2203a51534-1

April 3, 2024

```python
[1]: import pandas as pd
     from sklearn.model_selection import train_test_split
     from sklearn.tree import DecisionTreeClassifier
     from sklearn.metrics import accuracy_score, precision_score, recall_score,
      ↪f1_score


     dataset_path = 'path/to/dataset.csv'
     df = pd.read_csv('/content/archive (3).zip')


     df['Gender'] = df['Gender'].map({'MALE': 0, 'FEMALE':1})
     df['Tumour_Stage']=df['Tumour_Stage'].map({'I':1,'II':2,'III':3})
     df['Patient_Status']=df['Patient_Status'].map({'Alive':1,'Dead':0})

     print(df)
     df=df.drop('Histology',axis=1)
     df=df.drop('ER status',axis=1)
     df=df.drop('PR status',axis=1)
     df=df.drop('HER2 status',axis=1)
     df=df.drop('Surgery_type',axis=1)
     df=df.drop('Date_of_Surgery',axis=1)
     df=df.drop('Date_of_Last_Visit',axis=1)
     df=df.fillna(0)
     y = df['Patient_Status']
     X = df.drop(columns=['Patient_Status'])
     print(X)
     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
      ↪random_state=42)
     tree_classifier = DecisionTreeClassifier(random_state=42)
     tree_classifier.fit(X_train, y_train)

     y_pred = tree_classifier.predict(X_test)
```

```
     Age  Gender  Protein1  Protein2  Protein3  Protein4  Tumour_Stage  \
0    42       1  0.952560   2.15000  0.007972 -0.048340             2
1    54       1  0.000000   1.38020 -0.498030 -0.507320             2
2    63       1 -0.523030   1.76400 -0.370190  0.010815             2
```

```
3      78      1 -0.876180   0.12943 -0.370380  0.132190                1
4      42      1  0.226110   1.74910 -0.543970 -0.390210                2
..     …       …     …          …        …         …                    …
329    59      1  0.024598   1.40050  0.024751  0.280320                2
330    41      1  0.100120  -0.46547  0.472370 -0.523870                1
331    54      1  0.753820   1.64250 -0.332850  0.857860                2
332    74      1  0.972510   1.42680 -0.366570 -0.107820                2
333    66      1  0.286380   1.39980  0.318830  0.836050                2
```

```
                               Histology ER status PR status HER2 status  \
0      Infiltrating Ductal Carcinoma  Positive  Positive     Negative
1      Infiltrating Ductal Carcinoma  Positive  Positive     Negative
2      Infiltrating Ductal Carcinoma  Positive  Positive     Negative
3      Infiltrating Ductal Carcinoma  Positive  Positive     Negative
4      Infiltrating Ductal Carcinoma  Positive  Positive     Positive
..                                …       …         …           …
329    Infiltrating Ductal Carcinoma  Positive  Positive     Positive
330    Infiltrating Ductal Carcinoma  Positive  Positive     Positive
331    Infiltrating Ductal Carcinoma  Positive  Positive     Negative
332   Infiltrating Lobular Carcinoma  Positive  Positive     Negative
333    Infiltrating Ductal Carcinoma  Positive  Positive     Negative
```

```
                     Surgery_type Date_of_Surgery Date_of_Last_Visit  \
0                           Other       20-May-18          26-Aug-18
1                           Other       26-Apr-18          25-Jan-19
2                      Lumpectomy       24-Aug-18          08-Apr-20
3                           Other       16-Nov-18          28-Jul-20
4                      Lumpectomy       12-Dec-18          05-Jan-19
..                              …             …                  …
329                    Lumpectomy       15-Jan-19          27-Mar-20
330   Modified Radical Mastectomy       25-Jul-18          23-Apr-19
331              Simple Mastectomy       26-Mar-19          11-Oct-19
332                    Lumpectomy       26-Nov-18          05-Dec-18
333   Modified Radical Mastectomy       04-Feb-19          10-Aug-19
```

```
      Patient_Status
0                1.0
1                0.0
2                1.0
3                1.0
4                1.0
..                …
329              1.0
330              1.0
331              0.0
332              1.0
333              0.0
```

```
[334 rows x 15 columns]
     Age  Gender  Protein1  Protein2  Protein3   Protein4  Tumour_Stage
0     42       1  0.952560   2.15000  0.007972  -0.048340             2
1     54       1  0.000000   1.38020 -0.498030  -0.507320             2
2     63       1 -0.523030   1.76400 -0.370190   0.010815             2
3     78       1 -0.876180   0.12943 -0.370380   0.132190             1
4     42       1  0.226110   1.74910 -0.543970  -0.390210             2
..   ...     ...       ...       ...       ...        ...           ...
329   59       1  0.024598   1.40050  0.024751   0.280320             2
330   41       1  0.100120  -0.46547  0.472370  -0.523870             1
331   54       1  0.753820   1.64250 -0.332850   0.857860             2
332   74       1  0.972510   1.42680 -0.366570  -0.107820             2
333   66       1  0.286380   1.39980  0.318830   0.836050             2

[334 rows x 7 columns]
```