A REVIEW OF LIVER PATIENT ANALYSIS METHODS USING MACHINE LEARNING

SUBMITTED BY

DHILIPAN.R

INTRODUCTION

The liver is the largest solid organ and the largest gland in the human body, that sits on the right side of the belly. Weighing about 3 pounds, the liver is reddish-brown in colour and feels rubbery to the touch. The liver has two large sections, called the right and the left lobes. The gallbladder sits under the liver, along with parts of the pancreas and intestines. The liver and these organs work together to digest, absorb, and process food [1]. Health care and medicine handles huge data on daily basis. Liver failure means that your liver is losing or has lost all of its function. It is a life-threatening condition that demands urgent medical care [2]. Liver disease is also referred to as hepatic disease. Liver disease is a large term that covers all the potential problems that cause the liver to fail to perform its designated functions. Usually, more than 75% or three quarters of liver tissue needs to be affected before a decrease in function occurs [3]. The liver's main job is to filter 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 secretes bile that ends up back in the intestines. The liver also makes proteins important for blood clotting and other functions [4]. Figure 1 refers the structure of liver

Overwiew

the liver patients with the help of machine learning algorithms using the ILPD data set. Further this paper is organized with the following sections such as related work Liver Diseases Liver disease is the occurrence of any trouble of liver function that causes sickness. The liver is responsible for most important functions of the body. If the liver fails to do those functions, it can cause significant injury to the body [7]. Liver disease is also referred as hepatic disease. The different types of liver diseases are largely classified according to the cause of the specific problem. some of which are acute and not serious while others are chronic and may be life-threatening [8]. The most common liver diseases [7] are: Acute (sudden) hepatitis (inflammation): Acute hepatitis C is a contagious disease caused by the hepatitis C virus• (HCV), which is spread through contact with infected blood and bodily fluids Chronic (long duration) hepatitis: This long-lasting liver infection is caused by the hepatitis C virus. It begins as an• acute hepatitis that starts within the first 6 months of exposure to the virus. Fatty liver disease: steatosis, is a broad term that describes the build-up of fats in the liver. When too much fat builds up• in your liver, that's fatty liver disease. Cirrhosis (scarring): Cirrhosis is a late stage of scarring (fibrosis) of the liver caused by many forms of liver diseases• and conditions, such as hepatitis and chronic alcoholism. Cancer: Cancers that affect the liver are most commonly metastatic cancers that have spread via the bloodstream to the• liver from other sites in the body. However, primary cancers (cancers that arise in the liver) can also occur. The most common type of primary liver cancers is known as hepatocellular carcinomas. The main objective of this research work is to classify, machine learning techniques used, experimental evaluation and conclusion.

PURPOSE

If you google out some basic questions as such:

1. How many liver deaths take place every year in India?

Answer: Liver cirrhosis is the biggest health problem posed by alcohol use, with 1.4 lakh **deaths every year**.

2. Is **liver cirrhosis** a lifestyle disease?

Answer: Sadly, no. **In** fact, it is getting more common **in younger people** than ever before. Dr. Amrish said that **liver disease** can set **in** childhood too as it can pass through genes.

3. Is liver cirrhosis **treatable**?

Answer: **Cirrhosis** isn't **curable**, but it's **treatable**. Alcohol abuse, hepatitis, and fatty **liver** disease are some of the main causes.

Then you people will get answers like these as I mentioned above, So the **purpose** and **inspiration** of this project clearly simplifies the devastating answers from the **data available with Google**. We do need a system that in some stage reduces the burden on doctors, and today in this article I'll try to frame a practical logic that will help our healthcare system in a long run.

LITERATURE SURVEY

Paul Mangiameli et al., [2] proposed model selection affects the decision support systems accurately. In their model selection, how to affects the accuracy of decision support system hydrides by single model and ensembles. They proposed single model is not more accurate than ensembles. Ahmed M. Hashem et al., [18] proposed to predict Liver Cirrhosis or fibrosis single stage classification model and multistage classification model. In their model based on Decision Tree, Neural Network, Nearest Neighborhood clustering and Logistic Regression. Ziol.M et al.,[3] proposed to evaluated liver fibrosis with chronic hepatitis C for patients using liver stiffness measurement (LSM).Z. Jiang.Z.,[4] proposed for discovering the corresponding degree of fibrosis by support vector machine (SVM). Kemal Polat et al.,[22] proposed resource

allocation mechanism of AIRS was changed with a new one decided by Fuzzy-Logic. This approach called as Fuzzy- AIRS was used as a classifier in the diagnosis of Liver Disorders. In this Classification accuracies were evaluated by comparing them with reported classifier's accuracy, time and number of resources. Piscaglia et al.,[6] proposed to predict Liver cirrhosis and other liver-related diseases used by Artificial neural network. Dong-Hoi Kim et al.,[19] proposed machine learning technique and decision tree(C4.5).In this method is used for to predict the susceptibility to two liver diseases such as chronic hepatitis and cirrhosis from single nucleotide polymorphism(SNP) data . They also used to identify a set of SNPs relevant to those diseases. Anh Pham,[8] developed optimizing the classification accuracy when analyzing some medical datasets. This proposed work done by new meta-heuristic approach, called the Homogeneity-

Based Algorithm (or HBA).This approach used to predict error rates and associated penalty costs. These costs may be dramatically different in medical applications as the implications of having a false-positive and a false-negative case may be tremendously

THEORITICAL ANALYSIS

This data set contains **416 liver patient** records and **167 non-liver patient** records collected from **North East of Andhra Pradesh, India**. The "Dataset" column is a class label used to divide groups into a liver patient (liver disease) or not (no disease). This data set contains **441 male patient** records and **142 female patient records.**

Note: We have not started any data analysis yet, this is just to show you all the authenticity of the dataset**.**

3.2 HARDWARE / SOFTWARE DESIGNING

The hardware required for the development of this project is:

Processor : Intel CoreTM i5-9300H

Processor speed : 2.4GHz

RAM Size : 8 GB DDR

System Type : X64-based processor

SOFTWARE DESIGNING:

The software required for the development of this project is:

Desktop GUI : Anaconda Navigator

Operating system : Windows 10

Front end : HTML, CSS,JAVASCRIPT

Programming : PYTHON

Cloud Computing Service : IBM Cloud Services

EXPERIMENTAL INVESTIGATION IMPORTING AND READING THE DATASET

Importing the Libraries First step is usually importing the libraries that will be needed in the program.

Pandas: It is a python library mainly used for data manipulation.

NumPy: This python library is used for numerical analysis.

Matplotlib and Seaborn: Both are the data visualization library used for plotting graph which will help us for understanding the data. csr_matrix() :A dense matrix stored in a NumPy array can be converted into a sparse matrix using the CSR representation by calling the csr_matrix() function.

Train_test_split: used for splitting data arrays into training data and for testing data. Pickle: to serialize your machine learning algorithms and save the serialized format to a file.

