Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Machine Learning Model for Prediction of Electric Vehicle Prices

[1]Dr. Shiksha Dubey, [2]Dr. Ashwini Renavikar and [3]Dr. Sonal Kanungo
[1]Assistant Professor, [2]Professor and [3]Associate professor, Thakur Institute of Management Studies, Career Development & Research (TIMSCDR)
[1]shiksha.dubey@timscdrmumbai.in, [2]ashwini.renavikar@timscdrmumbai.in and [3]sonal@timscdrmumbai.in

ARTICLE INFO

Received: 26 Dec 2024

Revised: 14 Feb 2025

Accepted: 22 Feb 2025

ABSTRACT

This research paper explores the use of machine learning models for predicting the price of electric vehicles (EVs), focusing on improving forecast accuracy and understanding market dynamics. The growing demand for EVs, driven by environmental concerns, technological advancements, and government incentives, underscores the need for accurate price prediction to support decision-making by businesses, policymakers, and consumers. The research applies and compares the performance of three machine learning models — Linear Regression, Decision Tree, and Random Forest — using a dataset of electric vehicles registered with the Washington State Department of Licensing. Results demonstrate that Random Forest outperforms other models in terms of predictive accuracy, highlighting its capacity to handle difficult, non-linear relationships and reduce variance. The study provides valuable insights for improving market strategies and enhancing the adoption of electric vehicles.

Keywords: Electric Vehicle, Machine learning models etc.

I. INTRODUCTION

The transportation sector is widely acknowledged as a major source of CO_2 emissions, a key driver of climate change [1]. The shift toward electric vehicles (EVs) has gained significant attention as a strategy to reduce the environmental impact of transportation. EVs have the potential to lower CO_2 emissions and improve air quality compared to traditional internal combustion engine vehicles. Transitioning to electric mobility is essential for achieving sustainable transportation and addressing the challenges of climate change and air pollution.

In recent years, interest in EV adoption has increased, with a focus on reducing the life-cycle CO_2 footprint. Studies have compared the reduction in CO_2 emissions from hybrid and electric buses within transit networks, providing insights into the environmental benefits of these technologies [2]. Research has also explored the benefits of solar-powered EV charging stations, highlighting both economic advantages and reductions in CO_2 emissions [3]. Additionally, the decrease in harmful emissions is a key benefit of EV adoption. Studies on carbonyl emissions from gasoline and diesel vehicles underscore the importance of transitioning to cleaner alternatives like EVs [4]. Further, the development of zero-emission drive units for battery electric vehicles has shown a significant reduction in tire emissions through the use of advanced filter systems [5]

This study conducts a comparative analysis of machine learning models to predict expected price of EV vehicles, focusing on global market trends, particularly in US. The goal is to evaluate the performance of different models in forecasting EV sales and understanding market dynamics. The demand for EVs has been rising steadily in recent years, driven by environmental concerns, government incentives, and advancements in EV technology. This growth is expected to continue as EVs become more affordable and diverse models enter the market. Accurately forecasting EV prices can provide valuable insights for businesses, consumers, financial institutions, and policymakers. This is shown in Fig.1.

