Statistical Machine Learning Approaches to Liver Disease Prediction

TeamID:PNT2022TMID49098

TeamLeader: Manivanitha

TeamMember: Logapriya

TeamMember: Nisha

TeamMember: Sakthisubashini

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. Withpictures, maps and graphs, the human mind has an easier time processingand understanding any given data. Data visualization plays a significant role in there presentation of both small and large datasets, 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 isthe simplest form of data analysis where the data being analyzed 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

```
# Counting patients who are Male and who are Female
sns.countplot(data=data, x = 'Gender', label='Count')
m,f=data['Gender'].value_counts()
print("No of Males:",m)
print("No of Females:",f)
```

No of Males: 441
No of Females: 142

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

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

```
# Counting patients who are diagnosed and not diagnoised with liver disease
sns.countplot(data=data, x = 'Dataset')
LD,NLD=data['Dataset'].value_counts()
print("liver disease patinets:",LD)
print("Non-liver disease patinets:",NLD)
```

liver disease patinets: 416
Non-liver disease patinets: 167