International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 04 | April - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Role of Effective Model to Assure Price of Used Cars

*** Dr Rajeev Singh,**Dr.Pradeep Kr Singh Bhaduria,***Er.Dileep Verma**

*** Associate Professor(Business Management),**Assistant Professor(Civil Engg.),***Assistant Professor (C.S.Engg.)**

Faculty of Technology,Etawah C.S. Azad University of Agriculture & Technology, Kanpur, U.P.

Introduction

Predicting the price of used cars is both an important and interesting problem. According to data obtained from the National Transport Authority, the number of cars registered between 2003 and 2013 has witnessed a spectacular increase of 234%. From 68, 524 cars registered in 2003, this number has now reached 160, 701. With difficult economic conditions, it is likely that sales of second-hand imported (reconditioned) cars and used cars will increase. It is reported that the sales of new cars have registered a decrease of 8% in 2013. In many developed countries, it is common to lease a car rather than buying it outright. A lease is a binding contract between a buyer and a seller (or a third party – usually a bank, insurance firm or other financial institutions) in which the buyer must pay fixed installments for a predefined number of months/years to the seller/financier. After the lease period is over, the buyer has the possibility to buy the car at its residual value, i.e. its expected resale value. Thus, it is of commercial interest to 754 Sameerchand Pudaruth sellers/financers to be able to predict the salvage value (residual value) of cars with accuracy. If the residual value is under-estimated by the seller/financer at the beginning, the instalments will be higher for the clients who will certainly then opt for another seller/financer. If the residual value is overestimated, the instalments will be lower for the clients but then the seller/financer may have much difficulty at selling these high-priced used cars at this over-estimated residual value. Thus, we can see that estimating the price of used cars is of very high commercial importance as well. Manufacturers' from Germany made a loss of 1 billion Euros in their USA market because of mis-calculating the residual value of leased cars . Most individuals in Mauritius who buy new cars are also very apprehensive about the resale value of their cars after a certain number of years when they will possibly sell it in the used cars market. Predicting the resale value of a car is not a simple task. It is trite knowledge that the value of used cars depends on a number of factors. The most important ones are usually the age of the car, its make (and model), the origin of the car (the original country of the manufacturer), its mileage (the number of kilometers it has run) and its horsepower. Due to rising fuel prices, fuel economy is also of prime importance.

Unfortunately, in practice, most people do not know exactly how much fuel their car consumes for each km driven. Other factors such as the type of fuel it uses the interior style, the braking system, acceleration, the volume of its cylinders (measured in cc), safety index, its size, number of doors, paint colour, weight of the car, consumer reviews, prestigious awards won by the car manufacturer, its physical state, whether it is a sports car, whether it has cruise control, whether it is automatic or manual transmission, whether it belonged to an individual or a company and other options such as air conditioner, sound system, power steering, cosmic wheels, GPS navigator all may influence the price as well. Some special factors which buyers attach importance to in Mauritius are the location of previous owners, whether the car had been involved in serious accidents and whether it is a lady-driven car. The look and feel of the car certainly contributes a lot to the price. As we can see, the price depends on a large number of factors. Unfortunately, information about all these factors is not always available and the buyer must make the decision to purchase at a certain price based on a few factors only. In this work, we have considered only a small subset of the factors mentioned above. More details are provided in Section III. This paper is organised as follows. In the next section, a review of related work is provided. Section III describes the methodology while in section IV, we describe, evaluate and compare different machine learning techniques to predict the price of used cars. Finally, we end the paper with a conclusion with some pointers towards future work.

