LITERATURE SURVEY

DATE	19.09.2022
TEAM ID	PNT2022TMID52881
TEAM MEMBERS	Abitha.S
	Brindha.k
	Lokesh.T
	Nandana.R
TOPIC	Car resale value prediction

[1] Prediction of Prices for Used Car by Using Regression Models

ABSTRACT:

For this research, we conducted a comparative study on performance of regression based on supervised machine learning models. Each model is trained using data of used car market collected from German e-commerce website. As a result, gradient boosted regression trees gives the best performance with mean absolute error (MSE) = 0.28. . Followed by random forest regression with MSE = 0.35 and multiple linear regression with MSE = 0.55 respectively.

Keywords—comparative study, multiple linear regression, random forest, gradient boosting, supervised learning.

INTRODUCTION:

Considering the demand for private car all around the world, the demand of second hand car market has been rising and creating a chance in business for both buyer and seller. In several countries, buying a used car is the best choice for customer because its price is reasonable and affordable by buyer. After few years of using them, it may get a profit from resell again. However, various factors influence the price of a used car such as how old of those vehicles and the condition in current scenario of them. Normally, the price of used cars in the market is not constant. Thus, car price evaluation model is required for helping in trading. In this paper, we conducted a comparative study using multiple linear regression, random forest regression and gradient boosted regression trees to build a price model of used car. Each algorithm used data scraped from e-commerce website. The primary objective of this paper is to find the best predictive model for predicting used car price.

Noor and Jan [2] used multiple linear regression to predict vehicle car price. They performed variable selection technique to find the most influencing variables then eliminate the rest. The data contain only selected variable that used to form the linear regression model. The result was impressive with R-square = 98%.

Peerun et al. [3] did a research to evaluate the performance of the neural network in used car price prediction. The predicted value, however, are not very

close to the actual price, especially on cars with a higher price. They concluded that support vector machine regression slightly outperform neural network and linear regression in predicting used car price.

Sun et al. [4] proposed the application of online used car price evaluation model using the optimized BP neural network algorithm. They introduced a new optimization method called Like Block-Monte Carlo Method (LB-MCM) to optimize hidden neurons. The result shown that the optimized model yielded higher accuracy when it compared to the non-optimized model Based on the previous related works, we realized that none of them had implemented gradient boosting technique in the prediction of used car price yet. Thus, we decided to build a used car price evaluation model using gradient boosted regression trees.

RESULTS:

The following results are evaluating using testing data as input to multiple linear regression, random forest regression, and gradient boosted regression trees. The results are then compared using mean absolute error as a criterion. Table III shows mean absolute error (5) of multiple linear regression, random forest regression, and gradient boosted regression trees, in order. Gradient boosted regression yield the best performance with only MSE =0.28. Random forest regression is in second place with MSE = 0.35. Multiple linear regression has relatively large MSE of 0.55 when compared with the other.

CONCLUSION:

In this research, authors conducted a comparative study on regression based model performance. Data used in this research are scraped from German e-commerce site and then data preparation processed by using python programming language. As a result, final data have 304,133 rows and 11 attributes. We tested data by using multiple linear regression, random forest regression, and gradient boosted regression trees on that particular dataset. Each model was evaluated by using the same testing data. The results are then compared by using mean absolute error as a criterion. With gradient boosted regression trees gave the highest performance with only MAE =0.28. Followed by random forest regression with 0.35 errors, and multiple linear regression with 0.55 errors. Thus, we concluded that gradient boosted regression trees is recommended to develop the price evaluation model. The future work can be developed from this research by ¿ne tuning each model parameter. More appropriate data engineering can be utilize to create the better training data.

[2] Prediction of Used Car Price Based on Supervised Learning Algorithm

ABSTRACT:

In this paper, we use machine learning algorithms to predict the price of used cars with less human intervention to make the results more objective. The method used is to preprocess the dataset through Python's Pycaret package and compare the performance of each algorithm through the algorithm comparison function, in this study Extra Trees Regressor, Random Forest Regressor performs relatively well. Finally, the algorithm was optimized by using the hyperparameter function. The results show that R2 = 0.9807 obtained from extreme random numbers is the best performance. The algorithm was obtained and validated with new data to derive the final algorithm model. When new used car data flows into the used car system, used car prices will be automatically generated by this algorithm, which will make the workflow of the used car market faster and more competitive for that used car market.

Keywords:machine learning, supervised learning, used car price, prediction

INTRODUCTION:

According to relevant reports, in the next five years, the annual growth rate of automobiles in China will be 3.5%, while the annual growth rate of used cars will be 5%. The annual growth rate of used cars and automobiles is constantly expanding. Therefore, consumers think that when buying a new car, they will also consider the price of the same type of used car, especially some value-preserving brand cars are more worthy of consumers' attention, which is a change of value, and consumers can get the best return on investment. Faced with this situation, companies operating the used car market use traditional marketing methods (consulting prices for many times) to deal with business, which greatly increases the company's operating costs. This paper will predict the used car prices through various supervised learning algorithms in machine learning, and the used car companies can directly publish the predicted prices through Internet channels, so that consumers can know the used car prices at a glance and provide operational efficiency of the company

RESULTS:

Through the comparison of a series of algorithms of supervised learning regression algorithm, it is concluded that et algorithm has the best performance. through data preprocessing, super-parameter adjustment and other operations, R2=0.9807 is finally obtained, and 20% of the reserved data is tested to meet the expected standard [9]. Next, we can automatically get the predicted price of used cars by et algorithm after collecting all the 11 feature values on the Internet. The price of new used cars on the Internet platform is obtained by ET algorithm, and consumers can know the price directly by looking at the data.

CONCLUSION:

In this study, the author makes a series of performance comparisons based on supervised learning algorithms. The data set used here comes from the price of used cars, and python language is used to predict the data set [11]. It can be seen from the results that we compare the performance by using several algorithms, such as ET, rf, ridge, and so on. Each model is tested by using the same training data. The results are compared with the average absolute error and further demonstrated by the multi-dimensional evaluation model. Then the best performance model is selected as the prediction model, and finally verified by the new used car data. The result given from the best performing algorithm model is R2=0.9807, and the final verification of new data shows that ET alalgorithm is the best model for the second-hand car price prediction, and it will be more in line with the daily operation by adjusting the super parameters in the future work [11]. In practical application, inputting all kinds of characteristic data through the Internet port will directly display the prediction results on the port interface, which greatly improves the working efficiency of the used car market, thus improving its market competitiveness.

[3]Prediction Of Used Car Prices Using Artificial Neural Networks And Machine Learning

ABSTRACT:

With the extensive growth in usage of cars, the newly produced cars are unable to reach the customers for various reasons like high prices, less availability, financial incapability, and so on. Hence the used car market is escalated across the globe but in India, the used car market is in a very nascent stage and mostly dominated by the unorganized sector. This gives chance for fraud while buying a used car. Hence a high precision model is required which will estimate the price of an used car with none bias towards customer or merchandiser.In this model, A Supervised learning-based Artificial Neural Network model and Random Forest Machine Learning model are developed which can learn from the car dataset provided to it. This project presents a working model for used car price prediction with a low error value. A considerable number of distinct attributes are examined for reliable and accurate predictions. The results obtained agree with theoretical predictions and have shown improvement over models which use simple linear models. An ANN (Artificial Neural Network) is built by using Keras Regression algorithm namely Keras Regressor and other Machine Learning Algorithms namely Random Forest, Lasso, Ridge, Linear regressions are built. These algorithms are tested with the car dataset. Experimental results have shown that the Random Forest model with a Mean Absolute Error value of 1.0970472 and R2 error value of 0.772584 has given the less error among all the other algorithms.The work presented here has shown profound implications for future studies of Used Cars price Prediction using Random Forest and might one day help to solve the problem of frauds with one hundred percent accuracy.

INTRODUCTION:

The manufacturing rate of cars has been increasing notably during the past decade, with almost 90 million cars being manufactured in 2020. Car production rates have been expanding dynamically during the previous decade, with right around 90 million vehicles being delivered in the year 2020. The total number of vehicles that are sold in a year all throughout the planet is actually an enormous number. With the rapid growth of population rate, the individual's rate who wants buy a car is also increasing. So, the used car market presently comes into picture as a well-developing industry. There is a percentage of increase in the used cars purchase from the year 2019 to 2020 in India which is expected to increase by 10 percent by the year 2024.

[1] The first paper is based on predicting the Price of Secondhand Cars using Artificial Neural Networks. In this paper they used various supervised techniques to estimate the price of used cars in Mauritius. They used the data from daily newspapers. Various Machine Learning algorithms like Support Vector Regression, Linear Regression and K-Nearest Neighbor were used.

[2] The third paper is Used Cars Price Prediction using Supervised Learning Techniques wherein error rates were calculated for models like Lasso Regression, Multiple Regression, Regression Tree and ANOVA based comparison was made for the models if they significantly differed from each other.

[3] The fourth paper is Prediction of Car Price Using Linear Regression. The data set was taken from an online ecommerce website quickr.com, in this paper comparison was made for various regression based models like Multiple Linear Regression, Random Forest Regression and was concluded by choosing the best algorithm.

RESULT:

The performance metrics of all the algorithms are mentioned below in the table:

Algorithm type	Mean Absolute Error	R-Squared Error
DEEP NEURAL NETWORK (with 10 layers)	0.766	0.842
LINEAR REGRESSION	1.152	0.837
LASSO REGRESSION	1.051	0.871
RIDGE REGRESSION	1.143	0.809
RANDOM FOREST	0.746	0.917

Among all the proposed models, Random Forest determines the price of a used car with minimum possible error.

[4]Second Sale Car Price Prediction using Machine Learning Algorithm

ABSTRACT:

Every business firm recognizes the need of making sound and challenging decisions. Poor decisions can lead to substantial losses and even the demise of a firm. This paper is focused on one of the retail enterprises, which deals with the used car sales. The major goal is to develop a prediction model that can estimate the selling price of used cars based on key factors. Machine learning techniques such as Random Forest Regression, Feature engineering technique such as Extra Trees Regression are employed to accomplish the goal as Random Forest Regression is modeled for prediction analysis and Extra Trees Regression fits the number of decision trees. The results are so encouraging with our approach.

Keywords: Hyperparameter Tuning, Categorical data, RandomisedSearchCV, Prediction Model

INTRODUCTION:

Nowadays the Indian Used Car Market become so exuberant, the sales figure of secondhand cars has grown to be double when comparing with the size of the new purchased cars. Automatically used car sales have become more organized in recent years. Over the last three years, it was witnessed amazing progress, and the prices are expected to rise by up to 17% or even more. To effectively determine the car's worthiness, a prudent mechanism is essential. In this paper, a prediction model is built so that it estimates the selling price of already used cars based on their features. This in turn reduces the burden and risk from the seller, consumer and also

provides a positive firsthand knowledge about price and low financing cost for used cars.

But there are only a few records and features in the collection. This is considered as the fault. The proposed model which is shown is a combination of the two machine learning algorithms i.e. Random Forest Algorithm and Extra Trees Regression algorithm.

Random Forest Regression: The Random Forest algorithm is used to construct decision trees at the training phase. This method is also known as an Ensemble-Bagging method[15]. Random Forest is a mixture of both regression and classification techniques. It also works better for large datasets.The term Ensemble is derived from the word "assemble", which refers to the process of combining several elements into a single working model or product. Bagging is a type of Ensemble Technique where the result is obtained by using the voting-win method .

RESULT:

A. Comparison between the Linear Regression and Random Forest Regression Algorithms

ALGORITHM	MAE	MSE	RMSE
Linear Regression	9.5351	3.1199	6.4762
Random Forest Regression	4.1918	0.4069	2.4082

Finally to conclude that Extra Trees Regression gives good result in feature engineering and also Random Forest Regression algorithm is best suited for forecasting the price of the second cars with much accuracy.