Car Evaluation Dataset - CM3111

Muhammed Umair Iqbal

December 12, 2017

1 - Data Exploration

1.1 Dataset Choice

I have selected a data set that is related to cars as I have a passion for cars and thought that it would be a good exploring a data set related to cars via R Studio. I got my dataset from the UCI Machine Learning Repository from "http://archive.ics.uci.edu/ml/datasets/Car+Evaluation".

1.2 - Technology Platform

The application I have used to explore my data is R Studio and carry out experiments on. After researching hadoop i have found that i do not need to be using it. For many reasons, one of them being that Hadoop can store massive amounts of data, Petabytes to be exact. Moreover Hadoop should be used when the users requires thousands of operations per second on terabytes of data at a time, which is not what is required in this instance which is why i haven't incorporated Hadoop or any other Big Data Technologies.

1.3 Problem Statement Data Exploration

```
set.seed(1121)
(x=rnorm(20))
```

```
##  [1]  0.1449583  0.4383221  0.1531912  1.0849426  1.9995449 -0.8118832
##  [7]  0.1602680  0.5858923  0.3600880 -0.0253084  0.1508809  0.1100824
## [13]  1.3596812 -0.3269946 -0.7163819  1.8097690  0.5084011 -0.5274603
## [19]  0.1327188 -0.1559430
```

```
mean(x);var(x)
```

```
## [1] 0.3217385
## [1] 0.5714534
```

The dataset I have choosen is a Car Evaluation dataset. It has 6 Attributes; Price, Maint(Maintence of the car), Doors(Number of Doors), Persons(Number of people the car can seat at once),

Boot(The size of the boot), Safety(The safety rating of the car).The aim is to build a predictive model, to predict what the Evaluation is of the car.