International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 02 , February - 2026

Analysis and Prediction of Electric Vehicle Cost Using Machine Learning

Chiluka Nathanil Kumar[1], Jangiti Srikanth[2], Kandumalla Ram[3], Karamala Hanish[4],

Ms. Yedelli Nithya[5], Dr. Bandaru Venkataramana[6]

[1]Student, BTech CSE(AI&ML) 4th Year,Holy Mary Inst. Of Tech. and Science, Hyderabad, TG, India,
[2]Student, BTech CSE(AI&ML) 4th Year,Holy Mary Inst. Of Tech. and Science, Hyderabad, TG, India,
[3]Student, BTech CSE(AI&ML) 4th Year,Holy Mary Inst. Of Tech. and Science, Hyderabad, TG, India,
[4]Student, BTech CSE(AI&ML) 4th Year,Holy Mary Inst. Of Tech. and Science, Hyderabad, TG, India,
[5]Asst. prof , CSE(AI&ML), Holy Mary Inst. Of Tech. and Science , Hyderabad , TG, India,
[6]Assoc. prof , CSE , Holy Mary Inst. Of Tech. and Science , Hyderabad , TG, India,

Abstract - Electric Vehicles (EVs) are gaining significant attention as a sustainable alternative to conventional fuel-based vehicles. Despite technological advancements and government incentives, the high purchase cost of EVs remains a major challenge for widespread adoption. Accurate cost prediction can assist consumers, manufacturers, and policymakers in effective decision-making. This paper presents a machine learning–based approach for analyzing and predicting electric vehicle costs using technical specifications and market-related factors. Regression models such as Linear Regression, Decision Tree, Random Forest, and XGBoost are implemented and evaluated using performance metrics including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 score. The experimental results indicate that ensemble learning models provide superior prediction accuracy. The proposed system demonstrates the effectiveness of machine learning techniques in EV cost analysis and supports the growth of sustainable transportation systems.

Keywords: Electric Vehicle, Cost Prediction, Machine Learning, Regression, Sustainable Transport

1 INTRODUCTION

Electric Vehicles (EVs) have emerged as a transformative solution to address the growing concerns of environmental pollution, climate change, and depletion of fossil fuel resources. The transportation sector is one of the largest contributors to greenhouse gas emissions, and the shift from internal combustion engine vehicles to electric mobility is considered essential for achieving sustainable development goals. Governments across the world, including India, are promoting EV adoption through subsidies, tax benefits, and supportive policies.

Despite these initiatives, the high initial purchase cost of electric vehicles remains a major barrier to widespread adoption. Factors such as battery capacity, driving range, motor power, brand value, and market demand significantly influence EV pricing. For consumers, the lack of transparent and accurate cost estimation creates uncertainty, while manufacturers face challenges in competitive pricing and cost optimization.

Traditional EV cost estimation methods are largely based on static assumptions and linear cost models, which fail to capture the complex and nonlinear relationships between technical specifications and market dynamics. With the availability of large datasets and advancements in computational power, machine learning (ML) techniques have gained attention as effective tools for predictive analysis. ML models can learn patterns from historical data and provide accurate cost predictions by considering multiple influencing parameters simultaneously.

This research proposes a machine learning–based framework for the analysis and prediction of electric vehicle costs. By implementing and comparing multiple regression algorithms, the study aims to identify the most suitable model for EV cost prediction.

Objectives

The key contributions of this study are as follows:

1. Development of a machine learning–based framework for predicting electric vehicle costs.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 02 , February - 2026

2. Comparative evaluation of multiple regression models for cost estimation.
3. Identification of key technical and market-related factors influencing EV pricing.
4. Demonstration of practical applicability for consumers, manufacturers, and policymakers.

2 LITERATURE REVIEW

The increasing adoption of electric vehicles (EVs) has accelerated research in cost analysis and prediction models. Battery technology has been identified as the primary cost-driving component of EVs. Nykvist and Nilsson [9] reported a rapid decline in lithium-ion battery prices, significantly improving EV affordability. Reports from BloombergNEF [10] and the International Energy Agency (IEA) [11] further confirm that battery price reductions and technological advancements play a crucial role in EV market expansion.

