

dataset

In the provided dataset, there are two vehicle models. The purpose is to make a model predicting its trim and price. The trims are different for both vehicle models. Additionally, all cars are from the USA, thus the models from American modification that limits engine characteristics. One should note that cars' prices depend on the state where they are being sold.

https://www.isecars.com/used-car-prices-by-state-study?_isctk=l3vzzw

The dataset was split by model. Separately for each vehicle model was built a ML model.

nlp model

Firstly, I build a trim prediction linguistic model (tf-idf + naive bayes) for predicting model. I noticed that the most useful data for that prediction is contained in the column "VehFeats".

regression model

Secondly, I used obtained Trim and state, history with continues data to make price prediction. For each vehicle model I found the best parameters for the DecisionTreeRegressor algorithm using the GridSearchCV algorithm.