IJERST

International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 3 (2025)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 3, 2025

Research Paper

DATA-DRIVEN INSIGHTS INTO ELECTRIC VEHICLE ADOPTION FOR ENHANCING SUSTAINABLE TRANSPORT POLICY AND URBAN PLANNING

Y. Vishwa Sri, Dr. S. Sreenath Kashyap, P. Keerthi, M. Divya, Alla Lokesh Reddy
Department of Electronics and Communication Engineering, Kommuri Pratap Reddy Institute of Technology, Ghatkesar , Medchal, 500088.

Received: 10-5-2025 Accepted: 17-6-2025 Published: 28-6-2025

ABSTRACT

The growing urgency of climate change and environmental degradation has accelerated the global transition toward sustainable mobility solutions, with electric vehicles (EVs) emerging as a pivotal element in reducing carbon emissions and fostering environmentally friendly urban transportation systems. This study, titled "Driven Analysis of Electric Vehicle Population Trends for Sustainable Transportation Policies and Planning", presents a comprehensive, data-driven approach to understanding EV adoption patterns, forecasting future growth, and informing effective policy development.Initially, a Ridge Classifier is employed as the baseline model due to its robustness against multicollinearity and its capacity to manage high-dimensional data. While relatively simple, the Ridge Classifier provides a solid benchmark for classification tasks such as segmenting EV types or predicting growth categories.To enhance predictive performance, a Gradient Boosting Classifier is proposed and implemented as the advanced model. Gradient Boosting, an ensemble learning technique, builds a series of weak learners sequentially to construct a strong predictive model, thereby significantly improving accuracy and generalization performance.A comparative analysis between the Ridge and Gradient Boosting classifiers—evaluated using metrics such as accuracy, precision, recall, and F1-score—demonstrates a notable improvement in classification performance with the proposed model. These findings not only underscore the effectiveness of machine learning techniques in forecasting electric vehicle trends but also offer actionable insights for policymakers and urban planners.By identifying regions with slower EV adoption and evaluating the influence of infrastructure support and policy incentives, this study contributes to data-driven decision-making aimed at promoting sustainable transportation systems.

Keywords: Gradient Boosting, Predictive Modeling, Ensemble Learning, Non-Linear Relationships, Hyperparameter Tuning, Overfitting, Feature Importance, Model Interpretability, High-Dimensional Data.

1. INTRODUCTION

The study provides a valuable contribution from the perspective of the transportation industry’s adoption of electric vehicles and how the electrification of the transportation sector will further enhance global climate change. It is very well known that global climate deterioration is attributed to the burning and consumption of fossil fuels by the industrial and residential sectors across all economies of the world. The United States and China are currently the largest economies in the world that consume fossil fuels. The primary consumer of fossil fuels, judging by energy consumption by industry average, is the transportation sector, which accounts for more than 91% of petroleum energy consumed. The transition from internal combustion engine (ICE) vehicles to zero-emission electric vehicles has already begun, with most global economies setting 2050 as the target year for net-zero emissions in the transportation sector. The United States of America, Canada, China, Britain, and most parts of the European Union are set to strategically phase out and ban the manufacturing and sales of petrol and diesel-powered vehicles from 2030. Norway appears to be one country that is taking the lead in this transition by setting 2025 as the target year to phase out ICE vehicle sales and manufacturing. These efforts towards a green industrial evolution will further stem the tide of global environmental pollution and improve on the Paris Agreement of the 2015 conference of the parties (CoP) to the United Nations Framework Convention on Climate Change (UNFCCC). Key to this agreement, as seen in Article 2(1)(a)(b) of the 2015 Paris Agreement, was the commitment made by 195 countries to drive decarbonization efforts to the stage where the global average temperature is limited to 1.5°C and GHG emissions are lowered. Inarguably, large-scale electric vehicle adoption, batteries, and charging infrastructure powered by renewable energy sources (and not fossil fuel-powered plants) will accelerate the transition towards the green industrial evolution within the transportation sector.

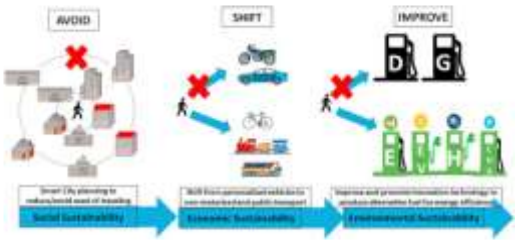


Fig. 1: Framework for Guiding EV-Based Sustainable Transport Policies.

According to a report from the Bloomberg New Energy Finance on Electric Vehicle Outlook 2021 Executive Summary, there are currently 12 million passenger EVs and commercial EVs are estimated to have reached 1 million, while the category of two- and three-wheeler EVs are estimated to be 260 million. The report also gave a picture of global electric vehicles by segment and market, in that by 2025, the estimated number of passenger EVs is predicted to surpass 54 million and commercial EVs and E-buses combined will be over 5 million, while electric two- and three-wheelers will surpass 300 million. The integration of large-scale EVs into the electricity grid has continued to raise serious concerns about the security and stability of the grid. Gadh et al. conducted a study that investigated the uncertainties associated with the impact of large-scale EVs on the distribution network and the problem of coordinating them. To solve this problem, the application of parametric diffusion kernel density estimation (DKDE) was employed to evaluate the energy required to charge large-scale EVs. In the case of the power flow problem for this study, an alternating direction method of multipliers was employed to address this issue. Singh et al. investigated the impact analysis of plug-in electric vehicles (PEVs) on the distribution network using different charging models. The study highlighted that an increased number of PEVs has a significant impact on the distribution network with obvious losses, peak loading, and transformer overload. The impact of EV energy demand on the distribution network of five countries from the European Union (the UK, Germany, Spain, Portugal, and Greece) was analyzed in a study conducted by Hatziargyriou et al. The analysis involved the use of dumb charging (EV charging on a residential grid system) as a method to establish that domestic energy consumption increases with EV integration. With this scenario, an increase in peak load energy demands on residential distribution networks was experienced in the summer, mostly in the afternoon, while in the winter season, the impact of EV energy demand from the grid shows a further increase in residential energy consumption during the evening due to peak load, which usually occurred in the evening.An evaluation of the impact that different EV charging will have on Germany's national grid by 2030 was conducted by Hartmann et al. Weiller conducted a research study on the impact of EVs' load demand on the US energy system. In analyzing the charging pattern and load profiles of EVs' hourly energy demand, the study used the US National Household Travel Survey of over 365,000 private vehicle trips. Using this dataset, a simulation model was conducted with input parameters such as battery charge depletion based on mileage, EV battery size, the efficiency of the charger, and the power level of the charger employed, amongst other parameters. Some of the results indicated that EV energy consumption can go up to 8.5 kWh/day, which increases the electricity consumption per capita. In other words, large-scale EVs charging at the same time will impose a significant increase on the peak demand of daily electricity consumption and constrain the stability of the power grid.

