International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Accurate Car_Price Prediction Using Machine Learning Model

Sweta Umakant Mungde

Swami Ramanad Teerth University, Nanded

Abstract

The Research paper used to build a Machine Learning Model on price prediction of used car.The car price depends on multiple factors like car -type,fuel type,urban area,local area ,year of purchasing,model of car,how many kilometre car has been driven.This model builds used car on reasonable prices.This prediction are useful to sellers, buyers, and car manufacturers in the used cars market.For this use of artificial intelligence comes in picture,machine learning model is used to measure accurate prediction .Machine learning model gives approximate accurate price prediction based on the information that I scrap it from data of used cars. Different categorical regression models are used such as linear regression,ridge and lasso regression,random forest regressor,Extra tree regressor and support vector regression are apply on model to achieve the highest accuracy.Catboost Regression Metrics are used to measure mathematical result of model such as 1.Mean Squared Error 2.Mean Absolute Error 3.Root Mean Squared Error(RMSE).These gives statistical arithmetic mean group of values.To build Car price prediction model python Jupyter notebook,numpy,Pandas and to fetch dataset mysql are used. Exploratory Data Analysis and feature selection technique is used to clean and fit data for testing and traing purposes.The dataset is divided and modified to fit the regression.To evaluate the performance of each regression mean squred gives highest accuracy MAE gives 0.8568669682476366 was calculated. Random forest algoritham achieved the highest score.The resulting model includes more aspects and higher prediction accuracy.

Keywords: Here,for building “Car-price-prediction” model mentioned some important keywords and programming language such as Python,pandas,anaconda navigator,My sql server,Data Analysis,Business Intelligence,Radom Forest algoritham, Ensemble Learning,Bagging method to train decision trees

Introduction

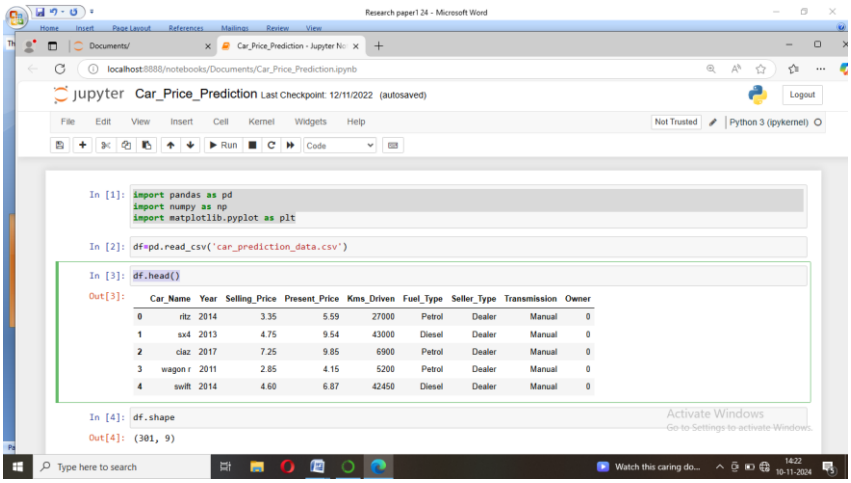
Car price prediction especially when car is used and not coming directly from factory is both critical and Important task.As the automotive market continues to expand ability to precisely predict car prices becomes cruicial for range of stakeholders from individual buyers and sellers to dealerships and insurance companies.The customer who doesn’t know about the technical specification and other prices of spare parts and how to deduct price will easily be cheated with high price .I wanted to solve this type of problem where customer has to know exact price car is worth for.using machine learning it is possible to predict correct price and other technical issues.This can be done by training the model using used cars dataset which has several features and parameters.There are many features from which the cars price can

be predicted. And also we can add if there is any damage or is it flood affected or accidental damaged car these factors can also be considered for predicting correct and exact price of car

METHODOLOGY 1

I have selected required used car prices data and some required features and parameter from kaggle dataset.To build model we i used Google colab with there inbuilt librararies-

import pandas as pd
import numpy as np
import matplotlib pyplot as plt




```
feat_importances = pd.Series(model.feature_importances_, index=X.columns)
feat_importances.nlargest(5).plot(kind='barh')
plt.show()
```

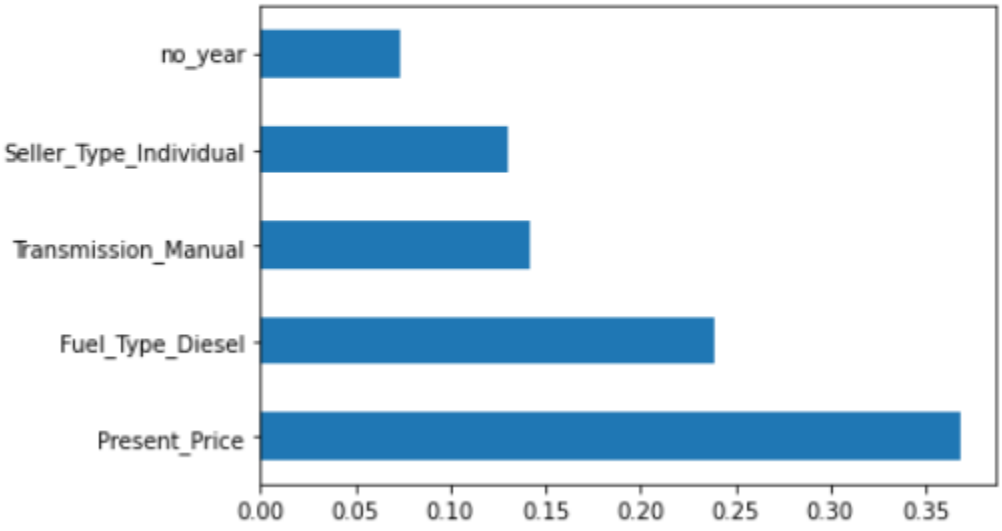


Figure shows plot diagram of fuel_type

Algorithms

1. Linear Regression

Linear Regression was chosen as the first model due to its simplicity and comparatively small training time. The features, without any feature mapping, were used directly as the feature vectors. No regularization was used since the results clearly showed low variance.

2. Random Forest

It is an ensemble learning based regression model. It uses a model called decision tree, specifically as the name suggests, multiple decision trees to generate the ensemble model which collectively produces a prediction. The benefit of this model is that the trees are produced in parallel and are relatively uncorrelated, thus producing good results as each tree is not prone to individual errors of other trees. This uncorrelated behavior is partly ensured by the use of Bootstrap Aggregation or bagging providing the randomness required to produce robust and uncorrelated trees.

3. Gradient Boost

Gradient Boosting is another decision tree based method that is generally described as “a method of transforming weak learners into strong learners”. This means that like a typical boosting method, observations are assigned different weights and based on certain metrics, the weights of difficult to predict observations are increased and then fed into another tree to be trained.

4. XGBoost Extreme

Gradient Boosting or XGBoost is one of the most popular machine learning models in current times. XGBoost is quite similar at the core to the original gradient boosting algorithm but features many additive features that significantly improve its performance such as built in support for regularization, parallel processing as well as giving additional hyperparameters to tune such as tree pruning, sub sampling and number of decision trees. A maximum depth of 16 was used and the algorithm was run on all cores in parallel.

Using LGBM Regressor

Here i used XGBOOST which is an extream gradient boosting machine but lightlime we cam can achieve much better results by traing our dataset in less time.

It play a cruitial part in rasing accuracy and dependability of data, which eventually improves effectiveness of machine learning model/

params = {
"gamma": uniform(0, 0.5),
"learning_rate": uniform(0.03, 0.3), # default 0.1
"max_depth": randint(2, 6), # default 3
"n_estimators": randint(100, 150), # default 100
"subsample": u

```
plt.scatter(y_test,predictions)
```

```
<matplotlib.collections.PathCollection at 0x7f0da8b9a160>
```

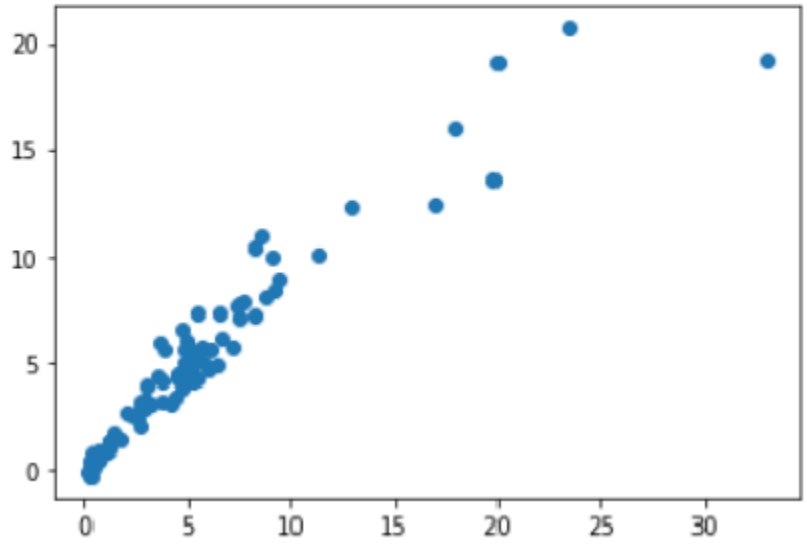


Figure shows scattered diagram

CatBoostRegressor():

```
In [83]: print('MAE:', metrics.mean_absolute_error(y_test, predictions))
         print('MSE:', metrics.mean_squared_error(y_test, predictions))
         print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, predictions)))
```

```
MAE: 0.8568669682476366
MSE: 3.4731227057809524
RMSE: 1.8636315906801302
```

References

1. https://www.kaggle.com/datasets/zafarali27/car-price-prediction
2. https://www.statista.com/topics/1487/automotive-industry/
3. https://stats.stackexchange.com/questions/394705/classification-xgboost-vs-logistic-regression
4. https://www.kaggle.com/nehalbirla/vehicle-dataset-from-cardekhohttps://www.kaggle.com/nehalbirla/vehicle-dataset-from-cardekho
5. Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. ACM, 2016.
6. Ke, Guolin, et al. "Lightgbm: A highly efficient gradient boosting decision tree." Advances in Neural Information Processing Systems. 2017. 6. Fisher, Walter D. "On grouping for maximum homogeneity." Journal of the American statistical Association 53.284 (1958): 789-798.