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

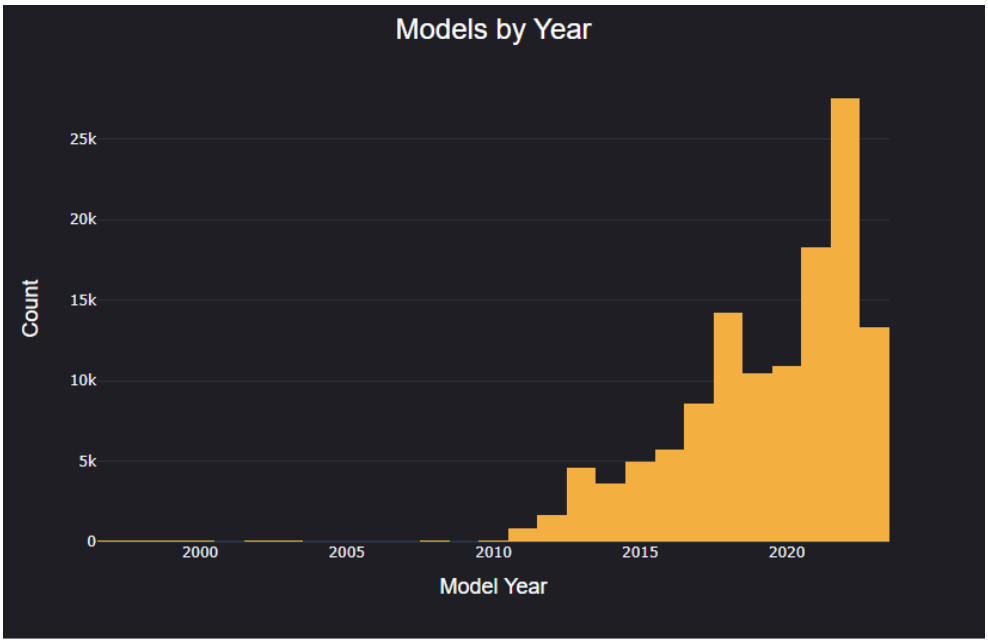


Fig.1:-Increasing trends of EV models in the recent years

Over the past year, the market for electric vehicles (EVs) has expanded. About 25,000 EVs were sold in CY2020, which is a notable tenfold increase over CY2010. The majority of the increase is attributable to the two- and three-wheeler segments, which are known as the "close to the bottom peaches" of the EV industry. Since they are less expensive than the electric passenger or commercial vehicle segments, they are the primary forces behind EV sales.

Objectives of the research

- To understand the rise in the use and sales of Electric vehicles globally particularly US
- To pre-process and analyse the dataset based on different feature variables.
- A comparative analysis on the performance of ML models based on different evaluation metrics

II. LITERATURE REVIEW

The purpose of machine learning (ML) models for predicting electric vehicle (EV) prices has gained significant attention in recent years due to the growing demand for EVs, driven by environmental concerns, technological advancements, and government incentives. For companies, legislators, financial institutions, and consumers to make well-informed decisions about EV investments and market strategies, accurate pricing forecast is essential. Regression models, artificial neural networks (ANN), support vector machines (SVM), and ensemble learning techniques are some of the machine learning (ML) approaches that have been investigated to increase the precision and dependability of EV price forecasts.

A brief overview on the use of ML models for prediction is discussed in the Table 1 below:-

Table 1: - Literature Summary

Author(s)	Year	Model Type	Key Attributes	Outcome	Limitations/ Recommendations
Noor and Jan[6]	2017	Multiple Linear Regression	Size, engine capacity, exterior color, posting date, number of ad views, power steering, mileage,	98% prediction accuracy	Proposed integrating many machine learning techniques into a group to increase the precision of

Research Article

			transmission type, engine type, location, registration area, layout, edition, make, model year		predictions.
Gonggie [7]	2011	Artificial Neural Networks (ANN)	Mileage, estimated car lifespan, brand	Improved accuracy over linear models due to better handling of non-linear data interactions	ANN-based model outperformed linear regression models
Wu et al.[8]	2009	Knowledge-Based Neuro-Fuzzy Method	Model, production year, engine size	Comparable results to simple regression; Developed ODAV (Optimal Distribution of Auction Vehicles) to help auto dealers maximize returns at lease end	Used k-nearest neighbor-based regression to predict vehicle speed; Processed over two million vehicle transactions
Richard son[9]	2009	Multiple Regression Analysis	Electric vs conventional vehicles	Electric vehicles retained value longer due to urban warming issues and better fuel efficiency	Suggested that automakers should focus on building more durable vehicles

MACHINE LEARNING APPROACHES

Through the use of mathematical models and data processing, machine learning (ML) enables computers to learn from and get better at exploiting data without explicit programming. Machine learning is part of the larger field of artificial intelligence (AI). Machine learning (ML) uses pattern recognition algorithms to analyze data and create predictive models. Similar to human learning, ML models can improve their predictions with more data and experience. This adaptability allows ML models to handle dynamic data and situations where coding a direct solution is impractical.

Machine Learning Types: - There are four primary categories of machine learning:

- Unsupervised learning (association and clustering)

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

- Supervised learning (classification and regression)
- Learning Under Semi-Supervision
- Learning Reinforcement

Supervised and unsupervised learning are the most commonly used approaches, while reinforcement learning is used for sequential decision-making tasks [10]. Currently, computers require training before they can make decisions independently. The term "supervised" refers to the presence of a teacher or expert guiding the learning process.The system is trained using labelled data in supervised learning, where the right answers (class labels) are pre-provided. Support Vector Machines (SVM), Random Forest, and Decision Trees are examples of popular supervised learning algorithms. Unsupervised learning, on the other hand, is applied when there are no class labels in the input data. Finding the data's underlying structure in order to categorize or organize it is the aim. Unsupervised learning can be divided into two primary categories: association and clustering. Affinity Propagation and K-means clustering are two well-known unsupervised methods.

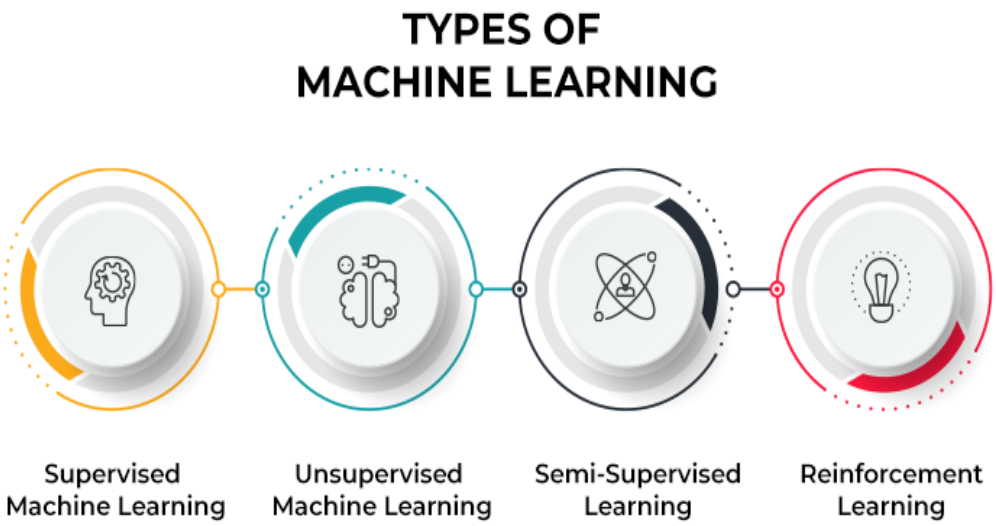


Fig 2: - Types of ML algorithms

III. DATASET PRE-PROCESSING

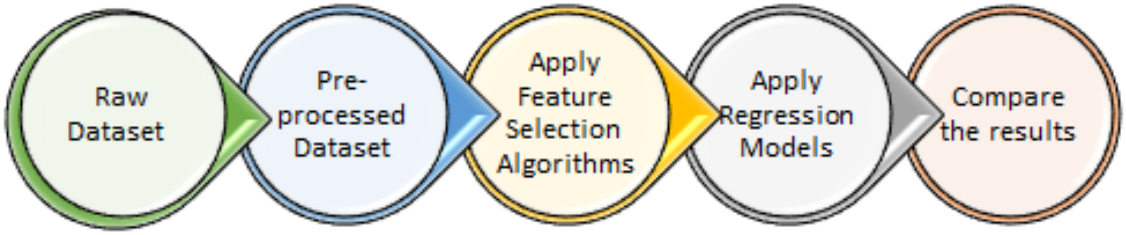


Fig.3:- Block Diagram of ML based Prediction for EV dataset

The dataset shows battery electric cars (BEVs) and plug-in hybrid electric vehicles (PHEVs) that are currently registered with the Washington State Department of Licensing (DOL). There are over two million records in this collection.

A brief description of the columns included is discussed in the below Table 2.

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 2: - Dataset Description

Column	Description
VIN (1-10)	The Vehicle Identification Number is a unique alphanumeric code assigned to each vehicle for identification purposes. This column represents the first 10 characters of the VIN.
County	The county where the vehicle is registered in Washington State.
City	The city where the vehicle is registered in Washington State.
State	The state where the vehicle is registered, which is Washington in this case.
Postal Code	The postal code of the registration address for the vehicle.
Model Year	The year in which the vehicle was manufactured.
Make	The specific model or name of the vehicle.
Electric Vehicle Type	Indicates whether the vehicle is a Battery Electric Vehicle (BEV), which runs solely on electricity, or a Plug-in Hybrid Electric Vehicle (PHEV), which combines electricity and an internal combustion engine.
Clean Alternative Fuel Vehicle (CAFV) Eligibility	Indicates if the vehicle meets the eligibility criteria for Clean Alternative Fuel Vehicle incentives or benefits.
Electric Range	The distance the vehicle can travel on electric power alone, typically measured in miles.
Base MSRP	The Manufacturer's Suggested Retail Price, which is the starting price set by the vehicle manufacturer.
Legislative District	The legislative district associated with the vehicle's registered address.
DOL Vehicle ID	A unique identifier assigned by the Washington State Department of Licensing (DOL) for each registered vehicle.
Vehicle Location	The precise location of the vehicle, which could be the address or coordinates.
Electric Utility	The name of the electric utility company associated with the vehicle, if applicable.
2020 Census Tract	The census tract associated with the vehicle's registered address, based on the 2020 Census data.

