iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78021

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

Electric Vehicle Analytics and Sales Performance Prediction Using Machine Learning

P. Pradeep[1], P. Santhosh Kumar Reddy[2], R. Maithri[3], R. V. Ravi Tez[4], Raj Priyadarshi[5], M. C. Leelu Mamakarni[6]
[1, 2, 3, 4, 5]*Department of Computer Science and Engineering, (Data Science) Sri Venkateswara College of Engineering and Technology, RVS Nagar, Chittoor, India 517127*
[6]*Assistant Professor, Sri Venkateswara College of Engineering And Technology, (Data Science), RVS Nagar, Chittoor 517127*

Abstract: Electric Vehicles (EVs) are rapidly gaining global attention as an environmentally friendly alternative to conventional fuel-powered vehicles. With the rapid expansion of the EV industry, large volumes of vehicle specification data and market statistics are generated. Analyzing this data manually is difficult and inefficient. This research proposes an Electric Vehicle Analytics and Sales Performance Prediction system using machine learning techniques. The system analyzes EV datasets containing battery capacity, charging time, vehicle price, safety ratings, and other technical features. Exploratory Data Analysis (EDA) is performed to identify patterns and correlations within the dataset. Machine learning algorithms such as Linear Regression and Random Forest Regression are used to predict country-wise future EV sales. Experimental results demonstrate that the Random Forest model achieves high prediction accuracy and provides meaningful insights for manufacturers, investors, consumers, and policymakers. The proposed system supports data-driven decision-making for sustainable transportation planning and EV market growth.

Keywords: Electric Vehicles, Machine Learning, Data Analytics, Random Forest, Linear Regression, Sales Prediction

I. INTRODUCTION

Electric vehicles have emerged as one of the most promising solutions for reducing environmental pollution and dependency on fossil fuels. Governments across the world are encouraging the adoption of EVs through subsidies, incentives, and infrastructure development.As EV adoption increases, large datasets related to vehicle specifications, charging infrastructure, and market sales trends become available. Analyzing these large-scale datasets using traditional methods is difficult and time-consuming. Data nalytics and machine learning techniques provide efficient solutions to extract insights from EV data. Machine learning models can identify relationships between factors such as battery capacity, vehicle cost, charging time, and driving range. This research focuses on building a machine learning-based EV analytics system that can analyze electric vehicle data and predict country-wise EV sales trends. The proposed system helps stakeholders understand market patterns and make informed decisions regarding EV adoption strategies.

II. LITERATURE REVIEW

Several studies have explored the use of machine learning techniques in the electric vehicle domain. De Cauwer et al. proposed regression models to analyze electric vehicle energy consumption patterns. Their research demonstrated that battery capacity and driving conditions significantly influence energy consumption.

Wu et al. applied machine learning algorithms such as Support Vector Machines and Decision Trees to predict EV energy usage. Their work showed that machine learning models could accurately forecast vehicle energy efficiency.

Severson et al. used Random Forest and Gradient Boosting techniques to analyze battery health in lithium-ion batteries used in electric vehicles. Their study highlighted the importance of predictive maintenance in EV systems.

Recent research has also focused on EV market analysis. Wang et al. applied time-series forecasting models to analyze electric vehicle market growth trends. However, limited research has integrated EV performance analytics with machine learning-based sales prediction.

The proposed research addresses this gap by combining exploratory data analysis with machine learning models to predict EV sales performance.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

III. PROPOSED METHODOLOGY

The proposed EV analytics system follows a structured machine learning pipeline. The workflow includes dataset collection, preprocessing, exploratory data analysis, machine learning model training, and prediction generation. The methodology is designed to analyze electric vehicle datasets and identify key factors influencing EV market growth.

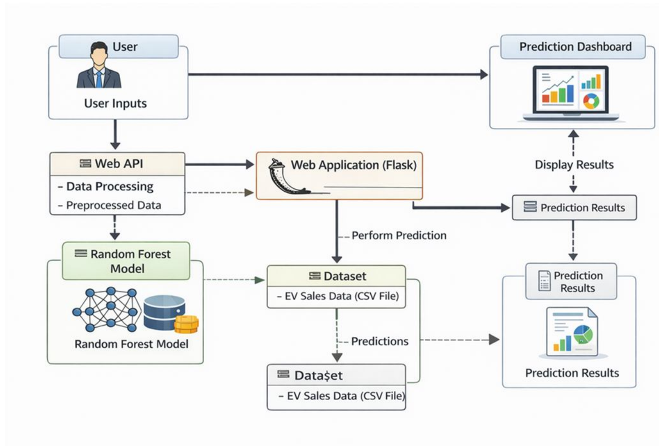


Figure 1: EV Analytics Workflow

IV. SYSTEM ARCHITECTURE

The architecture of the proposed system consists of five major modules that work together to analyze EV data and generate predictions.

1) Data Collection Module – Loads EV datasets from CSV or public EV data sources.
2) Data Processing Module – Cleans and preprocesses data by removing duplicates, handling missing values, and normalizing attributes.
3) Machine Learning Module – Trains predictive models using Linear Regression and Random Forest algorithms.
4) Backend Module – Flask framework processes requests and executes prediction models.
5) User Interface Module – Interactive dashboard displays insights and prediction results.

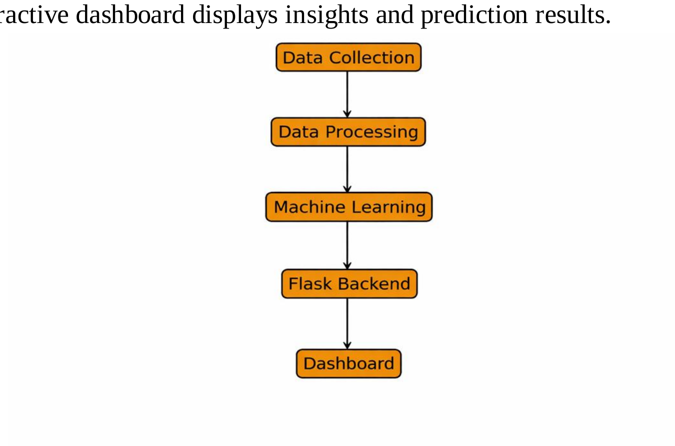


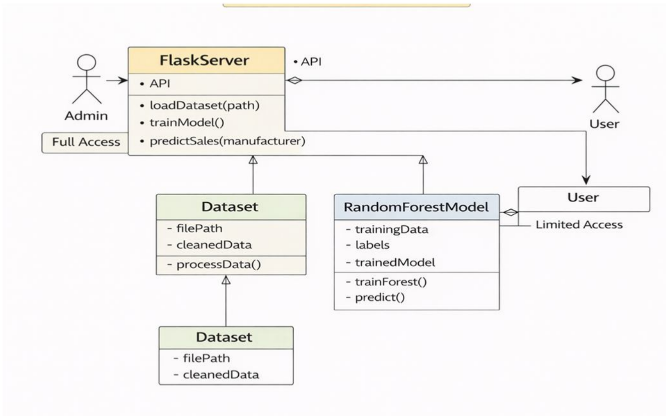
Figure 2: System Block Diagram

V. MACHINE LEARNING MODELS

Two machine learning algorithms were implemented in this research.

A. Linear Regression

Linear Regression is used as a baseline predictive model to identify linear relationships between EV features and sales performance. It is simple, interpretable, and computationally efficient.



B. *Random Forest Regression*

Random Forest is an ensemble machine learning algorithm that combines multiple decision trees to improve prediction accuracy. It handles nonlinear relationships and provides feature importance analysis. Random Forest demonstrated higher prediction accuracy compared to Linear Regression in EV sales forecasting.

VI. EXPERIMENTAL RESULTS

The EV dataset was used to train and evaluate the machine learning models. The performance of the models was evaluated using standard metrics such as R^2 Score, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

Results indicate that the Random Forest model achieved approximately 98% prediction accuracy. Battery capacity, vehicle price, and charging time were identified as key factors influencing EV sales performance.

