DATA PRE-PROCESSING

CHECKING FOR NULL VALUES AND HANDLING NULL VALUES

Date	19 November 2022
Team ID	PNT2022TMID04221
Project Name	Project – Statistical Machine Learning Approaches to Liver Disease Prediction

1. After loading it is important to check the complete information of data as it can indication many of the hidden information such as null values in a column or a row

2. Check whether any null values are there or not. if it is present then following can be done,

   a. Imputing data using Imputation method in sklearn

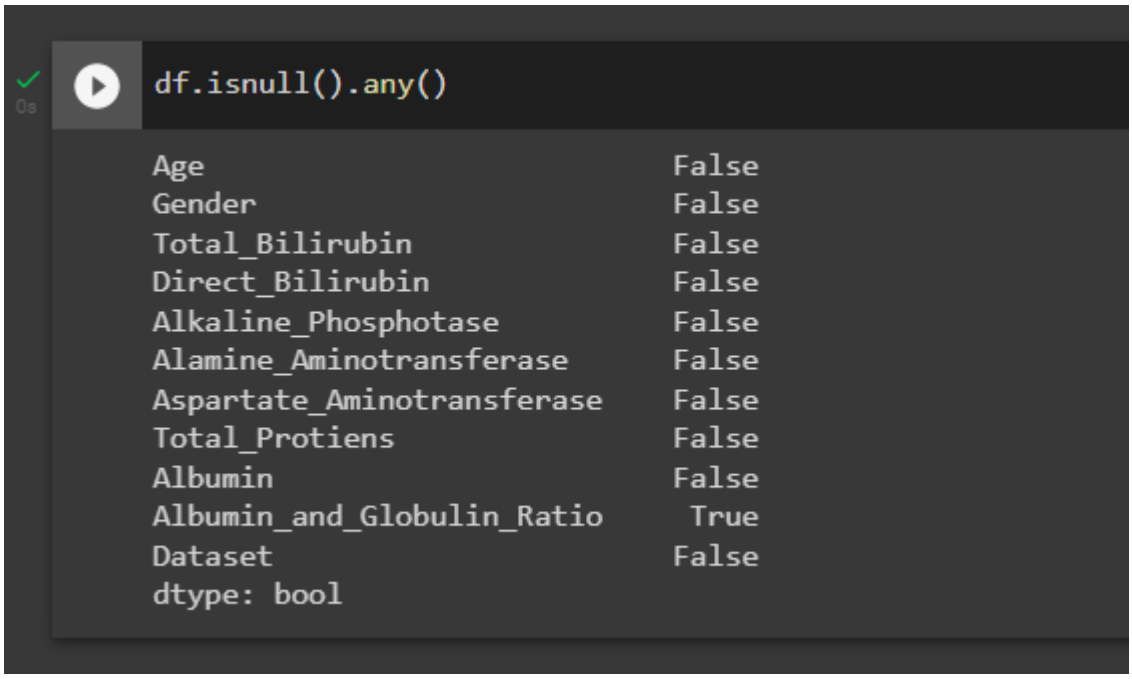
   b. Filling NaN values with mean, median and mode using fillna() method.

We will be using isnull().any() method to see which column has missing values.

This isnull().any() method return two values, False and True.

False return that Column has No Null Values.

True return that Column has Null values.



We can see that there are null values in the Albumin_and_Globulin_Ration Column.

Let us check how many numbers of null records present in the Closing Value column using sum() function.

```
#Check for any null values
df.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
```

We can notice that, there are 4 null values are there in the column Albumin_and_Globulin_Ratio. Now will handle or fill that null values with the help of fillna() method.

```
df["Albumin_and_Globulin_Ratio"] =df.Albumin_and_Globulin_Ratio.fillna(df['Albumin_and_Globulin_Ratio'].mean())
df.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
dtype: int64
```