Traditional statistical methods such as Linear Regression have been widely used for price prediction tasks. Linear models attempt to establish relationships between vehicle attributes and cost using mathematical formulations, as discussed in classical statistical learning literature [14]. However, EV pricing depends on complex and nonlinear interactions among features such as battery capacity, driving range, motor power, and brand value, which limits the effectiveness of purely linear approaches.

To address nonlinear relationships, machine learning algorithms have been extensively explored. Decision Tree models [3] provide rule-based prediction structures that are easy to interpret but may suffer from overfitting. Random Forest, introduced by Breiman [2], improves prediction accuracy by combining multiple decision trees through ensemble learning. Gradient Boosting techniques, particularly XGBoost proposed by Chen and Guestrin [1], have demonstrated superior performance in structured data prediction tasks due to efficient handling of feature interactions and regularization mechanisms. Friedman's work on Gradient Boosting Machines [5] further established boosting as a powerful predictive framework.

Support Vector Machines (SVM), introduced by Cortes and Vapnik [4], have also been applied to regression and price prediction tasks, offering strong performance in high-dimensional feature spaces. Additionally, Artificial Neural Networks and deep learning methods have gained attention for modeling complex nonlinear relationships. Foundational works in deep learning [6], [7] describe the ability of neural networks to capture hierarchical feature representations, although such models often require large datasets and substantial computational resources.

Several studies have also focused on broader economic evaluation frameworks such as Total Cost of Ownership (TCO), incorporating maintenance, energy consumption, and long-term operational costs [12]. These approaches provide comprehensive economic insights but often rely on region-specific assumptions or limited data samples.

Despite the availability of diverse predictive techniques, many existing studies focus on a single algorithm rather than conducting comparative evaluation across multiple regression models. Furthermore, integration of both technical vehicle specifications and market-related factors into a unified prediction framework remains limited.

Therefore, this study builds upon established machine learning foundations [2], [5], [14] and state-of-the-art boosting techniques [1] to develop a scalable and modular EV cost prediction system. By performing comparative evaluation using multiple regression algorithms and incorporating real-world EV market data, the proposed framework aims to enhance prediction accuracy and practical applicability.

3 SYSTEM ARCHITECTURE

The proposed Electric Vehicle (EV) Cost Prediction System follows an Input–Process–Output (IPO) based architecture designed to ensure accurate and scalable cost estimation. The system takes structured EV specifications as input, processes the data through a machine learning pipeline, and generates the predicted EV cost as output. The architecture is modular, allowing flexibility in integrating new datasets or algorithms without affecting the overall workflow.

