kamarajar Arts and Science college

surandai

Tenkasi - District.

Machine Learn With Python

April - 2023

A REVIEW OF LIVER PATIENT ANALYSIS METHODS USING MACHINE LEARNING

Team Size: 4

Team Leader: S.Sivanandh (Re.no: 20201061506141)

Team Members: S.Santhosh Sivakumar (Re.no: 20201061506134)

M.Mohamed Sahith (Re.no: 20201061506124)

M.Muthukutty Raja (Re.no: 20201061506127)

Table of Contents

Introduction

Liver disease averts the normal function of the liver. This disease is caused by a 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 infection. 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 algorithm is an efficacious task that can help the doctors to diagnose the disease within short duration of time. In this project we will analysis the parameters of various classification algorithms and compare their predictive accuracies so as to find out the best classifier 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 techniques. Based on this study, Random Forest with the highest accuracy outperformed the order algorithm and can be further utilized in the prediction of the liver disease and can be recommended to the user.

The liver is a critical organ in the human body that is responsible for an array of functions that help support metabolism, immunity, digestion, detoxification, vitamin storage among other functions. It comprises around 2% of an adult's body weight. The liver is a unique organ due to its dual blood supply from the portal vein (approximately 75%) and the hepatic artery (approximately 25%).

1.1 overview

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 liver is the largest glandular organ of the body. It weighs about 3 lb (1.36 kg) .The liver's main job is to strain the blood coming from the digestive tract, before passing it to the rest of the body. The liver also detoxifies chemicals and metabolizes drugs. As it does so, the liver hides bile that ends up back in the intestines. The liver also makes proteins important for blood clotting and other functions. The liver supports almost every organ in the body and is vital for our survival. Liver disease may not cause any symptoms at earlier stage or the symptoms may be

vague, like weakness and loss of energy. Symptoms partly depend on the type and the extent of liver disease. Liver diseases are diagnosed based on the liver functional test. Several diseases states can disturb the liver. Some of the diseases are Wilson's disease, hepatitis (an inflammation of the liver), liver cancer, and cirrhosis (a chronic inflammation that progresses ultimately to organ failure). Alcohol alters the metabolism of the liver, which can have on the whole detrimental effects if alcohol is taken over long periods of time. Hemo-chromatos is can cause liver problems.

Classification algorithm is one of the greatest significant and applicable data techniques used to apply in disease prediction. Classification algorithm is the most common in several automatic medical health diagnoses. Many of them show good classification

