Prediction of Millage Per Gallon with ML and using User Interface

K.Venkateswara Rao[1], Bharathi S[2], Nasreen Sulthana Sk[3], Bhargavi S[4], Sri Venkata AvinashT[5]
[1]Professor, Department of CSE, Vignan's Lara Institute of Technology & Science, Vadlamudi, Guntur, AP, India
[2,3,4,5]IV CSE, Department of CSE, Vignan's Lara Institute of Technology & Science, Vadlamudi, Guntur, AP, India

ABSTRACT:

Now a days there is a huge competition in the automobile industries and every leading company want to provide their customers with the efficient products which meets their requirements. The basic requirement for any automobile starts with MPG, Price, Model etc. Further in the manufacturing of the vehicle or the product one need to check the MPG and it is mostly done manually. However, in this paper a model is implemented with machine learning approach which is used to calculate the fuel efficiency automatically by considering the basic parameters. As all the parameters are not going to affect the final outcome some selected parameters are used like displacement, number of cylinders, horsepower, weight, acceleration etc. The MPG and also termed as fuel efficiency will be calculated using XGBOOST with GridSearchCV and linear regression algorithms and improved accuracy up to 90%.

KEYWORDS: XGBOOST with GridSearchCV, MPG(mileage per gallon), fuel efficiency, Linear Regression.

INTRODUCTION:

According to the 19-20 survey researches of fuel economy in India 3.2 million light-duty vehicles (LDV) have been sold out because 57% of people are showing interest in buying cars. Because of this Industries are competitive and introducing new models with new functionalities for every month along with that fuel consumption of cars enhancing on an average of 6% on every year from last few years. Comparing the both reasons industries need to develop the vehicle with best features which consumes less fuel.

Estimating of MPG for every new model come in to the market takes lot of resources like fuel, time and other testing devices etc. Sometimes we may not get best result after completing the whole model because the features developed may not be compatible with one another. For example, the weight of the car may lead to cause more fuel consumption or cylinders may affect. Therefore, implementing the model by observing the content based on corresponding output will built a perfect model other than implementing without knowing the results.

As discussed above there are a lot of factors going to affect the Fuel efficiency which is nothing but ratio of distance travelling per unit of fuel consumption. The mentioned parameters may not be accessible and some are difficult for collection. Considering these reasons we are calculating the fuel efficiency of cars with the available data. XGBOOST with GridsearchCV is suitable to know the main factors which are a effecting the MPG. A database is added in this project for future information gathering and in order to get accurate results as the considered dataset is very small.

Literature Survey:

Shirbhayye, V et al. [1] recommended Linear Regression using machine learning(ML) to calculate millage per gallon by pre-processing the data. Despite the fact that a few predictions are a long way far from the real cost, similarly inspection of the dataset leads me to accept as true with that some of the MPG values are wildly erroneous.

Meng, J et al. [2] We use the information mining principle to assemble a BP neural network model to expect MPG.Completely based on the non - linear relationships of the total factors provided, taking into account the flaw in using crucial variables at the same time. There is a multi-layer community in the BP neural community that does load education using quasi discrete abilities.

Karpate, Y et al[3] Proposed simulation method which fashions the fuel efficiency of the present fleet of HDVs and compares it with the world's excellent practices. The model evaluates fuel economy with technological information from the existing Heavy Duty Vehicles fleet and mandated usage intervals.

Yin, X. et al [4]regarded the mutual information index(MII) and the enlightening automotive database in which it is handled (MII)is hired to discover a set of characteristics that significantly have an effect on gas performance. Fuel mileage forecasting patterns are also built using five remarkable device learning strategies.

Shalini, L et al[5]recommended two techniques one to be able to be implemented via system studying strategies like linear regression, one another in utilizing the optimization techniques such as stochastic gradient descent and gradient descent.

Yao, Y et al[6] installed On-board diagnostic device (OBD) to extract driving conduct Back propagation (BP) neural networks, assist vector regression (SVR), and random forest were used to calculate fuel usage using facts and statistics on fuel intake.

Wickramanayake, S. et al [7]Used machine learning algorithms like random forest, gradient boosting to calculate fuel intake by means of gathering the information of a bus through GPS trackers.

Rusiman, M. et al [8]On this examine, the gasoline consumption of various automobiles in miles per gallon (MPG) with 8 unbiased variables were anticipated the usage of the OFCRM fashions. Synthetic information is used to evaluate the OFCRM algorithms.

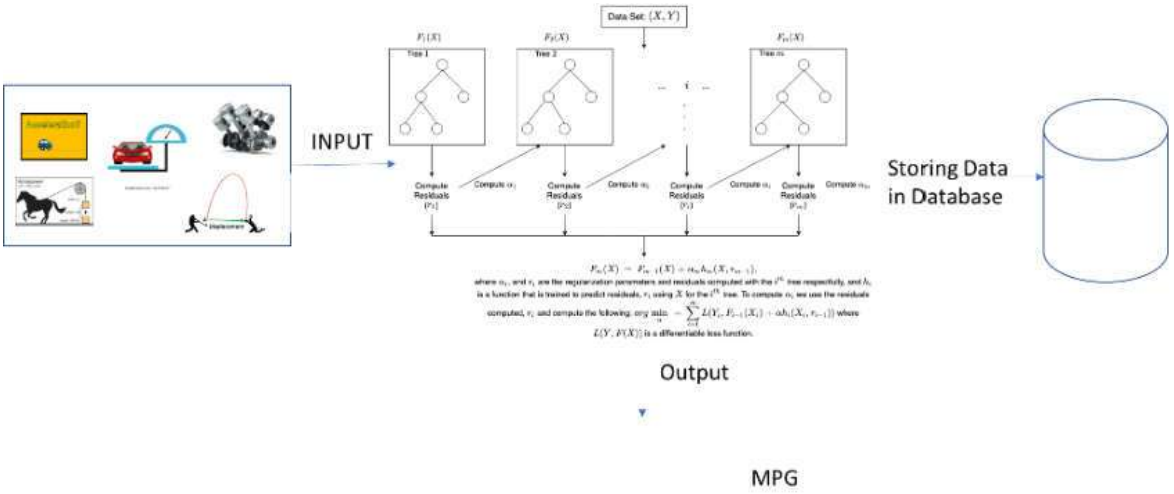
Model Architecture:

The proposed model starts working from the server and then predict the mpg using the values given through the server and background machine learning algorithms. The system will store the finished product. These below main functions are conducted by the framework.

The first part where system starts its execution is webpage. It supports HTML and CSS for user interface. A webpage will be visible bypasting the URL provided by flask frame work.

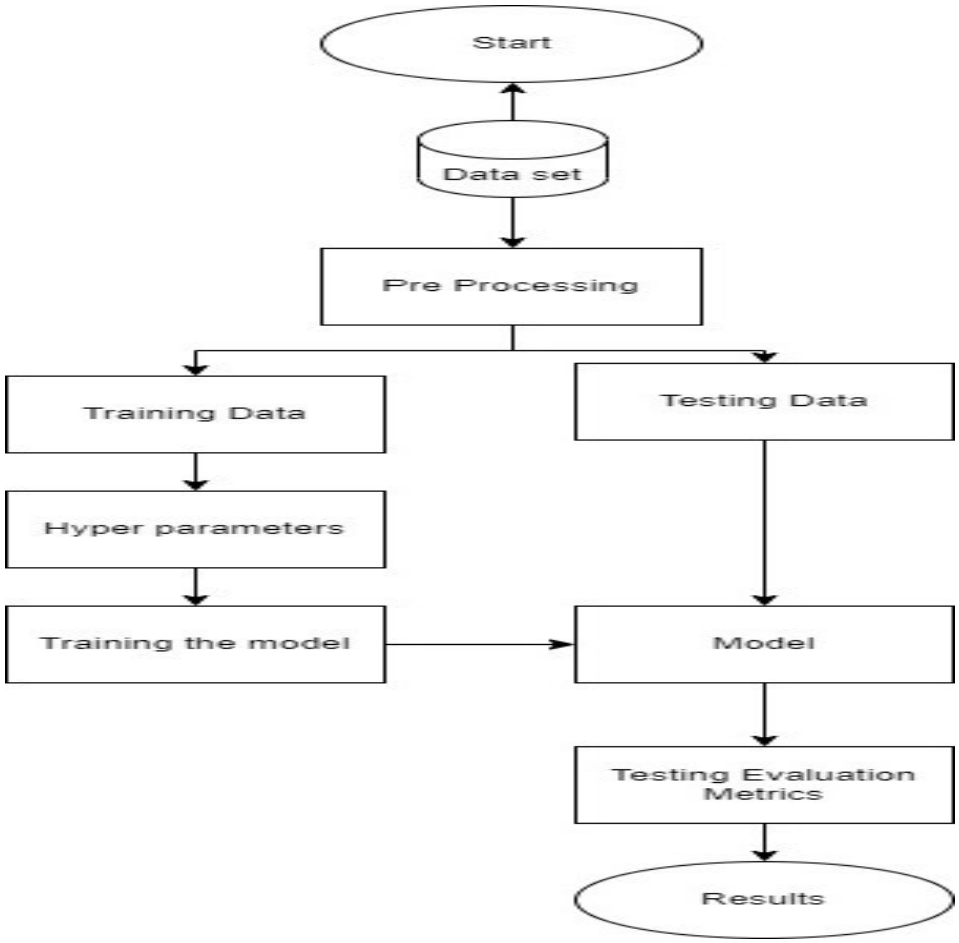
The provided values through webpage will be taken by machine learning model which is already trained and will predict the MPG. For predicting, XGBOOST with GridsearchCV and Linear Regression procedures are adopted.

Finally, a database is added using Xampp control panel which supports MYSQL and stores newly predicted value in it.



Methodology:

As the Collected data is unprocessed pre-processing and cleaning operations will be performed. The system would be constructed progressively by partitioning the dataset into two halves and incorporating hyper parameters. The findings will be presented using key metrics.



Dataset

The name of the dataset is auto-mpg which is collected from the Kaggle contains eight input parameters and one outcome described in detail as follows:

Cylinders:

Piston is an integral part of every internal combustion engine. It is a region where combustion occurs and generates electricity. A piston is located at the top of the cylinder, along with two valves one is intake and another one is exhaust valves. The piston goes down and up generates some energy that propels your car forward.

Displacement:

The power needed to transport anything from one point to another. The overall consumption of energy would be controlled by the quantity of movement.

Weight:

Motion of a particular body will depend on the weight it has and gravitation force acting upon it. The greater the mass of a substance, more the energy it takes to reach it.

Acceleration:

The variation in velocity profile is typically referred to as acceleration in regards including both strength and acceleration.

Model year:

The year where the automotive development model would be accomplished.

Origin:

The base for the car where it was created like country. someone can expect the features of cars based on the region.

MPG:

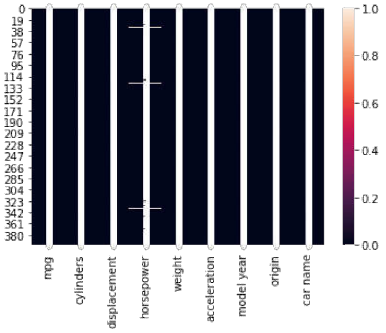
MPG stands for miles per gallon and refers to the distance a vehicle can go per gallon of gasoline. It is a popular measure for evaluating fuel savings:The more the fuel efficient of a vehicle means the greater the MPG it has.

The dataset's contents have the following features:

	MPG	Cylinders	Displacement	Weight	acceleration	Model year	Origin
Count	398.000000	398.000000	398.000000	398.000000	398.000000	398.000000	398.000000
Mean	23.514573	5.454774	193.425879	2970.424623	15.568090	76.010050	1.572864
Std	7.815984	1.701004	104.269838	846.841774	2.757689	3.697627	0.802055
Min	9.000000	3.000000	68.000000	1613.000000	8.000000	70.000000	1.000000
Max	46.600000	8.000000	455.000000	5140.000000	24.800000	82.000000	3.000000

Pre-Processing:

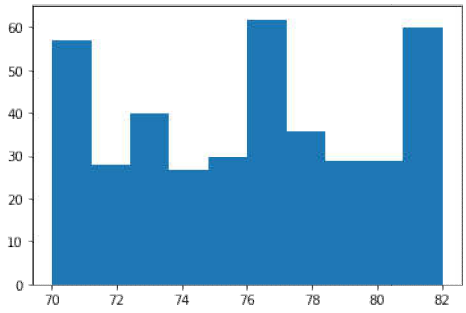
The foremost thing after collecting the required data is analysing.Studying is the way of transforming raw information into meaningful content. The result from descriptive data analysis is there are no null values presented in the collected data but there had some non-null and non-numerical values(?) in the dataset feature called horsepower which is a datatype of object. These are removed and converted to float data type which makes the prediction convenient. These missing hypotheses can be seen from the chart below.

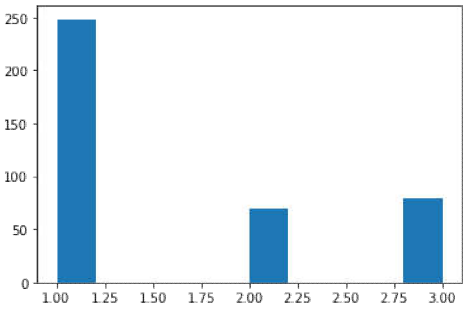


Correlations among the different variables need to find in order to justify which parameter is effecting the other parameters.The heat map includes that mass and displacement had the strongest adverse relationship with Efficiency. In contrast, the quantity of cylinders its horsepower are inversely related.

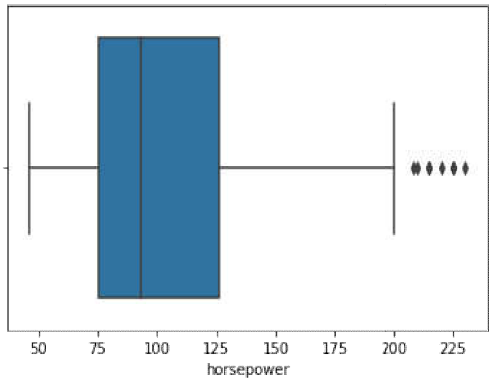
The originand model year both are categorical numerical values and can be visualized using histograms. The conclusions are shown as cars in the origin one aremostly suitedto mpg and model years 70’s, 76’s,

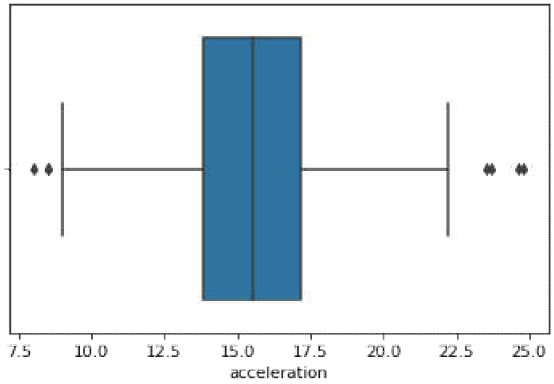
82's has more impact than others.





There are some outliers present in the horsepower and acceleration and can be seen by using boxplot as follows and can be detected and removed by dividing the data in to quartiles.





Training and Testing:

The much more important phase in learning algorithms is retraining.The information can be passed to model which finds some patterns and make predictions. These patterns can be applied on newly given data to the model and returns the output. The available information in this study was split up into two sections: 30% was employed for testing, while the leftover 70% has been used for teaching.

Hyper parameters:

The consideration of parameters where our model mainly depends on, and used to improve the accuracy are called hyper parameters.GridSearchCv presently includes 6 relevant features in its feature matrix. Tree depth, learning rate, thread count, as well as other factors are among them. The parameter grid is as follows

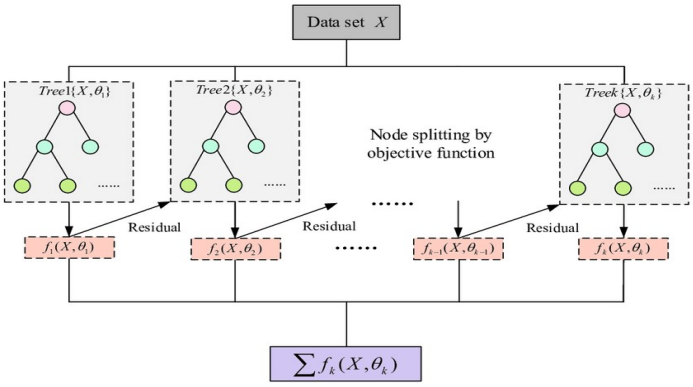
- Number of threads: [4]
- Learning rate: [.03, 0.05, .07]
- Maximum depth: [5, 6, 7]
- Minimum child weight: [4]
- Subsample: [0.7]
- colsample bytree: [0.7]
- Number of estimators: [500,1000]

Models:

Using past information, the models generated can also be used to make future analyses.It reduces forecast time while using the fewest resources possible. The model was created using the XGBOOST with GridSearchCv and Linear Regression methods.

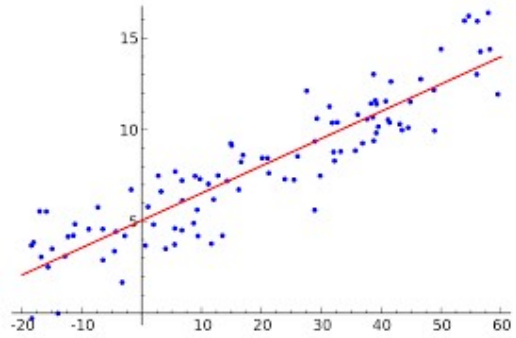
XGBOOST:

XG Boost refers to extreme gradient boosting trees. It is indeed a choice ensemble learning method for classification based on a gradient boosting structure. This can solve linear extrapolation, categorization, classifying, and consumer prediction problems. It is an outfit method that integrates an amount of poor adapts and chronologically established narrow decision trees to obtain the result, and even a scalability learning strategic plan that evades outfits and enables concurrent boosting. It continues to dominate systematic or form of tables sets of data in regression and classification computational modelling concerns. In respect of speed and reliability, it beats other machine learning techniques. It tends to work very well data containing alike mathematical and category aspects, and also data which only contains arithmetical characteristics.



Linear Regression:

It's far a popular gadget learning set of policies used for regression. Linear regression suggests the linear courting between impartial and dependent variable. Consequently, it is termed as linear regression. It manner the price of based totally variable is converting with respect to the unbiased variable. The final results of the linear regression is of numeric kind.



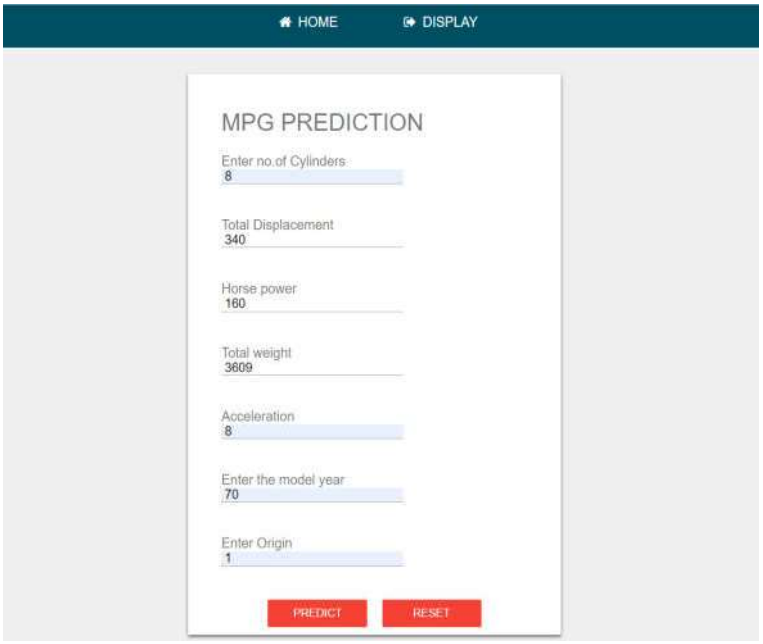
Testing Evaluation Metrics:

Arithmetic multiple regressions have been used so effective event seems to be numerical. These forecasting model' performance indicators diverge from categorization performance measures. Such metrics are made reference to as error metrics because there is a gap between the real and predictor variables.

That the very first error criterion is root mean square error, which would be the sum of the squares of the error which also determines the spacing of sample points from the linear interpolation. Another metric used in this study is mean absolute error calculates the error value based on the average. It doesn't give distinct sorts of errors just about weight; Somewhat more, as that of the quantity of damage increases, the results progressively increase. The r2-squared method, which takes into account the proportion of dependent variable can be explained by an individual entity or predictor variables and a dependent for a dependent variable, is the final metric. It clarifies the interaction of one attribute with the other.

Results:

Running of this project will start with the web service as it displays a web page asking for some information regarding future modelling car data as depicted below



The entered data will be processed and the outcomes numeric value can be predicted using the mentioned above two machine learning models. Those details will be recorded in the system for clarification purposes in the future.

+ Options

mpg	cylinders	displacement	horsepower	weight	acceleration	modelyear	origin
17.72	8	307	130	3504	12	70	1
13.94	8	340	160	3609	8	70	1
13.72	8	350	165	3693	11	70	1
17.46	8	318	150	3436	11	70	1
16.22	8	304	150	3433	12	70	1

After the comparing the results with different metrics the conclusion is XGBOOST gives better results than Linear Regression with r2-suare of 0.90.

Table 1. Comparison of Results

Model	RMSE	MAE	R2-square

XGBOOST	2.351	1.656	0.90
Linear Regression	3.401	2.591	0.80

Conclusion:

On the basis of current crisis everyone understand that how important the fuel efficiency is for a vehicle. A model was developed to calculate fuel efficiency with minimum number of parameters.When compared to the current models, it outperforms them by 90 percent.As already discussed the existing dataset contain some inaccurate data, a new database id added so that our model will perform better and significantly more reliable.

References:

1. Shirbhayye, V., Kurmi, D., Dyavanapalli, S., Prasad, A. S. H., & Lal, N. (2020, January). An accurate prediction of MPG (Miles per Gallon) using linear regression model of machine learning. In *2020 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-5). IEEE.
2. Meng, J., & Liu, X. (2006, May). MPG prediction based on BP Neural Network. In *2006 1ST IEEE Conference on Industrial Electronics and Applications* (pp. 1-3). IEEE.
3. Karpate, Y., Sharma, S., & Sundar, S. (2018). Modeling fuel efficiency for heavy duty vehicles (HDVs) in India. *Energy Efficiency*, *11*(6), 1483-1495.
4. Yin, X., Li, Z., Shah, S. L., Zhang, L., & Wang, C. (2015, October). Fuel efficiency modeling and prediction for automotive vehicles: A data-driven approach. In *2015 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 2527-2532). IEEE.
5. Shalini, L., Naveen, S., & Ashwinkumar, U. M. (2021, August). Prediction of Automobile MPG using Optimization Techniques. In *2021 IEEE Madras Section Conference (MASCON)* (pp. 1-6). IEEE.
6. Yao, Y., Zhao, X., Liu, C., Rong, J., Zhang, Y., Dong, Z., & Su, Y. (2020). Vehicle fuel consumption prediction method based on driving behavior data collected from smartphones. *Journal of Advanced Transportation*, *2020*.
7. Wickramanayake, S., & Bandara, H. D. (2016, April). Fuel consumption prediction of fleet vehicles using machine learning: A comparative study. In *2016 Moratuwa Engineering Research Conference (MERCon)* (pp. 90-95). IEEE.
8. Rusiman, M. S., Nasibov, E., & Adnan, R. (2011, December). The optimal fuzzy c-regression models (OFCRM) in miles per gallon of cars prediction. In *2011 IEEE Student Conference on Research and Development* (pp. 333-338). IEEE.
9. Jamala, M. N., & Abu-Naser, S. S. (2018). Predicting MPG for automobile using artificial neural network analysis. *International Journal of Academic Information Systems Research (IJAISR)*, *2*(10), 5-21.
10. Alsaadi, N. (2021). Comparative analysis and statistical optimization of fuel economy for sustainable vehicle routings. *Sustainability*, *14*(1), 64.
11. ”Prediction of Dengue Disease Cases by ML Techniques”, International Journal of Data Science and Machine Learning (IJDSML), ISSN : 2692-5141, Vol-1 Issue-1, Sep 2020, Page No: 1-6.
12. ”An Approach For Detecting Phishing Attacks Using Machine Learning Techniques”, Journal of Critical Reviews (JCR), ISSN : 2394-5125, Vol-7 Issue-18, Jun 2020, Page No: 321-324.
13. ”Disease Prediction and Diagnosis Implementing Fuzzy Neural Classifier based on IoT and Cloud”, International Journal of Advanced Science and Technology (IJAST), ISSN : 2005-4238, Vol-29 Issue-5, May 2020, Page No: 737-745.