International Journal of Advance and Innovative Research
Volume 12, Issue 2 (XXI): April - June 2025
ISSN 2394 - 7780

CAR PRICE PREDICTION USING MACHINE LEARNING

[1]Aman Kumar, [2]Adityalok, [3]Dr. Pooja Kapoor, [4]Rashmi Richhariya, [5]Kamlesh Kumar Yadav, [6]Devesh
[1, 2, 5, 6]Department of Computer Science, MIET, Greater Noida, India
[3, 4] Assistant Professor, Department of Computer Science, MIET, Greater Noida, India

ABSTRACT
The main goal of this project is to know the actual car prices, check prices, and estimate the lifespan of a certain including vehicle's mileage, manufacturing year, fuel consumption, transmission, road tax, type of fuel, and engine size.

A new car is reported to lose 10% of its value when driven out of the car's showroom. In this, the no. of kilo-meter's the car has been driven is the most essential factor in figuring out its price. As other factors include, it's necessary to understand that different car manufacturers price their vehicles differently, which results in price differences in models.

Therefore, to find the car price that every factor suited for the buyer, Keywords — ***component, formatting, style, styling, insert (key words)***

Keywords – *car price, machine learning, prediction, python, vehicles and how these factors can be integrated into the prediction process.*

I. INTRODUCTION

This industry is one of the largest and most dynamic sectors of the global economy, with a wide range of vehicle models available to consumers. However, understanding the price of a car can be challenging for both buyers and sellers. Many factors come into play such as the make, model, year mileage, condition and market demand. Estimating the value of cars becomes complex due to these variables.

In years there have been advancements in using machine learning and artificial intelligence to predict car prices with great precision. By analyzing data and employing algorithms and computational power car price prediction models provide valuable insights into a vehicle's likely cost.

These models serve purposes for buyers and sellers alike by assisting in negotiations setting prices and enabling informed decision making.

This Introduction aims to explore the concept of car price prediction models; their significance and the methodologies that underpin them. a key factor in enhancing overall productivity. We will delve into components such, as data collection and preprocessing techniques feature selection methods, model training approaches and evaluation processes. Further, we will discuss how various factors impact car prices.

we are going to predict its cost by using different Machine Learning algorithms available in the Python Environment. Our dataset consists of data related to different car brands with a set of parameters (Name, Location, Year, Fuel Type, Transmission, Owner Type, Mileage, Engine, Power, Seats, Price).

The primary purpose is to build a model for a given dataset and predict the car price, check which model has the most accuracy and make sure that the money spent on the car is a good investment for anyone.

II. RELATED WORK

Car price prediction models have become increasingly popular in times because of their usage, in the automotive industry and the wider realm of data-driven decision- making. Experts and analysts have extensively studied methodologies and strategies to enhance the precision and dependability of these models. In this section, we will examine some of the research conducted in the field of car price prediction. In the existing system, many data mining algorithms and machine learning algorithms are widely used to predict the price of 2-wheelers and 4-wheelers. The biggest drawback of the current system is that it requires a lot of behavior to predict the car price. More comparative methods should be used to obtain better results. It is very difficult to access the necessary information shared worldwide. Information can only be collected online. But not in offline mode. Especially in regions, everyone can't collect information over the Internet. Vehicles that have not been used for a long time will not be included in the configuration data. Vehicles belonging to the model may or may not be included in the configuration data.

The main disadvantage of the current system is that the system is very slow since most of the key questions attempt to identify only one point and are not suitable for many applications that require analysis of various

ISSN 2394 - 7780

categories of vehicle content. There is no query speed retrieval method, and since there are no constrained support vector machines (SVM), the retrieval speed is slow. Some of the current studies include:

In a research paper titled "Machine Learning, for Predicting Used Car Prices" published in 2018 the authors **i**mportance of ensuring data quality performing feature**.**

In 2017 researchers carried out a study called "Forecasting Car Prices Using Time Series Analysis" where they explored time series forecasting methods to predict car prices. By considering price trends and seasonality patterns their aim was to enhance price predictions for new cars.