Review of Literature

Predicting the price of a used car has been studied extensively in various researches. Listian discussed, in her paper written for Master thesis , that a regression model that was built using Support Vector Machines (SVM) can predict the price of a car that has been leased with better precision than multivariate regression or some simple multiple regression. This is on the grounds that a Support Vector Machine (SVM) is better in dealing with datasets with more dimensions and it is less prone to overfitting and underfitting. The weakness of this research is that a change of simple regression with more advanced SVM regression was not shown in basic indicators like mean, variance or standard deviation. Another approach was given by Richardson in his thesis work [3]. His theory was that car producers produce more durable cars. Richardson applied multiple regression analysis and demonstrated that hybrid cars retain their value for a longer time than traditional cars. This has roots in environmental concerns about the climate and it gives higher fuel efficiency. Wu et al. conducted a car price prediction study, by using a neuro-fuzzy knowledge-based system. They took into consideration the following attributes: brand, year of production and type of engine. Their prediction model produced similar results as the simple regression model. Moreover, they made an expert system named ODAV (Optimal Distribution of Auction Vehicles) as there is a high demand for selling the cars at the end of the leasing year by car dealers. This system gives insights into the best prices for vehicles, as well as the location where the best

price can be gained. Regression model based on the k-nearest neighbor machine learning algorithm was used to predict the price of a car. This system has a tendency to be exceptionally successful since more than two million vehicles were exchanged through it . Gonggie proposed a model that is built using ANN (Artificial Neural Networks) for the price prediction of a used car. He considered several attributes: miles passed, estimated car life and brand. The proposed model was built so it could deal with nonlinear relations in data, which was not the case with previous models that were utilizing the simple linear regression techniques. The non-linear model was able to predict prices of cars with better precision than other linear models. Furthermore, Pudaruth applied various machine learning algorithms, namely: k-nearest neighbors, multiple linear regression analysis, decision trees and naïve bayes for car price prediction in Mauritius. The dataset used to create a prediction model was collected manually from local newspapers in less than one month, as time can have a noticeable impact on the price of the car. He studied the following attributes: brand, model, cubic capacity, mileage in kilometers, production year, exterior color, transmission type and price. However, the author found out that Naive Bayes and Decision Tree were unable to predict and classify numeric values. Additionally, a limited number of dataset instances could not give high classification performances, i.e. accuracy less than 70%. Noor and Jan [8] build a model for car price prediction by using multiple linear regression. The dataset was created during the two-months period and included the following features: price, cubic capacity, exterior color, date when the ad was posted, number of ad views, power steering, mileage in kilometer, rims type, type of transmission, engine type, city, registered city, model, version, make and model year. 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, the authors proposed a prediction model based on the single machine learning algorithm. However, it is noticeable that the single machine learning algorithm approach did not give remarkable prediction results for various machine learning methods .

Objectives

- To develop an efficient and effective model which predicts the price of a used car according to the user's inputs.
- To achieve good accuracy.
- To develop a User Interface(UI) which is user-friendly and takes input from the user and predicts the price.

Methodology

We propose a methodology using a Machine Learning model namely random forest to predict the prices of used cars given the features. The price is estimated based on the number of features such as name , year , selling_price , km_driven , fuel , seller_type , transmission and Owner . The intricate details about this model on the used car's data set along with the accuracy are narrated in depth in Section V. We then deploy a website to display our results which are capable of predicting the price of a car given so many features of it. This deployed service is a result of our work, and it incorporates the data ML model with the features.

To summarize,

- First, we collect the data about used cars, and identify important features that reflect the price.
- Second, we preprocess and remove entries with NA values. Discard features that are not relevant for the prediction of the price.
- Third, we apply a random forest model on the preprocessed dataset with features as inputs and the price as output.
- Finally, we deploy a web page as a service which incorporates all the features of the used cars and the random forest model to predict the price of a car.

The first paper is predicting the price of Used Cars Using Machine Learning Techniques. In this paper, they investigate the application of supervised machine learning techniques to predict the price of used cars in Mauritius. The predictions are based on historical data collected from daily newspapers. Different techniques like multiple linear regression analysis, k-nearest neighbours, naïve bayes and decision trees have been used to make the predictions.

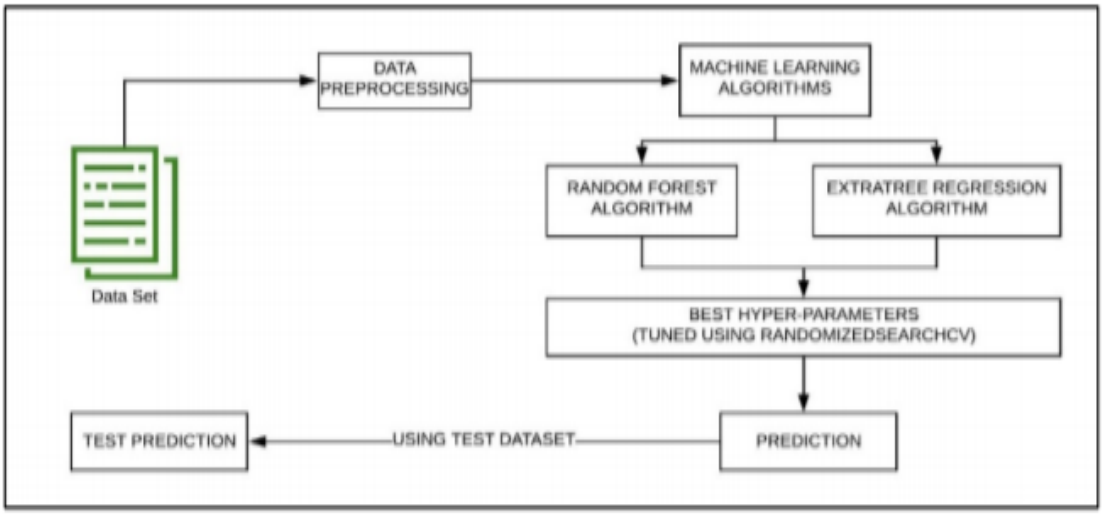
The Second paper is Car Price Prediction Using Machine Learning Techniques. Considerable number of distinct attributes are examined for the reliable and accurate prediction. To build a model for predicting the price of used cars in Bosnia and Herzegovina, they have applied three machine learning techniques (Artificial Neural Network, Support Vector Machine and Random Forest).

The Third paper is a Price Evaluation model in a second hand car system based on BP neural networks. In this paper, the price evaluation model based on big data analysis is proposed, which takes advantage of widely circulated vehicle data and a large number of vehicle transaction data to analyze the price data for each type of vehicle by using the optimized BP neural network algorithm. It aims to establish a second-hand car price evaluation model to get the price that best matches the car.

General Requirements

- The new system must be cost-effective.
- To increase and improve productivity and services.
- Enhancing the user interface.
- To improve information presentation and durability.

Proposed System



The proposed model is an application of the two machine learning algorithms i.e. Random Forest Algorithm and Extra Tree Regression algorithm. In this model first, the dataset is loaded for further exploration.

In this specific model, we used a Dataset available at Kaggle . The dataset can be accessed using the given link (https://www.kaggle.com/nehalbirla/vehicle-dataset-from-cardekho).

After performing the Data preprocessing steps on this dataset such as handling missing values, Hot encoding of Categorical Values, we start training the model for a distributed dataset into two

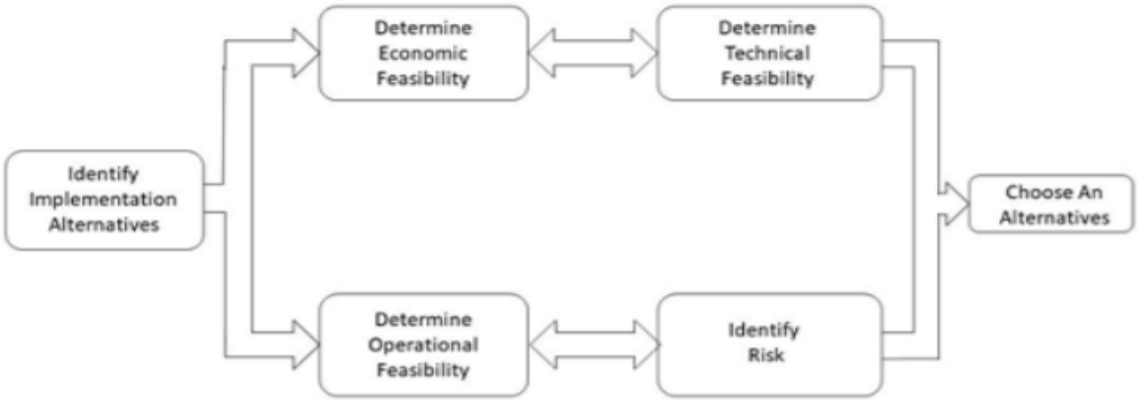
1. Training Dataset
2. Test Dataset.

This test data is picked randomly from the original dataset. Applied the two machine Learning algorithms i.e. Random Forest Algorithm and ExtraTree Regression Algorithm and done tuning of the Hyperparameters using RandomizedSearchCV to get the best Hyper-Parameters for result prediction. Once the model predicts a result, we test the prediction using a test dataset created using the scikit-Learn library and calculate its accuracy.

A feasibility study is an analysis of how successfully a project can be completed, accounting for factors that affect it such as economic, technological, legal and scheduling factors. Project managers use feasibility studies to determine potential positive and negative outcomes of a project before investing a considerable amount of time and money into it.

A feasibility study tests the viability of an idea, a project or even a new business. The goal of a feasibility study is to place emphasis on potential problems that could occur if a project is pursued and determine if, after all significant factors are considered, the project should be pursued.

A feasibility study is conducted to select the best system that meets performance requirements. This entails an identification description, an evaluation of the candidate system and the selection of the best system for the job. The system required a statement of constraints; the identification of specific system objectives and a description of output to define performance etc.



The key considerations in feasibility analysis are:

- Technical Feasibility
- Operational Feasibility
- Economic Feasibility

Technical Feasibility

Evaluating the technical feasibility is the trickiest part of a feasibility study. This is because at this point in time, detailed design of the system is unknown, making it difficult to access issues like performance, costs (on account of the kind of technology to be deployed) etc. It is a measure of the practicality of a specific technical solution and the availability of technical resources and expertise. The proposed system uses Web Pages (built with HTML, CSS and a bit of JavaScript) for graphics user interface (GUI) as front-end, Python, Flask is as back-end server and Scikit-Learn, Pandas and NumPy for Machine Learning, Business Intelligence and Data Analysis. We also used Matplotlib and Seaborn for Data Visualization and Plotting. MySQL is a popular tool used to design and develop database objects such as table views, indexes. Almost all the above tools are open-source, free, readily available, easy to work with and widely used for developing distributed and commercial applications.

To make this software technically feasible we need to consider the following "Technical Issues".

- Understand the different technologies involved in the proposed system.
- Before commencing the project, we have to be very clear about what are the technologies that are required to be available within the organization?
- Find out whether the organization currently possesses the required technologies.
- Is the required technology available with the organization?
- If so, is the capacity sufficient?
- For instance – "Will the current printer be able to handle the new reports and forms required for the new system?"

Technical feasibility is concerned with specifying equipment and software that will successfully satisfy the user requirement.

This can be further classified into two types as follows:

Time Based

In this world of busy schedules with which the industrial professionals are getting through, this kind of system is born for the kind of information they can readily access at the tip of their fingers.

Cost Based

If the physical system is established through a manual process, there is much need for stationery that has to be managed and maintained as files. The overall system once implemented as an internet-based web application not only saves the time but also eliminates the latency that can exist within the system, and saves the costs of stationary that is an unforeseen overhead within the system.

Operational Feasibility

Proposed projects are beneficial only if they can be turned into information systems that will meet the organization operating requirements. Simply stated, this test of feasibility asks if the system will work when it is developed and deployed. Here are questions that will help test the operational feasibility of a project:

- Is there sufficient support for the project from management and users? If the current system is well liked and used to the extent that persons will not be able to see reasons for change, there may be resistance.
- Are there major barriers to implementation?
- Are the current methods acceptable to the user? If they are not, Users may welcome a change that will bring about a more operational and useful system.
- Has the user been involved in the planning and development of the projects?
- Early involvement reduces the chances of resistance to the system and in General, increases the likelihood of successful projects.
- The system will be used if it is developed well then be resistant for users that are undetermined.
- No major training and new skills are required.
- It will help in the time saving and fast processing and dispersal of user requests .
- New product will provide all the benefits of the present system with better performance.
- Improved information, better management and collection of their ports.
- User support.
- User involvement in the building of the present system is sought to keep in mind the user specific requirement and needs.

Since the proposed system is to help and reduce the hardships encountered in the existing manual system, the new system is considered to be operationally feasible.

Economic Feasibility

Economic feasibility attempts to weigh the costs of developing and implementing a new system, against the benefits that would accrue from having the new system in place. This feasibility study gives the top management the economic justification for the new system. A simple economic analysis which gives the actual comparison of costs and benefits is much more meaningful in this case. In addition, this proves to be a useful point of reference to compare actual costs as the project progresses. There could be various types of intangible benefits on account of automation. These could include increased customer satisfaction, improvement in product quality, better decision making, timeliness of information, expanding activities, improved accuracy of operations, better documentation and record keeping, faster retrieval of information and better employee morale.

It looks at the financial aspects of the project. It determines whether the management has enough resources and budget to invest in the proposed system and the estimated time for the recovery of cost incurred. It also determines whether it is worthwhile to invest the money in the proposed project. Economic feasibility is determined by the means of cost benefit analysis. The proposed system is economically feasible because the cost involved in purchasing the hardware and the software are within approachable distance. The operating environment costs are marginal. The less time involved also helped in its economic feasibility. It was observed that the organization is already using mobiles, laptops, Pc's for other purposes incurred for adding this system.

Cost effectiveness depends on the following factors:

Potential Costs

- Hardware/ Software upgrades
- Support cost for application
- Expected operational cost
- Training costs for users to learn the application
- Training costs train developers in new/updated technologies.

Examples of things to consider:

- Hardware /Software selections
- How to convince management to develop new system

- Selection among alternative financing arrangements (rent/lease/purchase)
- Difficulties – discovering and assessing benefits and costs; they can both be intangible, hidden
- and /or hard to estimate, it's also
- Hard to rank multi-criteria alternatives.

Since there are no hidden costs in developing the new system, instead it reduces the costs incurred in manual proceedings, it is considered to be economically feasible.

SDLC (System Development Life Cycle) is a process followed for a software project within a software organization. It consists of a detailed plan describing how to develop, maintain, replace and alter or enhance specific software. The life cycle defines a methodology for improving the quality of software and the overall development process. Any SDLC results in a high-quality system that meets or exceeds customer expectations, reaches completion within time and cost estimates, works effectively and efficiently in the current and planned Information Technology Infrastructure, and is inexpensive to maintain and cost effective to enhance.

Product life Cycle Phases

The various stages in product life cycle are:

- Initiation/Planning
- Requirement gathering and analysis
- Design
- Coding
- Integration and Testing
- Implementation and Maintenance

Figure 4.1: Software Development Life Cycle

Stage 1: Planning and Requirement Analysis

Requirement analysis is the most important and fundamental stage in SDLC. It is performed by the senior members of the team with inputs from the customer, the sales department, market surveys and domain experts in the industry. This information is then used to plan the basic project approach and to conduct product feasibility studies in the economical, operational and technical areas.

Stage 2: Defining Requirements

Once the requirement analysis is done the next step is to clearly define and document the product requirements and get them approved from the customer or the market analysts. This is done through an **SRS (Software Requirement Specification)** document which consists of all the product requirements to be designed and developed during the project life cycle.

Designing the Product Architecture

SRS is the reference for product architects to come up with the best architecture for the product to be developed. Based on the requirements specified in SRS, usually more than one design approach for the product architecture is proposed and documented in a DDS - Design Document Specification.

Stage 4: Building or Developing the Product

In this stage of SDLC the actual development starts and the product is built. The programming code is generated as per DDS during this stage. Developers must follow the coding guidelines defined by their organization and programming tools like compilers, interpreters, debuggers, etc. are used to generate the code. Different high-level programming languages such as C, C++, Pascal, Java and PHP are used for coding.

