Thyroid Disease Classification Using ML
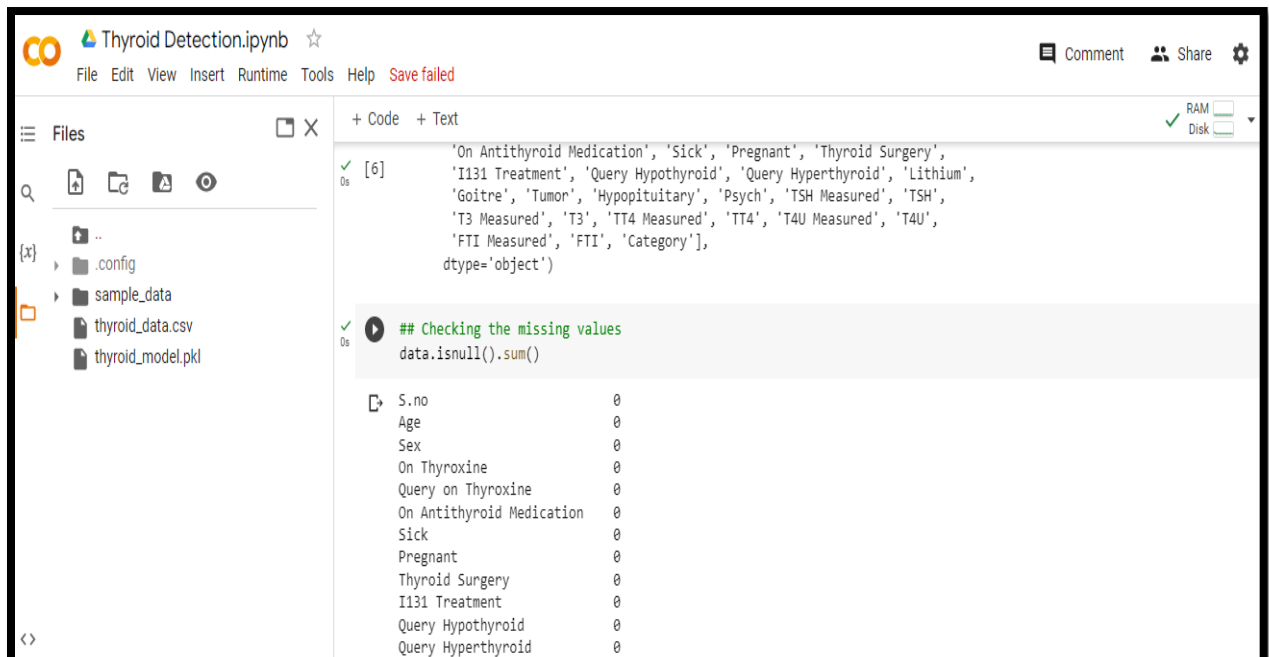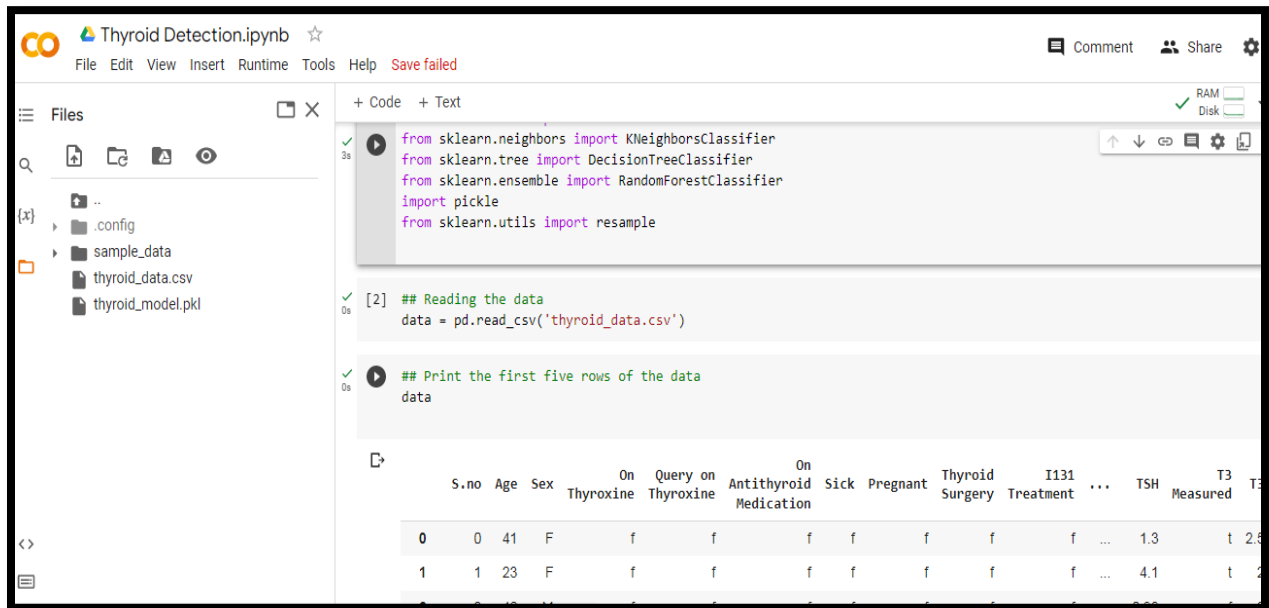
Sample Output

