Car Price Detection using Machine Learning Techniques

Dr. S. Kayalvili
M.E., M.B.A., Ph.D.,
Department of Computer Science and Engineering
Velalar College of Engineering and Technology
Thindal, Erode - 638012

Hariprasath M.
Department of Computer Science and Engineering
Velalar College of Engineering and Technology
Thindal, Erode - 638012

Harish K. K.
Department of Computer Science and Engineering
Velalar College of Engineering and Technology
Thindal, Erode – 638012

Kavya P.
Department of Computer Science and Engineering
Velalar College of Engineering and Technology
Thindal, Erode – 638012

Abstract-This paper represents a machine learning-based car price detection system that helps the consumer to detect the price of cars. We have used the car price dataset from Gaggle. Features in the dataset that are used for the detection of car price include: Car size, Car width, Engine size, Engine type, Horsepower, etc. We used linear regression and also determine the relationship between price and these features.

Keywords—Car price detection, Machine Learning, Regression.

I. INTRODUCTION

With new automobile models being released each year, the automotive industry has seen substantial expansion and change throughout the years. The cost of an automobile is a significant aspect that affects consumers' decisions. Several elements, like a car's features, mileage, year of production, model, horsepower, fuel type, etc., have an impact on its pricing. Car purchasers want precise information about the cost of automobiles depending on their characteristics to make informed purchases. Several sectors, including the automobile industry, have found applications for machine learning techniques. Machine learning techniques may be applied in the automobile sector to create models that can forecast car pricing depending on their qualities. This can give automobile buyers and sellers important knowledge to help them make wise decisions. A statistical method for determining a link between dependent and independent variables is linear regression. The price of the automobile is the dependent variable in the instance of car price detection, while the attributes of the car are the independent variables. Based on the correlation between the price and the attributes of the automobile, the application of linear regression in the construction of a car price detection model can produce precise predictions.

A. Machine Learning

Computer science and artificial intelligence's branch of machine learning focuses on creating models and algorithms that let computers learn and make judgments or predictions. Large datasets are analyzed to find patterns or insights that may be utilized to make precise predictions or judgments. In supervised learning, the machine learning algorithm attempts to learn how to translate inputs to outputs by being trained on labelled data, when the right response is already known. In unsupervised learning, the algorithm searches for patterns and correlations in the data on its own without being provided any labelled data. Training an agent to operate in a way that maximizes a reward signal is a part of reinforcement learning.

B. Industry of Mobiles

A significant worldwide industry, the automobile sector manufactures a variety of vehicles, including automobiles, trucks, buses, and motorbikes. The sector has experienced tremendous technical developments and advancements over the past several decades, including the creation of electric and autonomous cars as well as a rise in the usage of cutting-edge materials and production processes.

The development of autonomous vehicles is one of the key areas where machine learning is being employed in the automotive sector. Self-driving cars can now make judgments based on real-time data inputs from sensors, cameras, and other sources. For instance, a self-driving car may utilize machine learning to detect and identify items like other cars, people, and traffic signals. Self-driving cars can utilize machine learning to make judgments on how to handle tricky road situations, such merging into highways, changing lanes, and dodging objects. Predictive maintenance is another area in which machine learning is being used in the car industry. To assist save downtime and lower maintenance costs, machine learning algorithms may be used, for instance, to evaluate data from sensors on industrial equipment to spot possible problems. In order to save waste and boost overall efficiency, machine learning may also be utilized to optimize production scheduling and inventory management. In general, the use of machine learning in the automotive sector is fostering important improvements and innovations that will ultimately increase vehicle efficiency, safety, and dependability.

II. PROPOSED ALGORITHM

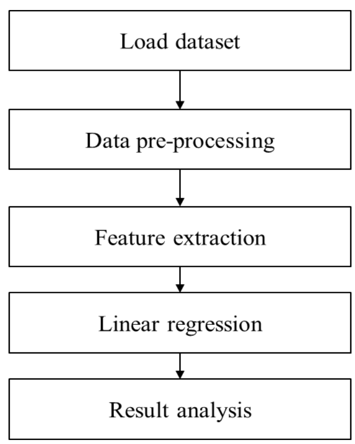

Fig. 1. System Architecture

A. Load input data

Using a library or package like pandas, jumpy, or sickie-learn, you can load data from a file, such a CSV or Excel spreadsheet. These libraries support a variety of file types and include methods for reading and processing the data. The data set was obtained from Gaggle and is intended to be used as input. Creating a dataset of automotive specs and costs is the first step. The brand, model, year, number of doors, engine size, horsepower, and cost of the car should all be included in this information. The process of gathering information on automobile pricing and their characteristics from dependable sources entails loading input data for a machine learning model for the identification of car price. To guarantee the model's accuracy and dependability, the dataset must be representative of the whole automobile population.

B. Data Preprocessing

Before input data is utilized, it must be cleaned, transformed, and normalized as part of the machine learning process. It is crucial because the performance of the model might be adversely affected by noise, outliers, or inconsistencies in the raw data. The dataset is then prepared and cleaned to make sure it can be used to train the model. This could entail categorizing coding, normalizing characteristics, and managing missing values. To prepare the input data for model training, it comprises cleaning and transformation. Throughout the dataset, look for any missing values and decide how to manage them. Rows without values can either be deleted or have an appropriate value, such the mean or median of the column, imputed to them.

C. Feature Extraction

The process of choosing and modifying the input characteristics that will be used to generate predictions is known as feature extraction, and it is a crucial stage in the development of a machine learning model. Finding the characteristics that will have the most influence on the target variable and are the most pertinent and informative is the aim of feature extraction. This is the procedure for selecting the model's most relevant attributes. This can be accomplished using methods like correlation analysis, principal component analysis, or feature significance estimations using decision-tree-based algorithms.

D. Linear regression applied to prediction

A continuous target variable may be predicted using the supervised learning process of linear regression using a collection of input characteristics. Finding a linear relationship between the input characteristics and the target variable is the aim, and predictions about brand-new data are to be made using this relationship. Separate the data into two groups: a training set for the model's training and a testing set for the model's performance assessment. Construct a linear regression model, then fit the test data to it. This entails determining the optimum linear equation coefficients for the training data. Predict the car's price based on its attributes using the trained linear regression model. The anticipated price will be produced as a continuous number by the model. It entails loading the trained model, processing the test data beforehand, forecasting the price, assessing the model's performance, and, if required, improving the model.

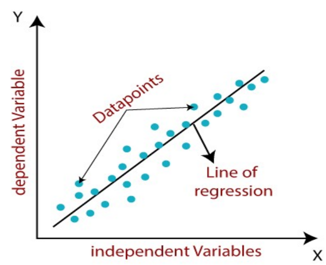

Fig. 2. Graphical representation of linear regression

III. Experiment and Result

The dataset for this experiment is taken from Gaggle. The dataset is split into two halves after the feature extraction. The training set consists of seventy percent of the data and the testing set consists of thirty percent of data. The training set consist of a lot to make sure that the model makes better prediction than the previous model demising auto encoder with convolution operation(DAECO). Linear regression is applied as the final part and the accuracy of the training and testing set are predicted. A graph is plotted for visualization.

```
#Defined X value and y value , and split the data train
X = data1.drop(columns="price")
y = data1["price"]    # y = price

# split the data train and test
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.70)

print("X Train : ", X_train.shape)
print("X Test  : ", X_test.shape)
print("Y Train : ", y_train.shape)
print("Y Test  : ", y_test.shape)
```

```
X Train :  (143, 13)
X Test  :  (62, 13)
Y Train :  (143,)
Y Test  :  (62,)
```

Fig. 3. Splitting the dataset

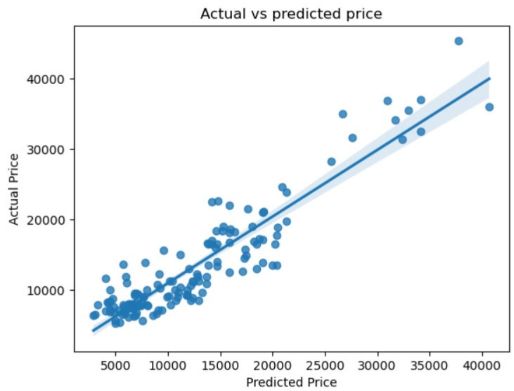

Fig. 4. Graph visualization

A graph is plotted between the actual price and the predicted price for both the testing and the training data. R squared error is calculated to analyze the model's efficiency. The accuracy of the training and testing set are calculated. The training dataset gives the accuracy around 85% and the testing set has the accuracy around 83%. The speed and accuracy of the model is compared with the previous model. The DAECO model requires larger dataset for prediction. The Linear regression model can perform well on smaller datasets and speed of execution is compared to the previous model.

Building a car price detection model using machine learning methods and linear regression requires careful consideration of the results. To make sure the model is accurate and dependable in forecasting automobile prices, it is crucial to assess its performance on test data once the model has been trained. To make sure the model is accurate and dependable in forecasting automobile prices once it has been refined, it is essential to validate it using fresh data. Building a car price identification model using machine learning methods and linear regression requires a vital step called result analysis. It entails assessing the model's effectiveness, identifying potential areas for improvement, improving the model, validating the model, and informing stakeholders of the findings.

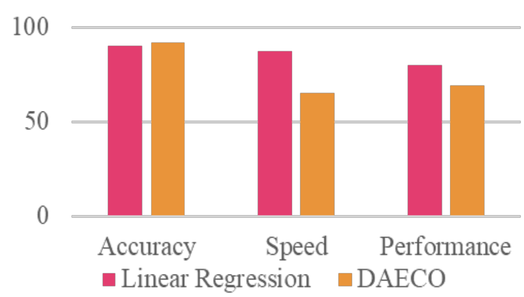

Fig. 3. Graph comparison

IV. Conclusion

In conclusion, developing a machine learning-based linear regression model for automobile pricing identification is a challenging but crucial procedure. The model depends on reliable input data, efficient feature extraction methods, efficient pre-processing methods, and a trained linear regression model. To make sure the model is accurate and dependable in forecasting automobile prices, the performance must be examined and improved. It is feasible to develop an accurate and dependable automobile price detection model that may assist stakeholders in making decisions based on actual data, despite the fact that linear regression is a common and efficient machine learning approach. However, it is essential to keep in mind that the application of machine learning algorithms to the predictions of car price has some drawbacks. It is necessary to carefully consider and address potential biases in the data that is utilized to train the algorithms.

Acknowledgments

We would acknowledge our guide for development of the Car Price Detection model with full support and guidance. We would like to thank the user who contribute their data to study.

REFERENCES

[1] Kumar, B. and Sarkar, P. (2016) "Prediction of future car forms based on historical trends," Perspect. Sci., vol. 8, pp. 764–766.

[2] Lessmann, S. and Voß, S. (2017) "Car resale price forecasting: The impact of regression method, private information, and heterogeneity on forecast accuracy," Int. J. Forecasting, vol. 33, no. 4, pp. 864–877.

[3] Nagarajan C., Neelakrishnan G., Akila P., Fathima U., Sneha S. "Performance Analysis and Implementation of 89C51 Controller Based Solar Tracking System with Boost Converter" Journal of VLSI Design Tools & Technology. 2022; 12(2): 34–41p.

[4] C. Nagarajan, G.Neelakrishnan, R. Janani, S.Maithili, G. Ramya "Investigation on Fault Analysis for Power Transformers Using Adaptive Differential Relay" Asian Journal of Electrical Science, Vol.11 No.1, pp: 1-8, 2022

[5] Noor, K. and Jan, S. (2017) "Vehicle price prediction system using machine learning techniques," Int. J. Comput. Appl., vol. 167, no. 9, pp. 27–31.

[6] Pal, N., Arora, P., Kohli, P., Sundararaman, D. and Palakurthy, S. S. (2018) "How much is my car worth? A methodology for predicting used cars' prices using random forest," in Proc. Future Inf. Commun. Conf. Singapore: Springer, pp. 413–422.

[7] Samruddhi, K. and Ashok Kumar, R. (2020) "Used Car Price Prediction using K-Nearest Neighbor Based Model", International Journal of Innovative Research in Applied Sciences and Engineering, vol. 4, no. 3, pp. 686-689

[8] G.Neelakrishnan, K.Anandhakumar, A.Prathap, S.Prakash "Performance Estimation of cascaded h-bridge MLI for HEV using SVPWM" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:750-756

[9] G.Neelakrishnan, S.N.Pruthika, P.T.Shalini, S.Soniya, "Perfromance Investigation of T-Source Inverter fed with Solar Cell" Suraj Punj Journal for Multidisciplinary Research, 2021, Volume 11, Issue 4, pp:744-749

[10] C.Nagarajan and M.Madheswaran - 'Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques'- Taylor & Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011.

[11] Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.

[12]

[13] M.Kannan, R.Srinivasan and G.Neelakrishnan, "A Cascaded Multilevel H-Bridge Inverter for Electric Vehicles with Low Harmonic Distortion", International Journal of Advanced Engineering Research and Science, November 2014; 1(6): 48-52.

[14] G.Neelakrishnan, M.Kannan, S.Selvaraju, K.Vijayraj, M.Balaji and D.Kalidass, "Transformer Less Boost DC-DC Converter with Photovoltaic Array", IOSR Journal of Engineering, October 2013; 3(10): 30-36.

[15] M.Kannan, R.Srinivasan and G.Neelakrishnan, "A Cascaded Multilevel H-Bridge Inverter for Electric Vehicles with Low Harmonic Distortion", International Journal of Advanced Engineering Research and Science, November 2014; 1(6): 48-52.

[16] G.Neelakrishnan, M.Kannan, S.Selvaraju, K.Vijayraj, M.Balaji and D.Kalidass, "Transformer Less Boost DC-DC Converter with Photovoltaic Array", IOSR Journal of Engineering, October 2013; 3(10): 30-36.

[17] Wu, J. D., Hsu, C. C. and Chen, H. C. (2009) "An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference," Expert Syst. Appl., vol. 36, no. 4, pp. 7809–7817.

[18] Xiaoyun Yue, Ling Xue, and Tianhao Ji (2022), "A Novel Used Vehicles Price Prediction Model Based on Denoising Autoencoder With convolution operation", IEEE trans. on intelligent transportation systems.