CAR RESALE VALUE

1. INTRODUCTION

1.1 Overview:

The study on automotive resale value aims to forecast and evaluate the resale value of autos. Its main objective is to give automobile buyers and sellers knowledge and advice about the anticipated depreciation and future value of a vehicle.
As part of the project, a variety of data factors that affect a car's resale value must be gathered and analysed. The make, model, year, mileage, condition, prior ownership, location, market trends, and other pertinent variables are just a few examples of the data pieces that could be included.
The analysed data is then placed into a predictive model that uses cutting-edge statistical and machine learning methods to produce precise forecasts of resale value. The model considers previous market patterns, economic variables, and particular vehicle characteristics.

1.2 Purpose:

The Car Resale Value project offers a number of benefits and practical applications for various automobile prospects and car owners. Here are some of the key uses and results that can be achieved with this project: 1. Informed buying and selling decisions: The car resale value project allows potential buyers to estimate the future value of the vehicle before making a purchase. This allows them to assess whether a particular car is a good investment and make informed decisions based on budgets and long-term plans. Likewise, sellers can determine a fair price for their vehicle, taking into account resale value factors, and negotiate better deals.
2. Financial Planning: Knowing the resale value of your car allows for effective financial planning. Car owners can estimate how their vehicle will depreciate in value over time, allowing them to budget for future expenses or plan for a possible trade-in or upgrade. Understanding expected resale value can also be helpful in determining insurance coverage and calculating the total cost of owning a car.
3.Trade-ins and Upgrades: The Car Resale Value project can help people who are thinking about replacing their current vehicle with a new one. By knowing the approximate value of their existing car, they can negotiate a fair exchange deal and understand the financial implications of upgrading to a newer model.

4. Used Car Market Analysis: Automotive professionals such as car dealerships and car rental companies can use the used car market analysis project. This information can help them with inventory management and pricing strategies, and make informed decisions about buying or selling used cars.
5. Residual Value Determination: The project can be used by leasing companies and financial institutions to estimate the residual value of leased vehicles. This allows you to estimate the future value of the vehicle at the end of the lease term, allowing you to determine suitable rental rates and terms.
6. Customization and Modification Planning: Car enthusiasts or owners interested in customising or modifying their vehicles can benefit from a project overview.You can evaluate how certain changes may affect resale value and make decisions accordingly.

2. LITERATURE SURVEY

2.1 Existing problem :

To solve the problem of predicting car resale value, the following methodology can be used:

1. Data collection: Collect a comprehensive data set containing historical information on car sales and resale value. The record should include relevant attributes such as make, model, year, mileage, condition, previous owner, location, and any other resale value factors.

2. Data Preprocessing: Cleans and preprocesses the collected data to handle missing values, outliers, and inconsistencies.This step can include data normalisation, feature scaling, and coding of categorical variables.

3. Feature Engineering: Extract relevant features from your dataset that can help predict a car's resale value. This can include creating new variables from existing ones, such as calculating a car's age or deriving additional functions that capture the vehicle's condition or service history.

4. Data Split: Split your data set into training and test sets. The training set is used to train the predictive model, while the test set evaluates the model's performance using invisible data.

5. Model Selection: Select the appropriate predictive modelling technique based on the nature of the problem and the characteristics of the data. Commonly used models for regression tasks include linear regression, decision trees, random forests, gradient improvement, and neural networks.

6. Train model: Train the selected model using the training data set. The model learns patterns and relationships between the input features and the car's corresponding resale values.

7. Model Evaluation: Evaluate the trained model against a set of test data to assess its performance and accuracy.Common evaluation metrics for regression tasks include mean square error (MSE), mean square error (MAE), and R-squared value.

8. Model Refinement: Adjust the model by optimising its hyperparameters or considering ensemble techniques to improve its predictive performance. This step may involve cross-validation or more advanced optimization methods.

9. Prediction and Interpretation: Once the model has been trained and refined, it can be used to predict the resale value of new and unreleased vehicle data. Model predictions can provide valuable insights into the expected value of a car given its specific characteristics.

10. Continuous Improvement: Regularly update the model with new data to ensure it remains current and reflects the latest market trends and dynamics affecting car resale value. Refine the model as needed based on feedback and new insights from real-world predictions.

2.2 Proposed solution:

Decision Trees: Decision trees are a machine learning algorithm that can be used for regression tasks. They partition the data based on different attributes and create a tree-like structure to make predictions. Decision trees are interpretable and can capture non - linear relationships between car attributes and resale value. Techniques like random forests and gradient boosting, which combine multiple decision trees, are also commonly used.

Support Vector Regression (SVR): SVR is a variant of support vector machines (SVM) that is used for regression tasks. It aims to find a hyperplane that best fits the data while maximizing the margin. SVR can handle non-linear relationships by applying kernel functions to transform the data into higher-dimensional space.

