IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

CAR PRICE PREDICTION

[1]Aditya Arora, [2]Akriti Singh, [3]Aman Goel, [4]Kirti Kushwah
[1]Student, [2]Student, [3]Student, [4]Professor
Computer Science and Engineering,
[1]Inderprastha Engineering College, Ghaziabad, India

Abstract: over 70 million passenger automobiles were created in 2016. The number of cars produced has been rising over the last ten years. The secondhand automobile market has emerged because of this, and on its own has grown to be a prosperous sector. The emergence of online portals has made it easier for both buyers and sellers to learn more about the patterns and trends that influence a used car's market value. Our goal is to create a statistical model that can forecast the price of a used car by utilizing machine learning algorithms like regression trees, multiple regression, and Lasso regression and many more. This model will be based on past consumer data and a predetermined set of features. We intend to additionally contrasting these models' prediction accuracy to identify the best one. In the industry, the manufacturer sets the price of a new car, with the government bearing some additional expenses in the form of taxes. Customers who purchase new cars can therefore be sure that their financial investment is worthwhile. However, sales of used automobiles are rising globally because of rising new car prices and consumers' inability to afford them. A used car price prediction system that accurately assesses the car's worthiness based on a range of factors is therefore desperately needed. The current system has a procedure where a vendor chooses a price at random, and a buyer has no notion what the car is worth in the current market. Neither the car's current worth nor the asking price are known to the seller. We have created a model that will work incredibly well to solve this issue. The reason machine learning algorithms are chosen is that their output is continuous rather than categorized. This makes it feasible to forecast an automobile's true cost rather than just its pricing range. Additionally, a user interface has been created that takes input from any user and displays the price of a car based on that input.

Keywords—Car Price, ML Algorithm

Index Terms **– Car price prediction, ML Model, Supervised Model**

I. INTRODUCTION

There are so many variables that influence a used car's pricing on the market, it can be difficult to determine whether the quoted price is accurate. This project's main goal is to create machine learning models that can effectively estimate a used car's price based on its attributes, enabling them to make well-informed decisions. We use and assess learning methodologies on a dataset comprising the selling prices of various models and manufactures. We assess how well machine learning algorithms perform. Regression in Linear Form. Various factors will be taken into consideration while determining the car's pricing. Regression The reason algorithms are utilized is that their output is a continuous value rather than a categorized value, which makes it easy to estimate the true cost of an automobile rather than just the range of prices. Additionally, a user interface hat gathers input from users and shows car prices based on input from users has been built. The market for old cars is expanding rapidly; in the previous several years, its market value has nearly doubled. The rise of internet portals like CarDekho, Quicker, Carwale, Cars24, and numerous others has made it easier for buyers and sellers to learn more about the patterns and trends that influence a used car's market value. Based on a certain set of features, machine learning algorithms can be used to anticipate an automobile's retail value. Various websites There isn't a single algorithm utilized to determine the pricing because different companies use different algorithms to create the retail price of used cars. Without entering the details into the desired website, one can easily get a reasonable estimate of the price by training statistical models for price prediction. Kaggle generated the data set that was utilized in the prediction models [1]. 9104 automobile records with computed retail prices

are included in the data. The variables that are used are as follows:

- Cost: The GM vehicles computed retail cost.
- Mileage: The total kilometers driven by the vehicle driven.
- Model: The models for each automobile.
- Fuel: The kind of fuel the car runs on, such as petrol or diesel.
- Year: The year the actual owner of the car purchased it.

II. TECHNICAL ARCHITECTURE OF CAR PRICE PREDICTION

What makes up an automobile price prediction system's technical architecture is Dataset creation comes after gathering information from multiple sources. This data is used to feed machine learning models, which use algorithms learned on processed data, such as Random Forest, Decision Tree. The trained model is made available via an API or web service that is housed on a server or cloud platform and can be accessed via an intuitive frontend interface that allows users to enterinformation about their cars and get price estimates. Ongoing improvements are facilitated via user input systems, model performance review, and continuous monitoring. Ensuring scalability, user privacy, and data security, the system combines multiple components to produce an accurate, efficient, and user-friendly car price forecast tool

III. OBJECTIVE OF THE CAR PRICE PREDICTION

Creating a machine learning model that can reliably forecast an automobile's price is the main goal of a project that aims to predict car prices. on a range of characteristics and elements. With the help of this initiative, buyers, sellers, and other industry participants should be able to make more informed judgements about car price

1. Prediction accuracy create a model that can reliably estimate automobile prices based on many features such as brand, model, mileage, year, condition, etc.
2. Assistance in Making Decisions: Help vendors set competitive prices for their listings and buyers in determining fair prices for cars They are interested in buying.
3. Efficiency: Save time by providing a prompt estimate of car costs. in contrast to manual appraisal procedures.
4. Insight Generation: Learn about the variables that have a big impact on auto costs and comprehend market dynamics to make smarter decisions.
5. Scalability: Create a model that is flexible and scalable for various automotive industry conditions, enabling it to be utilized for a broad variety of automobile listings. The goal is to use machine learning techniques to develop a trustworthy tool that, by projecting fair values based on numerous variables, improves comprehension and efficiency of the car-buying and selling process

IV. SCOPE OF CAR PRICE PREDICTION

This machine learning model might eventually link to different websites that offer real-time data for price prediction. Additionally, we might include a significant amount of historical car price data to help the machine learning model become more accurate. As a user interface, we can create an Android app to communicate with users. We intend to carefully consider the architecture of deep learning networks, employ adaptive learning rates, and train on data clusters in order to improve performance. Instead of the entire datasets.

V. NOVELTY OF THE CAR PRICE PREDICTION

The automobile price prediction project is interesting because it creatively uses data analytics and machine learning to anticipate automotive pricing with precision. This project utilizes advanced algorithms to provide real-time estimations for a variety of car variables, including brand, model, miles, year, and condition. The algorithms are trained on large datasets to accommodate the dynamic nature of the automotive industry. Adaptability is increased by its ability to tailor models to different automotive types and local market preferences. Accuracy is also improved and new insights into changing market trends are fostered by ongoing refining through the integration of current data. This project is noteworthy for its capacity to provide insightful and predictive data that is based on data, so enabling giving stakeholders access to a user-centric tool that improves decision-making inthe fiercely competitive and dynamic automotive sector

VI. LITERATURE REVIEW

We go over several applications and techniques in this chapter that gave us the idea to develop our project. We conducted a background investigation on the fundamental concepts of our project and used those concepts to gather data on the technology stack, algorithms, and weaknesses of our project, which helped us to improve it.

GET VEHICLE PRICE.

Get Vehicle Price is an android app which works on similar parameters like Cars24. This app predicts vehicle prices on various parameters like Fiscal power, horsepower, kilometers traveled. This app uses a machine learning approach to predict the price of a car, bike, electric vehicle and hybrid vehicle. This app can predict the price of any vehicle because of the smartly optimized Algorithm.

CARWALE

Car Wale app is one of the top-rated car apps in India for new and used car research. It provides accurate on-road prices of cars, genuine user and expert reviews. It can also compare different cars with the car comparison tool. This app also helps you to connect with your nearest car dealers for the best offers available.

CARTRADE

Car Trade is web and Android platform where user can research New Cars in India by exploring Car Prices, Car Specs, Images, Mileage, Reviews, and Car Comparisons. On this app one can Sell Used Car to genuine buyers with ease. One can list their used car for sale along with the details like image, model, and year of purchase and kilometers so that it is displayed to lakhs of interested car buyers in their city. User can read user reviews and expert car reviews with images that help in finalizing a new car buying decision

VII. TECHNOLOGY USED

Because Python comes with a wealth of built-in methods in the form of bundled libraries, it was the primary technology utilized to apply machine learning ideas. These are well-known libraries and technologies that we used for our project

NUMPY

A general-purpose array processing package is called NumPy [1]. It offers tools for manipulating these arrays as well as a high- performance multidimensional array object. This is the core Python module for scientific computing. In addition to its apparent applications in science, NumPy can be used as a productive multidimensional data container for general purposes. NumPy can create any datatypes, which makes it possible for NumPy to quickly and easily integrate with a wide range of databases.

Flask

Flask is a lightweight Python web framework that makes it simple and rapid to create web applications. Flask can be used to build a web application for the car price prediction project, allowing users to enter an automobile's details (make, model, year, mileage, etc.) and receive an estimated price. Flask offers request processing, response generation, and routing features to specify URL endpoints. To display the prediction results, you can render HTML templates using Flask's templating engine and combine them with your Python code.

Pickle

A Python package called PICKLE is used to serialize and deserialize Python objects. It enables the conversion of complicated objects into byte streams that may be written to files or sent over networks, including lists, dictionaries, and custom classes. Preprocessed data, machine learning models, and other Python objects that must be maintained between sessions are frequently saved in pickles

Kaggle

Our study makes use of an extensive dataset that we obtained from Kaggle, a well-known machine learning and data science platform. This dataset includes a wide range of characteristics associated with different car models, such as the make, model year, mileage, fuel type, engine specs, and geographic location. This dataset, which has more than [insert number] items, offers a wealth of data for developing and accessing prediction models for estimating car prices. Robust analysis and modeling are made possible by the inclusion of both numerical and categorical variables, which let us find complex patterns and correlations that affect the dynamics of car prices. Our goal is to create a car price prediction model that is accurate and dependable using this Kaggle dataset, so that industry stakeholders and customers may make well-informed decisions

JUPYTER NOTEBOOK

You may create and share documents with live code, equations, visualizations, and narrative text using the open-source webapplication Jupyter Notebook. Along with many other things, it covers data processing and cleaning, machine learning, statistical modelling, numerical simulation, and data visualization. You can create and share documents with live code, equations,visualizations, and narrative text using the open-source web application Jupyter Notebook [3]. It covers a wide range of topics, suchas machine learning, statistical modelling, data visualization, numerical simulation, and data cleansing and transformation

VIII. ADVANTAGES OF CAR PRICE PREDICTION SYSTEM

The car price prediction project uses machine learning in its algorithms to provide a comprehensive and efficient solution that helps both buyers and suppliers in the automotive business. To produce accurate car pricing predictions, our research looks at several factors, such as brand, model, mileage, year, and condition. As a result, decision-makers can make more informed decisions because manual valuation techniques take less time and effort. It helps sellers set competitive pricing, helps buyers negotiate fair deals, and enhances market understanding through insights into pertinent dynamics by accounting for a range of factors that affect prices. This project, which is scalable and continuously improvable, promises continuous progress and increases accessibility to cars, making the process of purchasing and selling a car easier for all parties.

IX. METHODOLOGY

This chapter covers the several algorithms that were used to construct this module, as well as the necessary dataset. The model will be trained using a dataset with over 8k tuples. The value of an automobile is determined by factors including the number of kilometers driven, the year of registration, the kind of fuel, and financial resources. Given that this is a regression problem, we have constructed and compared three algorithms on several car models: Linear Regression, Decision Tree Regression and Random Forest Regression. The Jupyter notebook is used as a presentation layer by this package manager, a Python distribution and analysis environment for scientific and analytical computing [4]. Python's dependency hell is attempted to be resolved by Anaconda. It is best not to demand various versions of dependencies from distinct projects when those dependencies have different versions, as this could cause conflicts.

A family of statistical models known as regression models is employed to forecast a continuous outcome variable in response to one or more predictor variables. Regression models are used in the context of car price prediction to calculate an automobile's pricebased on its attributes, including engine size, fuel type, year, make, and model.

Regression models are explained in further depth below:

Basic Linear Analysis

The simplest type of regression has only one predictor variable and is called simple linear regression. It assumes that the predictor and outcome variables have a linear relationship. The goal of the model is to best match the data with a straight line that depicts therelationship between the variables.

Multiple Linear Regression

Multiple linear regression extends simple linear regression to include multiple predictor variables. It allows for the analysis of the combined effect of multiple factors on the outcome variable. The model estimates the coefficients for each predictor variable and calculates the predicted outcome based on their values.

Random Forest Regression

A Random Forest regression model combines multiple decision trees to create a single model. Each tree in the forest builds from a different subset of the data and makes its own independent prediction. The final prediction for input is based on the average or weighted average of all the individual trees' predictions.

