

# *CS 699: FINAL PROJECT*

High Crime Area Classification

*Emily Darigo*

12/11/2019

## Table of Contents

<i>Introduction .....</i>	<b>2</b>
<i>Chapter 1 - Group 1.....</i>	<b>3</b>
<i>Chapter 2 - Group 2.....</i>	<b>9</b>
<i>Chapter 3 - Group 3.....</i>	<b>15</b>
<i>Chapter 4 - Best Test Result.....</i>	<b>21</b>
<i>Chapter 5 - Discussion .....</i>	<b>22</b>
<i>Appendix I - Chapter 1 Training Results.....</i>	<b>24</b>
<i>Appendix II - Chapter 2 Training Results.....</i>	<b>34</b>
<i>Appendix III - Chapter 3 Training Results .....</i>	<b>44</b>

# Introduction

## Creating Test and Training Datasets

Weka was used to create the training and test dataset with the Resample filter under the supervised category. To avoid duplicate instances in the dataset, the resample filter option of noReplacement was set to TRUE. Since about 600 instances were needed for the test dataset, the sampleSizePercent option was set to 33% (600 is about 33% of the total dataset). The same percentage was used, but the inverseSelection option was set to TRUE in order to obtain the training dataset.

## Model Application

The dataset has two classifiers {1,0}. An instance classified as 1 is in a low crime area. Those classified as 0 are in a high crime area. This model could be used in law enforcement to decide how many officers should be on patrol and in which areas. If certain neighborhoods start to have similar socio-economic traits as other high crime areas, police can be more proactive and increase police presence.

## Performance Evaluation

Since the focus of this model is on identifying high crime areas (classifier 0), we will use the performance measures that suggest the model is correctly classifying a tuple as 0. Two measures we can use is precision and recall. In order to consider both measures, we can use the F-measure.

Although the focus is on correctly identifying high crime areas, we want to avoid a model that incorrectly identifies low crime areas. Therefore, we will use the weighted F-measure to account for both the 0 and 1 classifier.

The second measure will be the area under the ROC curve. Not only does this demonstrate the relationship between the true positive and false positive rate, it can also be used to describe the accuracy of the model. This measure will be used as a tie break in the case that two models have the same weighted F-measure.

If a model has the same weighted F-measure and ROC curve, the model with the higher F-measure for Class 0 will be chosen.

## Chapter 1 – Group 1

### Attribute Selection Method

The first group of attributes were selected with the CorrelationAttributeEval method. The attributes were ranked based on the correlation calculation. Those attributes with a value above 0.4 were chosen for a total of 11 attributes.

### Attribute List

Ranking	Attribute Number	Attribute
0.518	37	pctFemDivorc
0.496	35	pctMaleDivorc
0.484	40	pctKidsBornNevrMarr
0.462	26	pctPoverty
0.439	66	pctHousW0phone
0.439	4	pctWhite
0.434	57	pctPersOwnOccup
0.429	3	pctBlack
0.423	15	pctPubAsst
0.421	13	pctWdiv
0.403	59	pctSmallHousUnits

### Performance Measures Overview

Model	Weighted F-measure	ROC Curve
Naïve Bayes	0.823	0.878
J48	0.807	0.799
IBk	0.819	0.878
Bagging – Naïve Bayes	0.822	0.879
Bagging – J48	0.819	0.871
Bagging – IBk	0.820	0.886
AdaBoostM1 – Naïve Bayes	0.830	0.864
AdaBoostM1 – J48	0.797	0.837
AdaBoostM1 – IBk	0.819	0.813
MultilayerPerceptron	0.827	0.873

## Classifier Models

Results from training set evaluation and stratified cross-validation are available in Appendix I. Models below were built using the stratified cross-validation method.

### 1. Naïve Bayes

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacer
Instances: unknown (yet). Reading incrementally
Attributes: 12

== Summary ==

Correctly Classified Instances      492          82.1369 %
Incorrectly Classified Instances   107          17.8631 %
Kappa statistic                   0.6095
Mean absolute error               0.1895
Root mean squared error          0.3963
Total Number of Instances        599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
      0.786    0.161    0.712     0.786   0.747     0.611   0.878    0.775     0
      0.839    0.214    0.886     0.839   0.862     0.611   0.878    0.929     1
Weighted Avg.      0.821    0.196    0.827     0.821   0.823     0.611   0.878    0.877

== Confusion Matrix ==

  a   b   <-- classified as
158 43 |  a = 0
 64 334 |  b = 1
```

### 2. J48

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement
Instances: unknown (yet). Reading incrementally
Attributes: 12

== Summary ==

Correctly Classified Instances      484          80.8013 %
Incorrectly Classified Instances   115          19.1987 %
Kappa statistic                   0.5646
Mean absolute error               0.2627
Root mean squared error          0.3969
Total Number of Instances        599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
      0.692    0.133    0.724     0.692   0.707     0.565   0.799    0.654     0
      0.867    0.308    0.848     0.867   0.857     0.565   0.799    0.852     1
Weighted Avg.      0.808    0.250    0.806     0.808   0.807     0.565   0.799    0.785

== Confusion Matrix ==

  a   b   <-- classified as
139 62 |  a = 0
 53 345 |  b = 1
```

### 3. IBk (k = 10)

== Re-evaluation on test set ==

User supplied test set  
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replace  
Instances: unknown (yet). Reading incrementally  
Attributes: 12

== Summary ==

Correctly Classified Instances	490	81.803 %
Incorrectly Classified Instances	109	18.197 %
Kappa statistic	0.5964	
Mean absolute error	0.2495	
Root mean squared error	0.3565	
Total Number of Instances	599	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.751	0.148	0.719	0.751	0.735	0.597	0.878	0.777	0
1	0.852	0.249	0.871	0.852	0.861	0.597	0.878	0.913	1
Weighted Avg.	0.818	0.215	0.820	0.818	0.819	0.597	0.878	0.868	

== Confusion Matrix ==

a	b	<-- classified as
151	50	a = 0
59	339	b = 1

### 4. Bagging with Naïve Bayes

== Re-evaluation on test set ==

User supplied test set  
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replace  
Instances: unknown (yet). Reading incrementally  
Attributes: 12

== Summary ==

Correctly Classified Instances	491	81.9699 %
Incorrectly Classified Instances	108	18.0301 %
Kappa statistic	0.6054	
Mean absolute error	0.1905	
Root mean squared error	0.3951	
Total Number of Instances	599	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.781	0.161	0.710	0.781	0.744	0.607	0.879	0.775	0
1	0.839	0.219	0.884	0.839	0.861	0.607	0.879	0.929	1
Weighted Avg.	0.820	0.199	0.825	0.820	0.822	0.607	0.879	0.877	

== Confusion Matrix ==

a	b	<-- classified as
157	44	a = 0
64	334	b = 1

## 5. Bagging with J48

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replace
Instances: unknown (yet). Reading incrementally
Attributes: 12

== Summary ==

Correctly Classified Instances      491          81.9699 %
Incorrectly Classified Instances   108          18.0301 %
Kappa statistic                   0.5947
Mean absolute error               0.2535
Root mean squared error          0.3591
Total Number of Instances        599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
      0.726    0.133    0.734     0.726    0.730     0.595   0.871    0.805     0
      0.867    0.274    0.863     0.867    0.865     0.595   0.871    0.905     1
Weighted Avg.      0.820    0.226    0.819     0.820    0.819     0.595   0.871    0.871

== Confusion Matrix ==

  a   b   <-- classified as
146 55 |  a = 0
 53 345 |  b = 1
```

## 6. Bagging with IBk (k = 10)

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replace
Instances: unknown (yet). Reading incrementally
Attributes: 12

== Summary ==

Correctly Classified Instances      492          82.1369 %
Incorrectly Classified Instances   107          17.8631 %
Kappa statistic                   0.5919
Mean absolute error               0.2494
Root mean squared error          0.3518
Total Number of Instances        599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
      0.697    0.116    0.753     0.697    0.724     0.593   0.886    0.807     0
      0.884    0.303    0.852     0.884    0.868     0.593   0.886    0.927     1
Weighted Avg.      0.821    0.240    0.819     0.821    0.820     0.593   0.886    0.887

== Confusion Matrix ==

  a   b   <-- classified as
140 61 |  a = 0
 46 352 |  b = 1
```

## 7. AdaBoostM1 with Naïve Bayes

== Re-evaluation on test set ==

User supplied test set  
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replace  
Instances: unknown (yet). Reading incrementally  
Attributes: 12

== Summary ==

Correctly Classified Instances	497	82.9716 %
Incorrectly Classified Instances	102	17.0284 %
Kappa statistic	0.6191	
Mean absolute error	0.2533	
Root mean squared error	0.3541	
Total Number of Instances	599	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.751	0.131	0.744	0.751	0.748	0.619	0.864	0.759	0
1	0.869	0.249	0.874	0.869	0.872	0.619	0.864	0.895	1
Weighted Avg.	0.830	0.209	0.830	0.830	0.830	0.619	0.864	0.849	

== Confusion Matrix ==

a	b	<-- classified as
151	50	a = 0
52	346	b = 1

## 8. AdaBoostM1 with J48

== Re-evaluation on test set ==

User supplied test set  
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replace  
Instances: unknown (yet). Reading incrementally  
Attributes: 12

