A Review of liver patient analysis using Machine Learning

TEAM MEMBERS

K.N.S.S.L RESHMA (20NM1A1224)

K. VIJAYA LAKSHMI (20NM1A1225)

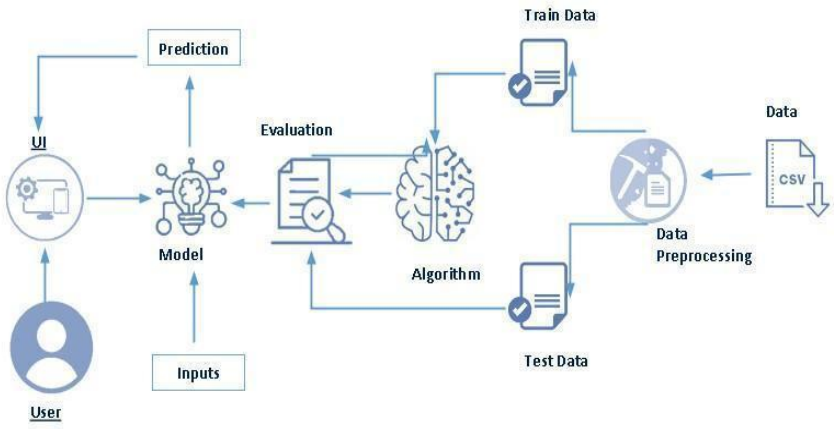
K. HARIKA GOWTHAMI (20NM1A1227)

M.V. S. S NAVYATHA (20NM1A1229)

A Review of Liver Patient Analysis Methods Using Machine Learning

Liver diseases averts the normal function of the liver. This disease is caused by an assortment of elements that harm the liver. Diagnosis of liver infection at the 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 liver disease diagnosis. 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. In this project we will 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 compares various classification algorithms such as Random Forest, Logistic Regression, KNN and ANN Algorithm with an aim to identify the best technique. Based on this study, Random Forest with the highest accuracy outperformed the other algorithms and can be further utilised in the prediction of liver disease and can be recommended.

Technical Architecture:



Project Flow

- User interacts with the UI to enter the input.
- Entered input is analysed by the model which is integrated.
- Once model analyses the input the prediction is showcased on the UI

TO accomplish this, we have to complete all the activities listed below,

- Define Problem / Problem Understanding
    - Specify the business problem
    - Business requirements
    - Literature Survey
    - Social or Business Impact.
- Data Collection & Preparation
    - Collect the dataset
    - Data Preparation
- Exploratory Data Analysis
    - Descriptive statistical
    - Visual Analysis
- Model Building
    - Training the model in multiple algorithms
    - Testing the model
- Performance Testing & Hyperparameter Tuning
    - Testing model with multiple evaluation metrics
    - Comparing model accuracy before & after applying hyperparameter tuning
- Model Deployment
    - Save the best model
    - Integrate with Web Framework
- Project Demonstration & Documentation
    - Record explanation Video for project end to end solution
    - Project Documentation-Step by step project development procedure

Understanding Liver Diseases

Classifications

Learn about different liver diseases' types, including hepatitis, cirrhosis, and fatty liver.

Risk Factors

Environmental and lifestyle factors play a role in liver diseases. diseases. Familiarize yourself with them here.

Symptoms

Early diagnosis of liver diseases is challenging. Explore the signs and symptoms to look out for.

Importance of Machine Learning in Liver Patient Analysis Analysis

1

Automated Diagnosis

Machine learning algorithms can quickly analyze medical and clinical data to accurately diagnose liver diseases.

2

Patient Monitoring

Machine learning can monitor patients' progress, predict disease progression and identify adverse outcomes early on.

3

Optimization of Clinical Trials

Machine learning can optimize the design and analysis of clinical trials for liver disease treatments.

Data Collection & Preparation

ML depends heavily on data. It is the most crucial aspect that makes algorithm training possible. So this section allows you to download the required dataset.

In this project we have used .csv data. This data is downloaded from kaggle.com. Please refer to the link given below to download the dataset.

