LIVER DISEASE PREDICTION USING MACHINE LEARNING

E Snayan Yadav[1], Bachepalli Sanhitha[2], and M Krishna Prasanna[3,4]Department of Information Technology,
B V Raju Institute of Technology, Narsapur, Telangana, India
[1]19211A1229@bvrit.ac.in [2]19211A1213@bvrit.ac.in
K r i s h n a p r a s a n n a . m @ b v r i t . a c . i n

Abstract

Liver diseases are becoming one of the most fatal diseases in several countries. Patients with Liver disease have been continuously increasing because of excessive consumption of alcohol, inhale of harmful gases, intake of contaminated food, pickles and drugs. The risk of liver disease was predicted using various machine learning algorithms. The final output is predicted based on the most accurate machine learning algorithm.

Based on the accurate model we designed a system which asks a person to enter the details of his/her blood test report. Then the system uses the most accurate model which is trained to predict, whether a person has risk of liver disease or not. The programming language which was used is python and machine learning Sklearn was used to build the model using classification algorithms like Logical regression, SVM.

Keywords: Liver Disease, SVM, Logistic Regression,

Introduction:

A. Background Information:

Liver diseases cause millions of deaths every year. Problems with liver patients are not easily discovered in an early stage as it will be functioning normally even when it is partially damaged. Liver failures are at high rate of risk among Indians. It is expected that by 2025 India may become the World Capital for Liver Diseases. So Early diagnosis of liver disease is very important to prevent from serious damage. Using ML, we can predict the liver disease in advance by applying different machine learning algorithms.

B. Objective

Liver is the largest internal organ in the human body, it is essential for digesting food and releasing the toxic element of the body and plays a major role in metabolism and serving several vital functions. The motivation behind this study is that liver disease has become a common disease around the world. The death rate due to the disease is becoming alarming. Early detection of the disease may reduce the complication of the disease misfortune on patients. The ease of use of inventive technologies such as the one anticipated in this research may help in alleviating the troubles of holdup in the uncovering and treatment of liver cirrhosis. Also, the data mining tools can be used to assist physicians in predicting and diagnosing the disease to enhance necessary treatment. One more significant drive behind this study is to advance on the works of previous researchers who make their own contributions in this particular field of study.

C. Dataset

The system being proposed here uses concept of machine learning, and the models are first trained, then tested. Finally the most accurate model will predict the final result. At first, the system asks you to enter your details including age, gender, total Bilirubin, direct Bilirubin, Alkaline Phosphatase, Alamine Amino transferase, Aspartate Aamino transferase, total proteins, albumin, A/ G ratio. Values of above parameters mentioned here, can be known by blood test report of the After taking these inputs from the user, the system compares the data input with the training dataset of most accurate model and then predicts the result accordingly as risk or no risk. The results are evaluated in terms of accuracy and confusion matrix. user

Literature Survey :

Logistic Regression:

Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist. In regression analysis, logistic regression [1] (or logit regression) is estimating the parameters of a logistic model (a form of binary regression). Mathematically, a binary logistic model has a dependent variable with two possible values, such as pass/fail which is represented by an indicator variable, where the two values are labeled "0" and "1". In the logistic model, the log odds (the logarithm of the odds) for the value labeled "1" is a linear combination of one or more independent variables.24 The corresponding probability of the value labeled "1" can vary between 0 (certainly the value "0") and 1 (certainly the value "1"), hence the labeling; the function that converts log odds to probability is the logistic function, hence the name. The unit of measurement for the log-odds scale is called a logit, from logistic unit, hence the alternative names. Analogous models with a different sigmoid function instead of the logistic function can also be used, such as the prohibit model; the defining characteristic of the logistic model is that increasing one of the independent variables multiplicatively scales the odds of the given outcome at a constant rate, with each independent variable having its own parameter; for a binary dependent variable this generalizes the odds ratio.

In a binary logistic regression model, the dependent variable has two levels (categorical). Outputs with more than two values are modeled by multinomial logistic regression and, if the multiple categories are ordered, by ordinal logistic regression (for example the proportional odds ordinal logistic model). The logistic regression model itself simply models probability of output in terms of input and does not perform statistical classification (it is not a classifier), though it can be used to make a classifier, for instance by choosing a cutoff value and classifying inputs with probability greater than the cutoff as one class, below the cutoff as the other; this is a common way to make a binary classifier.

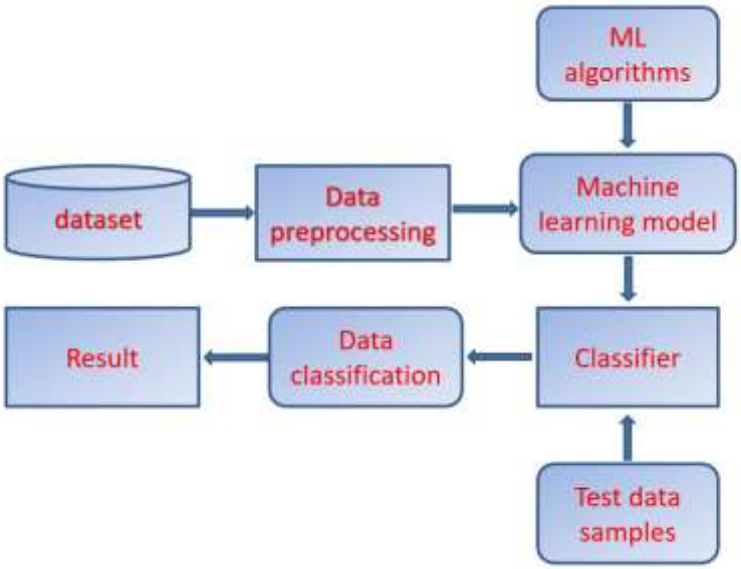
Steps in Logistic Regression: We will follow these steps in order to successfully implement the Logistic Regression using Python. The following are the stages:

- The step of data pre-processing
- The fitting of logistic regression to the training set
- The step of predicting the test result
- The step of testing the accuracy of the result (the creation of the confusionmatrix)
- The step of visually representing the result of the test set

Support Vector Machine :

SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two-dimensional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side. Suppose you are given plot of two label classes. It fairly separates the two classes. Any point that is left of line falls into black circle class and on right falls into blue square class. Separation of classes. That’s what SVM does. It finds out a line/ hyper-plane

D. Architecture:



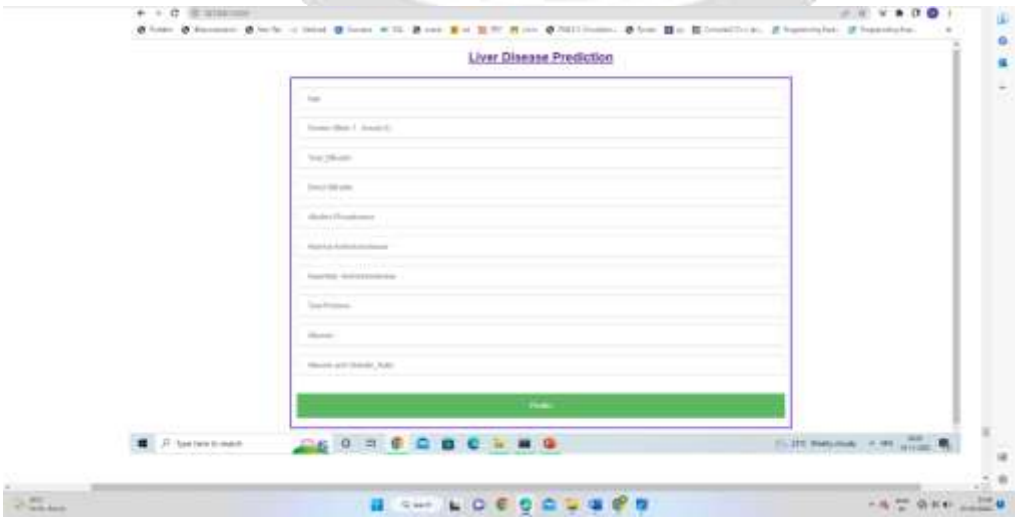
Description of Modules:

Modules present in this Liver disease prediction system project are:

- User Module
- Prediction Module

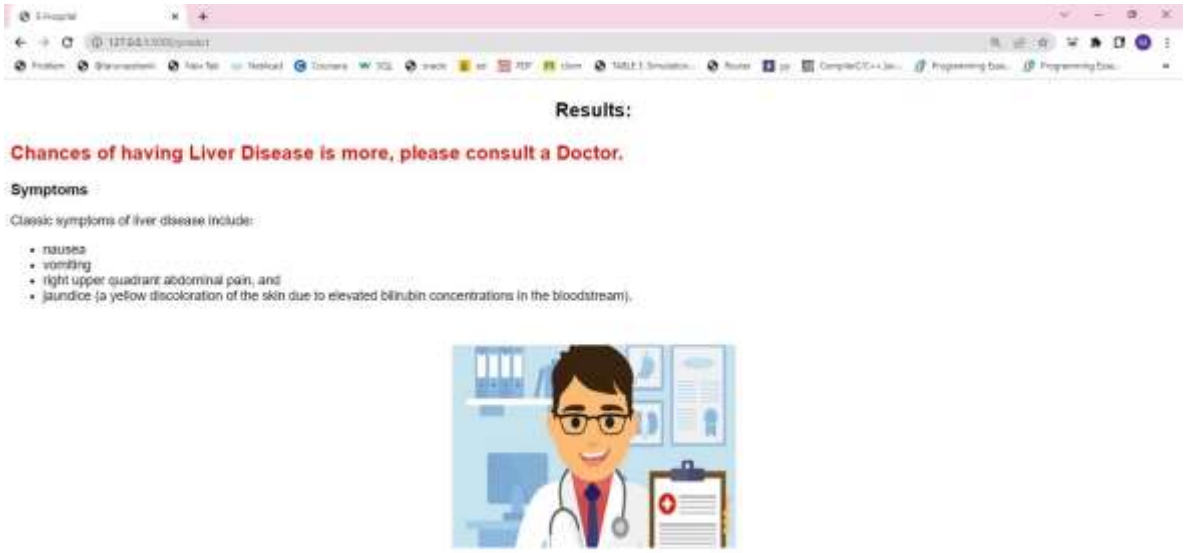
1. **User module:** Takes the input from the users the features/attributes considered for predicting the disease
2. User can be anyone visiting the application. The user should open the web application. They need to enter the required details from his/her blood test report asked by the application. They need to submit those details and can view the results

Prediction module: Responsible for building the classification model that is used for prediction of disease

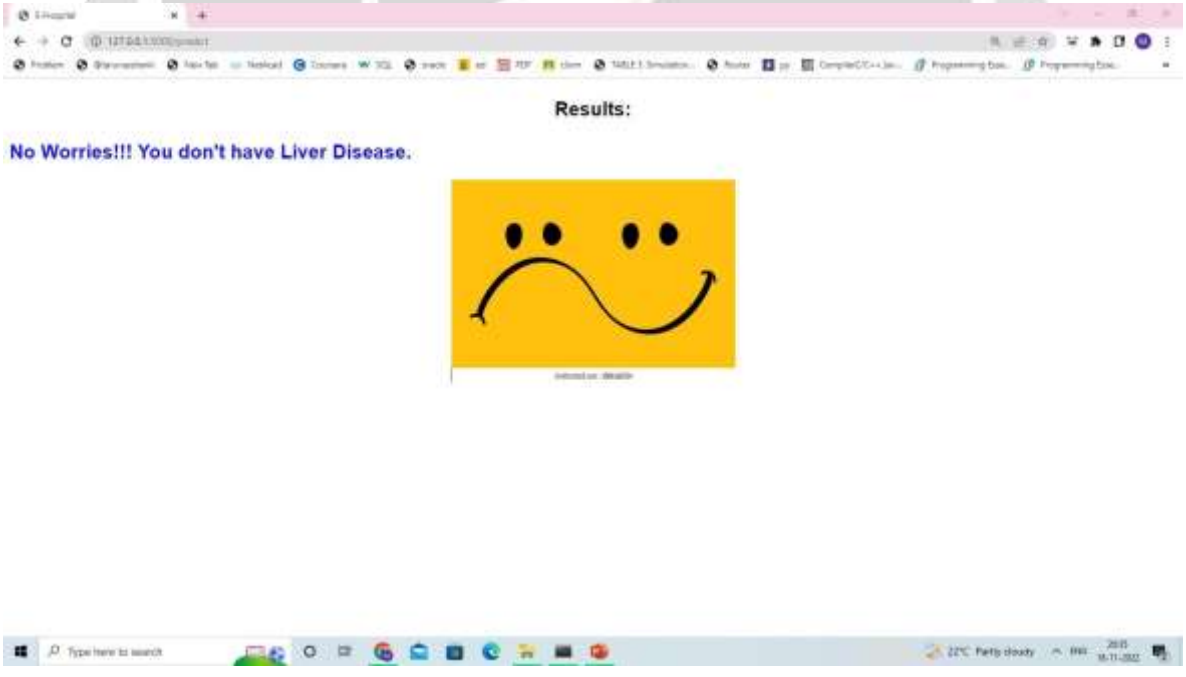


UI of Liver Disease Predictor :

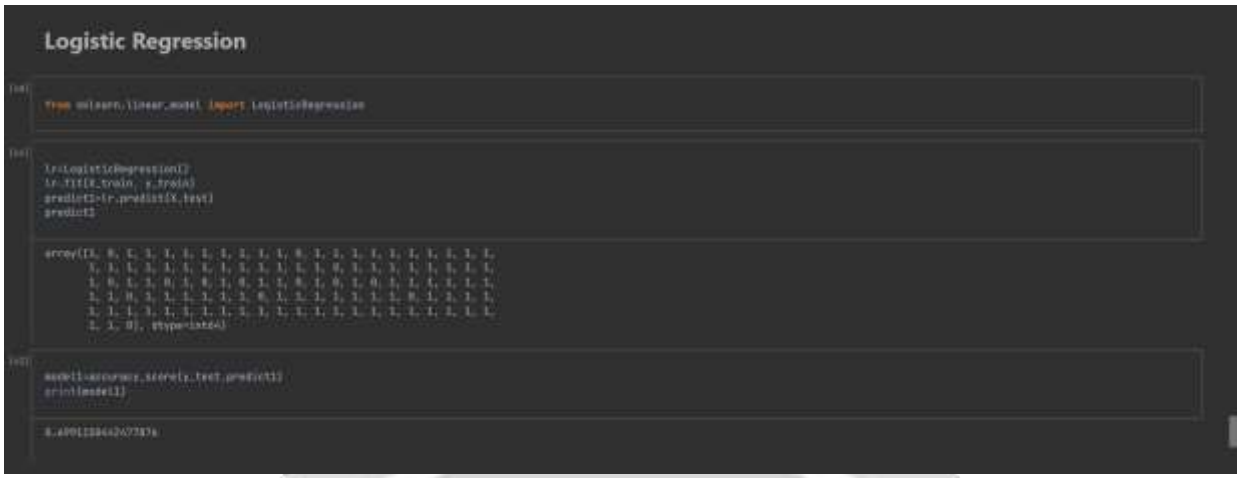
Test case screenshot :

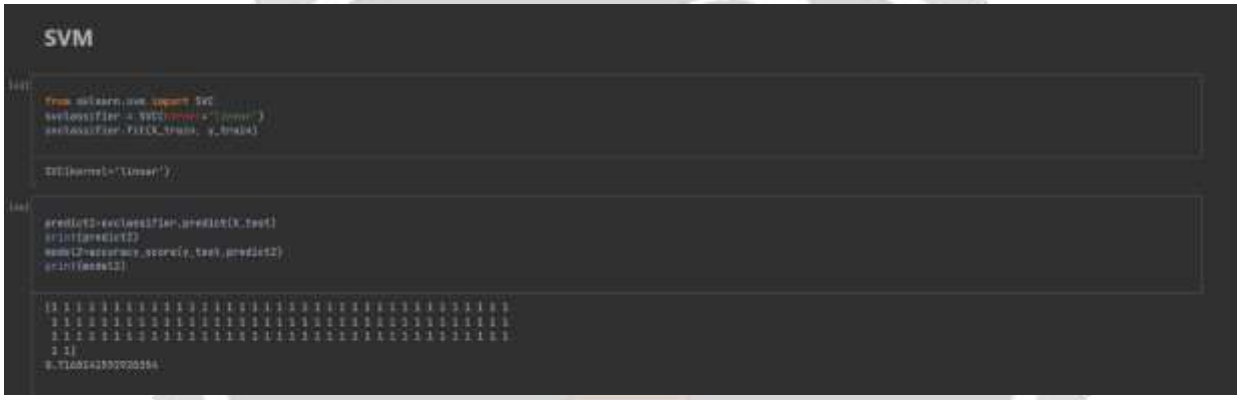


Output:



Accuracy of Training and Testing:





Conclusion:

The proposed project has been implemented in Anaconda for application development is developed using python. The tasks involved in this work are divided into modules. The proposed system is efficient and has friendly user interface. In this project, we have proposed methods for diagnosing liver disease in patients using machine learning techniques.

• The machine learning technique that were used are Logistic regression and SVM.

• From the project's results, we can conclude that we can predict the risk of liver diseases with more accuracy using SVM.

• A GUI, which can be used as a medical tool by hospitals and medical staff was implemented using SVM.

References:

[1] AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms. Zeel Doshi , Subhash Nadkarni , Rashi Agrawal, Prof. Neepa Shah.

[2] Kalimuthu M, Vaishnavi P and Kishore M 2020 Crop Prediction Using Machine Learning 2020 Third International Conference on Smart Systems and Inventive Teachnology (ICSSIT) pp 926-3210.1109/ICSSIT 10.1109/ICSSIT48917.2020.9214190.

[3] Prof. D.S. Zingade ,Omkar Buchade ,Nilesh Mehta ,Shubham Ghodekar ,Chandan Mehta "Crop Prediction System Using Machine Learning".

[4] https://www.ijert.org/crop-prediction-using-machine-learning-approaches