Webology (ISSN: 1735-188X)
Volume 18, Number 6, 2021

Used Car Price Prediction Using Machine Learning Techniques

Mrs Shyamali Das[1] , Mr Ananta Laha[2] , Mr Alok Jena[3] , Ms Priyadarshini Samal[4]

[1]Asst Professor in BPUT Affiliated Engineering College -Bhubaneswar.

[2]Associeate Consultant (TCS) -Bhubaneswar.

[3]Asst Professor in Andhra Affiliated Eng College – Vijayanagaram.

[4] Asst Professor in BPUT Affiliated Engineering College , Bhubaneswar.

Abstract: The price of a new car is being set by the manufacture by having a significant profit value with inclusive to government forms and taxes. Sometime, customers start thinking on value for price on the money while the even buy a car, especially the middle/economic class or some of the travel agents and assistance for which this detailed analysis would be very much helpful. This also covers the accuracy and precisions into prediction of used cars into great extent which can help the travel industry audiences. This also explains the comparisons between linear regression, lasso regression and logistic regression results on same data sets and their respective accuracies and precisions. Post these comparisons among all these techniques, we also detailed about heat map i.e. relative study of dependent and independent variables in same topic, classifications using Random Forest and

This comparison studies would help to get a better precision and accuracy with a predicted value rather than a range of values with a continuous output value. This also covers the details about the various outliers for the prediction methodologies being used.

General Terms: Data Mining, Machine Learning, Artificial Intelligence

Keywords— Used Car Prediction, Lasso Regression, Logistic Regression, Random Forest Classifier, Heat Map. Supervised Learning.

1. Introduction:

Predicting price of a car when it is not coming directly from a factory is really a challenging and critical task, with demand of used cars resale and buying is quiet increasing now-a-days and recent era the fuel price challenges are significantly increasing which becomes more challenging for the used car seller. People and organization prefer to have a legal between seller and buyer with a business part on the estimated price, thus finding a fair estimation is important and crucial step for buyers and predicting them with highest accuracy is really a great. Predicting the actual price of a car with a greater extent of precision would help really the buyers, so we had used the various supervised learning algorithms such as

1- Liner Regression Model
2- Lasso Regression Model
3- Stats Regression Model
4- Random Forest ML Classification Techniques
5- Heat Map correlation and Further scope

1. **Linear Regression:**
Linear Regression is most used Machine Learning supervised algorithm which works on train to predict a well established output that is dependent on the input data. These algorithms generally trains the set and results the output. Regression Analysis is about a predictive modeling methodology that has a objective to investigate the relation ship between various input data. For simple regression problem (a single x and single y) the format model follows as

$$\boxed{Y = B0 + B1 * X}$$

When we move on higher model and discuss on complexity of the model that varies as per B0 and B1 Values.

Example : Weight =B0 +B1 * height

Using the coefficient values will help you predicting the Weight values as per the height which falls into Linear Regression model.

2. **Lasso Regression Model**
Regularization is very effective model of Linear Regression adding with penalty term when the test and trained data is varying significantly. Usually, we reduce the magnitude of the coefficients for different ML techniques. Lasso Regression is being used for accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.

Mathematical Equation:

Residual Sum Squares + λ * (Sum of Absolute Value of the magnitude coefficient)

$$\sum_{i=1}^{n}\left(y_i - \sum x_{ij} B_j\right)^2 + \lambda \sum_{j=1}^{P}\left|B_j\right|$$

Where λ denotes the amount of shrinkage

λ = 0 Refers to all features considered and equivalent to linear regression and sum of squares to be considered

λ = ∞ Refers no features considered, closes to infinity reduce features

Variance to be increased or decreased by λ value.

3. **Statistics Regression Model**

This model used for statistical methodology to explore the data with a descriptive statistical analysis, This used with scikit-learn but may not be hard core statistics in python , This uses the simple methodologies (fit->transform-> Predict)

4. **Random Forest ML Classification Techniques**

A random forest is a supervised machine learning algorithm that is constructed from decision tree algorithms. This is a classification of regression trees. This algorithm works on possible illustrated graphs and decision trees where set of trees would move to the actual predicted output closer. Decision trees are in parallel and used with bagging techniques. This works efficiently as classification algorithms with large data set with better accuracy but where as it is slow in training and may create a biased values on categorized variables. Maximum depth of trees determines the overfitting.

5. **Heat Map:**
Heatmap is a way to show some sort of matrix plot. To use a heatmap the data should be in a matrix form. By matrix we mean that the index name and the column name must match in some way so that the data that we fill inside the cells are relevant.

1.1.Literature Survey

The price of a pre-owned car depends on various factors including model year, mileage, condition, equipment, etc. Since the price depends on so many factors, it is difficult to estimate it directly using rule-based algorithms. A more feasible strategy is to use inductive based learning to learn the price from the dataset. Hence, a machine learning approach is very suitable for this application. References to authors of “Application of ML techniques to predict the price of Pre-owned cars in

Bangladesh” have been performed with only restricted fields analysis and this is only restricted into Bangladesh.The second paper explains about Support Vector machine through the accurate price and the third paper is based on big data analysis and ANN neural networks which largely varies for vehicles The fourth and further papers are only explains the results Linear Regression, Ridge Regression , Lasso Regression and this paper aims also includes comparative study of Random Forest including with all other mentioned machine learning techniques.

1.2.Hardware /Software Requirements

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

Software Requirements : Python Pycharm PIP 2.7 Jupyter Notebook Chrome

2. Methodologies

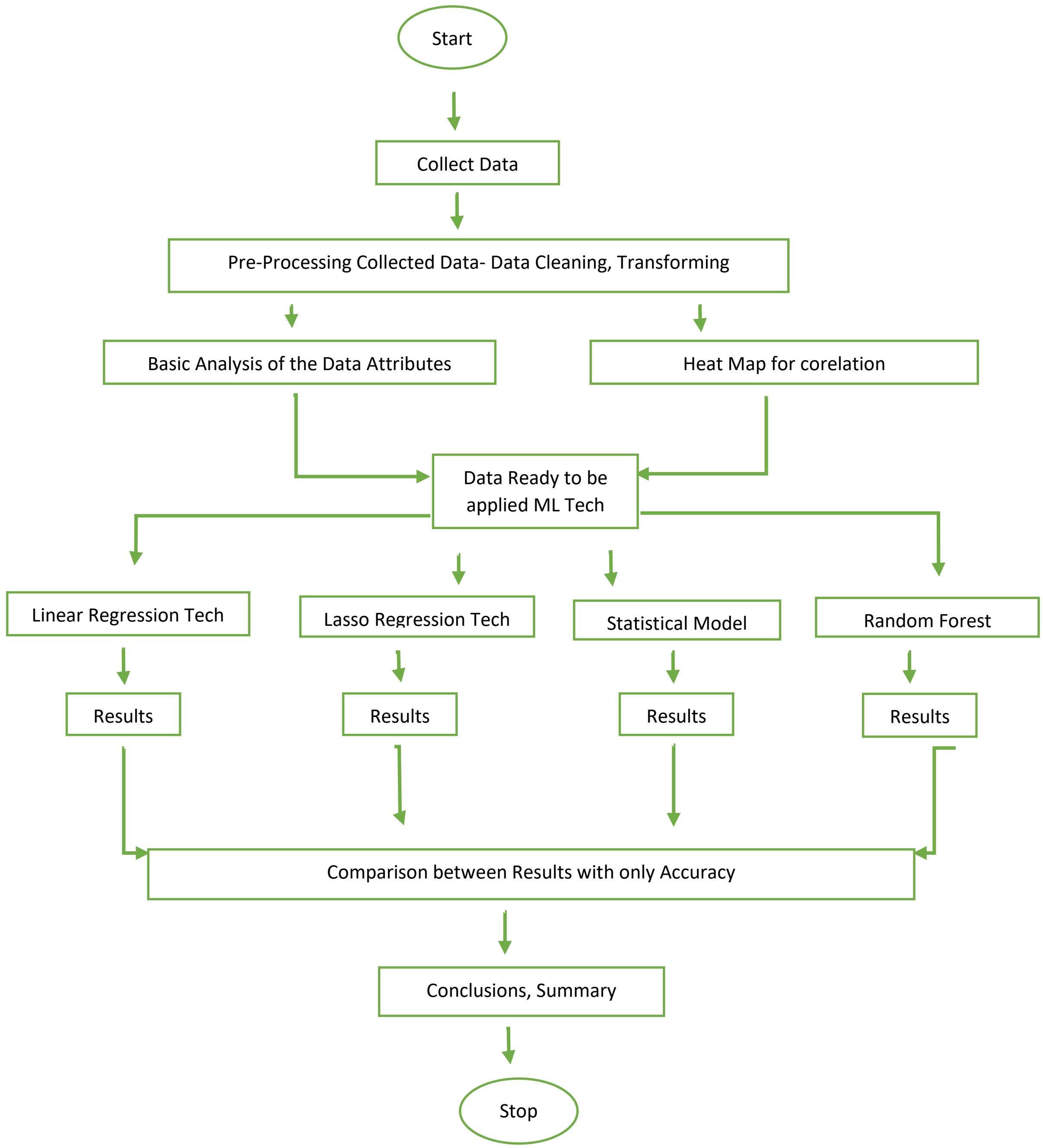
2.1.Background

We started collecting the regular data by Kaggle and data crawled to prepare the data set for training which took almost one month and prior to this literature survey took 2-3 weeks and a team of 4 people have been contributed as follows.

Mr Anant & Alok has contributed with Linear and Lasso regression techniques which consumed one additional month where the results were not much satisfied hence we got into a decision with all other co-authors where Mrs Shayamali & Priyadarshini has been contributed to Random forest post heat map derived which has better results than linear regression results.

2.2.Step Wise Process for Purposed Model

Figure :1



2.3.Collected Data Sample

The data have been collected over 20K Indian data samples where we have collected with various open source types and classified as below variables

Table-1

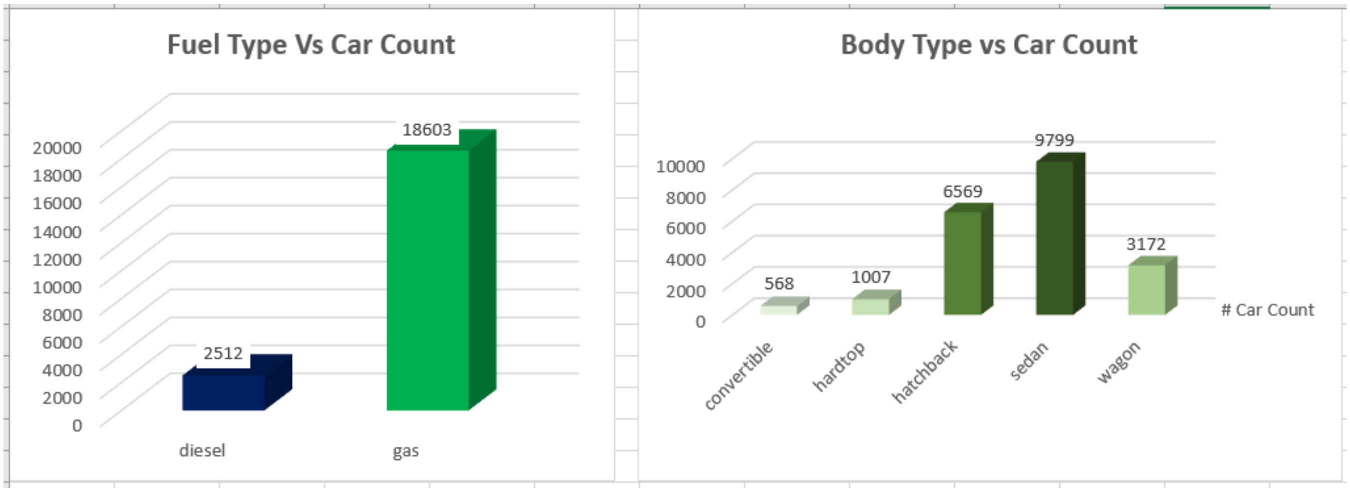
Sl No	Variable Name	Description
1	Car Name	Name of the Car

2	Fuel type	Type of the Fuel being Used
3	aspiration	Std /Turbo - IC aspiration
4	Door number	Number of doors
5	Car body	Body of Car
6	Drive wheel	Wheel details
7	Engine location	Location of Engines
8	wheelbase	Base of Wheel
9	Car length	Length of Car
10	Car width	Width of Car
11	Car height	Height of Car
12	Curb weight	Weight of Cube
13	Engine type	Type of Engine
14	Cylinder number	Cylinder No
15	Engine size	Size of Engine
16	Fuel system	Fuel System details
17	Bore ratio	Bore Ratio for the care
18	stroke	Stroke Ratio of Car
19	Compression ratio	Compression Ratio
20	horsepower	Horse Power
21	Peak rpm	Peak RPM
22	City mpg	Lowest MPG Rating for Car
23	Highway mpg	Average MPG Rating of Car
24	price	Price of Car

2.4.Basic Data Analysis & Visuals

Fuel Type Vs Car count and Body Type Vs Car Count can be presented as below with actual data.

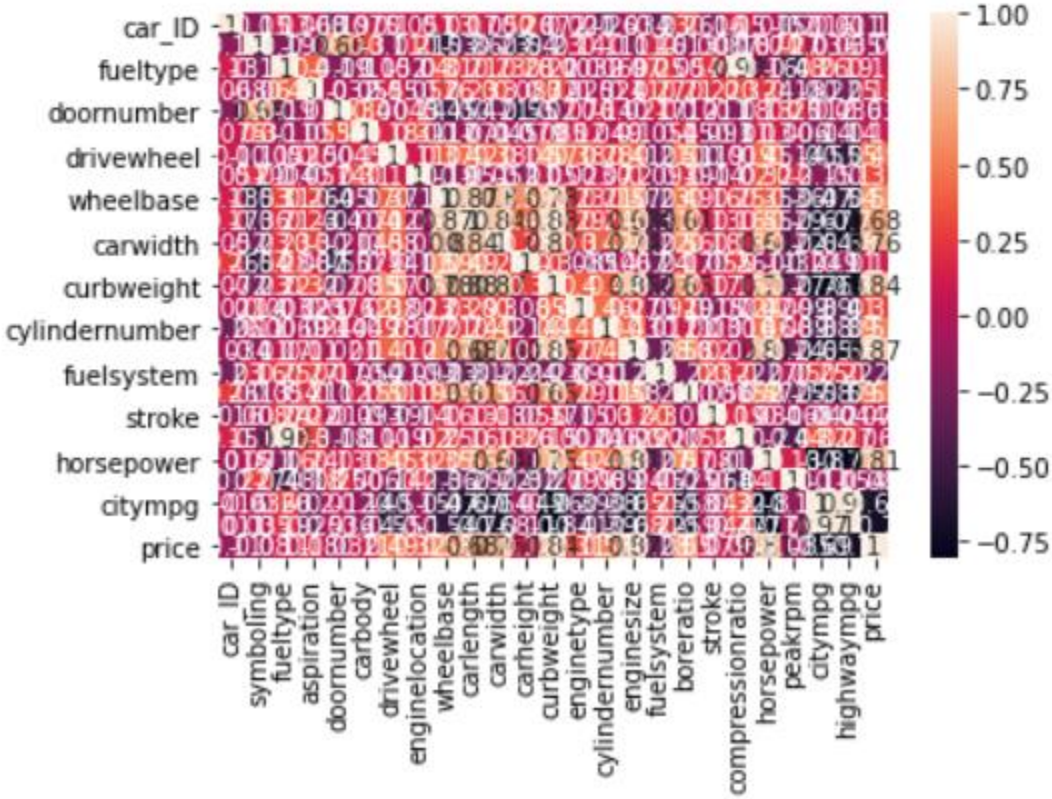
Figure -2



This has been performed to understand very basic feasibility on the relation towards targeted price vs Fuel Type and Body Type and then we decided to follow heat map to find the better relation ship between target (Price) and dependent variables (all other)

2.5.Heat Map

This is a matrix or graphical way of presentations where the data can be utilized color-coded systems, The primary purpose of heat map to locate and assist the data sets to viewers. This corelates between the target and dependent variables. As per below heat map Target Variable (Price) is negatively correlated with “city mpg” ,” highway mpg” and positively corelated with “Wheel base” and “Car Width “ ,”Car Length” , and e.t.c.

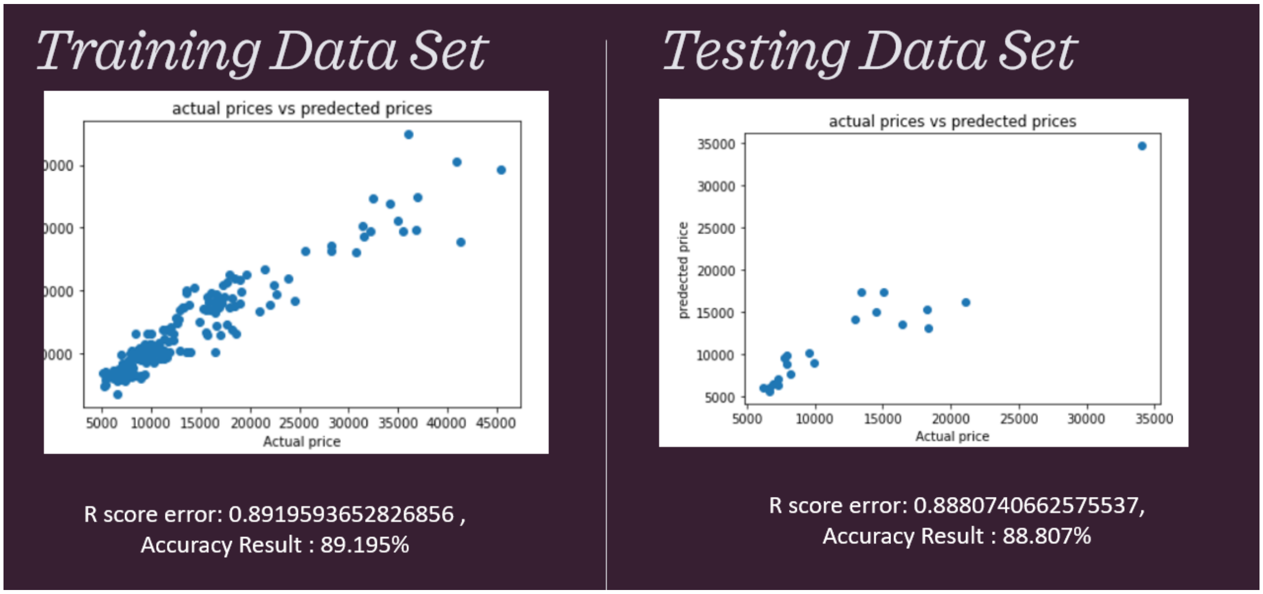


3. Outputs with Accuracy Results

3.1.Linear Regression:

Applied Linear Regression Algorithms with Training Data Set and Testing data set results as followed: Linear Regression is a type of supervised machine learning algorithm which is used to predict the value of a dependent variable based on the value of another independent variable. Here the model finds the best fit linear line between the independent and dependent variable.

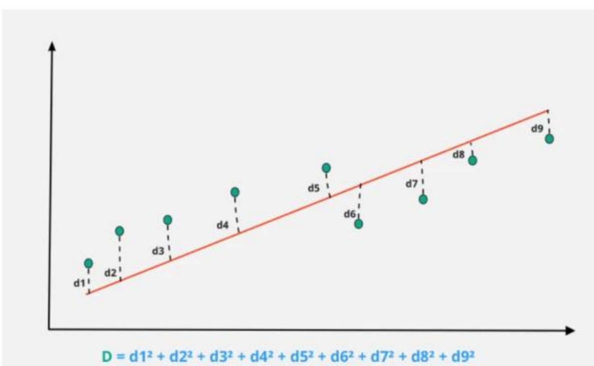
1. Data Sourcing, Data Understanding
2. Data cleaning, Manipulation, Visualization and Detecting Outliers
3. Perform EDA on Prepared Dataset (Univariate and Bivariate Analysis)
4. Model Preparation
5. Training and Testing set Data Split
6. Model Building
7. Residual Analysis of the Train Data
8. Making Predictions
9. Model Evaluation
10. Final Interface



3.2.Lasso Regression - Least Absolute Shrinkage and Selection Operator

This is L1 regularization method

The Statistics Of Lasso Regression



d1, d2, d3, etc., represents the distance between the actual data points and the model line in the above graph.

Least-squares is the sum of squares of the **distance between the points** from the plotted curve.

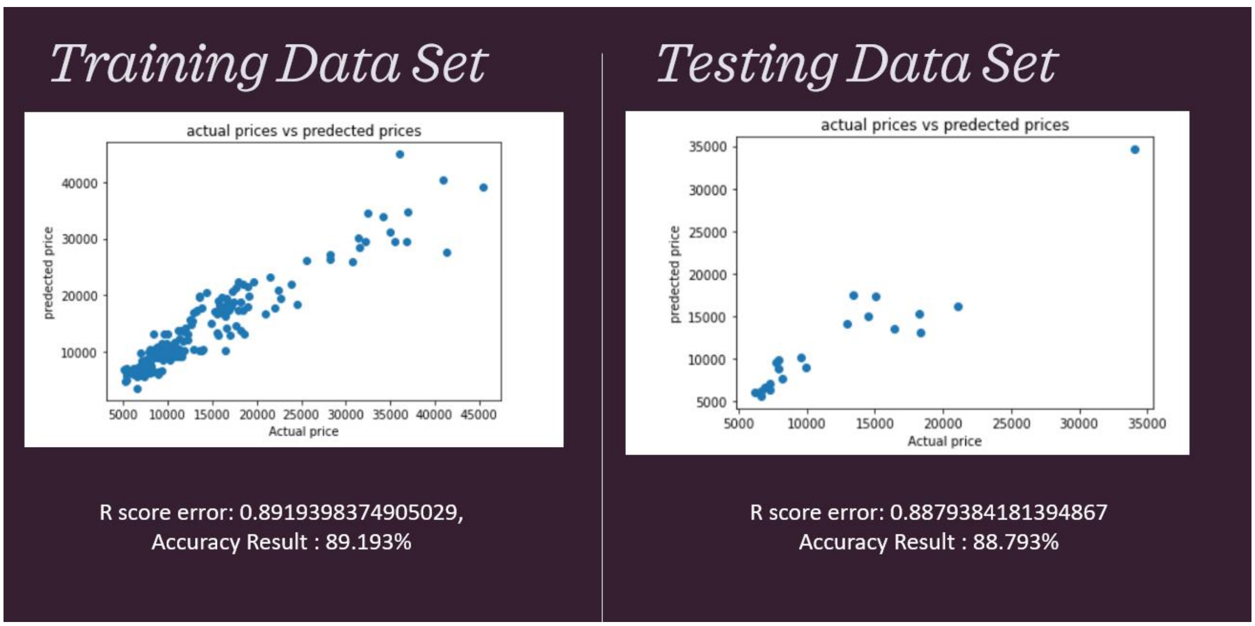
In linear regression, the best model is chosen in a way to **minimize** the least-squares.

While performing lasso regression, we add a penalizing factor to the least-squares. That is, the model is chosen in a way to reduce the below loss function to a minimal value.

D = least-squares + lambda * summation (absolute values of the magnitude of the coefficients)

Lasso regression penalty consists of all the estimated parameters. Lambda can be any value between zero to infinity. This value decides how aggressive regularization is performed. It is usually chosen using cross-validation.

Lasso penalizes the sum of absolute values of coefficients. As the lambda value increases, coefficients decrease and eventually **become zero**. This way, lasso regression eliminates insignificant variables from our model. Our regularized model may have a slightly high bias than linear regression but less variance for future predictions.



3.3.Statistical Model

ML includes random forests, recursive partitioning (CART), bagging, boosting, support vector machines, neural networks, and deep learning .This consists multiple iterations to get the better accuracy by removing one by one attributes that are not needed. Statistical modeling is the process of applying statistical analysis to a dataset. A statistical model is a mathematical representation (or mathematical model) of observed data.

Dep. Variable:	price	**R-squared:**	0.885
Model:	OLS	**Adj. R-squared:**	0.873
Method:	Least Squares	**F-statistic:**	70.73
Date:	Sat, 07 May 2022	**Prob (F-statistic):**	8.22E-68
Time:	15:39:26	**Log-Likelihood:**	-1717.8

No. Observations:	184	**AIC:**	3474
Df Residuals:	165	**BIC:**	3535
Df Model:	18		
Covariance Type:	nrobust		

Iteration-1

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-7.00E+04	1.42E+04	-4.933	0	-9.81E+04	-4.20E+04
aspiration	2034.8216	847.73	2.4	0.017	361.025	3708.618
doornumber	199.7214	605.634	0.33	0.742	-996.07	1395.513
carbody	90.6465	252.42	0.359	0.72	-407.743	589.036
drivewheel	1148.2433	502.632	2.284	0.024	155.824	2140.663
enginelocation	1.28E+04	2490.133	5.156	0	7923.546	1.78E+04
wheelbase	111.5045	94.53	1.18	0.24	-75.141	298.15
carlength	39.3484	57.126	0.689	0.492	-73.444	152.141
carwidth	712.8731	237.65	3	0.003	243.647	1182.1
cylindernumber	1104.1492	278.862	3.959	0	553.551	1654.747
enginesize	117.7557	14.96	7.871	0	88.218	147.294
fuelsystem	-152.854	218.778	-0.699	0.486	-584.82	279.112
boreratio	-1592.6456	1357.128	-1.174	0.242	-4272.221	1086.929
stroke	-1833.0059	904.835	-2.026	0.044	-3619.553	-46.459
compressionratio	56.7956	85.934	0.661	0.51	-112.876	226.467
horsepower	-5.5863	17.874	-0.313	0.755	-40.877	29.704
peakrpm	2.2502	0.723	3.113	0.002	0.823	3.677
citympg	-192.6521	186.045	-1.036	0.302	-559.987	174.683
highwaympg	211.6861	172.246	1.229	0.221	-128.405	551.777

Next Attribute to be removed as" door number"

Next "Car Body"

Next "Compression Ratio"

Next "Highwaympg"

Iteration-2

coef	std err	t	P>|t|	[0.025	0.975]	
Intercept	-6.95E+04	1.41E+04	-4.943	0	-9.72E+04	-4.17E+04
aspiration	2029.837	845.316	2.401	0.017	360.88	3698.794
carbody	115.3606	240.39	0.48	0.632	-359.256	589.977
drivewheel	1156.8401	500.606	2.311	0.022	168.464	2145.216
enginelocation	1.27E+04	2463.793	5.17	0	7872.682	1.76E+04
wheelbase	109.0143	93.975	1.16	0.248	-76.526	294.555
carlength	33.2603	53.916	0.617	0.538	-73.188	139.709
carwidth	719.8038	236.082	3.049	0.003	253.693	1185.915
cylindernumber	1110.6266	277.422	4.003	0	562.897	1658.356
enginesize	118.0021	14.901	7.919	0	88.582	147.423
fuelsystem	-148.4609	217.785	-0.682	0.496	-578.447	281.525
boreratio	-1562.3482	1350.375	-1.157	0.249	-4228.471	1103.775
stroke	-1794.3371	894.793	-2.005	0.047	-3560.978	-27.696
compressionratio	57.0137	85.7	0.665	0.507	-112.189	226.217
horsepower	-4.8894	17.701	-0.276	0.783	-39.837	30.058
peakrpm	2.2595	0.72	3.137	0.002	0.837	3.682
citympg	-190.8985	185.469	-1.029	0.305	-557.08	175.283
highwaympg	208.8422	171.568	1.217	0.225	-129.894	547.578

Iteration-3

coef	std err	t	P>|t|	[0.025	0.975]	
Intercept	-6.80E+04	1.38E+04	-4.912	0	-9.53E+04	-4.06E+04
aspiration	1955.7068	835.185	2.342	0.02	306.897	3604.516
drivewheel	1153.9207	494.236	2.335	0.021	178.208	2129.634
enginelocation	1.30E+04	2391.886	5.414	0	8227.065	1.77E+04
wheelbase	136.631	70.519	1.938	0.054	-2.586	275.848
carwidth	744.9677	232.329	3.207	0.002	286.307	1203.628
cylindernumber	1086.9441	272.997	3.982	0	547.998	1625.891
enginesize	118.3352	14.826	7.981	0	89.065	147.605
fuelsystem	-179.8639	212.149	-0.848	0.398	-598.686	238.958
boreratio	-1362.455	1318.049	-1.034	0.303	-3964.527	1239.617
stroke	-1720.7381	885.486	-1.943	0.054	-3468.852	27.376
compressionratio	71.8082	81.499	0.881	0.38	-89.087	232.703
horsepower	-5.5433	17.545	-0.316	0.752	-40.181	29.094
peakrpm	2.2069	0.714	3.092	0.002	0.798	3.616
citympg	-213.4806	177.928	-1.2	0.232	-564.744	137.783
highwaympg	203.8794	168.856	1.207	0.229	-129.474	537.232

Iteration-4

coef	std err	t	P>|t|	[0.025	0.975]	
Intercept	-7.18E+04	1.31E+04	-5.466	0	-9.77E+04	-4.59E+04
aspiration	2164.6108	800.293	2.705	0.008	584.751	3744.47
drivewheel	1212.8491	489.365	2.478	0.014	246.793	2178.906
enginelocation	1.35E+04	2316.245	5.815	0	8897.118	1.80E+04
wheelbase	141.0595	70.293	2.007	0.046	2.295	279.824
carwidth	795.6409	224.949	3.537	0.001	351.569	1239.713
cylindernumber	1124.2214	269.52	4.171	0	592.162	1656.281
enginesize	119.1	14.791	8.052	0	89.901	148.299
fuelsystem	-156.095	210.288	-0.742	0.459	-571.224	259.034
boreratio	-1298.5385	1315.18	-0.987	0.325	-3894.835	1297.758
stroke	-1619.809	877.464	-1.846	0.067	-3352.012	112.394
horsepower	-7.6816	17.365	-0.442	0.659	-41.962	26.599
peakrpm	2.1181	0.706	3	0.003	0.724	3.512
citympg	-189.3441	175.691	-1.078	0.283	-536.175	157.487
highwaympg	206.7427	168.713	1.225	0.222	-126.314	539.799

Next "Citympg"

Next "Horse Power"

Next "Fuel System"

Iteration-5

coef	std err	t	P>|t|	[0.025	0.975]	
Intercept	-7.04E+04	1.31E+04	-5.371	0	-9.62E+04	-4.45E+04
aspiration	1864.7717	763.094	2.444	0.016	358.412	3371.132
drivewheel	1040.2299	469.343	2.216	0.028	113.738	1966.722
enginelocation	1.38E+04	2306.454	5.971	0	9219.068	1.83E+04
wheelbase	122.5523	68.752	1.783	0.076	-13.166	258.271
carwidth	819.3144	224.448	3.65	0	376.249	1262.379
cylindernumber	1099.9531	269.188	4.086	0	568.572	1631.334
enginesize	112.431	13.774	8.163	0	85.241	139.62
fuelsystem	-169.0383	210.332	-0.804	0.423	-584.237	246.16
boreratio	-1132.5522	1310.114	-0.864	0.389	-3718.74	1453.635
stroke	-1414.1482	862.534	-1.64	0.103	-3116.806	288.509
horsepower	-1.366	16.607	-0.082	0.935	-34.149	31.416
peakrpm	1.9655	0.696	2.824	0.005	0.592	3.339
citympg	12.5237	61.161	0.205	0.838	-108.21	133.257

Iteration-6

coef	std err	t	P>|t|	[0.025	0.975]	
Intercept	-6.91E+04	1.15E+04	-6.018	0	-9.17E+04	-4.64E+04
aspiration	1881.5951	756.53	2.487	0.014	388.255	3374.935
drivewheel	1018.9958	456.46	2.232	0.027	117.974	1920.018
enginelocation	1.38E+04	2293.589	6.02	0	9279.846	1.83E+04
wheelbase	120.1918	67.589	1.778	0.077	-13.224	253.608
carwidth	815.1187	222.884	3.657	0	375.16	1255.077
cylindernumber	1099.1854	268.407	4.095	0	569.369	1629.002
enginesize	112.7639	13.639	8.268	0	85.841	139.686
fuelsystem	-169.6223	209.723	-0.809	0.42	-583.601	244.356
boreratio	-1187.394	1278.848	-0.928	0.354	-3711.756	1336.968
stroke	-1427.514	857.648	-1.664	0.098	-3120.455	265.427
horsepower	-2.6839	15.266	-0.176	0.861	-32.818	27.45
peakrpm	1.9411	0.684	2.839	0.005	0.591	3.291

Iteration-7

coef	std err	t	P>|t|	[0.025	0.975]	
Intercept	-6.85E+04	1.09E+04	-6.288	0	-8.99E+04	-4.70E+04
aspiration	1814.5382	651.502	2.785	0.006	528.57	3100.507
drivewheel	1006.189	449.339	2.239	0.026	119.261	1893.117
enginelocation	1.38E+04	2282.348	6.038	0	9276.191	1.83E+04
wheelbase	124.524	62.759	1.984	0.049	0.646	248.402
carwidth	807.8299	218.377	3.699	0	376.787	1238.873
cylindernumber	1090.1158	262.659	4.15	0	571.665	1608.566
enginesize	111.0106	9.278	11.965	0	92.697	129.324
fuelsystem	-158.7282	199.794	-0.794	0.428	-553.092	235.636
boreratio	-1254.4472	1217.209	-1.031	0.304	-3657.037	1148.143
stroke	-1424.4705	855.054	-1.666	0.098	-3112.222	263.281
peakrpm	1.8625	0.516	3.61	0	0.844	2.881

Next "Bore Ratio"

Next "Stroke"

Next "Wheel base "

Iteration-8						
coef	std err	t	P>|t|	[0.025	0.975]	
Intercept	-6.96E+04	1.08E+04	-6.45	0	-9.09E+04	-4.83E+04
aspiration	1664.2041	622.751	2.672	0.008	435.036	2893.372
drivewheel	1019.9378	448.526	2.274	0.024	134.649	1905.226
enginelocation	1.38E+04	2279.851	6.051	0	9294.74	1.83E+04
wheelbase	127.8147	62.556	2.043	0.043	4.344	251.286
carwidth	824.9565	217.078	3.8	0	396.494	1253.419
cylindernumber	1040.8741	254.969	4.082	0	537.623	1544.125
enginesize	112.6064	9.048	12.445	0	94.747	130.465
boreratio	-1284.3822	1215.328	-1.057	0.292	-3683.161	1114.397
stroke	-1626.7566	815.393	-1.995	0.048	-3236.156	-17.357
peakrpm	1.8815	0.515	3.655	0	0.865	2.898

Iteration-9						
coef	std err	t	P>|t|	[0.025	0.975]	
Intercept	-7.10E+04	1.07E+04	-6.636	0	-9.21E+04	-4.99E+04
aspiration	1618.7001	621.469	2.605	0.01	392.112	2845.289
drivewheel	891.1626	431.802	2.064	0.041	38.918	1743.407
enginelocation	1.34E+04	2256.367	5.958	0	8990.851	1.79E+04
wheelbase	126.3131	62.561	2.019	0.045	2.838	249.789
carwidth	773.5629	211.632	3.655	0	355.867	1191.259
cylindernumber	1118.2123	244.324	4.577	0	635.993	1600.432
enginesize	110.0064	8.71	12.63	0	92.815	127.198
stroke	-1383.7879	782.573	-1.768	0.079	-2928.345	160.769
peakrpm	1.9355	0.512	3.777	0	0.924	2.947

Iteration-10						
coef	std err	t	P>|t|	[0.025	0.975]	
Intercept	-7.53E+04	1.05E+04	-7.178	0	-9.60E+04	-5.46E+04
aspiration	1362.728	608.036	2.241	0.026	162.701	2562.755
drivewheel	1152.0178	408.28	2.822	0.005	346.232	1957.803
enginelocation	1.41E+04	2242.399	6.272	0	9639.277	1.85E+04
wheelbase	119.9254	62.835	1.909	0.058	-4.086	243.937
carwidth	793.741	212.604	3.733	0	374.143	1213.339
cylindernumber	1123.6581	245.784	4.572	0	638.575	1608.741
enginesize	105.8511	8.438	12.545	0	89.198	122.504
peakrpm	1.8406	0.513	3.59	0	0.829	2.853

Iteration -11						
coef	std err	t	P>|t|	[0.025	0.975]	
Intercept	-7.96E+04	1.03E+04	-7.713	0	-1.00E+05	-5.92E+04
aspiration	1351.0973	612.553	2.206	0.029	142.203	2559.992
drivewheel	1255	407.725	3.078	0.002	450.34	2059.66
enginelocation	1.34E+04	2227.106	5.993	0	8951.109	1.77E+04
carwidth	1055.7247	163.568	6.454	0	732.918	1378.531
cylindernumber	1035.5292	243.214	4.258	0	555.539	1515.519
enginesize	105.9679	8.501	12.466	0	89.191	122.745
peakrpm	1.6185	0.503	3.217	0.002	0.626	2.611

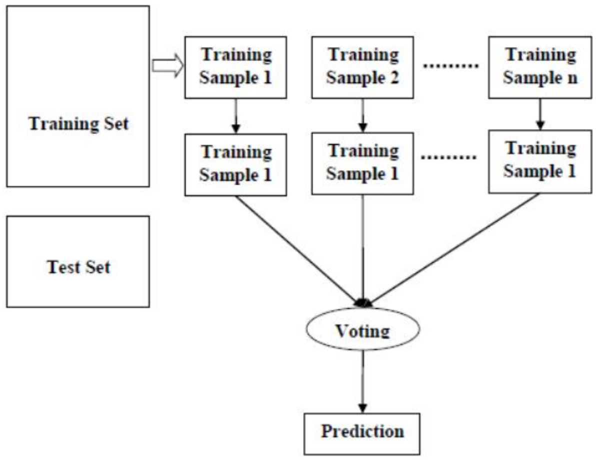
3.4.Random Forest Implementations

It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result. We can understand the working of Random Forest algorithm with the help of following steps

We can understand the working of Random Forest algorithm with the help of following steps −

- Step 1 − First, start with the selection of random samples from a given dataset.
- Step 2 − Next, this algorithm will construct a decision tree for every sample. Then it will get the prediction result from every decision tree.
- Step 3 − In this step, voting will be performed for every predicted result.
- Step 4 − At last, select the most voted prediction result as the final prediction result.

The following diagram will illustrate its working –



Benefits:

- Overcomes the problem of overfitting of combining decision trees.
- Works for large set of data
- This has less variance to single decision tree
- Proves high accuracy

Implementing Random forest on last Statistics modeling resulted as
R score error: 0.9143581729539816 Accuracy: 91.435%

Accuracy by Implementation of Random Forest 91. 435 %

3.5.Comparison between all Results

This comparative study evidence that application of "Random Forest with Statistical Modeling proved the better accuracy on old car price prediction over other supervised machine learning techniques.

4. Conclusion

The detailed study of the Machine Learning Techniques used with prediction of used Car Prices through various Supervised Learning approaches as Linear , Lasso, Statistical and Random forest model applied with Training and test set of data and Random forest over multiple iteration produced a great accuracy approx. 91.5% and it also leaves the further research methodologies to be applied as deep learning systems like ANN , B-Networks methodologies. This analysis definitely help the researcher and users widely on determination of prices for old cars in India.

The current analysis has been done with open source data base but if could be improvised by association of “ True Value” or similar industry player who can provide the recent actual data set to be trained and tested then that could result better class definition with greater accuracy.

5. References

[1] “Vehicle Price Prediction using SVM Techniques” in IJITEE on June, 2020 by S.E.Viswapriya, Durbaka Sai Sandeep Sharma, Gandavarapu Sathya kiran.

[2] “Application of Machine Learning Techniques to Predict the Price of Pre-Owned Cars in Bangladesh” published on MDPI by Fahad Rahman Amik , Akash Lanard , Ahnaf Ismat and Sifat Momen

[3] “USED CAR PRICE PREDICTION” published in IRJET on 4th APR 2021 by Praful Rane, Deep Pandya, Dhawal Kotak

[4] “Vehicle Price Prediction System using Machine Learning Techniques” published in IJCA on 9th June 2017 by Kanwal Noor, Sadaqat Jan

[5] “Car’s Selling Price Prediction using Random Forest Machine Learning Algorithm.” Published by Abhishek Pandey , Vanshika Rastogi , Sanika Singh

[6] “Car Price Prediction using Machine Learning Techniques” published in TEM Journal on Feb 2019 by Enis Gegic, Becir Isakovic, Dino Keco, Zerina Masetic, Jasmin Kevric

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

[8] 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)

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

[10] 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)