iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52309

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Smart Pricing: Estimating Price of Car Using Enhanced Features and Knowledge-Based System

Dhanshree Konde[1], Pooja Kachare[2], Prajakta Yadav[3], Rohini Reddy[4], Shrushti Kunjeer[5], Prof. Sourabh Natu[6]
[1, 2, 3, 4, 5]*Student TSSM's Bhivarabai Sawant College of Engineering And Research, Narhe, Pune*
[6]*Assistant Professor*

Abstract: ***The price of a new car in the industry is fixed by the manufacturer with some additional costs incurred by the Government in the form of taxes. So, customers buying a new car can be assured of the money they invest to be worthy. But due to the increased prices of new cars and the financial incapability of the customers to buy them, Used Car sales are on a global increase.***
Therefore, there is an urgent need for a Used Car Price Prediction system which effectively determines the worthiness of the car using a variety of features. Existing System includes a process where a seller decides a price randomly and buyer has no idea about the car and its value in the present-day scenario. In fact, seller also has no idea about the car's existing value or the price he should be selling the car at.
To overcome this problem we have developed a model which will be highly effective. Random Forest Algorithm is used because it provides us with continuous value as an output and not a categorized value. Because of which it will be possible to predict the actual price a car rather than the price range of a car. User Interface has also been developed which acquires input from any user and displays the Price of a car according to user's inputs.
Keywords: Machine Learning, Random Forest, Dataset Processing, Feature Extraction.

I. INTRODUCTION

The global automobile industry has expanded and grown every year, becoming increasingly competitive. Consequently, it is crucial to establish accurate pricing that is equitable for both clients and manufacturers in this fierce auto industry. However, both clients and manufacturers may have divergent opinions on how much a car should cost or be traded for. Therefore, both consumers and producers seek counsel from automobile dealers, auto publications, or websites, but obtaining this information may take a while, leading to bewilderment among market buyers. Given the significance of the transportation sector, it is not surprising that the automobile industry is crucial to developing nations. In India, where urban public transit is underdeveloped, automobiles and motorbikes are the primary modes of mobility. Compared to motorcycles, vehicles are becoming increasingly prevalent. As a result, with a considerable size and numerous transactions, buying and selling cars is a vital component of the economy. The sale and purchase of vehicles between the customer and the seller have become substantially simpler with the growth of e-commerce. We can categorize the information we obtain on automobiles into two categories: structured data and unstructured data. There are two types of structured data: quantitative and qualitative. Nevertheless, it is essential to discern the specific circumstances of the automobile to comprehend the observations regarding unstructured data, which refer to the perceptions of those who view the material or the evaluations of the seller.

A. Problem Statement

Predicting the value of a pre-owned vehicle was a challenging task due to the myriad of factors that affect its market price. In the future, this study will develop machine learning models that can nearly prophesy the cost of a car based on its attributes and overall performance, thus enabling consumers to make enlightened choices. We are utilizing and evaluating various learning algorithms on a dataset that encompasses the selling prices of different brands and models in this research. This project will juxtapose all of this data to all Support Vector Machine (SVM) techniques as well as the efficacy of multiple machine learning algorithms such as SVM. The project will appraise the value of a car based on multiple parameters and contrast the prices of new vehicles.

B. Motivation

Forecasting the worth of a previously owned automobile was a formidable task owing to the multitude of factors that influence its market value. Hence, we developed this model to simplify this task of predicting the price of used cars.

©IJRASET: All Rights are Reserved

C. Goal

This project's primary goal is to forecast used automobile prices using Random Forest Algorithm.

D. Objective

1) A supervised machine learning methodology will be employed to predict automobile prices.
2) This can empower customers to make informed decisions based on various inputs or variables, such as:
   - Purchased year
   - Current showroom price
   - Fuel type
   - Kilometres driven
   - Existing owners
   - Transmission type
   - Seller type
3) Designing a proficient and efficacious model that predicts the value of a pre-owned vehicle based on user inputs.
4) Establishing a user-friendly User Interface (UI) that receives user input and forecasts the price.

II. SYSTEM PLANNING

A. Literature Review

Using machine learning to determine the prices of automobiles has a strong association with the process of acquiring knowledge for expert systems.

In recent times, the predominant method for acquiring knowledge has been the time-intensive procedure of seeking advice, advertising for car purchasing or selling on internet market websites. Once we obtain the data, we can classify it into two types: organized and disorganized that necessitate knowledge-based examination. This paper will involve the methods for meaning extraction, data deduction, and regulations for qualitative data.

The primary aim of the current investigation is to investigate different types of car data, and the goal is to design an automated method for predicting car prices.

III. FEASIBILITY STUDY

A. Technical Feasibility

In general, Random Forest has demonstrated promising outcomes in tasks related to forecasting car prices, exhibiting a superior level of precision and resilience to disturbances and anomalies in the information. Nonetheless, the performance might differ based on the particular challenge and the execution of the methodology.

B. Economic Feasibility:

The economic viability of predicting the price of used cars is dependent on several factors such as the availability of reliable data sources, the accuracy and efficiency of predictive models, the cost and resources required to develop and maintain such models, and the potential return on investment for businesses and consumers utilizing these predictions.

C. Operational Feasibility:

The operational feasibility of used car price predictions refers to the practicality and effectiveness of implementing such a system in real-world settings. This includes considerations such as the availability and quality of data sources, the suitability of predictive models for the specific context, the availability of skilled personnel to operate and maintain the system, and the compatibility of the system with existing infrastructure and processes. Other factors that can impact operational feasibility include cost, scalability, and the ability to adapt to changing market conditions and customer needs.

IV. REQUIREMENTS

A. Reliability External Interface Requirements

User Interface: Application Based on Used Car Price Prediction.
Hardware Interface: Intel core
Speed: 2.80Hz, RAM: 8GB, Hardisk:40GB

Software Interfaces:

- Operating System: Windows 10
- IDE: Anaconda
- Programming Language: Python, HTML, CSS, JavaScript
- Frameworks: Flask

B. Non-Functional Requirement

Non-functional requirements for used car price prediction refer to the criteria that cannot be directly measured by the performance of the system but are essential for ensuring its overall quality and usability. Examples of non-functional requirements include:

- *Accuracy:* The system must provide accurate predictions to ensure that users can make informed decisions based on reliable information.
- *Reliability:* The system must be reliable and stable, providing consistent and dependable results over time.
- *Scalability:* The system should be scalable to accommodate growing volumes of data and users, without sacrificing performance or accuracy.
- *Security:* The system must be secure to protect user data and prevent unauthorized access or malicious attacks.
- *Usability:* The system must be user-friendly, intuitive, and easy to use, with clear and concise interfaces and documentation.

C. System Features

Our software possesses various quality attributes that are mentioned below:

- *Adaptability:* This software is versatile and can be adapted by all users.
- *Availability:* The software is freely accessible to all users and its availability is effortless for everyone.
- *Maintainability:* If any error occurs after the software deployment, it can be easily maintained by the software developer.
- *Reliability:* The software's performance is superior, which enhances its reliability.
- *User Friendliness:* As the software is a GUI application, the generated output is highly user-friendly in its behavior.
- *Integrity:* Integrity refers to the degree to which unauthorized access to software or data can be controlled.
- *Security:* Users are authenticated using several security phases, ensuring reliable security.
- *Testability:* The software will be thoroughly tested, considering all aspects.

D. Performance

Generally, Random Forest has been shown to provide good results in car price prediction tasks, with high accuracy and robustness against noise and outliers in the data. However, the performance may vary depending on the specific problem and the implementation of the algorithm.

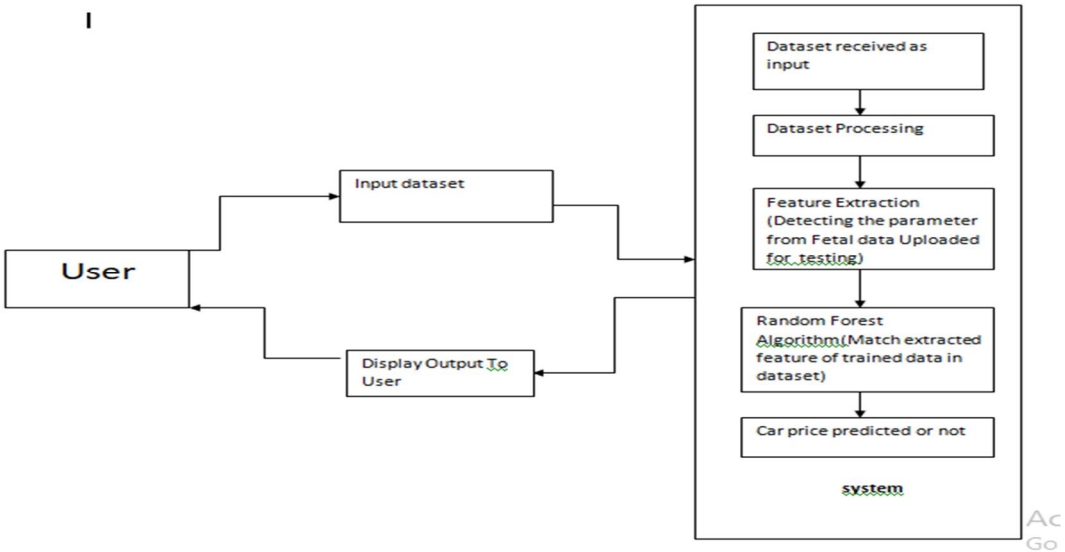


Fig. 1 System Architecture

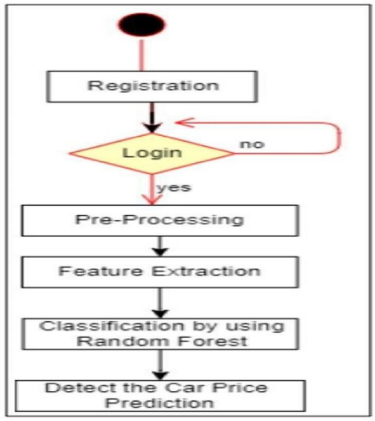


Fig. 1 Activity Diagram

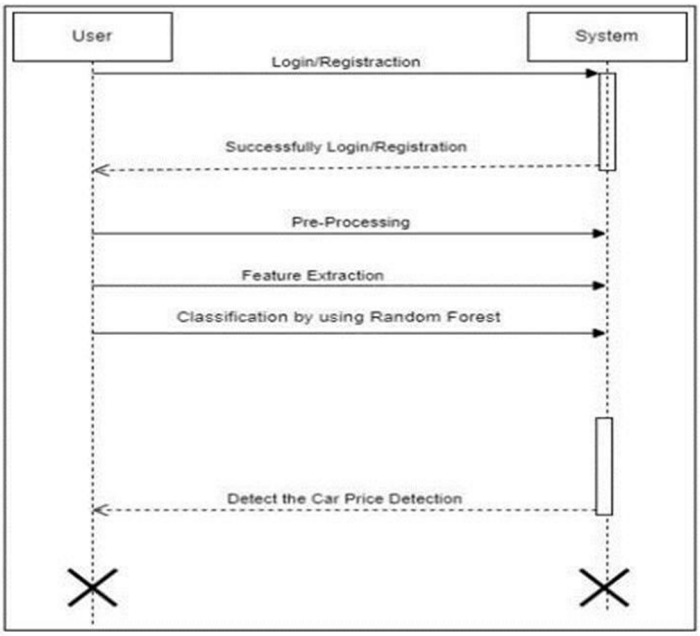


Fig. 2 Sequence Diagram

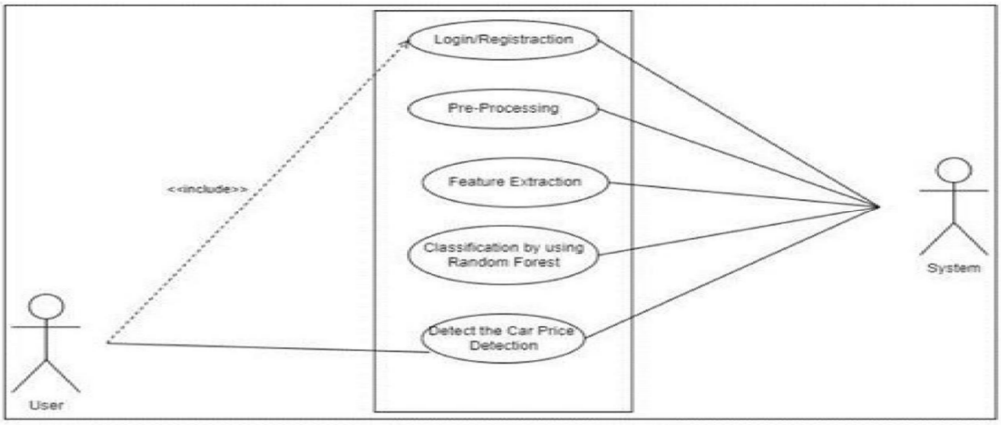


Fig. 3 Use Case Diagram

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

V. IMPLEMENTATION OF SYSTEM

A. Existing System

Existing System includes a process where a seller decides a price randomly and buyer has no idea about the car and its value in the present-day scenario. In fact, seller also has no idea about the car's existing value or the price he should be selling the car at. To overcome this problem we have developed a model which will be highly effective.

B. Limitations

While the random forest algorithm is a powerful tool for prediction, it does have some limitations when it comes to car price prediction. Some of the limitations include:

1) *Limited Input Features:* The accuracy of the random forest algorithm depends heavily on the input features. If the dataset does not contain sufficient relevant input features, the algorithm's predictive power may be limited.
2) *Overfitting:* Random forest models can overfit the training data, especially when the model has too many trees. Overfitting can lead to poor generalization and may result in inaccurate predictions for new data.
3) *Inability to Capture non-linear Relationships:* Random forest models are not able to capture complex non-linear relationships between the input features and the output variable. This can lead to underfitting and may result in inaccurate predictions.
4) *Difficulty Handling Missing Data:* Random forest models can struggle with missing data. If there are many missing values in the input data, the algorithm may not be able to provide accurate predictions.
5) *Sensitivity to Outliers:* Random forest models can be sensitive to outliers in the input data. Outliers can skew the results of the model, leading to inaccurate predictions.

Overall, while random forest models are a powerful tool for prediction, they do have some limitations when it comes to car price prediction. To mitigate these limitations, it is important to carefully select relevant input features, carefully tune the model parameters, and pre-process the data to handle missing values and outliers.

C. Proposed System

The proposed approach for car price prediction involves using the random forest algorithm to build a predictive model. The system will take into account various factors such as age of car, Purchasing year, Kilometre driven. Fuel type, etc. that influence the price of a used car. The dataset will be pre-processed to handle any missing or incorrect data, and feature engineering techniques will be applied to extract useful information from the data. The Random Forest algorithm will then be trained on this pre-processed data to predict the price of a used car.

VI. ADVANTAGES OF PROPOSED SYSTEM

There are various benefits of utilizing the random forest algorithm for predicting car prices:

1) *Precise Predictions:* Random forest models possess the capability to produce highly precise predictions. This is due to their capacity to capture intricate relationships between the input characteristics and the output variable, which can result in more precise predictions.
2) *Resilience:* Random forest models are less susceptible to overfitting compared to other machine learning algorithms. This is because they combine multiple decision trees, each of which is trained on a different subset of the data, which can help to reduce the impact of outliers and noisy data.
3) *Non-parametric:* Random forest models are non-parametric, which means they do not make any suppositions about the underlying distribution of the data. This can be an advantage when working with complex datasets where the relationship between the input features and the output variable is not well comprehended.
4) *Feature importance:* Random forest models can provide information on the importance of each input characteristic in the prediction. This can be useful in understanding which features are most relevant for predicting car prices.
5) *User-friendly:* Random forest models are comparatively effortless to use and require minimal parameter tuning. This makes them a prevalent choice for car price prediction, particularly for those with limited experience in machine learning. Overall, the random forest algorithm is a potent tool for car price prediction owing to its ability to provide accurate predictions, resilience to noisy data, non-parametric nature, feature importance analysis, and user-friendliness.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

VII. MODULE DESCRIPTION

A. Machine Learning and its Categories

Machine learning is a subfield of artificial intelligence that involves developing algorithms and statistical behavioural models that enable computer systems to inculcate from data and make decisions without being explicitly programmed. In understandable terms, machine learning involves using statistical techniques to enable Virtual Machines to improve their performance on a specific task module by learning from data. Machine learning algorithms are used in diversified applications, such as image and speech recognition, natural language processing, predictive analytics, and autonomous vehicles. To function, machine learning algorithms require three key components: data, models, and optimization. The data provides the algorithm with the information necessary to learn and make decisions. The model is the algorithm that is trained on the data, enabling the system to recognize patterns and make predictions. Optimization is the process of adjusting the model to improve its accuracy and effectiveness. The increasing amount of available data has made machine learning more critical in recent years. It has the potential to transform numerous industries, from healthcare and finance to manufacturing and transportation, by providing more accurate predictions and faster decision-making. However, it also raises significant ethical and social concerns, including privacy, bias, and job displacement, that require careful consideration and resolution.

B. Supervised Machine Learning

Supervised learning involves training a model on a labelled dataset, where the output is known for each input, while unsupervised learning involves training a model on an unlabelled dataset to discover patterns and structures in the data. Supervised learning is a commonly used machine learning algorithm that involves training a model on a labelled dataset. In supervised learning, the model is provided with input data and the corresponding correct output data, allowing it to learn the relationship between the two. Once the model is trained, it can be used to make predictions on new input data by applying the learned relationship to the new data. This makes supervised learning a useful tool for tasks such as classification and regression. In this example, the input data are purchased year, fuel type, kilometres driven, current showroom price, existing owner, etc., while the output data is the actual predicted price of the car. The model is trained using the labelled dataset, and once it has learned the relationship between the features and the selling price, it can be used to predict the selling price of car that it has not seen before.

C. Random Forest

Random forest regression is a machine learning algorithm that is utilized for prognostic modelling. It is an augmentation of the random forest algorithm, which is employed for categorization quandaries. In random forest regression, the algorithm assembles numerous decision trees and amalgamates their results to divine the target variable.

The random forest regression algorithm operates by selecting a haphazard sample of the data and utilizing it to erect a decision tree. This procedure is reiterated numerous times, with diverse arbitrary samples and different decision trees. Each decision tree is erected by culling a subset of the characteristics from the data, and dividing the data predicated on these characteristics to fabricate a tree-like structure. The final divination is made by amalgamating the prognostications of all the decision trees in the forest, utilizing techniques such as averaging or weighted averaging.

VIII. CONCLUSION

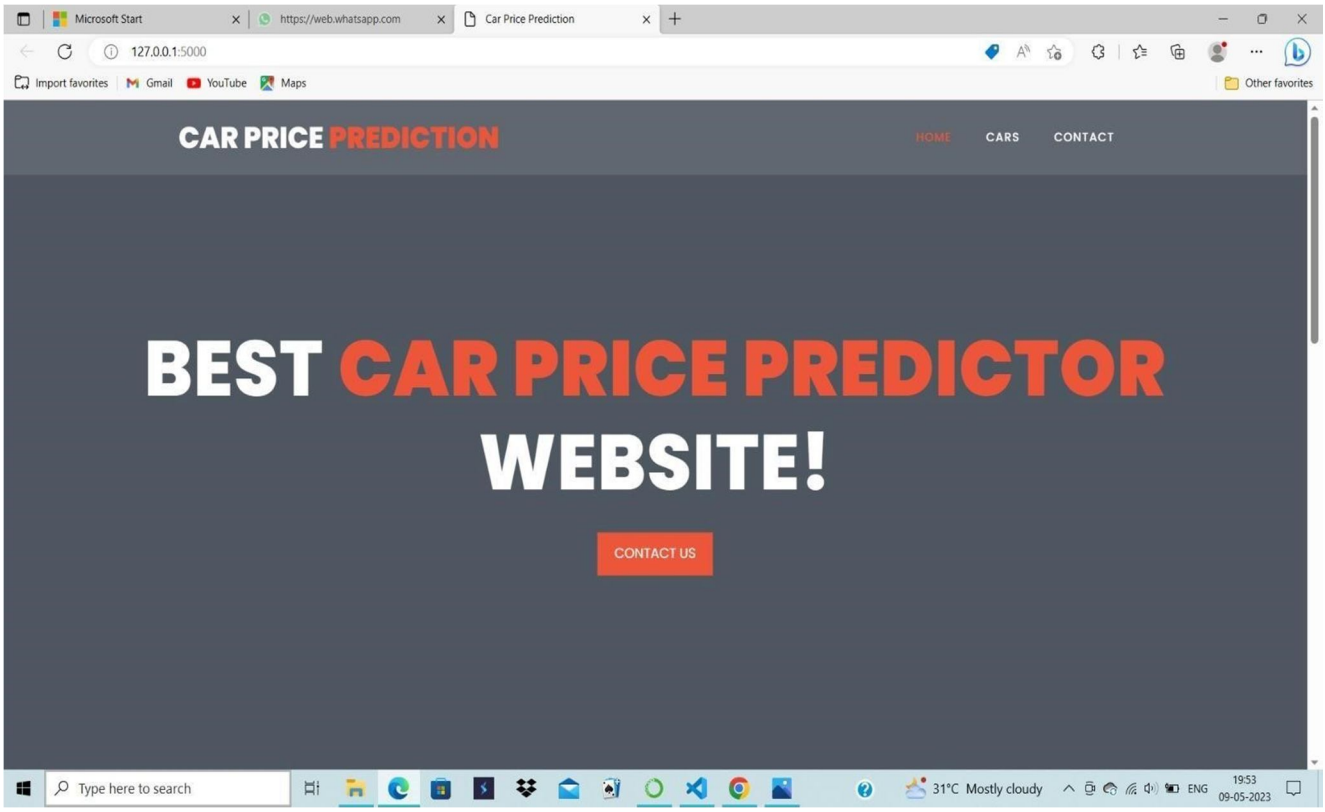
Machine Learning algorithms like Supervised Learning approaches are highly valuable in resolving pragmatic issues. Random forest is a powerful machine learning algorithm that can be used for car price prediction. By leveraging its ability to handle non-linear relationships and interactions between features, it can provide accurate predictions on the prices of cars based on various input variables such as purchased year, kilometre driven, showroom price, number of owners, etc.

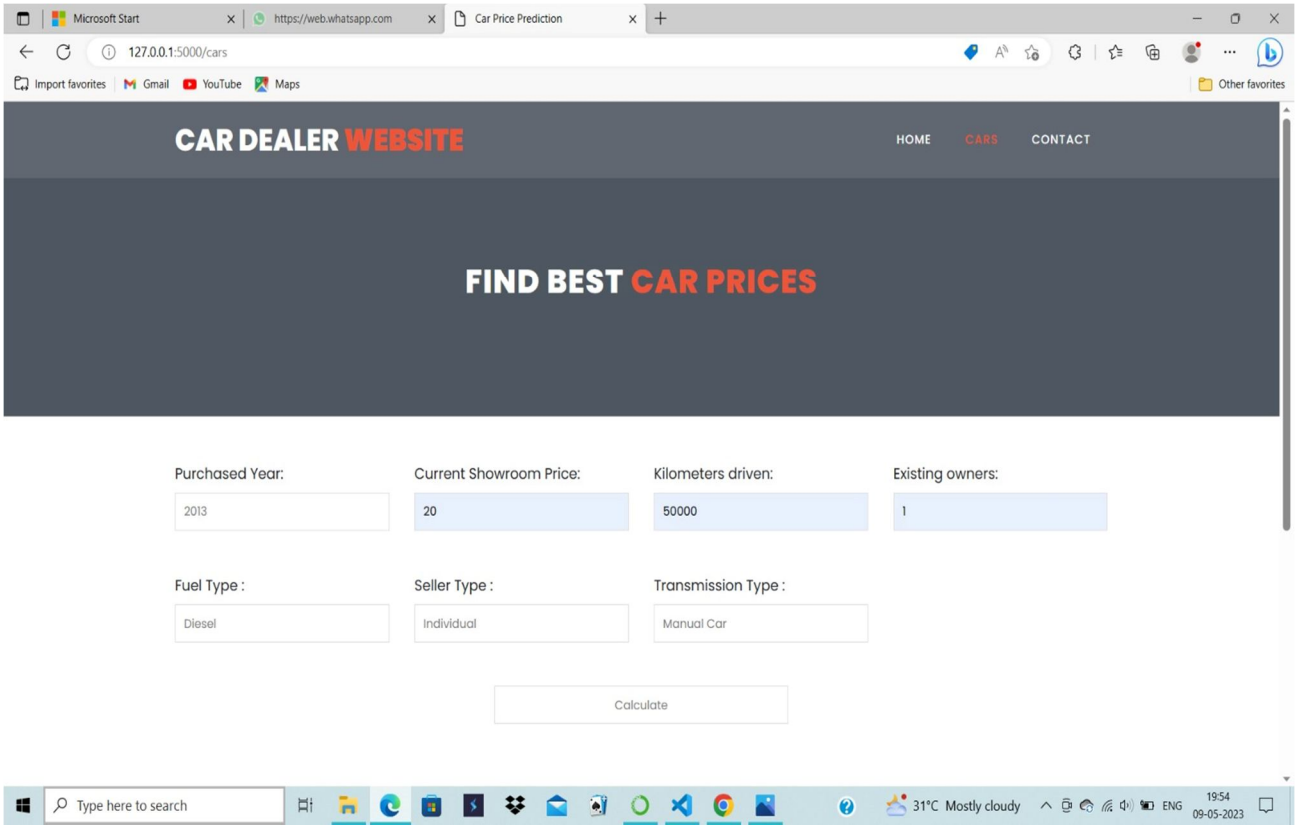
IX. FUTURE WORK

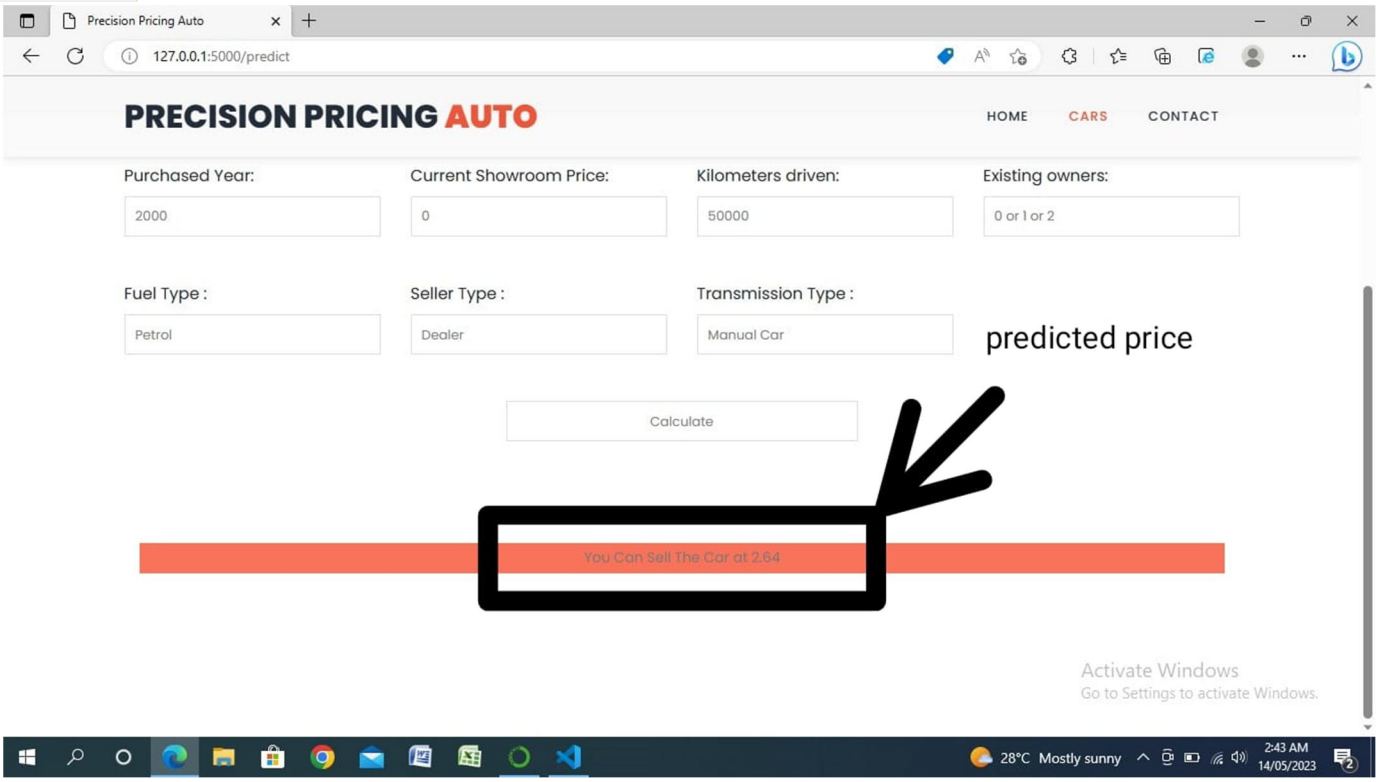
In this research, we have presented the pricing of vehicles, alongside a unique technique for measuring numerical data and a knowledge-based framework. We have proposed a new approach for quantifying numerical data and created a system that relies on knowledge. This model can be beneficial in accurately determining the price of a used car. We can construct a model using the SVM algorithm with the assistance of extensive survey papers. Therefore, there is a need for an Automobile Value Estimation system that can forecast the car's price based on various factors.

©IJRASET: All Rights are Reserved

X. RESULT

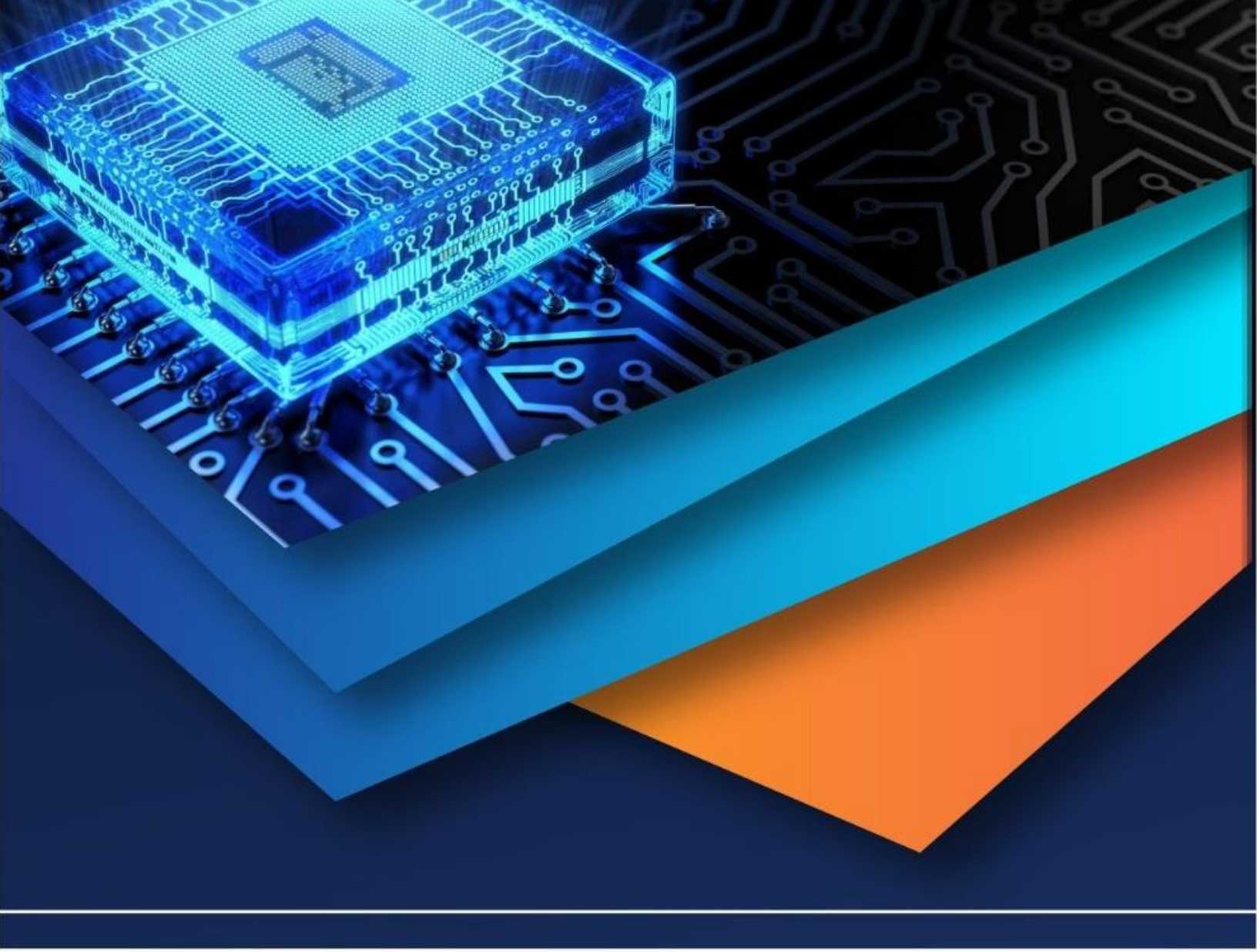

Microsoft Start
https://web.whatsapp.com
Car Price Prediction
127.0.0.1:5000
Import favorites
Gmail
YouTube
Maps
Other favorites
CAR PRICE PREDICTION
HOME
CARS
CONTACT
BEST CAR PRICE PREDICTOR WEBSITE!
CONTACT US
Type here to search
31°C Mostly cloudy
ENG
19:53
09-05-2023


Microsoft Start
https://web.whatsapp.com
Car Price Prediction
127.0.0.1:5000/cars
Import favorites
Gmail
YouTube
Maps
Other favorites
CAR DEALER WEBSITE
HOME
CARS
CONTACT
FIND BEST CAR PRICES
Purchased Year:
2013
Current Showroom Price:
20
Kilometers driven:
50000
Existing owners:
1
Fuel Type :
Diesel
Seller Type :
Individual
Transmission Type :
Manual Car
Calculate
Type here to search
31°C Mostly cloudy
ENG
19:54
09-05-2023



REFERENCES

[1] "Prediction of Prices for Used Car by using Regression Models" (ICBIR 2018). International Advanced Research Journal in Science, Engineering and Technology Impact Factor 8/IARJSET.2022.9457.
[2] Sameerchand Pudaruth, "Predicting the Price of Used Cars using Machine Learning Techniques"; (IJICT 2014).
[3] Enis gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric, "Car Price Prediction Using Machine Learning"; (TEM Journal 2019).
[4] Ning sun, Hongxi Bai, Yuxia Geng, Huizhu Shi, "Price Evaluation Model In Second Hand Car System Based On BP Neural Network Theory"; (Hohai University Changzhou, China).
[5] Nitis Monburinon, Prajak Chertchom, ThongchaiKaewkiriya, Suwat Rungpheung, Sabir Buya, PitchayakitBoonpou,
[6] Doan Van Thai, Luong Ngoc Son, Pham Vu Tien, NguyenNhat Anh, Nguyen Thi Ngoc Anh, "Prediction carpricesusing qualify qualitative data and knowledge-based system" (Hanoi National University).
[7] A. K. Elmagarmid, P. G. Ipeirotis, and V. S. Verykios, "Duplicate Record Detection: A Survey," IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. 1, pp. 1–16.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)