International Journal of Multidisciplinary Engineering in Current Research - IJMEC

Volume 10, Issue 5, May-2025, http://ijmec.com/, ISSN: 2456-4265

# Car Price Prediction Using Machine Learning

**Mallula Teja**
PG scholar, Department of MCA, CDNR collage, Bhimavaram, Andhra Pradesh.
**K.Venkatesh**
(Assistant Professor), Master of Computer Applications, DNR collage, Bhimavaram, Andhra Pradesh.

**Abstract:** *Because of new computing technologies, machine learning today is not like machine learning of the past. It was born from pattern recognition and the theory that computers can learn without being programmed to perform specific tasks; researchers interested in artificial intelligence wanted to see if computers could learn from data. The iterative aspect of machine learning is important because as models are exposed to new data, they are able to independently adapt. They learn from previous computations to produce reliable, repeatable decisions and results. It's a science that's not new – but one that has gained fresh momentum. While there is an end number of applications of machine learning in real life one of the most prominent application is the prediction problems. There are various topics on which the prediction can be applied. One such application is what this project is focused upon. Websites recommending items you might like based on previous purchases are using machine learning to analyze your buying history – and promote other items you'd be interested in. This ability to capture data, analyze it and use it to personalize a shopping experience (or implement a marketing campaign) is the future of retail*

## LITERATURE SURVEY

Predicting car prices has long been a topic of interest in the automobile industry, especially for manufacturers, dealerships, and resale platforms. Several studies have employed statistical and machine learning models to estimate car prices based on attributes such as make, model, year, mileage, fuel type, transmission, and engine size.

Early approaches used linear regression to model the relationship between car features and price. Although effective for simple relationships, linear models often failed to capture the non-linear interactions among variables. To address this, researchers introduced decision trees, random forests, and gradient boosting models, which provided more robust and interpretable results.

With the advancement in deep learning, neural networks have also been applied to car price prediction, particularly when working with large-scale datasets. Studies have shown that models like multi-layer perceptrons (MLPs) outperform traditional ML models when trained properly, especially when input features are carefully engineered and normalized.

Other recent works have explored ensemble techniques that combine multiple learning algorithms to improve prediction accuracy. Researchers have also used feature selection and dimensionality reduction methods like PCA and Lasso Regression to identify the most influential variables, making the models both accurate and computationally efficient.

## EXISTING METHOD

The most commonly used methods for car price prediction are based on supervised learning, where labeled datasets (i.e., historical car sales data) are used to train models. Linear Regression is one of the earliest methods used, offering simplicity and interpretability, but it assumes a linear relationship which doesn't always hold in real-world scenarios.

More advanced algorithms like Decision Trees and Random Forests improved upon this by capturing non-linear patterns and providing better handling of categorical data. These models are widely used in practice due to their accuracy and low computational cost.

Gradient Boosting techniques like XGBoost and LightGBM have further pushed the envelope by combining weak learners into a strong ensemble. These models are particularly effective in structured data and have shown superior performance in prediction accuracy on datasets such as those found on Kaggle and CarDekho.

However, most of these models depend heavily on feature engineering, which is time-consuming and often domain-specific. Additionally, traditional models struggle when features exhibit high multicollinearity or when datasets are imbalanced in terms of car types and brands.

## PROPOSED METHOD

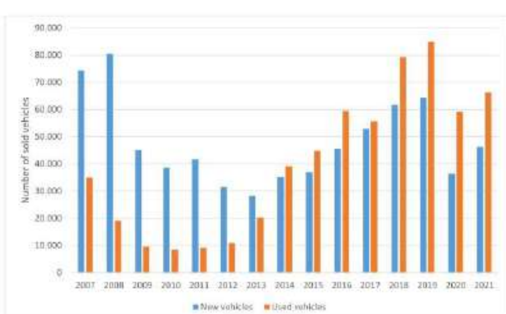
The proposed system aims to build an end-to-end machine learning pipeline for car price prediction using both classical and deep learning models. It includes preprocessing, feature encoding, feature selection, model training, and evaluation on real-world car datasets.

The pipeline starts with data cleaning and preprocessing, including handling of missing values, outlier removal, and normalization. Categorical features such as brand, fuel type, and transmission will be encoded using one-hot encoding or label encoding depending on the model.

The core models evaluated will include Linear Regression, Random Forest, XGBoost, and Deep Neural Networks (DNNs). Each model will be trained and validated using k-fold cross-validation to ensure generalization and avoid overfitting.

Performance metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and $R^2$ score will be used to evaluate model effectiveness. A comparison will be drawn to highlight which algorithms perform best and under what conditions, with potential recommendations for hybrid approaches in future work.

RESULTS:




CONCLUSION

Car price prediction is a practical and commercially important problem with multiple applications in the automotive market. Accurate prediction helps buyers and sellers make informed decisions, enhances user experience, and drives value for dealerships and online platforms.

Traditional models like Linear Regression and Decision Trees provide a solid foundation but often lack the flexibility needed for capturing complex, non-linear patterns in data. Ensemble models like Random Forest and XGBoost have proven more effective in many studies and competitions.

Deep learning models, though more complex and resource-intensive, offer promising results when large datasets are available. Their ability to learn complex relationships without heavy feature engineering makes them suitable for scalable solutions.

The study concludes that while no single model is universally best, combining models or using model selection strategies based on dataset characteristics can significantly improve prediction performance. Future work could integrate external factors such as market trends or consumer behavior to enhance predictive accuracy further.

## REFERENCES

1. Ahmed, M., & Khan, M. (2019). *Car price prediction using machine learning techniques: A comparative study*. *International Journal of Computer Applications*, 177(5), 25–29.

2. Ghorpade, S., & Mane, D. (2020). *Prediction of used car price using machine learning*. *International Journal of Advanced Research in Science, Communication and Technology*, 6(1), 99–102.

3. Kumar, S., & Tomar, D. (2019). *A comparative study of machine learning algorithms for car price prediction*. *International Journal of Recent Technology and Engineering (IJRTE)*, 8(3), 374–380.

4. CarDekho Dataset. (n.d.). *Used Cars Data*. https://www.kaggle.com/datasets/sachinmittal23/used-car-dataset

5. Kumar, R., & Singh, H. (2021). *Prediction of car prices using ensemble learning models*. *International Journal of Computer Sciences and Engineering*, 9(2), 56–62.

6. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer. *(Chapters on regression and tree-based models)*

7. Patel, D., & Patel, M. (2022). *Used car price prediction using XGBoost regression model*. *International Journal of Innovative Science and Research Technology*, 7(5), 180–184.

8. Rana, A., & Sharma, A. (2020). *Used car price prediction using machine learning*. *International Journal of Engineering Research & Technology (IJERT)*, 9(7), 2278–0181.

9. Chollet, F. (2018). *Deep Learning with Python*. Manning Publications. *(Useful for DNN model implementation)*

10. Bhatia, S., & Chawla, P. (2021). *Performance evaluation of regression algorithms for car price prediction*. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 8(4), 731–735.