### Milestone 2.Data collection and preparation 1.collect the dataset 1.1 Importing the libraries

```
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
import pickle
```

### 1.2 Read the dataset

```
from google.colab import files
uploaaded=files.upload()
```

Choose Files indian_liver_patient.csv

- **indian_liver_patient.csv**(text/csv) - 23930 bytes, last modified: 9/21/2019 - 100% done

```
Saving indian_liver_patient.csv to indian_liver_patient (1).csv
```

```
import pandas as pd
data=pd.read_csv('indian_liver_patient.csv')
```

```
data.head()
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

| | Age | Gender | Total_Bilirubin | Direct_Bilirubin | Alkaline_Phosphotase | Alamine_Aminotransferase | Aspartate_Aminotrans |
|---|---|---|---|---|---|---|---|
| **0** | 65 | Female | 0.7 | 0.1 | 187 | 16 | |
| **1** | 62 | Male | 10.9 | 5.5 | 699 | 64 | |

**

```
data.tail()
```

| | Age | Gender | Total_Bilirubin | Direct_Bilirubin | Alkaline_Phosphotase | Alamine_Aminotransferase | Aspartate_Aminotra |
|---|---|---|---|---|---|---|---|
| **578** | 60 | Male | 0.5 | 0.1 | 500 | 20 | |
| **579** | 40 | Male | 0.6 | 0.1 | 98 | 35 | |
| **580** | 52 | Male | 0.8 | 0.2 | 245 | 48 | |
| **581** | 31 | Male | 1.3 | 0.5 | 184 | 29 | |
| **582** | 38 | Male | 1.0 | 0.3 | 216 | 21 | |

```
data.describe()
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

| | Age | Total_Bilirubin | Direct_Bilirubin | Alkaline_Phosphotase | Alamine_Aminotransferase | Aspartate_Aminotr |
|---|---|---|---|---|---|---|
| **count** | 583.000000 | 583.000000 | 583.000000 | 583.000000 | 583.000000 | 5 |
| **mean** | 44.746141 | 3.298799 | 1.486106 | 290.576329 | 80.713551 | 1 |
| **std** | 16.189833 | 6.209522 | 2.808498 | 242.937989 | 182.620356 | 2 |

## 2. Data preparation 2.1 Handling missing values

```
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
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
Age                           False
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     True
Dataset                       False
dtype: bool
```

```
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
```

```
data[data['Dataset']==1]
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

| | Age | Gender | Total_Bilirubin | Direct_Bilirubin | Alkaline_Phosphotase | Alamine_Aminotransferase | Aspartate_Aminotra |
|---|---|---|---|---|---|---|---|
| **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 | |
| **...** | ... | ... | ... | ... | ... | ... | |
| **576** | 32 | Male | 15.0 | 8.2 | 289 | 58 | |
| **577** | 32 | Male | 12.7 | 8.4 | 190 | 28 | |

```
data_1=data.dropna()
```

```
data_1.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    0
Dataset                       0
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

## 2.2 Handling categorical values

```
from sklearn.preprocessing import LabelEncoder
lc = LabelEncoder()
```

**Converting textual data into num

```
data['Gender']=lc.fit_transform(data['Gender'])
```

```
data.head()
```

| | Age | Gender | Total_Bilirubin | Direct_Bilirubin | Alkaline_Phosphotase | Alamine_Aminotransferase | Aspartate_Aminotrans |
|---|---|---|---|---|---|---|---|
| **0** | 65 | 0 | 0.7 | 0.1 | 187 | 16 | |
| **1** | 62 | 1 | 10.9 | 5.5 | 699 | 64 | |
| **2** | 62 | 1 | 7.3 | 4.1 | 490 | 60 | |
| **3** | 58 | 1 | 1.0 | 0.4 | 182 | 14 | |
| **4** | 72 | 1 | 3.9 | 2.0 | 195 | 27 | |

```
data.info()
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
Data columns (total 11 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Age                         583 non-null    int64
 1   Gender                      583 non-null    int64
 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(6)
memory usage: 50.2 KB
```

