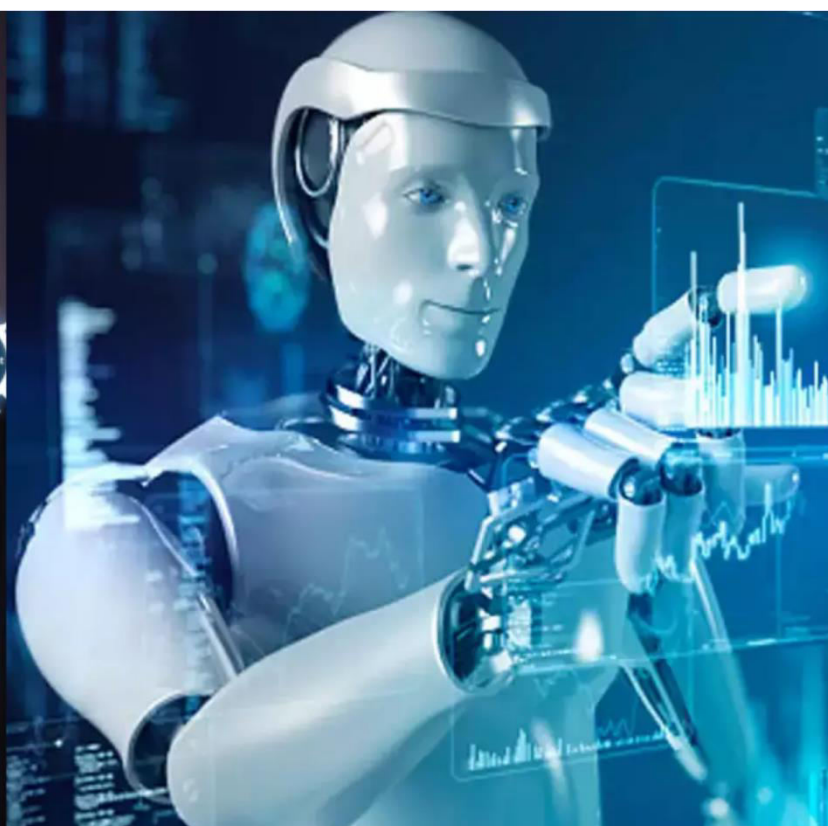




International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Auto Insights: Predicting Second-Hand Car Prices with Machine Learning

Dr. L.Nalini Joseph, Illandula Naveen, Jakkam Sravan, Jakkoju Mano Tej, Jedipally Sandeep Reddy

Head of the Department, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai,
Tamil Nadu, India

B. Tech Students, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, Tamil Nadu, India

ABSTRACT: This project is to create a precise car price prediction system with machine learning to assist buyers and sellers in establishing fair market prices. Conventional valuation techniques tend to be imprecise because there are various factors affecting them such as brand, mileage, fuel type, and transmission. The system uses gradient boosting for enhanced prediction accuracy by training the model on a dataset of historical car data. The model is implemented with Streamlit, where users can enter car specifications and get real-time price quotes. This computerized method increases transparency and accuracy in vehicle pricing to the advantage of consumers and the automotive sector.

KEYWORDS Machine Learning, XGBoost, Car Price Prediction, Data Preprocessing, Feature Engineering, Model Training, Model Evaluation, Web Application, Flask, Streamlit, Pandas, NumPy, Scikit-Learn

I. INTRODUCTION

The Car Price Estimation System is an application based on machine learning that predicts the resale price of pre-owned cars based on attributes like brand, model, mileage, fuel type, transmission type, and year of manufacturing. The conventional practices for estimating car prices involve manual evaluation, simple regression models, or user-submitted price data, which are prone to inconsistencies and errors. The objective of this project is to offer data-driven, automatic, and real-time price estimates through sophisticated machine learning methods.

The system uses XGBoost, a high-performance gradient boosting algorithm, that learns from historical sales data and becomes more accurate with time. Techniques of data preprocessing, feature engineering, and model evaluation are used to improve performance. The trained model is deployed in a Flask/Streamlit-based web application where users can input car specifications and get price predictions instantly. Through the use of AI and predictive analytics, this system favors car buyers, sellers, dealerships, and banks by providing transparent, fair, and efficient automobile market pricing decisions.

The project adheres to a well-organized machine learning pipeline, right from data fetching and preprocessing to model training, testing, and deployment. The Exploratory Data Analysis (EDA) is conducted to ascertain the prominent factors affecting car prices, including car age, mileage, and type of fuel. The model is trained on XGBoost as it is chosen for its good accuracy, high efficiency, and capability to tackle complex datasets. Following training, the model is run as a Streamlit web app, so that users can input car specifications and receive real-time price estimates. Future developments include adding real-time market trends, further development of location-based price adjustments, and creating a mobile app to improve usability and precision even more.

II. PROBLEM DEFINATION

It is a difficult and complex problem to determine the resale price of a used vehicle because various factors like brand, model, manufacturing year, mileage, fuel type, transmission, and market demand affect it. All current car valuation techniques are based on manual judgments, self-reported prices, and simple regression models, which result in inconsistent, inaccurate, and biased estimations. The absence of a standardized and data-based pricing method leads to mispricing, with the sellers undervaluing their vehicles and the buyers overpaying.

Market volatilities, regional demand changes, and seasonal price trends are also not accounted for in most current

pricing methods, which introduces more inaccuracies. The fact that there is no real-time, automated, and AI-driven prediction system renders it challenging for buyers, sellers, and dealerships to estimate a fair market price. The Car Price Prediction System is designed to address these problems through the application of machine learning algorithms, specifically XGBoost, to offer precise, transparent, and data-driven car price predictions. Through the utilization of historical sales data and predictive analytics, this system guarantees that prices are fair, reliable, and unbiased

III. RELATED WORK

A number of studies and applications in the industry have investigated applying machine learning algorithms to predict car prices. Manual pricing techniques, simple regression models, and expert judgment were used traditionally, but they produced inconsistent and inaccurate values. With improvements in data science and artificial intelligence, researchers have made data-driven techniques that use past sales data and predictive modeling to enhance accuracy.

Past research has used models like Linear Regression, Decision Trees, Random Forest, and Support Vector Machines (SVM) for estimating car prices. The above models have shortcomings in managing intricate, non-linear interactions between car attributes and prices. Current research has indicated that ensemble learning methods like Gradient Boosting and XGBoost perform better than conventional models because they can make predictions iteratively and minimize errors efficiently. Applications such as Kelley Blue Book (KBB), Edmunds, and CarDekho leverage machine learning-based pricing models to provide car valuations based on actual sales data, emphasizing how useful AI-powered price systems are.

Emulating such developments, this project applies XGBoost for the prediction of prices since it offers high accuracy, quicker processing, and improved management of missing values. In contrast to the earlier methods, this system also incorporates feature engineering methods and delves into real-time deployment through a Flask/Streamlit-based web interface, rendering it more user-friendly. Future updates, including real-time market data incorporation and deep learning-driven modeling, will make it even more accurate and flexible, making this system a holistic and smart car valuation system

IV. PROPOSED SYSTEM

Car Price Prediction System adopts a systematic machine learning approach to provide high-precision data-driven pricing predictions. The approach includes multiple steps ranging from data collection to model deployment, which ensure a well-structured and scalable prediction system.

It begins with data collection where past car sale history data is collected from online car resale websites, dealership databases, and public resources. The dataset has important features like brand, model, year, mileage, fuel type, transmission type, and sale price. After collection, the data is preprocessed and that involves missing values handling, categorical variable encoding, outliers removal, and feature engineering (e.g., determining car age from the manufacturing year). This makes the dataset clean and ready for machine learning.

Once preprocessed, Exploratory Data Analysis (EDA) is performed to find the trends and relationships between various car features and their effect on the price. Visualization in the form of scatter plots, histograms, and heatmaps assists in grasping the relationship between features such as mileage and depreciation rate. The data is then divided into training and testing sets so that the model generalizes well for unseen data.

The machine learning model is now trained and tuned. Several algorithms, such as Linear Regression, Decision Trees, Random Forest, and XGBoost, are compared in terms of performance, and the best model is chosen as XGBoost based on its high accuracy and effectiveness in dealing with structured data. Hyperparameter tuning is carried out to enhance the performance of the model and reduce errors. The model is tested with metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score to determine prediction accuracy.

After training and testing the model, it is saved and deployed as a web application based on Flask/Streamlit, where users can input car specifications and get immediate price predictions. The system is deployed on cloud servers such as Heroku or AWS, making it accessible to many users. Future improvements involve incorporating real-time market information, using deep learning models, and scaling the system to a mobile application for enhanced scalability and

user experience

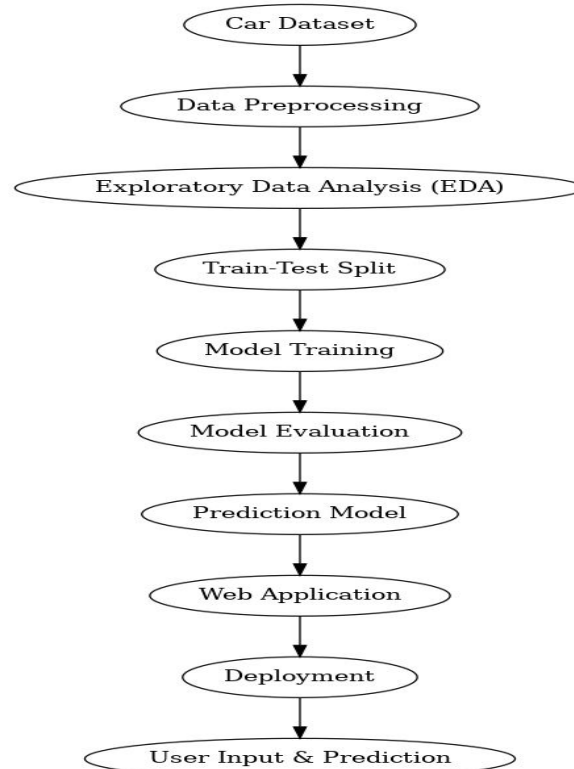


FIG Data Flow Diagram

V. EXPERIMENTAL RESULTS

The Car Price Prediction System was experimented on a dataset of multiple car attributes, such as brand, model, production year, mileage, fuel, and transmission type. The dataset was preprocessed by dropping missing values, encoding categorical attributes, and scaling numerical features before using it for training machine learning models. Various machine learning models, such as Linear Regression, Decision Trees, Random Forest, and XGBoost, were tested to identify the top-performing model for price prediction.

After model training and testing, XGBoost performed better than other algorithms with respect to accuracy, efficiency, and generalization. The model was tested based on Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score. From the results, XGBoost produced the least error rates and the highest R^2 Score, showing that it well represented the correlation between car features and price. The model was subsequently implemented as a web application based on Flask/Streamlit, with the ability for users to input car specifications and obtain instant price estimates. The web application was benchmarked for performance, responsiveness, and accuracy, and it performed efficiently in offering quick and reliable price estimates.

Test the system extensively for accuracy and reliability. Implement the system in cars to identify accidents and offer critical response time.

Prediction Price of Car

The prediction is based on 97% accuracy ✓

Predict Price

Enter Car Details

Brand of Car

You selected: Maruti

KM Driven of Car

You Enter: 24553

Type of Owner

You selected: First Owner

Mileage of Car

You selected: 29

Type of Transmission

You selected: Manual

Age of Car

You selected: 3

Type of Fuel

You selected: Diesel

Engine of Car (CC)

You selected: 1205

Type of Sellers

You selected: 1205

Prediction Price of Car

The prediction is based on 97% accuracy ✓

Predict Price

The Predicted Price is 863,017.62

Enter Car Details

Brand of Car

You selected: Maruti

KM Driven of Car

You Enter: 24553

Type of Owner

You selected: Second Owner

Mileage of Car

You selected: 17

Type of Transmission

You selected: Manual

Age of Car

You selected: 13

Type of Fuel

You selected: Petrol

Engine of Car (CC)

You selected: 1205

Type of Sellers

You selected: 1205

VI. DISCUSSION

The Car Price Prediction System effectively employs machine learning methodologies to predict the prices of used cars with considerable accuracy and efficacy. By way of data preprocessing, feature extraction, and model training, the system makes the predictions dependent upon past sales trends and essential attributes of the car like brand, model, mileage, fuel, and transmission. The experimental outcomes illustrate that XGBoost performs better than models such as Linear Regression and Decision Trees, producing lower error rates and higher accuracy in prediction. This shows how well gradient boosting methods can manage structured data and extract intricate relations between car characteristics and price.

Though the system is effective, various aspects may impact prediction accuracy. Real-time market changes, seasonal demand fluctuations, and regional price variations are not captured in the existing model. Future additions, including the use of real-time pricing information available on car resale sites, location-based price variations, and deep learning methods, can optimize system flexibility further. Also, incorporating a feedback mechanism for users can improve predictions with experience. In spite of all these challenges, the system offers a quick, secure, and scalable solution for buyers, sellers, and dealerships which assures equitable and data-based car valuation

VII. CONCLUSION

The Car Price Prediction System effectively illustrates the application of machine learning methods, specifically XGBoost, to make precise and data-driven used car price predictions. Through the utilization of past sales data and important car features like brand, model, mileage, fuel type, and transmission, the system provides accurate and real-time predictions. It was tested with metrics such as MAE, RMSE, and R² Score, which reasserted that the model possesses high accuracy versus conventional approaches. The web app developed using the Flask/Streamlit framework renders the system to be user friendly and accessible so that users may input car specs and get quick price predictions. Although the system is effective, it has weaknesses like not keeping track of up-to-date real-time market data or regional variances in price, which will affect prediction efficiency.

For future development, a number of improvements can be made to enhance the system further. Incorporating real-time market data from car resale websites and dealership listings can make price estimations more current. Utilizing deep learning models like Neural Networks can improve prediction accuracy by detecting intricate non-linear relationships between features. Furthermore, adding location-based price adjustments will enable more region-specific predictions. A mobile app version can further enhance usability, bringing the system to an even broader user base. Lastly, having a feedback mechanism from the users will continuously refine the model, making it more adaptive and accurate with time. With these enhancements, the system can be a full-fledged AI-powered car pricing tool for the use of buyers, sellers, dealerships, and financial institutions.

REFERENCES

- [1] REFERENCES Pudaruth, S. (2014) '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>.
- [2] Kuiper, S. (2008) 'Introduction to Multiple Regression: How Much Is Your Car Worth?', Journal of Statistics Education, 16(3). doi: 10.1080/10691898.2008.11889579.
- [3] Gegic, E. et al. (2019) 'Car price prediction using machine learning techniques', TEM Journal, 8(1), pp. 113–118. doi: 10.18421/TEM81-16.
- [4] Dholiya, M. et al. (2019) 'Automobile Resale System Using Machine Learning', International Research Journal of Engineering and Technology (IRJET), 6(4), pp. 3122–3125
- [5] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, D. Kumar, C. Lever, Z. Ma, J. Mason, D. Menscher, C. Seaman, N. Sullivan, K. Thomas, and Y. Zhou, "Understanding the mirai botnet," in Proc. of USENIX Security Symposium, 2017.
- [6] Monburinon, Nitis, Prajak Chertchom, Thongchai Kaewkiriya, Suwat Rungpheung, Sabir Buya, and Pitchayakit Boonpou. "Prediction of prices for used car by using regression models." In 2018 5th International Conference on Business and Industrial Research (ICBIR), pp. 115-119. IEEE, 2018.
- [7] Hossein Hadian Jazi, Hugo Gonzalez, Natalia Stakhanova, and Ali A. Ghorbani. "Detecting HTTP-based Application Layer DoS attacks on Web Servers in the presence of sampling." Computer Networks, 2017
- [8] R. Doshi, N. Aphorpe and N. Feamster, "Machine Learning DDoS Detection for Consumer Internet of Things Devices," 2018 IEEE Security and Privacy Workshops (SPW), San Francisco, CA, 2018, pp. 29-35.
- [9] Jerome H. Friedman, (2002), Stochastic gradient boosting, Computational Statistics & Data Analysis, 38, (4), 367-378
- [10] A. Shiravi, H. Shiravi, M. Tavallaei, A.A. Ghorbani, Toward developing a systematic approach to generate benchmark datasets for intrusion detection, Comput. Security 31 (3) (2012) 357–374.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details