



American International University-Bangladesh (AIUB)

Submitted by----

Name: Mohiduzzaman, Mohammad

ID: 15-30173-2

Section: A

Subject: Data Warehouse and Data Mining

Project on Supervised Learning (vehicle silhouettes)

Submitted to: TOHEDUL ISLAM

Submission Date: 30/04/2018

Problem:

The reason of this project is to classify a given silhouette as one of four types of vehicle using a set of attributes extracted from the silhouette. The vehicle might be seen from one of a wide range of edges. The first design was to discover a strategy for recognizing 3D questions inside a 2D picture by use of an outfit of shape highlight extractors to the 2D outlines of the articles. Proportions of shape highlights extricated from model outlines of articles to be separated were utilized to produce a grouping rule tree by methods for PC acceptance. This item acknowledgment procedure was effectively used to separate between outlines of model cars, vans and buses saw from compelled rise yet all points of turn. The principle center is to locate the best classifier utilizing WEKA for the given informational index.

Dataset Prepare:

This dataset taken from Turing Institute, Glasgow, Scotland is about the classification of four types of vehicle, using a set of features extracted from the silhouette. In this dataset there are 946 instances and 18 attributes excluding class attribute. Also, there is no missing values in the dataset. I used the attribute “Types of Vehicle” as a decision attribute or class attribute. All the value of class attribute here is numerical. There are 4 types of classes.

And the classes and no. in each class are:

No.	Type	Total No.
1.	OPEL	240
2.	SAAB	240
3.	BUS	240
4.	VAN	226

Dataset Conversion:

There was total 9 parts of that dataset. And all of them were in ‘.dat’ format. At first I have merged them and save them in ‘.csv’ format. Then I converted that ‘.csv’ file into ‘.arff’ formal. Then I worked on that ‘.arff’ file.

Run information:

Scheme: weka.classifiers.bayes.NaiveBayes

Relation: final

Instances: 846

Attributes: 19

COMPACTNESS

CIRCULARITY

DISTANCE CIRCULARITY

RADIUS RATIO

PR.AXIS ASPECT RATIO

MAX.LENGTH ASPECT RATIO

SCATTER RATIO

ELONGATEDNESS

PR.AXIS RECTANGULARITY

MAX.LENGTH RECTANGULARITY

SCALED VARIANCE(Major Axis)

SCALED VARIANCE(Minor Axis)

SCALED RADIUS OF GYRATION

SKEWNESS ABOUT(Major Axis)

SKEWNESS ABOUT(Minor Axis)

KURTOSIS ABOUT(Minor Axis)

KURTOSIS ABOUT(Major Axis)

HOLLOWS RATIO

Class

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

I used the following classifier in WEKA to find the best classifier for this data set:

1.Naive Bayes Classifier:

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	379	44.7991 %
Incorrectly Classified Instances	467	55.2009 %
Kappa statistic	0.2697	
Mean absolute error	0.2826	
Root mean squared error	0.462	
Relative absolute error	75.4027 %	
Root relative squared error	106.7136 %	
Total Number of Instances	846	

=== Detailed Accuracy By Class ===

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
van	0.874	0.411	0.395	0.874	0.545	0.393	0.821	0.529
saab	0.392	0.119	0.531	0.392	0.451	0.304	0.712	0.494
bus	0.147	0.027	0.653	0.147	0.240	0.224	0.843	0.609
opel	0.415	0.172	0.447	0.415	0.430	0.249	0.702	0.449
Weighted Avg.	0.448	0.177	0.510	0.448	0.413	0.291	0.769	0.521

=== Confusion Matrix ===

a b c d <-- classified as

174 7 14 4 | a = van
 67 85 3 62 | b = saab
 135 8 32 43 | c = bus
 64 60 0 88 | d = opel

2.IBK Classifier:

=== Stratified cross-validation ===

=== **Summary** ===

Correctly Classified Instances	591	69.8582 %
Incorrectly Classified Instances	255	30.1418 %
Kappa statistic	0.598	
Mean absolute error	0.1519	
Root mean squared error	0.3872	
Relative absolute error	40.5184 %	
Root relative squared error	89.4389 %	
Total Number of Instances	846	

=== **Detailed Accuracy By Class** ===

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
van	0.894	0.048	0.852	0.894	0.873	0.833	0.927	0.795
saab	0.498	0.170	0.502	0.498	0.500	0.329	0.669	0.391
bus	0.931	0.037	0.898	0.931	0.914	0.884	0.952	0.871
opel	0.481	0.148	0.520	0.481	0.500	0.342	0.666	0.396
Weighted Avg.	0.699	0.101	0.691	0.699	0.694	0.594	0.802	0.611

