**Journal of Computer Science and Technology Studies**
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

# Forecasting Electric Vehicle Adoption in the USA Using Machine Learning Models

Sanjib Kumar Shil[1], Muhammad Shoyaibur Rahman Chowdhury[2], Nikhil Rao Tannier[3],Mir Mohtasam Hossain Sizan[4] , Rabeya akter[5], Nisha Gurung[6], MD Tanvir Rahman Tarafder[7]

[1]MBA in Management Information System, International American University, Los Angeles, CA, USA
[2]Masters in information technology, Gannon University, Erie, PA, USA
[3]Masters in Artificial Intelligence, University of North Texas, Denton, Texas, USA
4Master of Science in Business Analytics, University of North Texas, Denton, Texas, USA
[5]Master of science in information technology. Washington University of Science and Technology, Alexandria, VA, USA
[6]MBA Business Analytics, Gannon University, Erie, PA, USA
[7]Master of Science in Information Technology, Westcliff University, Irvine, California, USA

**Corresponding Author**: Sanjib Kumar Shil, **E-mail**: sanjib.job.bd@gmail.com

## | ABSTRACT
Electric vehicles (Electric Vehicles) are at the vanguard of the global dispensation to sustainable transportation, depicting a pivotal step toward diminishing greenhouse gas emissions and reliance on fossil fuels. Notwithstanding, the adoption of Electric Vehicles has been growing in the USA, but their future remains at a crossroads. The objective of the research is to design and execute machine learning models capable of providing accurate predictions of future trends in electric vehicle adoption in the USA. The dataset gathered for analyzing EV adoption in the USA comprises data across three primary categories: environmental data, economic indicators, and policy-related data. The economic indicators include household income, fuel prices, electricity rates, and lithium battery costs that affect EV purchasing power obtained from the U.S. Census Bureau and the U.S. Energy Information Administration (EIA). Environmental data include greenhouse gas emissions and air quality indices from the EPA, providing information on regional environmental conditions that might affect EV attractiveness. Other policy data included federal and state incentives such as tax credits, rebates, and EV infrastructure data, collected from the U.S. Led by the U.S. Department of Energy's Alternative Fuels Data Center and the Energy Laboratory, additional EV sales trends were pulled from databases of the automotive industry. In this research project, credible and proven machine learning models were employed, most notably, Linear Regression, Random Forest, and XG-Boost. The performance of the models was tested for EV adoption prediction by considering a few important metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared ($R^2$). From the model performance metrics presented, the Gradient Boosting Regressor and Random Forest models performed far better by a big margin than Linear Regression. The prediction models, particularly, the Random Forest regressors and Gradient Boosting regressors demonstrated incredible forecasting of electric vehicle adoption. The model works excellently on the premise that historical data with relevant features can be utilized to gain some valuable insight into future trends. Policy-makers interested in stimulating the wider use of electric vehicles can ensure that targeted policies address both current barriers and future demands. Results of this analysis suggest incentives, such as tax credits, rebates, and subsidies, are some of the most common actions to reduce the upfront cost of an EV, a key circumventing factor in the choice that many consumers face.

## | KEYWORDS
Electric Vehicles Adoption; Machine Learning Models; EV Adoption Prediction; Greenhouse Gas Reduction; Random Forest; Gradient Boost Regressors

## | ARTICLE INFORMATION
**ACCEPTED:** 20 October 2024 **PUBLISHED:** 10 November 2024 **DOI:** 10.32996/jcsts.2024.6.5.6

## I. Introduction

### Background and Motivation

Afandizadeh et al., (2023), posit that electric vehicles (Electric Vehicles) are at the vanguard of the global dispensation to sustainable transportation, depicting a pivotal step toward diminishing greenhouse gas emissions and reliance on fossil fuels. Entry-level forecasting has become an indispensable tool for understanding the likely pace with which this transition would take place and identifying the factors that may encourage or hinder this shift. While environmental concerns and public awareness about climate change rise, there is a dire need for the transport industry to shift towards cleaner alternatives. Bas et al. (2023), contend that Electric Vehicles present an immediate solution to lowering carbon emissions, but the factors that determine their adoption are entangled in a complex web that spans from government policies to infrastructure and market economics. Predicting the growth trajectory of Electric Vehicles helps policymakers, industrial leaders, and investors make decisions that help further environmental sustainability goals, infrastructure needs, and economic growth in its entirety. Forecasting provides a window into possible future scenarios, pointing at everything from manufacturing strategies down to investment in EV-related technologies.

In that respect, Bampos et al. (2023), argues that the adoption of Electric Vehicles has been growing in the USA, but their future remains at a crossroads. While EV sales have seen a considerable rise in the last few years, their numbers are still a fraction of the entire vehicle market. That the USA is one of the most significant vehicle markets in the world opens up an entire can of worms: Full-on acceptance of Electric Vehicles would thus have a substantial effect on global carbon emissions, energy consumption, and environmental policies. However, Hitesh et al. (2024), uphold that such a transition into prominence for electric vehicles would depend on the pace and depth of government incentives, technology advances, prices of oil, and consumer acceptance. Although current growth in the adoption of Electric Vehicles is promising, it varies enough from state to state that developing a consistent national forecast is tricky. The divergence has once again reiterated the requirement of an effective forecasting model that would consider variation across regions and change with the economic, policy, and environmental conditions.

### *Objectives:*

The main objective of the research is to design and execute machine learning models capable of providing accurate predictions of future trends in Electric vehicle adoption in the USA. Most of the traditional forecasting methods may be ineffective in handling high-dimensional, nonlinear relationships between the variables affecting Electric Vehicles' adoption, including fuel prices, consumer income, government subsidies, and the availability of charging infrastructure. Therefore, the complication that a machine learning model will be in a better position to capture, resulting in an intricate and dynamic forecast. In this research project, we would like to use the latest machine learning algorithms for predicting EV adoption based on the wide range of factors acting as driving agents in consumer decisions and market trends.

## II. Literature Review

### *Overview of the various studies on the adoption of Electric Vehicles*

Dixit et al. (2022), articulated that the electric vehicle (EV) transition has been the subject of comprehensive research because of its implications for climate change mitigation, urban planning, and energy sustainability. Most early studies in this area were based on conventional statistical methods such as linear regression and time-series analysis for the forecasting of their adoption rates. Jia et al. (2020), asserted that While these models were useful for broad adoption trends, they often fell short in capturing complex relationships between the variables of policy incentives, charging infrastructures available, and consumer behavior. For instance, from the linear modeling, general correlations could be captured while nonlinear dynamics-such as the compound effect of rising fuel prices coupled with tax incentives in bringing up the adoption rate by consumers-could not be accounted for. Whereas these conventional approaches give very important insights into the patterns of adoption, they barely capture the contributions of manifold factors influencing the EV markets.

