iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.61441

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Car Price Prediction Using Machine Learning

Lakshmi Prasad Mudarakola[1], D Shabda Prakash[2], K L N Shashidhar[3], D Yaswanth[4]
Dept. of CSE Institute of Aeronautical Engineering Hyderabad, India

Abstract: ***The objective of the "Car Price Prediction Using Machine Learning" project is to anticipate car prices based on pertinent variables by utilizing predictive modelling and advanced data analytics. The rising need for precise and active pricing systems in the automobile sector is addressed by this initiative. The system will evaluate past vehicle data, including brand, model, manufacturing year, the mileage, type of fuel, and other details, by utilizing machine learning techniques. In order to produce specific predictions, the suggested model will be trained on an extensive dataset, noticing patterns and parallels within the data. In order to create a reliable prediction model, the research focuses on using regression techniques, such as ensemble approaches or linear regression. System of measurement like as absolute mean error and R-squared coefficient will be used to evaluate the predicted accuracy of the system in order to regulate its usefulness. If this resourcefulness is employed successfully, it will have a big impact on the automobile business for clients and venders alike by giving them a tool to evaluate fair market prices and helping them make decisions. This study enhances to the body of knowledge in price analysis using machine learning and lays the footing for future improvements in the prognostication of dynamic market trends in the automobile sector.***
Keywords: Car Price Prediction, Supervised Machine Learning, Linear Regression, Random Forest.

I. INTRODUCTION

A state-of-the-art usage of predictive analytics is machine learning for automotive price prediction. It uses cultured algorithms to study past vehicle sales data and other applicable data. The main goal is to estimate a car's cost constructed on its many features. For a diversity of automotive industry players, such as vehicle venders, clients, and dealerships, this predictive approach is useful because it gives them the ability to make well-informed judgements about pricing approaches and maintain their attractiveness in the market. The process relies on machine learning algorithms trained on a dataset, in details about cars. This dataset contains data such, as the brand, model, year of manufacture, mileage, aspiration, bore-ratio and other related factors. Machine learning algorithms will predict the prices of the cars after analyzing the dataset, by identifying the patterns and relationships present among them. The possible impact of using machine learning for predicting car prices covers to inducing decisions made by dealers, buyers, and sellers. By estimating the prices of the cars correctly helps customers to make better choices in purchasing the vehicle. Moreover, this technology allows sellers to launch prices that appeal customers while ensuring cost-effectiveness. By often changing their pricing strategies in reply to new customer likings and market changes, dealerships can gain. To predict the price of a car, we have to follow certain steps. Firstly, is data cleaning which involves picking the dataset that will address absent information, removing duplicates, and fixing inconsistencies. Following is feature engineering that involves selecting relevant features for prediction as well as coming up with additional different ones from old ones. Model selection comes next in which several algorithms are used like neural networks, random forests, decision trees and linear regression. This particular model is trained on past information to determine the complex relationships among costs and attributes of cars. There are various ways of estimating car prices owing to machine learning algorithms. A simple method for recognizing linear association between features and pricing is the application of linear regression. Alternatively, neural networks which are one type of deep learning can capture complex patterns in data while decision trees together with random forests excel at capturing complex relationships. To conclude, car price prediction based on machine learning provides a dynamic tool for all. Predictive analytics can enable them to make more informed, dynamic pricing changes and increase intelligent insights into market trends. This app allows the industry to be more open and competitive, thus improving overall efficiency of the car market. The precision and application of learning algorithms in predicting car prices are expected to improve further and transform the automotive industry as technology advances.

II. BACKGROUND & RELATED WORK

[1] The aim of this paper is therefore to examine some of the ways that machine learning can be used to predict car prices by applying various different pricing dynamics. Also, linear regression models dominate in the field of car price prediction as they provide strong evidence on how input variables affect automobile prices due to their linear relations with features. Support vector regression improves prediction accuracy because it can handle complex patterns and nonlinear effects.