So, we have replaced all such values with 'nan' values.

```python
##  Unique Values
for column in data.columns:
    print(column,(data[column].unique()))
```

```
0.015   10   168   0.85   9.2   80   0.02   0.39   9.1   5.9   52
'0.33' '31' '5.8' '0.28' '51' '6.3' '4.4' '9.6' '3.4' '0.09' '24' '0.76'
'42' '25' '10' '4.6' '8.6' '0.66' '6.2' '0.79' '28' '0.86' '9.7' '0.84'
'17' '18' '55' '14' '3.7' '0.87' '6.7' '0.74' '7.6' '0.065' '0.29' '0.37'
'8' '11' '0.48' '44' '7.9' '5' '7.2' '0.89' '0.93' '0.97' '0.12' '6.4'
'33' '0.85' '7.1' '0.73' '199' '8.2' '188' '0.22' '98' '22' '6.6' '5.1'
'0.06' '0.42' '3.8' '35' '4' '0.78' '0.63' '0.52' '60' '0.43' '5.6' '6.9'
'3.6' '29' '0.38' '4.9' '0.41' '9.9' '7.5' '34' '6.5' '4.7' '103' '0.95'
'0.14' '0.35' '4.2' '0.81' '0.54' '0.58' '8.9' '5.5' '0.34' '9.3' '0.13'
'54' '0.39' '8.3' '478' '21' '6.8' '0.32' '0.23' '0.24' '8.1' '0.91'
'5.3' '100' '0.27' '1.01' '58' '41' '183' '18.4' '0.47' '0.17' '12.1'
'0.19' '0.82' '43' '0.44' '70' '7.7' '8.4' '0.69' '8.5' '0.21' '82'
'0.055' '0.96' '0.71' '38' '0.36' '9.8' '7' '0.46' '11.1' '39' '76' '5.7'
'32' '126' '26.4' '0.53' '0.49' '36' '178' '145' '47' '4.8' '10.3' '89'
'7.4' '472' '0.51' '116' '61' '99' '46' '78' '468']
```

```python
for column in columns:
    ax = plt.subplot(3,2,plotnumber)
    sns.distplot(final_df[column])
    plt.xlabel(column,fontsize=10)
    plotnumber+=1
plt.show()
```

Screenshot 1

SVM:

Train Score:0.6165285503462368
Test Score:0.6391284611892873

KNN:

Train Score:0.8651379271199909
Test Score:0.8483885610531094

Decision Tree:

Train Score:0.9162220456351459
Test Score:0.8955969133000454

Random Forest:

Train Score:0.9162220456351459
Test Score:0.9024058102587381

[39] ## Performance Metrics

```
ForestClassifier(criterion = 'entropy')
ain,y_train)
```

Automatic saving failed. This file was updated remotely or in another tab.   Show diff

Screenshot 2

```
[42] print(accuracies.mean())

     0.8971514807553399
```

```
[43] import pickle
     filename = 'thyroid_model.pkl'
     pickle.dump(classifier_forest,open(filename,'wb'))
```

```
[44] model = open('thyroid_model.pkl','rb')
     forest = pickle.load(model)
```

```
[45] y_pred = forest.predict(X_test)
```

```
confusion_matrix(y_test, y_pred)
```

```
array([[563,   0,   0,   0],
       [  0, 526,   0,  27],
       [ 21,  48, 423,  40],
       [ 52,  29,   0, 474]])
```

[47] X columns

'T4U', 'FTI', 'Sex_M', 'Sick_t', 'Pregnant_t',

Automatic saving failed. This file was updated remotely or in another tab.   Show diff