== Summary ==

Correctly Classified Instances	478	79.7997 %
Incorrectly Classified Instances	121	20.2003 %
Kappa statistic	0.5442	
Mean absolute error	0.2135	
Root mean squared error	0.4347	
Total Number of Instances	599	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.687	0.146	0.704	0.687	0.695	0.544	0.837	0.727	0
1	0.854	0.313	0.844	0.854	0.849	0.544	0.837	0.895	1
Weighted Avg.	0.798	0.257	0.797	0.798	0.797	0.544	0.837	0.839	

== Confusion Matrix ==

a	b	<-- classified as
138	63	a = 0
58	340	b = 1

## 9. AdaBoostM1 with IBk (k = 10)

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacer
Instances: unknown (yet). Reading incrementally
Attributes: 12

== Summary ==

Correctly Classified Instances      490          81.803 %
Incorrectly Classified Instances    109          18.197 %
Kappa statistic                      0.5964
Mean absolute error                  0.2696
Root mean squared error              0.3877
Total Number of Instances           599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class
      0.751    0.148    0.719     0.751    0.735     0.597    0.813     0.642     0
      0.852    0.249    0.871     0.852    0.861     0.597    0.813     0.858     1
Weighted Avg.      0.818    0.215    0.820     0.818    0.819     0.597    0.813     0.785

== Confusion Matrix ==

      a   b   <-- classified as
151  50 |  a = 0
 59 339 |  b = 1
```

## 10. MultilayerPerceptron

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replace
Instances: unknown (yet). Reading incrementally
Attributes: 12

== Summary ==

Correctly Classified Instances      496          82.8047 %
Incorrectly Classified Instances    103          17.1953 %
Kappa statistic                      0.6091
Mean absolute error                  0.2284
Root mean squared error              0.3579
Total Number of Instances           599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class
      0.716    0.116    0.758     0.716    0.737     0.610    0.873     0.758     0
      0.884    0.284    0.861     0.884    0.872     0.610    0.873     0.918     1
Weighted Avg.      0.828    0.227    0.826     0.828    0.827     0.610    0.873     0.865

== Confusion Matrix ==

      a   b   <-- classified as
144  57 |  a = 0
 46 352 |  b = 1
```

## Chapter 2 – Group 2

### Attribute Selection Method

The second group of attributes were chosen with the GainRatioAttribute method. The attributes were ranked based on the gain ratio calculation. Those with a gain ratio value of 0.1 or greater were used. This gives a list of 8 attributes.

### Attribute List

Ranking	Attribute Number	Attribute
0.11939	57	pctPersOwnOccup
0.11855	40	pctKidsBornNevrMarr
0.11452	80	persHomeless
0.11283	39	kidsBornNevrMarr
0.10964	26	pctPoverty
0.10649	3	pctBlack
0.10603	37	pctFemDivorc
0.10361	35	pctMaleDivorc

### Performance Measures Overview

Model	Weighted F-measure	ROC Curve
Naïve Bayes	0.770	0.892
J48	0.767	0.811
IBk	0.819	0.879
Bagging – Naïve Bayes	0.770	0.891
Bagging – J48	0.819	0.868
Bagging – IBk	0.812	0.876
AdaBoostM1 – Naïve Bayes	0.770	0.720
AdaBoostM1 – J48	0.792	0.853
AdaBoostM1 – IBk	0.819	0.805
MultilayerPerceptron	0.809	0.886

## Classifier Models

Results from training set evaluation and stratified cross-validation are available in Appendix II. Models below were built using the stratified cross-validation method.

### 1. Naïve Bayes

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement
Instances: unknown (yet). Reading incrementally
Attributes: 9

== Summary ==

Correctly Classified Instances      473          78.9649 %
Incorrectly Classified Instances    126          21.0351 %
Kappa statistic                      0.4695
Mean absolute error                  0.2106
Root mean squared error              0.4426
Total Number of Instances            599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class
0  0.463    0.045    0.838     0.463    0.596     0.507   0.892    0.798     0
1  0.955    0.537    0.779     0.955    0.858     0.507   0.892    0.936     1
Weighted Avg.  0.790    0.372    0.799     0.790    0.770     0.507   0.892    0.890

== Confusion Matrix ==

  a   b  <-- classified as
93 108 |  a = 0
18 380 |  b = 1
```

### 2. J48

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement
Instances: unknown (yet). Reading incrementally
Attributes: 9

== Summary ==

Correctly Classified Instances      469          78.2972 %
Incorrectly Classified Instances    130          21.7028 %
Kappa statistic                      0.4616
Mean absolute error                  0.2605
Root mean squared error              0.3972
Total Number of Instances            599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class
0  0.483    0.065    0.789     0.483    0.599     0.488   0.811    0.677     0
1  0.935    0.517    0.782     0.935    0.851     0.488   0.811    0.856     1
Weighted Avg.  0.783    0.366    0.784     0.783    0.767     0.488   0.811    0.796

== Confusion Matrix ==

  a   b  <-- classified as
97 104 |  a = 0
26 372 |  b = 1
```

### 3. IBk (k = 10)

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement
Instances: unknown (yet). Reading incrementally
Attributes: 9

== Summary ==

Correctly Classified Instances      490           81.803 %
Incorrectly Classified Instances   109           18.197 %
Kappa statistic                   0.5954
Mean absolute error               0.2468
Root mean squared error          0.3599
Total Number of Instances        599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
      0.746    0.146    0.721     0.746    0.733     0.596   0.879    0.770     0
      0.854    0.254    0.870     0.854    0.862     0.596   0.879    0.921     1
Weighted Avg.      0.818    0.217    0.820     0.818    0.819     0.596   0.879    0.870

== Confusion Matrix ==

  a   b  <-- classified as
150  51 |  a = 0
 58 340 |  b = 1
```

### 4. Bagging with Naïve Bayes

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement
Instances: unknown (yet). Reading incrementally
Attributes: 9

== Summary ==

Correctly Classified Instances      473           78.9649 %
Incorrectly Classified Instances   126           21.0351 %
Kappa statistic                   0.4695
Mean absolute error               0.2121
Root mean squared error          0.4422
Total Number of Instances        599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
      0.463    0.045    0.838     0.463    0.596     0.507   0.891    0.795     0
      0.955    0.537    0.779     0.955    0.858     0.507   0.891    0.936     1
Weighted Avg.      0.790    0.372    0.799     0.790    0.770     0.507   0.891    0.888

== Confusion Matrix ==

  a   b  <-- classified as
93 108 |  a = 0
 18 380 |  b = 1
```

## 5. Bagging with J48

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement
Instances: unknown (yet). Reading incrementally
Attributes: 9

== Summary ==

Correctly Classified Instances      493          82.3038 %
Incorrectly Classified Instances   106          17.6962 %
Kappa statistic                   0.5879
Mean absolute error               0.2562
Root mean squared error          0.3643
Total Number of Instances        599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
          0.662   0.095    0.778     0.662   0.715     0.592   0.868    0.770     0
          0.905   0.338    0.841     0.905   0.872     0.592   0.868    0.911     1
Weighted Avg.      0.823   0.257    0.820     0.823   0.819     0.592   0.868    0.863

== Confusion Matrix ==

  a   b  <-- classified as
133  68 |  a = 0
 38 360 |  b = 1
```

## 6. Bagging with IBk (k = 10)

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement-
Instances: unknown (yet). Reading incrementally
Attributes: 9

== Summary ==

Correctly Classified Instances      487          81.3022 %
Incorrectly Classified Instances   112          18.6978 %
Kappa statistic                   0.5776
Mean absolute error               0.2538
Root mean squared error          0.3626
Total Number of Instances        599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
          0.706   0.133    0.728     0.706   0.717     0.578   0.876    0.783     0
          0.867   0.294    0.854     0.867   0.860     0.578   0.876    0.931     1
Weighted Avg.      0.813   0.240    0.812     0.813   0.812     0.578   0.876    0.881

== Confusion Matrix ==

  a   b  <-- classified as
142  59 |  a = 0
 53 345 |  b = 1
```

## 7. AdaBoostM1 with Naïve Bayes

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement
Instances: unknown (yet). Reading incrementally
Attributes: 9

== Summary ==

Correctly Classified Instances      473          78.9649 %
Incorrectly Classified Instances   126          21.0351 %
Kappa statistic                   0.4695
Mean absolute error               0.321
Root mean squared error          0.4064
Total Number of Instances        599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
      0.463    0.045    0.838     0.463    0.596     0.507   0.720     0.593     0
      0.955    0.537    0.779     0.955    0.858     0.507   0.720     0.781     1
Weighted Avg.      0.790    0.372    0.799     0.790    0.770     0.507   0.720     0.718

== Confusion Matrix ==

  a   b   <-- classified as
93 108 |  a = 0
18 380 |  b = 1
```

## 8. AdaBoostM1 with J48

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement
Instances: unknown (yet). Reading incrementally
Attributes: 9

== Summary ==

Correctly Classified Instances      475          79.2988 %
Incorrectly Classified Instances   124          20.7012 %
Kappa statistic                   0.5323
Mean absolute error               0.2109
Root mean squared error          0.4171
Total Number of Instances        599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
      0.677    0.148    0.697     0.677    0.687     0.532   0.853     0.748     0
      0.852    0.323    0.839     0.852    0.845     0.532   0.853     0.909     1
Weighted Avg.      0.793    0.265    0.792     0.793    0.792     0.532   0.853     0.855

== Confusion Matrix ==

  a   b   <-- classified as
136 65 |  a = 0
59 339 |  b = 1
```

