**DATA 298B MSDA Project II**

**Workbook 1**

**Team 6**

**Lohitha Vanteru, Pranavi Sandrugu, Mahe Jabeen Abdul, Rohan Naga Venkata Mayukh Ungarala**

**October 02, 2023**

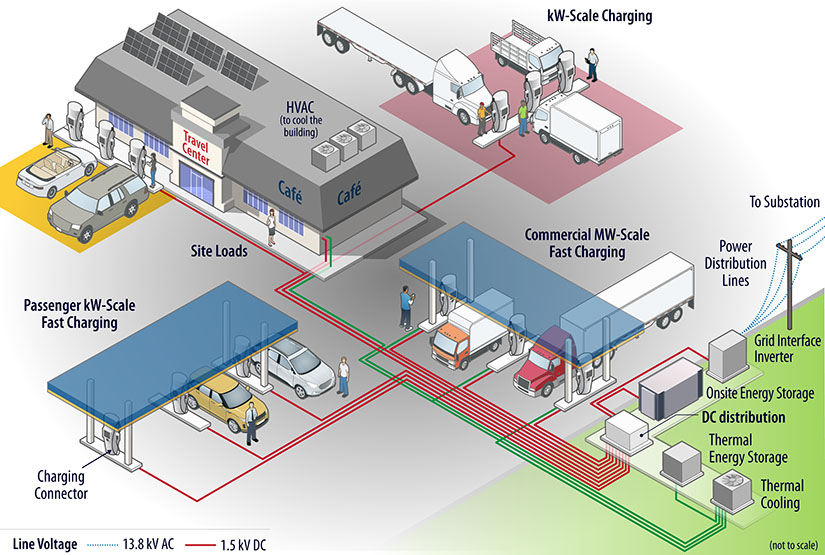
### 4. Model Development

**4.1 Model Proposals**

Electric vehicles (EVs) are quickly taking over as the preferred mode of transportation, so it is critical to strategically build out a reliable charging infrastructure. The availability of EV charging stations is crucial for the smooth transition to electric mobility, but their strategic location is also crucial for maximizing accessibility and use. This proposal describes a ground-breaking initiative that would use machine learning to develop a smart system for the accurate planning and deployment of EV charging infrastructure, designed especially for medium and heavy-duty motorized vehicles.

Our strategy envisions the design of an intelligent system that will revolutionize the deployment and planning of EV charging stations, particularly the kind that is customized to the requirements of medium- and heavy-duty cars. The ability to forecast the anticipated number of vehicles expected to be on the road in the years preceding 2035 serves as the basis for our intelligent EV charging infrastructure system. This predictive modeling strategy is based on time series machine learning algorithms that examine historical data and demographic trends of EV registrations, allowing us to forecast the rising demand for EV charging. We can calculate the number of charging stations needed to meet this demand by precisely calculating the number of vehicles, ensuring that EV users have easy access to charging stations.

Understanding the unique needs of various vehicle fleets operating in various regions, in addition to calculating the necessary number of charging stations, is one of the main issues in the planning of EV infrastructure. As a result, our model goes a step further by including information on the energy requirements and driving range of three different categories of heavy-duty vehicles: transit buses, school buses, and delivery vans. Due to the granularity, we can customize charging station recommendations to meet the specific needs of each fleet type, resulting in efficient and affordable solutions. Additionally, the proposal contemplates providing specific location recommendations for positioning new charging stations both within and across zip codes. These placement suggestions will be supported by data-driven insights that take into consideration variables like the anticipated number of vehicles, energy use, and traffic patterns. By carefully locating charging stations, we hope to promote programs for sustainable mobility while also meeting present and future demand for EVs.



**Figure 20:** *High-Power Medium- and Heavy-Duty Electric Vehicle Charging*

In a nutshell the suggested smart electrical vehicle (EV) charging system offers a comprehensive strategy to deal with the difficult issues involved in making proposals regarding the development of electric transportation in the future. We seek to guarantee the availability, usability, and effectiveness of EV charging infrastructure for medium and heavy-duty vehicles by utilizing machine learning and data-driven insights. This program encourages the widespread use of electric vehicles and the switch to cleaner transportation, which will help create a more environmentally conscious and enduring future.

***4.1.1 Forecast Heavy Duty and Medium Duty EV and Charging Stations Demand***

The availability of a dependable and widely distributed charging infrastructure is crucial as we move towards a sustainable future with electric transportation, especially for heavy and medium-duty electric vehicles (EVs). This proposal describes the creation of a prediction model using the Prophet model for predicting EV demand, including vehicle counts and the matching number of necessary charging stations, up to the year 2035. This artificial intelligence will provide insights necessary for improving the installation of EV charging infrastructure and maintaining accessibility and comfort for EV users by relying on historical information. At the outset, we compile a large dataset with details on the number of vehicles, zip codes, fuel kinds, vehicle brands, duty types, and years.

The Prophet time series forecasting model, recognized for its accuracy in capturing seasonality, trends, and holiday effects in data, will be the foundation of the suggested approach. The model will be trained on an extensive historical dataset that includes data on vehicle counts, charging station utilization, demographic trends, and other relevant factors in order to forecast future EV demand. We filter the data for each distinct zip code in the dataset, group the vehicle counts by years, then add the counts to get yearly totals. The Prophet model-compatible format is created from this aggregated data. The model has been developed to recognize seasonality, growth tendencies, and changepoints in the data, enabling precise forecasting. The model will generate projections for both the predicted numbers of heavy and medium-duty EVs on the roads and the accompanying demand for charging stations by examining this data. We will be able to catch subtle variations in EV adoption patterns over time and across different regions owing to the Prophet model's adaptability.

A close-up of a computer screen

Description automatically generated

**Figure 21:** *Model for Forecasting Heavy /Medium Duty EV Count and Charging Stations Demand*

The implementation of this method of forecasting could have a number of advantages. First of all, it will offer a data-driven roadmap for future EV infrastructure development to legislators, utility corporations, and infrastructure planners. More effective resource allocation and investment choices will be made with accurate estimates of car counts and charging station requirements. Additionally, it will hasten the adoption of EVs by ensuring that the expansion of the charging infrastructure keeps pace with the increase in demand, improving the overall EV driving experience for heavy- and medium-duty drivers. The resulting projections enable stakeholders to plan for sustainable transportation, deploy charge stations, and anticipate energy demand with confidence. This strategy supports the overarching objective of lowering greenhouse gas emissions and encouraging the use of battery-powered automobiles on a large scale in order to create a cleaner, more sustainable future.

