



# Price Estimation of Used Cars Using Machine Learning Algorithms

B. Valarmathi<sup>1</sup>, N. Srinivasa Gupta<sup>2</sup>, K. Santhi<sup>3</sup>, T. Chellatamilan<sup>3</sup>, A. Kavitha<sup>4</sup>, Armaan Raahil<sup>5</sup>, and N. Padmavathy<sup>6</sup>✉

<sup>1</sup> Department of Software and Systems Engineering, School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India

<sup>2</sup> Department of Manufacturing Engineering, School of Mechanical Engineering, Vellore Institute of Technology, Tamil Nadu Vellore, India

<sup>3</sup> Department of Analytics, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India  
santhikrishnan@vit.ac.in

<sup>4</sup> Department of Mathematics, School of Advanced Sciences, Vellore Institute of Technology, Vellore, Tamil Nadu, India  
kavitha@vit.ac.in

<sup>5</sup> Department of Information Technology, School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India

<sup>6</sup> Department of Electronics and Communication Engineering, Vishnu Institute of Technology, Bhimavaram, Andhra Pradesh, India  
padmavathy.n@vishnu.edu.in

**Abstract.** In this study, machine learning (ML) techniques are employed to predict used car prices. Several features are used to calculate the price of used cars, but in this paper, we find efficient ways to find the most precise car prices. Despite the fact that there are websites offering this service, they could not employ the most precise prediction system. It is also possible to predict a used car's true market value using a variety of models and techniques. It's important to understand their genuine market value before buying or selling. Both buyers and sellers will be benefitted from these accurate predictions. Support Vector regression, Random Forest regression, and CatBoost regression techniques are used in the proposed system. In the existing method [13], mean absolute error for decision tree regression was 0.6711, which was the least among other algorithms like Linear regression, Lasso regression, Ridge regression, Bayesian Ridge regression, and etc., they used. In the proposed system, mean absolute error (MAE) for Support Vector regression, CatBoost regression and Random Forest regression techniques are 0.1459, 0.1371 and 0.1284 respectively. The prices of second hand/used cars are predicted using the CatBoost regression, Support Vector regression, and Random Forest regression techniques. The accuracy of these algorithms are 86.28%, 85.40% and 87.16%. Among these three algorithms, Random Forest regression gives the least MAE of 0.1317 and the highest accuracy of 87.16%.

**Keywords:** CatBoost regression · Support Vector regression · Random Forest regression · Machine learning

## 1 Introduction

Many sites are involved in selling the used cars. CASH MY CAR is one of the largest car selling sites in India whose aim is to revolutionize the traditional method of selling used cars in India. On the basis of the different features that are used to determine the price, these companies have a mechanism for doing business with the buyer. As machine learning techniques have advanced, new algorithms have been developed that make it easier to make such predictions. Since the dataset has more features that might be used to estimate car prices, we want to use three machine learning techniques like Support Vector regression, CatBoost regression and Random Forest regression.

Yandex created the open-source boosting library is known as CatBoost, or categorical boosting. CatBoost can be utilized in ranking, recommendation systems, forecasting, and even personal assistants in addition to regression and classification. CatBoost is primarily concerned with decision tree theory and gradient boosting. The primary objective of boosting is to create a strong, competitive predictive model through greedy search by deliberately combining numerous weak models or models that just marginally outperform chance.

Support Vector Regression (SVR) is a supervised learning method for predicting discrete values. The same principle behind support vector machines underlies support vector regression. The core idea of SVR is to find the line that fits the data the best. The best fit line is generally regarded as the hyperplane with the highest number of SVR points.

A supervised learning technique known as Random Forest Regression uses the ensemble learning method for regression. In order to produce predictions that are more accurate than those from a single model, the ensemble learning method integrates predictions from several machine learning algorithms.

## 2 Literature Survey

In this paper [1], the main aim of the present research was to investigate various types of car data with the aim of developing an automated method for predicting car prices. They proposed a brand-new method for quantitative knowledge-based systems and data. They utilized a suggested car price model and provided non-numerical data solutions for forecasting. They used a range of data to solve the big data problem as well. Empirical predictions outperform other models in terms of performance.

In this research [2], they investigated how to forecast the cost of used cars in Mauritius using supervised machine learning techniques. The forecasts were created using historical information gathered from daily newspapers. Numerous techniques, such as K-nearest neighbours, Naive Bayes, and etc., were used for forecasting. The forecasts were then compared and assessed to see which ones delivered the best performances. The accuracy of handling a seemingly straightforward problem proved to be rather difficult. The performance produced by the four techniques was equal.

In this article [3], the need to estimate a used item's price according to its image and word-based description for different sets of item types, they provided a deep model architecture. For price prediction, this method used deep neural networks with long short-term

memory (LSTM) and convolutional neural networks (CNN). In terms of mean absolute error accuracy, the model outperformed the standard support vector machine model. They suggested predicting the second-hand item's minimum and maximum price. Prediction task models use linear regression, seasonal autoregressive integrated autoregressive moving average and LSTM methods.

In this paper [4], they discussed the development, deployment, and assessment of ForeXGBoost. To enhance data quality, ForeXGBoost made full use of precisely crafted data filling algorithms for missing values. ForeXGBoost could increase forecast accuracy by extracting features from historical sales and production data using the sliding window. After an extensive investigation was undertaken to assess the impact of many variables on vehicle sales through the collection of information and the correlation of data, the most indicative features of the predictive set of features had been selected. Additionally, they quickly and accurately predict vehicle sales using the XGBoost prediction algorithm. Numerous tests showed that ForeXGBoost may produce accurate predictions with little overhead.

In this paper [5], Artificial Neural Networks, Support Vector Machines, and Random Forest are three machine learning approaches that were used to create an accurate model to estimate the prices of secondhand cars. Instead of using these approaches on individual data items, they were applied to the full collection of data items. Through a web site, which was also utilized to make the forecast, this data set was gathered. Web scrapers created in PHP programming language must be used to gather the data. To obtain the optimum outcome from the available data set, many machine learning methods with diverse capabilities had been compared. The last prediction model was added to a Java program.

In this paper [6], the authors explore how current deep learning models can be used to estimate a secondhand car's pricing as well as rate its performance and pollution index services. In the study, it was addressed how judgments about a car's performance can be made using an on-board diagnostic system and a variety of sensors. This would entail a general improvement in cost, emission, and performance predictions for automobiles.

In this article [7], using more than 100,000 data of used car sales from all around China; the authors carried out an empirical analysis and compared linear regression with random forest in detail. Three different models were employed with these two algorithms to predict the cost of used cars: one for a particular car series, one for a particular car make. In pricing evaluation models for particular auto manufacturers, the universal model, the findings showed that random forest clearly outperformed linear regression, but its effect is consistent but not perfect. As a result, it was demonstrated that the random forest method was the most effective for complex models with numerous variables and samples.

In this article [8], The proposed price evaluation approach was used to assess the pricing information for each type of car, which made use of widely dispersed car data and a substantial amount of vehicle transaction data. The process was intended to be developed in order to evaluate used car pricing and determine the price that best suited the car. The appropriate numbers of hidden neurons in the BP neural network (BPN) were selected using the optimized BPN approach to increase the precision of the prediction model and speed the convergence of the network topology. Through sample simulation

trials, the actual transaction price generated from the improved model was compared with the fitting curve of the predicted price. Consequently, the optimized model's accuracy and fitting were both enhanced.

In this paper [9], they included feature screening and data preprocessing. Data transformation was used to standardize data formats in order to increase data quality. Data cleaning involved the removal of outliers and the filling in of missing values. The model building, training, and prediction processes all started with the screened features. Analysis of correlations and feature extraction using LightMBG were both part of the feature screening process. Five regression models were subsequently built, assessed, and the feature attributes collected through feature engineering for training. The result was a novel regression model that performed better than the five regression models when Random Forest and XGBoost were weighted and combined. Finally, the price of secondhand cars is predicted using the unique regression model.

In this paper [10], to forecast the cycle of trade in used cars, they applied machine learning. They combined the data, divided it into boxes, created additional features, dealt with outliers, and used principal component analysis to lower the dimension of the features in order to improve the expressivity of the model. Following that, the vehicle transaction cycle prediction data was fitted using the 50% discount cross-validation approach, and a random forest was employed for model training. It was discovered how each attribute related to the second-hand car transaction cycle. The cycle of transactions involving used cars was predicted and established. In both the test and the training sets, the fusion model's mean absolute error was 10.32 and 4.72, respectively.

In this paper [11], used car data was explored, and by examining the heat map, box line plot, and violin plot, features with correlations up to 99%, a few outliers, and missing values were discovered. Three distinct used car datasets were produced using three fill methods for missing values, namely, mode fill, median fill, and random forest fill, after the duplicate features and outliers were removed. Then, using the three used automobile datasets as input, four models—XGBoost, LightGBM, SVR, and BP neural network—were trained. The test set of used automobiles served as the subject of experiments, and the outcomes revealed that the LightGBM model performed well in both MAE and RA2, achieving results of 0.115350 and 0.965493, respectively. Which performed better on the LOSS of the training set than the other three models; yet, the BP neural network, perhaps as a result of the overfitting issue, on the test set performed poorly. The LightGBM improved the model on the XGBoost model to improve training speed and accuracy, so the results obtained by LightGBM model are slightly better than those of the LightBGM model. The SVR model performed best when the kernel selected was linear, but this did not apply to use car data, so the results obtained on MAE and R2 were worse than those of the other three models.

Developed a mathematical model that could estimate the cost of a used car based on its current attributes was the aim of this study [12]. It was challenging to estimate the cost of a used car because several variables, including current mileage, condition, make, and year, among others, might affect the price. Furthermore, precisely predicting the price of a used car became problematic from the standpoint of a seller. Therefore, the focus was on developing tools and researching models that could accurately predict a

used car's price based on its capabilities. As a result, a customer could make a purchase with considerably more knowledge.

In this article [13], a large number of features and parameters needed to be taken into account in order to produce accurate results made it difficult to predict used car prices. The initial and most crucial phase was data gathering and preprocessing. After that, a model for formulating algorithms and generating output was created. Different regression techniques were used on the model, and the best performance was found for the Decision Tree Algorithm.

### 3 Dataset Description and Sample Data

The dataset includes nine attributes namely Car Name, Year, Selling Price, Present Price, Kms Driven, Fuel Type, Seller Type, Transmission and Owner. 301 rows are present in this dataset. The description of the nine attributes of the dataset is shown in the Table 1. In the dataset, the attributes like Car Name, Fuel Type, Seller Type and Transmission contains the categorical values. The sample dataset before preprocessing is shown in the Table 2. For example, let's say that a dataset contains a column called Height. Height column has short, medium and tall labels. Label encoding is used to transform the height column into 0, 1 and 2. i.e., the labels for short, medium and tall heights are 0, 1 and 2 respectively, to convert the categorical values into the numerical data. The process of converting labels into a numeric format so that machines can read them is known as label encoding. The sample dataset after preprocessing is shown in the Table 3. The data used in this paper was downloaded from Kaggle (<https://www.kaggle.com/code/mdejazulhassan/vehicle-dataset-from-cardekho/data>).

**Table 1.** Attribute description of the dataset

Name of the attribute	Description of the attribute
Car name	It provides the car's name
Year	The year when the car was bought is specified, ranging from 2003 to 2018
Selling price	It specifies the price charged so that the owner sells the car
Present price	It specifies the actual price of the vehicle outside the showroom
Kms driven	It gives the car's total distance travelled in kilometres
Fuel type	It specifies the car's fuel type
Seller type	It states whether the vendor is either a merchant or a person
Transmission	It specifies if the vehicle is automatic/ manual
Owner	It identifies former owners of a vehicle

**Table 2.** Sample dataset before preprocessing

Car name	Year	Selling price	Present price	Kms driven	Fuel type	Seller type	Transmission	Owner
ritz	2014	3.35	5.59	27000	Petrol	Dealer	Manual	0
sx4	2013	4.75	9.54	43000	Diesel	Dealer	Manual	0
ciaz	2017	7.25	9.85	6900	Petrol	Dealer	Manual	0
wagon r	2011	2.85	4.15	5200	Petrol	Dealer	Manual	0
swift	2014	4.6	6.87	42450	Diesel	Dealer	Manual	0
vitarabrezza	2018	9.25	9.83	2071	Diesel	Dealer	Manual	0
ciaz	2015	6.75	8.12	18796	Petrol	Dealer	Manual	0
s cross	2015	6.5	8.61	33429	Diesel	Dealer	Manual	0
ciaz	2016	8.75	8.89	20273	Diesel	Dealer	Manual	0
ciaz	2015	7.45	8.92	42367	Diesel	Dealer	Manual	0
alto 800	2017	2.85	3.6	2135	Petrol	Dealer	Manual	0
ciaz	2015	6.85	10.38	51000	Diesel	Dealer	Manual	0
ciaz	2015	7.5	9.94	15000	Petrol	Dealer	Automatic	0
ertiga	2015	6.1	7.71	26000	Petrol	Dealer	Manual	0

**Table 3.** Sample dataset after preprocessing

S. No	Car Name_cat	Year	Present Price	KmsDriven	Fuel Type_cat	Seller Type_cat	Transmission_cat	Owner
0	90	2014	5.59	27000	2	0	1	0
1	93	2013	9.54	43000	1	0	1	0
2	68	2017	9.85	6900	2	0	1	0
3	96	2011	4.15	5200	2	0	1	0
4	92	2014	6.87	42450	1	0	1	0
...	...	...	...	...	...	...	...	...
296	69	2016	11.60	33988	1	0	1	0
297	66	2015	5.90	60000	2	0	1	0
298	69	2009	11.00	87934	2	0	1	0
299	69	2017	12.50	9000	1	0	1	0
300	66	2016	5.90	5464	2	0	1	0

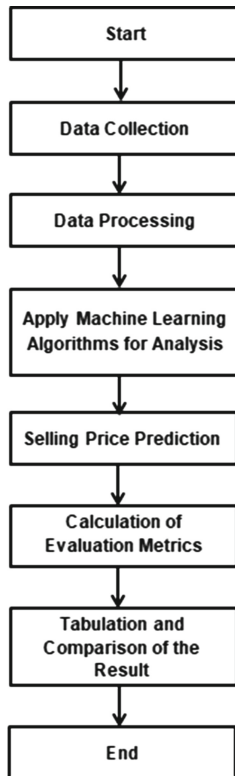
## 4 Methodology

The proposed workflow diagram is shown in the Fig. 1. There are six steps and these are described below.

**Step 1: Data Collection** – The dataset is collected from the Kaggle.

**Step 2: Data Processing** – By eliminating every row with a null value, the dataset is cleaned. Categorical values are present in the dataset for the attributes Car Name, Fuel Type, Seller Type, and Transmission. Using label encoding, these categorical variables are transformed into numerical values. Label coding is the process of converting labels into a numeric format that can be read by computers.

**Step 3: Applying ML Algorithms for Analysis** – An 80:20 ratio is used to divide the datasets into training and test data. In order to provide a flexible and consistent dataset, the data has been standardized. This was accomplished using Standard Scaler from the Scikit-Learn Library. It normalizes the features by deleting the mean and scaling the unit variance. After normalizing the range of features in the datasets, next is to forecast the selling price of used cars, a variety of techniques are used, such as CatBoost Regression, Random Forest Regression, and Support Vector Regression.



**Fig. 1.** Flowchart for the proposed work

**Step 4: Selling Price Prediction** – The accuracy of the used car selling price is estimated with the help of machine learning techniques.

**Step 5: Calculation of Evaluation Metrics** – R-SQUARE, MSE, RMSE, MAE, and the accuracy (1-MAE) of CatBoost Regression, Random Forest Regression, and Support Vector Regression are the metrics employed for the proposed work. The accuracy formula is shown in eq. (1).

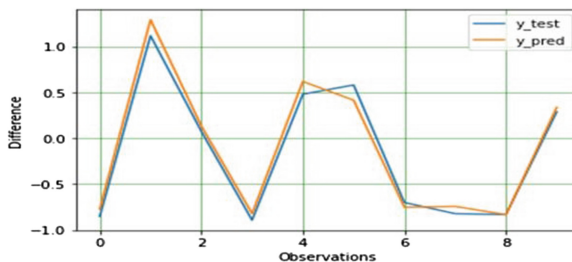
$$\text{Accuracy (A)} = 1 - \text{Mean Absolute Error (MAE)} \quad (1)$$

**Step 6: Tabulation and Comparison of the Result** – Taking into account the accuracy of the algorithms as well as R-SQUARE, MSE, RMSE, and all other metrics that were acquired after using machine learning techniques. A number of graphs, including line graphs, density plots, and scatter plots, are plotted for each model for the purpose of comparison tabulating. All metric values for each model are further tabulated, and bar graphs are generated to compare the accuracy of each model as well as the values for R-SQUARE, MSE, RMSE, and MAE. The next step after tabulating all the numbers is to compare the metric values of all the used algorithms to determine which one best fits the scenario.

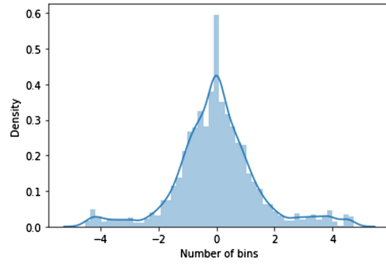
## 5 Result and Discussion

The proposed work uses the inbuilt Python libraries such as NUMPY, PANDAS (for linear algebra and scientific computing) and visualization libraries such as MATPLOTLIB, SEABORN. It uses the Label Encoder to convert all the other data types to numeric data type.

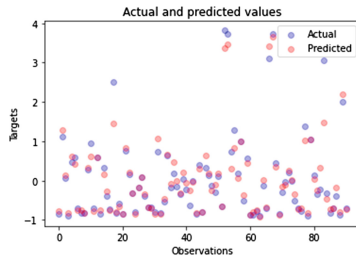
The closeness between predicted and actual data CatBoost Regression is shown in the Fig. 2. Density Plot to determine distribution of variables in the dataset is shown in Fig. 3. Figure 4 displays a scatter plot to show the correlation between the actual and expected values. The variance between actual and the predicted sales price (sample) is shown in the Table 4 and if it is accurate or not in the CatBoost Regression model. The evaluation metrics used in CatBoost Regression model are R2 score, MSE, RMSE, MAE and accuracy and their values are shown in the Fig. 5.



**Fig. 2.** Plot to determine the closeness between predicted and actual data using CatBoost Regression



**Fig. 3.** Density Plot to determine distribution of variables in the dataset using CatBoost Regression



**Fig. 4.** Scatter plot to determine correlation between actual and predicted values using CatBoost Regression

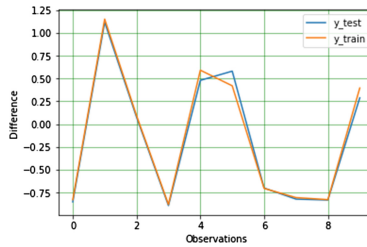
**Table 4.** The variance between actual and the predicted sales price (sample) and if it is Accurate or not in the CatBoost Regression model

S. No	Actual selling price	Predicted selling price	Difference	ACC or not	Percentage error
0	-0.852481	-0.773279	-0.079203	Accurate	-9.290831
1	1.117623	1.291965	-0.174343	Accurate	15.599416
2	0.076051	0.135253	-0.059202	Accurate	77.844952
3	-0.892852	-0.821449	-0.071404	Accurate	-7.997279
4	0.479761	0.618706	-0.138945	Accurate	28.961353
5	0.580689	0.414865	0.165824	Accurate	28.556397
6	-0.701090	-0.752449	0.051359	Accurate	-7.325561
7	-0.822203	-0.743478	-0.078726	Accurate	-9.574965
8	-0.832296	-0.832362	0.000066	Accurate	-0.007872
9	0.287999	0.336501	-0.048502	Accurate	16.841112
10	0.944028	0.618321	0.325706	Accurate	34.501768

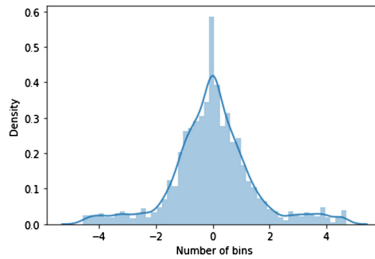
```
R2 SCORE is 0.9438514248527179
mean_sqrd_error is 0.06514319729907371
Root mean squared error of is 0.255231654187081
Mean Absolute error is 0.1371290433308082
Accuracy is 0.8628709566691918
```

**Fig. 5.** Evaluation metrics values for CatBoost Regression model

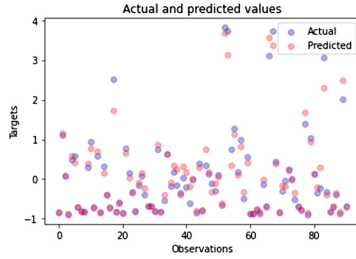
The closeness between predicted and actual data Random Forest Regression is shown in the Fig. 6. Density Plot to determine distribution of variables in the dataset is shown in Fig. 7. Scatter plot to determine correlation between actual and predicted values are shown in Fig. 8.



**Fig. 6.** Plot to determine the closeness between predicted and actual data using Random Forest Regression



**Fig. 7.** Density Plot to determine distribution of variables in the dataset using Random Forest Regression



**Fig. 8.** Scatter plot to determine correlation between actual and predicted values using Random Forest Regression

The Table 5 shows the variance between actual and the predicted sales price (sample) and if it is accurate or not in the Random Forest Regression model. The evaluation metrics used in Random Forest Regression model are R2 score, MSE, RMSE, MAE and accuracy and their values are shown in the Fig. 9.

**Table 5.** The variance between actual and the predicted sales price (sample) and if it is accurate or not in the Random Forest Regression model

S. no	Actual selling price	Predicted selling price	Difference	ACC or not	Percentage error
0	-0.852481	-0.828221	-0.024261	Accurate	-2.845921
1	1.117623	1.151512	-0.033889	Accurate	3.032278
2	0.076051	0.092846	-0.016794	Accurate	22.082912
3	-0.892852	-0.880713	-0.012140	Accurate	-1.359637
4	0.479761	0.591234	-0.111472	Accurate	23.234974
5	0.580689	0.420214	0.160475	Accurate	27.635238
6	-0.701090	-0.705139	0.004049	Accurate	-0.577559
7	-0.822203	-0.805542	-0.016661	Accurate	-2.026398
8	-0.832296	-0.828136	-0.004160	Accurate	-0.499850
9	0.287999	0.395880	-0.107881	Accurate	37.458948
10	0.944028	0.769627	0.174401	Accurate	18.474107

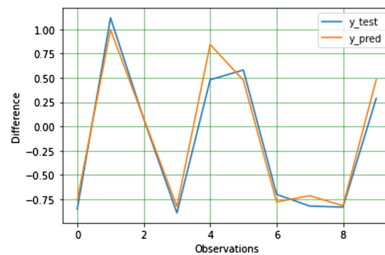
```

R2 SCORE is 0.9621157252658848
mean_sqrd_error is 0.04395307943368595
Root mean squared error of is 0.20964989728994848
Mean Absolute error is 0.1283807660158989
Accuracy is 0.8716192339841011

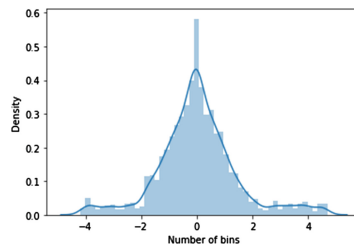
```

**Fig. 9.** Evaluation metrics values for random forest regression model

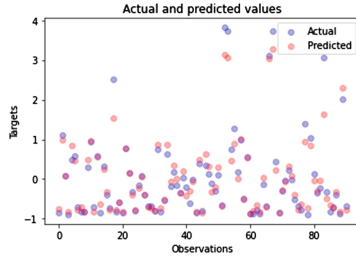
The closeness between predicted and actual data Support Vector Regression (SVR) is shown in the Fig. 10. Density Plot to determine distribution of variables in the dataset using SVR is shown in Fig. 11. Figure 12 displays a scatter plot using SVR to show the correlation between the actual and expected values. The variance between actual and the predicted sales price (sample) is shown in the Table 6 and if it is accurate or not in the Support Vector Regression model. The evaluation metrics used in Support Vector Regression model are R2 score, MSE, RMSE, MAE and accuracy and their values are shown in the Fig. 13.



**Fig. 10.** Plot to determine the closeness between predicted and actual data using Support Vector Regression (SVR)



**Fig. 11.** Density Plot to determine distribution of variables in the dataset using SVR



**Fig. 12.** Scatter plot to determine correlation between actual and expected values using SVR

**Table 6.** The variance between actual and the expected sales price (sample) and if it is Accurate or not in the SVR model

S. No	Actual selling price	Predicted selling price	Difference	ACC or not	Percentage error
0	-0.852481	-0.760063	-0.092418	Accurate	-10.841074
1	1.117623	0.995251	0.122372	Accurate	10.949334
2	0.076051	0.081998	-0.005947	Accurate	7.819079
3	-0.892852	-0.827224	-0.065628	Accurate	-7.350375
4	0.479761	0.843538	-0.363777	Accurate	75.824640
5	0.580689	0.474667	0.106022	Accurate	18.257982
6	-0.701090	-0.779501	0.078411	Accurate	-11.184135
7	-0.822203	-0.715066	-0.107138	Accurate	-13.030565
8	-0.832296	-0.818867	-0.013429	Accurate	-1.613529
9	0.287999	0.479512	-0.191513	Accurate	66.497896
10	0.944028	0.954947	-0.010920	Accurate	1.156733

```

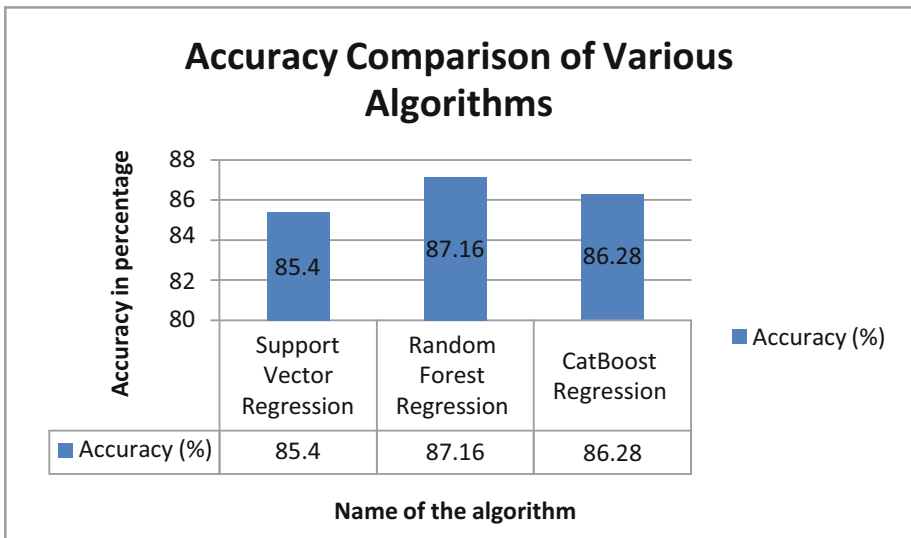
R2 SCORE is 0.9453731791724593
mean_sqrd_error is 0.06337766822497694
Root mean squared error of is 0.25174921693021596
Mean Absolute error is 0.14590172524212977
Accuracy is 0.8540982747578703
    
```

**Fig. 13.** Evaluation metrics values for SVR model

Comparison of various measures like R2 score, MSE, RMSE, MAE and accuracy using Support Vector Regression, Random Forest Regression, and CatBoost Regression and their values are shown in the Table 7. When compared to Support Vector Regression and CatBoost Regression, Random Forest Regression has the highest accuracy of 87.16% and it is shown in the Fig. 14. Random Forest Regression gives the lowest MAE of 0.1284 when compared to Support Vector Regression and CatBoost Regression and it is shown in the Fig. 15.

**Table 7.** Comparison of various measures used in various algorithms

	Support vector regression	Random forest regression	CatBoost regression
R2	0.9454	0.9621	0.9439
MSE	0.0634	0.044	0.0651
RMSE	0.2517	0.2096	0.2552
MAE	0.1459	0.1284	0.1371
Accuracy (%)	85.4	87.16	86.28



**Fig. 14.** Accuracy Comparison of Various Algorithms

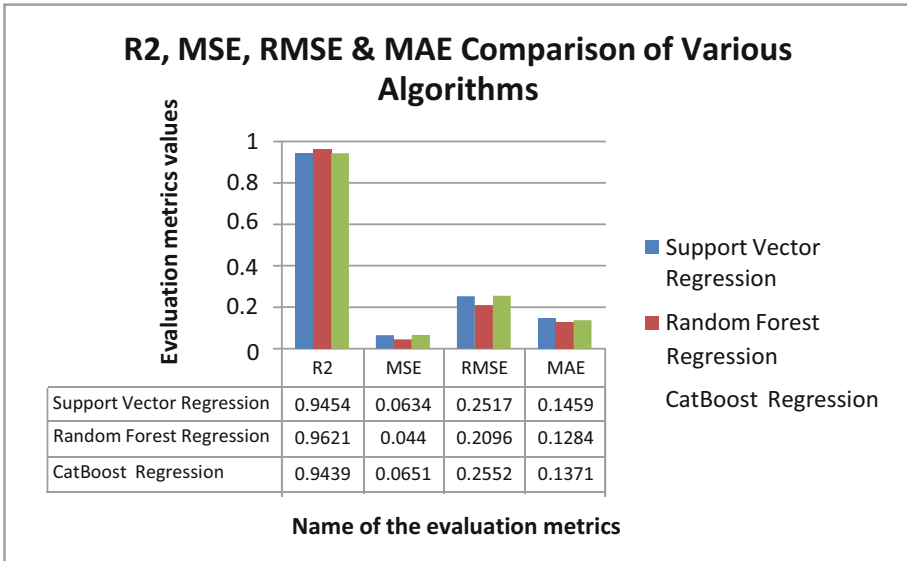


Fig. 15. R2, MSE, RMSE & MAE Comparison of Various Algorithms

## 6 Conclusion and Future Work

This article uses three different machine learning methods to predict the cost of used cars. In the existing method [13], mean absolute error for decision tree regression was 0.6711, which was the least used among the various algorithms that they employed, including linear regression, Lasso regression, Ridge regression, Bayesian Ridge regression, and others. The proposed system concludes that the Random Forest regression performs better than Support Vector Regression and CatBoost Regression. When compared to Support Vector Regression and CatBoost Regression, Random Forest Regression has the best accuracy (87.16%). Support Vector Regression gives the highest MAE of 0.1459 when compared to and Random Forest Regression and CatBoost Regression.

In this study, the primary problem is the little number of records used. We intend to collect additional data and use more advanced techniques, such as fuzzy logic, genetic algorithms, and deep learning approaches to forecast car costs in the future.

## References

1. Van Thai, D., Ngoc Son, L., Tien, P.V., Nhat Anh, N., Ngoc Anh, N.T.: Prediction car prices using quantify qualitative data and knowledge-based system. In: 2019 11th International Conference on Knowledge and Systems Engineering (KSE), Da Nang, Vietnam, pp. 1–5 (2019)
2. Sameerchand, P.: Predicting the price of used cars using machine learning techniques. *Int. J. Inform. Comput. Technol.* **4**, 753–764 (2014)
3. Fathalla, A., Salah, A., Li, K.: Deep end-to-end learning for price prediction of second- hand items. *KnowlInf Syst.* **62**, 4541–4568 (2020)

4. Xia, Z., Xue, S., Wu, L.: ForeXGBoost: passenger car sales prediction based on XGBoost. *Distrib. Parallel Databases* **38**, 713–738 (2020)
5. Viswapriya, S.E., Durbaka, S., Sandeep, S., Gandavarapu, S.K.: Vehicle price prediction using SVM techniques. *Int. J. Innovative Technol. Exploring Eng.* **9**, 398–401 (2020)
6. Aashay, P.H., Saurabh, S.: Inspecting & cost evaluation of pre-owned cars using deep learning. *Int. Res. J. Eng. Technol.* **7**, 7578–7582 (2020)
7. Chuancan, C., Lulu, H., Cong, X.: Comparative analysis of used car price evaluation models. In: 2017 AIP Conference Proceedings, vol. 1839, pp. 1–7 (2017). <https://doi.org/10.1063/1.4982530>
8. Sun, N., Bai, H., Geng, Y., Shi, H.: Price evaluation model in second-hand car system based on BP neural network theory. In: 2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), pp. 431–436 (2017). <https://doi.org/10.1109/SNPD.2017.8022758>
9. Han, S., Qu, J., Song, J., Liu, Z.: Second-hand car price prediction based on a mixed-weighted regression model. In: 2022 7th International Conference on Big Data Analytics (ICBDA), pp. 90–95 (2022). <https://doi.org/10.1109/ICBDA55095.2022.9760371>
10. Yu, L., Yan, H., Liu, L., Yi, S., Zhao, Y.: Car trading cycle prediction based on random forest algorithm. In: 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP). Pp. 639–642 (2022). <https://doi.org/10.1109/ICSP54964.2022.9778603>
11. Chen, S., Liu, Z.: Application of data mining technology in second-hand car price forecasting. In: 2022 3rd International Conference on Electronic Communication and Artificial Intelligence (WECAI), pp. 260–273 (2022). <https://doi.org/10.1109/TWECAI55315.2022.00058>
12. Satapathy, S.K., Vala, R., Virpariya, S.: An automated car price prediction system using effective machine learning techniques. In: 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), pp. 402–408 (2022). <https://doi.org/10.1109/CISES54857.2022.9844350>
13. Sharma, A.D., Sharma, V., Mittal, S., Jain, G., Narang, S.: Predictive analysis of used car prices using machine learning. *Int. Res. J. Modernization Eng. Technol. Sci.* **3**, 674–684 (2021)