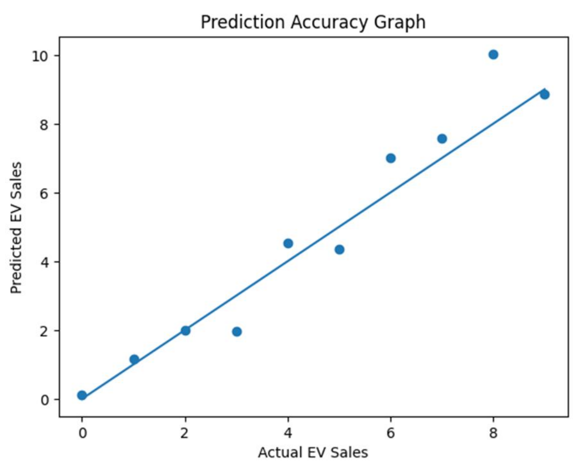


Figure 3: Prediction Accuracy Graph

VII. FUTURE SCOPE

Future research can extend this system by integrating deep learning models such as LSTM and XGBoost to improve prediction accuracy. Real-time EV datasets and cloud-based deployment can also be implemented to support large-scale EV analytics platforms. Additionally, the system can be expanded to predict battery health, charging demand, and energy consumption patterns.

VIII. DATASET DESCRIPTION

The dataset used in this research contains information about electric vehicles and their technical specifications along with market sales statistics. The dataset includes attributes such as battery capacity, charging time, vehicle price, driving range, energy consumption, safety ratings, and historical sales data. These attributes play a critical role in determining the performance and popularity of electric vehicles.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

The dataset was collected from publicly available EV databases and open data sources related to electric vehicle markets. The collected dataset was stored in CSV format and used as the input for data preprocessing and machine learning models.

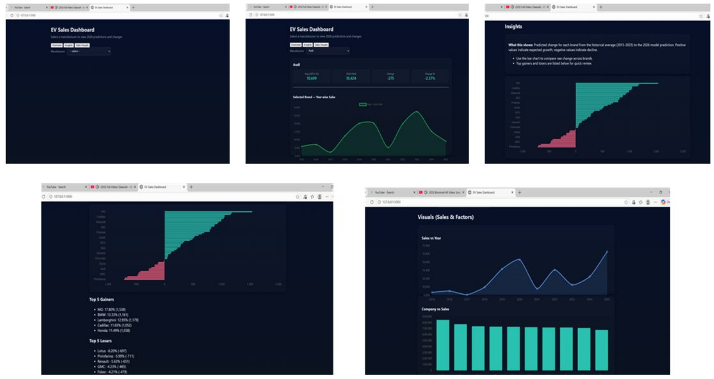


Figure 4: Output samples

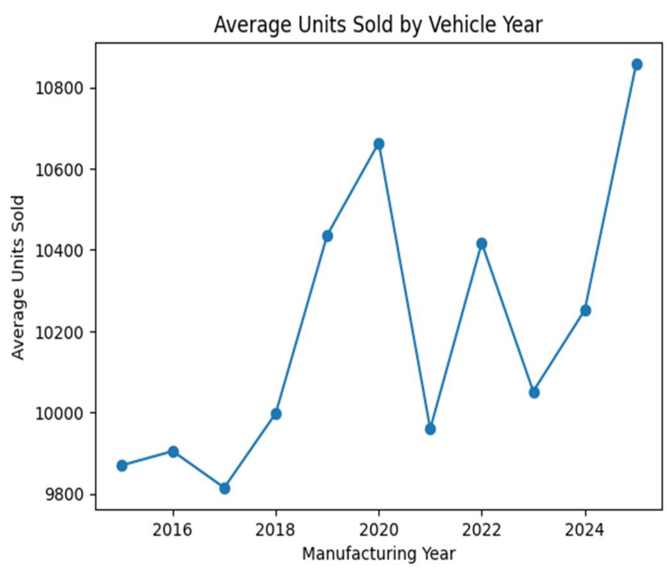
Each record in the dataset represents a specific electric vehicle model and its corresponding technical and market characteristics.

IX. EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis (EDA) is performed to understand the patterns, relationships, and distribution of data within the EV dataset. EDA helps identify important features that influence electric vehicle sales and performance.

Several visualization techniques are used in the analysis process including:

- Correlation matrices to identify relationships between variables
- Histogram plots to understand feature distribution
- Scatter plots to analyze relationships between EV attributes
- Box plots to detect outliers in the dataset



The EDA results revealed that battery capacity, vehicle price, and charging time have a strong correlation with EV sales trends. Vehicles with higher battery capacity and longer driving range tend to attract more consumers, which directly impacts sales performance.

X. ALGORITHM DETAILS

The machine learning models used in this project are Linear Regression and Random Forest Regression. These algorithms were selected based on their ability to handle structured numerical datasets and provide accurate predictions.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

A. Linear Regression Model

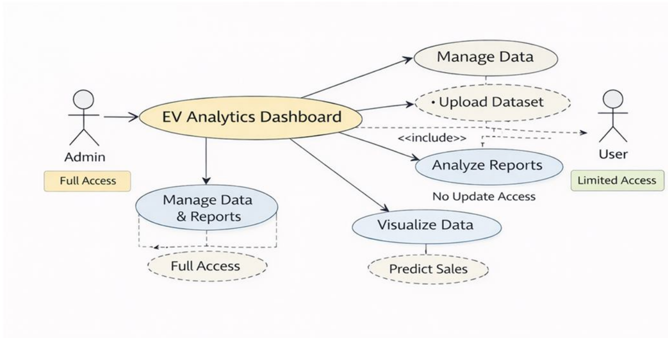
Linear Regression predicts the target variable by establishing a linear relationship between independent variables and the dependent variable. The mathematical representation of Linear Regression is:

$Y = \beta 0 + \beta 1X1 + \beta 2X2 + ... + \beta nXn$

Where Y represents predicted EV sales, X represents independent variables, and β represents model coefficients.

B. Random Forest Model

Random Forest is an ensemble learning technique that combines multiple decisiontrees to improve prediction accuracy. Each tree is trained using a random subsetof data and features. The final prediction is obtained by averaging the outputs ofall decision trees. Random Forest is effective in capturing complex relationshipsbetween EV features and sales patterns.



XI. SYSTEM TESTING

System testing ensures that all modules of the EV analytics system function correctly and efficiently.

1) Unit Testing: Each individual module such as data preprocessing, model training, and prediction generation was tested independently to verify correctness.
2) Integration Testing: Integration testing verified the communication between the machine learning models, backend processing system, and the user interface dashboard.
3) System Testing: The entire workflow from dataset input to prediction output was tested to ensure that the system performs accurately without runtime errors.

XII. ADVANTAGES OF THE SYSTEM

The proposed EV analytics and prediction system provides several advantages:

1) Accurate prediction of EV sales trends
2) Efficient processing of large datasets
3) Identification of key factors affecting EV market growth
4) Support for data-driven decision making
5) Scalable architecture that can handle future EV datasets
6) Interactive visualization dashboard for easy analysis

XIII. LIMITATIONS

Although the proposed system performs effectively, certain limitations exist.

1) The prediction accuracy depends on the quality of the dataset.
2) The system currently uses historical datasets rather than real-time data.
3) External factors such as government policies, fuel prices, and consumer behavior may influence EV sales but are not fully captured in the current dataset.

Future improvements can address these limitations by integrating real-time data sources and advanced deep learning models.

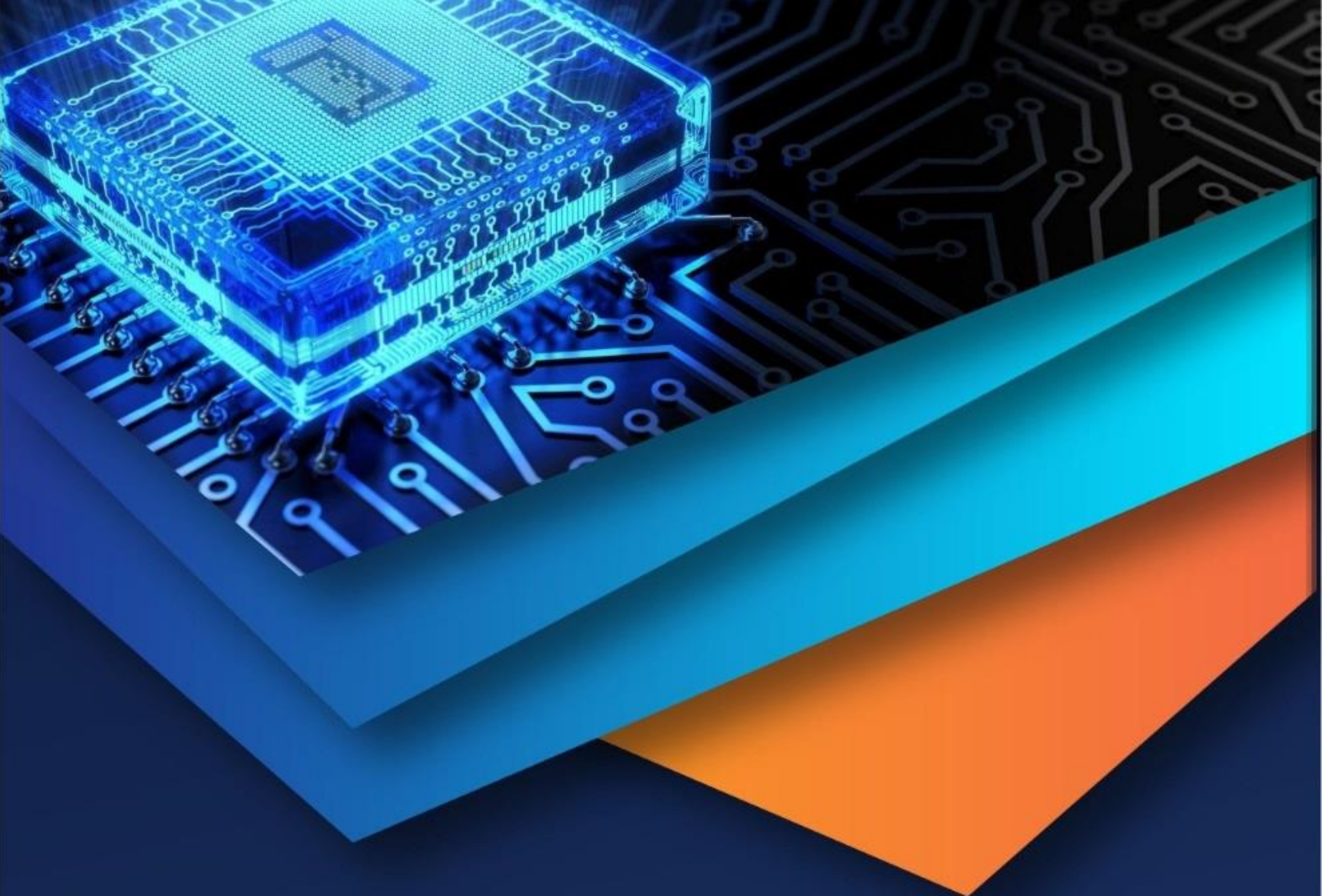
XIV. CONCLUSION

This research presented a machine learning-based Electric Vehicle Analytics and Sales Prediction system. The system successfully analyzes EV datasets and predicts future sales trends using Linear Regression and Random Forest lgorithms.

The results demonstrate that machine learning techniques can effectively identify patterns in EV datasets and provide accurate predictions. The proposed system supports data-driven decision making for EV manufacturers, investors, and policymakers while promoting sustainable transportation planning.

REFERENCES

[1] De Cauwer, J. Van Mierlo, and T. Coosemans, "Electric Vehicle Energy Consumption Prediction Using Machine Learning Techniques."
[2] J. Wu, H. Zhang, and Y. Li, "Machine Learning Based Energy Consumption Prediction of Electric Vehicles."
[3] S. Severson, P. Attia, and M. Chueh, "Prediction of Electric Vehicle Battery State of Health Using Machine Learning."
[4] Y. Wang, X. Ma, and F. Zhang, "Forecasting Electric Vehicle Sales Using Machine Learning Models."
[5] R. Mehta and S. Jain, "Performance Prediction of Electric Vehicles Using Random Forest and Neural Networks."



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY