECTION

Selecting the right problem for liver patient analysis using machine learning can have a significant impact on the success of your project. Here are a few potential problem statements for this type of analysis:

1. Predicting liver disease: Given a dataset of liver patient records, build a machine learning model that can accurately predict whether a patient has liver disease or not. This can be framed as a binary classification problem, where the goal is to distinguish between patients with liver disease and those without.

2. Severity prediction: Given a dataset of liver patient records, build a machine learning model that can predict the severity of the disease. This can be framed as a regression problem, where the goal is to predict a continuous value, such as the patient's liver function score.

3. Treatment recommendation: Given a dataset of liver patient records and their treatment outcomes, build a machine learning model that can recommend the most effective treatment for a new patient. This can be framed as a recommendation problem, where the goal is to suggest a treatment that maximizes the patient's chances of recovery.

4. Risk factor identification: Given a dataset of liver patient records, build a machine learning model that can identify the most significant risk factors for liver disease.

2. IDEATION

Analyzing liver disease patient data using machine learning can provide valuable insights into the factors that contribute to liver disease and help healthcare professionals identify patients at high risk for developing liver disease or its complications. Here are some ideas for analyzing liver patient data using machine learning

• Predicting liver disease: You can build a predictive model using machine learning algorithms that analyze patient data, such as age, gender, family history, blood test results, and lifestyle factors, to predict the likelihood of a patient developing liver disease. This can help healthcare professionals identify patients at high risk for liver disease and provide early intervention.

• Identifying risk factors: You can use machine learning to identify the risk factors that contribute to liver disease. This can be done by analyzing patient data to identify common patterns and trends that may be associated with liver disease, such as alcohol consumption, obesity, or viral infections.

• Early detection of liver disease: You can develop a machine learning model that analyzes patient data to detect early signs of liver disease. This can be done by analyzing blood test results, imaging scans, and other diagnostic tests to identify changes that may indicate the presence of liver disease.

REQUIREMENT ANALYSIS

Data Collection

For this study, the Indian Liver Patient Dataset (ILPD) was selected from the UCI Machine Learning repository. It is a sample of the whole Indian population taken from the area of Andhra Pradesh. There were 583 instances based on ten different biological parameters in the dataset. Based on these criteria, the class value was stated as either yes (416 cases) or no (167 cases), reflecting the liver.

Pre-processing and Feature selection

To normalize the missing values, pre-processing techniques have been introduced. The missing values were replaced by null values along with their instances. Feature selection was followed to classify the appropriate attribute for classification. Using both filter and wrapper approaches, feature selection was carried out. The attributes with more than 70% correlation were initially excluded by correlation analysis from the dataset. The algorithm was implemented to estimate the value of different features in a dataset on the basis of random forests .

Randomization and splitting of dataset

To build classification models, the features selected in the preceding phase were accepted. The dataset was initially randomized to produce an arbitrary sample permutation. Splitting of the dataset into training (70 percent of the dataset) and test (30 percent) sets was followed. The training set consisted of 389 cases and the evaluation set consisted of the remaining 194 cases.

Classification algorithms

Classification algorithm is one of the greatest significant and applicable data mining techniques used to apply in disease prediction. Classification algorithm is the most common in several automatic medical health diagnoses. Many of them show good classification accuracy. Different data mining algorithms like Naïve Bayes, Decision Tree, Logistic Regression, Random forest and Support vectormachine (SVM) were implemented

1. INPUT DESIGN

The input design is the process of entering data to the system. The input design goal is to enter to the computer as accurate as possible. Here inputs are designed effectively so that errors made by the operation are minimized.

The inputs to the system have been designed in such a way that manual forms and the inputs are coordinated where the data elements are

common to the sources document and to the input. The input is acceptable and understandable by the users who are using it.

Input design is the process of converting user-originated inputs to a computer-based format input data are collected and organized into group of similar data. Once identified, appropriate input media are selected for processing.

Input design means the physical and performance requirements of a device that are used as a basis for device design.Input is the raw data that is processed to produce output. During the input design, the developers must consider the input devices such as PC, MICR, OMR, etc

5.1. INPUT DESIGN DESCRIPTION

HOME

In the home page a simple concept of Liver patient is displayed and it shows a predict option to get to the predict page.

