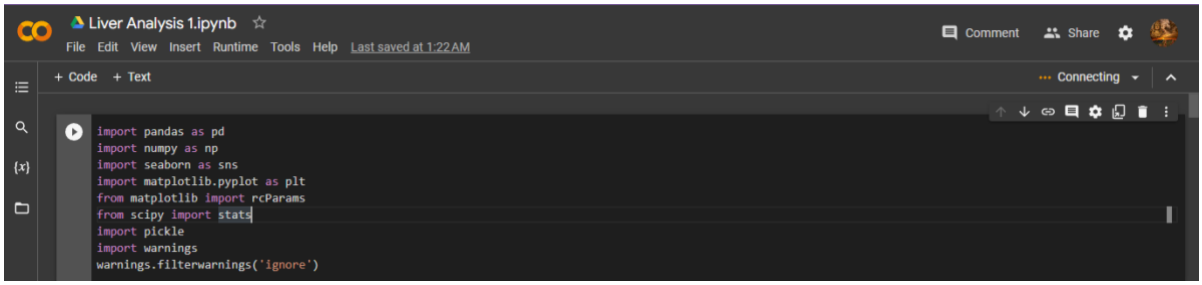
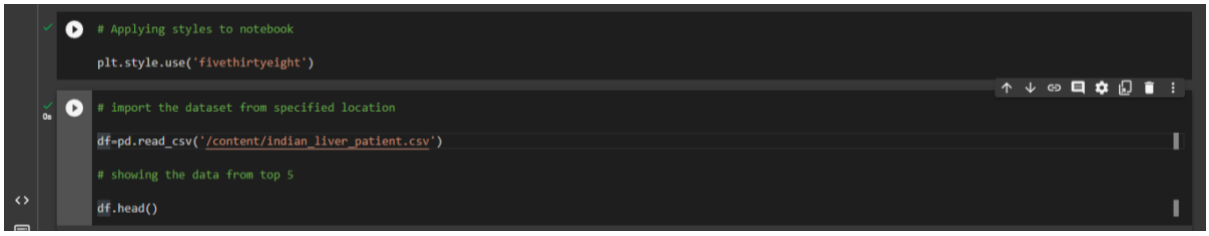
Liver Patient Analysis Methods using Machine Learning

Outputs:

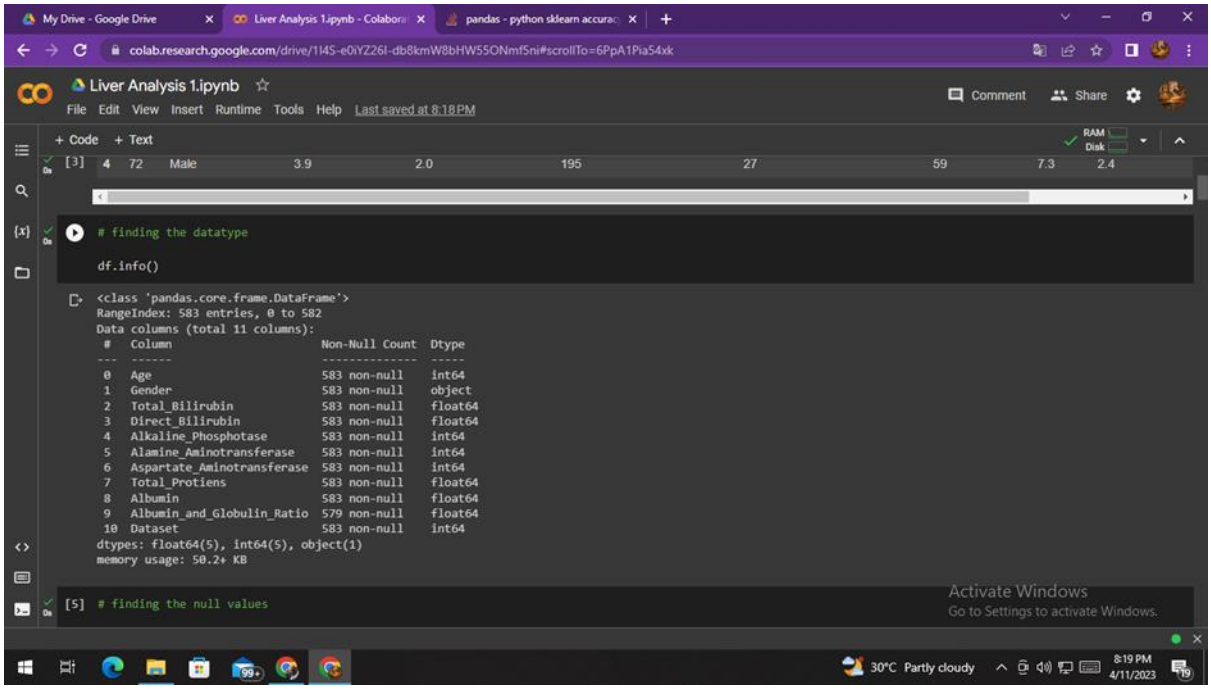
Importing required libraries:



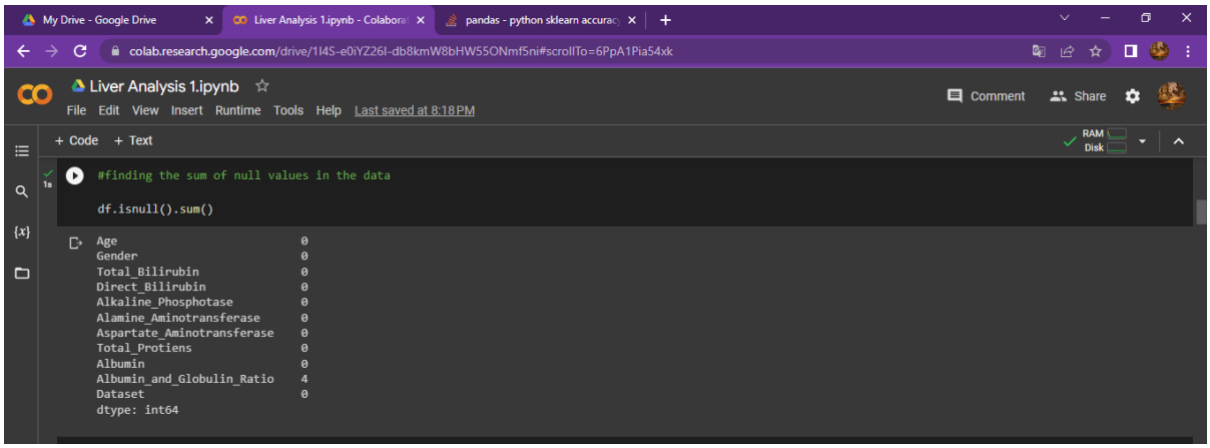
Read the csv file:



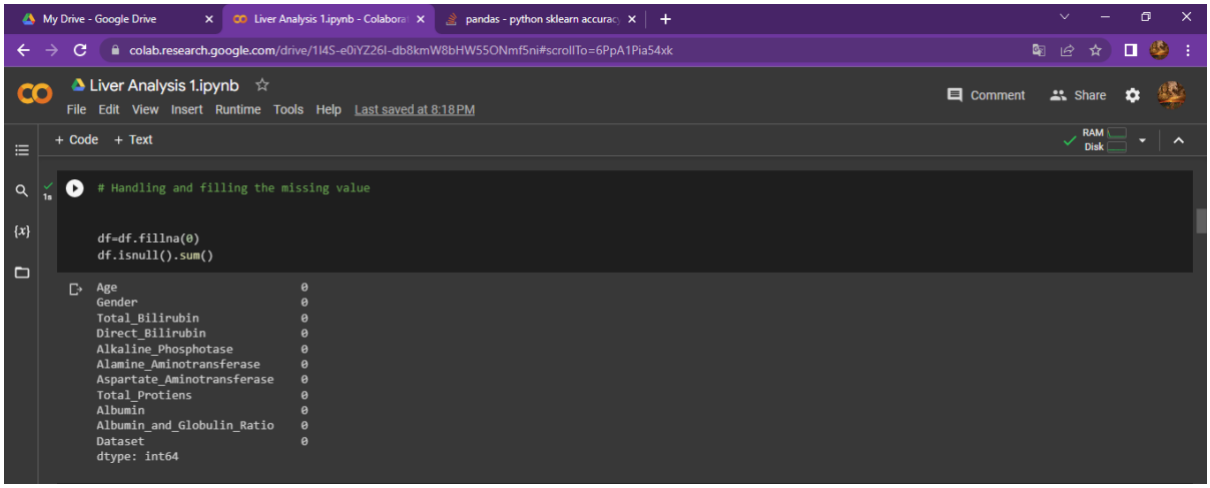
Getting the information using info() :



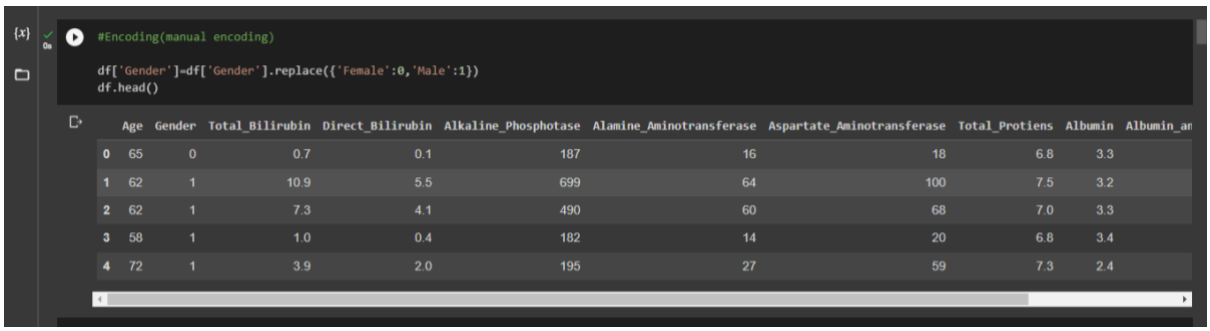
Finding sum of Null Values isnull().sum() function:



Handling and Filling the missing values:

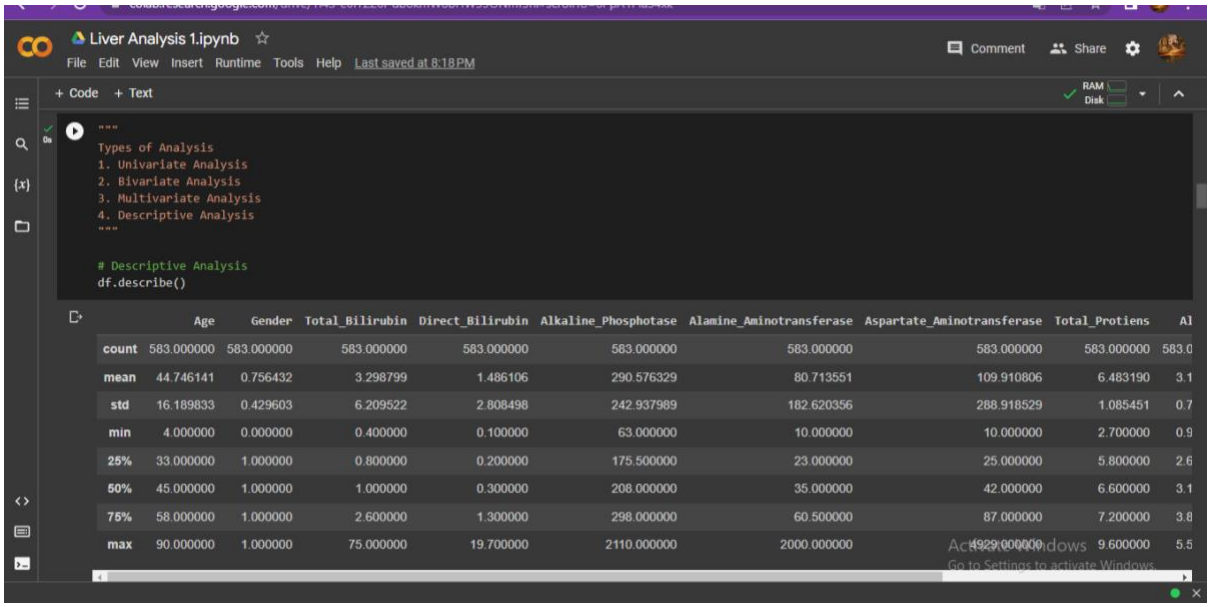


Manual Encoding:



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

TYPES OF ANALYSIS (Descriptive Analysis):



Liver Analysis 1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 8:18 PM

Comment Share

+ Code + Text

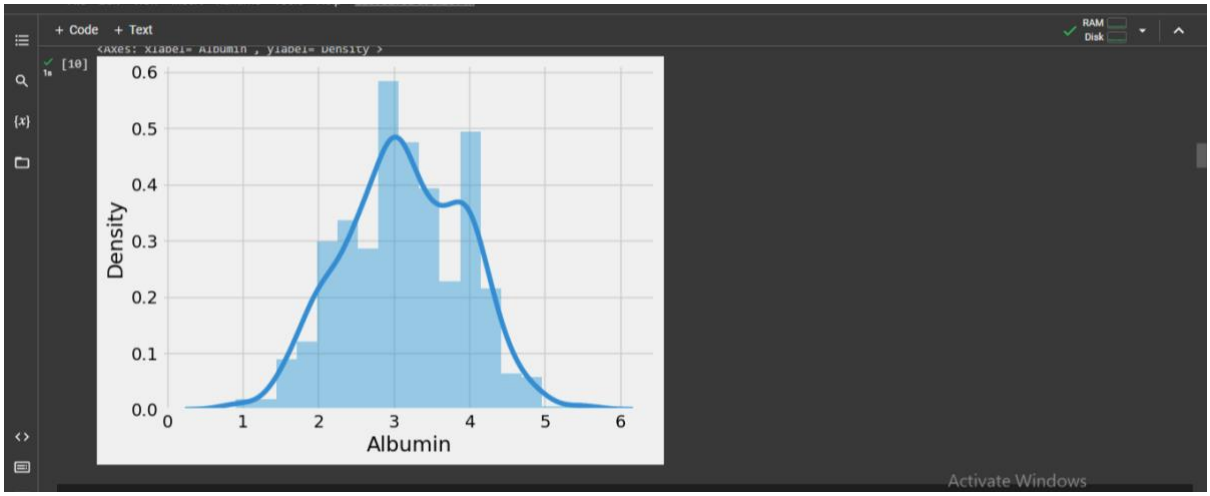
```
"""
Types of Analysis
1. Univariate Analysis
2. Bivariate Analysis
3. Multivariate Analysis
4. Descriptive Analysis
"""

# Descriptive Analysis
df.describe()
```

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	Al
count	583.000000	583.000000	583.000000	583.000000	583.000000	583.000000	583.000000	583.000000	583.0
mean	44.746141	0.756432	3.298799	1.486106	290.576329	80.713551	109.910806	6.483190	3.1
std	16.189833	0.429603	6.209522	2.808498	242.937989	182.620356	288.918529	1.085451	0.7
min	4.000000	0.000000	0.400000	0.100000	63.000000	10.000000	10.000000	2.700000	0.9
25%	33.000000	1.000000	0.800000	0.200000	175.500000	23.000000	25.000000	5.800000	2.6
50%	45.000000	1.000000	1.000000	0.300000	208.000000	35.000000	42.000000	6.600000	3.1
75%	58.000000	1.000000	2.600000	1.300000	298.000000	60.500000	87.000000	7.200000	3.8
max	90.000000	1.000000	75.000000	19.700000	2110.000000	2000.000000	4929.000000	9.600000	5.5

Activate Windows
Go to Settings to activate Windows.

i)Univariate Analysis:

- Univariate Analysis is defined as carried out on only one variable to summarize or describe the variable.



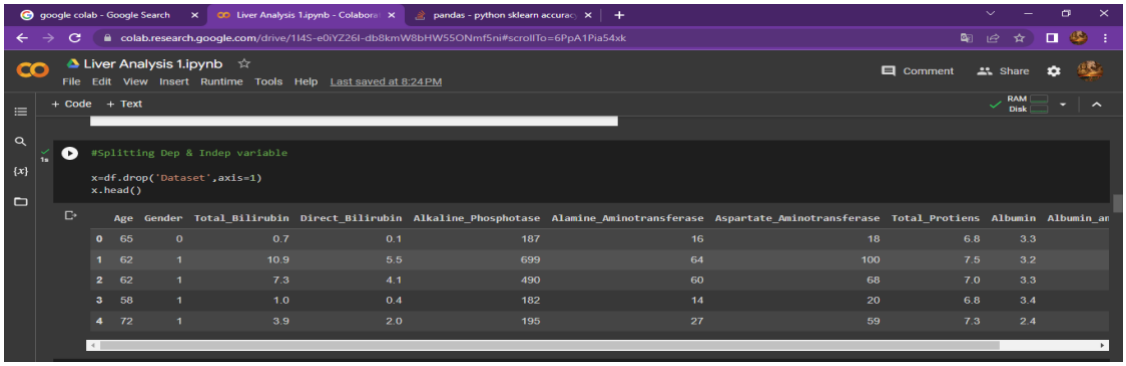
+ Code + Text

[10]

<Axes: xlabel='Albumin', ylabel='Density'>

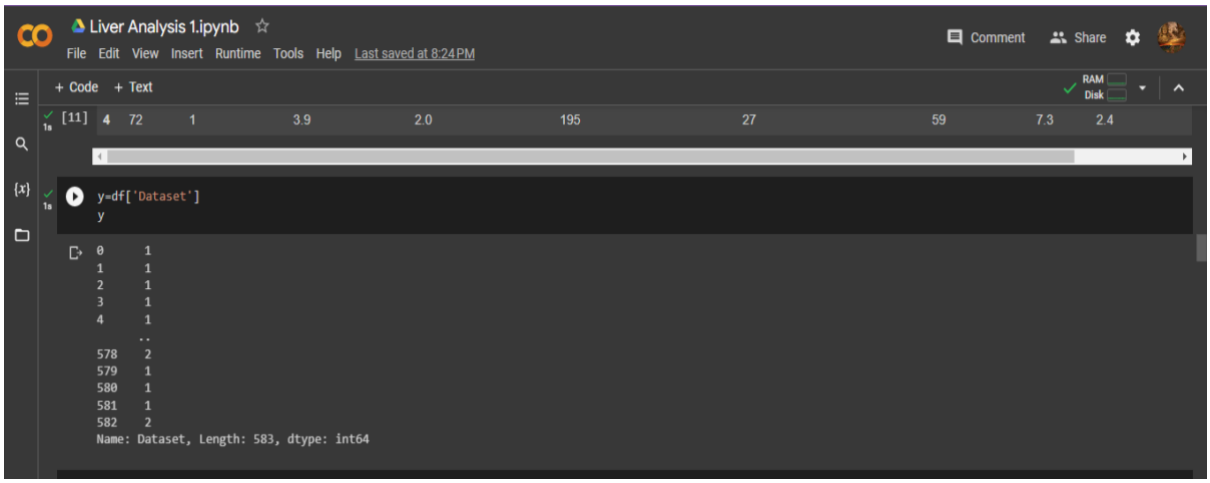
Activate Windows

Splitting the data into dependent and independent variable:



```
#Splitting Dep & Indep variable

x=df.drop('Dataset',axis=1)
x.head()
```

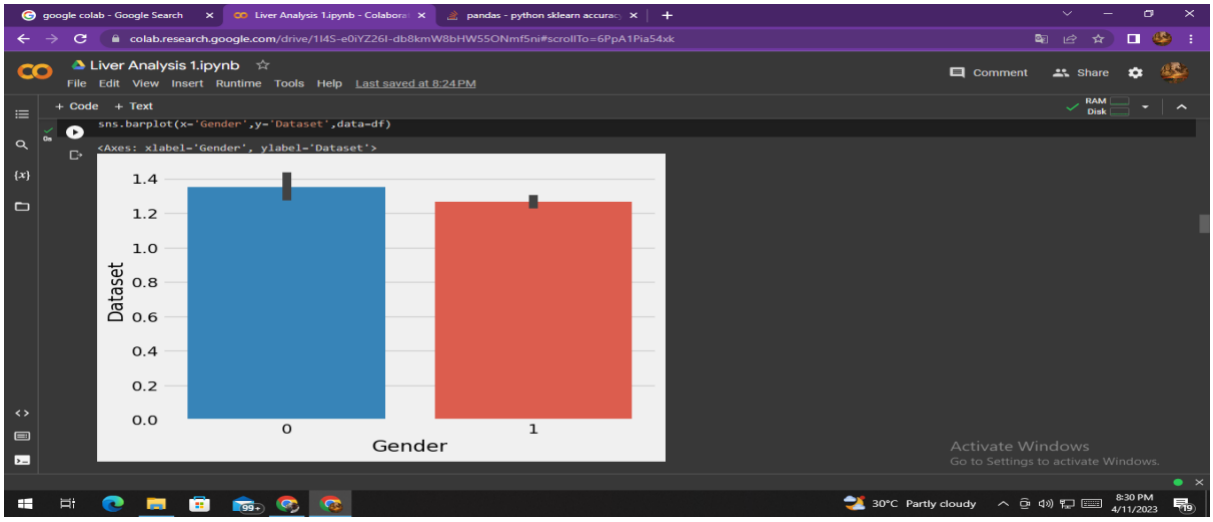
	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	Albumin	Albumin_ar
0	65	0	0.7	0.1	187	16	18	6.8	3.3	
1	62	1	10.9	5.5	699	64	100	7.5	3.2	
2	62	1	7.3	4.1	490	60	68	7.0	3.3	
3	58	1	1.0	0.4	182	14	20	6.8	3.4	
4	72	1	3.9	2.0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



```
[11]  4  72  1  3.9  2.0  195  27  59  7.3  2.4

y=df['Dataset']
y

0      1
1      1
2      1
3      1
4      1
      ..
578    2
579    1
580    1
581    1
582    2
Name: Dataset, Length: 583, dtype: int64
```

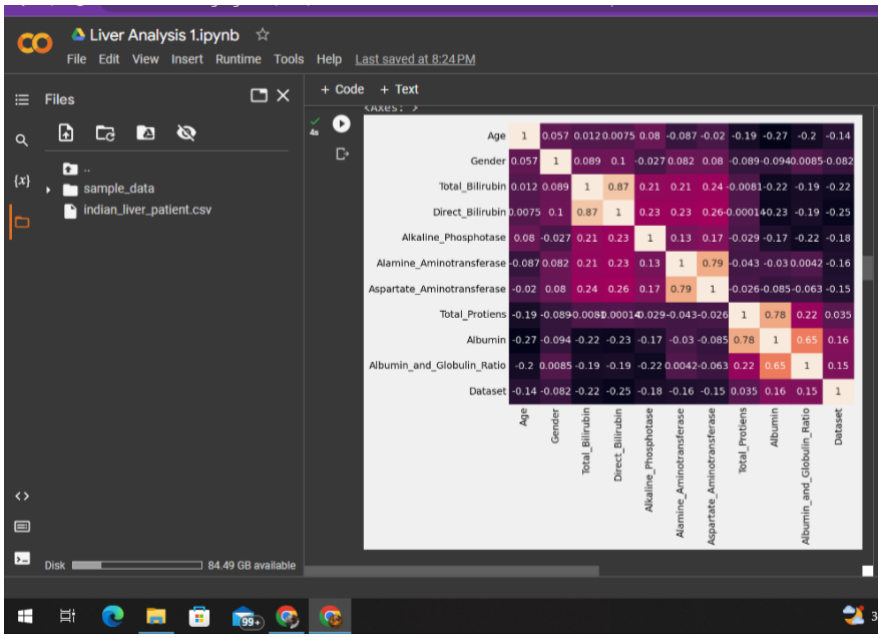
ii)Bivariate Analysis:

- Bivariate Analysis refers to the analysis of two variable to determine relationship between them.



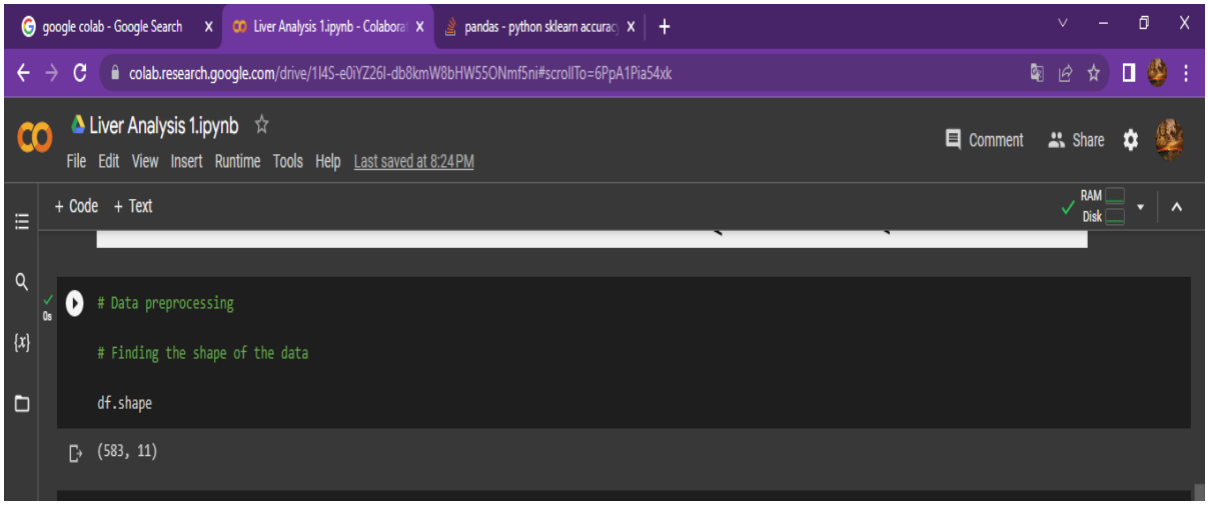
iii)Multivariate Analysis:

- Multivariate Analysis is based in observation and analysis of more than one statistical outcome variable at a time.

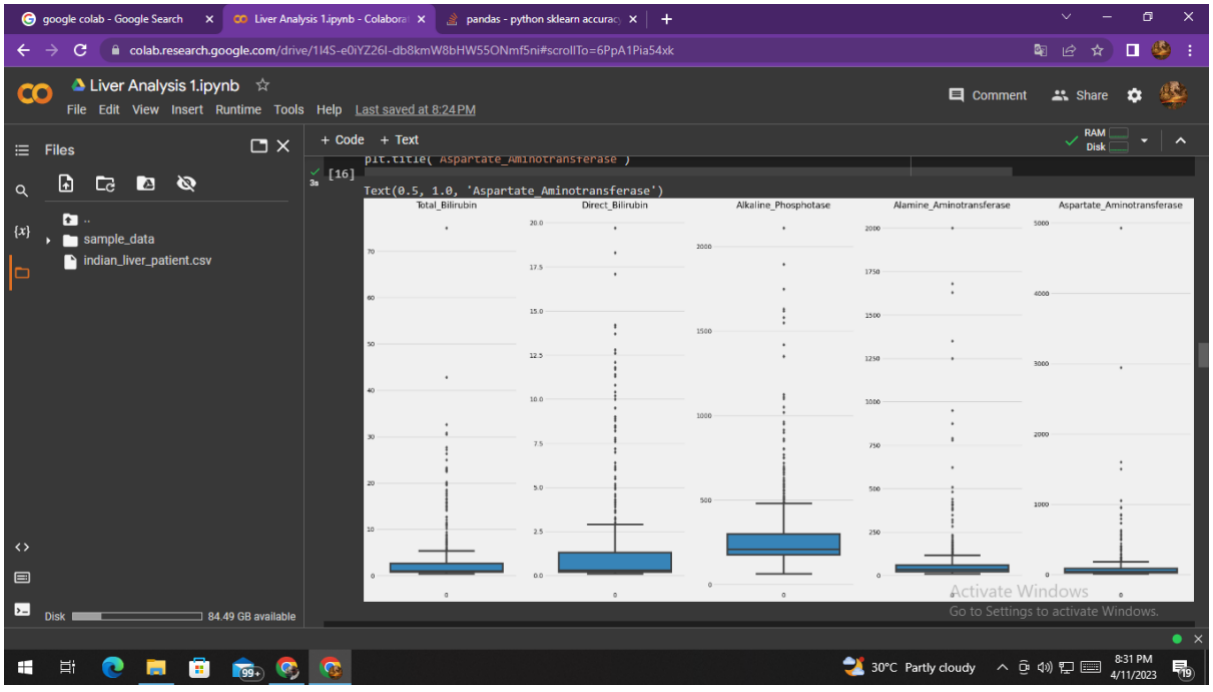


DATA PREPROCESSING:

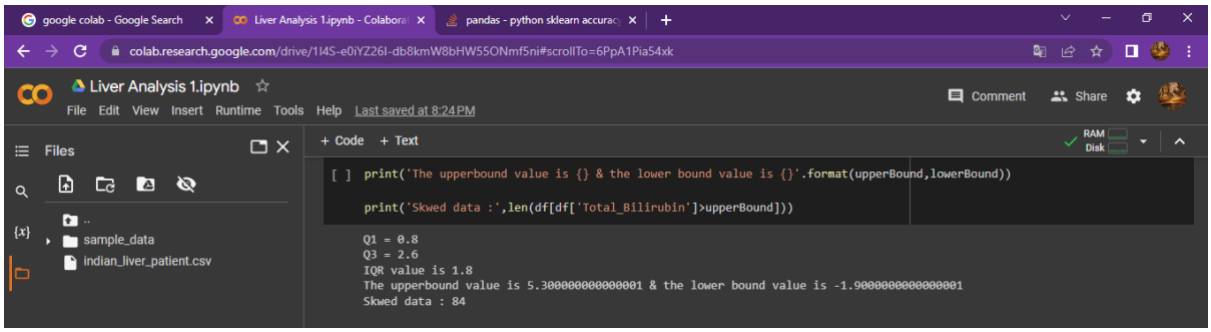
Finding the shape of the data

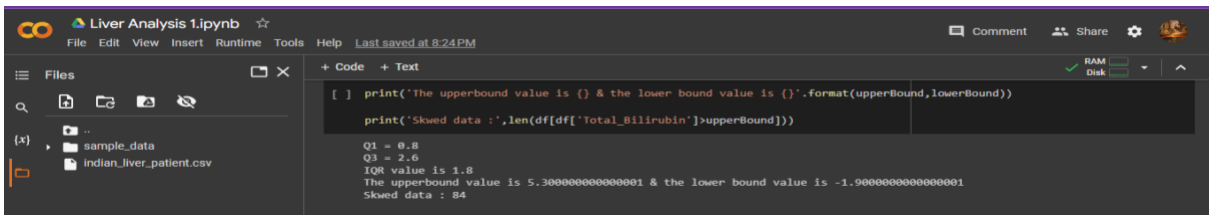


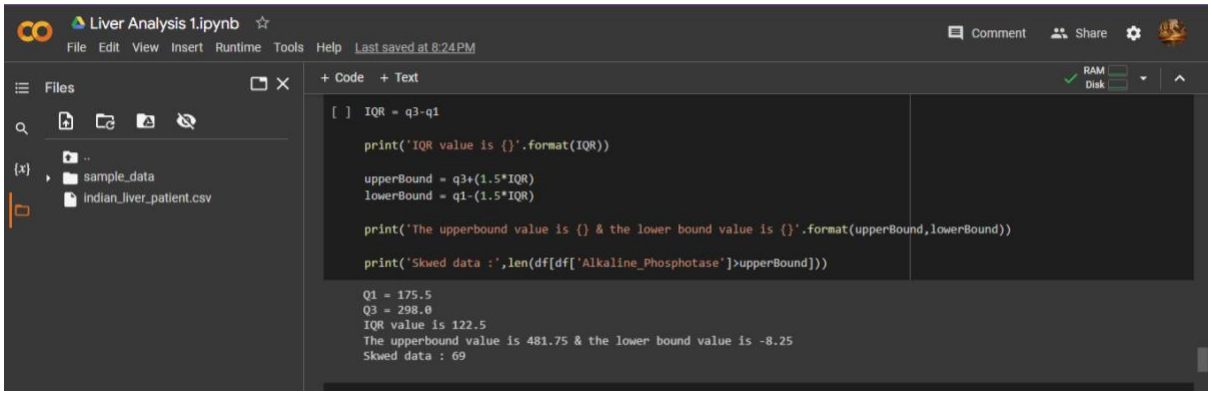
Finding outliers:

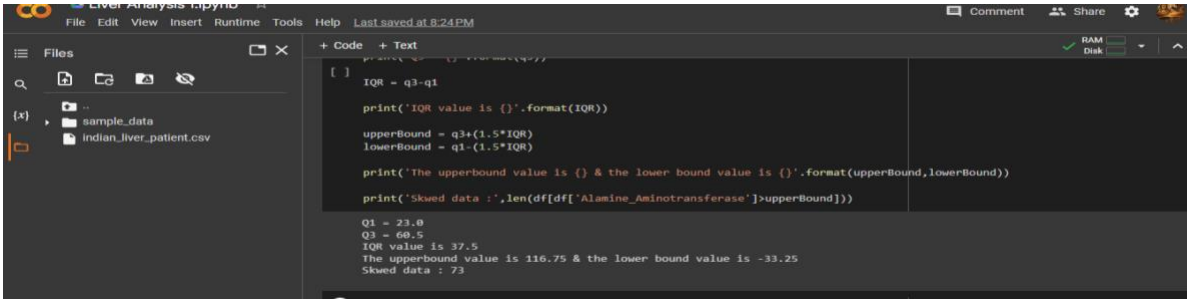


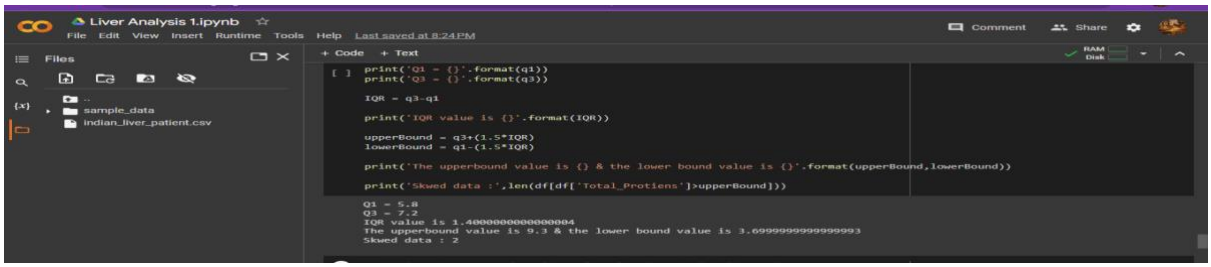
Count of Outliers:

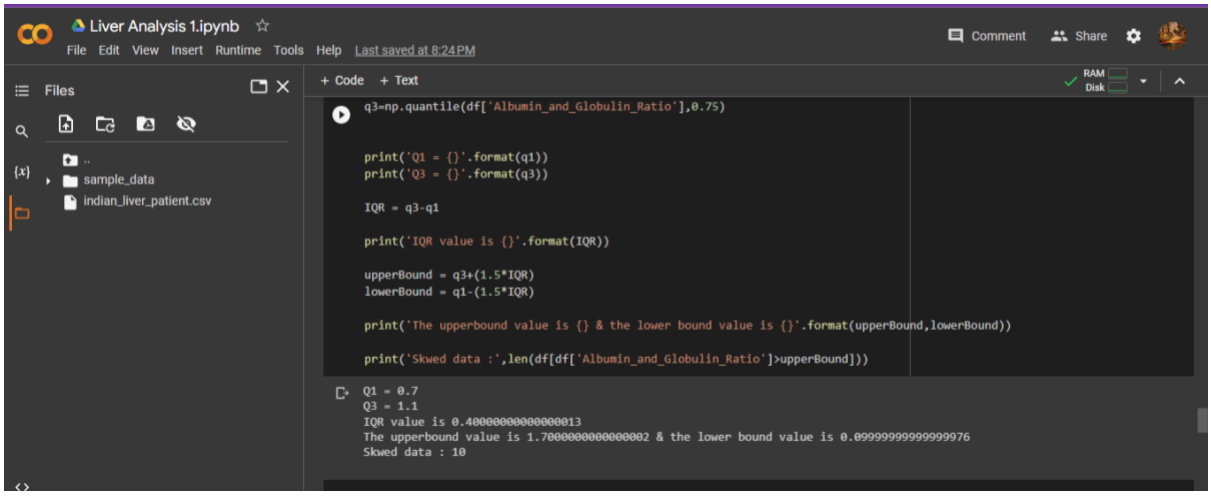


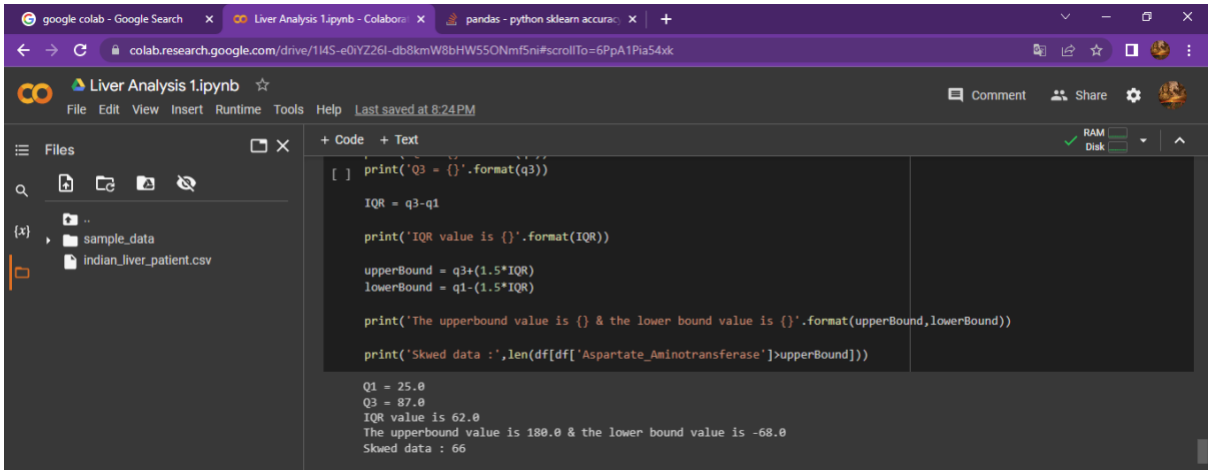



Liver Analysis 1.ipynb
File Edit View Insert Runtime Tools Help Last saved at 8:24 PM
Comment
Share
Files
+ Code
+ Text
RAM
Disk
sample_data
indian_liver_patient.csv
[ ] IQR = q3-q1
print('IQR value is {}'.format(IQR))
upperBound = q3+(1.5*IQR)
lowerBound = q1-(1.5*IQR)
print('The upperbound value is {} & the lower bound value is {}'.format(upperBound,lowerBound))
print('Skwed data :',len(df[df['Alkaline_Phosphotase']>upperBound]))
Q1 = 175.5
Q3 = 298.0
IQR value is 122.5
The upperbound value is 481.75 & the lower bound value is -8.25
Skwed data : 69

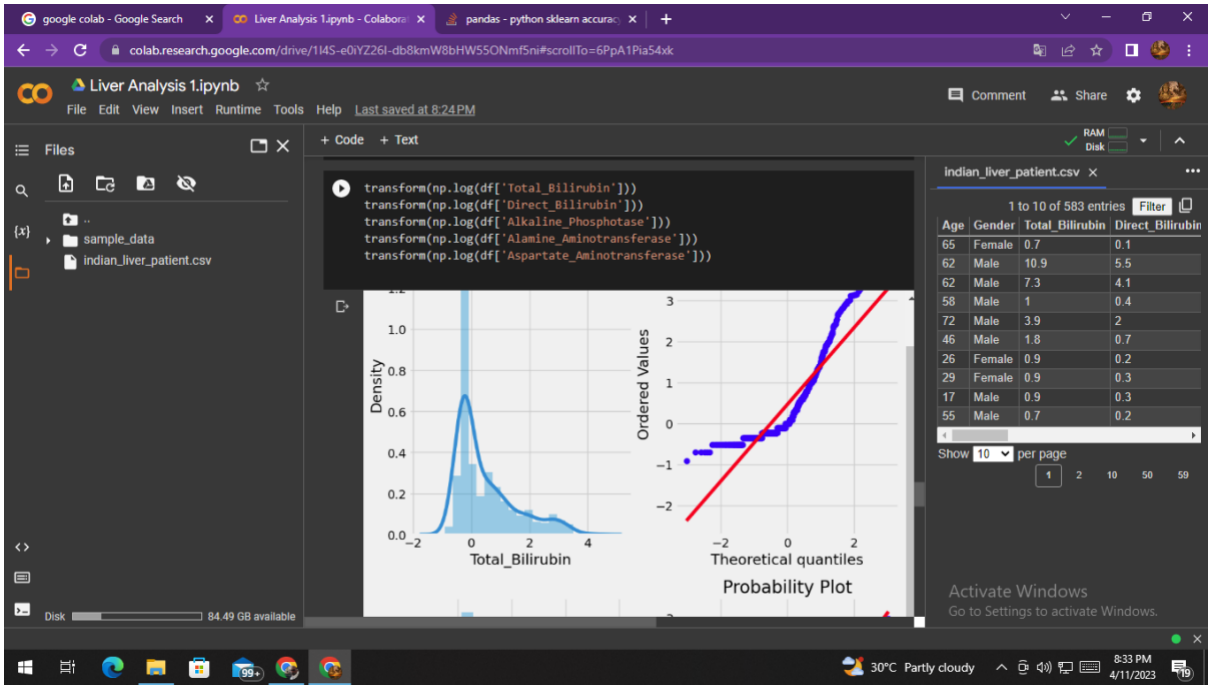

File Edit View Insert Runtime Tools Help Last saved at 8:24 PM
Comment
Share
Files
+ Code
+ Text
RAM
Disk
sample_data
indian_liver_patient.csv
[ ] IQR = q3-q1
print('IQR value is {}'.format(IQR))
upperBound = q3+(1.5*IQR)
lowerBound = q1-(1.5*IQR)
print('The upperbound value is {} & the lower bound value is {}'.format(upperBound,lowerBound))
print('Skwed data :',len(df[df['Alamine_Aminotransferase']>upperBound]))
Q1 = 23.0
Q3 = 60.5
IQR value is 37.5
The upperbound value is 116.75 & the lower bound value is -33.25
Skwed data : 73

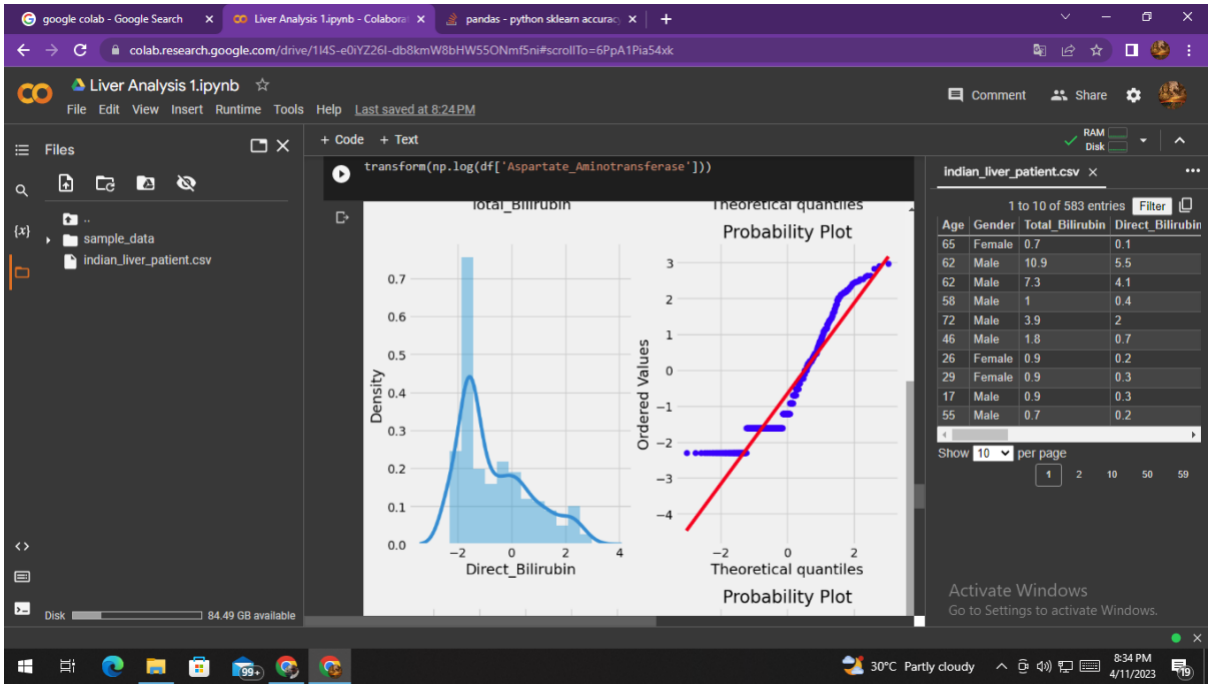

Liver Analysis 1.ipynb
File Edit View Insert Runtime Tools Help Last saved at 8:24 PM
Comment
Share
Files
+ Code
+ Text
RAM
Disk
sample_data
indian_liver_patient.csv
[ ] print('Q1 = {}'.format(q1))
print('Q3 = {}'.format(q3))
IQR = q3-q1
print('IQR value is {}'.format(IQR))
upperBound = q3+(1.5*IQR)
lowerBound = q1-(1.5*IQR)
print('The upperbound value is {} & the lower bound value is {}'.format(upperBound,lowerBound))
print('Skwed data :',len(df[df['Total_Protiens']>upperBound]))
Q1 = 5.8
Q3 = 7.2
IQR value is 1.4000000000000004
The upperbound value is 9.3 & the lower bound value is 3.6999999999999993
Skwed data : 2

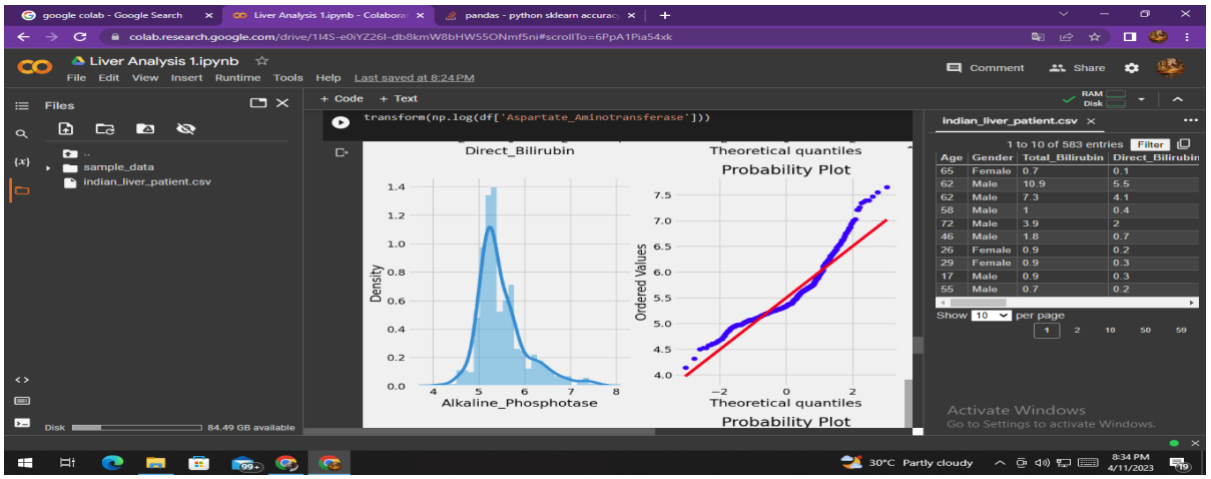

Liver Analysis 1.ipynb
File Edit View Insert Runtime Tools Help Last saved at 8:24 PM
Comment
Share
Files
+ Code
+ Text
RAM
Disk
sample_data
indian_liver_patient.csv
q3=np.quantile(df['Albumin_and_Globulin_Ratio'],0.75)
print('Q1 = {}'.format(q1))
print('Q3 = {}'.format(q3))
IQR = q3-q1
print('IQR value is {}'.format(IQR))
upperBound = q3+(1.5*IQR)
lowerBound = q1-(1.5*IQR)
print('The upperbound value is {} & the lower bound value is {}'.format(upperBound,lowerBound))
print('Skwed data :',len(df[df['Albumin_and_Globulin_Ratio']>upperBound]))
Q1 = 0.7
Q3 = 1.1
IQR value is 0.4000000000000013
The upperbound value is 1.7000000000000002 & the lower bound value is 0.09999999999999976
Skwed data : 10

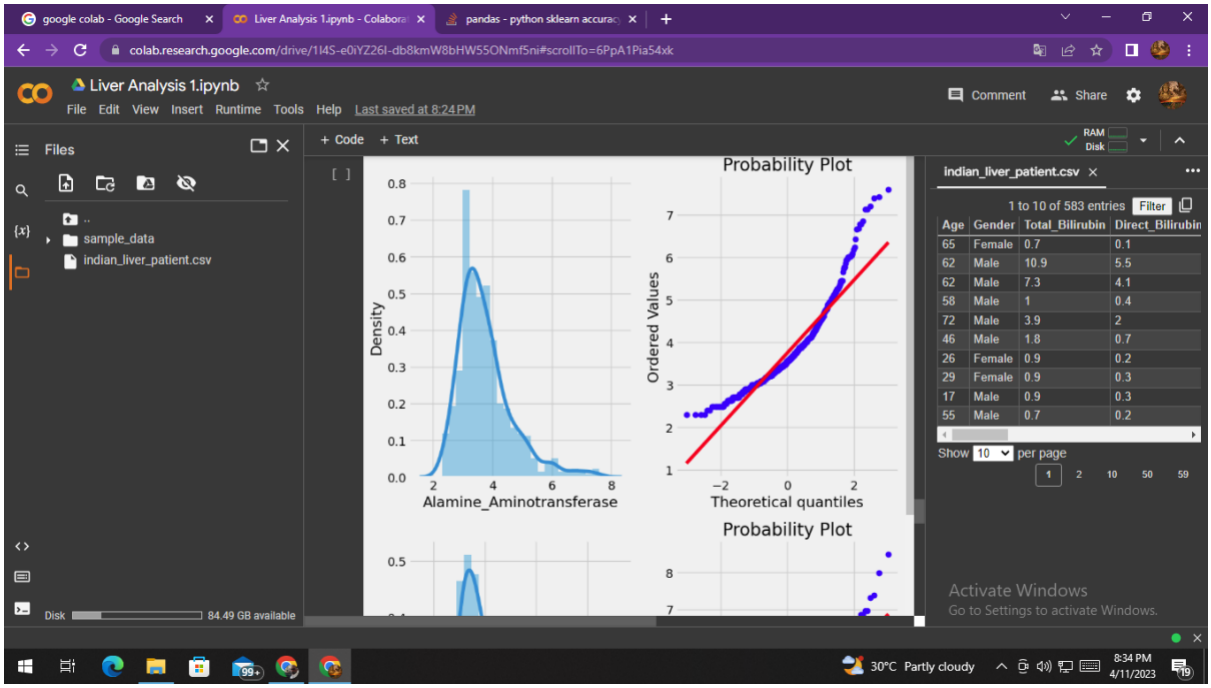


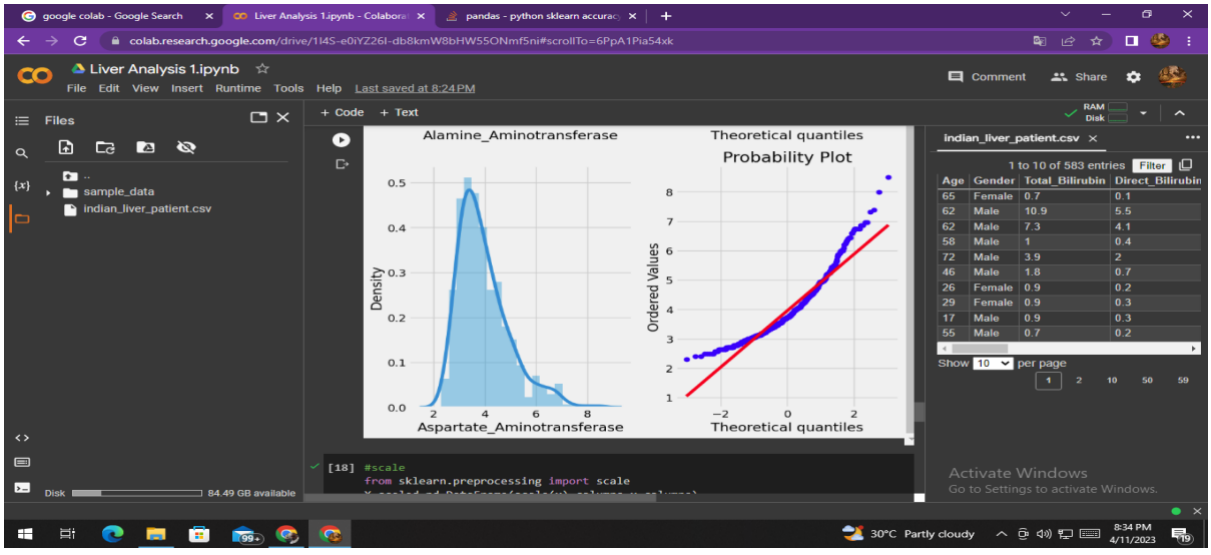
Transformation log:



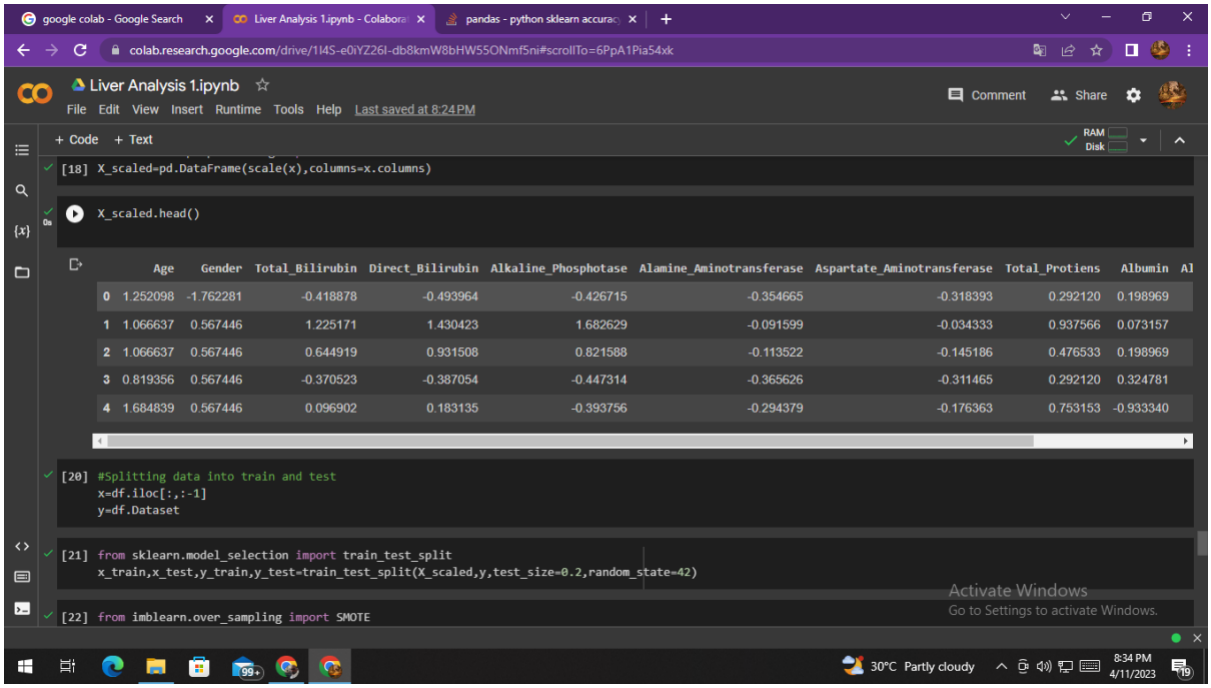

google colab - Google Search
Liver Analysis 1.ipynb - Colabora
pandas - python sklearn accurac
colab.research.google.com/drive/1I4S-e0iYZ26I-db8kmW8bHW55ONmf5ni#scrollTo=6PpA1Pia54xk
Liver Analysis 1.ipynb
File Edit View Insert Runtime Tools Help Last saved at 8:24 PM
Comment
Share
Files
sample_data
indian_liver_patient.csv
+ Code + Text
transform(np.log(df['Aspartate_Aminotransferase']))
Total_Bilirubin
Density
Direct_Bilirubin
Theoretical quantiles
Probability Plot
Ordered Values
Theoretical quantiles
Probability Plot
indian_liver_patient.csv
1 to 10 of 583 entries
Filter
Age Gender Total_Bilirubin Direct_Bilirubin
65 Female 0.7 0.1
62 Male 10.9 5.5
62 Male 7.3 4.1
58 Male 1 0.4
72 Male 3.9 2
46 Male 1.8 0.7
26 Female 0.9 0.2
29 Female 0.9 0.3
17 Male 0.9 0.3
55 Male 0.7 0.2
Show 10 per page
1 2 10 50 59
Disk 84.49 GB available
Activate Windows
Go to Settings to activate Windows.
30°C Partly cloudy
8:34 PM
4/11/2023


google colab - Google Search
Liver Analysis 1.ipynb - Colabora
pandas - python sklearn accurac
colab.research.google.com/drive/1I4S-e0iYZ26I-db8kmW8bHW55ONmf5ni#scrollTo=6PpA1Pia54xk
Liver Analysis 1.ipynb
File Edit View Insert Runtime Tools Help Last saved at 8:24 PM
Comment
Share
Files
sample_data
indian_liver_patient.csv
+ Code + Text
transform(np.log(df['Aspartate_Aminotransferase']))
Direct_Bilirubin
Density
Alkaline_Phosphotase
Theoretical quantiles
Probability Plot
Ordered Values
Theoretical quantiles
Probability Plot
indian_liver_patient.csv
1 to 10 of 583 entries
Filter
Age Gender Total_Bilirubin Direct_Bilirubin
65 Female 0.7 0.1
62 Male 10.9 5.5
62 Male 7.3 4.1
58 Male 1 0.4
72 Male 3.9 2
46 Male 1.8 0.7
26 Female 0.9 0.2
29 Female 0.9 0.3
17 Male 0.9 0.3
55 Male 0.7 0.2
Show 10 per page
1 2 10 50 59
Disk 84.49 GB available
Activate Windows
Go to Settings to activate Windows.
30°C Partly cloudy
8:34 PM
4/11/2023

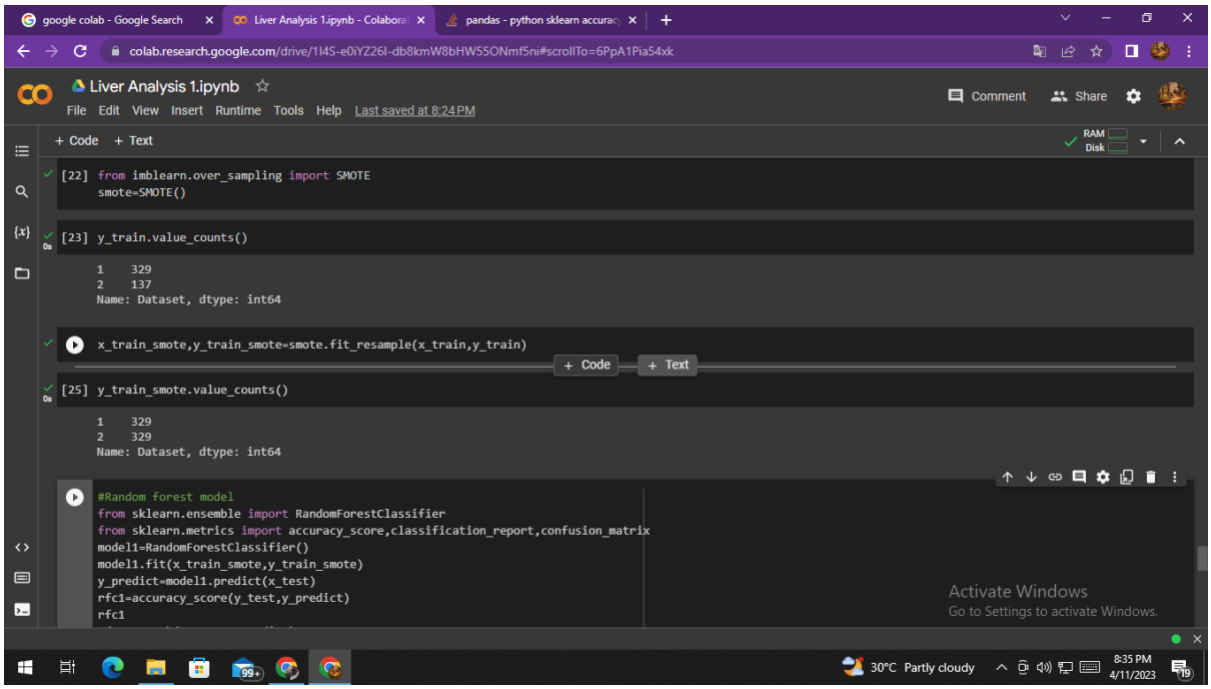

Liver Analysis 1.ipynb
Probability Plot
Density
Ordered Values
Alamine_Aminotransferase
Theoretical quantiles
indian_liver_patient.csv
1 to 10 of 583 entries
Age Gender Total_Bilirubin Direct_Bilirubin
65 Female 0.7 0.1
62 Male 10.9 5.5
62 Male 7.3 4.1
58 Male 1 0.4
72 Male 3.9 2
46 Male 1.8 0.7
26 Female 0.9 0.2
29 Female 0.9 0.3
17 Male 0.9 0.3
55 Male 0.7 0.2
Activate Windows


Liver Analysis 1.ipynb
Alamine_Aminotransferase
Theoretical quantiles
Probability Plot
Density
Ordered Values
Aspartate_Aminotransferase
Theoretical quantiles
indian_liver_patient.csv
1 to 10 of 583 entries
Age Gender Total_Bilirubin Direct_Bilirubin
65 Female 0.7 0.1
62 Male 10.9 5.5
62 Male 7.3 4.1
58 Male 1 0.4
72 Male 3.9 2
46 Male 1.8 0.7
26 Female 0.9 0.2
29 Female 0.9 0.3
17 Male 0.9 0.3
55 Male 0.7 0.2
[18] #scale
from sklearn.preprocessing import scale
Activate Windows

Scaling the Data:



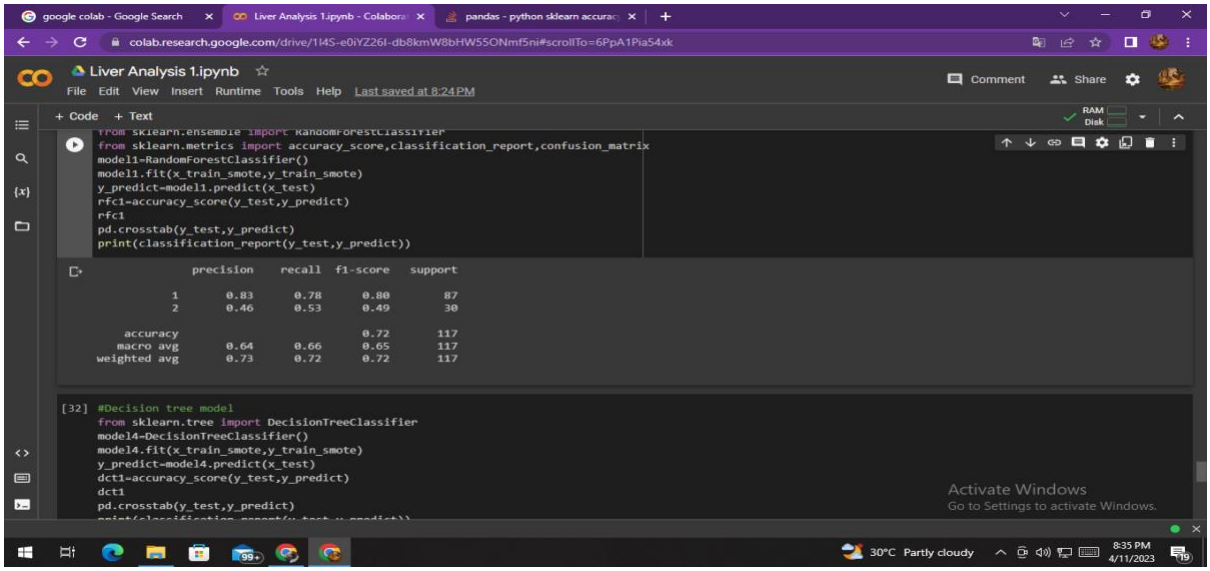
Splitting the data into training and testing data:



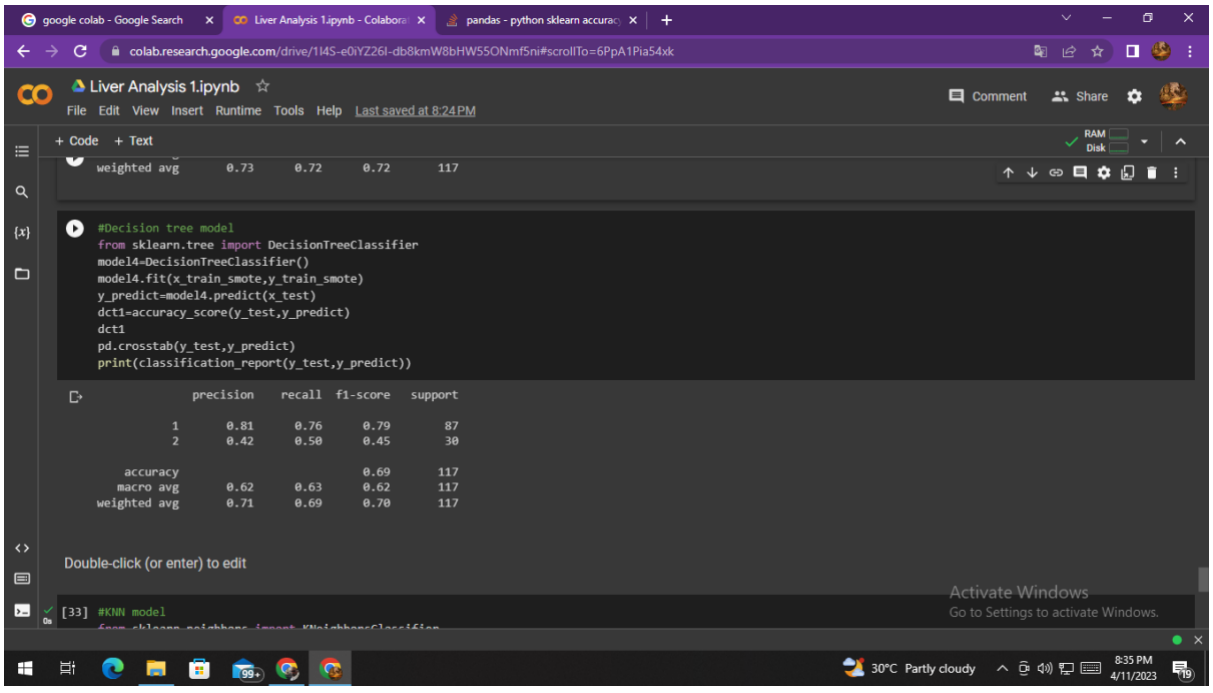
MODEL BUILDING:

Applying Algorithms

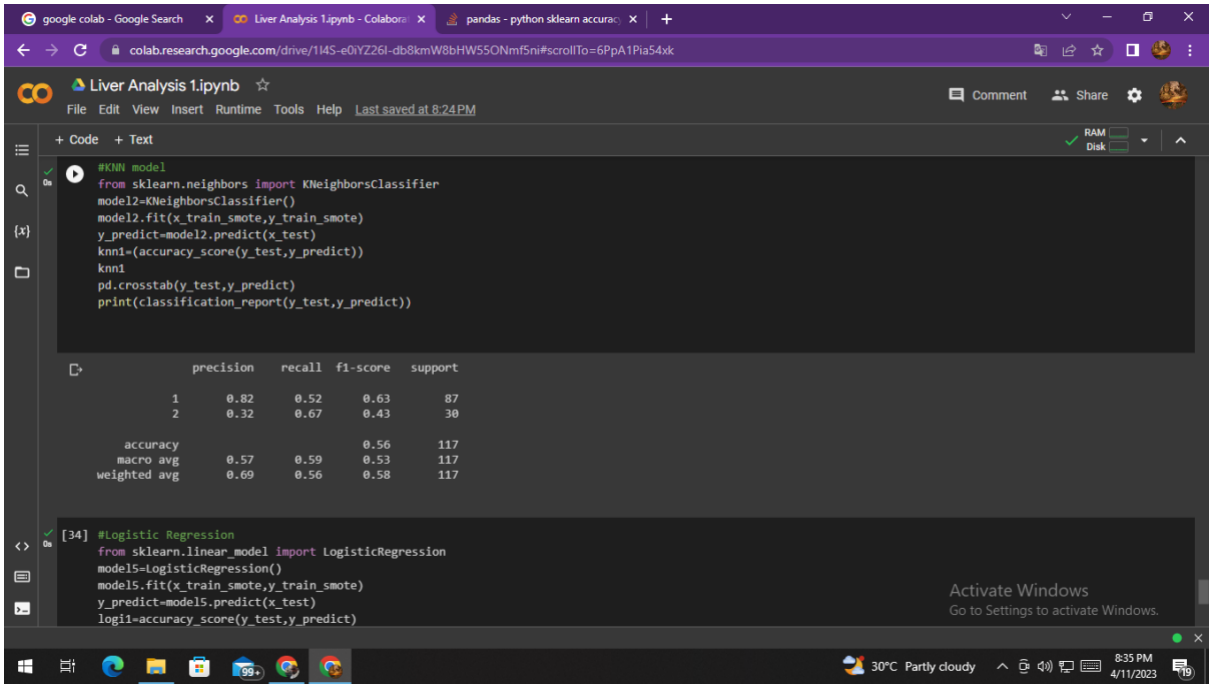
i) Random Forest Algorithm



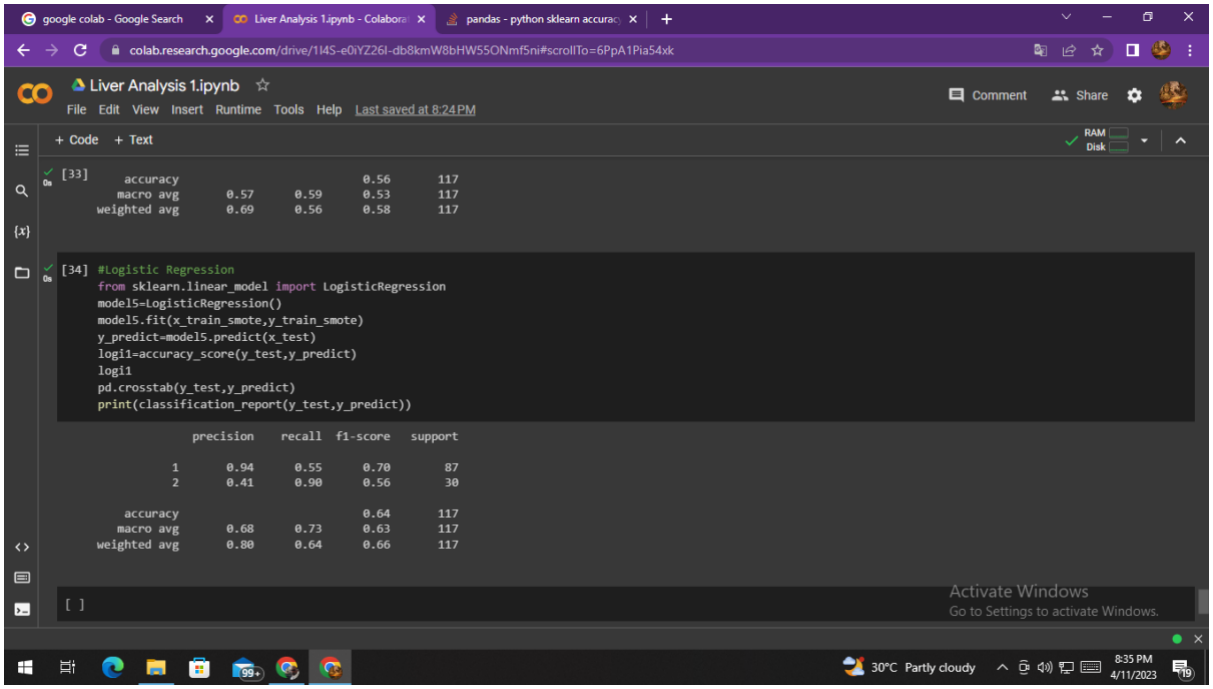
ii)Decision Tree



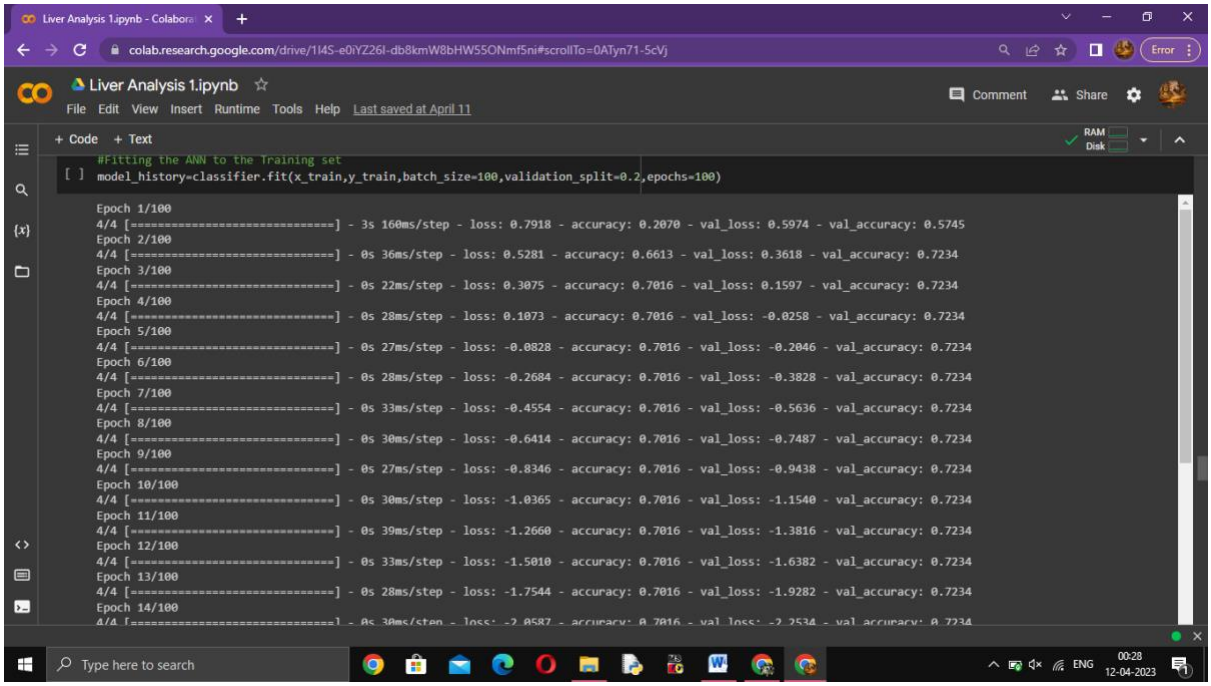
iii)K- Nearest Neighbour Algorithm



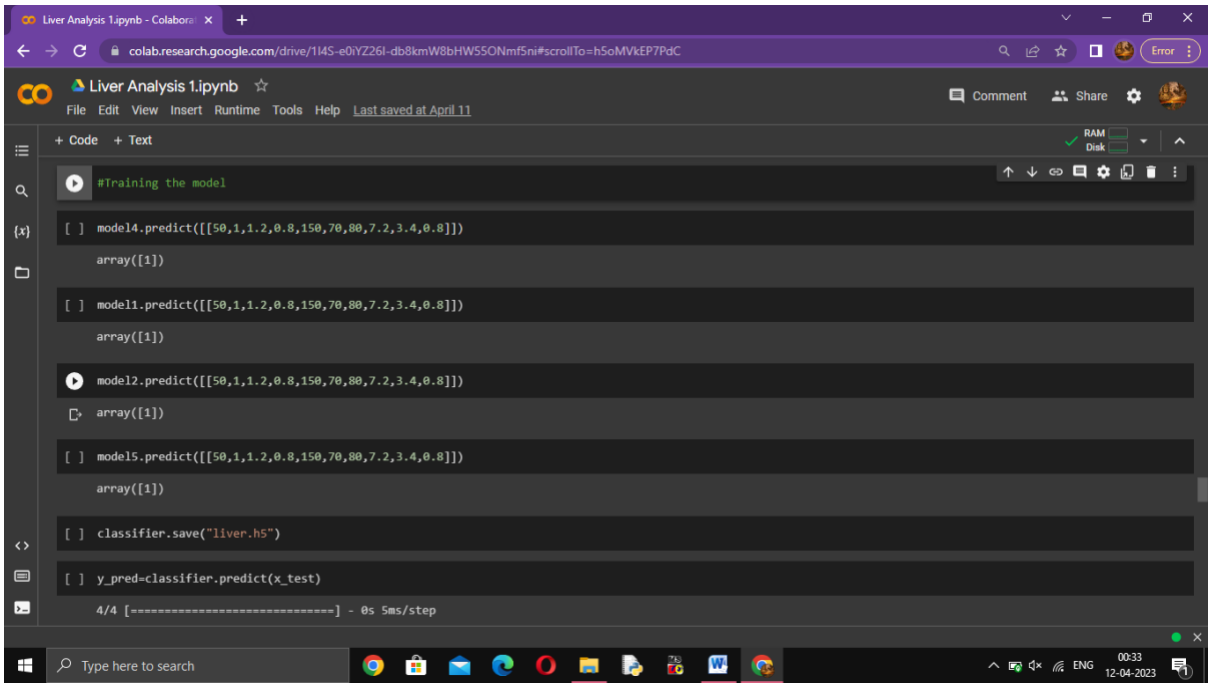
iv)Logistic Regression

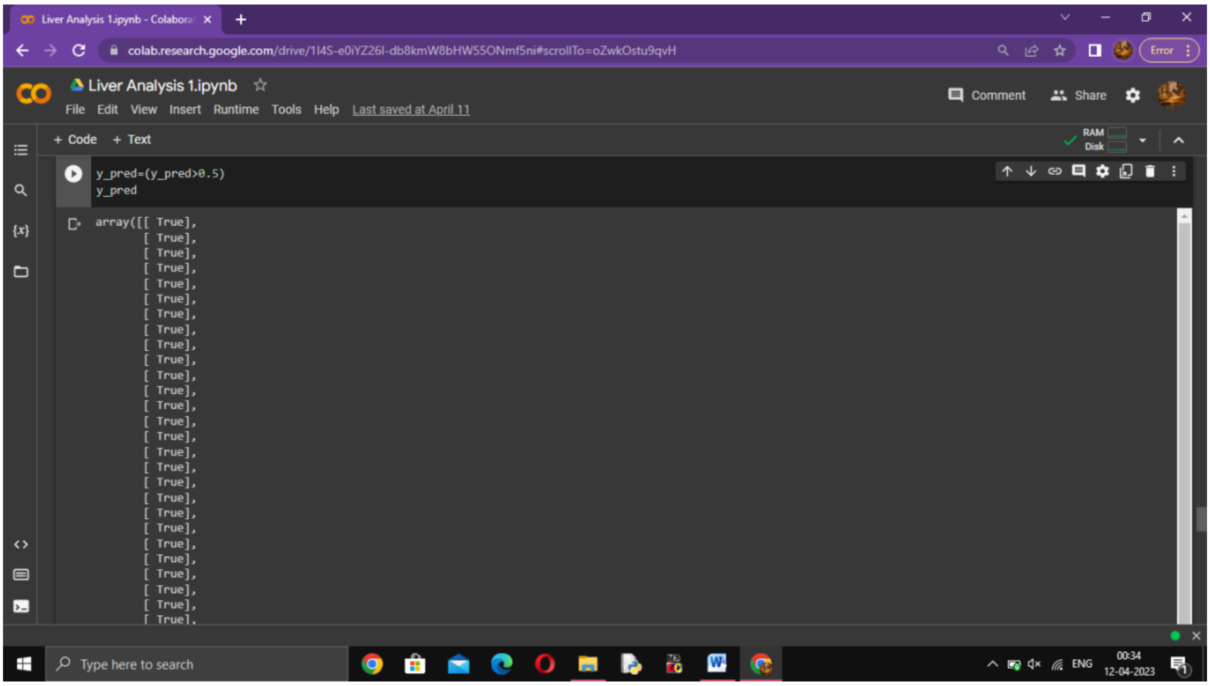


Fitting the ANN to the training set:

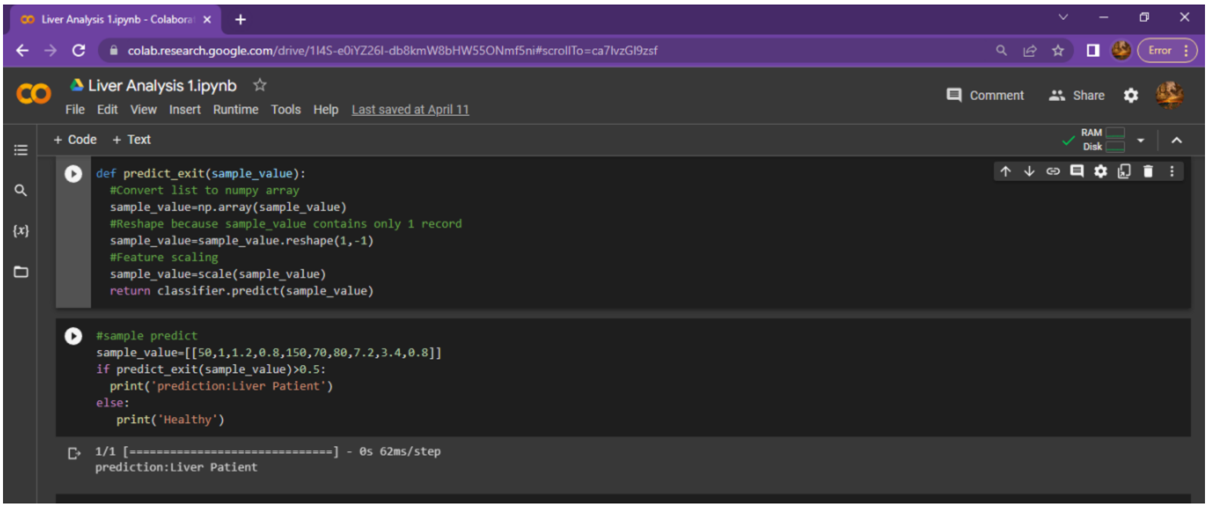


Testing the model:



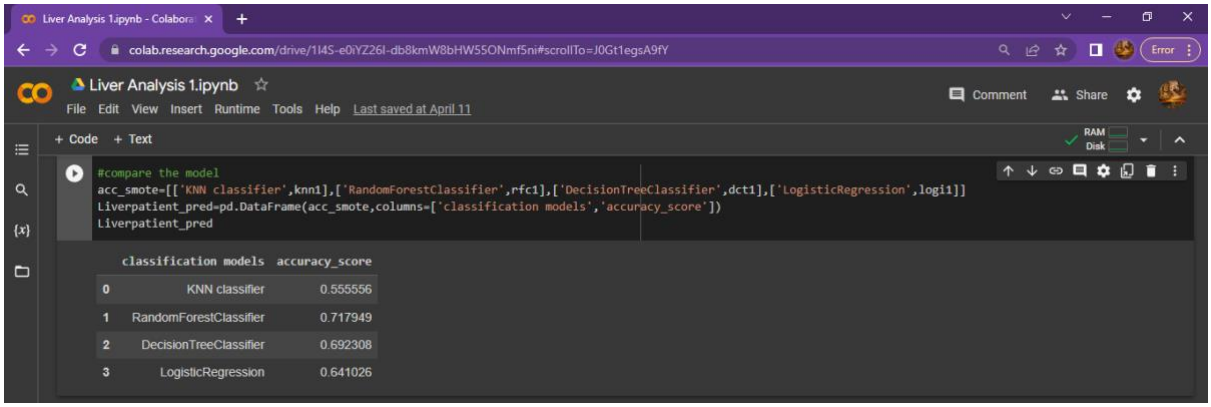


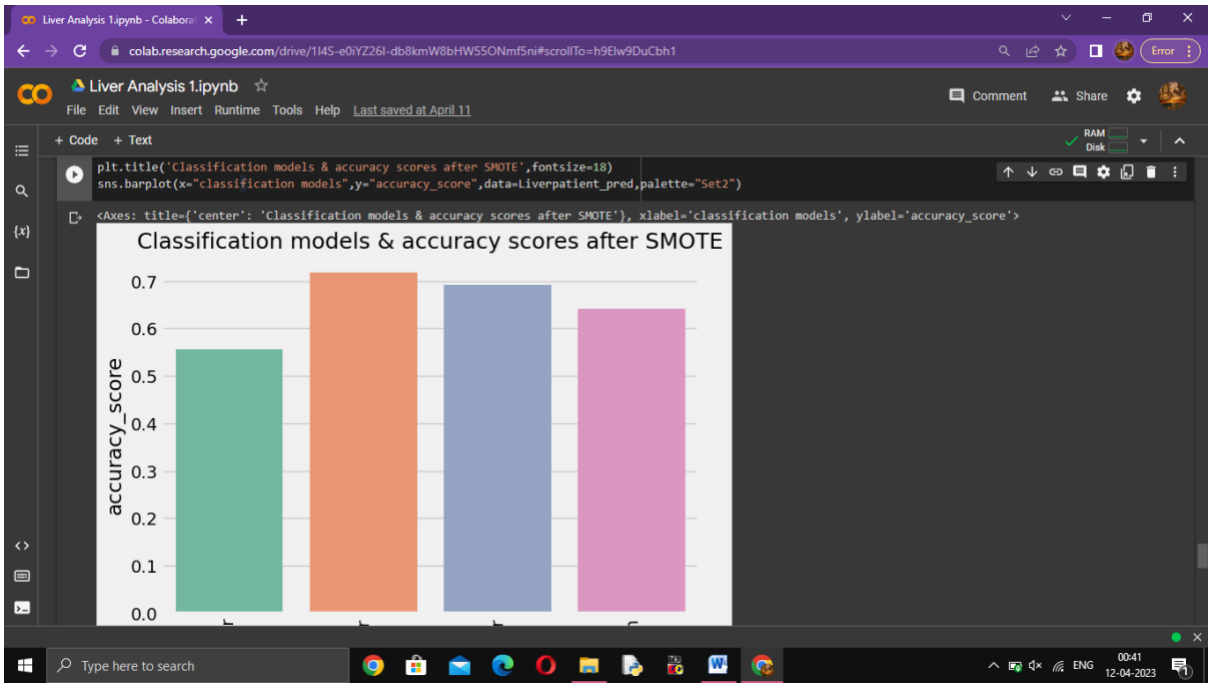
Predicting the Test data:

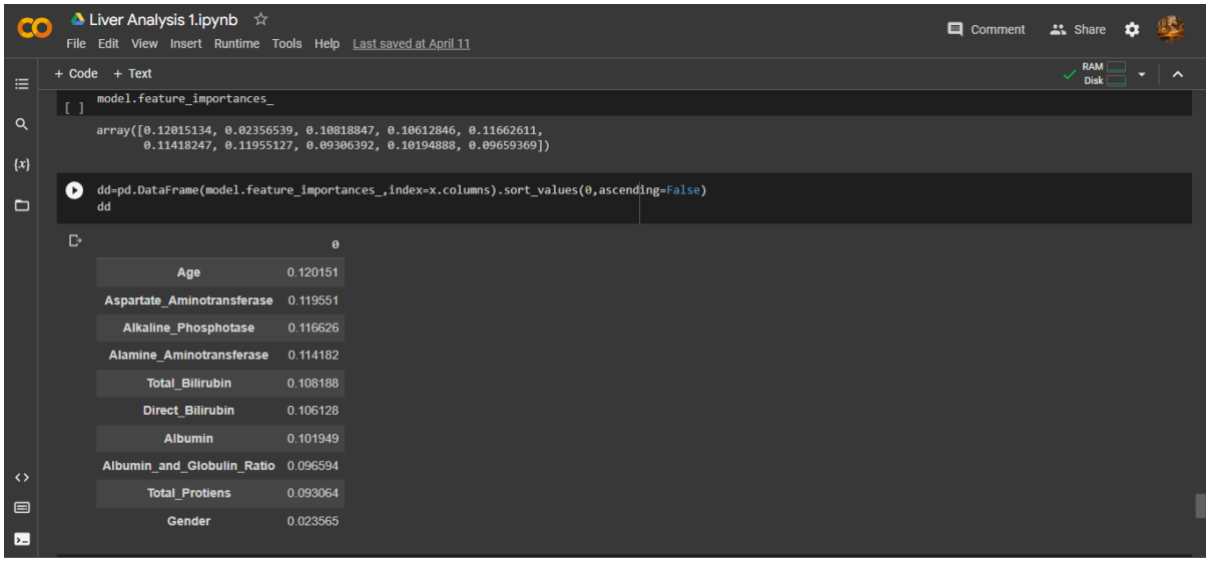


PERFORMANCE TESTING AND HYPER PARAMETER TUNING

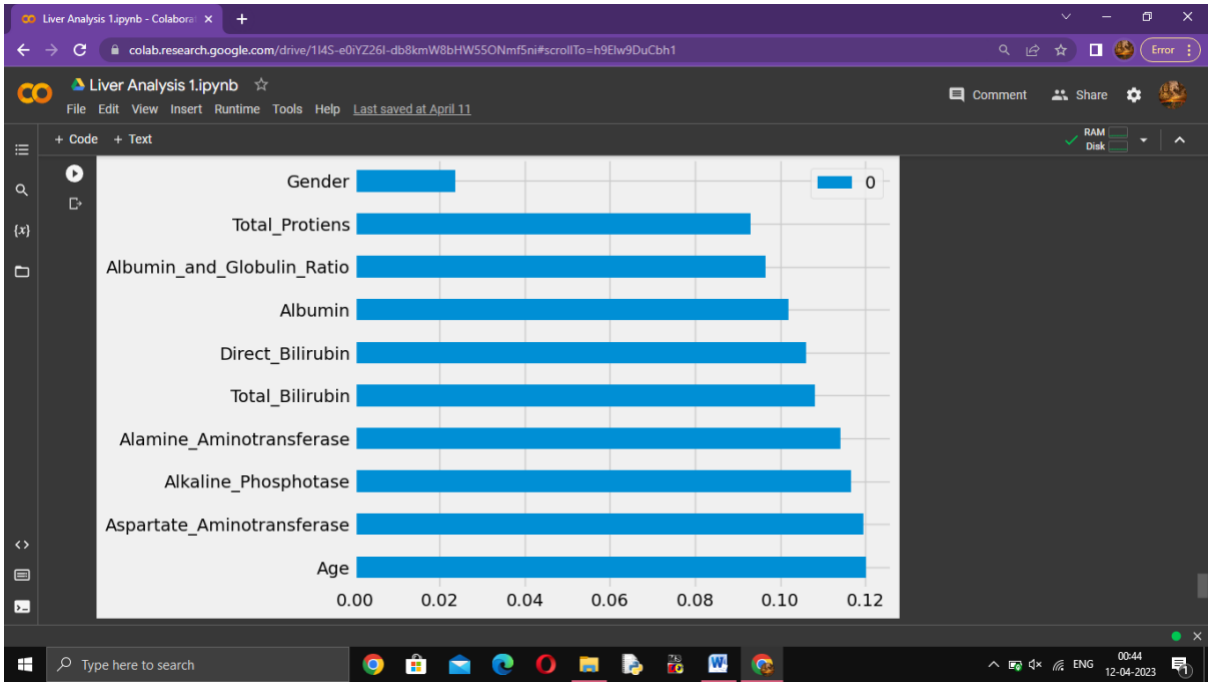
i)comparing the model





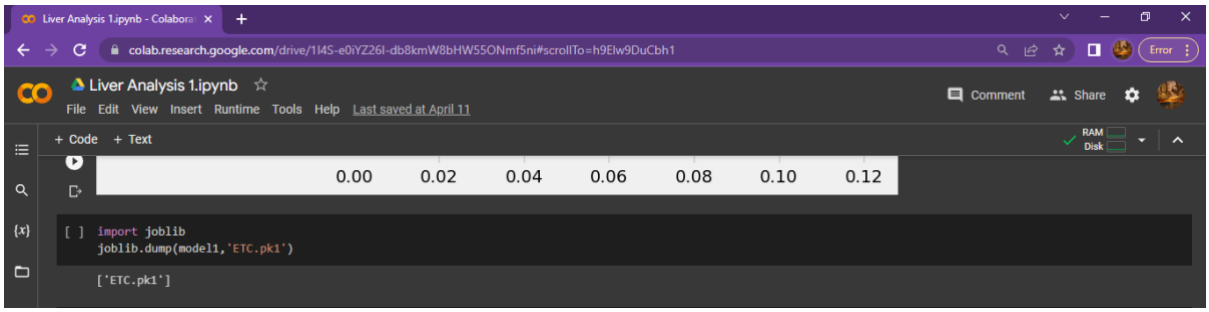


ii)Identifying important features



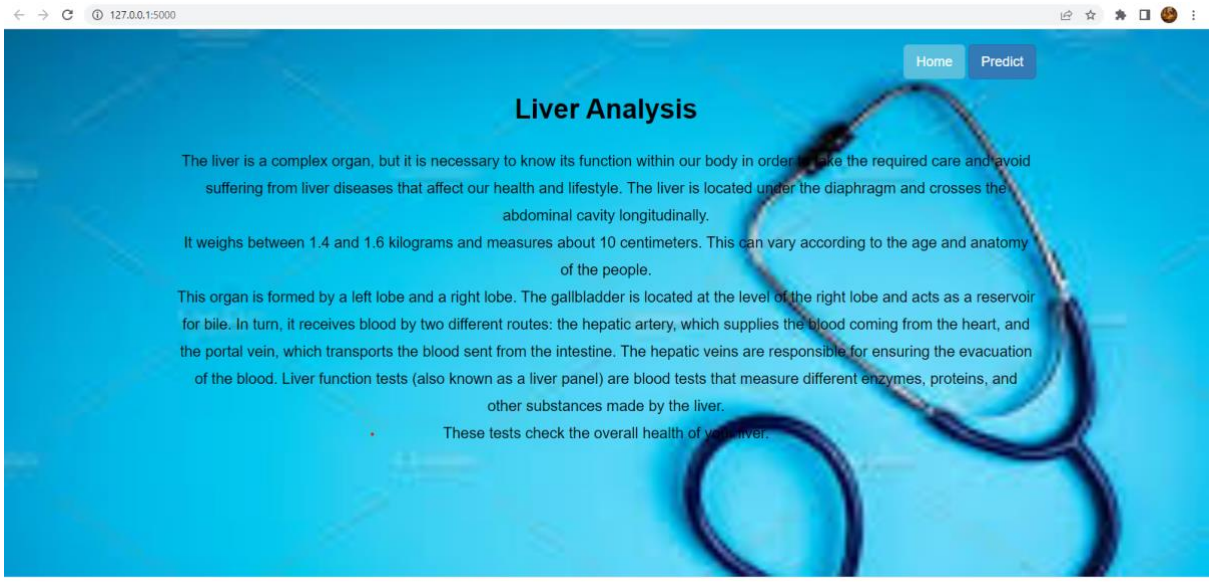
Model Deployment

i)Save the best model :

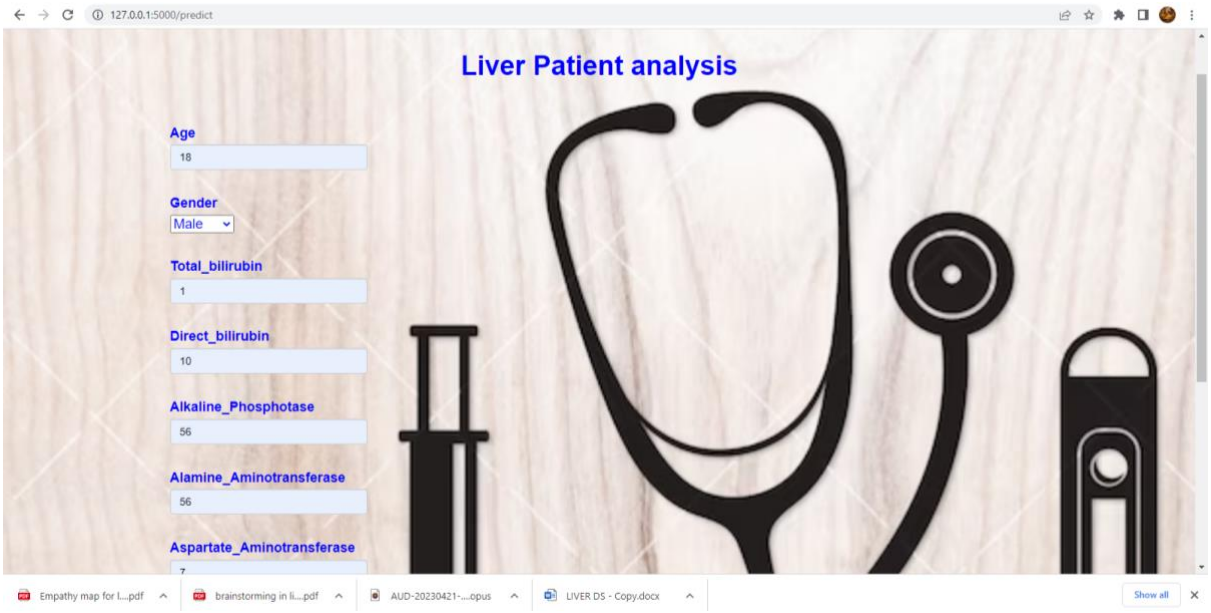


Flask Outputs:

Home page



Predict Page



Submit Page