### Milestone 3. Exploratory data analysis 1. Descriptive statistical

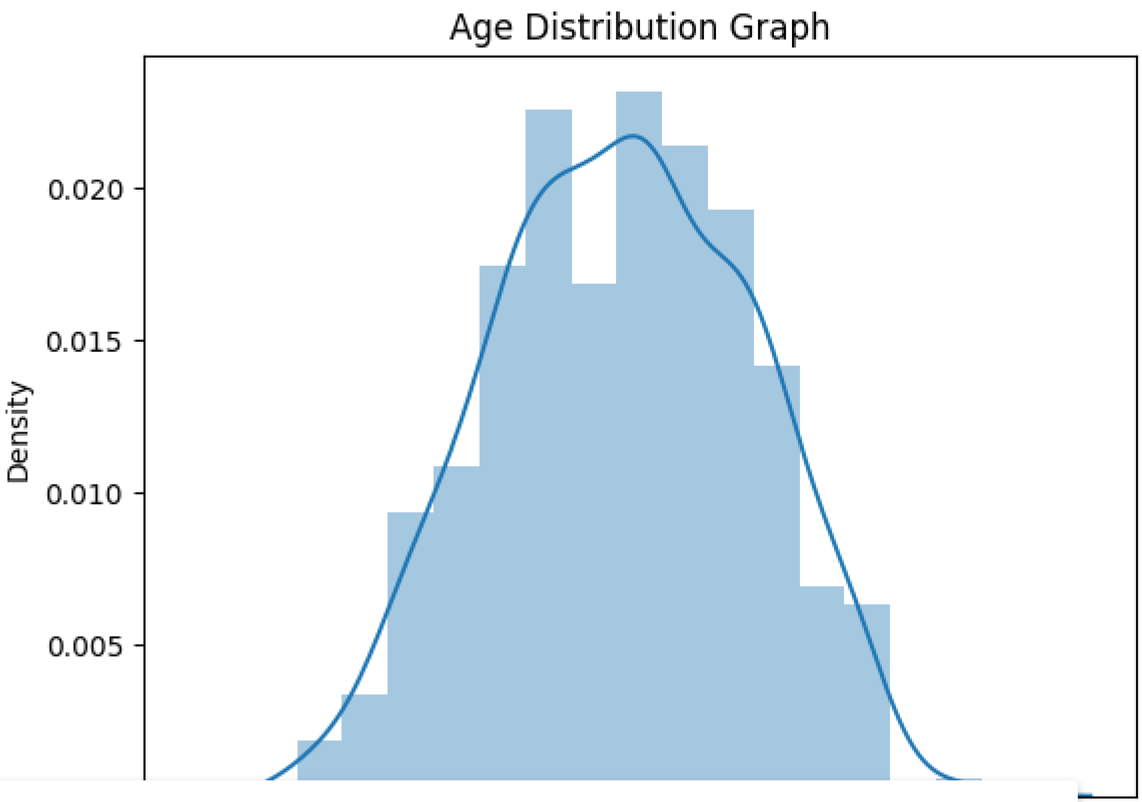
```
data.describe()
```

| | Age | Gender | Total_Bilirubin | Direct_Bilirubin | Alkaline_Phosphotase | Alamine_Aminotransferase | Aspart |
|---|---|---|---|---|---|---|---|
| **count** | 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 | |
| **std** | 16.189833 | 0.429603 | 6.209522 | 2.808498 | 242.937989 | 182.620356 | |
| **min** | 4.000000 | 0.000000 | 0.400000 | 0.100000 | 63.000000 | 10.000000 | |
| **25%** | 33.000000 | 1.000000 | 0.800000 | 0.200000 | 175.500000 | 23.000000 | |
| **50%** | 45.000000 | 1.000000 | 1.000000 | 0.300000 | 208.000000 | 35.000000 | |
| **75%** | 58.000000 | 1.000000 | 2.600000 | 1.300000 | 298.000000 | 60.500000 | |
| **max** | 90.000000 | 1.000000 | 75.000000 | 19.700000 | 2110.000000 | 2000.000000 | |

Automatic saving failed. This file was updated remotely or in another tab. Show diff

### 2. Visual analysis 2.1 Univariate analysis

```
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
sns.distplot(data_1['Age'])
plt.title('Age Distribution Graph')
plt.show()
```

