**Project Proposal: Auto Selector**

## Project Details

Project Title : Auto Selector  
Team Name : Team Torque  
Team Information : Hsiaoan Wang, Priyanka Annapureddy, Rani Sebastian   
Web page :   
<https://github.com/rani009/Team-Torque>

## Project Description

Shopping for a vehicle is quite an overwhelming task because of the sheer number of options we are surrounded with. Comparing fuel consumption is one of the quicker ways by which we can narrow down the choices based on our needs. This will be interesting and beneficial to not just a first-time car buyer, but also to anyone who is planning an upgrade. The project will also help one understand how the other attributes also play a role in determining fuel consumption of a model.

## Project Goals

Our goal is to come up with the best model that will help one determine the best pick among the various vehicles by comparing the various attributes to see which brand offers a better mpg value.

## Data Set

The dataset we will be using is the publicly available [Auto-mpg dataset](https://www.kaggle.com/uciml/autompg-dataset) which is originally maintained at the StatLib library at Carnegie Mellon University.

UCI link - <https://archive.ics.uci.edu/ml/datasets/auto+mpg>

The dataset has a total of 398 instances with 9 attributes, namely – MPG (miles per gallon), Cylinders, Displacement, Horsepower, Weight, Acceleration, Model Year, Origin and Car Name. The attributes are either continuous, multi-valued discrete or string. There are some missing values in the dataset that will require a clean-up.

## Tools

We plan to use Python for all stages of the project.

## Literature Review

Narrowing down your choice of vehicle based on your needs is no small feat and might take someone with minimal technical knowledge (automobiles related) weeks. The attributes in this auto-mpg dataset are simple and structured enough for novice students to apply and explore the concepts of machine learning.

While researching on the topic, we came across multiple resources which refer to the problem stated. Is fuel consumption that simple to predict given a set of attributes?

The ‘K-means clustering’ paper published by Prof A, John states that “*there are too many variables that separately affect the fuel economy of a vehicle*” [1] This supports our initial goal to prove that there is a significant enough relationship between mpg and other attributes, and that based on one’s needs (single attributes or a combination) we can narrow down the models which offers better mpg.

Since the dataset is a combination of discrete and continuous valued attributes, we plan to try various methods to reduce the error rate and get better predictions. This paper published by J. R Quinlan shows that a combination of instance and model-based learning offers better results. [2]

“Application of Neuro-Fuzzy method for prediction of vehicle fuel consumption” paper offers an informative read with a different methodology. The paper states *“predicting motor vehicle fuel consumption has become a strategic issue, because it is not only related to the issue of availability of fuel but also the problem of the environmental impact caused.*” [3] Even though the methods used are different from what we are going to do in this project – the lessons to take away remain the same, that mpg has strong ties with the other attributes and our model will be able to give us the best options based on the performance needs.

References:

[1] [K- Means Clustering Exercise – Prof Dr. John Aleshunas](http://mercury.webster.edu/aleshunas/Support%20Materials/K-Means/Newton-dominic%20newton%20MATH%203210%2001%20Data%20Mining%20Foundations%20Report%205%20%2828%20nov%2016%29%20COURSE%20PROJECT%20%28Autosaved%29.pdf)

[2] [Combining Instance-based and model-based learning- J.R. Quinlan](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.34.6358&rep=rep1&type=pdf) –



[3] [Application of Neuro-fuzzy method for prediction of vehicle fuel consumption - Ramadoni Syahputra](http://www.jatit.org/volumes/Vol86No1/15Vol86No1.pdf)