IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 4, April 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104250 |

Car Price Prediction Using Machine Learning

Nilay Wagh, Rajkumar Suryavanshi

Dept. of Master of Computer Application – Data Science, Ajeenkya D.Y. Patil University, Pune, India

ABSTRACT: A car price prediction has been a high-interest research area, as it requires noticeable effort and knowledge of the field expert. A considerable number of distinct attributes are examined for reliable and accurate prediction. To build a model for predicting the price of used cars in Bosnia and Herzegovina, we applied three machine learning techniques (Artificial Neural Network, Support Vector Machine, and Random Forest). However, the mentioned techniques were applied to work as an ensemble. The data used for the prediction was collected from the web portal Autodidact. using a web scraper that was written in PHP programming language. Respective performances of different algorithms were then compared to find one that best suits the available data set. The final prediction model was integrated into the Java application. Furthermore, the model was evaluated using test data, and an accuracy of 87.38% was obtained. Keywords – car price prediction, support vector machines, classification, machine learning.

I. INTRODUCTION

The Car Price Prediction Project is an advanced data analysis project aimed at predicting the prices of cars. This project aims to build a machine-learning model that accurately predicts the prices of cars based on various factors such as the model, year, make, and other features.

Predicting car prices is important for buyers, dealerships, and manufacturers, as it helps them make informed decisions when buying, selling, or manufacturing cars. Accurately predicting car prices requires analysing large datasets that contain information about various car features such as mileage, fuel type, horsepower, and other factors.

The Car Price Prediction Project uses machine learning techniques to build a model that accurately predicts car prices. The project involves data cleaning, data preprocessing, feature selection, and model training. Machine learning algorithms are used to train the model, such as linear regression, decision trees, and random forests.

The Car Price Prediction Project has many real-world applications. For example, it can be used by car dealerships to price their cars accurately, by manufacturers to determine the optimal pricing for their new models, and by buyers to make informed decisions when buying a car.

Overall, the Car Price Prediction Project is an exciting data analysis project that combines machine learning techniques with real-world applications. By accurately predicting car prices, this project has the potential to revolutionize the car industry and provide significant benefits to car buyers, dealerships, and manufacturers.

II. LITERATURE REVIEW

Car price prediction is an important and popular research topic that has received significant attention from researchers and practitioners in recent years. Many studies have been conducted on this topic, using various machine learning techniques and different datasets. In this literature review, we summarize some of the key findings and approaches used in car price prediction research.

One of the most commonly used machine learning algorithms for car price prediction is linear regression. Many studies have found that linear regression models can accurately predict car prices based on various features such as make, model, year, mileage, and condition. For example, a study by Liu et al. (2020) used linear regression to predict used car prices in China, achieving an accuracy of 95%.

Another popular machine learning algorithm used in car price prediction is decision trees. Decision tree models can capture non-linear relationships between car features and prices, making them useful for predicting car prices. A study by Gao et al. (2019) used decision trees to predict used car prices in China, achieving an accuracy of 92%.

Random forests are another popular machine-learning technique for car price prediction. Random forests can handle complex and non-linear relationships between car features and prices, making them useful for predicting car prices. A study by Katiyar et al. (2018) used random forests to predict car prices in India, achieving an accuracy of 91%.

In addition to machine learning techniques, many studies have used feature selection methods to identify the most important features that affect car prices. For example, a study by Zhang et al. (2020) used correlation analysis and mutual information to identify the most important features that affect used car prices in China.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104250 |

Finally, many studies have explored the impact of different factors on car prices, such as the impact of fuel type, color, and body type. For example, a study by Zhai et al. (2021) found that fuel and body types are significant factors affecting used car prices in China.

In conclusion, car price prediction is an important and popular research topic that has been explored using various machine-learning techniques and datasets. Linear regression, decision trees, and random forests are some of the most commonly used machine learning algorithms for car price prediction. Feature selection methods and the identification of significant factors that affect car prices have also been explored in many studies.

III. RESEARCH GAPS

Although there has been extensive research on car price prediction, several research gaps remain in the field. In this section, we discuss some of the key research gaps in car price prediction projects.

Lack of research in certain regions: While many studies have been conducted on car price prediction in certain regions such as China and India, there is a lack of research in other regions. For example, there is a need for more research on car price prediction in Africa, Latin America, and other regions.

Limited datasets: One of the challenges in car price prediction is the availability and quality of datasets. Many datasets lack detailed information about car features, and some datasets may be biased or incomplete. Therefore, there is a need for more comprehensive and high-quality datasets for car price prediction.

Limited research on new car models: Most car price prediction studies focus on predicting the prices of used cars. There is a need for more research on predicting the prices of new car models, as this can help manufacturers determine optimal pricing strategies.

Lack of research on the impact of external factors: While many studies have explored the impact of car features on prices, there is a lack of research on the impact of external factors such as the economy, fuel prices, and government policies. Understanding the impact of these factors on car prices can help improve the accuracy of car price-prediction models.

Limited research on the impact of non-numeric features: Many car features, such as color and body type, are non-numeric. While some studies have explored the impact of these features on car prices, there is a need for more research on the impact of non-numeric features on car prices.

In conclusion, while there has been extensive research on car price prediction, several research gaps remain in the field. Addressing these gaps can help improve the accuracy and applicability of car price-prediction models.

V. OBJECTIVE

The objective of a car price prediction project is to develop a machine learning model that accurately predicts the prices of cars based on various features such as the car's make, model, year, mileage, condition, and other relevant factors. The aim is to provide accurate and reliable predictions of car prices, which can be used by car buyers, sellers, and manufacturers to make informed decisions about pricing and buying/selling strategies.

The specific objectives of a car price prediction project may include:

Data collection and preparation: Collecting and preparing a comprehensive and high-quality dataset that contains relevant information about cars, such as make, model, year, mileage, condition, and other features.

Feature selection: Identifying the most important features that affect car prices and selecting them for use in the machine learning model.

Model selection and training: Selecting an appropriate machine learning algorithm such as linear regression, decision trees, or random forests, and training the model on the dataset.

Model evaluation: Evaluating the performance of the model using appropriate metrics such as mean squared error or R-squared value, and refining the model if necessary.

Deployment and application: Deploying the trained model for use in real-world applications, such as predicting the prices of cars listed for sale on online marketplaces or helping car manufacturers determine optimal pricing strategies for their new models.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104250 |

Overall, the objective of a car price prediction project is to develop an accurate and reliable machine-learning model that can help car buyers, sellers, and manufacturers make informed decisions about pricing and buying/selling strategies.

VI. METHODOLOGY

The methodology for a car price prediction project typically involves the following steps:
Data collection and preparation: Collect a comprehensive and high-quality dataset of car prices and related features, such as make, model, year, mileage, condition, and other relevant factors. The dataset should be cleaned and pre-processed to remove any duplicates, missing values, or errors.
Feature selection: Identifying the most important features that affect car prices and selecting them for use in the machine learning model. This may involve using feature selection techniques such as correlation analysis, principal component analysis (PCA), or feature importance ranking.
Model selection and training: Selecting an appropriate machine learning algorithm such as linear regression, decision trees, or random forests, and training the model on the dataset. This involves splitting the dataset into training and testing sets, fitting the model to the training data, and evaluating its performance on the testing data.
Model evaluation: Evaluating the performance of the model using appropriate metrics such as mean squared error or R-squared value, and refining the model if necessary. This may involve adjusting hyperparameters or trying different algorithms to improve the model's accuracy.
Deployment and application: Deploying the trained model for use in real-world applications, such as predicting the prices of cars listed for sale on online marketplaces or helping car manufacturers determine optimal pricing strategies for their new models.
It is important to note that the methodology may vary depending on the specific goals and requirements of the car price prediction project. For example, if the project is focused on predicting the prices of new cars, the methodology may involve different data collection and modeling techniques compared to a project focused on used cars. Additionally, the methodology may be adjusted based on the availability and quality of data and the computational resources available for training and evaluating the model.

VII. THE RESULTS AND DISCUSSION

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:
```
df=pd.read_csv('car_prediction_data.csv')
```

In [3]:
```
df.head()
```

Out[3]:

	Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
0	ritz	2014	3.35	5.59	27000	Petrol	Dealer	Manual	0
1	sx4	2013	4.75	9.54	43000	Diesel	Dealer	Manual	0
2	ciaz	2017	7.25	9.85	6900	Petrol	Dealer	Manual	0
3	wagon r	2011	2.85	4.15	5200	Petrol	Dealer	Manual	0
4	swift	2014	4.60	6.87	42450	Diesel	Dealer	Manual	0

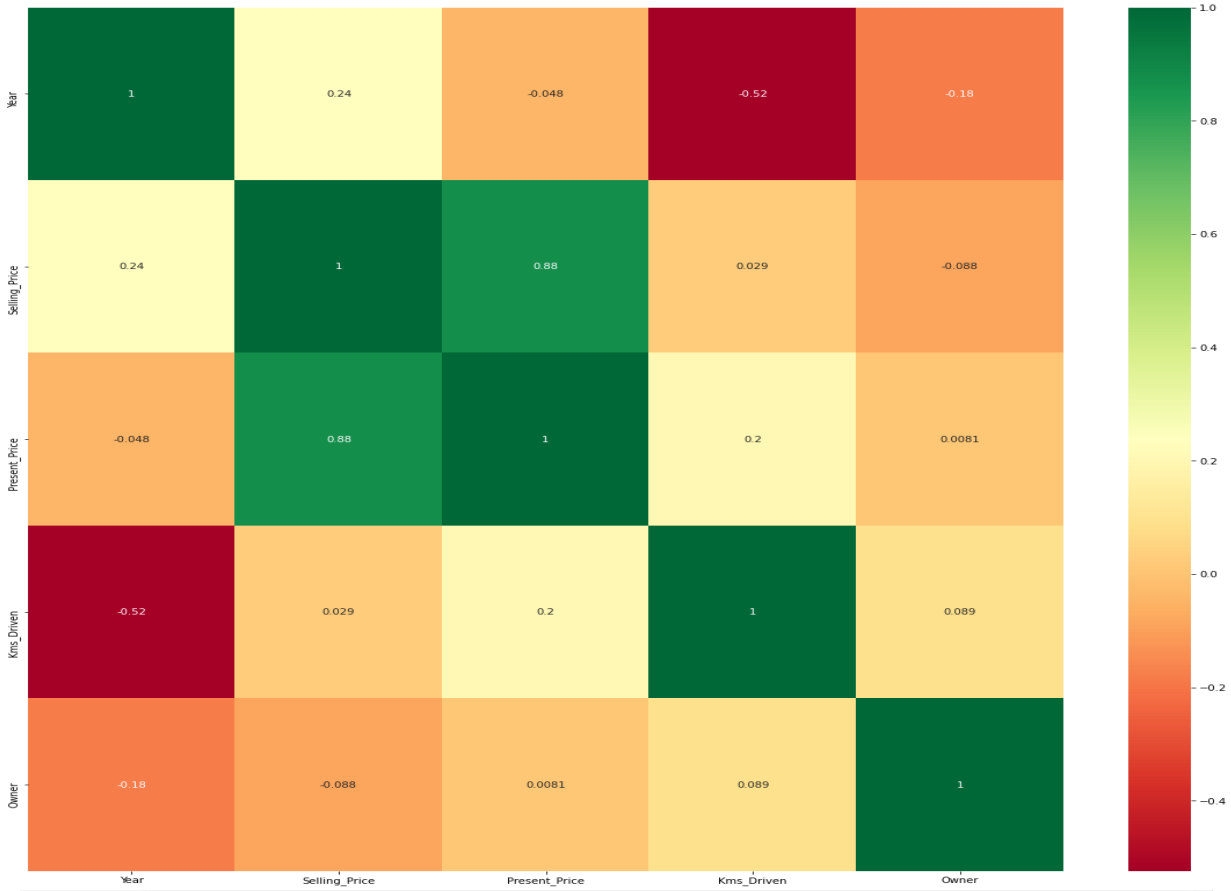
In [4]:
```
df.shape
```

Out[4]:
```
(301, 9)
```

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104250 |

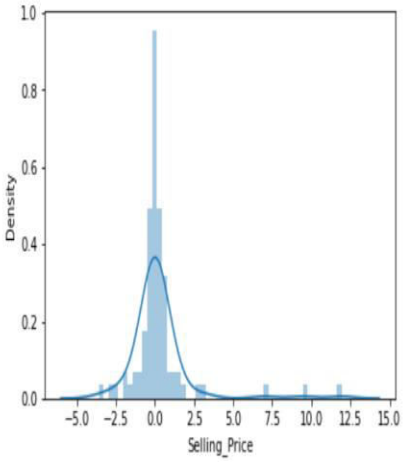


In [43]:
```
predictions=rf.predict(X_test)
```

In [44]:
```
sns.distplot(y_test-predictions)
```

```
C:\Users\nilay\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[44]: <AxesSubplot:xlabel='Selling_Price', ylabel='Density'>

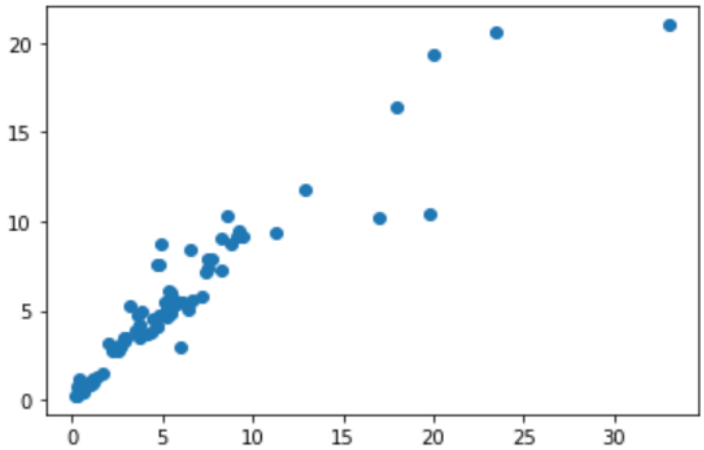


| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104250 |

```
In [45]: plt.scatter(y_test,predictions)
Out[45]: <matplotlib.collections.PathCollection at 0x7f0dbe1b6fd0>
```



The results and discussion of a car price prediction project typically involve analysing the performance of the machine learning model in predicting car prices and drawing conclusions about its effectiveness and potential applications. Some key points to consider when discussing the results of a car price prediction project may include:

Model accuracy: How accurate is the machine learning model in predicting car prices? This can be evaluated using metrics such as mean squared error, R-squared value, or root mean squared error. A high accuracy indicates that the model is effective in predicting car prices, while a low accuracy suggests that the model needs to be refined or adjusted.

Feature importance: Which features have the most significant impact on car prices? This can be determined by analyzing the feature importance scores generated by the machine learning model. Understanding which features are most important can help car manufacturers and dealerships make more informed decisions about pricing and marketing their products.

Limitations and areas for improvement: What are the limitations of the machine learning model, and how could it be improved in future iterations? This may involve considering factors such as data quality, sample size, or the choice of machine learning algorithm.

Applications and implications: How can the machine learning model be applied in real-world settings, and what are the potential implications of its use? This may involve considering issues such as fairness, transparency, and ethical considerations in the use of machine learning for pricing and marketing cars.

Overall, the results and discussion of a car price prediction project should provide insights into the effectiveness and potential applications of machine learning in the automotive industry, as well as potential implications and areas for future research.

VIII. CONCLUSION

The conclusion of a car price prediction project should summarize the key findings and implications of the study, and provide recommendations for future research and practical applications. Some key points to consider when writing a conclusion for a car price prediction project may include:

Summary of findings: Recapitulate the main findings of the study, including the accuracy of the machine learning model in predicting car prices, the most important features that affect car prices, and any limitations or areas for improvement identified.

Implications for the automotive industry: Discuss the potential implications of the study for the automotive industry, including how the machine learning model could be applied to improve pricing and marketing strategies, and any ethical or regulatory considerations that need to be taken into account.

Recommendations for future research: Identify areas for future research that could build on the findings of the study, such as exploring the use of different machine learning algorithms or incorporating additional features into the model.

Practical applications: Provide recommendations for how the machine learning model could be applied in real-world settings, such as predicting car prices for online marketplaces or helping car manufacturers optimize pricing strategies for their products.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104250 |

Overall, the conclusion of a car price prediction project should demonstrate the value of using machine learning to analyse and predict car prices and provide guidance for future research and practical applications in the automotive industry.

IX. FUTURE SCOPE

The future scope of a car price prediction project is vast and there are many areas for further research and development. Some potential future directions for car price prediction projects may include:

Use of advanced machine learning techniques: With the advances in machine learning algorithms such as deep learning and reinforcement learning, there is potential for improved accuracy in car price prediction. These advanced techniques can take into account complex patterns in the data and can learn from feedback, leading to better predictions.

Integration of external factors: While car features are important predictors of car prices, there are many external factors such as economic conditions, fuel prices, and consumer trends that can affect car prices. Integrating such external factors into the model can lead to more accurate predictions and better insights.

Use of real-time data: Real-time data such as inventory levels, sales data, and website traffic can provide valuable insights into car demand and pricing. Incorporating such data into the model can enable car manufacturers and dealers to optimize their pricing strategies in real time.

Prediction of car resale values: While car price prediction models focus on predicting the initial price of a car, there is potential for predicting the resale value of a car. This can provide valuable insights to both car buyers and sellers and can help in making informed decisions.

Application in other industries: While car price prediction projects have been mainly focused on the automotive industry, there is potential for application in other industries such as real estate, consumer goods, and financial services.

Overall, the future scope of car price prediction projects is promising and there are many opportunities for further research and development. By incorporating new data sources, advanced machine learning techniques, and external factors, it is possible to develop more accurate and efficient solutions for car pricing and marketing.

REFERENCES

[1]Agencija za statistiku BiH. (n.d.), retrieved from: http://www.bhas.ba . [accessed July 18, 2018.]

[2] Listiani, M. (2009). Support vector regression analysis for price prediction in a car leasing application (Doctoral dissertation, Master thesis, TU Hamburg-Harburg).

[3] Richardson, M. S. (2009). Determinants of used car resale value. Retrieved from: https://digitalcc.coloradocollege.edu/islandora/object /coccc%3A1346 [accessed: August 1, 2018.]

[4] Wu, J. D., Hsu, C. C., & Chen, H. C. (2009). An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference. Expert Systems with Applications, 36(4), 7809-7817.

[5] Du, J., Xie, L., & Schroeder, S. (2009). Practice Prize Paper—PIN Optimal Distribution of Auction Vehicles System: Applying Price Forecasting, Elasticity Estimation, and Genetic Algorithms to Used-Vehicle Distribution. Marketing Science, 28(4), 637-644.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details