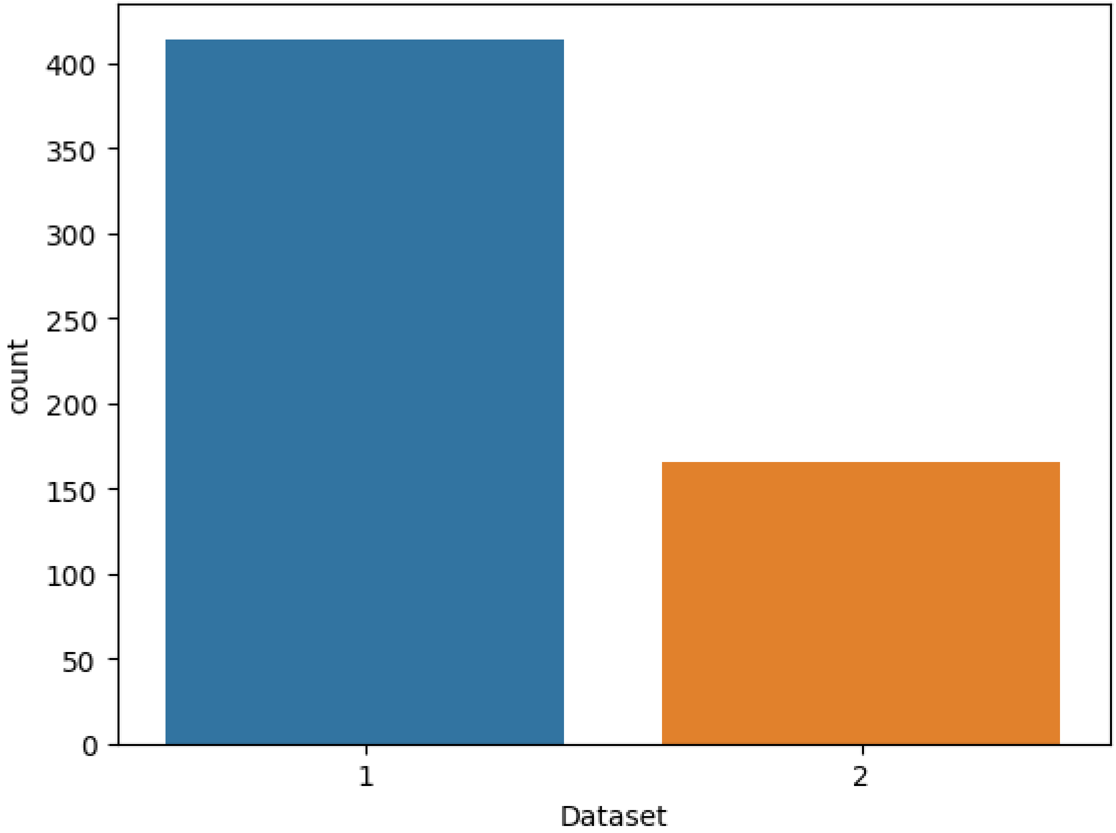


Automatic saving failed. This file was updated remotely or in another tab. Show diff

## counting patients who are dignosed and not diagnosed with liver disease

```
sns.countplot(data=data_1, x='Dataset')
LD,NLD=data_1['Dataset'].value_counts()
print("Liver disease patients:", LD)
print("NOn-liver disease patients:",NLD)
```

```
Liver disease patients: 414
NOn-liver disease patients: 165
```



Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
import seaborn as sns
import matplotlib.pyplot as plt
```

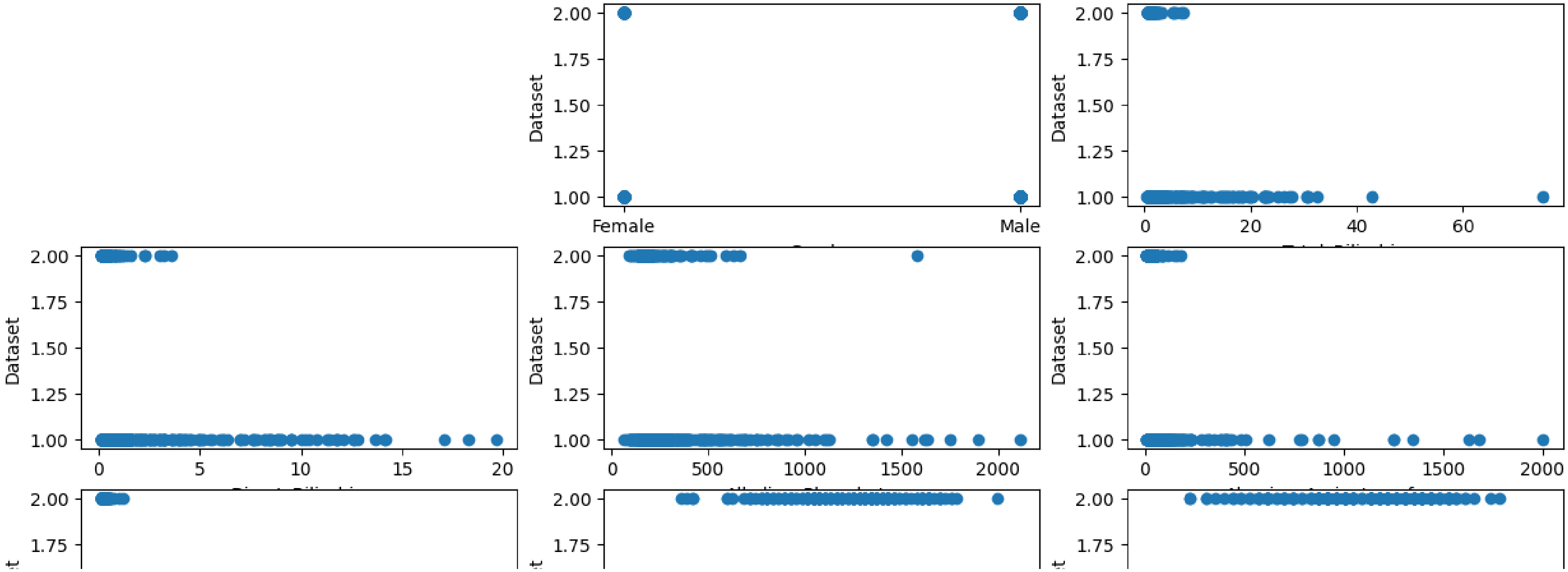
```
plt.figure(figsize=(15,7))
plt.subplot(1,3,1)
plt.scatter(data_1['Age'], data_1['Dataset'])
plt.ylabel('Dataset')
plt.xlabel('Age')


plt.subplot(3,3,2)
plt.scatter(data_1['Gender'], data_1['Dataset'])
plt.ylabel('Dataset')
plt.xlabel('Gender')


plt.subplot(3,3,3)
plt.scatter(data_1['Total_Bilirubin'], data_1['Dataset'])
plt.ylabel('Dataset')
plt.xlabel('Total_Bilirubin')


plt.subplot(3,3,4)
plt.scatter(data_1['Direct_Bilirubin'], data_1['Dataset'])
plt.ylabel('Dataset')
plt.xlabel('Direct_Bilirubin')


