

## Walchand College of Engineering, Sangli

#### (An Autonomous Institute)

**Department of Computer Science and Engineering**

TY CSE Mini Project 1

Report on

**Used-Car Price Predictor**

Submitted by

|  |  |
| --- | --- |
| **Aniket Rahul Khare** | **21510047** |
| **Prathamesh Sidram Basapure** | **21510044** |
| **Aditya Pralhad Kumbhar** | **21510052** |

Under the guidance of

**N. L. Mudegol**

Department of Computer Science and Engineering,

Walchand College of Engineering, Sangli

**2023 - 2024**



**Walchand College of Engineering, Sangli**

(An Autonomous Institute)

**Department of Computer Science and Engineering**

### CERTIFICATE

This is to certify that the Project Report entitled, **“USED CAR PRICE PREDICTOR”** submitted by Mr. Aniket Khare, Mr.Prathamesh Basapure, Mr.Aditya Kumbhar, to Walchand College of Engineering Sangli, India, is a record of bonafide project work of course **Mini Project-1** **6CS341**carried out by him/her under my/our supervision and guidance and is worthy of consideration for the award of the degree of Bachelor of Technology in Computer Science & Engineering of the Institute.

|  |  |  |
| --- | --- | --- |
| **N.L.Mudegol** |  | **Dr. M. A. Shah** |
| **Guide** | **External examiner** | **Head Of Department** |
| Department of Computer Science and Engineering, |  | Department of Computer Science and Engineering, |
| Walchand College of Engineering, Sangli |  | Walchand College of Engineering, Sangli |

# Acknowledgement

We would like to express our sincere appreciation to everyone who contributed to the successful completion of this used car price prediction project. Their support, guidance, and expertise were invaluable throughout the entire process.

We would like to extend our thanks and appreciation to Prof. N. L. Mudegol, our guide, for her continuous support and guidance during the entire project.

Moreover, we would like to extend our gratitude to Prof. S.D. Pujari for his guidance and valuable suggestions.

# Declaration

We hereby declare that work presented in this project report titled **“USED CAR PRICE PREDICTOR”** submitted by us in the partial fulfillment of the requirement of the award of the degree of **Bachelor of Technology (B. Tech)** in the **Department of Computer Science & Engineering, Walchand College of Engineering, Sangli** is an authentic record of my project work carried out under the guidance of Prof. N.L.Mudegol.

Date:8-12-2023

Place: Sangli

**Aniket Khare**

**Prathamesh Basapure**

**Aditya Kumbhar**

## Table of Contents

[Abstract](#_bookmark1) 6

1. [Introduction and](#_bookmark2) Related work 7
2. [Problem statement](#_bookmark3) 8
3. Objectives 8
4. [Methodology](#_bookmark4) 9
5. [Project Diagrams (UML diagrams, Flow chart etc.)](#_bookmark6) 11
6. Testing (Unit, System, Integration etc.) 13
7. Results and Conclusion 14
8. [References](#_bookmark11) 14

###### List Of Figures

* + figure 1 Methodology
  + figure 2 Flow Diagram
  + figure 3 Class Diagram

#### Abstract

**Predicting Second-Hand Car Prices with Machine Learning**

The second-hand car market is a substantial and complex industry, with various factors influencing the prices of used cars. Accurately estimating the resale value of a car can be challenging due to the interplay of multiple parameters such as the car's make, model, year of manufacture, mileage, fuel type, and various other features. This project aims to leverage machine learning and data science techniques to develop a regression model capable of predicting the second-hand price of cars based on historical data.

To accomplish this, data processing technology has been employed. Null, redundant, and missing values were removed from the dataset during pre-processing. In this supervised learning study, we have trained, tested, and compared against a benchmark dataset. Among all the experiments, finally Multiple Linear Regression had the highest score at 90%, followed by 0.14 MSE and 0.378 RMSE respectively. A train-test split of 80/20 with 40 random states was used in all experiments. Then the model is integrated into a web page for the general public to use.

#### Introduction and Related work

The automotive industry has witnessed a significant surge in the sale and purchase of used cars, driven by factors such as economic considerations, increased awareness of environmental sustainability, and a growing demand for affordable transportation solutions. One crucial aspect of the used car market is determining the fair and accurate pricing of pre-owned vehicles. Predicting the price of a used car is a complex task influenced by various factors, including the make and model of the vehicle, its age, mileage, condition, and market trends.

Studies and projects have delved into the realm of predicting used car prices using various machine learning techniques. Previous research in this domain has focused on feature engineering, dataset preparation, and model fine-tuning to enhance the predictive accuracy of linear regression models. Some studies have incorporated additional machine learning techniques, such as feature scaling, regularization, and ensemble methods, to improve the overall performance of used car price prediction models.

Moreover, researchers have investigated the impact of different features on the accuracy of linear regression models. Common features include mileage, vehicle age, brand reputation, fuel efficiency, and geographic location. Understanding the significance of these features is crucial for refining the model and ensuring its applicability across diverse markets.

#### Problem statement

To create a robust regression model using machine learning techniques to predict the second-hand price of cars. Given a dataset containing various parameters related to cars, such as the car's name, location, year of manufacture, Kilometers driven, fuel type, transmission, owner type, mileage, engine capacity, power, seats, and the second-hand price, the goal is to build a predictive model that can accurately estimate the resale value of a car based on these parameters.

The purpose of this study is to understand and evaluate used car prices in the India, and to develop a strategy that utilizes data mining techniques to predict used car prices.

#### Objectives

This project aims to deliver price prediction models to the public, to help guide the individuals looking to buy or sell cars and to give them a better insight into the automotive sector. Buying a used car from a dealer can be a frustrating and an unsatisfying experience as some dealers are known to deploy deceitful sale tactics to close a deal. Therefore, to help consumers avoid falling victims to such tactics, this study hopes to equip consumers with right tools to guide them in their shopping experience. Another goal of the project is to explore new methods to evaluate used cars prices and to compare their accuracies. Considering this is an interesting research topic in the research community, and in continuing their footsteps, we hope to achieve significant results using more advanced methods of previous work.

This project also aims to create a user-friendly User Interface (UI) that receives input from the user and forecasts the pricing effectively.

#### Methodology

#### 

After data collection the dataset was pre-processed to remove samples that have missing value, and remove non-numerical part from numerical attributes, converting categorical values into numerical (if needed), fix any discrepancies in the units, as well as removing attributes that doesn’t affect the price evaluations if needed to reduce the complexity of the model. Data Understanding and preparation is an essential part of building a model as it gives the insight into the data and what corrections or modifications shall be done before designing and executing the model, preliminary analysis of the data must be done to have deeper understanding into the quality of the data, in terms of outliers and the skewedness of the figures, descriptive Statistics of categorical and numerical variables was done for that to be achieved. As well as the ability to understand the main attributes that affect the results of the price. That was done through a correlation matrix for every attribute to understand the relations between the different factors.

Afterwards when the data is organized and transformed into a form that could be processed by the machine learning technique. In this study Multiple Linear Regression model is used for building prediction model. Firstly, the data was portioned into section for training and the other part for testing, portioning percentage can be tested with different ratios to analyse different results. All models were evaluated on four evaluation matrices known as model score, Mean Square Error (MSE),

Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). From all, the Random Forest Regressor outperformed.

**Frontend Framework:**

We will use HTML,CSS, javascript framework for frontend implementation.

**Backend Framework:**

We will use FLASK (python) framework for backend implementation.

**Development Environment:**

The code implementation will take place within a Jupyter Notebook environment.

**Programming Language:**

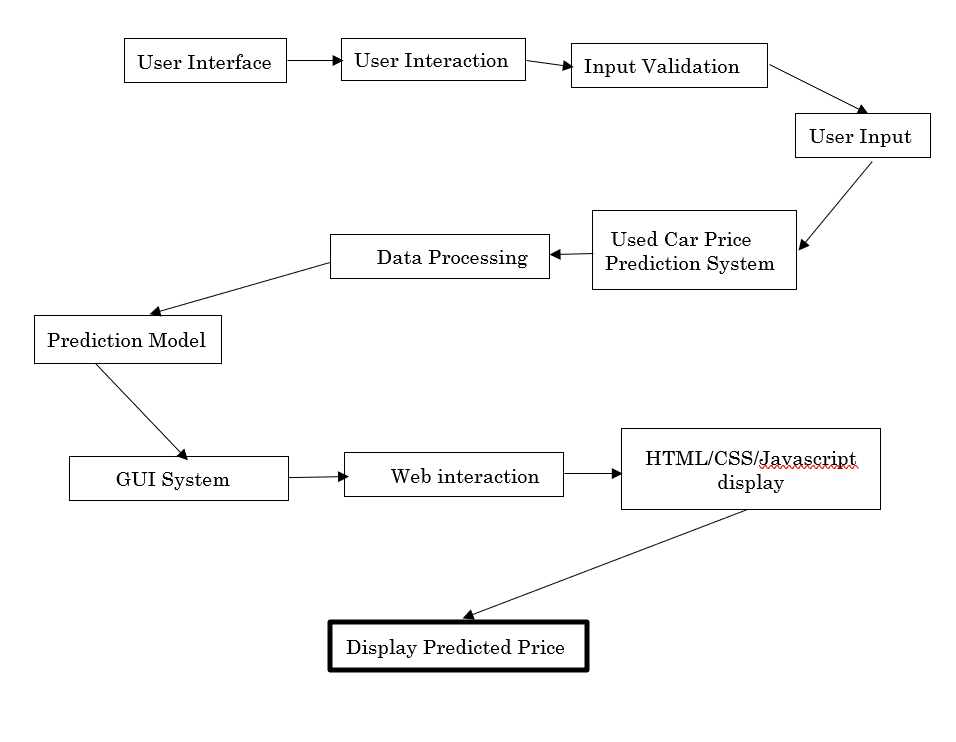
Python , javascript.

**Algorithms and Techniques:**

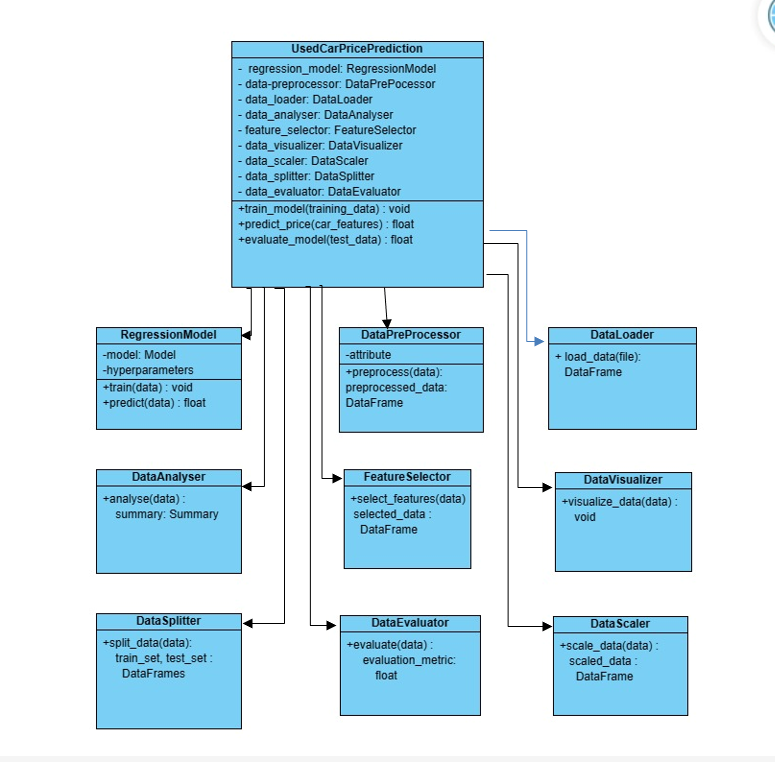
**Multiple Linear Regression**: A baseline linear regression model is implemented to establish a benchmark for comparison.

#### Project diagrams

1. **Functional Block Diagram :**

****

**2.Class Diagram:-**

****

**6. Testing**

#### Unit Testing:

#### Verified that pickling and unpickling the model using pickle does not result in data loss or corruption. Tested the log transform function for feature normalization to ensure it accurately scales input features. Confirmed the inverse transform function accurately recovers original feature values after log transformation.

#### Integration Testing:

#### Tested the integration of the pickled model within the system, ensuring seamless loading and utilization. Validated the consistent application of log transform across various stages of data processing. Ensure that inverse transform is appropriately applied to obtain interpretable predictions for users.

#### System Testing:

#### Conducted end-to-end testing to confirm the correct utilization of the pickled model in the prediction pipeline. Evaluate system behavior with varying log-transformed input features, ensuring stability and accuracy. Verify that the inverse transform effectively converts predicted log-scaled prices back to the original scale for user presentation.

#### Results and Conclusion

The used car price predictor simplifies transactions, promotes fair pricing, and aids decision-making for buyers and sellers. Its user-friendly approach and real-time insights contribute to a smoother and more transparent experience in the dynamic used car market as well it makes use of Machine Learning and Data Science algos to study the market trends and correctly predict the price of the used car.

R-square score, MSE and RMSE errors are used to evaluate the model. Following results are achieved from there.

R-square score: 0.9053702889269817

MSE: 0.140622580303109

MAE: 0.3749967737236002

The R-square of training data is almost 90% with a low RMSE which indicates that model is trained efficiently.

#### References

[1] Data Set

<https://www.kaggle.com/datasets/avikasliwal/used-cars-price-prediction?select=train-data.csv>

[2] Boston housing price prediction case study in python

<https://thinkingneuron.com/python-case-studies/>