Random Forest algorithms efficiently capture complex relationships among multiple features through a set of decision trees that produce accurate predictions. These wide arrays of algorithms guarantee a thorough examination of factors affecting auto prices that is crucial for an ever-changing automotive industry.

[2] This article gives an extensive measurement for comparing several types of regressions in forecasting motorcar prices. Like this, polynomial, linear and ANN regression are compared and contrasted as well. A polynomial regression model captures non-linear relationships whereas simple baseline is provided by linear regression. On the other hand, deep learning methods including Artificial neural networks have capabilities of detecting complicated designs. The current work observes how different models achieve these goals.

[3] Study on the use of data mining techniques in the field of auto price estimation such as naive Bayes, k-nearest neighbours and decision trees. K-Nearest Neighbors predict prices by finding similar incidents, Decision Trees map and analyze decision paths based on input features while Naive Bayes computes probability based on assumptions of independence among attributes. These methods are used to explore various algorithmic choices in detail thereby improving understanding of how well they can detect patterns and make accurate estimates concerning car prices. This study emphasizes the importance of involving a number of data mining techniques to enhance its adaptability.

III. EXISTING SOLUTION

In the automotive industry, machine learning (ML)-based car price prediction has become quite common providing shrewd insights to traders and customers. The present-day cars employ highly complex machine learning algorithms to analyze various factors that affect car prices that provide accurate predictions. Naturally, many of these systems use ensemble methods, regression models or deep learning techniques to do this. Machine learning (ML)-based automobile price forecasting has gained popularity in the automotive industry, providing customers and dealers with valuable data. Many of the systems in use today analyse a multitude of factors influencing car expenses and generate accurate projections using advanced machine learning algorithms. These descriptions often employ cooperative practices, regression models, or supervised learning techniques to discover patterns from large datasets. Techniques like Gradient Boosting and Random Forest are used to increase the prediction accuracy. In this situation, such methods create a robust predictor by combining together various weaker ones. For instance, while Random Forests can handle complicated feature connections and non-linearities; Gradient Boosting minimizes forecast errors in the model. Forecasting automobile prices, neural nets are appreciated for complex jobs owing to their ability to handle vast amounts of data and capture non-linear correlations well. Hyper-parameter adjustments and Cross-validation techniques are the authentication processes which ensures the reliability of the predictions. Cross-validation assurances that the model simplifies well to new instances by testing it on different segments of data. By hyper parameter alteration, the design of the ML model is improved leading to higher predictive capability. Car pricing varies significantly dependent on geographic reflections; some methods include location-specific factors or explanation for regional market variations. This localized approach enhances forecast accuracy given the large local discrepancies in pricing dynamics. Briefly stated, these machine learning approaches use a variety of techniques such as deep learning, advanced ensemble methods and conventional regression models for forecasting car costs today. To make accurate and dependable forecasts, these systems rely on a combination of effective feature engineering, powerful algorithms, and validation procedures Buyers as well as sellers gain from these models' constant evolution with fresh data and adaptation to market trends, which help to create a more transparent and knowledgeable automobile market.

IV. METHODOLOGY

The information was gathered from questionnaires administered by the RTO office and through the use of the Kaggle website. There are 26 attributes in all in the dataset. The following was included in the dataset:

attributes include the vehicle's name, fuel type, aspiration, symbolism, body, wheelbase, length, width, and height; engine characteristics include compression, bore ratio, and horse power.

An environment is developed by using Anaconda prompt. This environment would assist as a barrier between our project area and the base, additional default environment, or other environments that have previously been made. It helps us to install all the necessary packages and modules manually in the environment.

Importing and reading the research's CSV file is the initial stage. The dataset is thoroughly analysed in terms of its shape, columns, unique values for each feature, data information, and other aspects such as null values and numerical and categorical characteristics.

