```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [2]:
```python
data = pd.read_excel(r"C:\Users\DELL\Desktop\indian_liver_patient.csv")
```

In [3]:
```python
data
```

Out[3]:

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens
0	65	Female	0.7	0.1	187	16	18	6.8
1	62	Male	10.9	5.5	699	64	100	7.5
2	62	Male	7.3	4.1	490	60	68	7.0
3	58	Male	1.0	0.4	182	14	20	6.8
4	72	Male	3.9	2.0	195	27	59	7.3
...	...	...	...	...	...	...	...	...
78	60	Male	0.5	0.1	500	20	34	5.9
79	40	Male	0.6	0.1	98	35	31	6.0
80	52	Male	0.8	0.2	245	48	49	6.4
81	31	Male	1.3	0.5	184	29	32	6.8
82	38	Male	1.0	0.3	216	21	24	7.3

83 rows × 11 columns

In [4]:
```python
data.head()
```

Out[4]:

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens
0	65	Female	0.7	0.1	187	16	18	6.8
1	62	Male	10.9	5.5	699	64	100	7.5
2	62	Male	7.3	4.1	490	60	68	7.0
3	58	Male	1.0	0.4	182	14	20	6.8
4	72	Male	3.9	2.0	195	27	59	7.3

In [5]:
```python
data.shape
```

Out[5]:
```
(583, 11)
```

In [6]:
```python
data.index
```

Out[6]:
```
RangeIndex(start=0, stop=583, step=1)
```

In [7]:
```python
data.columns
```

Out[7]:
```
Index(['Age', 'Gender', 'Total_Bilirubin', 'Direct_Bilirubin',
       'Alkaline_Phosphotase', 'Alamine_Aminotransferase',
       'Aspartate_Aminotransferase', 'Total_Protiens', 'Albumin',
       'Albumin_and_Globulin_Ratio', 'Dataset'],
      dtype='object')
```

In [8]:
```python
data.dtypes
```

Out[8]:
```
Age                             int64
Gender                         object
Total_Bilirubin               float64
Direct_Bilirubin              float64
Alkaline_Phosphotase            int64
Alamine_Aminotransferase        int64
Aspartate_Aminotransferase      int64
Total_Protiens                float64
Albumin                       float64
Albumin_and_Globulin_Ratio    float64
Dataset                         int64
dtype: object
```

In [12]:
```python
data.nunique()
```

Out[12]:
```
Age                            72
Gender                          2
Total_Bilirubin               113
Direct_Bilirubin               80
Alkaline_Phosphotase          263
Alamine_Aminotransferase      152
Aspartate_Aminotransferase    177
Total_Protiens                 58
Albumin                        40
Albumin_and_Globulin_Ratio     69
Dataset                         2
dtype: int64
```

In [13]:
```python
data.count()
```

Out[13]:
```
Age                     583
Gender                  583
Total_Bilirubin         583
Direct_Bilirubin        583
Alkaline_Phosphotase    583
```

```python
In [1]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
        %matplotlib inline
```

```python
In [2]: data = pd.read_excel(r"C:\Users\DELL\Desktop\indian_liver_patient.csv")
```

```python
In [3]: data
```

Out[3]:

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Prot
0	65	Female	0.7	0.1	187	16	18	
1	62	Male	10.9	5.5	699	64	100	
2	62	Male	7.3	4.1	490	60	68	
3	58	Male	1.0	0.4	182	14	20	
4	72	Male	3.9	2.0	195	27	59	
...	...	...	...	...	...	...	...	
578	60	Male	0.5	0.1	500	20	34	
579	40	Male	0.6	0.1	98	35	31	
580	52	Male	0.8	0.2	245	48	49	
581	31	Male	1.3	0.5	184	29	32	
582	38	Male	1.0	0.3	216	21	24	

583 rows × 11 columns

```python
In [4]: data.head()
```

Out[4]:

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protien
0	65	Female	0.7	0.1	187	16	18	6.
1	62	Male	10.9	5.5	699	64	100	7.
2	62	Male	7.3	4.1	490	60	68	7.
3	58	Male	1.0	0.4	182	14	20	6.
4	72	Male	3.9	2.0	195	27	59	7.

```python
In [5]: data.shape
```

Out[5]: (583, 11)

```python
In [6]: data.index
```

Out[6]: RangeIndex(start=0, stop=583, step=1)

```python
In [7]: data.columns
```

Out[7]: Index(['Age', 'Gender', 'Total_Bilirubin', 'Direct_Bilirubin',
               'Alkaline_Phosphotase', 'Alamine_Aminotransferase',
               'Aspartate_Aminotransferase', 'Total_Protiens', 'Albumin',
               'Albumin_and_Globulin_Ratio', 'Dataset'],
              dtype='object')

```python
In [8]: data.dtypes
```

Out[8]:
```
Age                             int64
Gender                         object
Total_Bilirubin               float64
Direct_Bilirubin              float64
Alkaline_Phosphotase            int64
Alamine_Aminotransferase        int64
Aspartate_Aminotransferase      int64
Total_Protiens                float64
Albumin                       float64
Albumin_and_Globulin_Ratio    float64
Dataset                         int64
dtype: object
```

```python
In [12]: data.nunique()
```

Out[12]:
```
Age                            72
Gender                          2
Total_Bilirubin               113
Direct_Bilirubin               80
Alkaline_Phosphotase          263
Alamine_Aminotransferase      152
Aspartate_Aminotransferase    177
Total_Protiens                 58
Albumin                        40
Albumin_and_Globulin_Ratio     69
Dataset                         2
dtype: int64
```

```python
In [13]: data.count()
```

Out[13]:
```
Age               583
Gender            583
Total_Bilirubin   583
```

```
In [4]:    data.head()
```

```
Out[4]:
```

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protien
0	65	Female	0.7	0.1	187	16	18	6.
1	62	Male	10.9	5.5	699	64	100	7.
2	62	Male	7.3	4.1	490	60	68	7.
3	58	Male	1.0	0.4	182	14	20	6.
4	72	Male	3.9	2.0	195	27	59	7.

```
In [5]:    data.shape
```

```
Out[5]:    (583, 11)
```

```
In [6]:    data.index
```

```
Out[6]:    RangeIndex(start=0, stop=583, step=1)
```

```
In [7]:    data.columns
```

```
Out[7]:    Index(['Age', 'Gender', 'Total_Bilirubin', 'Direct_Bilirubin',
                  'Alkaline_Phosphotase', 'Alamine_Aminotransferase',
                  'Aspartate_Aminotransferase', 'Total_Protiens', 'Albumin',
                  'Albumin_and_Globulin_Ratio', 'Dataset'],
                  dtype='object')
```

```
In [8]:    data.dtypes
```

```
Out[8]:    Age                             int64
           Gender                         object
           Total_Bilirubin               float64
           Direct_Bilirubin              float64
           Alkaline_Phosphotase            int64
           Alamine_Aminotransferase        int64
           Aspartate_Aminotransferase      int64
           Total_Protiens                float64
           Albumin                       float64
           Albumin_and_Globulin_Ratio    float64
           Dataset                         int64
           dtype: object
```

```
In [12]:   data.nunique()
```

```
Out[12]:   Age                            72
           Gender                          2
           Total_Bilirubin               113
           Direct_Bilirubin               80
           Alkaline_Phosphotase          263
           Alamine_Aminotransferase      152
           Aspartate_Aminotransferase    177
           Total_Protiens                 58
           Albumin                        40
           Albumin_and_Globulin_Ratio     69
           Dataset                         2
           dtype: int64
```

```
In [13]:   data.count()
```

```
Out[13]:   Age                            583
           Gender                         583
           Total_Bilirubin               583
           Direct_Bilirubin              583
           Alkaline_Phosphotase          583
           Alamine_Aminotransferase      583
           Aspartate_Aminotransferase    583
           Total_Protiens                583
           Albumin                        583
           Albumin_and_Globulin_Ratio    579
           Dataset                        583
           dtype: int64
```

```
In [16]:   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)
```

```
dtype: object
```

In [12]: `data.nunique()`

Out[12]:
```
Age                          72
Gender                        2
Total_Bilirubin             113
Direct_Bilirubin             80
Alkaline_Phosphotase        263
Alamine_Aminotransferase    152
Aspartate_Aminotransferase  177
Total_Protiens               58
Albumin                      40
Albumin_and_Globulin_Ratio   69
Dataset                       2
dtype: int64
```

In [13]: `data.count()`

Out[13]:
```
Age                          583
Gender                       583
Total_Bilirubin              583
Direct_Bilirubin             583
Alkaline_Phosphotase         583
Alamine_Aminotransferase     583
Aspartate_Aminotransferase   583
Total_Protiens               583
Albumin                      583
Albumin_and_Globulin_Ratio   579
Dataset                      583
dtype: int64
```

In [16]: `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
```

In [18]: `data.isnull().sum()`

Out[18]:
```
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
```

In [19]: `data.notnull().sum()`

Out[19]:
```
Age                          583
Gender                       583
Total_Bilirubin              583
Direct_Bilirubin             583
Alkaline_Phosphotase         583
Alamine_Aminotransferase     583
Aspartate_Aminotransferase   583
Total_Protiens               583
Albumin                      583
Albumin_and_Globulin_Ratio   579
Dataset                      583
dtype: int64
```

In [ ]:
```python
data.dropna(inplace=true)
data=data['-data.duplicated(subset=none,keep='first')]
```

In [20]: `data.head(2)`

Out[20]:

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens
0	65	Female	0.7	0.1	187	16	18	6.8
1	62	Male	10.9	5.5	699	64	100	7.5

In [ ]:

```
dtype: int64
```

In [19]:
```python
data.notnull().sum()
```

Out[19]:
```
Age                           583
Gender                        583
Total_Bilirubin               583
Direct_Bilirubin              583
Alkaline_Phosphotase          583
Alamine_Aminotransferase      583
Aspartate_Aminotransferase    583
Total_Protiens                583
Albumin                       583
Albumin_and_Globulin_Ratio    579
Dataset                       583
dtype: int64
```

In [ ]:
```python
data.dropna(inplace=true)
data=data['-data.duplicated(subset=none,keep='first')]
```

In [20]:
```python
data.head(2)
```

Out[20]:

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens
0	65	Female	0.7	0.1	187	16	18	6.8
1	62	Male	10.9	5.5	699	64	100	7.5

In [ ]:

In [21]:
```python
data.rename(columns ={'indian_liver_patient':'liver'})
```

Out[21]:

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protier
0	65	Female	0.7	0.1	187	16	18	6
1	62	Male	10.9	5.5	699	64	100	7
2	62	Male	7.3	4.1	490	60	68	7
3	58	Male	1.0	0.4	182	14	20	6
4	72	Male	3.9	2.0	195	27	59	7
...	...	...	...	...	...	...	...	
578	60	Male	0.5	0.1	500	20	34	5
579	40	Male	0.6	0.1	98	35	31	6
580	52	Male	0.8	0.2	245	48	49	6
581	31	Male	1.3	0.5	184	29	32	6
582	38	Male	1.0	0.3	216	21	24	7

583 rows × 11 columns

In [22]:
```python
data.head()
```

Out[22]:

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens
0	65	Female	0.7	0.1	187	16	18	6.8
1	62	Male	10.9	5.5	699	64	100	7.5
2	62	Male	7.3	4.1	490	60	68	7.0
3	58	Male	1.0	0.4	182	14	20	6.8
4	72	Male	3.9	2.0	195	27	59	7.3

In [29]:
```python
data_new = data.dropna()
```

In [34]:
```python
data_new.head(2)
```

Out[34]:

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens
0	65	Female	0.7	0.1	187	16	18	6.8
1	62	Male	10.9	5.5	699	64	100	7.5

In [ ]:
```python
data['class'].value_counts()
```

In [35]:
```python
data =data.dropna()
```

In [36]:
```python
data.isnull().any()
```

Out[36]:
```
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    False
Dataset                       False
dtype: bool
```

In [37]:
```python
x=data.iloc[:,0:10]
x.head()
```

Out[37]:

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens

| 4 | 72 | Male | 3.9 | 2.0 | 195 | 27 | 59 | 7.3 |

```
In [29]:   data_new = data.dropna()
```

```
In [34]:   data_new.head(2)
```

Out[34]:

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens
0	65	Female	0.7	0.1	187	16	18	6.8
1	62	Male	10.9	5.5	699	64	100	7.5

```
In [ ]:   data['class'].value_counts()
```

```
In [35]:   data =data.dropna()
```

```
In [36]:   data.isnull().any()
```

Out[36]:
```
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  False
Dataset                     False
dtype: bool
```

```
In [37]:   x=data.iloc[:,0:10]
           x.head()
```

Out[37]:

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens
0	65	Female	0.7	0.1	187	16	18	6.8
1	62	Male	10.9	5.5	699	64	100	7.5
2	62	Male	7.3	4.1	490	60	68	7.0
3	58	Male	1.0	0.4	182	14	20	6.8
4	72	Male	3.9	2.0	195	27	59	7.3

```
In [38]:   y=data.iloc[:,-1]
           y.head()
```

Out[38]:
```
0    1
1    1
2    1
3    1
4    1
Name: Dataset, dtype: int64
```

```
In [39]:   x.Gender = x.Gender.map( {'Male':0 ,'Female':1})
```

```
In [40]:   x.head()
```

Out[40]:

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens
0	65	1	0.7	0.1	187	16	18	6.8
1	62	0	10.9	5.5	699	64	100	7.5
2	62	0	7.3	4.1	490	60	68	7.0
3	58	0	1.0	0.4	182	14	20	6.8
4	72	0	3.9	2.0	195	27	59	7.3

```
In [ ]:   from sklearn.modal_selection import train_test_split
          x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.25, random_state=0)f
```

```
In [ ]:   from sklearn.linear_modal import logisticRegression
          classifier =logisticRegression()
          classifier.fit(x_train,y_train)
```

```
In [ ]:   data.Gender =data.Gender.map( {'Male':0, 'Female':1} )
          df_majority =data[data.Class==1]
          df_majority =data[data.Class==2]
```

```
In [ ]:   rom sklearn.utils import resample
          df_minority_upsampled = resample(df_minority,
                                           replace=True,
                                           n_sample=414,
                                           random_state=123)
```

```
In [ ]:   df_upsampled = pd.concat([df_majority,df_majority_upsampled])
          df_upsampled.Class.value_counts()
```

```
In [ ]:   print(no_disease.shape,disease.shape)
```

```
In [ ]:   from imblearn.over_sampling import RandomOverSampler
```

```
In [57]:   from sklearn .model_selection import train_test_split
```