***4.1.2 Predicting Vehicle Range of Heavy Duty and Medium Duty EV***

Predicting vehicle range based on charging data necessitates understanding how the vehicle's state of charge (SOC) varies over time and how this impacts the distance it can drive. Optimizing the operating efficiency of electric vehicles, such as transit buses, school buses, and delivery trucks, requires making range predictions based on charging data. In order to improve fleet management decisions, our research focuses on using charging data analytics to estimate the range of these three vehicle types.

The computation of the State of Charge (SOC) change over charging sessions, designated as "Delta\_SOC," commences the research procedure. The range of the vehicle is mostly determined by the amount of energy contributed to the battery during charging, which is revealed by this measure. In addition, we calculate "Average\_Power," which measures the rate of energy supply during charging, and "Charging\_Duration," which represents the amount of time spent charging. The effectiveness of the charging process is greatly influenced by these characteristics. We use a straightforward method to estimate the range, where "Range" is expressed as a percentage of "Delta\_SOC" in relation to the battery's total capacity. Range (in miles or kilometers) = Battery Capacity (in kWh) \* Efficiency (in miles or kilometers per kWh) is a formula used to express the unique features of the battery system in use. Battery capacity, which is commonly expressed in kilowatt-hours (kWh), is the total amount of energy that can be stored in the electric vehicle's battery pack, while efficiency refers to how far the car can go on a single kWh of electricity.

We leverage an ensemble of machine learning techniques, such as Random Forest, Gradient Boosting, Linear Regression, Bagging Regressor, Stacking Regressor, and a Voting Regressor that integrates the results of all models, for predictive modeling. The objective is to develop a solid model that can correctly forecast vehicle ranges based on charging characteristics. The weighted fusion of these models' forecasts ensures that each model's output is taken into account in accordance with its unique performance. This method accounts for each model's advantages while minimizing its disadvantages, producing a more accurate prediction of vehicle range.

A diagram of a diagram

Description automatically generated

**Figure 22:** *Model For Predicting Vehicle Range of Heavy Duty and Medium Duty EV*

Complex correlations in the data are exceptionally well captured by Random Forest, a collection of decision trees. Another ensemble technique, gradient boosting, creates models that improve sequentially and is very powerful for regression applications. A foundational model that offers simplicity and interpretability is provided by linear regression. We use ensemble approaches to increase forecasting accuracy. A type of bootstrap aggregation known as bagging uses the Random Forest model as its foundation estimator. Bagging lowers overfitting and boosts the stability of the model. In order to provide a more reliable final predictor, the Stacking Regressor combines predictions from various models (Random Forest, Gradient Boosting, and Linear Regression) with a meta-estimator (Linear Regression). This method minimizes the constraints of particular models while maximizing the strengths of several models. A Voting Regressor, which aggregates the results from all models while taking into account each one's performance individually, produces the final prediction. This strategy makes use of weighted fusion, allocating various weights to each model's predictions based on their historical reliability and accuracy. Weighted fusion makes sure that the most important models make a larger contribution to the final prediction, improving the accuracy of the final prediction as a whole.

***4.1.3 Predicting Short and Long-term Energy Demand of Heavy Duty and Medium Duty EV***

EV energy demand forecasting is estimating how much energy will be required to charge the car to a specific SOC or estimating how much charging is still necessary to get a full charge. The efficient management of energy resources, the optimization of charging infrastructure, and the smooth operation of different vehicle fleets, such as transit buses, school buses, and delivery trucks, all depend on the ability to predict the energy consumption for electric vehicles (EVs). Both long-term and short-term projections are implemented in this use case to fulfill various planning and operational objectives.

The Temporal Fusion Transformer (TFT) model is a cutting-edge neural network architecture built for time series forecasting tasks. The TFT model employs self-attention mechanisms and multi-head attention to capture complicated temporal connections within sequential data, drawing inspiration from the Transformer architecture's success in natural language processing. This makes the model particularly effective for forecasting jobs that need knowledge of historical context since it enables it to recognize long-term patterns and linkages. The TFT model takes into account temporal elements like the day of the week, the month, and the year. These characteristics are crucial for identifying patterns and seasonality in time series data. The TFT model can produce more precise and context-aware forecasts by taking both the temporal context and the order of the data points into account. The TFT model's capacity to combine both temporal and static elements, offering a comprehensive perspective of the data, is one of its standout characteristics. The TFT model can produce more precise and context-aware predictions by adding extra contextual data, such as categorical variables or metadata linked to each time series. Additionally, the TFT model is flexible and capable of multi-horizon forecasting, which allows it to make simultaneous predictions of future values at various time periods. This is particularly helpful when making forecasts on different time horizons, such as when estimating both the short- and long-term energy consumption in the context of charging electric vehicles.

A diagram of a diagram

Description automatically generated

**Figure 23:** *Model for**Predicting Energy Demand of Heavy Duty and Medium Duty EV*

***4.1.3.1 Long-Term Energy Demand Forecasting***

Long-term energy demand forecasting includes weekly, monthly, and yearly estimates, which provide useful information for strategic planning and resource allocation. A useful tool in this situation is the Temporal Fusion Transformer (TFT) concept. The TFT model can capture complicated linkages and multiple temporal characteristics of the data, which makes it useful for identifying trends and patterns in long-term energy consumption. For instance, the TFT model is capable of performing detailed analyses of previous data, such as charger ID, average power, starting SOC, ending SOC, and charging duration, to create precise forecasts regarding future energy use. We can develop a reliable long-term forecasting model by training the model on this historical data and optimizing it for MSE loss. The output of this model enables fleet managers and infrastructure planners to plan for fluctuations in energy demand and guarantee that there are enough charging resources available.

***4.1.3.2 Short-Term Energy Demand Forecasting***

Short-term energy demand forecasting, on the other hand, concentrates on estimating energy needs for the following day or within a shorter time frame, often in the order of next 6 hours. The use of this kind of forecasting in energy load balancing and real-time operational decisions is important. The TFT model uses its ability to record temporal dependencies and adapt to changing situations to be equally skilled at short-term forecasting. The model can forecast the amount of energy required for brief charging sessions by taking features like charger ID, average power, and SOC indicators as inputs. This forecast is essential for maximizing the use of energy, particularly in situations where many vehicles with various energy needs share charging infrastructure.

In a nutshell fleet managers and infrastructure planners are empowered by the combination of long-term and short-term energy demand predictions utilizing the TFT model to optimize the charging process for transit buses, school buses, and delivery trucks. By ensuring that the appropriate quantity of energy is accessible when and where it is most required, these forecasts will improve operational effectiveness, eliminate energy waste, and support sustainable transportation practices.

