SciencePG
Science Publishing Group

Research Article

Optimizing EV Charging Infrastructure: A Data-Driven Approach to Predicting Power Demand and Analyzing Geographic Disparities

Archita Ruby Arun[*] (iD)

Department of Research Initiatives, Mission: MathMinds, Chicago, The United States

Abstract

This paper presents a data-driven approach to optimizing electric vehicle (EV) charging infrastructure using a stacked ensemble learning model, which predicts power demand (kWh) per session to address challenges like long wait times, geographic disparities, and uneven resource allocation. Leveraging data from 85 EV drivers across 105 charging stations, the study identifies critical factors influencing station performance, such as session duration, time of day, and regional demand. Extensive preprocessing steps, including cyclical encoding of time-related variables, one-hot encoding of categorical features, and standardization of numerical variables, ensured the dataset was properly prepared for machine learning analysis. The stacked ensemble model combines Random Forest, XGBoost, and Neural Network models, effectively capturing both linear and non-linear relationships in the data. The results highlight significant urban-rural disparities in charging infrastructure. Urban stations exhibit higher and more consistent demand, whereas rural areas show sporadic and limited usage, underscoring the need for targeted infrastructure investment in underserved regions. Temporal patterns further reveal peak charging demand during business hours at workplace stations, emphasizing the potential for dynamic optimization of station placement and operational capacity based on usage trends. The model achieved a low Mean Squared Error (MSE) on training data (0.1577 kWh), but a higher MSE on test data (1.7875 kWh) indicates overfitting, suggesting the need for further refinement. Despite this limitation, the model offers valuable insights into optimizing EV charging networks, enabling policymakers and developers to improve infrastructure planning and reduce geographic inequities. Future work will focus on expanding the dataset to include residential and public charging scenarios, incorporating additional variables like weather and traffic patterns, and refining model architecture to improve generalization. This study contributes to building equitable and efficient EV charging networks, supporting the growing adoption of sustainable transportation.

---

[*]Corresponding author: rarun@missionmathminds.org (Archita Ruby Arun)

1. Introduction

1.1. Background

The rapid adoption of electric vehicles (EVs) has created an urgent need for a well-distributed charging infrastructure. This need has become increasingly apparent as EV sales continue to rise, driven by growing consumer interest. According to the U.S. Department of Energy [1], battery electric vehicles (BEVs) are capturing a growing share of total light-duty vehicle sales, with many new car shoppers considering an EV for their next purchase.

As more consumers transition to EVs, the demand for reliable, fast-charging stations is outpacing supply, particularly in rural areas. Pew Research Center [2] highlights the significant geographic disparities, noting that only 17% of rural residents have access to a public charging station within a mile of their residence. This imbalance persists despite efforts from the Biden-Harris Administration, which has doubled the number of publicly available EV chargers since taking office (U.S. Department of Energy, [4]). By 2024, there were over 190,000 public charging ports, with approximately 1,000 new public chargers being added each week (U.S. Department of Energy, [4]).

However, reliability remains a major concern. J.D. Power [3] reports a 25% to 46% increase in broken charging stations, leading to a 20% drop in consumer satisfaction with public DC fast-charging networks between 2022 and 2024. This dissatisfaction underscores the need for further investments in charger maintenance and technology improvement, which the administration has begun addressing through new funding initiatives, including $325 million in January 2024 (U.S. Department of Transportation, 2024).

These efforts aim to support the transition to a zero-emission federal fleet, with light-duty zero-emission vehicle acquisition targets by 2027 and medium- and heavy-duty targets by 2035 (U.S. Department of Transportation, [5]). Nonetheless, challenges such as high installation costs, grid capacity issues, and geographic disparities remain, making it essential to address these barriers for an equitable and effective EV charging network.

1.2. Problem Statement

Despite increased efforts to expand charging networks, many EV owners still face challenges such as long wait times, geographic limitations, and unreliable charging stations. These issues are often the result of uneven infrastructure distribution and a lack of accurate power demand prediction. To address these problems, there is a need for more precise demand forecasting and optimization of station placement and operational features. Predictive models can play a crucial role in helping decision-makers allocate resources more effectively, improving both the availability and reliability of charging stations. This paper aims to fill these gaps by developing a machine learning model that optimizes EV charging station features, such as power usage patterns, station location, and maintenance scheduling. By doing so, the model seeks to reduce wait times, mitigate geographic disparities, and ensure the charging network keeps pace with the growing adoption of EVs.

