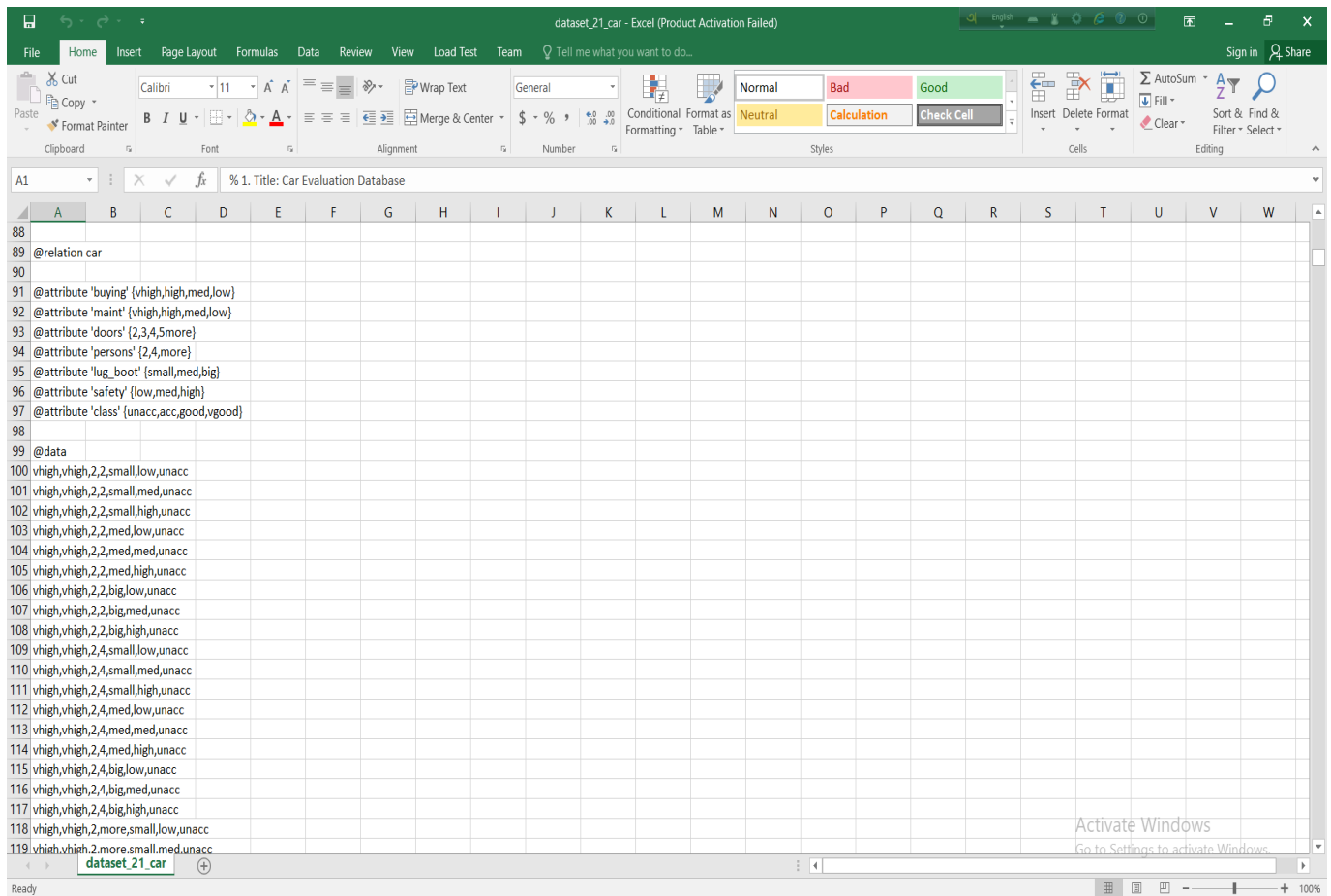


## Car Evaluation dataset:



The Car Evaluation Database contains examples with the structural information removed, i.e., directly relates CAR to the six input attributes: buying, maint, doors, persons, lug\_boot, safety & lastly classification attribute namely Class.