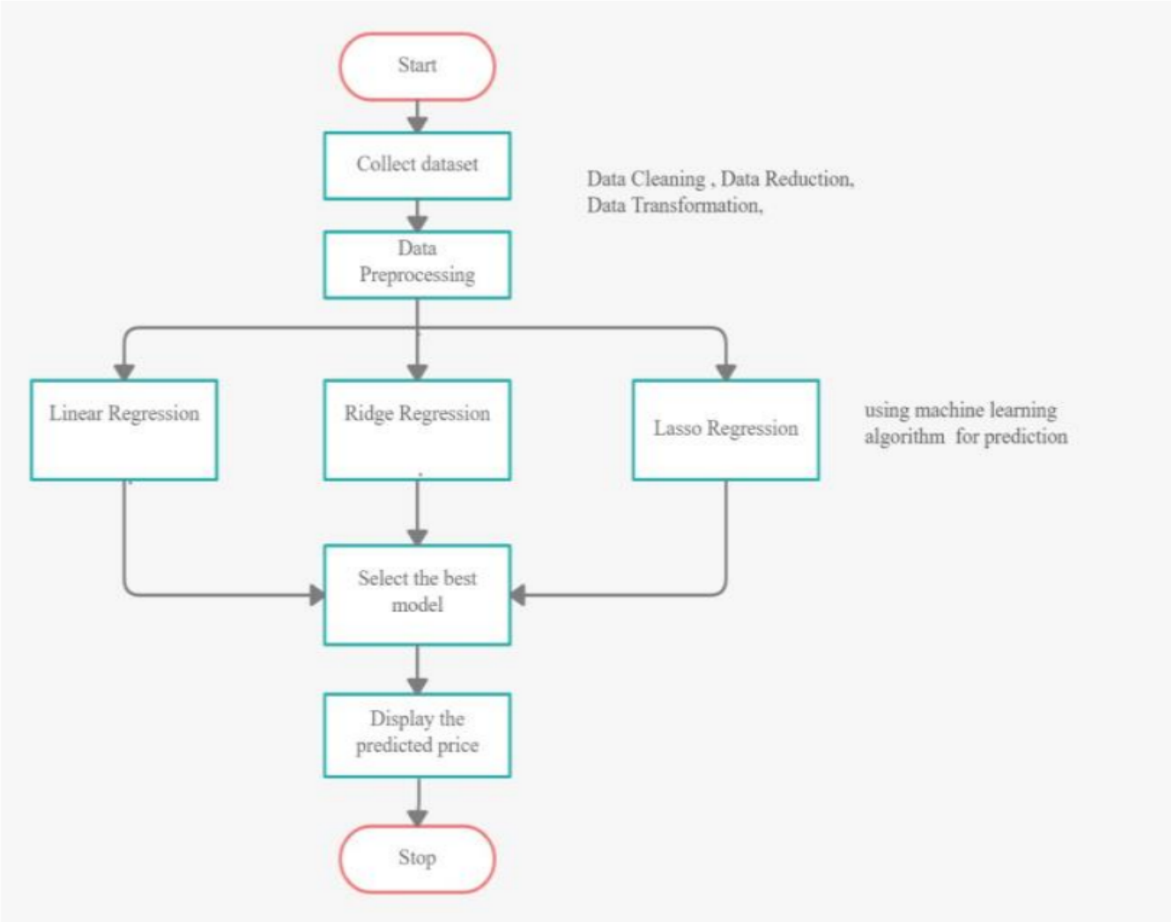
Multi-linear regression: Multi-linear regression is a statistical modelling technique used to estimate the relationship between a dependent variable and multiple independent variables. In the context of car resale value prediction, multi-linear regression can be applied to understand how various car attributes or factors influence the resale value.

The k-nearest neighbours (k-NN): The k-nearest neighbours (k-NN) algorithm is a machine learning technique used for classification and regression tasks. It is a non-parametric algorithm that makes predictions based on the similarity of instances in a training dataset.

Random forest method: The random forest method is a popular machine learning algorithm that is widely used for both classification and regression tasks. It is an ensemble learning technique that combines multiple decision trees to make predictions. Random forests are known for their accuracy, robustness, and ability to handle complex datasets.

3.THEORETICAL ANALYSIS

3.1 Block diagram:



3.2 Hardware / Software designing:

Hardware requirements
1. Operating system- Windows 7,8,10
2. Processor- dual core 2.4 GHz (i5 or i7 series Intel processor or equivalent AMD) RAM-4GB

Software Requirements
1. Python in Jupyter Notebook
2. Flask Framework
3. HTML CSS in spyder
4. Command prompt

4. EXPERIMENTAL INVESTIGATIONS:

During the method involved with chipping away at an answer for vehicle resale esteem expectation, a few examinations and examinations are ordinarily performed. These expect to grasp the information, recognize pertinent highlights, select suitable models, and assess the exhibition of the arrangement. Here are a few normal examinations and examinations led during the improvement of a vehicle resale esteem arrangement:

1. EDA, or exploratory data analysis,: Perform a preliminary examination of the dataset to comprehend the relationships between various features, discover missing values or outliers, and gain insight into the distribution of the data. In order to comprehend the characteristics of the data, EDA uses techniques for data visualisation, statistical analysis, and summary statistics.

2. Engineering and Selection of Features: Examine the pertinence and meaning of various elements in foreseeing vehicle resale esteem. Correlation analysis, feature importance measures from machine learning models, or domain knowledge expertise may be used to select the most informative features in this situation. In addition, feature engineering can be used to enhance the predictive power of existing features or to develop brand-new ones.

3. Choosing a Model: Assess different AI models or calculations reasonable for vehicle resale esteem expectation. Different models, such as linear regression, decision trees, random forests, support vector machines (SVM), and neural networks, can be compared in terms of performance, strengths, and weaknesses. In the selection process, take into account interpretability, computational efficiency, and scalability.

4. Model Assessment: Utilise appropriate evaluation metrics to evaluate the chosen model's performance. For relapse errands, measurements like mean squared mistake (MSE), mean outright blunder (MAE), or R-squared worth can be utilised to gauge the precision of the expectations. For more accurate performance estimates, cross-validation methods like k-fold cross-validation can be used.

5. Tuning the Hyperparameters: Examine the performance effects of various hyperparameters of the chosen model(s). Use methods like framework search or arbitrary inquiry to efficiently investigate various blends of hyperparameter esteems and select the ideal ones that yield the best outcomes.

6. Analyses of the Rest: To comprehend the model's performance, examine the residuals—the differences between the predicted and actual values. Look at the circulation of residuals, check for any

examples or inclinations, and guarantee the shortfall of critical mistakes or exceptions.

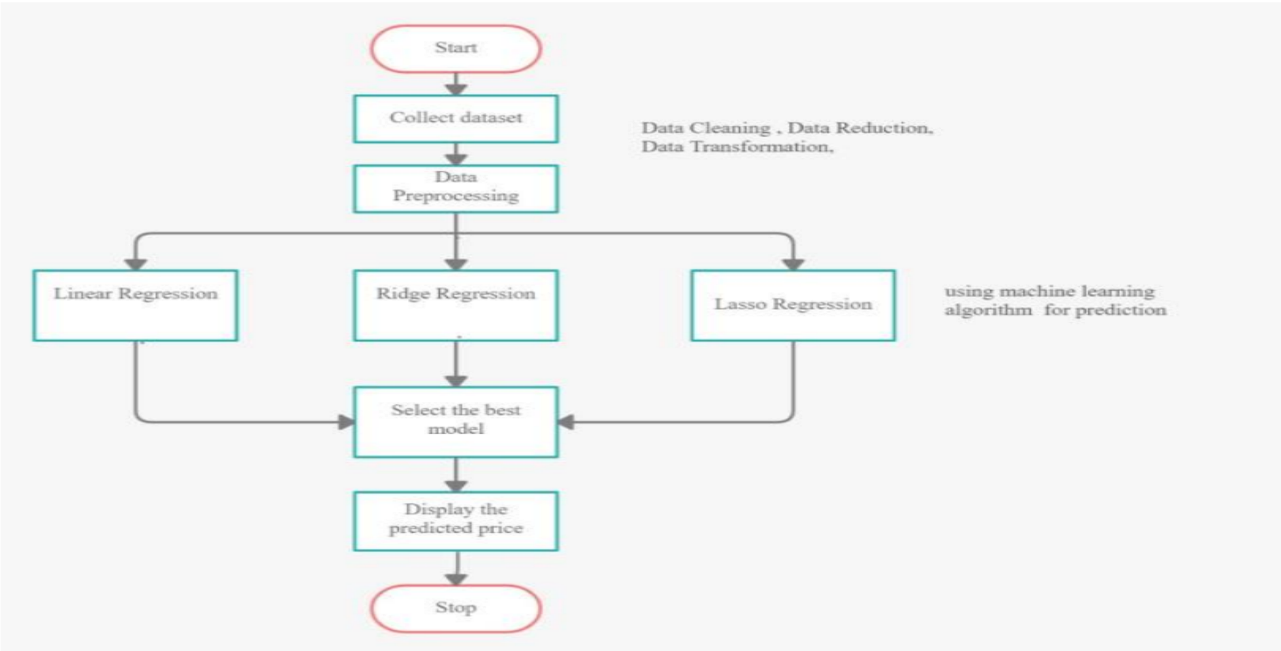
7. Explanation and Interpretability: Examine strategies to decipher and make sense of the expectations made by the model. This might include procedures, for example, highlight significance investigation, halfway reliance plots, or model-sceptic understanding strategies like LIME or SHAP values.

8. Responsiveness Investigation: Lead awareness examination to evaluate the vigour of the model to changes in input factors. Decide what varieties in specific elements or their qualities mean for the anticipated resale esteem and grasp the model's aversion to various variables.

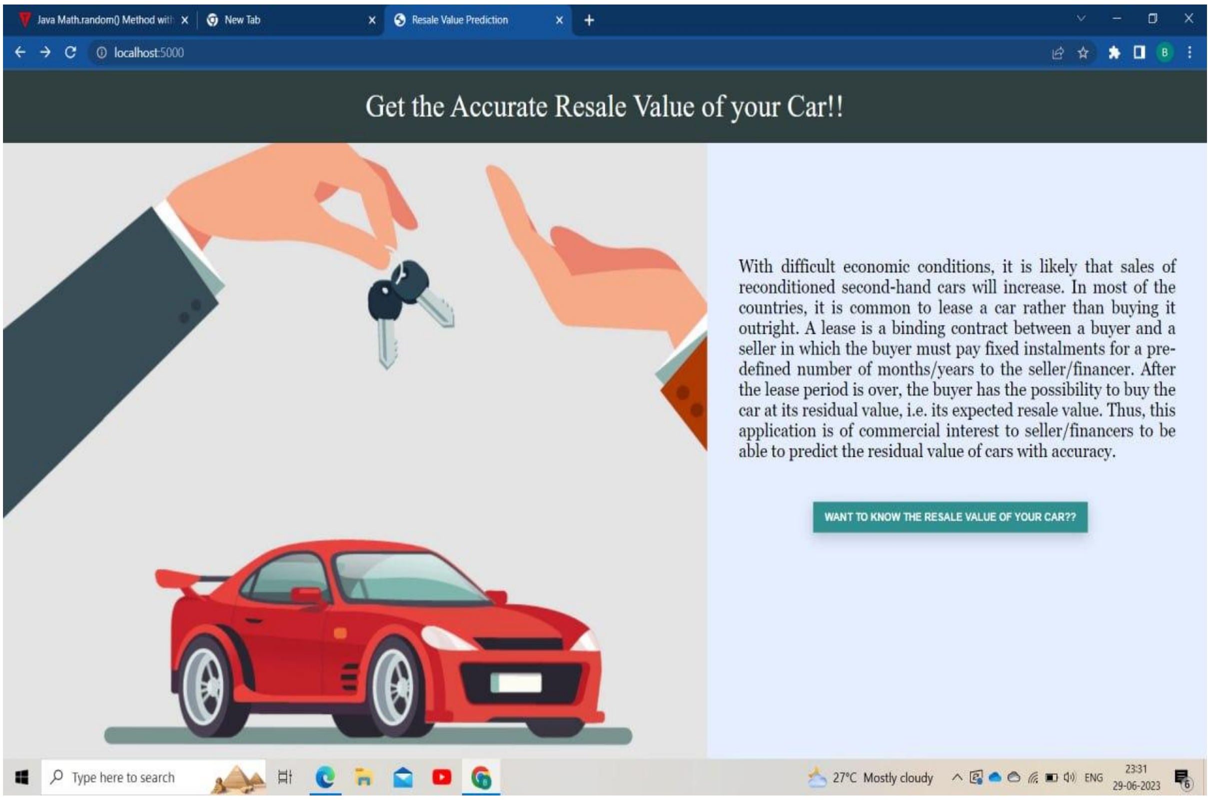
9. Analyses Comparatively: Analyse the exhibition of the created model with other existing techniques or benchmarks in the space of vehicle resale esteem expectation. The solution's competitiveness and effectiveness can be assessed with the assistance of this analysis.

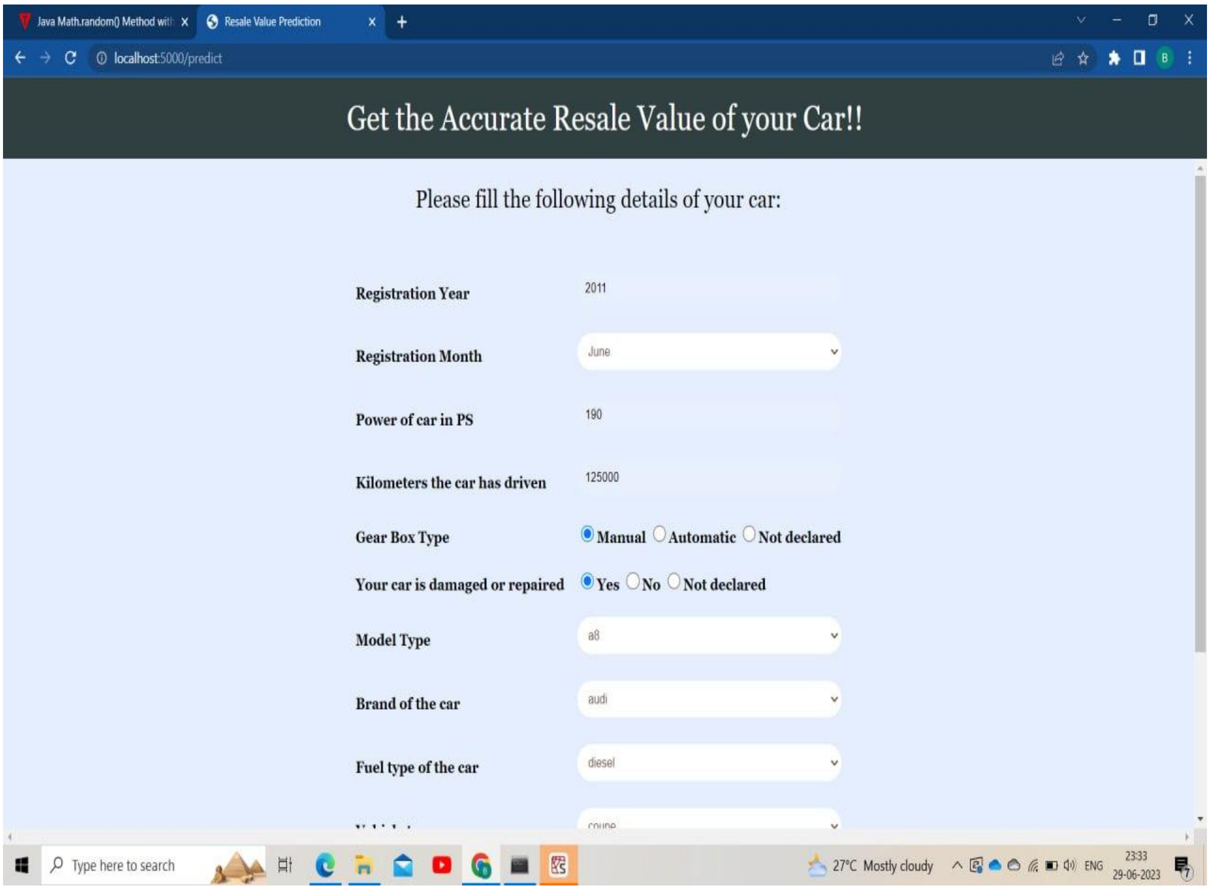
10. Mistake Investigation: Research the examples and wellsprings of blunders or errors in the expectations. Recognize any deliberate inclinations, constraints, or regions where the model may perform inadequately. This examination can direct further enhancements or refinements to the arrangement.

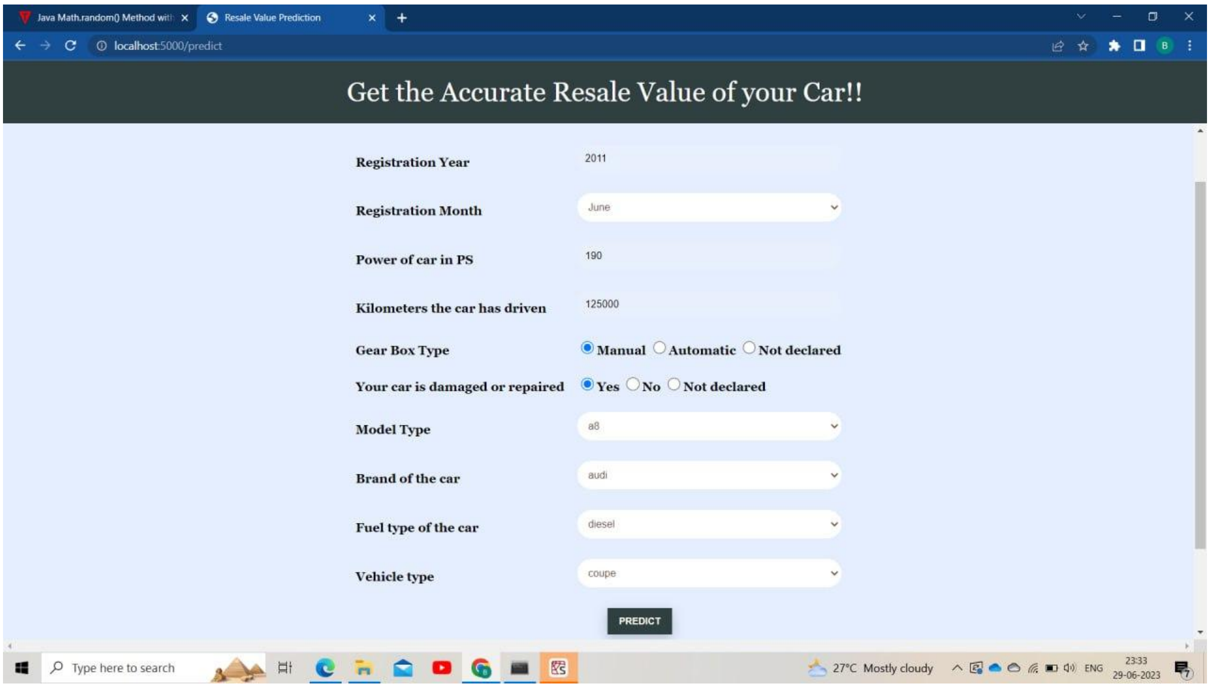
5. FLOWCHART:

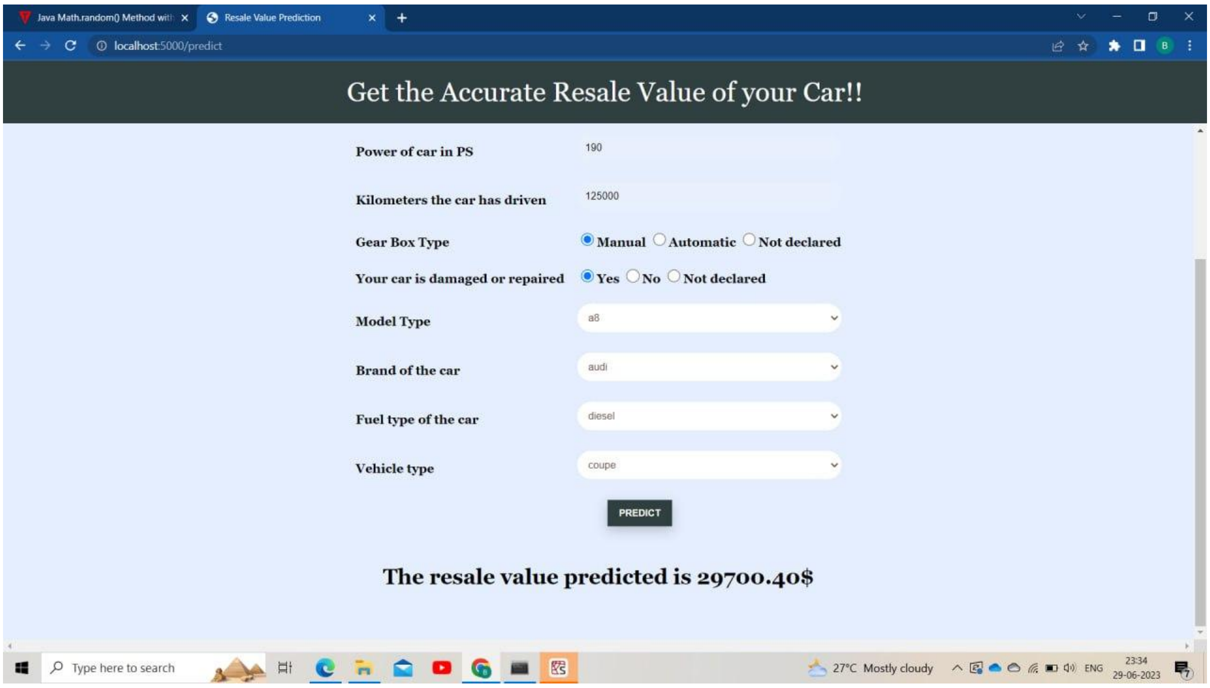


6.RESULT:


Java Math.random() Method wit
New Tab
Resale Value Prediction
localhost:5000
Get the Accurate Resale Value of your Car!!
With difficult economic conditions, it is likely that sales of reconditioned second-hand cars will increase. In most of the countries, it is common to lease a car rather than buying it outright. A lease is a binding contract between a buyer and a seller in which the buyer must pay fixed instalments for a pre-defined number of months/years to the seller/financer. After the lease period is over, the buyer has the possibility to buy the car at its residual value, i.e. its expected resale value. Thus, this application is of commercial interest to seller/financers to be able to predict the residual value of cars with accuracy.
WANT TO KNOW THE RESALE VALUE OF YOUR CAR??
Type here to search
27°C Mostly cloudy
ENG
23:31
29-06-2023


Java Math.random() Method wit
Resale Value Prediction
localhost:5000/predict
Get the Accurate Resale Value of your Car!!
Please fill the following details of your car:
Registration Year
2011
Registration Month
June
Power of car in PS
190
Kilometers the car has driven
125000
Gear Box Type
Manual
Automatic
Not declared
Your car is damaged or repaired
Yes
No
Not declared
Model Type
a8
Brand of the car
audi
Fuel type of the car
diesel
Type here to search
27°C Mostly cloudy
ENG
23:33
29-06-2023





7.ADVANTAGES & DISADVANTAGES

Advantages of Car Resale Value Project:

7.1: Financial Planning: By having insights into the resale value of a car, individuals can better plan their finances. They can consider the potential depreciation and factor it into their budgeting and long-term financial goals.

7.2 : Enhanced Negotiation Power: Knowledge of the car's resale value gives individuals an advantage during negotiations. They can use this information to negotiate a fair price when buying or selling a vehicle, ensuring they get the best deal possible.

7.3 : Residual Value Estimation: Car resale value projects enable the estimation of a car's residual value, which is particularly useful for businesses or individuals who lease vehicles. It helps them assess the future worth of the car and make informed decisions regarding leasing terms.

Disadvantages of Car Resale Value Project:

7.4 : Data Availability and Quality: Obtaining accurate and reliable data for car resale value predictions can be difficult. The availability of comprehensive and up-to-date data can be limited, making it challenging to build robust models. Additionally, the quality of data, including missing values or inconsistencies, can impact the accuracy of the predictions.

7.5 : Model Limitations: The accuracy of car resale value predictions heavily depends on the chosen modelling techniques and algorithms. No model can perfectly predict future car values, and different models may have their limitations and biases. It is crucial to understand the assumptions and limitations of the chosen model and interpret the results accordingly.

7.6 : Individual Factors: Car resale values can vary based on individual factors such as the car's condition, maintenance history, and unique features. These factors are not always captured in general models and may require additional manual evaluation or expert judgement.

8. APPLICATIONS:

The solution for car resale value prediction can be applied in various areas related to the automotive industry and car ownership. Here are some key areas where this solution can be beneficial:

1. Car Dealerships: Car dealerships can utilize the car resale value prediction solution to provide accurate estimates to customers who are trading in their current vehicles. This helps in determining fair trade-in values and facilitating smoother transactions.

2. Car Buyers: Individuals looking to purchase a used car can benefit from the resale value prediction. It enables them to assess the potential depreciation and estimate the future value of the vehicle they intend to buy, helping them make informed decisions and negotiate prices.

3. Car Sellers: Sellers can leverage the car resale value prediction to determine the appropriate selling price for their vehicles. It provides insights into the estimated value of their cars, considering factors such as age, mileage, condition, and market trends, enabling them to set competitive prices.

4. Car Leasing Companies: Car leasing companies can use the solution to estimate the residual value of vehicles. This information helps them determine lease terms and pricing, facilitating better financial planning and decision-making.

5. Insurance Companies: Insurance companies can incorporate car resale value predictions into their risk assessment and pricing models. Accurate estimation of a vehicle's value aids in setting appropriate insurance premiums and managing claims related to total loss or theft.

6. Automotive Manufacturers: Manufacturers can benefit from car resale value predictions to understand market trends and consumer preferences. This information helps them in product planning, determining the residual value of leased vehicles, and assessing the competitiveness of their models in the used car market.

7. Financial Institutions: Banks and lending institutions can consider car resale value predictions when evaluating loan applications related to car financing. It helps them assess the collateral value of the vehicle and make informed decisions regarding loan terms and interest rates.

8. Fleet Management: Companies managing vehicle fleets can utilize car resale value predictions to optimise their fleet turnover. It assists in determining the optimal time to replace vehicles, balancing operational costs with resale values.

9. Car Valuation Services: Organisations providing car valuation services can incorporate the car resale value prediction solution to enhance the accuracy of their valuation models. It helps them provide more reliable and comprehensive reports to their customers.

9. CONCLUSION:

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. The proposed system will help to determine the accurate price of used car price prediction. This paper compares 3 different algorithms for machine learning : Linear Regression, Lasso Regression and Ridge Regression.

10.FUTURE SCOPE:

In future this machine learning model may bind with various website which can provide real time data for price prediction. Also we may add large historical data of car price which can help to improve accuracy of the machine learning model. We can build an android app as user interface for interacting with user. For better performance, we plan to judiciously design deep learning network structures, use adaptive learning rates and train on clusters of data rather than the whole dataset.

11. BIBLIOGRAPHY:

[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)

APPENDIX:

Code screenshots :

Java Math.random() Method wit | Resale Value Prediction | Downloads/car_resale/ | car_resale_value_notebook

localhost:8888/notebooks/Downloads/car_resale/car_resale_value_notebook.ipynb

jupyter car_resale_value_notebook Last Checkpoint: Yesterday at 10:01 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help — Trusted — Python 3

In [1]:
```python
import pandas as pd
import numpy as np
import matplotlib as plt
from sklearn.preprocessing import LabelEncoder
import pickle
```

In [2]:
```python
df = pd.read_csv(r"C:\Users\mutyam bhargav reddy\Downloads\car_resale\autos.csv", header=0, sep=',', encoding='Latin1',)
```

In [3]:
```python
df
```

Out[3]:

	dateCrawled	name	seller	offerType	price	abtest	vehicleType	yearOfRegistration	gearbox	pov
0	2016-03-24 11:52:17	Golf_3_1.6	privat	Angebot	480	test	NaN	1993	manuell	
1	2016-03-24 10:58:45	A5_Sportback_2.7_Tdi	privat	Angebot	18300	test	coupe	2011	manuell	
2	2016-03-14 12:52:21	Jeep_Grand_Cherokee_"Overland"	privat	Angebot	9800	test	suv	2004	automatik	
3	2016-03-17 16:54:04	GOLF_4_1_4__3TÜRER	privat	Angebot	1500	test	kleinwagen	2001	manuell	
4	2016-03-31 17:25:20	Skoda_Fabia_1.4_TDI_PD_Classic	privat	Angebot	3600	test	kleinwagen	2008	manuell	

In [4]:
```python
df.head
```

Out[4]: <bound method NDFrame.head of dateCrawled \

Type here to search — 27°C Mostly cloudy — ENG 23:58 29-06-2023

Java Math.random() Method wit | Resale Value Prediction | Downloads/car_resale/ | car_resale_value_notebook

localhost:8888/notebooks/Downloads/car_resale/car_resale_value_notebook.ipynb

jupyter car_resale_value_notebook Last Checkpoint: Yesterday at 10:01 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help — Trusted — Python 3

In [4]: df.head

Out[4]:
```
<bound method NDFrame.head of                  dateCrawled  \
0       2016-03-24 11:52:17
1       2016-03-24 10:58:45
2       2016-03-14 12:52:21
3       2016-03-17 16:54:04
4       2016-03-31 17:25:20
5       2016-04-04 17:36:23
6       2016-04-01 20:48:51
7       2016-03-21 18:54:38
8       2016-04-04 23:42:13
9       2016-03-17 10:53:50
10      2016-03-26 19:54:18
11      2016-04-07 10:06:22
12      2016-03-15 22:49:09
13      2016-03-21 21:37:40
14      2016-03-21 12:57:01
15      2016-03-11 21:39:15
16      2016-04-01 12:46:46
17      2016-03-20 10:25:19
```

In [5]:
```python
#now printing the coulmn name and the shape of the dataset
print(df.columns ,df.shape)
```

```
Index(['dateCrawled', 'name', 'seller', 'offerType', 'price', 'abtest',
       'vehicleType', 'yearOfRegistration', 'gearbox', 'powerPS', 'model',
       'kilometer', 'monthOfRegistration', 'fuelType', 'brand',
       'notRepairedDamage', 'dateCreated', 'nrOfPictures', 'postalCode',
       'lastSeen'],
      dtype='object') (371528, 20)
```

In [6]:
```python
#printing the information of the salers
print(df.seller.value_counts())
```

Type here to search — 27°C Mostly cloudy — ENG 23:58 29-06-2023

jupyter car_resale_value_notebook Last Checkpoint: Yesterday at 10:01 PM (autosaved)

In [6]:
```python
#printing the information of the salers
print(df.seller.value_counts())
```

```
privat        371525
gewerblich         3
Name: seller, dtype: int64
```

In [7]:
```python
#removing the details of the saler who has only three cars
df[df.seller != 'gewerblich']
```

Out[7]:

	dateCrawled	name	seller	offerType	price	abtest	vehicleType	yearOfRegistration	gearbox	pov
0	2016-03-24 11:52:17	Golf_3_1.6	privat	Angebot	480	test	NaN	1993	manuell	
1	2016-03-24 10:58:45	A5_Sportback_2.7_Tdi	privat	Angebot	18300	test	coupe	2011	manuell	
2	2016-03-14 12:52:21	Jeep_Grand_Cherokee_"Overland"	privat	Angebot	9800	test	suv	2004	automatik	
3	2016-03-17 16:54:04	GOLF_4_1_4__3TÜRER	privat	Angebot	1500	test	kleinwagen	2001	manuell	
4	2016-03-31 17:25:20	Skoda_Fabia_1.4_TDI_PD_Classic	privat	Angebot	3600	test	kleinwagen	2008	manuell	

In [8]:
```python
# we have all the details of saller are same so we have to remove the column
df=df.drop('seller',1)
```

In [9]:
```python
#print all the different sellers
print(df.offerType.value_counts())
```


jupyter car_resale_value_notebook Last Checkpoint: Yesterday at 10:01 PM (autosaved)

In [9]:
```python
#print all the different sellers
print(df.offerType.value_counts())
```


```
Angebot    371516
Gesuch         12
Name: offerType, dtype: int64
```

In [10]:
```python
#remove the Offer Type having only 12 listings
df[df.offerType != 'Gesuch']
```

Out[10]:

	dateCrawled	name	offerType	price	abtest	vehicleType	yearOfRegistration	gearbox	powerPS
0	2016-03-24 11:52:17	Golf_3_1.6	Angebot	480	test	NaN	1993	manuell	0
1	2016-03-24 10:58:45	A5_Sportback_2.7_Tdi	Angebot	18300	test	coupe	2011	manuell	190
2	2016-03-14 12:52:21	Jeep_Grand_Cherokee_"Overland"	Angebot	9800	test	suv	2004	automatik	163
3	2016-03-17 16:54:04	GOLF_4_1_4__3TÜRER	Angebot	1500	test	kleinwagen	2001	manuell	75
4	2016-03-31 17:25:20	Skoda_Fabia_1.4_TDI_PD_Classic	Angebot	3600	test	kleinwagen	2008	manuell	69

In [11]:
```python
#now all the offers are same so we can get rid of this column
df=df.drop('offerType',1)
```

In [12]:
```python
# Cars having power less than 50ps and above 900ps are little bit suspious so we have to get ride of thes column
print(df.shape)
```

car_resale_value_notebook Last Checkpoint: Yesterday at 10:01 PM (autosaved)

```
In [12]: # Cars having power less than 50ps and above 900ps are little bit suspious so we have to get ride of thes column
         print(df.shape)

(371528, 18)

In [13]: df = df[(df.powerPS > 50) & (df.powerPS < 900)]
         print(df.shape)

(319709, 18)

In [14]: #simlarly, filtering our the cars having registeration years not in the mentioned range
         print(df.shape)

(319709, 18)

In [15]: df = df[(df.yearOfRegistration >= 1950) & (df.yearOfRegistration < 2017)]
         print(df.shape)

(309171, 18)

In [16]: #the details which are irrelivent should be removed
         df.drop(['name','abtest', 'dateCrawled', 'nrOfPictures', 'lastSeen',
                  'postalCode', 'dateCreated'], axis='columns', inplace=True)

In [17]: print(df.shape)

(309171, 11)

In [18]: #final data that we have now
         print(df.shape)

(309171, 11)

In [19]: #dropping the duplicates from the dataframe and stroing it in a new df.
         #here all rows having same value in all the mentioned columns will be deleted and by default,
```

car_resale_value_notebook Last Checkpoint: Yesterday at 10:01 PM (autosaved)

```
In [19]: #dropping the duplicates from the dataframe and stroing it in a new df.
         #here all rows having same value in all the mentioned columns will be deleted and by default,
         #only first occurance of anysuch row is kept
         new_df = df.copy()
         new_df = new_df.drop_duplicates(['price','vehicleType','yearOfRegistration'
                                     ,'gearbox','powerPS','model','kilometer','monthOfRegistration','fuelType'
                                     ,'notRepairedDamage'])

In [20]: #after removing duplicates
         print(new_df.shape)

(285145, 11)

In [21]: #As the dataset contained some german words for many features, cahnging them to english
         new_df.gearbox.replace(('manuell', 'automatik'), ('manual','automatic'), inplace=True)
         new_df.fuelType.replace(('benzin','andere','elektro'),('petrol','others','electric'),inplace=True)
         new_df.vehicleType.replace(('kleinwagen', 'cabrio','kombi','andere'),
                                    ('small car','convertible','combination','others'),inplace=True)
         new_df.notRepairedDamage.replace(('ja','nein'),('Yes','No'),inplace=True)

In [22]: #### Removing the outliers
         new_df = new_df[(new_df.price >= 100) & (new_df.price <= 150000)]

In [23]: #Filling NaN values for columns whose data might not be there with the information provider,
         #which might lead to some variance but our model
         #but we will still be able to give some estimate to the user
         new_df['notRepairedDamage'].fillna(value='not-declared', inplace=True)
         new_df['fuelType'].fillna(value='not-declared', inplace=True)
         new_df['gearbox'].fillna(value='not-declared', inplace=True)
         new_df['vehicleType'].fillna(value='not-declared', inplace=True)
         new_df['model'].fillna(value='not-declared', inplace=True)

In [24]: #can save the csv for future purpose.
         new_df.to_csv("autos_preprocessed.csv")
```

car_resale_value_notebook Last Checkpoint: Yesterday at 10:01 PM (autosaved)

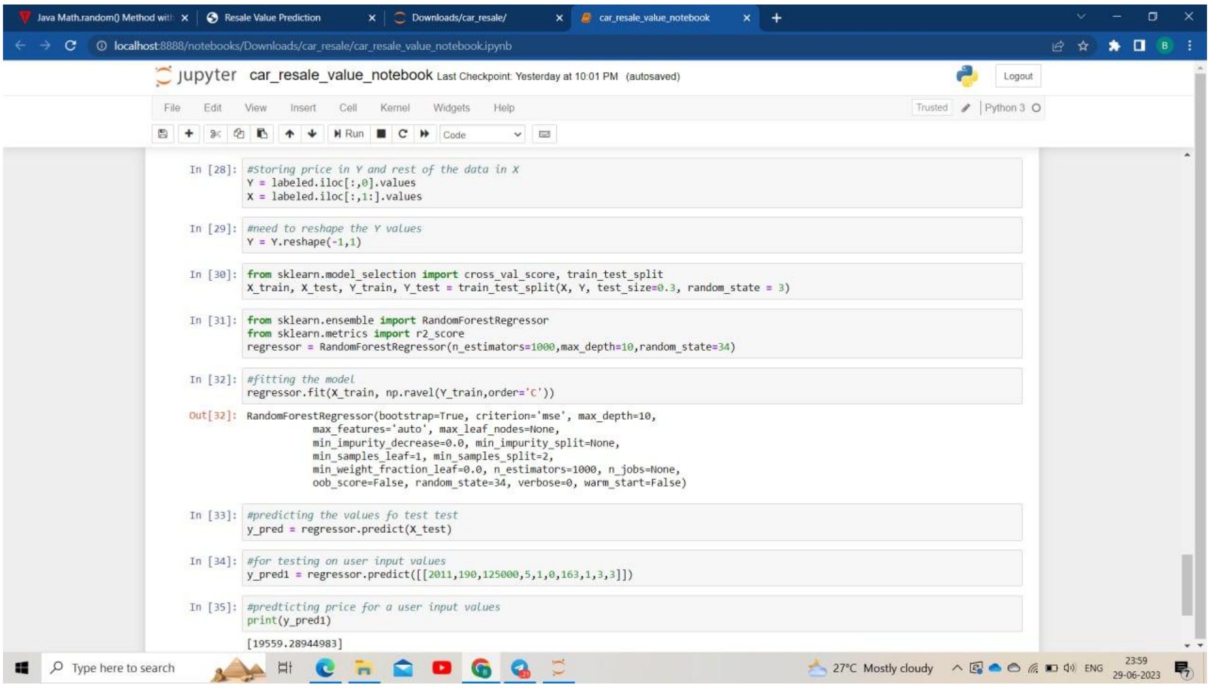
In [25]:
```python
#Columns which contain categorical values, which we'll need to convert via label encoding
labels = ['gearbox', 'notRepairedDamage', 'model', 'brand', 'fuelType', 'vehicleType']
```

