A Review of Liver Patient Analysis Methods Using Machine Learning

INTRODUCTION:

The liver is one of the most critical organs of the human body.

It plays an essential role in the body's function.

Primary purposes include removing toxins from the body, fighting against infections, and balancing the hormones and secretion of bile juice (Devikanniga et al., 2020).

If these functions are not performed by the liver correctly,

it will result in several complications and liver diseases.

OVERVIEW:

Liver disease is one of the most chronic and threatening diseases globally that can cause various side effects

if not treated early (Dutta et al., 2022). According to World Health Organization (WHO) report in 2018,

the number of deaths due to liver diseases is around one million and

ranked 11th in the world with a critical number of fatalities (World Total Deaths, n.d.).

As the symptoms of liver diseases cannot be visible until the condition becomes

chronic, it is challenging and daunting for medical health professionals to identify liver disease

at its early stages (Devikanniga et al., 2020). In addition,

the traditional testing methods like sonography,

MRI scans and CT scans that are available for detecting liver diseases are expensive and

harmful with numerous side effects (Joloudari et al., 2019).

Thus, a significant constraint found by health care workers is to predict liver diseases at an early stage,

at minimal cost and at the same time provide a better health care system to treat liver diseases

PURPOSE

The class distribution of the dataset is imbalanced, with 416 having liver disease and 167 without liver disease.

This class distribution is imbalanced and balanced by applying ROSE using R.

ROSE is a bootstrap method that produces balanced synthetic samples to balance the data (Lunardon et al., 2014).

The reason for choosing ROSE is that the dataset is small, and the most reliable information might be lost if undersampling is conducted.

The other reason for considering ROSE is that it generates samples similar to the rare class samples,

which is also a consideration for an effective method for getting reliable accuracy

from the balanced dataset as this study's main aim is the accuracy metrics of the algorithms (Lunardon et al., 2014).

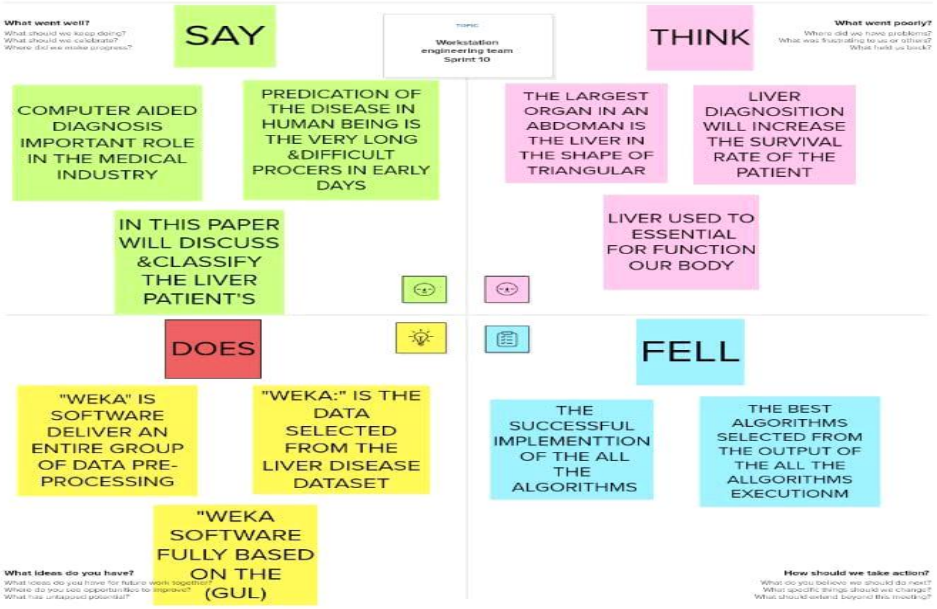
After applying the ROSE method on the class attribute, the sample generated is 520 having liver disease denoted by '1' and 480 without liver disease

PROBLEM DEFINITION & DESIGN THINKING

Accuracy is the value of correctly classified instances in both classes (Wu et al., 2019).

Accuracy = TN+TP/(TP+FP+TN+FN)

