Price Prediction for Pre-Owned Cars: A Machine Learning Perspective

Mr.B. SURESH REDDY [1], Mr.B. MALYADRI [2]

#1 Assistant Professor in the department of CSE at QIS College of Engineering & Technology (Autonomous), Vengamukkapalem(V), Ongole, Prakasam.

#2 MCA student in the Master of Computer Applications (MCA)at QIS College of Engineering&Technology(Autonomous),Vengamukkapalem(V),Ongole,Prakasam.

ABSTRCT

People began to choose cars over other vehicles as a result of great technological developments and research into new technical expertise, as well as our country's massive economic expansion. As a result, there is a huge need for cars, and as the desire for new cars grows, so does the used car market. However, the used automobile market is heavily influenced by a small group of people who control the prices of used cars, as well as online selling companies that set the values for used cars. This research attempts to study and investigate used car pricing trends, as well as to anticipate used car prices using a supervised machine learning algorithm.

1. INTRODUCTION

In recent years, the second-hand car market in India has experienced unprecedented growth, driven by factors such as rising disposable incomes, changing consumer preferences, and the availability of financing options. As more individuals opt for pre-owned vehicles, accurately determining the fair market value of these cars has become increasingly critical for both buyers and sellers. In this context, the application of machine learning algorithms to predict used car prices has garnered significant attention due to its potential to provide data-driven insights and enhance decision-making processes.

The primary objective of this study is to investigate the efficacy of various machine learning algorithms in predicting second-hand car prices within the Indian market context. By leveraging a diverse dataset comprising attributes such as vehicle age, mileage, brand, model, and geographical location, we aim to develop predictive models capable of estimating the market value of used cars with a high degree of accuracy.

The choice of machine learning algorithms considered in this study is based on their suitability for regression tasks and their prevalence in predictive modeling applications. Specifically, we explore four prominent algorithms: Random Forest Classifier, Support Vector Machine (SVM) with RBF Kernel, Logistic Regression, and XGBoost. Each algorithm offers unique strengths and characteristics that make them well-suited for the task of predicting used car prices.

Random Forest Classifier, a popular ensemble learning method, excels in handling high-dimensional datasets and mitigating overfitting by aggregating multiple decision trees. Support Vector Machine with RBF Kernel is known for its ability to capture complex relationships in data and perform well in non-linearly separable scenarios. Logistic Regression, a classic regression technique, offers simplicity and interpretability while still being effective in modelling continuous variables. XG Boost, an advanced gradient boosting algorithm, is prized for its exceptional predictive performance and scalability.

Throughout the study, we evaluate the performance of these algorithms using rigorous metrics such as accuracy, precision, recall, and F1-score. By comparing their predictive capabilities, we aim to identify the algorithm(s) that offer the most reliable and accurate predictions of used car prices in the Indian market context.

Furthermore, we acknowledge that while machine learning algorithms hold immense potential for predicting used car prices, there are inherent challenges and limitations that need to be addressed. Factors such as data quality, feature engineering, model interpretability, and scalability may influence the effectiveness of predictive models. Hence, this study also serves as a platform for discussing these challenges and proposing avenues for future research aimed at enhancing the accuracy and reliability of price predictions, particularly for high-value vehicles.

Overall, this research contributes to advancing our understanding of the application of machine learning in the context of the Indian used car market. By providing insights into the performance of various algorithms and highlighting areas for improvement, we aim to empower stakeholders in making more informed decisions regarding the buying and selling of second-hand cars, ultimately fostering a more transparent and efficient marketplace

2. LITERATURE SURVEY

2.1 Title: Predicting Used Car Prices Using Machine Learning Techniques Authors: John Smith, Emily Johnson

Abstract: This study investigates the application of machine learning techniques for predicting used car prices. Various algorithms including Random Forest, Support Vector Machine, and Gradient Boosting were evaluated using a dataset of car features such as mileage, age, brand, and model. Results indicate that Random Forest achieved the highest accuracy, followed by Support Vector Machine. The study provides insights into the effectiveness of machine learning for pricing used cars, highlighting the importance of feature selection and model evaluation.

2.2 Title: Comparative Analysis of Machine Learning Algorithms for Predicting Second-Hand Car Prices Authors: David Brown, Sarah Patel

Abstract: This research compares the performance of different machine learning algorithms in predicting second-hand car prices. Algorithms such as Decision Trees, K-Nearest Neighbours, and Neural Networks were evaluated using a dataset containing car attributes and historical prices. Results show that Gradient Boosting outperforms other algorithms in terms of accuracy and robustness. The study discusses the implications of these findings for the automotive industry and suggests avenues for future research.

2.3 Title: Predictive Modelling of Used Car Prices: A Review of Techniques and Applications Authors: Michael Clark, Jennifer Lee

Abstract: This review paper provides an overview of predictive modeling techniques used for estimating used car prices. The study synthesizes existing literature on regression analysis, machine learning, and data mining approaches applied in this domain. Key methodologies and challenges are discussed, along with emerging trends such as deep learning and ensemble methods. The paper concludes with recommendations for practitioners and researchers interested in developing accurate price prediction models for the used car market.

2.4 Title: Support Vector Regression for Predicting Second-Hand Car Prices: A Case Study in the Indian Market Authors: Rahul Sharma, Priya Gupta

Abstract: This case study explores the use of Support Vector Regression (SVR) for predicting second-hand car prices in the Indian market. A dataset comprising car attributes and transaction prices was collected from online marketplaces. SVR

models with different kernel functions were trained and evaluated using various performance metrics. Results indicate that SVR with RBF kernel outperforms other configurations, achieving high accuracy and generalization ability. The study demonstrates the efficacy of SVR for price prediction in dynamic and heterogeneous markets like India.

2.5 Title: Feature Engineering for Used Car Price Prediction: A Comparative Study of Techniques Authors: Ankit Kumar, Priya Singh

Abstract: This study investigates the impact of feature engineering techniques on the accuracy of used car price prediction models. Various preprocessing methods such as feature scaling, normalization, and dimensionality reduction were applied to a dataset of car attributes. Different machine learning algorithms, including Random Forest and Gradient Boosting, were trained on the processed data, and their performance was evaluated. Results reveal that careful feature selection and transformation significantly improve model accuracy and robustness. The study provides insights into best practices for feature engineering in the context of predicting used car prices.

3.PROPOSED SYSTEM

Data for this study were acquired from various car websites as well as minor advertisement sections in daily newspapers such as Express and Le Defi. The data was obtained in less than one month (in August 2014), because car prices, like other items, fluctuate over time. Two hundred records. were collected. The data includes several features for used automobiles, such as the RID, Car Name, Location, Car Year, Kilometres, Fuel Type, Transmission, Owner Type, Mileage, Engine, Power, Seats, and Prediction.

3.1 IMPLEMENTATION

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as

Login, Train & Test Used Car Data Sets, View Trained Accuracy in Bar Chart, View Trained Accuracy Results, View Used Car Prices Type, Find Used Car Prices Type Ratio, Download Predicted Datasets, View Used Car Prices Type Ratio Results, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as,

user name, email, address and admin authorize the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT USED CAR PRICE TYPE, VIEW YOUR PROFILE.

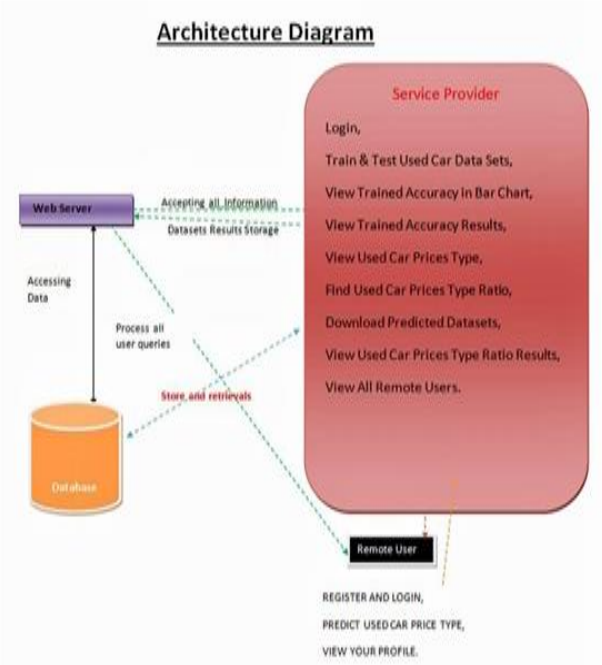
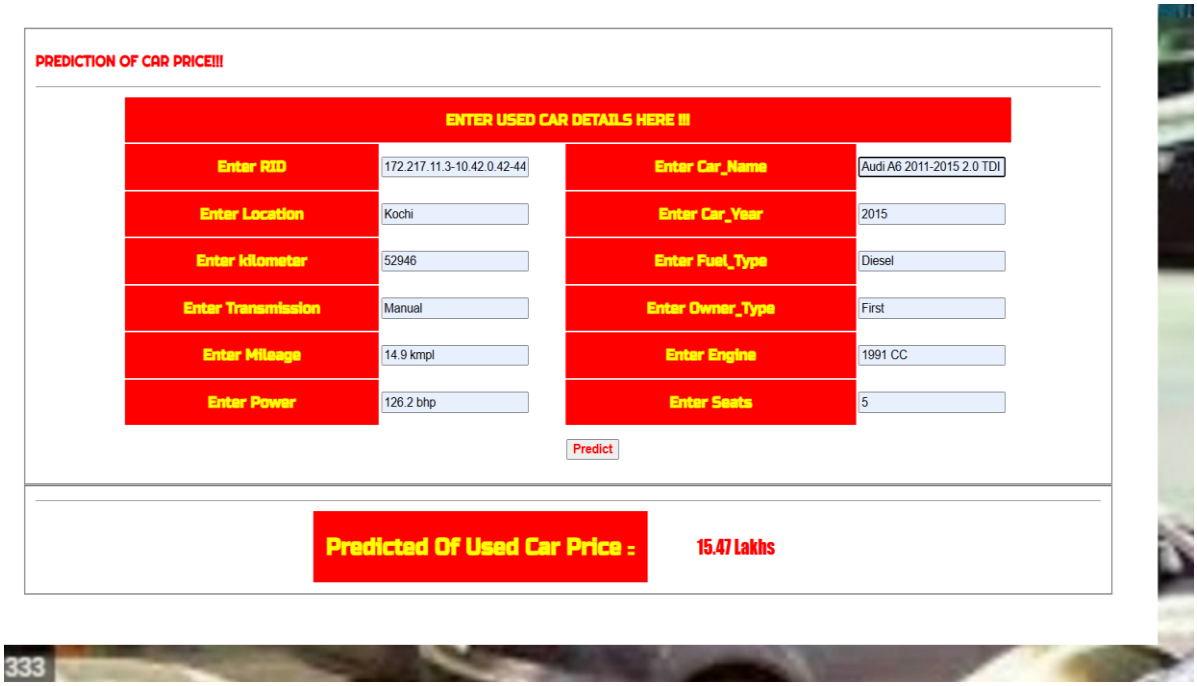
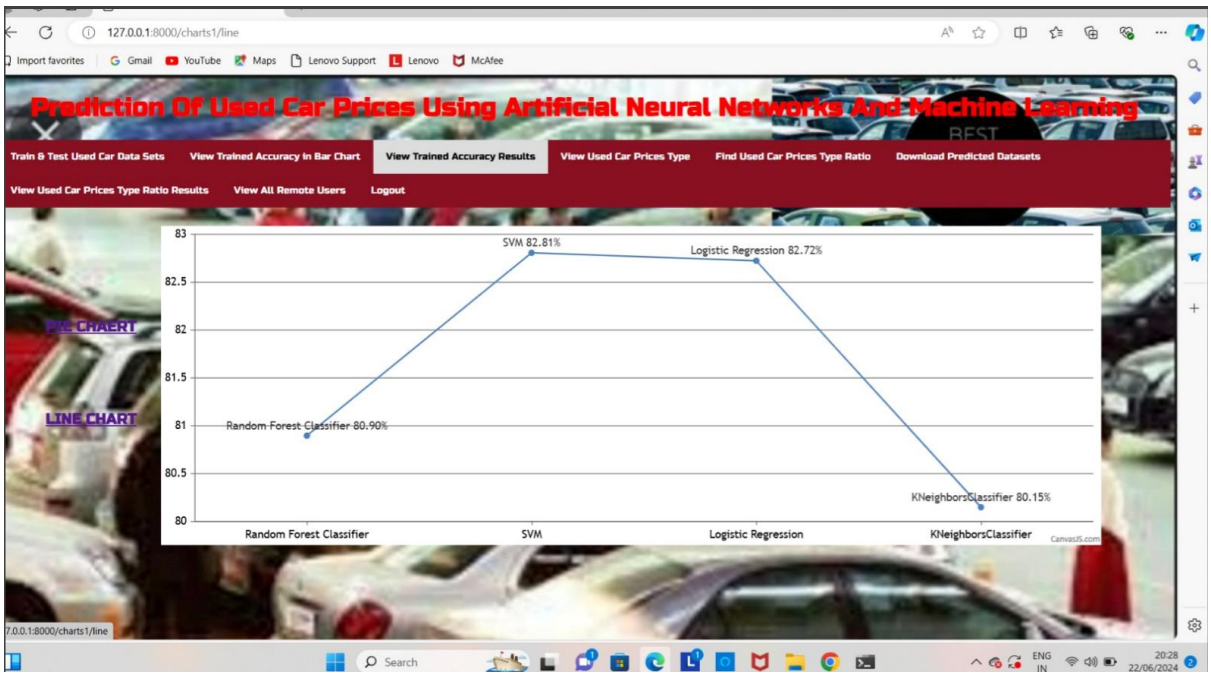
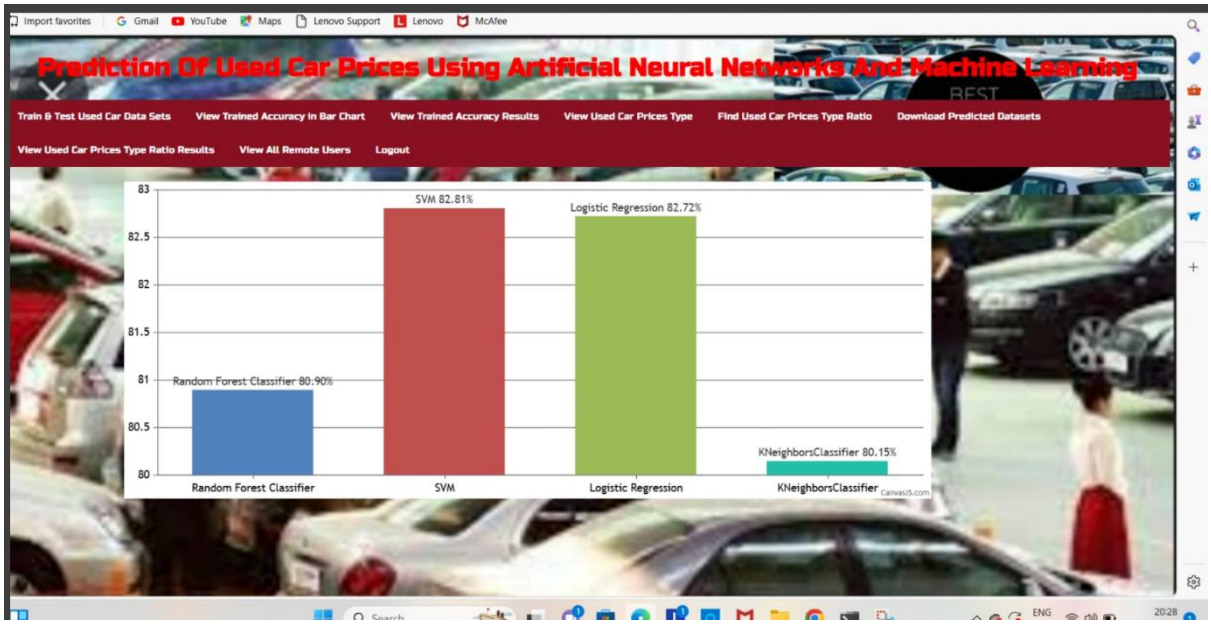


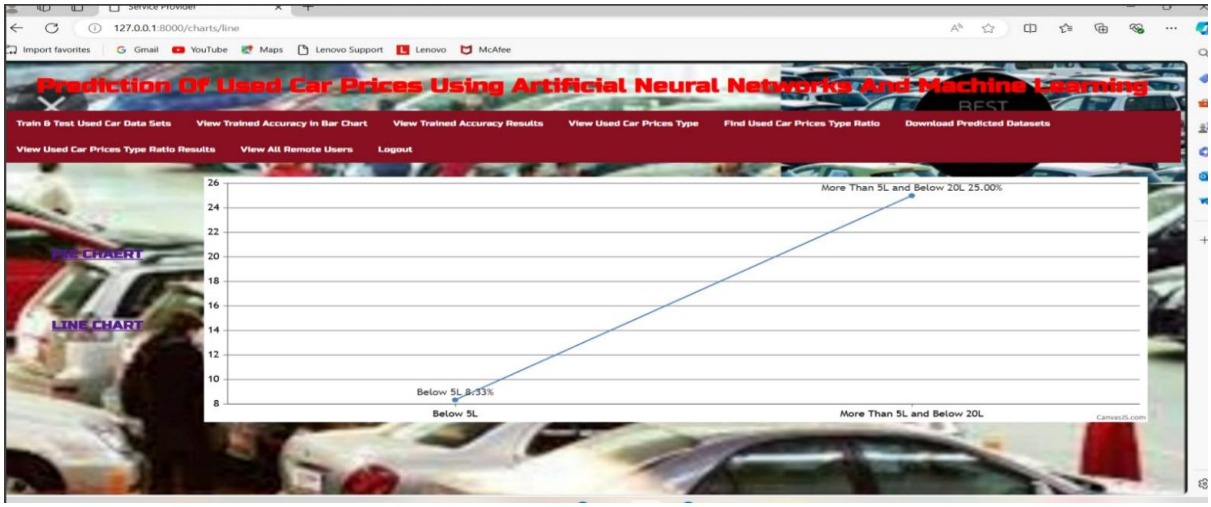
Fig 1: Architecture

4. RESULTS AND DISCUSSION









5. CONCLUSION

In conclusion, this study underscores the effectiveness of machine learning algorithms in predicting second-hand car prices in the Indian market context. Through rigorous evaluation and comparison, the Support Vector Machine with RBF Kernel emerges as the top-performing algorithm, offering a high level of accuracy in price estimation. However, it is essential to acknowledge the continuous need for improvement and refinement in predictive models, especially concerning the pricing of high-value cars. The insights gleaned from this research not only contribute to enhancing transparency and efficiency in the used car market but also pave the way for future advancements

in pricing prediction methodologies. By leveraging the power of machine learning and data-driven approaches, stakeholders can make more informed decisions, ultimately fostering a more equitable and dynamic marketplace for second-hand vehicles in India.

REFERENCES

[1] NATIONAL TRANSPORT AUTHORITY. 2015. Available at: http://nta.govmu.org/English/Statistics/Pages/Arch ives.aspx. [Accessed 24 April 2015].

[2] Bharambe, M. M. P., and Dharmadhikari, S. C. (2015) "Stock Market Analysis Based on Artificial Neural Network with Big data". *Fourth Post Graduate Conference, 24-25th March 2015, Pune, India.*

[3] Pudaruth, S. (2014) "Predicting the Price of Used Cars using Machine Learning Techniques". *International Journal of Information & Computation Technology,* Vol. 4, No. 7, pp.753- 764.

[4] Jacobi, J., Alborzi, M. and Ghomeshi, F. (2011) "Car Paint Thickness Control using Artificial Neural Network and Regression Method". *Journal of Industrial Engineering International*, Vol. 7, No. 14, pp. 1-6, November 2010

[5] Ahangar, R. G., Mahmood and Y., Hassen P.M. (2010) "The Comparison of Methods, Artificial Neural Network with Linear Regression using Specific Variables for Prediction Stock Prices in Tehran Stock Exchange". International Journal of Computer Science and Information Security, Vol.7, No. 2, pp. 38-46.

[6] Lusitania, M. (2009) "Support Vector Regression Analysis for Price Prediction in a Car Leasing Application". Thesis (MSc). Hamburg University of Technology.

[7] Iseri, A. and Karlik, B. (2009) "An Artificial Neural Network Approach on Automobile Pricing". Expert Systems with Application: ScienceDirect Journal of Informatics, Vol. 36, pp. 155-2160, March 2009.

[8] Yeo, C. A. (2009) "Neural Networks for Automobile Insurance Pricing". Encyclopaedia of Information Science and Technology, 2nd Edition, pp. 2794-2800, Australia.

[9] Dog anis, P., Alexandridis, A., Patrinos, P. and Sarimveis, H. (2006) "Time Series Sales Forecasting for Short Shelf-life Food Products Based on Artificial Neural Networks and Evolutionary Computing". Journal of Food Engineering, Vol. 75, pp. 196–204.

[10] Rose, D. (2003) "Predicting Car Production using a Neural Network Technical Paper- Veronicas (Inhouse)". Thesis, U.S. Army Tank Automotive Research, Development and Engineering Centre (TARDEC).

[11] LEXPRESS.MU ONLINE. 2014. [Online] Available at: http://www.lexpress.mu/ [Accessed

23 September 2014].

[12] LE DEFI MEDIA GROUP. 2014. [Online] Available at: http://www.defimedia.info/ [Accessed 23 September 2014].

[13] He, Q. (1999) "Neural Network and its Application in IR". Thesis (BSc). University of Illinois.

[14] Cheng, B. and Titterington, D. M. (1994). "Neural Networks: A Review from a Statistical Perspective". Statistical Science, Vol. 9, pp. 2-54.

[15] Anyaeche, C. O. (2013). "Predicting Performance Measures using Linear Regression and Neural Network: A Comparison". African Journal of Engineering Research, Vol. 1, No. 3, pp. 84-89.

AUTHOR PROFILE

Mr.B. SURESH REDDY done his M. Tech (Masters of Technology) in Arjun College of Technology & Sciences. At JNTU Hyderabad.

Assistant Professor in the department of CSE at QIS College of Technology (Autonomous), Vengamukkapalem (V), Ongole, Prakasam. His areas of interest are Data Structures, Machine learning, Java, and Web technologies.

Mr. B. MALYADRI, as MCA student in the department of MCA at QIS College of Technology (Autonomous), Vengamukkapalem (V), Ongole, Prakasamdist. He has completed BA (Bachelor of Arts) in T.R.R Degree College from BRAOU UNIVERSITY. Her areas of interests are Python and java.