Notwithstanding, Kamis & Abraham (2024), indicated that recently, the trend has shifted to modern methods based on more sophisticated analytical techniques such as logistic regression, random forest, XG-Boost, agent-based models, and system dynamics. Although these approaches have allowed researchers to consider a greater variety of influencing factors and achieve more subtlety in understanding adoption trajectories, agent-based models, for instance, simulate interactions among individual consumers, governments, and manufacturers' diverse decision processes together to drive EV adoption. Khusanboev et al. (2023), affirmed that these models are computationally intensive and generally require a great deal of data about consumer behavior, which is usually not available. Other recent studies employed econometric models to analyze how specific variables-such as subsidies or fuel affect EV market share. Although such econometric methods enable the analysis of a cause-effect relationship, those models are in most cases based on assumptions about the independence of the variables, a supposition that goes against the complex interdependencies inherent in an adoption process.

***Gaps in Recent Research***

Gerossier et al. (2024), states that despite noteworthy advancements, existing predictive methods for Electric vehicle adoption present a myriad of gaps, particularly in their capability to account for high-dimensional and interdependent variables that shape the EV market. Traditional methods can barely handle the complexity and nonlinearity of factors affecting the diffusion of Electric Vehicles, such as fluctuating fuel prices, variable government subsidies, and technological developments regarding pledged and commercially available batteries. These time-series models, for example, while they do predict from the historical data, cannot easily capture sudden changes in policy and technology. Naseri et al. (2023), point out that traditional approaches also rely on aggregate data and linear assumptions, which obscure variations in regional rates of adoption and mask the underlying trends in consumer behavior and preference. This is especially true in the context of the USA, with wide variation in EV adoption rates across states that are influenced by great differences in policy, climate, and socioeconomic conditions.

This current literature review gives evidence of the limited scope of machine learning techniques in predicting EV diffusion. The advantages provided by the Machine Learning model can be pointed out for processing large volumes of data and extracting from them hidden patterns that are not so clearly visible with traditional statistical analysis. By employing algorithms to handle complex nonlinear relationships among variables, machine learning approaches can give more accurate and foregoing forecasts (Vishnu et al. 2024). While some studies have now begun to explore applications of machine learning, these are sparse and often further constrained by limited access to high-quality large-scale data. Much of the existing machine learning work tends to concentrate on short-term prediction rather than providing longer-term insights into adoption trends, which are crucial for policy-making and infrastructure planning ( Ullah et al., 2022).

Yaghoubi et al. (2024), opines that in the face of these gaps, there is a need for an advanced model that could incorporate a wide range of factors affecting EV adoption, including dynamic market conditions. Some of these limitations could be overcome with advanced machine learning models: ensemble methods, neural networks, and gradient boosting. These can operate on high-dimensional datasets, learn from new data inputs, and provide robust predictions allowing for variability and complexity of the EV market. Moreover, the applications of such models address the lacunae in the traditional and econometric modeling approaches and further our understanding of interdependencies among economic, environmental, and policy-related factors at work in the adoption of Electric Vehicles (Yi et al., 2022). It thereby intends to critically bridge the information gap in EV diffusion forecasts and help decision-makers formulate targeted policies and strategies, evidenced by data, to expedite the transition to greener transport.

## III. Data Collection and Preprocessing

***Data Source***

The dataset gathered for analyzing EV adoption in the USA is comprised of data across three primary categories: environmental data, economic indicators, and policy-related data. The economic indicators include household income, fuel prices, electricity rates, and lithium battery costs that affect EV purchasing power obtained from the U.S. Census Bureau and the U.S. Energy Information Administration (EIA). Environmental data include greenhouse gas emissions and air quality indices from the EPA, providing information on regional environmental conditions that might affect EV attractiveness. Other policy data included federal and state incentives such as tax credits, rebates, and EV infrastructure data, collected from the U.S. Led by the U.S. Department of Energy’s Alternative Fuels Data Center and the Energy Laboratory, additional EV sales trends were pulled from databases of the automotive industry (Pro-AI-Robikul, 2024). Combined, these sources provide a comprehensive and region-specific dataset that allows for an in-depth analysis of the key drivers of EV adoption in the U.S.

***Data Pre-Processing***

The preparation of the EV adoption dataset for analysis involves several steps in data preprocessing: first, making sure that qualitatively the data is appropriate and compatible with any model. Cleaning the data first involved the elimination of incorrect entries, such as duplicate records and mislabeled data entries. Missing values in such datasets, coming from diverse sources, were imputed based on neighbors in the case of time-series variables or on median/mean values for economic and environmental ones. In the process, some highly missing data-containing rows are removed to maintain data integrity. General outliers, but most especially in economic indicators of fuel prices and household incomes, were identified using IQR and then capped to mitigate their skewing effect or analyzed for their potential impact on the model. Min-max scaling normalization was applied to economic and policy variables to scale values for features between 0 and 1 so that all features have an equal scale for better convergence during model training (Pro-AI-Robikul, 2024). These include preprocessing steps that result in a cleaned and normalized dataset, retaining important information while enhancing model readiness.

**Exploratory Data Analysis (EDA)**

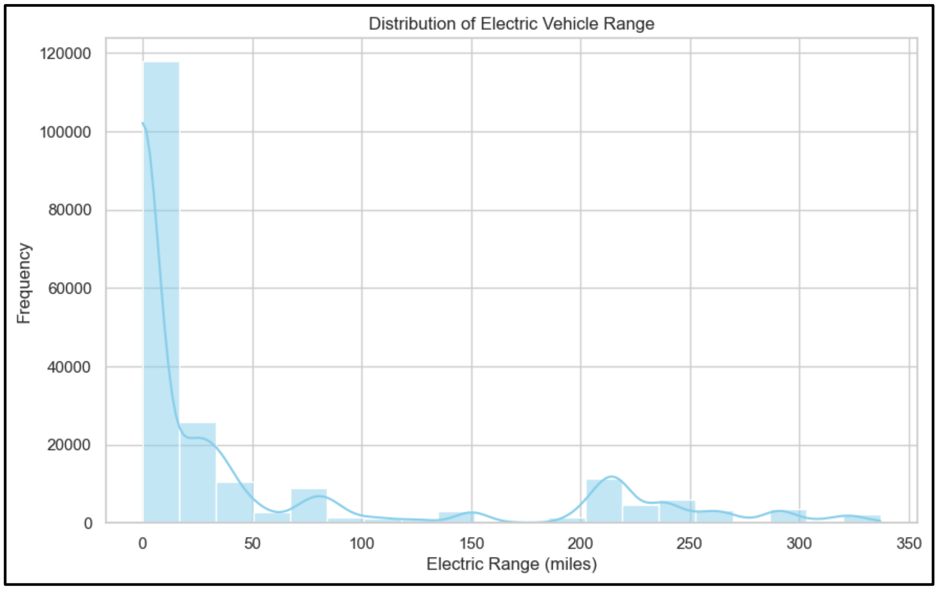


***Figure 1: Visualizes Distribution of Electric Vehicle Range***

The histogram above showcases the distribution of the electric vehicle range. The x-axis represents the electric range in miles, and the y-axis represents frequency. The data in this graph is right-skewed with a peak at around the 25-mile mark. This shows that most electric vehicles have a relatively low range. Similarly, a notable secondary peak comes at around 225 miles, which shows that a cluster of longer-range vehicles also exists. The overall distribution is multimodal; there are several distinct peaks, suggesting that there may be subcategories or segments in the electric vehicle market. The superimposed density curve consists of a smooth approximation of the distribution and underlines the right-skewed nature of the distribution, with multiple modes.

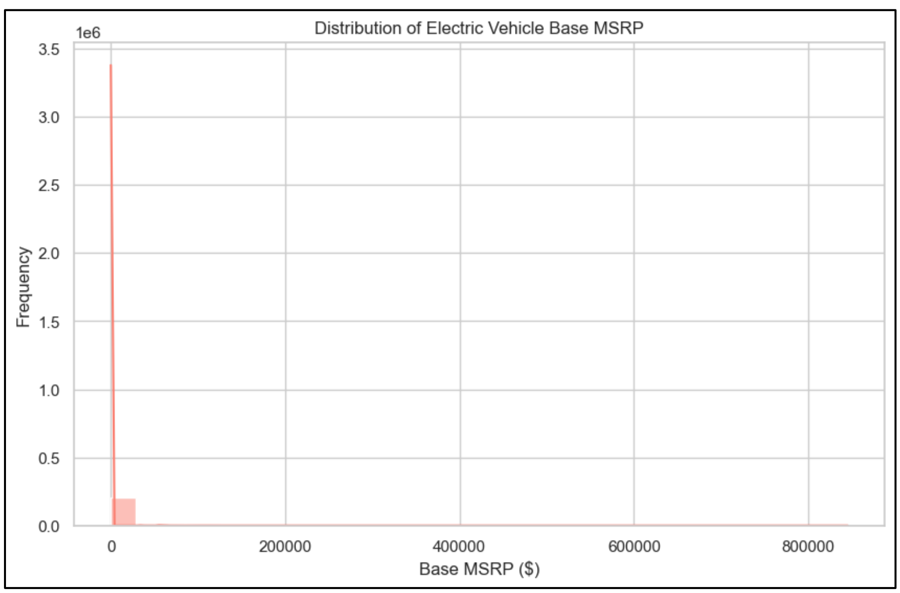


***Figure 2: Depicts the Distribution of Electric Vehicles Base MSRP***

This histogram portrays the distribution of the electric vehicle base Manufacturer's Suggested Retail Price (MSRP). The horizontal axis is MSRP in dollars, and the vertical axis represents frequency. The data distribution seems right-skewed (a long tail points toward higher values of MSRP). This would indicate that most electric vehicles have a far lower MSRP, and just a few vehicles are high-priced. It is peaked and crowded around the low end of the price range, meaning there are a great number of vehicles in the lower price range. This right-skewed graph illustrates that a few high-priced models create this shape.

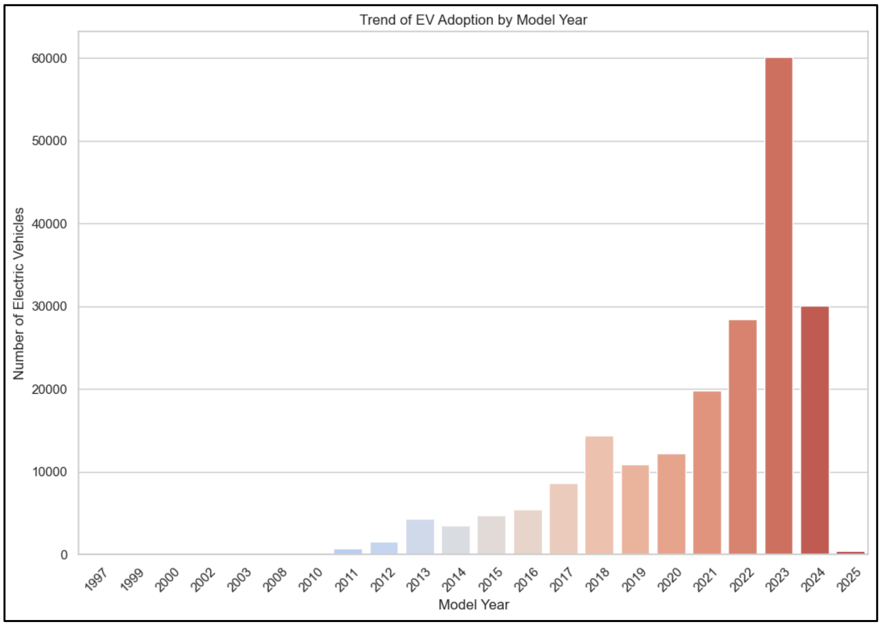


***Figure 3: Visualizes the Trend of EV Adoption by Model Year***

The bar chart above displays the trend in the adoption of electric vehicles, sorted by model year. The x-axis reflects the model year, while the y-axis represents the number of electric vehicles. From the trend of the data, it can be observed that this has been a growing trend over time. Initially, the number of Electric Vehicles was relatively low with certain years recording spikes here and there. Although the rate of adoption began to pick up some steam in about 2010, the number of Electric Vehicles dramatically increased each year. This growth continued unabated until 2023 when the number reached a peak. According to the chart, the electric vehicle market-popularly known as ELECTRIC VEHICLES has grown by leaps and bounds over the past couple of years due to improvements in technology, consumer awareness, and government policies.