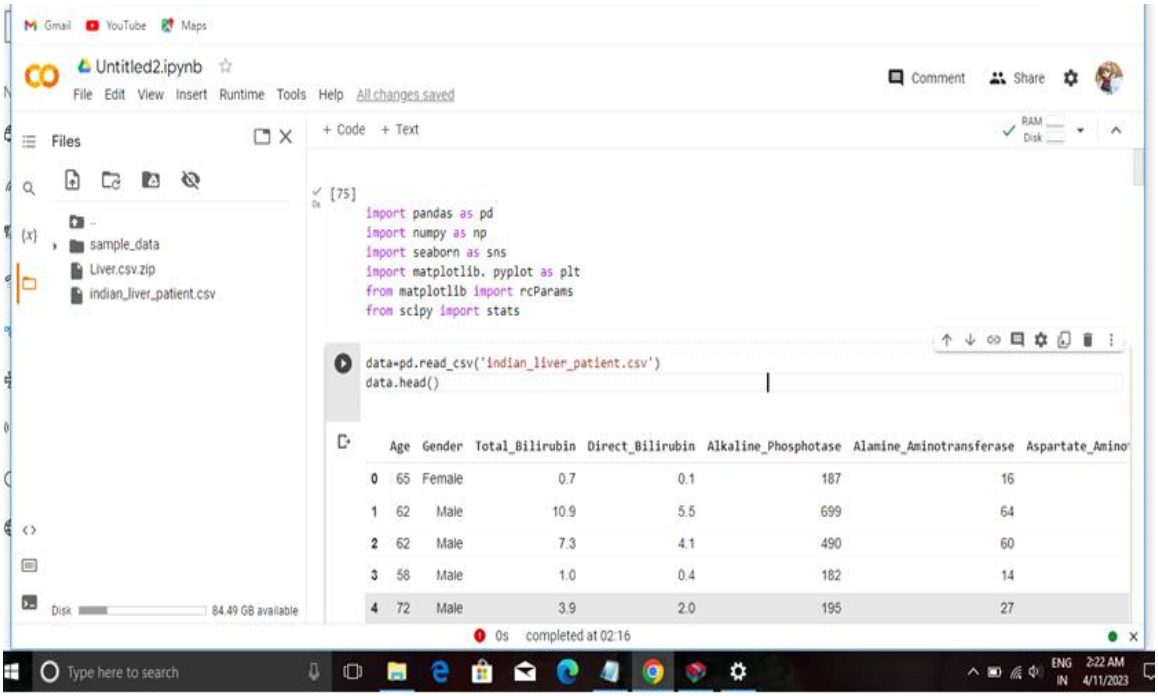
EMPATHY MAP

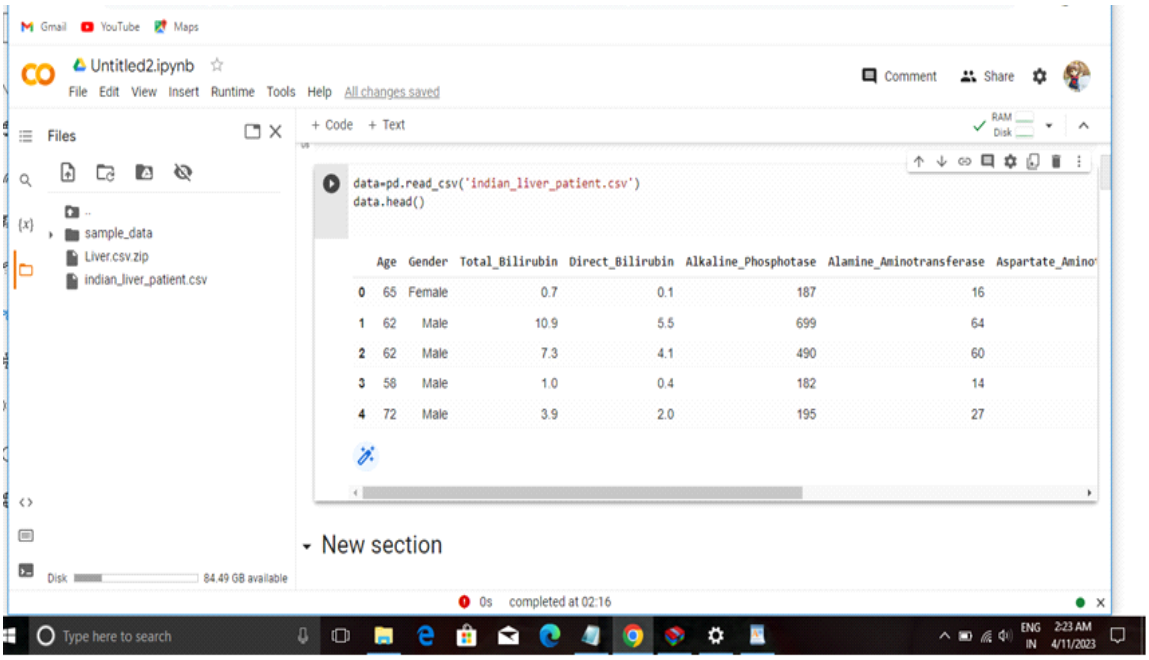


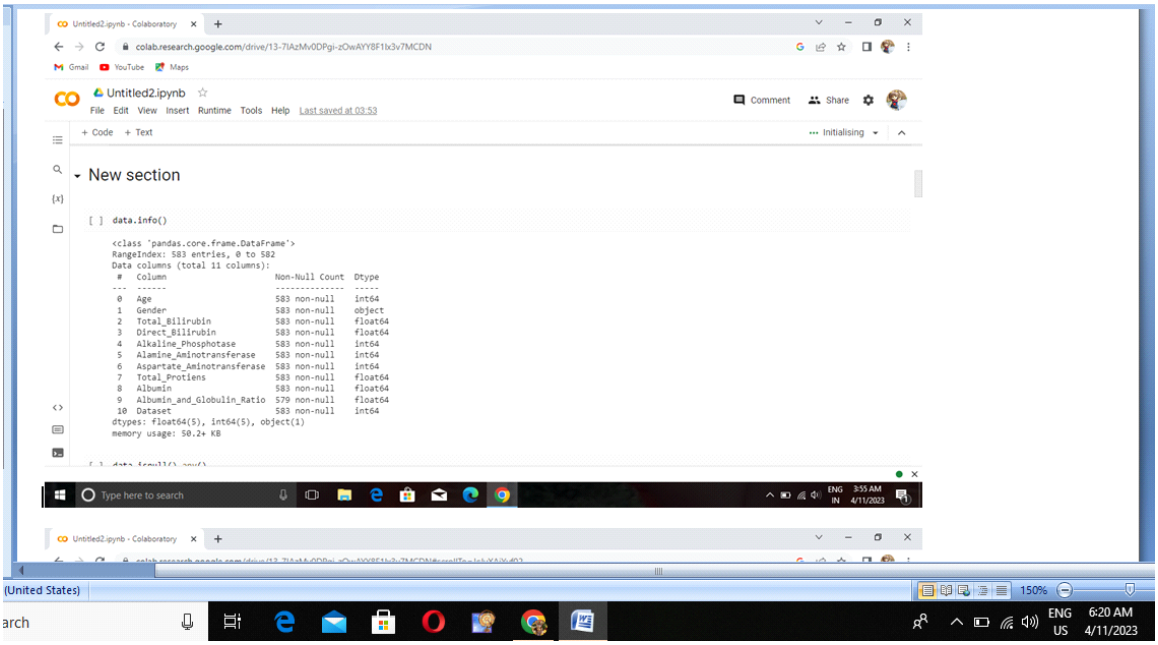
IDEATION & BRINSTORMING MAP

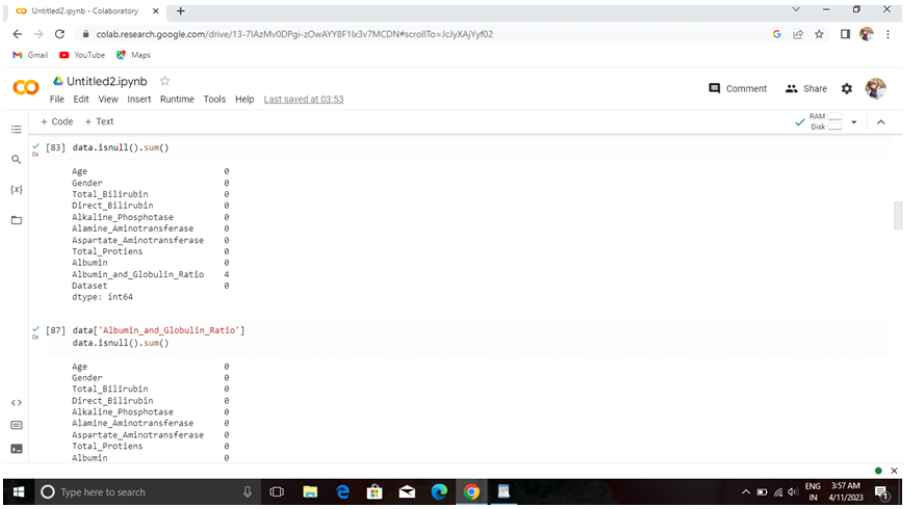
RESULT

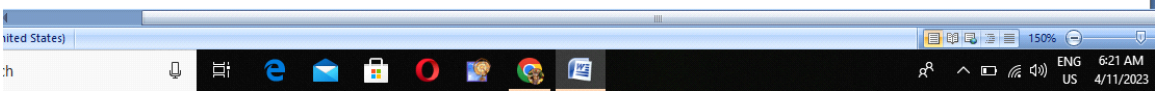
The seventh stage of the proposed LDP method is 'Results'. The results stage involves presenting the results after assessing the data. All the results of accuracies and confusion matrix metrics will be described .