In the research article titled "Enhancing Car Price Prediction, with Hybrid Models" published in 2020 experts have investigated the effectiveness of models that integrate regression techniques, with machine learning and deep learning methodologies. The study showcases how ensemble methods can be leveraged to enhance prediction accuracy highlighting their advantages.

Recent studies have incorporated sentiment analysis techniques to assess market sentiment by analyzing reviews and social media data for predicting car prices. This approach provides insights, into market trends and consumer opinions.

In summary, the field of car price prediction has seen significant advancements through the adoption of various machine learning and data analysis techniques. Researchers continue to explore innovative approaches to enhance the accuracy and practicality of these models, offering valuable tools for buyers, sellers, and industry professionals in the automotive market. These studies highlight the significance of data quality, feature engineering, and model evaluation in achieving accurate and reliable car price predictions. The evolution of these models holds the potential to revolutionize the way we assess and negotiate car prices, making transactions more transparent and efficient.

III. METHODOLOGY

We have created a very good model to overcome this problem. Machine learning algorithms are used because they give us continuous results as output instead data preprocessing step was applied. Tools with unexpected benefits will be managed accordingly; for example, in our case, we replaced them with the instruments with the highest return value in the instrument. Vehicles that have no value will be disposed of early.

To remove the competition of mileage of different cars, all the mileages of cars are scaled to a kmpl due to the car’s records are in km. To change divided data conducted a study on the application of machine learning algorithms such as Random Forests and Gradient Boosting to predict the prices of used cars. They highlighted the significance of selecting features and preprocessing the data to enhance the accuracy of their models.

A survey conducted in 2020 titled "Utilizing Machine Learning Techniques for Car Price Prediction" examined machine learning algorithms utilized for predicting car prices with a focus on regression-based models. The survey discussed the engineering and evaluating models accurately.

Therefore, it is possible to estimate the real value of the car rather than its price. A user interface was also created that can receive feedback from all users and display the price of the vehicle based on the user's input. Vehicle price estimation is done accurately based on different features and qualities and with the help of experienced experts. The most important factors to estimate are the model type the usage of the vehicle, and the mileage of the vehicle. Since fuel prices change frequently, the type of fuel used and the mileage of the fuel affect the cost of the vehicle. Different features such as exterior color, number of doors, transmission type, size, security, air conditioning, interior and navigation also affect the price of the vehicle. In this article, we use various methods and techniques to obtain more accurate vehicle cost estimations.

The following attributes were captured for each car: Name, Location, Year, Fuel Type, Transmission, Owner Type, Mileage, Engine, Power, Seats, and Price expressed in Indian rupees.

MODEL TRAINING

LINEAR REGRESSION

After collecting and storing the data, the data, linear regression attempts to model the relationship between two variables. The term "dependent variable" refers to the other variable. A statistical technique called linear regression is used to forecast or determine the connection between two distinct variables. Finding the best-fitting line to represent the connection between the independent and dependent variables is the goal, supposing a linear relationship between them. In data science and machine learning, linear regression is utilized for analysis and prediction. For example, you could use linear regression to predict weight if you knew an individual's height. In this example, if an individual was 70 inches tall, you would predict their weight:

ISSN 2394 - 7780

Weight=80+2x (70) = 220lbs. values into numeric attributes like (Company, Name, Location, Fuel, Transmission, and Owner) we have used a encoding approach: Linear Regression By fitting a linear equation to observable.

Random forest Regressor

The Random Forest Regressor is a potent machine learning instrument. It resembles a group of decision trees collaborating to provide predictions. Every tree contributes to the final prediction as it is trained on a distinct subset of the data. Imagine it as a group of experts voting on a decision. Combining the votes of all the experts (trees), each of which has an opinion based on a part of the data, yields a prediction that is more accurate.