A cumulative summary of the dataset is visualized in the below dataset in Table 3 .

Table 2: Cumulative descriptive summary based on some significant set of features

	Postal Code	Model Year	Electric Range	Base MSRP	Legislative District	DOL Vehicle ID	2020 Census Tract
count	177861.000000	177866.000000	177866.000000	177866.000000	177477.000000	1.778660e+05	1.778610e+05
mean	98172.453506	2020.515512	58.842162	1073.109363	29.127481	2.202313e+08	5.297672e+10
std	2442.450668	2.989384	91.981298	8358.624956	14.892169	7.584987e+07	1.578047e+09
min	1545.000000	1997.000000	0.000000	0.000000	1.000000	4.385000e+03	1.001020e+09
25%	98052.000000	2019.000000	0.000000	0.000000	18.000000	1.814743e+08	5.303301e+10
50%	98122.000000	2022.000000	0.000000	0.000000	33.000000	2.282522e+08	5.303303e+10
75%	98370.000000	2023.000000	75.000000	0.000000	42.000000	2.548445e+08	5.305307e+10
max	99577.000000	2024.000000	337.000000	845000.000000	49.000000	4.792548e+08	5.603300e+10

The above table includes a summative summary statistic based on different parameters as reflected in the above table.

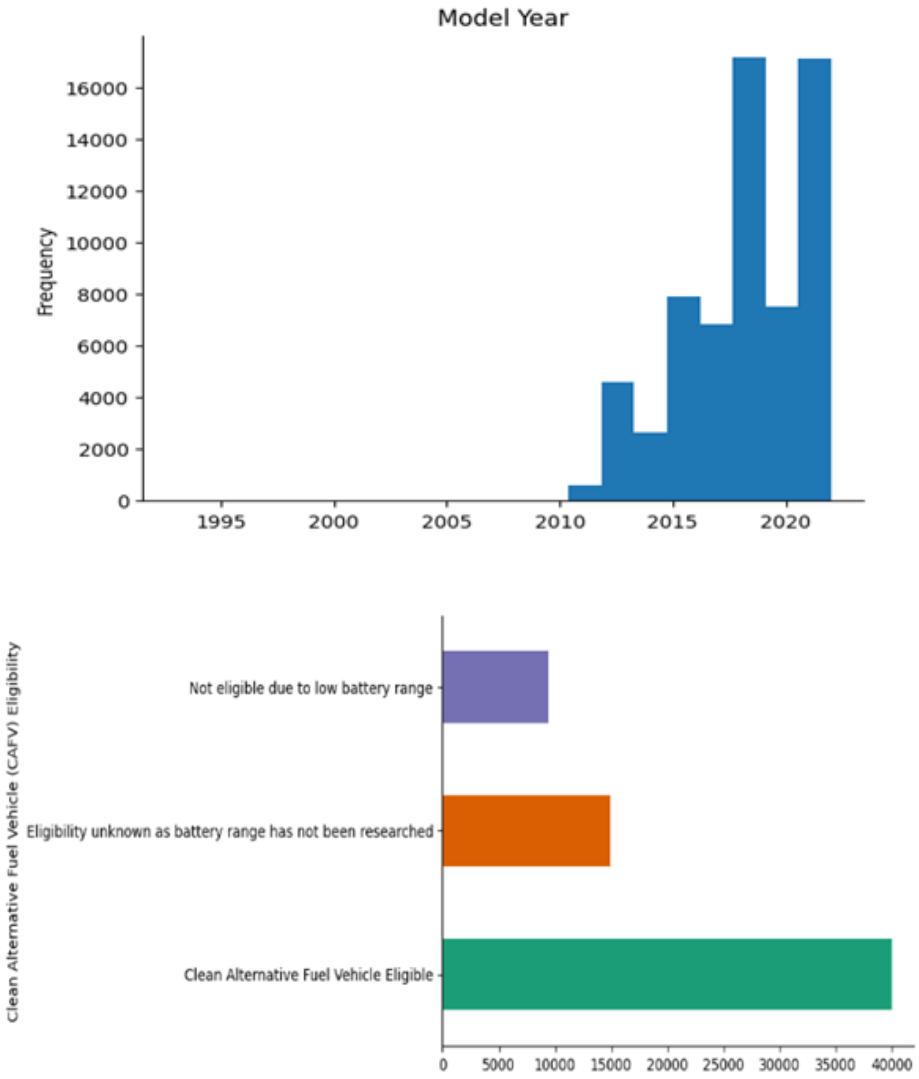
Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Univariate Analysis

Analyzing a single variable to determine its distribution and properties is known as univariate analysis. Because the word "uni" means "one," this analysis only looks at one variable at a time. The primary goal is to summarize and describe the variable's distribution without exploring relationships with other variables. Some significant features variables are analysed in the Fig. 4.

MODEL OF DIFFERENT EV VEHICLES



Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

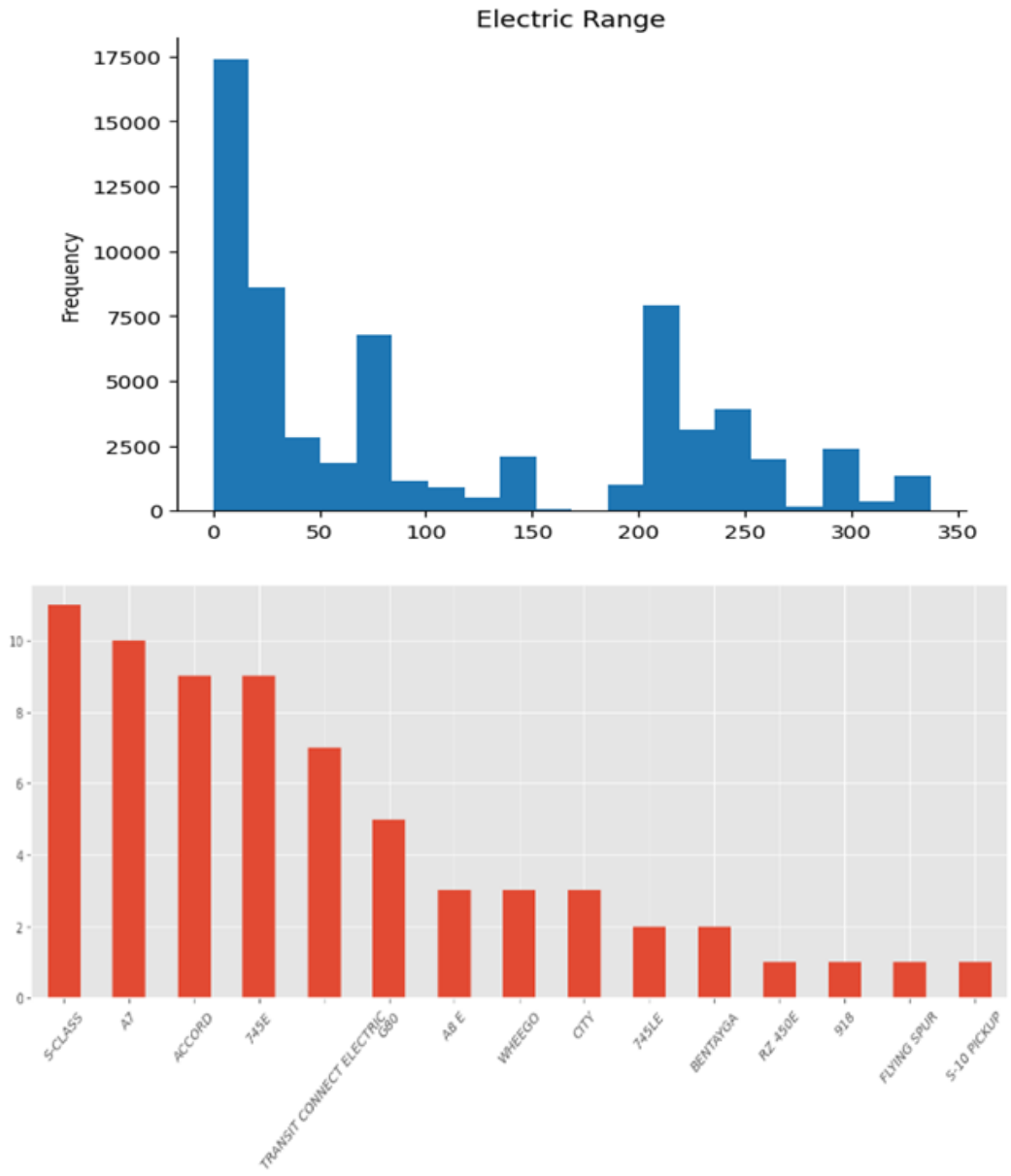


Fig. 4:- Univariate Analysis of feature variable

The data types of the dataset for the given features variable is shown as follows in Fig. 5: -

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

VIN (1-10)	object
County	object
City	object
State	object
Postal Code	float64
Model Year	int64
Make	object
Model	object
Electric Vehicle Type	object
Clean Alternative Fuel Vehicle (CAFV) Eligibility	object
Electric Range	int64
Base MSRP	int64
Legislative District	float64
DOL Vehicle ID	int64
Vehicle Location	object
Electric Utility	object
2020 Census Tract	float64
dtype: object	

Fig.5:- Datatypes of feature variable

Since majority feature variables are of object type which needs be converted into integer types for further pre-processing. The label encoding technique is applied and the results of conversation is shown below in Fig. 6.

	0
ID	int64
VIN (1-10)	int64
County	int64
City	int64
State	int64
ZIP Code	float64
Model Year	float64
Make	int64
Model	int64
Electric Vehicle Type	int64
Clean Alternative Fuel Vehicle (CAFV) Eligibility	int64
Electric Range	int64
Base MSRP	int64
Legislative District	float64
DOL Vehicle ID	int64
Vehicle Location	int64
Electric Utility	int64

Fig. 6:- Pre-processed datatypes of feature variables

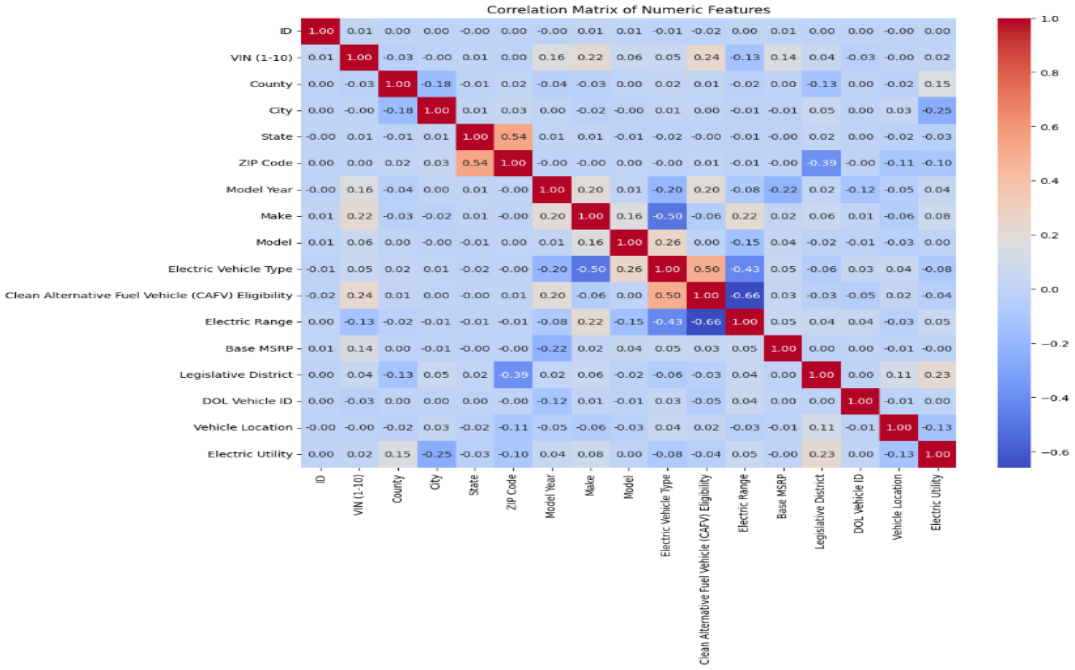
BIVARIATE ANALYSIS

A statistical method for analyzing the relationship between two variables is called bivariate analysis. We can better grasp how one variable affects or is related to another thanks to this analysis. It is especially helpful when

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

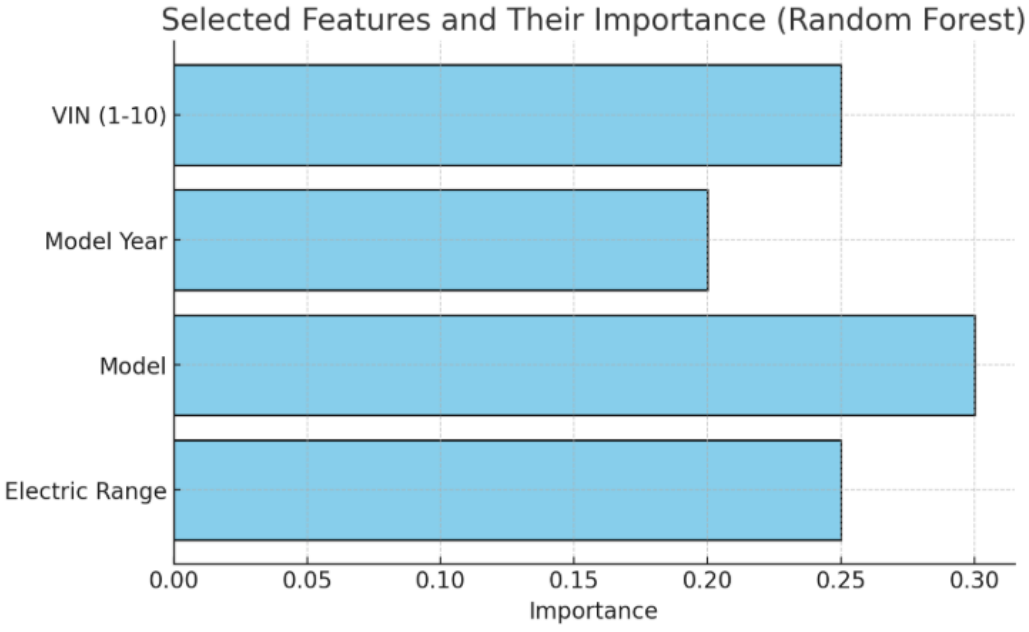
Research Article

attempting to investigate the link between one or more independent feature variables (predictor variables) and a target variable (dependent variable).



The correlation matrix reveals key relationships among numeric features in the dataset. **State and ZIP Code** show a moderate positive correlation (-0.66), indicating that higher electric range tends to reduce eligibility for clean alternative fuel vehicle benefits. Most other variables show weak or no correlation, suggesting little to no linear relationship.

Feature selection algorithms are used to identify the most relevant features in a dataset, improving model performance and reducing overfitting. Feature selection helps by removing irrelevant or redundant features, which can make the model simpler and more interpretable. Embedded method has been applied on the given the dataset and the results obtained are as follows



Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

IV. RESULTS

Machine Learning Models

The following machine learning models are applied on the dataset obtained by eliminating the irrelevant features and selecting only the relevant ones.

- By drawing a straight line through the data points, linear regression aims to model the connection between one or more independent variables (X) and a dependent variable (y). It minimizes the **Mean Squared Error (MSE)** to estimate the best-fit line.

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + \epsilon$$

- b_0 = Intercept
- b_n = Coefficients (weights)
- ϵ = Error term

- Decision Tree Regression uses a tree-like structure to divide the data into subsets, with each leaf representing the expected output value and each internal node representing a choice based on a feature..

1. Split data based on the feature that reduces variance the most.
2. Repeat until a stopping criterion is met (e.g., max depth or min samples).

$$y = \frac{1}{N} \sum_{i=1}^{N} y_i$$

- Random Forest is an ensemble learning technique that minimizes overfitting and increases accuracy by combining several decision trees. It functions by:

1. Building multiple trees using bootstrap sampling.
2. Averaging the predictions from all trees to reduce variance.

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t(x)$$

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

where:

- T = number of trees
- $h_t(x)$ = prediction from tree t

The above mentioned regression algorithms were applied on the pre-processed dataset and the results obtained are shown in the following Table for predicting the price of electric vehicles.

Model	MSE ↓ (Lower is Better)	R-squared ↑ (Higher is Better)
Linear Regression	488.73	0.3087
Decision Tree	146.55	0.7927
Random Forest	144.13	0.7961

The evaluation of machine learning models for electric vehicle price prediction reveals that **Random Forest Regression** demonstrates the best overall performance among the models tested. The Random Forest model leverages the combined strength of multiple decision trees, which allows it to strike an optimal balance between **bias** and **variance**. This ensemble-based approach enhances accuracy and reduces overfitting, making it the most effective model for predicting EV prices. **Decision Tree Regression** also performs well in capturing complex, non-linear relationships within the dataset. However, it is more prone to **overfitting**, where the model becomes too tailored to the training data, reducing its generalization ability on new data.

In contrast, **Linear Regression** struggles to handle the complexity and non-linearity present in the dataset. Its poor performance is reflected in a high **Mean Squared Error (MSE)** and low **R-squared value**, indicating that it cannot accurately capture the underlying patterns in the data. Overall, the results highlight that ensemble methods like Random Forest are better suited for complex, structured data, while simple linear models are less effective when dealing with intricate relationships in the dataset.

CONCLUSION

With the lowest Mean Squared Error (MSE) of 144.13 and the highest R-squared value of 0.7961, Random Forest Regression performs the best overall when machine learning models are evaluated and compared for predicting the pricing of electric vehicles. This outcome confirms that Random Forest's ensemble learning approach effectively balances bias and variance, leading to improved accuracy. Decision Tree Regression also shows strong performance with an MSE of 146.55 and R-squared of 0.7927, capturing complex patterns but with a higher risk of overfitting. In contrast, Linear Regression struggles with the dataset's complexity, yielding a high MSE of 488.73 and a low R-squared of 0.3087, indicating poor fit. The findings emphasize the importance of using ensemble-based models for handling complex, non-linear data and provide strategic insights for businesses and policymakers in the growing EV market.

[1] Babcock, H. M. (2009). Global Climate Change: A Civic Republican Moment for Achieving Broader Changes in Environmental Behavior. *Pace Envtl. L. Rev.*, *26*, 1.

[2] Tripathi, S., Gorbatenko, I., Garcia, A., & Sarathy, S. M. (2025). Life cycle environmental and cost impacts of bus transit networks: A real drive cycle evaluation of different powertrains. *Energy Conversion and Management*, *326*, 119459.

[3] Khan, S., Ahmad, A., Ahmad, F., Shafaati Shemami, M., Saad Alam, M., & Khateeb, S. (2018). A comprehensive review on solar powered electric vehicle charging system. *Smart Science*, *6*(1), 54-79.

[4] Pacura, W., Szramowiat-Sala, K., & Gołaś, J. (2023). Emissions from light-duty vehicles—From statistics to emission regulations and vehicle testing in the European Union. Energies, 17(1), 209.

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[5] Wieser, S., Reiland, S., Bondorf, L., Löber, M., Schripp, T., & Philipps, F. (2022, November). Development and testing of a zero emission drive unit for battery electric vehicles. In 2022 Second International Conference on Sustainable Mobility Applications, Renewables and Technology (SMART) (pp. 1-6). IEEE.

[6] Noor, K., Jan, S. (2017). Vehicle Price Prediction System using Machine Learning Techniques. International Journal of Computer Applications, 167(9), 27-31.

[7] Gongqi, S., Yansong, W., Qiang, Z. (2011). New Model for Residual Value Prediction of the Used Car Based on BP Neural Network and Nonlinear Curve Fit. In Measuring Technology and Mechatronics Automation (ICMTMA), 2011 Third International Conference, Vol. 2, pp. 682-685, IEEE

[8] Wu, J.D., Hsu, C.C., Chen, H.C. (2009). An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference. Expert Systems with Applications, 36(4), 7809-817. 10.

[9] Richardson, M. S. (2009). Determinants of used car resale value. Retrieved from: https://digitalcc.coloradocollege.edu/islandora/ 3A1346 [accessed: August 1, 2020.]

[10] Kulkarni, P. (2012). Reinforcement and systemic machine learning for decision making. John Wiley & Sons.