Number of Attributes: 6

<u>Attribute name</u>	<u>Refers to</u>	<u>Attribute values</u>	<u>Data type</u>
1. buying-	buying price of cars	v-high, high, med, low	Nominal
2. maint-	price of the maintenance.	v-high, high, med, low	Nominal
3. doors-	number of doors	2, 3, 4, 5-more	Nominal
4. persons-	capacity in terms of persons to carry	2, 4, more	Nominal

5. lug_boot-	the size of luggage boot	small, med, big	Nominal
6. safety-	estimated safety of the car	low, med, high	Nominal

Class Distribution (number of instances per class)

class	N	N[%]
-------	---	------

-----

**unacc** 1210 (70.023 %) – value of Car’s quality (**unacceptable**) in terms of buying & maintenance price, number of doors, person’s carrying capacity, luggage boot size & safety

**acc** 384 (22.222 %)- value of Car’s quality (**acceptable**) in terms of buying & maintenance price, number of doors, person’s carrying capacity, luggage boot size & safety

**good** 69 ( 3.993 %)- value of Car’s quality (**good**) in terms of buying & maintenance price, number of doors, person’s carrying capacity, luggage boot size & safety

**v-good** 65 ( 3.762 %)- value of Car’s quality (**very good**) in terms of buying & maintenance price, number of doors, person’s carrying capacity, luggage boot size & safety

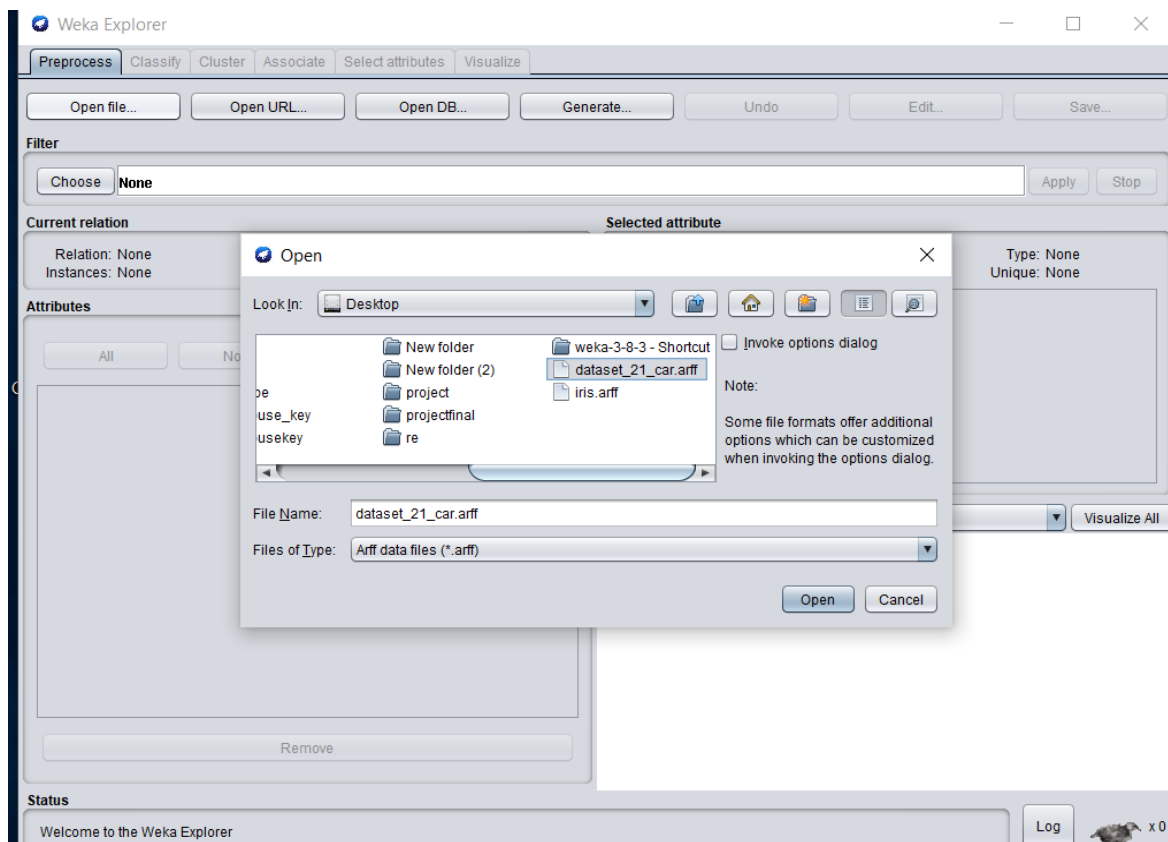
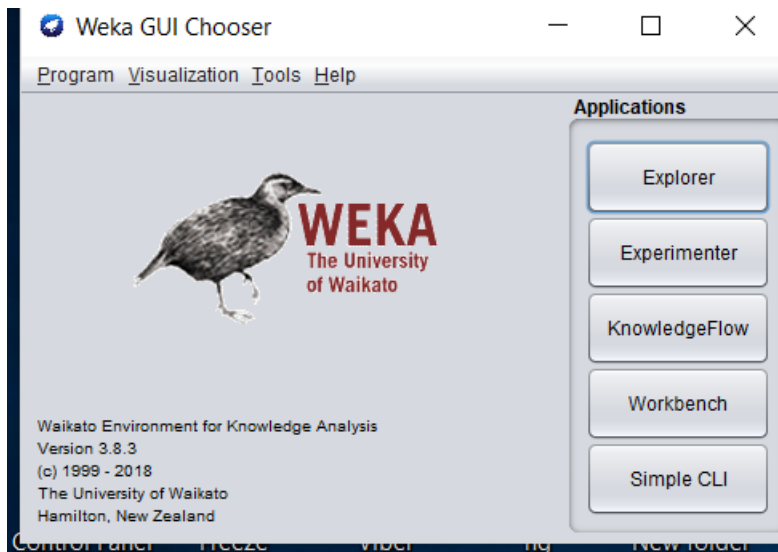
Missing Attribute Values: none

Number of Instances: 1728

Here each instances with values of 6 attribute falls into a specific categorical classification: unacceptable, acceptable, good or very good

After downloading the car evaluation dataset from uci repository, dataset has been pre-processed to be loaded into weka. The dataset was saved into arff format.

After opening weka in explore mode, open file button was clicked to load the dataset in weka. From there, dataset was chosen for further application.



After opening the selected dataset , the window looked like this:-

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter

Choose None Apply Stop

Current relation

Relation: car Instances: 1728 Attributes: 7 Sum of weights: 1728

Attributes

All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> buying
2	<input type="checkbox"/> maint
3	<input type="checkbox"/> doors
4	<input type="checkbox"/> persons
5	<input type="checkbox"/> lug_boot
6	<input type="checkbox"/> safety
7	<input type="checkbox"/> class

Remove

Selected attribute

Name: buying Missing: 0 (0%) Distinct: 4 Type: Nominal Unique: 0 (0%)

