

Zain Al Abidin

21L-6260

DM Lab 4

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      4074           88.546 %
Incorrectly Classified Instances    527           11.454 %
Kappa statistic                     0.7568
Mean absolute error                  0.1183
Root mean squared error              0.3147
Relative absolute error              24.7678 %
Root relative squared error          64.4068 %
Total Number of Instances           4601

=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
	0.931	0.185	0.886	0.931	0.908	0.758	0.953	0.967
	0.815	0.069	0.885	0.815	0.849	0.758	0.953	0.936
Weighted Avg.	0.885	0.139	0.885	0.885	0.885	0.758	0.953	0.955

```
=== Confusion Matrix ===

  a    b  <-- classified as
2596 192 |    a = 0
 335 1478 |   b = 1
```

Achieved an accuracy of 88.5 %

Possible Reason for good performance :

Because we binarized the classes meaning that the independence of data increases, making the model simpler. This also leads to increased bias and eventually better accuracy. The downside however is that there will be a lot of variance and the model will start to overfit.

Scalability :

```
Time taken to build model: 0.04 seconds
```

Took 0.04 second for 4,601 values, meaning that this is highly scalable. However due to it being so fast, sometimes it does not capture complex relationships between features like the way SVM and neural networks do. Therefore we can say that it is scalable but has some downsides as well.

```
=== Confusion Matrix ===  
      a    b  <-- classified as  
2596  192 |    a = 0  
 335 1478 |    b = 1
```

Using the confusion matrix:

Total count: $1904 + 884 + 78 + 1735 = 4601$

prior probability of class 0: $1904/4601 = 0.413$

prior probability of class 1: $1735/4601 = 0.377$