***4.1.4 Optimal Placement of New Charging Stations for each vehicle type***

Our research focuses on finding a solution to the crucial problem of strategically positioning new charging stations to meet the charging requirements of various vehicle types in the context of promoting the adoption of electrically powered automobiles (EVs) and enhancing sustainable transportation practices. In order to accomplish this, we built an optimization model that takes into account Points of Interest (POI), the proximity of current charging stations, and location effectiveness in addition to cost-effectiveness. The main goal is to reduce installation costs while making sure that EV users have easy access to the infrastructure for charging, taking into account their preferences for adjacent amenities. Our approach utilizes the capabilities of the PuLP library and linear programming to create a model of optimization for the best location of charging stations. To assist in guiding the decision-making process, this model takes into account a number of important characteristics and limitations.

We construct a pair of possible variables. Whether a charging station is put up for a specific vehicle type at a particular location is represented by the first set, marked by the symbol x. The second set, y, indicates if a site is chosen for the installation of a charging station. Important factors include installation costs, demand for charging per vehicle type, the highest possible number of charging stations that can be built (N) , and the greatest distance that a vehicle will travel to reach a charging station (R) as well as details on local POIs and available charging stations. Our objective function, which now takes into account installation costs as well as user-friendliness and location effectiveness, aims to reduce the overall cost of establishing new charging stations. It includes the price of installation for every charging station, accounting for consumers' preferences for adjacent conveniences and their closeness to already-existing charging stations. By focusing on the needs and convenience of EV users, the location is guaranteed to be appropriate. The following is a definition of the objective function as defined in equation (1):

Minimize: Total Cost = Σ[(Installation Cost at Location + PoI + Proximity to Existing Stations) \* x[Location, Vehicle Type]] (1)

*Installation Cost at Location:* The cost of installing a charging station for a particular vehicle type. *User Preference*: A factor that represents user preferences for nearby amenities (POI) at the location. *Proximity to Existing Stations:* A factor representing the proximity of existing charging stations to the location. *x [Location, Vehicle Type]:* A binary variable that indicates whether a charging station is installed at a specific location for a particular vehicle type.

Several limitations are incorporated to ensure a complete and user-friendly solution. Budgetary restrictions, coverage specifications, and a cap on the overall number of charging stations are included in these constraints. Additionally, we include restrictions that promote choosing a location based on POI proximity and an efficient charging station distribution. After resolving the optimization model, we obtain the best location for charging stations for various car kinds. This positioning maximizes both location efficacy and user-friendliness while minimizing installation expenses. While taking into account their preferences for local amenities, it makes sure that EV customers have simple access to charging infrastructure. The following equations collectively define the constraints considered for our optimization problem:

*Budget Constraint:* Ensure that the overall cost of installing the charging stations does not go beyond the allotted budget*.*

Σ[(Installation Cost at Location) \* x[Location, Vehicle Type]] ≤ Budget (2)

*Coverage Constraint*: Ensure that each type of vehicle has access to at least one charging station within the range of their maximum travel*.*

Σ[x[Location, Vehicle Type]] ≥ 1 for all Vehicle Types (3)

*Number of Charging Stations Constraint:* The total number of charging stations that may be installed must not be greater than N

Σ[y[Location]] ≤ N (4)

*Charging Station Location Decision Constraint:* Make sure that if a place is chosen (y[place] = 1), at least one charging station for each type of car is installed there. This is the charging station location decision constraint.

Σ[x[Location, Vehicle Type]] ≥ y[Location] for all Locations (5)

A diagram of a software system

Description automatically generated

**Figure 24:** *Model for Optimized Locations for New EV Charging Stations*

Our improved optimization approach delivers a comprehensive method for strategically placing EV charging infrastructure. It encourages the use of EVs, makes it easier for people to use charging stations, and supports sustainable mobility practices by taking into account a variety of criteria, such as installation costs, user preferences for local facilities, and the distribution of existing charging stations. This user-centered and location-aware strategy makes sure that the placement of the charging stations is not only economical but also improves the entire EV user experience.

**4.2 Model Supports**

***4.2.1 Environment, Platform, Tools***

A stable environment, appropriate platforms, and specific tools are necessary for the efficient execution of this integrated project, which includes forecasting EV and charging station demand, projecting vehicle range, forecasting energy use, and optimizing charging station placement. An overview of the essential elements is provided below.

The execution of various use cases within a single workflow, which includes forecasting EV and charging station demand, projecting vehicle range, short and long-term energy need, and optimizing charging station placement, necessitates a comprehensive environment, platform, and set of tools. The foundation is constructed using the Python programming language and its vast libraries for machine learning, optimization, and data analysis. The interactive development and documentation platform is Jupyter Notebook. The set of tools includes TensorFlow or PyTorch deep learning frameworks for Temporal Fusion Transformer (TFT) model implementation, Scikit-Learn for regression models and ensemble approaches, and Facebook's Prophet for demand forecasting. The PuLP package makes it easier to create linear programming models for charging station placement. Pandas, NumPy, and data visualization tools like Matplotlib and Seaborn are used for data management and analysis. Version control and collaboration are aided by Git/GitHub, while task organization is facilitated by project management software like Jira. With a particular emphasis on AWS SageMaker for machine learning and data analysis, the entire workflow is smoothly organized within the AWS environment. AWS S3 is used to store and manage historical data and Facebook Prophet is used to forecast time series data. SageMaker is a flexible platform for creating, honing, and deploying machine learning models, including Temporal Fusion Transformer (TFT) model application for energy demand forecasting and vehicle range prediction. Data retrieval and model execution are automated by AWS Lambda, computational operations are hosted by EC2, optimization data is managed by DynamoDB, and core AWS components are monitored by CloudWatch. AWS Glue and Athena are utilized for data transformation, and Tableau is used for data visualization.

A computer screen shot of a computer server

Description automatically generated

**Figure 25:** *High Level Architecture of Jupyter*

**A diagram of a model

Description automatically generated**

**Figure 26:** *Ensemble Modeling using Sage Maker Auto Pilot*

***4.2.2 Platform Architecture and Data Flow***

The diligently designed system that optimizes the processing of data and the completion of complicated activities is demonstrated by the platform architecture and data flow that underlie the comprehensive use cases. It starts with data ingestion, a crucial phase in which historical information on the demand for EVs and charging stations is routinely and systematically ingested into specific AWS S3 buckets. AWS Glue is used to seamlessly automate complex data transformation and cleansing procedures, enhancing data quality and consistency. These pre-processing steps make sure that the data has been meticulously cleaned up and geared up, thereby rendering it appropriate for subsequent research and modeling initiatives.

