INDEX

LIST OF FIGURES

LIST OF TABLE

CHAPTER 1
INTRODUCTION

Car Price Prediction Model: -

The used car market is an ever-rising industry, which has almost doubled its market value in the last few years. The emergence of online portals such as Car Dekho, Quikr, Carwale, Cars24, and many others has facilitated the need for both the customer and the seller to be better informed about the trends and patterns that determine the value of the used car in the market. Machine Learning algorithms can be used to predict the retail value of a car, based on a certain set of features. Different websites have different algorithms to generate the retail price of the used cars, and hence there isn't a unified algorithm for determining the price. By training statistical models for redacting the prices, one can easily get a rough estimate of the price without actually entering the details into the desired website. The main objective of this paper is to use three different prediction models to predict the retail price of a used car and compare their levels of accuracy. The data set used for the prediction models was created in 2020 by Car Dekho uploaded now on Kaggle, whose retail prices have been calculated. The dataset primarily comprises of categorical attributes along with two quantitative attributes.

1.1 Purpose : -

The Main purpose of creating this project is to create a smart way to identify the prices and values of commodities online by some practical inputs. By using this application on can predict their cars and bike prices online on the basis of machine learning model and also there is no need of any offline broker or agent to identify their price or values.

1.2 Problem Statement : -

The prices of new cars in the industry are fixed by the manufacturer with some additional costs incurred by the Government in the form of taxes. So, customers buying a new car can be assured of the money they invest to be worthy. But due to the increased price of new cars and the incapability of customers to buy new cars due to the lack of funds, used cars sales are on a global increase. There is a need for a used car price prediction system to effectively determine the worthiness of the car using a variety of features. Even though there are websites that offers this service, their prediction method may not be the best. Besides, different models and systems may contribute on predicting power for a used car's actual market value. It is important to know their actual market value while both buying and selling.

1.3 Overview :-

Required of prediction system: To effectively determine the worthiness of things by giving accurate result by prediction for better decision.
Target Business: Online Used-car dealers.
Target Customers: Buyers who would like to purchase used cars via online portal.

1.4 Objective:-

The main objective of the project is to create an online predictor which can predict different prices, size, occurrences, probability that allows users to find out best prices or accuracy of the things on which they are searching for.

1.5 Tools and Technology:

- Python.
- Google Colab.
- Machine Learning.
- Flask
- Visual Studio Code
- Github

CHAPTER 2
FEASIBLITY STUDY

2.1 Research Paper-1:-

The first paper is Predicting the worth of Used automobile exploitation Machine Learning techniques. In this paper, they investigate the application of supervised machine learning techniques to predict the worth of used cars in Mauritius. The predictions are supported historical knowledge collected from daily newspapers. Different techniques like multiple regression toward the mean analysis-nearest neighbours, native Bayes and call trees have been accustomed create the predictions.

Automobile value Prediction victimization Machine Learning Techniques considerable range of distinct attributes square measure examined for the reliable and correct prediction. to create a model for predicting the worth of used cars in Bosnia-Herzegovina and Herzegovina,[1] they need applied machine learning techniques (Artificial Neural Network, Support Vector Machine and Random Forest).

The worth analysis model in used car system supported neural networks. during this paper, the price analysis model supported massive information analysis is proposed, that takes advantage of wide circulated vehicle data and an oversized variety of car group action information to analyse the value information for every sort of vehicles by victimization the optimized neural network rule. It aims to ascertain a second-hand automobile worth analysis model to induce the value that best matches the automobile.

A wide variety of approaches, algorithms, statistical software, and validation strategies were employed in the application of machine learning methods to inform provider for decision making. There is a need to ensure that multiple machine learning approaches are used, the model selection strategy is clearly defined, and both internal and external validation are necessary to be sure that decisions for predictions are being made with the highest quality evidence.

2.2 Research Paper-2 :-

Overfitting and underfitting come into picture when we create our statistical models. The models might be too biased to the training data and might not perform well on the test data set. This is called overfitting. Likewise, the models might not take into consideration all the variance present in the population and perform poorly on a test data set. This is called underfitting. A perfect balance needs to be achieved between these two, which leads to the concept of Bias-Variance tradeoff.
Pierre Geurts [2] has introduced and explained how bias-variance tradeoff is achieved in both regression and classification. The selection of variables/attribute plays a vital role in influencing both the bias and variance of the statistical model. Robert Tibshirani
[3] proposed a new method called Lasso, which minimizes the residual sum of squares. This returns a subset of attributes which need to be included in multiple regression to get the minimal error rate. Similarly, decision trees suffer from overfitting if they are not

pruned/shrunk. Trevor Hastie and Daryl Pregibon. [4] have explained the concept of pruning in their research paper. Moreover, hypothesis testing using ANOVA is needed to verify whether the different groups of errors really differ from each other. This is explained by TK Kim and Tae Kyun in their paper [5]. A Post-Hoc test needs to be performed along with ANOVA if the number of groups exceeds two. Tukey's Test has been explored by Haynes W. in his research paper.

2.3 Machine Learning Model Structure :-

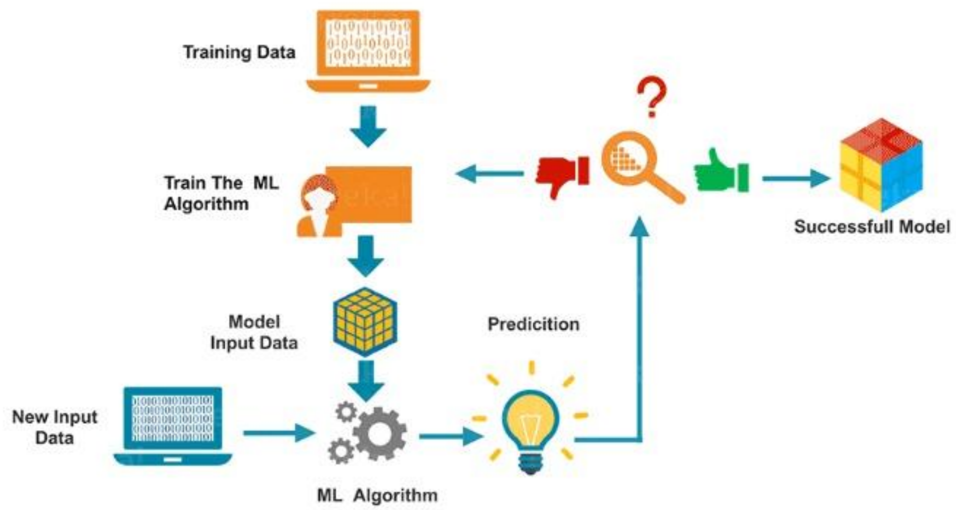

Figure 1 Machine Learning Structure

2.4 Implementation Steps: -

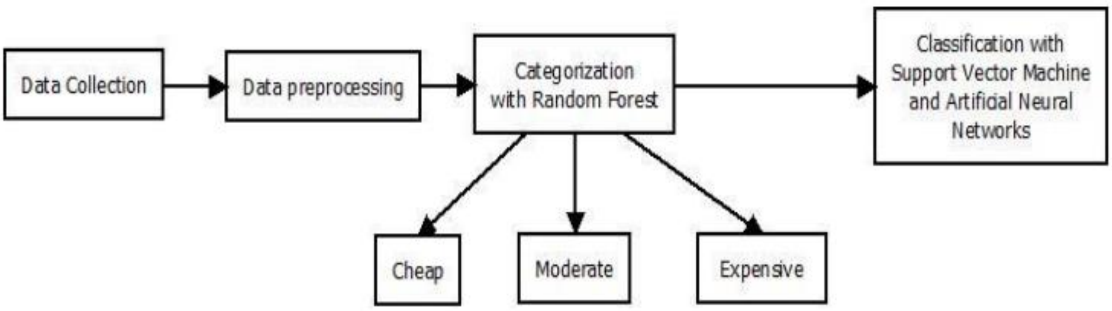

Figure 2 Block Diagram for all Classification Process

CHAPTER 3
SYSTEM REQUIREMENTS STUDY

3.1 Hardware & Software Requirements : -

	Hardware	**Software**
Developers	1. 8 GB RAM 2. 256 GB Storage 3. Intel i5 10th Gen + Processor	1. Collab or Python 2. Pycharm IDE 3. Vscode
Users	1. Windows PC, Mobiles, Tablets. 2. Min 4 GB RAM	1. Chrome, Edge, Firefox.

Table 1 : Hardware and Software Requirement

CHAPTER 4
PROTOTYPE

4.1 Design :-

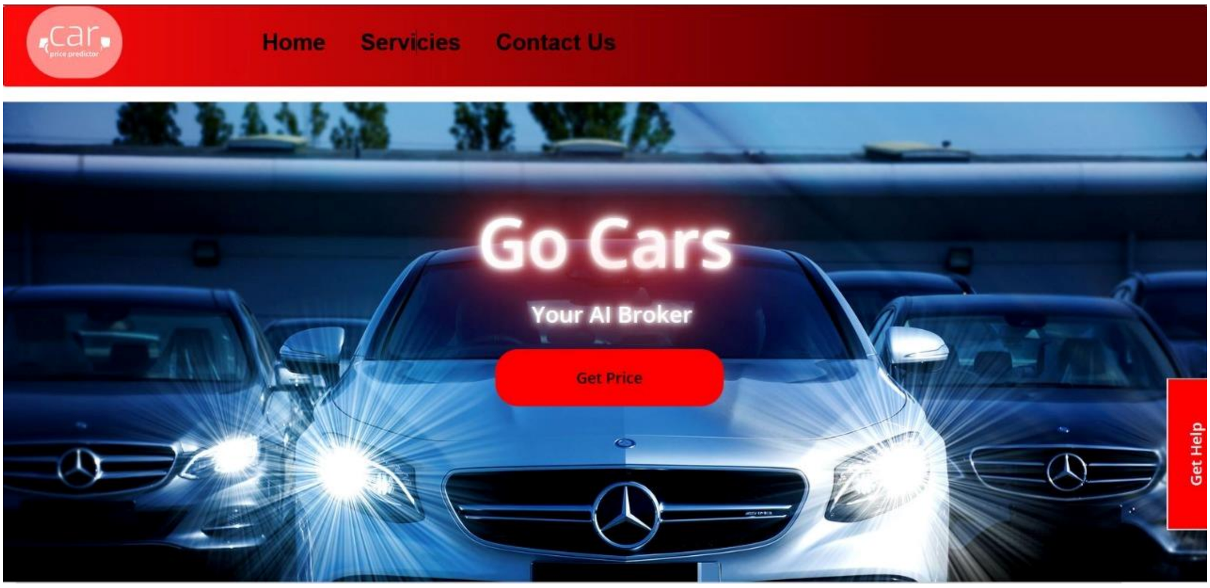

Figure 3 Frontend UI

4.2 Designing Tools :-

HTML5
CSS 3.0
BOOTSTRAP
Javascript
Django else Flask

CHAPTER 5
DATABASE STRATEGY

Database:-

5.1 Rejected Databases :-

A) Car price Assignment (Kaggle) :-

car_ID	symboling	CarName	fueltype	aspiration	doornumbe	carbody	drivewheel	engineloca	wheelbase	carlength	carwidth	carheight	curbweight	enginetype	cylindernu	enginesize	fuelsystem	boreratio	stroke	compressic	ho
1	3	alfa-romer	gas	std	two	convertible	rwd	front	88.6	168.8	64.1	48.8	2548	dohc	four	130	mpfi	3.47	2.68	9	
2	3	alfa-romer	gas	std	two	convertible	rwd	front	88.6	168.8	64.1	48.8	2548	dohc	four	130	mpfi	3.47	2.68	9	
3	1	alfa-romer	gas	std	two	hatchback	rwd	front	94.5	171.2	65.5	52.4	2823	ohcv	six	152	mpfi	2.68	3.47	9	
4	2	audi 100 ls	gas	std	four	sedan	fwd	front	99.8	176.6	66.2	54.3	2337	ohc	four	109	mpfi	3.19	3.4	10	
5	2	audi 100ls	gas	std	four	sedan	4wd	front	99.4	176.6	66.4	54.3	2824	ohc	five	136	mpfi	3.19	3.4	8	
6	2	audi fox	gas	std	two	sedan	fwd	front	99.8	177.3	66.3	53.1	2507	ohc	five	136	mpfi	3.19	3.4	8.5	
7	1	audi 100ls	gas	std	four	sedan	fwd	front	105.8	192.7	71.4	55.7	2844	ohc	five	136	mpfi	3.19	3.4	8.5	
8	1	audi 5000	gas	std	four	wagon	fwd	front	105.8	192.7	71.4	55.7	2954	ohc	five	136	mpfi	3.19	3.4	8.5	
9	1	audi 4000	gas	turbo	four	sedan	fwd	front	105.8	192.7	71.4	55.9	3086	ohc	five	131	mpfi	3.13	3.4	8.3	
10	0	audi 5000s (	gas	turbo	two	hatchback	4wd	front	99.5	178.2	67.9	52	3053	ohc	five	131	mpfi	3.13	3.4	7	
11	2	bmw 320i	gas	std	two	sedan	rwd	front	101.2	176.8	64.8	54.3	2395	ohc	four	108	mpfi	3.5	2.8	8.8	
12	0	bmw 320i	gas	std	four	sedan	rwd	front	101.2	176.8	64.8	54.3	2395	ohc	four	108	mpfi	3.5	2.8	8.8	
13	0	bmw x1	gas	std	two	sedan	rwd	front	101.2	176.8	64.8	54.3	2710	ohc	six	164	mpfi	3.31	3.19	9	
14	0	bmw x3	gas	std	four	sedan	rwd	front	101.2	176.8	64.8	54.3	2765	ohc	six	164	mpfi	3.31	3.19	9	
15	1	bmw z4	gas	std	four	sedan	rwd	front	103.5	189	66.9	55.7	3055	ohc	six	164	mpfi	3.31	3.19	9	
16	0	bmw x4	gas	std	four	sedan	rwd	front	103.5	189	66.9	55.7	3230	ohc	six	209	mpfi	3.62	3.39	8	
17	0	bmw x5	gas	std	two	sedan	rwd	front	103.5	193.8	67.9	53.7	3380	ohc	six	209	mpfi	3.62	3.39	8	
18	0	bmw x3	gas	std	four	sedan	rwd	front	110	197	70.9	56.3	3505	ohc	six	209	mpfi	3.62	3.39	8	
19	2	chevrolet i	gas	std	two	hatchback	fwd	front	88.4	141.1	60.3	53.2	1488	l	three	61	2bbl	2.91	3.03	9.5	
20	1	chevrolet n	gas	std	two	hatchback	fwd	front	94.5	155.9	63.6	52	1874	ohc	four	90	2bbl	3.03	3.11	9.6	
21	0	chevrolet v	gas	std	four	sedan	fwd	front	94.5	158.8	63.6	52	1909	ohc	four	90	2bbl	3.03	3.11	9.6	
22	1	dodge ram	gas	std	two	hatchback	fwd	front	93.7	157.3	63.8	50.8	1876	ohc	four	90	2bbl	2.97	3.23	9.41	
23	1	dodge chall	gas	std	two	hatchback	fwd	front	93.7	157.3	63.8	50.8	1876	ohc	four	90	2bbl	2.97	3.23	9.4	
24	1	dodge d20C	gas	turbo	two	hatchback	fwd	front	93.7	157.3	63.8	50.8	2128	ohc	four	98	mpfi	3.03	3.39	7.6	
25	1	dodge mon	gas	std	four	hatchback	fwd	front	93.7	157.3	63.8	50.6	1967	ohc	four	90	2bbl	2.97	3.23	9.4	
26	1	dodge colt	gas	std	four	sedan	fwd	front	93.7	157.3	63.8	50.6	1989	ohc	four	90	2bbl	2.97	3.23	9.4	
27	1	dodge colt	gas	std	four	sedan	fwd	front	93.7	157.3	63.8	50.6	1989	ohc	four	90	2bbl	2.97	3.23	9.4	
28	1	dodge coro	gas	turbo	two	sedan	fwd	front	93.7	157.3	63.8	50.6	2191	ohc	four	98	mpfi	3.03	3.39	7.6	
29	-1	dodge dart	gas	std	four	wagon	fwd	front	103.3	174.6	64.6	59.8	2535	ohc	four	122	2bbl	3.34	3.46	8.5	

Reason's Why Rejected:

- We rejected this dataset because it doesn't have the cars showroom price. So without the comparison between cars road price and showroom price we cannot predict the price of cars by just cars showroom price.
- In this Dataset we don't get accuracy as per mark.
- This dataset contain so 26 columns which includes various unwanted columns and by removing those columns affects the accuracy and coorelation for heatmap

Rows And Columns :-

No. of Rows	205
No. of Cols.	26

A) Car price assignment features :-

car_ID
symboling
CarName
Fueltype
aspiration
doornumber
carbody
drivewheel
enginelocation
wheelbase
carlength
curbweight
enginetype
cylindernumber
enginesize
boreratio
stroke
compression
horsepower
highwaympg
peakrpm
Price
citympg

B. New Used car Dataset (Kaggle) :-

Car Names	Mileages	Ratings	Reviews	used/certif	price drop	Price
2019 Honda	Mileage	3.7	(3 reviews)	Used	$501 price c	$27,995
2015 Chevr	29,192 mi.	3.7	(3 reviews)	Used	$277 price c	$52,000
2020 Jeep G	53,561 mi.	4.8	(1,147 revie	Used	$2,000 price	$41,998
2019 Audi A	46,664 mi.	4.9	(1,831 revie	Used	$2,090 price	$62,750
2017 Ford N	21,726 mi.	4	(409 review	Used	$200 price c	$29,995
2016 Jeep G	66,514 mi.	4.6	(1,348 revie	Used	$2,000 price	$26,499
2018 Jeep G	52,607 mi.	3.3	(329 review	Used	$2,074 price	$39,944
2016 Ford F	45,046 mi.	4.3	(73 reviews	Ford Certifi	$2,250 price	$14,756
2021 Ford F	43,502 mi.	3.5	(193 review	Used	$890 price c	$123,500
2019 GMC Y	5,386 mi.	4.7	(267 review	Used	$5,000 price	$61,974
2022 Kia Te	62,129 mi.	4.8	(1,305 revie	Used	$110 price c	$54,500
2018 Jeep V	10,100 mi.	4.6	(317 review	Jeep Certifi	$2,000 price	$36,277
2015 Toyota	68,884 mi.	4.4	(580 review	Used	$3,200 price	$33,050
2020 Toyota	57,311 mi.	5	(3 reviews)	Used	$738 price c	$33,677
2016 INFINI	21,713 mi.	4.7	(1,338 revie	Used	$1,489 price	$28,900
2017 Ford F	74,765 mi.	5	(1 review)	Used	$340 price c	$56,999
2013 Lexus	66,000 mi.	4.7	(659 review	Used	$670 price c	$16,990
2021 Mitsul	48,605 mi.	3.3	(239 review	Used	$499 price c	$15,165
2022 Subaru	19,844 mi.	4.6	(2,411 revie	Used	$400 price c	$39,747
2017 Lexus	28,271 mi.	3.7	(3 reviews)	Used	$240 price c	$29,500
2021 Toyota	56,748 mi.	4.1	(57 reviews	Used	$500 price c	$97,900
2017 BMW I	11,194 mi.	4.7	(743 review	Used	$284 price c	$53,789
2011 Toyota	19,847 mi.	4.6	(33 reviews	Used	$1,896 price	$8,999
2020 Ford N	178,400 mi.	4.7	(738 review	Used	$2,469 price	$30,223
2019 Acura	45,111 mi.	4.7	(2,792 revie	Used	$620 price c	$24,995
2019 Subaru	41,045 mi.	4.4	(39 reviews	Used	$248 price c	$32,900
2018 Nissar	19,181 mi.	4.7	(1,338 revie	Used	$1,067 price	$29,997
2020 Lambc	21,256 mi.	4.8	(1,666 revie	Used	$1,000 price	$280,000

Reason's Why Rejected :-

- This Dataset contain foreign countries cars descriptions and models that's why that dataset dosen't fit in our model.
- This Dataset doesn't contain our requirements to predict the car prices and Dosent give Correlation for Heatmap.

New Used car Dataset Columns Name :-

Car Name
Mileages
Ratings
Reviews
used/certified
price drop

Rows And Columns :

No. of Rows	9330
No. of Cols.	6

5.1 Final Databases :-

full_name	selling_price	new_price	year	seller_type	km_driven	owner_type	fuel_type	transmission_t	mileage	engine	max_power	seats
Maruti Alto Std	1.2 Lakh*	null	2012	Individual	1,20,000 kms	First Owner	Petrol	Manual	Mileage19.7 kmpl	Engine796 CC	Max Power46.3	Seats5
Hyundai Grand i10 Asta	5.5 Lakh*	New Car (On-	2016	Individual	20,000 kms	First Owner	Petrol	Manual	Mileage18.9 kmpl	Engine1197 CC	Max Power82 bh	Seats5
Hyundai i20 Asta	2.15 Lakh*	null	2010	Individual	60,000 kms	First Owner	Petrol	Manual	Mileage17.0 kmpl	Engine1197 CC	Max Power80 bh	Seats5
Maruti Alto K10 2010-20	2.26 Lakh*	null	2012	Individual	37,000 kms	First Owner	Petrol	Manual	Mileage20.92 kmpl	Engine998 CC	Max Power67.1	Seats5
Ford Ecosport 2015-202	5.7 Lakh*	New Car (On-	2015	Dealer	30,000 kms	First Owner	Diesel	Manual	Mileage22.77 kmpl	Engine1498 CC	Max Power98.5	Seats5
Maruti Wagon R VXI BS	3.5 Lakh*	New Car (On-	2013	Individual	35,000 kms	First Owner	Petrol	Manual	Mileage18.9 kmpl	Engine998 CC	Max Power67.1	Seats5
Hyundai i10 Sportz 1.2	3.15 Lakh*	New Car (On-	2013	Dealer	40,000 kms	First Owner	Petrol	Manual	Mileage20.36 kmpl	Engine1197 CC	Max Power78.9	Seats5
Maruti Wagon R VXI	4.1 Lakh*	New Car (On-	2018	Dealer	17,512 kms	First Owner	Petrol	Manual	Mileage20.51 kmpl	Engine998 CC	Max Power67.0	Seats5
Hyundai Venue SX Plus	10.5 Lakh*	New Car (On-	2019	Individual	20,000 kms	First Owner	Petrol	Automatic	Mileage18.15 kmpl	Engine998 CC	Max Power118.	Seats5
Mahindra TUV 300 T6	5.75 Lakh*	null	2017	Dealer	70,000 kms	First Owner	Diesel	Manual	Mileage18.49 kmpl	Engine1493 CC	Max Power100	Seats7
Tata Indigo LX (TDI) BS-	3.05 Lakh*	null	2015	Individual	50,000 kms	First Owner	Diesel	Manual	Mileage19.09 kmpl	Engine1405 CC	Max Power69.01	Seats5
Renault Captur Platine	11.5 Lakh*	null	2019	Individual	18,000 kms	First Owner	Diesel	Manual	Mileage20.37 kmpl	Engine1461 CC	Max Power108.	Seats5
Maruti Swift VXI with A	5.11 Lakh*	New Car (On-	2017	Dealer	28,321 kms	First Owner	Petrol	Manual	Mileage16.6 kmpl	Engine1197 CC	Max Power85 bh	Seats5
Nissan Micra XL CVT	4.1 Lakh*	null	2016	Dealer	27,000 kms	First Owner	Petrol	Automatic	Mileage19.34 kmpl	Engine1198 CC	Max Power76 bh	Seats5
Hyundai Verna 1.6 SX	4.25 Lakh*	New Car (On-	2013	Dealer	65,278 kms	First Owner	Diesel	Manual	Mileage22.32 kmpl	Engine1582 CC	Max Power126.	Seats5
Renault Duster 110PS D	7.5 Lakh*	null	2016	Individual	50,000 kms	First Owner	Diesel	Manual	Mileage19.64 kmpl	Engine1461 CC	Max Power108.	Seats5
Mini Cooper Cooper S	32.5 Lakh*	null	2017	Dealer	6,000 kms	First Owner	Petrol	Automatic	Mileage14.41 kmpl	Engine1998 CC	Max Power189.	Seats5
Maruti Ciaz ZDi Plus SH	6.5 Lakh*	null	2016	Dealer	76,000 kms	First Owner	Diesel	Manual	Mileage28.09 kmpl	Engine1248 CC	Max Power88.5	Seats5
Maruti Swift VDI BSIV	6.27 Lakh*	null	2016	Individual	20,000 kms	First Owner	Diesel	Manual	Mileage25.2 kmpl	Engine1248 CC	Max Power74 bh	Seats5
Mercedes-Benz C-Class	14.25 Lakh*	New Car (On-	2014	Dealer	65,000 kms	First Owner	Diesel	Automatic	Mileage19.27 kmpl	Engine2143 CC	Max Power170	Seats5
Maruti Swift VDI	4.25 Lakh*	null	2014	Dealer	62,200 kms	First Owner	Diesel	Manual	Mileage28.4 kmpl	Engine1248 CC	Max Power74 bh	Seats5
Toyota Innova 2.5 GX (D	6.05 Lakh*	null	2013	Individual	1,10,000 kms	First Owner	Diesel	Manual	Mileage12.99 kmpl	Engine2494 CC	Max Power100.	Seats8
Maruti Baleno Zeta 1.2	6 Lakh*	New Car (On-	2015	Individual	20,000 kms	First Owner	Petrol	Manual	Mileage21.4 kmpl	Engine1197 CC	Max Power83.1	Seats5
Maruti Swift Dzire VXI	5.75 Lakh*	New Car (On-	2016	Individual	40,000 kms	First Owner	Petrol	Manual	Mileage20.85 kmpl	Engine1197 CC	Max Power83.1	Seats5
Fiat Grande Punto 1.3 E	2.6 Lakh*	null	2012	Individual	97,000 kms	First Owner	Diesel	Manual	Mileage17.8 kmpl	Engine1248 CC	Max Power76 bh	Seats5
Volkswagen Vento 1.6	4.25 Lakh*	New Car (On-	2013	Dealer	47,000 kms	First Owner	Petrol	Manual	Mileage16.09 kmpl	Engine1598 CC	Max Power103.	Seats5
Maruti Alto K10 2010-20	2.3 Lakh*	null	2013	Dealer	25,000 kms	First Owner	Petrol	Manual	Mileage20.92 kmpl	Engine998 CC	Max Power67.1	Seats5
Hyundai Creta 1.6 CRDi	12.25 Lakh*	New Car (On-	2019	Individual	15,000 kms	First Owner	Diesel	Manual	Mileage19.67 kmpl	Engine1582 CC	Max Power126.	Seats5
Mahindra Xylo H9	3.75 Lakh*	null	2010	Dealer	56,823 kms	First Owner	Diesel	Manual	Mileage14.02 kmpl	Engine2179 CC	Max Power118.	Seats8
Honda City i VTEC VX	7.5 Lakh*	New Car (On-	2015	Individual	50,000 kms	First Owner	Petrol	Manual	Mileage17.4 kmpl	Engine1497 CC	Max Power117.	Seats5

Reason Why Accepted :-

- This Dataset mainly contains cars descriptions and models which perfectly fits into our model.
- Pricing of cars like Showroom price and On-Road price are given perfectly in each rows so that we can correlate.

CHAPTER 6
CAR PRICE PREDICTION MODEL

6.1 Flowchart :-

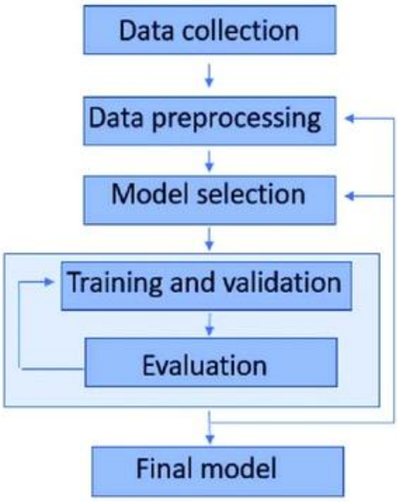

Figure 4 Flowchart

6.2 Datasets :-

full_name	selling_price	new_price	year	seller_type	km_driven	owner_type	fuel_type	transmission_t	mileage	engine	max_power	seats
Maruti Alto Std	1.2 Lakh*	null	2012	Individual	1,20,000 kms	First Owner	Petrol	Manual	Mileage19.7 kmpl	Engine796 CC	Max Power46.3	Seats5
Hyundai Grand i10 Asta	5.5 Lakh*	New Car (On-	2016	Individual	20,000 kms	First Owner	Petrol	Manual	Mileage18.9 kmpl	Engine1197 CC	Max Power82 bh	Seats5
Hyundai i20 Asta	2.15 Lakh*	null	2010	Individual	60,000 kms	First Owner	Petrol	Manual	Mileage17.0 kmpl	Engine1197 CC	Max Power80 bh	Seats5
Maruti Alto K10 2010-20	2.26 Lakh*	null	2012	Individual	37,000 kms	First Owner	Petrol	Manual	Mileage20.92 kmpl	Engine998 CC	Max Power67.1	Seats5
Ford Ecosport 2015-202	5.7 Lakh*	New Car (On-	2015	Dealer	30,000 kms	First Owner	Diesel	Manual	Mileage22.77 kmpl	Engine1498 CC	Max Power98.5	Seats5
Maruti Wagon R VXI BS	3.5 Lakh*	New Car (On-	2013	Individual	35,000 kms	First Owner	Petrol	Manual	Mileage18.9 kmpl	Engine998 CC	Max Power67.1	Seats5
Hyundai i10 Sportz 1.2	3.15 Lakh*	New Car (On-	2013	Dealer	40,000 kms	First Owner	Petrol	Manual	Mileage20.36 kmpl	Engine1197 CC	Max Power78.9	Seats5
Maruti Wagon R VXI	4.1 Lakh*	New Car (On-	2018	Dealer	17,512 kms	First Owner	Petrol	Manual	Mileage20.51 kmpl	Engine998 CC	Max Power67.0	Seats5
Hyundai Venue SX Plus	10.5 Lakh*	New Car (On-	2019	Individual	20,000 kms	First Owner	Petrol	Automatic	Mileage18.15 kmpl	Engine998 CC	Max Power118.	Seats5
Mahindra TUV 300 T6	5.75 Lakh*	null	2017	Dealer	70,000 kms	First Owner	Diesel	Manual	Mileage18.49 kmpl	Engine1493 CC	Max Power100 b	Seats7
Tata Indigo LX (TDI) BS-	3.05 Lakh*	null	2015	Individual	50,000 kms	First Owner	Diesel	Manual	Mileage19.09 kmpl	Engine1405 CC	Max Power69.01	Seats5
Renault Captur Platine	11.5 Lakh*	null	2019	Individual	18,000 kms	First Owner	Diesel	Manual	Mileage20.37 kmpl	Engine1461 CC	Max Power108.	Seats5
Maruti Swift VXI with A	5.11 Lakh*	New Car (On-	2017	Dealer	28,321 kms	First Owner	Petrol	Manual	Mileage16.6 kmpl	Engine1197 CC	Max Power85 bh	Seats5
Nissan Micra XL CVT	4.1 Lakh*	null	2016	Dealer	27,000 kms	First Owner	Petrol	Automatic	Mileage19.34 kmpl	Engine1198 CC	Max Power76 bh	Seats5
Hyundai Verna 1.6 SX	4.25 Lakh*	New Car (On-	2013	Dealer	65,278 kms	First Owner	Diesel	Manual	Mileage22.32 kmpl	Engine1582 CC	Max Power126.	Seats5
Renault Duster 110PS D	7.5 Lakh*	null	2016	Individual	50,000 kms	First Owner	Diesel	Manual	Mileage19.64 kmpl	Engine1461 CC	Max Power108.	Seats5
Mini Cooper Cooper S	32.5 Lakh*	null	2017	Dealer	6,000 kms	First Owner	Petrol	Automatic	Mileage14.41 kmpl	Engine1998 CC	Max Power189.	Seats5
Maruti Ciaz ZDi Plus SH	6.5 Lakh*	null	2016	Dealer	76,000 kms	First Owner	Diesel	Manual	Mileage28.09 kmpl	Engine1248 CC	Max Power88.5	Seats5
Maruti Swift VDI BSIV	6.27 Lakh*	null	2016	Individual	20,000 kms	First Owner	Diesel	Manual	Mileage25.2 kmpl	Engine1248 CC	Max Power74 bh	Seats5
Mercedes-Benz C-Class	14.25 Lakh*	New Car (On-	2014	Dealer	65,000 kms	First Owner	Diesel	Automatic	Mileage19.27 kmpl	Engine2143 CC	Max Power170 b	Seats5
Maruti Swift VDI	4.25 Lakh*	null	2014	Dealer	62,200 kms	First Owner	Diesel	Manual	Mileage28.4 kmpl	Engine1248 CC	Max Power74 bh	Seats5
Toyota Innova 2.5 GX (D	6.05 Lakh*	null	2013	Individual	1,10,000 kms	First Owner	Diesel	Manual	Mileage12.99 kmpl	Engine2494 CC	Max Power100.	Seats8
Maruti Baleno Zeta 1.2	6 Lakh*	New Car (On-	2015	Individual	20,000 kms	First Owner	Petrol	Manual	Mileage21.4 kmpl	Engine1197 CC	Max Power83.1	Seats5
Maruti Swift Dzire VXI	5.75 Lakh*	New Car (On-	2016	Individual	40,000 kms	First Owner	Petrol	Manual	Mileage20.85 kmpl	Engine1197 CC	Max Power83.1	Seats5
Fiat Grande Punto 1.3 E	2.6 Lakh*	null	2012	Individual	97,000 kms	First Owner	Diesel	Manual	Mileage17.8 kmpl	Engine1248 CC	Max Power76 bh	Seats5
Volkswagen Vento 1.6 (	4.25 Lakh*	New Car (On-	2013	Dealer	47,000 kms	First Owner	Petrol	Manual	Mileage16.09 kmpl	Engine1598 CC	Max Power103.	Seats5
Maruti Alto K10 2010-20	2.3 Lakh*	null	2013	Dealer	25,000 kms	First Owner	Petrol	Manual	Mileage20.92 kmpl	Engine998 CC	Max Power67.1	Seats5
Hyundai Creta 1.6 CRDi	12.25 Lakh*	New Car (On-	2019	Individual	15,000 kms	First Owner	Diesel	Manual	Mileage19.67 kmpl	Engine1582 CC	Max Power126.	Seats5
Mahindra Xylo H9	3.75 Lakh*	null	2010	Dealer	56,823 kms	First Owner	Diesel	Manual	Mileage14.02 kmpl	Engine2179 CC	Max Power118.	Seats8
Honda City i VTEC VX	7.5 Lakh*	New Car (On-	2015	Individual	50,000 kms	First Owner	Petrol	Manual	Mileage17.4 kmpl	Engine1497 CC	Max Power117.	Seats5

Proposed dataset information: -

full_name
selling_price
new_price
Year
seller_type
km_driven
owner_type
fuel_type
transmission_type
Mileage engine
Mileage engine
seatstransmission_type

No of Rows.	19974
No of Cols.	12

6.3 Preprocessing Steps :-

A. Heatmap :-

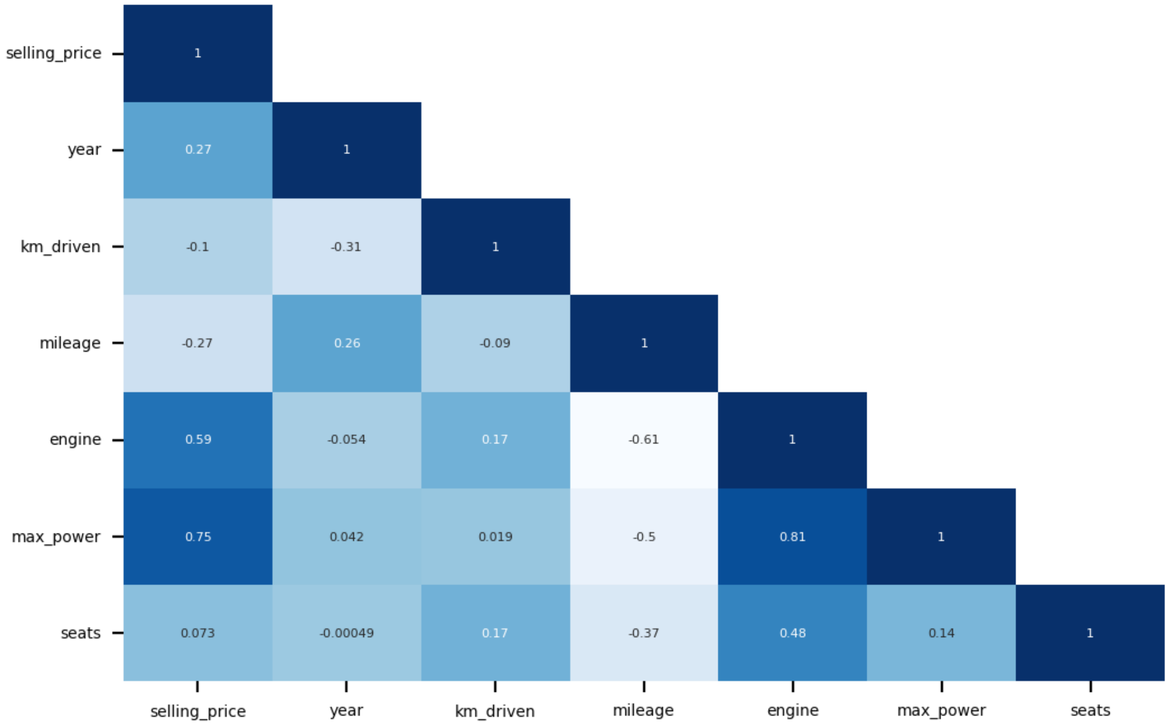

Figure 5 Heatmap

B. Relational Heatmap :-

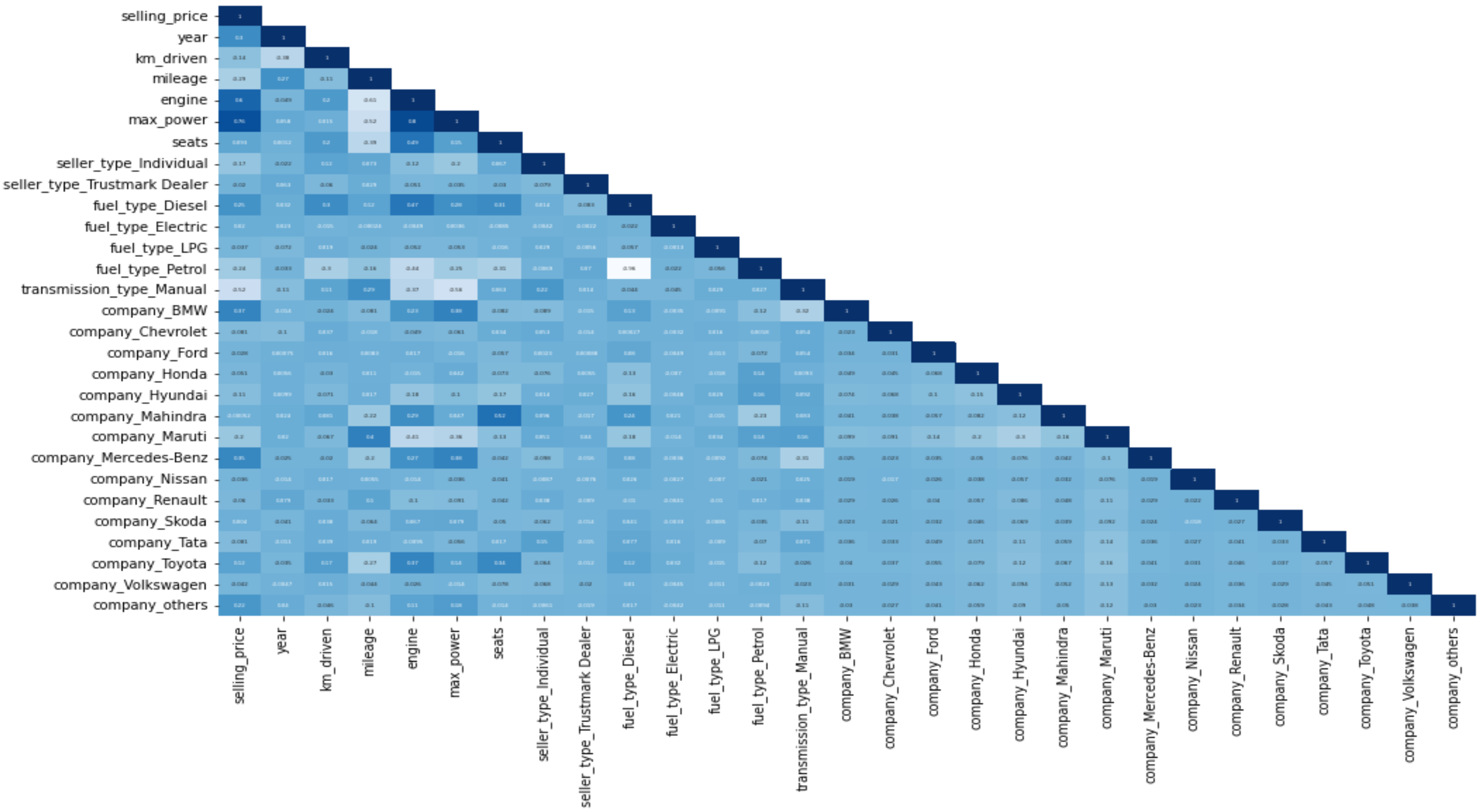

Figure 6 Relational Heatmap

6.4 Methodology :-

We utilized several classic and state-of-the-art methods, including ensemble learning techniques, with a 90% - 10% split for the training and test data. To reduce the time required for training, we used 500 thousand examples from our dataset. Linear Regression, Random Forest and Gradient Boost were our baseline methods. For most of the model implementations, the open-source Scikit-Learn package was used.

Linear Regression :-
Linear Regression was chosen as the first model due to its simplicity and comparatively small training time. The features, without any feature mapping, were used directly as the feature vectors. No regularization was used since the results clearly showed low variance.

Explanation :-

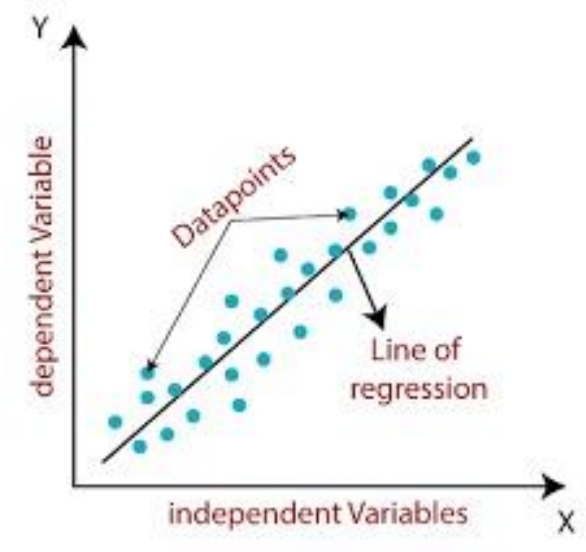

Figure 7 Linear Regression

Equation :-

$$\overset{\text{Dependent Variable}}{Y_i} = \underbrace{\overset{\text{Population Y intercept}}{\beta_0} + \overset{\text{Population Slope Coefficient}}{\beta_1}\overset{\text{Independent Variable}}{X_i}}_{\text{Linear component}} + \underbrace{\overset{\text{Random Error term}}{\varepsilon_i}}_{\text{Random Error component}}$$

Random Forest : -
Forest is an ensemble learning based regression model. It uses a model called decision tree, specifically as the name suggests, multiple decision trees to generate the ensemble model which collectively produces a prediction. The benefit of this model is that the trees are produced in parallel and are relatively uncorrelated, thus producing good results as each tree is not prone to individual errors of other trees. This uncorrelated behavior is partly ensured by the use of Bootstrap Aggregation or bagging providing the randomness required to produce robust and uncorrelated trees. This model was hence chosen to account for the large number of features in the dataset and compare a bagging technique with the following gradient boosting methods.

Explanation :-
When using the Random Forest Algorithm to solve regression problems, you are using the mean squared error (MSE) to how your data branches from each node.

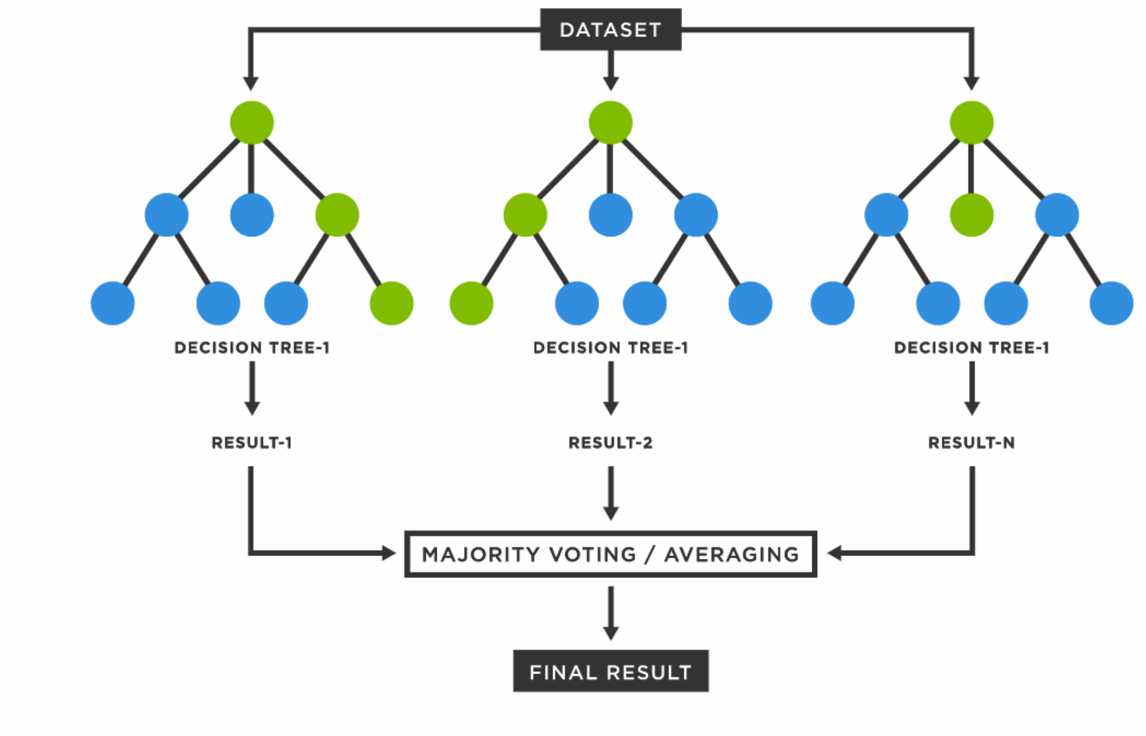

Figure 8 Random Forest

Equation :-

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (fi - yi)^2$$

Where N is the number of data points,
fi is the value returned by the model and
yi is the actual value for data point i.

6.5 Result :-

A. Accuracy of Random Forest.

```
[ ] model_3 = RandomForestRegressor()
    error, score = do_prediction(model_3)

    print('Random Forest Regressor MAE: {}'.format(round(error,2)))
    print('Cross validation score: {}'.format(round(score,2)))

    Random Forest Regressor MAE: 102617.79
    Cross validation score: 0.92
```

B. Accuracy of Linear Regression.

```
[ ] model_1 = LinearRegression()
    error, score = do_prediction(model_1)

    print('Linear Regression MAE: {}'.format(round(error,2)))
    print('Cross validation score: {}'.format(round(score,2)))

    Linear Regression MAE: 235187.72
    Cross validation score: 0.71
```

Accuracy of Linear Regression.	**0.71**
Accuracy of Random Forest.	**0.92**

Table 2: Accuracy Table

Q.Why the Accuracy of Random Forest is Greater than Linear Regression?

Linear Models have very few parameters, Random Forests a lot more. That means that Random Forests will overfit more easily than a Linear Regression.

CHAPTER 7
CONCLUSION

7.1 Conclusion: -

After this project work and research our conclusion is that we can identify the best prices by using the trained machine learning model. But for some old things or products it is very difficult to identify their prices but by extending the databases we can reduce the difficulty. Also, when number of inputs is more in the frame then it’s very tough to getting best prices of them. But we learnt a lot from it. Our research still is going on it. Another thing is that we need a lot of positive and true values in dataset for training the machine learning Model.

Bibliography :-

- **Cars24 :-** Sell Your Car Safely with Cars24. Book Home Inspection or Visit a Cars24 Store Near You. Sell Your Car at the Comfort of Your Home.

  Website: **https://www.scars24.com/**

- **Car Dekho :-** Thinking of buying a car? At CarDekho.com, buy new and used cars, search by filter and preferences, compare cars, read latest news and updates.

  Website: **https://www.cardekho.com**

- **Spinny: -** Spinny is the most trusted way of buying and selling used cars. Choose from over 5000 fully inspected second-hand car models.

  Website: **https://www.spinny.com/**

References :-

[1] TEM Journal. Volume 8, Issue 1, Pages 113-118, ISSN 2217-8309, DOI:10.18421/TEM81-16, February 2019.
TEM Journal – Volume 8 / Number 1 / 2019. Car Price Prediction using Machine Learning Techniques Enis Gegic, Becir Isakovic, Dino Keco.

[2] Geurts P. (2009) Bias vs Variance Decomposition for Regression and Classification. In: Maimon O., Rokach L. (eds) Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA

[3] Robert T. (1996) Regression Shrinkage and Selection Via the Lasso. In: Journal of the Royal Statistical Society: Series B Methodological) Volume 58, Issue 1

[4] Hastie, Trevor, and Daryl Region. Shrinking trees. AT & T Bell Laboratories, 1990. [5] Kim, Tae Kyun. "Understanding one-way ANOVA using conceptual figures." Korean journal of anesthesiology 70.1 (2017).

[5] Kim, Tae Kyun. "Understanding one-way ANOVA using conceptual figures." Korean journal of anesthesiology 70.1 (2017): 22.

Figure 1 : https://images.app.goo.gl/ss2iRrEELe1P7ohr6
Figure 2 : https://images.app.goo.gl/SdaZRtYMzPF4zvXdA
Figure 4 : https://images.app.goo.gl/SdaZRtYMzPF4zvXdA
Figure 7 : https://images.app.goo.gl/yVnVz3zuxEWwtkau9 .
Figure 8 : https://www.tibco.com/reference-center/what-is-a-randoforest