1.3. Purpose and Objectives

This study develops a stacked ensemble approach model that predicts power (kWh) required for each charging session, aiming to minimize wait times and address geographic disparities in EV infrastructure. Current charging networks struggle to meet rising EV demand, particularly in underserved areas, resulting in long wait times and uneven access. By accurately predicting power usage and station load, the model helps optimize station placement and operational capacity, ensuring charging stations are located in high-demand areas and equipped to handle peak usage.

Trained on data from 85 EV drivers across 105 charging stations, the model identifies key factors affecting station performance, such as peak usage and regional demand. This data-driven approach provides a framework for policymakers and developers to improve infrastructure deployment and reliability, reducing geographic disparities and minimizing congestion. By guiding optimal station placement and operation, the model offers a solution for creating a more efficient and accessible EV charging network.

2. Literature Review

The rapid growth of electric vehicle (EV) adoption has sparked extensive research into optimizing charging infrastructure. This literature review examines recent studies on EV charging demand prediction, infrastructure planning, and socioeconomic considerations, with a focus on linear regression models and geographic disparity analysis.

2.1. Charging Demand Prediction

Recent advancements in machine learning have significantly improved EV charging demand forecasting. Wang et al. [6] and Yi et al. [7] employed various predictive models to forecast EV charging demand at station and commercial levels, respectively. These approaches have shown promise in capturing complex temporal patterns in charging behavior. Koohfar et al. [9, 10] conducted a comparative analysis of various machine learning algorithms for EV charging demand prediction, providing insights into the effectiveness of different techniques, including linear regression.

Linear regression models, in particular, have been widely used due to their simplicity and effectiveness in handling numerical data with linear relationships between input fea-

tures and output. These models have demonstrated success in predicting power demand and optimizing charging station performance, especially when dealing with time-series data and multiple variables such as charging duration, time of day, and geographic location.

2.2. Infrastructure Planning and Optimization

Optimizing the location and capacity of EV charging stations is crucial for efficient infrastructure development. Hu et al. [14] proposed a model that considers both point and path demands for charging station location optimization. Ge et al. [12] developed a data-driven multi-objective optimization approach, integrating simulation and optimization techniques to balance multiple objectives in EV charging infrastructure planning.

Shahraki and Pantoš [15] focused on optimal locations for public charging stations based on real-world vehicle travel patterns, emphasizing the importance of data-driven approaches. Their work highlighted the potential of using historical charging data and geographic information to inform infrastructure planning decisions, aligning closely with the objectives of linear regression-based predictive models.

2.3. Socioeconomic Considerations and Accessibility

Several studies have addressed the socioeconomic aspects of EV charging infrastructure, which are particularly relevant to geographic disparity analysis. Khan et al. [11] analyzed the distribution of EV charging stations in New York City, revealing correlations between socio-demographic features and charging station accessibility. Lou et al. [13] conducted a comprehensive national study in the United States, uncovering income and racial disparities in access to public EV charging infrastructure.

These studies underscore the importance of considering geographic and socioeconomic factors in predictive models for EV charging infrastructure. Linear regression models can be particularly useful in quantifying these disparities and informing policy decisions to ensure equitable access to charging stations across different regions and demographic groups.

2.4. Challenges in Data Quality and Availability

A recurring theme in the literature is the challenge of data quality and availability for EV charging infrastructure optimization. Many studies highlight the limitations of publicly available datasets, which often lack granularity or critical variables necessary for accurate predictions. For instance, Cui et al. [8] noted the importance of capturing fast-charging behavior at public stations, emphasizing the need for comprehensive data collection strategies.

The success of linear regression models, like the one pro-posed in this study, heavily depends on the quality and comprehensiveness of the input data. Future research [17-19] should focus on improving data collection methods and developing techniques to handle incomplete or noisy data effectively.

This literature review demonstrates the multifaceted nature of EV charging infrastructure optimization [20, 21], encompassing demand prediction, location optimization, and socioeconomic considerations. The use of linear regression models, as proposed in this study, offers a promising approach to address these challenges, particularly in predicting power demand and analyzing geographic disparities [16]. The field continues to evolve rapidly, with new methodologies and approaches being developed to address the complex challenges of creating an efficient and equitable EV charging network.

3. Methodology

3.1. Dataset Description

The dataset used in this study consists of charging session data from 85 EV drivers across 105 charging stations located in 25 workplace sites. Key variables include total kilowatt-hour (kWh) charged (the target variable), charge duration, platform used, distance traveled by the user, and station ID. These features allow for an in-depth analysis of charging patterns, station performance, and geographic variations.

To prepare the data for machine learning, several preprocessing steps were applied. Time-related variables, such as startTime and endTime, were encoded using cyclical encoding to handle the circular nature of time, ensuring smooth numerical relationships between times close to each other. A new feature, session duration, was derived to capture charging session length.

Categorical variables like day of the week and platform used were transformed into one-hot encoded vectors for numerical interpretation, while numerical features, including total kWh charged and charge duration, were standardized to maintain consistent scaling across all features.

Station ID was retained for station-specific and geographic pattern analysis. Distance traveled was included to provide context on the vehicle's state of charge and urgency. Columns irrelevant to the analysis, such as session ID and user ID, were dropped to reduce noise. Missing data was handled using appropriate imputation techniques.

This streamlined preprocessing approach ensured the dataset was well-prepared for machine learning, with cyclical encoding, one-hot encoding, and standardization enabling more accurate predictions and insights into EV charging behavior across locations.

3.2. Model Architecture

This study employed a stacked ensemble learning approach

to predict the total kilowatt-hours (kWh) consumed in each electric vehicle (EV) charging session. Initially, linear regression was considered due to its simplicity and effectiveness in modeling linear relationships between input features and output [21]. However, to improve predictive accuracy and capture potential non-linear interactions in the data, the final model used was an ensemble of machine learning algorithms. This method combines multiple models, each designed to capture different patterns within the data, resulting in a more robust prediction framework.

The dataset, consisting of charging session data from 85 EV drivers across 105 charging stations at 25 different workplace sites, contained several key input features. These included the day of the week, platform used for charging initiation, charging session duration in hours, distance traveled prior to charging, and a unique identifier for each charging station. These features were carefully selected for their hypothesized impact on charging behavior and power demand, supported by previous research in the field of EV infrastructure optimization.

To preprocess the data, several steps were taken to ensure it was properly formatted for machine learning. Cyclical encoding was applied to time-related variables, such as start-Time and endTime, to account for the circular nature of time. A new feature representing the session duration was also created. Numerical variables, such as total kWh charged and charge duration, were standardized to ensure all features contributed equally to the model. Categorical variables, such as the platform type and the day of the week, were one-hot encoded to allow for numerical interpretation. Irrelevant features like session ID and user ID were removed to reduce noise in the data.

The stacked ensemble approach combined multiple base models, each capturing different aspects of the data [22-24]:

1) Random Forest Regressor (RF): A robust non-linear model capable of capturing complex interactions between input features and output variables through decision trees.
2) XGBoost Regressor (XGB): A gradient-boosting algorithm known for its effectiveness in handling structured and time-dependent data.
3) Neural Network (NN): A Multi-Layer Perceptron (MLP) model with two hidden layers, designed to capture non-linear relationships that might be missed by traditional machine learning models.

The outputs from these three base models were then used as input features for a meta-model. This meta-model, another neural network, learned from the outputs of the base models to make more accurate predictions. By leveraging the strengths of each individual model, the ensemble approach improved the overall performance and generalizability of the prediction system.

3.3. Preprocessing and Feature Engineering

Prior to model training, comprehensive preprocessing and feature engineering steps were implemented to optimize the dataset for machine learning analysis. These steps were critical in transforming raw data into a format suitable for the prediction task and enhancing the model's overall performance.

Categorical variables, such as the platform used for charging initiation and the day of the week, underwent one-hot encoding. This process transformed non-numeric categorical data into binary vectors, enabling the model to interpret these features effectively. For example, the 'day of the week' variable was expanded into seven binary columns, one for each day, with a value of 1 indicating the presence of that day and 0 otherwise.

Time-related features, specifically the start and end times of charging sessions, underwent cyclical encoding to account for their repetitive nature. Since time follows a cyclical pattern (e.g., 23:00 is close to 00:00), sine and cosine transformations were applied to these features, capturing this cyclic behavior. This transformation allowed the model to accurately interpret temporal data without introducing artificial discontinuities. Additionally, we calculated the charging duration as a new feature, representing the total time spent charging during each session.

Numerical variables, such as charge duration and distance traveled, were standardized to have a mean of zero and a standard deviation of one. This was essential to ensure consistency across different scales, preventing variables with larger magnitudes from disproportionately influencing the model's learning process. Standardization was especially crucial for the linear regression model, which assumes equal contribution from all features.

Further feature extraction was conducted on datetime variables, converting them into useful components such as hour of day and month. These components helped the model capture temporal patterns associated with specific times or days that may influence charging behavior and power consumption.

In an effort to reduce noise and simplify the model, several columns were eliminated from the dataset, including 'sessionId', 'userId', and 'stationId'. These identifiers, while potentially useful in other types of analyses, did not provide meaningful contributions to the prediction of kWh usage in this context. Their removal served to reduce model complexity and minimize the risk of overfitting, allowing the model to generalize better to new data.

Collectively, these preprocessing and feature engineering steps transformed the raw dataset into a machine learning-ready format. By encoding categorical variables, standardizing numerical features, applying cyclical encoding to time-related variables, and removing non-contributory columns, the dataset was optimized for training the predictive model. These steps provided a solid foundation for develop-

ing a robust machine learning model capable of accurately predicting power demand at EV charging stations.

3.4. Model Training and Evaluation

The model training and evaluation were conducted using a systematic approach to ensure optimal performance and generalizability. Initially, a linear regression model was considered due to its simplicity and effectiveness in handling linear relationships between input features and the output. However, to enhance predictive accuracy and account for potential non-linear interactions in the data, a stacked ensemble learning approach was ultimately employed. This ensemble approach combined multiple models to capture different aspects of the data, resulting in a more robust prediction framework.

Key features such as the day of the week, platform used for charging initiation, charging session duration in hours, distance traveled prior to charging, and a unique identifier for each charging station were included in the model. These features were carefully selected based on their hypothesized impact on charging behavior and power demand, supported by prior research. Cyclical encoding was applied to time-related variables such as startTime and endTime to capture their repetitive nature. One-hot encoding was applied to categorical variables like platform type and day of the week. Numerical variables such as charge duration and distance traveled were standardized to ensure equal contribution from all features. These preprocessing steps helped optimize the data for machine learning.

The predictions from these base models were then used as input for a meta-model, which was another neural network. This meta-model learned from the outputs of the base models to generate more accurate final predictions. By combining the strengths of each base model, the stacked ensemble approach improved the overall performance and generalization ability of the prediction system.

An 80-20 train-test split was used, with 80% of the dataset allocated for training and 20% for testing. This ensured that enough data was available for training while allowing the model's performance to be evaluated on unseen data. Mean Squared Error (MSE) was used as the primary evaluation metric. MSE quantifies the average squared difference between the predicted and actual kWh values, providing a comprehensive measure of the model's accuracy. The optimization process focused on minimizing MSE to enhance the predictive accuracy of the model.

Initially, the linear regression model struggled with underfitting, as it was unable to capture the complexities of EV charging data. To address this, additional feature engineering steps were undertaken, such as creating interaction terms and applying polynomial transformations. However, it was ultimately the ensemble learning approach that effectively captured both linear and non-linear relationships, leading to improved model performance.

4. Results and Analysis

4.1. Model Performance

The performance of the stacked ensemble model was rigorously evaluated using Mean Squared Error (MSE), a commonly used metric for regression tasks. MSE measures the average squared difference between the predicted and actual values, with lower values indicating better model performance. This metric was particularly suited for the task of predicting the total kilowatt-hours (kWh) consumed during electric vehicle (EV) charging sessions, as it provides a comprehensive assessment of the prediction accuracy by penalizing larger errors more heavily than smaller ones.

The stacked ensemble model employed in this study integrated the predictions from three base models: Random Forest Regressor, XGBoost Regressor, and a Neural Network (Multi-Layer Perceptron). These models, each trained to capture different relationships and interactions within the dataset, provided input to a meta-model. The meta-model, a secondary neural network, combined the outputs of the base models to generate final predictions.

The performance of the meta-model was evaluated on both the training and test datasets, yielding the following results:

1) Meta-Model Train MSE: 0.1577 kWh
2) Meta-Model Test MSE: 1.7875 kWh

The relatively low train MSE indicates that the model successfully captured the underlying patterns in the training data. A value of 0.1577 kWh suggests that, on average, the model's predictions during training deviated only slightly from the actual kWh values, highlighting its ability to learn from the data effectively. However, the higher test MSE of 1.7875 kWh reflects a notable gap in performance when applied to unseen data.

This disparity between training and test MSE suggests the presence of overfitting, a condition where the model becomes too specialized in learning the idiosyncrasies of the training data, capturing noise or minor variations that do not generalize well to new datasets. Overfitting is a common challenge in machine learning, particularly when models are overly complex or when the training data is not fully representative of future observations.

The higher test MSE indicates that the stacked ensemble model, while effective on the training set, struggles to maintain the same level of accuracy on unseen data. This can be attributed to several factors. First, the use of multiple base models in the ensemble may have led to over-complexity, where the meta-model integrated minor variances from each base learner into the final predictions. Second, the dataset may contain certain noise or specific patterns that, although significant during training, do not persist across the broader dataset.

To mitigate overfitting, future iterations of the model could explore regularization techniques, such as L1 or L2 regularization, to penalize overly complex models and encourage

simpler, more generalizable patterns. Additionally, employing cross-validation techniques during training, such as k-fold cross-validation, could help ensure that the model is trained and validated on diverse subsets of the data, leading to better generalization to unseen data. Furthermore, simplifying the architecture of the meta-model or reducing the number of base models in the ensemble may also improve generalizability by reducing model complexity.

While the current stacked ensemble model offers promising predictive accuracy, improvements can be made to enhance its robustness on test data. Future work could involve experimenting with alternative base models and adjusting hyperparameters to reduce overfitting. Moreover, augmenting the dataset with additional features or data points may help capture a more complete representation of the EV charging behavior, thereby improving model generalization.

In addition, the potential to explore more advanced ensemble techniques, such as blending or stacking with different algorithms, could further optimize performance. Exploring other loss functions beyond MSE, such as Mean Absolute Error (MAE), could offer additional insights into the model's predictive capabilities by evaluating errors in absolute terms rather than squared differences.

In summary, while the stacked ensemble model demonstrated strong learning capabilities on the training data, its performance on the test set suggests that further refinement is necessary to improve its ability to generalize to unseen data. The insights gained from this analysis provide a strong foundation for future efforts to refine predictive models for EV charging behavior, with potential applications in optimizing energy consumption and improving the efficiency of charging infrastructure.

4.2. Key Findings

The ensemble approach provided several valuable insights into EV charging patterns, offering a deeper understanding of factors influencing charging behavior and demand.

First, the duration of the charging session was found to be strongly correlated with the total kWh consumed, confirming the expected linear relationship between session length and power usage. This relationship is intuitive, as longer charging sessions naturally result in higher energy transfer. However, the ensemble model's ability to capture this correlation across varying conditions and stations reinforces the reliability of this variable as a key predictor of power consumption.

Second, geographic disparities were observed in the data, with significant differences in charging station usage between urban and rural areas. Urban stations exhibited more consistent usage patterns and higher demand, suggesting that charging infrastructure is better established and utilized in these regions. In contrast, rural stations saw lower and more sporadic demand, indicating that these areas may remain underserved by the current charging infrastructure. This finding points to the need for targeted expansion of charging networks in rural areas to meet growing EV adoption.

Third, clear time-based patterns emerged from the analysis. Charging stations located at workplaces demonstrated peak usage during business hours, aligning with standard work schedules. This suggests that demand for charging infrastructure is highly influenced by temporal factors, and highlights the potential for dynamic optimization of charging station availability and capacity based on usage patterns throughout the day. For instance, increasing charging availability during peak business hours and reducing it during off-peak times could optimize resource allocation and improve station efficiency.

Finally, the meta-model demonstrated the value of combining the strengths of different machine learning algorithms—specifically, Random Forest, XGBoost, and Neural Networks. Each of these models excels at capturing different relationships within the data: Random Forests handle complex feature interactions, XGBoost effectively models time-dependent and structured data, and Neural Networks excel at identifying non-linear relationships. By integrating these diverse strengths, the meta-model achieved superior predictive accuracy compared to any single model alone. This ensemble approach highlights the importance of leveraging multiple perspectives in complex predictive tasks, allowing for a more comprehensive understanding of EV charging behavior.

These insights have important implications for future infrastructure planning and optimization, indicating opportunities to improve charging availability and efficiency based on geographic, temporal, and behavioral factors.

4.3. Limitations

Despite the promising results of this study, several limitations must be acknowledged. First, the disparity between the training and testing MSE suggests potential overfitting in the meta-model. This issue could be addressed in future research by fine-tuning hyperparameters, implementing regularization techniques, or employing cross-validation strategies to improve the model's ability to generalize to unseen data.

Second, the dataset used, while informative, was relatively limited in both size and scope, consisting primarily of workplace charging data. This homogeneity may limit the model's ability to generalize effectively to other charging contexts, such as residential or public charging stations. Future research should focus on incorporating more diverse datasets that capture a broader range of charging behaviors, which would enhance the model's robustness and applicability across different environments.

Lastly, the model's predictive accuracy could potentially be improved by including additional features. Variables such as weather conditions, which can influence charging patterns, or traffic data, which may affect travel distances and charging needs, could be integrated into future iterations of the model. Exploring and incorporating such features could significantly refine the model's predictive capabilities and increase its overall explanatory power.

5. Discussion

5.1. Contextualizing Through Prior Research

This study contributes to the field of EV infrastructure optimization by employing a stacked ensemble learning model to predict power demand (kWh) per charging session, addressing challenges such as geographic disparities and efficient resource allocation. Previous research has explored various approaches to charging demand prediction and infrastructure planning. Wang et al. [6] and Yi et al. [7] demonstrated the potential of machine learning for forecasting EV charging demand at station and commercial levels, using algorithms like linear regression and deep learning models. Koohfar et al. [9] further compared the performance of multiple machine learning algorithms, providing insights into their strengths and limitations. However, these studies often relied on single-model approaches that, while effective in capturing specific patterns, may fail to integrate the diverse relationships within EV charging data.

Additionally, research by Hu et al. [14] and Ge et al. [12] examined infrastructure planning using optimization models that balance multiple objectives, including location and capacity considerations. Shahraki and Pantoš [15] highlighted the importance of data-driven approaches for determining optimal public charging station locations based on vehicle travel patterns. These works underscore the growing reliance on data analytics for EV infrastructure development but often lack the focus on combining multiple model strengths to address non-linear and dynamic relationships.

This study builds on these foundations by integrating Random Forest, XGBoost, and Neural Network models into a stacked ensemble, capturing both linear and non-linear relationships in the data. Unlike prior work, it also highlights geographic disparities by analyzing rural and urban utilization patterns and incorporating advanced preprocessing techniques like cyclical encoding and one-hot encoding. By situating itself within the broader context of machine learning-based infrastructure optimization, this research offers a comprehensive framework to enhance predictive accuracy and inform equitable infrastructure planning.

5.2. Addressing Key Challenges

The predictive efficacy of the ensemble model in this study underscores the advantages of integrating diverse machine learning algorithms to encapsulate the multifaceted nature of EV charging behaviors. However, the persistent challenge of overfitting, as evidenced by the disparity between training and testing MSE, necessitates further refinement. Future research should explore the implementation of regularization techniques, such as Lasso or Ridge regression, or the adoption of cross-validation methodologies, like k-fold cross-validation, to enhance the model's capacity for generalization to previously unseen data.

Furthermore, the exploration of more sophisticated non-linear models, including deep neural networks or gradient boosting machines, could potentially improve the model's ability to discern intricate interactions within the data, thereby mitigating the observed test MSE. Additionally, the incorporation of a more heterogeneous dataset encompassing a wider array of charging scenarios, such as public charging stations or residential charging, could augment the model's robustness and applicability across diverse contexts.

5.3. Comparison to Existing Models

In contrast to heuristic models or simplistic single-model approaches, such as standalone linear regression, the proposed stacked ensemble model exhibits markedly superior performance in the intricate task of predicting kWh consumption at EV charging stations. This empirical superiority can be attributed to the ensemble's inherent capacity to synthesize the predictive strengths of diverse machine learning algorithms, thereby capturing a broader spectrum of linear and non-linear relationships within the data. While the exploration of more intricate models, such as deep neural networks with elaborate architectures, holds promise for further performance gains, the current ensemble model strikes a judicious balance between predictive efficacy and interpretability, rendering it a pragmatic and insightful tool for real-world applications. Moreover, the model's relative computational efficiency, compared to more complex deep learning alternatives, enhances its feasibility for deployment in resource-constrained environments.

5.4. Implications for Policy and Infrastructure Planning

The insights from this model can guide policymakers in prioritizing infrastructure development in under-served areas, particularly rural regions where access to charging stations is limited. The model's ability to predict temporal demand patterns can help optimize station placement and energy distribution, reducing wait times and improving station utilization.

Furthermore, the meta-model's predictions can support dynamic pricing strategies at charging stations, encouraging more efficient use of energy resources and aligning with smart grid technologies. The insights gleaned from this model can offer invaluable guidance to policymakers in strategically prioritizing the development of EV charging infrastructure in currently underserved regions, particularly in rural areas where access to such facilities remains limited. Furthermore, the model's capacity to accurately forecast temporal fluctuations in charging demand can significantly contribute to the optimization of both station placement and energy distribution strategies, thereby mitigating wait times and enhancing overall station utilization.

Moreover, the predictive capabilities of the meta-model can

serve as a robust foundation for the implementation of dynamic pricing mechanisms at charging stations, a measure that has the potential to incentivize more judicious and efficient consumption of energy resources, aligning seamlessly with the broader objectives of smart grid technologies.

6. Conclusion and Future Work

This study establishes a robust framework for predicting power demand at EV charging stations using a stacked ensemble model, demonstrating its potential to optimize infrastructure and reduce wait times. The model integrates Random Forest, XGBoost, and Neural Networks, effectively capturing the complex, non-linear relationships inherent in EV charging behavior. Key insights gleaned from the model include a strong correlation between charge duration and kWh consumption, the existence of geographic disparities in station utilization, and the identification of distinct temporal charging patterns. These insights can be leveraged to inform both infrastructure planning and policy decisions.

The study acknowledges certain limitations, including the restricted size of the dataset and its primary focus on workplace charging. Future research aims to address these limitations by expanding and diversifying the dataset to encompass charging data from residential, public, and fleet charging stations, thereby enhancing the model's generalizability. Furthermore, the incorporation of additional variables, such as weather conditions, local events, traffic patterns, real-time grid conditions, and user demographics, is envisaged to improve the model's predictive accuracy. The exploration of more advanced models, including deep learning architectures like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), as well as hybrid models, is planned to capture intricate spatial and temporal patterns.

Future research will also focus on integrating the model with smart grid technologies, enabling real-time monitoring, demand-response strategies, and vehicle-to-grid (V2G) integration for optimized energy management. Addressing ethical considerations, such as ensuring equitable access to charging infrastructure and mitigating potential biases in the model's predictions, will be a key priority. Additionally, efforts will be made to enhance the user experience by leveraging the model's insights for dynamic pricing mechanisms and user-friendly interface design.

In conclusion, this study lays a strong foundation for the utilization of ensemble-based models in the optimization of EV charging infrastructure. By addressing the outlined future research directions, the model can be further refined to support the development of a charging network that is not only reliable and efficient but also equitable, thereby facilitating the widespread adoption of electric vehicles and contributing to a more sustainable transportation future.

Abbreviations

EV	Electric Vehicle
BEV	Battery Electric Vehicle
DC	Direct Current
kWh	Kilowatt-Hour
RF	Random Forest
XGB	XGBoost
NN	Neural Network
MSE	Mean Squared Error
MAE	Mean Absolute Error
V2G	Vehicle-to-Grid
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network

Acknowledgments

Author Contributions

Archita Ruby Arun is the sole author. The author read and approved the final manuscript.

Funding

Data Availability Statement

Not applicable.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1]   U.S. Department of Energy. (2023). "Electric vehicle sales trends and projections." Vehicle Technologies Office.

[2]   Pew Research Center. (2024). "Electric vehicle charging access in rural areas: A study of geographic disparities."

[3]   J. D. Power. (2024). "U.S. electric vehicle experience public charging study: Reliability and consumer satisfaction."

[4]   U.S. Department of Energy. (2024). "Federal investment in EV infrastructure: A report on charging station growth and funding."

[5] U.S. Department of Transportation. (2024). "Charging and Fueling Infrastructure (CFI) Grant Program: Building a national network."

[6] Wang, Y., et al. (2023). "LSTM-based EV charging demand prediction at station level." Energy and Buildings, 278, 112608. https://doi.org/10.1016/j.enbuild.2023.112608

[7] Yi, Z., et al. (2022). "Time-series forecasting of commercial EV charging demand using deep learning." Applied Energy, 308, 118340. https://doi.org/10.1016/j.apenergy.2021.118340

[8] Cui, Q., et al. (2024). "Understanding fast-charging behavior of EV users at public stations using the Variational-Bayesian Gaussian-mixture model." Transportation Research Part C: Emerging Technologies, 140, 103741. https://doi.org/10.1016/j.trc.2022.103741

[9] Koohfar, F., et al. (2023). "Machine learning algorithms for EV charging demand prediction." Transportation Research Part D: Transport and Environment, 106, 103289. https://doi.org/10.1016/j.trd.2022.103289

[10] Brockway, A. M., et al. (2022). "Inequitable access to distributed energy resources due to grid infrastructure limits in California." Nature Energy, 7(9), 876-885. https://doi.org/10.1038/s41560-022-01065-6

[11] Khan, H. A. U., et al. (2022). "Inequitable Access to EV Charging Infrastructure." Transportation Research Record, 2676(5), 135-147. https://doi.org/10.1177/03611981221090212

[12] Ge, Y., et al. (2024). "Data-driven multi-objective optimization for electric vehicle charging infrastructure planning." Applied Energy, 335, 120801. https://doi.org/10.1016/j.apenergy.2023.120801

[13] Lou, Y., et al. (2024). "Income and racial disparity in household publicly available electric vehicle charging accessibility in the United States." Nature Communications, 15(1), 1-12. https://doi.org/10.1038/s41467-023-38206-0

[14] Hu, D., et al. (2023). "Electric Vehicle Charging Infrastructure Location Optimization with Mixed and Forecasted Charging Requirements." Scientific Programming, 2023, 9567183. https://doi.org/10.1155/2023/9567183

[15] Shahraki, N., et al. (2015). "Optimal locations of electric public charging stations using real world vehicle travel patterns." Transportation Research Part D: Transport and Environment, 41, 165-176. https://doi.org/10.1016/j.trd.2015.09.011

[16] Davidov, S., & Pantoš, M. (2017). "Planning of electric vehicle infrastructure based on charging reliability and quality of service." Energy, 118, 1156-1167. https://doi.org/10.1016/j.energy.2016.11.024

[17] Pagany, R., et al. (2019). "Spatial analysis of charging demand and optimal charging infrastructure placement." Energy, 180, 1008-1019. https://doi.org/10.1016/j.energy.2019.05.117

[18] Zhu, Z. H., et al. (2018). "Charging station location problem of plug-in electric vehicles." Journal of Transport Geography, 68, 160-168. https://doi.org/10.1016/j.jtrangeo.2018.03.014

[19] Xiang, Y., et al. (2016). "Economic planning of electric vehicle charging stations considering traffic constraints and load profile templates." Applied Energy, 178, 647-659. https://doi.org/10.1016/j.apenergy.2016.06.101

[20] Mak, H. Y., et al. (2013). "Infrastructure planning for electric vehicles with battery swapping." Management Science, 59(7), 1557-1575. https://doi.org/10.1287/mnsc.1120.1665

[21] Ghamami, M., et al. (2016). "A survey of models and algorithms for optimizing shared mobility." Transportation Research Part B: Methodological, 87, 115-134.

[22] Jia, L., et al. (2018). "Optimal siting and sizing of electric vehicle charging stations." IEEE Transactions on Power Systems, 33(3), 2721-2732.

[23] Xiong, Y., et al. (2018). "Optimal electric vehicle charging station placement." Transportation Research Part B: Methodological, 116, 340-359. https://doi.org/10.32657/10356/74999

[24] Dong, X., et al. (2019). "Charging electric vehicles in the smart city: A survey of economy-driven approaches." IEEE Transactions on Intelligent Transportation Systems, 20(6), 2184-2196.

Biography

Archita Ruby Arun is the founder and CEO of Mission: MathMinds, a nonprofit dedicated to applying technology, artificial intelligence (AI), and data-driven strategies to address critical challenges in infrastructure optimization and sustainability. By leveraging advanced machine learning techniques, Ruby has spearheaded impactful initiatives in electric vehicle (EV) infrastructure, predictive analytics, and equitable technology deployment, creating significant value and promoting accessibility in underserved regions. Ruby collaborates with leading institutions (Stanford, Yale) and organizations (Nexus Edge, Kumon International) to design scalable AI-driven models that address the intersection of technology, sustainability, and economic equity. Her work focuses on reducing systemic disparities through innovative and sustainable solutions, driving long-term advancements in AI and EV technology adoption.

Research Fields

Archita Ruby Arun: Artificial Intelligence, Electric Vehicles, Government Policy, Climate Change Action