APPLIED DATA SCIENCE

CAR RESALE VALUE

TEAM MEMBERS :
1. NAGANATH S (20BEC1232)
2. KHAVIN C (20BEC1234)
3. VIMAL RAJ G (20BEC1298)
4. SANDEEP S (20BEC1086)

1. INTRODUCTION

1.1. OVERVIEW

In this project, we propose the development of an intelligent, adaptable, and efficient system that utilizes regression algorithms to forecast the resale value of vehicles. By considering the key factors that influence a vehicle's resale value, a regression model will be constructed to provide the closest estimate of the vehicle's future worth. We will employ the Random Forest Regressor regression algorithm and find out the accuracy of the solution. This model will be integrated into a web-based application, allowing users to receive updates on the status of their product. A HTML page will be created by us where the users can give the details about the cars that they are planning to buy. Upon submitting the details, the resale value of the vehicle will be displayed to them.

1.2. PURPOSE

In light of challenging economic conditions, it is expected that the demand for second-hand imported (reconditioned) cars and used cars will rise. Many developed countries commonly opt for leasing cars instead of purchasing them outright. There are several independent dealers who sell second hand cars. Each of them value the vehicles differently. There will be a lack of uniform appraisal. Our project will also give a clear idea to the customers before they commit to the purchase. They can obtain and compare the resale values of several vehicles at the same time which makes price evaluation a time efficient process. Moreover, we can also compare the price ranges of different dealers and bring up a common evaluation system. Therefore, our project is aimed at improving the market of used cars and the issues associated with it.

2. LITERATURE SURVEY

2.1. EXISTING PROBLEM

S.NO	JOURNAL DETAILS	INFERENCE
1	Car resale price forecasting: The impact of regression method, private information, and heterogeneity on forecast accuracy - Stefan Lessmann, Stefan Voß (2017).	This paper examines statistical models used for predicting the resale prices of used cars. Through an empirical study, the paper explores how different modeling approaches impact the accuracy of the forecasts. The analysis compares various prediction methods and finds that random forest regression is particularly effective for resale price forecasting, outperforming the previously popular method of linear regression. Furthermore, the study reveals the presence of heterogeneity in resale price forecasting and identifies methods that can automatically overcome its negative impact on forecast accuracy. Lastly, the research confirms that sellers of used cars have informational advantages over market research agencies, allowing them to make more accurate resale price forecasts.
2	Comparison of Various Regression Techniques and Predicting the Resale Price of Cars - Iqbal Singh Saini, Navneet Kaur (2023).	This paper is to compare five regression techniques of machine learning using a dataset of car resale prices. The dataset is sourced from Kaggle. The paper primarily discusses the process of cleaning and transforming the unstructured dataset into useful data. The study involves a comparative analysis of different algorithms including linear regression, decision tree, random forest, k nearest neighbor, and extreme gradient boost. It was discovered that the random forest regression algorithm exhibited the best accuracy when tested with the test dataset. Consequently, the random forest algorithm was utilized to develop a model that could predict the resale value of a car. The model is saved using the pickle library of Python.
3	Predicting the Prices of Used Cars using Machine Learning for Resale - B Hemendiran, P N Renjith (2023).	The paper discusses the utilization of machine learning models for accurately predicting the price of used cars based on their attributes and features. According to the study, supervised machine learning models were applied to forecast the prices of second-hand cars in India. The forecasts were

		generated using historical data gathered from daily news articles, magazines, and various standard websites. Several models, including Random Forest Regressor, Extra Tree Regressor, Bagging Regressor, Decision Tree, and XG Boost, were employed to make predictions. After comparing the predictions of these models, the most accurate ones were selected. The paper states that there were slight variations in their performance, and the obtained results were comparable. The researchers made significant efforts to measure various distinct properties and enhance the reliability and accuracy of the predictions. The models were evaluated based on their prediction accuracy, leading to the conclusion that the Random Forest model produced the best outcomes.

2.2. PROPOSED SOLUTION

In our project, we implemented the Random Forest Regressor algorithm to estimate the prices of used cars. The Random Forest Regressor is a machine learning algorithm that leverages an ensemble of decision trees to make predictions. By training the model on a dataset containing various features of used cars and their corresponding prices, we were able to create a model capable of predicting the resale price of a vehicle based on its characteristics.

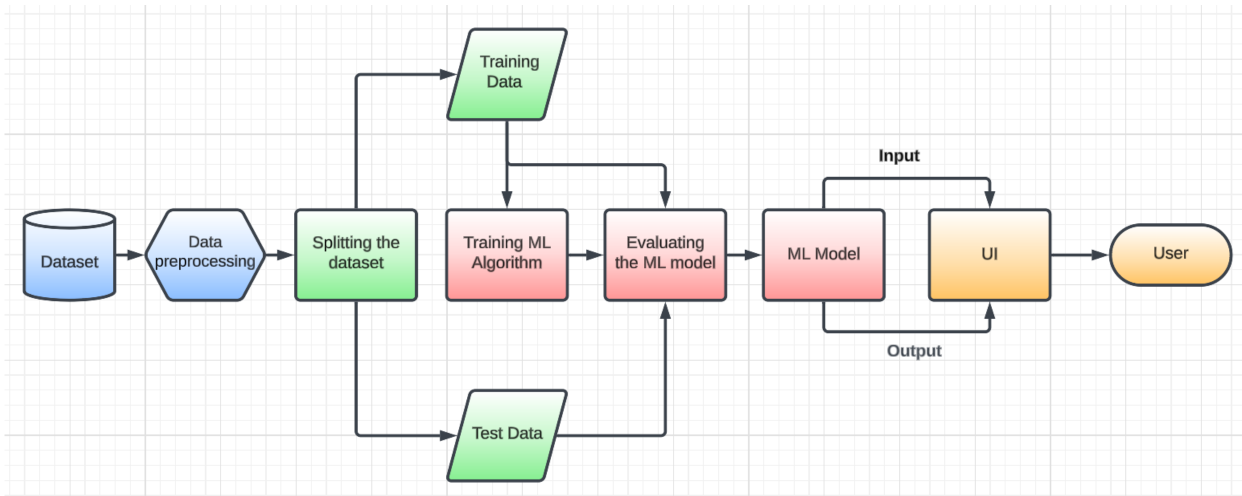
To ensure the reliability and accuracy of our model, we followed standard procedures for data preprocessing. This involved handling missing values, dealing with categorical variables through encoding techniques such as one-hot encoding, and scaling numerical features as needed. The preprocessed dataset was then split into training and testing sets. The training set was used to train the Random Forest Regressor model, while the testing set was used to evaluate its performance and calculate accuracy.

To provide a user-friendly interface for individuals seeking to determine the resale price of a car, we developed a web application using Python Flask and HTML. The web page allowed users to input details about the vehicle they were interested in, forming the test dataset. Using the trained model, we then predicted the output, which represented the estimated resale price of the car based on the provided information.

This integration of machine learning and web development enables users to conveniently access our model and obtain accurate price predictions for used cars.

3. THEORETICAL ANALYSIS

3.1 BLOCK DIAGRAM



3.2 SOFTWARE DESIGNING

This project mainly focuses on predicting the value of a car which undergoes resale using a Machine Learning model and all of its requirements come under the software side.

Major software requirements of the project:

- Dataset
- Libraries to run the code of Machine Learning model.
- Python notebook

- **Dataset:**
  We need a proper dataset which contains extensive data regarding the car which has to be resold including the year of first buy, kilometers ran, model and its condition. The above necessity has to be satisfied fully so that we can observe the trends in the cost of resale and build the predictive model based on the necessary parameters. The data should be uniformly distributed and understandable for efficient building of models.

- **Libraries:**
  To build a Machine Learning model, we need to have the required libraries included for functions ranging from uploading the dataset to determining model parameters. If certain libraries are not available, they need to be installed so that the functions that require them can be implemented. Libraries allow us to use pre written functions and avoid code repetition.

Libraries used:

- Numpy: For mathematical operations on arrays.
- Pandas: For machine learning models involving dataframes.
- Matplotlib: For visualizations of data like line plot, barplot etc.
- Sklearn: For accessing various machine learning algorithms such as regression, classification, SVM etc.
- Pickle: For pickling each tuple separately and to call the callable on provided arguments.
- Flask: It is a micro web framework written in python.

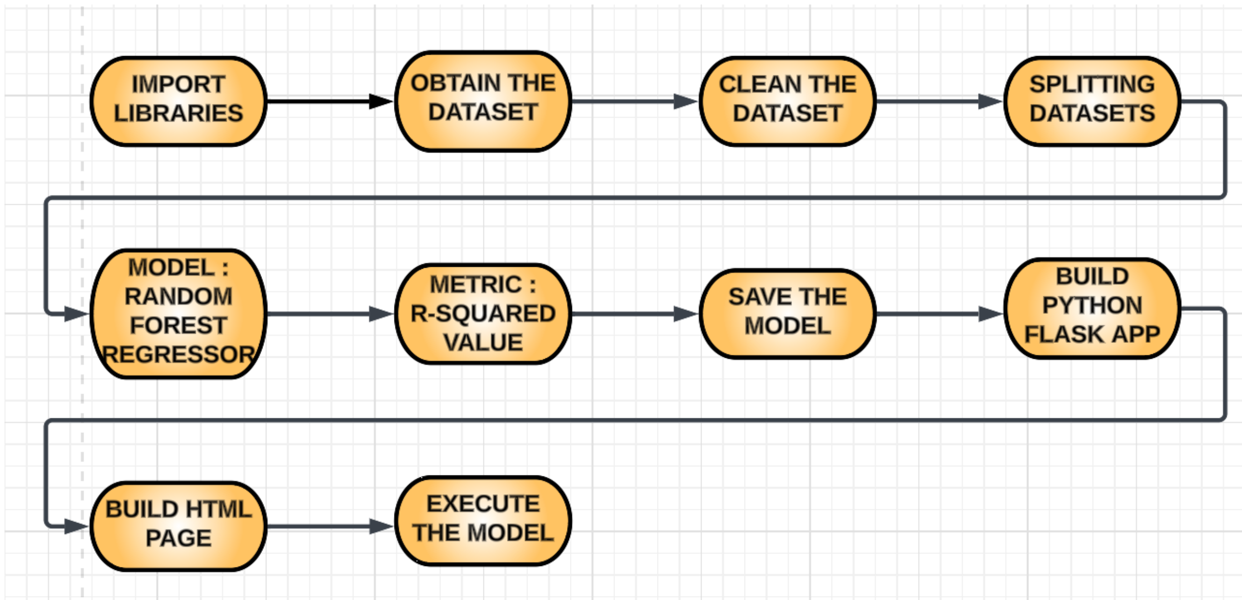
- **Python notebook:**

  Machine learning models are usually implemented in python notebooks which can run blocks of code where the codes can access the dataset and perform preprocessing, splitting the dataset and then build the Machine Learning model and even a web application. It has a flexible interface that allows users to configure and arrange workflows in data science and machine learning. Its modular design allows extensions to expand and enhance functionality.

4. EXPERIMENTAL INVESTIGATIONS

We analyzed the considerations that are necessary when designing a machine learning model that predicts the car resale value. We determined the parameters that are very crucial such as vehicle type, year of registration, gearbox and many more. Also, we found the parameters that are irrelevant to the end result such as name, abtest, dateCrawled, postalCode etc. In some of the parameters, we found out that we need to remove the outliers present in it since it could affect the resale price prediction of the car. In crucial columns, we figured out that there are duplicates which are required to be removed. We also found out some columns which had missing values such as notReparitedDamage, fuel type, gearbox and likewise. With the help of these investigations, the model can be built smoothly with the help of libraries and functions.

5. FLOWCHART



6. RESULT

Displaying Head of the dataset:

	price	vehicleType	yearOfRegistration	gearbox	powerPS	model	kilometer	fuelType	brand	notRepairedDamage
1	18300	coupe	2011	manuell	190	NaN	125000	diesel	audi	ja
2	9800	suv	2004	automatik	163	grand	125000	diesel	jeep	NaN
3	1500	kleinwagen	2001	manuell	75	golf	150000	benzin	volkswagen	nein
4	3600	kleinwagen	2008	manuell	69	fabia	90000	diesel	skoda	nein
5	650	limousine	1995	manuell	102	3er	150000	benzin	bmw	ja

Statistical Analysis:

	price	yearOfRegistration	powerPS	kilometer	monthOfRegistration	nrOfPictures	postalCode
count	2.279100e+04	22791.000000	22791.000000	22791.000000	22791.000000	22790.0	22790.000000
mean	7.361659e+03	2004.657233	113.487341	125601.553245	5.754421	0.0	50737.071566
std	1.291478e+05	92.133562	115.606172	40096.008466	3.714944	0.0	25910.935074
min	0.000000e+00	1000.000000	0.000000	5000.000000	0.000000	0.0	1067.000000
25%	1.150000e+03	1999.000000	69.000000	125000.000000	3.000000	0.0	30167.000000
50%	2.990000e+03	2003.000000	105.000000	150000.000000	6.000000	0.0	49530.500000
75%	7.250000e+03	2008.000000	150.000000	150000.000000	9.000000	0.0	71665.000000
max	1.234568e+07	9999.000000	10317.000000	150000.000000	12.000000	0.0	99994.000000

Summary of the Data Frame:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 22791 entries, 0 to 22790
Data columns (total 20 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   dateCrawled          22791 non-null  object
 1   name                 22791 non-null  object
 2   seller               22791 non-null  object
 3   offerType            22791 non-null  object
 4   price                22791 non-null  int64
 5   abtest               22791 non-null  object
 6   vehicleType          20411 non-null  object
 7   yearOfRegistration   22791 non-null  int64
 8   gearbox              21572 non-null  object
 9   powerPS              22791 non-null  int64
 10  model                21574 non-null  object
 11  kilometer            22791 non-null  int64
 12  monthOfRegistration  22791 non-null  int64
 13  fuelType             20705 non-null  object
 14  brand                22791 non-null  object
 15  notRepairedDamage    18357 non-null  object
 16  dateCreated          22790 non-null  object
 17  nrOfPictures         22790 non-null  float64
 18  postalCode           22790 non-null  float64
 19  lastSeen             22790 non-null  object
dtypes: float64(2), int64(5), object(13)
memory usage: 3.5+ MB
```

Data after removing the Irrelevant variables:

	dateCrawled	name	price	abtest	vehicleType	yearOfRegistration	gearbox	powerPS	model	kilometer	monthOfRegistration	fuelType
0	2016-03-24 11:52:17	Golf_3_1.6	480	test	NaN	1993	manuell	0	golf	150000	0	benzin
1	2016-03-24 10:58:45	A5_Sportback_2.7_Tdi	18300	test	coupe	2011	manuell	190	NaN	125000	5	diesel
2	2016-03-14 12:52:21	Jeep_Grand_Cherokee_"Overland"	9800	test	suv	2004	automatik	163	grand	125000	8	diesel
3	2016-03-17 16:54:04	GOLF_4_1_4__3TÜRER	1500	test	kleinwagen	2001	manuell	75	golf	150000	6	benzin
4	2016-03-31 17:25:20	Skoda_Fabia_1.4_TDI_PD_Classic	3600	test	kleinwagen	2008	manuell	69	fabia	90000	7	diesel

```
(22791, 18)
(19502, 18)
(19502, 18)
(18855, 18)
```

Removing Irrelevant Columns:

	price	vehicleType	yearOfRegistration	gearbox	powerPS	model	kilometer	fuelType	brand	notRepairedDamage
1	18300	coupe	2011	manuell	190	NaN	125000	diesel	audi	ja
2	9800	suv	2004	automatik	163	grand	125000	diesel	jeep	NaN
3	1500	kleinwagen	2001	manuell	75	golf	150000	benzin	volkswagen	nein
4	3600	kleinwagen	2008	manuell	69	fabia	90000	diesel	skoda	nein
5	650	limousine	1995	manuell	102	3er	150000	benzin	bmw	ja

Independent Variables:

	0	1	2	3	4	5	6	7	8
0	2011	190	125000	1	0	155	1	3	3
1	2004	163	125000	0	3	116	14	3	8
2	2001	75	150000	1	2	115	37	1	4
3	2008	69	90000	1	2	100	31	3	4
4	1995	102	150000	1	0	11	2	1	6

Fitting the Model:

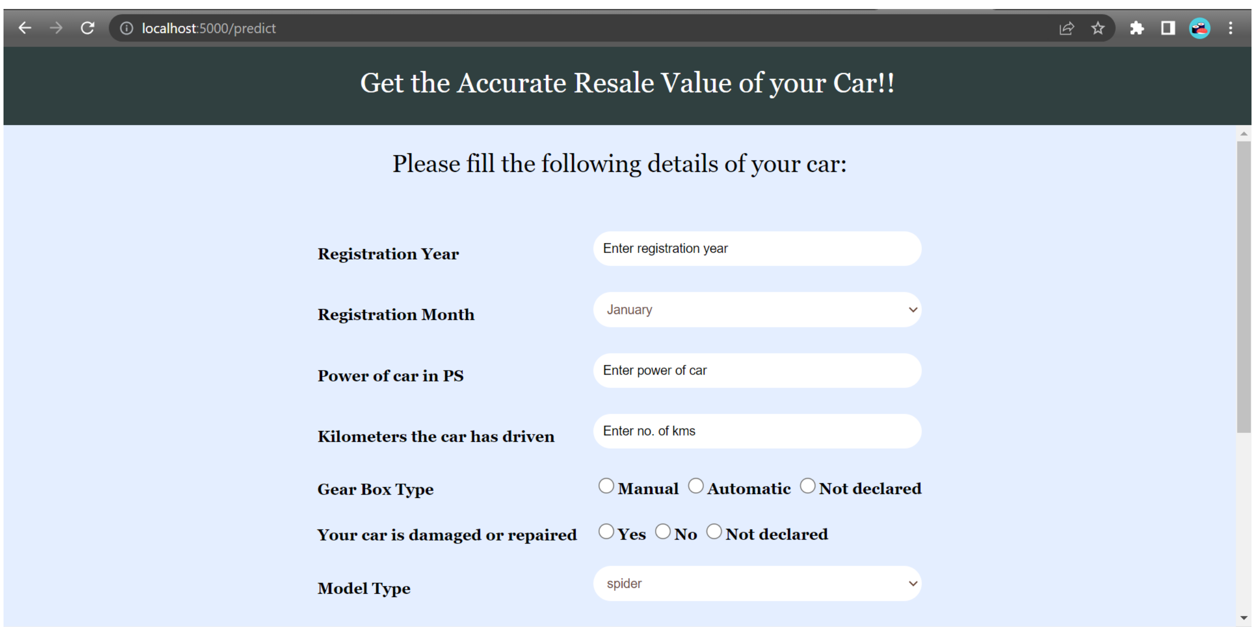
```
RandomForestRegressor
RandomForestRegressor(max_depth=10, n_estimators=1000, random_state=34)
```

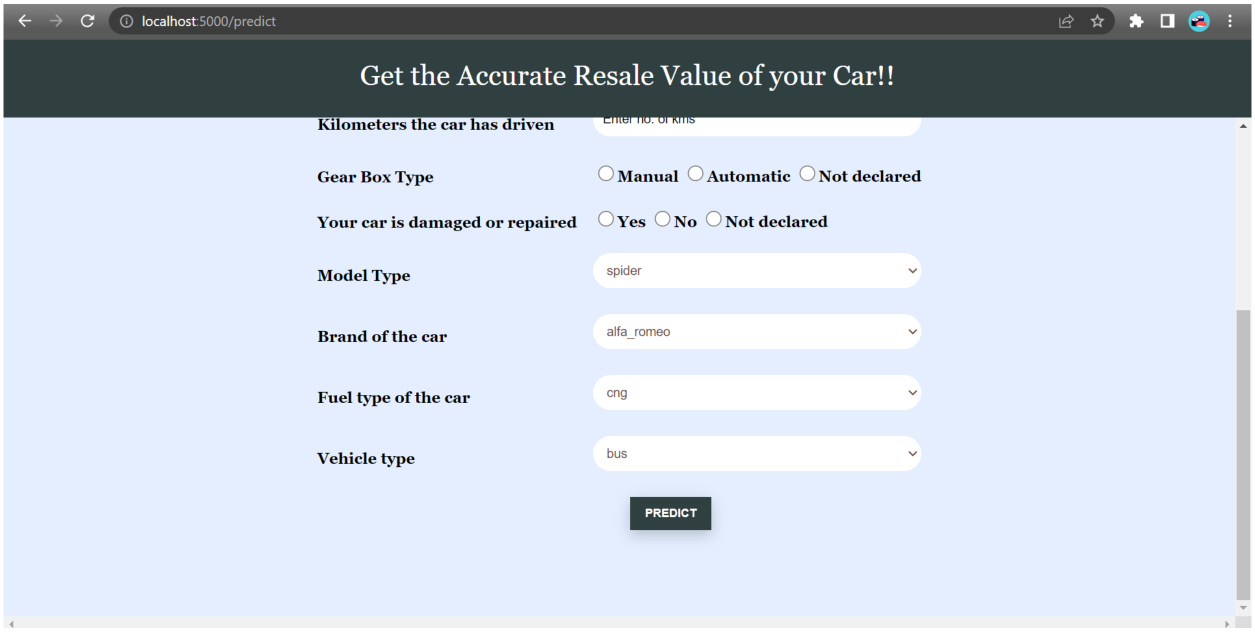
R_2 Score:

```
0.8459768282230389
```

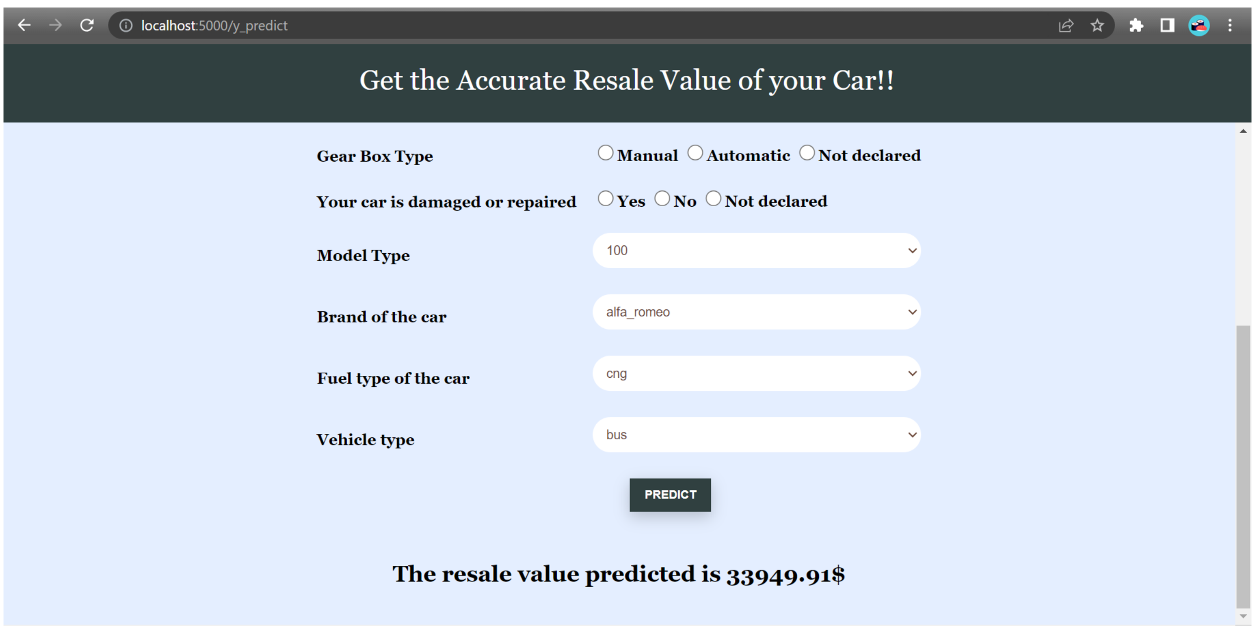
Resaleintro HTML page:

Resalepredict HTML page :




localhost:5000/predict
Get the Accurate Resale Value of your Car!!
Kilometers the car has driven
Enter no. of kms
Gear Box Type
Manual Automatic Not declared
Your car is damaged or repaired
Yes No Not declared
Model Type
spider
Brand of the car
alfa_romeo
Fuel type of the car
cng
Vehicle type
bus
PREDICT

Output page :


localhost:5000/y_predict
Get the Accurate Resale Value of your Car!!
Gear Box Type
Manual Automatic Not declared
Your car is damaged or repaired
Yes No Not declared
Model Type
100
Brand of the car
alfa_romeo
Fuel type of the car
cng
Vehicle type
bus
PREDICT
The resale value predicted is 33949.91$

7. ADVANTAGES AND DISADVANTAGES

Advantages of Car Resale Value Prediction :

1) Informed Decision-making: Predictions of resale value help potential automobile purchasers make wise selections when choosing a vehicle. Buyers can evaluate the long-term cost of ownership and select a car that holds its worth well by understanding the expected future value. This aids them in avoiding any monetary losses brought on by quick depreciation.

2) Financial Planning: Car resale value predictions are beneficial for individuals who plan to finance their vehicle purchase. Knowing the expected future value can help them estimate the equity they will have in the car at the end of their loan term. This information is crucial when considering factors such as refinancing, trading in for a new vehicle, or selling the car to settle the remaining loan balance.

3) Negotiation Power: When selling their vehicles, sellers might use estimates of resale value to their advantage. Sellers can set a competitive asking price and defend it to prospective purchasers provided they have correct information about the predicted depreciation of the vehicle. This can raise the probability of a profitable sale at a reasonable price.

4) Brand and Model Comparison: Buyers can evaluate various automobile brands and models based on their ability to retain value over time thanks to resale value estimates. With the help of this knowledge, buyers may choose a car that fits their needs and budget while taking into account its prospective resale value.

Disadvantages of Car Resale Value Prediction :

1) Individual Vehicle Factors: Resale value forecasts provide broad estimations, but they might not take into account certain elements that influence the value of a particular car. The resale value can be considerably impacted by elements like mileage, maintenance history, condition, extra options, and regional market variances. Estimates may be less accurate if predictions do not take these distinctive characteristics into account.

2) Technological Advancements: The automotive industry is evolving rapidly, with new technologies and features being introduced regularly. Resale value predictions may struggle to account for the impact of emerging technologies on the value of older vehicles. Electric vehicles, autonomous driving capabilities, and advanced safety features can influence the desirability and depreciation of cars over time.

3) Limited Data Accuracy: Resale value forecasts provide broad estimations, but they might not take into account certain elements that influence the value of a particular car. The resale value can be considerably impacted by elements like mileage, maintenance history, condition, extra options, and regional market variances. Estimates may be less accurate if predictions do not take these distinctive characteristics into account.

It's important to remember that estimates of a car's resale value are just intended to provide a general reference. To make educated judgments, buyers and sellers should weigh these aspects alongside others including their own wants and preferences as well as the state of the local market.

8. APPLICATIONS

Predicting the resale value of used cars is a valuable application of data science in the automotive industry. Here are some practical applications for "Used Car Resale Value Prediction" using applied data science:

1) **Inventory Management:** By predicting the resale value of different car models and factors affecting their depreciation rates, dealerships can optimize their inventory management. They can make informed decisions about which vehicles to acquire, how long to hold onto them, and when to sell them to minimize losses.

2) **Consumer Guidance:** Car buyers often consider the potential resale value when making a purchase. A predictive model can provide consumers with insights into how a particular make, model, or feature affects the depreciation rate. This information assists buyers in making more informed decisions about their purchase and potentially minimizing the cost of ownership.

3) **Financial Planning:** Used car resale value prediction can be valuable for financial institutions, leasing companies, and insurance providers. It helps them assess the value of vehicles as collateral, calculate leasing rates, and determine insurance premiums based on the estimated future value of the car.

4) **Fleet Management:** Companies that manage vehicle fleets, such as car rental agencies or transportation services, can benefit from accurate resale value predictions. By understanding the depreciation rates and expected resale values of their vehicles, they can plan their fleet replacement cycles, optimize maintenance schedules, and make informed decisions about selling or retiring vehicles.

5) **Market Analysis:** Manufacturers and market researchers can leverage resale value prediction models to analyze market trends, customer preferences, and the impact of different factors on

the value of used cars. This information can inform product development, marketing strategies, and investment decisions.

6) **Risk Assessment:** Resale value prediction models can aid financial institutions in assessing the risk associated with lending or financing used cars. By estimating the future value of a vehicle, lenders can evaluate the potential loan-to-value ratio and determine appropriate interest rates, ensuring responsible lending practices.

These applications demonstrate how "Used Car Resale Value Prediction" using applied data science can benefit various stakeholders in the automotive industry, including dealerships, buyers, financial institutions, and market researchers.

9. CONCLUSION

In this Project we have used Machine Learning Algorithms to Predict a car's resale value by considering some crucial parameters. it has a number of benefits and can be a useful tool for both buyers and sellers in the automotive market. It enables purchasers to negotiate favorable terms, arrange their finances more effectively, and make informed decisions. It helps merchants maximize returns and set pricing that are competitive. These forecasts are, however, subject to a number of restrictions, such as data inaccuracy, market volatility, specific vehicle characteristics, and technical developments.The accuracy of resale value projections may fluctuate as the automobile industry develops, with new technologies and market forces impacting the value of vehicles. Consequently, it is necessary to utilize these forecasts as a reference while taking into account other crucial elements including the condition of the vehicle, its maintenance history, its optional features, and local market trends.

Ultimately, car resale value prediction serves as a useful tool to enhance decision-making in the automotive market, but it should be complemented with critical thinking and a comprehensive assessment of the specific vehicle and prevailing market conditions.

10. FUTURE SCOPE

The future scope for "Used Car Resale Value Prediction" using applied data science is promising and offers several potential advancements and opportunities. Here are some aspects that could shape its future:

1) **Incorporation of Unstructured Data**: While existing models utilize structured data like vehicle attributes and historical transaction data, incorporating unstructured data sources such as online reviews, social media sentiments, and automotive news articles can provide richer insights into the factors influencing resale value. Natural language processing and sentiment analysis techniques can be employed to process and extract relevant information from textual data.

2) **Integration of IoT and Telematics Data:** Connected car technologies and IoT devices can provide valuable data on vehicle usage, maintenance history, driving patterns, and more. Integrating such data into resale value prediction models can offer a more comprehensive understanding of a vehicle's condition and its impact on its future value.

3) **Real-Time Market Monitoring:** The resale value of used cars is influenced by dynamic market conditions, including changes in supply and demand, economic factors, and consumer preferences. Future models can leverage real-time data streams and develop algorithms that continuously monitor and adapt to market fluctuations, enabling more accurate and up-to-date predictions.

4) **Integration of Advanced Machine Learning Techniques:** Currently, predictive models for resale value estimation primarily rely on traditional machine learning algorithms. However, the future scope involves exploring advanced techniques such as deep learning, reinforcement learning, and ensemble methods to enhance the accuracy and reliability of predictions.

5) **Enhanced Data Visualization and Interpretability:** As the complexity of predictive models increases, it becomes essential to improve data visualization techniques to effectively communicate the predictions and underlying factors to end-users. Interactive dashboards and visual representations can assist users in understanding and interpreting the resale value predictions.

6) **Geographic-Specific Models:** Resale value can vary significantly based on geographical location, regional preferences, and market dynamics. Developing location-specific models that consider local factors and regional market trends can provide more accurate predictions for specific areas, facilitating better decision-making for both buyers and sellers.

Overall, the future scope for "Used Car Resale Value Prediction" using applied data science involves leveraging advanced techniques, incorporating diverse data sources, adapting to real-time market conditions, and ensuring ethical considerations. These advancements have the potential to enhance accuracy, provide better decision support, and enable more personalized and tailored experiences for stakeholders in the used car market.

11. BIBLIOGRAPHY

- https://ieeexplore.ieee.org/abstract/document/10112343

- https://www.sciencedirect.com/science/article/abs/pii/S016920701730050X

- https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=car+resale+value+projects&btnG=

- https://www.w3schools.com/tags/tag_output.asp#:~:text=The%20tag%20is%20used,one%20performed%20by%20a%20script

- https://www.techiedelight.com/set-width-input-textbox-html-css-javascript/#:~:text=In%20HTML%2C%20you%20can%20use,the%20width%20of%20an%20element

SOURCE CODE :

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import LabelEncoder
import pickle

df                                                                    =
pd.read_csv(r"C:\Users\user\Desktop\CAR_RESALE_PROJECT\autos.csv
", sep=',',encoding='Latin1',)

df.head()

df.describe()

df.info()

#print al the different sellers
print(df.seller.value_counts())
#remove the seller type having only 3 cars
df[df.seller!='gewerblich']
#now all the sellers are same so we can get rid of this column
df=df.drop('seller', 1)

#print al the different sellers
print(df.offerType.value_counts())
```

```
#remove the Offer Type having only 12 listings
df[df.offerType!='Gesuch']
#now all the offers are same so we can get rid of this column
df=df.drop('offerType',1)

print(df.shape)

df.head()

#Cars having power less than 50ps and above 900ps seems a little
suspicious,
#let's remove them and see what we 've got now
print (df. shape)
df = df[(df.powerPS > 50) & (df.powerPS < 900)]
print (df. shape)
#around 50000 cars have been removed which could have introduced
error to our data
#similarly, filtering our the cars having registration years not
in the mentioned range
print(df. shape)
df = df[(df.yearOfRegistration > 1950) & (df.yearOfRegistration
< 2017)]
print (df. shape)
#not much of a difference but still, 10000 rows have been
reduced. it's bette
#get rid of faulty data instead of keeping them just to increase
the size.
#to

df.yearOfRegistration.value_counts()

#removing irrelevant columns which are
#introduce bias, so removing them too-
df.drop(['name','abtest','dateCrawled','postalCode','dateCreated
','nrOfPictures','lastSeen','postalCode','dateCreated'],axis='co
lumns',inplace=True)

df.head()

#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
#only first occurrence of any such row is kept
new_df = df. copy()
new_df =
new_df.drop_duplicates(['price','vehicleType','yearOfRegistration','gearbox','powerPS','model','monthOfRegistration','kilometer','fuelType','notRepairedDamage'])

#As the dataset contained some german words for many features, changing 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)

#### Removing the outliers
new_df = new_df[(new_df.price >= 100) & (new_df.price <= 150000)]
#Fi11ing NaN values for columns whose data might not be
#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 ('not-declared',inplace=True)
new_df['gearbox'].fillna('not-declared',inplace=True)
new_df['vehicleType'].fillna('not-declared',inplace=True)
new_df['model'].fillna('not-declared',inplace=True)

#can save the csv for future purpose-
new_df.to_csv("autos_preprocessed.csv")

#Columns which contain categorical values, which we'll need to convert via labe1 encoding
```

```
labels = [ 'gearbox', 'notRepairedDamage' , 'model' ,'brand',
'fuelType', 'vehicleType']

#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)
#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)

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

#need to to reshape the Y values
Y=Y.reshape(-1,1)

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)
d=pd.DataFrame(X)
d.head()
#d.isnull().sum()

#Model building and Fitting
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import r2_score
```

```
regressor =
RandomForestRegressor(n_estimators=1000,max_depth=10,random_stat
e=34)

#fitting the model
regressor.fit(X_train, np.ravel(Y_train,order='C'))

#predicting the values for testset
y_pred = regressor.predict(X_test)

#printing the Accuracy for test set
print(r2_score(Y_test,y_pred))

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

app.py CODE :

```
import pandas as pd
import numpy as np
from flask import Flask, render_template, Response, request
import pickle
from sklearn.preprocessing import LabelEncoder

app = Flask(__name__,
template_folder=r'C:\Users\user\Desktop\CAR_RESALE_PROJECT\Flask
\templates',
static_folder=r'C:\Users\user\Desktop\CAR_RESALE_PROJECT\Flask\s
tatic')


filename =
r'C:\Users\user\Desktop\CAR_RESALE_PROJECT\Flask\resale_model.sa
v'
model_rand = pickle.load(open(filename, 'rb'))

@app.route('/',methods=['GET'])
def index():
    return render_template('resaleintro.html')
```

```
@app.route('/predict')
def predict():
    return render_template('resalepredict.html')

@app.route('/y_predict', methods=['POST'])
def y_predict():
    regyear = int(request.form['regyear'])
    powerps = float(request.form['powerps'])
    kms = float(request.form['kms'])
    regmonth = int(request.form.get('regmonth'))
    gearbox = request.form['gearbox']
    damage = request.form['dam']
    model = request.form.get('modeltype')
    brand = request.form.get('brand')
    fuelType = request.form.get('fuel')
    vehicletype = request.form.get('vehicletype')
     new_row = {'yearOfRegistration':regyear, 'powerPS':powerps,
'kilometer':kms,
              'monthOfRegistration':regmonth,  'gearbox':gearbox,
'notRepairedDamage':damage,
       'model':model, 'brand':brand, 'fuelType':fuelType,
       'vehicleType':vehicletype}
    print(new_row)
            new_df   =   pd.DataFrame(columns   =['vehicleType',
'yearOfRegistration', 'gearbox',
                                 'powerPS', 'model', 'kilometer',
'monthOfRegistration', 'fuelType',
                                'brand', 'notRepairedDamage'] )
    new_df = new_df.append(new_row,ignore_index = True)
     labels = ['gearbox', 'notRepairedDamage', 'model', 'brand',
'fuelType', 'vehicleType']
    mapper = {}
    for i in labels:
        mapper[i] = LabelEncoder()
                                        mapper[i].classes_      =
np.load(str('classes'+i+'.npy'),allow_pickle=True)
        tr = mapper[i].fit_transform(new_df[i])
                    new_df.loc[:,i+'_labels']   =   pd.Series(tr,
index=new_df.index)
    labeled = new_df[ ['yearOfRegistration'
                        ,'powerPS'
```

```
                        ,'kilometer'
                        ,'monthOfRegistration'
                        ]
                    + [x+'_labels' for x in labels]]
    X = labeled.values
    print(X)
    y_prediction = model_rand.predict(X)
    print(y_prediction)
       return  render_template('resalepredict.html',ypred  =  'The
resale value predicted is {:.2f}$'.format(y_prediction[0]))

if __name__ == '__main__':
    app.run(host='localhost', debug=True, threaded=False)
```