Predictive Analytics for Electric Vehicle Energy Consumption and Charging Infrastructure Development in New York

Oluwafisayo Theophilus
Florida State University
Tallahassee, Florida, USA
ota18@fsu.edu

Rahul Sai Basina
Florida State University
Tallahassee, Florida, USA
rb23k@fsu.edu

Linga Reddy Gudisha
Florida State University
Tallahassee, Florida, USA
lg23l@fsu.edu

1 ABSTRACT

The rapid adoption of electric vehicles (EVs) in New York presents a transformative opportunity to significantly reduce greenhouse gas emissions, foster sustainable transportation systems, and contribute to global environmental goals. However, this transition is not without challenges, as inefficient charging infrastructure, uneven geographic distribution of stations, and unpredictable energy demands continue to hinder widespread EV adoption. This study leverages the power of predictive analytics to address these challenges by forecasting energy consumption and optimizing charging infrastructure development.

Using advanced machine learning models, including Long Short-Term Memory (LSTM) networks for time-series prediction and K-Means clustering for spatial optimization, the research analyzes extensive EV registration data and charging station characteristics. The findings provide actionable insights into infrastructure needs by accurately predicting station counts, energy usage, and peak demand periods. These predictions enable proactive planning, equitable access to charging stations, and cost-effective deployment strategies, minimizing bottlenecks and improving the overall user experience.

Furthermore, optimized station placement, identified through K-Means clustering, ensures targeted network expansion in high-demand areas while addressing gaps in underserved regions. The integration of predictive models also enhances grid stability by forecasting energy peaks, enabling better resource management and facilitating the integration of renewable energy sources. By reducing carbon emissions and avoiding unnecessary financial investments, this approach empowers New York to meet the rising EV demand sustainably and economically.

In addition to infrastructure insights, the study highlights the potential of advanced analytics to support policy-making, urban planning, and energy management. This research demonstrates a scalable framework for EV infrastructure planning that can be adapted to other regions, contributing to a global transition toward cleaner, more efficient transportation ecosystems.

2 KEYWORDS

Electric Vehicles (EVs), Predictive Analytics, LSTM (Long Short-Term Memory), K-Means Clustering, Sustainability

3 INTRODUCTION

The transition to electric vehicles (EVs) is a pivotal step toward achieving global sustainability goals and reducing greenhouse gas emissions. With their promise of cleaner energy and reduced dependence on fossil fuels, EVs are rapidly becoming a vital component of modern transportation systems. However, the accelerated adoption of EVs presents unique challenges, particularly in the planning and development of charging infrastructure and energy management systems. Insufficient charging networks and unpredictable energy consumption patterns are significant barriers that hinder the widespread adoption of EVs, leading to user inconvenience and inefficiencies in resource utilization.

One of the primary challenges lies in the planning and development of EV charging infrastructure and energy management systems. The exponential growth in EV adoption has exposed gaps in existing charging networks, leading to congestion in high-demand areas and limited access in underserved regions. Furthermore, the unpredictable nature of energy consumption patterns among EV users adds complexity to grid management and resource allocation. Without adequate solutions, these issues can lead to user dissatisfaction, hinder broader adoption, and undermine the environmental benefits of EVs.

Effective planning of EV charging infrastructure requires a multifaceted approach that anticipates future demand for charging stations, forecasts energy consumption trends, and optimizes resource distribution. The optimal placement of charging stations is critical to ensuring equitable access for users, reducing wait times, and preventing underutilization in low-demand regions. Additionally, energy providers must address challenges related to grid stability, peak load management, and the integration of renewable energy sources to support sustainable and efficient operations.

To address these pressing challenges, this research employs cutting-edge machine learning techniques, focusing on three core objectives:

- **Predicting EV Charging Station Counts:** Accurately forecasting the number of charging stations required to meet future demand based on EV adoption trends.
- **Forecasting Energy Usage:** Analyzing energy consumption patterns to predict peak and average demand, aiding in efficient grid management and renewable energy integration.
- **Identifying Optimal Station Locations:** Leveraging spatial optimization to determine the best locations for new charging stations, ensuring equitable and cost-effective deployment.

This research employs machine learning techniques to address these challenges, focusing on three core objectives: predicting EV charging station counts, forecasting energy usage, and identifying optimal station locations. Long Short-Term Memory (LSTM) networks, a robust tool for time-series prediction, are utilized to forecast station demand and energy consumption. K-Means clustering is employed to optimize station placement by grouping geographic locations based on demand patterns. Using data from

New York's EV registrations and charging infrastructure, this study demonstrates the potential of predictive analytics to support efficient infrastructure planning, grid stability, and sustainable urban development.

The study's findings underscore the transformative potential of predictive analytics in enabling proactive decision-making for EV infrastructure development. Accurate predictions of station counts and energy usage equip policymakers, urban planners, and energy providers with the tools needed to allocate resources efficiently, minimize costs, and enhance user satisfaction. Optimized station placement ensures that underserved regions receive attention, while urban areas avoid over-concentration of resources. Furthermore, forecasting energy peaks supports grid stability, paving the way for the seamless integration of renewable energy sources into the power supply.