In [26]:
```python
#looping over the labels to do the label encoding for all at once and
#saving the LABEL ENCODING FILES
mapper = {}
for i in labels:
    mapper[i] = LabelEncoder()
    mapper[i].fit(new_df[i])
    tr = mapper[i].transform(new_df[i])
    np.save(str('classes'+i+'.npy'), mapper[i].classes_)
    print(i,":",mapper[i])
    new_df.loc[:, i + '_labels'] = pd.Series(tr, index=new_df.index)
```

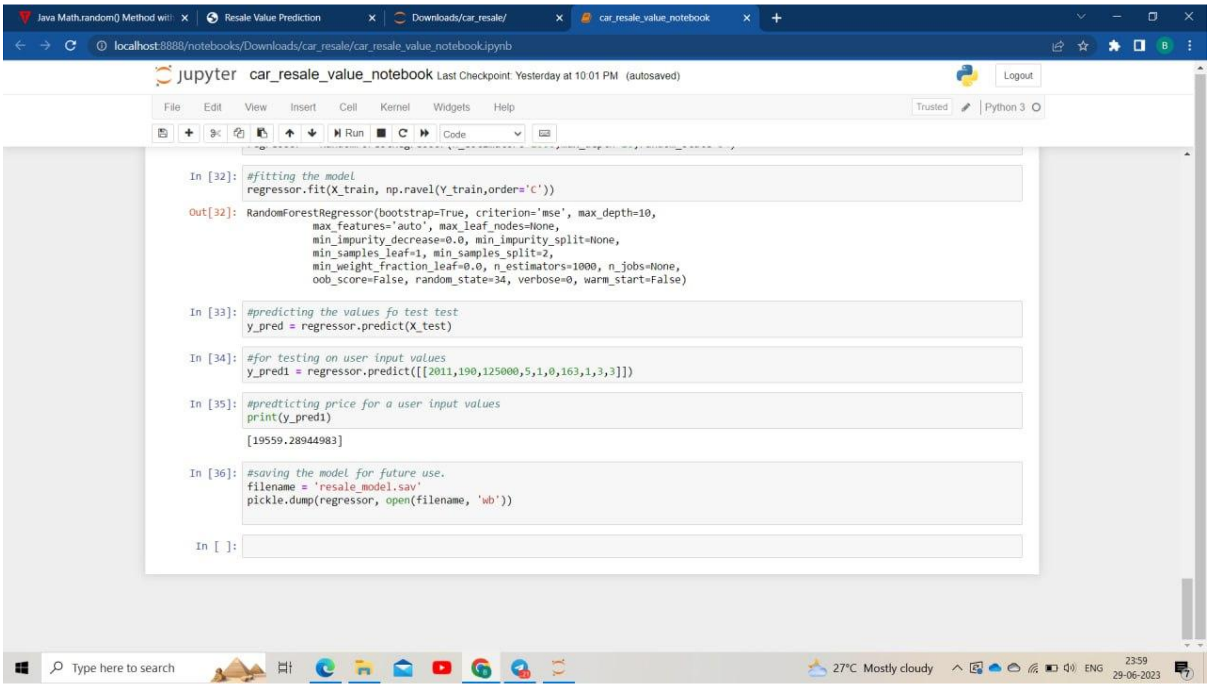
```
gearbox : LabelEncoder()
notRepairedDamage : LabelEncoder()
model : LabelEncoder()
brand : LabelEncoder()
fuelType : LabelEncoder()
vehicleType : LabelEncoder()
```

In [27]:
```python
#Final data to be put in a new dataframe called "LABELED",
labeled = new_df[ ['price'
                        ,'yearOfRegistration'
                        ,'powerPS'
                        ,'kilometer'
                        ,'monthOfRegistration'
                        ]
                    + [x+"_labels" for x in labels]]

print(labeled.columns)
```

```
Index(['price', 'yearOfRegistration', 'powerPS', 'kilometer',
       'monthOfRegistration', 'gearbox_labels', 'notRepairedDamage_labels',
       'model_labels', 'brand_labels', 'fuelType_labels',
       'vehicleType_labels'],
      dtype='object')
```



```
In [28]: #Storing price in Y and rest of the data in X
         Y = labeled.iloc[:,0].values
         X = labeled.iloc[:,1:].values

In [29]: #need to reshape the Y values
         Y = Y.reshape(-1,1)

In [30]: from sklearn.model_selection import cross_val_score, train_test_split
         X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.3, random_state = 3)

In [31]: from sklearn.ensemble import RandomForestRegressor
         from sklearn.metrics import r2_score
         regressor = RandomForestRegressor(n_estimators=1000,max_depth=10,random_state=34)

In [32]: #fitting the model
         regressor.fit(X_train, np.ravel(Y_train,order='C'))

Out[32]: RandomForestRegressor(bootstrap=True, criterion='mse', max_depth=10,
                    max_features='auto', max_leaf_nodes=None,
                    min_impurity_decrease=0.0, min_impurity_split=None,
                    min_samples_leaf=1, min_samples_split=2,
                    min_weight_fraction_leaf=0.0, n_estimators=1000, n_jobs=None,
                    oob_score=False, random_state=34, verbose=0, warm_start=False)

In [33]: #predicting the values fo test test
         y_pred = regressor.predict(X_test)

In [34]: #for testing on user input values
         y_pred1 = regressor.predict([[2011,190,125000,5,1,0,163,1,3,3]])

In [35]: #predtcting price for a user input values
         print(y_pred1)

         [19559.28944983]
```




```
In [32]: #fitting the model
         regressor.fit(X_train, np.ravel(Y_train,order='C'))

Out[32]: RandomForestRegressor(bootstrap=True, criterion='mse', max_depth=10,
                    max_features='auto', max_leaf_nodes=None,
                    min_impurity_decrease=0.0, min_impurity_split=None,
                    min_samples_leaf=1, min_samples_split=2,
                    min_weight_fraction_leaf=0.0, n_estimators=1000, n_jobs=None,
                    oob_score=False, random_state=34, verbose=0, warm_start=False)

In [33]: #predicting the values fo test test
         y_pred = regressor.predict(X_test)

In [34]: #for testing on user input values
         y_pred1 = regressor.predict([[2011,190,125000,5,1,0,163,1,3,3]])

In [35]: #predtcting price for a user input values
         print(y_pred1)

         [19559.28944983]
```


```
In [36]: #saving the model for future use.
         filename = 'resale_model.sav'
         pickle.dump(regressor, open(filename, 'wb'))

In [ ]:
```