=== Confusion Matrix ===

a b c d <-- classified as

178 10 6 5 | a = van

13 108 11 85 | b = saab

8 3 203 4 | c = bus

10 94 6 102 | d = opel

3.Kstar Classifier:

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	604	71.3948 %
Incorrectly Classified Instances	242	28.6052 %
Kappa statistic	0.6186	
Mean absolute error	0.1492	
Root mean squared error	0.353	
Relative absolute error	39.8121 %	
Root relative squared error	81.5301 %	
Total Number of Instances	846	

=== Detailed Accuracy By Class ===

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
van	0.955	0.042	0.876	0.955	0.913	0.992	0.973	
saab	0.493	0.173	0.495	0.493	0.494	0.789	0.447	

	0.991	0.016	0.956	0.991	0.973	0.964	0.999	0.998	bus
	0.429	0.151	0.487	0.429	0.456	0.290	0.788	0.463	opel
Weighted Avg.	0.714	0.096	0.701	0.714	0.707	0.612	0.891	0.717	

=== Confusion Matrix ===

a b c d <-- classified as

190 5 0 4 | a = van

13 107 5 92 | b = saab

2 0 216 0 | c = bus

12 104 5 91 | d = opel

4.JRIP Classifier:

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	584	69.0307 %
Incorrectly Classified Instances	262	30.9693 %
Kappa statistic	0.5868	
Mean absolute error	0.1914	
Root mean squared error	0.3323	
Relative absolute error	51.0598 %	
Root relative squared error	76.7524 %	
Total Number of Instances	846	

=== Detailed Accuracy By Class ===

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
van	0.864	0.043	0.860	0.864	0.862	0.820	0.927	0.861
saab	0.484	0.146	0.533	0.484	0.507	0.349	0.793	0.501
bus	0.940	0.099	0.768	0.940	0.845	0.792	0.940	0.812
opel	0.481	0.126	0.560	0.481	0.518	0.374	0.785	0.545
Weighted Avg.	0.690	0.105	0.677	0.690	0.680	0.580	0.860	0.677

=== Confusion Matrix ===

a b c d <-- classified as

172 15 8 4 | a = van

11 105 27 74 | b = saab

5 6 205 2 | c = bus

12 71 27 102 | d = opel

5.OneR Classifier:

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 426 50.3546 %

Incorrectly Classified Instances 420 49.6454 %

Kappa statistic 0.3391

Mean absolute error 0.2482

Root mean squared error 0.4982

Relative absolute error 66.2212 %

Root relative squared error 115.083 %

Total Number of Instances 846

=== Detailed Accuracy By Class ===

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
van	0.709	0.196	0.526	0.709	0.604	0.467	0.756	0.441
saab	0.392	0.173	0.438	0.392	0.414	0.227	0.609	0.328
bus	0.569	0.137	0.590	0.569	0.579	0.437	0.716	0.447
opel	0.358	0.155	0.437	0.358	0.394	0.219	0.602	0.317
Weighted Avg.	0.504	0.165	0.498	0.504	0.496	0.336	0.669	0.383

=== Confusion Matrix ===

```
a  b  c  d  <-- classified as
141 15 35  8 | a = van
46  85 26 60 | b = saab
37  27 124 30 | c = bus
44  67 25 76 | d = opel
```

6.ZeroR Classifier:

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	217	25.6501 %
Incorrectly Classified Instances	629	74.3499 %
Kappa statistic	-0.0014	
Mean absolute error	0.3748	

Root mean squared error	0.4329
Relative absolute error	100 %
Root relative squared error	100 %
Total Number of Instances	846

=== Detailed Accuracy By Class ===

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
van	0.000	0.000	?	0.000	?	?	0.496	0.234
saab	0.097	0.100	0.250	0.097	0.140	-0.005	0.494	0.254
bus	0.899	0.901	0.257	0.899	0.400	-0.003	0.495	0.256
opel	0.000	0.000	?	0.000	?	?	0.493	0.248
Weighted Avg.	0.257	0.258	?	0.257	?	?	0.495	0.248

=== Confusion Matrix ===

```

a  b  c  d  <-- classified as
0 20 179  0 | a = van
0 21 196  0 | b = saab
0 22 196  0 | c = bus
0 21 191  0 | d = opel

```

7.J48 Classifier:

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	614	72.5768 %
--------------------------------	-----	-----------

Incorrectly Classified Instances	232	27.4232 %
Kappa statistic	0.6343	
Mean absolute error	0.1415	
Root mean squared error	0.3355	
Relative absolute error	37.7493 %	
Root relative squared error	77.4887 %	
Total Number of Instances	846	

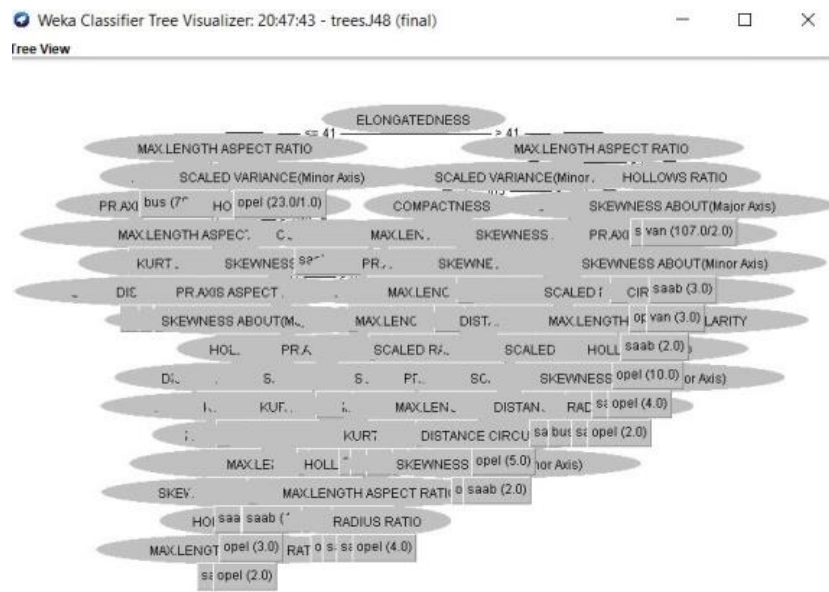
=== Detailed Accuracy By Class ===

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
van	0.894	0.040	0.873	0.894	0.883	0.847	0.932	0.793
saab	0.461	0.141	0.529	0.461	0.493	0.335	0.758	0.473
bus	0.950	0.024	0.932	0.950	0.941	0.920	0.977	0.945
opel	0.608	0.161	0.558	0.608	0.582	0.435	0.784	0.544
Weighted Avg.	0.726	0.092	0.721	0.726	0.723	0.631	0.862	0.687

=== Confusion Matrix ===

a b c d <-- classified as

178	11	6	4	a = van
16	100	6	95	b = saab
4	4	207	3	c = bus
6	74	3	129	d = opel



8.RandomTree Classifier:

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	602	71.1584 %
Incorrectly Classified Instances	244	28.8416 %
Kappa statistic	0.6152	
Mean absolute error	0.1442	
Root mean squared error	0.3797	
Relative absolute error	38.4714 %	
Root relative squared error	87.7165 %	
Total Number of Instances	846	

=== Detailed Accuracy By Class ===

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
van	0.874	0.040	0.870	0.874	0.872	0.833	0.917	0.790
saab	0.576	0.180	0.525	0.576	0.549	0.385	0.698	0.411
bus	0.894	0.029	0.915	0.894	0.905	0.872	0.933	0.846
opel	0.509	0.137	0.554	0.509	0.531	0.383	0.686	0.405
Weighted Avg.	0.712	0.097	0.714	0.712	0.712	0.615	0.807	0.611

