USED CARS PRICE PREDICTION

Presented by:
Faisal Mirza

OUTLINE

1. Problem statement
2. Data
3. Data cleaning
4. Model experiments
5. Conclusion

PROBLEM STATEMENT:

- Carsgurus.com wants to know if a new post of a used car has a fair price or not
- Focused on used SUV/ Crossovers

DATA

- Dataset taken from Kaggle.com
- 3 million rows, 60 columns
- Used columns:

	city_fuel_economy	engine_displacement	highway_fuel_economy	horsepower	is_new	mileage	owner_count	price	year	make	type
0	18.0	3600.0	27.0	310.0	False	36410.0	1.0	23723.0	2018	Chevrolet	SUV / Crossover
1	15.0	3600.0	22.0	281.0	False	36055.0	1.0	22422.0	2017	Chevrolet	SUV / Crossover
2	18.0	3600.0	25.0	295.0	False	25745.0	1.0	29424.0	2018	Jeep	SUV / Crossover

DATA CLEANING

- Drop nulls
- Removing unwanted columns
- Showing only used SUV/ Crossovers
- Adding dummy variable for (make)

make_Alfa Romeo	make_Audi	...	make_Pontiac	make_Porsche	make_Rolls-Royce	make_Saab	make_Saturn	make_Subaru	make_Suzuki	make_Toyota	make_Volkswagen
0	0	...	0	0	0	0	0	0	0	0	0
0	0	...	0	0	0	0	0	0	0	0	0
0	0	...	0	0	0	0	0	0	0	0	0
0	0	...	0	0	0	0	0	0	0	0	0
0	0	...	0	0	0	0	0	0	0	0	0

MODEL EXPERIMENTS

1) Baseline model without dummy variables:

On training: 0.58

On validation: 0.67

2) Baseline model with dummy variables:

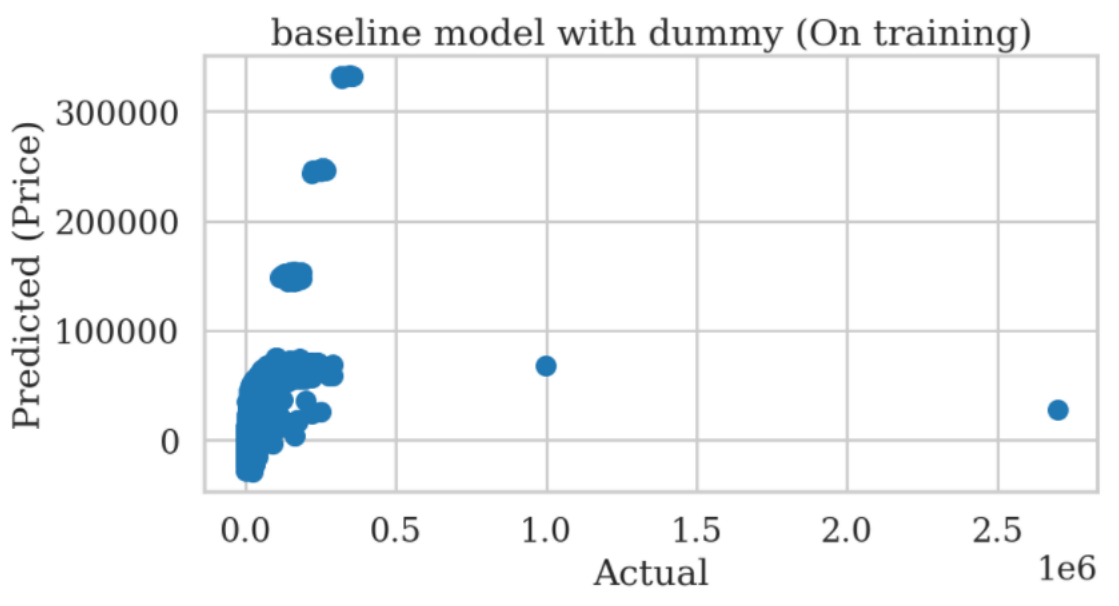
On training: 0.67

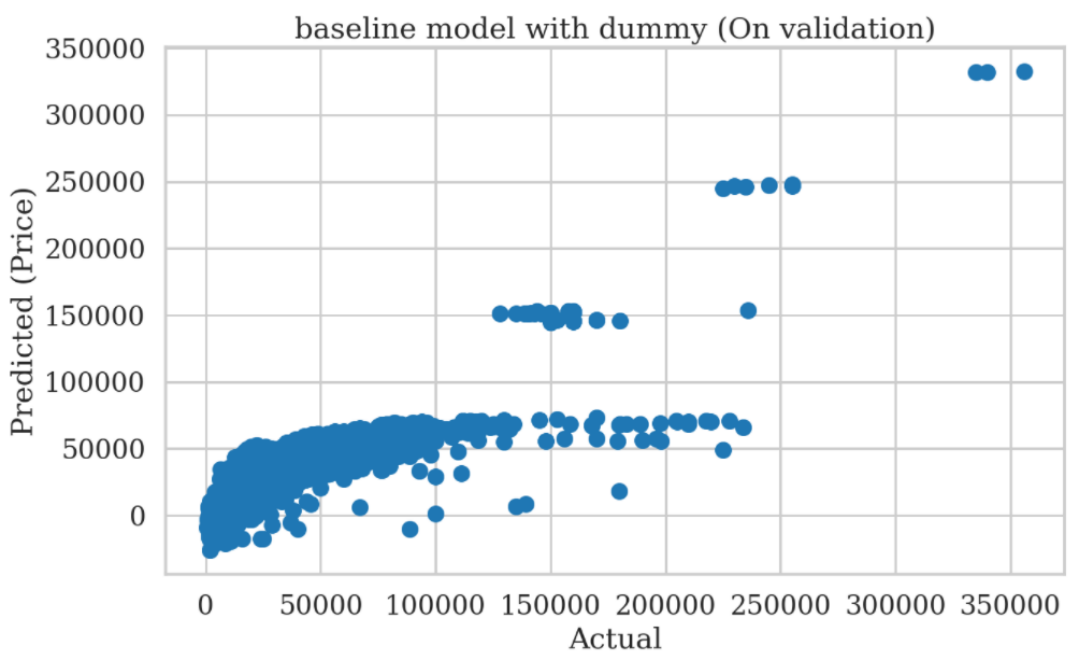
On validation: 0.78

MODEL EXPERIMENTS

2) Baseline model with dummy variables:

On training: 0.67 On validation: 0.78





MODEL EXPERIMENTS

3) polynomial without dummy variables:

On training score: 0.841

On validation score: 0.842

4) polynomial with dummy variables:

On training score: 0.912

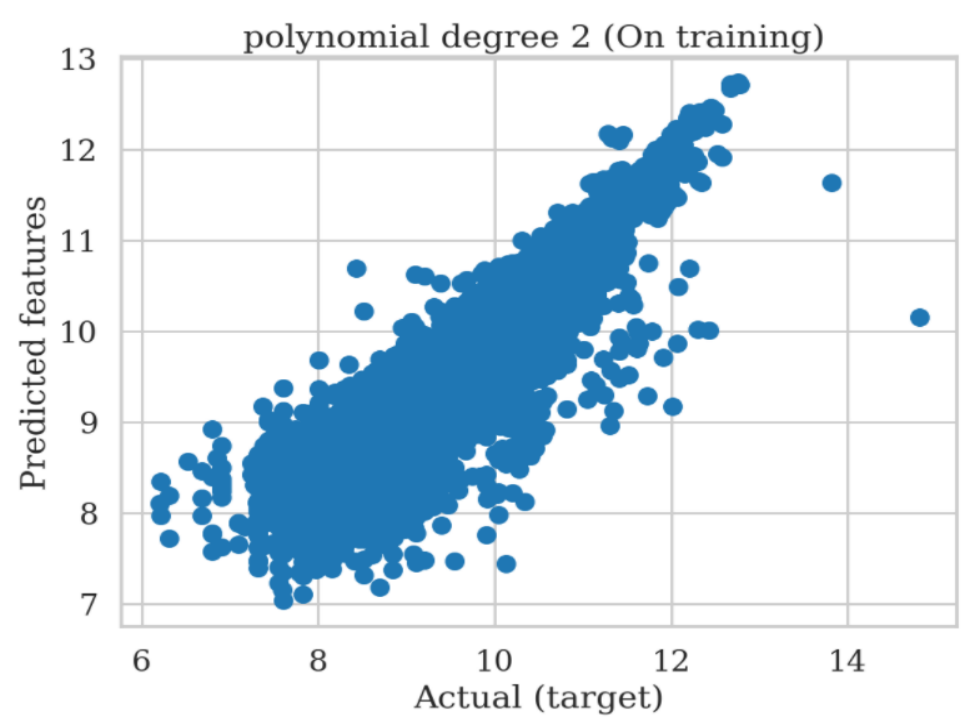
On validation score: 0.901

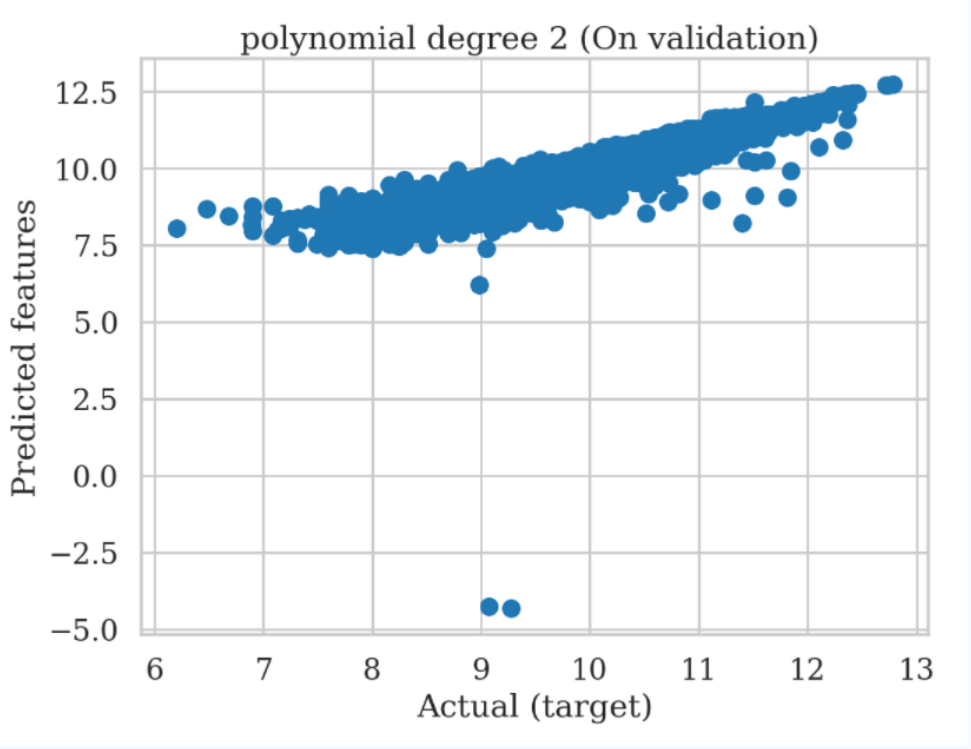
MODEL EXPERIMENTS

4) polynomial with dummy variables:

On training score: 0.912

On validation score: 0.901





MODEL EXPERIMENTS

5) Cross validation with dummy variables:

K = 5

with kfolds: 0.716

6) lasso on polynomial with dummy variables:

training score : 0.7788

validation score : 0.7781

CONCLUSION

- Best model was :

polynomial with dummy variables:

On training score: 0.912

On validation score: 0.901

THANK YOU