Link: https://www.kaggle.com/datasets/uciml/indian-liver-patient-records.

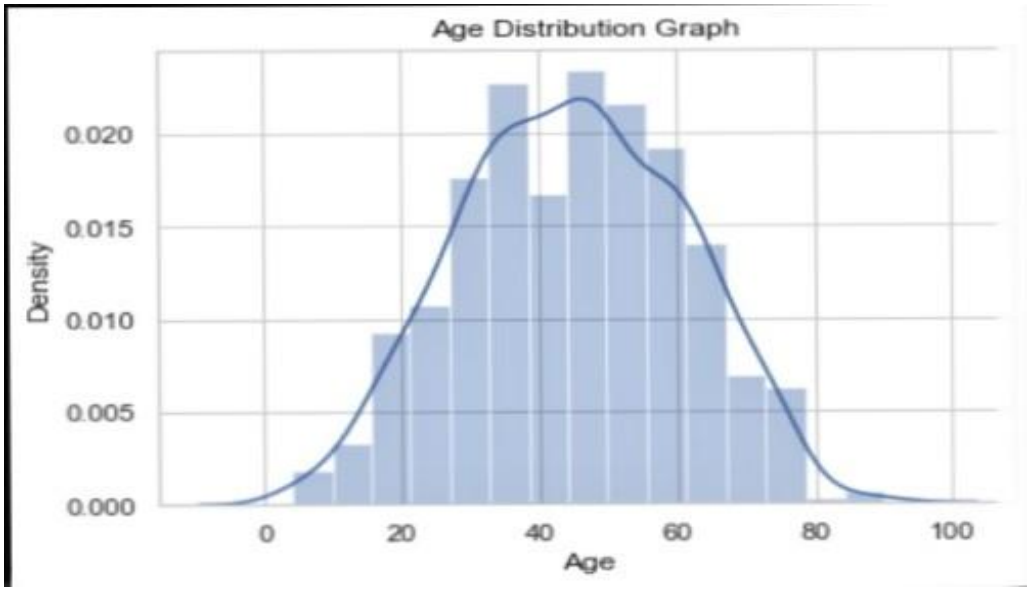
Data Preparation

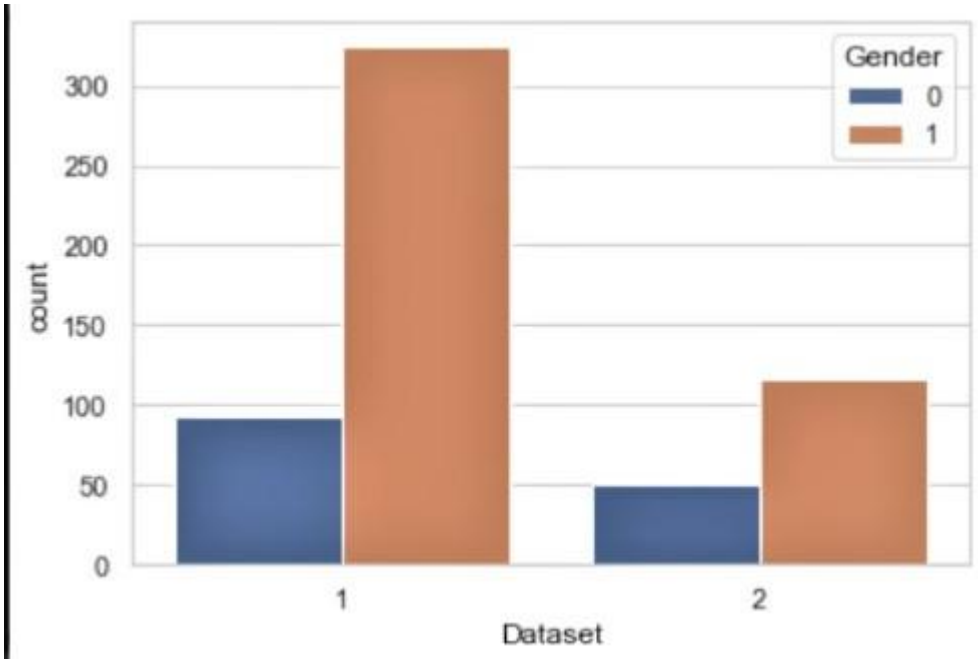
- As we have understood how the data is, let's pre-process the collected data.
- The download data set is not suitable for training the machine learning model as it might have so much randomness so we need to clean the dataset properly in order to fetch good results. This activity includes the following steps.
- Handling missing values
- Handling categorical data
- Note: These are the general steps of pre-processing the data before using it for machine learning. Depending on the condition of your dataset, you may or may not have to go through all these steps.