Clearer names were given to some data features (Present Price = Starting Price, Owner = Former Owners), and features that weren't necessary for analysis were eliminated.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Statistical graphics and other visualisation techniques are used in exploratory data analysis to describe the significant features of the data. The most popular cars, the year compared to the number of cars available, the selling price compared to the initial price, the fuel type, the transmission type, the seller type, the age, the selling price in relation to age, the seller type, the transmission, the fuel type, the selling price in relation to previous owners, the initial price compared to the selling price, and the selling price in relation to kilometres To comprehend data better, driven, pairplot, heat map visualisations, and other visualisations are employed.

Subsequently, many plots are created and used to develop and comprehensively examine the dataset's correlation properties. The dependent feature (selling cost) and independent characteristics (original price, kilometres driven, previous ownership, age, and so forth) are then assigned for additional processing in the features allocation of the data.

We divide the dataset into training and testing data once both dependent and independent characteristics have been assigned. We train our model on 80% of the data, and we test it on 20%.

After the Train-Test split, the model building procedure starts after the data modelling is finished. For later use, the model and a few of its parameters are specified. The final results are produced using a variety of techniques once the model is constructed. The following methods are used for predictive modelling once the model is built.

Linear Regression: In statistics, linear regression is a methodical technique used to predict the correlations between scalar responses and independent and dependent variables. Relationships are described in linear regression using functions like the linear predictor, and data is used to estimate unknown model parameters.

Random Forest: The ensemble learning technique is used by the supervised learning algorithm Random Forest for both regression and classification. Trees in random forests grow in parallel to one another without any contact throughout development. A meta estimator called Random Forest combines the results of many forecasts. Additionally, it combines many decision trees with specific adjustments

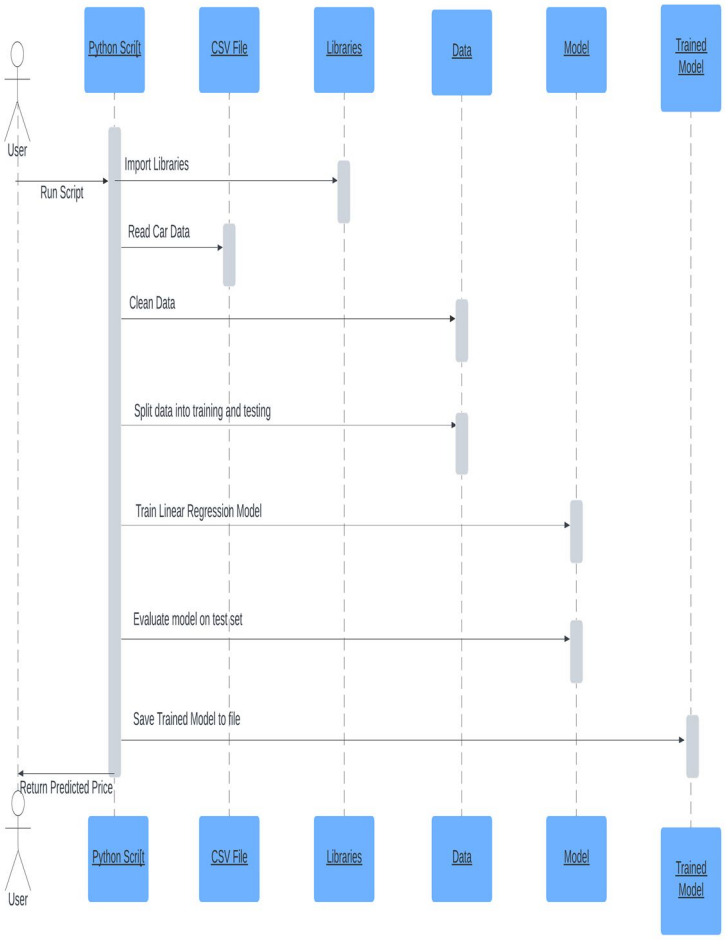


Figure 1 : Sequence Diagram

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

V. IMPLEMENTATION

Adding new features like citympg and highwaympg. Citympg is the mileage given by the car in the city and highwaympg is the mileage of the car when used on the highway.

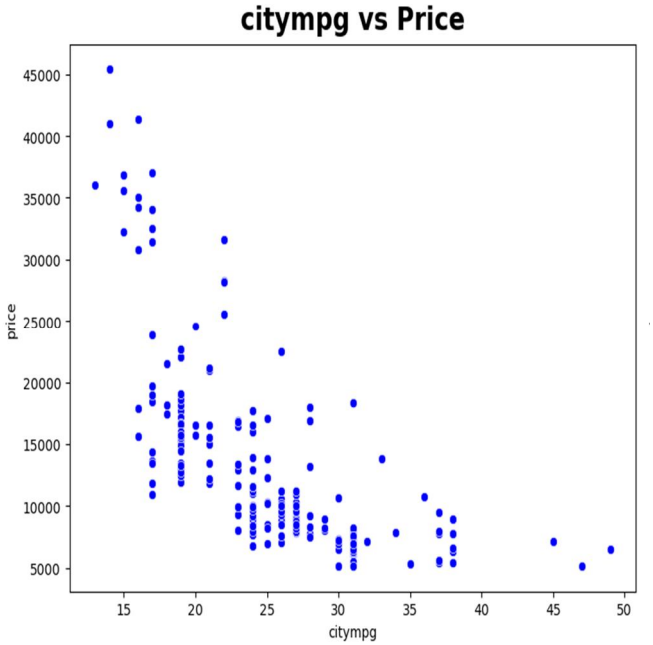


Figure 2: Visualizing "Citympg vs Price" Features

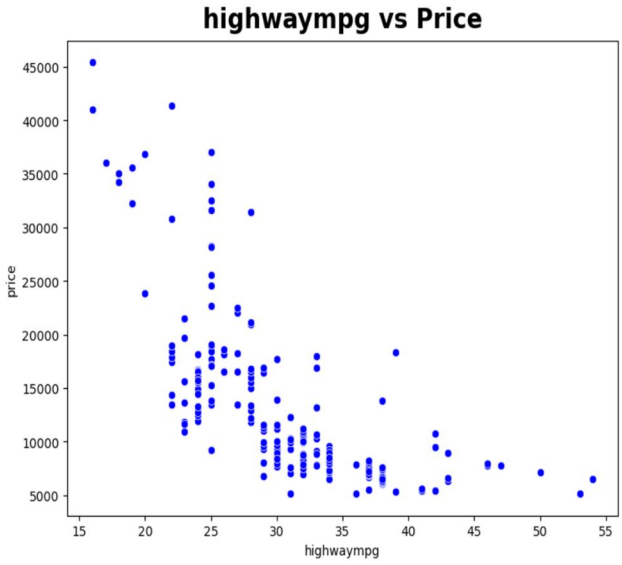


Figure 3: Visualizing "Highwaympg vs Price" Features

Figure 2 & 3 : These scattered plots show the comparison between the prices vs citympg and highwaympg, where both the citympg and highwaympg are dependent variables. The price of the car goes up when the citympg and highwaympg decreases.

A graph type called a scattered plot, or simply scatter plot, shows discrete data points as dots on the two-dimensional coordinate system. It is frequently used to see how two variables relate to one another and to find trends or correlations in the data. With one variable represented on the x-axis and other on the y-axis, each dot on the plot symbolises a single observation.

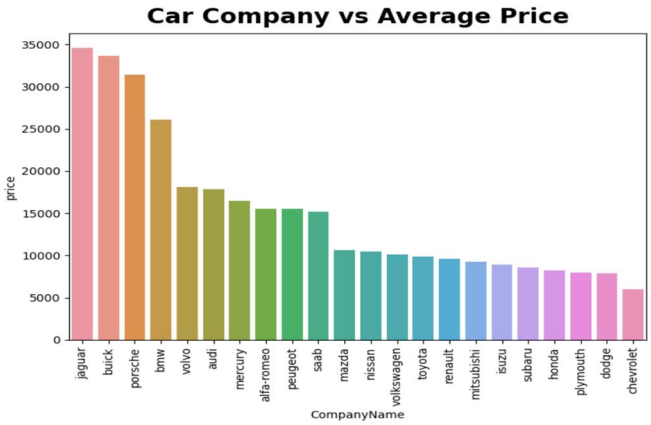


Figure 4: Visualizing Car Company w.r.t Price

Figure 4: The above bar graph shows the price of the car with respect to the car company. It can be seen that the Jaguar company has more average price of the cars compared to other car companies.

The feature significance approach assigns a score to each feature in a feature set according to how well it can predict the target variable. First in the dataset that is supplied, Price is the most significant attribute

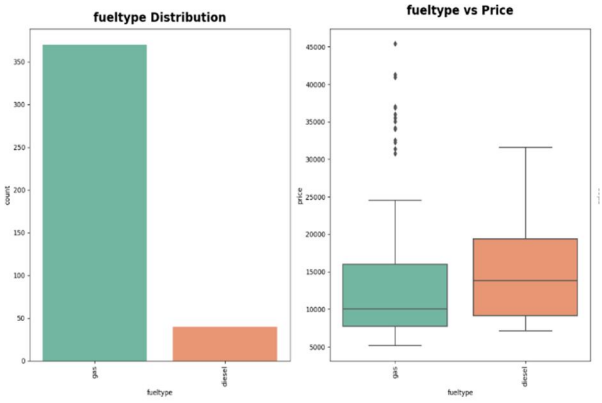


Figure 5: Visualizing Car Fuel Type Feature

Figure 5: The first bar graph shows the relationship between the fuel type and count of the cars. As we can see there are more number of gas cars than diesel cars. The second graph shows the relation of fuel type and price. The cars using gas is less cost than diesel.

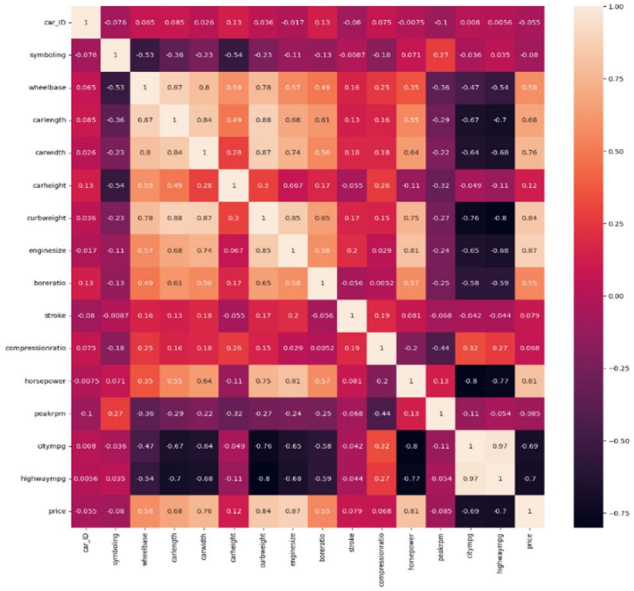
VI. RESULTS

When regression techniques are applied, r2 score, Pearson correlation and Cosine similarity evaluation metrics are obtained. These allowed for a comparison of the effectiveness of each method.

Table 1

Model	R-Square Score	Pearson Correlation	Cosine Similarity
Linear Regression	89.20	0.9464	0.9844
Random Forest	96.32	0.9861	0.9955

The Correlation Features of Heatmap of the Final Dataset's: Correlation characteristics defines the degree of two variables which are linearly related to one another. The variables which are linearly dependent on one another have same influence on the correlation features. If the other two have a significant association, then one of the variables will always be eliminated. Lighter colours indicate low correlation and darker colours indicate strong correlation in the correlation heatmap.



VII. CONCLUSION

In summary, the usage of machine learning to evaluation automobile prices is a big step forward for the automotive sector since it delivers a data-driven method for defining and understanding market values. ML models may yield precise forecasts by analyzing a variety of data, comprising market trends, model specs, mileage, and brand reputation.

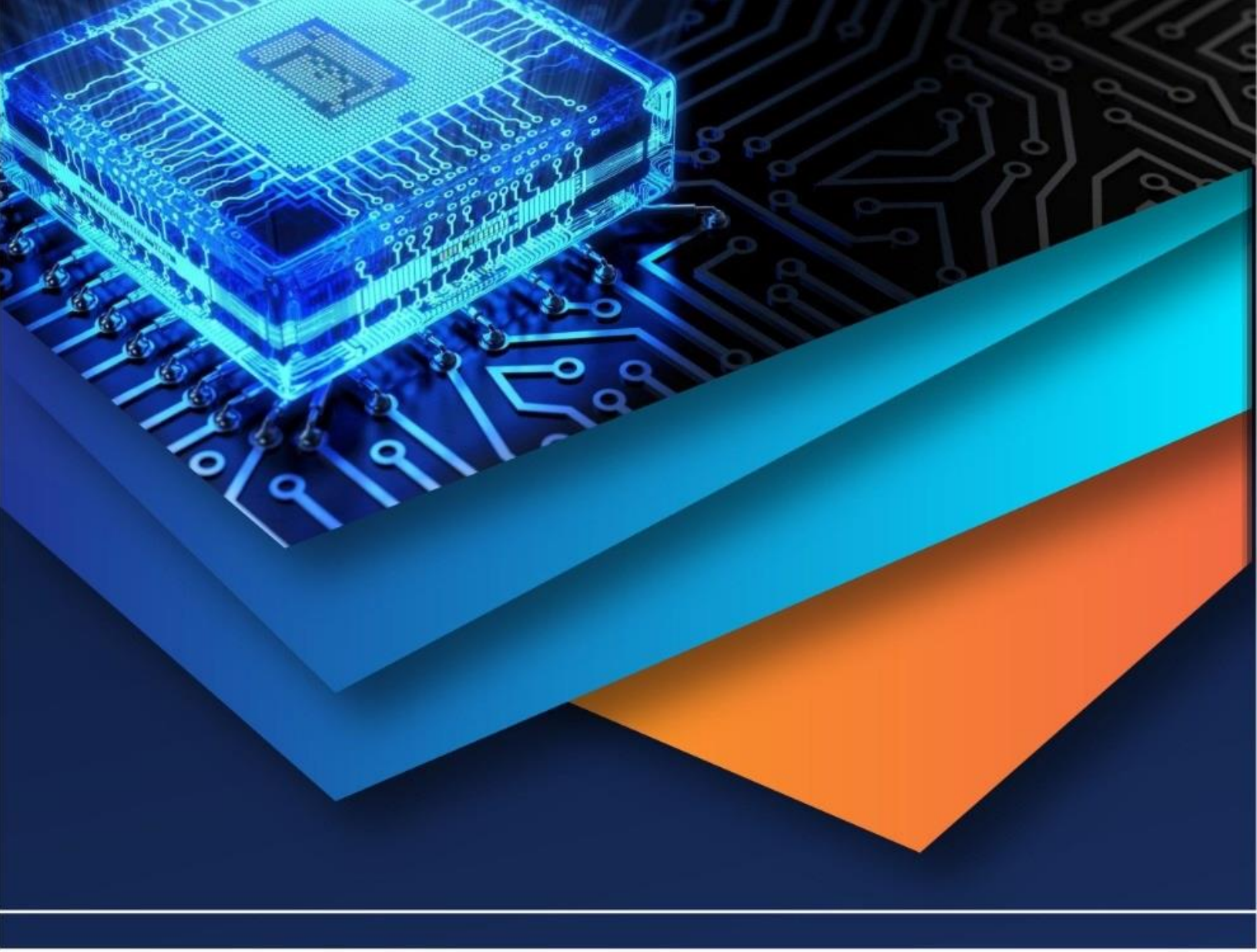
This helps buyers and sellers in making well-versed decisions. One of the main advantages of machine learning- based automobile price prediction is its capacity to adapt to changing market conditions. By following additional approaches, it is difficult to assimilate real-time data as they react slow to the changes occurred in demand and supply. On the other hand, machine learning algorithms are always a step ahead with new information which allows them to deliver up-to-date predictions that resemble to the dominant marketing trends. This elasticity changes the accuracy of pricing estimate and decreases chances of either overpricing or underpricing vehicles. Furthermore, ML models' transparency makes for a more liable car buying process. By considerate what factors impact the price of a car may make customers feel more confident in their purchase decisions. Correspondingly, sellers might increase their chances of pulling in possible clients by keeping low prices that are built on comprehensive research data. But it must be borne in mind that machine learning models are prone to problems too. The accuracy of forecasts depends significantly on how much and well quality data is inputted into the system. Additionally, external factors like economic shifts or unexpected events may lead to additional uncertainty.