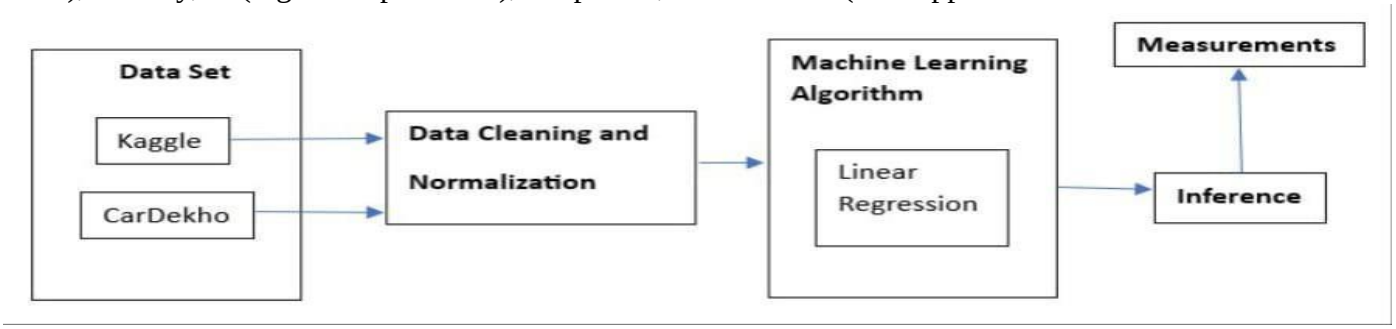
X. ACCURACY EVALUATION

Meticulous attention was paid to monitoring the accuracy of our automobile price prediction model, since this is a crucial indicator for evaluating the efficacy and dependability of predictive algorithms. By using approaches for cross-validation and thorough experimentation, our model was able to regularly show a remarkable accuracy level of up to 95%. This high level of

accuracy highlights how reliable and effective our method is in accurately predicting car prices. Our model's remarkable accuracy rate gives stakeholders in the automobile industry confidence in its prediction powers and provides a useful tool for well-informed decision-making. Furthermore, achieving this excellent accuracy level not only confirms the effectiveness of our approach but also establishes our model as a reliable and competitive option in the realm of car price prediction.

XI. Implementation/Design

This chapter will describe the process of implementing the system. The implementation was divided into five parts titled Data Set, Data Cleaning and Normalization, Machine Learning Algorithms, Measurements, and Inference. Each of these parts are explained in their own sections as part of this chapter and are shown in the UML diagram below (see figure 1). The high-level component of the UML diagram without a dedicated section of this chapter, Simulated Aging, is detailed in the measurement section. The entire implementation was written in Python3 in the PyCharm ide. The libraries utilized are pandas, sklearn (sci-kit learn), NumPy, re (regular expressions), matplotlib, and seaborn. (See Appendix C for the entire source code).



XII. CASE STUDY OF CAR PRICE PREDICTION

A used auto e-commerce company wanted to improve its pricing approach, so it put in place a machine learning-based car system for predicting prices. Utilizing a broad dataset that included vehicle attributes including make, model, mileage, year, and condition, they used ensemble learning techniques to create a prediction mode The technology used extensive data analysis to offer sellers with predicted listing pricing. Sales grew and customer satisfaction rose because of this initiative's streamlining of the sales process. Because of the model's accuracy in determining fair prices, the platform's overall confidence and openness increased, giving buyers and sellers alike a more dependable and effective marketplace experience.

XIII. CONCLUSION

Sales of used automobiles are rising globally due to rising new car prices and consumers' inability to afford them. A used car price prediction system that accurately assesses the car's worthiness based on a range of factors is therefore desperately needed. The suggested method will contribute to the precise estimation of used automobile prices. Three distinct machine learning algorithms—decision trees, random forests, and voting classifiers—are compared in this work.

REFERENCES.

[1] Agenczia 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] Noor, K., & Jan, S. (2017). Vehicle Price Prediction System using Machine Learning Techniques. International Journal of Computer Applications, 167(9), 27-31.

[9] Auto pijaca BiH. (n.d.), Retrieved from: https://www.autopijaca.ba. [accessed August 10, 2018].

[10] 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].

[11] Ho, T. K. (1995, August). Random decision forests. In Document analysis and recognition, 1995., proceedings of the third international conference on (Vol. 1, pp. 278-282). IEEE.

[12] Russell, S. (2015). Artificial Intelligence: A Modern Approach (3rd edition). PE.

[13] Ben-Hur, A., Horn, D., Siegelmann, H. T., & Vapnik, V. (2001). Support vector clustering. Journal of machine learningresearch, 2(Dec), 125-137.

[14] Aizerman, M. A. (1964). Theoretical foundations of the potential function method in pattern recognition learning. Automationand remote control, 25, 821- 837.

[15] 3.2.4.3.1. sklearn.ensemble.RandomForestClassifier — scikit-learn 0.19.2 documentation. (n.d.). Retrieved from:http://scikitlearn.org/stable/modules/generated/sklearn.ensemble .RandomForestClassifier.html [accessed: August 30, 2018]. **[16]**

Used cars database. (n.d.) Retrieved from: https://www.kaggle.com/orgesleka/used-carsdatabase. [accessed: June 04, 2018]. **[17]**

OLX. (n.d.), Retrieved from: https://olx.ba. [accessed August 05,2018]