Grenze International Journal of Engineering and Technology, June Issue

Auto Value Estimation: Car Price Prediction

Anant Kumar[1], Ashish Kumar[2] and Vandana Sharma[3]
[1-3]Department of Computer Science, ABES ENGINEERING COLLEGE, Ghaziabad, UP, India
Email: Anant.20b0121021@abes.ac.in, {Ashish.20b0121118, Vandana.sharma}@abes.c.in

***Abstract*— As a result of incredible technological advancements and research of new technical expertise and huge economic growth of our country, people started to buy cars more than other vehicles. Therefore, there arises an enormous demand for cars, as the demand increases for new cars the used car market also booms alike. But the used car market is highly manipulated by few numbers of people who govern the rates of the used cars and also online selling websites designate the values for the used cars. This paper tries to study and investigate the trends in used car prices and predicts the price of used cars with the help of supervised machine learning algorithms. And to suggest which machine learning algorithm performs well among the selected methods for predicting the cars price. There has been related work done with machine learning algorithms like linear regression, multiple regression, random forest and so on. We wanted to study which algorithm predicts the car price more reliably and accurately**

***Index Terms*— Supervised Machine Learning Algorithms, Linear Regression, Multiple Regression, Random Forest.**

I. INTRODUCTION

Currently, the automotive industry is experiencing a transformation marked by rapid technological advancements, evolving consumer preferences, and a focus on sustainability. With the industry's continued innovation, researchers and industry experts are increasingly focused on predicting new car prices. Optimizing production and marketing strategies requires accurate price predictions not only for automakers but also for consumers to make informed purchasing decisions. In this study, we explore innovative methodologies and advanced computational models to enhance the precision of forecasting new car prices. Market trends and historical data were traditionally used to inform pricing strategies. It is necessary, however, to take a more nuanced and sophisticated approach to understanding the current automotive landscape, integrating factors such as advanced technologies, fuel efficiency, and geopolitical influences. As we embark on this research journey, our goal is to unravel the complexities of predicting new car prices. We will achieve this by considering a wide range of variables and utilizing advanced data analytics techniques. Our aim is to provide valuable insights that will empower both industry stakeholders and consumers to confidently navigate the ever-changing automotive market. in updating the ambulance location in the database. It is also possible to control the traffic signal based on the approaching ambulance vehicle, resulting in a less congested route.

II. OBJECTIVES

1. Providing a standardized pricing mechanism for new car designs based on objective data and statistical analysis.
2. Help the designer team to prioritize design features and make more informed decisions about where to allocate

Grenze ID: 01.GIJET.10.2.50

resources and effort.

3. To provide a fair estimate of the market value of a car based on objective data and advanced statistical techniques.

4. Helping the team to optimize the price of the new car design based on various factors such as production costs, market demand, and profitability.

III. Literature Review

[1] In the September 2009 edition of the Journal of Theoretical and Applied Information Technology, Durgesh K. Shrivastava and Lekha Bhambhu present a study on "Data Classification Using Support Vector Machine." This paper explores the application of Support Vector Machine (SVM) techniques for effective data classification. The authors contribute insights into the utilization of SVM in the context of data analysis, providing valuable information on its potential and applications in the field of information technology. [2] In their 2019 paper published in the TEM Journal, Gegic et al. present a comprehensive review on the prediction of car prices using machine learning techniques. This study, which emphasizes innovative methodologies, contributes to the field of automotive pricing by exploring advanced data-driven approaches. Despite being confined to a concise single page, the authors manage to provide valuable insights into the application of machine learning in accurately and efficiently predicting car prices, highlighting the continuous evolution of predictive modelling in the automotive industry. [3] Published in the International Journal of Computer Sciences and Engineering in 2019, the review paper by Chandak et al. makes a significant contribution to the field of automotive pricing. Their study, which spans an extensive 444-450 pages, delves into advanced data-driven techniques, shedding light on the practical application of machine learning in accurately forecasting car prices. [4] In the 2020 edition of the International Journal of Innovative Research in Applied Sciences and Engineering, Samruddhi and Kumar present a study on how technology is being integrated into predictive car pricing models. Their research, which spans 629-632 pages, explores the use of a specific machine learning approach called k-nearest neighbour for accurate forecasting in the used car market. By examining the efficacy of this model, the authors provide valuable insights into the field of predictive modelling in the automotive sector and its potential for precise price predictions. [5] In a 2019 paper published in the International Journal of Computer Sciences and Engineering, Ashish Chandak, Prajwal Ganorkar, Shyam Sharma, Ayushi Bagmar, and Soumya Tiwari delve into the topic of "Car Price Prediction Using Machine Learning." This research, featured in Volume 7, Issue 5, May 2019, explores innovative machine learning techniques that can be used to forecast car prices. Through their study, the authors contribute valuable insights to the constantly evolving world of automotive pricing, shedding light on advancements that improve the accuracy and efficiency of predictive modelling. [6] In a 2014 review paper by Pudaruth in the International Journal of Information and Computer Technology, the focus is on using machine learning techniques to predict used car prices. This paper delves into different approaches to using machine learning for accurately predicting the prices of used cars. By doing so, it contributes to the expanding knowledge base on how advanced technologies can improve the accuracy and efficiency of assessing the value of pre-owned vehicles. [7] In the September 2009 issue of the Journal of Theoretical and Applied Information Technology, Durgesh K. Shrivastava and Lekha Bhambhu share their research on "Data Classification Using Support Vector Machine." Their study focuses on the practical use of Support Vector Machine (SVM) techniques for accurate data classification. The authors offer valuable knowledge on how SVM can be applied in data analysis, shedding light on its potential and relevance in the realm of information technology.

IV. Proposed Methodology

I have obtained the necessary dataset of used car prices from Kaggle, a platform that provides data and resources for data scientists and analysts. To ensure accuracy in predicting prices, I have cleaned and pre-processed the data using machine learning techniques. Following this, I divided the data into two parts - one for training and the other for validation - using the train test split method. Next, I applied a simple linear regression model to predict the output and assessed its accuracy using the roc auc score. Additionally, I trained and tested the data using multiple linear regression models to validate their accuracies. To further enhance the prediction, I employed clustering methods, logistic regression methods, and kNN methods. Furthermore, I experimented with random forests and decision tree algorithms. Finally, I compared the accuracies of all the machine learning algorithms and selected the best one for the prediction task

V. Modelling And Analysis

Visualizing the features or attributes related for predicting the price of the new car.

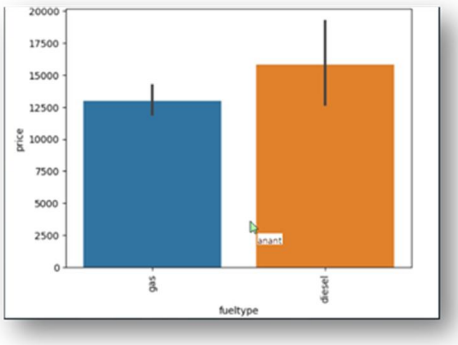

Figure 1. Note how the Fuel type affect price

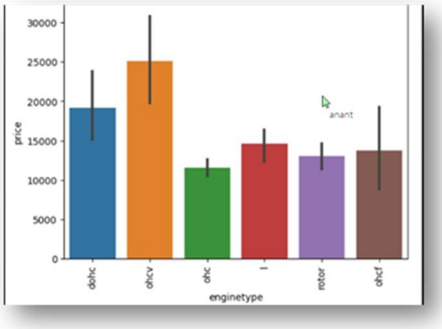

Figure 2. Note how the Engine type affect price

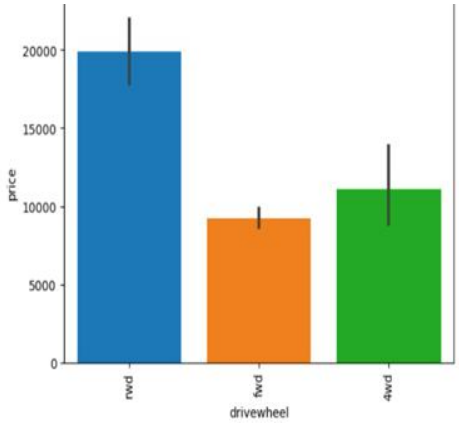

Figure 3. Note how the Drive wheel affect price

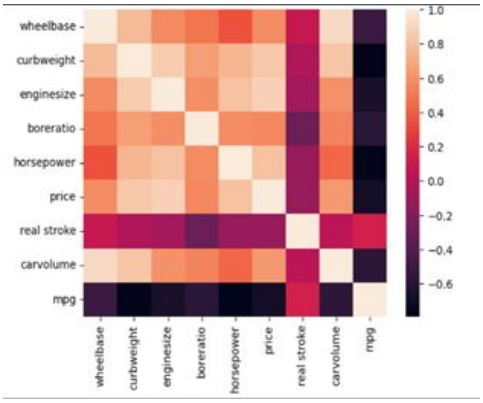

Figure 4. Note Heatmap of the attributes

VI. Result And Conclusion

```
step1 = ColumnTransformer(transformers=[
    ('col_tnf',OneHotEncoder(sparse=False,drop='first'),[0,1,2,3,6])
],remainder='passthrough')

step2 = LinearRegression()

pipe = Pipeline([
    ('step1',step1),
    ('step2',step2)
])

pipe.fit(x_train,y_train)

y_pred = pipe.predict(x_test)
print("Linear Regression Testinng !")
print('R2 score=',r2_score(y_test,y_pred))
print('MAE=',mean_absolute_error(y_test,y_pred))
print('MAPE=',mean_absolute_percentage_error(y_test,y_pred))
```

```
Linear Regression Testinng !
R2 score= 0.7229624250624016
MAE= 2573.0575339488714
MAPE= 0.19042070239205675
```

Figure 5. Note Accuracy of Linear Regression

```
from sklearn.linear_model import  BayesianRidge
step1 = ColumnTransformer(transformers=[
    ('col_tnf',OneHotEncoder(sparse=False,drop='first'),[0,1,2,3,6])
],remainder='passthrough')

step2 = BayesianRidge()

pipe = Pipeline([
    ('step1',step1),
    ('step2',step2)
])
0
pipe.fit(x_train,y_train)

y_pred = pipe.predict(x_test)
print("Bayesian Ridge Testinng !")
print('R2 score=',r2_score(y_test,y_pred))
print('MAE=',mean_absolute_error(y_test,y_pred))
print('MAPE=',mean_absolute_percentage_error(y_test,y_pred))
```

```
Bayesian Ridge Testinng !
R2 score= 0.5950146888563029
MAE= 3306.8136748058355
MAPE= 0.23554951651187528
```

Figure 6. Note Accuracy of Bayesian Ridge

```
from sklearn.ensemble import RandomForestRegressor
step1 = ColumnTransformer(transformers=[
    ('col_tnf',OneHotEncoder(sparse_output=False,drop='first'),[0,1,2,3,6])
],remainder='passthrough')

step2 = RandomForestRegressor(n_estimators=100,
                              random_state=10,
                              max_samples=0.9,
                              max_features=0.3,
                              max_depth=15)

pipe = Pipeline([
    ('step1',step1),
    ('step2',step2)
])

pipe.fit(x_train,y_train)

y_pred = pipe.predict(x_test)
print("Random Forest Regressor Testinng !")
print('R2 score=',r2_score(y_test,y_pred))
print('MAE=',mean_absolute_error(y_pred,y_test))
print('MAPE=',mean_absolute_percentage_error(y_pred,y_test))
```

```
Random Forest Regressor Testinng !
R2 score= 0.9059370262957341
MAE= 1500.8650011382113
MAPE= 0.10372643966484509
```

Figure 7. Note Accuracy of Random Forest Regressor

When comparing the R2 scores, Mean Absolute error and Mean Absolute Percentage Error scores of Linear Regression model (Fig 5), Bayesian Regression model (Fig 6) and Random Forest Regressor model (Fig 7), we can infer than the random forest model performs far better with high accuracy score of 91%. and Mean Absolute Percentage Error is 10.32 which is the best value for saying the model is the greatest model, therefore it is better to use random forest model for this car price prediction project.

VII. Conclusion

I acquired the necessary dataset of used car prices, including the desired features and parameters, from Kaggle. Kaggle serves as an open-source platform for data scientists and analysts, offering a range of data and notebooks. Before utilizing any prediction algorithms, the data underwent cleaning and preprocessing using various machine learning techniques. Once the data was prepared, it was divided into two parts for training and validation purposes, employing the train-test split method. Initially, a simple linear regression model was applied to predict the output, and its accuracy was assessed using the roc auc score. Subsequently, a multiple linear regression model was trained and tested, and its accuracies were validated. Additionally, clustering methods, logistic regression methods, and KNN methods were utilized to predict car prices. Random forests and decision tree algorithms were also considered. Finally, all the accuracies obtained from the various machine learning algorithms were compared, and the most effective algorithms were chosen for prediction purposes.

References

[1] Gajera, p., gondaliya, a. And kavathiya, j., 2021. Old car price prediction with machine Learning. Int. Res. J. Mod. Eng. Technol. Sci, 3, pp.284-290.

[2] Pudaruth, s., 2014. Predicting the price of used cars using machine learning techniques. Int. J. Inf. Comput. Technol, 4(7), pp.753-764.

[3] Gegic, e., isakovic, b., keco, d., masetic, z. And kevric, j., 2019. Car price prediction using Machine learning techniques. Tem journal, 8(1), p.113.

[4] Chandak, a., ganorkar, p., sharma, s., bagmar, a. And tiwari, s., 2019. Car price prediction Using machine learning. International journal of computer sciences and engineering, 7(5), Pp.444-450.

[5] Samruddhi, k. And kumar, r.a., 2020. Used car price prediction using k-nearest neighbor based model. Int. J. Innov. Res. Appl. Sci. Eng.(ijirase), 4, pp.629-632.

[6] Ashish chandak, prajwal ganorkar, shyam sharma, ayushi bagmar, soumya tiwari, car price prediction using machine learning, international journal of computer sciences and engineering, volume 7, issue 5, may 2019.

[7] Durgesh k. Shrivstava, lekha bhambhu, “data classification using support vector machine”, journal of theoretical and applied information technology, sep. 2009.

[8] Kumar, N., chauhan, r., dubey, g. (2021). Forecasting of stock price using lstm and prophet algorithm. In: choudhary, a., agrawal, a.p., logeswaran, r., unhelkar, b. (eds) applications of artificial intelligence and machine learning. Lecture notes in electrical engineering, vol 778. Springer, singapore. Https://doi.org/10.1007/978-981-16-3067-5_12.

[9] R. Pramanik, r. Agrawal, m. Kumar gourisaria and p. Kumar singh, "comparative analysis of mobile price classification using feature engineering techniques," 2021 5th international conference on information systems and computer networks (iscon), mathura, india, 2021, pp. 1-7, doi: 10.1109/iscon52037.2021.9702320.

[10] Rastogi, r., chaturvedi, d.k., satya, s., arora, n. (2020). Intelligent heart disease prediction on physical and mental parameters: a ml based iot and big data application and analysis. In: jain, v., chatterjee, j. (eds) machine learning with health care perspective. Learning and analytics in intelligent systems, vol 13. Springer, cham. Https://doi.org/10.1007/978-3-030-40850-3_10

[11] A. Verma, s. Goyal, s. K. Thakur, a. Gupta and i. Gupta, "mortality prediction using machine learning techniques: comparative analysis," 2019 ieee 9th international conference on advanced computing (iacc), tiruchirappalli, india, 2019, pp. 230-234, doi: 10.1109/iacc48062.2019.8971566.

[12] Kumar, S., syed, m.h. (2023). House pricing prediction based on composite facility score using machine learning algorithms. In: singh, p.k., wierzchoń, s.t., tanwar, s., rodrigues, j.j.p.c., ganzha, m. (eds) proceedings of third international conference on computing, communications, and cyber-security. Lecture notes in networks and systems, vol 421. Springer, singapore. Https://doi.org/10.1007/978-981-19-1142-2_18