Project 3
Used vehicles price prediction model

In this project you will explore the provided vehicle prices dataset, verify data validity, perform model selection, create a **regression** model to predict the **price** and analyze its performance.

Objectives

During this project you will have the opportunity to apply all the accumulated machine learning knowledge from our course and build a price prediction model. As a ML consultant, you must back your choices with data and are expected to provide an analysis of your model's performance and recommendations.

Dataset

This dataset is a compilation of several files that contain information scraped from multiple websites in the UK. The entire dataset contains 108,540 observations and 10 features.

Dataset Columns

`model:` Model of the car.
`year:` Year the car was built.
`price:` Price of sale (**Target variable**).
`transmission:` Type of transmission of the vehicle.
`mileage:` Current mileage of the vehicle.
`fuelType:` Fuel type of the vehicle.
`tax:` Current tax value of the vehicle.
`mpg:` Motor efficiency in miles per gallon.
`engineSize:` Size of the engine.
`make:` Vehicle manufacturer.

This is not a curated dataset so watch out for missing values, outliers and duplicated observations during your exploratory analysis.

Project requirements

Your project is required to follow these guidelines:

1. Deliver your code within a jupyter notebook.
2. Implement a classification model using only one of the algorithms presented during the course.
3. Provide evidence of your algorithm selection process and data-backed decisions. Include commented code of your model selection process.
4. Keep 30 percent of the observations for testing purposes (Shuffle the data).
5. Use Mean Absolute Error (MAE) to evaluate your model's performance.
6. Discuss your results, performance metrics and implications of your data pre-processing.
7. Discuss the pros and cons of your model.
8. Clearly state your conclusions and recommendations based on your results.

Expected delivery

You are expected to deliver your project before May 18 1:00 PM. in a .zip file named as **name_lastname.zip**:
1. A .pdf report with the sections described below.
2. Your exploratory analysis, model selection, model training and model evaluation code in a jupyter notebook file.
3. Presentation slides.

Grading

Code		20	
	- High-level overview of dataset done - Initial data visualizations performed - Cleaned dataset used for further analysis - Sections and comments that make it easy to understand what's being done. - Algorithm selection reasons and interpretations are presented - Hyper parameter tuning for best performance done		
Report		55	
	Introduction: - Summarize the purpose of the report and summarize the data / subject. - Include important contextual information about the reason for the report. - Summarize your analysis questions, your conclusions, and briefly outline the report.		5
	Data: - Include written descriptions of dataset(s). - Analyze any bias and specific attention to features.		5
	Methods: - Explain how you gathered and analyzed data. - Explain the algorithms you used.(Analysis, preprocessing, models) - Explain your training and parameter selection process.(Description and reasoning)		15
	Results:		20

	- Describe the results of your analysis. - Provide all relevant metrics to evaluate the performance of your model. - Present metrics in an easy-to-understand format.		
	Conclusion: - Restate the questions from your introduction. - Restate important results. - Include any recommendations for additional data as needed.		10
Presentation		25	
	Comprehensibility to those outside DS		5
	Logic of presentation flow		5
	Clear recommendations		5
	Validity of ideas and strength of recommendation support		10