

Random Forest Classification

Confusion Matrix:

```
In [37]: from sklearn.metrics import confusion_matrix  
cm = confusion_matrix(y_test, y_pred)
```

```
In [51]: print(cm)
```

```
[[62 10]  
 [ 5 43]]
```

Purchased

Confusion Matrix for Purchased **True** is 62

Confusion Matrix for Purchased **False** is 10

Not Purchased:

Confusion Matrix for Not Purchased **True** is 43

Confusion Matrix for Not Purchased **False** is 5

Classification Report:

```
In [53]: from sklearn.metrics import classification_report  
clf_report = classification_report(y_test, y_pred)
```

```
In [55]: print(clf_report)
```

	precision	recall	f1-score	support
0	0.93	0.86	0.89	72
1	0.81	0.90	0.85	48
accuracy			0.88	120
macro avg	0.87	0.88	0.87	120
weighted avg	0.88	0.88	0.88	120

Accuracy:

What is the Accuracy of the Model?

0.88%

Explanation:

	Green color denotes True	Lightblue color denotes False
Purchased	62	10
Not Purchased	5	43

Formula:

T(Purchased) + T(Not-Purchased)

T(Purchased) + T(Not-Purchased) + F(Purchased) + F(Not-Purchased)

62 + 43

105

----- = ----- = **0.88%**

62 + 43 + 10 + 5

120

Precision:

	Green color denotes True	Lightblue color denotes False
Purchased	62	10
Not Purchased	5	43

```
In [53]: from sklearn.metrics import classification_report
         clf_report = classification_report(y_test, y_pred)
```

```
In [55]: print(clf_report)
```

```

              precision    recall  f1-score   support

     0       0.93      0.86      0.89         72
     1       0.81      0.90      0.85         48

 accuracy          0.88
 macro avg          0.87
 weighted avg       0.88
```

Total Count in the test set = 120

Total Count of Purchased in the test set = 72

Total Count of Not Purchased in the test set = 48

Questions:

What is the percentage of correct classification of Purchased to sum of correctly classified as Purchased and wrongly classified as Not-purchased in the test set.

Formula:

T(Purchased)

T(Purchased) + F (Not Purchased)

$$\frac{62}{62 + 5} = \frac{62}{67} = 0.93\%$$

Questions:

What is the percentage of correct classification of Not-Purchased to sum of correctly classified as Purchased and wrongly classified as Purchased in the test set.

Formula:

T(Not-Purchased)

T(Not-Purchased) + F (Purchased)

$$\frac{43}{43 + 10} = \frac{43}{53} = 0.81\%$$

Recall:

	Green color denotes True	Lightblue color denotes False
Purchased	62	10
Not Purchased	5	43

```
In [53]: from sklearn.metrics import classification_report
clf_report = classification_report(y_test, y_pred)
```

```
In [55]: print(clf_report)
```

	precision	recall	f1-score	support
0	0.93	0.86	0.89	72
1	0.81	0.90	0.85	48
accuracy			0.88	120
macro avg	0.87	0.88	0.87	120
weighted avg	0.88	0.88	0.88	120

Total Count in the test set = 120

Total Count of Purchased in the test set = 72

Total Count of Not Purchased in the test set = 48

Questions:

What is the percentage of correct classification of Purchased to the total input of Purchased in the test set

Formula:

T(Purchased)

Total count of (Purchased) in the test set

62

--- = 0.86%

72

Questions:

What is the percentage of correct classification of Not-Purchased to the total input of Not-Purchased in the test set

Formula:

T(Not-Purchased)

Total count of (Not-Purchased) in the test set

43

--- = 0.90%

48

F1-score:

	Green color denotes True	Lightblue color denotes False
Purchased	62	10
Not Purchased	5	43

```
In [53]: from sklearn.metrics import classification_report
         clf_report = classification_report(y_test, y_pred)
```

```
In [55]: print(clf_report)
```

	precision	recall	f1-score	support
0	0.93	0.86	0.89	72
1	0.81	0.90	0.85	48
accuracy			0.88	120
macro avg	0.87	0.88	0.87	120
weighted avg	0.88	0.88	0.88	120

