OPEN ACCESS
Indo-JC Indonesia Journal of Computing
ISSN 2460-9056
socj.telkomuniversity.ac.id/indojc

Ind. Journal on Computing
Vol. 9, Issue. 3, December 2024. pp. 198-207
doi:10.34818/indojc.2024.9.3.975

Car Price Prediction Using Artificial Neural Networks: A Data-Driven Approach

Lukman A. Ogundele [#1], Adekunle O. Ejidokun [*2], Femi E. Ayo [#3], Abass I. Taiwo [#4]

[#] *Department of Mathematical Sciences, Olabisi Onabanjo University, Ago-Iwoye, Nigeria*
[1] ogundele.lukman@oouagoiwoye.edu.ng
[3]ayo.femi@oouagoiwoye.edu.ng
[4] taiwo.abass@oouagoiwoye.edu.ng

[*] *Department of Computer Sciences, Kampala International University, Kampala, Uganda*
[2] ejidokun.olugbenga@kiu.ac.ug

Abstract

Used cars suffer from depreciation and require reevaluation from time to time to ascertain the actual price at which the car can be purchased or sold by buyers and sellers. Car price prediction is important because of the increase in the rate of purchase of used car compared with that of new cars due to inflation, fluctuation in exchange rates, currency devaluation and so on. To address the issues of accuracy and error rate, this work suggests a hybrid feature selection approach that extracts the most crucial properties from the dataset. The most important attributes in the dataset were then used as input for the prediction phase using deep learning approach. The deep learning model's output is contrasted with that of other machine learning techniques to identify the most effective approach. In comparison to the Decision Tree and Support Vector Machine (SVM) models, which performed at 87.8% and 88.3%, respectively, the suggested hybrid feature selection using deep learning model attained an accuracy of 96.9%, according to the evaluation data. However, the other two classifiers indicate a lower error rate as compared to the ANN model.

Keywords: Hybrid feature selection, Correlation, Prediction, Performance, Used car data

Abstrak

A Mobil bekas mengalami depresiasi dan memerlukan penilaian ulang dari waktu ke waktu untuk menentukan harga sebenarnya di mana mobil dapat dibeli atau dijual oleh pembeli dan penjual Prediksi harga mobil penting karena peningkatan tingkat pembelian mobil bekas dibandingkan dengan mobil baru akibat inflasi, fluktuasi nilai tukar, devaluasi mata uang, dan sebagainya. Untuk mengatasi masalah akurasi dan tingkat kesalahan, pekerjaan ini menyarankan pendekatan pemilihan fitur hibrida yang mengekstrak properti paling penting dari himpunan data. Atribut terpenting dalam dataset kemudian digunakan sebagai input untuk fase prediksi menggunakan pendekatan deep learning. Hasil dari model deep learning dibandingkan dengan metode machine learning lainnya untuk menentukan metode dengan performa optimal. Dibandingkan dengan model Decision Tree dan Support Vector Machine (SVM), yang masing-masing berkinerja 87,8% dan 88,3%, model pembelajaran mendalam yang disarankan berdasarkan pemilihan fitur hibrida mencapai akurasi 96,9%, menurut data evaluasi. Namun, dua pengklasifikasi lainnya menunjukkan tingkat kesalahan yang lebih rendah dibandingkan dengan model ANN.

Kata Kunci: Pemilihan fitur hybrid, Korelasi, Prediksi, Performa, Data mobil beka

Received on August, 2024. Accepted on December, 2024

I. INTRODUCTION

The automobile market plays a significant role in the global economic activity. This is because rapid technological development has changed people's way of life and influenced how various items ar transported from one place to another [1]. The demand for new and used automobiles has soared worldwide [2]. Retailers are in charge of selling both new and used cars, and a recent study by [3, 4] found that the used cars market has grown more than that of new cars.

New cars of the same make, model, and year have their prices set by manufacturers, which are fairly the same except for optional features [5]. However, it is difficult to determine the price of used cars, which leads to cheating and manipulation by retailers of the product [2, 6, 7, 8]. This is because they are subject to supply and demand pricing, and additional attributes including the number of years the vehicle was manufactured, brand, distance (the number of kilometers driven), condition, repair history, engine size, type, and capacity, version, city, colour, alloy rims, power steering, seat type, and fuel type, among others. Therefore, there is an urgent need for a standardised system for predicting used cars to avoid cheating and manipulation of customers that increases accuracy and reduces error rates.

Prediction can be done based on various machine learning techniques using certain parameters can be helpful in this regard [9, 10, 11]. Diverse machine learning methods are available for forecasting. Supervised and unsupervised learning are the two categories of machine learning algorithms. The input and desired output must be provided by the user in the supervised learning. The algorithm must be trained using the data, and new inputs must be predicted using the system's newly acquired knowledge. The actual value of the output values is used in supervised learning for training. Therefore, given a sample of data and actual outputs, the idea of supervised learning is to produce a model that best predict the actual output value based on a given input value. On the other hand, unsupervised learning does not require training in order to get the desired result. Its objective is to predict output based on the structure of the attributes in the dataset.

A basic example of a supervised learning system is the Neural Network Classifier, which classifies instances that are unclassified though weight adjustment using the errors in the trained network. Learners who utilize this categorization system are referred to as memory-based classifiers, lazy classifiers, and instance-based classifiers. They are all affected by the same issue. Indiscriminate storage is done using the examples used to train the classifier [12]. There is no process of selection with instance-based learning, and as a result, unnecessary instances are stored, which results in a less accurate classification scheme and a high error rate. This is because the data associated with car price prediction is very large, and there are thousands of used cars. Since the dataset for each car has a huge number of features, a feature selection technique is required [9, 13, 14]. In order to overcome the difficulties found, a hybrid feature selection method for car price prediction employing an artificial neural network classifier is required.

II. LITERATURE REVIEW

In [9], research was carried out on a vehicle price prediction system with the aid of multiple linear regression which offered a prediction accuracy of 95% measured in terms of Predicted Interval (PI) and Confidence Interval (CI). According to [15], an ensemble model for forecasting used cars price in Bosnia and Herzegovina was constructed using random forest, support-vector-machine, and artificial neural network. The accuracy of 87.38%, which is lower than that of multiple linear regression, was found when the model was assessed using test data. A correlation matrix is utilized to identify and extract features, and a linear regression approach is used to estimate pricing for used automobiles, comparing its accuracy with the classification algorithm. By using the second data set to train the model, prediction accuracy rises to 95%. [16] determined used car sales prices using colour, transmission, and city variables and had an accuracy of 76.2%. It was suggested that more variables should be used for prediction which ultimately will lead to higher accuracy in predicting prices. [17] when further employing several supervised machine learning-based regression algorithms to project the reselling value of used cars using several determinant factors including the car's model, year of manufacturing, fiscal power, mileage, and fuel type. In every model that was examined, the gradient boosting regression tree showed a low error and a high regression. [18] developed a hybrid model for car price prediction. The model was trained on a public dataset. The study determined the value of an automobile by factors including the number of kilometres driven, the year of registration, the kind of fuel, and financial resources. The study used Regression Tree (CART) and K Nearest Neighbor (KNN) and compared the two on various car models. According to the study, the actual price was 4999, however the error in regression for k-NN with k = 7 is 5581.96 and for CART is 4961.6. [19] developed a feature-based machine learning method for car price prediction. The study employed linear regression and designed a p-value for obtaining the most suitable features. The study used variance inflation factor (VIF) for the Ordinary Least Squares regression (OLS) after using the Recursive feature elimination (RFE) to determine the best features. The simulation outcomes of the study demonstrated the effectiveness of their method. [20] proposed a K-NN regression model for used car price prediction. The study trained their proposed model on used automobile data. In their simulation outcome, several percentages of the data were used to evaluate their method. Therefore, the simulation outcome showed the effectiveness of their method with improved accuracy for prediction of used car price. [21] proposed a used car prediction based on a deep learning method and using a publicly available used car data to test their method. The performance of their method was verified using a test set of 35,000 old vehicles. Several characteristics are looked at to provide precise and dependable predictions. The Keras regression technique was utilized to develop an Artificial Neural Networks (ANN), and their effectiveness is evaluated against more fundamental models like random forests, gradient boosting, decision tree algorithms, and linear regression. To enhance prediction performance, embedding techniques were applied to categorical data. The results showed a considerable improvement over the baseline model and are consistent with actual values. An ANN model with regression value of 0.96 and regression error of 11% outperformed other methods. In [22], the study seeks to determine the key variables influencing electric vehicle sales and to determine how important these variables are for predicting electric vehicle sales. The study used the literature to identify twelve important criteria driving electric vehicle sales. These elements were also ranked according to expert judgment and the DEMATEL technique, and the dataset for electric vehicle sales forecasts included the influential weights of these components. Additionally, a variety of machine learning methods were used to forecast future electric vehicle sales, including XGBoost, Random Forest, ARIMA, and LSTM. Out of all the machine learning techniques, the Random Forest method was shown to work the best, capturing complicated correlations and allowing for accurate prediction. [23] employed decision tree and random forest techniques for the analysis of automobile prediction. The accuracy for the decision tree analysis showed a value of 0.6721, while the random forest showed a better value of 0.7213. [24] observed that maintaining a vehicle's high quality will prolong its lifespan, improve customer satisfaction, reduce maintenance issues and increase the price of the used vehicle. The study introduces a computational framework that utilizes interpretable machine learning methods in measuring the quality of vehicles. The suggested interpretable machine learning method was found to performed well when evaluated on a publicly available dataset.

III. Research Method

A. Dataset Description

The University of California, Irvine C.A. Center (UCI) for Intelligent Systems provided the dataset used in this project. There are 205 cases in the collection. There are 26 attributes in total for every instance in the dataset. Table 1 provides a comprehensive explanation of the dataset.

TABLE 1
PREDICTION DATASET ON CAR PRICE

S/N	ATTRIBUTE	DATA TYPE	ATTRIBUTE RANGE
1.	Symboling	Numeric	-3 to +3
2.	NormalizedLosses	Numeric	65 to 256
3.	Model	Symbolic	Alfa-romero, Audi, BMW, Chevrolet, Dodge, Honda, Isuzu, Jaguar, Mazda, Mercedes-Benz, Mercury, Mitsubishi, Nissan, Peugeot, Plymouth, Porsche, Renault, SAAB, Subaru, Toyota, Volkswagen, Volvo.
4.	FuelType	Symbolic	Diesel, Gas

5.	Aspiration	Symbolic	Std, Turbo
6.	NumberOfDoors	Numeric	Four, Two.
7.	BodyStyle	Symbolic	Hardtop, Wagon, Sedan, Hatchback, Convertible.
8.	DriveWheels	Symbolic	4wd, fwd, rwd
9.	EngineLocation	Symbolic	Front, Rear.
10.	WheelBase	Numeric	86.6 to 120.9
11.	Distance	Numeric	141.1 to 208.1
12.	Breadth	Numeric	60.3 to 72.3
13.	Altitude	Numeric	47.8 to 59.8
14.	CurbWeight	Numeric	1488 to 4066
15.	EngineType	Symbolic	Dohc, dohcv, l, ohc, ohcf, ohcv, rotor.
16.	EngineSize	Numeric	61 to 326
17.	NumberOfCylinders	Numeric	Eight, five, four, six, three, twelve, two.
18.	FuelSystem	Symbolic	1bbl, 2bbl, 4bbl, idi, mfi, mpfi, spdi, spfi.
19.	Bore	Numeric	2.54 to 3.94.
20.	Stroke	Numeric	2.07 to 4.17
21.	CompressionRatio	Numeric	7 to 23
22.	Horsepower	Numeric	48 to 288
23.	PeakRPM	Numeric	4150 to 6600
24.	CityMPG	Numeric	13 to 49
25.	HighwayMPG	Numeric	16 to 54
26.	Amount	Numeric	5118 to 45400

B. Dataset Preprocessing

Using the hybrid feature selection strategy, the training dataset was reduced from 26 characteristics to 8 attributes. Improving the outcome of the forecast phase is the goal of the reduction. A hybrid feature selection and artificial neural network (ANN) were the foundations of the study effort on car price prediction. In the hybrid feature selection process, the optimal feature subset was identified using correlation-based feature selection (CFS), which was then coupled with genetic search as the search method. Selecting the most significant features from the dataset is how CFS and Genetic Algorithms help to increase prediction accuracy.

Utilising the Genetic Algorithm (GA) as the search technique, CFS was employed as the subset evaluator (fitness function). CFS was used to evaluates the value of a subset characteristics by considering the inter-correlations among the features and the link between a subset of attributes and the target class label, as well as the individual predictive power and amount of redundancy of each feature. As intercorrelation increases, a set of features' relevance diminishes and increases with the connection between the attributes and classes [25]. Next, the neural network algorithm that will estimate automobile prices uses the features that have been chosen as input.

C. Mathematical Model

This part is used to discuss the mathematical models and formulations used in this research. These are as follows:

1) Attribute: An attribute A_i will be considered important if and only if there is a_i and class (d) for which $p(A_i = a_i) > 0 \ni$

$$p(D = d|A_i = a_i) \neq p(D = d) \tag{1}$$

2) Correlation: As seen in (2), if one knows the correlation between each component in the test and the external variable as well as the inter-correlation between every pair of components, one can predict the correlation between a composite test composed of the summed components and the external variable.

$$x_{zc} = \frac{y\bar{x}_{zi}}{\sqrt{y + y(y-1)\bar{x}_{ii}}} \tag{2}$$

where x_{zc} denote the connection among the totaled components and the external variable, y represents the number of components, $\bar{x}_{zi}$ is the mean of the connections between the components and the outside attribute, and $\bar{x}_{ii}$ is the mean inter-correlation among components.

3) Genetic search: This is a search stimulated by natural processes. The fitness function employed in this genetic search is a linear mixture of an accuracy term and a simplicity term:

$$f(x) = \frac{3}{4}g + \frac{1}{4}\left(1 - \frac{E+A}{2}\right) \tag{3}$$

where A is the number of subset features, g is the average cross-validation accuracy of an ANN, E is the number of instances or training samples, and x is a feature subset.

4) Price prediction: This is the weighted sum of the product of inputs x and weights w for the suggested price prediction using ANN. A bias b_k can be added to normalize the net of the neural network.

$$net_k = x_1 w_{k1} + x_2 w_{k2} + \cdots x_m w_{km} + b_k \tag{4}$$

5) Output: The predicted output y_k of the suggested ANN from the net_k using some specified activation function f is computed using

$$y_k = f(net_k) \tag{5}$$

6) The error: The error e_k of the suggested ANN for weight adjustment by comparing the target output d_k with the predicted output y_k is defined as

$$e_k = d_k - y_k \tag{6}$$

7) *The weight adjustment*: According to the delta rule, the weight adjustment is defined as

$$\Delta w_{kj}(n) = \eta . e_k . x_j(n) \tag{7}$$

where $\Delta w_{kj}(n)$ is the change in weight at the input instance n, η is the learning rate of the network, e_k is the error of the suggested ANN and $x_j(n)$ is the input at node j.

8) The change in weight: $w_{kj}(n+1)$ is the change in weight computed from the addition of the original weight $w_{kj}(n)$ and the new weight $\Delta w_{kj}(n)$ of the suggested ANN defined as

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n) \tag{8}$$

D. Architectural Framework

The steps of the CFS method and its application with the Genetic Search algorithm are depicted in Figure 1. The CFS receives a copy of the Car training data first. Following the computation of feature-class and feature-feature correlations, CFS searches the feature subset space using the Genetic Search method. The full training and testing data are made less dimensional by using the subset with the highest merit that was discovered throughout the search. The ANN algorithm may then be given both reduced datasets for testing and training. Ultimately, the final forecast is generated by the ANN algorithm once it has evaluated the dataset.

E. Implementation

There are three stages in the implementation of this car price prediction system. They include:

1) Preprocessing phase: This is the phase where feature evaluation takes place. Here the car dataset is fed into the CfsSubsetEval and the Genetic algorithm (hybrid feature selectors) for the extraction of the most essential attributes. For the ANN prediction to become more accurate, two feature selectors are required.

2) Processing phase: The resulting attributes from the preprocessing phase are used to train the ANN algorithm.

3) Prediction phase: The ANN model is built with Java and then used for the final evaluation and prediction of car prices based on the reduced dataset.

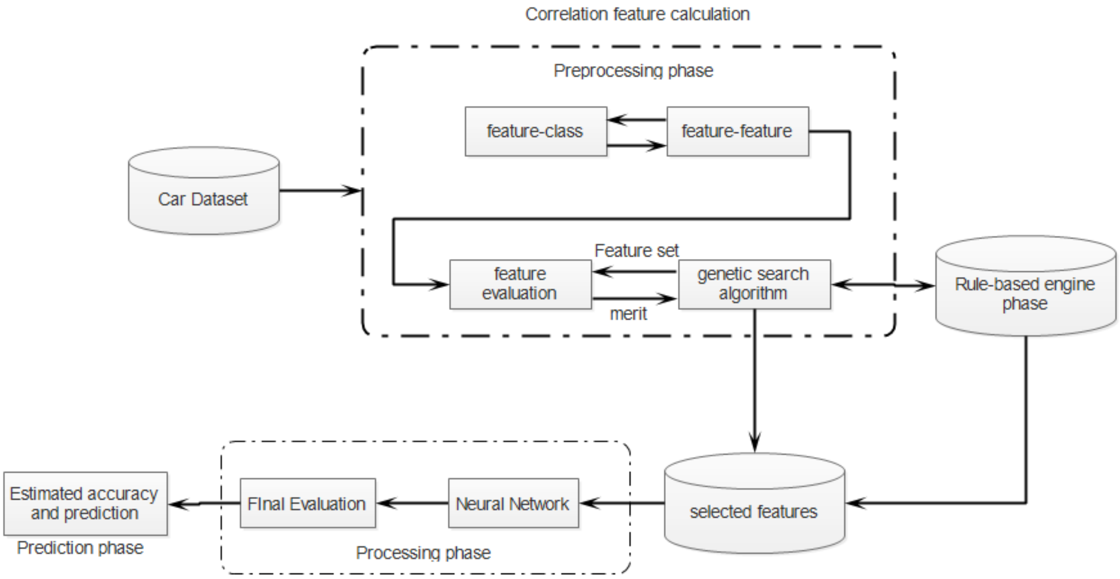


Fig. 1. Proposed System Architecture

F. Algorithms Used

The algorithms for hybrid feature selection, creating training dataset and test dataset, mlp-Algorithm for training the Multilayer perceptron network and price prediction is presented as follows:

1) Algorithms for Hybrid Feature Selection:

Input: $S(F_1, F_2, \ldots, F_k, F_c)$ *//car training dataset*
Output: S_{best} *//selected features*
1. Start by creating an initial population P at random.
2. Find the value of f(x:CfsSubsetEval) for every member x in P.
3 Over the members of P, define a probability distribution p such that p(x) / f(x).
4. Pick two individuals from the population, x and y, with relation to p.
5. Utilize crossover on x and y to create x' *and* y'*, new population members.*
6. Modify x' *and* y' *using mutation.*
7. Add x' *and* y' *into the next generation,* P'.
8. If $|P'| < |P|$*, goto 4.*
9. Let $P \leftarrow P'$
10. Go to Step 2 if there are more generations to process.
11. Return back $x \in P$ *where f(x) is largest.*

2) Algorithm to Create Training Dataset and Test Dataset:

Input: DataSet

Output: TrainingSet, TestSet

1. *SET DataSet.Count.FirstInstances (15000) as trainingSetSize*
2. *SET DataSet.Count – trainingSetSize as testSetSize*
3. *SET DataSet.Take(trainingSetSize) as TrainingSet*
4. *SET DataSet.Skip(trainingSetSize).Take(testSetSize) as TestSet*

3) mlp-Algorithm for training the Multilayer perceptron network

Input: Input instances p
Output: Outputweight Weight(t)

```
1. Begin with a random initial weight (uniform random in [-.3,.3], for example).
2. For all instances p
3.     For All Output Node j
4.         Calculate Activation(j)
5.          For All Input Nodes i To Output Node j
6.             Calculate DeltaWeight = LearningConstant * Error_j * Activation_i
7.             Weight (t) = Weight + DeltaWeight until Error is sufficiently small
8. Return Weight(t)
```

4) Algorithm for price prediction

Input: Input instances p
Output: Output class Ci

```
1. p ← Read input instances
2. //CALL mlp-Algorithm
3. For i = length of p
4.     For j = length of text t
5.         if (j = i) and Ci = t[j]
6.             assign text pattern to a class
7.         Else
8.             increment i by 1
9. Return Ci
```

IV. Results And Discussion

The ANN classifier correctly classified 205 instances with 26 attributes with 96.9% classification accuracy. The prediction results showed that the Artificial Neural Network is a trustworthy technique for predicting car prices based on the test dataset. The output of the proposed ANN algorithm with hybrid feature selection approach was evaluated using the Support Vector Machine (SVM) and Decision Tree methods. The comparison of three distinct classifiers (ANN, SVM, and Decision Tree techniques) is shown in Table 2. Figure 2 show the dataset sample editor for the car price prediction with instances corresponding to the attributes described in Table 1. The majority of the attributes in the dataset sample shown in Fig. 2 are numeric while a few others are symbolic attributes.

Additionally, Table 2 was used to compare the classification accuracy of several classifiers and reveals that the suggested ANN classifier has superior accuracy to the other two classifiers. Table 3 indicate the prediction outcomes of the system as compared to the target values. The results validate the correctness of the developed system since it is very close to the actual values. In order to further justify the accuracy of the ANN over the other two traditional classifiers, Figure 3 displays a comparison graph of classification accuracy between the two classical classifiers and the artificial neural network. The evaluation comparison indicates that the developed ANN approach achieved a better performance accuracy of 96.9% as compared to the other two traditional classifiers with a reduced error rate. This proposed system will address the problems of price prediction of used cars which is difficult to achieve because the safety and roadworthiness of used cars cannot be ascertained easily.

This work's limitation is that its correlation-based selection of features method might not be able to detect significant interactions within the attributes in the dataset. In order to operate, the process also needs more memory and processing power.

Figure 4 show the comparison of the correlation coefficient with scatter plot of the proposed feature selection-based ANN with related machine learning algorithms in order to increase the evidence of the study. The correlation coefficient with scatter plot showed that the proposed feature selection-based ANN attained the best prediction value with 96.9% correlation coefficient when compared to decision tree and SVM with 87.8% and 88.3% respectively.

TABLE 2
COMPARISON OF CLASSIFIERS

S/N	EVALUATION METRIC	ANN(%)	DECISION TREE(%)	SVM(%)
1.	Correlation coefficient	96.9	87.8	88.3
2.	Mean absolute error	2	0.1	0.3
3.	Root mean squared error	2	0.3	0.3

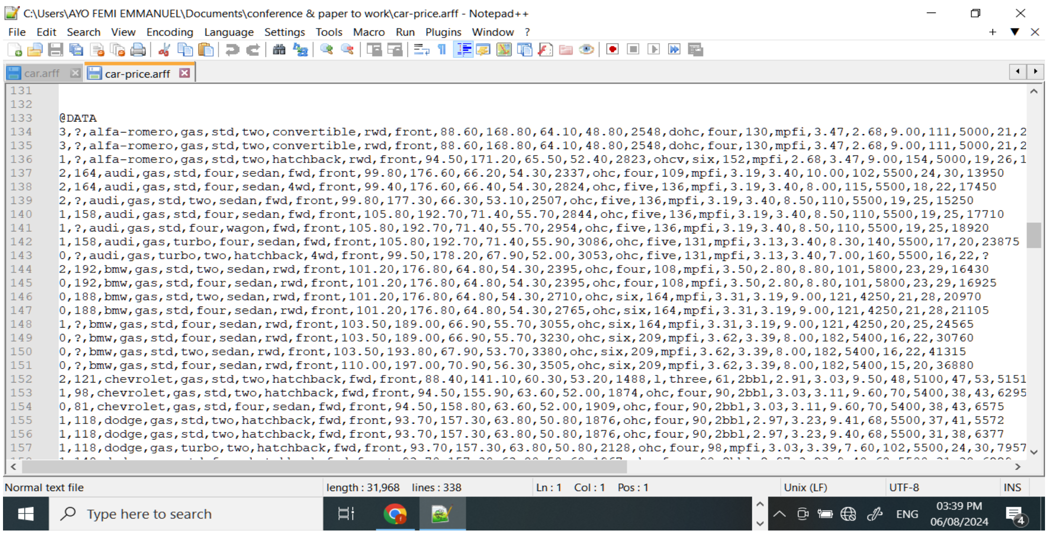
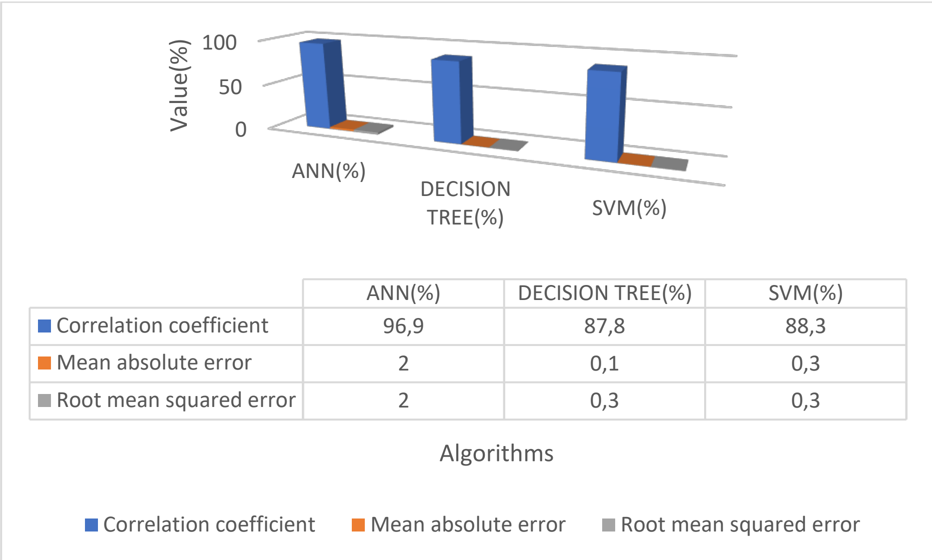


Fig. 2. Dataset sample editor

TABLE 3
PREDICTION RESULT COMPARISON

S/N	TARGET PREDICTION	ACTUAL PREDICTION
1.	13495	14428.94593127
2.	7463	7981.917281814676
3.	24565	17330.672968295952
4.	12964	12665.321555773817
5.	11048	10461.02254511755
6.	8495	8905.128452193101
7.	16503	16607.90855582892
8.	5499	7505.63078734426
9.	17199	16299.810817619253
10.	12170	12884.225999254018



	ANN(%)	DECISION TREE(%)	SVM(%)
Correlation coefficient	96,9	87,8	88,3
Mean absolute error	2	0,1	0,3
Root mean squared error	2	0,3	0,3

Fig. 3. Evaluation of Classifier Performance

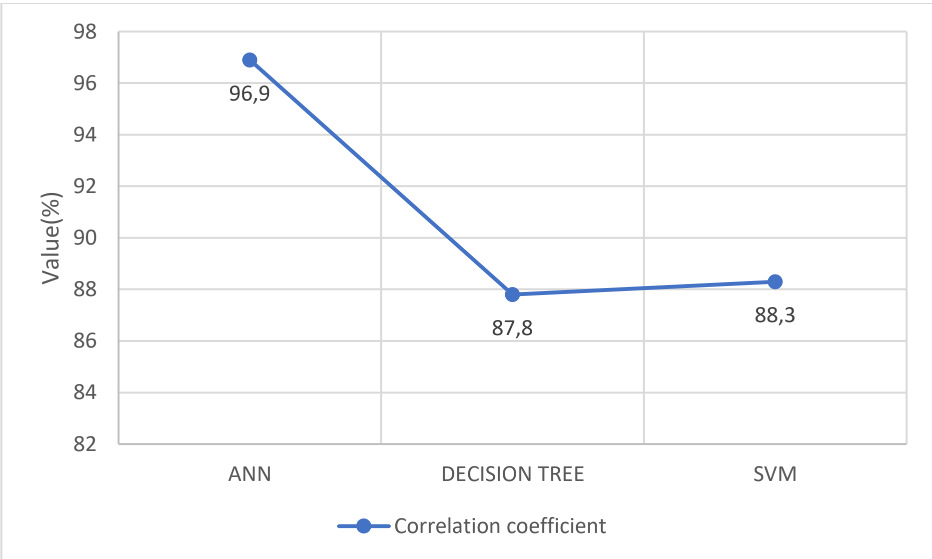


Fig. 4. Correlation coefficient with scatter plot

V. Conclusion

This study was used to discuss used car prediction price where the issues of accuracy and error rate was addressed using a hybrid feature selection approach that extracts the most crucial properties from the dataset using deep learning approach. The adopted hybrid approach was based on a correlation-based feature selection and Genetic Algorithm. The result from the deep learning model signified an optimal performance than the other machine learning methods considered. In comparison to the Decision Tree and Support Vector Machine (SVM) models, which performed at 87.8% and 88.3%, respectively, the suggested a hybrid feature selection-based deep learning model attained an accuracy of 96.9%, according to the evaluation data. However, the other two classifiers indicate a lower error rate as compared to the ANN model. Therefore, this study showcases the ability of hybrid feature selection that includes the evaluation of the relationship between the attributes and a genetic search method since the results obtained was better when compared with previously used approaches when predicting used car price. The study has added to the body of scientific knowledge with a hybrid feature selection approach for used car price prediction.

REFERENCES

[1] A. A. Alhakamy, A. Alhowaity, A. A. Alatawi and H. Alsaadi, "Are Used Cars More Sustainable? Price Prediction Based on Linear Regression," *Sustainability,* vol. 15, no. 2, p. 911, 2023.

[2] Abishek, "Car Price Prediction using Machine Learning Techniques," *International Research Journal of Modernization in Engineering Technology and Science,* vol. 4 , no. 2, pp. 45-50, 2022.

[3] E. Costa, P. Wells, L. Wang and G. Costa, " The electric vehicle and renewable energy: Changes in boundary conditions that enhance business model innovations," *Journal of cleaner production,* vol. 333, p. 130034., 2022.

[4] D. M. &. N. R. C. Gomes, "Techno-economic analysis of vehicle-to-grid technology: Efficient integration of electric vehicles into the grid in Portugal.," *Journal of Energy Storage,* vol. 97, p. 112769, 2024.

[5] M. Collard, "Price prediction for used cars: a comparison of machine learning regression models.," Mid Sweden University,Bsc. Thesis Computer engineering, 2022.

[6] C. Trivendra, " The Price Prediction for Used Cars using Multiple Linear Regression Model.," *International Journal for Research in Applied Science and Engineering Technology,* pp. 1801-1804, 2020.

[7] J. Nilsson, J. Carlsson Hauff and A. Carlander, " How consumers evaluate costs in complex services: the case of mutual fund fees.," *European Journal of Marketing,* vol. 58, no. 13, pp. 56-80, 2024.

[8] A. Goel and G. Sahil, " An analysis of policy prospective of taxi aggregators and consumers in digital eco-system.," *International Journal of Intelligent Enterprise,* vol. 11, no. 2, pp. 120-140., 2024.

[9] K. Noor and S. Jan, "Vehicle price prediction system using machine learning techniques.," *International Journal of Computer Applications,* vol. 167, no. 9, pp. 27-31, 2017.

[10] M. Méndez, M. G. Merayo and M. Núñez, "Machine learning algorithms to forecast air quality: a survey.," *Artificial Intelligence Review,* vol. 56, no. 9, pp. 10031-10066., 2023.

[11] J. Zhao, H. Ling, J. Liu, J. Wang, A. F. Burke and Y. Lian, "Machine learning for predicting battery capacity for electric vehicles.," *ETransportation,* vol. 15, p. 100214, 2023.

[12] M. A. Kutlugün and Y. Şirin, "Augmenting the training database with the method of gradual similarity ratios in the face recognition systems.," *Digital Signal Processing,* Vols. 135,, p. 103967, 2023.

[13] D. Kshirsagar and S. Kumar, "Towards an intrusion detection system for detecting web attacks based on an ensemble of filter feature selection techniques.," *Cyber-Physical Systems,* vol. 9, no. 3, pp. 244-259., 2023.

[14] F. E. Ayo, L. A. Ogundele, S. Olakunle, J. B. Awotunde and F. A. Kasali, " A hybrid correlation-based deep learning model for email spam classification using fuzzy inference system," *Decision Analytics Journal,* vol. 10, p. 100390, 2024.

[15] E. Gegic, B. Isakovic, D. Keco, Z. Masetic and J. Kevric, "Car price prediction using machine learning techniques.," *TEM Journal,* vol. 8, no. 1, p. 113, 2019.

[16] C. K. Puteri and L. N. Safitri, "Analysis of linear regression on used car sales in Indonesia," *In Journal of Physics: Conference Series ,* vol. 1469, no. 1, pp. 012143, IOP Publishing., 2020.

[17] M. Hankar, M. Birjali and A. Beni-Hssane, "Used car price prediction using machine learning: A case study.," in *In 2022 11th International symposium on signal, image, video and communications (ISIVC) (pp. 1-4)*, IEEE, 2022.

[18] A. Chandak, P. Ganorkar, S. Sharma, A. Bagmar and S. Tiwari, "Car price prediction using machine learning.," *International Journal of Computer Sciences and Engineering,* vol. 7, no. 5, pp. 444-450., 2019.

[19] M. Asghar, K. Mehmood, S. Yasin and Z. M. Khan, "Used cars price prediction using machine learning with optimal features.," *Pakistan Journal of Engineering and Technology,* vol. 4, no. 2, pp. 113-119., 2021.

[20] K. Samruddhi and R. A. Kumar, "Used car price prediction using K-nearest neighbor based model," *Int. J. Innov. Res. Appl. Sci. Eng.(IJIRASE),* vol. 4, no. 3, pp. 2020-686, 2020.

[21] A. S. Pillai, "A Deep Learning Approach for Used Car Price Prediction," *Journal of Science & Technology,* vol. 3, no. 3, pp. 31-50, 2022.

[22] S. Kumar, V. Singh and R. Goel, "Strategic Forecasting for Electric Vehicle Sales: A Cutting Edge Holistic Model Leveraging Key Factors and Machine Learning Technique," *Transportation in Developing Economies,* vol. 10, no. 2, p. 23, 2024.

[23] P. H. Putra, A. Azanuddin, B. Purba and Y. A. Dalimunthe, "Random forest and decision tree algorithms for car price prediction.," *Jurnal Matematika Dan Ilmu Pengetahuan Alam LLDikti Wilayah 1 (JUMPA),* vol. 3, no. 2, pp. 81-89., 2023.

[24] M. Alwadi, G. Chetty and M. Yamin, "A Framework for Vehicle Quality Evaluation Based on Interpretable Machine Learning," *International Journal of Information Technology,* vol. 15, no. 1, pp. 129-136., 2023.

[25] M. A. Hall, "Correlation-based feature selection for machine learning," *(Doctoral dissertation, The University of Waikato).,* 1999.