e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

USED OLD CAR SELLING PRICE PREDICTION USING MACHINE LEARNING TECHNIQUES

Soham S. Ghume[*1], Pranav V. Danavale[*2], Tanvi N. Hagawane[*3], Nikhil V. Shelar[*4], Prathamesh S. Waghmode[*5], Prof. Komal B. Babar[*6]

[*1]Computer Engineering, Zeal Polytechnic, Pune, Maharashtra, India.

[*2,3,4,5,6]Zeal Polytechnic, Pune, Maharashtra, India.

Abstract: India's automobile market is huge, with many people selling their used cars to new buyers as second or third owners. Platforms like Cars24, CarDekho, and OLX help connect sellers and buyers, but figuring out the right price for a used car is challenging. Machine Learning (ML) can help by using past car sales data to predict a fair price. In this case, Supervised Learning techniques, including Random Forest and Extra Tree Regression algorithms, were applied using the Scikit-Learn library. These algorithms proved to be highly accurate in predicting prices, regardless of the dataset size.

General Terms: Machine Learning, Artificial Intelligence, Data Mining.

Keywords: Used Car Selling Price Prediction Using. Logistic Regression, Lasso Regression, Heat Map, Angular Framwork.

I. INTRODUCTION

Predicting the price of a used car, which doesn't have the standard pricing of a new vehicle directly from the factory, is indeed a multi-faceted and challenging task. The complexity arises from numerous factors like the car's age, model, condition, mileage, and even external factors like fluctuating fuel prices, which have increasingly become a concern in today's economic landscape. As the resale market for used cars grows, so does the need for accurate pricing mechanisms, because buyers and sellers are more reliant on these estimates to make informed decisions. In many cases, legal agreements between the two parties hinge on the estimated price of the car, which is why it's critical for these estimates to be as accurate as possible.

Accurate price prediction becomes even more significant as buyers are looking for a fair valuation that considers every aspect of the car, while sellers want to ensure they get the best possible price in a highly competitive market. Using data-driven methods like machine learning to predict prices with the highest precision not only builds trust in the process but also creates a more efficient marketplace where transactions can occur with greater confidence. In essence, the ability to predict used car prices accurately benefits all parties involved by ensuring that both the seller and buyer have a fair basis for their transactions, thus reducing disputes and increasing satisfaction in the buying and selling experience ccurate price prediction becomes even more significant as buyers are looking for a fair valuation that considers every aspect of the car, while sellers want to ensure they get the best possible price in a highly competitive market. Using data-driven methods like machine learning to predict prices with the highest precision not only builds trust in the process but also creates a more efficient marketplace where transactions can occur with greater confidence. In essence, the ability to predict used car prices accurately benefits all parties involved by ensuring that both the seller and buyer have a fair basis for their transactions, thus reducing disputes and increasing satisfaction in the buying and selling experience

Predicting the price of a car that is not coming directly from the factory is a complex and critical task, especially as the demand for used cars in the resale market continues to grow. This challenge has been further compounded by rising fuel prices, which add another layer of complexity for used car sellers. Both individuals and organizations often prefer to conduct transactions with legal agreements based on an estimated price, making it crucial to find an accurate and fair price estimation. Achieving a high degree of precision in predicting the actual price of a used car would not only benefit sellers but also give buyers confidence in the fairness of the transaction. By leveraging machine learning algorithms to predict prices with greater accuracy, it becomes possible to streamline the buying and selling process, ensuring both parties have a reliable basis for negotiation and decision-making so we had used the various supervised learning algorithms such as

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

1. LINER REGRESSION MODEL

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

$$Y = B0 + B1*X$$

When we move on higher model and discuss on com plexity 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

II. LITERATURE SURVEY

The price of a pre-owned car depends on various factors such as the model year, mileage, overall condition, and the equipment or features it possesses. With so many variables influencing the price, it becomes difficult to estimate it accurately using traditional rule-based algorithms. Instead, a more effective strategy is to employ inductive learning methods, which allow the system to learn from a dataset and predict the price based on the patterns within that data. This makes machine learning particularly suitable for this application.

In existing literature, for example, a study on the "Application of ML techniques to predict the price of pre-owned cars in Maharashtra" demonstrates a restricted field analysis within that region. However, the scope of this research extends far beyond simple geographic limitations. Another referenced paper discusses the use of Support Vector Machines (SVM) to achieve accurate price predictions. A third study explores the use of Big Data and artificial neural networks (ANN), which takes into account more complex and variable data for vehicles. In contrast, other papers focus primarily on traditional methods like Linear Regression, Ridge Regression, and Lasso Regression.

This research aims to broaden the comparative analysis by incorporating Random Forest algorithms alongside these other machine learning techniques. By extending the study to include a wider range of algorithms, the paper will offer a more comprehensive understanding of which methods provide the most accurate and reliable predictions for pre-owned car prices, particularly within the context of Maharashtra's used car market.

2.1 HARDWARE /SOFTWARE REQUIREMENTS:

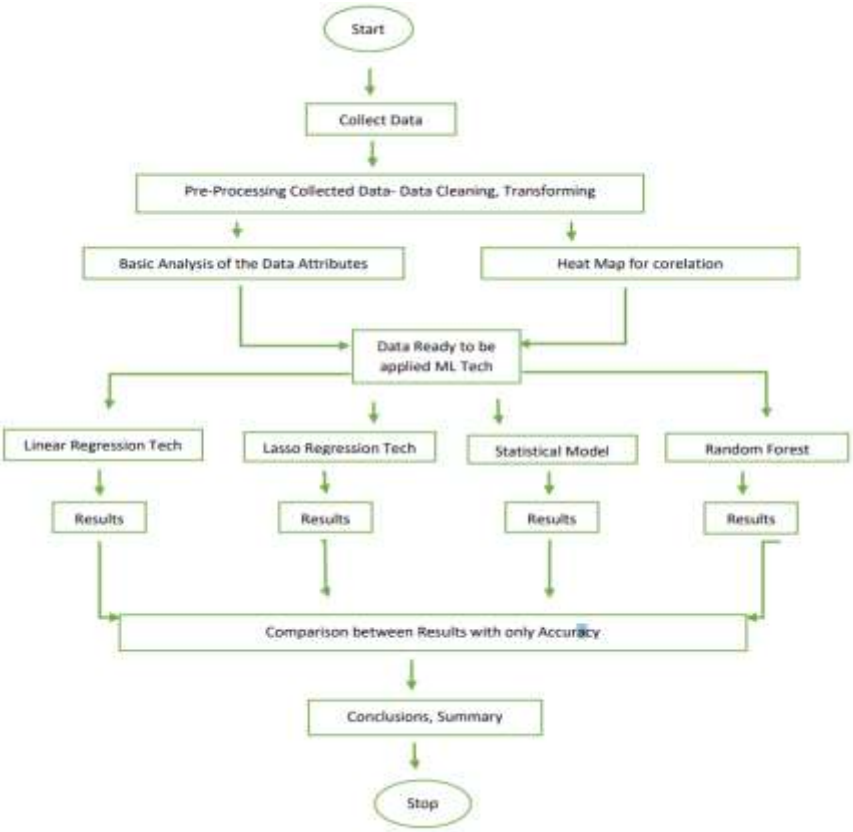
Hardware requirements Operating system- Windows 7,8,10 Processor- dual core 2.4 GHz (i5 or i7 series Intel processor or equivalent AMD) RAM-4GB Software Requirements : Google Colab, Python Pycharm PIP 2.7 Jupyter Notebook Chrome.

III. METHODOLOGIES

3.1 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 3 people have been contributed as follows. Mr Soham & Pranav, Prathamesh has contribud with Linear and Lasso regression techniques which consumed one additional month where the results were not much satisfied hence we got into a decision Model with all other co-authors where Mrs Tanvi & Nikhil has been contributed than linear regression results.

STEP WISE PROCESS FOR PURPOSED MODEL:

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com



3.2 COLLECTED DATA SAMPLE:

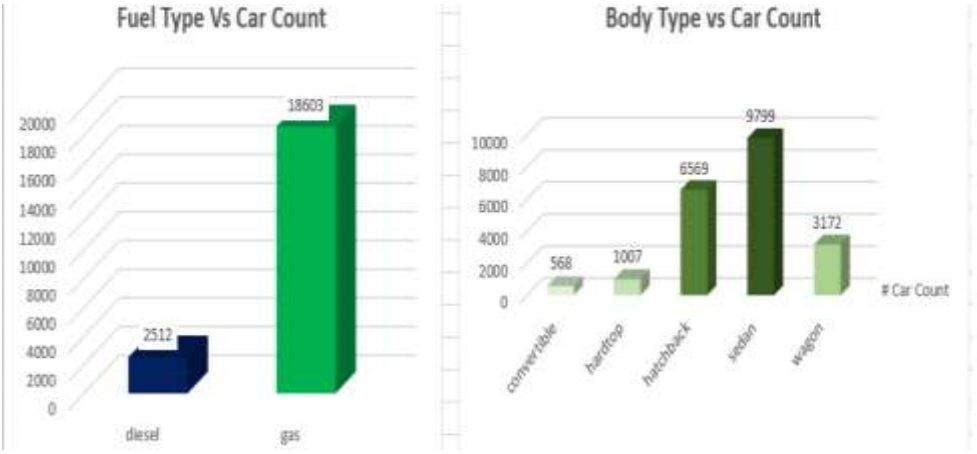
Data have been collected over 20K Indian data samples where we have collected with various open source types and classified as below variables

Sr.No	Variable Name	Description
1	Car Name	Name Of The Car
2	Year	Year Of Ve-hical
3	Selling Price	Selling Price of the car
4	Km Driven	Km Driven of the car
5	Fuel	Fuel of the car
6	Seller Type	Seller Type of the car
7	Owner	Owner of the car
8	Mileage	Mileage of the car
9	Engine	Engine of the car
10	Max Power	Max Power the car
11	Seats	Seat of the car

3.3 BASIC DATA ANALYSIS & VISUALS

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

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com



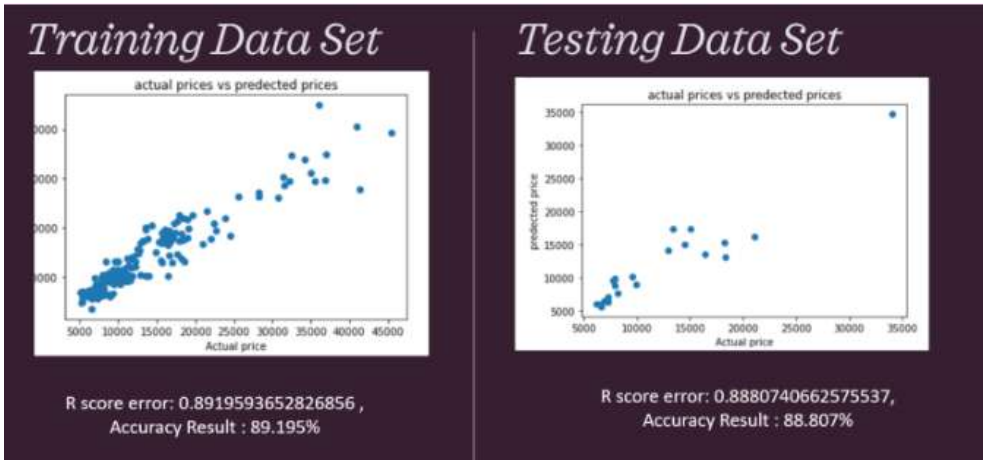
This has been performed to understand very basic feasibility on the relation towards targeted price vs Fuel Type and Body Type and then we decided to follow heat map to find the better relation ship between target (Price) and dependent variables (all other)

IV. OUTPUTS WITH ACCURACY RESULTS

4.1 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

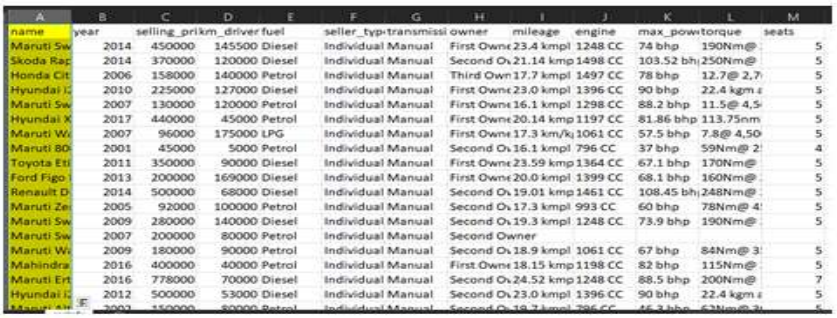


4.2 STATISTICAL MODEL:

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.

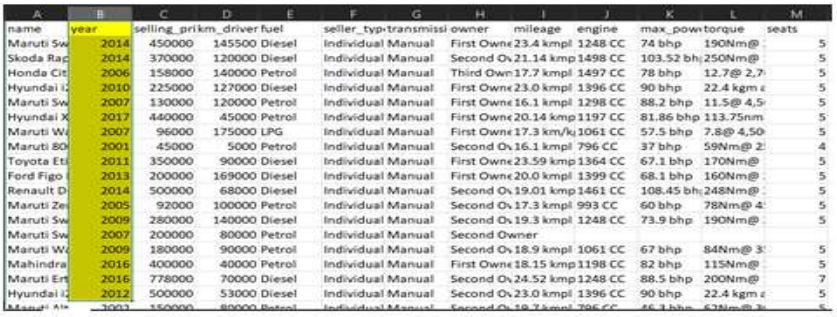
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

4.3 REQUIRED DATA COLLETION:



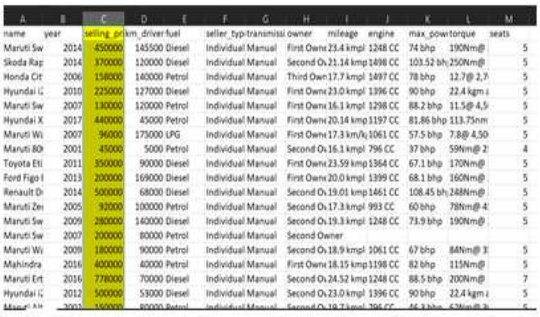
name	year	selling_price	km_driven	fuel	seller_type	transmission	owner	mileage	engine	max_power	torque	seats
Maruti Sw	2014	450000	145500	Diesel	Individual	Manual	First Owne	23.4 kmpl	1248 CC	74 bhp	190Nm@	5
Skoda Rap	2014	370000	120000	Diesel	Individual	Manual	Second Ow	21.14 kmp	1498 CC	103.52 bh	250Nm@	5
Honda Cit	2006	158000	140000	Petrol	Individual	Manual	Third Own	17.7 kmpl	1497 CC	78 bhp	12.7@ 2,7	5
Hyundai i2	2010	225000	127000	Diesel	Individual	Manual	First Owne	23.0 kmpl	1396 CC	90 bhp	22.4 kgm a	5
Maruti Sw	2007	130000	120000	Petrol	Individual	Manual	First Owne	16.1 kmpl	1298 CC	88.2 bhp	11.5@ 4,5	5
Hyundai X	2017	440000	45000	Petrol	Individual	Manual	First Owne	20.14 kmp	1197 CC	81.86 bhp	113.75nm	5
Maruti Wa	2007	96000	175000	LPG	Individual	Manual	First Owne	17.3 km/kg	1061 CC	57.5 bhp	7.8@ 4,50	5
Maruti 80	2001	45000	5000	Petrol	Individual	Manual	Second Ow	16.1 kmpl	796 CC	37 bhp	59Nm@ 2	4
Toyota Eti	2011	350000	90000	Diesel	Individual	Manual	First Owne	23.59 kmp	1364 CC	67.1 bhp	170Nm@	5
Ford Figo I	2013	200000	169000	Diesel	Individual	Manual	First Owne	20.0 kmpl	1399 CC	68.1 bhp	160Nm@	5
Renault D	2014	500000	68000	Diesel	Individual	Manual	Second Ow	19.01 kmp	1461 CC	108.45 bh	248Nm@	5
Maruti Ze	2005	92000	100000	Petrol	Individual	Manual	Second Ow	17.3 kmpl	993 CC	60 bhp	78Nm@ 4	5
Maruti Sw	2009	280000	140000	Diesel	Individual	Manual	Second Ow	19.3 kmpl	1248 CC	73.9 bhp	190Nm@	5
Maruti Sw	2007	200000	80000	Petrol	Individual	Manual	Second Owner					
Maruti Wa	2009	180000	90000	Petrol	Individual	Manual	Second Ow	18.9 kmpl	1061 CC	67 bhp	84Nm@ 3	5
Mahindra	2016	400000	40000	Petrol	Individual	Manual	First Owne	18.15 kmp	1198 CC	82 bhp	115Nm@	5
Maruti Ert	2016	778000	70000	Diesel	Individual	Manual	Second Ow	24.52 kmp	1248 CC	88.5 bhp	200Nm@	7
Hyundai i2	2012	500000	53000	Diesel	Individual	Manual	Second Ow	23.0 kmpl	1396 CC	90 bhp	22.4 kgm a	5

Next Attribute to be removed as " Cars name" into number

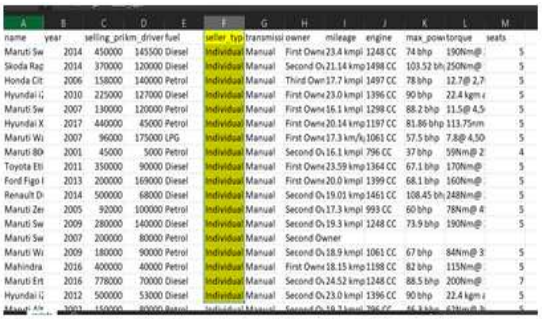


name	year	selling_price	km_driven	fuel	seller_type	transmission	owner	mileage	engine	max_power	torque	seats
Maruti Sw	2014	450000	145500	Diesel	Individual	Manual	First Owne	23.4 kmpl	1248 CC	74 bhp	190Nm@	5
Skoda Rap	2014	370000	120000	Diesel	Individual	Manual	Second Ow	21.14 kmp	1498 CC	103.52 bh	250Nm@	5
Honda Cit	2006	158000	140000	Petrol	Individual	Manual	Third Own	17.7 kmpl	1497 CC	78 bhp	12.7@ 2,7	5
Hyundai i2	2010	225000	127000	Diesel	Individual	Manual	First Owne	23.0 kmpl	1396 CC	90 bhp	22.4 kgm a	5
Maruti Sw	2007	130000	120000	Petrol	Individual	Manual	First Owne	16.1 kmpl	1298 CC	88.2 bhp	11.5@ 4,5	5
Hyundai X	2017	440000	45000	Petrol	Individual	Manual	First Owne	20.14 kmp	1197 CC	81.86 bhp	113.75nm	5
Maruti Wa	2007	96000	175000	LPG	Individual	Manual	First Owne	17.3 km/kg	1061 CC	57.5 bhp	7.8@ 4,50	5
Maruti 80	2001	45000	5000	Petrol	Individual	Manual	Second Ow	16.1 kmpl	796 CC	37 bhp	59Nm@ 2	4
Toyota Eti	2011	350000	90000	Diesel	Individual	Manual	First Owne	23.59 kmp	1364 CC	67.1 bhp	170Nm@	5
Ford Figo I	2013	200000	169000	Diesel	Individual	Manual	First Owne	20.0 kmpl	1399 CC	68.1 bhp	160Nm@	5
Renault D	2014	500000	68000	Diesel	Individual	Manual	Second Ow	19.01 kmp	1461 CC	108.45 bh	248Nm@	5
Maruti Ze	2005	92000	100000	Petrol	Individual	Manual	Second Ow	17.3 kmpl	993 CC	60 bhp	78Nm@ 4	5
Maruti Sw	2009	280000	140000	Diesel	Individual	Manual	Second Ow	19.3 kmpl	1248 CC	73.9 bhp	190Nm@	5
Maruti Sw	2007	200000	80000	Petrol	Individual	Manual	Second Owner					
Maruti Wa	2009	180000	90000	Petrol	Individual	Manual	Second Ow	18.9 kmpl	1061 CC	67 bhp	84Nm@ 3	5
Mahindra	2016	400000	40000	Petrol	Individual	Manual	First Owne	18.15 kmp	1198 CC	82 bhp	115Nm@	5
Maruti Ert	2016	778000	70000	Diesel	Individual	Manual	Second Ow	24.52 kmp	1248 CC	88.5 bhp	200Nm@	7
Hyundai i2	2012	500000	53000	Diesel	Individual	Manual	Second Ow	23.0 kmpl	1396 CC	90 bhp	22.4 kgm a	5

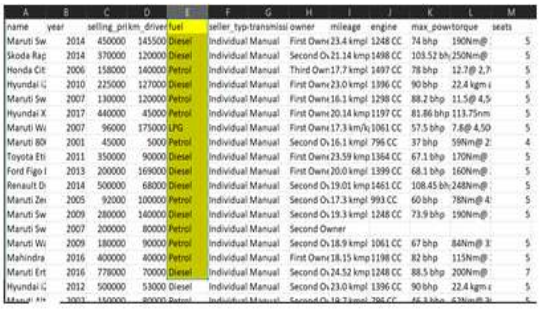
Next"Year"



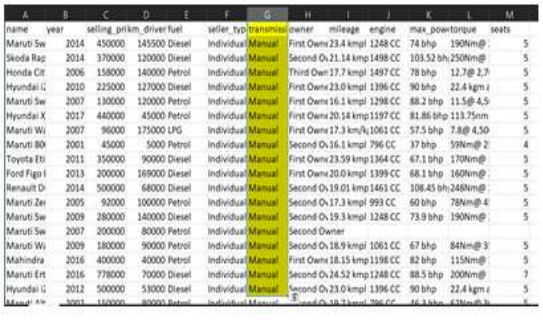
Next"selling Price"



Next"Selling Type"

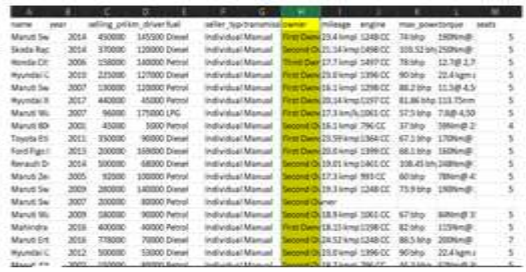


Next"Fuel"

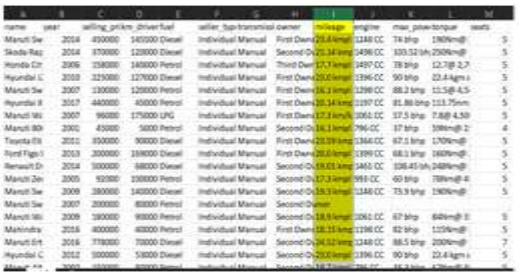


Next"Transmission"

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com



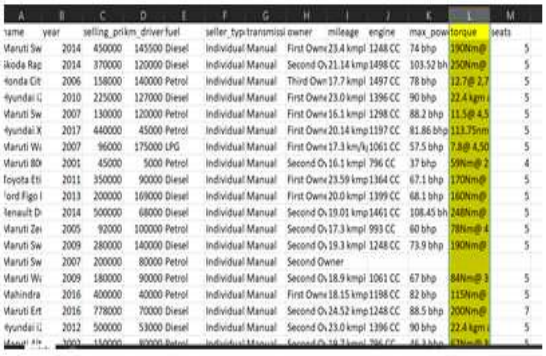
Next"Owner"



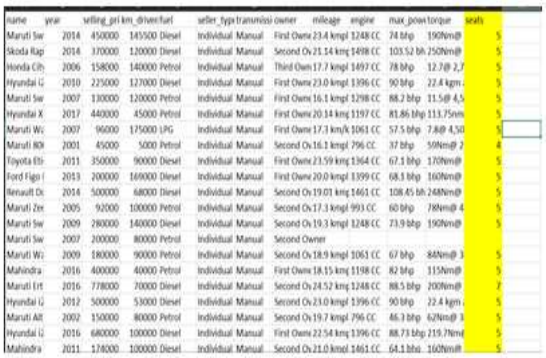
Next"Mileage"

Next"Engine"

Next"Max Power"



Next"Torque"



Next"Seat"

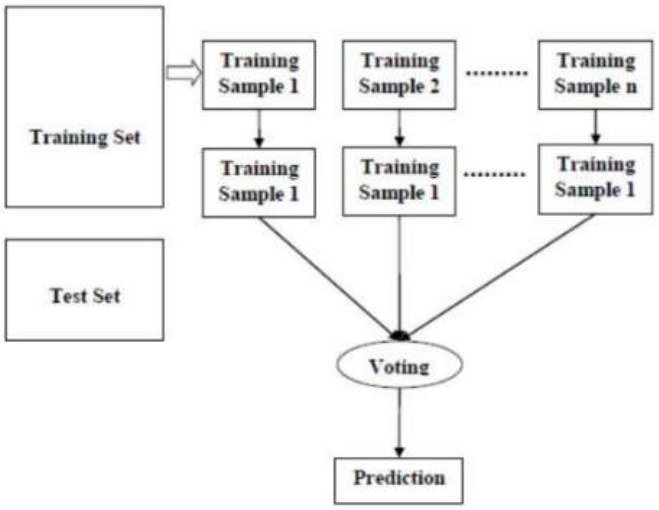
4.4 RANDOM FOREST IMPLEMENTATIONS:

It is an ensemble method which is better than a single decision tree because it reduces the overfitting 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

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

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
- R score error: 0.9143581729539816 Accuracy: 91.435%

4.5 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

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.

V. 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.

VI. REFERENCES

[1] "Used Car Price Predication Using Machine Learning Techniques"in Google Scholar on 2021 by Mrs Shyamali Das, Mr Ananta Laha, Mr Alok Jena, Ms Priyadarshini Samal.

[2] "Car's Selling Price Prediction using Random Forest Machine Learning Algorithm." Abhishek Pandey, Vanshika Rastogi, Sanika Singh.