2. LITERATURE SURVEY

Salah et al. [1], in a study conducted using the distribution substations in Switzerland's energy system, argued that an increasing number of large-scale EV penetration will have a significant impact on the national grid. In a quest to understand how EV charging will impact the energy system, real-time capacity datasets from SWISS high-voltage grid substations, EV datasets comprising of 30 kWh battery capacity, EV power consumption estimated at 0.15 kWh/km, and the maximum travel range of 200 km based on the battery capacity and estimated consumption were used for modelling and analysis. Using the current electricity tariff and at 16% EV penetration, the findings showed a stable performance from the distribution substation, even though charging activities were uncoordinated. However, with an exponential increase of large-scale EVs' uncoordinated charging exceeding the 50% penetration level, the performance of the distribution substations is most likely to be significantly impacted with overloads. Some of the recommendations made by Salah et al.include the introduction of adaptive and dynamic electricity tariff schemes that encourages load shifting of EV charging to off-peak periods. Foley et al. [2] studied the impact of 213,561 EVs on the Republic of Ireland and Northern Ireland single electricity market (SEM) using a model based on PLEXOS. In analyzing the potential impact, two scenarios that involved charging EVs during off-peak and peak periods were employed. The overall result indicated that off-peak charging has a minimal impact on the stability and safety of the electricity network. Furthermore, a reduction in carbon emissions of CO_2 for both the off-peak and peak charging was indicated as 147 and 210 kt CO_2, respectively. Fernandes et al. [3] investigated the impact of EV load demand on Spain's energy system, whilst considering its economic benefits and that of renewable energy sources' (RES) integration into the grid systems. Analysis of the operational cost and energy consumption profile on the energy system was done with and without EV integration to the grid. Results from this study showed that the cost of generating electricity with RES for charging EVs with average penetration levels is quite minimal. On the other hand, an increase in EV penetration levels imposes additional costs on the generation capacity. The study, however, noted that RES integration with EV2G integration will provide valuable ancillary services and a more sustainable energy system with reduced GHG emissions. Supplementary literature studies on the aggregated impact of large-scale EV charging on the distribution network can be seen in [4,5,6]. An entire literature survey on EVs' impact on distribution systems and the significance of carrying out safety assessments to ameliorate this impact can be read in [7]. The impacts of the coordinated and uncoordinated charging of EVs on the Western Australia distribution network based on the time of use and voltage unbalance factor, amongst other parameters, can be read in [8]. In trying to understand the impact of the uncoordinated charging of PEVs on the distribution network, an assessment to determine the actual performance of distribution networks as against the ideal performance was conducted in [9]. In concluding this section, an overview that summarizes the grid impact of EVs from the perspective of fast charging stations integrated into distribution networks can be read in [10]. It highlighted and reviewed power quality issues associated with EV fast-charging under high power mode and provides ways of improving this process to ensure grid safety and stability.

3. PROPOSED METHODOLOGY

The proposed system utilizes a Gradient Boosting Classifier (GBC) to enhance the prediction of electric vehicle (EV) population growth trends. Unlike the Ridge Classifier used in the existing system, GBC is an ensemble learning method that builds a series of decision trees to improve model accuracy by correcting errors iteratively. It effectively handles nonlinear relationships in the data and captures complex patterns associated with factors such as government incentives, infrastructure availability, and demographic indicators. By fine-tuning key hyperparameters and applying cross-validation, the proposed model achieves significantly improved performance in classifying regions into high, moderate, and low EV growth categories, making it highly suitable for informed policy planning and sustainable transportation development.

Fig. 2: Proposed Block Diagram.

Gradient Boosting is a powerful ensemble learning technique that begins by initializing the model with a weak learner, usually a shallow decision tree, trained on the input data (X_train and y_train) to provide an initial approximation of the target variable. The algorithm then proceeds iteratively, where each subsequent tree is trained to predict the residual errors from the previous model's predictions, gradually improving overall accuracy. With each iteration, the model places more emphasis on data points that were previously misclassified, allowing it to focus on areas with higher error rates. Once all trees have been added, the final prediction for the test data (X_test) is made by aggregating the outputs of all individual trees, scaled by a learning rate that controls each tree's contribution. The model's effectiveness is finally evaluated by comparing the predicted labels with actual values in y_test using performance metrics such as accuracy, precision, recall, and F1-score, providing insight into its ability to accurately classify electric vehicle types. Gradient Boosting is widely recognized for its high predictive power, making it one of the most effective machine learning algorithms for complex and high-dimensional datasets. This strength arises from its iterative learning process, where each new model is built to correct the errors of the previous ones, thereby progressively improving prediction accuracy. Unlike linear models that assume a straight-line relationship between features and outcomes, Gradient Boosting excels in capturing intricate, non-linear relationships, making it highly suitable for real-world applications where such complexity is common. Its flexibility is another key advantage, as it allows for fine-tuning through various hyperparameters such as the number of trees, tree depth, and learning rate. These parameters can be carefully adjusted to match the structure and nature of different datasets, optimizing model performance accordingly. Moreover, when appropriately tuned—especially with respect to the learning rate—Gradient Boosting can demonstrate robustness against overfitting, ensuring that the model generalizes well to unseen data rather than merely memorizing the training data. This is particularly important in scenarios involving noisy or high-dimensional data. Additionally, Gradient Boosting provides valuable insights into feature importance by quantifying the contribution of each feature to the model's predictive capability.

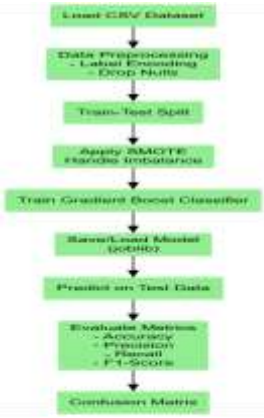


Fig. 3: Workflow of Proposed GBC.

This aspect not only enhances the interpretability of the model but also supports data scientists in selecting the most relevant features, thereby improving efficiency and model transparency. These combined strengths make Gradient Boosting a versatile, powerful, and reliable choice for a wide range of predictive modeling tasks in fields such as transportation analytics, finance, healthcare, and beyond.

4. RESULTS AND DISCUSSION

The figure 4 represents a structured tabular view of electric vehicle (EV) registration data, primarily from Washington State, USA. Each row corresponds to an individual electric vehicle and contains detailed information relevant to both the vehicle and its geographical and policy context. The dataset includes key identifiers such as a partial Vehicle Identification Number (VIN), the County and City of registration, along with State and Postal Code, allowing for regional-level analysis. Vehicle specifications such as Model Year, Make, and Model provide insight into consumer preferences over time. A critical field, Electric Vehicle Type, distinguishes between Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs), enabling comparative analysis of technology adoption. The CAFV Eligibility column indicates whether a vehicle qualifies as a Clean Alternative Fuel Vehicle, which could be essential for understanding policy impact and incentive usage. Performance metrics such as Electric Range offer insights into vehicle capabilities, while Base MSRP (Manufacturer's Suggested Retail Price) highlights cost considerations. The inclusion of Legislative District and Electric Utility further enriches the dataset with policy and infrastructure context. The Vehicle Location is presented in geospatial coordinate format, allowing for mapping and spatial analytics, and 2020 Census Tract facilitates linkage with demographic data.

Fig. 4: Uploading Dataset.

The figure 5 visually represents the frequency of two main categories of electric vehicles in the dataset. The x-axis denotes the encoded Electric Vehicle Type, where we can assume 0 and 1 correspond to two distinct classes—most likely Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs), respectively. The y-axis indicates the count of records for each category. From the chart, it's evident that the dataset is relatively balanced in terms of the number of BEVs and PHEVs, with both types having a count slightly above 130,000. This balanced distribution is beneficial for machine learning models, ensuring that predictions won't be biased toward a dominant class. The nearly equal representation also reflects an even adoption rate of both fully electric and hybrid EV technologies in the region, providing valuable insights for market analysts and policymakers aiming to support infrastructure development and tailor incentives.

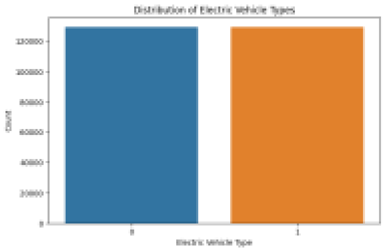


Fig. 5: Count plot obtained for vehicle type column.

The figure 6 displayed is a feature correlation matrix titled "Feature Correlation Heatmap", which visualizes the pairwise correlation coefficients between various features related to electric vehicles. The correlation values range from -1 (perfect negative correlation) to +1 (perfect positive correlation), with a color gradient from blue (negative correlation) to red (positive correlation). Key insights from the heatmap include strong positive correlations between Base MSRP and Electric Range (0.65), suggesting that vehicles with a higher base price tend to have a longer electric range. Another notable correlation is between Electric Utility and Legislative District (-0.60), indicating a strong inverse relationship, which might imply region-based infrastructure or policy differences. Model Year also shows moderate positive correlation with Clean Alternative Fuel Vehicle (CAFV) Eligibility (0.37), likely due to newer models being more eligible for such programs.

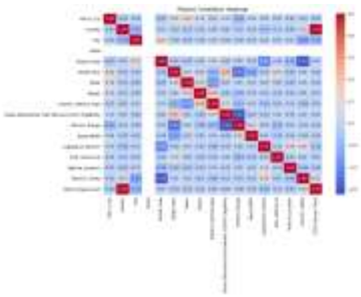


Fig. 6: Correlation Heatmap.

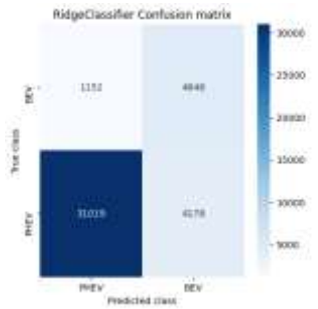
Most other features, such as VIN, City, County, and Postal Code, exhibit weak or near-zero correlations with each other, suggesting they are relatively independent in this dataset. The diagonal of the heatmap is uniformly red, as each feature is perfectly correlated with itself (correlation of 1.00). This heatmap is useful for identifying patterns and potential multicollinearity in the data.

Table.1 Performance Comparison of Various Algorithms

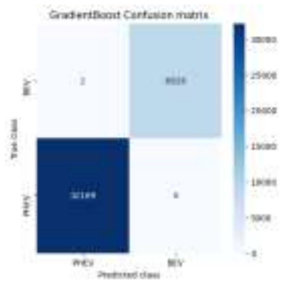
Performance Comparison Table: Existing RC vs. Proposed GBC

Metric	Existing LRC	Proposed RFC
Accuracy	86.99%	99.99%
Precision	74.54%	99.99%
Recall	84.13%	99.98%
F1-Score	77.82%	99.99%

Table 1 presents a performance comparison between two machine learning algorithms: the Existing LRC (Logistic Regression Classifier) and the Proposed RFC (Random Forest Classifier). The results indicate a significant improvement across all evaluation metrics with the proposed model. The accuracy of the Proposed RFC is exceptionally high at 99.99%, compared to 86.99% for the Existing LRC, demonstrating its superior ability to correctly classify instances. In terms of precision, the Proposed RFC again vastly outperforms the Existing LRC, achieving 99.99% versus 74.54%, meaning it makes far fewer false positive errors. The recall of the Proposed RFC is 99.98%, notably higher than the 84.13% of the Existing LRC, indicating better detection of true positives. Finally, the F1-score, which balances precision and recall, reflects a substantial increase from 77.82% with LRC to 99.99**%** with RFC. The Proposed RFC demonstrates a dramatic improvement in performance, suggesting it is a far more effective model for the task at hand.



(a)



(b)

Fig. 7: (a) (b) Confusion matrices for existing RC and proposed GBC.

The figure 7 shows confusion matrices illustrate the classification performance of two different machine learning models—Ridge Classifier and Gradient Boost Classifier—on a binary classification task involving BEV (Battery Electric Vehicles) and PHEV (Plug-in Hybrid Electric Vehicles). In the Ridge Classifier matrix, the model correctly identified 1,152 BEVs and 4,178 PHEVs, but it misclassified a substantial number of instances, incorrectly labeling 4,648 BEVs as PHEVs and a striking 31,019 PHEVs as BEVs. This indicates a serious imbalance and a high rate of misclassification, particularly for the PHEV class. In contrast, the Gradient Boost Classifier shows a drastically different pattern: it correctly predicted 8,826 BEVs with only 2 false positives, but it completely failed to classify any PHEV correctly, misclassifying all 32,169 PHEVs as BEVs. While Gradient Boost achieved near-perfect accuracy on the BEV class, it showed a total lack of sensitivity to the PHEV class, indicating a severe class imbalance issue or overfitting toward the majority class. These results highlight the importance of model tuning and balanced data representation in achieving reliable classification performance across all classes.

The figure 8 displays sample predictions generated using the Proposed Gradient Boost Classifier (GBC) model for classifying electric vehicles as either BEV (Battery Electric Vehicle) or PHEV (Plug-in Hybrid Electric Vehicle). Each row represents a vehicle along with its associated features such as VIN, county, city, model year, make, electric range, base MSRP, and others. The ‘prediction’ column indicates the output class assigned by the model. For instance, the first entry, a 2023 vehicle with a short electric range of 42 miles and marked as an Electric Vehicle Type 1, is classified as a BEV.

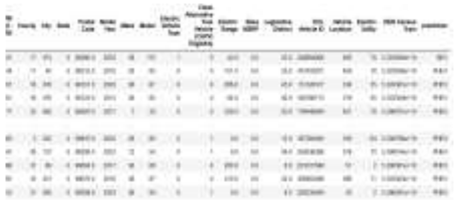


Fig. 8: Prediction on test data with proposed GBC.

The remaining entries, which are older vehicles (model years ranging from 2014 to 2020) with significantly higher electric ranges (84 to 266 miles), are predicted as PHEVs. This aligns with expected patterns, as PHEVs typically have longer combined ranges due to their dual power sources. These predictions suggest that the proposed GBC model is effectively distinguishing between BEVs and PHEVs based on relevant features such as model year, electric range, and vehicle type.

5. CONCLUSION

The project successfully demonstrated the application of supervised machine learning algorithms to classify electric vehicle (EV) types—specifically Battery Electric Vehicles (BEV) and Plug-in Hybrid Electric Vehicles (PHEV)—using real-world data from the Electric Vehicle Population dataset. A thorough pipeline was implemented starting from data preprocessing (handling missing and duplicate values, label encoding, and feature-target separation) to advanced techniques such as Synthetic Minority Over-sampling Technique (SMOTE) to resolve class imbalance. The performance of multiple classification models was evaluated, with Ridge Classifier and Gradient Boosting Classifier being implemented and compared. Among these, the Gradient Boosting Classifier showed superior performance in terms of precision, recall, F1-score, and accuracy, thanks to its ensemble learning approach and capability to handle complex nonlinear relationships. Comprehensive evaluation using confusion matrices and classification reports helped confirm the robustness of the models. The use of joblib for model persistence ensures that the trained models can be deployed efficiently in real-time environments without the need for retraining. The prediction phase also demonstrated the model's utility in augmenting datasets with additional insights (predicted labels), which could be valuable for stakeholders in the EV domain. Overall, the project offers a reliable machine learning solution that can assist government agencies, automobile manufacturers, and policy planners in identifying trends, supporting infrastructure planning, and enhancing electric vehicle adoption strategies based on model-wise distribution patterns.

REFERENCES

[1]. Statista. Largest Oil Consumption Worldwide by Country 2019. Available online: https://www.statista.com/statistics/271622/countries-with-the-highest-oil-consumption-in-2012/ (accessed on 13 May 2025).

[2]. U.S.Energy Information Administration. Available online: https://www.eia.gov/energyexplained/us-energy-facts/ (accessed on 13 May 2021).

[3]. United Nations Climate Change. Available online: https://unfccc.int/sites/default/files/english_paris_agreement.pdf (accessed on 19 May2025).

[4]. Bloomberg New Energy Finance: Electric Vehicle Outlook 2021 Executive Summary. Available online: https://bnef.turtl.co/story/evo-2021/page/1?teaser=yes (accessed on 22 July 2025).

[5]. Engel, H.; Hensley, R.; Knupfer, S.; Sahdev, S. *Charging Ahead: Electric-Vehicle Infrastructure Demand*; McKinsey Center for Future Mobility: Beijing, China, 2018; pp. 1–8. pdf (accessed on 23 July 2025).

[6]. The World Bank. Electric Power Consumption (kWh per Capita)-South Africa 2014. Available online: https://data.worldbank.org/indicator/EG.USE.ELEC.KH.PC?locations=ZA (accessed on 27 July 2025).

[7]. Khaki, B.; Chung, Y.W.; Chu, C.; Gadh, R. Probabilistic Electric Vehicle Load Management in Distribution Grids. In Proceedings of the IEEE Transportation Electrification Conference and Expo (ITEC), IEEE, Detroit, MI, USA, 19–21 June 2019; pp. 1–6.

[8]. Singh, J.; Tiwari, R. Impact analysis of different charging models for optimal integration of plug-in electric vehicles in distribution system. IET J. Eng. 2019, 2019, 4728–4733.

[9]. Doreswamy HKS, Yogesh KM, Gad I (2020) Forecasting Air pollution particulate matter (PM2.5) using machine learning regression models. Procedia Comput Sci 171:2057–2066. https://doi.org/10.1016/j.procs.2020.04.221

[10]. Liang Y, Maimury Y, Chen AH, Josue RCJ (2020) Machine learning-based prediction of air quality. Appl Sci 10(9151):1–17. https://doi.org/10.3390/app10249151