## 9. AdaBoostM1 with IBk (k = 10)

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement
Instances: unknown (yet). Reading incrementally
Attributes: 9

== Summary ==

Correctly Classified Instances      490          81.803 %
Incorrectly Classified Instances    109          18.197 %
Kappa statistic                      0.5954
Mean absolute error                  0.2786
Root mean squared error              0.3929
Total Number of Instances           599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class
0  0.746    0.146    0.721     0.746    0.733     0.596   0.805     0.650     0
1  0.854    0.254    0.870     0.854    0.862     0.596   0.805     0.849     1
Weighted Avg.  0.818    0.217    0.820     0.818    0.819     0.596   0.805     0.783

== Confusion Matrix ==

      a   b  <-- classified as
150  51 |  a = 0
 58 340 |  b = 1
```

## 10. MultilayerPerceptron

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement
Instances: unknown (yet). Reading incrementally
Attributes: 9

== Summary ==

Correctly Classified Instances      482          80.4674 %
Incorrectly Classified Instances    117          19.5326 %
Kappa statistic                      0.5869
Mean absolute error                  0.2879
Root mean squared error              0.3728
Total Number of Instances           599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class
0  0.831    0.209    0.668     0.831    0.741     0.596   0.886     0.776     0
1  0.791    0.169    0.903     0.791    0.843     0.596   0.886     0.938     1
Weighted Avg.  0.805    0.182    0.824     0.805    0.809     0.596   0.886     0.884

== Confusion Matrix ==

      a   b  <-- classified as
167  34 |  a = 0
 83 315 |  b = 1
```

## Chapter 3 – Group 3

### Attribute Selection Method

The third group uses the InfoGainAttribute method, which is then ranked based on info gain. Attributes that have more than 0.15 info gain and above are in the list below. There are 15 attributes in this group.

### Attribute List

Ranking	Attribute Number	Attribute
0.23202	37	pctFemDivorc
0.22885	40	pctKidsBornNevrMarr
0.223	25	persPoverty
0.20192	35	pctMaleDivorc
0.17893	66	pctHousWOPhone
0.17827	57	pctPersOwnOccup
0.17536	58	pctPopDenseHous
0.1733	4	pctWhite
0.15916	10	medIncome
0.15744	13	pctWdiv
0.15462	15	pctPubAsst
0.15395	59	pctSmallHousUnits
0.15287	17	medFamIncome
0.15281	63	pctHousOwnerOccup
0.15028	3	pctBlack

### Performance Measures Overview

Model	Weighted F-measure	ROC Curve
Naïve Bayes	0.832	0.889
J48	0.820	0.769
IBk	0.823	0.881
Bagging – Naïve Bayes	0.832	0.891
Bagging – J48	0.815	0.879
Bagging – IBk	0.837	0.885
AdaBoostM1 – Naïve Bayes	0.832	0.853
AdaBoostM1 – J48	0.794	0.857
AdaBoostM1 – IBk	0.823	0.812
MultilayerPerceptron	0.825	0.875

## Classifier Models

Results from training set evaluation and stratified cross-validation are available in Appendix III. Models below were built using the stratified cross-validation method.

### 1. Naïve Bayes

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement
Instances: unknown (yet). Reading incrementally
Attributes: 16

== Summary ==

Correctly Classified Instances      498          83.1386 %
Incorrectly Classified Instances   101          16.8614 %
Kappa statistic                   0.6251
Mean absolute error               0.1738
Root mean squared error           0.3861
Total Number of Instances         599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class
      0.766    0.136    0.740     0.766    0.753     0.625    0.889    0.797     0
      0.864    0.234    0.880     0.864    0.872     0.625    0.889    0.933     1
Weighted Avg.      0.831    0.201    0.833     0.831    0.832     0.625    0.889    0.887

== Confusion Matrix ==

  a   b  <-- classified as
154  47 |  a = 0
 54 344 |  b = 1
```

### 2. J48

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement
Instances: unknown (yet). Reading incrementally
Attributes: 16

== Summary ==

Correctly Classified Instances      492          82.1369 %
Incorrectly Classified Instances   107          17.8631 %
Kappa statistic                   0.5919
Mean absolute error               0.2313
Root mean squared error           0.4025
Total Number of Instances         599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class
      0.697    0.116    0.753     0.697    0.724     0.593    0.769    0.631     0
      0.884    0.303    0.852     0.884    0.868     0.593    0.769    0.827     1
Weighted Avg.      0.821    0.240    0.819     0.821    0.820     0.593    0.769    0.761

== Confusion Matrix ==

  a   b  <-- classified as
140  61 |  a = 0
 46 352 |  b = 1
```

### 3. IBk (k = 10)

== Re-evaluation on test set ==

User supplied test set  
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement-  
Instances: unknown (yet). Reading incrementally  
Attributes: 16

== Summary ==

Correctly Classified Instances	493	82.3038 %
Incorrectly Classified Instances	106	17.6962 %
Kappa statistic	0.6032	
Mean absolute error	0.2478	
Root mean squared error	0.3535	
Total Number of Instances	599	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.736	0.133	0.736	0.736	0.736	0.603	0.881	0.788	0
1	0.867	0.264	0.867	0.867	0.867	0.603	0.881	0.915	1
Weighted Avg.	0.823	0.220	0.823	0.823	0.823	0.603	0.881	0.873	

== Confusion Matrix ==

a	b	<-- classified as
148	53	a = 0
53	345	b = 1

### 4. Bagging with Naïve Bayes

== Re-evaluation on test set ==

User supplied test set  
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement-  
Instances: unknown (yet). Reading incrementally  
Attributes: 16

== Summary ==

Correctly Classified Instances	498	83.1386 %
Incorrectly Classified Instances	101	16.8614 %
Kappa statistic	0.6251	
Mean absolute error	0.174	
Root mean squared error	0.3824	
Total Number of Instances	599	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.766	0.136	0.740	0.766	0.753	0.625	0.891	0.804	0
1	0.864	0.234	0.880	0.864	0.872	0.625	0.891	0.935	1
Weighted Avg.	0.831	0.201	0.833	0.831	0.832	0.625	0.891	0.891	

== Confusion Matrix ==

a	b	<-- classified as
154	47	a = 0
54	344	b = 1

## 5. Bagging with J48

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement-V
Instances: unknown (yet). Reading incrementally
Attributes: 16

== Summary ==

Correctly Classified Instances      490          81.803 %
Incorrectly Classified Instances   109          18.197 %
Kappa statistic                   0.5811
Mean absolute error               0.2467
Root mean squared error           0.3543
Total Number of Instances         599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
      0.677    0.111    0.756     0.677    0.714     0.583   0.879    0.816     0
      0.889    0.323    0.845     0.889    0.867     0.583   0.879    0.919     1
Weighted Avg.      0.818    0.252    0.815     0.818    0.815     0.583   0.879    0.884

== Confusion Matrix ==

  a   b  <-- classified as
136  65 |  a = 0
 44 354 |  b = 1
```

## 6. Bagging with IBk (k = 10)

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement-I
Instances: unknown (yet). Reading incrementally
Attributes: 16

== Summary ==

Correctly Classified Instances      503          83.9733 %
Incorrectly Classified Instances   96          16.0267 %
Kappa statistic                   0.6296
Mean absolute error               0.2485
Root mean squared error           0.3501
Total Number of Instances         599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
      0.701    0.090    0.797     0.701    0.746     0.632   0.885    0.817     0
      0.910    0.299    0.858     0.910    0.883     0.632   0.885    0.927     1
Weighted Avg.      0.840    0.229    0.837     0.840    0.837     0.632   0.885    0.890

== Confusion Matrix ==

  a   b  <-- classified as
141  60 |  a = 0
 36 362 |  b = 1
```

## 7. AdaBoostM1 with Naïve Bayes

== Re-evaluation on test set ==

User supplied test set  
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement-  
Instances: unknown (yet). Reading incrementally  
Attributes: 16

== Summary ==

Correctly Classified Instances	498	83.1386 %
Incorrectly Classified Instances	101	16.8614 %
Kappa statistic	0.6251	
Mean absolute error	0.2375	
Root mean squared error	0.3532	
Total Number of Instances	599	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.766	0.136	0.740	0.766	0.753	0.625	0.853	0.746	0
1	0.864	0.234	0.880	0.864	0.872	0.625	0.853	0.880	1
Weighted Avg.	0.831	0.201	0.833	0.831	0.832	0.625	0.853	0.835	

== Confusion Matrix ==

a	b	<-- classified as
154	47	a = 0
54	344	b = 1

## 8. AdaBoostM1 with J48

== Re-evaluation on test set ==

User supplied test set  
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement-  
Instances: unknown (yet). Reading incrementally  
Attributes: 16

== Summary ==

Correctly Classified Instances	476	79.4658 %
Incorrectly Classified Instances	123	20.5342 %
Kappa statistic	0.5367	
Mean absolute error	0.204	
Root mean squared error	0.4317	
Total Number of Instances	599	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.682	0.148	0.699	0.682	0.690	0.537	0.857	0.769	0
1	0.852	0.318	0.841	0.852	0.846	0.537	0.857	0.904	1
Weighted Avg.	0.795	0.261	0.793	0.795	0.794	0.537	0.857	0.859	

== Confusion Matrix ==

a	b	<-- classified as
137	64	a = 0
59	339	b = 1

## 9. AdaBoostM1 with IBk (k = 10)

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement-V
Instances: unknown (yet). Reading incrementally
Attributes: 16

== Summary ==

Correctly Classified Instances      493          82.3038 %
Incorrectly Classified Instances   106          17.6962 %
Kappa statistic                   0.6032
Mean absolute error               0.2737
Root mean squared error           0.3839
Total Number of Instances         599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
      0.736    0.133    0.736     0.736   0.736    0.603   0.812    0.652     0
      0.867    0.264    0.867     0.867   0.867    0.603   0.812    0.854     1
Weighted Avg.      0.823    0.220    0.823     0.823   0.823    0.603   0.812    0.786

== Confusion Matrix ==

      a   b   <-- classified as
148 53 |   a = 0
 53 345 |   b = 1
```

## 10. MultilayerPerceptron

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement-V
Instances: unknown (yet). Reading incrementally
Attributes: 16

== Summary ==

Correctly Classified Instances      493          82.3038 %
Incorrectly Classified Instances   106          17.6962 %
Kappa statistic                   0.6127
Mean absolute error               0.2323
Root mean squared error           0.3649
Total Number of Instances         599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
      0.786    0.158    0.715     0.786   0.749    0.614   0.875    0.764     0
      0.842    0.214    0.886     0.842   0.863    0.614   0.875    0.925     1
Weighted Avg.      0.823    0.195    0.829     0.823   0.825    0.614   0.875    0.871

== Confusion Matrix ==

      a   b   <-- classified as
158 43 |   a = 0
 63 335 |   b = 1
```

## Chapter 4 – Best Test Result

### The Best Test Result

Group	Model	Weighted F-measure	ROC Curve
1	AdaBoostM1 – Naïve Bayes	0.830	0.864
2	IBk	0.819	0.879
3	Bagging – IBk	0.837	0.885

The best model from each group is in the table above. Based on the performance measures discussed in the introduction section, the third group has the best model, with the highest weighted F-measure and ROC Curve area. The full test results for the best model are displayed below.

### Group 3 – Bagging with IBk (k = 10)

```
== Re-evaluation on test set ==

User supplied test set
Relation: project-crime-binary-weka.filters.supervised.instance.Resample-B0.0-S1-Z14.16-no-replacement-
Instances: unknown (yet). Reading incrementally
Attributes: 16

== Summary ==

Correctly Classified Instances      503           83.9733 %
Incorrectly Classified Instances    96            16.0267 %
Kappa statistic                   0.6296
Mean absolute error               0.2485
Root mean squared error          0.3501
Total Number of Instances         599

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class
      0.701    0.090    0.797     0.701    0.746     0.632    0.885    0.817    0
      0.910    0.299    0.858     0.910    0.883     0.632    0.885    0.927    1
Weighted Avg.      0.840    0.229    0.837     0.840    0.837     0.632    0.885    0.890

== Confusion Matrix ==

  a   b  <-- classified as
141  60 |  a = 0
 36 362 |  b = 1
```

## Chapter 5 - Discussion

### Best Model Criteria

The best model was chosen based on the weighted F-Measure and ROC curve performance measures. The F-measure was chosen as an evaluation between the precision and recall measures. The weighted F-measure evaluates the performance for both classes, as the model aims to ensure exactness and completeness.

The ROC curve was used to evaluate the accuracy of the model as a whole. The closer the area is to 1, the better fit of the model. In the case that multiple models had the same F-measure, it was used as a secondary measurement to determine the better model.

Compared to other models, Bagging with IBk in Group 3 had both the best weighted F-measure and best ROC curve. Group 3 also had the most attributes (total of 15), which may have attributed to the improvement in the model, as it had the most data to use.

### Most Relevant Attributes

The five attributes that are most relevant to the model are the following:

1. pctFemDivorc
2. pctMaleDivorc
3. pctKidsBornNevrMarr
4. persPoverty
5. pctBlack

Most of these attributes had a higher information gain and all of the attributes were chosen across the three groups. This indicates that these attributes were continuously chosen with different attribute selection methods. Particularly the percentage of kids born to unmarried parents (pctKidsBornNevrMarr) was ranked high in all three attribute selection methods.

When these attributes were removed from the dataset and re-evaluated, the F-measure and ROC curve area both decreased. This indicates that the attributes are important to the model performance.

## **Project Observations**

Most of the models performed quite well, there weren't any models that were extremely low performing. However, less models had a high precision/recall/F-measure for the 0 class. Models seemed to be able to classify tuples as 1 better than classifying tuples as 0.

Although we did not do clustering in this project, there does seem to be a theme among the attributes chosen, not only in group 3, but in all groups. Firstly, it seems that the family situation (couples married/not married/divorced) has an impact on if an area is classified as a high or low crime area. In relation, the children involved are also a factor, specifically children in non-married households.

Less surprising, race and measures of income were common attributes. I did expect a wider variety of races to be chosen, not just the amount of Black and White people in an area.

It was interesting that group 3 included the attribute percentage of houses without a phone (pctHousWOPhone). This would certainly be something that may not apply today with the commonality of smart phones.

## **Other Lessons and Observations**

I learned how to use Weka at each stage of the project process from preprocessing to model testing. In the preprocessing stage, I was able to split the data with filters. At first, I had attempted to split the data through a different method (not using Weka). I quickly saw that transferring the dataset from .arff to .csv and back had affected any prior discretization.

Other than learning how to use Weka in general, I learned how each classifier model can differ and the number of potential classifiers there are to choose from. Probably the biggest lesson is that choosing a certain classifier or performance measure always depends on the goal of the model.

## Appendix I – Chapter 1 Training Results

### Naïve Bayes Training Model Results

```
Time taken to build model: 0.01 seconds
== Evaluation on training set ==
Time taken to test model on training data: 0.02 seconds
== Summary ==
Correctly Classified Instances      953          78.1148 %
Incorrectly Classified Instances   267          21.8852 %
Kappa statistic                   0.5191
Mean absolute error               0.2198
Root mean squared error          0.4324
Relative absolute error           49.2459 %
Root relative squared error     91.5357 %
Total Number of Instances        1220

== Detailed Accuracy By Class ==
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
          0.715    0.185    0.661    0.715    0.687    0.520    0.845    0.714    0
          0.815    0.285    0.849    0.815    0.832    0.520    0.845    0.904    1
Weighted Avg.      0.781    0.252    0.786    0.781    0.783    0.520    0.845    0.840

== Confusion Matrix ==
      a   b  <-- classified as
293 117 |  a = 0
150 660 |  b = 1

== Stratified cross-validation ==
== Summary ==
Correctly Classified Instances      956          78.3607 %
Incorrectly Classified Instances   264          21.6393 %
Kappa statistic                   0.5248
Mean absolute error               0.2209
Root mean squared error          0.4339
Relative absolute error           49.4908 %
Root relative squared error     91.8587 %
Total Number of Instances        1220

== Detailed Accuracy By Class ==
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
          0.720    0.184    0.664    0.720    0.691    0.526    0.843    0.707    0
          0.816    0.280    0.852    0.816    0.834    0.526    0.843    0.901    1
Weighted Avg.      0.784    0.248    0.789    0.784    0.786    0.526    0.843    0.836

== Confusion Matrix ==
      a   b  <-- classified as
295 115 |  a = 0
149 661 |  b = 1
```

## J48 Training Model Results

Time taken to build model: 0.1 seconds

== Evaluation on training set ==

Time taken to test model on training data: 0.03 seconds

== Summary ==

Correctly Classified Instances	1062	87.0492 %
Incorrectly Classified Instances	158	12.9508 %
Kappa statistic	0.7037	
Mean absolute error	0.2109	
Root mean squared error	0.3248	
Relative absolute error	47.2612 %	
Root relative squared error	68.7535 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.766	0.077	0.835	0.766	0.799	0.705	0.895	0.807	0
1	0.923	0.234	0.886	0.923	0.904	0.705	0.895	0.918	1
Weighted Avg.	0.870	0.181	0.869	0.870	0.869	0.705	0.895	0.881	

== Confusion Matrix ==

a	b	<-- classified as
314	96	a = 0
62	748	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	943	77.2951 %
Incorrectly Classified Instances	277	22.7049 %
Kappa statistic	0.4878	
Mean absolute error	0.2956	
Root mean squared error	0.4246	
Relative absolute error	66.226 %	
Root relative squared error	89.8982 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.649	0.164	0.667	0.649	0.658	0.488	0.765	0.588	0
1	0.836	0.351	0.825	0.836	0.830	0.488	0.765	0.815	1
Weighted Avg.	0.773	0.288	0.772	0.773	0.772	0.488	0.765	0.739	

== Confusion Matrix ==

a	b	<-- classified as
266	144	a = 0
133	677	b = 1

## IBk (k=10) Training Model Results

```
Time taken to build model: 0 seconds
== Evaluation on training set ==
Time taken to test model on training data: 0.32 seconds
== Summary ==
Correctly Classified Instances      1014          83.1148 %
Incorrectly Classified Instances   206           16.8852 %
Kappa statistic                   0.6234
Mean absolute error               0.2402
Root mean squared error          0.3446
Relative absolute error           53.8171 %
Root relative squared error     72.9593 %
Total Number of Instances        1220

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
      0.759    0.132    0.744     0.759    0.751     0.624   0.901     0.807    0
      0.868    0.241    0.877     0.868    0.872     0.624   0.901     0.939    1
Weighted Avg.      0.831    0.205    0.832     0.831    0.832     0.624   0.901     0.895

== Confusion Matrix ==

      a   b   <-- classified as
 311  99 |  a = 0
107 703 |  b = 1

== Stratified cross-validation ==
== Summary ==
Correctly Classified Instances      965          79.0984 %
Incorrectly Classified Instances   255          20.9016 %
Kappa statistic                   0.5308
Mean absolute error               0.2681
Root mean squared error          0.3807
Relative absolute error           60.0595 %
Root relative squared error     80.6036 %
Total Number of Instances        1220

== Detailed Accuracy By Class ==

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
      0.685    0.156    0.690     0.685    0.688     0.531   0.848     0.726    0
      0.844    0.315    0.841     0.844    0.843     0.531   0.848     0.893    1
Weighted Avg.      0.791    0.261    0.791     0.791    0.791     0.531   0.848     0.837

== Confusion Matrix ==

      a   b   <-- classified as
 281 129 |  a = 0
126 684 |  b = 1
```

## Bagging – Naïve Bayes Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.1 seconds

== Summary ==

Correctly Classified Instances	954	78.1967 %
Incorrectly Classified Instances	266	21.8033 %
Kappa statistic	0.5229	
Mean absolute error	0.2204	
Root mean squared error	0.4313	
Relative absolute error	49.3788 %	
Root relative squared error	91.3154 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.724	0.189	0.660	0.724	0.691	0.524	0.846	0.714	0
1	0.811	0.276	0.853	0.811	0.832	0.524	0.846	0.904	1
Weighted Avg.	0.782	0.246	0.788	0.782	0.784	0.524	0.846	0.840	

== Confusion Matrix ==

a	b	<-- classified as
297	113	a = 0
153	657	b = 1

== Classifier model (full training set) ==

Bagging with 10 iterations and base learner

weka.classifiers.bayes.NaiveBayes

Time taken to build model: 0.04 seconds

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	955	78.2787 %
Incorrectly Classified Instances	265	21.7213 %
Kappa statistic	0.5216	
Mean absolute error	0.2216	
Root mean squared error	0.4334	
Relative absolute error	49.6394 %	
Root relative squared error	91.7554 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.712	0.181	0.665	0.712	0.688	0.522	0.843	0.710	0
1	0.819	0.288	0.849	0.819	0.833	0.522	0.843	0.900	1
Weighted Avg.	0.783	0.252	0.787	0.783	0.785	0.522	0.843	0.836	

== Confusion Matrix ==

a	b	<-- classified as
292	118	a = 0
147	663	b = 1

## Bagging – J48 Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.01 seconds

== Summary ==

Correctly Classified Instances	1147	94.0164 %
Incorrectly Classified Instances	73	5.9836 %
Kappa statistic	0.8639	
Mean absolute error	0.1537	
Root mean squared error	0.2352	
Relative absolute error	34.4309 %	
Root relative squared error	49.7974 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.880	0.030	0.938	0.880	0.908	0.865	0.978	0.967	0
1	0.970	0.120	0.941	0.970	0.956	0.865	0.978	0.986	1
Weighted Avg.	0.940	0.089	0.940	0.940	0.940	0.865	0.978	0.979	

== Confusion Matrix ==

a	b	<-- classified as
361	49	a = 0
24	786	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	964	79.0164 %
Incorrectly Classified Instances	256	20.9836 %
Kappa statistic	0.5176	
Mean absolute error	0.2709	
Root mean squared error	0.3875	
Relative absolute error	60.6983 %	
Root relative squared error	82.031 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.637	0.132	0.709	0.637	0.671	0.519	0.843	0.730	0
1	0.868	0.363	0.825	0.868	0.846	0.519	0.843	0.903	1
Weighted Avg.	0.790	0.286	0.786	0.790	0.787	0.519	0.843	0.845	

== Confusion Matrix ==

a	b	<-- classified as
261	149	a = 0
107	703	b = 1

## Bagging – IBk (k=10) Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 1.26 seconds

== Summary ==

Correctly Classified Instances	1005	82.377 %
Incorrectly Classified Instances	215	17.623 %
Kappa statistic	0.5976	
Mean absolute error	0.2424	
Root mean squared error	0.3432	
Relative absolute error	54.3105 %	
Root relative squared error	72.6562 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.700	0.114	0.757	0.700	0.728	0.599	0.904	0.834	0
1	0.886	0.300	0.854	0.886	0.870	0.599	0.904	0.951	1
Weighted Avg.	0.824	0.237	0.821	0.824	0.822	0.599	0.904	0.911	

== Confusion Matrix ==

a	b	<-- classified as
287	123	a = 0
92	718	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	965	79.0984 %
Incorrectly Classified Instances	255	20.9016 %
Kappa statistic	0.5174	
Mean absolute error	0.2692	
Root mean squared error	0.3807	
Relative absolute error	60.3027 %	
Root relative squared error	80.6006 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.629	0.127	0.715	0.629	0.669	0.520	0.848	0.743	0
1	0.873	0.371	0.823	0.873	0.847	0.520	0.848	0.897	1
Weighted Avg.	0.791	0.289	0.787	0.791	0.787	0.520	0.848	0.845	

== Confusion Matrix ==

a	b	<-- classified as
258	152	a = 0
103	707	b = 1

## AdaBoostM1 – Naïve Bayes Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.07 seconds

== Summary ==

Correctly Classified Instances	976	80	%
Incorrectly Classified Instances	244	20	%
Kappa statistic	0.5518		
Mean absolute error	0.28		
Root mean squared error	0.3839		
Relative absolute error	62.739 %		
Root relative squared error	81.2725 %		
Total Number of Instances	1220		

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.702	0.151	0.702	0.702	0.702	0.552	0.827	0.694	0
1	0.849	0.298	0.849	0.849	0.849	0.552	0.827	0.871	1
Weighted Avg.	0.800	0.248	0.800	0.800	0.800	0.552	0.827	0.812	

== Confusion Matrix ==

a	b	<-- classified as
288	122	a = 0
122	688	b = 1

Time taken to build model: 0.12 seconds

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	963	78.9344 %
Incorrectly Classified Instances	257	21.0656 %
Kappa statistic	0.5242	
Mean absolute error	0.277	
Root mean squared error	0.3926	
Relative absolute error	62.0666 %	
Root relative squared error	83.1152 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.671	0.151	0.693	0.671	0.682	0.524	0.815	0.673	0
1	0.849	0.329	0.836	0.849	0.843	0.524	0.815	0.857	1
Weighted Avg.	0.789	0.269	0.788	0.789	0.788	0.524	0.815	0.795	

== Confusion Matrix ==

a	b	<-- classified as
275	135	a = 0
122	688	b = 1

## AdaBoostM1 – J48 Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.01 seconds

== Summary ==

Correctly Classified Instances	1220	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.0003		
Root mean squared error	0.0016		
Relative absolute error	0.0573 %		
Root relative squared error	0.3437 %		
Total Number of Instances	1220		

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	0
1	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

== Confusion Matrix ==

a	b	<-- classified as
410	0	a = 0
0	810	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	947	77.623 %
Incorrectly Classified Instances	273	22.377 %
Kappa statistic	0.4934	
Mean absolute error	0.2283	
Root mean squared error	0.4545	
Relative absolute error	51.1509 %	
Root relative squared error	96.2153 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.646	0.158	0.674	0.646	0.660	0.494	0.814	0.722	0
1	0.842	0.354	0.825	0.842	0.833	0.494	0.814	0.875	1
Weighted Avg.	0.776	0.288	0.774	0.776	0.775	0.494	0.814	0.823	

== Confusion Matrix ==

a	b	<-- classified as
265	145	a = 0
128	682	b = 1

## AdaBoostM1 – IBk (k=10) Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.24 seconds

== Summary ==

Correctly Classified Instances	1014	83.1148 %
Incorrectly Classified Instances	206	16.8852 %
Kappa statistic	0.6234	
Mean absolute error	0.2539	
Root mean squared error	0.3625	
Relative absolute error	56.8817 %	
Root relative squared error	76.7366 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.759	0.132	0.744	0.759	0.751	0.624	0.856	0.717	0
1	0.868	0.241	0.877	0.868	0.872	0.624	0.856	0.886	1
Weighted Avg.	0.831	0.205	0.832	0.831	0.832	0.624	0.856	0.829	

== Confusion Matrix ==

a	b	<-- classified as
311	99	a = 0
107	703	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	965	79.0984 %
Incorrectly Classified Instances	255	20.9016 %
Kappa statistic	0.5308	
Mean absolute error	0.29	
Root mean squared error	0.4116	
Relative absolute error	64.9802 %	
Root relative squared error	87.14 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.685	0.156	0.690	0.685	0.688	0.531	0.774	0.620	0
1	0.844	0.315	0.841	0.844	0.843	0.531	0.774	0.828	1
Weighted Avg.	0.791	0.261	0.791	0.791	0.791	0.531	0.774	0.758	

== Confusion Matrix ==

a	b	<-- classified as
281	129	a = 0
126	684	b = 1

## MultilayerPerceptron

== Evaluation on training set ==

Time taken to test model on training data: 0.01 seconds

== Summary ==

Correctly Classified Instances	1051	86.1475 %
Incorrectly Classified Instances	169	13.8525 %
Kappa statistic	0.6898	
Mean absolute error	0.2107	
Root mean squared error	0.3321	
Relative absolute error	47.216 %	
Root relative squared error	70.2961 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.795	0.105	0.793	0.795	0.794	0.690	0.897	0.844	0
1	0.895	0.205	0.896	0.895	0.896	0.690	0.897	0.922	1
Weighted Avg.	0.861	0.171	0.862	0.861	0.862	0.690	0.897	0.896	

== Confusion Matrix ==

a	b	<-- classified as
326	84	a = 0
85	725	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	971	79.5902 %
Incorrectly Classified Instances	249	20.4098 %
Kappa statistic	0.5351	
Mean absolute error	0.2513	
Root mean squared error	0.3929	
Relative absolute error	56.3106 %	
Root relative squared error	83.1762 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.663	0.137	0.710	0.663	0.686	0.536	0.836	0.737	0
1	0.863	0.337	0.835	0.863	0.849	0.536	0.836	0.885	1
Weighted Avg.	0.796	0.270	0.793	0.796	0.794	0.536	0.836	0.836	

== Confusion Matrix ==

a	b	<-- classified as
272	138	a = 0
111	699	b = 1

## Appendix II – Chapter 2 Training Results

### Naïve Bayes Training Model Results

```
== Evaluation on training set ==
Time taken to test model on training data: 0.01 seconds
== Summary ==
Correctly Classified Instances      935          76.6393 %
Incorrectly Classified Instances    285          23.3607 %
Kappa statistic                      0.4011
Mean absolute error                  0.2361
Root mean squared error              0.4648
Relative absolute error              52.9063 %
Root relative squared error         98.3996 %
Total Number of Instances           1220

== Detailed Accuracy By Class ==
      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class
          0.398    0.047    0.811    0.398    0.534    0.446    0.855    0.742     0
          0.953    0.602    0.758    0.953    0.844    0.446    0.855    0.907     1
Weighted Avg.      0.766    0.416    0.776    0.766    0.740    0.446    0.855    0.852

== Confusion Matrix ==
      a   b  <-- classified as
163 247 |  a = 0
 38 772 |  b = 1

== Stratified cross-validation ==
== Summary ==
Correctly Classified Instances      931          76.3115 %
Incorrectly Classified Instances    289          23.6885 %
Kappa statistic                      0.3952
Mean absolute error                  0.2373
Root mean squared error              0.4655
Relative absolute error              53.1612 %
Root relative squared error         98.5391 %
Total Number of Instances           1220

== Detailed Accuracy By Class ==
      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class
          0.400    0.053    0.792    0.400    0.532    0.437    0.853    0.737     0
          0.947    0.600    0.757    0.947    0.841    0.437    0.853    0.904     1
Weighted Avg.      0.763    0.416    0.769    0.763    0.737    0.437    0.853    0.848

== Confusion Matrix ==
      a   b  <-- classified as
164 246 |  a = 0
 43 767 |  b = 1
```

## J48 Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0 seconds

== Summary ==

Correctly Classified Instances	1059	86.8033 %
Incorrectly Classified Instances	161	13.1967 %
Kappa statistic	0.6802	
Mean absolute error	0.1944	
Root mean squared error	0.3118	
Relative absolute error	43.5562 %	
Root relative squared error	66.0036 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.651	0.022	0.937	0.651	0.768	0.702	0.911	0.857	0
1	0.978	0.349	0.847	0.978	0.908	0.702	0.911	0.929	1
Weighted Avg.	0.868	0.239	0.877	0.868	0.861	0.702	0.911	0.905	

== Confusion Matrix ==

a	b	<-- classified as
267	143	a = 0
18	792	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	964	79.0164 %
Incorrectly Classified Instances	256	20.9836 %
Kappa statistic	0.4964	
Mean absolute error	0.2736	
Root mean squared error	0.4008	
Relative absolute error	61.3002 %	
Root relative squared error	84.848 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.554	0.090	0.757	0.554	0.639	0.508	0.808	0.678	0
1	0.910	0.446	0.801	0.910	0.852	0.508	0.808	0.849	1
Weighted Avg.	0.790	0.327	0.786	0.790	0.781	0.508	0.808	0.792	

== Confusion Matrix ==

a	b	<-- classified as
227	183	a = 0
73	737	b = 1

## IBk (k=10) Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.13 seconds

== Summary ==

Correctly Classified Instances	1012	82.9508 %
Incorrectly Classified Instances	208	17.0492 %
Kappa statistic	0.6161	
Mean absolute error	0.2427	
Root mean squared error	0.3472	
Relative absolute error	54.3863 %	
Root relative squared error	73.5135 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.737	0.123	0.751	0.737	0.744	0.616	0.898	0.803	0
1	0.877	0.263	0.868	0.877	0.872	0.616	0.898	0.937	1
Weighted Avg.	0.830	0.216	0.829	0.830	0.829	0.616	0.898	0.892	

== Confusion Matrix ==

a	b	<-- classified as
302	108	a = 0
100	710	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	952	78.0328 %
Incorrectly Classified Instances	268	21.9672 %
Kappa statistic	0.5054	
Mean absolute error	0.2726	
Root mean squared error	0.3861	
Relative absolute error	61.0694 %	
Root relative squared error	81.7281 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.663	0.160	0.677	0.663	0.670	0.505	0.839	0.714	0
1	0.840	0.337	0.831	0.840	0.835	0.505	0.839	0.887	1
Weighted Avg.	0.780	0.277	0.779	0.780	0.780	0.505	0.839	0.829	

== Confusion Matrix ==

a	b	<-- classified as
272	138	a = 0
130	680	b = 1

## Bagging – Naïve Bayes Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.09 seconds

== Summary ==

Correctly Classified Instances	929	76.1475 %
Incorrectly Classified Instances	291	23.8525 %
Kappa statistic	0.3877	
Mean absolute error	0.2372	
Root mean squared error	0.4638	
Relative absolute error	53.1486 %	
Root relative squared error	98.1909 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.388	0.049	0.799	0.388	0.522	0.433	0.856	0.744	0
1	0.951	0.612	0.754	0.951	0.841	0.433	0.856	0.907	1
Weighted Avg.	0.761	0.423	0.769	0.761	0.734	0.433	0.856	0.852	

== Confusion Matrix ==

a	b	<-- classified as
159	251	a = 0
40	770	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	935	76.6393 %
Incorrectly Classified Instances	285	23.3607 %
Kappa statistic	0.4109	
Mean absolute error	0.2318	
Root mean squared error	0.4485	
Relative absolute error	51.9426 %	
Root relative squared error	94.9463 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.427	0.062	0.778	0.427	0.551	0.445	0.852	0.739	0
1	0.938	0.573	0.764	0.938	0.842	0.445	0.852	0.903	1
Weighted Avg.	0.766	0.401	0.769	0.766	0.744	0.445	0.852	0.848	

== Confusion Matrix ==

a	b	<-- classified as
175	235	a = 0
50	760	b = 1

## Bagging – J48 Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.01 seconds

== Summary ==

Correctly Classified Instances	1134	92.9508 %
Incorrectly Classified Instances	86	7.0492 %
Kappa statistic	0.8373	
Mean absolute error	0.1653	
Root mean squared error	0.2468	
Relative absolute error	37.042 %	
Root relative squared error	52.2383 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.837	0.023	0.948	0.837	0.889	0.841	0.977	0.963	0
1	0.977	0.163	0.922	0.977	0.948	0.841	0.977	0.988	1
Weighted Avg.	0.930	0.116	0.931	0.930	0.928	0.841	0.977	0.979	

== Confusion Matrix ==

a	b	<-- classified as
343	67	a = 0
19	791	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	970	79.5082 %
Incorrectly Classified Instances	250	20.4918 %
Kappa statistic	0.5248	
Mean absolute error	0.2644	
Root mean squared error	0.3827	
Relative absolute error	59.2293 %	
Root relative squared error	81.0127 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.627	0.120	0.726	0.627	0.673	0.528	0.849	0.748	0
1	0.880	0.373	0.823	0.880	0.851	0.528	0.849	0.910	1
Weighted Avg.	0.795	0.288	0.791	0.795	0.791	0.528	0.849	0.855	

== Confusion Matrix ==

a	b	<-- classified as
257	153	a = 0
97	713	b = 1

## Bagging – IBk (k=10) Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 1 seconds

== Summary ==

Correctly Classified Instances	1002	82.1311 %
Incorrectly Classified Instances	218	17.8689 %
Kappa statistic	0.5861	
Mean absolute error	0.2466	
Root mean squared error	0.3488	
Relative absolute error	55.2553 %	
Root relative squared error	73.8357 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.668	0.101	0.770	0.668	0.715	0.589	0.897	0.823	0
1	0.899	0.332	0.843	0.899	0.870	0.589	0.897	0.946	1
Weighted Avg.	0.821	0.254	0.818	0.821	0.818	0.589	0.897	0.905	

== Confusion Matrix ==

a	b	<-- classified as
274	136	a = 0
82	728	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	960	78.6885 %
Incorrectly Classified Instances	260	21.3115 %
Kappa statistic	0.5076	
Mean absolute error	0.2728	
Root mean squared error	0.3841	
Relative absolute error	61.1089 %	
Root relative squared error	81.3069 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.622	0.130	0.708	0.622	0.662	0.510	0.844	0.727	0
1	0.870	0.378	0.820	0.870	0.844	0.510	0.844	0.902	1
Weighted Avg.	0.787	0.295	0.782	0.787	0.783	0.510	0.844	0.843	

== Confusion Matrix ==

a	b	<-- classified as
255	155	a = 0
105	705	b = 1

## AdaBoostM1 – Naïve Bayes Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.04 seconds

== Summary ==

Correctly Classified Instances	935	76.6393 %
Incorrectly Classified Instances	285	23.3607 %
Kappa statistic	0.4011	
Mean absolute error	0.3345	
Root mean squared error	0.4219	
Relative absolute error	74.9417 %	
Root relative squared error	89.3067 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.398	0.047	0.811	0.398	0.534	0.446	0.704	0.579	0
1	0.953	0.602	0.758	0.953	0.844	0.446	0.704	0.771	1
Weighted Avg.	0.766	0.416	0.776	0.766	0.740	0.446	0.704	0.706	

== Confusion Matrix ==

a	b	<-- classified as
163	247	a = 0
38	772	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	931	76.3115 %
Incorrectly Classified Instances	289	23.6885 %
Kappa statistic	0.3952	
Mean absolute error	0.3295	
Root mean squared error	0.4217	
Relative absolute error	73.8317 %	
Root relative squared error	89.2781 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.400	0.053	0.792	0.400	0.532	0.437	0.708	0.623	0
1	0.947	0.600	0.757	0.947	0.841	0.437	0.708	0.786	1
Weighted Avg.	0.763	0.416	0.769	0.763	0.737	0.437	0.708	0.731	

== Confusion Matrix ==

a	b	<-- classified as
164	246	a = 0
43	767	b = 1

## AdaBoostM1 – J48 Training Model Results

Time taken to build model: 0.19 seconds

== Evaluation on training set ==

Time taken to test model on training data: 0.01 seconds

== Summary ==

Correctly Classified Instances	1220	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.0037		
Root mean squared error	0.0139		
Relative absolute error	0.8232 %		
Root relative squared error	2.9405 %		
Total Number of Instances	1220		

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	0
0	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

== Confusion Matrix ==

	a	b	<-- classified as
410	0		a = 0
0	810		b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	938	76.8852 %
Incorrectly Classified Instances	282	23.1148 %
Kappa statistic	0.4731	
Mean absolute error	0.2353	
Root mean squared error	0.4478	
Relative absolute error	52.7179 %	
Root relative squared error	94.8079 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.622	0.157	0.668	0.622	0.644	0.474	0.813	0.710	0
1	0.843	0.378	0.815	0.843	0.829	0.474	0.813	0.878	1
Weighted Avg.	0.769	0.304	0.765	0.769	0.767	0.474	0.813	0.821	

== Confusion Matrix ==

	a	b	<-- classified as
255	155		a = 0
127	683		b = 1

## AdaBoostM1 – IBk (k=10) Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.3 seconds

== Summary ==

Correctly Classified Instances	1012	82.9508 %
Incorrectly Classified Instances	208	17.0492 %
Kappa statistic	0.6161	
Mean absolute error	0.2587	
Root mean squared error	0.3619	
Relative absolute error	57.9642 %	
Root relative squared error	76.6122 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.737	0.123	0.751	0.737	0.744	0.616	0.863	0.735	0
1	0.877	0.263	0.868	0.877	0.872	0.616	0.863	0.901	1
Weighted Avg.	0.830	0.216	0.829	0.830	0.829	0.616	0.863	0.845	

== Confusion Matrix ==

a	b	<-- classified as
302	108	a = 0
100	710	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	952	78.0328 %
Incorrectly Classified Instances	268	21.9672 %
Kappa statistic	0.5054	
Mean absolute error	0.3006	
Root mean squared error	0.4192	
Relative absolute error	67.3392 %	
Root relative squared error	88.7492 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.663	0.160	0.677	0.663	0.670	0.505	0.757	0.614	0
1	0.840	0.337	0.831	0.840	0.835	0.505	0.757	0.814	1
Weighted Avg.	0.780	0.277	0.779	0.780	0.780	0.505	0.757	0.746	

== Confusion Matrix ==

a	b	<-- classified as
272	138	a = 0
130	680	b = 1

## MultilayerPerceptron

== Evaluation on training set ==

Time taken to test model on training data: 0 seconds

== Summary ==

Correctly Classified Instances	968	79.3443 %
Incorrectly Classified Instances	252	20.6557 %
Kappa statistic	0.5639	
Mean absolute error	0.2889	
Root mean squared error	0.3753	
Relative absolute error	64.7176 %	
Root relative squared error	79.4443 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.817	0.219	0.654	0.817	0.727	0.573	0.875	0.791	0
1	0.781	0.183	0.894	0.781	0.834	0.573	0.875	0.920	1
Weighted Avg.	0.793	0.195	0.813	0.793	0.798	0.573	0.875	0.876	

== Confusion Matrix ==

a	b	<-- classified as
335	75	a = 0
177	633	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	962	78.8525 %
Incorrectly Classified Instances	258	21.1475 %
Kappa statistic	0.5215	
Mean absolute error	0.2765	
Root mean squared error	0.3837	
Relative absolute error	61.9475 %	
Root relative squared error	81.2396 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.666	0.149	0.693	0.666	0.679	0.522	0.846	0.730	0
1	0.851	0.334	0.834	0.851	0.842	0.522	0.846	0.899	1
Weighted Avg.	0.789	0.272	0.787	0.789	0.787	0.522	0.846	0.842	

== Confusion Matrix ==

a	b	<-- classified as
273	137	a = 0
121	689	b = 1

## Appendix III – Chapter 3 Training Results

### Naïve Bayes Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.02 seconds

== Summary ==

Correctly Classified Instances	972	79.6721 %
Incorrectly Classified Instances	248	20.3279 %
Kappa statistic	0.5439	
Mean absolute error	0.2099	
Root mean squared error	0.4288	
Relative absolute error	47.0321 %	
Root relative squared error	90.772 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.695	0.152	0.699	0.695	0.697	0.544	0.848	0.738	0
1	0.848	0.305	0.846	0.848	0.847	0.544	0.848	0.903	1
Weighted Avg.	0.797	0.253	0.796	0.797	0.797	0.544	0.848	0.848	

== Confusion Matrix ==

a	b	<-- classified as
285	125	a = 0
123	687	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	969	79.4262 %
Incorrectly Classified Instances	251	20.5738 %
Kappa statistic	0.5398	
Mean absolute error	0.211	
Root mean squared error	0.4306	
Relative absolute error	47.2633 %	
Root relative squared error	91.1667 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.698	0.157	0.692	0.698	0.695	0.540	0.845	0.731	0
1	0.843	0.302	0.846	0.843	0.845	0.540	0.845	0.899	1
Weighted Avg.	0.794	0.253	0.795	0.794	0.794	0.540	0.845	0.842	

== Confusion Matrix ==

a	b	<-- classified as
286	124	a = 0
127	683	b = 1

## J48 Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0 seconds

== Summary ==

Correctly Classified Instances	1124	92.1311 %
Incorrectly Classified Instances	96	7.8689 %
Kappa statistic	0.8198	
Mean absolute error	0.1394	
Root mean squared error	0.264	
Relative absolute error	31.2298 %	
Root relative squared error	55.8892 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.839	0.037	0.920	0.839	0.878	0.822	0.933	0.889	0
1	0.963	0.161	0.922	0.963	0.942	0.822	0.933	0.944	1
Weighted Avg.	0.921	0.119	0.921	0.921	0.920	0.822	0.933	0.925	

== Confusion Matrix ==

a	b	<-- classified as
344	66	a = 0
30	780	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	945	77.459 %
Incorrectly Classified Instances	275	22.541 %
Kappa statistic	0.4859	
Mean absolute error	0.2874	
Root mean squared error	0.4265	
Relative absolute error	64.3942 %	
Root relative squared error	90.2893 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.629	0.152	0.677	0.629	0.652	0.487	0.777	0.597	0
1	0.848	0.371	0.819	0.848	0.833	0.487	0.777	0.828	1
Weighted Avg.	0.775	0.297	0.771	0.775	0.772	0.487	0.777	0.750	

== Confusion Matrix ==

a	b	<-- classified as
258	152	a = 0
123	687	b = 1

## IBk (k=10) Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.17 seconds

== Summary ==

Correctly Classified Instances	1000	81.9672 %
Incorrectly Classified Instances	220	18.0328 %
Kappa statistic	0.593	
Mean absolute error	0.2416	
Root mean squared error	0.3485	
Relative absolute error	54.1292 %	
Root relative squared error	73.7827 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.717	0.128	0.739	0.717	0.728	0.593	0.896	0.803	0
1	0.872	0.283	0.859	0.872	0.865	0.593	0.896	0.936	1
Weighted Avg.	0.820	0.231	0.818	0.820	0.819	0.593	0.896	0.891	

== Confusion Matrix ==

a	b	<-- classified as
294	116	a = 0
104	706	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	952	78.0328 %
Incorrectly Classified Instances	268	21.9672 %
Kappa statistic	0.5036	
Mean absolute error	0.2713	
Root mean squared error	0.3861	
Relative absolute error	60.7756 %	
Root relative squared error	81.746 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.656	0.157	0.679	0.656	0.667	0.504	0.840	0.720	0
1	0.843	0.344	0.829	0.843	0.836	0.504	0.840	0.888	1
Weighted Avg.	0.780	0.281	0.779	0.780	0.779	0.504	0.840	0.832	