1.*2* ***Purpose***

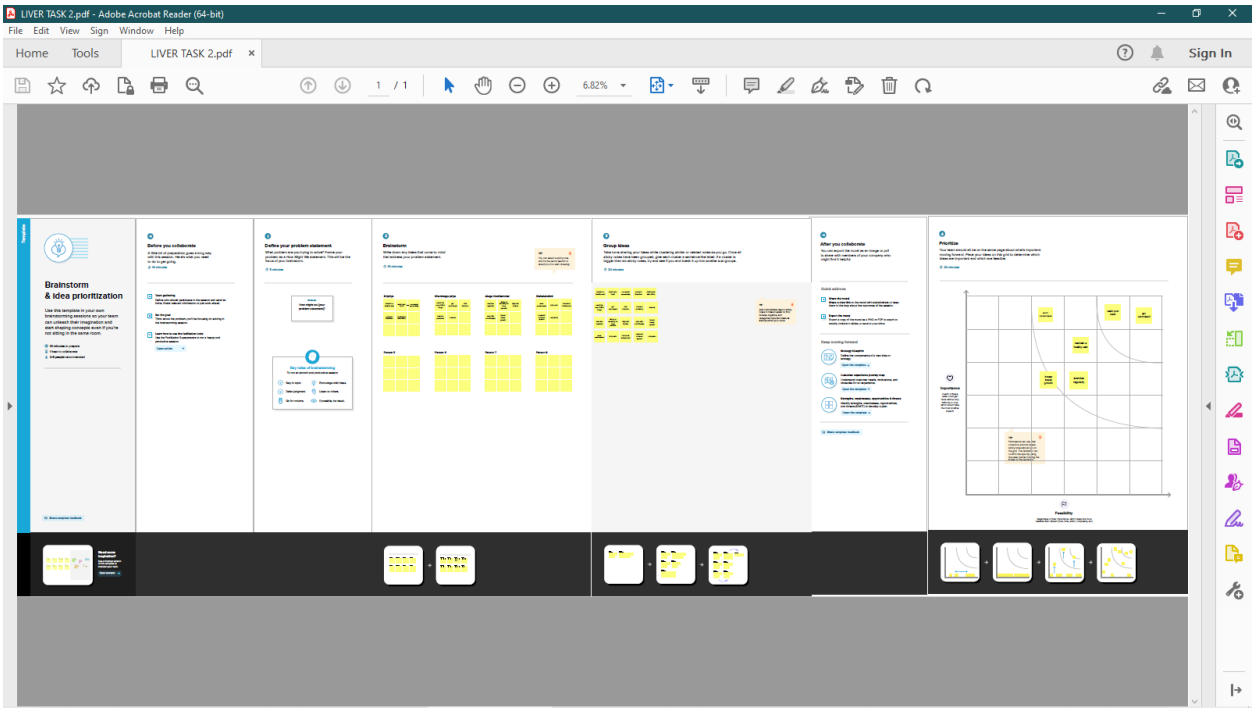
The liver is essential for digesting food and ridding your body of toxic substances. Liver disease can be inherited (genetic). Liver problems can also be caused by a variety of factors that damage the liver, such as viruses, alcohol use and obesity. It helps your body digest food, store energy, and remove poisons. There are many kinds of liver diseases: Diseases caused by viruses, such as hepatitis A, hepatitis B, and hepatitis C. Diseases caused by drugs, poisons, or too much alcohol.

It regulates blood sugar, makes vitamins, maintains blood at the right thickness and keeps muscles from rumoring. It filters toxins from the blood and clears medications and alcohol. When your muscles grow with regular exercise, your liver makes it possible. It's very important to heart health. The liver must process everything that the stomach and intestines digest and absorb. For nutrition, one of its main jobs is to produce bile, a chemical that converts fat into fuel the body can use. Without your liver, eating a cheeseburger and milkshake would make you very sick.

While the liver is responsible for cleaning toxins from the blood, overexposure to toxins can be harmful. Read warning labels on chemicals you use around the house, and wash fruits and vegetables before consumption to ensure you’re not digesting pesticides.

2 Problem Definition & Design Thinking

2.1 Brainstorm map



Maintain a healthy weight**.** Obesity can cause nonalcoholic fatty liver disease

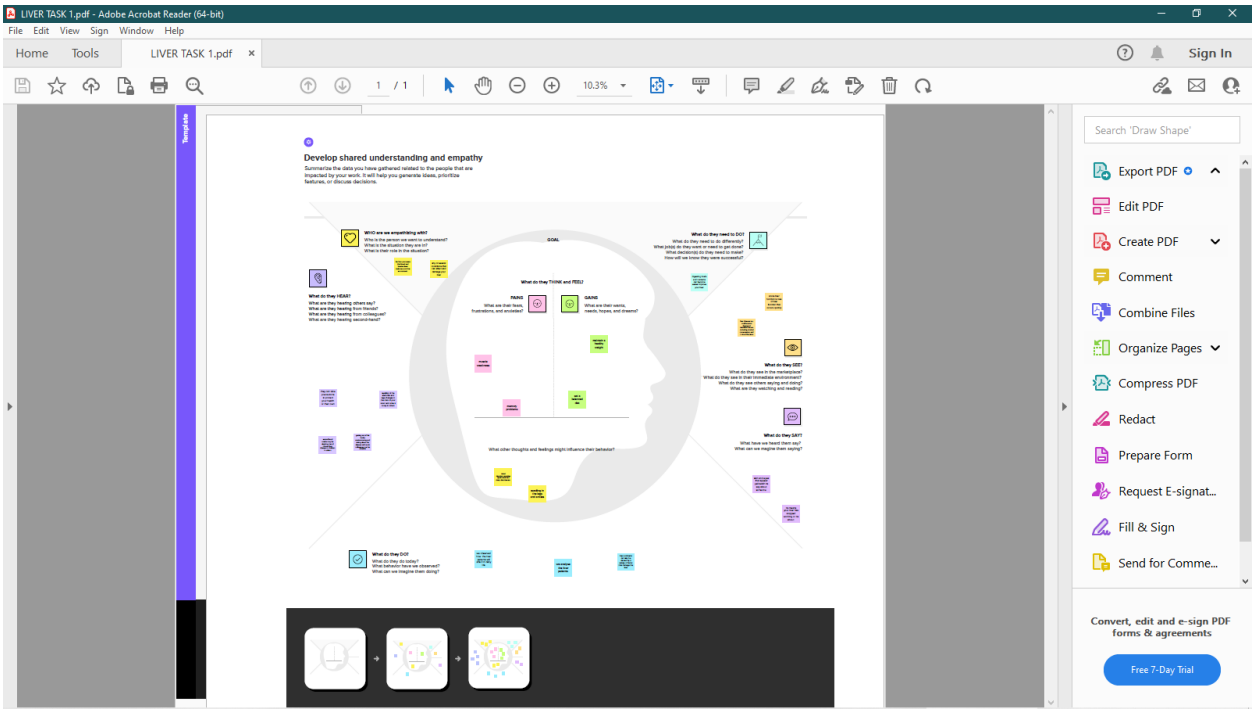
Hormonal imbalance can affect people with liver conditions, where symptoms can include a loss of libido, fatigue and mood symptoms.