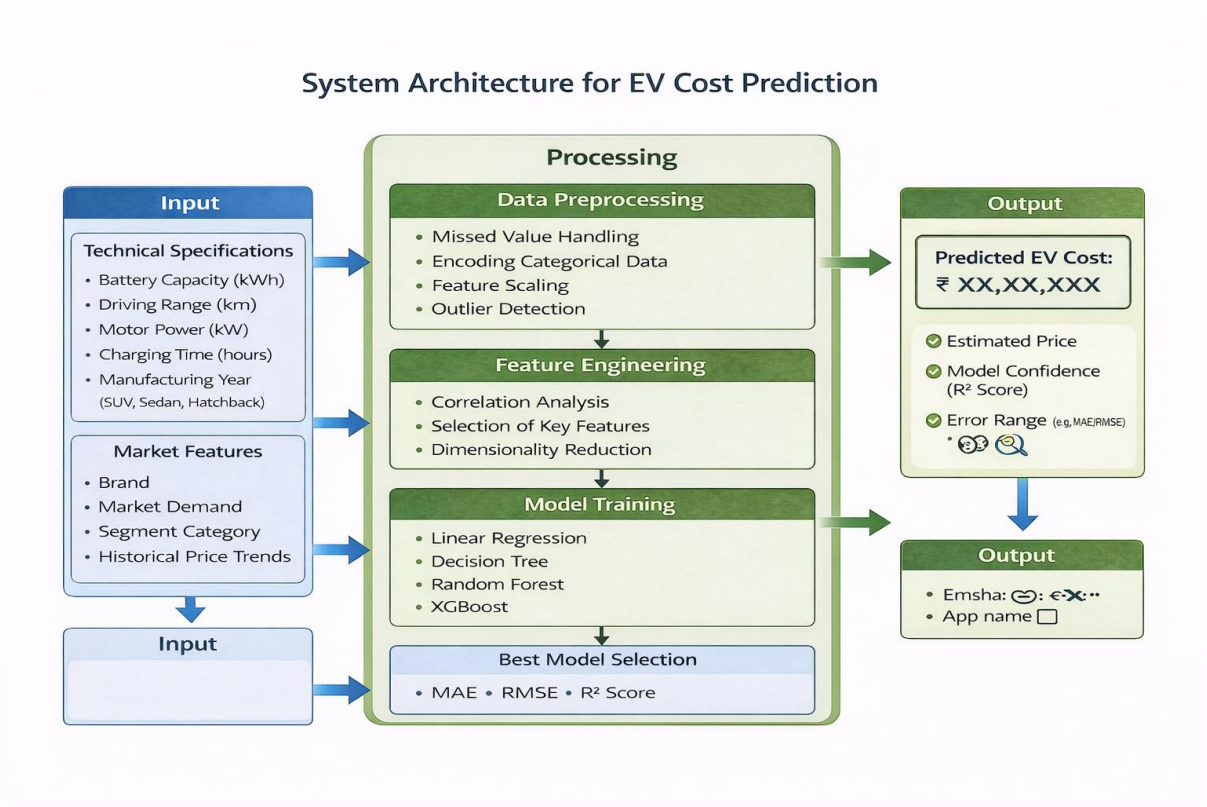
In the input stage, the system receives both technical and market-related attributes of electric vehicles. The technical specifications include battery capacity (kWh), driving range (km), motor power (kW), charging time, manufacturing year, and vehicle type. In addition, market-related features such as brand, vehicle segment, demand trends, and historical pricing information are considered. These inputs are either collected from verified datasets (manufacturer websites, automotive marketplaces, and

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 02 , February - 2026

government EV databases) or entered by users through an application interface. The inclusion of both technical and economic parameters ensures comprehensive cost analysis.

Once the input data is collected, it enters the processing stage, which consists of multiple sub-modules. Initially, data preprocessing is performed to improve data quality. This involves handling missing values using statistical imputation methods, removing duplicate records, detecting and treating outliers, normalizing numerical features, and encoding categorical variables such as brand and vehicle type. These steps ensure that the dataset becomes clean, consistent, and suitable for machine learning algorithms.

Following preprocessing, feature engineering is carried out to identify the most influential attributes affecting EV cost. Correlation analysis and feature importance techniques are applied to eliminate redundant or weakly contributing variables. This step reduces multicollinearity, improves computational efficiency, and enhances model performance. The refined dataset is then divided into training and testing subsets, typically using an 80:20 ratio.



In the model training phase, multiple regression algorithms such as Linear Regression, Decision Tree Regression, Random Forest Regression, and XGBoost Regression are implemented. Hyperparameter tuning and cross-validation techniques are applied to optimize model parameters and prevent overfitting. Each model learns the relationship between input features and vehicle price during the training process.

The evaluation stage assesses model performance using statistical metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2 score). A comparative analysis is conducted to determine the most accurate and reliable model. The best-performing model is selected for final deployment.

In the output stage, the system generates the predicted electric vehicle cost based on the provided specifications. The output includes the estimated market price along with model performance indicators to ensure transparency and reliability. When deployed as a web-based or application-based tool, users can input EV details and receive real-time cost predictions.

Overall, the system architecture ensures a structured data flow from input collection to prediction generation. The modular design supports scalability, adaptability to new datasets, and integration of advanced machine learning models, making it suitable for real-world EV cost analysis and prediction applications.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 02 , February - 2026

4 IMPLEMENTATION

The implementation of the Electric Vehicle (EV) Cost Prediction system was carried out using a structured machine learning pipeline consisting of data acquisition, preprocessing, model development, evaluation, and deployment stages. The system was developed using Python programming language due to its extensive machine learning libraries and data processing capabilities.

The dataset was collected from reliable EV data sources including manufacturer websites, online automotive marketplaces, government EV reports, and publicly available market research datasets. The collected dataset contained technical specifications such as battery capacity (kWh), driving range (km), motor power (kW), charging time, manufacturing year, and vehicle brand, along with pricing information. The dataset was stored in CSV format and processed using the Pandas library for efficient data manipulation.

During the preprocessing phase, missing values were handled using mean and median imputation techniques for numerical attributes, while categorical variables were encoded using label encoding and one-hot encoding methods. Duplicate records were removed to ensure data integrity. Outliers were detected using statistical techniques such as the Interquartile Range (IQR) method and treated appropriately to prevent model distortion. Numerical features were normalized using feature scaling techniques to improve model performance and convergence.

Feature engineering was implemented to identify the most influential variables affecting EV cost. Correlation analysis and feature importance techniques were used to remove redundant features and reduce multicollinearity. This step improved computational efficiency and enhanced prediction accuracy. The final feature set was then divided into training and testing subsets using an 80:20 split ratio to ensure unbiased model evaluation.

Multiple regression algorithms were implemented for cost prediction, including Linear Regression, Decision Tree Regression, Random Forest Regression, and XGBoost Regression. These models were developed using the Scikit-learn library and the XGBoost framework. Hyperparameter tuning was performed using Grid Search Cross-Validation to identify optimal model parameters and avoid overfitting. Cross-validation ensured that the models generalized well to unseen data.

Model evaluation was conducted using performance metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2 score). Comparative analysis showed that ensemble-based models, particularly Random Forest and XGBoost, achieved higher accuracy and better generalization performance compared to traditional regression techniques.

The best-performing model was selected for deployment. The trained model was saved using serialization techniques and integrated into a web-based user interface. The interface allows users to input EV specifications and obtain real-time cost predictions. The system is designed to be modular and scalable, allowing future integration of real-time market data and advanced machine learning models.

Overall, the implementation ensures a robust, efficient, and scalable EV cost prediction system capable of supporting accurate market price estimation and decision-making processes.

5 EXPERIMENTAL SETUP

The dataset consists of more than 500 electric vehicle records collected from manufacturer websites, online repositories, and market platforms. The data is divided into training and testing sets using an 80:20 split ratio. Cross-validation is applied to improve generalization. Experiments are conducted in a Python-based environment on a standard computing system.

Dataset Description

Feature	Description
Battery Capacity	Energy storage capacity in kWh
Driving Range	Distance per full charge (km)
Motor Power	Motor output power (kW)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 02 , February - 2026

Manufacturing Year	Vehicle production year
Brand	Vehicle manufacturer
Price	Electric vehicle cost (target variable)

6 RESULTS

The performance of the proposed electric vehicle cost prediction system is evaluated using multiple regression models, including Linear Regression, Decision Tree Regression, Random Forest Regression, and XGBoost Regression. The models are assessed using standard evaluation metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2), which collectively measure prediction accuracy and model reliability.

Experimental results indicate that Linear Regression provides a baseline performance but struggles to capture nonlinear relationships among features, resulting in higher error values. Decision Tree Regression improves prediction accuracy by modeling nonlinear patterns; however, it is prone to overfitting when trained on complex datasets.

Ensemble learning models demonstrate superior performance compared to individual models. Random Forest Regression achieves lower MAE and RMSE values due to its ability to reduce variance by aggregating multiple decision trees. XGBoost Regression delivers the best overall performance, achieving the highest R^2 score, indicating a strong correlation between predicted and actual EV costs. The boosting mechanism enables XGBoost to efficiently learn complex feature interactions and minimize prediction errors.

Feature importance analysis reveals that battery capacity and driving range are the most influential parameters affecting EV cost, followed by motor power and vehicle brand. These findings align with real-world market trends, where higher battery capacity and extended range significantly increase vehicle pricing.

The results confirm that machine learning–based ensemble models are highly effective for EV cost prediction and can be reliably applied to real-world scenarios. The comparative analysis highlights the suitability of XGBoost and Random Forest models for accurate and robust cost estimation.

Model Comparison

S.No	Model	Accuracy Level	Speed	Handles Complex Data	Easy to Understand	Overall Performance
1	Linear Regression	Medium	Very Fast	No (only simple relationships)	Very Easy	Basic Model
2	Decision Tree	Good	Fast	Yes	Easy	Better than Linear
3	Random Forest	Very Good	Medium	Yes (better handling)	Moderate	Strong & Stable
4	XGBoost	Excellent	Medium	Very Good	Slightly Complex	Best Performance

7 CHALLENGES

Several challenges were encountered during the development of the proposed EV cost prediction system. One major challenge was the limited availability of high-quality and standardized EV cost datasets. Data collected from different sources exhibited inconsistencies, missing values, and variations in pricing structures.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 02 , February - 2026

Another challenge involved handling outliers and multicollinearity among features without introducing bias into the model. Preventing overfitting while maintaining high prediction accuracy was also a critical concern, particularly for complex ensemble models. Additionally, rapidly changing market conditions and policy variations posed difficulties in ensuring model generalization.

8 CONCLUSION

This paper presented a machine learning–based approach for predicting electric vehicle costs using technical and market-related features. The results demonstrate that ensemble regression models provide accurate and reliable predictions.

Ethical and Sustainability Impact

The proposed system supports transparent pricing and informed decision-making, contributing to sustainable transportation adoption. Accurate cost prediction can encourage EV usage, thereby reducing environmental impact and promoting cleaner mobility solutions.

9 FUTURE WORK

The proposed electric vehicle cost prediction system can be extended in several promising directions to enhance its accuracy, applicability, and real-world impact. One important future enhancement is the adoption of advanced deep learning techniques such as Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) models. These models can better capture complex nonlinear relationships and temporal price variations in EV markets.

Another significant extension involves integrating real-time data sources, including battery raw material prices, charging infrastructure costs, government subsidy policies, and regional tax structures. Incorporating such dynamic data will enable more accurate and up-to-date cost predictions, reflecting real market conditions.

Future research may also focus on estimating the total cost of ownership (TCO) rather than only the purchase price. This includes maintenance costs, charging expenses, battery degradation, resale value, and lifecycle emissions costs. Such an approach would provide a more comprehensive evaluation of EV affordability.

Additionally, the system can be expanded to support global EV markets by incorporating region-specific parameters and policy variations. Explainable AI (XAI) techniques may be applied to improve model transparency and trust by clearly explaining the factors influencing cost predictions. These enhancements will further strengthen the system's usefulness for consumers, manufacturers, and policymakers.

REFERENCES

[1] Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. Proceedings of the 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining.
[2] Breiman, L. (2001). *Random forests*. Machine Learning, 45(1), 5–32.
[3] Quinlan, J. R. (1986). *Induction of decision trees*. Machine Learning, 1(1), 81–106.
[4] Cortes, C., & Vapnik, V. (1995). *Support-vector networks*. Machine Learning, 20, 273–297.
[5] Friedman, J. H. (2001). *Greedy function approximation: A gradient boosting machine*. Annals of Statistics, 29(5), 1189–1232.
[6] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
[7] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
[8] Hittinger, E., & Azevedo, I. (2015). *Bulk energy storage increases United States electricity system emissions*. Environmental Science & Technology, 49(5), 3203–3210.
[9] Nykvist, B., & Nilsson, M. (2015). *Rapidly falling costs of battery packs for electric vehicles*. Nature Climate Change, 5, 329–332.
[10] BloombergNEF (2023). *Battery Price Survey Report*.
[11] IEA (International Energy Agency). (2023). *Global EV Outlook*.
[12] Lutsey, N., & Nicholas, M. (2019). *Update on electric vehicle costs in the United States*. International Council on Clean Transportation (ICCT).
[13] He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. IEEE CVPR.
[14] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.
[15] Scikit-learn Developers (2023). *Scikit-learn: Machine Learning in Python*.