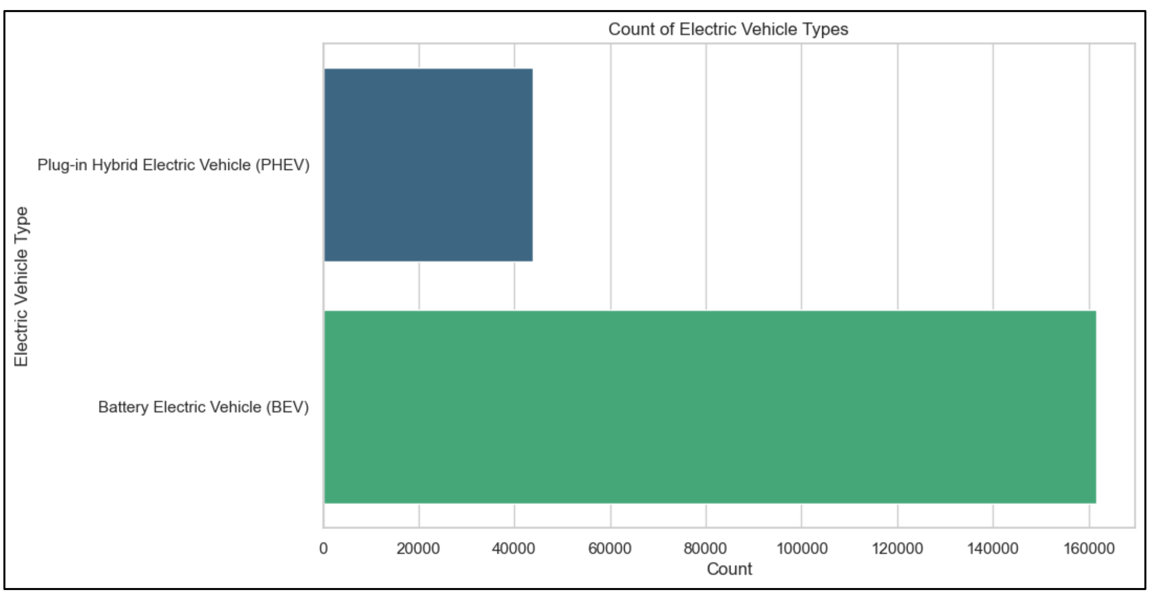


***Figure 4: Exhibits Count of Electric Vehicle Types***

This bar chart compares the count of different types of electric vehicles. The x-axis shows the count, while the y-axis refers to the electric vehicle type. It can be obtained from this data that BElectric Vehicles, by far, outnumber the PHElectric Vehicles in the dataset. The length of the bar for BElectric Vehicles is much longer compared to that for PHElectric Vehicles, which further highlights the discrepancy. Therefore, BElectric Vehicles have a higher market presence compared to PHElectric Vehicles which may be on account of several issues related either to driving range, operating cost, or infrastructure charging.

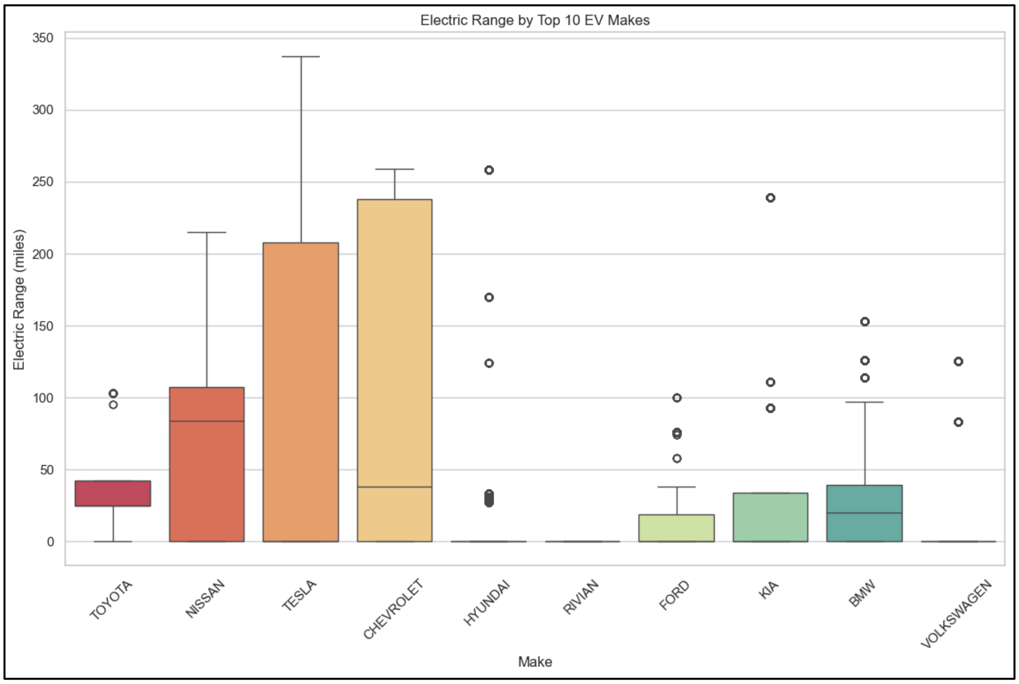


***Figure 5:Showcases Electric Range by Top 10 EV Makes***

This boxplot shows the distribution of electric range across the top 10 EV makes. The x-axis reflects the car's make, and the y-axis reflects the electric range in miles. There is a huge variation in electric range across different car makes. Tesla, Chevrolet, and Rivian have the highest median electric ranges, indicating that they offer models with longer driving capabilities. On the other hand, Toyota, Nissan, and Hyundai have comparatively lower median ranges. It is also observable that within the makes, there are outliers and whiskers, signifying true variability within each make due to some model varieties that can give considerably higher or lower ranges than the median. Overall, the box plot shows the range of electric vehicles available for different needs and preferences regarding driving distance.

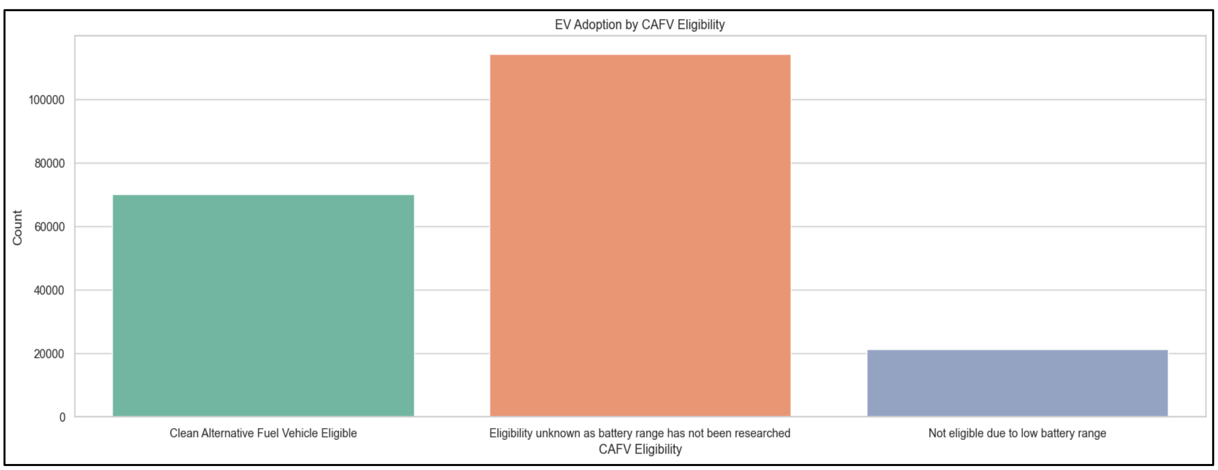


***Figure 6: Displays EV Adoption by CAFV Eligibility***

The histogram above reveals the split of matched electric vehicles into their CAFV program eligibility: the x-axis represents CAFV eligibility categories, and the y-axis represents the count of electric vehicles. As indicated, the highest number of electric vehicles pertains to matched Electric Vehicles eligible for the CAFV program, while the second pertains to unknown categories because of the lack of research in the field of the battery range. Fewer are ineligible due to their insufficient battery range. That is, most of the Electric Vehicles qualify for CAFV incentives, undetermined, or remain ineligible due to their respective ranges.

## IV. Methodology

### *Model Selection*

Model selection is a very crucial step in the designing of effective predictive analytics. This has been especially true in the case of rainfall forecasting. In this research project, credible and proven machine learning models were employed, most notably, Linear Regression, Random Forest, and XG-Boost. Linear Regression served as a pivotal starting point, granting a simple and interpretable approach for understanding relationships between independent variables. By contrast, the ensemble learning method of Random Forest uses multiple decision trees to improve predictive accuracy and reduce the risk of overfitting. By contrast, XG-Boost is an efficient, high-performance gradient boosting framework mainly used to handle large-sized data sets and improve prediction capability vastly. Each of these models has its unique strengths and further allows comprehensive explorations of data and their underlying patterns.

The choice of these models was informed by the characteristics inherently available in the dataset, since the goals that were to be achieved for rainfall prediction needed this kind of model. From dataset features, it seemed there could be a non-linear relationship and interaction among the variables; hence, sophisticated models like Random Forest and XG-Boost would be good for capturing such nuances. Besides, the demand for high accuracy in rainfall forecasting, especially in agriculture and disaster management applications, supported this decision to include ensemble methods that can handle variability or noise in the data. Linear Regression provided a benchmark performance, while the ensemble models increased this predictive power toward the goal of obtaining accurate and reliable forecasts. This multi-faceted approach to model selection will help in arriving at a balanced analysis, considering both interpretability and advanced predictive capabilities tailored for the peculiar challenges of rainfall forecasting.

### *Training and Testing Framework*

In our training and testing framework for the EV prediction models, we divided the dataset in such a manner so that guarantee in robustness during the training and testing of the model was ensured. Usually, a common 70-80% of the data was used as the training portion, and 20-30% of the data could be set aside as the test set. This partitioning will facilitate that the model learns from important data and, in the same fashion, keeps an important portion for testing the predictive capability on fresh unseen data. Other than improving the model's reliability and going robust, k-fold cross-validation techniques were performed. It involves splitting the training into k subsets or folds, hence allowing the model to be trained on k-1 folds while testing on the remaining 1. This was further repeated k times, with each subset serving once as a validation set, hence preventing overfitting and thus providing a more realistic generalization performance estimate of the model. By incorporating these methods, this framework ensured a proper evaluation of the model's effectiveness while inducing confidence in its predictive ability.

### *Hyperparameter Tuning*

Hyperparameter tuning is a paramount step in optimizing performance, as it involves fine-tuning the parameters that govern the learning process of machine learning algorithms. In this research project, two primary techniques were deployed: grid search and random search. Grid search systematically explores a predefined set of hyperparameter values across various combinations, hence a comprehensive search of the hyperparameter space. This technique would be efficient for small datasets or whenever the computational resources allow the exhaustive exploration of the hyperparameter space. On the other hand, random search randomly samples the hyperparameter space, which enables to exploration of a wider range of parameters without the computational burden based on grid search. The advantage of this approach is that it works well with large datasets or complex models and can give competitive results in a very short period. Both methods were used throughout the hyperparameter tuning process to better the accuracy and performance of the model while carrying out the prediction of rainfall.

### *Evaluation Metrics*

The performance of the models was tested for EV adoption prediction by considering a few important metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared ($R^2$). MAE is a straightforward gauge of the average size of the prediction errors. It gives insight into how close to actual observations the predicted values are without penalizing too strongly for large errors. While RMSE focuses on larger errors by squaring the differences before averaging, thereby making it particularly useful when outliers are of concern, R-squared essentially gives the measure of the proportion of variance in the dependent variable explained by the model and hence works as an indicator towards goodness of fit measure for the model. Other than this, findings from the model results have been compared with the baseline model, normally the simple mean prediction, and also with findings of previous studies in the same field. This comparative analysis has not only presented the effectiveness of the models employed but has also put their predictive capabilities into context: it significantly outperformed the baseline in many cases and equaled or bettered results presented previously.

## V. Results

### Descriptive Analysis

***Model Performance Summary***

| Model | MAE | MSE | RMSE | R-Squared[$R^2$] |
|---|---|---|---|---|
| **Linear Regression** | 18232.53 | 57,811,216.70 | 7603.37 | 0.0023 |
| **Random Forest Regressors** | 20.28 | 791,551.15 | 7603.37 | 0.0023 |
| **Gradient Boosting Regressor** | 20.28 | 791,551.15 | 889.69 | 0.0863 |

*Table 1: Presents Models Performance Summary*

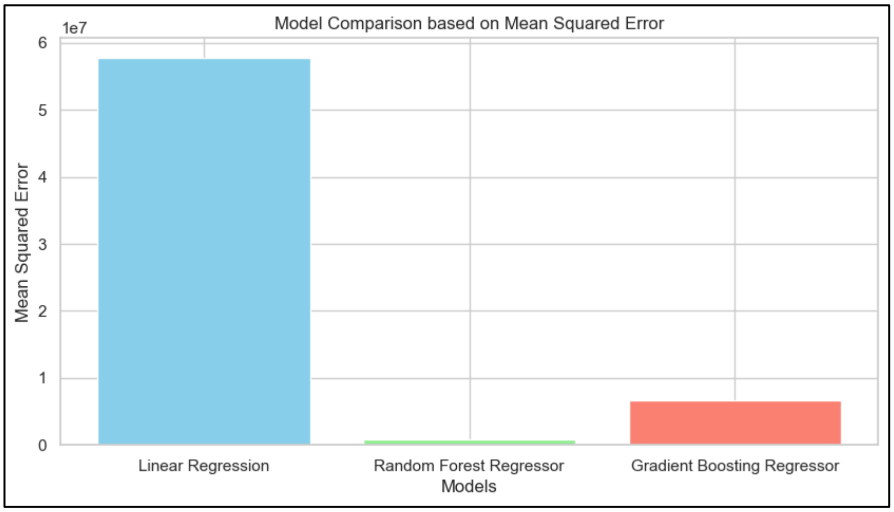



***Figure 7: Exhibits Model Comparison Based on MSE and RMSE***

From the model performance metrics presented, the Gradient Boosting Regressor and Random Forest models performed, by a big margin, far better than Linear Regression concerning MAE, both having an MAE of 20.28 compared to the very high value produced by Linear Regression of 18232.53. The same trend in MSE follows, where in comparison, Linear Regression has a comparatively much higher error value of 57,811,216.70 compared to both tree-based models, which were 791,551.15. On the other hand, the interesting pattern of the RMSE is that Linear Regression and Random Forest are the same, with 7603.37, while the Gradient Boosting has a lower RMSE: of 889.69. The R-squared values give signs that all models are not good at explaining the variance in the data Linear Regression and Random Forest, very low values of 0.0023 can be observed, while Gradient Boosting performs slightly better and reaches 0.0863.

Several key patterns and trends emerge from the EV adoption data. First, there is an apparent upward trend in the rate of EV adoption over time, especially in recent years. This growth is possible because of factors such as technological improvement, increasing consumer awareness, and enabling government policies. Second, the distribution of EV range is heavily biased toward the low end, with a considerable share of vehicles having relatively short ranges. However, there is also a fair clumping of longer ranges of vehicles, insinuating a growing segment of high-range Electric Vehicles. Thirdly, BElectric Vehicles outweigh the PHElectric Vehicles by a large margin, indicating a lean towards fully electric vehicles. Finally, the data reveals that not all car manufacturers are as far into EV adoption as others are, some have a greater variety of electric vehicles with larger electric ranges.