PREDICT

In this page it asks for input of certain data that is necessary to predict the Liver disease type it also shows a submit button which redirects it to the submit page.

2. OUTPUT DESIGN

A design output is a drawing or specification or manufacturing instruction. Design outputs describe all the components, parts, and pieces that go into your device. Design outputs describe all assemblies and subassemblies of product.

Output design is the process of converting data into hard copy that is understood by all. The various outputs have been in such a way that they represent the same format that the office and management used to.

Computer output is the most important and direct source of information to the user. Efficient, intelligible output design should improve the systems relationships with the user and help in decision making. A major form of out is the hardcopy from the printer. Output requirements are designed during system analysis.

PROJECT DESIGN

Designing a project to analyze liver patient data using machine learning can be a valuable contribution to healthcare. Here is a general outline of the project design: ➢ Data Collection: Collect relevant data related to liver patients from multiple sources. This may include hospital records, clinical trials, or publicly available datasets. It is essential to ensure that the data is comprehensive and includes various aspects such as demographic information, medical history, lab results, and imaging reports

7. DESCRIPTION OF MODULES

Modules are unit of code written in access basic language.

- HOME
- PREDICT
- SUBMIT

HOME

- In the home page a simple concept of Liver patient is displayed and it shows a predict option to get to the predict page.

PREDICT

- In this page it asks for input of certain data that is necessary to predict the Liver disease type.

SUBMIT

- In this page the result of the predicted Liver disease type is displayed like

8. PROJECT PLANNING PHASE

Definine the project :

Collect relevant data related to liver patients from multiple sources. This may include hospital records, clinical trials, or publicly available datasets. It is essential to ensure that the data is comprehensive and includes various aspects such as demographic information, medical history, lab results, and imaging reports.

.

Gather Data:

Cleaning and pre-processing the data is a critical step in the machine learning pipeline. This may include data cleaning, data normalization, missing data handling, and outlier detection

.**Identify Machine Learning Algorithm:**

After analyzing the data, you will need to identify the appropriate machine learning algorithms that can be used to train the model. This could include supervised learning algorithms such as logistic regression, decision trees, and neural networks.

Train the model:

Once you have chosen an algorithm, you will need to train the model on the data. This involves dividing the data into training and validation sets, and using the training set to optimize the model's parameters.

Evaluate the model:

After training the model, you will need to evaluate its performance on the validation set. This will give you an idea of how well the model will perform on new data.

Deploy the model:

Once you have a model that performs well on the validation set, you can deploy it in a clinical setting. This may involve integrating it into an electronic health record system, or creating a standalone application.

9. PROJECT DEVELOPMENT PHASE

The development phase for a liver patient analysis project can be broken down into several key steps:

1. **Defining the problem**:

   The first step us to clearly define the problem you are trying to solve.

   the key factors that contribute to liver disease and developing a model to predict the likelihood of a patient developing The first step is to clearly define the problem you are trying to solve. For a liver patient analysis project, this may involve identifying

2. **Gathering data:**

   The next step is to gather relevant data for your project. This may include medical records, lab test results, patient demographics, and other relevant information.

3. **Data preprocessing**:

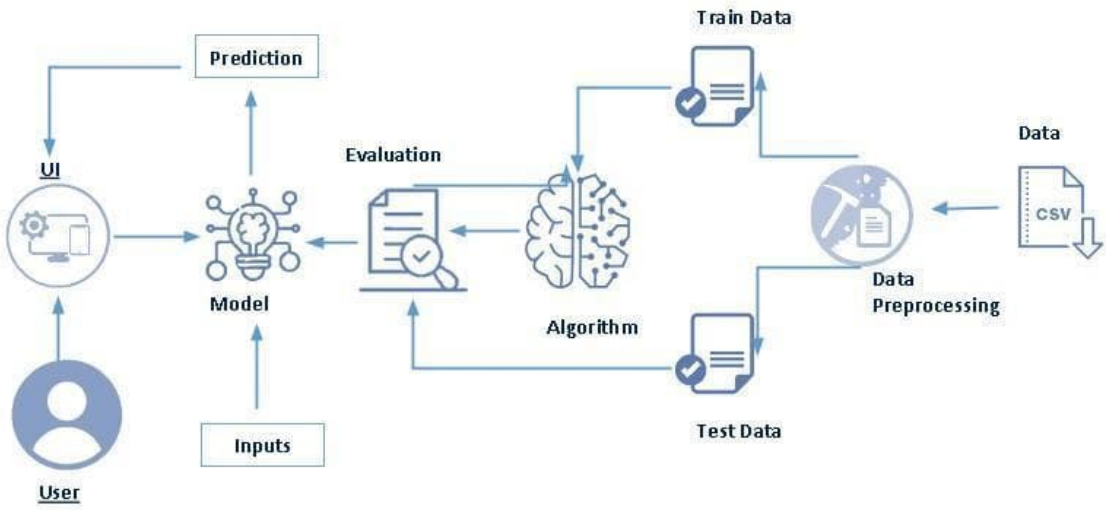
   Once you have gathered the data, you will need to preprocess it to make it suitable for analysis. This may involve cleaning the data, removing

10. CONCLUSION

Machine learning plays a very deciding role in diseases of the liver and heart are becoming increasingly widespread. These are solely going to get worse in the future, thanks to ongoing technological improvements. Despite the truth that people are becoming more health-conscious and enrolling in yoga and dancing classes, the sedentary lifestyle and facilities that are continuously being delivered and improved will proceed to be an issue. As a result, in this situation, our project will be rather recommended to society.

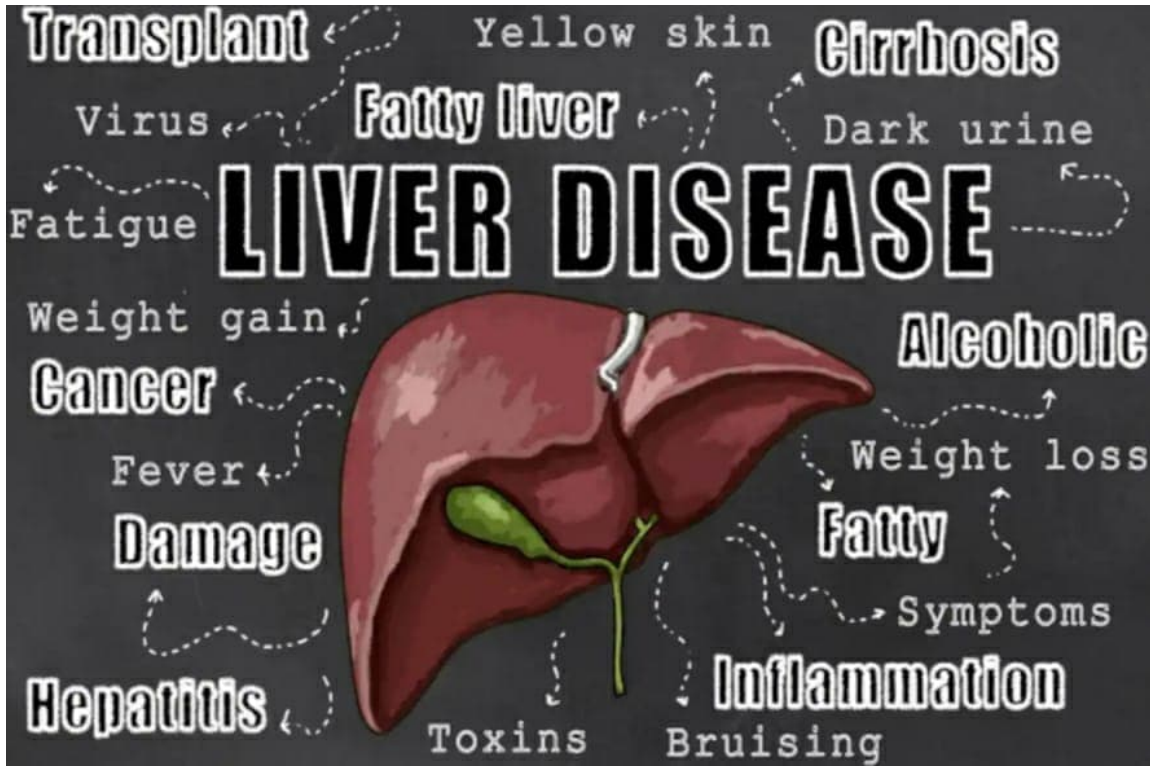
APPENDICES

A.TECHICAL FLOW



B.SAMPLE INPUT

HOME

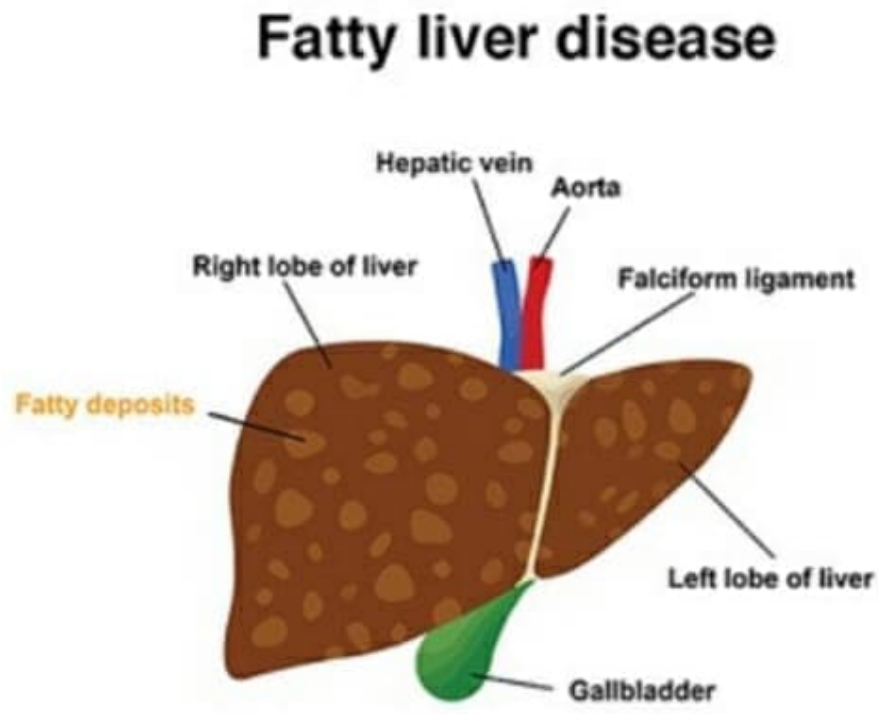


PREDICT


Liver Patient Prediction
Age:
Gender:
Enter 0 as male, 1 as female
Total_Bilirubin:
Direct_Bilirubin:
Alkaline_Phosphotase:
Alamine_Aminotransferase:
Aspartate_Aminotransferase:
Total_Protiens:
Albumin:
Albumin_and_Globulin_Ratio:
Predict

C.SAMPLE OUTPUT

SUBMIT



D.SAMPLE CODING

```
#import the dataset from specified location
data = pd.read_csv('E:/Datascience/Datasets/indian_liver_patient.csv')
```

```
# showing the data from top 5
data.head()
```

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferas
0	65	Female	0.7	0.1	187	16	18
1	62	Male	10.9	5.5	699	64	100
2	62	Male	7.3	4.1	490	60	68
3	58	Male	1.0	0.4	182	14	20
4	72	Male	3.9	2.0	195	27	59

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 583 entries, 0 to 582
Data columns (total 11 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Age                         583 non-null    int64
 1   Gender                      583 non-null    object
 2   Total_Bilirubin             583 non-null    float64
 3   Direct_Bilirubin            583 non-null    float64
 4   Alkaline_Phosphotase        583 non-null    int64
 5   Alamine_Aminotransferase    583 non-null    int64
 6   Aspartate_Aminotransferase  583 non-null    int64
 7   Total_Protiens              583 non-null    float64
 8   Albumin                     583 non-null    float64
 9   Albumin_and_Globulin_Ratio  579 non-null    float64
 10  Dataset                     583 non-null    int64
dtypes: float64(5), int64(5), object(1)
memory usage: 50.2+ KB
```

```
data.isnull().any()
```

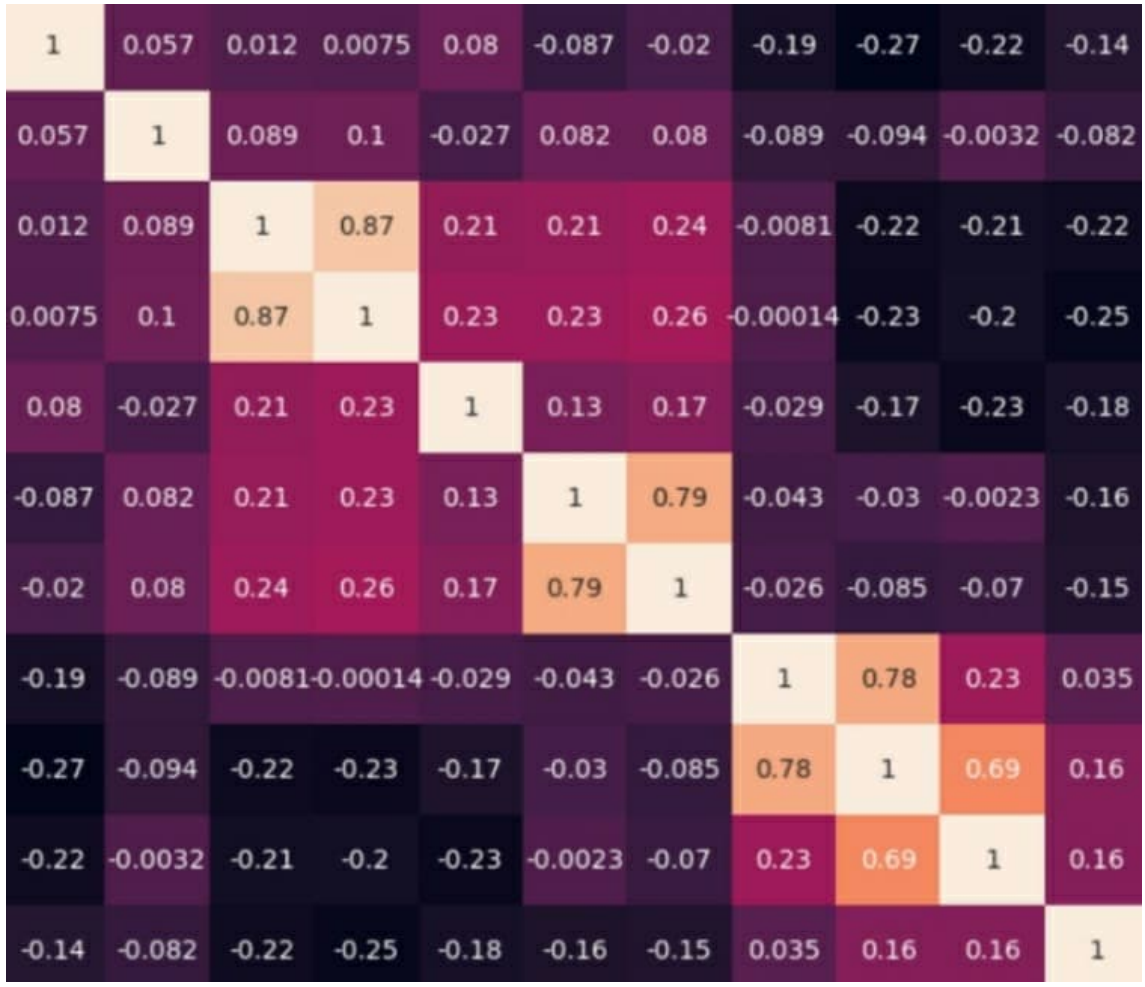
```
Age                         False
Gender                      False
Total_Bilirubin             False
Direct_Bilirubin            False
Alkaline_Phosphotase        False
Alamine_Aminotransferase    False
Aspartate_Aminotransferase  False
Total_Protiens              False
Albumin                     False
Albumin_and_Globulin_Ratio   True
Dataset                     False
dtype: bool
```

```
data.isnull().sum()
```

```
Age                         0
Gender                      0
Total_Bilirubin             0
Direct_Bilirubin            0
Alkaline_Phosphotase        0
Alamine_Aminotransferase    0
Aspartate_Aminotransferase  0
Total_Protiens              0
Albumin                     0
Albumin_and_Globulin_Ratio  4
Dataset                     0
dtype: int64
```

```
#checking for the missing data after cleaning data
data['Albumin_and_Globulin_Ratio'] = data.fillna(data['Albumin_and_Globulin_Ratio'].mode()[0])
data.isnull().sum()
```

```
Age                         0
Gender                      0
Total_Bilirubin             0
Direct_Bilirubin            0
Alkaline_Phosphotase        0
Alamine_Aminotransferase    0
Aspartate_Aminotransferase  0
Total_Protiens              0
Albumin                     0
Albumin_and_Globulin_Ratio  0
Dataset                     0
dtype: int64
```



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

```
app=Flask(__name__) # our flask app

@app.route('/') # rendering the html template
def home():
    return render_template('home.html')
@app.route('/predict') # rendering the html template
def index() :
    return render_template("index.html")
```

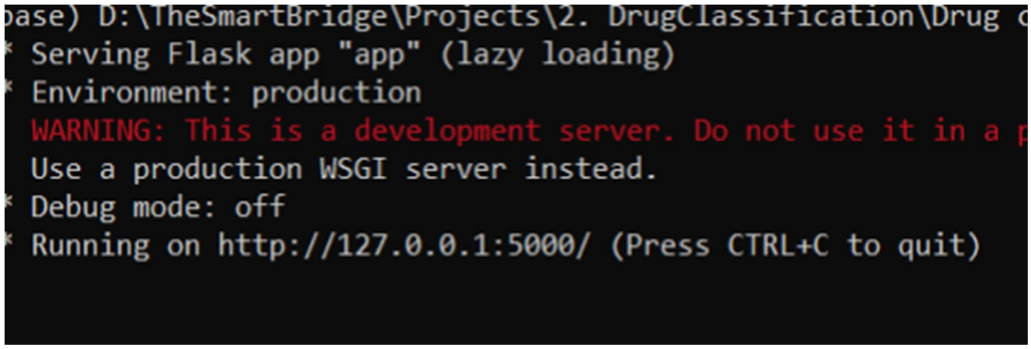
```
@app.route('/data_predict', methods=['POST']) # route for our prediction
def predict():
    age = request.form['age'] # requesting for age data
    gender = request.form['gender'] # requesting for gender data
    tb = request.form['tb'] # requesting for Total_Bilirubin data
    db = request.form['db'] # requesting for Direct_Bilirubin data
    ap = request.form['ap'] # requesting for Alkaline_Phosphotase data
    aa1 = request.form['aa1'] # requesting for Alamine_Aminotransferase data
    aa2 = request.form['aa2'] # requesting for Aspartate_Aminotransferase data
    tp = request.form['tp'] # requesting for Total_Protiens data
    a = request.form['a'] # requesting for Albumin data
    agr = request.form['agr'] # requesting for Albumin_and_Globulin_Ratio data

    # coverting data into float format
    data = [[float(age), float(gender), float(tb), float(db), float(ap), float(aa1), float(aa2), float(tp),

    # loading model which we saved
    model = pickle.load(open('liver_analysis.pkl', 'rb'))

    prediction= model.predict(data)[0]
    if (prediction == 1):
        return render_template('noChance.html', prediction='You have a liver desease problem, You must and
    else:
        return render_template('chance.html', prediction='You dont have a liver desease problem')

if __name__ == '__main__':
    app.run()
```



Liver Patient Prediction

You have a liver desease problem, You must and should consult a doctor. Take care