**LIVER PATIENT ANALYSIS USING IN MACHINE LEARNING**

**INTRODUCTION**

Prediction of the disease in the human being is the very long and difficult process in early days. Now a days,

computer aided diagnosis is the important role in the medical industry for predicting, analyzing and storing

medical information with the images. In this paper will discuss and classify the liver patients with the help of the

liver patient dataset with the help of the machine learning algorithms. WEKA is the software used here for

implement the some of the classification algorithms with the data selected from the liver disease dataset. After

the successful implementation of the all the algorithms, the best algorithms selected from the output of the all the

algorithms execution.

* 1. **Overview**

The largest organ in an abdoman is the liver in the shape of

triangular. The two parts of the livr is left and right hemi liver. It is

a single organ. Liver used to essential for function our body. This

is the primary organ for maintaining the chemicals like glucose,

balancing the so many nutirents, fat, vitamis, choleserol and

hormones.[1] In an early stage of the liver problem diagnosition

will increase the the survival rate of the patient. Analyzing the

1. enzymesin’s levels will lead to diagnose the liver diseases from

In Knowledge Discovery Process, Data mining methods are

partitioned into two noteworthy classes. These are expressive

compose and forecast write. Every one of the sort will have

distinctive kind of the methodologies[8]. Information Mining is a

procedure of separating conceivably helpful, already obscure data

from the crude information. Information mining is one stage in the

KDD procedure. It is the most explored piece of the procedure.

Information mining is characterized as a "sort of database

investigation that endeavors to find valuable examples or

connections in a gathering of information. The examination

utilizes progressed measurable strategies, for example, bunch

investigation, and in some cases utilizes counterfeit consciousness

or neural system methods. A noteworthy objective of information

mining is to find beforehand obscure connections among the

information, particularly when the information originate from

various databases [9]."Classification algorithm that is generally

utilized as a part of expectations basing on chronicled information

.

Classification is a class expectation system, which is regulated in

nature. This system has the capacity to anticipate the name for

classes, gave that adequate quantities of preparing cases are access

**purpose**

Liver function tests can be used to:

* Screen for liver infections, such as hepatitis
* Monitor the progression of a disease, such as viral or alcoholic
* hepatitis, and determine how well a treatment is working
* Measure the severity of a disease, particularly scarring of the liver (cirrhosis)
* Monitor possible side effects of medications

Liver function tests check the levels of certain enzymes and proteins in

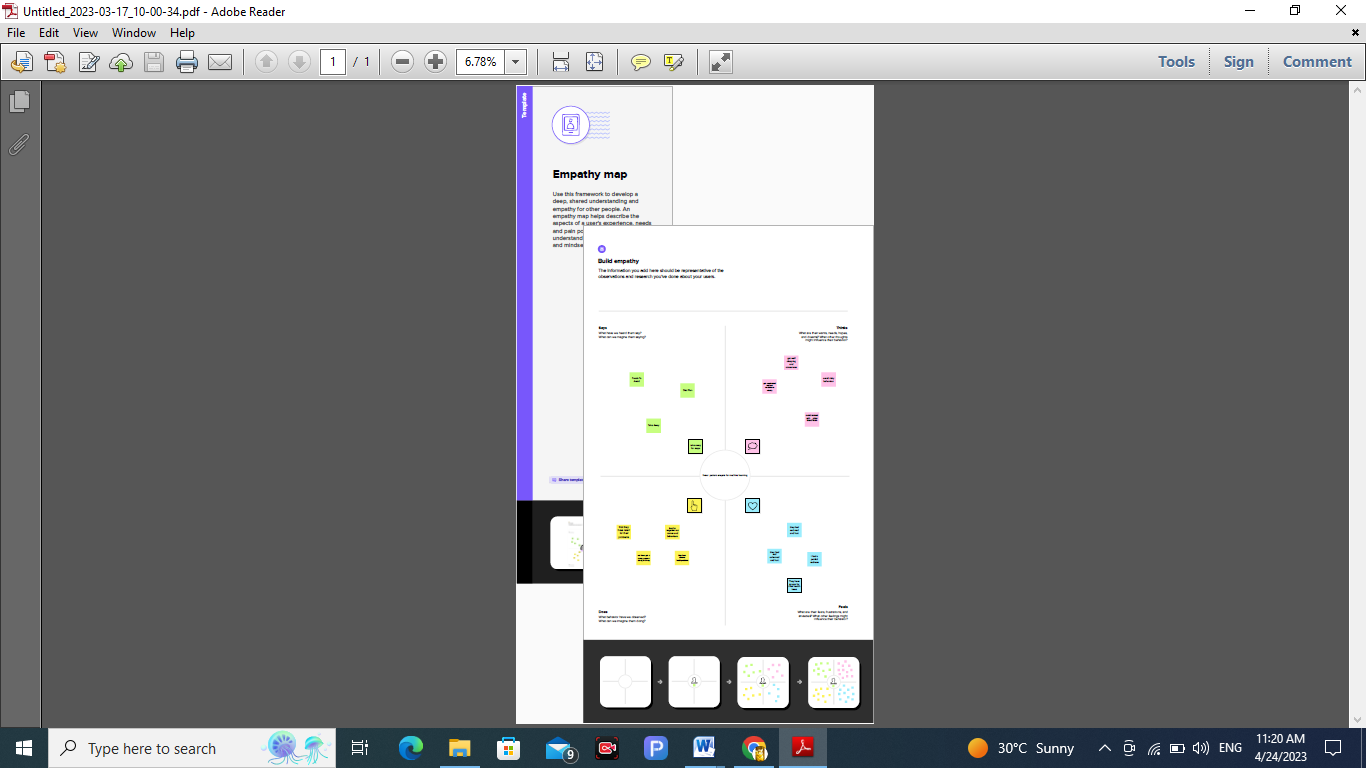
your blood. Levels that are higher or lower than normal can indicate live

r problems. Some common liver function tests include:

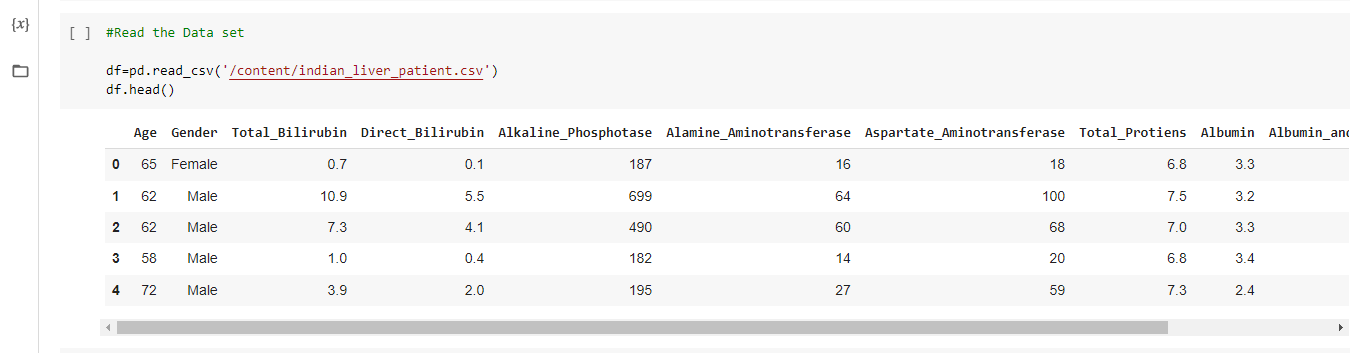
* **Alanine transaminase (ALT).** ALT is an enzyme found in the liver
* that helps convert proteins into energy for the liver cells. When the
* liver is damaged, ALT is released into the bloodstream and levels
* increase.
* **Aspartate transaminase (AST).** AST is an enzyme that helps
* metabolize amino acids. Like ALT, AST is normally present in blood at
* low levels. An increase in AST levels may indicate liver damage,
* disease or muscle damage.
* **Alkaline phosphatase (ALP).** ALP is an enzyme found in the liver
* and bone and is important for breaking down proteins. Higher-than
* -normal levels of ALP may indicate liver damage or disease, such as
* a blocked bile duct, or certain bone diseases.
* **Albumin and total protein.** Albumin is one of several proteins made
* in the liver. Your body needs these proteins to fight infections and to
* perform other functions. Lower-than-normal levels of albumin and
* total protein may indicate liver damage or disease.
* **Bilirubin.** Bilirubin is a substance produced during the normal
* breakdown of red blood cells. Bilirubin passes through the liver and is
* excreted in stool. Elevated levels of bilirubin (jaundice) might indicate
* liver damage or disease or certain types of anemia.
* **Gamma-glutamyltransferase (GGT).** GGT is an enzyme in the
* blood. Higher-than-normal levels may indicate liver or bile duct
* damage.
* **L-lactate dehydrogenase (LD).** LD is an enzyme found in the liver.
* Elevated levels may indicate liver damage but can be elevated in
* many other disorders.
* **Prothrombin time (PT).** PT is the time it takes your blood to clot.
* Increased PT may indicate liver damage but can also be elevated if
* you're taking certain blood-thinning drugs, such as warfarin.

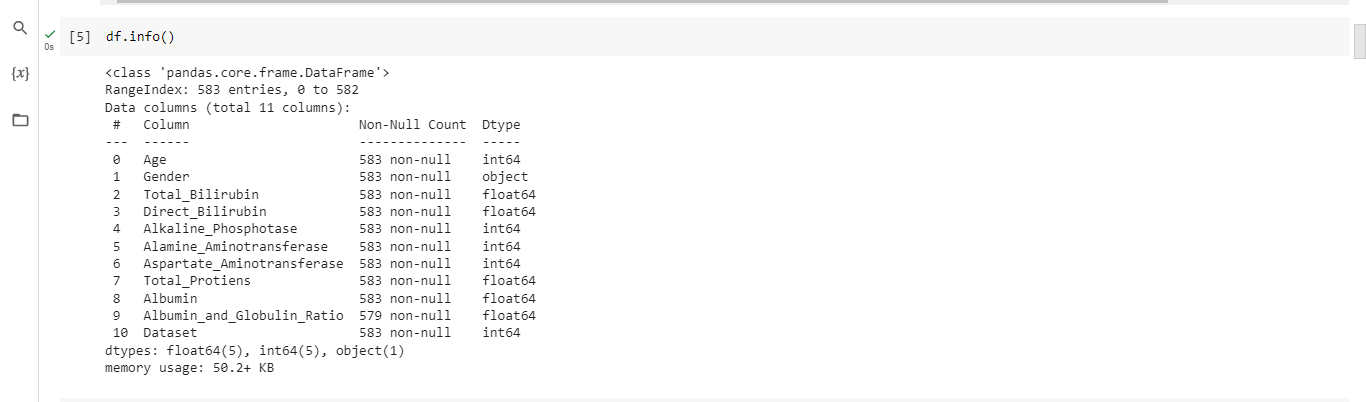
**2.Problem defining and design thinking**

* 1. **Empathy map**

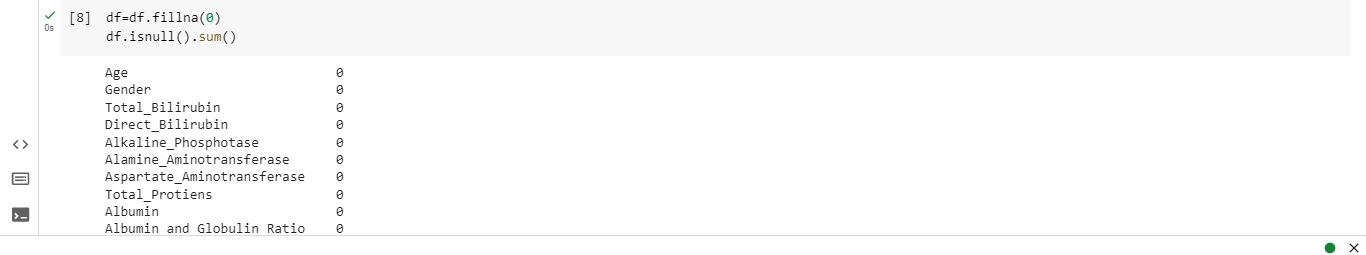


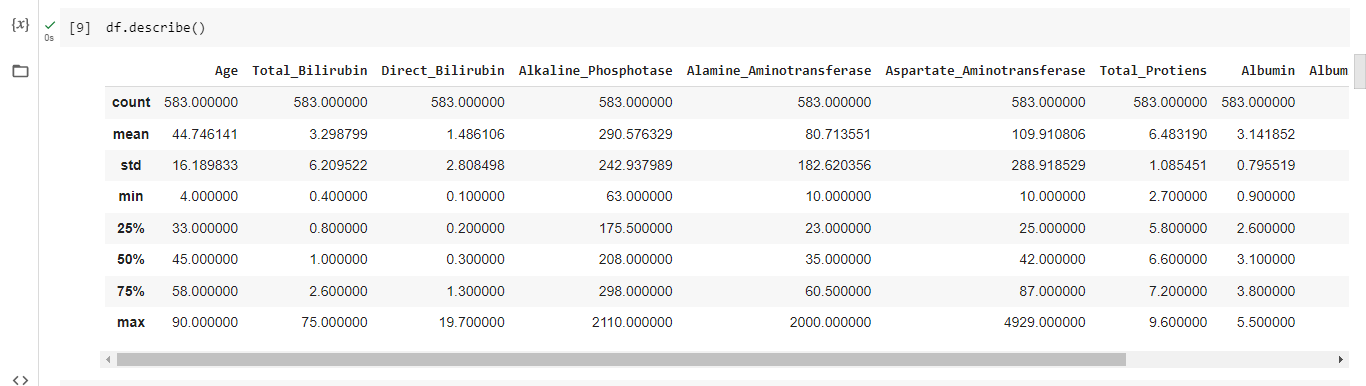
**RESULT**

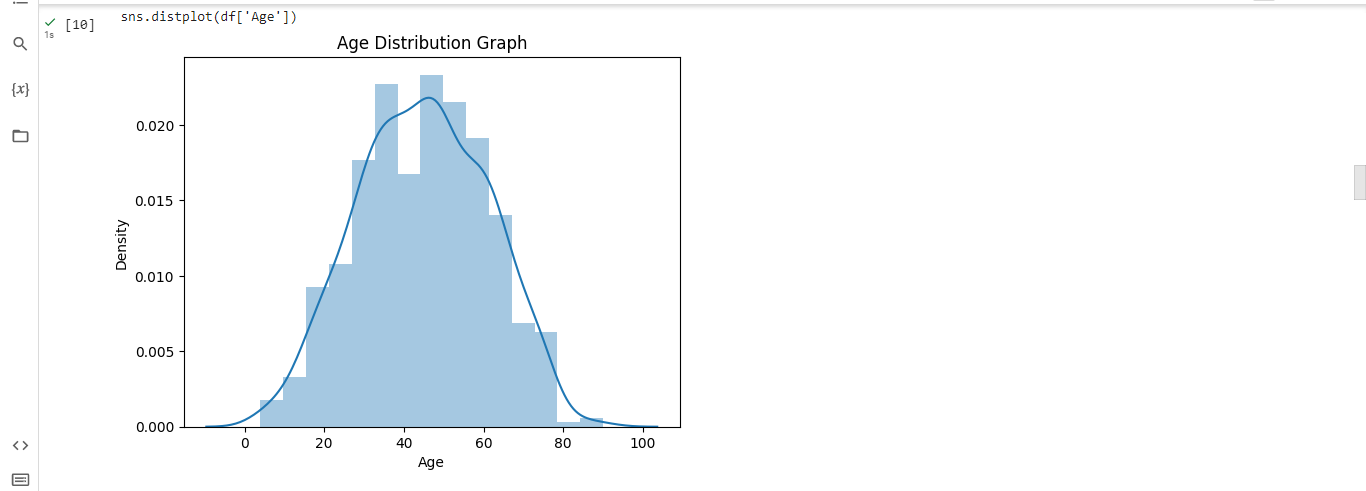
****

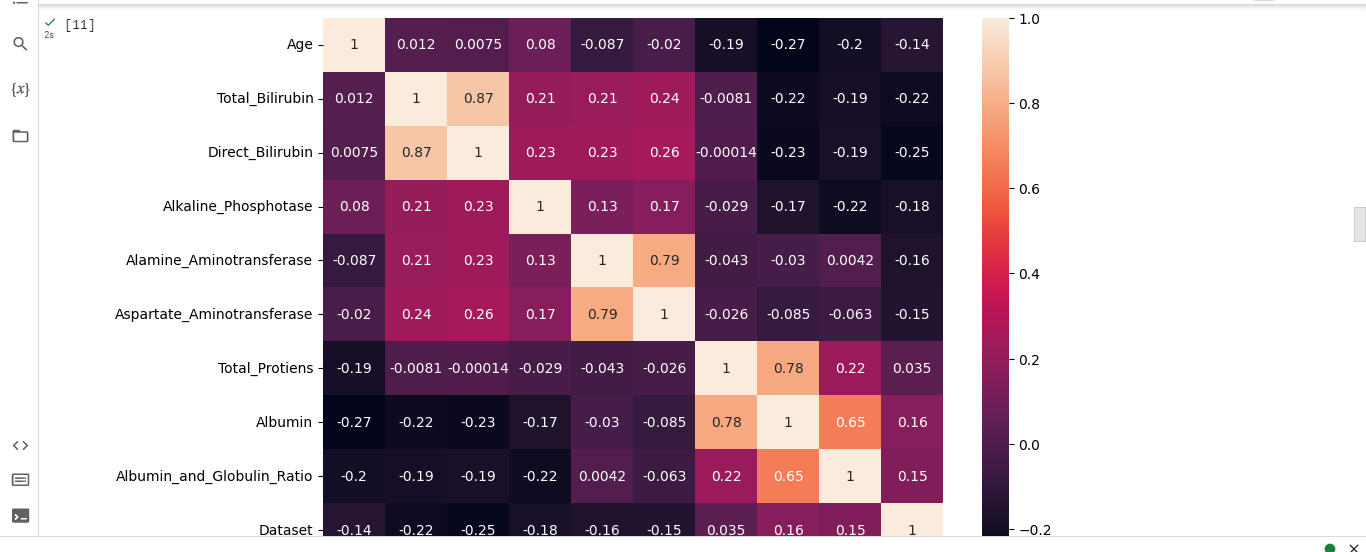
****

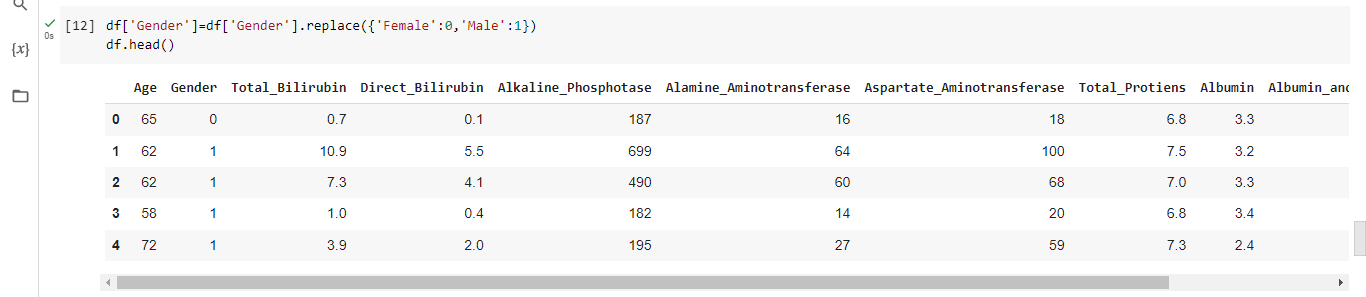
****

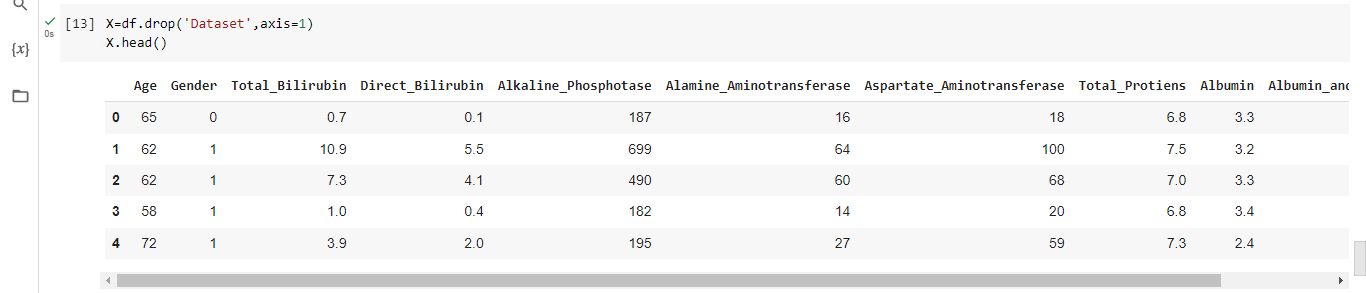
****

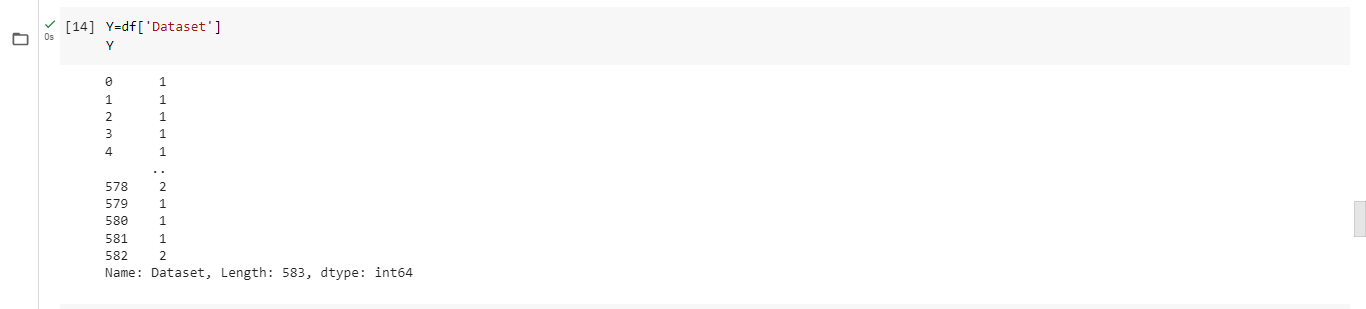
****

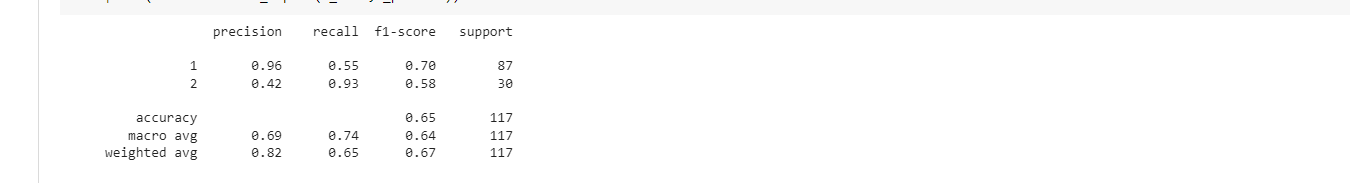
****

****

****

****

****

****

**4.Advantages and disadvantages**

1

| **Advantages** | **Disadvantages** |
| --- | --- |
| Diagnostic criterion standard |  |
| Confirmed diagnostic value | Highly invasive test |
| Etiologic suggestion |  |
| Differential diagnosis | The potential complications include death |
| Grade and stage evaluation |  |
| Therapeutic decision | Significant sampling error |
| (eligibility) | High cost |
| Treatment evaluation | Inter-observer variation |
| (effectiveness) |  |
| Follow-up comparison of treated and untreated patients |  |
| **APPLICATION** |  |

The first stage in the LDP method proposed in this study is ‘Sample’. After deciding on the topic

for the study, the first step is data collection. It is referred to as data collection

and considering the part

of data useful for the study (Azevedo & Santos, 2008). So, the data sets related to

liver diseases are

searched on different platforms named UCI repository and Kaggle. The suitable

dataset is found from

the platform Kaggle, a binary classification dataset that determines whether the

patient has liver disease.

After observing the credibility of the dataset, this dataset named ‘Indian Liver Patient Records’ is

selected.

3.2 Explore

The second stage is ‘Explore’. Exploring the data stage involves data

understanding. This

exploration stage also comprises finding the surprising trends and patterns

present in the data to generate

new ideas (Azevedo & Santos, 2008). In this study, exploring the data is at two

stages. One is the data

exploration on the background of liver disease. The other stage is exploring the

dataset, which shows

the details regarding the attributes present and how these attributes are correlated

with each other and

how these input attributes are correlated with the output attribute are studied. In

addition, missing values

are also identified. This analysis is performed using R.

Modify

The third stage is ‘Modify’. Modify refers to data transformation (Azevedo & Santos, 2008). In this

study, the attributes in the dataset are not in the same format, and the attribute’s data type restricts the

analysis to be done on the attribute. So, some of the features having the data type

integers are converted

into numerical, which makes all the attributes have the same numerical data type

and makes the analysis

be done efficiently.

3.4 Data Preprocessing

The fourth stage is ‘Data preprocessing’. This data preprocessing refers to cleaning and preparing

the data for modelling (Azevedo & Santos, 2008). This data preprocessing

involves replacing the

missing values and balancing the dataset as the class distribution of the dataset is

imbalanced. This

balancing is done using the Random Over Sampling Example (ROSE) (Menardi

& Torelli, 2014).

3.5 Model

The fifth stage is the ‘Model’. The modelling stage means applying the selected techniques or the

algorithms to the data (Azevedo & Santos, 2008). The five algorithms, SVM,

Naïve Bayes, LDA,

CART and K-NN, are applied.

3.6 Assess

Assess stage, which is the sixth stage, involves assessing the data by deciding

whether the data

produced from modelling techniques are reliable and accurate. This stage also

evaluates how well the

algorithms performed on the data (Azevedo & Santos, 2008).

3.7 Results

The seventh stage of the proposed LDP method is ‘Results’. The results stage involves presenting

the results after assessing the data. All the results of accuracies and confusion

matrix metrics will be

described.

4 Performance Analysis

4.1 Descriptive Analysis

The dataset selected for this study is the liver disease dataset. This dataset, named ‘Indian Liver

Patient Records’, is obtained from Kaggle. The data from the dataset is collected from the North-East

part of Andhra Pradesh, India (Indian Liver Patient Records, n.d.). It is a binary

classification dataset

predicting whether the patient has liver disease or not. As stated in Table 1, the

dataset contains 583

instances and 11 attributes. Of those 11 attributes, one of the attributes is class

which denotes whether

**Conclusion**

This persistence of this study is to assess and examined the data

composed from the Andhra Pradesh’s North East area in India for

applying the machine learning algorithm with the help of the

machine learning Tool. All the selected algorithms are executed

and classified the liver disease patients as well the non-diseased

patient from the selected data. The highest accuracy given with

nominal execution time taken is the Random Tree algorithm. The

time taken for the executing this algorithm is higher than the all

other selected algorithms for this study. But the comparison

between the all other algorithm, the Random Tree algorithm given

higher accuracy. The liver disease predicted with the few machine

learning algorithms. This study will helpful to the medical area

people for the easy predictions

**FUTURE SCOPE**

Diseases related to liver and heart are becoming more and more

common with time. With continuous technological advancements, these

are only going to increase in the future. Although people are becoming

more conscious of health nowadays and are joining yoga classes, dance

classes; 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.

With the dataset that we used for this project, we got 100 % accuracy for

SVM model, and though it might be difficult to get such accuracies with

very large datasets, from this projects results, one can clearly conclude

that we can predict the risk of liver diseases with accuracy of 90 % or

more.

Today almost everybody above the age of 12 years has smartphones

with them, and so we can incorporate these solutions into an android app

or ios app. Also it can be incorporated into a website and these app and

website will be highly beneficial for a large section of society.

**APPENDIX**

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from matplotlib import rcParams

from scipy import stats

df=pd.read\_csv('/content/indian\_liver\_patient.csv')

df.head()

df.info()

df.isnull().any()

df.isnull().sum()

df=df.fillna(0)

df.isnull().sum()

df.describe()

sns.distplot(df['Age'])

plt.title('Age Distribution Graph')

plt.show()

plt.figure(figsize=(10,7))

sns.heatmap(df.corr(),annot=True)

df['Gender']=df['Gender'].replace({'Female':0,'Male':1})

df.head()

X=df.drop('Dataset',axis=1)

X.head()

Y=df['Dataset']

Y

from sklearn.preprocessing import scale

X\_scaled=pd.DataFrame(scale(X),columns=X.columns)

X\_scaled.head()

X=df.iloc[:,:-1]

Y=df.Dataset

from sklearn.model\_selection import train\_test\_split

X\_train,X\_test,Y\_train,Y\_test=train\_test\_split(X\_scaled,Y,test\_size=0.2,random\_state=42)

pip install imblearn

from imblearn.over\_sampling import SMOTE

smote=SMOTE()

Y\_train.value\_counts()

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy\_score,classification\_report,confusion\_matrix

model1=RandomForestClassifier()

model1.fit(X\_train\_smote, Y\_train\_smote)

Y\_predict=model1.predict(X\_test)

rfc1 = accuracy\_score(Y\_test,Y\_predict)

rfc1

pd.crosstab(Y\_test,Y\_predict)

print(classification\_report(Y\_test,Y\_predict))

import tensorflow.keras

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Dense

classifier=Sequential()

classifier.add(Dense(units=100,activation='relu',input\_dim=10))

classifier.add(Dense(units=50,activation='relu'))

classifier.add(Dense(units=1,activation='sigmoid'))

classifier.compile(optimizer='adam',loss='binary\_crosssentropy',metrics=['accuracy'])