

Homework #5: Regressions and Predicting Automobile MPG

In this assignment, you have **two** tasks to complete using the provided *auto-mpg* dataset:

- 1) Use your ML / DM tool of choice (e.g., KNIME) to experiment with various **regression** classifiers on the *training* dataset (i.e., *auto-mpg-train.data*). Specifically, you must train a classifier to **predict** the **MPG** of a car¹, given the attributes. Submit a short description (with screenshots) explaining your experimentation. E.g.,
 - What *algorithms* and parameters did you try?
 - Did you use **K-fold cross-validation**?
 - What were the results (e.g., R^2 , mean squared error)?
- 2) Once you determine the “best” algorithm and parameters, implement your workflow in **Python** to predict the *test* dataset (i.e., *auto-mpg-test.data*). Submit your Python code.
 - You can *hard-code* the input file names
 - You must write the **MPG predictions** to a file named **<lastname>.csv**
 - In the given *test* dataset, the MPG attribute has been replaced with an ID
 - i. Please include this ID number in your output

EXAMPLES (task 2 only)

```
//E.g., Datsun 310 gets 37.2 MPG
UNIX> head -n 3 auto-mpg-train.data
37.2 4. 86.00 65.00 2019. 16.4 80. 3. "datsun 310"
34.2 4. 105.0 70.00 2200. 13.2 79. 1. "plymouth horizon"
37.7 4. 89.00 62.00 2050. 17.3 81. 3. "toyota tercel"
```

```
UNIX> head -n 3 auto-mpg-test.data
1 4. 98.00 65.00 2045. 16.2 81. 1. "ford escort 4w"
2 4. 140.0 90.00 2264. 15.5 71. 1. "chevrolet vega 2300"
3 4. 121.0 115.0 2671. 13.5 75. 2. "saab 99le"
```

```
//run the regression workflow..
UNIX> python reg_workflow.py
```

```
//examine the 1st prediction (e.g., ford escort 4w gets 30.3 mpg)
UNIX> head -n 1 rubin.csv
1.0 30.3
```

¹ Note that **MPG** is the *first* column in the provided training dataset: *auto-mpg-train.data*