=== Confusion Matrix ===

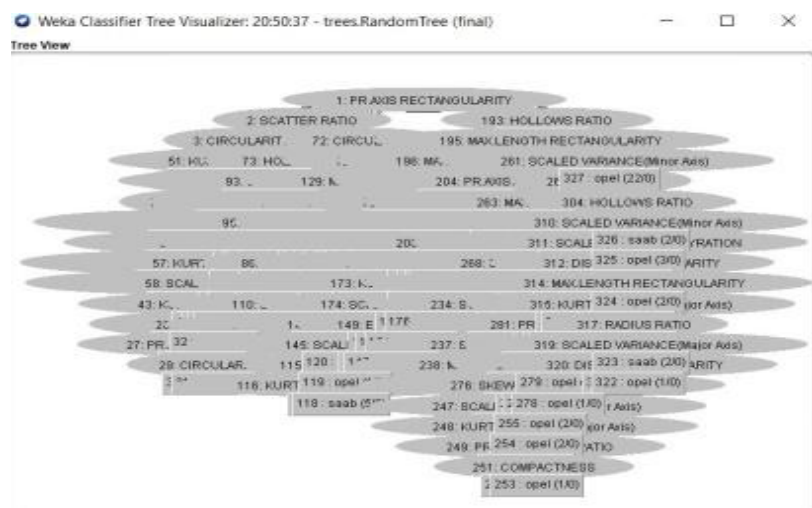
a b c d <-- classified as

174 14 2 9 | a = van

15 125 11 66 | b = saab

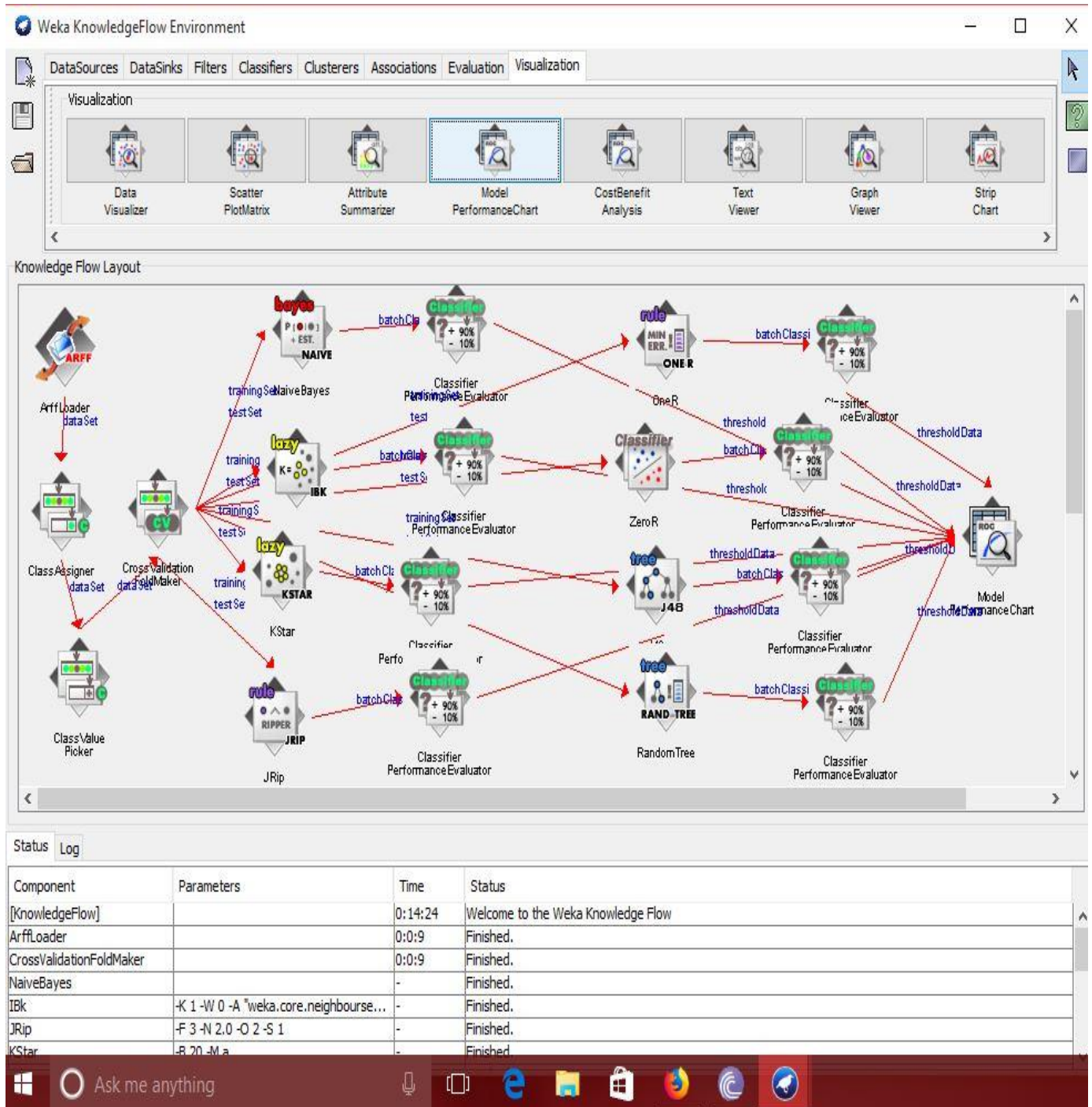
3 8 195 12 | c = bus

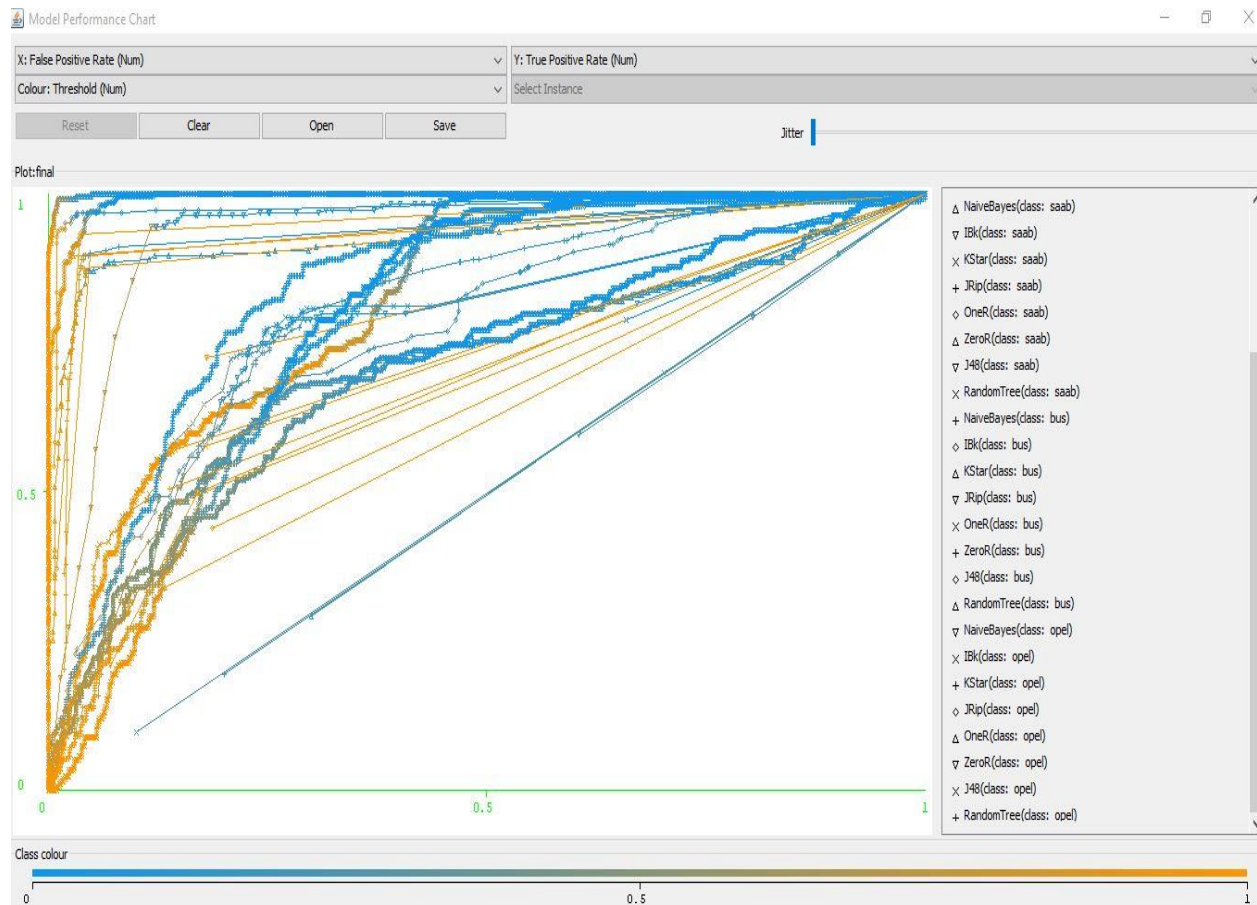
8 91 5 108 | d = opel



ROC Curve:

Here I draw the ROC curve for the eight classifier I used above and for all four classes.





Analysis the Result:

Analyzing all 8 different algorithms for the given specification and data set was tested, it has been seen that **Jrip** is the best classifier for this dataset. According to analysis **Jrip** has the highest correctly Classified Instances Correctly Classified Instances is 614 which is = 72.5768 % and Incorrectly Classified Instances is 232 which is 27.4232 %.

Now 2nd nearest best algorithm Correctly Classified Instances for **Kstar** is 604 which is 71.3948 % and Incorrectly Classified Instances=242 which is 28.6052 %.

Now 3rd nearest best algorithm Correctly Classified Instances for **Random Tree** is 602 which is 71.1584 % and Incorrectly Classified Instances=244 which is 28.8416 %.

ZeroR and **NaiveBayes** classifiers are worst for this dataset. ZeroR correctly classified only 217 instance and NaiveBayes only 379 instances.

So, any of these three classifier (Jrip, Kstar, Random Tree) can be chosen as the best classifier because other of the five classifiers points are comparatively away from best point.

So, for this dataset **Jrip** is the best classifier.