



BCSE351E

Foundations of Data Analytics
DIGITAL ASSIGNMENT -2

Disease Prediction of Liver Patients

Akruti Somkuwar – 21BLC1024
Devarinti Jyothika – 21BLC1050
Vamsi Krishna K – 21BLC1065

The background is a light blue gradient. It features several decorative elements: a large cyan oval in the top left, a pink oval in the middle left, a large pink abstract shape in the top right, and a pink abstract shape in the bottom left. Scattered throughout are various colored circles in shades of cyan, pink, orange, and red.

01

Problem Statement

>>>Liver Patient Analysis

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. Diagnosis at an early stage of this deadly disease increases the safety and overall well being of the patient.

Thus,there is a need to create data mining algorithms and techniques to study medical data which saves time of doctors and automatic prediction of the disease.



02

Objectives/Motivation

>>> Aim:

Early prediction of liver disease is very important to save human life and take proper steps to control the disease. Thus we aim to evaluate blood markers and predict various diseases in an effort to reduce burden on doctors.

03

*Literature Survey /
Related Work*

>>>Dataset

Patient records collected from North East of Andhra Pradesh

Click-> [Indian Liver Patient Record](#)

CODE:

```
import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

# Load the liver disease dataset

dataset = pd.read_csv('indian_liver_patient.csv')

dataset.head()
```



```
dataset.Dataset.value_counts()
```

```
dataset["Liver_Disease"] = dataset['Dataset'].map({1:1,2:0})
```

```
dataset.head()
```

```
dataset.isnull().sum()
```

```
dataset.Albumin_and_Globulin_Ratio.fillna(dataset['Albumin_and_Globulin_Ratio'].mean(),inplace=True)
```

```
dataset.isnull().sum()
```

- from sklearn.preprocessing import LabelEncoder



```
dataset['Gender'] = dataset['Gender'].map({"Male":1,"Female":0})
```

```
sns.heatmap(data = dataset.corr())
```

```
plt.show()
```

```
val = dataset['Liver_Disease'].value_counts()
```

```
val.plot(kind="bar")
```




```
plt.show()
```

```
from sklearn.feature_selection import SelectKBest
```

```
from sklearn.feature_selection import chi2
```

```
# Extract the features and the target variable
```



```
features=dataset[['Total_Bilirubin','Direct_Bilirubin','Alkaline_Phosphotase','Alamine_Aminotransferase','Aspartate_Aminotransferase', 'Total_Protiens', 'Albumin', 'Albumin_and_Globulin_Ratio']]
```

```
target = dataset['Liver_Disease']
```

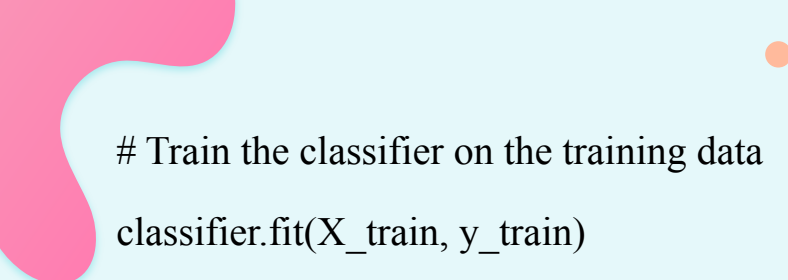
```
# Split the dataset into training and testing sets
```

```
X_train, X_test, y_train, y_test = train_test_split(features, target, test_size=0.2, random_state=42)
```

```
# Create a random forest classifier
```

```
classifier = RandomForestClassifier(n_estimators=70)
```



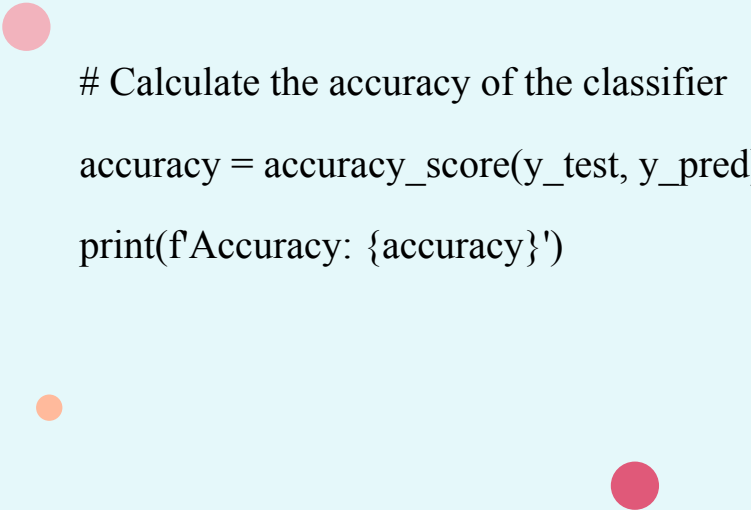


Train the classifier on the training data

```
classifier.fit(X_train, y_train)
```

Make predictions on the testing data

```
y_pred = classifier.predict(X_test)
```




Calculate the accuracy of the classifier

```
accuracy = accuracy_score(y_test, y_pred)
```

```
print(f'Accuracy: {accuracy}')
```





04

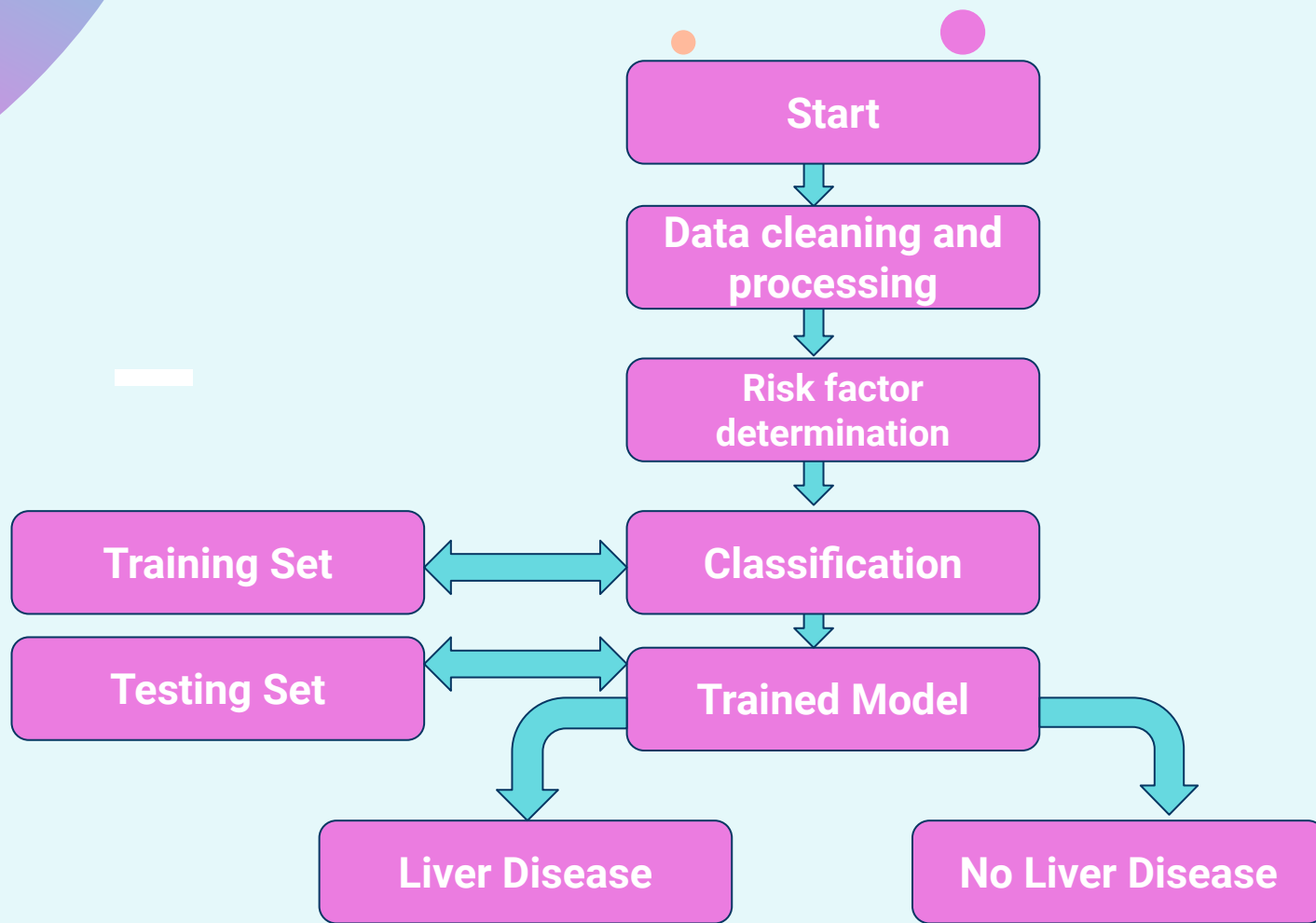
Research Gap

>>>Gaps are as follows:

- 1) Diseases related to Liver are major cause of recurrent hospitalizations in those population with a history of heavy alcohol consumption. Unfortunately, despite its clinical relevance, there are many gaps in knowledge related to this syndrome that represent barriers to the development of effective preventive surveillance, early detection, and therapeutic strategies.
- 2) In this analysis, various parameter and tests are to be carried out. One or two tests are not enough for prediction of the disease. Thus we cannot pinpoint the exact disease with just few blood markers given as there are same parameters across various disease.
- 3) Also, the patient may get diagnosed with multiple diseases which will lead to medical inaccuracy. Thus this analysis demands more parameters to be accurate.

05

***Block Diagram/
Methodology***



The background is a light blue gradient. It features several large, organic, blob-like shapes in shades of pink, orange, and purple. Scattered throughout are numerous small, solid-colored circles in pink, orange, and teal.

06

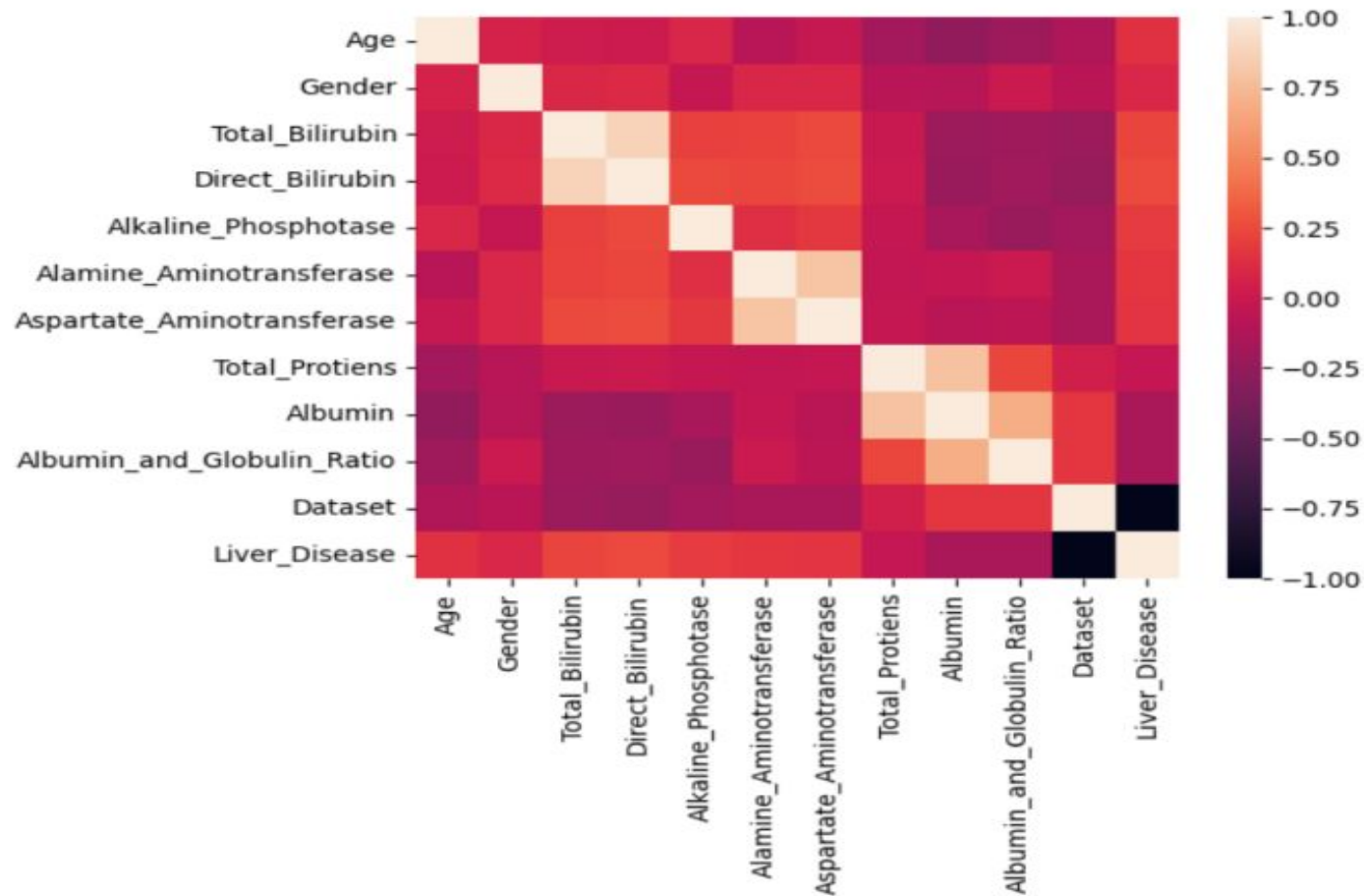
Results

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	Albumin	Albumin_and_Globulin_Ratio	Data
0	65	Female	0.7	0.1	187	16	18	6.8	3.3	0.90	
1	62	Male	10.9	5.5	699	64	100	7.5	3.2	0.74	
2	62	Male	7.3	4.1	490	60	68	7.0	3.3	0.89	
3	58	Male	1.0	0.4	182	14	20	6.8	3.4	1.00	
4	72	Male	3.9	2.0	195	27	59	7.3	2.4	0.40	

Null values Sum

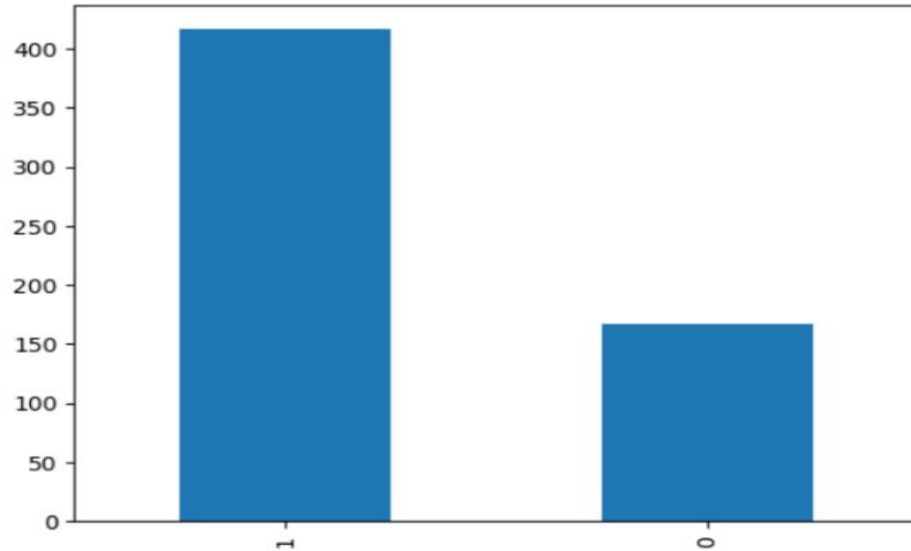
```
dataset.Albumin_and_Globulin_Ratio.fillna(dataset['Albumin_and_Globulin_Ratio'].mean(),inplace=True)
dataset.isnull().sum()
```

```
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
Liver_Disease  0
dtype: int64
```



It shows Normal people and Liver disease patients

```
[93] val = dataset['Liver_Disease'].value_counts()  
val.plot(kind="bar")  
plt.show()
```



With a Accuracy of 74.32

Publications

INTERNATIONAL JOURNALS

- “Journals of hepatology” : published in the year 1985.
- “Liver International” : published in the year 2000.

INTERNATIONAL CONFERENCES

1. “Machine Learning “(ICML) in Proc International Conferences held at Long Beach and Sydney.
2. “Data Science and Advanced Analytics “(DSAA) in Proc International Conference at Tokyo, Washington.

References

LINKS:

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5510030/>
- Liver disease analysis

BOOKS:

- “Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning” By Geeta Rani & Pradeep Kumar Tiwari.
- "Machine Learning Techniques for Predictive Maintenance" by Chintan Bhatt and Nehal Shah

JOURNALS:

- “A Data Mining Approach for Early Detection and Prediction of Liver Disease”By John Smith.
- “Intelligent Techniques and Comparative Performance Analysis of Liver Disease Prediction” By Sreenivasa Rao Veeranki and Manish Varshney.

CONFERENCE PAPER:

- “Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms” By M.Banu Priya,P.Laura Juliet,P.R.Tamilselvi.
- “Liver Disease Prediction Using Machine Learning Classification Techniques” By Ketan Gupta, Nasmin Jiwani, Neda Afreen, Divyarani D,Proceeding of the IEEE

The background is a light blue gradient. It features several small, solid-colored circles in teal, pink, and orange scattered across the frame. There are also larger, abstract, organic shapes in shades of pink, purple, and teal, primarily located in the corners and along the edges.

Thank you