Two methods are used to add randomness: first, each tree is trained on a random subset of the data (bagging); second, only a random subset of characteristics is taken into account at each decision point in a tree. This reduces overfitting and improves the usability of the model. Random forest regression is an attempt to describe the connection between two variables after gathering and storing the data. Random Forest's ability to handle both regression and classification jobs is one of its amazing features. Regression is useful for applications like quantity or price prediction since it predicts a continuous result.

Using it in Python with a library like Scikit-Learn involves creating a Random Forest Regressor, fitting it to your training data, and then making predictions on new data. So, in a nutshell, the Random Forest Regressor is like a wise crowd of decision-makers, each with its own perspective, coming together to give you a solid prediction for your regression problem.

Gradient boosting Regressor

Gradient Boosting Regressor is a group of specialists working on solving a problem. Each specialist is like a mini- expert, and they team-up to improve their collective performance. This is how it operates: The first expert (the tree) attempts to forecast the result, but it may be inaccurate. Rather than giving up, the subsequent expert arrives, recognises those errors, and concentrates on fixing them. Each specialist refines and enhances the forecasts produced by the preceding ones in this process, which is repeated. It's like learning from mistakes and getting better with each attempt. These specialists are humble – they're weak learners, not trying to do everything on their own. But when they team up, their collective wisdom becomes a strong predictive force. Create a Gradient Boosting Regressor, train it on your data, and then let it make predictions on new data to utilise this team in Python with Scikit-Learn.

After collecting and storing the data, the data, gradient- boosting regression attempts to model the relationship between two variables. One technique that stands out for its accuracy and speed of prediction, especially when working with big and complicated datasets, is gradient boosting. This algorithm has yielded the greatest results across several platforms, including Kaggle contests and corporate machine learning solutions. We can reduce the bias error of the model by using the gradient boost approach. It successively aggregates the predictions of several weak learners, usually decision trees. By steadily lowering prediction errors and raising the model's accuracy, it seeks to increase overall predictable performance by optimising the model's weights based on the errors of prior iterations.

Extreme Gradient Boosting

In the field of machine learning, Extreme Gradient Boosting (XG-Boost) is a superstar because to its accuracy and efficiency.

Speed is another superpower. XG-Boost is designed to be lightning-fast, thanks to its ability to handle tasks in parallel. It's like having multiple teammates working on different parts of the problem at the same time. This superstar also knows how to handle missing information gracefully, saving you from a headache during data processing. It prunes unnecessary branches from the decision trees as they grow, ensuring the team focuses on what really matters. When it comes to playing with data, especially structured or tabular data, XG-Boost shines. It's like having a maestro who understands the rhythm of your data and orchestrates it to perfection.

To use XG-Boosting in Python, we can create an XG- Boost Regressor, train it on your data, and it will make predictions. Fine-tuning is crucial; adjusting parameters like the learning rate ensures the model gives its best performance. XG-Boost is essentially your machine learning team's MVP—it is productive, efficient, and constantly aiming for perfection when it comes to making precise predictions.

Reg: linear is the most often used loss function in XG-Boost for regression issues, while reg: logistical is the most often used loss function for binary classification. XG-Boost is one of the ensembles learning techniques. Ensemble learning entails training and integrating individual models (referred to as base learners) to obtain a single prediction.

ISSN 2394 - 7780

After training and testing the data we got the accuracy of the model are-

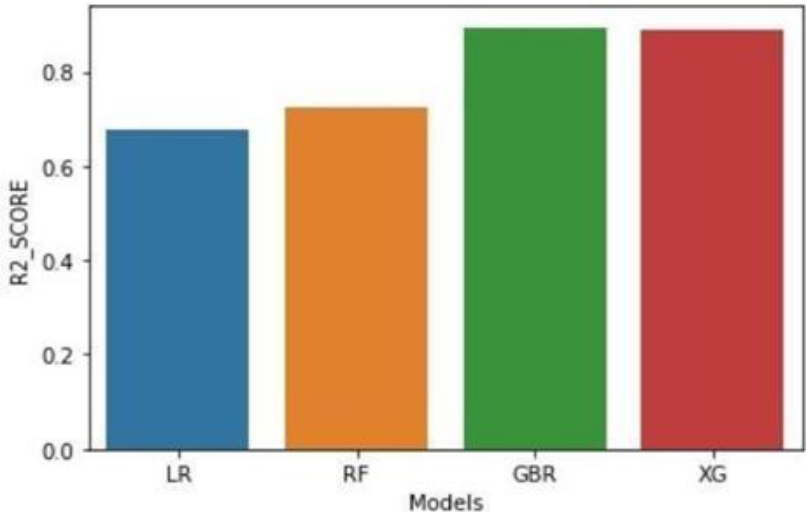
```
final_data
```

	Models	R2_SCORE
0	LR	67.908850
1	RF	73.450048
2	GBR	89.501617
3	XG	88.874718

Final Data Accuracy (%)

RESULTS AND DISCUSSION

After applying the different regression algorithm models, it has shown different accuracy results according to the dataset we can apply the completed model for the prediction of car prices from manual input of various data of the car such as the Car name, selling price, owner type, fuel type, etc. after that we get the output as follows:



IV. CONCLUSION

Car prices can be a difficult task because an accurate estimate requires taking into account many features. The key steps in the forecasting process are data collection and prioritization. As new car prices increase in the market, there is a need for second-hand car sales at all levels for people who cannot afford new car prices.

Therefore, there should be a car price estimate that will estimate the price of the car based on many factors. Applying this modeling will help determine accurate traffic forecasting. With the help of a lot of research data, we developed a model using a different regression algorithm and managed to create the model.

FUTURE WORK

In the future, this machine learning model will be connected to many websites to provide instant data for price prediction. We can also add more older data on car prices, which will help improve the efficiency and accuracy of machine- learning models.

We can create an Android application that works as a user interface for user interaction.

We plan to develop a deep learning model for communication integrity, use adaptive learning, and train sets of data instead of entire data to achieve better performance. The purpose of the machine learning model will be to connect to various datasets and websites to provide real-time information for cost estimation. We may also send large amounts of traffic data to help improve the accuracy of machine- learning models.

ISSN 2394 - 7780

REFERENCES

1) Sameerchand Pudaruth, ―Predicting the Price of Used Cars using Machine Learning Techniques‖; (IJICT 2014)

2) https://en.wikipedia.org/wiki/Machine_learning.

3) https://www.kaggle.com/jpayne/852k-used-car- listings

4) Enis gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric, ‖Car Price Prediction Using Machine Learning‖; (TEM Journal 2019)

5) Google, Youtube

6) Gegic, E., Isakovic, B., Keco, D., Masetic, Z., & Kevric, J. (2019). Car price prediction using machine learning techniques. TEM Journal, 8(1), 113–118ddddddd

7) Bukvić, L., Pašagić Škrinjar, J., Fratrović, T., & Abramović, B. (2022). Price prediction and classification of used vehicles using supervised machine learning. Sustainability, 14(24), 17034.

8) Kaur chitranjanjit, kapoor pooja, kaur Gurjeet(2023), "image recognition(soil feature extraction)using Metaheuristic technique and artificial neural network to find optimal output.Eur. Chem. Bull.2023(special issue 6).

9) Maheshwari Chanana shalu, Kapoor pooja,kaur chitranjanjit(2023),"Data mining techniques adopted by google: A study.: Empirical Economics Letters,22(special issue 2).

10) Jiang, X. (2024). Research for car price prediction based on machine learning. Transactions on Computer Science and Intelligent Systems Research, 5, 1608–1617.

11) Venkatasubbu, P., & Ganesh, M. (2023). Used cars price prediction using supervised learning techniques. International Journal of Innovative Research in Applied Sciences and Engineering.

12) Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM.