HARNESSING ARTIFICIAL NEURAL NETWORKS AND MACHINE LEARNING FOR USED CAR PRICE PREDICTION

Dr.O.Sampath, Associate Professor
Appaiah gari Anusha,MCA Student
Department of Master of Computer application

Rajeev Gandhi Memorial College of Engineering and Technology, Nandyal, 518501, Andhra Pradesh, India.

ABSTRACT

Predicting automobile prices has drawn a lot of attention in research since it takes observable effort and subject-matter expertise. A large variety of unique characteristics are looked at in order to make a prediction that is accurate and trustworthy. To develop a model for used automobile pricing prediction. We used machine learning techniques, including Random Forest, Support Vector Machine, Logistic Regression, and Linear Regression. Nonetheless, the aforementioned methods were used for regression analysis and group work. The information gathered for the forecast was The optimal algorithm for the given data set was then determined by comparing the respective performances of the several methods. We must forecast the optimal % with the aid of algorithms and also disclose the results using the appropriate techniques.

Keywords: machine learning, classification, support vector machines, and car price prediction.

I.INTRODUCTION

Predicting car prices is a popular and somewhat fascinating issue. Eighty-four percent of the vehicles registered in 2014 were for personal use, according to information obtained from the Agency.

Since 2013, this number has climbed by 2.7%, and it is expected that this trend will continue and that there will be an increase in cars in the future.

This gives the challenge of predicting car prices even more importance. Because the price of a car typically depends on a number of unique features and elements, accurate car price prediction requires specialist expertise. The most important ones are usually the age, horsepower, mileage, and brand and model.

Because gasoline prices fluctuate frequently, the type of fuel a car runs on as well as its fuel consumption per mile have a significant impact on its pricing.

The price of a car will also be influenced by several aspects, such as the outside color, door number, type of transmission, dimensions, safety, air conditioning, interior, and whether or not it includes navigation. In order to increase the used car price prediction's precision, we used a variety of strategies and tactics in this research.

II. LITERATURE SURVEY

Using historical data from daily newspapers, the first study investigates the use of supervised machine learning techniques to predict used automobile prices in Mauritius.

The predictions are created using a variety of methods, including decision trees, k-nearest neighbours, naïve Bayes, and multiple linear regression analysis.

The ensuing study aims to forecast used automobile prices in Bosnia and Herzegovina by examining a large number of different characteristics. In order to create a trustworthy model for price prediction, the study uses three different machine learning techniques: Random Forest, Support Vector Machine, and Artificial Neural Network.

The project concludes with a novel strategy that suggests a BP neural network-based price evaluation model for the secondary auto market.

The model makes use of a large amount of big data analysis performed on a large number of transaction records and widely distributed vehicle data.

It applies an optimized BP neural network technique to assess price data across numerous vehicle kinds, with the purpose of building an appropriate model for secondhand car price appraisal adapted to particular automobiles.

III. EXISTING SYSTEM:

It's difficult to estimate an automobile's resale value. It goes without saying that a variety of factors affect a used car's value.

The most crucial ones are typically the vehicle's age, make (and model), origin (the nation in which the manufacturer originally located the vehicle), mileage (the total number of kilometers driven), and horsepower. gasoline economy is particularly crucial because gasoline prices are rising. Regretfully, in real life, most people are unaware of the precise amount of fuel their vehicle uses each kilometer of travel.

The price may also be affected by other elements like the fuel type, interior design, braking system, acceleration, cylinder volume (measured in cc), safety index, size, number of doors, paint color, weight of the vehicle, customer reviews, prestigious awards the car manufacturer has won, physical condition, whether the vehicle is a sports car, whether cruise control is installed, whether it is an automatic or manual transmission, and whether it belonged to a person or a business.

Some unique considerations that purchasers in Mauritius place a premium on include the previous owner's residence, whether the vehicle has been in any major collisions, and whether the driver is a woman. The price of the car is undoubtedly greatly influenced by its appearance and feel. As we can see, a lot of things influence the pricing.

Unfortunately, it's not always possible to have knowledge about all of these criteria, so the buyer has to base his or her decision to buy at a certain price on a limited number of considerations.

Disadvantages:

Dependency on Data: These systems are highly dependent on large, high-quality datasets for training, which can be costly or challenging to acquire. Predictions that are not accurate can result from biased or small datasets.

Overfitting: When neural networks overfit to training data, they have a hard time generalizing to new input. This happens when the model becomes adept at memorization of the training data instead of identifying underlying patterns.

IV. PROPOSED SYSTEM:

Numerous studies have looked closely at predicting the price of used cars. The price of a leased car can be predicted more accurately by a regression model developed using machine learning techniques than by multivariate or even basic multiple regression.

This is justified by the fact that machine learning algorithms are less likely to overfit or underfit datasets and perform better when working with datasets having multiple dimensions. This study's flaw is that fundamental metrics like mean, variance, and standard deviation did not display a comparison between basic and more sophisticated regression.

In his thesis, Richardson took a different tack [3]. According to his theory, automakers make vehicles that are more resilient. Using multivariate regression analysis, Richardson showed that hybrid automobiles outlive conventional cars in terms of market value. This results in increased fuel efficiency and has its roots in environmental worries about the climate.

carried out a study on car price prediction using a knowledge-based neuro-fuzzy system. The brand, the year of manufacturing, and the kind of engine were all taken into account. Similar outcomes were obtained by their prediction model and the basic regression model. Additionally, they developed an expert system called ODAV (Optimal Distribution of Auction Vehicles) because car dealers have a strong desire to sell the vehicles at the conclusion of the leasing term.

This system provides information on the place where the best deal may be found as well as the best prices for cars. The price of an automobile was predicted using a regression model based on the k-nearest neighbor machine learning technique. With almost two million cars exchanged through it, this system has the potential to be incredibly successful. It offers a model for used car price prediction that is constructed using artificial neural networks, or ANNs. He took into account a number of factors, including brand, expected car life, and miles driven. In contrast to earlier models that relied solely on basic linear regression techniques, the suggested model was designed to handle nonlinear relations in data.

Compared to other linear models, the non-linear model performed better in predicting car prices.

Moreover, many machine learning methods were utilized for car price prediction in Mauritius, including k-nearest neighbours, multiple linear regression analysis, decision trees, and naïve bayes.

The dataset used to create a prediction model was collected manually from local newspapers in period less than one month, as time can have a noticeable impact on price of the car. He studied the following attributes: brand, model, cubic capacity, mileage in kilometers, production year, exterior colour, transmission type and price. However, the author found out that Naive Bayes and Decision Tree were unable to predict and classify numeric values. Additionally, limited number of dataset instances could not give high classification performances, i.e. accuracies less than 70%.

Noor and Jan [8] build a model for car price prediction by using multiple linear regression. The price, cubic capacity, exterior color, date the ad was posted, number of ad views, power steering, mileage in kilometers, type of rims, type of transmission, engine type, city, registered city, model, version, make, and model year were all included in the dataset that was created during the two-month period.

After applying feature selection, the authors considered only engine type, price, model year and model as input features. With the given setup authors were able to achieve prediction accuracy of 98%. In the related work shown above, authors proposed prediction model based on the single machine learning algorithm. However, it is noticeable that single machine learning algorithm approach did not give remarkable prediction results and could be enhanced by assembling various machine learning methods in an ensemble.

We applied a number of traditional and cutting-edge approaches, such as ensemble learning strategies, using 90% of the training data and 10% of the test data. We used 500 thousand instances from our dataset in order to shorten the training time. Our baseline techniques were Gradient Boost, Random Forest, and Linear Regression. Utilizing the open-source Scikit-Learn software, the majority of the model implementations were performed.

Linear Regression: Because of its ease of use and short training period, Linear Regression was selected as the initial model. The features were used as the feature vectors directly, without the need for feature mapping. Regularization was not applied because the data were obviously low variance.

Random Forest: Random Forest is a regression model based on ensemble learning. It creates the ensemble model, which generates a prediction collectively, using a decision tree model, more precisely, several decision trees, as the name implies. This model's advantage is that the trees are generated in parallel and with no correlation, which leads to high-quality results because no single tree is susceptible to the individual errors of other trees.

The use of bagging or bootstrap aggregation, which provide the necessary randomness to create resilient and uncorrelated trees, helps to partially ensure this uncorrelated behavior. Because of this, this model was selected to compare a bagging strategy with the subsequent gradient boosting methods and account for the enormous number of features in the dataset.

IV. ALGORITHM

Gradient Boost :

Another decision tree-based technique called "a method of transforming weak learners into strong learners" is gradient boosting. This means that, similar to a standard boosting strategy, observations are given varying weights. The weights of observations that are hard to forecast are subsequently increased based on specific criteria, and the resulting data is sent into another tree for training. The gradient of the loss function serves as the metric in this instance. Because of the non-linear correlations between the features and the expected price, this model was selected.

KMeans :

The Linear Regression An ensemble approach that employed KMeans clustering of the features and linear regression on each cluster was utilized in order to take advantage of the linear regression results and the apparent categorical linearity in the data as shown in. A three-cluster approach was employed because of the lengthy training period. After the dataset was divided into these three clusters, a linear regressor that had been trained on each of the three training sets was used.

Advantages:

Accuracy: Compared to conventional regression models, neural networks can identify intricate patterns in data, producing predictions that are more accurate.

Adaptability: As new data becomes available, machine learning models can adjust to it, enabling the system to continuously improve its predictions.

Efficiency: The system can expedite the used automobile pricing process by rapidly analyzing massive datasets and making predictions in real-time once it has been trained.

Feature Extraction: Neural networks have the ability to automatically extract pertinent features from data, eliminating the need for human feature engineering and possibly capturing subtle associations that people could miss. This process is known as feature extraction.

Scalability: The system's ability to manage a sizable amount of data makes it appropriate for price prediction across a range of marketplaces and automobile types.

Decreased Bias: By using data-driven algorithms, the system can lessen biases that could be present in conventional pricing techniques, resulting in more fair and impartial pricing.

Integration: By integrating the technology with already-existing platforms or services, customers would have easy access to precise used car price estimates.

V. SYSTEM DESIGN AND MODEL

The design phase's goal is to plan a solution to the issue that the requirement document outlines. It is the process of defining software objects, methods, and functions as well as the general structure and interaction of your code to ensure that the functionality that is produced will meet the needs of your consumers. It enables you to do the finest abstraction, comprehend the needs more fully, and satisfy them more effectively. Reusability is increased and redundancy is avoided as a result.

Transitioning from the problem domain to the solution domain begins with this phase. To put it another way, designing from the needs leads to the methods of meeting those demands. Perhaps the most important aspect influencing software quality is the system's design, which has a big influence on later stages, especially testing and maintenance.

The design document is the result of this step. This document, which is utilized subsequently for implementation, testing, and maintenance, resembles a blueprint for the solution. System Design and Detailed Design are the two distinct phases into which the design process is often divided. The goal of system design, also known as top-level design sign, is to specify which modules should be in the system, how they should interact with one another to achieve the intended outcomes, and which modules should be included. The internal logic of each module specification in the system design is chosen during detailed design.

In this stage, the data specifications are typically made in a high-level design description language that is separate from the target language that will be used to implement the software in the end. While creating the logic for each module is the main goal of detailed design, system design is more concerned with identifying the modules.

Developers fill in the gaps between the requirements specification generated by requirements elicitation and analysis and the system that is provided to the user during the system design operations.

SYSTEM MODEL

The world will run on artificial intelligence in the future. It is growing quickly across all industry verticals. Artificial intelligence has a bright future as a result. AI is a technique that enables machines to mimic human behavior. Artificial Intelligence is the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making and translation between languages. AI is the simulation of human intelligence done by machines programmed by us. The machines need to learn how to reason and do some self-correction as needed along the way.

The process of creating artificial intelligence involves first understanding how the human brain functions to solve problems and then applying the knowledge gained to create intelligent software and systems.

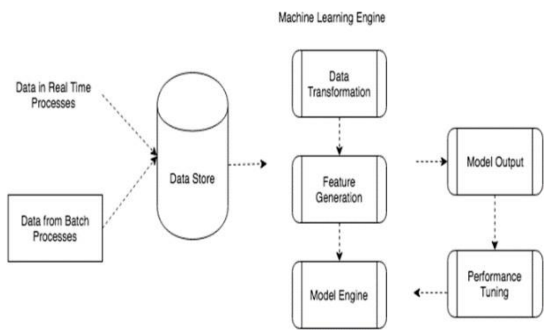
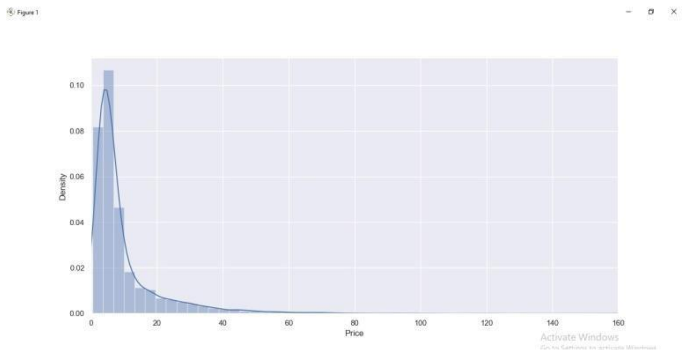
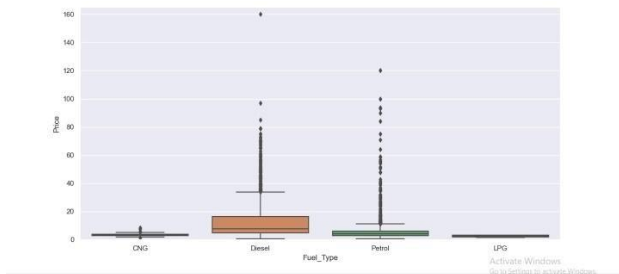
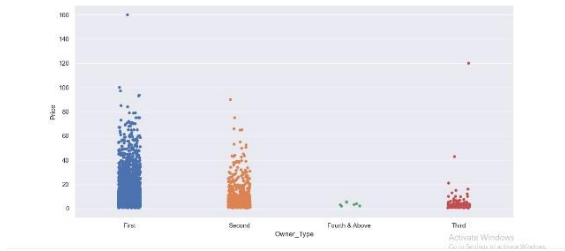


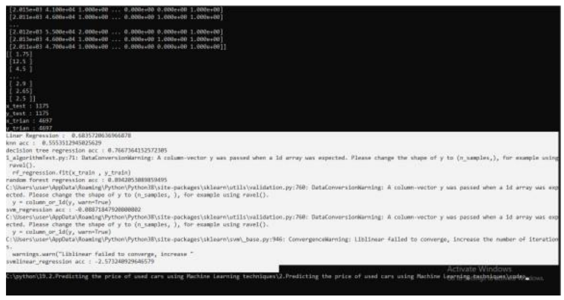
Fig:- Block diagram flow of architecture for Machine learning systems

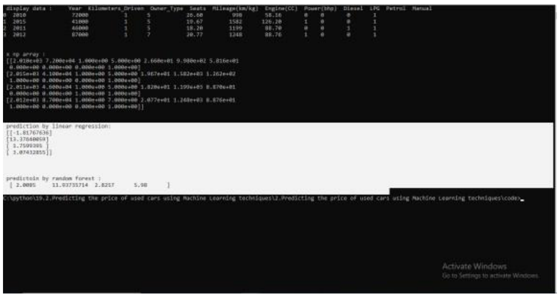
VI. SNAP SHOTS











VII. CONCLUSION

Predicting the price of a car can be difficult because there are so many factors to take into account in order to make an accurate estimate. The gathering and preliminary analysis of the data is a crucial phase in the prediction process.

Standardize and purify the data in this study to keep machine learning algorithms free of needless noise. Although one method that improves prediction performance is data cleansing, it is insufficient when dealing with complicated data sets like the one used in this study.

The accuracy of using a single machine algorithm on the data set was less than half. As a result, more algorithms must be used in order to anticipate a higher proportion. When compared to the single machine learning method approach, this represents a huge improvement. The suggested system's disadvantage is that it uses a lot more computer power than a single machine learning algorithm.

Despite the system's astounding performance in the automobile price prediction challenge, our goal for future study is to see if it can also function well with different kinds of data.

FEATURE ENHANCEMENT

Our machine learning algorithm may eventually be coupled with numerous websites that offer real-time pricing prediction data. To increase the machine learning model's accuracy, we might also make use of a sizable amount of historical data on car prices.

An Android application may serve as a UI for interacting with others. We aim to carefully construct deep learning network designs, use variable learning rates, and train on data clusters rather than the entire dataset in order to increase performance.

REFERENCES

[1] 2015. National Transportation Authority [1]. At http://nta.govmu.org/English/Statistics/Pages/Arch ives.aspx, it is accessible. [As of April 24, 2015]

[2] Reference: Bharambe, M. M. P. and Dharmadhikari, S. C. (2015) "An Artificial Neural Network-Based Approach to Stock Market Analysis Using Large Data." Fourth Post Graduate Conference, Pune, India, March 24, 25.

[3] "Predicting the Price of Used Cars using Machine Learning Techniques," Pudaruth, S. (2014). Volume 4, Issue 7, pages 753–764, International Journal of Information & Computation Technology.

[4] Jassibi, J., Alborzi, M. and Ghoreshi, F. (2011) “Car Paint Thickness Control using Artificial Neural Network and Regression Method”. Journal of Industrial Engineering International, Vol. 7, No. 14, pp. 1-6, November 2010

[5] Ahangar, R. G., Mahmood and Y., Hassen P.M. (2010) “The Comparison of Methods, Artificial Neural Network with Linear Regression using Specific Variables for Prediction Stock Prices in Tehran Stock Exchange”. International Journal of Computer Science and Information Security, Vol.7, No. 2, pp. 38-46.

[6] Listiani, M. (2009) “Support Vector Regression Analysis for Price Prediction in a Car Leasing Application”. Thesis (MSc). Hamburg University of Technology.

[7] Iseri, A. and Karlik, B. (2009) “An Artificial Neural Network Approach on Automobile Pricing”. Expert Systems with Application: ScienceDirect Journal of Informatics, Vol. 36, pp. 155-2160, March 2009.

[8] Yeo, C. A. (2009) “Neural Networks for Automobile Insurance Pricing”.

Encyclopedia of Information Science and Technology, 2nd Edition, pp. 2794-2800, Australia.

[9] Doganis, P., Alexandridis, A., Patrinos, P. and Sarimveis, H. (2006) "Time Series Sales Forecasting for Short Shelf-life Food Products Based on Artificial Neural Networks and Evolutionary Computing". Journal of Food Engineering, Vol. 75, pp. 196–204.