**Project proposal – Computational data analysis**

**Linear regression – Auto MPG**

**Problem Statement**:

To build a model fitting on Auto Miles per gallon dataset to predict mpg variable using predictor variables of dataset.

**Dataset**:

Revised from CMU StatLib library, data concerns city-cycle fuel consumption. Auto MPG dataset is a slightly modified version of the dataset provided in the StatLib library. In line with the use by Ross Quinlan (1993) in predicting the attribute "mpg”. "The data concerns city-cycle fuel consumption in miles per gallon, to be predicted in terms of 3 multivalued discrete and 5 continuous attributes." (Quinlan, 1993)

**Features information**:

1. mpg: continuous

2. cylinders: multi-valued discrete

3. displacement: continuous

4. horsepower: continuous

5. weight: continuous

6. acceleration: continuous

7. model year: multi-valued discrete

8. origin: multi-valued discrete

9. car name: string (unique for each instance)

It is publicly available at:

**Dataset**: <https://archive.ics.uci.edu/ml/datasets/Auto+MPG>

**Proposal**:

The proposal is to fit linear regression model on Auto MPG dataset and will explore the dataset and the relation among each independent variable with dependent variable MPG. Will clean the data if required and will fit the linear regression on this dataset later will try to generalize and fit the regularization techniques like lasso or ridge to improve the performance of the predicted model.