IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, November 2024

An Analytical Study on Car Price Prediction

Pratiksha V Sonde[1] and Dr. Manjunath N[2]
Student, Department of MBA[1]
Assistant Professor, Department of MBA[1]
RNS Institute of Technology, Bengaluru, Karnataka, India

Abstract: *This study on car price prediction explores the important role of accurate price forecasting in the growing automotive market, particularly in India. Precise predictions benefit for stakeholders such as buyers, sellers, and dealerships. The research examines key variables influencing car prices, including brand reputation, vehicle specifications, market trends, and economic factors. Utilizing machine learning techniques like multiple and logistic regression, the study analyses approximately 5,000 car records to determine the impact of various predictors. Incorporating both domestic and global insights, it addresses factors such as fuel type, engine size, and emerging technologies like electric vehicles. The study emphasizes the importance of developing accurate prediction models to support informed decision-making and suggests enhancing model accuracy by integrating regional market trends and advanced hybrid models to capture complex pricing dynamics.*

Keywords: Car price prediction, Automotive market, Market trends, Economic factors, Used cars, new cars, Consumer preference, Multiple and Logistic regression analysis

I. INTRODUCTION

The automotive industry is an important driver of the global economy, and car pricing has a significant impact on consumer decisions, market trends, and business strategies. Accurate car price forecast are essential for consumers, sellers, dealerships, and finance institutions to make informed choices. This study focuses on forecasting automobile prices in India's rapidly expanding automotive market, where both new and second-hand vehicles are in high demand. The study combines machine learning techniques, such as multiple and logistic regression models, to examine major aspects impacting car prices, such as brand reputation, fuel type, engine size, vehicle condition, and market dynamics. This study intends to increase price forecast accuracy while also providing practical insights into market behaviour. The findings highlight the need of integrating regional trends with advanced modelling techniques to improve the precision and reliability of cars price forecasts.

OBJECTIVES

- To understand the essentials of car price prediction and its relevance
- To identify the factor influencing car price prediction
- To analyse the impact of Length, MPG Highway, MPG City, Horsepower, Wheelbase, Cylinders, Engine Size on MSRP
- To Analyse the impact of Fuel Type, Selling Price and Kms Driven on car price prediction which may be classified as sale by individual and sale by dealer.

II. LITERATURE REVIEW

In their 2023 study, Sumeyra Muti and Kazim Yildiz explore the process of removing illogical values from a dataset to enhance the model's accuracy in estimating actual and predicted values for car prices. They analyzed variables such as model, year, mileage, gear type, damage history, and color to estimate the selling price of used vehicles using machine learning techniques. Their research underscores the importance of accurate pricing to ensure fair market values for second-hand cars. The study finds that linear regression (LR) is an effective model for predicting car prices, although additional data and preprocessing could further improve its accuracy.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, November 2024

According to Enis Getic, Zerina Masetic, Jasmin Kevric(2019), shows that merging multiple machine learning classifiers improves forecast accuracy for car prices. The final model was integrated into a user-friendly Java application for practical use, and this explore provides valuable information into how machine learning can be effectively applicable to predict car prices, taking into account the complexity and variety of influencing factors like brand, model, age, horsepower, mileage, exterior colour, type of transmission, safety, etc.

V.C. Sanap, Mohammed Munawwar Rangila(2022), the researchers' goal is to deliver a realistic assessment of used cars values to help buyers and sellers save time and money. Variables such as car company, model, year of purchase, fuel type, and number of kilometres driven are used. The study underlines the need of managing massive, complicated datasets and generating reliable results regardless of data volume.

III. RESEARCHER METHODOLOGY

The descriptive methodology examines key factors affecting car prices, including variables like fuel type, mileage, brand, and transmission, using a sample of over 5000 cars from Indian and international markets. Data analysis employs descriptive statistics and regression in Excel and SPSS, identifying trends and relationships crucial for accurate price prediction.

SAMPLING

The study uses simple random sampling to ensure a representative subset of the overall population of 5,000+ car records. This dataset includes both Indian and foreign car brands.

STATISTICAL TOOLS

- **Microsoft Excel:** Data organization, visualization, and basic statistical analysis, such as creating charts and calculating descriptive statistics.
- **SPSS Software:** Multiple regression and logistic regression analysis to examine the relationships between independent variables and dependent variables.

HYPOTHESIS TEST

- **Null Hypothesis (H_0)**– Length, MPG Highway, MPG City, Horsepower, Wheelbase, Cylinders, Engine Size do not influence MSRP.
- **Alternative Hypothesis (H_1)**– Length, MPG Highway, MPG City, Horsepower, Wheelbase, Cylinders, Engine Size will influence MSRP.

ANNOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	181.057	7	25.865	71.184	.000[b]
	Residual	151.884	418	.363		
	Total	332.941	425			

a. Dependent Variable: MSRP

b. Predictors: (Constant), Length, MPG City, MPG Highway, Horsepower, Wheelbase, Cylinders, Engine Size.

Inference: As sig value of 0.000 is less than the standard P value of 0.05, we reject H_0 and conclude that Length, MPG City, MPG Highway, Horsepower, Wheelbase, Cylinders, Engine Size will Influence MSRP.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, November 2024

Coefficients[a]

Model		B	Std. Error	Standardized Coefficients Beta	t	Sig.	Results
	(Constant)	.285	.357		.798	.425	-
	Engine Size	-.056	.065	-.071	-.860	.390	Insignificant
	Cylinders	.507	.093	.425	5.421	.000	Significant
	Horsepower	.556	.064	.434	8.712	.000	Significant
	MPG City	.136	.091	.063	1.504	.133	Insignificant
	MPG Highway	-.204	.084	-.104	-2.443	.015	Significant
	Wheelbase	-.068	.051	-.065	-1.333	.183	Insignificant
	Length	-.098	.108	-.038	-.843	.400	Insignificant

a. **Dependent Variable:** MSRP

Inference: From the above coefficient table, it is the evident that as far as determining the MSRP is concerned, the sig. value of Cylinder is .000, sig. value of Horsepower is .000, sig. value of MPG Highway is 0.183. The Sig value of these variables is less than .000 so these are Significant Variables.

Logistic Regression:

Null Hypothesis (H_0) - Kms Driven, Fuel Type, and Selling Price do not influence the profile of the sales person. Which is classified as Individual and dealer.

Alternative Hypothesis (H_1) - Kms Driven, Fuel Type and Selling Price will influence the profile of the sales person. Which is classified as Individual and Dealer.

Classification Table[a,b]

Predicted					
Seller Type					Percentage Correct
Observed			0	1	
Step 0	Seller Type	0	0	1281	.0
		1	0	5564	100.0
	Overall Percentage				81.3

a. Constant is included in the model

b. The cut value is .500

Variables In the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	1.469	.031	2246.027	1	.000	4.343

Inference: As the Sig. value of 0.000 is less than the standard P value of 0.05, we reject H_0 and conclude that Kms Driven, Fuel Type, Selling Price all together will influence the profile of the sales person. Which is classified as Individual and Dealer.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, November 2024

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	Result
Step 1[a]	Kms Driven	.597	.051	136.706	1	.000	1.817	Significant
	Fuel Type	-.014	.069	.039	1	.843	.986	Insignificant
	Selling Price	-1.844	.111	275.804	1	.000	.158	Significant
	Constant	2.535	.151	281.451	1	.000	12.621	-

Variable(s) entered on step 1: Kms Driven, Fuel Type, Selling Price

Inference: The following independent variables influences the profile of the sales person. Which is classified as Individual and Dealer. The Kms Driven Sig. Value is .000. Selling Price Sig. Value is 0 .000. The Sig value of these variables is less than .000 so these are Significant Variables.

IV. DISCUSSION AND RESULTS

The use of descriptive statistics, Microsoft excel and SPSS software provides insights into making informed decision about car price and accuracy of predicting car prices. The analysis reveals that fuel type, transmission type, and mileage significantly impact car prices, with diesel and automatic transmission vehicles typically valued higher. Popular models, such as Maruti Swift Dzire, show lower depreciation, suggesting brand reputation plays a key role in resale value. Additionally, the regression analysis confirms that engine size, horsepower, and seller type are strong predictors of car price, enhancing model accuracy for price estimation. Here key findings and implication are discussed in more details.

V. KEY FINDINGS

Positive Impact

Market Transparency: By identifying key factors like brand, mileage, and fuel type, the research promotes transparency in the used car market, helping both buyers and sellers assess fair values and avoid price manipulation.

Improved Pricing Models: The use of multiple and logistic regression models improves the accuracy of price forecasting, making it easier to predict car prices based on a range of variables and boosting model reliability.

Negative Impact

Market Disruptions: External factors like economic downturns, supply chain disruptions, and changes in fuel prices can drastically alter car prices, making it difficult for the model to remain accurate over time.

Inadequate Handling of Outliers: Outlier data points, such as luxury cars or vehicles with unusual features, may distort predictions. If not properly managed, these outliers can skew the results, making the model less robust for mid-range or common car models.

VI. CONCLUSION

This study on car price prediction highlights the critical role of accurate forecasting in the automotive industry, particularly for stakeholders such as buyers, sellers, and dealerships. By analysing key factors like brand, fuel type, engine size, and market trends, the research provides valuable insights into the dynamics influencing car prices. The use of machine learning techniques, including regression models, improves prediction accuracy. However, limitations such as regional variability, economic fluctuations, and data constraints must be addressed to enhance model robustness. Overall, the study underscores the importance of data-driven approaches for informed decision-making in the automotive market.

LIMITATIONS

- Reliable and diverse data may be difficult to source, especially for underrepresented regions or vehicle types.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, November 2024

- Rapid changes in market conditions, such as supply chain disruptions or regulatory changes, may render the model's predictions outdated.
- Emerging automotive technologies like electric vehicles or autonomous driving may not be fully accounted for in traditional models.

REFERENCES

[1]. Muti, S., & Yildiz, K. (2023). Using linear regression for used car price prediction. International Journal of Computational and Experimental Science and Engineering,9(2)149-153.https://dergipark.org.tr/en/pub/ijcesen/article/1070505

[2]. Gegic, E., &Kevric, J. (2019). Car price prediction using machine learning techniques. Central and Eastern European Online Library.

[3]. https://www.ceeol.com/search/article-detail?id=746689