Stomach Disorders Pain in your liver itself can feel like a dull throbbing pain or a stabbing sensation in your right upper abdomen just under your ribs.

High salt index It is well known that consuming too much salt is linked to high blood pressure. Now, a new study suggests it may also lead to liver damage in adults and developing embryos.

Regular exercise is key to a healthy liver. Exercise decreases stress on the liver, increases energy levels and helps to prevent obesity – a risk factor for liver disease. Aim for a total of 150 minutes of exercise, such as brisk walking or swimming per week

2.2 Empathy Map



3 ADVANTAGES & DISADVANTAGES

ADVANTAGES:

Maintain a healthy weight**.** Obesity can cause nonalcoholic fatty liver disease.

Get vaccinated**.** If you're at increased risk of contracting hepatitis or if you've already been infected with any form of the hepatitis virus, talk to your doctor about getting the hepatitis A and hepatitis B vaccines.

Keep your food safe**.** Wash your hands thoroughly before eating or preparing foods. If traveling in a developing country, use bottled water to drink, wash your hands and brush your teeth.

Use medications wisely**.** Take prescription and nonprescription drugs only when needed and only in recommended doses. Don't mix medications and alcohol. Talk to your doctor before mixing herbal supplements or prescription or nonprescription drugs.

Avoid risky behavior**.** Use a condom during sex. If you choose to have tattoos or body piercings, be picky about cleanliness and safety when selecting a shop. Seek help if you use illicit intravenous drugs, and don't share needles to inject drugs.

DISADVANTAGES:

- High invasive test
- The Potential complication include Death
- Significant Sampling error
- High cost Inter-observer
- It's possible, and dangerous, to get too much vitamin A. Eating large amounts of Liver can lead to symptoms of vitamin A toxicity, which happens when your own liver can't process the excess vitamin A quickly enough.
- Acute liver failure, also known as hepatic failure, can cause serious complications, including bleeding and increased pressure in the brain. It's a medical emergency that requires hospitalization.
- Pain and distention of the abdomen due to release of fluid from the liver Confusion or forget fullness. When the liver isn't functioning properly, Toxins build up in the blood and can travel to the brain, affecting brain function
- Pale colored stool
- Chronic fatigue
- Jaundice or yellowing of the eyes or skin

4 APPLICATION:

The area offering the most exciting new applications in healthcare is ML. Many studies in recent years have suggested that ML technology has many potential uses in heap topology, ranging from exploring new noninvasive means to predict or diagnose different liver diseases to automated image analysis. From the identification of liver areas at risk of radiation toxicity to the use of drug structures to predict the risk of liver injury, the accuracy of diagnosis and the effectiveness of treatment can be improved, and the efficiency can also be improved through

automation. Although promising data from preclinical studies are now available, the application of AI in liver disease is far from being applied in clinical practice, so the application of AI in liver disease and other diseases remains challenging and deserves further study.

