2) Model Measurement Analysis: Using *any dataset* and *any classifier* do the following:
   a) Calculate TP, TN, FP and FN from sklearn library functions
   b) Calculate different metrics (Accuracy, Precision, Recall(Sensitivity), F1-Score, MCC, Specificity, Negative Predictive Value) by defining our own functions
   c) Get the sklearn metrics of these values
   d) Verify them by comparing with scikit-learn's library functions.
   e) Get the result of Confusion Matrix using sklearn
   f) Using sklearn, plot the ROC Curve of the probability values in our test data
   g) Using sklearn, plot the ROC Curve of random probabilities
   h) Calculate the AUC of our test data using sklearn
   i) Calculate the AUC of random probabilities using sklearn
   j) Interpret the results. Write the inference/analysis of each output.

   Dataset Used: In-built sklearn's Breast Cancer

OUTPUT ANALYSIS

```
True Negatives 68
True Positives 123
False Positives 2
False Negatives 2
```
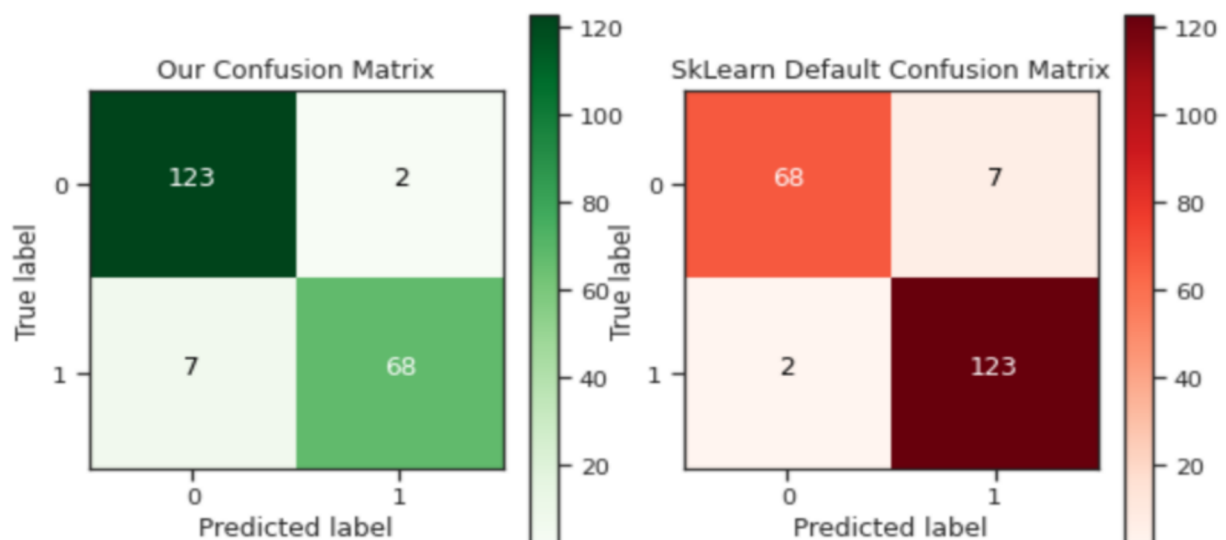

Refer Code: Run this *Handcoded_ConfusionMatrix_ROC_AUC.ipynb* to get the above output. This depicts the difference in Confusion Matrix between sklearn and what we want.

- In this program *(Prg#2_Model Measurement Analysis.ipynb)* the following code is used to get the Confusion Matrix we want:
- `conf=metrics.confusion_matrix(y_test, y_preds,labels=[1,0])`
- `#Note to change the labels from the default 0,1 to 1,0`


```
Calculated and scikit-learn Accuracy:                        0.979,  0.955
Calculated and scikit-learn Precision score:                 0.984,  0.946
Calculated and scikit-learn Recall score:                    0.984,  0.984
Calculated and scikit-learn F1 score:                        0.984,  0.965
Calculated and scikit-learn Matthew's correlation coefficient: 0.955,  0.904
```

Sample test data probabilities

```
2D to 1D reshaped Probability of benign.  [0.99697262 0.08136644 0.99999168 0.96809392 0.99999907]
```

Random probabilities

```
Randon Probablities of Being Benign are:  [0.37772889 0.53432747 0.49656119 0.38961809 0.29763517]
```
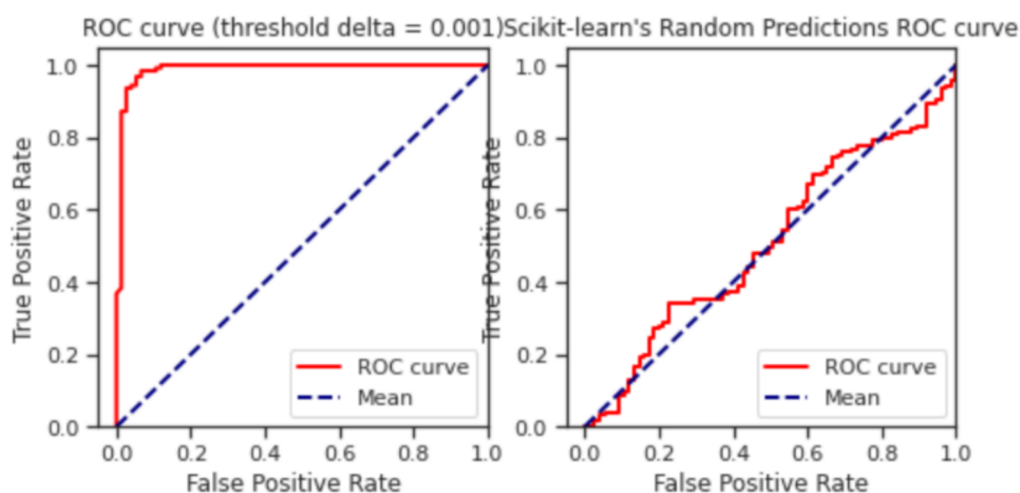

ROC curve (Receiver Operating Characteristic curve)

A receiver operating characteristic curve, i.e. ROC curve, is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied.

The ROC curve is created by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) *at various threshold settings. *

Analysis: The above is the ROC curve for the Breast Cancer Dataset.

1. The peak towards left-most corner means near perfect classifier

2. Random prediction will have the curve as the blue dotted straight line.

3. This ROC curve tells us that our model is nearly perfect classifier, with high accuracy!

ROC for Random Predictions

Analysis It is very close to the "Guess" line

Indicator of a Bad Classifier

Will give a low value of AUC

AUC

```
Scikit's ROC-AUC score of SVC model is
0.9872
Scikit's ROC-AUC score of random
predictions is:  0.5111
```