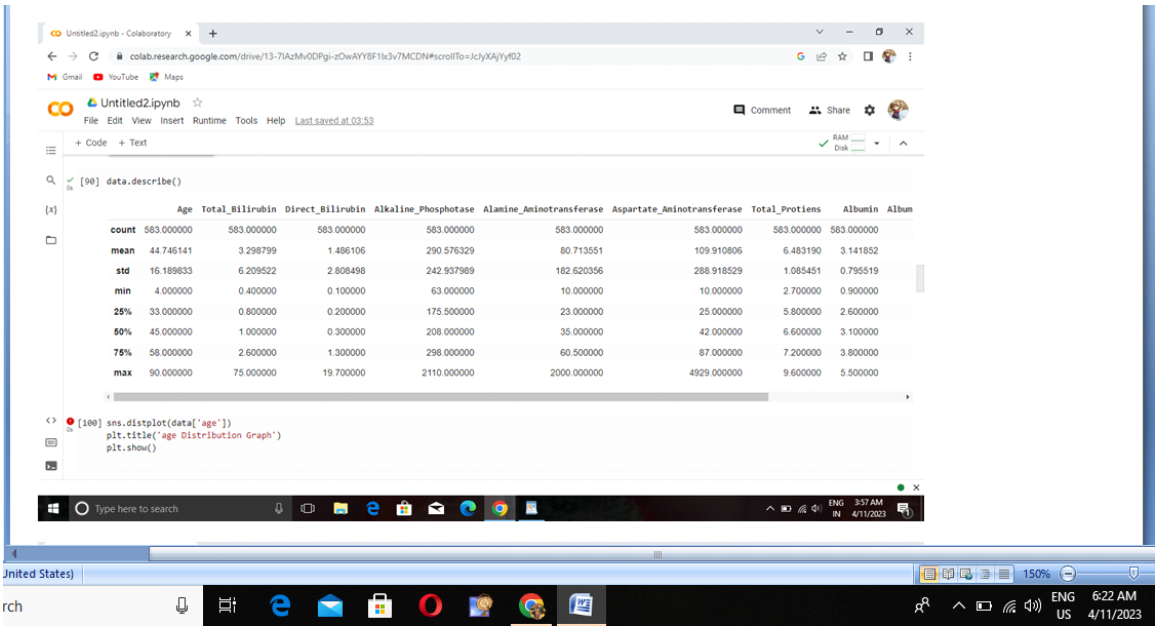



Untitled2.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
Files
sample_data
Liver.csv.zip
indian_liver_patient.csv
+ Code + Text
data=pd.read_csv('indian_liver_patient.csv')
data.head()
Age Gender Total_Bilirubin Direct_Bilirubin Alkaline_Phosphotase Alamine_Aminotransferase Aspartate_Amino
0 65 Female 0.7 0.1 187 16
1 62 Male 10.9 5.5 699 64
2 62 Male 7.3 4.1 490 60
3 58 Male 1.0 0.4 182 14
4 72 Male 3.9 2.0 195 27
New section
Disk 84.49 GB available
0s completed at 02:16
Type here to search
ENG IN 2:23 AM 4/11/2023

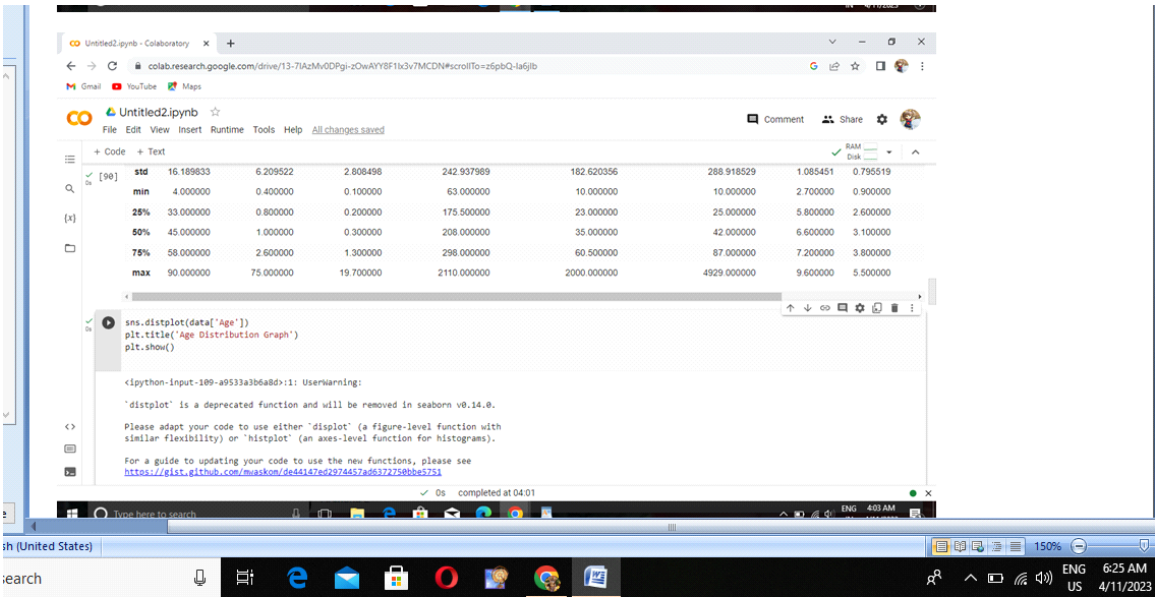

Untitled2.ipynb
File Edit View Insert Runtime Tools Help Last saved at 03:53
Comment Share
+ Code + Text
Initialising
New section
[ ] data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 583 entries, 0 to 582
Data columns (total 11 columns):
Column Non-Null Count Dtype
0 Age 583 non-null int64
1 Gender 583 non-null object
2 Total_Bilirubin 583 non-null float64
3 Direct_Bilirubin 583 non-null float64
4 Alkaline_Phosphotase 583 non-null int64
5 Alamine_Aminotransferase 583 non-null int64
6 Aspartate_Aminotransferase 583 non-null int64
7 Total_Protiens 583 non-null float64
8 Albumin 583 non-null float64
9 Albumin_and_Globulin_Ratio 579 non-null float64
10 Dataset 583 non-null int64
dtypes: float64(5), int64(5), object(1)
memory usage: 50.2+ KB
Type here to search
ENG IN 3:55 AM 4/11/2023
(United States)
150%
ENG US 6:20 AM 4/11/2023