== Confusion Matrix ==

a	b	<-- classified as
269	141	a = 0
127	683	b = 1

## Bagging – Naïve Bayes Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.16 seconds

== Summary ==

Correctly Classified Instances	969	79.4262 %
Incorrectly Classified Instances	251	20.5738 %
Kappa statistic	0.5359	
Mean absolute error	0.2103	
Root mean squared error	0.4258	
Relative absolute error	47.1059 %	
Root relative squared error	90.1464 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.680	0.148	0.699	0.680	0.690	0.536	0.850	0.743	0
1	0.852	0.320	0.840	0.852	0.846	0.536	0.850	0.904	1
Weighted Avg.	0.794	0.262	0.793	0.794	0.794	0.536	0.850	0.850	

== Confusion Matrix ==

a	b	<-- classified as
279	131	a = 0
120	690	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	968	79.3443 %
Incorrectly Classified Instances	252	20.6557 %
Kappa statistic	0.5399	
Mean absolute error	0.2122	
Root mean squared error	0.4278	
Relative absolute error	47.5378 %	
Root relative squared error	90.5719 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.705	0.162	0.688	0.705	0.696	0.540	0.848	0.742	0
1	0.838	0.295	0.849	0.838	0.843	0.540	0.848	0.900	1
Weighted Avg.	0.793	0.250	0.795	0.793	0.794	0.540	0.848	0.847	

== Confusion Matrix ==

a	b	<-- classified as
289	121	a = 0
131	679	b = 1

## Bagging – J48 Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.01 seconds

== Summary ==

Correctly Classified Instances	1164	95.4098 %
Incorrectly Classified Instances	56	4.5902 %
Kappa statistic	0.896	
Mean absolute error	0.136	
Root mean squared error	0.2111	
Relative absolute error	30.4675 %	
Root relative squared error	44.6955 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.910	0.023	0.952	0.910	0.930	0.897	0.988	0.982	0	
0.977	0.090	0.955	0.977	0.966	0.897	0.988	0.993	1	
Weighted Avg.	0.954	0.068	0.954	0.954	0.954	0.897	0.988	0.989	

== Confusion Matrix ==

a	b	<-- classified as
373	37	a = 0
19	791	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	973	79.7541 %
Incorrectly Classified Instances	247	20.2459 %
Kappa statistic	0.5371	
Mean absolute error	0.2656	
Root mean squared error	0.3794	
Relative absolute error	59.5058 %	
Root relative squared error	80.325 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.659	0.132	0.716	0.659	0.686	0.538	0.847	0.755	0	
0.868	0.341	0.834	0.868	0.851	0.538	0.847	0.896	1	
Weighted Avg.	0.798	0.271	0.794	0.798	0.795	0.538	0.847	0.849	

== Confusion Matrix ==

a	b	<-- classified as
270	140	a = 0
107	703	b = 1

## Bagging – IBk (k=10) Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 1.43 seconds

== Summary ==

Correctly Classified Instances	1000	81.9672 %
Incorrectly Classified Instances	220	18.0328 %
Kappa statistic	0.5823	
Mean absolute error	0.244	
Root mean squared error	0.3469	
Relative absolute error	54.6662 %	
Root relative squared error	73.4327 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.666	0.102	0.767	0.666	0.713	0.585	0.899	0.831	0
1	0.898	0.334	0.841	0.898	0.869	0.585	0.899	0.947	1
Weighted Avg.	0.820	0.256	0.816	0.820	0.816	0.585	0.899	0.908	

== Confusion Matrix ==

a	b	<-- classified as
273	137	a = 0
83	727	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	960	78.6885 %
Incorrectly Classified Instances	260	21.3115 %
Kappa statistic	0.5051	
Mean absolute error	0.2707	
Root mean squared error	0.3826	
Relative absolute error	60.6561 %	
Root relative squared error	80.992 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.612	0.125	0.713	0.612	0.659	0.508	0.846	0.749	0
1	0.875	0.388	0.817	0.875	0.845	0.508	0.846	0.905	1
Weighted Avg.	0.787	0.299	0.782	0.787	0.782	0.508	0.846	0.853	

== Confusion Matrix ==

a	b	<-- classified as
251	159	a = 0
101	709	b = 1

## AdaBoostM1 – Naïve Bayes Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.04 seconds

== Summary ==

Correctly Classified Instances	972	79.6721 %
Incorrectly Classified Instances	248	20.3279 %
Kappa statistic	0.5439	
Mean absolute error	0.2625	
Root mean squared error	0.3863	
Relative absolute error	58.8065 %	
Root relative squared error	81.7744 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.695	0.152	0.699	0.695	0.697	0.544	0.813	0.689	0
1	0.848	0.305	0.846	0.848	0.847	0.544	0.813	0.850	1
Weighted Avg.	0.797	0.253	0.796	0.797	0.797	0.544	0.813	0.796	

== Confusion Matrix ==

a	b	<-- classified as
285	125	a = 0
123	687	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	969	79.4262 %
Incorrectly Classified Instances	251	20.5738 %
Kappa statistic	0.5398	
Mean absolute error	0.2659	
Root mean squared error	0.3906	
Relative absolute error	59.5744 %	
Root relative squared error	82.6956 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.698	0.157	0.692	0.698	0.695	0.540	0.796	0.712	0
1	0.843	0.302	0.846	0.843	0.845	0.540	0.796	0.836	1
Weighted Avg.	0.794	0.253	0.795	0.794	0.794	0.540	0.796	0.794	

== Confusion Matrix ==

a	b	<-- classified as
286	124	a = 0
127	683	b = 1

## AdaBoostM1 – J48 Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.01 seconds

== Summary ==

Correctly Classified Instances	1220	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Relative absolute error	0.0011 %		
Root relative squared error	0.0051 %		
Total Number of Instances	1220		

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

== Confusion Matrix ==

a	b	<-- classified as
410	0	a = 0
0	810	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	932	76.3934 %
Incorrectly Classified Instances	288	23.6066 %
Kappa statistic	0.4613	
Mean absolute error	0.2341	
Root mean squared error	0.4613	
Relative absolute error	52.4588 %	
Root relative squared error	97.657 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.612	0.159	0.661	0.612	0.635	0.462	0.810	0.703	0	
0.841	0.388	0.811	0.841	0.825	0.462	0.810	0.874	1	
Weighted Avg.	0.764	0.311	0.760	0.764	0.762	0.462	0.810	0.817	

== Confusion Matrix ==

a	b	<-- classified as
251	159	a = 0
129	681	b = 1

## AdaBoostM1 – IBk (k=10) Training Model Results

== Evaluation on training set ==

Time taken to test model on training data: 0.28 seconds

== Summary ==

Correctly Classified Instances	1000	81.9672 %
Incorrectly Classified Instances	220	18.0328 %
Kappa statistic	0.593	
Mean absolute error	0.2677	
Root mean squared error	0.3741	
Relative absolute error	59.9729 %	
Root relative squared error	79.1997 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.717	0.128	0.739	0.717	0.728	0.593	0.839	0.694	0
1	0.872	0.283	0.859	0.872	0.865	0.593	0.839	0.873	1
Weighted Avg.	0.820	0.231	0.818	0.820	0.819	0.593	0.839	0.813	

== Confusion Matrix ==

a	b	<-- classified as
294	116	a = 0
104	706	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	952	78.0328 %
Incorrectly Classified Instances	268	21.9672 %
Kappa statistic	0.5036	
Mean absolute error	0.3018	
Root mean squared error	0.4185	
Relative absolute error	67.6095 %	
Root relative squared error	88.597 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.656	0.157	0.679	0.656	0.667	0.504	0.756	0.612	0
1	0.843	0.344	0.829	0.843	0.836	0.504	0.756	0.809	1
Weighted Avg.	0.780	0.281	0.779	0.780	0.779	0.504	0.756	0.743	

== Confusion Matrix ==

a	b	<-- classified as
269	141	a = 0
127	683	b = 1

## MultilayerPerceptron

== Evaluation on training set ==

Time taken to test model on training data: 0 seconds

== Summary ==

Correctly Classified Instances	1022	83.7705 %
Incorrectly Classified Instances	198	16.2295 %
Kappa statistic	0.6449	
Mean absolute error	0.2174	
Root mean squared error	0.3439	
Relative absolute error	48.7127 %	
Root relative squared error	72.7978 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.807	0.147	0.736	0.807	0.770	0.647	0.896	0.855	0
1	0.853	0.193	0.897	0.853	0.875	0.647	0.896	0.923	1
Weighted Avg.	0.838	0.177	0.843	0.838	0.839	0.647	0.896	0.900	

== Confusion Matrix ==

a	b	<-- classified as
331	79	a = 0
119	691	b = 1

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	970	79.5082 %
Incorrectly Classified Instances	250	20.4918 %
Kappa statistic	0.5397	
Mean absolute error	0.2538	
Root mean squared error	0.4054	
Relative absolute error	56.8649 %	
Root relative squared error	85.8264 %	
Total Number of Instances	1220	

== Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0	0.690	0.152	0.697	0.690	0.694	0.540	0.819	0.721	0
1	0.848	0.310	0.844	0.848	0.846	0.540	0.819	0.873	1
Weighted Avg.	0.795	0.257	0.795	0.795	0.795	0.540	0.819	0.822	

== Confusion Matrix ==

a	b	<-- classified as
283	127	a = 0
123	687	b = 1