Formula:

$$2 * \frac{\text{Recall (Purchased)} * \text{precision (Purchased)}}{\text{Recall (Purchased)} + \text{precision (Purchased)}}$$

$$2 * \frac{0.86 * 0.93}{0.86 + 0.93} = 2 * \frac{0.80}{1.79} = 0.89\%$$

Formula:

$$\text{Recall (Not-Purchased)} * \text{precision (Not-Purchased)}$$

2 * -----

Recall (Not-Purchased) + precision (Not-Purchased)

$$2 * \frac{0.90 * 0.81}{0.90 + 0.81} = 2 * \frac{0.80}{1.85} = 0.85\%$$

Questions:

What is the overall performance for Purchase: **0.89%**

What is the overall Not-performance for Purchase: **0.85%**

Macro avg:

	Green color denotes True	Lightblue color denotes False
Purchased	62	10
Not Purchased	5	43

```
In [53]: from sklearn.metrics import classification_report
         clf_report = classification_report(y_test, y_pred)
```

```
In [55]: print(clf_report)
```

	precision	recall	f1-score	support
0	0.93	0.86	0.89	72
1	0.81	0.90	0.85	48
accuracy			0.88	120
macro avg	0.87	0.88	0.87	120
weighted avg	0.88	0.88	0.88	120

Formula:

Precision (Purchased) + Precision (Not-Purchased)

2

$$\frac{0.93 + 0.81}{2} = \frac{1.74}{2} = 0.87\%$$

Recall (Purchased) + Recall (Not-Purchased)

2

$$\frac{0.86 + 0.90}{2} = \frac{1.76}{2} = 0.88\%$$

F1-score (Purchased) + F1-score (Not-Purchased)

2

$$\frac{0.89 + 0.85}{2} = \frac{1.74}{2} = 0.87\%$$

Questions:

What is the average performance (correctly & wrongly classified) precision = 0.87%

What is the average performance (correctly & wrongly classified) recall = 0.88%

What is the average performance (correctly & wrongly classified) f1-score = 0.87%

Weighted avg:

```
In [53]: from sklearn.metrics import classification_report  
clf_report = classification_report(y_test, y_pred)
```

```
In [55]: print(clf_report)
```

	precision	recall	f1-score	support
0	0.93	0.86	0.89	72
1	0.81	0.90	0.85	48
accuracy			0.88	120
macro avg	0.87	0.88	0.87	120
weighted avg	0.88	0.88	0.88	120

Formula:

Total Count in the test set = 120

Total Count of Purchased in the test set = 72

Total Count of Not Purchased in the test set = 48

Precision (Purchased) * (Total count of purchased/ total count) + Precision (Not-Purchased) * (Total count of not-purchased/ total count) +

$0.93 * (72/120) + 0.81 * (48/120)$

$0.93 * 0.6 + 0.81 * 0.4$

0.88%

Formula-Recall:

Total Count in the test set = 120

Total Count of Purchased in the test set = 72

Total Count of Not Purchased in the test set = 48

recall (Purchased) * (Total count of purchased/ total count) + recall (Not-Purchased) * (Total count of not-purchased/ total count) +

0.86 * (72/120) + 0.90 * (48/120)

0.93 * 0.6 + 0.81 * 0.4

0.88%

Formula-F1-Score:

Total Count in the test set = 120

Total Count of Purchased in the test set = 72

Total Count of Not Purchased in the test set = 48

F1Score (Purchased) * (Total count of purchased/ total count) + F1Score (Not-Purchased) * (Total count of not-purchased/ total count) +

0.89 * (72/120) + 0.85 * (48/120)

0.93 * 0.6 + 0.81 * 0.4

0.88%

Questions

What is the sum of precision rate (weight) of each class= 0.88%

What is the sum of recall rate (weight) of each class= 0.88%

What is the sum of f1-score rate (weight) of each class= 0.88%