MARMARA UNIVERSITY
FACULTY OF ENGINEERING

CSE4062 – S25

Data Science Project Delivery #1

CAR PRICE PREDICTION SYSTEM

Group 8:

Muhammed Furkan Kahyaoğlu (ME)–150420058, furkankahyaoglu@marun.edu.tr

Özlem Demirtaş (IE) – 150320006, demirtasozlem444@gmail.com

Niyazi Ozan Ateş (CSE) – 150121991, niyaziozanates@gmail.com

Doğukan Onmaz (CSE) – 150120071, dogukanonmaz@marun.edu.tr

Şükrü Can Mayda (CSE) – 150120031, canmayda@marun.edu.tr

Instructor: Dr. Murat Can Ganiz

Project Description

For this project, we selected a dataset containing car listings with various attributes from kaggle, the owner of our dataset is wspirat and the dataset has an usability of 9.41 with 39 upvotes. This dataset includes data for vehicles manufactured between 1995 and 2023 and contains core attributes such as brand, model, color, fuel type, engine power, transmission type, price, mileage, and more.

We chose this dataset because:

- It provides a diverse set of features suitable for both regression and classification tasks.
- Predicting car prices is a real-world problem with practical applications.
- The subject is appropriate for computer science engineering, mechanical engineering and industrial engineering.
- The dataset is large enough (251,078 records) to train and test models effectively.

In this project, our goals are:

- To predict car prices using different machine learning models.
- To analyze which features have the most impact on car prices.
- To visualize the data and discover insights in the used car market.

Tools and technologies used:

- Python
- pandas, numpy, matplotlib, seaborn for data analysis and visualization

Dataset Statistics

This dataset includes 251,078 rows and 15 columns including index columns. Each row represents a unique car offer with various attributes related to vehicle specifications and market data.

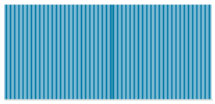
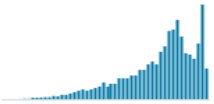
Target attributes can be used for classification:

- fuel_type: To predict fuel type based on technical specs.
- transmission_type: To classify cars as manual or automatic

Target attributes can be used for regression:

- price_in_euro: To predict prices based on features.
- mileage_in_km: To estimate the distance based on other attributes.

Some statistics and overview about our dataset:

# index	brand Alfa Romeo, Aston Martin, Audi, Bentley, BMW, Cadillac, Chery, Chevrolet, Chrysler, Citroen, Dacia, Daewoo,	model model of a car	color color of the car	registration_date month, year
 0 251k	volkswagen 13% mercedes-benz 11% Other (190572) 76%	Volkswagen Golf 3% Opel Astra 2% Other (239127) 95%	black 23% grey 19% Other (145573) 58%	 1995-01-01 2023-11-01
0	alfa-romeo	Alfa Romeo GTV	red	10/1995
1	alfa-romeo	Alfa Romeo 164	black	02/1995
2	alfa-romeo	Alfa Romeo Spider	black	02/1995
3	alfa-romeo	Alfa Romeo Spider	black	07/1995

# year year of production from 1995 to 2023	# price_in_euro price	# power_kw	# power_ps	transmission_type transmission type manual, semi automatic or automatic
 1995 27.4k	 120 5.89m	 1 2023	 1 2022	Automatic 52% Manual 47% Other (1461) 1%
1995	1300	148	201	Manual
1995	24900	191	260	Manual
1995	5900	110	150	Unknown
1995	4900	110	150	Manual

brand

Alfa Romeo, Aston Martin, Audi, Bentley, BMW, Cadillac, Chery, Chevrolet, Chrysler, Citroen, Dacia, Daewoo, Daihatsu, DFM, Dodge, Ferrari, Fiat, Ford, Geely, Honda, Hyundai, Infiniti, Isuzu, Jaguar, Jeep, Kia, Lada, Lamborghini, Lancia, Land Rover, Maserati, Mazda, Mercedes, Mini, Mitsubishi, Nissan, Opel, Peugeot, Porsche, Proton, Renault, Rover, Saab, Seat, Skoda, Smart, SsangYong, Subaru, Suzuki, Tata, Tofaş, Toyota, Volkswagen, Volvo.

volkswagen	13%
mercedes-benz	11%
Other (190572)	76%

Valid	251k	100%
Mismatched	0	0%
Missing	0	0%
Unique	47	
Most Common	volkswagen	13%

model

model of a car

Volkswagen Golf	3%

Valid	251k	100%
Mismatched	0	0%

The features, their data types(general) and their description:

Feature	Type	Description
brand	Nominal	The brand or manufacturer of the car
model	Nominal	The specific model of the car
color	Nominal	The color of the car's exterior
registration_date	Date	The registration date in MM/YYYY format
year	Numeric	The year of production
price_in_euro	Numeric	The price of the car in Euros
power_kw	Numeric	Numeric Power of the car in kilowatts
power_ps	Numeric	Power of the car in horsepower
transmission_type	Nominal	The type of transmission (e.g., Manual, Automatic)
fuel_type	Nominal	Type of fuel the car uses (e.g., Petrol, Diesel)
fuel_consumption_l_100 km	Text	Fuel consumption in liters per 100 km
fuel_consumption_g_km	Text	CO_2 emissions in grams per km
mileage_in_km	Numeric	Distance the car has traveled in kilometers
offer_description	Text	Free-text description of the offer

References

- https://www.kaggle.com/datasets/wspirat/germany-used-cars-dataset-2023
- https://www.ooyyo.com/germany/used-cars-for-sale/c=CDA31D7114D3854F111BFE6FAA651453/
- https://autoline.info/-/cars/Germany--c1169cntDE