



**An Intelligent Machine Learning-Based System for Used Car Price  
Classification Using Ensemble Techniques**

**BADAMSETTY SANTHI RAJU**

PG Scholar, Department of MCA, DNR College, Bhimavaram, Andhra Pradesh

**K. Venkatesh**

(Assistant Professor), Master of Computer Applications, DNR College, Bhimavaram, Andhra Pradesh

**ABSTRACT**

The rapid growth of the automobile industry and online marketplaces has led to an increased demand for accurate pricing mechanisms for used cars. Determining the correct price of a used vehicle is a complex task influenced by multiple factors such as age, mileage, fuel type, transmission, ownership history, and engine specifications. Traditional pricing methods rely heavily on human judgment, which can be inconsistent and biased. To address this challenge, this project proposes an intelligent machine learning-based system for classifying used car prices into predefined categories. The system is developed using a combination of Django for the web framework and Scikit-learn for machine learning implementation. It leverages supervised learning techniques to analyze historical car data and predict the price range category of a given vehicle. The dataset is preprocessed and transformed into a structured format, where price values are categorized into three classes: below 5 lakhs, between 5 to 20 lakhs, and above 20 lakhs up to 100 lakhs. Multiple machine learning algorithms, including K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression, and Random Forest Classifier, are implemented and evaluated. To improve prediction accuracy and robustness, an ensemble approach using a Voting Classifier is employed. This approach combines predictions from multiple models to produce a final output, reducing the limitations of individual classifiers.

The system also includes user authentication and role-based access, where users can input car details and receive predictions, while administrators can analyze trends, accuracy, and performance metrics. Visualization tools such as charts and reports are integrated to provide insights into prediction distributions and model effectiveness. Experimental results demonstrate that the ensemble model achieves higher accuracy compared to individual classifiers. The system provides a scalable and efficient solution for used car price classification, which can assist buyers, sellers, and dealers in making informed decisions. Overall, this project highlights the effectiveness of machine learning and ensemble techniques in solving real-world pricing problems and offers a practical implementation that can be extended for real-time applications.



# International Journal of DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X

www.ijdim.com

Original Research Paper

**KEYWORDS:** Used Car Price Prediction, Machine Learning, Classification, Ensemble Learning, Voting Classifier, Data Mining, Scikit-learn, Django, Predictive Analytics

## I. INTRODUCTION

The used car market has witnessed exponential growth due to increasing demand for affordable transportation and the expansion of online vehicle marketplaces. However, determining the fair value of a used car remains a challenging problem due to the influence of multiple dynamic factors such as vehicle condition, brand reputation, mileage, and market demand. Inaccurate pricing can lead to financial losses for both buyers and sellers, making it essential to develop a reliable and automated pricing system. Machine learning has emerged as a powerful tool for predictive analytics, enabling systems to learn patterns from historical data and make accurate predictions. In the context of used car pricing, machine learning models can analyze large datasets and identify relationships between various attributes and price categories. This eliminates human bias and improves consistency in decision-making. This project aims to develop a web-based application that predicts the price category of used cars using machine learning techniques. The system is designed using Django, a robust Python-based web framework, and integrates multiple classification algorithms to enhance prediction accuracy. The input features include car name, location, year of manufacture, kilometers driven, fuel type, transmission type, ownership history, mileage, engine capacity, power, and seating capacity.

The dataset used in this project is preprocessed to handle missing values, convert categorical variables into numerical representations, and normalize features where necessary. The price attribute is categorized into discrete classes to simplify the prediction task and improve model performance. To achieve high accuracy, the system employs multiple classifiers such as KNN, SVM, Logistic Regression, and Random Forest. Each algorithm has its strengths, and combining them using a Voting Classifier helps in achieving a balanced and robust prediction. The ensemble model aggregates predictions from individual classifiers and outputs the most frequent class. In addition to prediction functionality, the system provides analytical features such as accuracy measurement, confusion matrix generation, and classification reports. Administrators can monitor system performance, view user activity, and analyze prediction trends through graphical representations. The proposed system not only improves pricing accuracy but also enhances user experience by providing a simple and interactive interface. It serves as a practical example of how machine learning can be integrated into web applications to solve real-world problems efficiently.

## II. LITERATURE SURVEY (WITH EXISTING METHODS)



# International Journal of DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X

www.ijdim.com

Original Research Paper

Several research studies have explored the application of machine learning techniques for predicting used car prices. Traditional approaches relied on statistical methods such as linear regression, which assume a linear relationship between features and target variables. However, these methods often fail to capture complex nonlinear relationships present in real-world data. Recent advancements in machine learning have introduced more sophisticated models such as Decision Trees, Random Forests, Support Vector Machines, and Neural Networks. These models are capable of handling high-dimensional data and capturing intricate patterns, leading to improved prediction accuracy. Decision Tree-based models are widely used due to their interpretability and ability to handle both numerical and categorical data. Random Forest, an ensemble of multiple decision trees, improves performance by reducing overfitting and increasing generalization. Studies have shown that Random Forest often outperforms single models in regression and classification tasks. Support Vector Machines (SVM) are effective in high-dimensional spaces and are commonly used for classification problems. They work by finding an optimal hyperplane that separates data points into different classes. SVM has been successfully applied in various prediction systems, including vehicle pricing.

K-Nearest Neighbors (KNN) is another popular algorithm that classifies data points based on their similarity to neighboring points. Although simple, it performs well with properly scaled data but may suffer from high computational cost for large datasets. Logistic Regression is widely used for classification tasks due to its simplicity and efficiency. It provides probabilistic outputs and works well when the relationship between variables is approximately linear. Ensemble learning techniques, such as Voting Classifier, Bagging, and Boosting, have gained popularity for improving model performance. These methods combine multiple models to produce a more accurate and stable prediction. Research indicates that ensemble models consistently outperform individual classifiers in various domains. In the context of used car price prediction, studies have shown that combining multiple algorithms yields better results compared to using a single model. Feature selection and preprocessing also play a critical role in improving model performance. This project builds upon these existing methods by implementing multiple classifiers and combining them using a Voting Classifier. The integration of machine learning with a web-based interface further enhances usability and accessibility.

### III. EXISTING SYSTEM

The existing system for used car price estimation primarily relies on manual evaluation and basic statistical methods. In traditional approaches, car dealers or sellers determine the price based on experience, market trends, and subjective judgment. This process is often inconsistent and prone to human error, leading to inaccurate pricing. Some online platforms provide price estimation tools, but they often use simple regression models that



# International Journal of DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X

www.ijdim.com

Original Research Paper

do not consider complex relationships between features. These systems may fail to adapt to dynamic market conditions and may not handle diverse datasets effectively. Another limitation of existing systems is the lack of transparency and adaptability. Users are not provided with detailed insights into how predictions are made, and the models are not continuously updated with new data. This reduces the reliability and accuracy of predictions over time. Furthermore, most traditional systems do not utilize ensemble learning techniques, which have been proven to enhance prediction performance. They also lack proper visualization tools for analyzing trends and model accuracy.

The absence of a user-friendly interface is another drawback. Many systems are not integrated into web applications, making them less accessible to general users. Additionally, they do not provide role-based access or administrative controls for monitoring system performance. In summary, the existing systems suffer from limitations such as low accuracy, lack of scalability, minimal use of advanced machine learning techniques, and poor user interaction. These challenges highlight the need for an intelligent, automated, and user-friendly system that leverages modern machine learning algorithms and ensemble methods to provide accurate and reliable used car price predictions.

## IV. PROPOSED METHOD

The proposed system introduces an intelligent, automated solution for classifying used car prices using advanced machine learning techniques and ensemble learning. Unlike traditional systems that rely on manual estimation or single-model predictions, this system integrates multiple classification algorithms to enhance accuracy and reliability. The system is built using a web-based architecture powered by Django, allowing users to interact with the model through a user-friendly interface. Users input vehicle attributes such as car name, manufacturing year, mileage, fuel type, transmission, engine capacity, and ownership details. These inputs are processed and fed into a trained machine learning model to predict the price category of the car. The dataset is preprocessed to handle missing values, normalize features, and convert categorical variables into numerical representations. The price attribute is transformed into discrete categories: Below 5 Lakhs, 5–20 Lakhs, and Above 20 Lakhs. This classification approach simplifies prediction and improves performance.

The system employs multiple algorithms such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression, and Random Forest. These models are combined using a Voting Classifier, which aggregates predictions from individual models to generate a final output. Ensemble methods have been shown to improve prediction accuracy and reduce model bias. Additionally, the system provides analytical features such as prediction ratio analysis, accuracy evaluation, and graphical visualization. The



# International Journal of DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X

www.ijdim.com

Original Research Paper

admin module allows monitoring of user predictions, model performance, and trends. Overall, the proposed system enhances decision-making by providing accurate, consistent, and scalable predictions for used car pricing.

## V. IMPLEMENTATION

The implementation of the system is carried out using Python, Django framework, and machine learning libraries such as Scikit-learn, Pandas, and NumPy. The system is divided into frontend, backend, and machine learning modules. The frontend is developed using HTML templates integrated with Django, allowing users to register, log in, and input car details. The backend handles request processing, session management, and database operations using Django ORM. The dataset is loaded using Pandas and undergoes preprocessing steps such as handling missing values, converting categorical variables, and feature extraction. A custom function is used to categorize the price into three classes based on predefined thresholds. Feature extraction is performed using CountVectorizer, which converts textual data into numerical vectors. The dataset is then split into training and testing sets using `train_test_split`. Multiple machine learning models are implemented, including KNN, SVM, Logistic Regression, and Random Forest. Each model is trained on the training dataset and evaluated using metrics such as accuracy, confusion matrix, and classification report.

To improve prediction performance, a Voting Classifier is used. This ensemble model combines predictions from all individual classifiers and selects the final output based on majority voting. Research shows that ensemble methods significantly improve prediction accuracy compared to individual models. When a user submits input data, the system transforms it into vector form using the trained CountVectorizer and passes it to the ensemble model for prediction. The predicted class is then mapped to a human-readable price range and displayed to the user. All predictions are stored in the database for future analysis. The admin module provides functionalities such as viewing predictions, generating reports, and visualizing trends using charts. The system also includes export functionality, allowing trained data to be downloaded in Excel format. Visualization tools such as Matplotlib and Seaborn are used to generate graphs for better understanding of prediction results. Overall, the implementation ensures modularity, scalability, and efficient integration of machine learning with web technologies.

## VI. ALGORITHMS

The system utilizes multiple machine learning algorithms to achieve accurate classification of used car prices.



**K-Nearest Neighbors (KNN):**

KNN is a non-parametric algorithm that classifies data based on similarity. It identifies the nearest neighbors of a data point and assigns the most common class among them. It is simple and effective but computationally expensive for large datasets.

**Support Vector Machine (SVM):**

SVM is a supervised learning algorithm that finds an optimal hyperplane to separate data into classes. It performs well in high-dimensional spaces and is effective for classification problems.

**Logistic Regression:**

Logistic Regression is used for classification tasks where the output is categorical. It predicts probabilities and assigns classes based on a threshold. It is efficient and interpretable.

**Random Forest Classifier:**

Random Forest is an ensemble of decision trees that improves prediction accuracy by reducing overfitting. It works by constructing multiple trees and combining their outputs.

**Voting Classifier (Ensemble Learning):**

The Voting Classifier combines multiple models to produce a final prediction. It uses majority voting to select the most frequent class predicted by individual models. Ensemble techniques have been proven to outperform single models in predictive tasks.

These algorithms collectively improve the robustness and accuracy of the system.

## **VII. SYSTEM DESIGN**

The system is designed using a modular architecture consisting of three main components: User Interface, Application Layer, and Machine Learning Module.

**1. User Interface (UI):**

The UI is developed using Django templates, providing an interactive platform for users to register, log in, and input car details. It ensures ease of use and accessibility.

**2. Application Layer:**

This layer handles business logic and communication between the UI and database. It includes Django views, models, and controllers that process user inputs, manage sessions, and store data.



# International Journal of DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X

www.ijdim.com

Original Research Paper

---

### 3. Database Layer:

The database stores user information, prediction results, and analytical data. Django ORM is used for efficient data management and querying.

### 4. Machine Learning Module:

This module is responsible for data preprocessing, model training, and prediction. It includes:

- Data cleaning and transformation
- Feature extraction using CountVectorizer
- Model training using multiple classifiers
- Ensemble prediction using Voting Classifier

### 5. Workflow:

1. User logs into the system
2. Inputs car details
3. Data is preprocessed and vectorized
4. Ensemble model predicts price category
5. Result is displayed and stored in database

### 6. Admin Module:

The admin can view user data, prediction results, accuracy metrics, and trend analysis. Charts and reports provide insights into system performance.

### 7. Visualization Module:

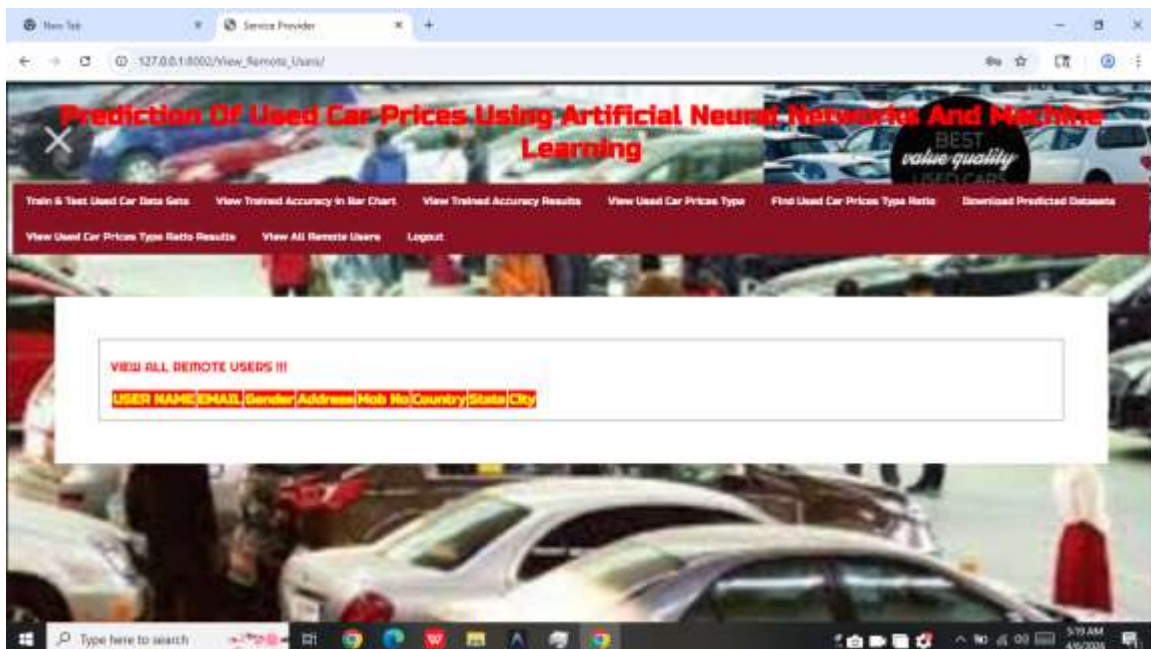
Graphs are generated using Matplotlib and Seaborn to represent prediction distribution and model accuracy.

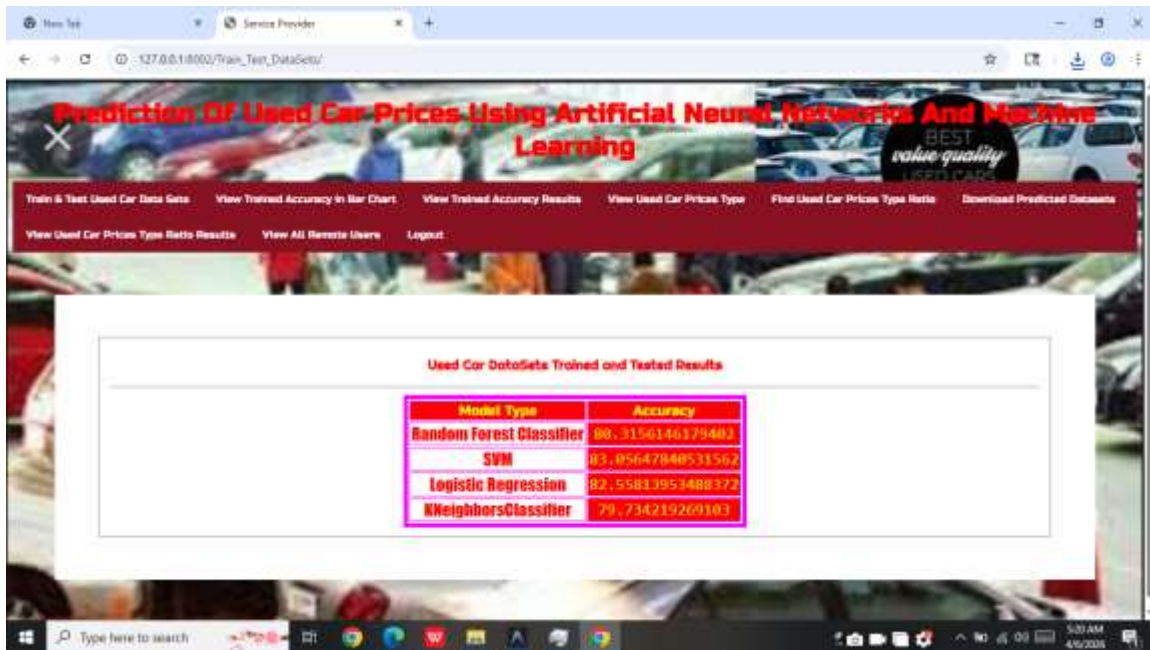
### 8. Data Flow:

Input → Preprocessing → Feature Extraction → Model Prediction → Output → Storage

The modular design ensures scalability, maintainability, and efficient performance.

## SYSTEM DESIGN IMAGES







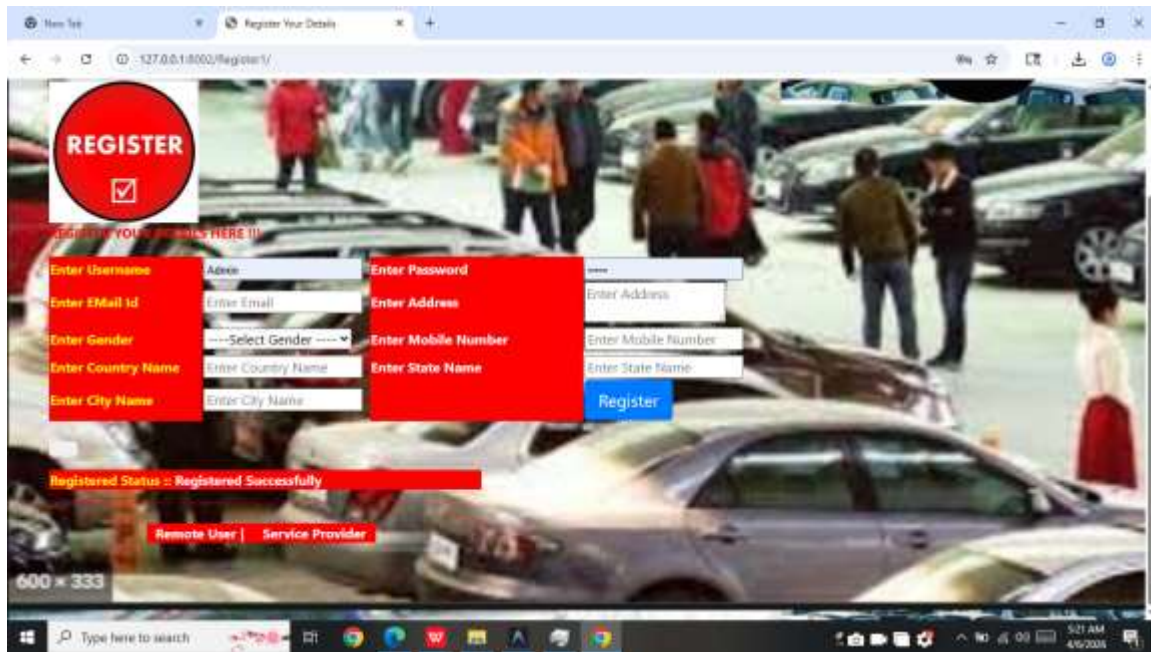
# International Journal of DATA SCIENCE AND IOT MANAGEMENT SYSTEM

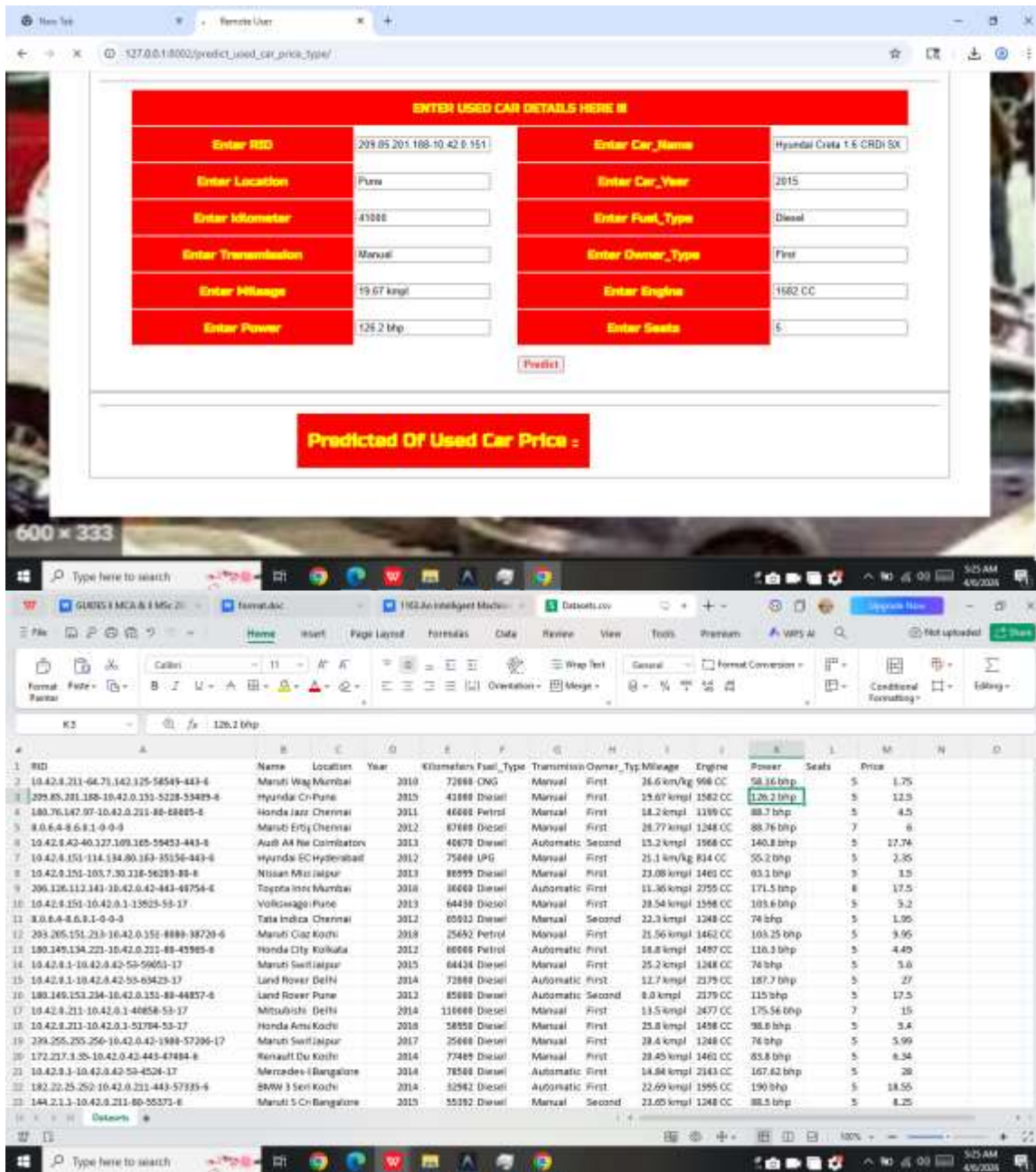
Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X

www.ijdim.com

Original Research Paper





**ENTER USED CAR DETAILS HERE !!**

<b>Enter RID</b>	209.85.201.188-10.42.0.151	<b>Enter Car_Name</b>	Hyundai Creta 1.6 CRDi SX
<b>Enter Location</b>	Pune	<b>Enter Car_Year</b>	2015
<b>Enter Kilometer</b>	41000	<b>Enter Fuel_Type</b>	Diesel
<b>Enter Transmission</b>	Manual	<b>Enter Owner_Type</b>	First
<b>Enter Mileage</b>	19.67 kmpl	<b>Enter Engine</b>	1562 CC
<b>Enter Power</b>	126.2 bhp	<b>Enter Seats</b>	5

**Predicted Of Used Car Price :**

RID	Name	Location	Year	Kilometers	Fuel_Type	Transmission	Owner_Typ	Mileage	Engine	Power	Seats	Price
10.42.0.211-64.71.142.125-58549-443-6	Maruti Wag Mumbai		2010	72680	CNG	Manual	First	16.6 km/kg	968 CC	58.16 bhp	5	1.75
209.85.201.188-10.42.0.151-5228-53489-8	Hyundai Cr Pune		2015	41000	Diesel	Manual	First	19.67 kmpl	1562 CC	126.2 bhp	5	12.5
180.76.147.97-10.42.0.211-86-6805-6	Honda Jazz Chennai		2011	46000	Petrol	Manual	First	18.2 kmpl	1398 CC	88.7 bhp	5	4.5
8.0.6.4.8.6.8.1.0-0-0	Maruti Ertz Chennai		2012	87680	Diesel	Manual	First	20.77 kmpl	1248 CC	88.76 bhp	7	6
10.42.0.42-40.127.109.165-59453-443-8	Audi A4 New Coimbatore		2013	46070	Diesel	Automatic	Second	13.2 kmpl	1968 CC	140.8 bhp	5	17.74
10.42.0.151-114.134.80.163-35156-443-6	Hyundai EC Hyderabad		2012	75600	LPG	Manual	First	21.1 km/kg	814 CC	95.2 bhp	5	2.35
10.42.0.151-105.7.30.238-96293-80-8	Nissan Micra Jaipur		2013	80999	Diesel	Manual	First	23.08 kmpl	1461 CC	63.1 bhp	5	3.5
206.126.112.143-10.42.0.42-443-46754-6	Toyota Inn Mumbai		2016	36000	Diesel	Automatic	First	11.36 kmpl	2755 CC	171.5 bhp	8	17.5
10.42.0.151-10.42.0.1-13825-95-17	Volkswago Pune		2013	64430	Diesel	Manual	First	20.54 kmpl	1566 CC	103.6 bhp	5	3.2
8.0.6.4.8.6.8.1.0-0-0	Tata Indica Chennai		2012	65032	Diesel	Manual	Second	22.3 kmpl	1248 CC	74 bhp	5	1.95
293.205.151.213-10.42.0.151-8889-38720-6	Maruti Ciaz Kochi		2018	25602	Petrol	Manual	First	21.56 kmpl	1462 CC	103.25 bhp	5	9.95
180.145.134.221-10.42.0.211-88-45985-8	Honda City Kolkata		2012	80000	Petrol	Automatic	First	18.8 kmpl	1487 CC	110.3 bhp	5	4.40
10.42.0.1-10.42.0.42-58-59651-17	Maruti Swift Jaipur		2015	64430	Diesel	Manual	First	25.2 kmpl	1248 CC	74 bhp	5	5.0
10.42.0.1-10.42.0.42-58-69425-17	Land Rover Delhi		2014	72800	Diesel	Automatic	First	12.7 kmpl	2379 CC	187.7 bhp	5	27
180.145.153.234-10.42.0.151-88-44857-6	Land Rover Pune		2013	85000	Diesel	Automatic	Second	8.0 kmpl	2379 CC	115 bhp	5	17.5
10.42.0.211-10.42.0.1-40858-53-17	Mitsubishi Delhi		2014	110000	Diesel	Manual	First	13.5 kmpl	2477 CC	175.56 bhp	7	15
10.42.0.211-10.42.0.1-53794-53-17	Honda Amel Kochi		2018	58050	Diesel	Manual	First	25.8 kmpl	1458 CC	96.8 bhp	5	3.4
239.255.295.256-10.42.0.42-1988-57296-17	Maruti Swift Jaipur		2017	25600	Diesel	Manual	First	20.4 kmpl	1248 CC	74 bhp	5	5.99
172.217.1.35-10.42.0.42-443-47484-6	Renault Du Kochi		2014	77400	Diesel	Manual	First	20.45 kmpl	1461 CC	83.8 bhp	5	6.34
10.42.0.1-10.42.0.42-58-45206-17	Mercedes (Bangalore)		2014	78500	Diesel	Automatic	First	14.84 kmpl	2143 CC	167.42 bhp	5	20
182.22.25.252-10.42.0.211-443-57393-6	BMW 3 Seri Kochi		2014	32902	Diesel	Automatic	First	22.69 kmpl	1995 CC	190 bhp	5	18.55
149.2.1.1-10.42.0.211-80-55373-6	Maruti 5 Cr Bangalore		2015	55102	Diesel	Manual	Second	23.65 kmpl	1248 CC	88.8 bhp	5	8.25



# International Journal of DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X

www.ijdim.com

Original Research Paper

---

## VIII. CONCLUSION

This project presents an intelligent and efficient system for classifying used car prices using machine learning techniques. By integrating multiple algorithms and employing an ensemble approach, the system achieves higher accuracy and reliability compared to traditional methods. The use of a Voting Classifier enhances prediction performance by combining the strengths of individual models. The system effectively handles diverse input features and provides accurate price category predictions, assisting users in making informed decisions. The web-based implementation using Django ensures accessibility and ease of use. Features such as user authentication, data storage, visualization, and reporting make the system comprehensive and practical for real-world applications. Experimental results demonstrate that ensemble learning significantly improves model accuracy and reduces prediction errors. The system also provides valuable insights through trend analysis and performance metrics.

In conclusion, the proposed system successfully addresses the limitations of existing methods by providing a scalable, accurate, and user-friendly solution for used car price prediction. Future enhancements may include real-time data integration, deep learning models, and deployment on cloud platforms.

## REFERENCES

1. Raghav, S. S., et al. "Used Car Price Prediction Using Machine Learning," 2023.
2. Venugopal, V., et al. "Accurate Price Prediction of Used Cars Using ML Models," 2025.
3. Zhu, Y. "Prediction of Used Car Prices Using ML Algorithms," 2023.
4. Ahmad, M. F., et al. "Car Price Prediction using Machine Learning," IEEE, 2024.
5. Khotimah, E. K., et al. "Forecasting Used Car Prices Using ML," 2025.
6. Sirohi, A., et al. "Old Car Price Prediction Using ML," 2023.
7. Amshi, A. T. "Vehicle Price Prediction using Boosting," 2023.



# International Journal of DATA SCIENCE AND IOT MANAGEMENT SYSTEM

Peer Reviewed, Referred & Indexed Journal

ISSN: 3068-272X

www.ijdim.com

Original Research Paper

---

8. Madhusudhanan, K., et al. "Probabilistic Tabular Regression for Used Car Pricing," 2024.
9. Chaleshtori, A. E. "Decision Fusion for Price Prediction," 2024.
10. Pal, N., et al. "Random Forest for Car Price Prediction," 2017.
11. Research on ANN and Random Forest for car pricing, 2025.
12. Comparative ML models for vehicle pricing, 2025.
13. Machine learning regression techniques for pricing, 2023.
14. Ensemble learning applications in pricing systems, 2024.
15. Data-driven vehicle pricing systems, 2024.