VIII. FUTURE SCOPE

In the upcoming years, Machine Learning based automobile pricing prediction will reach greater heights and achieve unparalleled levels of efficiency. Deep Machine Learning and Neural Networks which are sub-field of Machine Learning can easily analyze complex datasets that contains past sales information, economic indicators, and technical advancements. As these models get better by continuously evolving and have the ability to recognize minute correlations or patterns that would be missed by normal techniques. Presence of actual data streams for instance macroeconomic variables, customer preferences or opinions stated in social media develops prediction accuracy. This dynamic procedure makes ML-based automobile predicting systems more robust and approachable by allowing them to adjust to rapidly changing market necessities. Besides, another layer can be added to price forecasting with development of electric and driverless cars. Machine learning algorithms for battery technology enhancements, modifications in regulations or competitors entering a market can make more complete predictions and integrates the possible effects of international politics, environmental laws, and global economics.

REFERENCES

[1] "A Comparative Study of Car Price Prediction Using Different Regression Techniques" by Mishra et al. (2018)
[2] "Predicting Car Prices Using Machine Learning Techniques" by Liang and Zhang (2017)
[3] "Car Price Prediction Using Data Mining Techniques" by Sahu and Patra (2019)
[4] Pudaruth, S. (2018) 'Predicting the Price of Used Cars using Machine Learning Techniques', International Journal of Information & Computation Technology, 4(7), pp. 753–764. Available at: http://www.irphouse.com.
[5] Kuiper, S. (2018) 'Introduction to Multiple Regression: How Much Is Your Car Worth?', Journal of Statistics Education, 16(3). doi: 10.1080/10691898.2008.11889579.
[6] Pal, N. et al. (2019) 'How Much is my car worth? A methodology for predicting used cars' prices using random forest', Advances in Intelligent Systems and Computing, 886, pp. 413–422. doi: 10.1007/978-3- 030-03402-3_28.
[7] Gegic, E. et al. (2019) 'Car price prediction using machine learning techniques', TEM Journal, 8(1), pp. 113–118. doi: 10.18421/TEM81-16.
[8] Dholiya, M. et al. (2019) 'Automobile Resale System Using Machine Learning', International Research Journal of Engineering and Technology(IRJET), 6(4), pp. 3122–3125.
[9] "Chronic Kidney Disease Risk Prediction Using Machine Learning Techniques. Journal" by D, B., K, C., Prasad, M. L., N, P., Kiran, A., N, P., & Shaker Reddy, P. C. (2024).
[10] Listiani, M. (2020) Support Vector Regression Analysis for Price Prediction in a Car Leasing App.
[11] Richardson, M. (2019) Determinants of Used Car Resale Value. The Colorado College
[12] "Car Price Prediction: A Hybrid Machine Learning Approach" by Patel et al. (2021)
[13] "An Accurate Prediction of Coronary Heart Disease Using Ensemble Algorithms" by K. P. Joshi, M. L. Prasad, R. Natchadalingam, P. C. S. Reddy, S. Mukherjee and G. C. Babu
[14] "Car Price Prediction Using Deep Learning" by Prakash et al. (2020)



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY