Bharati Vidyapeeth's College of Engineering for Women

Liver Disease Diagnosis Using ML
Domain of Idea: ML

Presented By:
Anushka Sawant(B150344232)
Manasi Khopade(B150344235)
Pratiksha Kawale(B150344228)
Yashika Palan(B150344243)

Introduction

- The liver is well-thought-out to be one of the central organs in any living body.
- It perform fundamental functions such as processing leftover products, generating enzymes, and eliminating exhausted tissues or cells.
- We can stay alive merely a couple of days if our liver shuts down.
- Fortunately, the liver can continue its role even when up to 75% of it is contaminated or removed.
- According to World Health Organization (WHO) and World Gastroenterology Organization (WGO), 35 million individuals pass away due to chronic diseases, and liver failure is one of the apprehensive disease stated.
  -

Problem Statement

- To develop a web based model for liver disease diagnosis using ML

Objectives

- To present, a Machine Learning Web based system for liver disease diagnosis and risk prediction.
- To use classification algorithms to identify the liver patients from healthy individuals.

Attributes in Dataset	Description	Normal Range
Age	A numeric value having range [4 -90] In the year	
Gender	Having two nominal value “male” or “female”	
TB (Total Bilirubin)	A numeric value having range [0.4-75]	0.1-1.2 mg/dL
DB (Direct Bilirubin)	Numeric value having range [0.1-19.7]	0-0.4 mg/dL
Alkphos (Alkaline Phosphatase)	A numeric value having range [63-2110]	44-147 IU/L
Sgpt (Alanine Aminotransferase)	Numeric value having range [10-2000]	7 to 56 IU/L of serum
Sgot (Aspartate aminotransferase)	A numeric value having range [10-4929]	5-40 IU/L of serum
TP (Total Proteins)	A numeric value having range [2.7 - 9.6]	6 and 8.3 g/dL (grams of deciliter)
ALB (Albumin)	Numeric value having range [0.9-5.5]	3.5 to 5.5 g/dL
Albumin and Globulin Ratio (A/G Ratio)	A numeric value having range [0.3 - 2.8]	0.8-2.0
Class	Having the class value “1” represents Liver Disease present and “2” represent Liver Disease not present.	

The results of using different classifiers on Liver Patient Diseases Dataset

- A model for liver prophecy is proposed.
- Evaluated, and Validated to test and compare results of five various ML classification algorithms.
- Which includes SVM, KNN, Neural Network, Decision Tree and RF.
- And as the results revealed that RF is best suitable classifier in the environment related to liver prediction.

Accuracy of all the models

Model Name	Test Accuracy	Train Accuracy
Logistic Regression	60	70
SVC	80	50
Decision Tree	72	65
KNN	68	60
RF	87	70

Logistic Regression

- Logistic Regression-

  The outcome variable is binary,and the purpose of the analysis is to assess the effects of multiple explanatory variables,which can be numeric and on categorial on the outcome variable.

- E.g. The presence of the absence of the symptom,presence absence of disease.

SVM

- A classification problem for both linear and non linear data.
- It uses a nonlinear mapping to transform the original training data into a higher dimension.
- With the new dimension,it searches for the linear optimal separating hyperplane(i.e., “decision boundary”)
- With an appropriate non-linear mapping to a sufficiently high dimension,data from two classes can always be separated by a hyperplane.

Decision Tree

- A decision tree learning is a method for approximating discrete-valued target functions, in which the learned function is represented by a decision tree.
- Decision tree algorithm is a popular method for inductive inference.
- Decision tree algorithm induces concepts from examples.

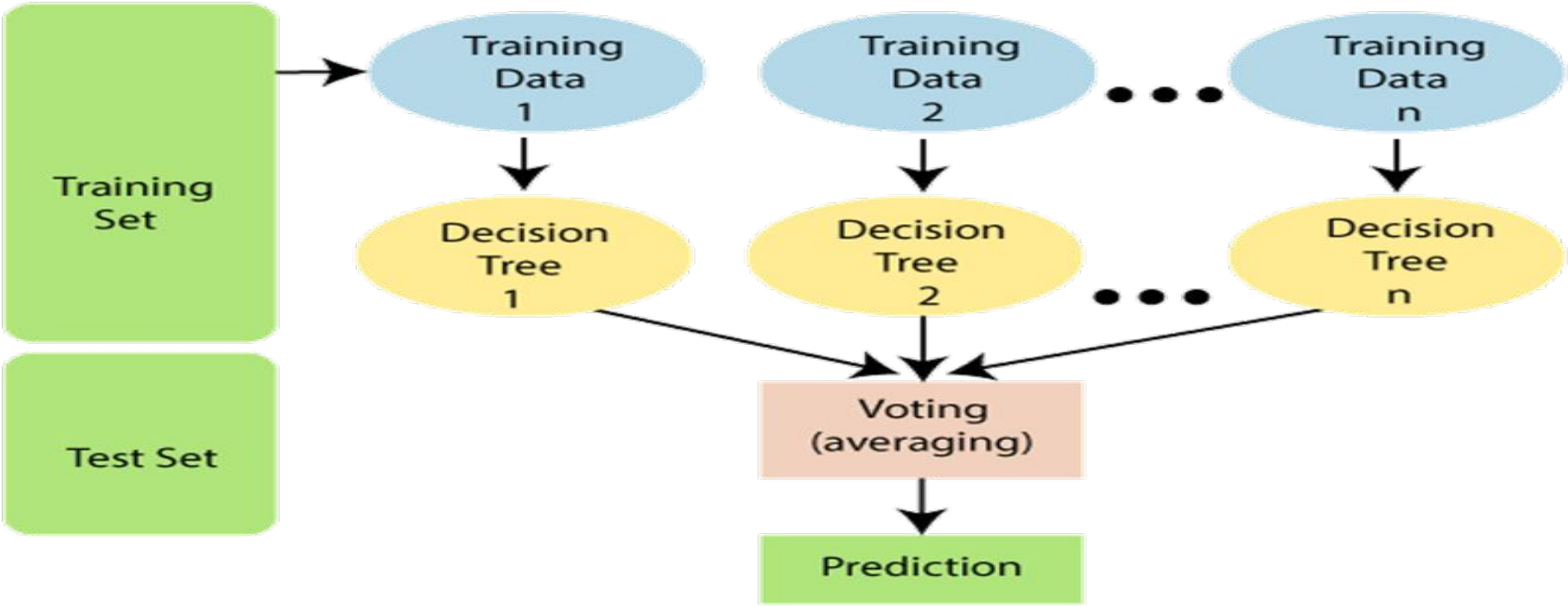
KNN

- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
- K-NN is a non-parametric algorithm (An instance based learning method)

RANDOM FOREST

- An ensemble classifier using many decision tree models.
- Can be used for classification or regression.
- Accuracy and variable importance information is provided with the result.
- Implementation Steps are given below:
    - Data Pre-processing step
    - Fitting the random forest algorithm to the Training set
    - Predicting the test result
    - Test accuracy of the result (Creation of Confusion matrix)
    - Visualizing the test set result

Working of Random Forest



Working Steps

The Working process can be explained in the below steps and diagram:

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

Step-3: Choose the number N for decision trees that you want to build.

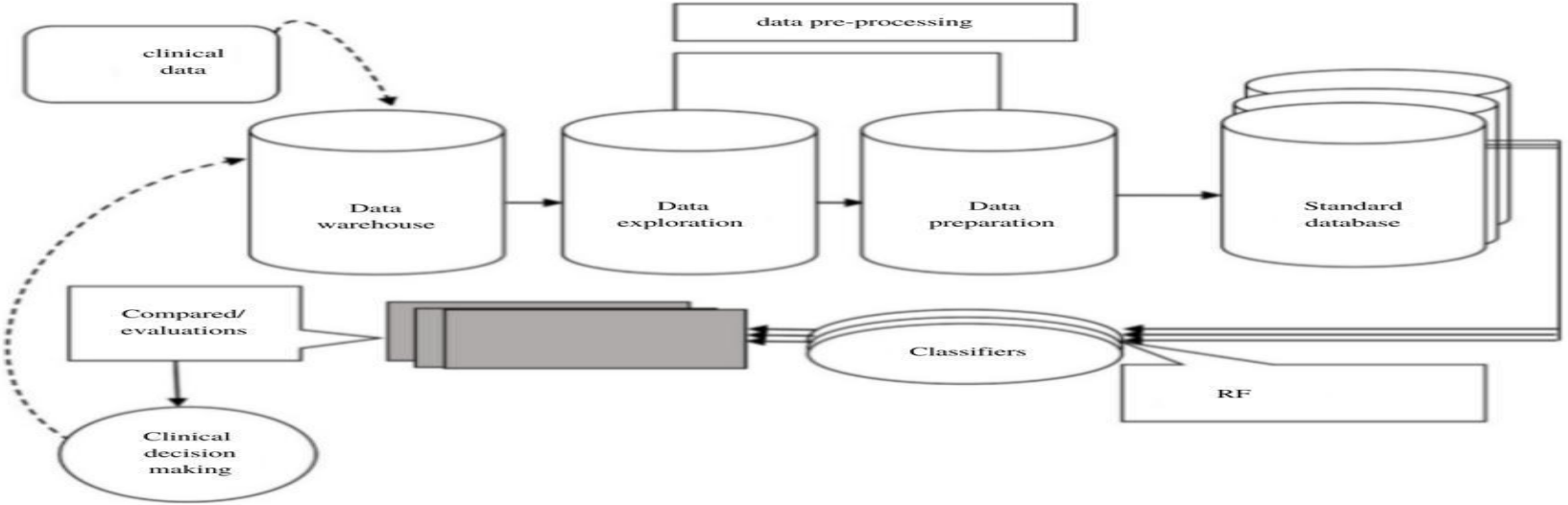
Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

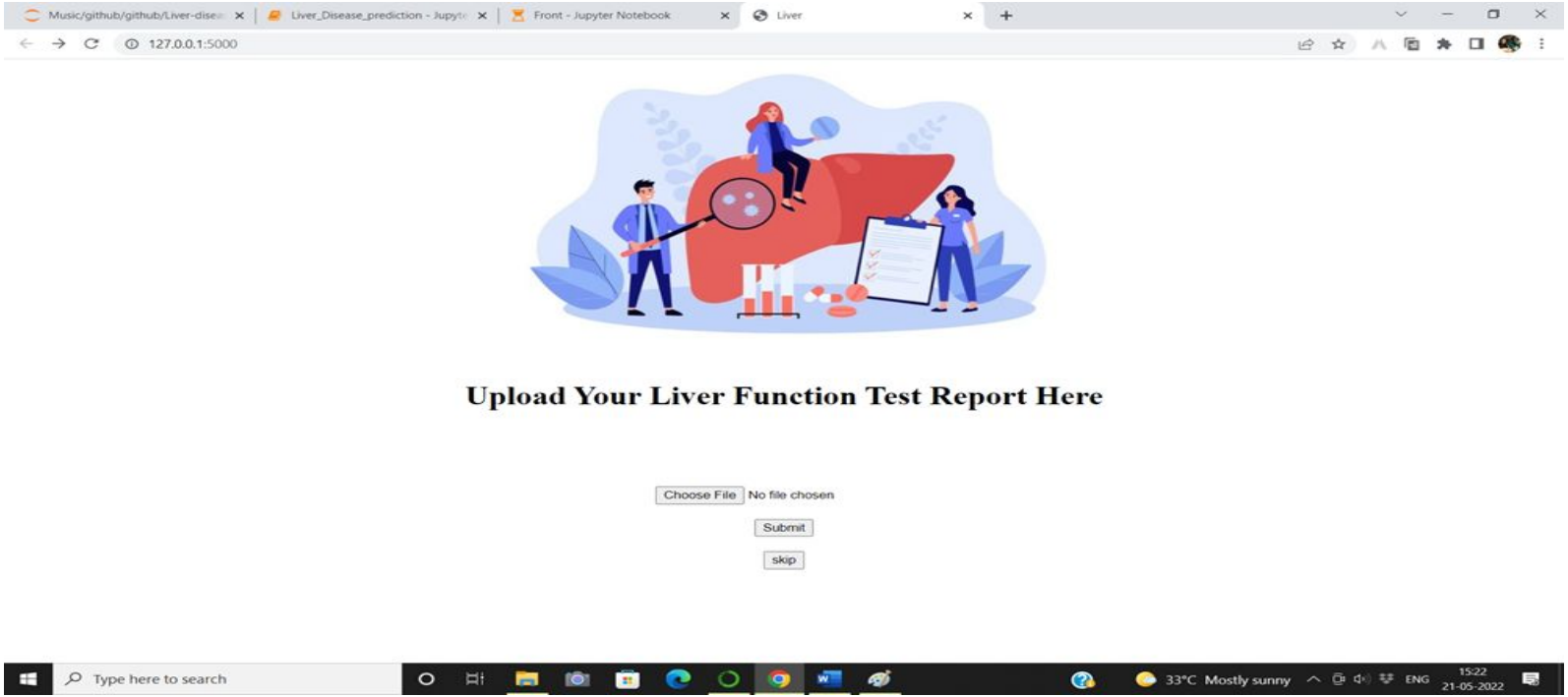
PYTHON-TESSERACT

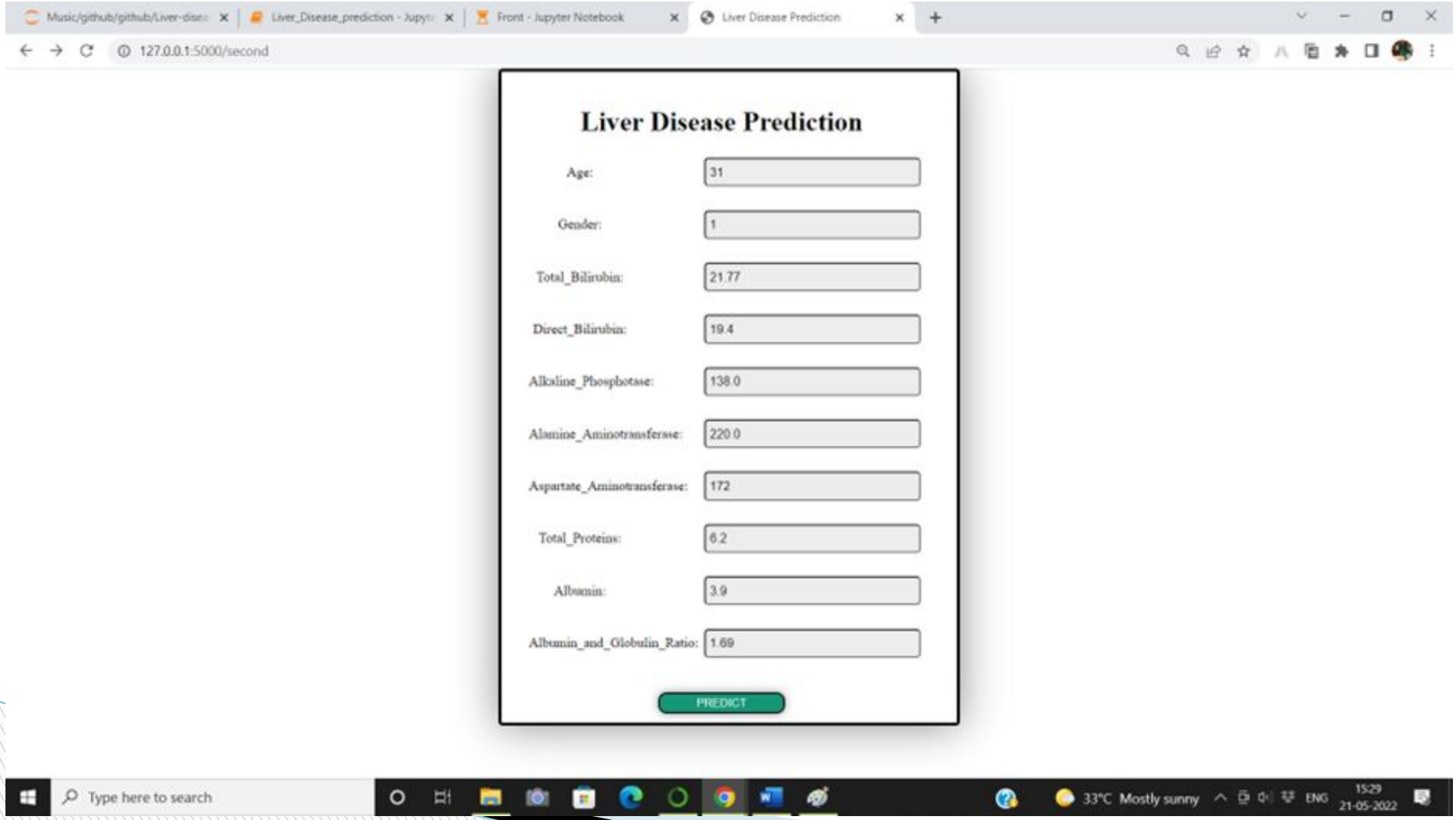
- Python-tesseract is an optical character recognition (OCR) tool for python.
- It will recognize and "read" the text embedded in images.
- Python-tesseract is a wrapper for Google's Tesseract-OCR Engine.
- It can read all image types supported by the Pillow and Leptonica imaging libraries, including jpeg, png, gif.

SYSTEM ARCHITECTURE



DEMO


Front - Jupyter Notebook
Liver
127.0.0.1:5000
Upload Your Liver Function Test Report Here
Choose File
No file chosen
Submit
skip
Type here to search
33°C Mostly sunny
ENG
15:22
21-05-2022


Liver_Disease_prediction - Jupyt
Front - Jupyter Notebook
Liver Disease Prediction
127.0.0.1:5000/second
Liver Disease Prediction
Age: 31
Gender: 1
Total_Bilirubin: 21.77
Direct_Bilirubin: 19.4
Alkaline_Phosphotase: 138.0
Alamine_Aminotransferase: 220.0
Aspartate_Aminotransferase: 172
Total_Proteins: 6.2
Albumin: 3.9
Albumin_and_Globulin_Ratio: 1.69
PREDICT
Type here to search
33°C Mostly sunny
ENG
15:29
21-05-2022

Conclusion:

- Although people are becoming more conscious of health nowadays. 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 and doctors as well.