



## STEERS4U – Car Price Estimator

---

R Deeptha, Keshav S Vardhan, Sachin N Srinivas and E Balaji

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

November 7, 2022

# STEERS4U – CAR PRICE ESTIMATOR

<sup>1</sup> R. Deeptha, <sup>2</sup> Keshav Vardhan S, <sup>3</sup> Sachin Srinivas N, <sup>4</sup> Balaji E

<sup>1, 2, 3, 4</sup> Department of Information Technology, SRM Institute of Science and Technology, Ramapuram, Chennai

E-mail: <sup>1</sup> [deepthar@srmist.edu.in](mailto:deepthar@srmist.edu.in), <sup>2</sup> [sk7032@srmist.edu.in](mailto:sk7032@srmist.edu.in), <sup>3</sup> [ss3903@srmist.edu.in](mailto:ss3903@srmist.edu.in),  
<sup>4</sup> [be5688@srmist.edu.in](mailto:be5688@srmist.edu.in)

## ABSTRACT

Buying a car has been a dream for many middle class people. Some of them may not afford a brand new car. Some of them may need a luxury car for a cheaper price [5]. That's where the second hand car market comes into the play. Many of the customers are very much confused about the genuine pricing of the cars in the second hand market. We are here to deliver a brand new Web-Application which predicts the price of the desired car the customer wishes to buy. In this project we are planning to include almost all the top manufacturers in the market and all the models present in the second hand market. The primary focus of this project is developing a Machine Learning model that can accurately predict the price of a used car based on its purchase history and kms completed. We have used various datasets to train a Machine Learning model and have deployed it as a Web-Application with seamless User-Interface and User-Experience.

**Key:** *Flask, Numpy, SciPy, Pandas, sklearn, Bootstrap, User-Interface/User-Experience.*

## 1.INTRODUCTION

For daily use, many people in the nation choose to purchase used vehicles over brand-new ones. This is a result of the abrupt increase in the cost of brand-new, future automobile models [1]. Middle class people might not be able to purchase a brand-new car, which costs more than 5 Lakhs Rupees on average, despite the fact that every manufacturer releases new models and facelifts with their own distinctive and genuinely fantastic features. Because of this, the used automobile market is booming and has recently seen a value increase of 2x. Up until October 2021, about 2,60,067 previously owned vehicles were registered, according to 2021 statistics [4]. The true pricing of a particular car model is a big source of confusion for new car buyers. For their business profit,

several third-party sellers offer very high prices. Due to the numerous elements that influence the used market price of the vehicle, determining whether the used car's quoted price is accurate is a challenging undertaking. The goal of this research is to create machine learning models that can precisely predict a used car's price based on its attributes so that buyers can make educated decisions. On a dataset made up of sales prices for various brands and models, we put several learning techniques into practise and evaluate their effectiveness. Different machine learning techniques including linear regression, ridge regression, lasso regression, and elastic net will be compared for performance. This project's best algorithm is linear regression. We calculate the cost of the car based on a number of factors. Regression techniques are used because they produce continuous values, not classified ones, which allow us to anticipate actual automobile prices rather than just price ranges and models. A user interface that accepts input from any user and displays the car's price in accordance with the user inputs has also been built.

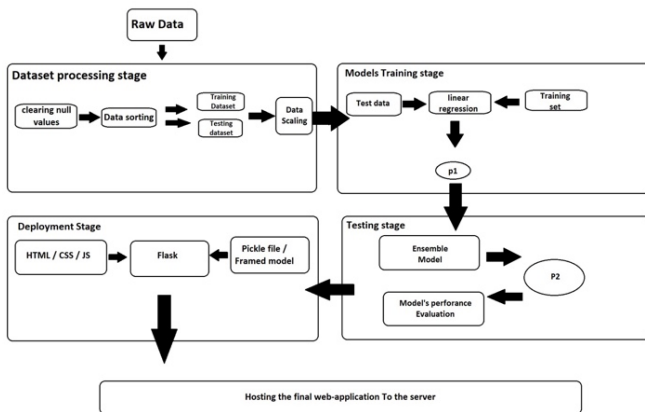
## 2.SYSTEM ANALYSIS

The proposed system of analysing the market's available inventory of cars yields unambiguous knowledge of a car's pricing depending on its model, condition, and total kilo meters driven. The high prices charged by third party sellers/dealers have not yet satisfied the automobile industry. After the COVID-19 outbreak, there is an increasing market for old cars as everyone wants to drive safely without using public transportation. By examining the model, year of registration, and other factors, the conclusion will be given that contains the best estimation of a car's price [2][3]. As a result, the data in the datasets are sorted using modern technologies like Python, Flask, etc., and a model is trained using the Linear Regression algorithm. The Flask Framework is then used to

deploy the trained model as a web application. Technologies including HTML5, CSS3, JavaScript, and Bootstrap were used to create the Web Application's front end.

- The model trained using the Linear Regression algorithm helps us in understanding and analyzing the huge amount of data.
- The Python libraries Pickle, Pandas and NumPy helps for processing and sorting the data.
- The trained model provides accurate results for the user's input.
- The User Interface works seamless which allows the user to enter the information of the vehicle and get the prediction as soon as possible.

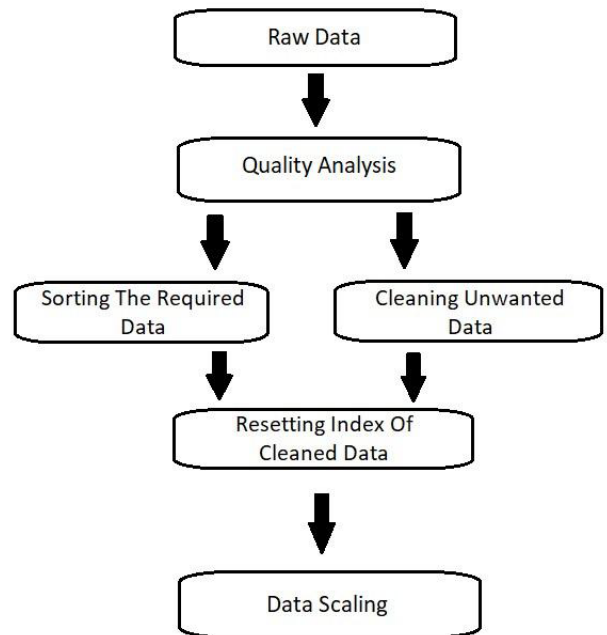
### 3.SYSTEM ARCHITECTURE



**Fig.5.1. Steers4U System Architecture Diagram**

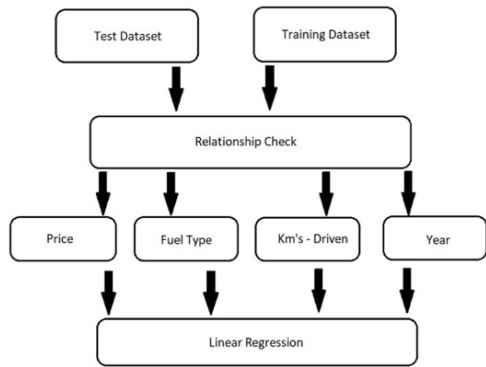
The system architecture diagram is a split into four phases, data processing phase, model training phase, testing phase and deployment phase. In the above diagram **Fig.5.1**, the raw data is processed in the first stage by clearing of null values, sorting of data and scaling the data in the dataset. The processed dataset is used to train the Machine Learning model using Linear Regression Algorithm and then the trained model is tested completely in the third phase. The trained model is then deployed as a Web Application using Flask (Python Framework) and other front-end languages like HTML3, CSS5, JavaScript and Bootstrap. Finally, the Web Application is hosted to the server.

### 4.SYSTEM MODULES



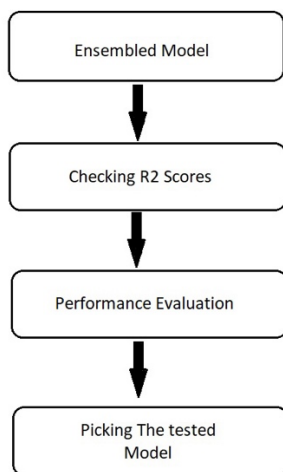
**Fig.6.1. Dataset Processing Stage**

- Given **Fig.6.1** illustrates the first step of the process begins with collection of latest datasets from various sources.
- The dataset is then imported into Jupyter Notebook by using Pandas.
- The quality of the data in the dataset is analyzed thoroughly.
- The redundancies in the dataset are cleaned by removing all the null values, changing the datatypes (object to integer), and by clearing the unwanted attributes.
- The index of the cleaned data is reset.
- The cleaned dataset is updated and saved in the computer.



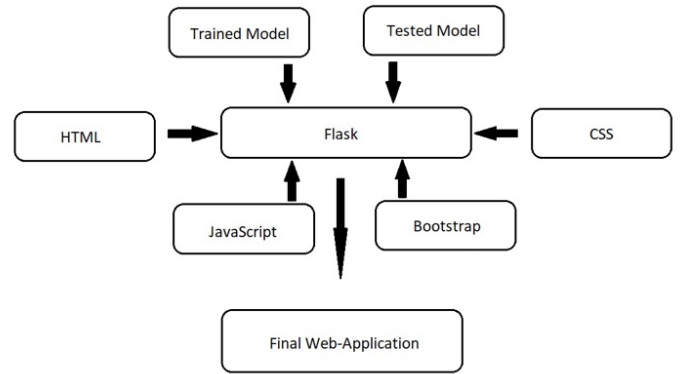
**Fig.6.2. Model Training Stage**

- As shown in the Fig.6.3 mentioned above, the relationship of company is checked with price and kms driven.
- The relationship of the year is checked with price and fuel type.
- The relationship of the price with fuel type, year and company are checked together.
- The training data is then extracted and Train-Test Split is applied using the package called sklearn.
- By using the Linear Regression Algorithm, the model is trained successfully.



**Fig.6.3. Model Testing Stage**

- In the testing stage, the ensembled model undergoes various score tests as shown in the above mentioned Fig.6.3.
- The model's performance is evaluated before it is processed to the next stage.
- Finally, the trained model is dumped as a .pkl file by using Pickle (Python Package).



**Fig.6.4. Deployment Stage**

- Given Fig.6.4 illustrates the ensembled model being deployed as a Web Application by using the Flask framework.
- HTML5, CSS3, JavaScript and Bootstrap are used for developing the User Interface of the application.
- The front-end and back-end are integrated together using Flask, and the final Web Application can be hosted to the server.

#### 4.SYSTEM IMPLEMENTATION

Collect the required datasets from the internet. Open Jupyter Notebook and import the following packages. (Pandas, Numpy, Matplotlib) Using pandas, read and cross-check the Dataset file. Call the shape and info functions to get a clear idea about the datatypes and number of rows & columns in the dataset. Create a backup copy of the dataset. Clean the unwanted attributes and null values from the dataset, sort only the required attributes – Name, Company, Year, Price, Kms Driven and Fuel Type. Save the file as “Cleaned\_Car\_Data.csv”. Call the cleaned dataset by using ‘describe’ keyword and all the statistics of the dataset are displayed. Check the relationship of the company with price. Import seaborn package and plot a graph to display the relationship of the company with price. Plot a graph to check and display the relationship of the price with year. Plot a graph to check and display the relationship of the price with kms driven. Plot a graph to check and display the relationship of the price with fuel type. Plot a graph to check and display the relationship

of the price with company. Extract the training data. Declare the variable X to the attributes – name, company, kms driven and fuel type. Declare the variable Y to the attribute – price.

```
Out[22]:
```

	name	company	year	Price	kms_driven	fuel_type
count	816	816	816.000000	8.160000e+02	816.000000	816
unique	254	25	NaN	NaN	NaN	3
top	Maruti Suzuki Swift	Maruti	NaN	NaN	NaN	Petrol
freq	51	221	NaN	NaN	NaN	428
mean	NaN	NaN	2012.444853	4.117176e+05	46275.531863	NaN
std	NaN	NaN	4.002992	4.751844e+05	34297.428044	NaN
min	NaN	NaN	1995.000000	3.000000e+04	0.000000	NaN
25%	NaN	NaN	2010.000000	1.750000e+05	27000.000000	NaN
50%	NaN	NaN	2013.000000	2.999990e+05	41000.000000	NaN
75%	NaN	NaN	2015.000000	4.912500e+05	56818.500000	NaN
max	NaN	NaN	2019.000000	8.500003e+06	400000.000000	NaN

From the package sklearn, import train\_test\_split, LinearRegression, OneHotEncoder, R2-score and make\_pipeline. Apply the Train Test Split. Create an OneHotEncoder object to contain all the possible categories. Create a column transformer to transfer all the categorical columns. Using make\_pipeline, fit the model. Check the R2 score using the package. The best model is found in certain random state. Import the pickle package. Reshape and Dump the trained model as a pickle file which will be used to provide data to the Web Application. The pickle file will be stored inside the project folder. Design a logo for the web application using Adobe Photoshop 2021 and Adobe Illustrator. Create icons and buttons which should be added on the User Interface. Create an animation for the background of the web page using Adobe After Effects 2021 to make the User Interface much more attractive to the users. Open a new HTML file in Pycharm Professional and name it as 'index.html'. Link all the CSS and JavaScript files to the HTML file in the head section. Create a folder named 'Assets' and store all the required logos, background images etc. Integrate Bootstrap CSS framework inside the HTML file. Add the background image, logo, buttons and all the required fields as forms. Finally, add the JavaScript code for communicating the inputs with the database and displaying an output to the end user. Create a virtual environment in Pycharm Professional and install the Flask framework by entering 'pip install flask' in the terminal. Name the Python file as 'application.py'. Import the packages – flask, numpy, pandas and pickle. Using pickle, load the trained model. Using pandas, read the csv file. Create a function and link

all the HTML forms to the function. Add the parameters of 'GET' & 'POST' and route the functions to the respective pages. Fetch the user's input and provide predictions using pandas. Return the render template to 'index.html' and its other attributes. Finally, run the code. The website will be running on the URL - <http://127.0.0.1:5000/>

## 5.RESULT AND DISCUSSION

From the Fig.8.1 mentioned below, we have completely analyzed and tested this project with various algorithms such as Gradient Boost and Random Forest. By clearly analyzing the R2 scores of each algorithm's test data, training data and training time, all the decision-tree algorithms did not perform well. Thus, Linear Regression becomes the most efficient algorithm.

Learning Algorithm	R <sup>2</sup> Score on Test Data	R <sup>2</sup> Score on Training Data	Training Time
Linear Regression	0.87	0.87	15 minutes
Gradient Boost	0.64	0.64	130 minutes
Random Forest	0.88	0.98	75 minutes

Fig.8.1. Comparison in efficiency of algorithms

From Fig.8.2 mentioned below, X-axis is considered as the actual price of the vehicle, Y-axis is considered as the predicted price of the vehicle and the dots are considered as data in the dataset. By looking at the figure, we can conclude that the predicted price is linearly increasing with the actual price.

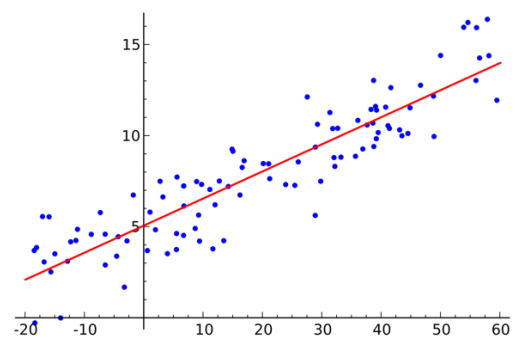


Fig.8.2. Graph of Linear Regression

Therefore, based upon the R2 scores, there is a clear statistical representation of the predictions which are fitted to regression line.

## 6.CONCLUSION

'Steers4U' helps all the used car buyers and sellers who are the end users by providing them the prediction of the price by thoroughly analyzing the model, condition, features, year of registration and kilometers completed of the vehicle which the user gives all these fields as an input. Linear Regression algorithm is used in this project which provides predictions more accurately than any other algorithm. This Application is developed in a User-Friendly way, by which any new user can also understand the key features and get benefited. This system is easy to maintain and can be easily assessed by all form of users. This application also works seamless, fast and responsive.

## 7.FUTURE ENHANCEMENTS

In the upcoming years, the electric vehicles are going to create a revolutionary change in the automobile industry. Many automobile experts predict that India will be having over 50 million electric vehicles or EVs registered by 2030 [6]. In future updates of the web application, we will try to predict the cost of EVs based on their condition. Therefore, even people who wish to buy used electric cars can also be benefited. New datasets can be updated for each and every 1 year. The User-Interface will be updated for a better experience. Integration of technologies like ReactJS or AngularJS to this project will allow users to find results much more faster. The performance of the website can be boosted by optimizing the search engines and by using APIs. Steers4U Application will also be build using technologies like Kotlin and Swift UI, which allows regular users to download it on their Android/iOS devices. Much more detailed features like "Inspect Vehicle" will be added to the application which gets more precise or accurate predictions.

## 8.REFERENCES

- [1] Deeptha, R. and Rajeswari Mukesh. (2015). Single Sign-On Mechanism for Secure Web Service Access through ISSO. Journal of communications Software and Systems. 11(1): 8-14. (ISSN: 1845-6421/03/8373 @ 2015 CCIS)
- [2] Ketan Agrahari, Ayush Chaubey, Mamoor Khan and Manas Srivastava, "Car Price Prediction Using Machine Learning"  
<https://ijirt.org/Article?manuscript=151705>
- [3] Chuyang Jin, University of Sydney, Sydney, NSW, "Price Prediction of Used Cars Using Machine Learning"  
<https://ieeexplore.ieee.org/document/9696839>
- [4] "India Used Car Market Research Report"  
<https://www.psmarketresearch.com/market-analysis/india-used-car-market>
- [5] Yuthika Bharghava, "Price Rise keeps consumers off entry-level cars"  
<https://www.thehindu.com/business/Economy/price-rise-deters-small-car-buyers-as-premium-cars-sustain-demand/article65494447.ece>
- [6] "India Used Car Market – Growth, Trends, Covid-19 Impact and Forecast (2022-2027)"  
<https://www.mordorintelligence.com/industry-reports/india-used-car-market>