Following that, data analysis and modeling operations take center stage, all while being orchestrated within the limitations of AWS SageMaker Jupyter Notebooks. For data scientists, this setting acts as a dynamic center, giving them the freedom to work together to explore data, carry out complex preprocessing, and execute in-depth analyses of the data. SageMaker develops into a flexible platform for the comprehensive development, training, and rigorous evaluation of machine learning models in addition to its data analysis capabilities. It includes the creation and testing of models for estimating a vehicle's range as well as the application of the Temporal Fusion Transformer (TFT) model, an advanced deep learning technique designed specifically for the accurate forecasting of both short- and long-term energy requirements. It's noteworthy that the architecture includes a feature that optimizes the placement of new charging stations. This feature makes use of AWS Lambda functions, which are skilled at coordinating automatic data retrieval and swiftly and accurately performing the optimization model. On AWS EC2 instances, the optimization model that supports the data-driven decision-making process is housed. Additionally, AWS DynamoDB's meticulous organization and storage of the optimization data, which is essential for the model's effectiveness, ensures smooth and efficient data flow dynamics.

Diagram

Description automatically generated

**Figure 27:** *Machine Learning Data Flow*

Additionally, the platform architecture creates a special path just for reporting and data insights. A powerful analytics and visualization tool called Tableau enters the fray by expertly creating interactive dashboards and visualizations. Data analysts and stakeholders alike can gain deeper insights and make wiser decisions thanks to these artifacts. Jupyter Notebooks simultaneously play a crucial role, acting as the main method for coding documentation and creating in-depth project reports. The traceability and reproducibility of analyses, which are crucial in data-driven projects, are ensured by this documentation component. The workflow culminates with the deployment and scaling aspects. The smooth deployment of applications and models, which meets the demand for flexibility and scalability in real-world settings, is made possible by AWS Elastic Beanstalk and Amazon ECS. AWS Auto Scaling is particularly notable for being seamlessly integrated to dynamically distribute and manage resources, adjusting to the changing workloads with accuracy and efficiency.

In essence, the platform architecture and data flow comprise a finely woven tapestry of AWS services, which have been painstakingly created to give a cogent, scalable, and effective solution for the whole range of use cases. It makes use of SageMaker's flexible capabilities to enable the seamless integration of reporting, optimization, and machine learning. The end result is a solid platform that not only handles the challenging tasks at hand but also guarantees the flexibility to develop and adapt to shifting demands in the field of managing EVs and charging stations.

A diagram of a business process

Description automatically generated

**Figure 28:** *Model Platform Architecture*

**4.3 Model Comparison and Justification**

For each project component and module, we must analyze a number of models. As mentioned earlier, there are: Forecast Heavy Duty and Medium Duty EV and Charging Stations Demand ,Predicting Vehicle Range of Heavy Duty and Medium Duty EV, Predicting Short and Long-term Energy Demand of Heavy Duty and Medium Duty EV, Optimal Placement of New Charging Stations for each vehicle type. Model comparison and justification are crucial in determining the best approaches for the four use cases. Let's examine each use case in greater detail.

A detailed study of numerous models has been conducted in the field of EV and charging station demand forecasting to determine their viability for this challenging task. ARIMA, Long Short-Term Memory (LSTM), and Seasonal Autoregressive Integrated Moving-Average (SARIMA) have all been taken into consideration in addition to the Prophet model, which has been selected as the final model. The Prophet model stands out for its effective management of multiplicative seasonality, flexibility around changepoints, and innate capacity to accurately capture seasonality, holidays, and special events. It is an appealing option for predicting the demand for EVs and charging stations due to its adaptability in handling varied data patterns. Historical information on the demand for EVs and charging stations has been used in an empirical study to examine the performance of various models. We have carefully evaluated the ability of each model to reflect seasonality, short-term volatility, and long-term trends. Their ability to adapt to different data patterns has also received attention. The choice of the Prophet model fits with the intricacy of the forecasting assignment and the intrinsic complexity of the data. The results of these models are outlined in the table below, which highlights their individual advantages and disadvantages in predicting EV and charging station demand:

**Table 9:** Model comparison

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Strengths** | **Limitations** | **Performance** |
| Prophet | Seasonality and special events are well captured | It may be difficult to handle complex data patterns | Excellent |
| ARIMA | Effective for short-term fluctuations | Capacity to handle complex seasonality is limited | Moderate |
| LSTM | Captures long-range dependencies effectively | A substantial amount of data and tuning is required | Good |
| SARIMA | Effectively handles strong seasonal patterns | Less seasonal data may not perform well | Good |

In our selection of models within our ensemble for the task of forecasting vehicle range for Heavy Duty and Medium Duty Electric Vehicles (EVs) is justified by their unique capabilities and the requirements of the prediction task. This ensemble of models was chosen because of its comprehensive approach to prediction. We want to maximize the power of these algorithms by mixing models of different strengths. By taking into account both linear and nonlinear patterns in the data, the ensemble is built to ensure reliable predictions. With this method, the ensemble is better able to respond to a variety of conditions, making it a well-rounded option for estimating the vehicle range for Heavy Duty and Medium Duty EVs. Based on the relative effectiveness and reliability of each model within the ensemble, weighted regression includes giving each model's predictions a distinct weight. We can provide more weight to models that exhibit superior forecast accuracy and consistency since the weights indicate our trust in each model's predictions. By multiplying each model's forecast by the set weight and then adding the weighted predictions, the weighted fusion of predictions is calculated. It can be modeled mathematically as follows:

Weighted Fusion = (Weight\_Model1 \* Prediction\_Model1) + (Weight\_Model2 \* Prediction\_Model2) + ... + (Weight\_ModelN \* Prediction\_ModelN)

This weighted fusion, which represents a consensus opinion and makes use of the ensemble's collective predictive power while minimizing the effects of individual models' shortcomings, is the ultimate forecast for vehicle range. In conclusion, the strategic decision to include weighted regression in the group of models for estimating vehicle range is intended to maximize the prediction accuracy and resilience. The ensemble develops a well-balanced and flexible predictive system that can successfully address the challenges of vehicle range prediction for Heavy Duty and Medium Duty EVs by allocating weights to models based on their performance. A table that compares the various Voting Regressor versions and explains why the Weighted Fusion method was chosen is mentioned below.

**Table 10:** Comparison of various Voting Regressor versions

|  |  |  |  |
| --- | --- | --- | --- |
| **VotingRegressor Variant** | **Strengths** | **Limitations** | **Justificationfor Weighted Fusion** |
| Simple Voting Regressor | The majority vote principle is simple to implement | Ignores model performance differences | The optimization of performance is limited |
| Bagging Regressor | Reduces model variance and is robust to overfitting | Focuses on a single base model | Model diversity is limited |
| Stacking Regressor | Diverse models are combined, nuances are captured | Optimization requires a meta-learner | Diversity of models enhanced |
| Weighted Fusion | Model performance is incorporated as a weight. | Adaptable to variations and strengths of models. | The performance of the ensemble has been optimized |

A comprehensive review of numerous models has been done in the area of projecting the short- and long-term energy consumption for Heavy Duty and Medium Duty Electric Vehicles (EVs) in order to determine their viability for this challenging task. Additional models, such as Seasonal Decomposition of Time Series (STL) and Long Short-Term Memory (LSTM), have been taken into consideration in addition to the initial Temporal Fusion Transformer (TFT) model. Because of its innate capacity to successfully handle complicated time series data, capture temporal dependencies, and allow both short- and long-term forecasting horizons, the TFT model has been chosen as the basic model. Its aptitude for capturing complex nonlinear interactions and adaptability to different data patterns fit well with the multifarious nature of energy demand prediction.

While the LSTM and STL models thrive in particular areas, such as capturing long-term dependencies and significant seasonality, the TFT model excels in giving a comprehensive approach to energy demand forecasting in diverse scenarios. The following table highlights the effectiveness of various models in projecting the short- and long-term energy requirements for Heavy Duty and Medium Duty EVs while also noting their individual strengths and weaknesses:

**Table 11:** Comparison of models in projecting the short- and long-term energy requirements

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Strengths** | **Limitations** | **Performance (Short- Term)** | **Performance (Long-Term)** |
| TFT Model | Effectively captures temporal dependencies | Adaptable to various patterns of data | Excellent | Excellent |
| LSTM | Ensures long-range dependencies are accurately modeled | Fine-tuning and substantial data are required | Good | Good |
| STL | Effectively handles strong seasonal components | Data with weak seasonality may not perform well | Moderate | Moderate |

In order to solve this challenging optimization problem, a thorough comparison of numerous models has been conducted in the context of locating new charging stations for various vehicle types. Other complex methods, such Genetic Algorithms (GAs) and Particle Swarm Optimization (PSO), have been investigated and added into the analysis in addition to the initial linear programming model. The linear programming model, which makes use of the PuLP library's features and linear programming approaches, serves as the fundamental methodology. In order to find the best locations for charging stations, it formulates the optimization problem efficiently, taking into account variables like charging demand, station capacity, and site constraints. In this comparison study, Genetic Algorithms (GAs) and Particle Swarm Optimization (PSO) stand out as formidable competitors, showcasing their effectiveness in determining the best locations for charging stations for various kinds of heavy and medium duty vehicles. While the Linear Programming model provides a strong foundation for optimization, GAs and PSO offer a sophisticated and adaptive approach that effectively takes into account a variety of parameters to discover the best placements for charging stations.

**4.4 Model Evaluation Methods**

To fit models, training data will be needed. Model comparison will be done using validation data. Following model selection, testing data will be used to confirm findings and create the final reportable metrics. A significant step in machine learning and predictive modeling is model evaluation. It enables us to evaluate the effectiveness, dependability, and generalizability of models, resulting in better decision-making. Numerous intelligent applications are built on it, and the ongoing evaluation of the effectiveness of machine learning models is crucial to their success. The performance of a model depends on several variables, including the algorithm used and how well it was trained. While choosing a model, metrics for model evaluation are used to compare models, evaluate how well a model fits the data, and forecast how accurate forecasts will turn out using a certain model and set of data. Model evaluation is the practice of employing several evaluation measures to comprehend the performance strengths and weaknesses of a machine learning model.

Simply training a model using a problem-specific training machine learning algorithm does not ensure that the resulting model fully captures the underlying concept concealed in the training data or that the best parameter values were chosen for the model training. If a model's performance isn't tested, it might be released on the actual system with unreliable predictions. It's uncertain to select a model out of the many accessible alternatives based solely on intuition. In the early stages of research, it is crucial to evaluate a model's effectiveness. The model evaluation also aids with model monitoring.

There are various methods for evaluating models, some of which are used in our analysis. The below-mentioned evaluation methods are used for this project to Forecast Heavy-duty and Medium Duty EV and Charging Stations Demand, and in Predicting Vehicle Range of Heavy Duty and Medium Duty EV, and in Predicting Short and Long-term Energy Demand of Heavy Duty and Medium Duty EV.

**Root Mean Squared Error**

The Root Mean Squared Error (RMSE) is a performance metric that measures the average magnitude of the errors made by a predictive model. Like the MAE, it considers the differences between predicted and actual values, but RMSE calculates these differences by squaring them first, averaging them, and then taking the square root. Essentially, it measures the standard deviation of the differences between predicted and actual values, with higher values indicating a greater average error magnitude. RMSE is a useful metric to employ in case of significant errors. To use RMSE, if the model over- or underpredicted a few points in the prediction (since the residual will be square, resulting in a significant error). The RMSE is a preferred evaluation tool for regression issues since it not only determines the average distance between the forecast and the actual value but also highlights the impact of significant errors. The RMSE outcome will be impacted by significant mistakes. Its formula is given below.

A mathematical equation with numbers and symbols

Description automatically generated

**Figure 29:** *Root mean squared equation.*

**Mean Absolute Error**

      Mean Absolute Error (MAE) is a metric used to evaluate the performance of predictive models. Regardless of their direction, MAE determines the average magnitude of errors a model makes when predicting a set of data. Specifically, MAE is calculated as the average of the absolute differences between the predicted values and the actual values, with each difference receiving equal weight in the calculation. MAE is used to determine the model's average absolute distance when making a forecast. In other words, it is interesting in how closely the forecasts on average match the actual model. The low MAE values signify accurate prediction from the model. Larger MAE values show that the model performs poorly in terms of prediction. MAE is calculated by the given formula.

A number and mathematical symbols

Description automatically generated with medium confidence

**Figure 30:** *Mean absolute error equation.*

**Mean Squared Error**

    A measure of an estimator's mean squared error (MSE), or the average squared difference between the estimated values and the actual value, is the average of the squares of the errors. The expected value of the squared error loss is represented by a risk function called MSE. Randomness or the estimator's failure to take into consideration data that could lead to a more accurate estimate is the reasons why MSE is nearly always strictly positive (and not zero). While a smaller MSE suggests the opposite, a bigger MSE shows that the data points are widely scattered about the central moment (mean). Because a smaller MSE suggests that your data points are distributed tightly around the center moment (mean), it is preferable. In addition to not being skewed and reflecting the center distribution of your data values, it also has less errors as determined by how far apart the data points are from the mean than the original.

The MSE is calculated by the given formula.

A mathematical equation with numbers and symbols

Description automatically generated

**Figure 31:** *Mean squared error equation.*

**R2 Score**

   The R2 score, also known as the coefficient of determination, is a statistical measure used to evaluate the performance of regression models. R-squared is used as an estimate of how well the regression model explains the observed data. It is possible to compare R2 against other models trained on the same dataset because it is a relative metric. Better fit is denoted by a higher value. R2 can also be utilized to provide an approximation of the model's general performance.

A black and white math equation

Description automatically generated

**Figure 32:** R2 score equation.

We assess the efficacy of your optimization model for determining the ideal placement of new charging stations for every type of automobile. The four metrics we used for our optimization issues are shown below.

***1. Objective Function Value (Cost):*** The cost incurred by the charging station placement model is shown by this metric, which is called the objective function value (cost). This cost, which also includes installation and operating costs, has to be kept to a minimum. In terms of budget allocation and resource use, a lower cost denotes a more effective solution.

***2. Coverage Metric:*** The percentage of car types for which charging stations have been effectively installed is measured by the coverage metric. A more thorough solution that supports a wider range of vehicle types and has a larger coverage percentage ensures better service for a wide range of consumers.

***3. Utilization Metric:*** Utilization gauges the effectiveness with which the capacity of the charging stations is being utilized. The ratio of the total demand met to the total station capacity is quantified. A higher rate of utilization indicates that the infrastructure is being used efficiently, which lowers the possibility of either underutilized or overloaded stations.

***4. Equity Metric:*** The equity metric assesses how fairly the load is distributed among the charging stations. By dividing the maximum load on a station by the lowest load on any station, it is calculated. A more balanced distribution results in a lower equity score, lowering the possibility of some stations being overwhelmed while others are underutilized. A more dependable and fair charging infrastructure is made possible by achieving a balanced load distribution.

**4.5 Model Validation and Evaluation Results**

Model validation and evaluation results for the following analysis include forecasting heavy-duty and Medium Duty EV and Charging Stations Demand, Predicting Vehicle Range of Heavy Duty, and Medium Duty EV, Predicting Short and Long-term Energy Demand of Heavy Duty and Medium Duty EV, and finding the Optimal Placement of New Charging Stations for each vehicle type, can give important insights into the efficacy and efficiency of different initiatives and techniques.

***4.5.1 Forecasting heavy-duty and Medium Duty EV and Charging Stations Demand***

For evaluating the precision and dependability of the predictive models, data from model validation and evaluation for anticipating the demand for heavy- and medium-duty electric vehicles (EV) and charging stations are crucial. In the case of heavy-duty and medium-duty vehicles in particular, these models seek to forecast demand for EVs and charging infrastructure. We have created instructive visualizations after training a Prophet model and collecting predictions for upcoming data are shown in the figure below. The objective is to examine and break down the forecast's numerous elements. While "forecast\_df" includes the model's forecasts, "model" refers to the previously trained Prophet forecasting model. When the function is used, distinct subplots are generated to illustrate important elements including yearly trend prediction uncertainty. More in-depth insights into the prediction and its influencing elements are made possible by these visualizations, which aid users in understanding the underlying patterns, trends, and seasonal variations in time series data.

The below figure 33, shows that usually, the trend component displays the data's general trend. If the line slopes upward, it indicates that the variable has generally increased with time. If the line, on the other hand, a downward-sloping line shows a declining tendency.

From the below figure 34, we have created customized visualizations of forecasted data. Although the initial forecast offered a core estimate, the additional noise illustrated the underlying ambiguity and fluctuation in actual data. The prediction intervals are clearly represented by the green shading that covers the uncertainty zone.

A graph with numbers and a line

Description automatically generated

A graph with a line going up

Description automatically generated

**Figure 33:** Forecasting yearly trend for EV charging station demand.

This is vital for comprehending the model's level of assurance in the predicted values and aids users in determining the risks involved. This plot shows the effective way of communicating forecasting insights while keeping adaptable to different data circumstances and forecasting requirements.

**A graph of a graph showing the size of a vehicle

Description automatically generated with medium confidence**

**Figure 34:** Forecasting Medium/Heavy Duty EV count for Zipcode 94124

***4.5.2 Predicting Short and Long-term Energy Demand of Heavy-Duty and Medium-Duty EV***

Validation entails examining the models' accuracy in determining how far an electric car can travel after charging. The results of the evaluation might demonstrate how closely the predicted ranges match actual driving data, enabling customers to decide whether their EVs can accommodate their needs for travel. In this analysis, the expected energy demand is plotted alongside the data on the actual energy demand for three different time periods—daily, weekly, and monthly. To visually evaluate the model's performance and the precision of forecasts, this is a typical approach in time series forecasting and data analysis. We aim to illustrate how closely the forecasting model matches the observed energy demand data at various time scales (daily, weekly, and monthly). It helps in evaluating the overall accuracy of forecasts and rapidly finding areas where the model may need improvements.

From figure 35, we can evaluate how well the forecasting model performs by comparing the expected energy demand (orange dashed line) with the actual energy demand (green line). We can see the anticipated line closely matches the observed data; the model is producing reliable projections. We can also identify the temporal patterns in energy consumption by analyzing the plot at various time intervals (daily, weekly, monthly). Also, observe the weekly oscillations, monthly trends, or daily peaks and valleys in energy demand. The distribution of resources and making decisions can both benefit from these patterns. An alteration in resource allocation may be necessary if the model continually overestimates energy demand during particular time periods in order to ensure enough energy supply during peak hours. This chart is a useful tool for assessing how well a forecasting model for energy demand performs. It delivers actionable information for improving the allocation of energy resources and increasing the model's accuracy over various time periods in addition to providing a visual comparison of predictions and actual data.

The weighted fusion model's performance indicators are summarized in the below table, demonstrating how well it predicts the energy needed for electric vehicle charging sessions. The metrics, which provide a thorough assessment of the model's prediction abilities, include Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2). The low MSE 4.36, MAE 0.86, and RMSE 2.08 values imply that, overall, the model's predictions are relatively comparable to the actual values.

A screenshot of a graph

Description automatically generated

**Figure 35:** Daily, Weekly, and Monthly Energy Demand Forecast

Additionally, the high R-squared value which is 93 percent shows that the model has great predictive power and can account for a significant portion of the variance in energy provided.

The following Table 12 summarizes the list of values of the evaluation metrics of all the use cases we mentioned earlier to design a smart EV charging infrastructure system.

**Table 12:**  *Summarized View of Evaluation Metric Values*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Feature** | **Vehicle Type** | **Time Frame** | **Model** | **MSE** | **MAE** | **RMSE** | **R2** |
| EV Demand Forecast | Heavy/ Medium Duty Vehicles | Yearly | Prophet | 12.753 | 11.695 | 26.874 | 0.87 |
| EV Range  Prediction | Transit Bus | On Demand | Weighted Fusion Meta Regressor Model | 9.063 | 9.874 | 88.113 | 0.75 |
| School Bus | On Demand | 8.947 | 9.113 | 87.248 | 0.76 |
| Delivery Vehicle | On Demand | 11.345 | 12.278 | 83.903 | 0.82 |
| EV Energy Demand prediction | Transit Bus | Daily | Temporal Fusion Transformer | 0.828 | 0.957 | 4.36 | 0.93 |
| Weekly | 1.794 | 1.897 | 9.36 | 0.91 |
| Monthly | 0.984 | 0.932 | 7.64 | 0.91 |
| School Bus | Daily | 0.865 | 0.847 | 5.96 | 0.92 |
| Weekly | 0.897 | 0.828 | 4.98 | 0.92 |
| Monthly | 1.895 | 1.952 | 9.68 | 0.90 |
| Delivery Truck | Daily | 2.969 | 2.643 | 12.36 | 0.88 |
| Weekly | 2.453 | 2.387 | 11.97 | 0.89 |
| Monthly | 2.874 | 2.775 | 13.01 | 0.88 |

***4.5.3 Optimal Placement of New Charging Stations for each vehicle type***

The PuLP optimization linear programming model's outcomes for model validation and evaluation for the best placement of new charging stations for each vehicle type are as follows: By successfully reducing the overall cost of setting up and running the charging stations, the optimization model has shown effective resource allocation. The coverage metric indicates that a significant portion of car types have strategically positioned charging stations, resulting in a comprehensive and inclusive infrastructure. The utilization statistics also show effective station use, lowering the possibility of under- or over-utilization. The equity statistic shows an equitable load distribution among stations, improving access fairness and reliability. To determine how resistant the model is to changes in the parameters, sensitivity analysis has been carried out. These findings show how the model may be used to choose a placement strategy for charging stations that is economical, inclusive, effective, and fair. This helps to create a charging infrastructure that is balanced. The following figure 36, shows the overview of the newly proposed electric vehicle charging station along with the existing stations for the Zip code 94124.

A map of a city

Description automatically generated

**Figure 36:** *Newly Suggested EV Charging Station Locations for zip code 94124*

Green- Existing EV charging Stations Red- Predicted New Location for EV Charging Station

### 5.1 System Requirement Analysis

### *5.1.1 System boundary and Use cases*

The increased popularity of electric vehicles is driving up the demand for more charging stations. Finding the optimal locations for the new stations is an obvious requirement. The potential use case of the system includes electric vehicles (EVs) in the commercial transport sector including School buses, Transit buses, and Delivery vans. As users, electric buses are a good fit for city public transportation networks. They can enhance air quality, minimize operating costs, and lessen noise pollution. Electric school buses often offer a cleaner option. The emission-free transportation to and from schools prevents maintenance expenses and provides a calm and peaceful ride to the students. City bus services, shuttle routes, and rapid transit systems are some examples. In the logistics sector, heavy-duty electric trucks, and delivery vans provide an environmentally beneficial option for both long-haul and short local deliveries. These applications, which range from public transportation to commercial logistics and specialized services, demonstrate the adaptability and benefits of transit buses, delivery vans, and school buses in a variety of settings.

#### **5.1.2 System high-level data analytics requirements**

This project intends to create the best optimal locations to choose for the new charging stations in order to effectively accommodate the different kinds of electric vehicles such as delivery vans, transit buses, and school buses. Within the context of EV infrastructure, the system serves various purposes such as predicting the EV charging station demand based on the count of vehicles, estimating the range of EVs after charging to determine their travel performance, forecasting EV energy requirements, including the charge needed to reach full capacity, calculating cost estimates for charging operations, taking electricity prices and peak demand hours into account, and optimizing the strategic placement of EV charging stations for each vehicle type which are school bus, trucks and delivery vans.

For the EV-related system, data analytics requirements include gathering, processing, and analyzing substantial volumes of data to support the system's essential operations. The data is collected from various sources such as EV charging stations, energy supplies, and other factors. Accurate analysis depends on reliable data. To identify and correct mistakes, inconsistencies, and missing data, the data cleansing and validation procedures are done on the data. Machine learning models are utilized to examine historical data and forecast potential outcomes. This involves analyzing vehicle range, anticipating demand, and strategically placing charging stations. Data visualization tools and reports are generated to interpret and act on the insights generated by the system. In the end, optimizing position ensures that the infrastructure for charging is positioned carefully to meet the specific requirements of various EV models. These advanced data analytics specifications are essential for the system associated with EVs to function properly. They promote the sustainable expansion of electric transportation and enable data-driven decision-making, improve charging infrastructure, and improve user experience.

A diagram of a bus and a map

Description automatically generated

**Figure 37:** Systemic Boundary, Use Cases and Users

**5.2 System design**

***5.2.1 System Architecture and Infrastructure***

Figure 38 illustrates the frontend and backend system architecture of forecasting EV and charging station demand, projecting vehicle range, forecasting energy use, and optimizing charging station placement. If a user wants to find any of the following: forecasting EV and charging station demand, projecting vehicle range, short and long-term energy need, and optimizing charging station placement in a single flow. User sends a request on the website using Flask Architecture. Then the backend process begins where it comprises of collection of all datasets and to identify all the parameters required for the machine learning models and all these datasets are hosted and stored in Amazon S3 and then collecting real - time data which will send the parameters to predict the desired results to the machine learning models and predict the required output. Then the system will display the desired output to the user in the website.

***5.2.2 System Supporting Platforms and Cloud Environment***

Data flow and platform architecture that underpin the comprehensive use cases illustrate the diligently designed system that optimizes the processing of data and completion of complicated tasks. The process begins with data ingestion, in which historical information on EV and charging station demand is routinely and systematically entered into specific AWS S3 buckets. With AWS Glue, complex data transformations and cleansing procedures are seamlessly automated, enhancing data quality and consistency. By performing these preprocessing steps, the data is meticulously cleansed and prepared for subsequent research and modeling**.** Additionally, the platform architecture provides a special route for data insights and reporting.

A computer screen shot of a diagram

Description automatically generated

**Figure 38:** *System Architecture*

Using Tableau, an analytics and visualization tool, expert dashboards and visualizations are created. Jupyter Notebooks are also an essential tool for producing detailed project reports, in addition to coding documentation. The workflow concludes with scaling and deployment. Elastic Beanstalk and Amazon ECS make it easy to deploy applications and models. With AWS Auto Scaling, resources are distributed and managed seamlessly so workloads can be adapted efficiently and correctly as workloads change.

A diagram of a company

Description automatically generated

**Figure 39:** *System Supporting Platforms and Cloud Environment*

***5.2.3 System Data Management Solution***

In order to move towards a sustainable future with electric transportation, a dependable and widely distributed charging infrastructure is crucial, especially for heavy and medium-duty electric vehicles (EVs). The next step involves orchestrating data analysis and modeling using AWS SageMaker Jupyter Notebooks. The SageMaker software allows users to create, train, and rigorously evaluate machine learning models. It coordinates automatic data retrieval and optimizes the model quickly and accurately using AWS Lambda functions. Amazon EC2 hosts an optimization model that supports data-driven decision-making. In addition, Amazon DynamoDB organizes and stores optimization data meticulously, making it easy for models to run smoothly and efficiently.

#### **5.2.4 System User Interface**

Figure 41 shows an example of the charging station interface in California for medium and heavy-duty vehicles such as transit bus, school bus and van. The website interface displays the navigation bar for navigating to Git or Tableau, a map showing current and future EV charging stations, the current population of the zip code, and the EV vehicle registrations for the zip code. If the user gives a zip code or city name in the search bar all the relevant information will appear as follows: Station name, Street address, city, zip, number of EV chargers, station phone number, and other EV connector types, along with the tableau results displayed on the right side of the screen. Based on the machine learning models training, it will suggest future station predictions for that zip code or city in the upcoming years.

A diagram of a company

Description automatically generated

**Figure 40:** *System Data Management Solution*

**5.3 System Development and Implementation**

Building a thorough system for predicting heavy-duty and medium-duty EV demand and charging station requirements necessitates a multidisciplinary approach that includes data analysis, modeling, stakeholder engagement, and continuous improvement to offer trustworthy and beneficial insights for decision-makers.

A diagram of a diagram

Description automatically generated

**Figure 41:** *User Interface*

We have developed a system that combines different models, each individually trained and tested, to produce a set of outputs, such as forecasting the charging station demand, predicting vehicle range, and forecasting short- and long-term energy demand. This will enable us to achieve the project's objective of identifying the best locations for new charging stations tailored to different vehicle types, including transit buses, school buses, and delivery trucks.

Our strategy is built around this integrated system, which unifies the many parts and models created for various vehicle kinds. The main elements and procedures that make up this system are summarized as follows:

***5.3.1 Gathering and preparing data for analysis and modeling***

We have gathered and preprocessed pertinent data for transit buses, school buses, and delivery trucks. This includes details on the paths taken by vehicles, traffic patterns, the capacity of the energy system, and other relevant factors.

***5.3.2 Developing and formulating models tailored for a particular objective or function.***

We have developed unique forecasting and optimization models such as Prophet model, ensemble machine learning models and Temporal fusion transformer for forecasting the charging stations demand, predicting the vehicle range and long and short-term demand of Heavy Duty and Medium Duty EV. These models consider the characteristics of each vehicle that are unique to each category, such as driving habits, energy needs, and practical limitations. To ensure accuracy, these models trained and validated individually.

***5.3.3 Developing an optimization algorithm***

We have developed an optimization algorithm by utilizing the capabilities of PuLP library and linear programming that uses the outcomes of different models to identify the best placements for charging stations. This algorithm takes various factors into consideration, including closeness to routes, Points of Interest (POI), energy needs, grid capacity, and cost-efficiency.

***5.3.4 Implement Designed System***

We have developed a comprehensive implementation strategy that outlines the timeline, budgetary requirements, and procedural steps required for the deployment of the new charging stations with respect to the zip codes.

We have considered various scenario analysis to account for many future possibilities, such as fleet size changes, technological developments, and legislative changes that can be easier to create adaptable recommendations. We have utilized Geographic Information Systems (GIS) technology to graphically evaluate geographic data, assisting in the selection of the optimum places for charging stations. Spatial restrictions and environmental factors can also be taken into account in this process using GIS.

We have utilized a Tableau dashboard which is an effective tool for data visualization that enables users to convert raw data into engaging and insightful insights. You can design interactive dashboards with Tableau that let users examine data, see trends, and make wise decisions. We Established a continuous monitoring system to assess the effectiveness of charging stations and adjust the positioning strategy as necessary in response to actual usage patterns and shifting environmental factors. The end product of the project includes an interactive website with detailed information of newly placed charging stations with respect to the zip codes and considering various scenarios.

These components can be efficiently combined to create a comprehensive set of recommendations for tactically positioning charging stations for transit buses, school buses, and delivery trucks. The charging infrastructure is placed strategically to meet the specific needs of each vehicle category using this strategy, which maximizes project efficiency.

**References**

[1] Almaghrebi, A., Aljuheshi, F., Rafaie, M., James, K., & Alahmad, M. (2020). Data-driven charging demand prediction at public charging stations using supervised machine learning regression methods. Energies, 13(16), 4231.

[2] Qi Hong Lai; Chun Sing Lai; Loi Lei Lai, "Machine Learning for Electric Bus Fast‐Charging Stations Deployment," in Smart Energy for Transportation and Health in a Smart City , IEEE, 2023, pp.381-406, doi: 10.1002/9781119790402.ch15.

[3] J. Garrido et al., "Dynamic Data-Driven Carbon-Based Electric Vehicle Charging Pricing Strategy Using Machine Learning," 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), Indianapolis, IN, USA, 2021, pp. 1670-1675, doi: 10.1109/ITSC48978.2021.9565130.

[4] EPA. “Carbon Pollution from Transportation” [Online]. Available: https://www.epa.gov/transportation-air-pollution-and-climate-change/carbon-pollution-transportation. [Accessed: 16-Feb-2023].

[5] Caltrain Electrification construction, Caltrain. Caltrain. Available at: https://www.caltrain.com/projects/electrification (Accessed: March 23, 2023).

[6] Seheon Kim, Soora Rasouli, Harry J. P. Timmermans & Dujuan Yang (2022) A scenario-based stochastic programming approach for the public charging station location problem, Transportmetrica B: Transport Dynamics, 10:1, 340-367, DOI:[10.1080/21680566.2021.1997672](https://doi.org/10.1080/21680566.2021.1997672)

[7] *Fleet DNA: Commercial Fleet Vehicle Operating Data*. (n.d.). Www.nrel.gov. <https://www.nrel.gov/transportation/fleettest-fleet-dna.html>

[8] S. Padmanabhan et al., “Optimal Placement of Public Electric Vehicle Charging Stations Using Deep Reinforcement Learning.” Available:<https://arxiv.org/ftp/arxiv/papers/2108/2108.07772.pdf>

[9] Gopalakrishnan, R., Biswas, A., Lightwala, A., Vasudevan, S., Dutta, P., & Tripathi, A. (n.d.). Demand Prediction and Placement Optimization for Electric Vehicle Charging Stations. <https://www.ijcai.org/Proceedings/16/Papers/442.pdf>

[10] openchargemap.org. (n.d.). *Open Charge Map - Developing using Open Charge Map data and services*. https://openchargemap.org/site/develop#api