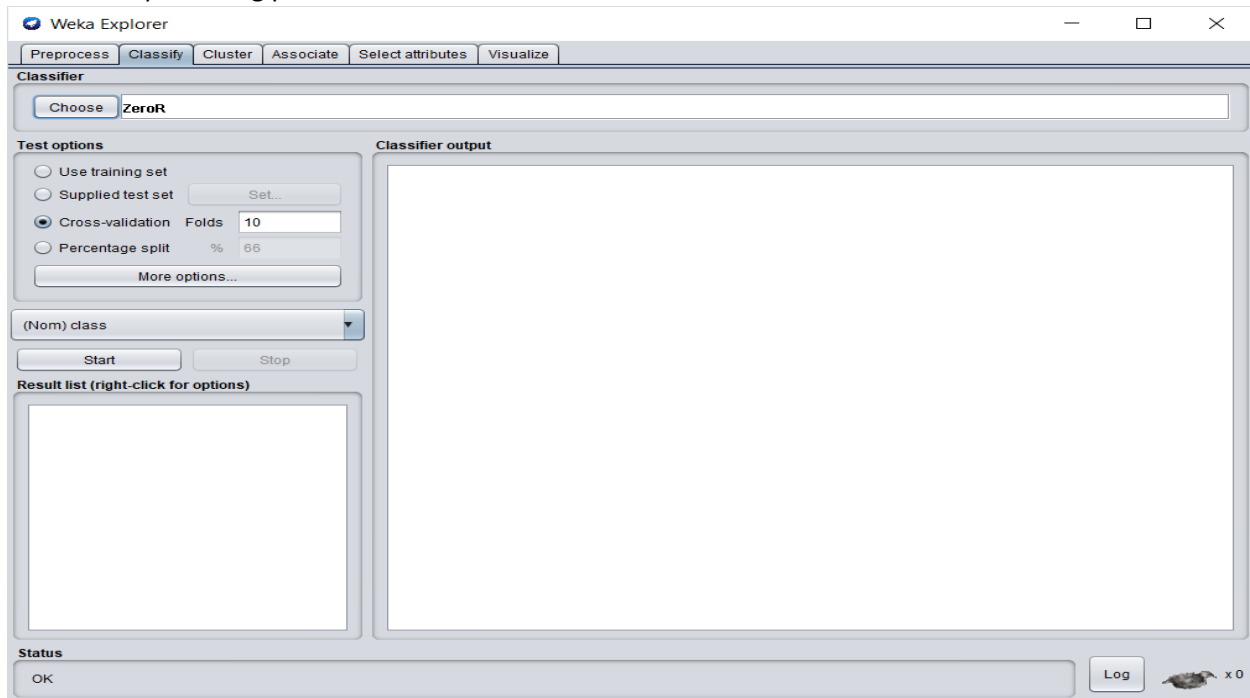
No.	Label	Count	Weight
1	vhigh	432	432.0
2	high	432	432.0
3	med	432	432.0
4	low	432	432.0

Class: class (Nom) Visualize All

Status

OK Log x 0

Then classify button was clicked to choose the classification algorithm that would be applied on the dataset for predicting performance measure



Naïve bayes classification under bayes folder was chosen as the first classification algorithm & the output was like this:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

19:15:16 - bayes NaiveBayes

Classifier output

more	323.0	187.0	34.0	36.0
[total]	1213.0	387.0	72.0	68.0
lug_boot				
small	451.0	106.0	22.0	1.0
med	393.0	136.0	25.0	26.0
big	369.0	145.0	25.0	41.0
[total]	1213.0	387.0	72.0	68.0
safety				
low	577.0	1.0	1.0	1.0
med	358.0	181.0	40.0	1.0
high	278.0	205.0	31.0	66.0
[total]	1213.0	387.0	72.0	68.0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1478	85.5324 %
Incorrectly Classified Instances	250	14.4676 %
Kappa statistic	0.6665	
Mean absolute error	0.1137	
Root mean squared error	0.2262	
Relative absolute error	49.6626 %	
Root relative squared error	66.9048 %	
Total Number of Instances	1728	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.960	0.203	0.917	0.960	0.938	0.783	0.982	0.993	unacc
	0.706	0.098	0.672	0.706	0.689	0.597	0.950	0.842	acc
	0.275	0.007	0.633	0.275	0.384	0.403	0.980	0.538	good
	0.415	0.001	0.931	0.415	0.574	0.613	0.998	0.953	vgood
Weighted Avg.	0.855	0.164	0.852	0.855	0.847	0.720	0.976	0.940	

=== Confusion Matrix ===

a	b	c	d	<-- classified as
1161	48	1	0	a = unacc
104	271	9	0	b = acc
1	47	19	2	c = good
0	37	1	27	d = vgood

As second classification algorithm, j48 tree algorithm was chosen. Outcome of j48.tree is:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

- 19:15:16 - bayes.NaiveBayes
- 19:32:56 - trees.RandomForest
- 19:33:18 - bayes.NaiveBayes
- 19:33:52 - trees.RandomForest
- 19:41:24 - trees.J48

Classifier output

```
| | | | lug_boot = small: acc (4.0/1.0)
| | | | lug_boot = med: vgood (4.0/1.0)
| | | | lug_boot = big: vgood (4.0)
| | | | maint = med
| | | | lug_boot = small: good (4.0/1.0)
| | | | lug_boot = med: vgood (4.0/1.0)
| | | | lug_boot = big: vgood (4.0)
| | | | maint = low
| | | | lug_boot = small: good (4.0/1.0)
| | | | lug_boot = med: vgood (4.0/1.0)
| | | | lug_boot = big: vgood (4.0)
```

Number of Leaves : 131

Size of the tree : 182

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1596	92.3611 %
Incorrectly Classified Instances	132	7.6389 %
Kappa statistic	0.8343	
Mean absolute error	0.0421	
Root mean squared error	0.1718	
Relative absolute error	18.3833 %	
Root relative squared error	50.8176 %	
Total Number of Instances	1728	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.962	0.064	0.972	0.962	0.967	0.892	0.983	0.992	unacc
	0.867	0.047	0.841	0.867	0.854	0.811	0.962	0.859	acc
	0.609	0.011	0.689	0.609	0.646	0.634	0.918	0.593	good
	0.877	0.010	0.770	0.877	0.820	0.814	0.995	0.808	vgood
Weighted Avg.	0.924	0.056	0.924	0.924	0.924	0.861	0.976	0.940	