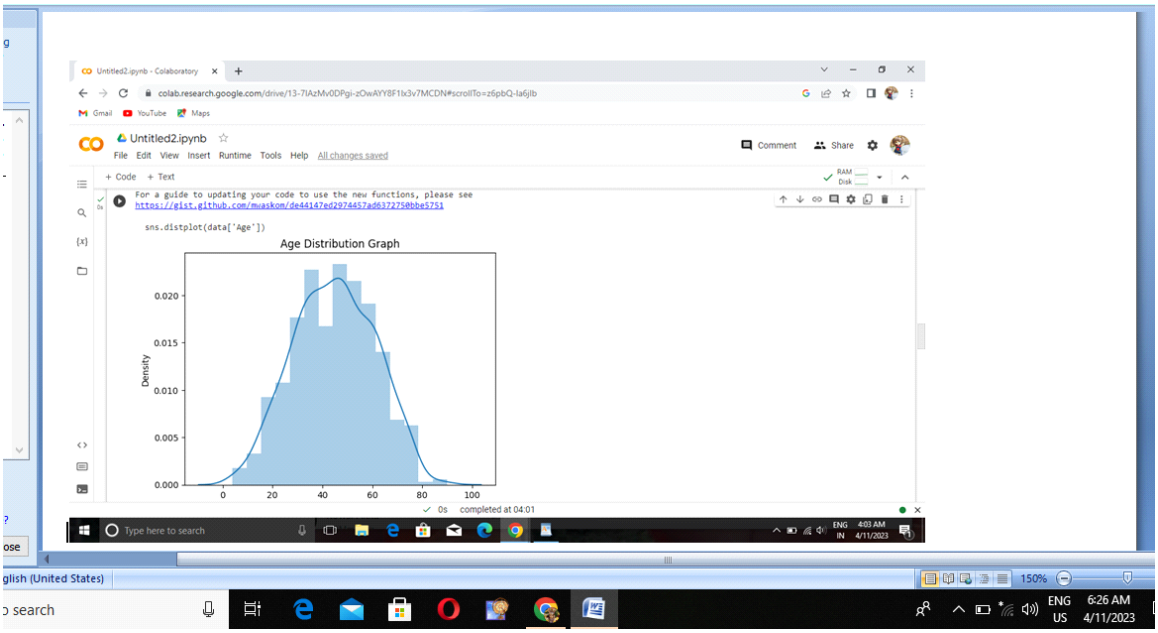

Untitled2.ipynb - Colaboratory
colab.research.google.com/drive/13-7lAzMv0DPgi-zOwAYY8F1lx3v7MCDN#scrollTo=JcJyXAjYyf02
Untitled2.ipynb
File Edit View Insert Runtime Tools Help Last saved at 03:53
Comment
Share
+ Code + Text
[83] data.isnull().sum()
Age 0
Gender 0
Total_Bilirubin 0
Direct_Bilirubin 0
Alkaline_Phosphotase 0
Alamine_Aminotransferase 0
Aspartate_Aminotransferase 0
Total_Protiens 0
Albumin 0
Albumin_and_Globulin_Ratio 4
Dataset 0
dtype: int64
[87] data['Albumin_and_Globulin_Ratio']
data.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
Type here to search
ENG IN 3:57 AM 4/11/2023


nited States)
150%
ENG US 6:21 AM 4/11/2023


Untitled2.ipynb - Colaboratory
colab.research.google.com/drive/13-7lAzMv0DPgi-zOwAYY8F1lx3v7MCDN#scrollTo=JcJyXAjYyf02
Untitled2.ipynb
File Edit View Insert Runtime Tools Help Last saved at 03:53
Comment
Share
+ Code + Text
[90] data.describe()
Age Total_Bilirubin Direct_Bilirubin Alkaline_Phosphotase Alamine_Aminotransferase Aspartate_Aminotransferase Total_Protiens Albumin Albumin
count 583.000000 583.000000 583.000000 583.000000 583.000000 583.000000 583.000000 583.000000
mean 44.746141 3.298799 1.486106 290.576329 80.713551 109.910806 6.483190 3.141852
std 16.189833 6.209522 2.808498 242.937989 182.620356 288.918529 1.085451 0.795519
min 4.000000 0.400000 0.100000 63.000000 10.000000 10.000000 2.700000 0.900000
25% 33.000000 0.800000 0.200000 175.500000 23.000000 25.000000 5.800000 2.600000
50% 45.000000 1.000000 0.300000 208.000000 35.000000 42.000000 6.600000 3.100000
75% 58.000000 2.600000 1.300000 298.000000 60.500000 87.000000 7.200000 3.800000
max 90.000000 75.000000 19.700000 2110.000000 2000.000000 4929.000000 9.600000 5.500000
[100] sns.distplot(data['age'])
plt.title('age Distribution Graph')
plt.show()
Type here to search
ENG IN 3:57 AM 4/11/2023

United States)
150%
ENG US 6:22 AM 4/11/2023

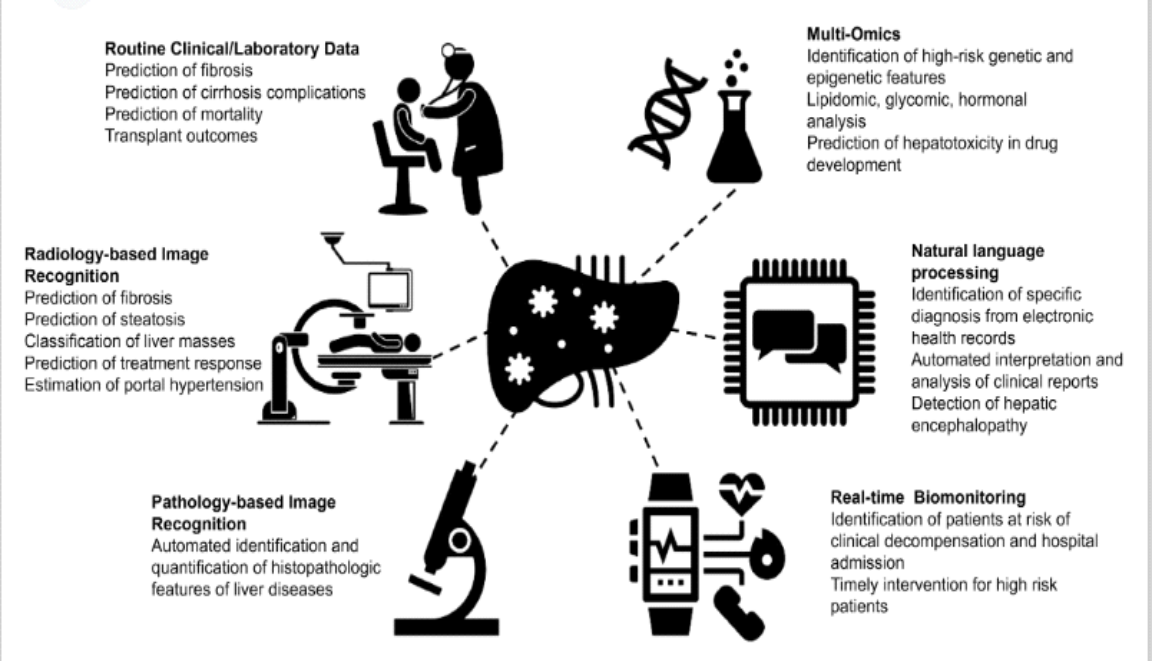




ADVANTAGES & DISADVANTAGES

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 (effectiveness)	Inter-observer variation
Follow-up comparison of treated and untreated patients	

APPLICATION:



CONCLUSION

Since the liver disease is not easy to diagnose, given the delicate nature of its signs, this research is pertinent in determining the algorithms that have better accuracy in predicting this dreadful

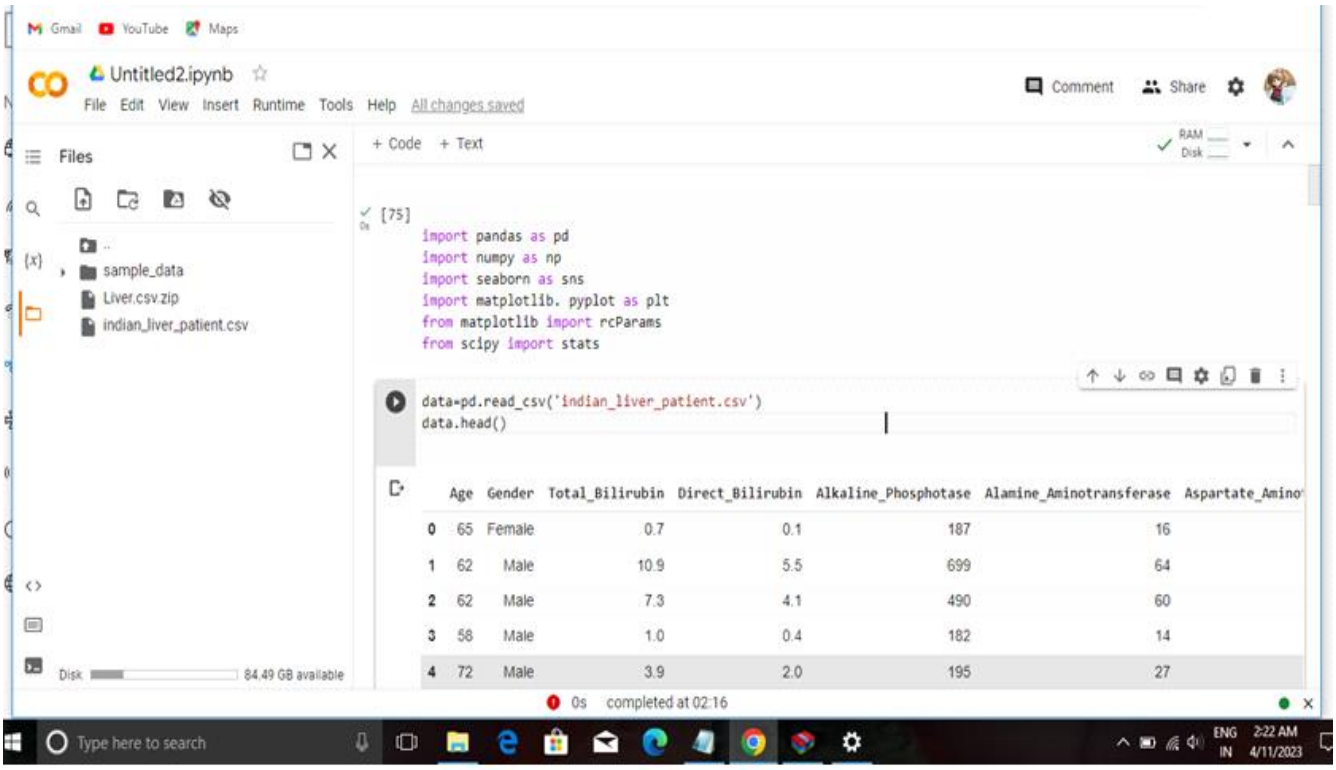
disease. The stages in the proposed LDP method provide a better alignment of each phase. Once the dataset is selected, the preprocessing step is conducted by replacing the missing values and balancing the dataset. After that, using R, five different supervised learning methods are applied (i.e., SVM, Naïve Bayes, K-NN, LDA, and CART), and the accuracy with confusion matrix metrics are recorded. The result shows that K-NN has a better accuracy of 91.7% for liver disease prediction

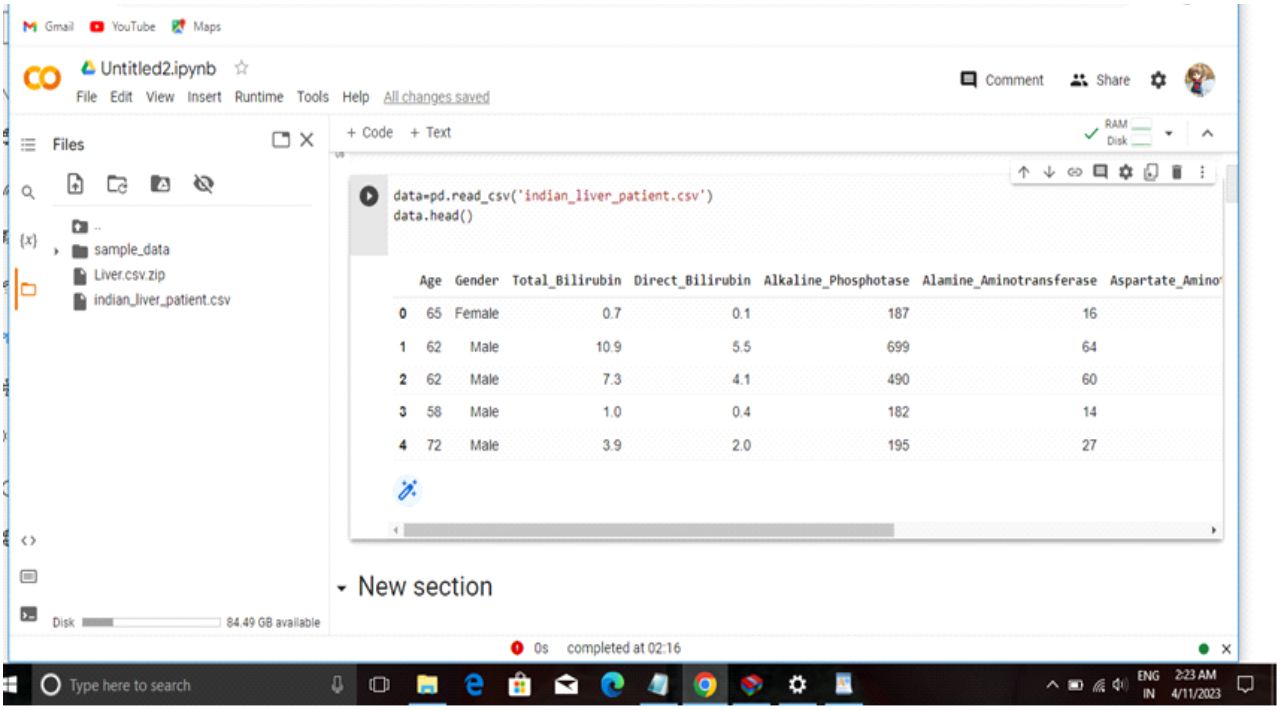
FUTURE SCOPE

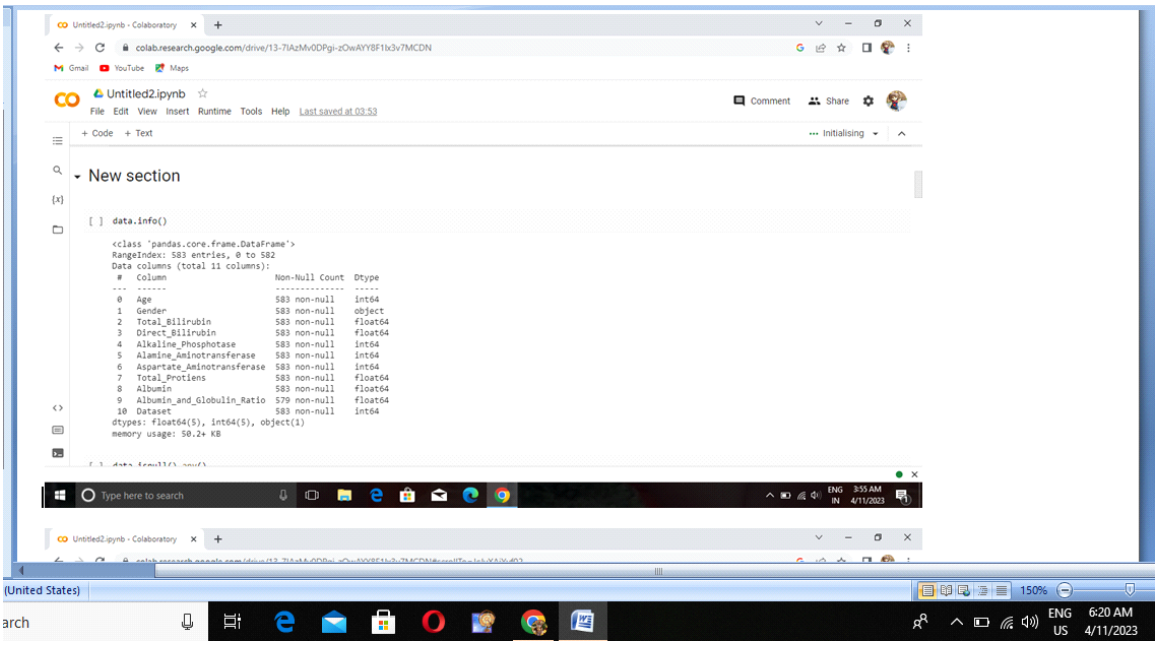
To conclude, the approach of the predicting analysis helps in predicting future possibilities by current data. The proposed model improved by applying a combination of three classifiers, Logistic regression, Random forest, and KNN algorithm. The python is employed for the implementation of the suggested model and the result proved regarding accuracy that is achieved 77.58 percent. For the future, the execution of the clustering algorithm is performed with the hybrid classifier technique for the division of data

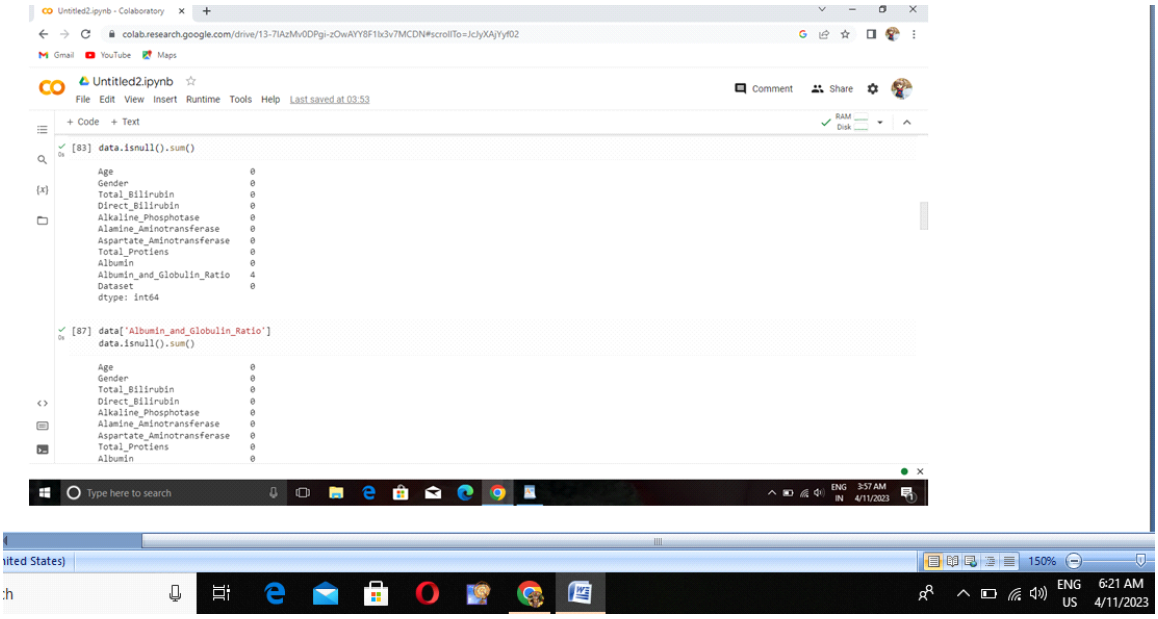
APPENDIX

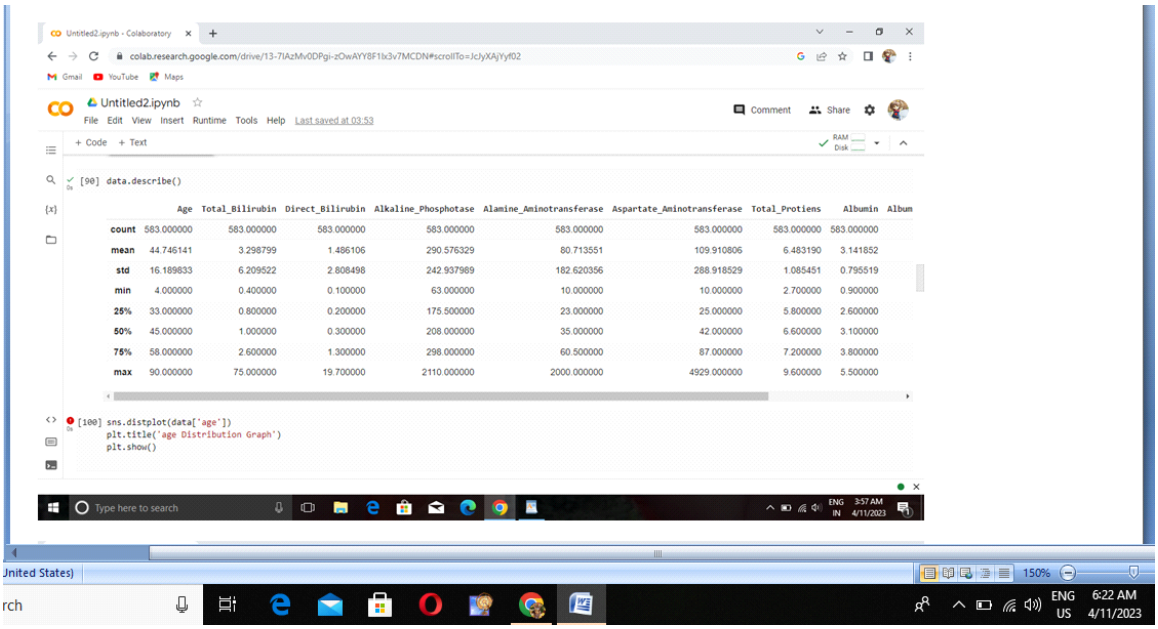
SOURSCE CODE

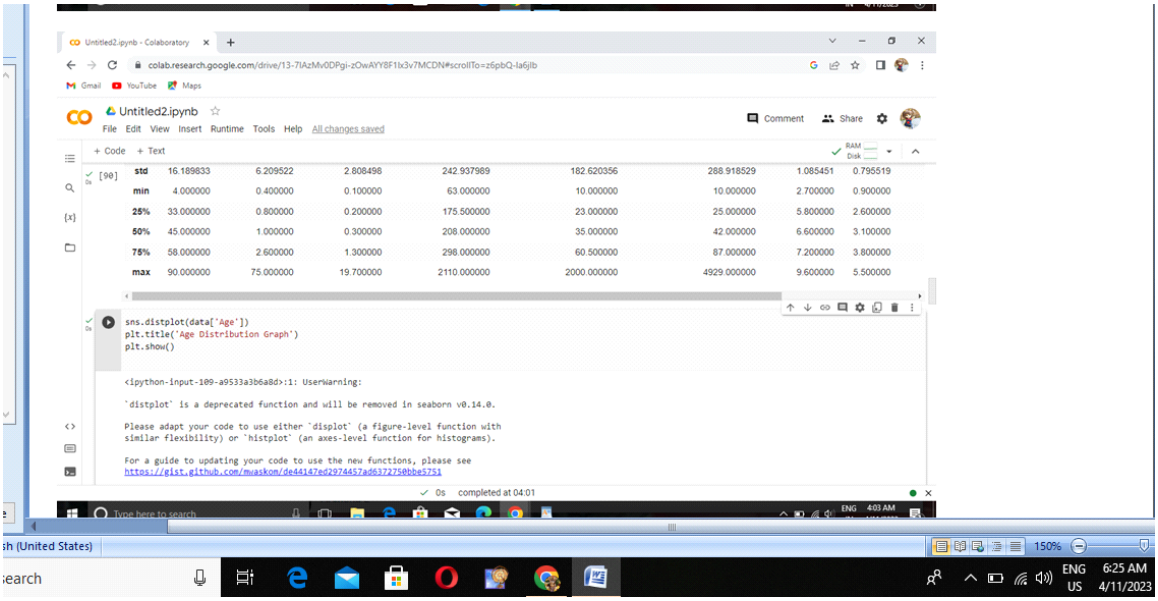

Gmail YouTube Maps
Untitled2.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
Files
+ Code + Text
RAM Disk
sample_data
Liver.csv.zip
indian_liver_patient.csv
[75]
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib. pyplot as plt
from matplotlib import rcParams
from scipy import stats
data=pd.read_csv('indian_liver_patient.csv')
data.head()
Age Gender Total_Bilirubin Direct_Bilirubin Alkaline_Phosphotase Alamine_Aminotransferase Aspartate_Amino
0 65 Female 0.7 0.1 187 16
1 62 Male 10.9 5.5 699 64
2 62 Male 7.3 4.1 490 60
3 58 Male 1.0 0.4 182 14
4 72 Male 3.9 2.0 195 27
Disk 84.49 GB available
0s completed at 02:16
Type here to search
ENG IN 2:22 AM 4/11/2023

