MachineLearning – Classification - Documentation

Model Taken in the problem: SVM – Support vector Machine – Model performance percentage is calculated using Confusion matrix below:

```
In [23]: conf_mx
Out[23]: array([[77,  3],
                [32, 20]])
```

Confusion matrix metrics to evaluate best model and calculation result is represented in table called classification_report below:

Class	Precision	Re-call	F1 score	Support	Accuracy
0 – Purchased	0.71	0.96	0.81	80	.73
1 – Not Purchased	0.87	0.38	0.53	52	
Macro Avg	0.79	0.67	0.67	132	
Weighted Avg	0.77	0.73	0.70	132	

Question & Answeres:

1. Types of Class invovled??

A. Two classes such as Purchased – 0, NotPurchased – 1

		Actual Class	
	Class	**Purchased**	**Notpurchased**
Predicted class	**Purchased**	**TP** (Correctly classified as Purchased)	**FNP** (Wrongly classified as Not purchased)
	Not purchased	**FP** (Wrongly classified as purchased).	**TNP** (Correctly classified as Not purchased)

Note:
TP – True Purchased, TNP – True Not-purchased, FP – False Purchased, FNP – False not purchased

2. What is the status of the problem Balanced or Imbalanced??

A. Imbalanced

```
dataset["Purchased"].value_counts()

0    257
1    143
Name: Purchased, dtype: int64
```

3**. Confusion matrix calculation consist of 6 parameter's as follows to find good model performance score.**

1. Accuracy
2. Re-call
3.Precision
4.F1 Measure
5.Macro-average
6.Weighted Average

```
In [23]: conf_mx

Out[23]: array([[77,  3],
                [32, 20]])
```

Class	Precision	Re-call	F1 score	Support	Accuracy
0 – Purchased	0.71	0.96	0.81	80	.73
1 – Not Purchased	0.87	0.38	0.53	52	
Macro Avg	0.79	0.67	0.67	132	
Weighted Avg	0.77	0.73	0.70	132	

Total number of purchased samples in dataset : 132
Total number of Purchase : 77
Total number of Wrong not purchase prediction : 3
Total number of Not Purchase : 20
Total number of Wrong purchase prediction : 32

Note:
TP – True Purchased, TNP – True Not-purchased, FP – False Purchased, FNP – False not purchased

3.a. What is Accuracy?? - percentage of overall performance of the total input of the test set.

TP+TNP/TP+FNP+TNP+FP = 77+20/77+3+32+20
= 97/132
=.73

3.b. What is Precision?? -

percentage of correctly classified purchase to the sum of correctly classified purchase and wrongly classified purchased

Purchased - TP/TP+FP = 77/77+32 = .71

percentage of correctly classified not purchase to sum of correctly classified not purchase and wrongly classified not purchased

Not-Purchased - TNP/TNP+FNP = 20/20+3 = .87

3.c. What is Recall??

percentage of correctly classified purchase to sum of correctly classified purchase and wrongly classified not purchased

Purchased - TP/TP+FNP = 77/77+3 = .96

percentage of correctly classified not purchase to sum of correctly not purchase and wrongly classified purchase

Not Purchased - TNP/TNP+FP = 20/20+32 = .38

3.d. What is F1 Measure?? - Recall value is high, Precision value is low, So calculating F1Measure for both Precision and Recall

overall performance of Purchased

2 * precison(.71) * recall(.96)/precision(.71) + recall(.96)

2 * .6816/1.67 = .82

overall performance of Not-purchased

2 * precsion(.87) * recall(.38)/precision(.87) + recall(.38)

2 * .3306/1.25 = .6612/1.25 = .53

3.c. What is Macro Average ?? - Calculating the Average of Precsion, Recall, F1-Measure

Precision:

purchased + notpurchased / 2 = .71 + .87 /2 = .79

Recall:

purchased + notpurchased / 2 = .96 + .38 / 2 = .67

F1-Measure:

purchased + notpurchased / 2 = .82 + .53 / 2 = .67

3.d. What is weighted Avg?? - Calculating the Proportion of Precision, Re-call, F1-Measure

Precsion:

precison(purchased) * 80/132 + precsion(notpurchase) * 52/132

.71 * 80/132 + .87 * 52/132

.43 + .34 = .77

Recall:

Recall(Purchased) * 80/132 + Recall(notpurchased) * 52/132

.96 * 80/132 + .38 * 52/132

.58 + .15 = .73

F1-Measure:

F1-measure(purchased) * 80/132 + F1-measure(notpurchased) * 52/132

.82 * 80/132 + .53 * 52/132

.49 + 21 = .70