International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

RETAIL PRICE PREDICTION USING LINEAR REGRESSION

MS. M BANUPRIYA [1] , CHETANAPPRIYA K.L [2] , MAHESWARAN M [3] , MAYURI K.B [4] , RAJARATNAM KAWSHIKA [5]

[1] *Assistant Professor, Dept of Computer Science and Engineering, Saranathan College of Engineering, Tiruchirappalli, Tamil Nadu, India*

[2-5] *Student, Dept of Computer Science and Engineering, Saranathan College of Engineering, Tiruchirappalli, Tamil Nadu, India*

Abstract — Machine Learning plays a vital role from past years in normal speech command, product recommendation as well as in the medical field. It also provides better customer services and a safer automobile system. This shows that ML is a trending technology in almost all the fields. People are looking forward to buy a new home and car with their budgets by analyzing the market. By using certain Machine learning algorithms, a price predicting model can be implemented. This model will help people to put resources into a bequest without moving towards a broker. This project demonstrates the usage of machine learning algorithm in the prediction of prices using dataset downloaded from Kaggle. This model provides a review of the usage of existing machine learning algorithms on a dataset and tries to implement this prediction engine for real-life usage by users. The findings are obtained using linear regression algorithm. Machine learning Algorithms are used because they provide us with continuous value as output and not a categorized value. Because of which it is possible to predict the actual price rather than the price range of a car or house.

Key Words: Machine learning, price prediction, linear regression, regression analysis, feature

1. INTRODUCTION

Retail markets have a term called retail price. This is nothing but the price paid by the customers for the products which are bought by them. People will have a dream to buy their own property within their budgets by analyzing the market instead of selling it onwards. In a retail market there are many types of prices which are used by retailers, manufacturers, and distributors. These prices have their own unique property and it changes between different sellers. Inside the market system, the final retailers have the opportunity to fix the product price based on its need and availability of products. The price fixed by the retailers will have the motive to increase the profit while making the consumers to pay for their product. As the next level of authority, producers of the products have the option to change the product price depending on the production cost.

Price prediction uses certain algorithm to analyze an outcome or service based on its behavior, need, and current market trends. Then the software sets a price at a level which both attract customers and increase sales. In certain circles, this practice is called as price prediction. When you come to release a new product Machine learning models fall into three primary categories. Entering to a new market, or exploring a new niche, you likely will not have such a good grasp of this new space. This is a challenge often referred to as a 'cold-start' — in other words, you have no real-world data. In these situations, you can use a method called transfer learning. The practice assumes the market will behave in a same way to your existing offerings. And it uses machine learning solutions, backed by insights from other domains, to find the right price point. Machine Learning is a type of domain which is used in artificial intelligence and computer science. This domain talks about how to train the machine automatically without any human mediation or any other programs. This domain is classified into three primary categories. They are supervised learning, unsupervised learning, and semi-supervised learning. This system uses linear regression which is a type of machine learning algorithm. Linear regression is used to measure, or visualize, a relationship between two different features. There two kinds of variables which are used by this regression type. They are dependent variable and independent variable. Machine learning models use both technical and fundamental analysis in the price process. Technical analysis looks at historical prices, economic growth rates, and other related factors, formulating an approximate price. Then, to get a more accurate picture of the market, the process turns to fundamental analysis. This step looks at various external and internal factors, including macro-factors like the season and micro-influencers like the time of the day, trying to figure out when a consumer is most likely to buy.

In mathematical terms, these processes are known as regression analysis. Which is a statistical way to predict the relationship between variables. In price prediction, price is the independent variable. And it is affected by several dependent variables. Beyond regression, price prediction uses descriptive and predictive analytics.

2. LITERATURE SURVEY

The related work on this project shows that there have been several methods of implementing the system under different domains:

Tarun Kumara and Rajiv Kumara (2020) proposed a model to foresee whether the pay of a person surpasses $50000 per year or not founded on statistics information gave by Census authority database, while considering different factors such as age, work class, gender, marital status, education, race, occupation etc. using exploratory analysis and classification algorithms. This investigation intends to utilize machine learning procedures in giving a response to the salary forecast. This model measures the accuracy of different machine learning models like Random Forest classifier, Naïve Bayes classifier and Support vector classifier. To find the relation between attributes co-relation matrix is calculated. The result can be used for social purposes, commercial purposes and economic study of a region for level of economic inequality.

Ashutosh Datt Sharma and Vibhor Sharma (2020) proposed a Linear Regression Model that was implemented by employing various prominent algorithms from the python libraries and modules. After the collection of data was done, further processing of data was done. The null entries and missing datapoints were removed from the dataset and the categorical variables were also processed using One Hot Encoding technique. The results showed that there is a positive correlation between selling price and present price while a negative correlation between selling price and Kms Driven, Years Used and Owner (Number of Previous Owners). Positive correlation can be referred to as direct proportion while negative correlation can be referred to as inverse proportion. Also, it was concluded that the selling price of cars were higher when sold by dealers when compared to individuals. It was also observed that the selling price of cars with fuel type diesel was higher than those having petrol and CNG as fuel type.

Smith Dabreo, Shaleel Rodrigues, Valiant Rodrigues and Parshvi Shah (2021) proposed a review of the usage of existing machine learning algorithms on two extremely different datasets and tries to implement this prediction engine for real-life usage by users. This proposed model allows people to buy houses and real estate at their rightful prices and ensures that they do not get tricked by sketchy agents who just are after their money. Additionally, this model will also help big companies by giving accurate predictions for them to set the pricing and save them from a lot of hassle and save a lot of precious time and money. The findings indicate that using different algorithms can drastically change accuracy. Also, a poor dataset can negatively affect predictions. Furthermore, it provides sufficient proof of what algorithm is most suitable for this task.

Mohamed Zaim Shahrel and Sofianita Mutalib and Shuzlina Abdul-Rahman (2021) proposed a Web application namely Price Cop that helps the user or customers monitor price of product listed at e-commerce, so that they are aware of the price changes. Price changes can be both good and bad. The price prediction model is developed by using Linear Regression (LR) technique. LR is commonly used to determine outcomes and used as predictors. Least Squares Support Vector Machine (LSSVM) and Artificial Bee Colony (ABC) are used as a comparison to evaluate the accuracy of the LR technique. LSSVM-ABC was initially proposed for stock market price predictions. The results show the accuracy of pricing prediction using LSSVM-ABC is 84%, while it is 62% when LR is employed. ABC is integrated into SVM to optimize the solution and is responsible for the best solution in every iteration. Even though LSSVM-ABC predicts product pricing more accurately than LR, this technique is best trained using at least a year's worth of product prices, and the data is limited for this purpose. In the future, the dataset can be collected daily and trained for accuracy.

Monish S, Mridul Mohta and Shanta Rangaswamy (2022) proposed model that compares different price prediction models and concludes that bidirectional LSTM is the best model among RNN, LSTM and Bi-LSTM to forecast the price of Ethereum. The model uses the closing price as the parameter to predict the price. When it comes to the financial market, getting to know the trends is very important. This model gives a good trend for the prices for longer periods of time (preferably 90 days). The model is also easily scalable, and the accuracy can further be improved with some tweaks. As noticed above, RNN fails in predicting the prices. LSTM and Bi-directional LSTM succeed in predicting the price of Ethereum but the better one is bidirectional LSTM. It achieves its purpose of predicting the prices and forecasting the price trends with reasonable accuracy. The model can further be enhanced by considering more parameters and changing various hyper parameters. There is a lot of scope and things to further explore with better models coming up each day.

3. PROPOSED SYSTEM

The proposed system focuses on forecasting house and car prices by inculcating machine learning algorithms like Linear Regression. The proposed system "Retail Price Prediction Using Machine Learning" is used to predict the prices of both house and car using multiple features. This proposed system can train models from various attributes. The 80% of data which is taken from previous data is used for the purpose of training while

the remaining 20% of data is used for the purpose of testing. Here, the raw data is stored in a '.csv' file. This system majorly uses two machine-learning libraries to solve these problems. The first one is 'pandas' and the other is 'NumPy'. The pandas used to load the '.csv' file into Jupiter notebook and used to clean the data as well as manipulate the data. Another was scikit-Learn, which is used for real analysis, and it contains various inbuilt functions which help to solve the problem. One more library is used which is nothing but NumPy. For the purpose of train-test splitting NumPy is used. System architecture is the conceptual model that defines the structure, behavior, and views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system. It consists of two sections: training and testing. The training has the following components: the label, input, feature extractor, and machine learning algorithm. The testing section has the following components in it: the input, feature extractor, regression model, and output label. This system undergoes six different modules and they are data collection, data preprocessing, feature selection, data visualization, model training and evaluation, and user interface which are explained further. Table 1 and 2 depicts the attributes which are used in both house and car price prediction respectively.

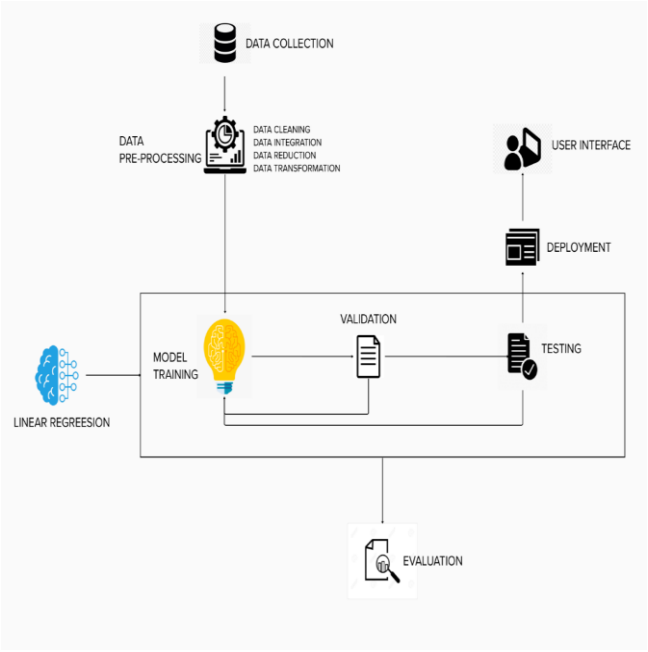

Fig-1: System Architecture

3.1 Data Collection

It is recommended to collect multiple datasets from various platform in a structured manner. This provides solution evaluating outcomes. Data collection is the way toward social events and estimating data on focused elements in a built-up framework, which at that point empowers one to address pertinent inquiries and assess results. It has been attempted for various datasets on Kaggle, which would suit the proposed system. After looking at a lot of datasets appropriate datasets are chosen for both house and car price prediction.

Table -1: Feature attributes of house model

ATTRIBUTE NAME	DESCRIPTION	TYPE
Number of bedrooms	Number of bedrooms in the house	Number
Number of bathrooms	Number of bathrooms in the house	Number
Square feet	Area of the house	Number
Location	Location of the house	Text

Table-2: Feature attributes of car model

ATTRIBUTE NAME	DESCRIPTION	TYPE
Company	Company the car belongs to	Text
Model	Model of the car	Text
Year of purchase	Year in which the car was purchased	Number
Fuel type	Type of fuel the car uses	Text
KM travelled	Distance travelled by the car	Number

3.2 Data Preprocessing

It is mandatory to modify the dataset before uploading. This process is used to change the clumsy data into a clean data. It is said as data mining strategy which cleans messy information into a justifiable organization. Result of this process is the cleaned dataset is used for further execution of the proposed system.

Data cleaning: This phase is nothing but detecting and deleting errors to maximize the worth of data. It is carried out with the help of data-wrangling tools. It is the way toward identifying and amending off-base records from a record set, table, or database. It finds the deficient information and replaces messy information. Principal objective of data cleaning is to distinguish and expel blunders to build the estimation of information dynamically.

3.3 Feature Selection

Feature Definition: The current work utilizes data from the web resource Kaggle.com and the dataset has been used for a competition hosted by that web application.
Feature Selection: This work utilizes feature selection methods such as variance influence factor, Information value, principal component analysis, and data transformation techniques such as outlier and missing

value treatment as well as box-cox transformation techniques for the feature selection and subsequent transformation process.

3.4 Data Visualization

This phase implied to grasp the visual meaning of dataset. It helps us to understand difficult theories also identifies new patterns. This phase provides many unique patterns obtained from the system through which many meaningful concepts can be identified in a pictorial form. To impart data effectively, this illustration utilizes measurable plots, data designs, and different apparatuses. Successful visualization assists customers with separating and reasoning about data and verification. It makes complex data progressively accessible, reasonable, and usable. Below fig 4 depicts the visual representation of price of cars over companies and fig 5 depicts the relationship of cars with years.

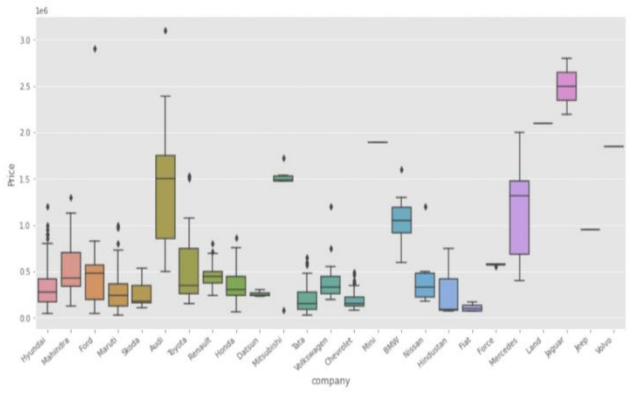

Fig -2: Box Plot

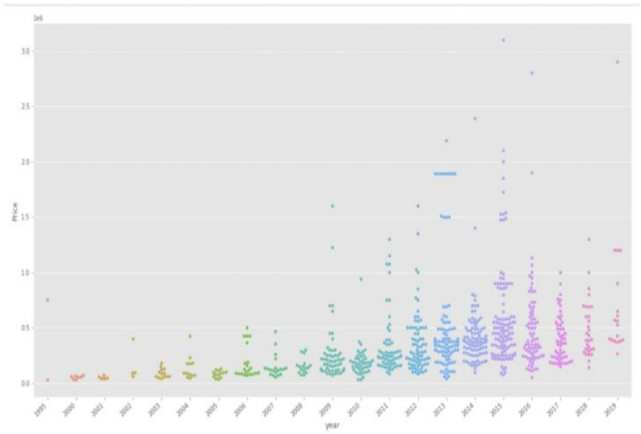

Fig -3: Swarm Plot

3.5 Model Training and Evaluation

This module is used for training and evaluation. Training is enabling the model to solve the problem using the given dataset. Evaluation is measuring the performance of the model by a given dataset.

3.6 User Interface

This module provides a user–friendly interface for the end-users to interact with the system. It is a web-based interface or a mobile application, depending on the target audience. The user will open the website and enter all the features/attributes of the house or car they wish to predict the price for. Furthermore, when the user clicks submit attributes it will be checked for null values and then all the attributes will be validated to check if they are in the same data type as necessary. Finally, after all the conditions are satisfied the data will be sent for prediction and the predicted price will be displayed to the user on the website. Below figures 6 and 7 are the user interface of both house and car price prediction respectively.

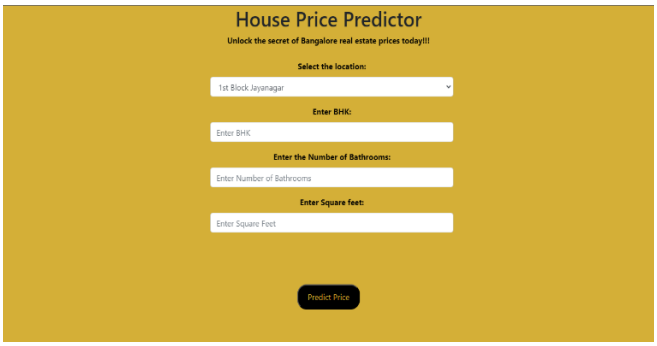

Fig -4: House Price Predictor Page

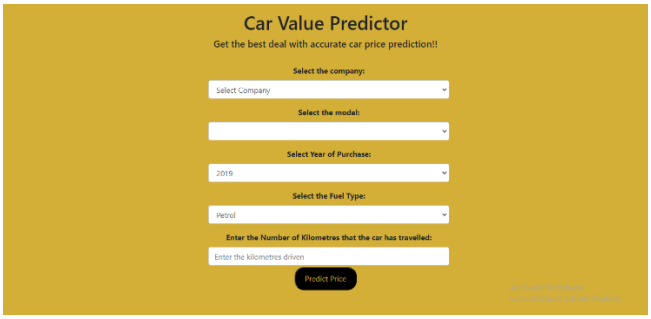

Fig -5: Car Price Predictor Page

4. RESULT AND DISCUSSION

The motive of this proposed system is to predict the prices of both house and car by grasping certain attribute values from the client based on their budget. These values are fed to the model and depending on these values this system predicts the prices of houses and cars. In order achieve this initially required and appropriate dataset is collected from certain platform, secondly these collected dataset will undergo the cleaning process. Moreover, these datasets are allowed to go through data pre-processing phase where missing datasets are undertaken, and if required label encoding will be

performed. For the purpose of training the model certain machine learning algorithms are used. Once the model is developed the data given by the users are applied to this system and within certain measured time the appropriate and approximate predicted price of the properties are displayed on the screen. With the help of this model users can easily predict the price of their dream house or cars within their budget by analyzing the current market of retail products and can plan to purchase.

This retail price prediction model is trained to predict the prices of both houses and cars.This model acheives an accuracy of 84% for the given datasets.Below tables 3 and 4 represents the input data used for training this model. Each row represents to a unique item while the columns represents different features. The actual prices of the items are labeled as target variables which is given as last column.

Table-3: Car Price Prediction Values

Company	Model	Year of purchase	Fuel type	No.of km travelled	Target Variable (Price)
Audi	Audi A4.18	2019	Petrol	2	926672.27
Audi	Audi Q5 2.0	2002	Diesel	4	2148214.5
Force	Force Motors Force	2012	Petrol	8	491358.61
Chevrolet	Chevrolet Beat	2019	Petrol	8	241454.54

Table-4: House Price Prediction Values

Location	No.of BHK	No.of Bathrooms	Square feet	Target Variable (Price)
1ST Block Jayanagar	3	2	600	14645996.11
5th Block Hbr Layout	4	4	5000	32859912.09
Abbigere	3	1	4000	325412198.3
Anandapura	3	1	4000	325768764.15

5. CONCLUSION

The project entitled “Retail Price Prediction” has presented to predict house or car price based on various features on given data. Users and companies will be able to fill a form about various attributes about their property that they want to predict the price of. It helps people to buy house or car in budget and reduce loss of money. This project is currently working on deployment using flask and automate the result file. The model uses single data set for prediction. Usage of other sectors and countries dataset, is yet to be explored.

ACKNOWLEDGEMENT

We extend our sincere appreciation to professors from the department of Computer Science at Saranathan College of Engineering for their guidance and expertise.

Their insightful suggestions and constructive feedback greatly influenced the direction and quality of this work.Lastly, we would like to express our deepest gratitude to our families and friends for their enduring support and encouragement throughout this research.

REFFERENCES

1. Anand G.Rawool, Dattatray V.Rogye, Sainath G.Rane, Dr.Vinayak A. Bharadi (2021), "House price prediction using machine learning", vol.4, pp. 29-33.
2. Ashutosh Datt Sharma and Vibhor Sharma (2020), "Used car price prediction using linear regression model", vol. 2, pp. 1-8.
3. Julakha Jahan Jui, M. M. Imran Molla, Bifta Sama Bari, Mamunur Rashid and Md Jahid Hasan (2020), "Flat Price Prediction Using Linear and Random Forest Regression Based on Machine Learning Techniques", pp. 1-14, doi: 10.1007/978-981-15-6025-5_19.
4. Ketan Agrahari, Ayush Chaubey, MamoorKhan and Manas Srivastava (2021), "Car price prediction using machine learning".
5. Mastan vali, K.Sankeerthana, B.Naveen and N.Vishal (2020), "Prediction of online sales using linear regression" vol. 8, pp. 1-3.
6. Mohamed Zaim Shahrel and SofianitaMutalib and Shuzlina AbdulRahman (2020), "PriceCop – Price Monitor and Prediction Using Linear Regression and LSVM-ABC Methods for E-commerce Platform", pp.1- 14, doi: 10.5815/ijieeb.2021.01.01.
7. Monish S, Mridul Mohta and Shanta Rangaswamy (2022), "Ethereum price prediction using machine learning techniques – A comparative study", vol.7, pp.137-142.
8. Patrick LoolaBokonda, Khadija Ouazzani and NissrineSouissi (2021), "Predictive analysis using machine learning: Review of trends and methods", doi: 10.1109/ISAECT50560.2020.9523703.
9. Sakshi Madose, Komal Mhaske, Akansha Narake and Pratiksha Moholkar (2021), "Stock price prediction using machine learning", vol.8.
10. Smith Dabreo, Shaleel Rodrigues, Valiant Rodrigues and Parshvi Shah (2021), "Real Estate Price Prediction", Vol. 10.
11. SussyBayona-Ore, RinoCerna and Eduardo Tirado Hinojoza (2021),"Machine learning for price

prediction for agricultural products", vol.18, doi:10.37394/23207.2021.18.92.

12. Tarun Kumara and Rajiv Kumara (2020), "Intelligent Income prediction model", vol.7, pp. 660-669.