Electric Vehicle Price Prediction Using Machine Learning Techniques

A. Rakul[1], V. Santhosh[1], A. Kabilesh[1], J. Mithun Raj[1], T. Arachelvi[2]
UG Students[1], Associate Professor[2]
Computer Science and Engineering[1], Electronics and Communication Engineering[2], Tamilnadu College of Engineering, Coimbatore, India

Abstract-This Streamlit-based application uses a trained Random Forest regression model to forecast the volume of electric vehicle sales worldwide. A CSV file with test data can be uploaded by users, and the application will process it by eliminating missing data, cleaning out unnecessary columns, and converting values to numeric representation. A pre-trained scaler is then utilized to scale the cleaned data to the format specified during model training. The application uses metrics like Mean Squared Error and R-squared score to assess model performance and makes predictions based on the processed data. Interactive outputs, such as data tables, prediction results, and various visualizations including feature importance, actual versus predicted values, residual plots, and distribution of prediction errors, are displayed using Streamlit. When combined, these elements offer a user-friendly interface for examining the efficacy and behavior of the model.

Keywords: Electric car sales prediction, random forest regression, data visualization, streamlit application

1.INTRODUCTION

The use of electric vehicles has increased demand for data-driven techniques to comprehend and forecast industry trends. Making predictions about sales volumes based on a variety of technical and performance-related characteristics enables better analysis and decision-making. In order to facilitate this, Streamlit has been used to create a web-based interface that combines machine learning methods with intuitive visualization tools. A Random Forest regression model, which was trained using structured input data to estimate global sales volume, forms the basis of the system. Datasets in CSV format are available for user upload; these are automatically processed through a number of processes, such as feature scaling using a pre-fitted scaler, column filtering, numerical conversion, and handling of missing information. Following the preparation of the data, the model makes predictions and assesses its accuracy using performance measures like R-squared score and Mean Squared Error. The system produces visual insights in addition to numerical outputs, such as distribution graphs, feature importance analysis, residual diagnostics, and comparisons of anticipated and actual values. Users can evaluate the impact of different factors on sales forecasts and interpret model behavior with the use of these capabilities. By improving accessibility and interactivity, Streamlit makes it possible to explore machine learning results in an organized and understandable manner.

1.1 ELECTRIC CAR SALES PREDICTION

Using input data including technical specifications, performance indicators, and market-related variables, predictive modeling techniques are used to estimate the potential number of sales of electric cars. In order to identify underlying trends and patterns that affect purchasing decisions, these models examine historical data. Commonly used input features include things like battery capacity, efficiency ratings, range between charges, and car price. The objective is to generate accurate forecasts that can aid in research, planning, and performance assessment of the product. In the automotive industry, analysts and developers may enhance the design and targeting of electric vehicles by knowing which qualities have a big impact on sales outcomes.

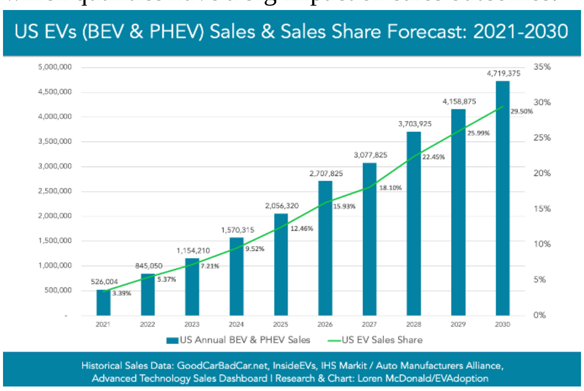

FIGURE 1. ELECTRIC VEHICLE PRICE PREDICTION

1.2 RANDOM FOREST REGRESSION

An ensemble-based machine learning approach called Random Forest Regression builds a number of decision trees during the training stage and combines their results to provide a final prediction. It works especially well with datasets that include a lot of characteristics, which improves accuracy and generalization. By adding randomness to the feature selection and training samples for every tree, this technique reduces the overfitting issue that single decision trees frequently face. It is a flexible tool for regression tasks since it performs well with both continuous and categorical variables. Furthermore, Random Forest can assess each input feature's relative importance, providing information about which variables have the most effects on the prediction outcomes.

1.3 DATA VISUALIZATION

The graphical display of datasets, statistical findings, and model outputs to improve information interpretability is known as data visualization. It facilitates the identification of patterns, correlations, or deviations and makes complex interactions between variables easier to detect. Histograms show the distribution qualities of variables like residuals, bar charts show the significance of features, and scatter plots assist in comparing expected and actual values. These graphic styles facilitate a more thorough examination of analytical results while streamlining the sharing of findings. Users can better grasp data behavior and analyze machine learning performance with the help of effective visualization.

FIGURE 2. DATA VISUALIZATION

1.4 STREAMLIT APPLICATION

Building interactive web-based tools that showcase the capabilities and outcomes of Python-based data science workflows is made possible by Streamlit applications. Features like uploading datasets, starting calculations, and displaying results like graphs and prediction metrics are all supported. Without the need for sophisticated web programming abilities, Streamlit offers a straightforward framework for integrating model predictions and UI components. This makes it possible to work with visualization tools and machine learning models in a same environment with ease. Developers can create apps that make data interpretation and model usage easier for people from different backgrounds by utilizing Streamlit.

2 LITERATURE REVIEW

In this study, Youssef Amry et al. have suggested The widespread use of electric and plug-in hybrid vehicles and the demand for more ecologically friendly modes of transportation have made electric vehicle (EV) charging stations a significant problem for automakers and a significant task for researchers worldwide. Indeed, a number of research projects focusing on advanced power electronics topologies and the optimization of EV charging stations in terms of power transfer and geographical location have their roots in the high cost of battery energy storage, the limited autonomy and battery lifespan of EVs, the battery charging time, the deployment cost of a fast charging infrastructure, and the substantial impact on the power grid. There are three distinct charge levels, each with a different output power and charging duration. Since more power is supplied to the car at the expense of power quality problems and disruptions, the charging process proceeds more quickly the higher the charge level. Additionally, there are three different kinds of charging methods: battery swapping, conductive charging systems, and inductive recharging (contactless power transmission) [1].

In this paper, Phiraphat Antarasee et al. have proposed The rapid advancement of electric vehicles (EVs) has led to a number of study topics in this field, including charging pricing strategy development, charging control, charging station location, and charging station layout. Using metaheuristic algorithms, this research suggests the best possible configuration for an EV fast-charging station (EVFCS) that is connected to a renewable energy source and battery energy storage systems (BESS). This structure's ideal design seeks to determine the quantity and strength of chargers. Furthermore, BESS and renewable energy sources can lessen their impact on the grid, which is why they are regarded in this work as two of the best-designed EVFCS structures. The best size for the renewable energy source, BESS, and grid power coupled to EVFCS must therefore be established. The station's profitability may increase with this ideal setup. Three metaheuristic algorithms—the arithmetic optimization algorithm (AOA), the Salp swarm algorithm (SSA), and particle swarm optimization (PSO)—are used to tackle the optimization problem [2].

In this research, J.A. Domínguez-Navarro et al. have claimed that the development of electric cars (EVs) is dependent on a number of criteria, including the cost of purchase, autonomy, charging infrastructure, and charging process. The design of an EV fast-charging station is the final component that is the subject of this research. The fast-charging station has a storage system and renewable generating (photovoltaic and wind) to increase its profitability and reduce the excessive energy demand on the grid. In contrast to previous studies, this one makes use of a comprehensive charging process model that takes into account the electric vehicles' arrival time and charge level. First, the demand for EVs and renewable energy are modeled using the Monte Carlo approach. The EV fast-charging station's installation and operation are then optimized by a genetic algorithm (GA). By calculating its net present value (NPV), it determines the best course of action that optimizes profit [3].

Hazardous features of on-road vehicle-based emissions have been postulated by Furkan Ahmad et al. in this paper, raising concerns for urban residents. Accordingly, emission-free electric cars guarantee a notable decrease in air pollution and enhance the environment. The reliability of the distribution networks is impacted by the significant increase in electric demand that comes with the widespread commercialization of electric vehicles. Therefore, a thorough framework for placing solar-powered charging stations in a distribution network with a better voltage profile, minimal power loss, and lower cost is suggested in this research. Additionally, the effects of EV load demand on the distribution network are investigated in terms of power loss, average voltage deviation index, voltage stability index, and per unit voltage profile. The charging stations are then positioned in the IEEE 33 bus system in the best possible way using a computer technique called improved chicken swarm optimization [4].

Willy In this study, Stephen Tounsi Fokui et al. have suggested In today's transportation industry, the growing number of electric vehicles (EVs) is causing petroleum-based vehicles to eventually be phased out. However, the coordinated and quick construction of EV charging stations (EVCSs) is crucial to the rapid deployment of EVs. Since EVCSs can result in high power losses and voltage deviations that are outside permissible bounds, their integration into the contemporary distribution network, which is marked by a rise in the penetration of randomly dispersed photovoltaic (PV) systems, is difficult. In order to put EVCSs in the distribution network with the highest penetration of randomly dispersed rooftop PV systems, a hybrid bacterial foraging optimization algorithm and particle swarm optimization (BFOA-PSO) technique is suggested in this publication [5].

3. EXISTING SYSTEM

Installing public electric vehicle charging stations (EVCS) is necessary to promote user convenience and assist users who do not have access to domestic charging due to the growing number of EVs. To handle the demand for EV traffic, which would otherwise cause congestion at the charging stations, the public electric vehicle charging infrastructures (EVCIs) must have a sufficient number of EVCSs with fast charging capabilities. The distribution system (DS) is greatly impacted by the location of these fast-charging infrastructures. In order to minimize power loss and voltage variations, we suggest employing multi-objective particle swarm optimization (MOPSO) to arrange fast-charging EVCIs at various points throughout the distribution system. MATLAB and OpenDSS are used to conduct time-series analysis of the fluctuations in DS and EV loads. We use an autoregressive integrated moving average (ARIMA) model to forecast the dynamic price and further examine the cost-benefits of the EVCIs under real-time pricing conditions.

4. PROPOSED SYSTEM

The suggested system is an interactive web-based platform that was created with Streamlit to help anticipate sales of electric vehicles utilizing structured input information. In order to produce precise forecasts of sales volume, it integrates a pre-trained Random Forest regression model that analyzes a number of electric car parameters, including battery capacity, energy usage, and vehicle specs. Users can first submit CSV datasets to the system, which cleans them up to get rid of incomplete records and non-numeric elements.

Following preprocessing, the data is scaled to match the model's training settings using a normalization strategy that has already been fitted. Predicted sales figures are then obtained by running the scaled data through the model. Along with several graphical analyses including feature importance, prediction comparisons, and residual distributions, the system also displays model evaluation metrics like mean squared error and R-squared values. All of these elements work together to create a thorough setting for investigating the various ways that factors affect sales of electric vehicles.

5. MODULE DESCRIPTION

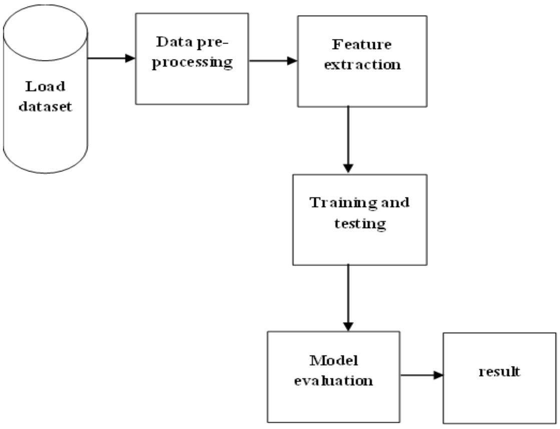

FIGURE 3. SYSTEM FLOW DIAGRAM

5.1 DATA LOADING

Importing and reading datasets into the system for additional processing is the responsibility of the data loading module. In order to make the dataset available for analysis, this phase usually entails loading data from several file formats, such as Excel or CSV. By arranging the data in a structured manner, the main objective is to get it ready for the next phases by making it simpler to manipulate, explore, and extract pertinent information.

5.2 DATA PRE-PROCESSING

A crucial step in making sure the dataset is clean and prepared for analysis is data preparation. This module manages missing values, eliminates superfluous or unnecessary columns, and transforms category data into numerical representations. Additionally, it entails scaling or normalizing data to guarantee that every feature is on a similar scale, which is essential for a lot of machine learning methods. Enhancing data quality and making sure the model can accurately understand the dataset without bias or errors are the goals.

5.3 FEATURE EXTRACTION

Finding and choosing the most pertinent variables or features from the dataset that will support the model's prediction performance is the main goal of feature extraction. In this module, the correlations between various features are analyzed, and the features that have the most effects on the target variable are chosen. In order to better depict the underlying patterns, it could employ strategies like dimensionality reduction, which involves eliminating superfluous features and creating new features from the available data. Making sure the model gets the most significant and instructive data is intended to improve its capacity to provide accurate predictions.

5.4 TRAINING AND TESTING

The dataset is separated into two subsets in the training and testing module: one for machine learning model training and another for performance evaluation. While the testing subset is used to evaluate how effectively the model generalizes to new, unknown data, the training subset is used to educate the model by modifying its internal parameters to minimize errors. This stage guarantees that the model has discovered the underlying patterns in the data and is capable of producing precise predictions for cases that haven't been observed before. Additionally, it aids in the detection of overfitting, guaranteeing that the model functions properly in real-world situations as well as on training data.

5.5 MODEL EVALUATION

The trained model's performance is evaluated using the model evaluation module. The degree to which the model's predictions match the actual values is assessed using a number of metrics, including accuracy, precision, recall, and mean squared error. This step offers information about the model's advantages and disadvantages, emphasizing its strong points and potential areas for development. Before the model is used to make predictions or provide insights in any real-world applications, model evaluation makes sure the model is resilient and dependable.

6.RESULT ANALYSIS

When appropriately trained on historical data, machine learning models—especially those based on regression techniques—can anticipate electric vehicle costs with a considerable degree of accuracy, according to the results analysis of the suggested method. The use of numerous regression models, including ensemble-based and basic models, made it possible to compare how well they predicted outcomes. The ability of ensemble models, like Random Forest and Gradient Boosting, to capture intricate, non-linear correlations between car attributes and pricing is demonstrated by their constant outperformance of simpler models. By guaranteeing that the data input into the models was relevant and clean, the preprocessing steps—such as feature selection and outlier removal—significantly improved the accuracy of the models. Performance was measured using evaluation metrics such Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared values; the findings showed a low margin of error and good explanatory power. According to these findings, machine learning can be a potent tool for comprehending pricing behavior and creating successful pricing strategies in the electric vehicle industry when combined with properly formatted data and suitable modeling methodologies.

