DOI: https://doi.org/10.51846/vol5iss4pp52-59

Pakistan Journal of Engineering and Technology, PakJET
ISSN (p): 2664-2042, ISSN (e): 2664-2050
Volume: x, Number: 5, Pages: 52- 59, Year: 2022

The Framework of Car Price Prediction and Damage Detection Technique

Muhammad Kashif Shaikh [1], Hassan Zaki [1], Muhammad Tahir [1], Muzammil Ahmed Khan [2], Osama Ahmed Siddiqui [2] and Inam Ur Rahim [3]

[1]Department of Software Engineering, Sir Syed University of Engineering & Technology, Karachi, Pakistan.
[2]Department of Computer Engineering, Sir Syed University of Engineering & Technology, Karachi, Pakistan.
[3]Vice President, IT Department, JS Bank Limited, Karachi, Pakistan.

Corresponding author: Muhammad Kashif Shaikh (e-mail: mkshaikh@gmail.com).

Received: 05/08/2022, Revised: 01/11/2022, Accepted: 10/12/2022

***Abstract*- In this paper, the research area has always been car price forecasting. We demonstrate that using the proposed object detection method, the type of damage can be categorized into two classes with good accuracy damaged and undamaged. So, when we discovered these issues, we decided to develop a mobile application called Car Price Prediction, which allows users to anticipate the price of a used car. So, we trained the damage identification model using our data using a state-of-the-art image detection method convolutional neural network and evaluated the accuracy on a GPU server and a smartphone.**

***Index Terms*-- Price Prediction, Used cars, Regression Model, Convolutional Neural Network.**

I. INTRODUCTION

The price of a car results from continuous volatile prices of a supply of energy and various city, fuel types, and owner types. When car enthusiasts wish to purchase a new vehicle, they travel to a dealership or another location where new vehicles are sold.

However, when it comes to a secondhand car, many people can afford one. Why? And how? If we compare the price of a new car to that of a used car, the price of a used car will always be half that of a new car. If someone wants to buy a brand new zero-kilometre car, the cost will be around 20 lacs, but if they buy a used car with 1000000 kilometres on it, the cost will be half or more than half, approximately 0.6 to 0.7 million.

So, we have Worked on a Mobile application that is needed for now a day so that people may be able to predict the price of a used car, and we have created car dents accuracy find feature. By using this feature, car buyers can check the accuracy of the damaged car by uploading an image and checking the result by selecting get damage result.

This paper examines the advancement of a multivariate relapse show to anticipate the retail cost of 2005 Common Engines (GM) cars. Measurable reading material regularly offers numerous little information sets chosen to demonstrate an assortment of issues and procedures that a client of relapse ought to know. The structure of this dataset permits understudies to work through the complete handle of demonstrating building and evaluation, hence giving a guided phone run some time recently handling a comprehensive dataset on their claim. Shockingly, data around all these components aren't continuously accessible, and the buyer must choose to buy at a certain price based on many variables, as it were. This inquiry is about connected machine learning strategies to make a demonstration that can foresee the cost of utilized cars. Although small data sets can offer the advantage of a sharp focus on issues, their narrow focus also carries disadvantages. A few extraordinary variables to which buyers connect significance in Mauritius are nearby past proprietors, whether the car had been included in genuine mishaps and whether it may be a lady-driven car. As we will see, the cost depends on numerous variables. The foremost imperative ones are ordinarily the age of the car, its make (and show), the beginning of the car (the first nation of the producer), its mileage (the number of kilometres it has run), and its drive. It is commonplace information that the esteem of utilized cars depends on a few components.

II. LITERATURE SURVEY

This paper uses various methods, including Multiple Linear Regression Analysis, K-Nearest Neighbor, nave Bayes, and decision trees, but he claims that finding the greatest accuracy result was tough [1].

Another researcher Shonda Kuiper also worked on this project [2]. He has collected the data from Kelly Blue Book for several hundred 2005 used General Motors and used a Multivariate

Regression algorithm to predict the value of the used car. And it used various attributes such as mileage, make, model, engine size, and interior style.

To anticipate the price of a secondhand car, Nabarun Pal employed the Random Forest algorithm [3]. This model was adopted after extensive exploratory data analysis to establish the impact of each attribute on price, he explains in his article. And after training his datasets, it discovered that the training accuracy was 95.82 percent, and the testing accuracy was 83.63 percent. And he claims that this model accurately predicted pricing based on given qualities. [3].

EnisGegic and his crew have also worked on a project to forecast car prices [4]. And he and his colleagues used three algorithms (Artificial Neural Network, Support Vector Machine, and Random Forest). Using a web scraper, they obtained the datasets from the file name (autopijaca.ba) web page. The web scrapper is only used to extract data into an excel spreadsheet. After completing the algorithm, they integrated it into a Java application, then tested the data for correctness, achieving an accuracy of 87 percent.

Multiple Linear Regression shown in Fig. 1 was utilized by Sagar Tanna and his team [5]. And he explains that if you must manually forecast the price, you go to the company, where they inspect the cars and produce the documentation, and then you get the price after a long process. However, in today's world, when no one has enough time to go through the lengthy method, they desperately need a short period to estimate the price to save time and acquire a used automobile. Using the existing system, a user can access an online portal where, if he is a buyer, he can look for a car and speak with the automobile owner. If the seller accesses the portal, he can also connect with the seller.

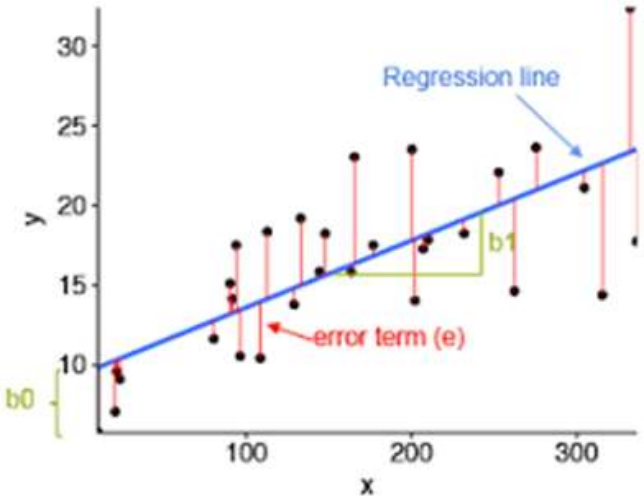

FIGURE 1: Multiple Linear Regression

Gonggi [6] predicted the price using the artificial neural network on the residual value of privately used cars. And he predicted the price by using a few attributes mileage, manufacturer, and estimated useful life. And the model has predicted the accurate value or price.

Richardson's [7] inquiry was based on the speculation that car firms are more energetic in delivering vehicles that don't deteriorate quickly. Utilizing numerous relapse examinations, he illustrated that crossover cars are more competent in keeping consistent speed. Their worth is higher than that of traditional automobiles. This is most likely the case. It is a better choice because of increased environmental concerns And its higher fuel efficiency and the climate. The significance of other parameters such as age, mileage, and MPG (miles per gallon) and make also considered. Considered in our research, He gathered all his information from various websites [7].

Another consideration that Listiani [8] details utilize is Back Vector Machines (SVM) to expect the values of rented autos. When a huge information set is accessible, SVM is impressively more exact in determining costs than numerous direct relapses, according to this think. SVM is also predominant in dealing with high-dimensional information and dodges under-fitting and over-fitting issues. SVM employments a hereditary calculation to discover fundamental characteristics. However, the technique does not demonstrate why SVM is better than ordinary multiple regression in terms of variance and mean standard deviation [8].

Pinheiro et al. [11] depicted a Profound Cover acknowledgment showing that yields estimation choice veils through events shown within the input picture to partition each illustration substance, but the boundary classification strategy is moo [12]. Li et al. [13] display the full convolutional case recognizable proof as the primary end-to-end semantic division strategy. It figures out the outline and occurrence division by making strides in position-sensitive rating mapping. Still, when analyzing numerous protest occasions, it seems as if it roughly distinguished the border of each case question [14].

Among the current segmentation techniques, He et al. [15] proposed the Mask RCNN structure, which is an algorithm with basically fine example classification performance [16]. Lin et al. [17] utilized Mask RCNN to classify wheat planthoppers and found that wheat planthoppers and non-rice planthoppers might be recognized fast and efficiently, with a mean recognition rate of 0.923. Wang et al. [18] used Masked RCNN to identify ship objects, showing that Mask RCNN works better while handling intimately connected objectives and cross objectives.

Li et al. [19] proposed a Mask RCNN-based structure focus detection algorithm. Building targets can be identified with 94.6 % in remotely sensed images of various scenes. The Cover RCNN calculation incorporates an assortment of applications, and no one has utilized it within vehicle harm discovery. Desktop Applications and Mobile Applications display the records related to the parameter captured by the corresponding screenshots [20]. The correlation coefficient of the new car price, engine power, and used car price is greater than 0.6, which has a certain linear correlation [21].

III. METHODOLOGY

Our initial step was to get the datasets for predicting the price and train a model for forecasting the price of a used automobile, as well as got the images and train photos to machine and apply

methods for predicting the exact price and determining the accuracy of damaged and undamaged cars by uploading an image. Following that, we examined the model that can more accurately estimate the price. The datasets were gathered from the Internet. After receiving the automobile dataset, the price was predicted using the following criteria: *brand, city, purchase year, kilometres driven, fuel type, transmission, mileage, engine, power, seats*, and *owner type*. When someone goes to buy a car, these factors are always considered. We also downloaded photos from many websites, such as Google. [9]. Because we need to teach the system how to distinguish between damaged and undamaged autos. As you can see in Fig. 2.

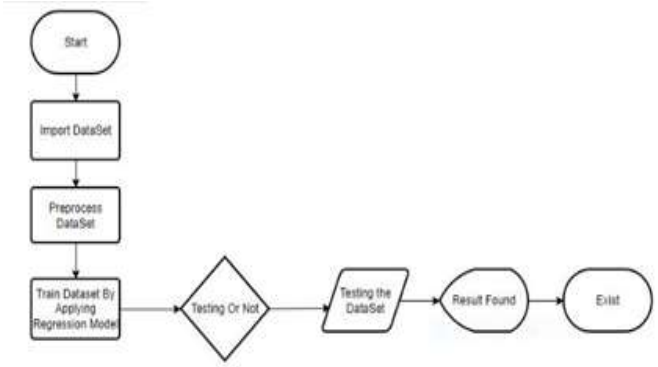

FIGURE 2. Flow Chart of Car Price Prediction

First Phase, as you can see in phase 1, we first obtained the dataset, then found too many outliers, empty fields, and the same value in the column, and finally implemented our algorithm to get a better score for predicting the price, which includes Regression Model for price. And we also compared other algorithms such as Long Short-Term Memory (LSTM), K-Th Nearest Neighbor (mathematics) (KNN), and Logistic Regression.

A logistic regression model always measures the link between a categorical dependent variable and an independent variable, and we chose this approach to get the best results.

Long-short term memory technique, for example, is commonly used in deep learning to categorize and forecast based on time series data and is frequently used for speech recognition and other tasks. Because it generates highly accurate predictions and is a nonparametric method, the k-nearest neighbour may complete the most accurate model.

We obtained accurate results after training and modelling the dataset using the technique described above, and we examined which algorithm was suitable for our project.

Following the implementation of the algorithm for estimating the pricing in a mobile application, we trained the machine using automobile photos to obtain the accuracy of damaged and undamaged cars by uploading an image.

The absence of publicly available datasets of vehicle-broken images with categorization is an essential ingredient though there are no available public datasets. The hardest challenge is building a model with such a small dataset. In this scenario [20], significant progress has been made in learning how and where to address classification tasks when a small dataset is inadequate to train a CNN model. This issue can be addressed using data mining algorithms, which entails proactively gathering and labelling information on the web [21].

We first obtained 400 car photos (damaged and undamaged) from different websites to train the model. Then we import our image folder into the vehicles to train them because machines always recognize angles from three sides (length, width, and height) and in RGB colour to gray colour and images from the X-axis and Y-axis [25]. After that, we used the CNN algorithm (Fig. 3, Fig. 4, and Table I).

FIGURE 3: Damage Cars [25]

FIGURE 4: Not Damage Cars [25]

TABLE I
DAMAGE AND UNDAMAGED CARS DATASET

	Train	**Test**	**Result**
Damage	200 Images	20 images	91.44%
Undamaged	79 images	16 images	92.36%

The CNN technique tried to compare three components of a picture (colour, placement, and colour scheme) to a picture that is always predicated on dots and has an RGB colour scheme (red, green, blue). CNN performs by exporting the following features:

1) It accepts grayscale images as input.
2) And then there's the binary output.
3) Convolution layers, ReLU (amended straight unit) layers, pooling layers, and a covered-up layer comprising of convolution layers, ReLU (amended straight unit) layers, and a covered-up layer comprising of convolution layers, ReLU (corrected straight unit) layers, and a completely associated Neural Organize [10].

An image is composed of the lowest interconnected parts of bitmap graphics, but every dot does have a resilience recognized

as the image pixels. When we investigate a digital photo, it typically has three colour features, such as the Red-Green-Blue channels, known as the "RGB" values. Why is RGB used? Because it's been revealed that trying to combine these three can produce a plausible palette. When we operate with a colour photo, the picture comprises various dots, each with three RGB stream values (see Fig. 5).

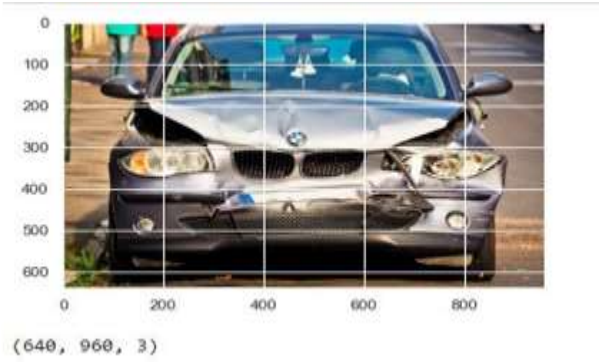

FIGURE 5: RGB Format

Image Shape's output is (640, 960, 3). The picture's sizes seem to be 640 x 960 x 3, with 640 expressing the height, 960 indicating the width, and 3 showing the number of source images (see Fig. 6).

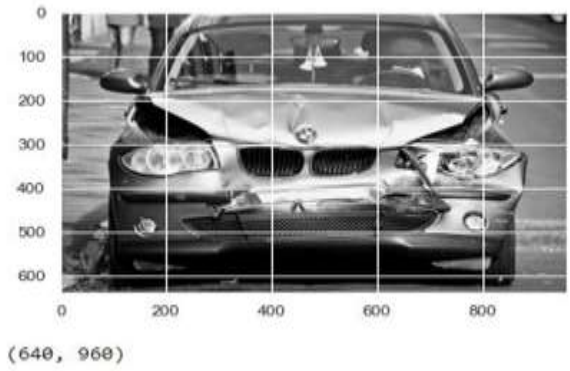

FIGURE 6: Gray yields

Grayshape yields 640 x 960 pictures. What we've had a look at up now is a picture of 614,400 dots, but still only each channel. This is what we see once we attempt to persuade the number of pixels from the grayscale image into a tabular format [23]. We took numerous of those kinds of pictures as well as began to classify them as various individuals by observing the image pixels and finding patterns in them. Because of the background, colour size, and wide variation from frame to frame, it was hard to find shapes by evaluating image pixels. Consequently, we used another advanced technology, such as CNN, to identify these corners or seek the innate behaviour of different functionalities in the car using these labelled images.

A. OBJECTIVES

- Users may be able to predict the price.
- Users may be able to check the accuracy of the damaged and undamaged car.
- Users can save time.
- Users can save money

B. COMPARISON

As applied, three algorithms were to compare a few attributes more commonly seen when buying a car, such as Brand, City, Purchase year, Kilometers driven, Fuel Type, Transmission, Mileage, Engine, Seats, and Owner type [27-30]. As we mentioned in our literature review for this project, many researchers have worked on it. Still, they have applied various algorithms on various attributes, and each of their works and methods to apply the algorithm were different (see Figs. 7, 8, 9 and Tables II, III, IV.

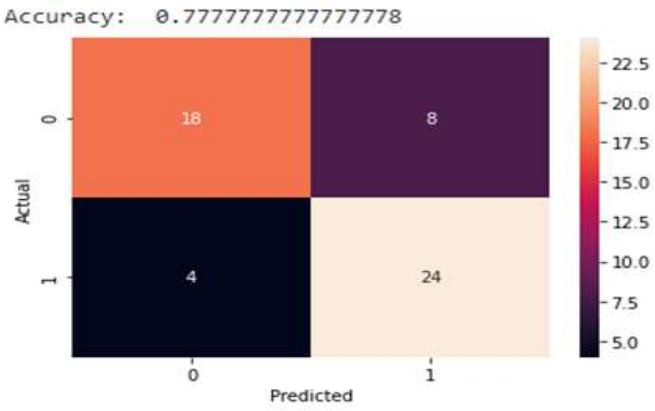

FIGURE 7: Accuracy of Logistic Regression

TABLE II
ACCURACY MODEL I

Accuracy	**Model**
0.77 %	Logistic Regression

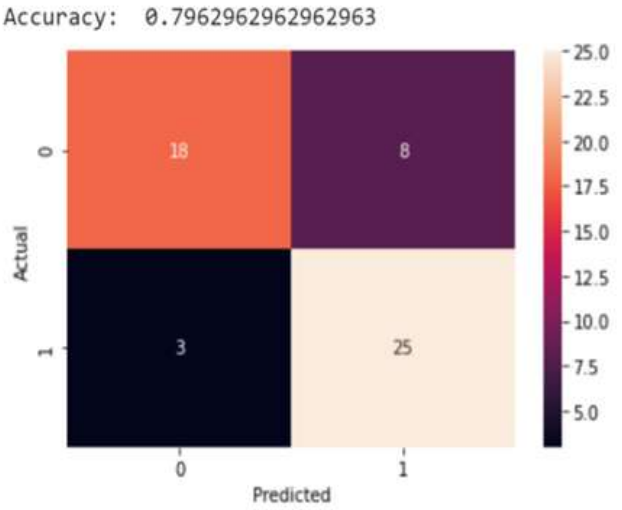

FIGURE 8: Accuracy of KNN algorithm

TABLE III
ACCURACY MODEL II

Accuracy	Model
0.79 %	KNN Algorithm

TABLE IV
ACCURACY MODEL III

Accuracy	Model
0.87 %1	Long short-term memory

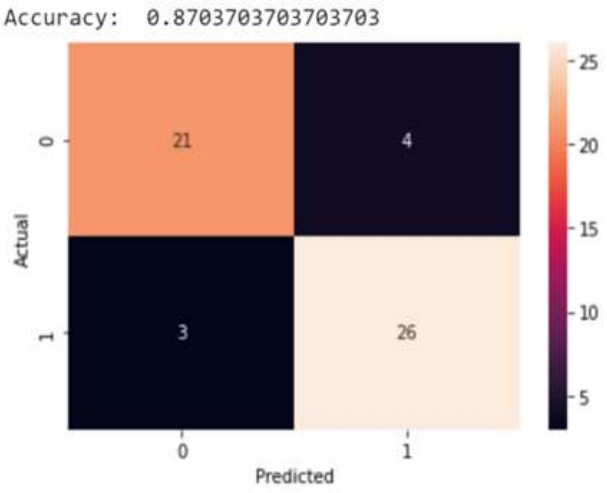

FIGURE 9: Accuracy of Long term-short memory

C. ALGORITHM

The Algorithm 1 pseudocode is as follows:
Input: Train data and Test data (No label);
Output: Prediction results Pr of Test data;
Training ResNet models by using Train data;
Train and Test data results were predicted using the trained ResNet model.
Add the prediction result as a new feature pre_result to the original feature.

min_loss =∞;
index = 1;
while loss < min_loss do:
min_loss = loss;
if index%2 == 1:
Training with XGB;
else:
Training with LGBM;
data;
The pre_result feature is replaced with the current round prediction Pr;
end

Output prediction results Pr;

D. SIMPLE STATISTICS

The basic characteristic properties of the data are shown in Table V. Through the simple statistics and observations of the data, and it is found that some serious missing characteristics need to be processed.

TABLE V
BASIC CHARACTERISTICS OF THE DATA[22]

ID	Features	Type
1	carid	int
2	tradeTime	string
3	brand	int
4	serial	int
5	model	int
6	mileage	float
7	colour	int
8	cityId	int
9	carCode	int
10	transferCount	int
11	seatings	int
12	registerDate	string
13	licenseDat	string
14	country	int
15	maketype	int
16	modelyear	int
17	displacement	float
18	gearbox	int
19	oiltype	int
20	newprice	float
21	anonymousFeature*15	int or string
22	price	float

IV. RESULT AND ANALYSIS

Following our efforts, we applied an algorithm to estimate the price and determine the accuracy of a car by uploading an image. We achieved a better result of the Regression Model Algorithm for forecasting the price of an automobile after applying the algorithm. To assess the accuracy of the car, we trained 90 percent of photographs and tested 10 percent of images by applying the CNN algorithm. Figure 10-15 are the mobile application screenshots, Showing the splash, login, register, and working of the application (see Table VI).

TABLE VI
CARS DATA BY DAMAGE AND UNDAMAGED

Training	Testing	Accuracy	Model
90 %	10%	92.36%	CNN

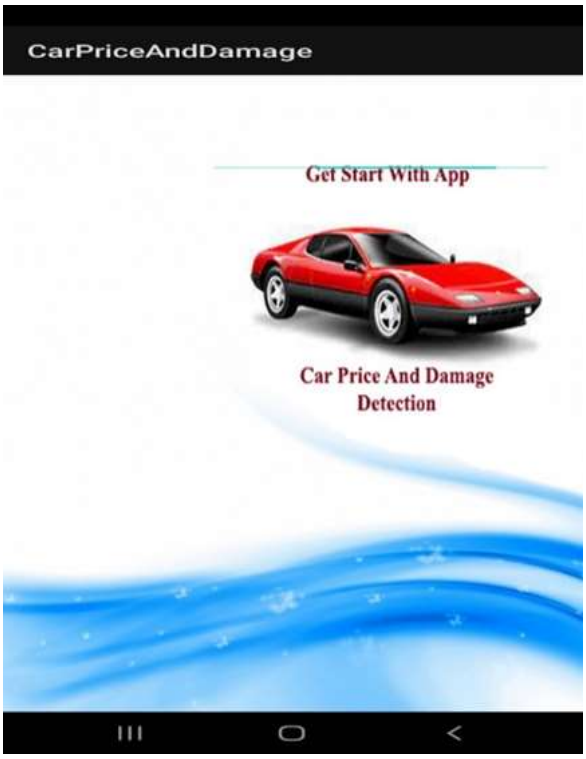

FIGURE 10: First You can see the Splash Screen

FIGURE 12: The registration and login process has connected the database in Firebase. Now the user will visit the home page.

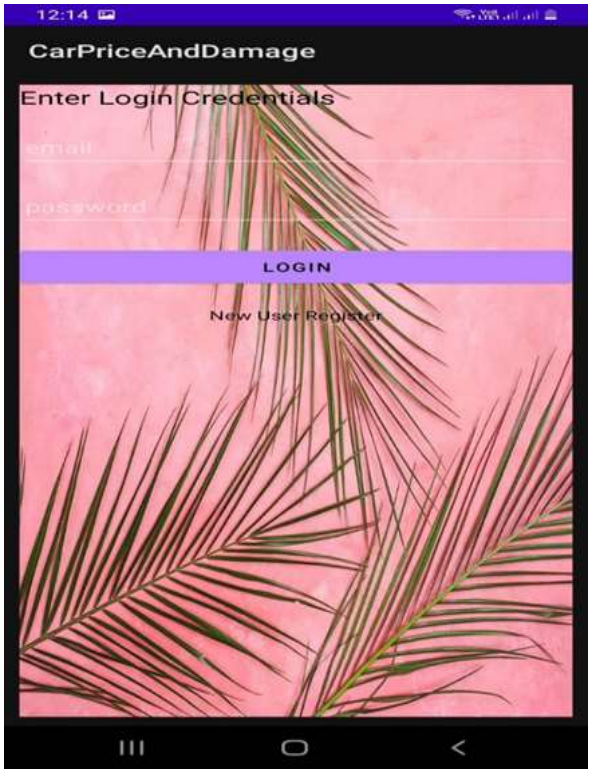

FIGURE 11: Now we have created the Login and Registration page
The user is not already login, so the user must register himself/herself.

FIGURE 13: If the user wants to predict the price, then will select the predict button and if the user wants to predict the dents on the car, then he or she will select the damage to detect button

FIGURE 14: To predict the price of a car user must fill in the car name, company, year, kms_driven, and fuel type and can predict the price of a car.

FIGURE 15: If the user wants to check the damage result of the car, then he/she must take a live picture with a better camera result; otherwise they cannot predict the better accuracy result.

VII. CONCLUSION

We aim to develop this project to predict the price of used cars and to get the accuracy of cars because so many people want to buy used cars due to less budget. When they buy used cars, see a different website or go to the market to buy the car, most people/owners sell the car at costly prices, which is not matched its price, and when people buy a car. Hence, people often see these features such as (*brand, city, purchase year, kilometres driven, fuel type, transmission, mileage, engine, seats, and owner*), and these attributes we have used for our application and applied algorithm. Still, we got the best result using by LSTM algorithm.

FUNDING STATEMENT

The authors declare they have no conflicts of interest to report regarding the present study.

CONFLICT OF INTEREST

The Authors declare that they have no conflicts of interest to report regarding the present study.

REFERENCES

[1] S. Pudaruth, "Predicting the Price of Used Cars using Machine Learning Techniques," International Journal of Information & Computation Technology, vol. 4, 2014.

[2] S. Kuiper, "Introduction to Multiple Regression: How Much Is Your Car Worth?" Journal of Statistics Education, 2008.

[3] P. A. S. S. P. D. S. P. K. Nabarun Pal, "How much is my car worth? A methodology for predicting used car prices using Random Forest," Future of Information and Communications Conference.

[4] B. I. D. K. Z. M. J. K. EnisGegic, "Car Price Prediction using Machine Learning Techniques," TEM Journal, vol. 8, no. 1, pp. 113-118, 2019.

[5] S. T. A. B., R. D., R. S. MehulDholiya, "Automobile Resale System Using Machine Learning," International Research Journal of Engineering and Technology (IRJET), vol. 6, no. 4, 2019.

[6] S. Gonggi, "New model for residual value prediction of used cars based on BP neural network and non-linear curve fit," International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), pp. 682-685, 2011.

[7] M. S. Richardson, "DETERMINANTS OF USED CAR RESALE VALUE," May 2009.

[8] M. Listiani, "Support Vector Regression Analysis for Price Prediction in a Car Leasing Application," March 2009.

[9] [Online].Available: https://www.google.com/search?q=damaged+cars&rlz=1C1BNSD_enPK967PK967&sxsrf=AOaemvL4_S3wNwgfcYoNZHM5j63EDoDlyA:1641571227415&source=lnms&tbm=isch&sa=X&ved=2ahUKEwi1ooivgaD1AhUs5uAKHaVMCbQQ_AUoAXoECAEQAw&biw=1366&bih=625&dpr=1.

[10] [Online]. Available: https://towardsdatascience.com/convolution-neural-network-for-image-processing-using-keras-dc3429056306.

[11] [11] OPinheiro, P. O., Collobert, R., &Dollár, P. (2015). Learning to segment object candidates. Advances in neural information processing systems, 28.

[12] Tang, W., Liu, H. L., Chen, L., Tan, K. C., & Cheung, Y. M. (2020). Fast hypervolume approximation scheme based on a segmentation strategy. Information Sciences, 509, 320-342.

[13] Li, Y., Qi, H., Dai, J., Ji, X., & Wei, Y. (2017). Fully convolutional instance-aware semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2359-2367).

[14] Rong, X., Yi, C., &Tian, Y. Unambiguous scene text segmentation with referring expression comprehension. IEEE Transactions on Image Processing, vol. 29, p. 591-601, 2019.

[15] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask RCNN," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 2980–2988.

[16] Qiao, Y., Truman, M., &Sukkarieh, S. Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming. Computers and Electronics in Agriculture, vol. 165, 104958, 2019.
[17] Lin, X., Zhu, S., Zhang, J., & Liu, D. Rice planthopper image classification method based on transfer learning and mask R-CNN. Trans. Chin. Soc. Agricult. Mach., vol. 13, no. 4, p. 181-184, 2019.
[18] Zhang, Q., Chang, X., & Bian, S. B. Vehicle-damage-detection segmentation algorithm based on improved mask RCNN. IEEE Access, vol. 8, p. 6997-7004, 2020.
[19] Shi, J., Zhou, Y., & Zhang, Q. Service robot item recognition system based on improved Mask RCNN and Kinect. Chinese Journal of Scientific Instrument, vol. 11, no. 9, p. 40-52, 2019.
[20] O. A. Siddiqui, M. K. Shaikh, M. Tahir, A. Urooj, M. A. Khan, and I. U. Rahim, "Atmospheric Environment Monitoring System Based on Wireless Sensor Prototype to Prevent Data Centers", Pakistan Journal of Engineering and Technology (PakJET), vol. 5, no. 2, pp. 79-86, Jun. 2022.
[21] Liu, E.; Li, J.; Zheng, A.; Liu, H.; Jiang, T. Research on the Prediction Model of the Used Car Price in View of the PSO-GRA-BP Neural Network. Sustainability vol. 14, p. 8993, 2022. https://doi.org/10.3390/su14158993
[22] Cui, B.; Ye, Z.; Zhao, H.; Renqing, Z.; Meng, L.; Yang, Y. Used Car Price Prediction Based on the Iterative Framework of XGBoost+LightGBM. Electronics vol. 11, p. 2932, 2022. https://doi.org/10.3390/electronics11182932
[23] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. How transferable are featured in deep neural networks? Advances in neural information processing systems, vol. 27, 2014.
[24] Dhieb, N., Ghazzai, H., Besbes, H., &Massoud, Y. A very deep transfer learning model for vehicle damage detection and localization. In 2019 31st International Conference on Microelectronics (ICM), 2019, (pp. 158-161). IEEE.
[25] Saamiyah Peerun, Nushrah Henna Chummun, and SameerchandPudaruth, University of Mauritius, Reduit, Mauritius. Predicting the Price of Second-hand Cars using Artificial Neural Networks. Proceedings of the Second International Conference on Data Mining, Internet Computing, and Big Data, Reduit, Mauritius 2015.
[26] Roorkee, Roorkee, India), Priya Arora (Department of Computer Science, Texas A & M University Texas, United States), Sai SumanthPalakurthy(Department of Computer Science and Engineering, IIT (ISM) Dhanbad, Dhanbad, India), Dhanasekar Sundararaman (Department of Information Technology.
[27] PattabiramanVenkatasubbu, Mukkesh Ganesh. Used Cars Price Prediction using Supervised Learning Techniques. International Journal of Engineering and Advanced Technology (IJEAT), December 2019
[28] Muhammad Kashif Shaikh; Sellappan Palaniappan; Touraj Khodadadi, An AI-driven automotive smart black box for accident and theft prevention. International Journal of Modelling, Identification, and Control (IJMIC), vol. 39, no. 4, 2021.
[29] Laveena D'Costa, Ashoka Wilson D'Souza, Abhijith K, Deepthi Maria Varghese. Predicting True Value of Used Car using Multiple Linear Regression Model. International Journal of Recent Technology and Engineering (IJRTE). January 2020.
[30] [S.E.Viswapriya, Durbaka Sai Sandeep Sharma, GandavarapuSathyaKiran. Vehicle Price Prediction using SVM Techniques. International Journal of Innovative Technology and Exploring Engineering (IJITEE), June 2020.