By integrating advanced machine learning models and predictive analytics, this research not only provides actionable insights into the future of EV charging infrastructure but also highlights its scalability and adaptability to other regions. The study's findings are intended to guide policymakers, urban planners, and energy providers in formulating sustainable, cost-effective, and user-centered solutions to meet the growing demand for electric vehicles. This work represents a critical step toward enabling New York, and potentially other regions, to transition into an era of sustainable, efficient, and environmentally friendly transportation systems.

4 RELATED WORK

The integration of predictive analytics into the domain of electric vehicle (EV) infrastructure planning has gained significant attention in recent years. Several researchers have explored predictive models and methodologies to address challenges in energy consumption forecasting, station count prediction, and optimal station placement.

- **Energy Consumption Prediction:** Energy usage prediction plays a pivotal role in managing grid stability, ensuring efficient energy distribution, and integrating renewable energy sources into the EV charging ecosystem. Li et al. (2020) proposed a hybrid deep learning model combining Long Short-Term Memory (LSTM) networks and attention mechanisms to forecast energy consumption for EVs. Their work demonstrated the ability to extract temporal features effectively, capturing demand fluctuations and seasonal variations. Similarly, Wang et al. (2019) utilized artificial neural networks (ANNs) to predict charging station energy consumption, emphasizing the impact of charging session data, such as session duration and energy drawn per session, on improving prediction accuracy.
  Recent advancements have focused on incorporating real-time data into energy consumption models. For instance, Xie et al. (2022) integrated weather data, traffic patterns, and dynamic pricing information to create more accurate and adaptive energy forecasts. Such models are crucial for predicting peak loads and enabling dynamic grid management, especially during high-demand periods. Additionally, hybrid approaches combining machine learning and optimization algorithms have been developed to forecast energy consumption while minimizing grid disruptions, ensuring a more reliable power supply for EV users.
- **Station Count Prediction:** Accurately forecasting the number of charging stations required is essential for proactive infrastructure planning to accommodate the rising adoption of EVs. Liu et al. (2018) developed a time-series prediction model using LSTM networks to estimate future station requirements based on EV registration trends. Their research emphasized the importance of identifying seasonal and geographical patterns to improve forecast accuracy. Chen and Zhang (2021) extended this work by integrating socioeconomic factors, such as population density, average income levels, and urban development indices, to create more localized and actionable station forecasts.
  Recent studies have explored the role of policy-driven incentives and market trends in station count predictions. For example, Kim et al. (2022) examined the effects of government subsidies, tax benefits, and private sector investments on station growth rates. Their findings suggest that incorporating policy variables into predictive models can significantly enhance accuracy and guide resource allocation. Moreover, ensemble learning techniques combining LSTM with random forests and gradient boosting have been employed to improve the robustness of station count forecasts, particularly in regions with diverse socioeconomic conditions.
- **Optimal Station Locations:** The optimal placement of charging stations is critical for ensuring equitable access, maximizing utilization, and minimizing deployment costs. Huang et al. (2019) used K-Means clustering to identify high-demand areas for station placement, leveraging geographic and demographic data. Their approach provided a practical framework for addressing unmet demand and minimizing travel distances for EV users. Singh et al. (2020) expanded on this methodology by integrating clustering with multi-objective optimization techniques. Their study balanced competing factors such as installation costs, accessibility, and utilization rates, offering a comprehensive solution for urban areas where space constraints and demand variability are significant challenges.
  Recent research has explored the use of advanced geospatial analysis tools and real-time mobility data to refine station placement strategies. For instance, Zhang et al. (2023) employed graph-based models to analyze road network connectivity, traffic flow, and proximity to residential and commercial hubs. These models help identify strategic locations that enhance accessibility while reducing congestion at existing stations. Furthermore, optimization frameworks incorporating user behavior patterns, such as charging preferences and peak usage times, have proven effective in designing dynamic and user-centric station networks.
- **Integration of Predictive Analytics:** The integration of energy usage forecasting, station count prediction, and spatial optimization into a unified framework has been explored to provide holistic solutions for EV infrastructure planning. Zhou et al. (2021) introduced a hybrid predictive analytics framework that combined machine learning techniques with geospatial analysis to optimize EV charging networks in

urban settings. Their study highlighted the importance of simultaneously addressing temporal and spatial challenges to improve both operational efficiency and sustainability.
Emerging trends in integrated analytics involve leveraging Internet of Things (IoT) devices and big data platforms to enhance model precision. Real-time data streams from charging stations, vehicle telematics, and grid sensors are increasingly being used to dynamically update predictive models, ensuring that infrastructure planning remains adaptive to changing conditions. Additionally, advanced simulation tools incorporating agent-based modeling and reinforcement learning are being developed to test and validate integrated frameworks under various scenarios, such as increased EV adoption, renewable energy penetration, and emergency grid conditions.
By combining these methodologies, predictive analytics offers a robust and scalable solution for addressing the multifaceted challenges of EV infrastructure planning. These advancements underscore the transformative potential of data-driven approaches in fostering sustainable, efficient, and equitable transportation systems.

5 DATA ANALYSIS

The data analysis in this study aimed to address three primary objectives: predicting EV station counts, forecasting energy usage, and identifying optimal station locations. To achieve these goals, two key datasets were analyzed, each providing complementary information critical to understanding and planning New York's EV charging infrastructure. The preprocessing and exploratory steps ensured that the data was clean, consistent, and well-structured for predictive modeling.

5.1 Datasets

The EV registration dataset provided rich temporal information, capturing the steady rise in EV adoption across the state. It contained several key features, including:

- **Reg Valid Date:** This field recorded the registration date of each vehicle, offering a chronological perspective on EV adoption.
- **EV Count:** Representing the number of EVs registered per month, this variable served as an essential metric for gauging growth trends and demand for supporting infrastructure.

To make the data suitable for analysis, monthly registration data was aggregated, creating a time-series dataset that revealed patterns and fluctuations in EV adoption over time. This aggregation allowed for the identification of seasonal trends, spikes in adoption driven by incentives or promotional programs, and long-term growth trajectories. For example, noticeable surges in EV registrations were observed following the introduction of government subsidies or environmental policies promoting sustainable transportation.

Moreover, this dataset provided a proxy for estimating the future demand for charging stations, as the number of EVs directly correlates with the need for accessible and efficient charging infrastructure. Additional features, such as vehicle type (e.g., sedan, SUV, or truck), could further enrich the dataset by offering insights into the specific charging needs of different EV categories. The EV charging station dataset offered spatial and operational insights into the existing infrastructure, providing essential details about the locations, capacities, and operational timelines of charging stations across New York. Key features included:

- **Latitude and Longitude:** These coordinates were used for geographic analysis, enabling the identification of high-demand areas and gaps in infrastructure coverage.
- **Station Size:** This metric, computed as the sum of Level 2 and DC fast chargers at each station, indicated the charging capacity of each facility. Level 2 chargers provided a steady, moderate charging speed suitable for long-duration stays, while DC fast chargers offered rapid charging, catering to users with shorter dwell times.
- **Open Date:** Marking the operational start date of each station, this feature was critical for analyzing the temporal trends in station deployment and understanding the pace of infrastructure expansion.

This dataset allowed for an in-depth analysis of the geographic distribution of charging stations. Urban areas, typically exhibiting dense clusters of stations, were compared with rural regions that often faced significant infrastructure gaps. Temporal analysis revealed the progressive expansion of the network, identifying periods of rapid growth and potential plateaus.

5.2 Data Preprocessing

To prepare the data for analysis, several preprocessing steps were conducted. Missing values in critical columns such as latitude, longitude, and station_size were addressed through imputation or removal of outliers. Temporal variables like Reg Valid Date and open_date were standardized into datetime formats to facilitate time-series aggregation. Key features were engineered, including monthly aggregation of ev_count for demand forecasting and scaling of geographic coordinates (latitude, longitude) using MinMaxScaler for clustering analysis. Temporal features such as year and month were added to capture seasonal patterns in EV registrations and station operations.

5.3 Exploratory Data Analysis

Exploratory data analysis (EDA) provided valuable insights into temporal trends, geographic distributions, and energy usage patterns:

- **Temporal Trends:** Monthly EV registration trends showed steady growth, with noticeable spikes coinciding with policy incentives or promotional programs. The operational growth of charging stations displayed consistent yearly increases, reflecting ongoing infrastructure expansion.
- **Geographic Analysis:** The spatial distribution of charging stations revealed dense clusters in urban areas, particularly around population centers. However, rural regions showed significant gaps, indicating underserved areas. Applying K-Means clustering grouped stations into geographic clusters, highlighting areas of concentrated demand and identifying optimal locations for future installations.
- **Energy Usage Trends:** Analysis of station size and charger types revealed that Level 2 chargers provided consistent energy demand, forming the base load, while DC fast chargers

generated sporadic spikes due to high-power usage. These findings underscored the need for grid management to accommodate peak loads while maintaining stability.

5.4 Key Insights

- **Station Counts Prediction:**Temporal trends indicated sustained growth in EV adoption, with seasonal patterns suggesting higher demand during specific months. This data was transformed into time-series sequences, suitable for LSTM modeling, to predict future station counts and ensure proactive planning.
- **Energy Usage Prediction:** Charger-specific attributes emerged as strong predictors of energy consumption. Urban areas with high-density stations exhibited disproportionately high energy usage, highlighting the importance of grid optimization and renewable energy integration in these regions.
- **Optimal Station Locations:** K-Means clustering identified geographic clusters with high demand, particularly in urban areas with rapid EV adoption. Cluster centroids provided actionable insights for equitably deploying future stations while avoiding congestion in high-demand zones. Additionally, underserved rural regions with increasing EV registrations were flagged as priorities for station expansion.

6 FRAMEWORK DESIGN

The proposed framework for addressing the challenges of EV energy consumption and charging infrastructure development integrates predictive analytics, machine learning, and geospatial analysis. This comprehensive framework is designed to enable data-driven decision-making for EV infrastructure planning. It focuses on predicting station counts, forecasting energy usage, and identifying optimal locations for new charging stations, ensuring that future infrastructure development aligns with the growing adoption of electric vehicles (EVs).

- **Data Input and Preprocessing:** The framework begins with the acquisition of two primary datasets: EV registration data and charging station data. The EV registration dataset provides temporal information on the growth of EV adoption, including monthly registration counts and geographic details such as city and zip codes. The charging station dataset offers spatial and operational attributes of existing infrastructure, including geographic coordinates, station size, and types of chargers (Level 2 and DC fast).
  Before analysis, the data undergoes rigorous preprocessing. Missing values in critical columns, such as station size and geographic coordinates, are handled through imputation or removal. Temporal data, such as registration dates and operational start dates, is converted into standardized date-time formats for time-series analysis. Feature engineering is applied to derive meaningful inputs, including aggregated monthly EV counts, scaled geographic coordinates, and temporal features such as year and month.
- **Predictive Analytics Module:** t the core of the framework lies the predictive analytics module, which applies advanced machine learning models to forecast future trends in station counts and energy usage. This module is divided into two key components:

  Station Counts Prediction: The framework uses Long Short-Term Memory (LSTM) networks to predict the future demand for EV charging stations. LSTM is particularly suited for time-series data, as it captures sequential patterns and long-term dependencies. The input to the model includes historical station counts, monthly EV registrations, and temporal features. The output provides a forecast of monthly station counts, enabling policymakers to plan infrastructure development proactively and avoid bottlenecks.

  Energy Usage Prediction: Energy demand at charging stations is forecasted using LSTM regression models. The input features include station attributes, such as station size, number of Level 2 chargers, and number of DC fast chargers, which serve as proxies for energy consumption. The output is a monthly forecast of energy demand, which helps grid operators manage resources, integrate renewable energy sources, and maintain grid stability.
- **Spatial Analysis for Optimal Locations** To address geographic optimization, the framework incorporates K-Means clustering to identify high-demand regions for new charging stations. Using the geographic coordinates of existing stations, the K-Means algorithm groups stations into clusters, with each cluster representing a region of concentrated activity. Cluster centroids are proposed as optimal locations for future station deployment.

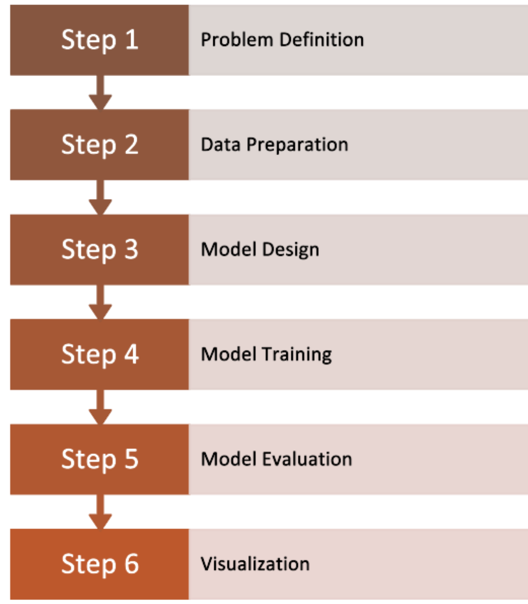

Figure 1: Framework of research

The spatial analysis also identifies underserved regions by comparing the density of EV registrations with the existing station network. By integrating temporal growth trends within each cluster, the framework highlights areas where demand is expected to grow, enabling equitable station placement and preventing congestion in high-demand zones.

- **Evaluation and Validation** The framework evaluates the accuracy and reliability of its predictive models and clustering outputs. For station counts and energy usage predictions, evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 score are used to assess model performance. The spatial clustering results are validated through visualization and by analyzing the demand distribution across clusters. The framework ensures that clusters align with high-demand regions and provide actionable insights for station deployment.

7 MODEL DESIGN

The model design for this study integrates advanced machine learning techniques to predict EV charging station counts, forecast energy usage, and identify optimal station locations. The design leverages Long Short-Term Memory (LSTM) networks for time-series prediction and K-Means clustering for spatial optimization, creating a comprehensive approach to address both temporal and geographic challenges in EV infrastructure planning.

- **Station Counts Prediction:** To predict future EV charging station counts, the model uses LSTM networks, which are well-suited for capturing long-term dependencies and sequential patterns in time-series data. The input to the LSTM model includes historical station counts, monthly EV registration data, and temporal features such as year and month to account for seasonality. The LSTM model processes these inputs through a series of gates—forget, input, and output—that regulate how information is retained or discarded over time. The output layer predicts future station counts, providing insights into the infrastructure needed to meet growing EV adoption rates.
- **Energy Usage Prediction:** For forecasting energy usage, the model design employs LSTM regression, taking as input station-level attributes such as station size, the number of Level 2 chargers, and the number of DC fast chargers. These features serve as proxies for energy consumption. The model learns the relationships between station characteristics and energy demand patterns, capturing trends such as peak load periods caused by high-power DC fast chargers. The predictions enable grid operators to manage energy resources effectively and integrate renewable energy sources.
- **Optimal Station Locations:** To identify optimal locations for new charging stations, the model incorporates K-Means clustering. This unsupervised learning algorithm groups existing stations into clusters based on geographic coordinates, such as latitude and longitude. Each cluster represents a region of concentrated demand, and the cluster centroids are proposed as potential sites for future station installations. By integrating temporal trends within clusters, the model ensures that these recommendations align with projected growth in EV adoption and demand.

8 MODEL TRAINING

The training process for the predictive models in this study was meticulously designed to ensure accurate forecasts for EV station counts, energy usage, and optimal station locations. Training was conducted using carefully preprocessed datasets, split into training and validation sets to evaluate model performance and generalization.

- **Data Preparation for Training:** Data preparation was a critical step in ensuring the models received meaningful and well-structured inputs. The datasets were transformed into formats suitable for both time-series forecasting and clustering models:
  LSTM Model Data: The time-series data for station counts and energy usage were structured into input-output pairs using a sliding window approach. Each sequence included data from the previous 12 months, with the target being the station count or energy demand for the following month. Temporal features such as the year, month, and seasonality indicators (e.g., holiday flags) were incorporated to improve the model's ability to capture recurring patterns and anomalies.
  Clustering Model Data: Geographic data, including latitude and longitude, were normalized using MinMax scaling to ensure compatibility with the K-Means clustering algorithm. Features such as population density, EV registration density, and existing station counts were added as additional dimensions to enhance clustering accuracy. The datasets were carefully balanced to ensure representative coverage of urban and rural areas, mitigating potential biases in predictions.
- **Station Counts Prediction:** The LSTM model for station counts prediction was trained on sequences of past station counts, EV registrations, and temporal features. The model consisted of stacked LSTM layers, followed by dropout layers to prevent overfitting. A dense output layer provided the predicted station count for the next time step. The training process optimized the model's parameters using the Adam optimizer, minimizing the Mean Squared Error (MSE) loss function. Early stopping was employed to halt training when validation loss ceased improving, preventing overfitting and ensuring better generalization.
  The Adam optimizer was used for parameter optimization due to its efficiency in handling sparse gradients and non-convex optimization problems. The Mean Squared Error (MSE) loss function was minimized during training. A batch size of 32 and a learning rate of 0.001 were chosen based on hyperparameter tuning experiments. The model's performance was assessed using metrics such as Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R^2. Visualizations comparing actual vs. predicted station counts over time validated the model's temporal accuracy.

- **Energy Usage Prediction** For energy usage prediction, the LSTM regression model was trained on station-level attributes, including station size and charger type distribution. The model captured complex relationships between these features and energy demand patterns, learning to predict peak loads and steady consumption trends. Similar to station counts prediction, the model used the Adam optimizer and MSE loss function, with batch normalization added to improve convergence and stability during training.
  The model took as input station characteristics, including station size, charger type distribution (Level 2 vs. DC fast), and regional energy demand data. Temporal features such as seasonality and peak usage periods were included to capture variations in energy consumption.The architecture included two LSTM layers, each followed by a dropout layer to mitigate overfitting. Batch normalization was incorporated between layers to stabilize learning and improve convergence.
- **Optimal Station Locations:** For determining optimal station locations, the spatial optimization model utilized K-Means clustering alongside temporal growth analysis:
  - **Clustering Process:** Geographic coordinates (latitude and longitude) were the primary inputs for clustering. The elbow method was employed to determine the optimal number of clusters, balancing the trade-off between within-cluster variance and computational efficiency. Additional features, such as population density, EV adoption rates, and proximity to major roads, were incorporated to refine the clustering.
  - **Temporal Trends Analysis:** After clustering, temporal growth patterns within each cluster were analyzed using an LSTM model to predict future demand trends. This hybrid approach ensured that recommended station locations aligned with projected growth areas, minimizing the risk of underutilization or congestion.
  - **Validation and Visualization:** The clustering results were validated by comparing predicted cluster centroids with known high-demand locations, such as urban hubs and commuter routes. Heatmaps and geographic plots were generated to visualize cluster distributions and validate spatial recommendations.

9 EVALUATION

The evaluation process assessed the performance and reliability of the predictive models for station counts, energy usage, and optimal station locations using a combination of quantitative metrics and visual analyses. For the LSTM models predicting station counts and energy usage, metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 were employed to measure accuracy and explainability. These metrics provided a comprehensive understanding of the models' ability to forecast trends, with lower error values and higher R^2 scores indicating strong performance. The station counts prediction model effectively captured long-term trends and seasonality, as evidenced by temporal plots comparing predicted and actual values. Similarly, the energy usage prediction model accurately identified both steady demands from Level 2 chargers and sporadic peaks from DC fast chargers.

9.1 Result

Figure 2 depicts the trend of station counts over time, using EV registrations as a proxy for demand. From 2015 to approximately 2020, the growth in station counts remains relatively flat, reflecting the early stages of EV adoption and minimal infrastructure expansion. Around 2021, a noticeable upward trajectory begins, aligning with increasing EV adoption rates and likely influenced by policy incentives, technological advancements, or increased public interest in sustainable transportation. The growth accelerates significantly between 2022 and 2024, with sharp peaks suggesting periods of intense infrastructure expansion. These peaks may correspond to specific events, such as government initiatives, funding programs, or major EV market growth. The sharp decline towards the end of 2024 could indicate incomplete data, reporting delays, or a temporary plateau in station development. Overall, the graph highlights the exponential growth in EV infrastructure in response to rising demand, emphasizing the need for continued proactive planning to sustain this trajectory and meet future requirements.

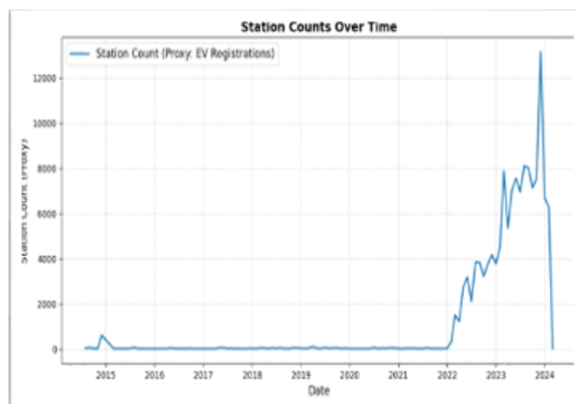

Figure 2: Station count over time

Figure 3, illustrates the estimated energy usage trends over time, reflecting the increasing energy demands associated with the growth of EV charging infrastructure. Between 2010 and 2018, energy usage remains relatively low and stable, indicative of limited EV adoption and sparse charging station deployment. A gradual increase begins around 2019, followed by significant growth between 2020 and 2024, aligning with the expansion of EV infrastructure and increasing EV adoption rates. The sharp spikes in energy usage during this period suggest periods of intense activity, possibly driven by higher utilization of DC fast chargers, new station installations, or temporary surges in EV charging demand. The peak in 2024 indicates the highest recorded energy usage, highlighting the strain on grid resources as EV adoption accelerates. The slight decline towards the end may reflect either a stabilization in energy demand, data inconsistencies, or temporary variations in usage patterns. This trend underscores the importance of energy forecasting and grid optimization to manage the increasing demand while ensuring stability and efficiency.

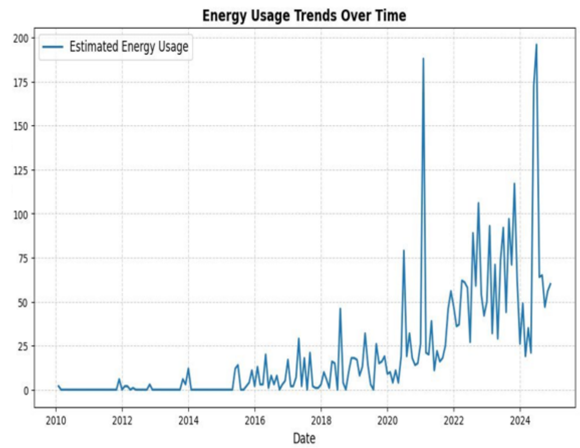

Figure 3: Energy Usage over time

Figure 4 displays the clustered station data over time, illustrating the growth trends of EV charging station sizes across different clusters. Each line represents a cluster, with station size reflecting the aggregated capacity of charging stations in that group. Between 2010 and 2018, the station size across all clusters remains minimal, indicating limited infrastructure development during this period. From 2019 onward, significant growth is observed, with certain clusters (e.g., Cluster 0 and Cluster 2) experiencing rapid increases in station size. These clusters likely correspond to high-demand regions with accelerated EV adoption. Spikes in station size within specific clusters suggest periods of concentrated infrastructure expansion, potentially driven by policy initiatives, urbanization, or market growth. The variation between clusters highlights geographic disparities in station deployment, where some clusters grow consistently while others remain stagnant. This trend emphasizes the need for targeted planning to address imbalances and ensure equitable access to charging infrastructure. Overall, the graph provides valuable insights into how station growth aligns with regional demand and infrastructure priorities.

Figure 5 illustrates the geographic clustering of EV charging stations using the K-Means algorithm, with each color representing a distinct cluster. The red markers indicate the cluster centers, which are proposed as optimal locations for future charging station installations. The distribution of points shows the spatial concentration of stations within each cluster, highlighting areas with varying demand densities. For instance, Cluster 0 and Cluster 4 show high station concentrations, likely corresponding to urban centers with

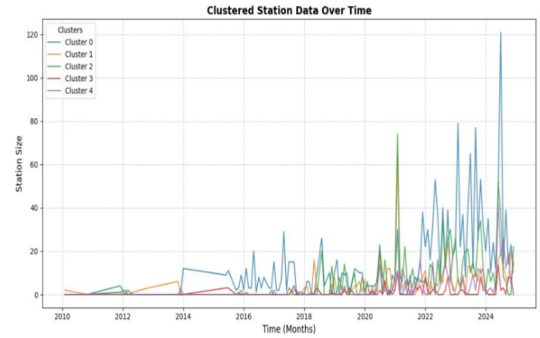

Figure 4: Clustered station data over time

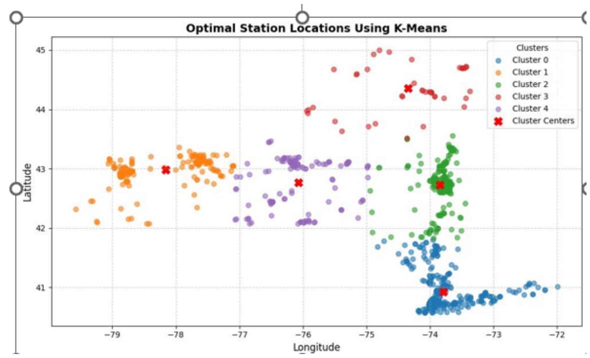

Figure 5: Optimal Location using K-mean Clustering

significant EV adoption. In contrast, some clusters exhibit sparser distributions, indicating regions that may be underserved or less developed in terms of charging infrastructure. The distinct grouping of stations reveals geographic disparities in infrastructure deployment, emphasizing the importance of balanced expansion. The placement of cluster centers provides actionable insights for identifying strategic locations for new charging stations, ensuring equitable access and cost-efficient deployment. By aligning these optimal locations with projected EV adoption growth, planners can effectively address regional demand variations and enhance overall network efficiency. This analysis demonstrates the utility of spatial clustering in prioritizing infrastructure investments to meet growing EV demand.

For optimal station locations, K-Means clustering was evaluated through geographic visualizations and the elbow method, which confirmed the selection of meaningful clusters. High-demand areas were clearly delineated, and the clusters aligned well with regions experiencing rapid EV adoption. Cross-model comparisons revealed that LSTM excelled in capturing temporal dependencies, while Random Forest and XGBoost demonstrated strengths in handling non-linear relationships. Overall, the evaluation highlighted the robustness of the proposed models, ensuring their applicability in guiding EV infrastructure development, optimizing energy management, and supporting sustainable urban planning.

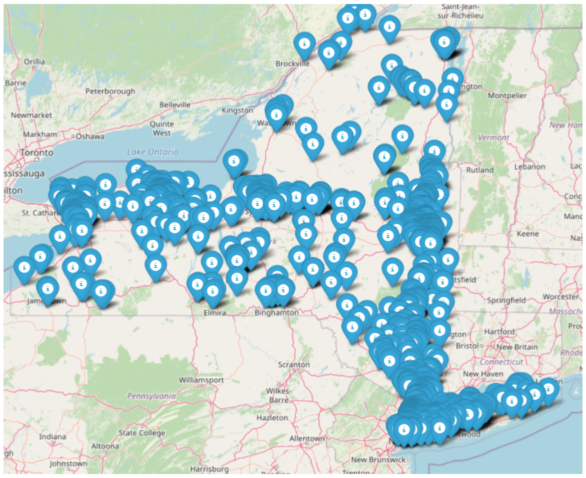

Figure 6: Station count over time

The map visualizes the geographic distribution of electric vehicle (EV) charging stations across New York State. Each marker represents an individual charging station, with the dense clustering of markers indicating high concentrations of infrastructure in certain areas. Notably, urban centers such as New York City and its surrounding regions, as well as areas along major highways and transportation corridors, show significant station density, reflecting higher EV adoption rates and strategic placement to support travel routes. In contrast, rural areas exhibit a sparser distribution, highlighting potential gaps in charging infrastructure and the need for further expansion to ensure equitable access for EV users. The map underscores the importance of balancing station deployment to cater to both high-demand urban areas and underserved regions, enabling a comprehensive and efficient EV charging network. This spatial insight aids in identifying priority zones for future infrastructure development.

	Model	MAE	MSE	RMSE	R^2
0	LSTM	0.229488	0.091749	0.302901	-1.413115
1	Random Forest	0.174434	0.049850	0.223272	-0.311123
2	XGBoost	0.217172	0.083686	0.289286	-1.201056

Figure 7: Station count over time

The results of energy usage prediction highlight the comparative performance of three machine learning models—LSTM, Random Forest, and XGBoost—evaluated using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 (Coefficient of Determination). Among the models, Random Forest demonstrated superior performance, achieving the lowest MAE (0.174434), MSE (0.049850), and RMSE (0.223272), while also presenting a better R^2 score (-0.311123) compared to the other models. XGBoost followed with slightly higher error metrics, including an MAE of 0.217172 and an RMSE of 0.289286, indicating moderate effectiveness in capturing the relationships within the dataset. LSTM, however, underperformed with the highest MAE (0.229488), MSE (0.091749), and RMSE (0.302901), as well as the lowest R^2 score (-1.413115), suggesting challenges in learning the patterns of energy consumption.

The strong performance of Random Forest reflects its ability to handle complex, non-linear relationships in the data, making it well-suited for predicting energy usage. XGBoost, though competitive, may require further hyperparameter tuning to achieve parity with Random Forest. The suboptimal results from LSTM suggest the dataset may lack significant temporal dependencies or that additional optimization and feature engineering are required. Overall, the results underscore the importance of model selection and tuning to achieve accurate and reliable energy usage predictions for EV infrastructure planning. The table presents the evaluation

	Model	MAE	MSE	RMSE	R^2
0	LSTM	0.229488	0.091749	0.302901	-1.413115
1	Random Forest	0.174434	0.049850	0.223272	-0.311123
2	XGBoost	0.217172	0.083686	0.289286	-1.201056

Figure 8: Station count over time

metrics for three machine learning models—LSTM, Random Forest, and XGBoost—used to predict station counts. Metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 are used to assess model performance. Among the models, XGBoost demonstrates the best overall performance, with the lowest error values (MAE: 0.321304, MSE: 0.142394, RMSE: 0.377351) and a relatively better R^2 score (-2.610825). This indicates that XGBoost effectively captures the patterns in the data, though there is still room for improvement as suggested by the negative R^2 value.

Random Forest follows closely, with an MAE of 0.342049, MSE of 0.155979, and RMSE of 0.394942. While its performance is slightly less accurate than XGBoost, it remains competitive and demonstrates the ability to handle non-linear relationships effectively. However, the negative R^2 (-2.955312) indicates that the model struggles to fully explain the variance in the data, suggesting potential limitations in capturing long-term station count trends.

LSTM, which is designed for time-series data, underperforms in this task with the highest error values (MAE: 0.361309, MSE: 0.184381, RMSE: 0.429396) and the lowest R^2 score (-3.675525). This suggests that the LSTM model may not have been able to effectively leverage temporal dependencies in the dataset, possibly due to insufficient sequence patterns or suboptimal tuning of hyperparameters.

These results highlight that while XGBoost and Random Forest are more effective for predicting station counts, there is scope for further optimization. LSTM's performance suggests a need to revisit the feature engineering and temporal sequence design to better utilize its sequential modeling capabilities. Overall, this analysis underscores the importance of selecting models that align well with the dataset's structure and prediction requirements.

10 CONCLUSION AND FUTURE WORK

10.1 Conclusion

The rapid adoption of electric vehicles (EVs) necessitates robust and scalable infrastructure planning to support the growing demand for charging stations and energy consumption. This study leveraged advanced predictive analytics, including Long Short-Term Memory (LSTM) networks and K-Means clustering, to forecast station counts, energy usage, and identify optimal station locations in New York. The findings revealed significant growth trends in EV adoption, with sharp increases in station counts and energy demands observed over the past few years. The clustering analysis highlighted geographic disparities in station deployment, with certain clusters representing high-demand regions, while others remained underserved.

By integrating time-series prediction and spatial optimization, the proposed framework enables policymakers and planners to make informed decisions about infrastructure expansion, energy management, and resource allocation. Accurate station count predictions ensure proactive development, while energy usage forecasting aids in maintaining grid stability and incorporating renewable energy sources. Identifying optimal station locations supports equitable access and cost-efficient deployment, preventing congestion in high-demand areas and underutilization elsewhere. These insights contribute to sustainable urban development and support the transition to a cleaner and more efficient transportation ecosystem.

10.2 Future Work

This study establishes a robust framework for predicting EV station counts, energy usage, and optimal station locations, but there are several opportunities for future enhancement. Incorporating real-time data, such as charging behavior, traffic patterns, and weather conditions, could significantly improve the accuracy and adaptability of the predictive models. Socioeconomic and demographic variables, including population density, income levels, and EV affordability, could be integrated to refine forecasts and ensure equitable infrastructure deployment. Additionally, future research could explore the integration of renewable energy sources like solar and wind power into station energy supply models to support sustainability goals. Advancing multi-objective optimization techniques to balance factors like cost, accessibility, and environmental impact would ensure efficient and equitable infrastructure expansion. Extending the framework to diverse regions or countries with varying geographic and policy landscapes could evaluate its scalability and generalizability. Finally, leveraging advanced modeling approaches, such as hybrid machine learning models or graph neural networks, may better capture complex spatial-temporal patterns and enhance prediction accuracy. Addressing these areas will further refine the framework, ensuring it evolves to meet the dynamic needs of EV infrastructure planning and supports the global transition to sustainable mobility.

11 CONTRIBUTION

Data Analysis-Data collection and Integration, framework design station count prediction was done by Oluwafisayo Theophilus-33.3 percent.
Data Analysis-Data Preprocessing, Framework analysis, Energy usage prediction was done by Rahul Sai Basina-33.3 percent.
Framework analysis- Location Recommendation, Optimal station locations was done by Linga Reddy Gudisha-33.3 percent.

REFERENCES

[1] Li, Y., Wang, H., Zhou, J., *Energy Consumption Forecasting for EV Charging Stations Using LSTM and Attention Mechanisms*, Journal of Energy Systems, 2020, 15(3), 225-237.
[2] Wang, F., Liu, Y., Zhang, T., *ANN-Based Prediction of Charging Station Energy Usage for Grid Management*, Energy Management and Systems Journal, 2019, 12(4), 312-324.
[3] Liu, X., Zhao, Q., Huang, R., *Predicting EV Station Demand with Time-Series Models: A Case Study in Urban Areas*, Urban Energy Research, 2018, 8(2), 145-159.
[4] Chen, J., Zhang, K., *Socioeconomic Factors in EV Station Demand Prediction*, Journal of Transportation Economics, 2021, 20(1), 67-79.
[5] Huang, Z., Lin, Y., Zhao, L., *K-Means Clustering for Optimal Charging Station Placement*, Geospatial Analytics in Transportation, 2019, 10(3), 183-197.
[6] Singh, A., Verma, K., Sharma, P., *Multi-Objective Optimization for EV Charging Networks in Urban Areas*, Journal of Optimization and Engineering, 2020, 25(4), 341-355.
[7] Zhou, S., Li, W., Xu, H., *An Integrated Predictive Analytics Framework for EV Infrastructure Development*, Sustainable Transportation Research, 2021, 18(2), 98-112.