Car Price Prediction Through Cloud Server

1st Thasin Sahabuddin Sadia
dept. of Computer Science and Engineering
United International University
Dhaka,Bangladesh
tsadia181150@bscse.uiu.ac.bd

2nd Md Tanvir
dept. of Computer Science and Engineering
United International University
Dhaka,Bangladesh
mtanvir191258@bscse.uiu.ac.bd

3rd Saima Rahaman Eyana
dept. of Computer Science and Engineering
United International University
Dhaka,Bangladesh
seyana191252@bscse.uiu.ac.bd

4th Md Mohiuddin Munna
dept. of Computer Science and Engineering
United International University
Dhaka,Bangladesh
mmunna173035@bscse.uiu.ac.bd

5th MD Sabbir
dept. of Computer Science and Engineering
United International University
Dhaka,Bangladesh
msabbir172032@bscse.uiu.ac.bd

6th Md. Rafiqul Islam
dept. of Computer Science and Engineering
United International University
Dhaka,Bangladesh
mislam191047@bscse.uiu.ac.bd

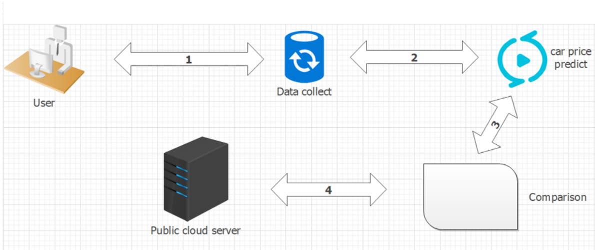
7th Dr. A.K.M. Muzahidul Islam
Professor,dept. of Computer Science and Engineering
United International University
Dhaka,Bangladesh
muzahid@cse.uiu.ac.bd

***Abstract*—The current study puts forth the more recent idea of anticipating the costs of specific items in this rapidly changing generation. We developed a method to obtain an accurate assessment applying machine learning techniques on someone's car since we wanted to help everyone and it would save them a lot of time and money. Predicting the price of cars has drawn a lot of interest in research because it necessitates significant effort and subject matter expertise. For the dependable and accurate prediction,a large number of unique attributes are considered. We used the Random Forest Regression model and the Lasso Regression model,two machine learning approaches, to create a model for forecasting the cost of cars. In order to increase the anticipated accuracy of the data set, the Random Forest classifier averages the results from multiple decision trees applied to various subsets of the input data set. The random forest uses predictions from each decision tree and predicts the outcome based on the votes of the majority of projections rather than relying solely on one decision tree. Higher accuracy and over-fitting are prevented by the larger number of trees in the forest.Additionally, the aim of lasso regression is to identify the subset of predictors that reduces prediction error for a quantitative response variable.Our study suggests a system in which all the data is manually collected, the automobile price is predicted, the accuracy of two M.L. models is then compared,and finally the system is hosted on a cloud server.**

***Index Terms*—data set train, car price prediction, cloud environment, context based**

I. Introduction

Due to the fact that the cost of an automobile is often determined by a number of distinctive features and elements, accurate car price prediction requires specialist expertise. Usually, brand and model, age, horsepower, year,and mileage are the most important ones.because of the changing requirements in gasoline prices, this kind of combustible(Petrol/Diesel) used in the automobile and energy consumption per mile have a significant impact on the price of a car. Consequently, the car's value will be impacted by many aspects such as the outside color, door count, transmission type, dimensions, safety, air conditioning, interior,and whether or not it includes navigation. We used a variety of strategies and techniques To ascertain the car's reliability, we carried out this experiment of price prediction.This project's main objective is to examine the accuracy of used automobile price predictions made using several Machine Learning (ML) models. By doing so, customers may be able to base their choices on many inputs or considerations. In addition, we will personally gather every piece of the data set, which will then be used to estimate the price. As the data-set grows, it will subsequently be hosted on a cloud server. The up-time of the server, cost-efficiency, increased security,location independence, increased group collaboration, backup, disaster recovery, and, most importantly, the use of the newest technology all directly affect the performance of the website.Here is the detailed work proposal diagram.



A. Motivation

It might be challenging to determine whether an automobile is worth the advertised price when someone sees the lists online.Numerous elements, including mileage, make, model, year, selling price, current pricing, etc., might affect a car's true value.Setting a fair price for a vintage car presents problems from the seller's side as well. The goal is to employ machine learning algorithms to create models for forecasting used automobile prices based on existing data. Additionally, cloud hosting scales more fast than traditional hosting. The cloud servers scale up and down automatically depending on whether a website or application experiences increased or decreased traffic. Unlike shared hosting, where server space must be manually added or removed,cloud hosting does not require this.

The remainder of the article is structured as follows. Section II presents an overview of previous efforts on car price prediction systems. Section III describes our proposed methodology for predicting car prices using a cloud server, and Section IV shows our experimental analysis and outcomes. Chapter V is the conclusion, which covers our future progress and improvement.

II. Related Works

A. literature review

1. With accuracy ranging between 50% to 90%, used car values have been forecasted using a range of studies and related works employing different techniques and approaches. The analyzer recommended in (Pudaruth, 2014) [8] to forecast used car values in Maldives. He grounded his results on historical data acquired from the journal and a range of machine learning approaches, such as decision trees, K-nearest neighbors, multiple regression, and Naive Bayes algorithms [16].

2. Statistical models were suggested by Nitis Monburinon et al. [10] to predict used car prices. The information used in this study was chosen by the researchers from a German attaining portal. Finding an appropriate prediction model to forecast the price of used cars is the primary objective of this endeavor. They compared the results using various machine learning algorithms, using the average absolute error as the measurement. In comparison to linear regression, which has an MAE value of ”point five five” and random forest, which has an MAE value of ”point three five”, they claimed that their model with gradient-boosted regression has a lower error with an MAE value of ”point two eight” and performs better overall.

3. In this study, the K-Nearest Neighbor technique is utilized to predict the pricing of secondhand cars using a machine learning model. Utilized automobile data for various trained and test ratios is used to train the algorithm. The effectiveness of the suggested model is then evaluated using the K layer approach in order to prevent over fitting. Therefore, thinking Patterns Cars data set was used as the basis for data processing to strip out extraneous information. To more accurately estimate the sales of used automobiles, the model was trained using the processed data and the KNN algorithm. [9].

4. Three alternative machine learning methods were employed by researcher to forecast used automobile values. After pre-processing, the used automobile data scraped from a local Bosnian website totaled 797 car samples. The methods suggested were assistance Vector Machine, Random Forest, and Artificial Neural Network. Results showed in order for employing a single machine learning algorithm alone produced results that were less than 50% accurate, however combining the formulas with the pre calculation of pricing using Random Forest produced outcomes that were up to eighty seven point three eight percent accurate. [17]

5. Researchers (Nabarun Pal, 2018) employed a be in charge of learning method known as Random Forest to forecast the price of used autos. Data set from Kaggle was used to forecast used car prices. Analyzing scientific questions with consideration was undertaken to identify the pricing effects of the each attribute. It is usually used during categorization, however they converted the issue to an equivalent prediction model to turn it into a multivariate regression. Validation accuracy was eighty three point six three percent, whereas training accuracy rate was ninety five point eight two percent. [15]

III. Methodology

In this section, we described our recommended method, as well as the list of parameters needed for our task. A summary of the full prototype design

A. Materials and method

The study's strategy for estimating car prices includes a number of steps, which are displayed in the diagram below. Figure 1.

Here,We locate the data-set for automobile price prediction on the regional website kaggle.com, where we gather the data. For each car, the pursuing characteristics were noted: ”Fuel Type,” ”Seller Type,” ”Transmission,” ”Owner,” ”Selling Price” ,”Present Price,” and ”Kms Driven” are the many variables. As part of the research process, a ”web scraper” is created to automate this task and reduce data gathering time

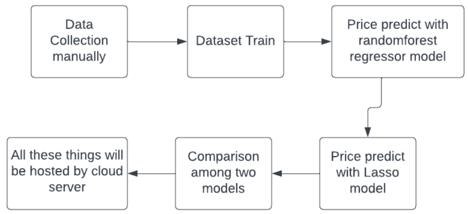


Fig. 1.

because manually data gathering requires a lot of time, particularly when there are numerous records to process. A well-known method for obtaining data from websites and storing it in a local file or database is web scraping. Web scrapers are used to complete manual data extraction in a fraction of the time it would take by hand. Web scrapers are designed specifically for a given website and are capable of imitating normal visitors from the website's perspective. Data preparation was carried out after raw data had been gathered and saved in a local database. Numerous qualities lacked essential data and could not be used to make predictions [11].In attempt to make the model simpler complex, the data-set was pre-processed after data collection to remove samples with missing values, add some random samples for the non numerical parts, fix any discrepancies in the units, and remove attributes that don't affect price evaluations when necessary. Data preparation and understanding is a crucial component of constructing a model since it provides a preliminary examination of the data in addition to highlighting outliers and skewedness. Additionally, it helps to be aware of the key factors that influence how prices are determined. To understand the relationships between the various components, a correlation matrix for each attribute was used. The data is then organized and translated into a format that the data train approach can process. In this work, two models are suggested to be constructed using the Random Forest Regressor and the Lasso Regression Model approach. First, the data was divided into sections for testing and training. Portioning percentages can be tried with various ratios to comparison various results. All three models were assessed through using three evaluation indices known as model score, mean square error (MSE), mean absolute error (MAE), and root mean square error (RMSE). The Random Forest Regressor performed best across the board.
1. Mean Square Error (MSE):
The Mean Squared Error (MSE) is likely the preeminent basic and time-honored loss function, and it is frequently taught in introductory Machine Learning courses. To compute the MSE, divide the difference between the model's predictions and the classification algorithm by two, square it, and average it over the entire data-set.
2. Root Mean Square Error (RMSE):
The nonlinear Root mean square assessment algorithm also establishes the average error magnitude. It is the square root of the average of the squares difference between both the prediction and the observed data.
3.Mean Absolute Error (MAE):
This metric symbolizes the average absolute difference between the dataset's actual and anticipated values. It represents the average residual of the dataset.
Then,it will host on a cloud.Here we need to follow these things :
1.Context setting.
2.Source repository cloned.
3.Docker containers can be made using Cloud Build.
4.Container deployment on Cloud Run.
5.New revision should be created with less concurrency.
6.Website updates should be made.
7.website update urgently with no downtime.
8.Website updates should be made.

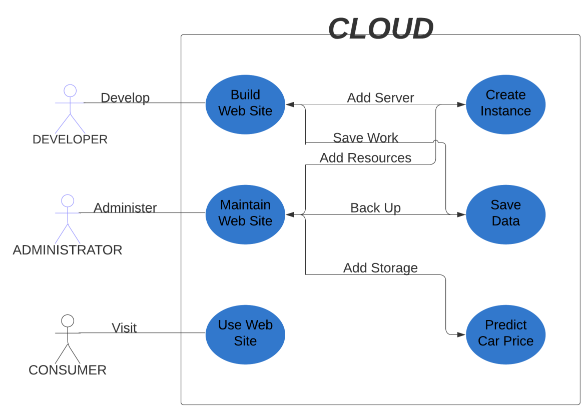


Fig. 2.

As a result, it is determined to take them out of the dataset. here are (309,9) samples in the raw data set that was gathered from the dataset Figure 3. A sample of a processed data collection in CSV format

	Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
0	ritz	2014	3.35	5.59	27000	Petrol	Dealer	Manual	0
1	sx4	2013	4.75	9.54	43000	Diesel	Dealer	Manual	0
2	ciaz	2017	7.25	9.85	6900	Petrol	Dealer	Manual	0
3	wagon r	2011	2.85	4.15	5200	Petrol	Dealer	Manual	0
4	swift	2014	4.60	6.87	42450	Diesel	Dealer	Manual	0

Fig. 3.

IV. Results and Discussion

Car Name was dropped because it was a redundant property. The steps of exploratory data analysis are as follows:

1.The dataset was plotted in the following manner to display the missing values showing in fig 4

```
#Missing values(if any)
car.dropna(axis = 0,  inplace=True) ##
for column in car:
  if car[column].eq(' ').sum():
    print(column, ' has missing value: ', car[column].eq('' ).sum())
  elif car[column].eq('?').sum():
    print(column, 'has missing value: ', car[column].eq('?').sum())
  else:
    print(column, "has no missing values")
```

```
Car_Name has no missing values
Year has no missing values
Selling_Price has no missing values
Present_Price has no missing values
Kms_Driven has no missing values
Fuel_Type has no missing values
Seller_Type has no missing values
Transmission has no missing values
Owner has no missing values
```

Fig. 4.

2. It's time to look for links between attributes after data transformation. To ascertain whether relationships exist, a correlation matrix was used showing in fig 5

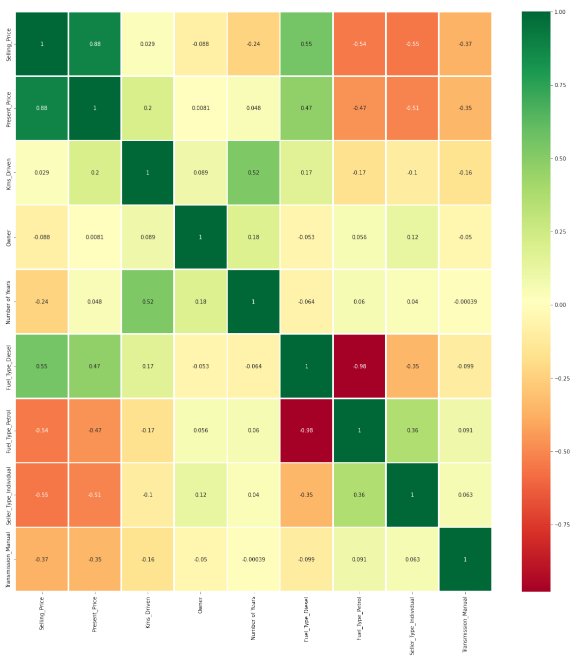


Fig. 5.

3. The top five features are listed below based on all the other features.showing in fig 6

4. The next step in visualization is to compare two machine learning models and evaluate how skewedly all attributes are, regardless of whether they are regularly distributed or not showing in fig 7 and 8

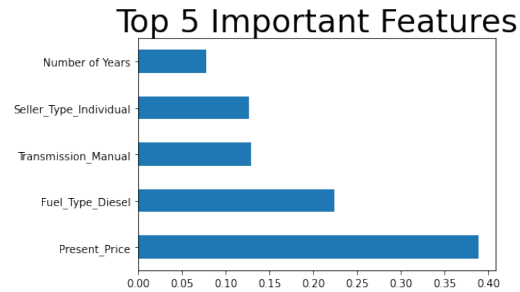


Fig. 6.

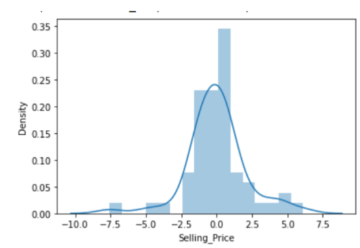


Fig. 7.

5. The outcomes of comparing actual prices to expected prices and the comparison between two models are as follows. showing in fig 9 and 10

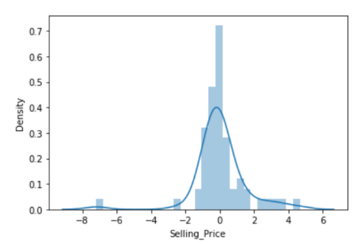


Fig. 8.

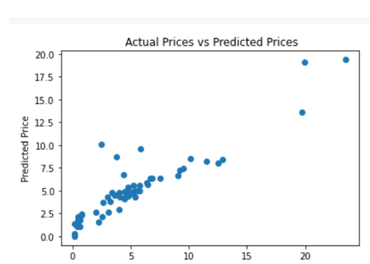


Fig. 9.

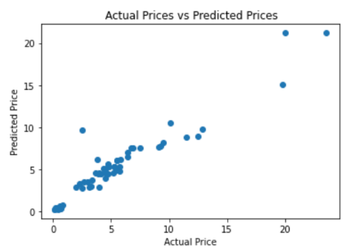


Fig. 10.

6. Here is the website where users will input the dataset manually on the basis of their requirements.

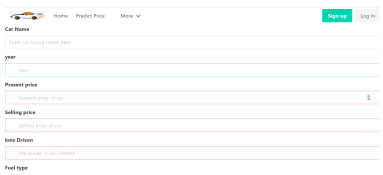


Fig. 11.

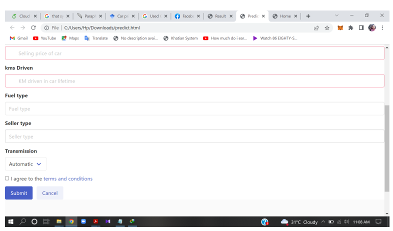


Fig. 12.

Fig. 13.

7.Now all off these things will hosted by cloud server in figure 14.

V. Conclusion

Because there are so many factors to take into account for an accurate projection, predicting car prices can be difficult.The gathering and preparation of the data is a crucial phase in the prediction process.To eliminate unneeded noise for machine learning algorithms, normalization, standardization, and data cleaning PHP scripts were written for this study.The accuracy of a first algorithm applied to the data set was under 80 As a result, Different approaches to machine learning integrated have been offered, and the mix of machine learning techniques

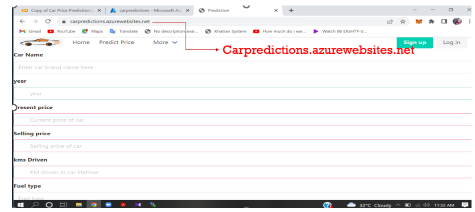


Fig. 14.

achieves accuracy of more than 98 When compared with a first machine learning method approach, it shows a huge improvement. The disadvantage of these systems are it uses significantly more computer expedient than a eligible machine learning method.Despite the apparatus remarkable execution in the car price prediction task, our aim for the coming exploration is to see how well it performs when tested against different data sets. With the help of the used automobile data sets from OLX [12] and eBay [13], we will expand our test data and confirm the suggested methodology.The consumers will benefit from getting updates on automobile price predictions and cloud hosting in addition to Cloud storage, in particular dedicated server hosting, gives consumers access to a cloud provider's own security measures to guard against cybercrime and other security risks.

VI. Future Work

Additional data will be gathered in the coming using various techniques, and new algorithms will be tried. The advantage of emerging technologies will then be included in smartphone platforms for wider populace use. Additionally, some prognosticators assert that edge computing will replace cloud computing after the period of data collection in opposition to the cloud hosting phase because computing will become decentralized and the need for the centralized cloud will disappear, allowing us to have an actual time processing scheme.

References

[1] Agencija za statistiku BiH. (n.d.), retrieved from: http://www.bhas.ba . [accessed July 18, 2018.]

[2] Listiani, M. (2009). Support vector regression analysis for price prediction in a car leasing application (Doctoral dissertation, Master thesis, TU Hamburg-Harburg).

[3] Richardson, M. S. (2009). Determinants of used car resale value. Retrieved from: https://digitalcc.coloradocollege.edu/islandora/object/coccc%3A1346 [accessed: August 1, 2018.]

[4] Wu, J. D., Hsu, C. C., Chen, H. C. (2009). An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference. Expert Systems with Applications, 36(4), 7809-7817.

[5] Du, J., Xie, L., Schroeder, S. (2009). Practice Prize Paper—PIN Optimal Distribution of Auction Vehicles System: Applying Price Forecasting, Elasticity Estimation, and Genetic Algorithms to Used-Vehicle Distribution. Marketing Science, 28(4), 637-644.

[6] Gongqi, S., Yansong, W., Qiang, Z. (2011, January). New Model for Residual Value Prediction of the Used Car Based on BP Neural Network and Nonlinear Curve Fit. In Measuring Technology and Mechatronics Automation (ICMTMA), 2011 Third International Conference on (Vol. 2, pp. 682-685). IEEE.

[7] Pudaruth, S. (2014). Predicting the price of used cars using machine learning techniques. Int. J. Inf. Comput. Technol, 4(7), 753-764.

[8] Pudaruth, S., 2014. Predicting the price of used cars using machine learning techniques. Int. J. Inf. Comput. Technol, 4(7), pp.753-764.
[9] Samruddhi, K. and Kumar, R.A., 2020. Used Car Price Prediction using K-Nearest Neighbor Based Model. Int. J. Innov. Res. Appl. Sci. Eng.(IJIRASE), 4, pp.629-632.
[10] Monburinon, N., Chertchom, P., Kaewkiriya, T., Rungpheung, S., Buya, S. and Boonpou, P., 2018, May. Prediction of prices for used car by using regression models. In 2018 5th International Conference on Business and Industrial Research (ICBIR) (pp. 115-119). IEEE.
[11] Weka 3 - Data Mining with Open Source Machine Learning Software in Java. (n.d.), Retrieved from: https://www.cs.waikato.ac.nz/ml/weka/. [August 04, 2018]
[12] Used cars database. (n.d.) Retrieved from: https://www.kaggle.com/orgesleka/used-cars-database. [accessed: June 04, 2018].
[13] OLX. (n.d.), Retrieved from: https://olx.ba.[accessed August 05,2018].
[14] Monburinon, N., Chertchom, P., Kaewkiriya, T., Rungpheung, S., Buya, S., Boonpou, P. (2018). Prediction of Prices for Used Car by Using Regression Models. 5th International Conference on Business and Industrial Research (ICBIR), (pp. 115-119). Bangkok.
[15] Nabarun Pal, P. A. (2018). How much is my car worth? A methodology for predicting used cars prices using Random Forest. Future of Information and Communications Conference (FICC) 2018 , 1-6.
[16] AlShared, Abdulla, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.Used Cars Price Prediction and Valuation using Data Mining Techniques, 2021
[17] Gegic, E., Isakovic, B., Keco, D., Masetic, Z. and Kevric, J., 2019. Car price prediction using machine learning techniques. TEM Journal, 8(1), p.113.