plt.subplot(3,3,5)
plt.scatter(data_1['Alkaline_Phosphotase'], data_1['Dataset'])
plt.ylabel('Dataset')
plt.xlabel('Alkaline_Phosphotase')
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
plt.subplot(3,3,6)
plt.scatter(data_1['Alamine_Aminotransferase'], data_1['Dataset'])
plt.ylabel('Dataset')
plt.xlabel('Alamine_Aminotransferase')
plt.subplot(3,3,7)
plt.scatter(data_1['Aspartate_Aminotransferase'], data_1['Dataset'])
plt.ylabel('Dataset')
plt.xlabel('Aspartate_Aminotransferase')

plt.subplot(3,3,8)
plt.scatter(data_1['Total_Protiens'], data_1['Dataset'])
plt.ylabel('Dataset')
plt.xlabel('Total_Protiens')



plt.subplot(3,3,9)
plt.scatter(data_1['Albumin'], data_1['Dataset'])
plt.ylabel('Dataset')
plt.xlabel('Albumin')
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
Text(0.5, 0, 'Albumin')
```

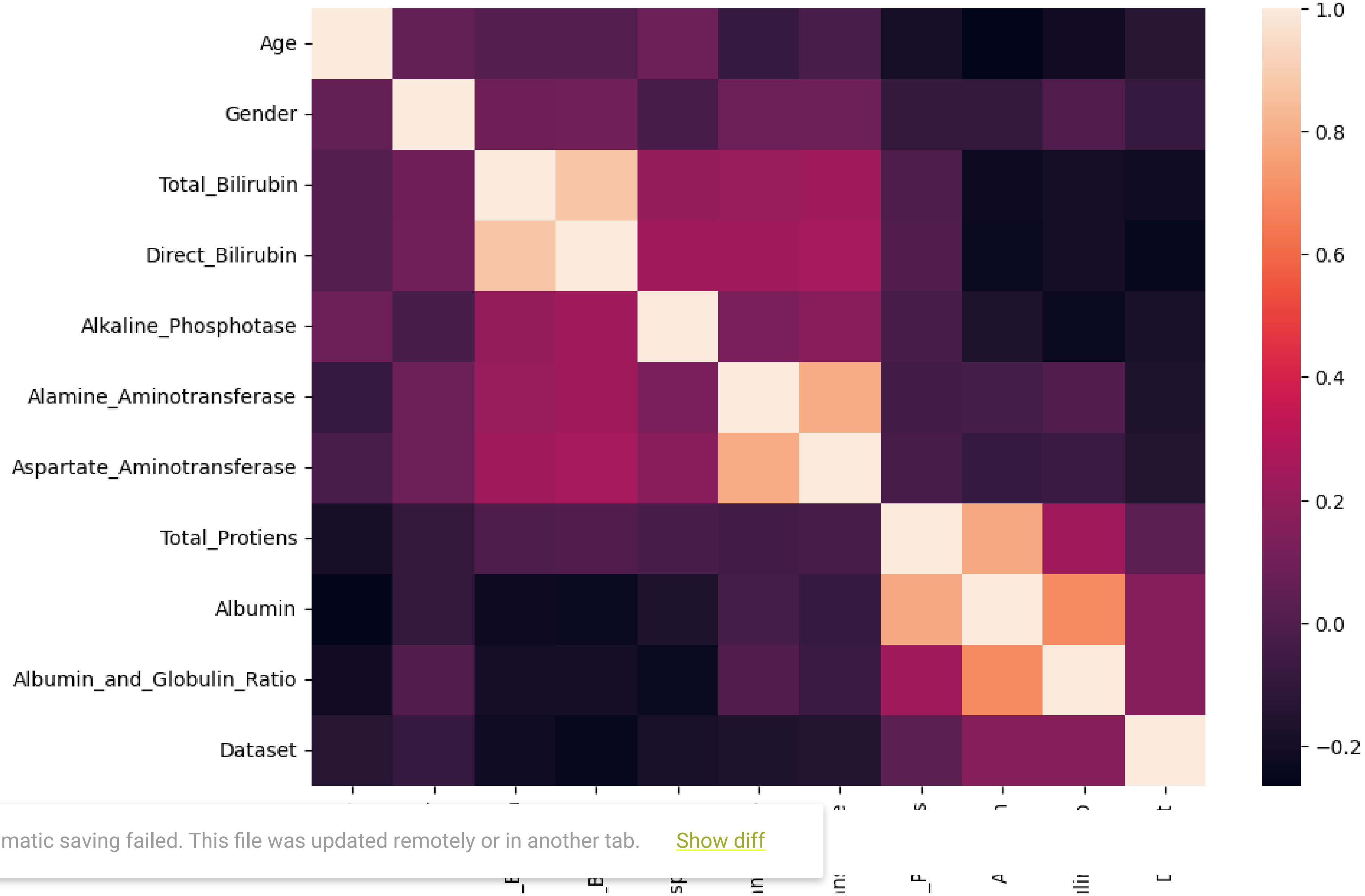


## 2.3 Multivariate analysis

```
import matplotlib.pyplot as plt
plt.figure(figsize=(10,7))
sns.heatmap(data.corr())
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
<Axes: >
```



Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
data_1['Gender'] = le.fit_transform(data_1['Gender'])
data_1.head()
```

| | Age | Gender | Total_Bilirubin | Direct_Bilirubin | Alkaline_Phosphotase | Alamine_Aminotransferase | Aspartate_Aminotrans |
|---|---|---|---|---|---|---|---|
| **0** | 65 | 0 | 0.7 | 0.1 | 187 | 16 | |
| **1** | 62 | 1 | 10.9 | 5.5 | 699 | 64 | |
| **2** | 62 | 1 | 7.3 | 4.1 | 490 | 60 | |
| **3** | 58 | 1 | 1.0 | 0.4 | 182 | 14 | |
| **4** | 72 | 1 | 3.9 | 2.0 | 195 | 27 | |

```
from sklearn.preprocessing import scale
x=data
x_scaled=pd.DataFrame(scale(x), columns=x.columns)
```

```
data.head()
```

| | Age | Gender | Total_Bilirubin | Direct_Bilirubin | Alkaline_Phosphotase | Alamine_Aminotransferase | Aspartate_Aminotrans |
|---|---|---|---|---|---|---|---|
| **0** | 65 | 0 | 0.7 | 0.1 | 187 | 16 | |
| **1** | 62 | 1 | 10.9 | 5.5 | 699 | 64 | |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 60 | |
| **3** | 58 | 1 | 1.0 | 0.4 | 182 | 14 | |
| **4** | 72 | 1 | 3.9 | 2.0 | 195 | 27 | |

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
x_scaled.head()
```

| | Age | Gender | Total_Bilirubin | Direct_Bilirubin | Alkaline_Phosphotase | Alamine_Aminotransferase | Aspartate_Ami |
|---|---|---|---|---|---|---|---|
| **0** | 1.252098 | -1.762281 | -0.418878 | -0.493964 | -0.426715 | -0.354665 | |
| **1** | 1.066637 | 0.567446 | 1.225171 | 1.430423 | 1.682629 | -0.091599 | |
| **2** | 1.066637 | 0.567446 | 0.644919 | 0.931508 | 0.821588 | -0.113522 | |
| **3** | 0.819356 | 0.567446 | -0.370523 | -0.387054 | -0.447314 | -0.365626 | |
| **4** | 1.684839 | 0.567446 | 0.096902 | 0.183135 | -0.393756 | -0.294379 | |

**~Spliting data into train and test**

# Divide the data into input and output

```
x=data_1.iloc[:,0:-1]
y=data_1.iloc[:,-1]
x
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

| | Age | Gender | Total_Bilirubin | Direct_Bilirubin | Alkaline_Phosphotase | Alamine_Aminotransferase | Aspartate_Aminotra |
|---|---|---|---|---|---|---|---|
| **0** | 65 | 0 | 0.7 | 0.1 | 187 | 16 | |
| **1** | 62 | 1 | 10.9 | 5.5 | 699 | 64 | |
| **2** | 62 | 1 | 7.3 | 4.1 | 490 | 60 | |
| **3** | 58 | 1 | 1.0 | 0.4 | 182 | 14 | |
| **4** | 72 | 1 | 3.9 | 2.0 | 195 | 27 | |
| **...** | ... | ... | ... | ... | ... | ... | |
| **578** | 60 | 1 | 0.5 | 0.1 | 500 | 20 | |
| **579** | 40 | 1 | 0.6 | 0.1 | 98 | 35 | |
| **580** | 52 | 1 | 0.8 | 0.2 | 245 | 48 | |

# importing train_test_split

```
from sklearn.model_selection import train_test_split
xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.3)
```

```
xtrain.shape
```

```
    (405, 10)
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

### 2.4 Handling imbalance data

```
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: numpy>=1.17.3 in /usr/local/lib/python3.9/dist-packages (from imbalanced-learn->imblearr
     Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.9/dist-packages (from imbalanced-learn->i
     Requirement already satisfied: scikit-learn>=1.0.2 in /usr/local/lib/python3.9/dist-packages (from imbalanced-learn->im
     Requirement already satisfied: scipy>=1.3.2 in /usr/local/lib/python3.9/dist-packages (from imbalanced-learn->imblearn)
     Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.9/dist-packages (from imbalanced-learn->imblearr
     Installing collected packages: imblearn
     Successfully installed imblearn-0.0
```

```
from imblearn.over_sampling import SMOTE
smote = SMOTE()
```

```
ytrain.value_counts()
```

```
     1    291
     2    114
     Name: Dataset, dtype: int64
```

```
xtrain.value_counts()
```

```
     Age  Gender  Total_Bilirubin  Direct_Bilirubin  Alkaline_Phosphotase  Alamine_Aminotransferase
     Aspartate_Aminotransferase  Total_Protiens  Albumin  Albumin_and_Globulin_Ratio
     38   0       2.6              1.2               410                   59                        57
     5.6             3.0      0.8                           2
     18   1       0.8              0.2               282                   72                        140
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
                                                                           32                        245
     36   1       5.3              2.3               145                   32                        92
     5.1             2.6      1.0                           2
     72   1       0.7              0.1               196                   20                        35
     5.8             2.0      0.5                           2

     ..
     37   1       1.3              0.4               195                   41                        38
```

```
        5.3             2.1      0.6                           1
                     0.8              0.2              195                     60                        40
        8.2             5.0      1.5                           1
                                                       147                     27                        46
        5.0             2.5      1.0                           1
                                                       125                     41                        39
        6.4             3.4      1.1                           1
        90   1       1.1              0.3              215                     46                        134
        6.9             3.0      0.7                           1
        Length: 399, dtype: int64
```

**Milestone 4. Model building 1. training the model in multipe algorithms 1.1 Random forest model**

## importing the classifier algorithms

```
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
```

```
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
```

## Initialize

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
RFmodel=RandomForestClassifier
KNmodel=KNeighborsClassifier
```

```
from sklearn.svm import SVC
svm=SVC()
```

```
svm.fit(xtrain, ytrain)
```

```
▾ SVC
SVC()
```

# train the data with svm

```
from sklearn.preprocessing import StandardScaler
ss=StandardScaler()
```

```
data1=ss.fit_transform(data)
```

```
data1=pd.DataFrame(data1, columns=data.columns)
data1.head()
```

| | Age | Gender | Total_Bilirubin | Direct_Bilirubin | Alkaline_Phosphotase | Alamine_Aminotransferase | Aspartate_Ami |
|---|---|---|---|---|---|---|---|
| 0 | 1.252098 | -1.762281 | -0.418878 | -0.493964 | -0.426715 | -0.354665 | |
| 1 | 1.066637 | 0.567446 | 1.225171 | 1.430423 | 1.682629 | -0.091599 | |
| | | | | | .821588 | -0.113522 | |
| | | | | | .447314 | -0.365626 | |
| 4 | 1.684839 | 0.567446 | 0.096902 | 0.183135 | -0.393756 | -0.294379 | |

Automatic saving failed. This file was updated remotely or in another tab. Show diff

# Random Forest Model

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
model1=RandomForestClassifier()
model1.fit(xtrain,ytrain)
y_predict=model1.predict(xtest)
rfc1=accuracy_score(ytest,y_predict)
rfc1
pd.crosstab(ytest,y_predict)
```

```
   col_0    1   2

 Dataset
    1     108  15
    2      35  16
```

### 1.2 Decision tree model

```
from sklearn.tree import DecisionTreeClassifier

model4=DecisionTreeClassifier()
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
dtc1=accuracy_score(ytest, y_predict)
dtc1
pd.crosstab(ytest, y_predict)
```

| col_0 | 1 | 2 |
|---|---|---|
| **Dataset** | | |
| **1** | 94 | 29 |

## 1.3 KNN model(K KNwighborsClassifier)

```
from sklearn.neighbors import KNeighborsClassifier
model2=KNeighborsClassifier()
model2.fit(xtrain, ytrain)
y_predict=model2.predict(xtest)
knn1=(accuracy_score(ytest, y_predict))
knn1
pd.crosstab(ytest, y_predict)
```

| col_0 | 1 | 2 |
|---|---|---|
| **Dataset** | | |
| **1** | 105 | 18 |
| **2** | 36 | 15 |

## 1.4 Logistic regression model

```
from sklearn.linear_model import LogisticRegression
model5=LogisticRegression()
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
logi1=accuracy_score(ytest,y_predict)
logi1
pd.crosstab(ytest, y_predict)
```

| col_0 | 1 | 2 |
|---|---|---|
| **Dataset** | | |
| **1** | 112 | 11 |

## 1.5 ANN model

```
import tensorflow
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_crossentropy', metrics=['accuracy'])


model_history=classifier.fit(xtrain, ytrain, batch_size=100, validation_split=0.2, epochs=100)
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
Epoch 98/100
4/4 [==============================] - 0s 21ms/step - loss: -53283.2539 - accuracy: 0.7377 - val_loss: -75568.7500 -
Epoch 99/100
```

## 2.Testing the model

```
model4.predict([[50,1,1.1,0.8,150,70,80,7.2,3.4,0.8]])
```

```
array([1])
```

```
model1.predict([[50,1,1.1,0.8,150,70,80,7.2,3.4,0.8]])
```

```
array([1])
```

```
classifier.save("liver.h5")
```

```
y_pred=classifier.predict(xtest)
```

```
6/6 [==============================] - 0s 3ms/step
```

```
y_pred
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
                [1.],
                [1.],
```

```
y_pred=(y_pred>0.5)
y_pred
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

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

```
def predict_exit(sample_value):


  sample_value=np.array(sample_value)

  sample_value=sample_value.reshape(1,-1)

  sample_value=scale(sample_value)
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
sample_value=[[50,1,1.1,0.8,150,70,80,7.2,3.4,0.8]]
if predict_exit(sample_value)>0.5:
  print('prediction: Liver patient')
else:
    print('prediction: Healthy')
```

```
1/1 [==============================] - 0s 72ms/step
prediction: Liver patient
```

### Milestone 5. performance testing and hyperparameter tuning 1. testing and model with multiple evaluation metrics 1.1 compare the model

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

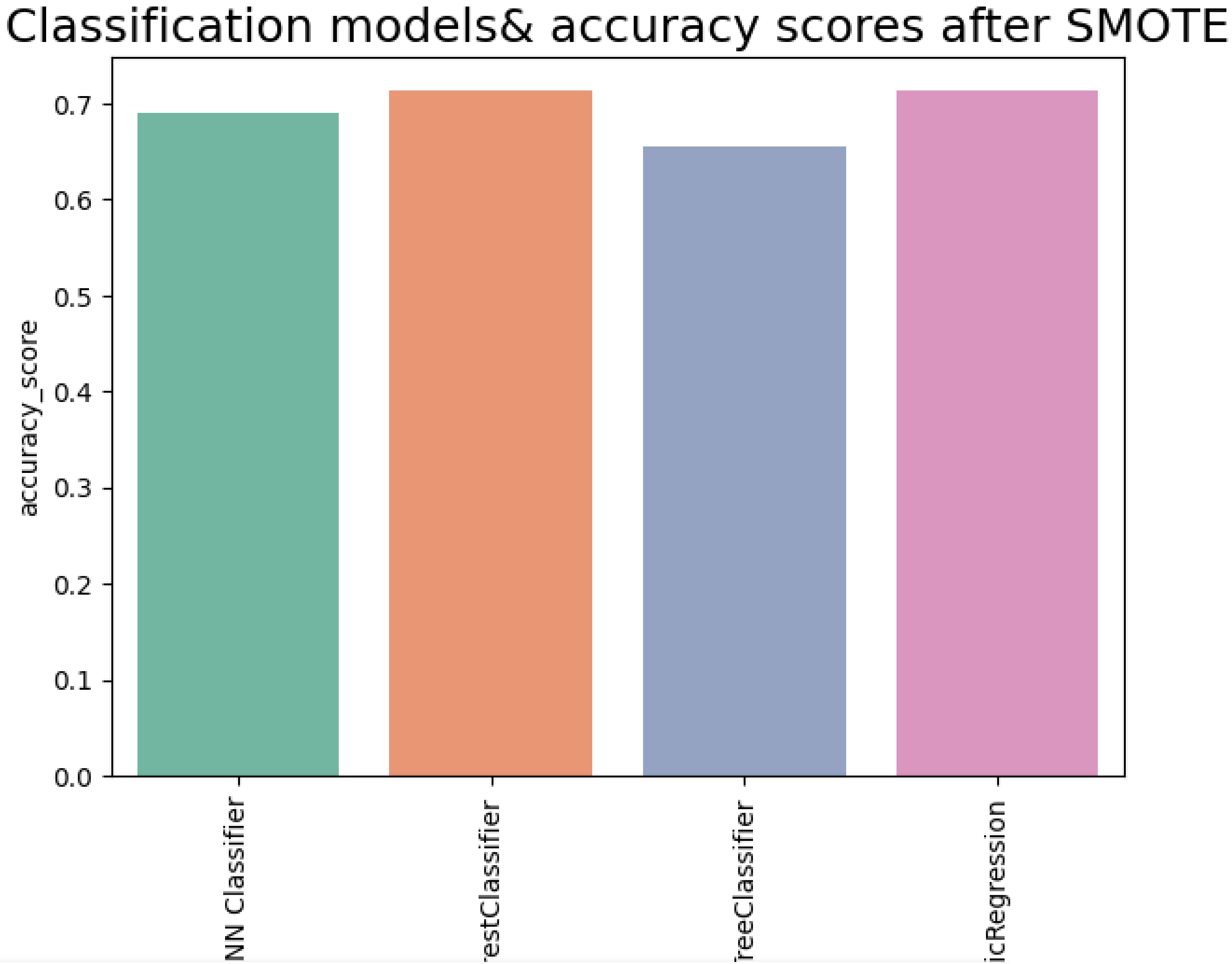
| | classification models | accuracy_score |
|---|---|---|
| **0** | KNN Classifier | 0.689655 |
| **1** | RandomForestClassifier | 0.712644 |
| **2** | DecisionTreeClassifier | 0.655172 |
| **3** | LogisticRegression | 0.712644 |

```
import matplotlib.pyplot as plt
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
plt.title("classification models& accuracy scores after SMOTE", fontsize=18)
sns.barplot(x="classification models", y="accuracy_score", data=Liverpatient_pred,palette="Set2")
```

```
<Axes: title={'center': 'Classification models& accuracy scores after SMOTE'}, xlabel='classification models',
ylabel='accuracy_score'>
```



Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
model=ExtraTreesClassifier()
model.fit(x,y)
```

```
▾ ExtraTreesClassifier
ExtraTreesClassifier()
```

```
model.feature_importances_
```

```
     array([0.11583568, 0.02876283, 0.11361285, 0.104553  , 0.11363613,
            0.11642062, 0.11490209, 0.09259462, 0.10122526, 0.09845691])
```

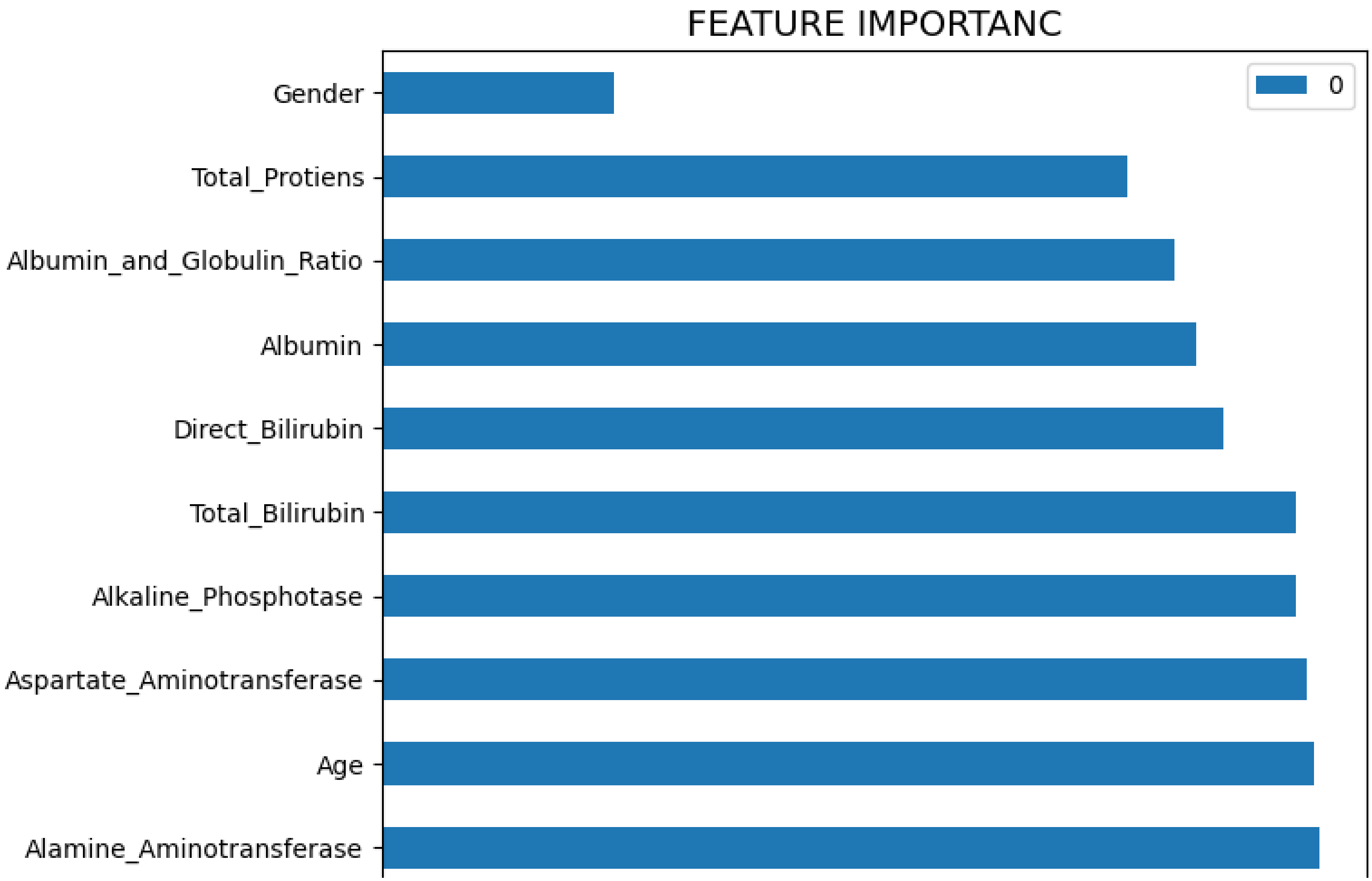
```
import pandas as pd
dd=pd.DataFrame(model.feature_importances_,index=x.columns).sort_values(0,ascending=False)
dd
```

| | 0 |
|---|---|
| **Alamine_Aminotransferase** | 0.116421 |
| **Age** | 0.115836 |
| **Aspartate_Aminotransferase** | 0.114902 |
| **Alkaline_Phosphotase** | 0.113636 |
| **Total_Bilirubin** | 0.113613 |
| **Direct_Bilirubin** | 0.104553 |
| **Albumin** | 0.101225 |
| **Albumin_and_Globulin_Ratio** | 0.098457 |
| **Total_Protiens** | 0.092595 |
| **Gender** | 0.028763 |

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
dd.plot(kind='barh', figsize=(7,6))
plt.title("FEATURE IMPORTANC", fontsize=14)
```

```
Text(0.5, 1.0, 'FEATURE IMPORTANC')
```



### Milestone 6. Model deployment 1. save the best model

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
['ETC.pk1']
```

Double-click (or enter) to edit

## 2. Integrate with web framework 2.2 build python code

```
from flask import Flask, render_template, request
import numpy as np
import pickle


app=Flask(__name__)
@app.route('/')
def home():
  return render_template('home.html')
  @app.route('/predict')
  def index():
    return render_template("index.html")
```

✓ 0s completed at 7:32 PM ● ×

Automatic saving failed. This file was updated remotely or in another tab. Show diff