**VEHICLE RESALE VALUE PREDICTION**

**INTRODUCTION:**

**Machine learning** is an application of artificial **intelligence** (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. **Machine learning** focuses on the development of computer programs that can access data and use it learn for themselves. Machine learning combines data with statistical tools to predict an output.

**For example, medical diagnosis, image processing, prediction , classfication ,learning association, regression , Financial Services,** **Marketing,automation,etc.,**

Machine learning is also used for a variety of task like fraud detection, predictive maintenance, portfolio optimization, automatize task and so on.



Supervised Learning:

An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output.

REGRESSION:

When the output is a continuous value, the task is a regression. The system will be trained to estimate the price of the stocks with the lowest possible error.

Multiple linear regression (MLR): also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression (MLR) is to model the [linear relationship](https://www.investopedia.com/terms/l/linearrelationship.asp) between the explanatory (independent) variables and response (dependent) variable.

### The Formula for Multiple Linear Regression Is:

*yi*​=*β*0​+*β*1​*xi*1​+*β*2​*xi*2​+...+*βp* ​*xip*​+*ϵ*

where, for *i*=*n* observations:

*yi*​=dependent variable

*xi*​=independent variables

*β*0​=y-intercept (constant term)

*βp*=slope coefficients for each explanatory variable

*ϵ*=the model’s error term (also known as the residuals)​

**Python** is a general-purpose high level programming language that is being increasingly used in data science and in designing machine learning algorithms.

Python is a popular platform used for research and development of production systems. It is a vast language with number of modules, packages and libraries that provides multiple ways of achieving a task.

Python and its libraries like NumPy, SciPy, Scikit-Learn, Matplotlib are used in data science and data analysis.

* **numpy** − is used for its N-dimensional array objects
* **pandas** − is a data analysis library that includes dataframes
* **matplotlib** − is 2D plotting library for creating graphs and plots
* **scikit-learn** − the algorithms used for data analysis and data mining tasks
* **seaborn** − a data visualization library based on matplotlib

To install Python and other scientific computing and machine learning packages simultaneously, we should install **Anaconda** distribution. It is a Python implementation for Linux, Windows and OSX, and comprises various machine learning packages like numpy, scikit-learn, and matplotlib. It also includes **Jupyter Notebook**, an interactive Python environment. We can install Python 2.7 or any 3.x version as per our requirement.

**OBJECTIVE:**

Main idea of this project is to build a model that determines if the asking price for a particular car is reasonable given the information provided in the listing. and also to analyze car ads posted on the website, and predict the price of a car in relation to all the other cars posted on the website.

**PROBLEM STATEMENT:**

The price of a new car is typically much more expensive than if you bought the vehicle used. Be prepared to pay much more in sales tax. New vehicles are said to lose up to 20 percent of their value as soon as you drive off the lot. so, we prepare a model for the customers to predict the price of the used cars depends on number of factors like vehicle type, brand, fuel type, kilometers, mileage, etc.,

**INDUSTRY PROFILE:**

In the automotive industry machine learning is most often associated with product innovations such as self-driving cars, used car resale value prediction etc.,Exploring ML in automotive marketing to reach customers at the right moment with a tailored, individual messages. Through programmatic advertising, with AI at it's very core, automotive players can let AI determine which customers they should target, with which messages, at what time, and via which channel.

**DATA COLLECTION:**

1. www.kaggle.com/orgesleka/used-cars-database
2. www.data.gov.in

**DATA SET:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Vehicle Type** | **yearOf**  **Registration** | **Gear**  **Box** | **kilometer** | **notRepairedDamage** | **fuelType** | **monthOf**  **Registration** | **brand** | **price** |
| kombi | 1993 | Manual | 150000 | yes | petrol | 0 | volkswagen | 480000 |
| coupe | 2011 | Manual | 125000 | yes | diesel | 5 | audi | 183000 |
| Suv | 2004 | automatic | 125000 | No | diesel | 8 | jeep | 98000 |
| Kleinwagen | 2001 | Manual | 150000 | no | petrol | 6 | volkswagen | 158000 |
| Kleinwagen | 2008 | Manual | 90000 | no | diesel | 7 | skoda | 360000 |
| Limousine | 1995 | Manual | 150000 | yes | petrol | 10 | bmw | 65000 |
| Cabrio | 2004 | Manual | 150000 | no | petrol | 8 | peugeot | 220000 |
| Limousine | 1980 | Manual | 40000 | no | petrol | 7 | volkswagen | 10000 |
| Bus | 2014 | Manual | 30000 | yes | petrol | 8 | ford | 145000 |

**METHODOLOGY:**

In this model we use "Supervised Learning Model" . In that we apply "Multiple Linear Regression model". Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression (MLR) is to model the [linear relationship](https://www.investopedia.com/terms/l/linearrelationship.asp) between the explanatory (independent) variables and response (dependent) variable.

**(i)EXPLORATORY DATA ANALYSIS:**

|  |  |
| --- | --- |
| Vehicle type | Different names of the cars |
| Year of registration | Defines the year of car registered |
| Gear box | Defines either the car gear box is manual or automatic |
| Kilometer | Defines the distance it covers |
| Not repaired damage | Defines either there is a damage of car or not |
| Fuel type | Defines either the fuel type is petrol or diesel |
| Month of registration | Defines the month of car registered |
| Brand | Brand of the car |
| Price | Gives the resale price of the car |

**(ii)Figures and Tables:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Vehicle Type** | **yearOf**  **Registration** | **Gear**  **Box** | **kilometer** | **notRepairedDamage** | **fuelType** | **monthOf**  **Registration** | **brand** | **price** |
| Kombi | 1993 | Manual | 150000 | yes | petrol | 0 | volkswagen | 480000 |
| Coupe | 2011 | Manual | 125000 | yes | diesel | 5 | audi | 183000 |
| Suv | 2004 | automatic | 125000 | No | diesel | 8 | jeep | 98000 |
| Kleinwagen | 2001 | Manual | 150000 | no | petrol | 6 | volkswagen | 158000 |
| Kleinwagen | 2008 | Manual | 90000 | no | diesel | 7 | skoda | 360000 |
| Limousine | 1995 | Manual | 150000 | yes | petrol | 10 | bmw | 65000 |
| Cabrio | 2004 | Manual | 150000 | no | petrol | 8 | peugeot | 220000 |
| Limousine | 1980 | Manual | 40000 | no | petrol | 7 | volkswagen | 10000 |
| Bus | 2014 | Manual | 30000 | yes | petrol | 8 | ford | 145000 |

**(iii)Data Model and visualization:**

In that we apply "Multiple Linear Regression model".

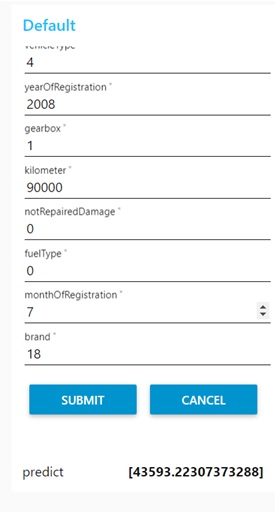
steps involved in doing this model:

1. #### Importing the Libraries ```python import numpy as np import matplotlib.pyplot as plt import pandas as pd ```
2. #### Importing the Dataset ```python dataset = pd.read\_csv('vehicle\_resale.csv’) X = dataset.iloc[:, :8].values , y = dataset.iloc[:, 8:].values ```
3. #### Encoding Categorical Data ```python from sklearn.preprocessing import LabelEncoder
4. #### Avoiding the Dummy Variable Trap ```python X = X[:, 1:] ```
5. #### Splitting the Dataset ```python from sklearn.cross\_validation import train\_test\_split X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size = 0.2, random\_state = 0) ```
6. #### Fitting Multiple Linear Regression to the Training set ```python from sklearn.linear\_model import LinearRegression regressor = LinearRegression() regressor.fit(X\_train, y\_train) ```
7. #### Predicting the Test Set Results```python y\_pred = regressor.predict(X\_test) ```
8. visualize the model
9. run the code in watson studio
10. construct the flow in NodeRed and predict the value.

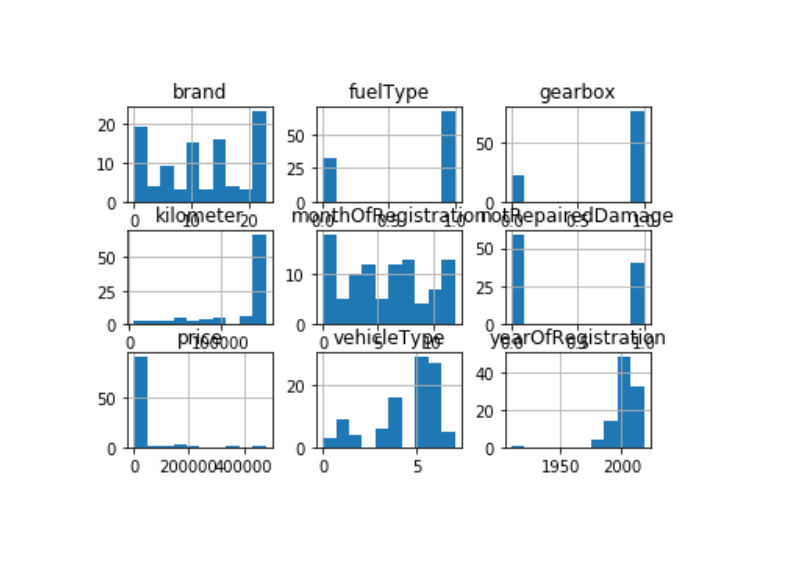
**NODE-RED:**

Node-RED is a **programming tool** for wiring together hardware devices, APIs and online services in new and interesting ways. It provides a browser-based editor that makes it easy to wire together flows using the wide range of nodes in the palette that can be deployed to its runtime in a single-click.

**OUTPUT:**



**Visualization**:



**FINDINGS AND SUGGESTIONS:**

The main limitation of this model is the low number of attributes that have been used. As future work, we intend to collect more data and to use more advanced techniques like artificial neural networks, fuzzy logic and genetic algorithms to predict used car prices.

**CONCLUSION:**

Car price prediction can be a challenging task due to the high number of attributes that should be considered for the accurate prediction. The major step in the prediction process is collection and preprocessing of the data.

**REVIEW OF LITERATURE:**

1. www.kaggle.com/orgesleka/used-cars-database
2. www.data.gov.in
3. [https://www.investopedia.com](https://www.investopedia.com/)
4. [https://www.ripublication.com](https://www.ripublication.com/)
5. https://www.researchgate.net