

## Section 1: Data Set

The Car Evaluation Database data set. I found the Car Evaluation Database data set from [UCI Machine Learning Repo](#). Which was created by Marko Bohanec. The data set has not had any pre-processing done to it. Cars are evaluated according to the buying price, price of maintenance, number of doors, capacity in terms of persons to carry, the size of the luggage boot/trunk, and the estimated safety of the car. Cars are then classified as unacceptable, acceptable, good, and very good.

The buying prices can be set as vhigh, high, med, and low. Price of maintenance ranges from vhigh, high, med, low. Number of doors on the vehicle can be set to 2, 3, 4, 5more. Person seating capacity can be set as 2, 4, more. Trunk size can be set to small, med, and big. While safety can hold the values low, med, and high. The system is trying to decide whether a car is acceptable according to these attributes compared to some standard that is given to the system.

## Section 2: Baseline Classifier

For the baseline classifier, ZeroR was used to classify the Car Evaluation Database data set. ZeroR is called the majority classifier. It focuses on the classification that is derived the most through the data set. With this baseline we are able to tell which classification should be observed the most out of all labels.

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

Correctly Classified Instances      843          69.6694 %
Incorrectly Classified Instances   367          30.3306 %
Kappa statistic                    0
Mean absolute error                0.2322
Root mean squared error            0.3404
Relative absolute error            100 %
Root relative squared error        100 %
Total Number of Instances         1210

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

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	1.000	0.697	1.000	0.821	?	0.496	0.695	unacc
	0.000	0.000	?	0.000	?	?	0.494	0.216	acc
	0.000	0.000	?	0.000	?	?	0.484	0.042	good
	0.000	0.000	?	0.000	?	?	0.491	0.041	vgood
Weighted Avg.	0.697	0.697	?	0.697	?	?	0.495	0.535	

```
=== Confusion Matrix ===

 a  b  c  d  <-- classified as
843  0  0  0 | a = unacc
264  0  0  0 | b = acc
 52  0  0  0 | c = good
 51  0  0  0 | d = vgood
```

Figure 1. ZeroR Summary

As can be seen in the report above, the label unacceptable holds the majority of the classifications with approximately 70% accuracy. There were 367 incorrectly classified instances given in the report.

### Section 3: Intelligent Classifier

Using J48, an extension of the Iterative Dichotomiser 3 algorithm, which is used to generate decisions trees. J48 can be thought of as a statistical classifier. J48's decision tree generating capability makes it a perfect algorithm to use with the decision problem given with the Car Evaluation Database data set.

```

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

Correctly Classified Instances      1084           89.5868 %
Incorrectly Classified Instances    126           10.4132 %
Kappa statistic                     0.776
Mean absolute error                  0.0583
Root mean squared error             0.1926
Relative absolute error              25.119 %
Root relative squared error          56.584 %
Total Number of Instances          1210

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.960	0.076	0.967	0.960	0.963	0.879	0.978	0.989	unacc
	0.818	0.063	0.783	0.818	0.800	0.743	0.960	0.848	acc
	0.365	0.014	0.543	0.365	0.437	0.425	0.936	0.546	good
	0.784	0.019	0.645	0.784	0.708	0.697	0.973	0.767	vgood
Weighted Avg.	0.896	0.068	0.895	0.896	0.894	0.822	0.972	0.930	

```

=== Confusion Matrix ===

```

a	b	c	d	<-- classified as
809	31	2	1	a = unacc
28	216	8	12	b = acc
0	24	19	9	c = good
0	5	6	40	d = vgood

Figure 2. J48 Summary

The report above displays J48 algorithm's results. J48 correctly classified 1084 instances out of a total of 1210, giving an accuracy rate of approximately 90%. Through figures 1 and 2 the data shows the correlation between the baseline's majority labeling of unacceptable. In ZeroR's report we were told that unacceptable was classified the majority of the instances. J48 backs this hypothesis and can be observed through the "Detailed Accuracy By Class" table which tells us unacceptable was classified with a 96.7% accuracy and the "Confusion Matrix" provides us with the number of instances that were labeled unacceptable, which does indeed have the highest number out of all instances labeled.

## Work Cited

[Bohanec, Marko. \*UCI Machine Learning Repository: Car Evaluation Data Set\*, 1997, archive.ics.uci.edu/ml/datasets/Car+Evaluation.](https://archive.ics.uci.edu/ml/datasets/Car+Evaluation)