Statistical Machine Learning Approaches to Liver Disease Prediction

Team ID: PNT2022TMID26382

Team Leader : Madhulika I
Team Member : Rithika AM
Team Member : Kavya A P
Team Member : Kirthiga S

Data Visualization

Data visualization is a field in data analysis that deals with visual representation of data. It graphically plots data and is an effective way to communicate inferences from data.

Using data visualization, we can get a visual summary of our data. With pictures, maps and graphs, the human mind has an easier time processing and understanding any given data. Data visualization plays a significant role in the representation of both small and large data sets, but it is especially useful when we have large data sets, in which it is impossible to see all of our data, let alone process and understand it manually.

Univariate Analysis:

Univariate analysis is the simplest form of data analysis where the data being analysed contains only one variable.

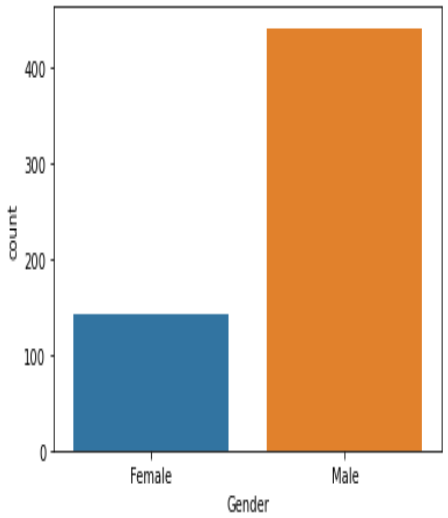
Bivariate Analysis:

It involves the analysis of two variables (often denoted as X, Y), for the purpose of determining the empirical relationship between them.

1. Bar plot between Gender and Count

```
In [13]: sns.countplot(data=data,x='Gender',label='Count')
         m,f=data['Gender'].value_counts()
         print("Number of males:",m)
         print("Number of females:",f)
```

```
Number of males: 441
Number of females: 142
```



We can observe from above bar plot is, count of liver disease is observed in male is higher compare to females.

2. Bar Plot which describes about the total number of disease.

In [11]:
```
sns.countplot(data=data,x='Dataset')
LD,NLD=data['Dataset'].value_counts()
print("Liver disease patients:",LD)
print("Non-Liver disease patients:",NLD)
```

```
Liver disease patients: 416
Non-Liver disease patients: 167
```