Northwestern

Used Car Price Predictor

MSIA 423 Final Presentation

Sophie Du

June 10, 2019

Motivation

- Predict used car prices based on user inputs of 8 significant car features
- Designed for users who
  - Plan to *SELL* a used car & need set a right price
  - Want to *BUY* a used car & evaluate how much it's worth

DEMO

Data Description

- From Kaggle, scraped from Ebay classifieds in Germany
- 370,000 used cars
- 70%-30% train-test split
- Response variable:

  Used car price

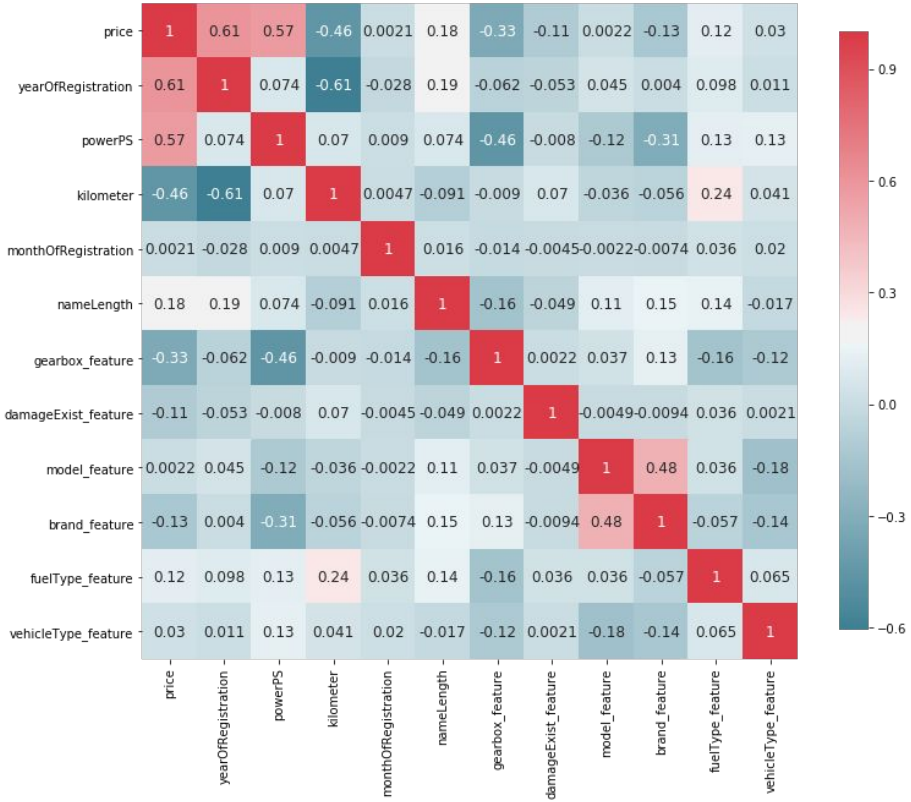
Predictors

- Registration Year
- Miles
- Car Brand
- Horsepower
- Vehicle Type (sedan, SUV, etc)
- Fuel Type (gas, diesel, etc)
- Automatic vs Manual
- Damage exists?

Modeling - Random Forest

- Random forest regressor ➔ Fit model
- Grid Search ➔ Set optimal parameters
  - N_estimators: 500 ➔ 500 trees in the forest
  - Max_depth: 10 ➔ 10 depth of each tree
  - Min_samples_split: 3 ➔ >= 3 samples split an internal node
  - Min_samples_leaf: 3 ➔ >= 3 samples required at a leaf node
- Mean Squared Error: 0.016
- R-squared: 0.87

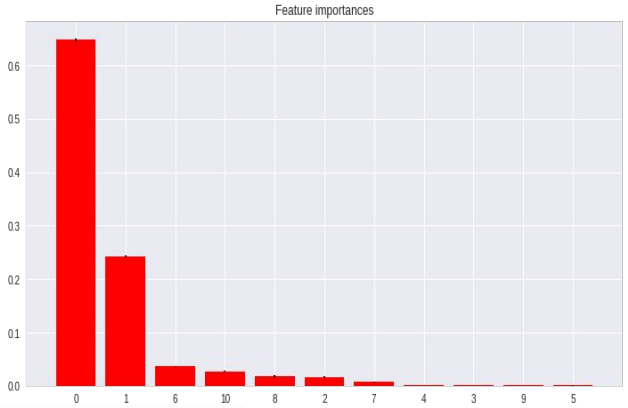
Interesting Insight from Data



Most highly correlated variables with price:

- Year of first registration
- Horsepower
- Miles the car has driven

Surprisingly, car model is NOT significant at all.



THANK YOU

chuandu2019@u.northwestern.edu