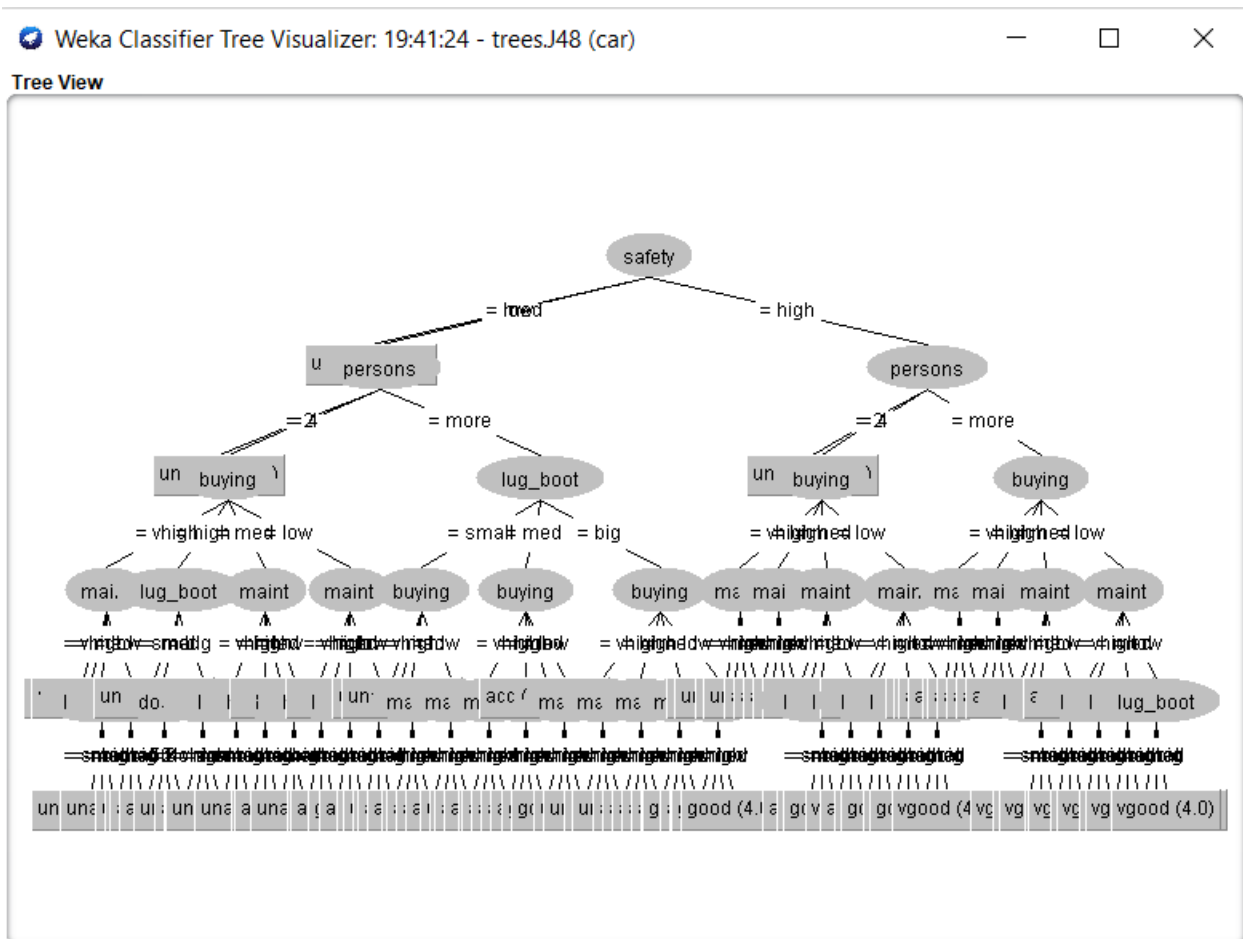
=== Confusion Matrix ===

	a	b	c	d	<-- classified as
1164	43	3	0	0	a = unacc
33	333	11	7	0	b = acc
0	17	42	10	0	c = good
0	3	5	57	0	d = vgood

Status

OK

As j48 is a decision tree algorithm, visualization of the decision tree is given below:





And as 3<sup>rd</sup> classification logistic function algorithm was chosen. The outcome after running the algorithm is:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose Logistic -R 1.0E-8 -M -1 -num-decimal-places 4

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

- 19:15:16 - bayes.NaiveBayes
- 19:32:56 - trees.RandomForest
- 19:33:18 - bayes.NaiveBayes
- 19:33:52 - trees.RandomForest
- 19:41:24 - trees.J48
- 19:47:59 - functions.Logistic

Classifier output

maint=med	0	0	2.2487
maint=low	0	0	2.4306
doors=2	709.9863	129.676	11.7163
doors=3	1.5511	1.8034	1.9388
doors=4	0.0301	0.0654	0.2098
doors=5more	0.0301	0.0654	0.2098
persons=2	9.5427725863447504E16	13945891.2554	4928.8553
persons=4	0	0.0005	0.0275
persons=more	0	0.0001	0.0074
lug_boot=small	4152369581.0901	357650150.4906	1845615.887
lug_boot=med	0.0008	0.0013	0.0039
lug_boot=big	0	0	0.0001
safety=low	5.864009034157339E10	133.0258	0.0071
safety=med	5826.5009	27374929.8414	69505821.253
safety=high	0	0	0

Time taken to build model: 0.47 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1609	93.1134 %
Incorrectly Classified Instances	119	6.8866 %
Kappa statistic	0.8504	
Mean absolute error	0.0428	
Root mean squared error	0.152	
Relative absolute error	18.6918 %	
Root relative squared error	44.9452 %	
Total Number of Instances	1728	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.960	0.071	0.969	0.960	0.965	0.884	0.991	0.997	unacc
	0.872	0.045	0.846	0.872	0.859	0.818	0.984	0.944	acc
	0.783	0.007	0.818	0.783	0.800	0.792	0.995	0.875	good
	0.892	0.005	0.866	0.892	0.879	0.874	0.999	0.963	vgood
Weighted Avg.	0.931	0.061	0.932	0.931	0.931	0.865	0.990	0.979	

=== Confusion Matrix ===

a	b	c	d	<-- classified as
1162	47	1	0	a = unacc
37	335	8	4	b = acc
0	10	54	5	c = good
0	4	3	58	d = vgood

After observing and analyzing the outcome of the above three classifier based on the measurement of the performance of classifier, it can be said that the best classifier among these three algorithm is logistic function.

True Positive Rate & False Positive Rate value of a classifier can define the performance of a classifier which are both proportions from 0 to 1 Inclusive

In first classifier( naïve bayes) TP Rate of unacc was 0.960 & FP Rate was 0.203. For acc TP Rate: 0.706 & FP Rate : 0.098, for good TP Rate : 0.275 FP Rate : 0.007 , for vgood TP Rate : 0.415 & FP Rate: 0.001.

Weighted Average TP Rate : 0.855 & weighted average FP Rate :0.164

WEIGHTED AVERAGE Precision : 0.852 ( proportion of instances classified as positive that are really positive)

It also can be seen that among 1728 instances 1478 instances were correctly classified- 85.53% ,  
incorrectly classified 250 instances – 14.47%

In 2nd classifier( trees.j48) TP Rate of unacc was 0.962 & FP Rate was 0.064. For acc TP Rate: 0.867 & FP Rate : 0.047, for good TP Rate : 0.609 FP Rate : 0.011 , for vgood TP Rate : 0.877 & FP Rate: 0.010.

Weighted Average TP Rate : 0.924 & Weighted average FP Rate :0.056

WEIGHTED AVERAGE Precision : 0.924 ( proportion of instances classified as positive that are really positive)

It also can be seen that among 1728 instances 1596 instances were correctly classified- 92.36% ,  
incorrectly classified 132 instances – 7.64%

In 3rd classifier( function.logistic) TP Rate of unacc was 0.960 & FP Rate was 0.071. For acc TP Rate: 0.872 & FP Rate : 0.045, for good TP Rate : 0.783 FP Rate : 0.007 , for vgood TP Rate : 0.892 & FP Rate: 0.005.

Weighted Average TP Rate : 0.931 & weighted average FP Rate :0.061

WEIGHTED AVERAGE Precision : 0.932 ( proportion of instances classified as positive that are really positive)

It also can be seen that among 1728 instances 1609 instances were correctly classified- 93.11% ,  
incorrectly classified 119 instances – 6.89%

Comparing among these three classifier :

	<u>Naïve Byes</u>	<u>J48</u>	<u>Logistic</u>
Weight. Avg TP Rate :	0.855	0.924	0.931
Weight. Avg FP Rate	0.164	0.056	0.061
Weight. Avg Precision	0.852	0.924	0.932

Correctly classified

Instances	1478(85.53%)	1596(92.36%)	1609(93.11%)
-----------	--------------	--------------	--------------

Incorrectly classified

Instances	250(14.47%)	132(7.64%)	119(6.89%)
-----------	-------------	------------	------------

As logistic function has the highest value for weighted average TP RATE , highest value for weighted average PRECISION, highest percentage for correctly classified instance number, lowest percentage for incorrectly classified instance number & comparatively low value for weighted average FP RATE, hence here among these three classifier **logistic function** is the **best classifier**.