Liver disease is not an independent disease. Because the specific types of lesions are different, the diagnostic methods disease.

Different examination methods can be selected according to the specific types of liver diseases to be examined. For example, at present, the common diagnostic method for nonalcoholic fatty liver disease (NAFLD) is liver ultrasound (US), the common diagnostic method for liver fibrosis is liver biopsy; the diagnosis of liver cancer (LC) mainly uses imaging images and biomarkers, and the staging mainly uses the Barcelona staging system. However, due to subjective and invasive factors, the current examination methods have certain limitations in the diagnosis of some liver diseases. The sensitivity and specificity of liver US decrease with increasing body mass index because US is subjective. As a solid tumor, heap to cellular carcinoma (HCC) has significant temporal and spatial heterogeneity, which can predict the treatment response and prognosis of HCC. The Barcelona staging system does not include the histological and molecular characteristics of tumors. The application of AI has filled the gaps in these respects. By designing noninvasive examination means to intelligently analyze images and pictures, AI has improved the diagnostic efficiency.

Liver fibrosis, regardless of the etiology, is believed to be key to the progression of any form of chronic liver disease (CLD), and persistent fibrosis is widely believed to be a major driver of the eventual development of cirrhosis and liver failure. Liver biopsy is considered to be the gold standard for staging liver fibrosis; however, it is invasive and is limited by sample error, inter observer variability and various potential complications. Radiological and serum markers of fibrosis are also used to assess liver fibrosis, and it is not reliable to accurately distinguish the stages of fibrosis in these patterns. There is a clear need for safe, effective and reliable noninvasive assessment modalities. A study that aimed to develop and validate a deep learning system (DLS) for staging liver fibrosis by using portal venous phase CT images demonstrated that a DLS trained by using a large amount of CT data allowed for highly accurate staging of liver fibrosis. In this study, DLS was superior to radiologists and serum fibrosis tests in diagnosing significant fibrosis, advanced fibrosis and cirrhosis. In addition, an existing model called deep learning radiomics of

elastography has shown the best overall performance in predicting liver fibrosis stage, which has certain value and practical value for the accurate noninvasive diagnosis of liver fibrosis stage in hepatitis-B-virus-infected patients.

5 CONCLUSION

Initially, the data set was studied and prepared for inclusion in the classifier. This was achieved by removing some rows containing zero values ,modifying some columns indicating the skew ness , and using appropriate techniques (a hot Coding) to make the labels more useful for classification purposes. The performance indicators for which the models will be evaluated have been resolved. The data set was then divided into a reading and testing package.

First, a simple predictive and base model ("Logistic Regression") was Developed in the data set to determine the value of the base accuracy. The biggest challenge in implementing this project was in two areas: defining learning algorithms and selecting the appropriate parameters for precise configuration. Initially, making a decision on 3 or 4 methods out of the many choices available at sklearn was very tedious.

AI has become an important part of liver disease research, improving diagnostic accuracy, improving decision-making by enhancing predictive power, increasing efficiency through automation, and even predicting liver disease prognosis. Analysis of key biomarkers using ML can also provide deeper insights into the path physiology of liver disease. Despite the challenges, the application of AI in the field of liver disease is promising and worthy of further study. Researchers need to further develop new models of AI in liver disease diagnosis and precise treatment and conduct clinical verification to improve the accuracy of the results and promote the clinical application of AI. However, we must also be wary of over-reliance on such algorithms. AI will support rather than replace doctors, although computers and healthcare workers will have to work together. Ultimately, healthcare workers will have to make decisions for their patients based on their preferences, circumstances and ethics

6 FUTURE SCOPE

This is a procedure that helps diagnose and treat problems in the liver, gallbladder, bile ducts, and pancreas. It uses X-rays and a long, flexible, lighted tube (endoscope).

The scope is put into your mouth and throat. Then it is guided down into your esophagus, stomach, and the first part of the intestines (duodenum). The provider can examine the inside of these organs and find any problems. A tube is then passed through the scope. The tube is put into the duct that drains the gallbladder and pancreas into the duodenum. A dye is injected to better see those organs on an X-ray.

CT scan. This is a test that uses X-rays and a computer to make detailed images of the body. A CT scan shows details of the bones, muscles, fat, and organs. CT scans are more detailed than general X-rays.

Liver biopsy. Tissue samples from the liver are removed with a needle or during surgery. They are checked under a microscope.

7 APPENDIX

```
import pandas
as pdimport
numpy as np
import seaborn as sns
import matplotlib.pyplot as
plt from  matplotlib  import
rcParams from  scipy  import
stats
from  google.colab  import
files
uploaded=files.upload()
```

Choose Files indian_liver.csv

- **indian_liver.csv**(text/csv) - 20725 bytes, last modified: 6/4/2023 - 100% done

```
Saving indian_liver.csv to indian_liver.csv
```

```
data=pd.read_csv('indian_liver.c
sv')data.head()
```

	Age	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_
0	65	0.7	0.1	187	
1	62	10.9	5.5	699	
2	62	7.3	4.1	490	
3	58	1.0	0.4	182	
4	72	3.9	2.0	195	

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>RangeIndex: 583 entries, 0 to 582
Data columns (total 10 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Age                         583 non-null    int64
 1   Total_Bilirubin             583 non-null    float64
 2   Direct_Bilirubin            583 non-null    float64
 3   Alkaline_Phosphotase        583 non-null    int64
 4   Alamine_Aminotransferase    583 non-null    int64
 5   Aspartate_Aminotransferase  583 non-null    int64
 6   Total_Protiens              583 non-null    float64
 7   Albumin                     583 non-null    float64
 8   Albumin_and_Globulin_Ratio  579 non-null    float64
 9   Dataset                     583 non-null    int64
dtypes: float64(5),
int64(5)memory usage:
45.7 KB
```

```
data.isnull().any()
```

```
Age                         False
Total_Bilirubin             False
Direct_Bilirubin            False
Alkaline_Phosphotase      False
Alamine_Aminotransferase  False
Aspartate_Aminotransferase
False            Total_Protiens
                            F
alse
Albumin                     False
Albumin_and_Globulin_Ratio  True
Dataset
                            F
alse
dtype: bool
```

```
data.isnull().sum()
```

```
Age                         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  4
Dataset                     0
dtype: int64
```

```
data['Albumin_and_Globulin_Ratio']=(data['Albumin_and_Globulin_Ratio'].mode()
[0])data.isnull().sum()
```

```
Age                         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
dtype: int64
```

```
from sklearn.preprocessing import
LabelEncoderlc=LabelEncoder()
#data['Gender']=lc.fit_transform(data['Gender'])
```

```
import pandas
as pdimport
numpy as np
import seaborn as sns
import matplotlib.pyplot as
plt from  matplotlib  import
rcParams from  scipy  import
stats
#from  google.colab  import
files
#uploaded=files.upload()
```

```
data=pd.read_csv('indian_liver.c
sv')data.describe()
```

	Age	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase
count	583.000000	583.000000	583.000000	583.000000
mean	44.746141	3.298799	1.486106	290.576329
std	16.189833	6.209522	2.808498	242.937989
min	4.000000	0.400000	0.100000	63.000000
25%	33.000000	0.800000	0.200000	175.500000
50%	45.000000	1.000000	0.300000	208.000000
75%	58.000000	2.600000	1.300000	298.000000
max	90.000000	75.000000	19.700000	2110.000000

```
sns.distplot(data['Age'])
plt.title('Age Distribution
Graph')plt.show()
```

```
<ipython-input-10-a9533a3b6a8d>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.
```

```
Please adapt your code to use either `displot` (a figure-level
function            similar   flexibility) or `histplot` (an axes-level
```

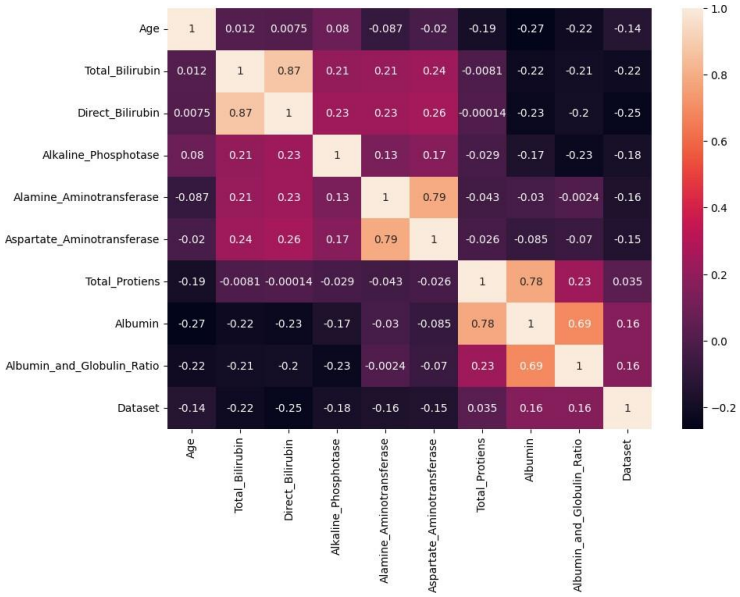
```
sns.countplot(data['Albumin'])
```

```
<Axes: ylabel='count'>
```


```
function for histograms
```

```
plt.figure(figsize=(10,7))
sns.heatmap(data.corr(),annot=True)
```

```
<Axes: >
```



```
X=
da
ta
.i
lo
c[
:,
:-
1]
y=
da
ta
.A
lb
um
in
```

```
from sklearn.preprocessing import scale
X_scaled=pd.DataFrame(scale
(X),columns=X.columns)
X_scaled.head()
```

	Age	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alam
0	1.252098	-0.418878	-0.493964	-0.426715	
1	1.066637	1.225171	1.430423	1.682629	
2	1.066637	0.644919	0.931508	0.821588	
3	0.819356	-0.370523	-0.387054	-0.447314	
4	1.684839	0.096902	0.183135	-0.393756	

```
X
=
d
a
t
a
.
i
l
o
c
[
:
,
:
-
1
]

y
=
d
a
t
a
```

.
A
l
b
u
m
i
n

```
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)
```

	Age	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alam
0	1.252098	-0.418878	-0.493964	-0.426715	
1	1.066637	1.225171	1.430423	1.682629	
2	1.066637	0.644919	0.931508	0.821588	
3	0.819356	-0.370523	-0.387054	-0.447314	
4	1.684839	0.096902	0.183135	-0.393756	

```
pip install imblearn
```

```
Looking in indexes: https://pypi.org/simple, https://us-
python.pkg.dev/colab-wheels/public/simple/Collecting imblearn
  Downloading imblearn-0.0-py2.py3-none-any.whl (1.9 kB)
Requirement already satisfied: imbalanced-learn in /usr/local/lib/python3.9/dist-packages (from imblearn)
(0.10.1)
Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.9/dist-packages (from imbalanced-
learn->imblearn) (1.1.1) Requirement already satisfied: numpy>=1.17.3 in /usr/local/lib/python3.9/dist-
packages (from imbalanced-learn->imblearn) (1.22.4)Requirement already satisfied: scipy>=1.3.2 in
/usr/local/lib/python3.9/dist-packages (from imbalanced-learn->imblearn) (1.10.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.9/dist-packages (from
imbalanced-learn->imblearn) (3 Requirement already satisfied: scikit-learn>=1.0.2 in
/usr/local/lib/python3.9/dist-packages (from imbalanced-learn->imblearn) (1. Installing collected packages:
imblearn
Successfully installed imblearn-0.0
```

```
from
imblearn.over_samp
ling import SMOTE
smote=SMOTE()
y_train.value_counts()
```

```
3.0    31
4.0    30
3.1    27
2.9    22
2.7    22
2.5    21
3.9    21
3.3    18
2.6    18
3.2    18
3.4    17
3.7    17
3.5    17
2.0    15
2.4    15
3.6    15
```

```
 3.8    13
 2.8    13

4.1     12
 4.3    12
 1.8    10
 4.2    10
 2.1    10
 2.2     9
 4.4     8
 1.9     7
 2.3     7
 1.6     6
 4.9     4
 4.5     4
 4.6     3
 1.7     3
 1.4     2
 4.7     2
 0.9     2
 5.5     2
 4.8     1
 5.0     1
 1.5     1
 Name: Albumin, dtype: int64
```