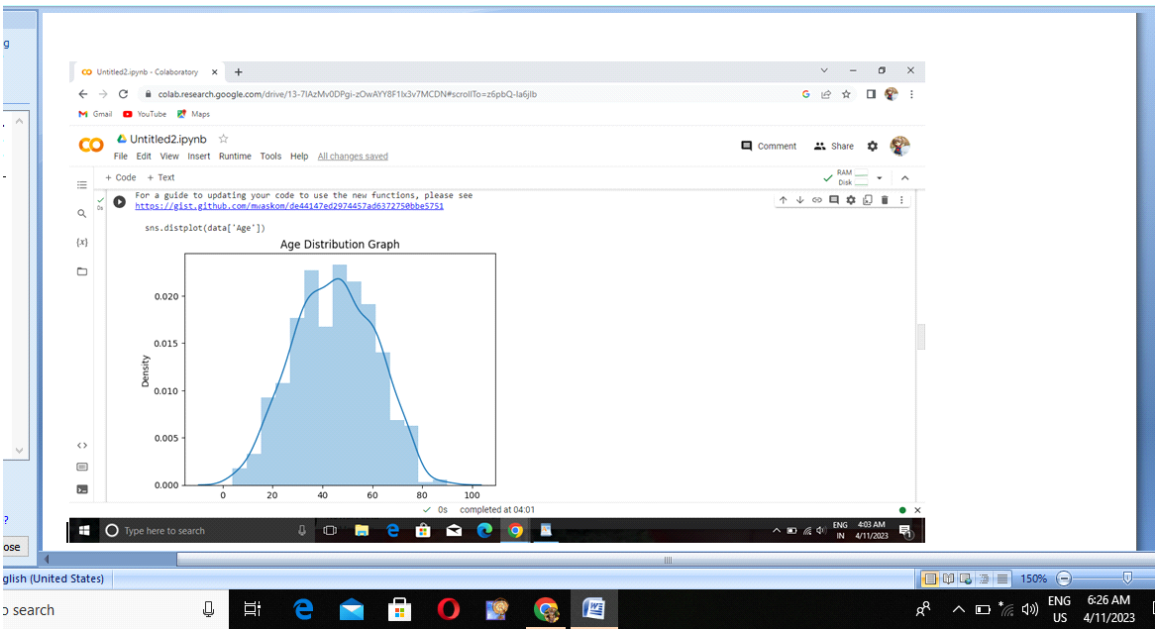

Gmail YouTube Maps
Untitled2.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
Files
+ Code + Text
RAM Disk
sample_data
Liver.csv.zip
indian_liver_patient.csv
data=pd.read_csv('indian_liver_patient.csv')
data.head()
Age Gender Total_Bilirubin Direct_Bilirubin Alkaline_Phosphotase Alamine_Aminotransferase Aspartate_Amino
0 65 Female 0.7 0.1 187 16
1 62 Male 10.9 5.5 699 64
2 62 Male 7.3 4.1 490 60
3 58 Male 1.0 0.4 182 14
4 72 Male 3.9 2.0 195 27
New section
Disk 84.49 GB available
0s completed at 02:16
Type here to search
ENG IN 2:23 AM 4/11/2023


Untitled2.ipynb - Colaboratory
colab.research.google.com/drive/13-7lAzMv0DPgi-zOwAYY8F1lx3v7MCDN
Untitled2.ipynb
File Edit View Insert Runtime Tools Help Last saved at 03:53
Comment
Share
+ Code + Text
Initialising
New section
[ ] data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 583 entries, 0 to 582
Data columns (total 11 columns):
Column Non-Null Count Dtype
0 Age 583 non-null int64
1 Gender 583 non-null object
2 Total_Bilirubin 583 non-null float64
3 Direct_Bilirubin 583 non-null float64
4 Alkaline_Phosphotase 583 non-null int64
5 Alamine_Aminotransferase 583 non-null int64
6 Aspartate_Aminotransferase 583 non-null int64
7 Total_Protiens 583 non-null float64
8 Albumin 583 non-null float64
9 Albumin_and_Globulin_Ratio 579 non-null float64
10 Dataset 583 non-null int64
dtypes: float64(5), int64(5), object(1)
memory usage: 50.2+ KB
Type here to search
ENG IN 3:55 AM 4/11/2023
(United States)
150%
ENG US 6:20 AM 4/11/2023


Untitled2.ipynb - Colaboratory
colab.research.google.com/drive/13-7lAzMv0DPgi-zOwAYY8F1lx3v7MCDN#scrollTo=JclyXAjYyf02
Gmail YouTube Maps
Untitled2.ipynb
File Edit View Insert Runtime Tools Help Last saved at 03:53
Comment
Share
+ Code + Text
RAM Disk
[83] data.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 4
Dataset 0
dtype: int64
[87] data['Albumin_and_Globulin_Ratio']
data.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
Type here to search
ENG IN 3:57 AM 4/11/2023
150%
ENG US 6:21 AM 4/11/2023


Untitled2.ipynb - Colaboratory
colab.research.google.com/drive/13-7lAzMv0DPgi-zOwAYY8F1lx3v7MCDN#scrollTo=JcJyXAjYyf02
Untitled2.ipynb
File Edit View Insert Runtime Tools Help Last saved at 03:53
Comment
Share
+ Code + Text
[90] data.describe()
Age Total_Bilirubin Direct_Bilirubin Alkaline_Phosphotase Alamine_Aminotransferase Aspartate_Aminotransferase Total_Protiens Albumin Album
count 583.000000 583.000000 583.000000 583.000000 583.000000 583.000000 583.000000 583.000000
mean 44.746141 3.298799 1.486106 290.576329 80.713551 109.910806 6.483190 3.141852
std 16.189833 6.209522 2.808498 242.937989 182.620356 288.918529 1.085451 0.795519
min 4.000000 0.400000 0.100000 63.000000 10.000000 10.000000 2.700000 0.900000
25% 33.000000 0.800000 0.200000 175.500000 23.000000 25.000000 5.800000 2.600000
50% 45.000000 1.000000 0.300000 208.000000 35.000000 42.000000 6.600000 3.100000
75% 58.000000 2.600000 1.300000 298.000000 60.500000 87.000000 7.200000 3.800000
max 90.000000 75.000000 19.700000 2110.000000 2000.000000 4929.000000 9.600000 5.500000
[100] sns.distplot(data['age'])
plt.title('age Distribution Graph')
plt.show()
Type here to search
ENG IN 3:57 AM 4/11/2023
United States)
150%
ENG US 6:22 AM 4/11/2023


Untitled2.ipynb - Colaboratory
colab.research.google.com/drive/13-7lAzMv0DPgi-zOwAYY8F1lx3v7MCDN#scrollTo=z6pbQ-la6jlb
Untitled2.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
[90]
std 16.189833 6.209522 2.808498 242.937989 182.620356 288.918529 1.085451 0.795519
min 4.000000 0.400000 0.100000 63.000000 10.000000 10.000000 2.700000 0.900000
25% 33.000000 0.800000 0.200000 175.500000 23.000000 25.000000 5.800000 2.600000
50% 45.000000 1.000000 0.300000 208.000000 35.000000 42.000000 6.600000 3.100000
75% 58.000000 2.600000 1.300000 298.000000 60.500000 87.000000 7.200000 3.800000
max 90.000000 75.000000 19.700000 2110.000000 2000.000000 4929.000000 9.600000 5.500000
sns.distplot(data['Age'])
plt.title('Age Distribution Graph')
plt.show()
<ipython-input-109-a9533a3b6a8d>:1: UserWarning:
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.
Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).
For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
0s completed at 04:01
Type here to search
ENG 4:03 AM
sh (United States)
150%
ENG US 6:25 AM 4/11/2023


Untitled2.ipynb - Colaboratory
colab.research.google.com/drive/13-7lAzMv0DPgi-zOwAlYY8F1tx3v7MCDN#scrollTo=z6pbQ-la6jlb
Gmail
YouTube
Maps
Untitled2.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
sns.distplot(data['Age'])
Age Distribution Graph
Density
0.020
0.015
0.010
0.005
0.000
0
20
40
60
80
100
0s completed at 04:01
Type here to search
ENG
IN
4:03 AM
4/11/2023

glish (United States)
150%
search
ENG
US
6:26 AM
4/11/2023