TECHNO INTERNATIONAL NEW TOWN

(Formerly known as Techno India College of Technology)

Block-DG, Action Area 1, New Town, Kolkata -700156, West Bengal, India

Department of Computer Science & Engineering

Sixth Semester Project-I Report (PROJ-CS601)

Topic Name

Prepared by

Debjyoti Basak (18700120158)

Rabi Majumdar (18700120141)

Tushar Singh (18700120146)

Shirsha Mullick (18700117032)

Academic Year: 2022-2023

Stream: - Computer Science & Engineering

Year of Study: 3rd

Affiliated to

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WESTBENGAL

(FORMERLY KNOWN AS WEST BENGAL UNIVERSITY OF TECHNOLOGY)

CERTIFICATE

This is to certify that the project report entitled "Old Vehicle Price Estimation using Machine Learning and Python" has been prepared by the following students of the Department of Computer Science & Engineering under my supervision in partial fulfillment for the degree of Bachelor of Technology (B.Tech.) in Computer Science & Engineering which is affiliated to Maulana Abul Kalam Azad University of Technology, West Bengal (Formerly known as West Bengal University of Technology) in the academic year 2022-2023.

It is to be understood that by this approval, the undersigned does not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn thereof, but approves the report only for the purpose for which it has been submitted.

Sl. No.	Name	Roll Number
1	Debjyoti Basak	18700120158
2	Rabi Majumdar	18700120141
3	Tushar Singh	18700120146
4	Shirsha Mullick	18700117032

…….
Prof. Bikash Sadhukhan,
HOD, Computer Science & Engineering,
Techno International New Town

Abstract

The manufacturer sets the price of a new car in the market, with some additional expenses paid by the government in the form of taxes. Customers who purchase a new car can be sure that their investment will be worthwhile. But because new cars are becoming more expensive and consumers can no longer afford to acquire them, used car sales are rising everywhere. Consequently, there is a critical need for a system that accurately assesses the value of the vehicle utilising a number of features.

The current system involves a procedure where a vendor chooses a price arbitrarily and the buyer is unaware of the car and its current market value. In actuality, neither the seller nor the price at which he ought to sell the car have any notion of the current value of the vehicle. We have created a model that will be quite effective in resolving this issue. Regression algorithms are employed because, as opposed to categorical values, their output is a continuous value.

As a result, it will be feasible to forecast the exact cost of an automobile rather than its price range. A user interface that accepts input from any user and displays the price of a car based on their inputs has also been developed.

CONTENTS

LIST OF FIGURES

1. Introduction

Due to the numerous variables that affect a used automobile's market price, figuring out whether the quoted price of a used car is accurate is a difficult undertaking. This project focuses on creating machine learning models that can precisely forecast the price of a used automobile depending on its qualities, allowing users to make educated decisions. On a dataset comprised of the sale prices of various brands and models, we apply and assess several learning methodologies. The performance of different machine learning algorithms, including Linear Regression, Ridge Regression, Lasso Regression, Elastic Net, and Decision Tree Regressor, will be compared, and the best one will be selected.

The cost of the automobile will be determined based on a number of factors. Because regression algorithms provide us a continuous number rather than a categorized value as an output, it is feasible to estimate a car's exact price rather than just its price range. A user interface that accepts input from any user and shows the price of a car based on their inputs has also been developed.

2. Problem Definition

- Problem Statement: The goal of this project is to create a machine learning model that can reliably justifies the price of a used car using historical data on old automobiles, including parameters like brand, model, manufacturing year, fuel type, distance driven and other important criterias.

- Input: A dataset containing historical data of old vehicles, including features such as company, model, manufacturing year, fuel type, distance driven and other relevant factors.

- Output: Predicted price as per the dataset model of the used vehicle.

- Challenges: Predicting the price of used cars can be difficult for a number of reasons, such as dealing with incomplete or missing data, handling categorical variables like make and model, managing outliers, dealing with potential multicollinearity among features, and choosing a machine learning algorithm that best fits the data.

- Objective: The main goal of this issue is to create a trustworthy machine learning model that can precisely forecast the price of used cars so that prospective buyers or sellers may make educated choices.

3. Related Work

The enormous amount of study into used-car price prediction systems, Listiani (2009) discussed a regression model based on machine learning methods like SVMs (Support Vector Machines) in his M.Sc. thesis report. To more accurately predict the residual value of a leased car than multivariate regression. The SVM approach can handle very high dimensional data and is accurate for existing datasets with few dimensions. There is less chance that issues like underfitting and overfitting would arise. According to Listiani (2009), they use a genetic algorithm to identify the best variable or parameters for their research investigation. According to Nur et al. (2021; Pérez et al. (2018), this research has a limitation in that the variation of regression using advanced Support Vector Machines regression was not expressed in terms of central tendency of metrics like mean, standard deviation, Moments, skewness, correlation, and variance.

Richardson (2009) outlined the work done on the construction of long-lasting, robust automobiles by automobile manufacturers in his thesis. In order to prove that hybrid automobiles are more value than other vehicles, a multiple regression analysis was utilized. Because of their greater fuel economy, hybrid engines are used to address environmental issues including climate change and global warming. During the thesis report investigation, they took into account variables including the car's age, manufacturer or brand, MPG (miles per gallon), and mileage (Lodarosi, 2020; Nyandra et al., 2018).

Wu et al. (2009) developed a neuro-fuzzy knowledge-based research used automobile pricing prediction system. Three factors were taken into consideration: brand, year of manufacture, and engine style. According to Du et al. (2009), leasing is a common method of selling automobiles at US automotive dealerships. They offered an ODAV (Optimal Distribution of Auction Vehicles) expert prediction method, and the findings of their model are quite similar to those of a straightforward regression. For the purpose of predicting the price of a leasing automobile, they employed the Regression model, which is based on the KNN (K Nearest Neighbor) ML (machine learning) approach. Over 2 million automobiles were distributed effectively using this technique (Shang et al., 2014; Bowling & Veloso, 2002).

4. Methodology and Architecture

Fig 1: Model Architecture

- Data collection:

  The technique employed for this project begins with the collecting of unprocessed data into an online dataset or logical dataset. The following step is making an environment for data cleansing and minimization during pre-processing. We train and test for data analysis after obtaining meaningful information from the dataset.
  Finally, a machine learning model is chosen for price prediction and categorization using artificial intelligence based on this result.

- Preparing Dataset:

  Here we have used dataset of the website "Quickrr" for the old vehicle. The followings are used in this dataset:

  - Name: The name is referred to as used car models.
  - Company: The company from which the car belongs to.
  - Year: The manufacturing year of the car.
  - Price: 2nd hand price of the car.
  - Kms_driven: The total number of kilometres the used car has been driven.
  - Fuel_type: the fuel used by car CNG, or Petrol, or Diesel.

	name	company	year	Price	kms_driven	fuel_type
0	Hyundai Santro Xing XO eRLX Euro III	Hyundai	2007	80,000	45,000 kms	Petrol
1	Mahindra Jeep CL550 MDI	Mahindra	2006	4,25,000	40 kms	Diesel
2	Maruti Suzuki Alto 800 Vxi	Maruti	2018	Ask For Price	22,000 kms	Petrol
3	Hyundai Grand i10 Magna 1.2 Kappa VTVT	Hyundai	2014	3,25,000	28,000 kms	Petrol
4	Ford EcoSport Titanium 1.5L TDCi	Ford	2014	5,75,000	36,000 kms	Diesel

Fig 2: Preparing Dataset

- Analyzing Data:

In this collected data, there are many inconsistencies, like:

- names are pretty inconsistent
- names have company names attached to it
- some names are spams like 'Maruti Ertiga showroom condition with' and 'Well maintained Tata Sumo'
- company: many of the names are not of any company like 'Used', 'URJENT', and so on.
- year has many non-year values
- year is in object. Change to integer
- Price has Ask for Price
- Price has commas in its prices and is in object
- kms_driven has object values with kms at last.
- It has nan values and two rows have 'Petrol' in them
- fuel_type has null values

Fig 3: Data Analysis

- Cleaning Data:

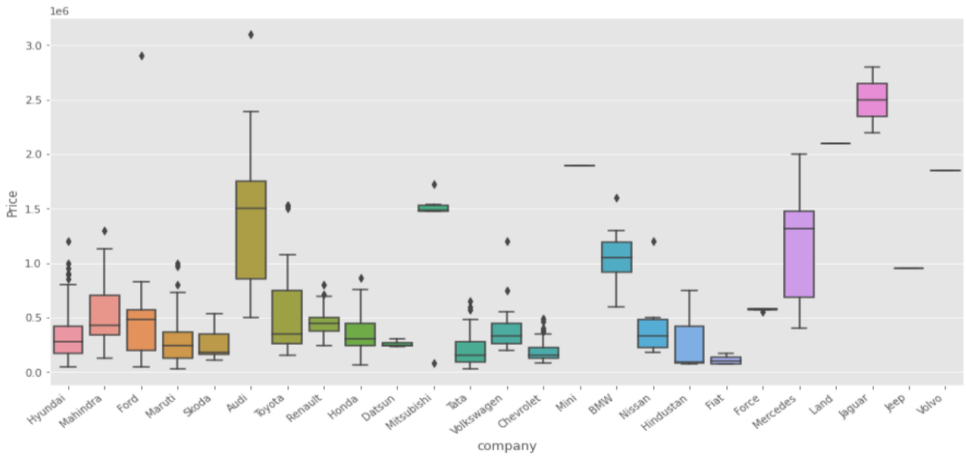
To avoid such inconsistencies, we need to clean the data and arrange it in a significant or desired way. We cleaned out all the data and now it looks like this:

	name	company	year	Price	kms_driven	fuel_type
0	Hyundai Santro Xing	Hyundai	2007	80000	45000	Petrol
1	Mahindra Jeep CL550	Mahindra	2006	425000	40	Diesel
2	Hyundai Grand i10	Hyundai	2014	325000	28000	Petrol
3	Ford EcoSport Titanium	Ford	2014	575000	36000	Diesel
4	Ford Figo	Ford	2012	175000	41000	Diesel
...	...	...	...	...	...	...
811	Maruti Suzuki Ritz	Maruti	2011	270000	50000	Petrol
812	Tata Indica V2	Tata	2009	110000	30000	Diesel
813	Toyota Corolla Altis	Toyota	2009	300000	132000	Petrol
814	Tata Zest XM	Tata	2018	260000	27000	Diesel
815	Mahindra Quanto C8	Mahindra	2013	390000	40000	Diesel

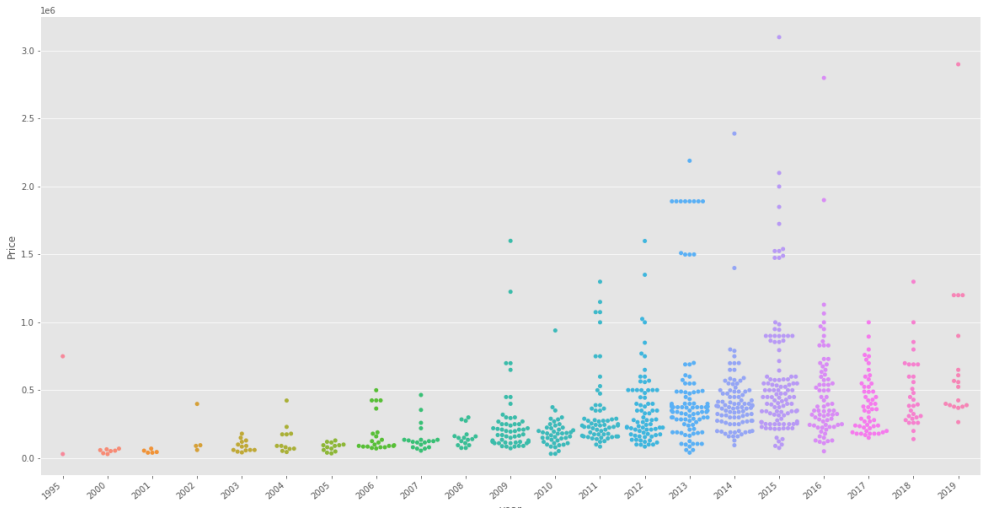
816 rows × 6 columns

Fig 4: Cleaned Data

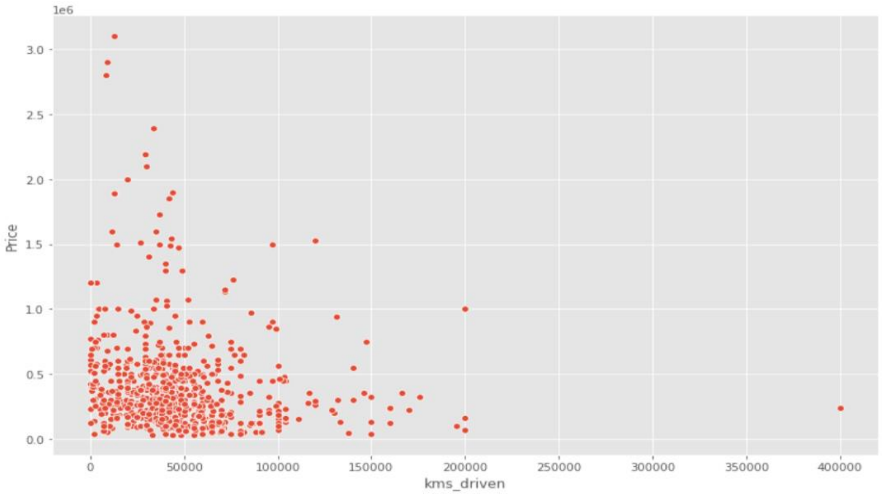
- Checking Relationship of Company with price:



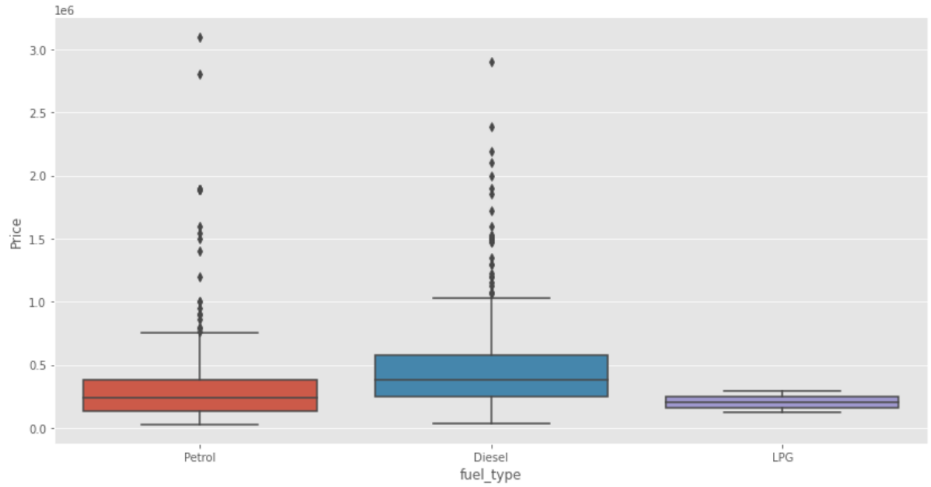
- Checking Relationship of Year with price:



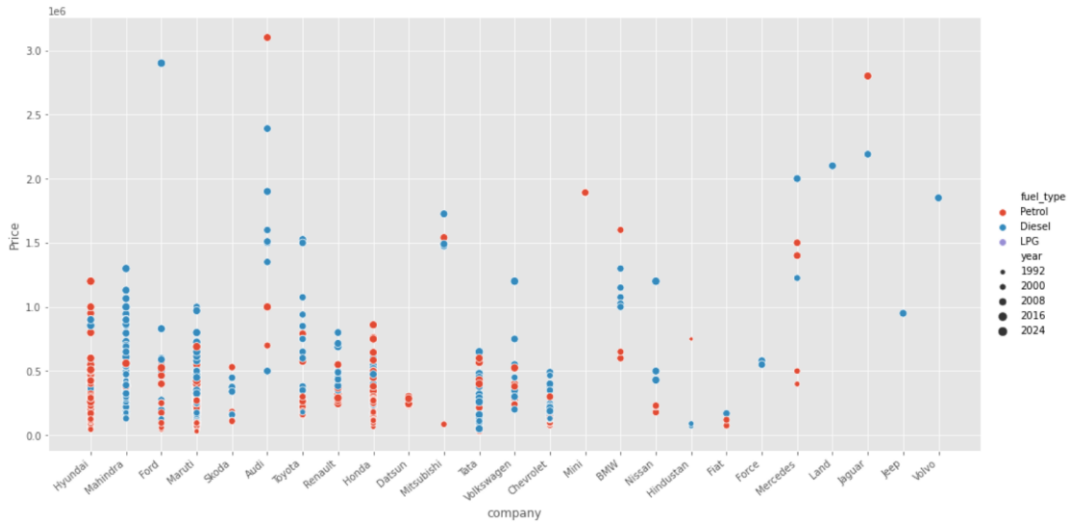
- Checking Relationship of kms_driven with price:



- Checking Relationship of fuel_type with price:



- Relationship of Price with FuelType, Year and Company mixed:



- Linear Regression:

Linear regression is a statistical tool used to model the relationship between two variables, where one variable is considered the dependent variable, and the other is considered the independent variable. The goal of linear regression is to fit a straight line that best represents the linear relationship between the variables, allowing for prediction of the dependent variable given the independent variable(s). Here we used to take the average of all the possible prices according to the data with the help of LRM.

5. Future Scope of the Project

The future scope of old vehicle price prediction using machine learning is very promising and may have applications beyond just the automotive industry. Potential growth areas include:

- Used car market:

  As the demand for used cars continues to grow, accurately predicting used car prices can provide valuable insight to buyers and sellers in the used car market. Machine learning models can analyze historical data about vehicle attributes such as make, model, year, mileage, condition, location, as well as market trends and economic factors to predict the optimal price range for a given vehicle. This allows buyers and sellers to make informed decisions and negotiate a fair price.

- Auto insurance:

  Predicting the value of older vehicles also benefits the auto insurance industry. Insurers can use machine learning models to accurately estimate the market value of vehicles. This helps determine premium and claim payments. Accurate price predictions also help detect fraud by identifying exaggerated claims related to vehicle value.

- Fleet management:

  Fleet managers often need to assess the value of vehicles in order to make decisions about buying, selling, or retaining them in their fleet. A machine learning-based pricing forecasting model provides fleet managers with data-driven insights into vehicle value, enabling them to optimize fleet operations and financial planning.

- Financial operations:

  Banks, credit unions, and other financial institutions often need to value older vehicles for loan or lease purposes. Machine learning models help provide accurate vehicle valuations based on historical data and current market trends, helping financial institutions make informed decisions about credit terms, interest rates, and collateral valuations.

- Personalized vehicle evaluation service:

  Advances in digital technology may lead to the proliferation of personalized vehicle valuation services. Users can enter specific details of their older vehicle into the online platform, and machine learning models can generate estimates based on historical data and market trends. This gives individual vehicle owners valuable insight when selling or trading vehicles.

Overall, the future scope of old vehicle price prediction using machine learning is vast, with potential applications in various industries, including the used car market, automotive insurance, fleet management, financial services, and personalized vehicle valuation services. Accurate price prediction can provide data-driven insights, facilitate fair transactions, and optimize decision-making processes in the automotive industry and beyond.

6. Conclusion

In summary, using machine learning to predict used car prices is promising for a variety of applications. By leveraging historical data, market trends, and vehicle attributes, machine learning models provide valuable insights to buyers, sellers, insurance companies, fleet managers, financial institutions, and individual vehicle owners. Accurate price forecasts promote fair trading, streamline fleet operations and financial planning, help detect fraud, and enable users to make data-driven decisions. As technology continues to advance and more data becomes available, the future scope of used car price prediction may expand, offering new opportunities for the automotive and related sectors.

References

[1] Sameerchand Pudaruth, “Predicting the Price of Used Cars using Machine Learning Techniques”;(IJICT 2014)

[2] Enis gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric, ”Car Price Prediction Using Machine Learning”; (TEM Journal 2019)

[3] 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)

[4] Nitis Monburinon, Prajak Chertchom, Thongchai Kaewkiriya, Suwat Rungpheung, Sabir Buya, Pitchayakit Boonpou, “Prediction of Prices for Used Car by using Regression Models” (ICBIR 2018)

[5] Doan Van Thai, Luong Ngoc Son, Pham Vu Tien, Nguyen Nhat Anh, Nguyen Thi Ngoc Anh, “Prediction car prices using qualify qualitative data and knowledge-based system” (Hanoi National University

----------X----------