Patrick Griffin

Machine Learning Report

**Section 1: Data Set**

The Car Evaluation Database data set. I found the Car Evaluation Database data set from [UCI Machine Learning Repo](http://archive.ics.uci.edu/ml/datasets/Car+Evaluation). 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.

A picture containing text, receipt, screenshot

Description automatically generated

*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.

A picture containing text, receipt, screenshot

Description automatically generated

*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.](http://archive.ics.uci.edu/ml/datasets/Car+Evaluation)