

# Hausarbeit im Modul „Data Science und Machine Learning“ WS 23/24: Assignment 2

## The Car Price Data

You have been hired by a US startup that helps consumers to predict the price of their used car. As a data scientist, your job is to build a machine learning model that predicts the price consumers can get for their used cars.

The following attributes are given:

<b>Car_ID</b>	<b>Unique id of each observation</b>
<b>Symboling</b>	<b>Its assigned insurance risk rating, A value of +3 indicates that the auto is risky, -3 that it is probably pretty safe.</b>
<b>CarName</b>	<b>Name of car company</b>
<b>fueltype</b>	<b>Car fuel type i.e gas or diesel</b>
<b>aspiration</b>	<b>Aspiration used in a car</b>
<b>doornumber</b>	<b>Number of doors in a car</b>
<b>carbody</b>	<b>Body of car</b>
<b>drivewheel</b>	<b>Type of drive wheel</b>
<b>engineloation</b>	<b>Location of car engine</b>
<b>wheelbase</b>	<b>Weelbase of car</b>
<b>carlength</b>	<b>Length of car</b>
<b>carwidth</b>	<b>Width of car</b>
<b>carheight</b>	<b>Height of car</b>
<b>curbweight</b>	<b>Weight of a car without occupants or baggage.</b>
<b>enginetype</b>	<b>Type of engine.</b>
<b>cylindernumber</b>	<b>cylinder placed in the car</b>
<b>enginesize</b>	<b>Size of car</b>
<b>fuelsystem</b>	<b>Fuel system of car</b>

<b>boreratio</b>	<b>Boreratio of car</b>
<b>stroke</b>	<b>Stroke or volume inside the engine</b>
<b>compressionratio</b>	<b>compression ratio of car</b>
<b>horsepower</b>	<b>Horsepower</b>
<b>peakrpm</b>	<b>car peak rpm</b>
<b>citympg</b>	<b>Mileage in city</b>
<b>highwaympg</b>	<b>Mileage on highway</b>
<b>price(Dependent variable)</b>	<b>Price of car</b>

### Tasks:

1. Train and evaluate a multiple regression model. Describe your approach and interpret the performance.
  - Note:
    - No regularization (lecture 3) is expected.
    - No hyperparameter tuning (lecture 04) is expected.
2. Train and evaluate at least two polynomial regression models with different polynomial degrees. Describe your approach and interpret the performance.
  - Note.
    - No regularization (lecture 3) is expected.
    - No hyperparameter tuning (lecture 04) is expected.
3. Compare the results to the chosen machine learning models from task 1 and 2. Which machine learning model would you select?
4. Focus on the machine learning model that you have selected in task 3. Now optimizing your results by using regularization (lecture 3) and hyperparameter tuning (lecture 04).
  - Note:
    - Use cross-validation (lecture 04) when evaluating your results.
    - Explain if your model is over or underfitted (lecture 3 and 4).
    - Explain your results and interpret them.
5. Train and evaluate a regression tree. Optimize your results by using hyperparameter tuning (lecture 04).
  - Note:
    - Use cross-validation (lecture 04) when evaluating your results.
    - Explain if your model is over or underfitted (lecture 3 and 4).
    - Explain your results and interpret them.
6. Compare the results to the chosen machine learning models from task 4 and 5. Which machine learning model would you select?
7. Apply an ensemble learning technique (lecture 05) that provides explainable results. Analyze if this technique leads to a better performance than in the previous models that you have selected in task 6.

- Note:
  - Use hyperparameter tuning (lecture 04)
  - Use cross-validation (lecture 04) when evaluating your results.
  - Explain if your model is over or underfitted (lecture 3 and 4).
  - Explain your results and interpret them.