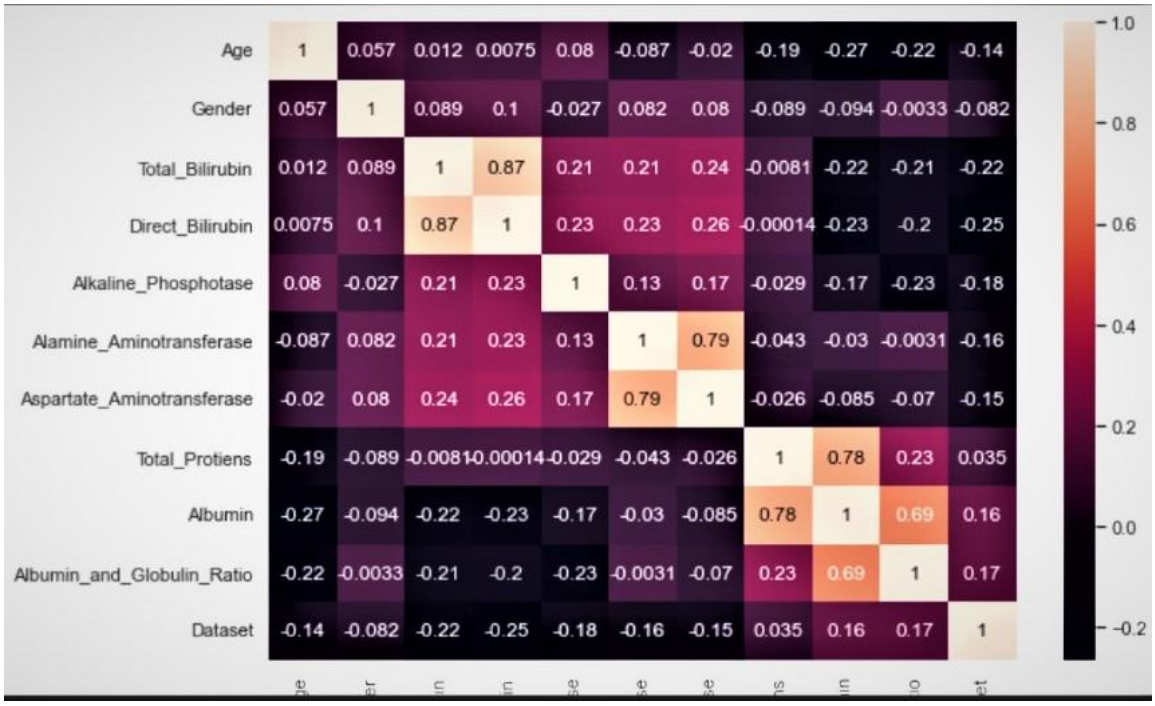
Exploratory Data Analysis

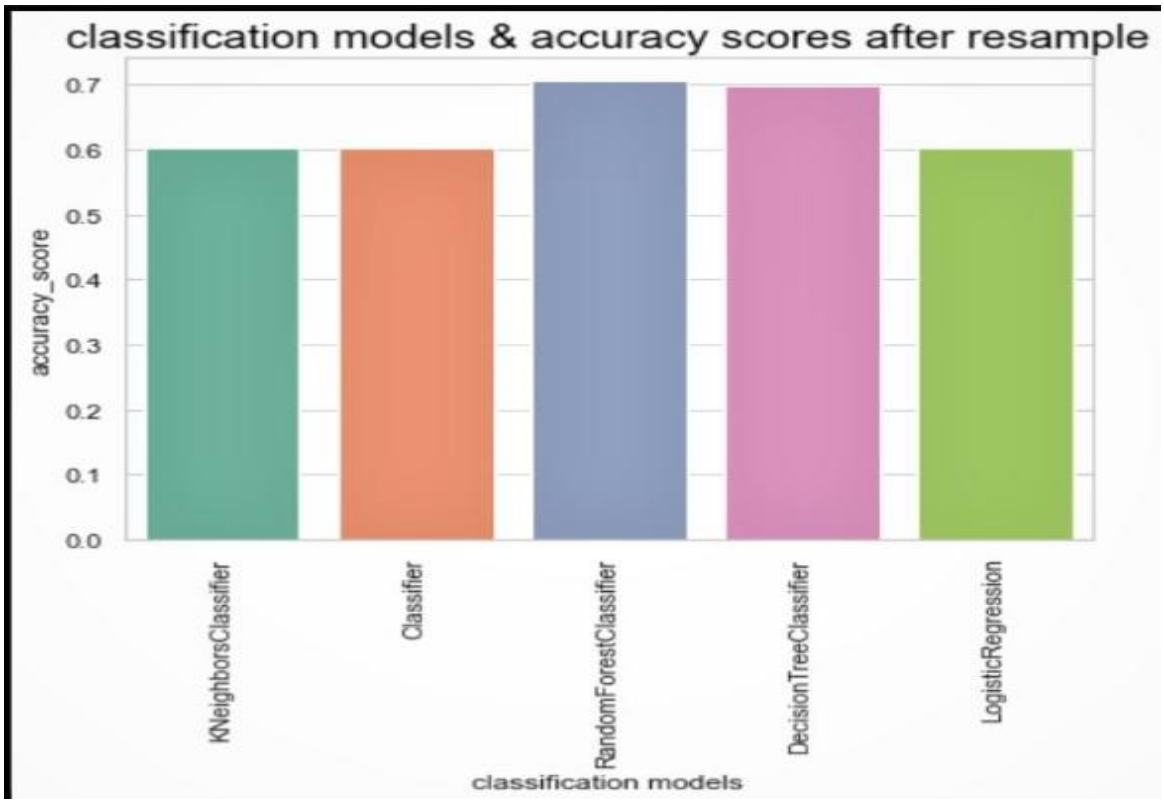
- Exploratory data analysis (EDA) is used by data scientists to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods.

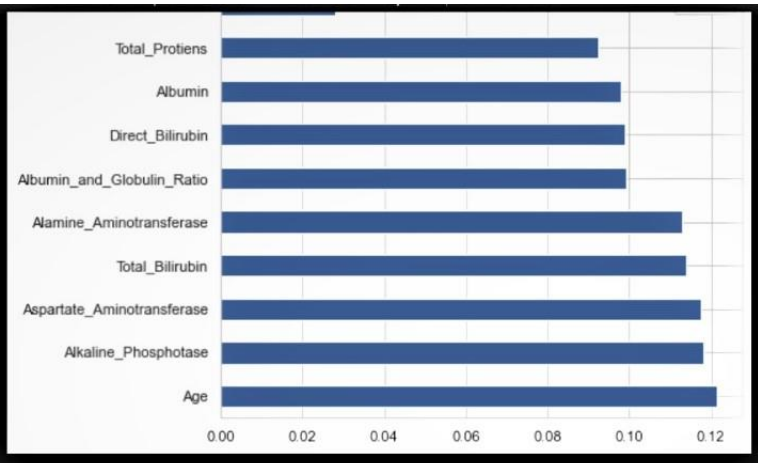
OUTPUT:


Age Distribution Graph
Density
0.020
0.015
0.010
0.005
0.000
0
20
40
60
80
100
Age


Gender
0
1
count
300
250
200
150
100
50
0
1
2
Dataset


Age 1 0.057 0.012 0.0075 0.08 -0.087 -0.02 -0.19 -0.27 -0.22 -0.14
Gender 0.057 1 0.089 0.1 -0.027 0.082 0.08 -0.089 -0.094 -0.0033 -0.082
Total_Bilirubin 0.012 0.089 1 0.87 0.21 0.21 0.24 -0.0081 -0.22 -0.21 -0.22
Direct_Bilirubin 0.0075 0.1 0.87 1 0.23 0.23 0.26 -0.00014 -0.23 -0.2 -0.25
Alkaline_Phosphotase 0.08 -0.027 0.21 0.23 1 0.13 0.17 -0.029 -0.17 -0.23 -0.18
Alamine_Aminotransferase -0.087 0.082 0.21 0.23 0.13 1 0.79 -0.043 -0.03 -0.0031 -0.16
Aspartate_Aminotransferase -0.02 0.08 0.24 0.26 0.17 0.79 1 -0.026 -0.085 -0.07 -0.15
Total_Protiens -0.19 -0.089 -0.0081 -0.00014 -0.029 -0.043 -0.026 1 0.78 0.23 0.035
Albumin -0.27 -0.094 -0.22 -0.23 -0.17 -0.03 -0.085 0.78 1 0.69 0.16
Albumin_and_Globulin_Ratio -0.22 -0.0033 -0.21 -0.2 -0.23 -0.0031 -0.07 0.23 0.69 1 0.17
Dataset -0.14 -0.082 -0.22 -0.25 -0.18 -0.16 -0.15 0.035 0.16 0.17 1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2


classification models & accuracy scores after resample
accuracy_score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
KNeighborsClassifier
Classifier
RandomForestClassifier
DecisionTreeClassifier
LogisticRegression
classification models

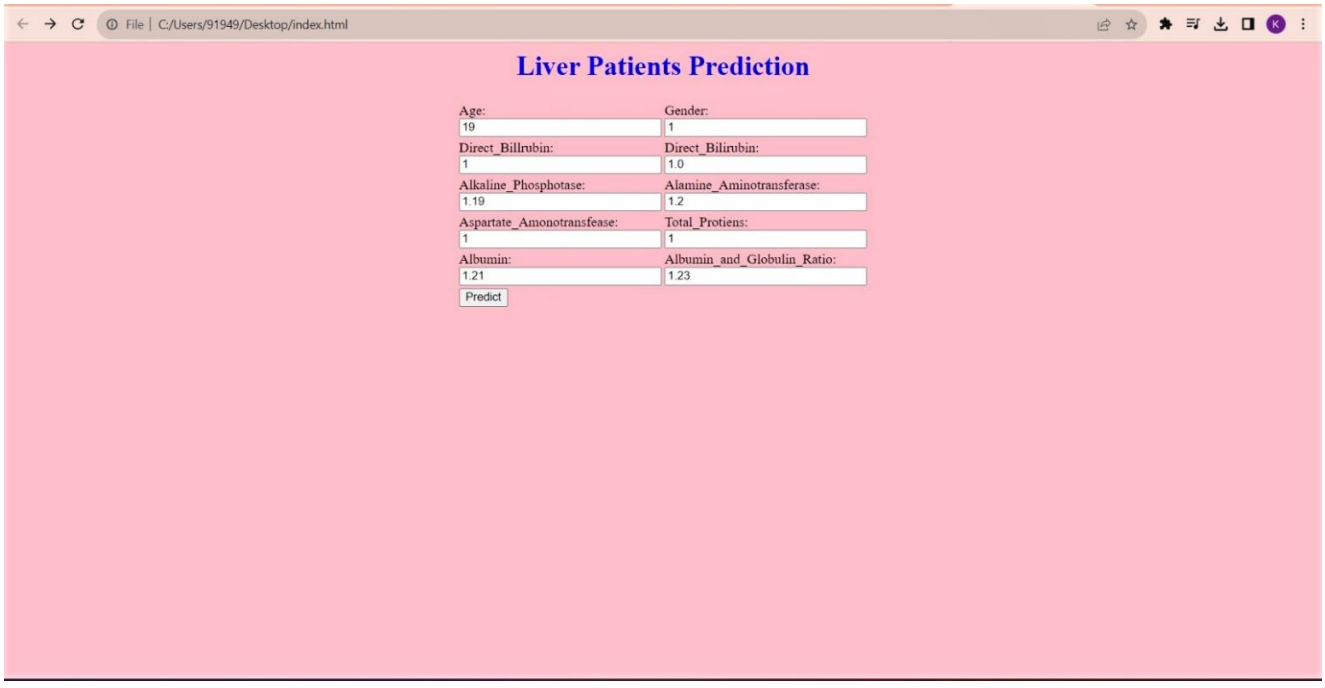


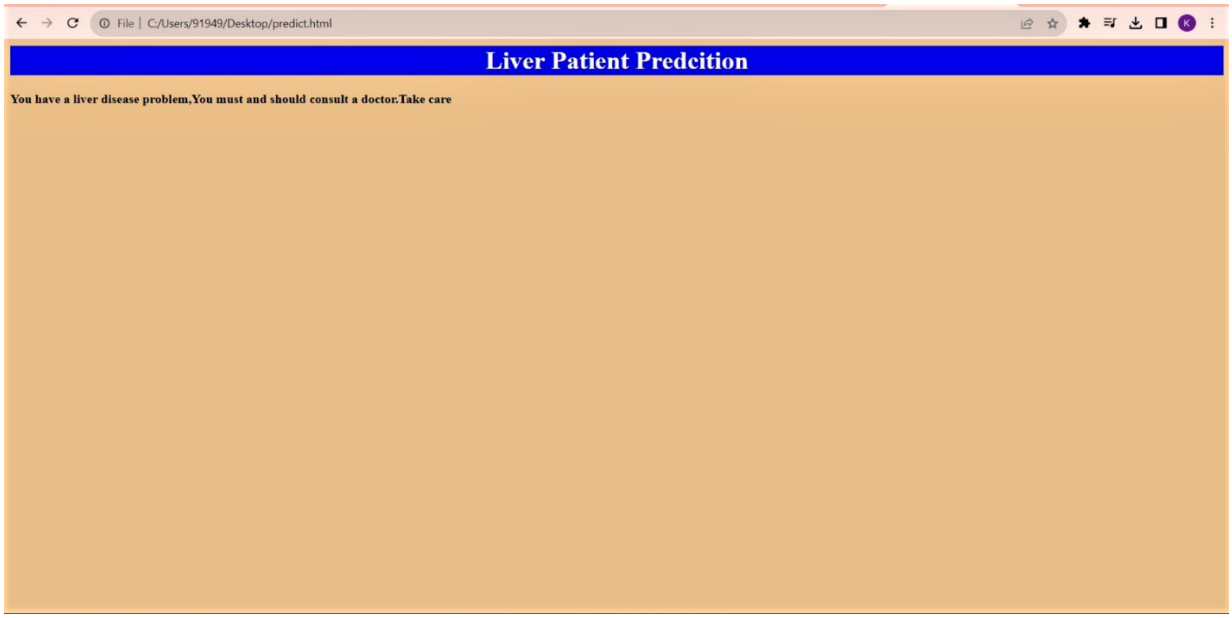
Direct_Billrubin & Total_Billrubin are the most important features to predict the outcomes

```
import joblib
joblib.dump(model1, 'ETC.pk1')
```

[277] Python

```
['ETC.pk1']
```





Conclusion:

In conclusion, the review of live patient analysis using machine learning techniques such as logistic regression, decision trees, and random forests underscores the significant contributions these methods have made to the field of healthcare.

Random Forest provides a solid foundation for predicting outcomes and understanding the relationships between variables in medical contexts. Its simplicity and interpretability make it a valuable tool for risk assessment and prognosis, aiding clinicians in making informed decisions about patient care.

Decision trees offer an intuitive approach to analyzing patient data, allowing for the identification of key factors that influence medical outcomes. The visual and easily understandable nature of decision trees aids in uncovering complex patterns and contributing to diagnostic and treatment strategies.