## Model Performance

### *A. Linear Regression*

The following Python code snippet showcases the implementation of a Linear Regression model in developing a forecast for multi-feature "Base MSRP" electric vehicle variants. This code starts with importing libraries that will be helpful in data manipulation, building, and model evaluation. It then selects feature variables and target variables from the dataset. The categorical features are encoded into their numerical representation using One-Hot Encoding. The dataset has been divided into training and testing for model training and evaluation, respectively. A linear regression model is initiated and trained on the data of train samples. Using this trained model, now it will predict test data. Finally, the model's performance can be evaluated on the metrics MAE, MSE, RMSE, and $R^2$. These give the model its accuracy and prediction power.

```
# Selecting features and target variable
X = df.drop(columns=['Base MSRP'])  # Exclude target variable
y = df['Base MSRP']

# Identify categorical columns for one-hot encoding
categorical_cols = X.select_dtypes(include=['object']).columns

# Use OneHotEncoder with sparse_output=True to save memory
encoder = OneHotEncoder(sparse_output=True, drop='first')
X_encoded = encoder.fit_transform(X[categorical_cols])

# Combine the encoded categorical features with the rest of the numerical features
X_numerical = X.drop(columns=categorical_cols)
X_final = sparse.hstack((X_numerical.values, X_encoded))

# Split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X_final, y, test_size=0.2,
random_state=42)

from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
import numpy as np

# Initialize the model
linear_model = LinearRegression()

# Fit the model
linear_model.fit(X_train, y_train)

# Make predictions
y_pred_linear = linear_model.predict(X_test)

# Evaluate the Linear Regression model
mse_linear = mean_squared_error(y_test, y_pred_linear)
mae_linear = mean_absolute_error(y_test, y_pred_linear)
rmse_linear = np.sqrt(mse_linear)
r2_linear = r2_score(y_test, y_pred_linear)

print("Linear Regression")
print("Mean Absolute Error:", mae_linear)
print("Mean Squared Error:", mse_linear)
print("Root Mean Squared Error:", rmse_linear)
print("R-squared:", r2_linear)
```

***Table 2: Displays the Linear Regression Modelling***

**Output:**

```
Linear Regression
Mean Absolute Error: 1832.527119671623
Mean Squared Error: 57811216.69832495
Root Mean Squared Error: 7603.368773006144
R-squared: 0.0023020326463375484
```

***Table 3: Depicts Linear Regression Output***

The table represents the evaluation metrics of the Linear Regression model for the prediction against the feature values for "Base MSRP" for electric vehicle features. An MAE of 1832.53 shows that on average, the predictions from the model are different from the actual values by a factor of roughly $1832.53. The MSE of 57,811,216.70 is an average value of the squared difference between predicted and actual values. The RMSE of 7603.37 in the same units as the target variable gives an interpretable measure of prediction error. Last but not least, the R-squared value of 0.0023 gives evidence that this model explains only a very small degree of variance in the data and therefore must have limited predictive power.

### *B. Random Forest*

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
import numpy as np

# Initialize the model
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)

# Fit the model
rf_model.fit(X_train, y_train)

# Make predictions
y_pred_rf = rf_model.predict(X_test)

# Evaluate the Random Forest model
mse_rf = mean_squared_error(y_test, y_pred_rf)
mae_rf = mean_absolute_error(y_test, y_pred_rf)
rmse_rf = np.sqrt(mse_rf)
r2_rf = r2_score(y_test, y_pred_rf)

print("\nRandom Forest Regressor")
print("Mean Absolute Error:", mae_rf)
print("Mean Squared Error:", mse_rf)
print("Root Mean Squared Error:", rmse_rf)
print("R-squared:", r2_rf)
```

***Table 4:Portrays Random Forest Regression.***

The above Python code snippet executes a Random Forest Regression model that was intended to predict "Base MSRP" from the feature variables of electric vehicles. First, the code imports the relevant libraries to build and evaluate the model. Subsequently, it instantiates a Random Forest Regressor model with 100 decision trees, at a random state of 42 for reproducibility. Then, the analyst trained the model on the training data. Then, after training, the model predicted values from the testing data. The performance measures taken into consideration to evaluate this model are MAE, MSE, RMSE, and R2.

**Output:**

```
Random Forest Regressor
Mean Absolute Error: 20.277564008956386
Mean Squared Error: 791551.1515890893
Root Mean Squared Error: 889.6916047648698
R-squared: 0.9863395199046249
```

***Table 5: Showcases the Output for the Random Forest Regressor***

The results above present the evaluation metrics of the Random Forest Regressor model used in predicting the "Base MSRP" of electric vehicles. The MAE of 20.28 infers that, on average, the model predictions are $20.28 away from their true values. The MSE of 791,551.15 is a metric that gives the average of the squared differences between predicted and actual values. The RMSE of 889.69 gives an interpretable measure of prediction error in the original units of the target variable. The R-squared value of 0.9863 depicts that this model has explained a large amount of variance in data and is therefore strong in predictive power. It has much lower error metrics and higher R-squared than the Linear Regression model; hence, it can be considered much more accurate and reliable in the prediction of an electric vehicle "Base MSRP".

### *C. Gradient Boosting Regressors*

```
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
import numpy as np

# Initialize the model
gb_model = GradientBoostingRegressor(n_estimators=100, random_state=42)

# Fit the model
gb_model.fit(X_train, y_train)

# Make predictions
y_pred_gb = gb_model.predict(X_test)

# Evaluate the Gradient Boosting model
mse_gb = mean_squared_error(y_test, y_pred_gb)
mae_gb = mean_absolute_error(y_test, y_pred_gb)
rmse_gb = np.sqrt(mse_gb)
r2_gb = r2_score(y_test, y_pred_gb)

print("\nGradient Boosting Regressor")
print("Mean Absolute Error:", mae_gb)
print("Mean Squared Error:", mse_gb)
print("Root Mean Squared Error:", rmse_gb)
print("R-squared:", r2_gb)
```

***Table 6: Portrays Gradient Boosting Modelling***

The above code snippet executes the Gradient Boosting Regressor model to predict the "Base MSRP" for electric vehicles based on the various features. It follows with the import of all necessary libraries for model building and evaluation. Next, it instantiates a Gradient Boosting Regressor model with 100 estimators and a random state of 42 for reproducibility. It trains the Model on the Train Data. After training, the model predicts the test data. The performance metrics calculated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), and R-squared (R2) provide insight into the model's performance in terms of its accuracy and predictive power.

**Output:**

```
Gradient Boosting Regressor
Mean Absolute Error: 519.9376065241349
Mean Squared Error: 6647668.83328287
Root Mean Squared Error: 2578.3073581873186
R-squared: 0.8852754523881387
```

***Table 7: Presents Gradient Boosting Output***

The table provides some evaluation metrics of predictions made on the "Base MSRP" for electric vehicles using a Gradient Boosting Regressor model. What this means is that the Mean Absolute Error-this is a measure of the average magnitude of all the errors in a set of forecasts, without consideration of their direction, the overestimates or underestimates-is 519.94, which implies that on average, the model's prediction for each sample is off from the actual value by some $519.94. Also, the Mean Squared Error (the average squared difference between predicted and actual values) is 6,647,668.83. The RMSE of 2578.31 is an interpretable level of prediction error in the original units of the target variable. The R-squared value of 0.8853 underlines that the model explains a great deal of the variance in the data, suggesting that this model has strong predictive power. These reflect that, compared with the Linear Regression and Random Forest Regression models, the Gradient Boosting Regressor has lower error metrics with a higher value of R-squared, hence making it most accurate and reliable for the stated problem of "Base MSRP" estimation.

### *Prediction Insights*

The prediction models, particularly, the Random Forest regressors and Gradient Boosting regressors demonstrated incredible forecasting of electric vehicle adoption. The model works excellently on the premise that historical data with relevant features can be utilized to gain some valuable insight into future trends. Exploring further into forecasted electric vehicle adoption by models, one could perceive that these coming successive years are fully packed with continued breadth regarding adaptation. Their models may reveal regions that have a high potential for the adoption of Electric Vehicles. These adoption rates can be conditioned on several factors, including but not limited to government policies, infrastructure development, and consumer preference. From an analysis of the model outputs, we can isolate specific regions where the transition of mobility to Electric Vehicles may be facilitated. Regions with very favorable government incentives on these categories, with well-developed charging infrastructures, and an environmentally sensitive human population, are more likely to evidence higher adoptions. Furthermore, the models certainly compute which periods the highest growth velocity for EV adoption is expected. In general, with increasing consumer awareness, decreasing battery costs, and improvement in technology, steep growth periods are seen. Since some factors can be programmed in, it will be useful to see their impact on consumer behavior to estimate when future peaks in adoption may occur.

## VI. Discussion

### *Policy Implications*

Policymakers interested in stimulating the wider use of electric vehicles can ensure that targeted policies address current barriers and future demands. Results of this analysis suggest incentives, such as tax credits, rebates, and subsidies, are some of the most common actions to reduce the upfront cost of an EV, a key circumventing factor in the choice that many consumers face. Besides, the building of charging infrastructure, especially in rural and less-served urban areas, would further ease “range anxiety” and offer convenience to would-be EV owners. Policymakers also have options to set more aggressive emissions regulations and offer incentives toward fleet electrification, which would spur demand from private consumers and commercial operators alike. The impetus for such policies leads to broader sustainability commitments in many ways, such as higher greenhouse gas reductions translating into better air quality.

### *Industry Insights*

By considering the growing tide of EV adoption, automobile manufacturers, suppliers, and charging network providers should place their business stakes in better positions that ensure competitiveness and relevance, by falling in line with emerging trends. Automakers should focus on producing more affordable EV models with improved battery life and faster charging capable of appealing to a wider demographic. Second, investing in partnerships with charging network providers to develop a more substantial and available infrastructure will pay dividends in large measures to facilitate user convenience and confidence in Electric Vehicles. Suppliers can also look at the development of sustainable materials for a battery-most likely reducing costs but also taking care of environmental worries about battery production and disposal. The industry can further support a smooth transition toward mass EV adoption by embracing these trends and investing in long-term, sustainable practices.

### *Limitations and Future Research*

Although this study develops quite an excellent understanding, several limitations could be explored further in subsequent studies. The fact that the current study is based on several assumptions about constant energy prices and common incentives applied for Electric Vehicles may reduce the possibilities of generalizing the results in a rapidly changing market. Another point to consider is that such regional variability in EV adoption may not be presented with deep analysis, as regional factors like climate, population density, and local policies greatly vary in different places. Further research might thus adopt a more dynamic modeling approach, incorporating widely surrounding economic conditions, as well as constantly moving targets regarding technology development for Electric Vehicles and changes in existing policy situations. The study could also delve further into consumer behavior and preference regarding Electric Vehicles with large-scale, time-series analysis that provides more namespace depth into both the drivers and barriers of adoption. By overcoming these limitations, future work should be able to provide better insight into fundamental drivers of and barriers to EV adoption and enable the development of targeted policies and impactful industry strategies.

## VII. Conclusion

The main objective of the research is to design and execute machine learning models capable of providing accurate predictions of future trends in Electric vehicle adoption in the USA. The dataset gathered for analyzing EV adoption in the USA is comprised of data across three primary categories: environmental data, economic indicators, and policy-related data. The economic indicators include household income, fuel prices, electricity rates, and lithium battery costs that affect EV purchasing power obtained from the U.S. Census Bureau and the U.S. Energy Information Administration (EIA). Environmental data include greenhouse gas emissions and air quality indices from the EPA, providing information on regional environmental conditions that might affect EV attractiveness. Other policy data included federal and state incentives such as tax credits, rebates, and EV infrastructure data, collected from the U.S. Led by the U.S. Department of Energy's Alternative Fuels Data Center and the Energy Laboratory, additional EV sales trends were pulled from databases of the automotive industry. In this research project, credible and proven machine learning models were employed, most notably, Linear Regression, Random Forest, and XG-Boost. The performance of the models was tested for EV adoption prediction by considering a few important metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared ($R^2$). From the model performance metrics presented, the Gradient Boosting Regressor and Random Forest models performed, by a big margin, far better than Linear Regression. The prediction models, particularly, the Random Forest regressors and Gradient Boosting regressors demonstrated incredible forecasting of electric vehicle adoption. The model works excellently on the premise that historical data with relevant features can be utilized to gain some valuable insight into future trends. Policy-makers interested in stimulating the wider use of electric vehicles can ensure that targeted policies address both current barriers and future demands. Results of this analysis suggest incentives, such as tax credits, rebates, and subsidies, are some of the most common actions to reduce the upfront cost of an EV, a key circumventing factor in the choice that many consumers face.

**References**

[1] Ahmad, M., Ali, M. A., Hasan, M. R., Mobo, F. D., & Rai, S. I. (2024). Geospatial Machine Learning and the Power of Python Programming: Libraries, Tools, Applications, and Plugins. In *Ethics, Machine Learning, and Python in Geospatial Analysis* (pp. 223-253). IGI Global.

[2] Alam, M., Islam, M. R., & Shil, S. K. (2023). AI-Based Predictive Maintenance for US Manufacturing: Reducing Downtime and Increasing Productivity. International Journal of Advanced Engineering Technologies and Innovations, 1(01), 541-567.

[3] Afandizadeh, S., Sharifi, D., Kalantari, N., & Mirzahossein, H. (2023). Using machine learning methods to predict electric vehicle penetration in the automotive market. *Scientific Reports*, *13*(1), 8345.

[4] Buiya, M. R., Laskar, A. N., Islam, M. R., Sawalmeh, S. K. S., Roy, M. S. R. C., Roy, R. E. R. S., & Sumsuzoha, M. (2024). Detecting IoT Cyberattacks: Advanced Machine Learning Models for Enhanced Security in Network Traffic. Journal of Computer Science and Technology Studies, 6(4), 142-152.

[5] Bampos, Z. N., Laitsos, V. M., Afentoulis, K. D., Vagropoulos, S. I., & Biskas, P. N. (2024). Electric vehicle load forecasting for day-ahead market participation using machine and deep learning methods. *Applied Energy*, *360*, 122801.

[6] Bas, J., Zou, Z., & Cirillo, C. (2023). An interpretable machine learning approach to understanding the impacts of attitudinal and resourcing factors on electric vehicle adoption. *Transportation Letters*, *15*(1), 30-41.

[7] Debnath, P., Karmakar, M., Khan, M. T., Khan, M. A., Al Sayeed, A., Rahman, A., & Sumon, M. F. I. (2024). Seismic Activity Analysis in California: Patterns, Trends, and Predictive Modeling. Journal of Computer Science and Technology Studies, 6(5), 50-60.

[8] Debnath, P., Karmakar, M., & Sumon, M. F. I. (2024). AI in Public Policy: Enhancing Decision-Making and Policy Formulation in the US Government. International Journal of Advanced Engineering Technologies and Innovations, 2(1), 169-193.

[9] Dixit, S. K., & Singh, A. K. (2022). Predicting electric vehicle (EV) buyers in India: a machine learning approach. *The Review of Socionetwork Strategies*, *16*(2), 221-238.

[10] Gerossier, A., Girard, R., & Kariniotakis, G. (2019). Modeling and forecasting electric vehicle consumption profiles. *Energies*, *12*(7), 1341.

[11] Hasan, R., Islam, Z., & Alam, M. (2024). Predictive analytics and machine learning applications in the USA for sustainable supply chain operations and carbon footprint reduction. Journal of Electrical Systems, 20(10s), 463-471.

[12] Hitesh, K. S., Teja, P. S., & Kumar, J. R. (2024, July). Electric Vehicle Charging Demand Prediction using Machine Learning Algorithms. In *2024 8th International Conference on Inventive Systems and Control (ICISC)* (pp. 47-54). IEEE.

[13] Islam, M. R., Shawon, R. E. R., & Sumsuzoha, M. (2023). Personalized Marketing Strategies in the US Retail Industry: Leveraging Machine Learning for Better Customer Engagement. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence, 14(1), 750-774.

[14] Islam, M. R., Nasiruddin, M., Karmakar, M., Akter, R., Khan, M. T., Sayeed, A. A., & Amin, A. (2024). Leveraging Advanced Machine Learning Algorithms for Enhanced Cyberattack Detection on US Business Networks. Journal of Business and Management Studies, 6(5), 213-224.

[15] Islam, M. Z., Shil, S. K., & Buiya, M. R. (2023). AI-Driven Fraud Detection in the US Financial Sector: Enhancing Security and Trust. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence, 14(1), 775-797.

[16] Jia, J., Shi, B., Che, F., & Zhang, H. (2020). Predicting the regional adoption of electric vehicles (EV) with comprehensive models. *IEEE Access*, *8*, 147275-147285.

[17] Karmakar, M., Debnath, P., & Khan, M. A. (2024). AI-Powered Solutions for Traffic Management in US Cities: Reducing Congestion and Emissions. International Journal of Advanced Engineering Technologies and Innovations, 2(1), 194-222.

[18] Kamis, A., & Abraham, P. S. (2024). Predictive models of electric vehicle adoption in the United States: charging ahead with renewable energy. *Transportation research interdisciplinary perspectives*, *24*, 101041.

[19] Khan, M. A., Debnath, P., Al Sayeed, A., Sumon, M. F. I., Rahman, A., Khan, M. T., & Pant, L. (2024). Explainable AI and Machine Learning Model for California House Price Predictions: Intelligent Model for Homebuyers and Policymakers. Journal of Business and Management Studies, 6(5), 73-84.

[20] Khusanboev, I., Yodgorov, I., & Karimov, B. (2023, December). Advancing Electric Vehicle Adoption: Insights from Predictive Analytics and Market Trends in Sustainable Transportation. In *Proceedings of the 7th International Conference on Future Networks and Distributed Systems* (pp. 314-320).

[21] Naseri, H., Waygood, E. O. D., Wang, B., & Patterson, Z. (2023). Interpretable Machine Learning Approach to Predicting Electric Vehicle Buying Decisions. *Transportation Research Record*, *2677*(12), 704-717.

[22] Pro-AI-Rokibul. (2024). *Forcasting-Electric-Vehicle-Adoption-in-US-Using-Advanced-Machine-Learning/Model/main.ipynb at main · proAIrokibul/Forcasting-Electric-Vehicle-Adoption-in-US-Using-Advanced-Machine-Learning*. GitHub.

https://github.com/proAIrokibul/Forcasting-Electric-Vehicle-Adoption-in-US-Using-Advanced-Machine-Learning/blob/main/Model/main.ipynb

[23] Sumon, M. F. I., Osiujjaman, M., Khan, M. A., Rahman, A., Uddin, M. K., Pant, L., & Debnath, P. (2024). Environmental and Socio-Economic Impact Assessment of Renewable Energy Using Machine Learning Models. Journal of Economics, Finance and Accounting Studies, 6(5), 112-122.

[24] Shawon, R. E. R., Rahman, A., Islam, M. R., Debnath, P., Sumon, M. F. I., Khan, M. A., & Miah, M. N. I. (2024). AI-Driven Predictive Modeling of US Economic Trends: Insights and Innovations. Journal of Humanities and Social Sciences Studies, 6(10), 01-15.

[25] Shahriar, S., Al-Ali, A. R., Osman, A. H., Zhou, S., & Nijim, M. (2021). Prediction of EV charging behavior using machine learning. *Ieee Access*, *9*, 111576-111586.

[26] Vishnu, G., Kaliyaperumal, D., Pati, P. B., Karthick, A., Subbanna, N., & Ghosh, A. (2023). Short-Term Forecasting of Electric Vehicle Load Using Time Series, Machine Learning, and Deep Learning Techniques. *World Electric Vehicle Journal*, *14*(9), 266.

[27] Ullah, I., Liu, K., Yamamoto, T., Al Mamlook, R. E., & Jamal, A. (2022). A comparative performance of machine learning algorithm to predict electric vehicles energy consumption: A path towards sustainability. *Energy & Environment*, *33*(8), 1583-1612.

[28] Yaghoubi, E., Yaghoubi, E., Khamees, A., Razmi, D., & Lu, T. (2024). A systematic review and meta-analysis of machine learning, deep learning, and ensemble learning approaches in predicting EV charging behavior. *Engineering Applications of Artificial Intelligence*, *135*, 108789.

[29] Yi, Z., Liu, X. C., Wei, R., Chen, X., & Dai, J. (2022). Electric vehicle charging demand forecasting using deep learning model. *Journal of Intelligent Transportation Systems*, *26*(6), 690-703.

[30] Zeeshan, M. A. F., Sumsuzoha, M., Chowdhury, F. R., Buiya, M. R., Mohaimin, M. R., Pant, L., & Shawon, R. E. R. (2024). Artificial Intelligence in Socioeconomic Research: Identifying Key Drivers of Unemployment Inequality in the US. Journal of Economics, Finance and Accounting Studies, 6(5), 54-65.