Reading the Dataset

For this project, we make use of three different datasets (Books_Ratings, Books, Users). We will be selecting the important features from these datasets that will help us in recommending the best results.

The next step is to read the dataset into a data structure that's compatible with pandas. Let's load a .csv data file into pandas. There is a function for it, called read_csv().We will need to locate the directory of the CSV file at first (it's more efficient to keep the dataset in the same directory as your program).If the dataset in same directory of your program, you can directly read it, without any path.

After the next Steps we made following bellow:

1.Data visualization

2.Collabrative and filtering

3.Creating the Model

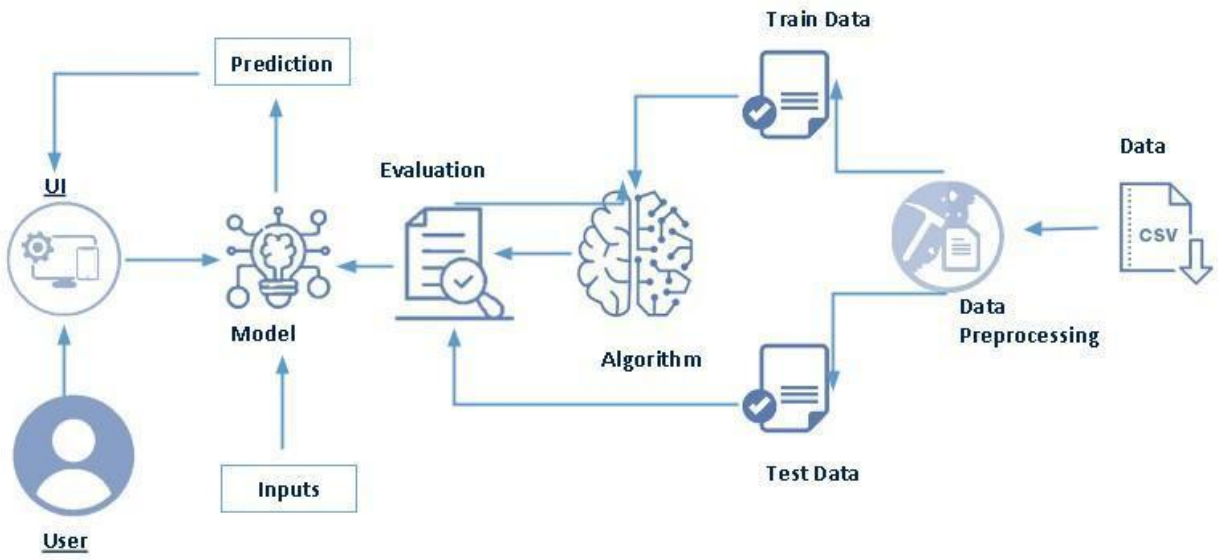
4.Test and save the model

5.Buil Python Code

6.Build HTML Code

7.Run the Application We are the following above sections we did and investigate it.

FLOW CHART



Project Flow:

User interacts with the UI (User Interface) to upload the input features. • Uploaded features/input is analysed by the model which is integrated. Once a model analyses the uploaded inputs, the prediction is showcased on the UI.

1. Data collection

• Collect the dataset or create the dataset

• Visualizing and analyzing data

• Importing Libraries

• Read the DataSet

2. Data pre-processing

• Checking for null values

• Handling outlier

• Handling categorical data

• Splitting data into train and test

3. Model building

• Import the model building libraries

• Initializing the model

• Training and testing the model

• Evaluating performance of model

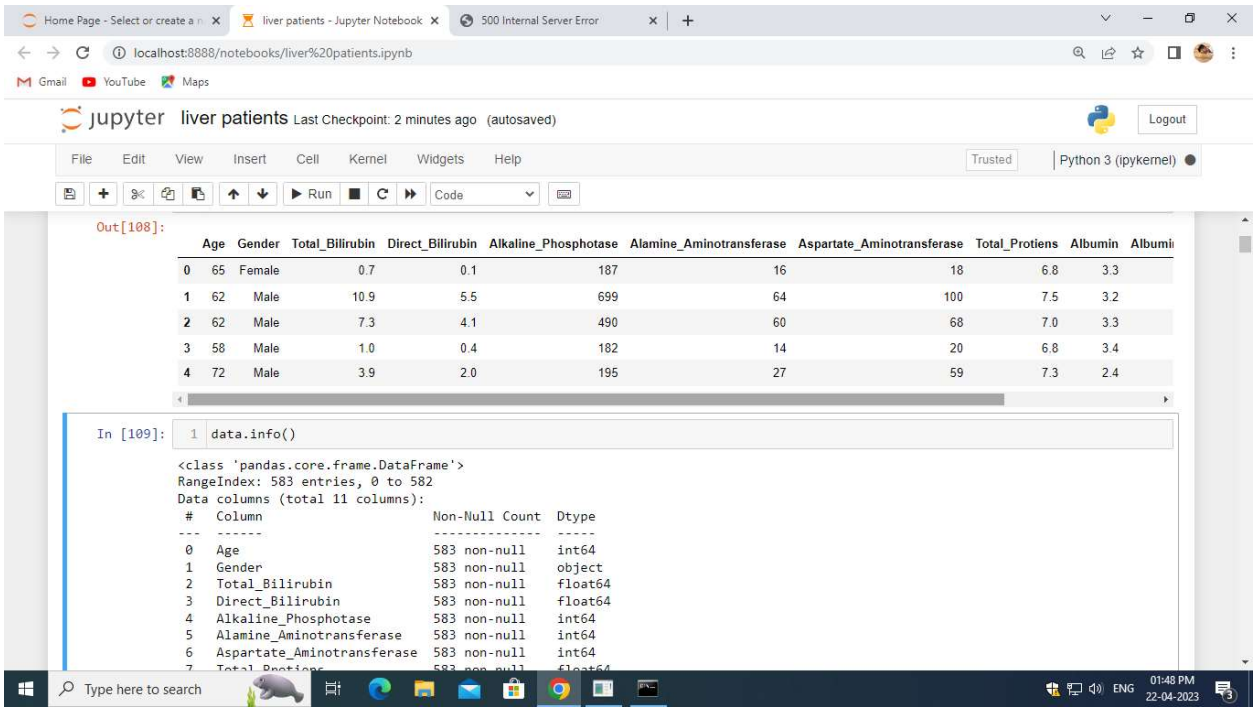
• Save the model

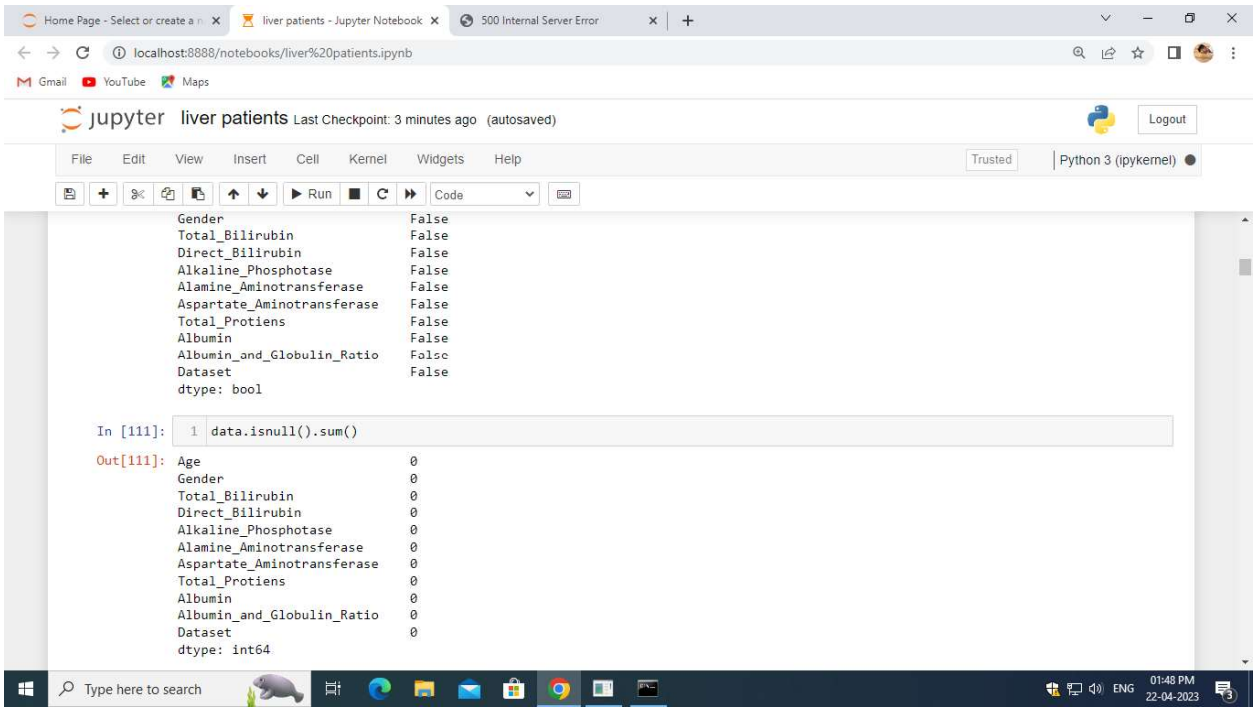
4. Application Building

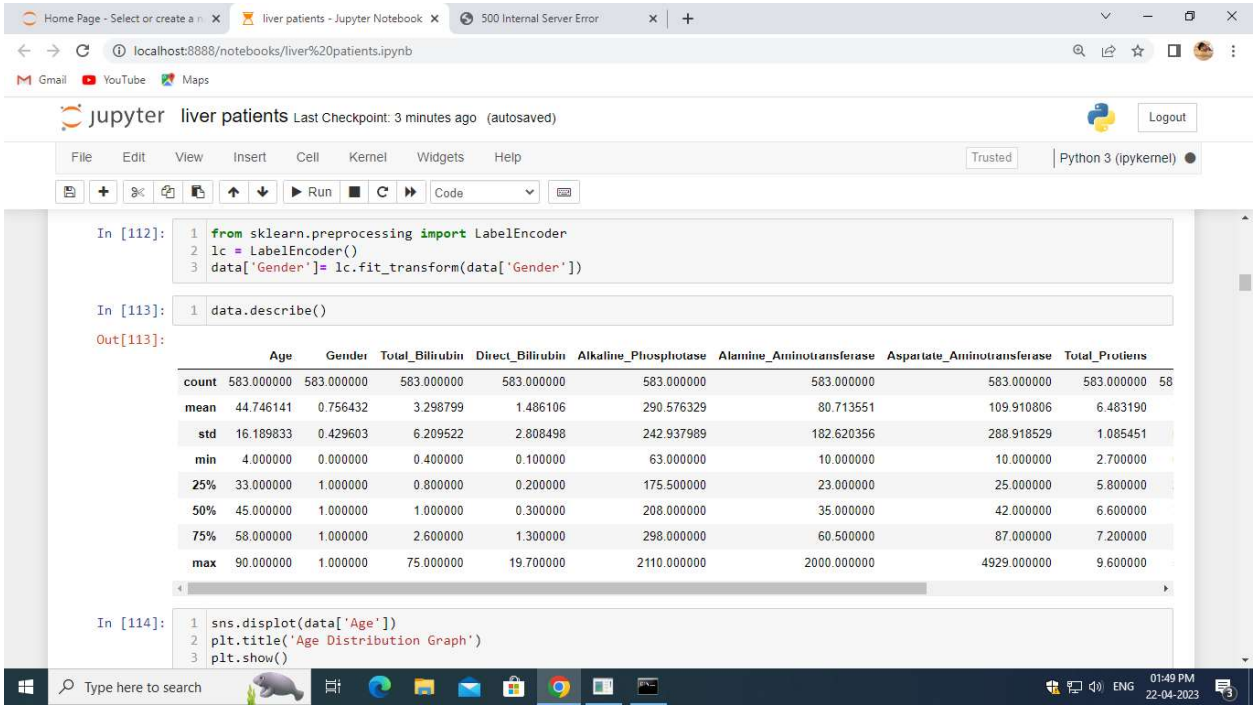
• Create an HTML file

• Build python code

RESULT




Home Page - Select or create a n
liver patients - Jupyter Notebook
500 Internal Server Error
localhost:8888/notebooks/liver%20patients.ipynb
Gmail
YouTube
Maps
jupyter liver patients Last Checkpoint: 3 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Gender False
Total_Bilirubin False
Direct_Bilirubin False
Alkaline_Phosphotase False
Alamine_Aminotransferase False
Aspartate_Aminotransferase False
Total_Protiens False
Albumin False
Albumin_and_Globulin_Ratio False
Dataset False
dtype: bool
In [111]:
data.isnull().sum()
Out[111]:
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
dtype: int64
Type here to search
ENG
01:48 PM
22-04-2023


Home Page - Select or create a n
liver patients - Jupyter Notebook
500 Internal Server Error
localhost:8888/notebooks/liver%20patients.ipynb
Gmail
YouTube
Maps
jupyter liver patients Last Checkpoint: 3 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [112]:
from sklearn.preprocessing import LabelEncoder
lc = LabelEncoder()
data['Gender']= lc.fit_transform(data['Gender'])
In [113]:
data.describe()
Out[113]:
Age Gender Total_Bilirubin Direct_Bilirubin Alkaline_Phosphotase Alamine_Aminotransferase Aspartate_Aminotransferase Total_Protiens
count 583.000000 583.000000 583.000000 583.000000 583.000000 583.000000 583.000000 583.000000
mean 44.746141 0.756432 3.298799 1.486106 290.576329 80.713551 109.910806 6.483190
std 16.189833 0.429603 6.209522 2.808498 242.937989 182.620356 288.918529 1.085451
min 4.000000 0.000000 0.400000 0.100000 63.000000 10.000000 10.000000 2.700000
25% 33.000000 1.000000 0.800000 0.200000 175.500000 23.000000 25.000000 5.800000
50% 45.000000 1.000000 1.000000 0.300000 208.000000 35.000000 42.000000 6.600000
75% 58.000000 1.000000 2.600000 1.300000 298.000000 60.500000 87.000000 7.200000
max 90.000000 1.000000 75.000000 19.700000 2110.000000 2000.000000 4929.000000 9.600000
In [114]:
sns.displot(data['Age'])
plt.title('Age Distribution Graph')
plt.show()
Type here to search
ENG
01:49 PM
22-04-2023

Home Page - Select or create a n | liver patients - Jupyter Notebook | 500 Internal Server Error

localhost:8888/notebooks/liver%20patients.ipynb

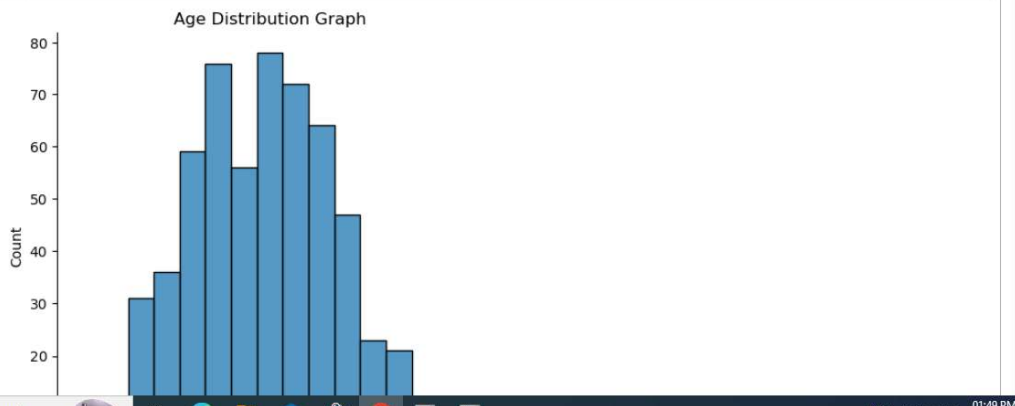
Gmail YouTube Maps

jupyter liver patients Last Checkpoint: 3 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

In [114]:
```
sns.displot(data['Age'])
plt.title('Age Distribution Graph')
plt.show()
```



Type here to search ENG 01:49 PM 22-04-2023

Home Page - Select or create a n | liver patients - Jupyter Notebook | 500 Internal Server Error

localhost:8888/notebooks/liver%20patients.ipynb

Gmail YouTube Maps

jupyter liver patients Last Checkpoint: 3 minutes ago (autosaved) Logout

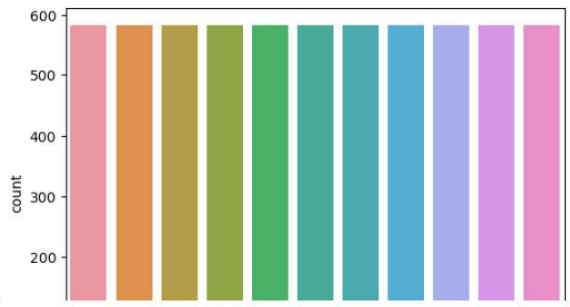
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

Age

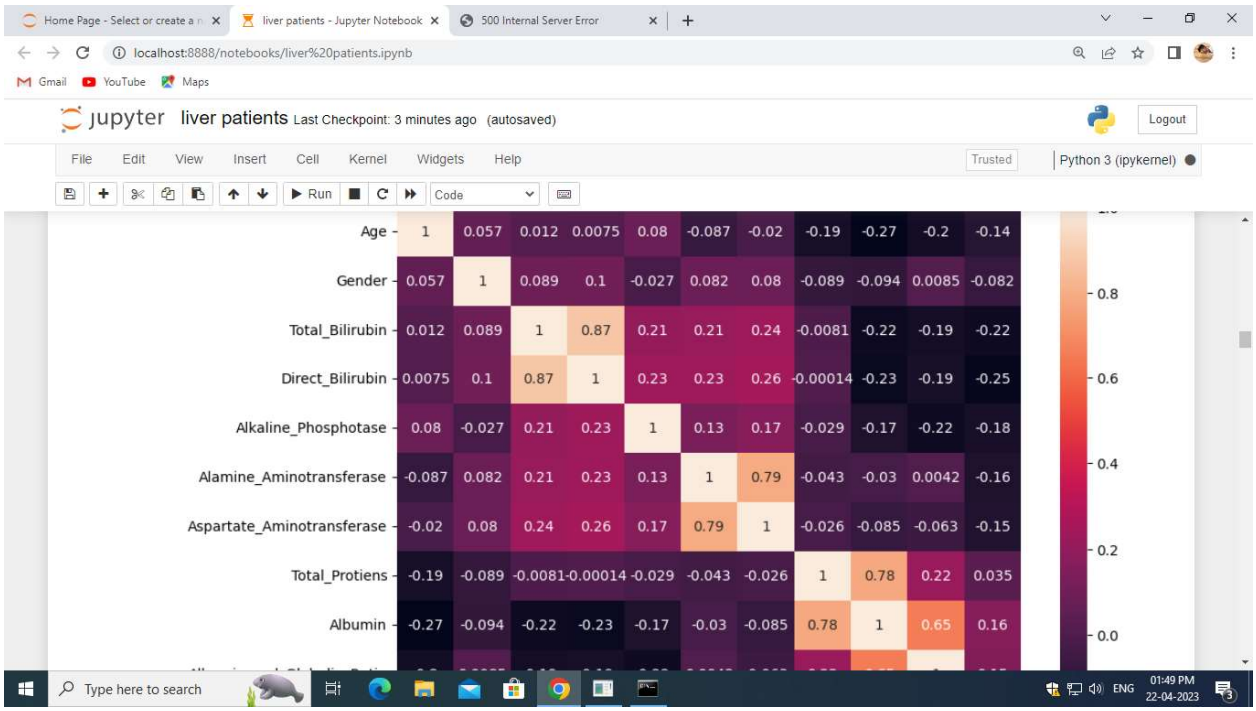
In [115]:
```
sns.countplot(data)
```

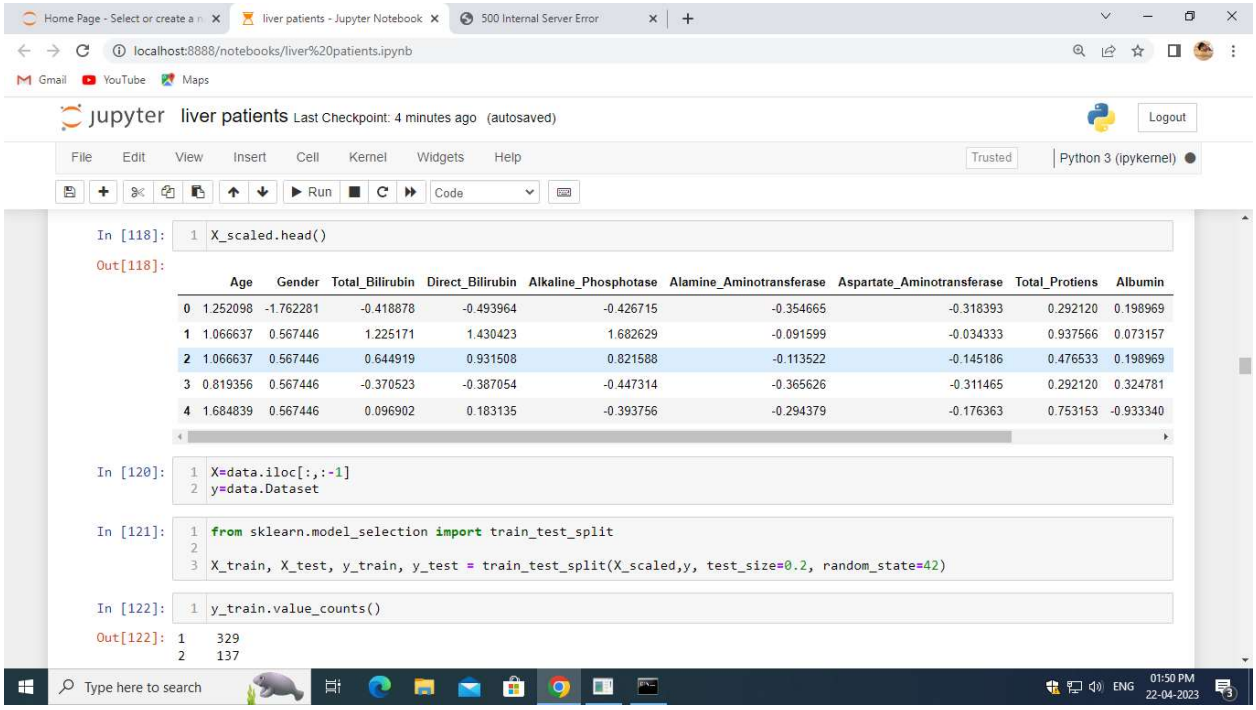
Out[115]: <Axes: ylabel='count'>

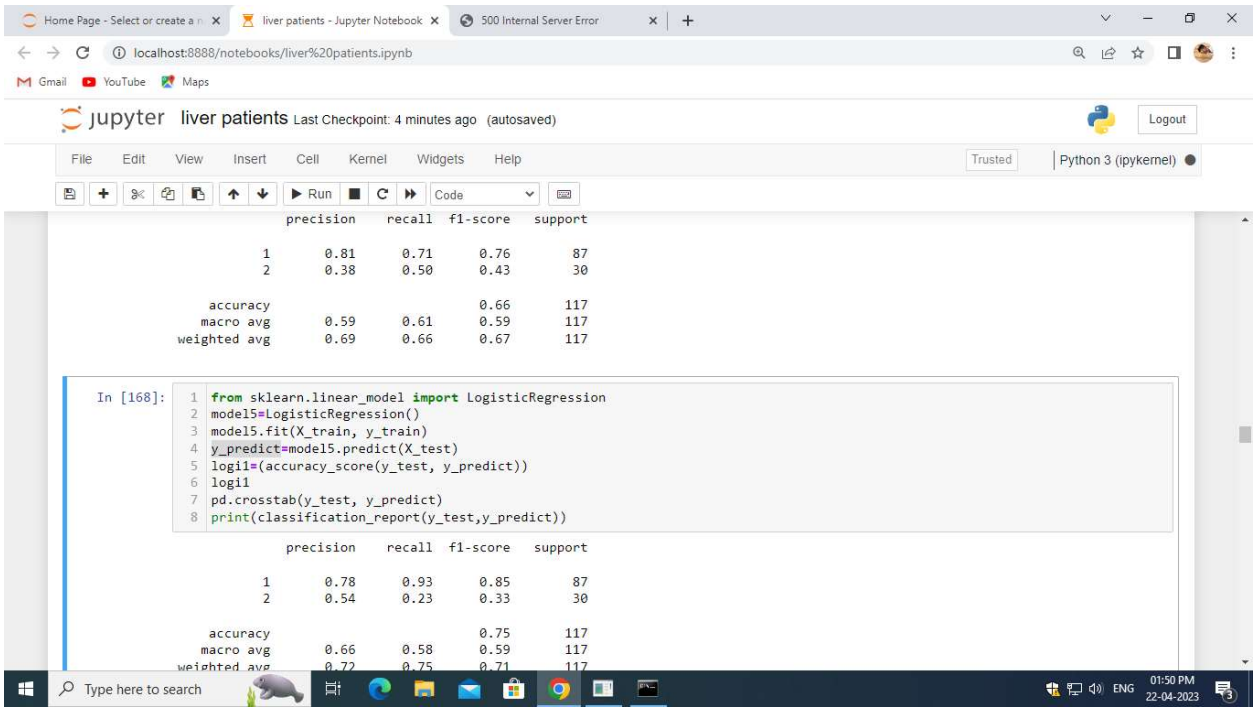


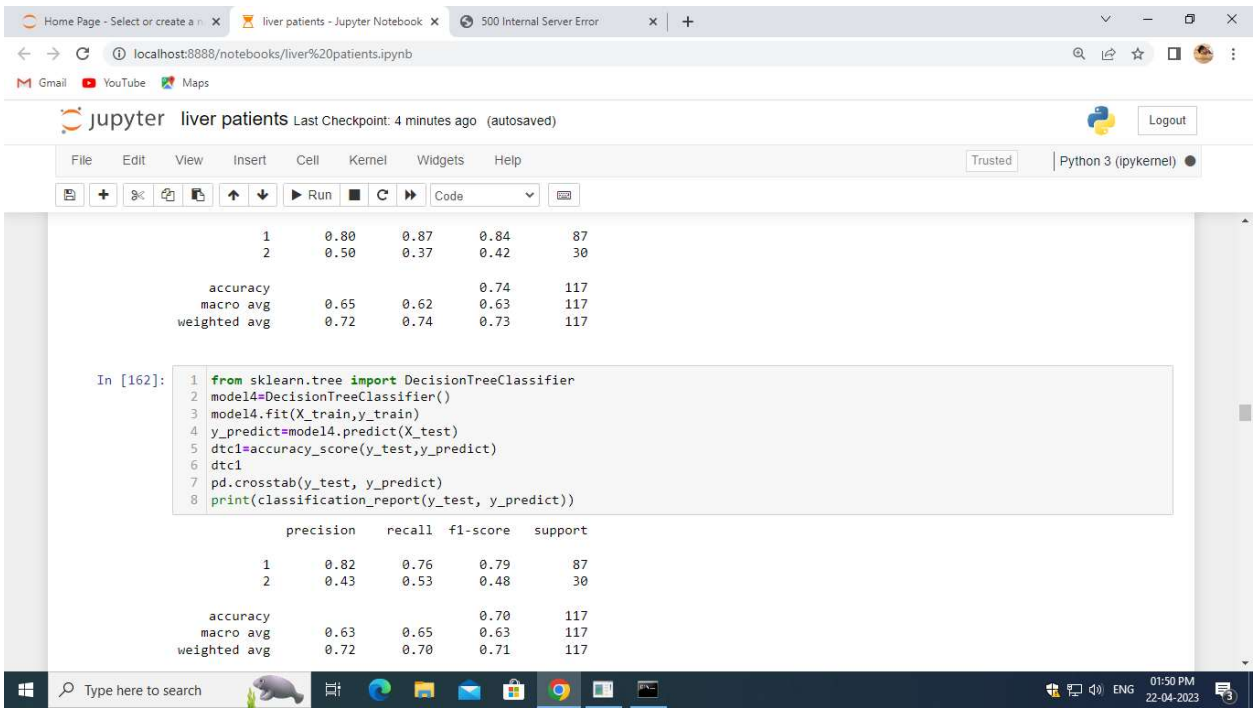
In [116]:
```
plt.figure(figsize=(10,7))
```

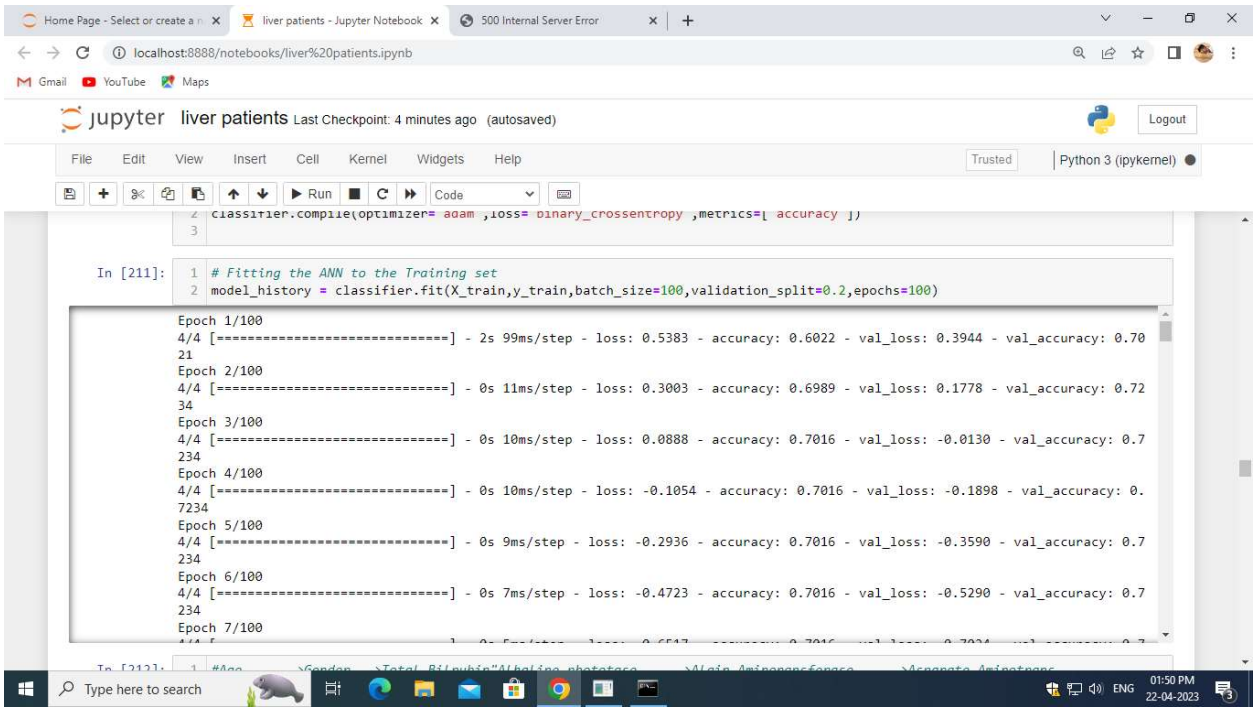
Type here to search ENG 01:49 PM 22-04-2023

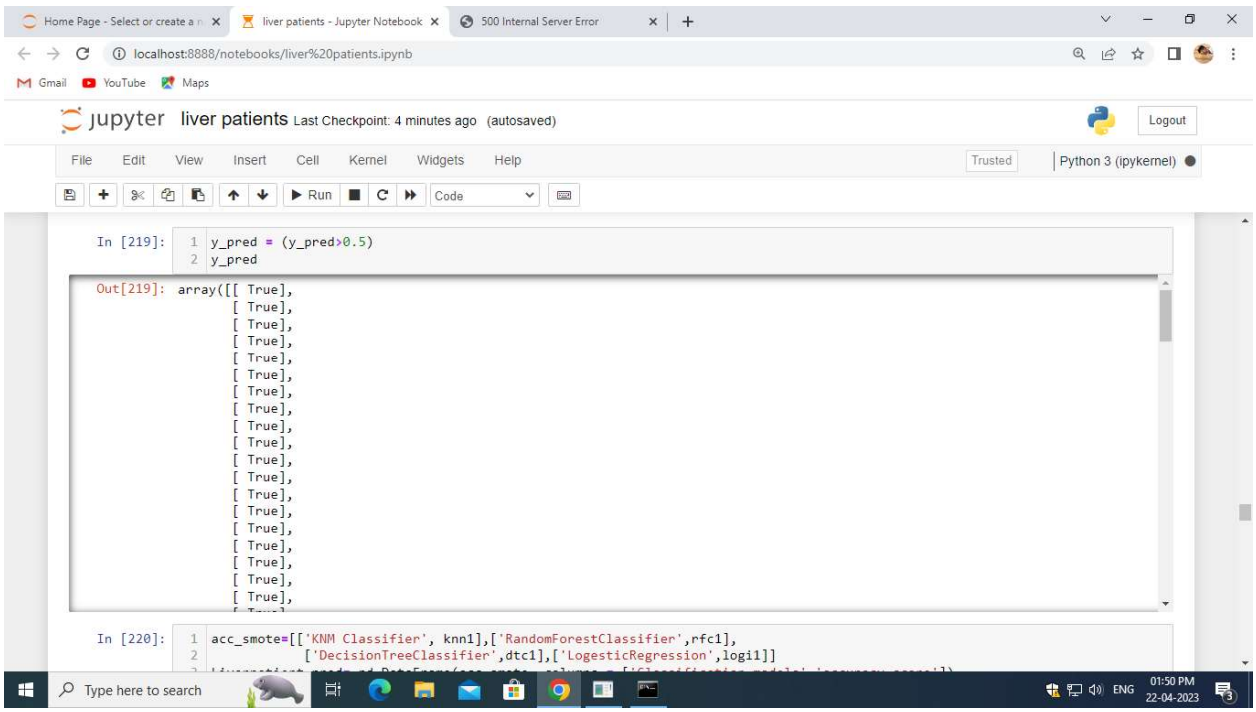

Home Page - Select or create a n
liver patients - Jupyter Notebook
500 Internal Server Error
localhost:8888/notebooks/liver%20patients.ipynb
Gmail
YouTube
Maps
jupyter liver patients Last Checkpoint: 3 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Age 1 0.057 0.012 0.0075 0.08 -0.087 -0.02 -0.19 -0.27 -0.2 -0.14
Gender 0.057 1 0.089 0.1 -0.027 0.082 0.08 -0.089 -0.094 0.0085 -0.082
Total_Bilirubin 0.012 0.089 1 0.87 0.21 0.21 0.24 -0.0081 -0.22 -0.19 -0.22
Direct_Bilirubin 0.0075 0.1 0.87 1 0.23 0.23 0.26 -0.00014 -0.23 -0.19 -0.25
Alkaline_Phosphotase 0.08 -0.027 0.21 0.23 1 0.13 0.17 -0.029 -0.17 -0.22 -0.18
Alamine_Aminotransferase -0.087 0.082 0.21 0.23 0.13 1 0.79 -0.043 -0.03 0.0042 -0.16
Aspartate_Aminotransferase -0.02 0.08 0.24 0.26 0.17 0.79 1 -0.026 -0.085 -0.063 -0.15
Total_Protiens -0.19 -0.089 -0.0081 -0.00014 -0.029 -0.043 -0.026 1 0.78 0.22 0.035
Albumin -0.27 -0.094 -0.22 -0.23 -0.17 -0.03 -0.085 0.78 1 0.65 0.16
0.8
0.6
0.4
0.2
0.0
Type here to search
ENG
01:49 PM
22-04-2023


Home Page - Select or create a n
liver patients - Jupyter Notebook
500 Internal Server Error
localhost:8888/notebooks/liver%20patients.ipynb
Gmail
YouTube
Maps
jupyter liver patients Last Checkpoint: 4 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [118]:
X_scaled.head()
Out[118]:
Age Gender Total_Bilirubin Direct_Bilirubin Alkaline_Phosphotase Alamine_Aminotransferase Aspartate_Aminotransferase Total_Protiens Albumin
0 1.252098 -1.762281 -0.418878 -0.493964 -0.426715 -0.354665 -0.318393 0.292120 0.198969
1 1.066637 0.567446 1.225171 1.430423 1.682629 -0.091599 -0.034333 0.937566 0.073157
2 1.066637 0.567446 0.644919 0.931508 0.821588 -0.113522 -0.145186 0.476533 0.198969
3 0.819356 0.567446 -0.370523 -0.387054 -0.447314 -0.365626 -0.311465 0.292120 0.324781
4 1.684839 0.567446 0.096902 0.183135 -0.393756 -0.294379 -0.176363 0.753153 -0.933340
In [120]:
X=data.iloc[:,:-1]
y=data.Dataset
In [121]:
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)
In [122]:
y_train.value_counts()
Out[122]:
1 329
2 137
Type here to search
ENG
01:50 PM
22-04-2023


Home Page - Select or create a n
liver patients - Jupyter Notebook
500 Internal Server Error
localhost:8888/notebooks/liver%20patients.ipynb
Gmail
YouTube
Maps
jupyter liver patients Last Checkpoint: 4 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
precision recall f1-score support
1 0.81 0.71 0.76 87
2 0.38 0.50 0.43 30
accuracy 0.66 117
macro avg 0.59 0.61 0.59 117
weighted avg 0.69 0.66 0.67 117
In [168]:
from sklearn.linear_model import LogisticRegression
model5=LogisticRegression()
model5.fit(X_train, y_train)
y_predict=model5.predict(X_test)
logi1=(accuracy_score(y_test, y_predict))
logi1
pd.crosstab(y_test, y_predict)
print(classification_report(y_test,y_predict))
precision recall f1-score support
1 0.78 0.93 0.85 87
2 0.54 0.23 0.33 30
accuracy 0.75 117
macro avg 0.66 0.58 0.59 117
Type here to search
ENG
01:50 PM
22-04-2023


Home Page - Select or create a n
liver patients - Jupyter Notebook
500 Internal Server Error
localhost:8888/notebooks/liver%20patients.ipynb
Gmail
YouTube
Maps
jupyter liver patients Last Checkpoint: 4 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
1 0.80 0.87 0.84 87
2 0.50 0.37 0.42 30
accuracy 0.74 117
macro avg 0.65 0.62 0.63 117
weighted avg 0.72 0.74 0.73 117
In [162]:
from sklearn.tree import DecisionTreeClassifier
model4=DecisionTreeClassifier()
model4.fit(X_train,y_train)
y_predict=model4.predict(X_test)
dtc1=accuracy_score(y_test,y_predict)
dtc1
pd.crosstab(y_test, y_predict)
print(classification_report(y_test, y_predict))
precision recall f1-score support
1 0.82 0.76 0.79 87
2 0.43 0.53 0.48 30
accuracy 0.70 117
macro avg 0.63 0.65 0.63 117
weighted avg 0.72 0.70 0.71 117
Type here to search
ENG
01:50 PM
22-04-2023


Home Page - Select or create a n
liver patients - Jupyter Notebook
500 Internal Server Error
localhost:8888/notebooks/liver%20patients.ipynb
Gmail
YouTube
Maps
jupyter liver patients Last Checkpoint: 4 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
classifier.compile(optimizer= 'adam' ,loss= 'binary_crossentropy' ,metrics=['accuracy'])
In [211]:
# Fitting the ANN to the Training set
model_history = classifier.fit(X_train,y_train,batch_size=100,validation_split=0.2,epochs=100)
Epoch 1/100
4/4 [==========] - 2s 99ms/step - loss: 0.5383 - accuracy: 0.6022 - val_loss: 0.3944 - val_accuracy: 0.7021
Epoch 2/100
4/4 [==========] - 0s 11ms/step - loss: 0.3003 - accuracy: 0.6989 - val_loss: 0.1778 - val_accuracy: 0.7234
Epoch 3/100
4/4 [==========] - 0s 10ms/step - loss: 0.0888 - accuracy: 0.7016 - val_loss: -0.0130 - val_accuracy: 0.7234
Epoch 4/100
4/4 [==========] - 0s 10ms/step - loss: -0.1054 - accuracy: 0.7016 - val_loss: -0.1898 - val_accuracy: 0.7234
Epoch 5/100
4/4 [==========] - 0s 9ms/step - loss: -0.2936 - accuracy: 0.7016 - val_loss: -0.3590 - val_accuracy: 0.7234
Epoch 6/100
4/4 [==========] - 0s 7ms/step - loss: -0.4723 - accuracy: 0.7016 - val_loss: -0.5290 - val_accuracy: 0.7234
Epoch 7/100
Type here to search
ENG
01:50 PM
22-04-2023


Home Page - Select or create a n
liver patients - Jupyter Notebook
500 Internal Server Error
localhost:8888/notebooks/liver%20patients.ipynb
Gmail
YouTube
Maps
jupyter liver patients Last Checkpoint: 4 minutes ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [219]:
y_pred = (y_pred>0.5)
y_pred
Out[219]:
array([[True],
[True],
[True],
[True],
[True],
[True],
[True],
[True],
[True],
[True],
[True],
[True],
[True],
[True],
[True],
[True],
[True],
[True],
[True],
In [220]:
acc_smote=[['KNM Classifier', knn1],['RandomForestClassifier',rfc1],
['DecisionTreeClassifier',dtc1],['LogesticRegression',logi1]]
Type here to search
ENG
01:50 PM
22-04-2023

Home Page - Select or create a n | liver patients - Jupyter Notebook | 500 Internal Server Error

localhost:8888/notebooks/liver%20patients.ipynb

Gmail YouTube Maps

jupyter liver patients Last Checkpoint: 4 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

```
 [ True],
 [ True],
 [ True],
 [ True],
 [ True],
 [ True],
 [ True],
```

In [220]:
```
acc_smote=[['KNM Classifier', knn1],['RandomForestClassifier',rfc1],
           ['DecisionTreeClassifier',dtc1],['LogesticRegression',logi1]]
Liverpatient_pred= pd.DataFrame(acc_smote, columns = ['Classification models','accuracy_score'])
Liverpatient_pred
```

Out[220]:

	Classification models	accuracy_score
0	KNM Classifier	0.658120
1	RandomForestClassifier	0.743590
2	DecisionTreeClassifier	0.700855
3	LogesticRegression	0.752137

In [223]:
```
plt.figure(figsize=(7,5))
plt.xticks(rotation=90)
plt.title('Classification models & accuracy scores after SMOTE',fontsize=18)
sns.barplot(Liverpatient_pred)
```

Type here to search ENG 01:50 PM 22-04-2023

Home Page - Select or create a n | liver patients - Jupyter Notebook | 500 Internal Server Error

localhost:8888/notebooks/liver%20patients.ipynb

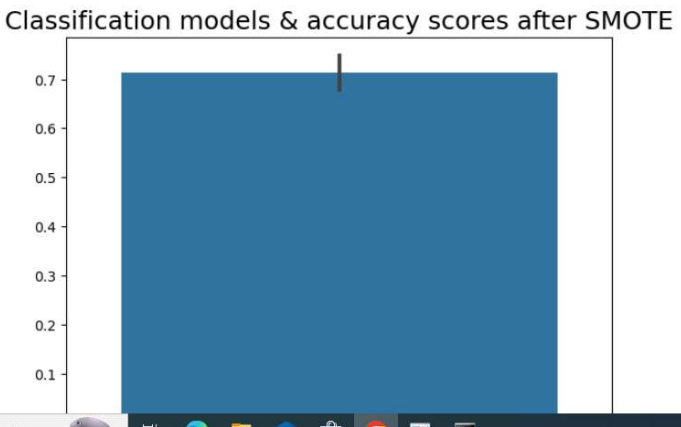
Gmail YouTube Maps

jupyter liver patients Last Checkpoint: 4 minutes ago (autosaved) Logout

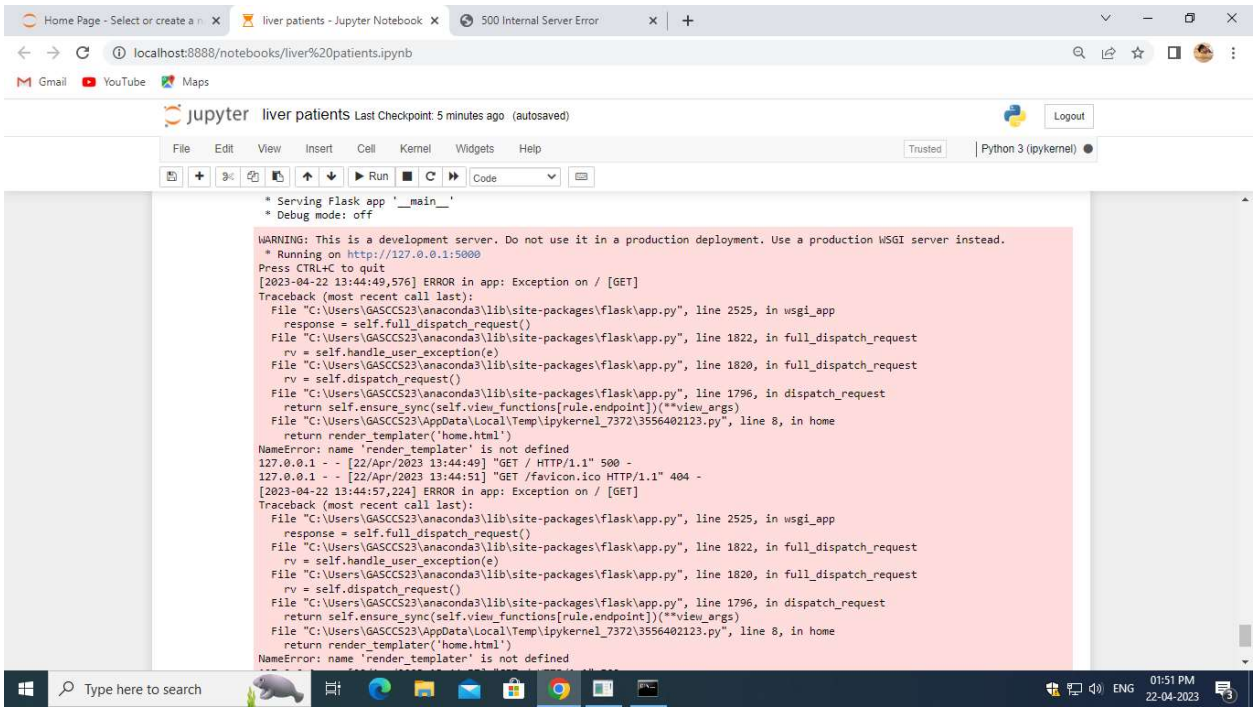
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

Out[223]: <Axes: title={'center': 'Classification models & accuracy scores after SMOTE'}>



Type here to search ENG 01:50 PM 22-04-2023



CONCLUSION

The comparative analysis employed with the machine learning algorithms such as Logistic Regression, Support Vector Machine, Random Forest and Decision Tree. These algorithms are used to predict the liver disease at an early stage. These algorithms were evaluated and compared based on performance metrics such as accuracy, precision, specificity, sensitivity. From the analysis, logistic regression outperforms well than the other algorithms with slight variation and its achieved accuracy is 81.9%. This comparative analysis will help to predict the liver disease and will benefit in managing the health of the individuals.