```
from sklearn.ensemple import
RandomForestClassifier
model1=RandomForestClassifier
()
model1.fit(X_train
_smote,
y_train_smote)
y_predict=model1.p
redict(X_test)
rfc1=accuracy_s
core(y_test,y_p
redict)rfc1
pd.crosstab(y_test,y_predict)
print(classification_report(y_test,y_predict))
```

```
from sklearn.tree import
DecisionTreeClassifier
model4=DecisionTreeClassi
fier()
model4.fit(X_trai
n_smote,y_train_s
mote)
y_predict=model4.
predict(X_test)
dtc1=accuracy_s
core(y_test,y_p
redict)dtc1
pd.crosstab(y_test,y_predict)
print(classification_report(y_test,y_predict))
```

```
from sklearn.neighbors
import KNeighborsClassifier
model2=KNeighborsClassifier(
)
model2.fit(X_trai
n_smote,y_train_s
mote)
y_predict=model2.
predict(X_test)
knn1=(accuracy_sc
ore(y_test,y_pred
ict))knn1
```

```
pd.crosstab(y_test,y_predict)
print(classification_report(y_test,y_predict))


from sklearn.linear_model
import LogisticRegression
model5=LogisticRegression()
model5.fit(X_trai
n_smote,y_train_s
mote)
y_predict=model5.
predict(X_test)
logi1=accuracy_s
core(y_test,y_pr
edict)logi1
pd.crosstab(y_test,y_predict)
print(classification_report(y_test,y_predict))


import tensorflow.keras
from
tensorflow.keras.models
import Sequentialfrom
tensorflow.keras.layers
import Dense


classifier=Sequential()
classifier.add(Dense(units=100,activation
='relu',input_dim=10))
classifier.add(Dens(units=50,activation='
relu'))
classifier.add(Dens(units=1,activation='sigmoid'))
classifier.compile(optimizer='adam',loss='binary_crossentropy',metrics=['accuracy'])


model4.predict([[50,1,1.2,0.8,150,70,80,7.2,3.4,0.8]])
model1.predict([[50,1,1.2,0.8,150,70,80,2.2,3.4,0.0]])
classifier.save("liver.h5")
y_pred=clas
sifier.pred
ict(X_test)
y_pred
y
_
p
r
e
d
=
(
y
_
p
r
e
d
>
0
.
5
)

y
_
p
r
e
d
```

```
def predict_exit(sample_value):
  sample_value=np.array(sample_value)
  sample_value=samp
  le_value.reshape(
  1,-1)
  sample_value=scal
  e(sample_value)
  return classifier.predict(sample_value)
sample_value=[[50,1,1.2,0.8,150,70,80,7.2,3.4,0.8]]
if predict_exit(sample_value)>0.5:
  print('Predi
ction: Liver
Patient')else:
  print('Prediction: Healthy')


acc_smote=[['KNN
Classifier',knn1],['RandomForestClassifier',rfc1],['DecisionTreeClassifier',dtc1],['LogisticRegression',lo
gi1]]LiverPatient_pred=pd.DataFrame(acc_smote,columns=['classification models','accuracy_score'])
LiverPatient_pred


plt
.fi
gur
e(f
igs
ize
=(7
,5)
)
plt
.xt
ick
s(r
ota
tio
n=9
0)
plt.title(X='classification models & accuracy scores after SMOTE',fontsize=18)
sns.barplot(X="classification
models",y="accuracy_score",data=LiverPatient_pred,palette="Set2")from
sklearn.ensemble import ExtraTreesClassifier
model.
ExtraT
reesCl
assifi
er()
model.
fit(X,
y)
ExtraTreesClassifier()
model.feature_importance

    array([0.1205029, 0.02063187, 0.10625368,
                       0.10648578, 0.1245292,
 0.11049943, 0.110963, 0.09033392, 0.09882431,
                                  0.09497621])
dd=pd.DataFrame(model.feature_importance_,index=X.columns).sort_val
ues(0.ascending=False)
dd
```