7. CONCLUSION

In conclusion, the study effectively illustrated the ability to use machine learning approaches, such as Random Forest models, to forecast the price of electric cars based on a variety of criteria. The system has been able to provide significant insights into the relationship between various variables and the goal variable, electric car sales, through efficient data preprocessing, feature extraction, and model evaluation. In addition to offering precise pricing predictions, the project's integration of the model into an intuitive user interface allowed for a clear depiction of feature importance and prediction performance. The outcomes demonstrate the model's efficacy as well as its room for improvement, including the addition of new features or the investigation of different methods. All things considered, this study demonstrates how data-driven methods may be used to tackle challenging issues, providing helpful assistance with prediction and decision-making duties. The results of this study offer a strong basis for upcoming advancements in the field of automotive predictive analytics.

8. FUTURE WORK

This project can be extended in the future to include new features and enhancements that might increase the precision and reach of the forecasts. To create a more complete model, one possible direction for improvement is to incorporate more varied datasets, such as customer behavior, economic indicators, or environmental elements. Additional machine learning algorithm modification, such as investigating more sophisticated methods or adjusting hyperparameters, may enhance model performance and prediction precision. Furthermore, incorporating more advanced visualization capabilities would enable users to engage with the outcomes in more significant ways, like making price predictions in response to dynamic input. Implementing an automated model retraining method could be another area for development. This would guarantee that the system stays correct and pertinent as market conditions change. In the end, these improvements will increase the system's adaptability and resilience, providing better predictive power for a larger variety of situations.

REFERENCES

[1] Amry, Y., Ghogho, M., El Hani, S., Le Gall, F., and Elbouchikhi, E., 2022. The present state and difficulties of electric car traction drives and charging station power electronics. Energies 1–30, 15 (16). doi: 10.3390/en15166037 (http://dx.doi.org). Siritaratiwat, A., Khunkitti, S., Antarasee, P., and Premrudeepreechacharn, S. (2023)

[2] Metaheuristic algorithms are used to optimize the building of electric vehicle fast-charging stations. Sustainability 15 (1), 22 (Switzerland). doi: 10.3390/su15010771 (http://dx.doi.org).

[3] Design of an Electric Vehicle Fast-Charging Station with Integration of Renewable Energy and Storage Systems by Domínguez-Navarro, J.A.; Dufo-López, R.; Yusta-Loyo, J.M.; Artal-Sevil, J.S.; and Bernal-Agustín, J.L. 2019 105,46–58; Int. J. Electr. Power Energy Syst. [Reference]

[4] Panigrahi, B.K.; Khalid, M.; Ahmad, F. A Better Method for Placing the Solar-Powered EV Charging Station in the Distribution Network. Energy Storage Journal, 42, 103090, 2021. [Reference]

[5] Tounsi Fokui, W.S.; Saulo, M.J.; Ngoo, L. A Distribution Network with Randomly Distributed Rooftop Photovoltaic Systems: The Best Location for EV Charging Stations. IEEE Access, 9, 132397–132411, 2021. [Reference]

[6] Using metaheuristic optimization techniques, Yenchamchalit, K.; Kongjeen, Y.; Prabpal, P.; and Bhumkittipich, K. optimally positioned distributed photovoltaic systems and electric vehicle charging stations. Symmetry 13, 2378 (2021). [Reference]

[7] Kusakana, K.; Vermaak, H.J. The design of a solar-powered wind charging station for small electric tuk-tuks in the Democratic Republic of the Congo. 40–45 in RenewEnergy 2014, 67. [Reference]

[8] Badea, G.; Zarlam, M.; Filote, C.; Culcer, M.; Iliescu, M.; Raboaca, M.S.; Felseghi, R.A. Romanian Solar Energy Charging Station Design and Simulation for Electric Vehicles. Energies 12, 74 (2019). [Reference]

[9] Optimal Charging Scheduling by Pricing for EV Charging Stations with Dual Charging Modes Zhang, Y.; You, P.; Cai, L. Intell. Transp. Syst. IEEE Trans. 2019, 20, 3386–3396. [Reference]

[10] Optimal Scheduling and Economic Analysis of Hybrid Electric Vehicles on a Microgrid, Savari, R.C.; Hung, J.J. 2017, 10, 483. [Reference]