Stage 5: Testing the Product

This stage is usually a subset of all the stages as in the modern SDLC models, the testing activities are mostly involved in all the stages of SDLC. However, this stage refers to the testing only stage of the product where product defects are reported, tracked, fixed and retested, until the product reaches the quality standards defined in the SRS.

Stage 6: Deployment in the Market and Maintenance

Once the product is tested and ready to be deployed it is released formally in the appropriate market. Sometimes product deployment happens in stages as per the business strategy of that organization. The product may first be released in a limited segment and tested in the real business environment (UAT- User acceptance testing).

There are various software development life cycle models defined and designed which are followed during the software development process. These models are also referred as Software Development Process Models". Each process model follows a Series of steps unique to its type to ensure success in the process of software development. The most used, popular and important SDLC models are given below:

- Waterfall Model
- Iterative Model
- Spiral Model
- V-Shaped Model
- Agile Model

In the “Car Price Prediction” project Agile MODEL is used.

The Agile SDLC model is a combination of iterative and incremental process models with focus on process adaptability and customer satisfaction by rapid delivery of working software products. Agile Methods break the product into small incremental builds. These builds are provided in iterations. Each iteration typically lasts from about one to three weeks. Every iteration involves cross functional teams working simultaneously on various areas like –

- Planning
- Requirement Analysis
- Design
- Coding
- Unit Testing
- Acceptance Testing.

The Agile model believes that every project needs to be handled differently and the existing methods need to be tailored to best suit the project requirements. In Agile, the tasks are divided into time boxes (small time frames) to deliver specific features for a release. Iterative approach is taken and a working software build is delivered after each iteration. Each build is incremental in terms of features; the final build holds all the features required by the customer.

Evaluation and Conclusion In this paper, four different machine learning techniques have been used to forecast the price of used cars in Mauritius. The mean error with linear regression was about Rs51, 000 while for kNN it was about Rs27, 000 for Nissan cars and about Rs45, 000 for Toyota cars. J48 and Naïve Bayes accuracy dangled between 60-70% for different combinations of parameters. The main weakness of decision trees and naïve bayes is their inability to handle output classes with numeric values. Hence, the price attribute had to be classified into classes which contained a range of prices but this evidently introduced further grounds for inaccuracies. The main limitation of this study is the low number of records that have been used. As future work, we intend to collect more data and to use more advanced techniques like artificial neural networks, fuzzy logic and genetic algorithms to predict car prices.

REFERENCES

[1] NATIONAL TRANSPORT AUTHORITY. 2014. Available from: http://nta.gov.mu/English/Statistics/Pages/Archive.aspx Accessed 15 January 2014].

[2] MOTORS MEGA. 2014. Available from: http://motors.mega.mu/news/2013/12/17/auto-market-8-decrease-sales-newcars/ [Accessed 17 January 2014].

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

[4] RICHARDSON, M., 2009. Determinants of Used Car Resale Value. Thesis (BSc). The Colorado College.

[5] 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 Systems with Applications. Vol. 36, Issue 4, pp. 7809-7817.

[6] DU, J., XIE, L. AND 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, Vol. 28, Issue 4, pp. 637-644.

[7] GONGGI, S., 2011. New model for residual value prediction of used cars based on BP neural network and non-linear curve fit. In: Proceedings of the 3 rd IEEE International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Vol 2. pp. 682-685, IEEE Computer Society, Washington DC, USA.

[8] LEXPRESS.MU ONLINE. 2014. Available from: http://www.lexpress.mu/ [Accessed 17 January 2014]

[9] LE DEFI MEDIA GROUP. 2014. Available from: http://www.defimedia.info/ [Accessed 17 January 2014]

[10] GELMAN, A. AND HILL, J., 2006. Data Analysis Using Regression and Multilevel Hierarchical Models. Cambridge University Press, New York, USA LI, Y. H. AND JAIN, A. K., 1998. Classification of Text Documents. The Computer Journal, Vol. 41, pp. 537-546.

[13] QUINLAN, J. R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann.

[14] MITCHELL, T. M., 1997. Machine Learning. McGraw-Hill, Inc. New York, NY, USA.