Predicting Charging Patterns of Electric Vehicles

Joseph Lozada, Leonardo Pinheiro, Christopher Thawley

Fall 2023

Abstract

We used a published dataset to predict patterns in electric vehicle charging, including energy usage and number of cars charging. We used a set of three different algorithms and a feature set including temporal, meteorological, and traffic variables to build models. We assessed model performance and feature importance, and, for the best models, predicted EV charging under different scenarios. We found that EV charging does vary strongly based on time of day, weather, and traffic and that these features should be incorporated in future models to predict charging patterns.

1 Introduction and Literature Review

Human-induced climate change is a serious threat to both the human species and the remainder of Earth's biodiversity. The primary driver of global warming is emissions of carbon dioxide from the burning of fossil fuels [4]. Vehicles with internal combustion engines which use gasoline or related compounds are one of the major sources of carbon dioxide emissions. Therefore, serious efforts to combat global warming rely on strategies that reduce these emissions [2]. One of the principal pathways to reducing these emissions from the transportation sector is via the use of electric vehicles (EVs) which are powered by rechargeable batteries. If the electricity used to charge these batteries is produced more cleanly (either in a more centralized, controlled manner) or via renewable (ie, non-fossil fuel) sources, then carbon dioxide emissions will be reduced, potentially be 10-12 times current production [11].

Electric vehicle adoption is increasing rapidly, and with this increase, concomitant changes in both the power grid and user behavior are needed [6]. Electric vehicle charging capacities can be quite high, and adding car-charging to household or business usage can cause major changes in overall patterns of electricity usage, including increasing overall demand as well as changes in temporal patterns of demand [3]; [8]. To meet both consumer demand and climate change goals, utilities that deliver electricity will need to be able to predict increases in demand over the long term, but also on finer (i.e. hourly) scales, to ensure the stability of the electrical grid [6].

Companies may also be able to optimize end user charging in other ways, such as offering discounted rates during off-peak hours to encourage users to spread car charging load into periods with lower overall demand [8]. "Smart" chargers can also determine lower-demand or cheaper energy periods (such as the middle of the night) and initiate charging during those times.

From a consumer perspective, optimizing individual charging behavior also requires knowledge about charging patterns among other users and availability of charging facilities. For instance, charging an EV on a standard home connection may take a long period of time (up to 2 days) which may not be feasible for many vehicles uses. Upgraded home electrical systems can support much quicker charging comparable to some commercial chargers, but these systems may not be available in some homes or may require costly upgrades (up to $5,000). Because of this, many EV drivers make use of charging stations at work, on the road (such as at interstate rest stops), or in communal spaces such as shopping or apartment complexes [7]. The number of available chargers, however, is generally less than the number of EV drivers. If drivers can understand general charging patterns, and how they vary with time and other factors, they can make more efficient decisions about when and where to charge [5]. Ultimately optimizing predictions for and behavior of both utilities and EV end users will make an EV charging system more efficient, increase EV adoption, and help meet society's goals for reducing carbon dioxide emissions.

We aimed to create a small set of models which predict electric car charging patterns (total energy usage/charging load and number of cars charging) based on readily available features including time period (hourly and daily), traffic, and weather. These models could allow users to make predictions about future car charging patterns as EV usage increases and utilities to make more informed decisions about how power demand will vary in the future. We created several specific scenarios and predicted charging load and number of cars charging under these to demonstrate how charging patterns vary depending on conditions. We also examined which features were most important in driving patterns of charging which can help future modelers identify features that will be most useful to include in models (and those which might be able to be ignored). Similar studies have been of interest for at least 10 years, e.g. [9], but electric vehicle capabilities and usage are changing rapidly, so continuing studies with current data are valuable.

We obtained a published dataset [10] to use in building our models. As a whole, the dataset describes charging activity for electric vehicles in a residential apartment complex in Trondheim, Norway for a 14-month period from 2018 to 2020. The data include instances of $6,878$ charging sessions by almost 100 ($n = 97$) individuals using both private and shared/communal chargers, with charging times, duration, and energy used. Additional files provide processed data on the hourly loads and unused charging capacity for the charging stations under different charging rates, and hourly data for local traffic density at five locations near the apartment complex where charging sessions occurred during the same time period during which charging sessions occurred as well as other data which we did not use. We also accessed data from the Norwegian Meteorological Institute (MET Norway) on 12/5/23, and use these data under a CC BY 4.0 license. These data were from three weather stations (Gloshaugen, Trondheim Voll, and Trondheim Voll Plu) and included ten variables: air temperature, precipitation in gauge, precipitation (1 hour), mean wind speed, max mean wind speed (1 hour), maximum wind gust (1 hour), maximum air temp, minimum air temp (1 hour), snow depth, and sunshine duration (1 hour).

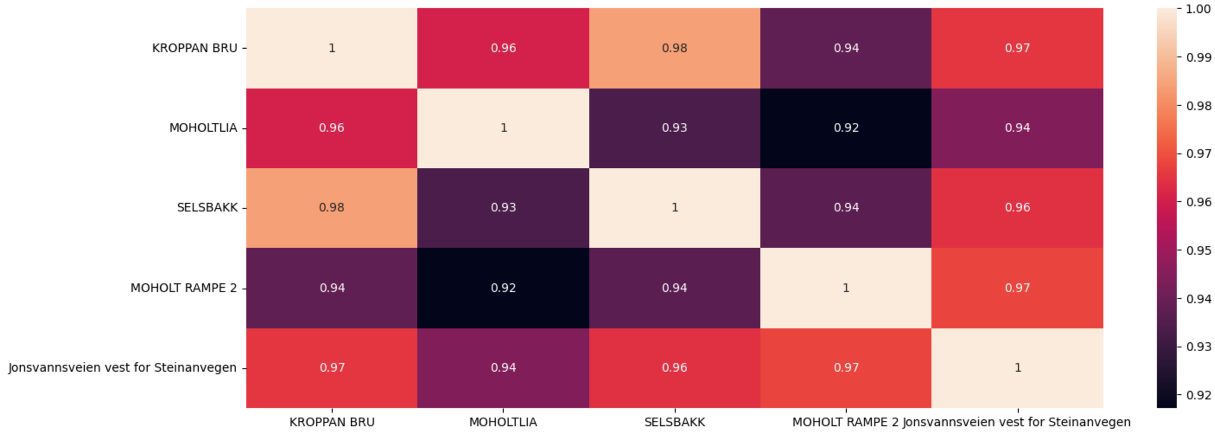

Figure 1: Correlation matrix for number of cars per hour at five intersections within 3 km of the focal apartment complex location.

2 Methods and Results

2.1 Preprocessing data

Prior to importing the data into Python, we conducted initial datawrangling using Excel. This included converting the data from European formatting (commas instead of periods), and calculating estimated charging loads for each time period by summing the values for shared and private chargers in the provided dataset. We also estimated the number of cars charging during each hourly period and rounded this to a whole number.

Prior to conducting analysis, we examined the provided traffic data and the separate weather datasets that we had downloaded. Our goals were to determine if any variables were strongly correlated, and, if so, to reduce the feature sets to aid both interpretation and efficiency of fitting models. Analysis of the traffic data showed that traffic at all five intersections was strongly correlated (Figure 1), with all correlations being greater than or equal to 0.92. As such, we used the mean value of traffic for each hour in our set of predictors.

We conducted a similar procedure with our meteorological data. We found that all variables of related measurements for temperature and wind were highly correlated.For temperature, all four variables were strongly positively correlated with $r \geq 0.98$, so we chose to use only hourly air temperature in our final feature set. For three variables measuring different aspects of wind speed, we found that all were strongly positively correlated with $r \geq 0.97$, so we chose to use only mean maximum wind speed for each 1 hour period in our final feature set.

We trimmed all data to a one year period from 1/31/19 to 1/30/20. This ensured that all times of year were equally represented to avoid bias due to seasonality. Prior to use in modeling, we standardized all numerical features to have a minimum value of 0 and a maximum value of 1 to improve model fitting. We also condensed a categorical feature, day of the week, to three categories: weekday (including Mondays-Fridays), Saturday, and Sunday to assess if charging patterns varied on weekdays or weekends.

We use Random Forest, Ada Boost, and Gradient Boosting methods to predict the total number of cars charging and the total load in kWh on the grid at any given hour of the day. We perform a grid search to tune parameters and then examine the feature importance for each method. All of our analysis was performed in Python with Pandas used for data manipulation and the machine learning algorithms coming from Scikit-Learn.

2.2 Random Forest

A random forest is an ensemble method that uses decision trees as weak learners. The algorithm trains each tree in the forest utilizing a random sample of the training data and a random subset of the features. This process allows the trees in the forest to be uncorrelated and the final prediction is then the mean value of the values predicted by the individual trees.

After tuning hyperparameters to predict total hourly load, we selected a random forest with 15 trees, each having maximum depth of 15. The optimal complexity parameter (c_α) used for minimal cost-complexity pruning was equal to 0.01.

We see that the random forest model returned R-squared values or $R^2_{\text{train}} = 0.79$ and $R^2_{\text{test}} = 0.60$ and root mean square deviations of $RMSE_{\text{tr ain}} = 5.28$ and $RMSE_{\text{test}} = 7.11$ respectively. This suggest some amount of overfitting. These values, are summarized in Table 1.

	$R2$	$RMSE$
Train	0.79	5.28
Test	0.60	7.11

Table 1: Random Forest
R^2 and $RMSE$ - Total Houry Load

Figure 2 shows us the feature importance for the load prediction. We see that the time of day (daily_hour), the depth of snow (Snow_depth) and air temperature (Air_temp) were the most important predictors for total load in our model. We were a little surprised to see that whether charging occurs on a weekday, Saturday, or Sundays has little importance for the model.

The results for the number of cars charging at any give hour were unsurprisingly very similar. The best model fitted used a forest with 15 trees, each having maximum depth of 10. The optimal complexity parameter (c_α) used for minimal cost-complexity pruning was equal to zero. The results are sumarized in table 2.

	$R2$	$RMSE$
Train	0.79	1.45
Test	0.61	1.96

Table 2: Random Forest
R^2 and $RMSE$ - Total Number of Cars

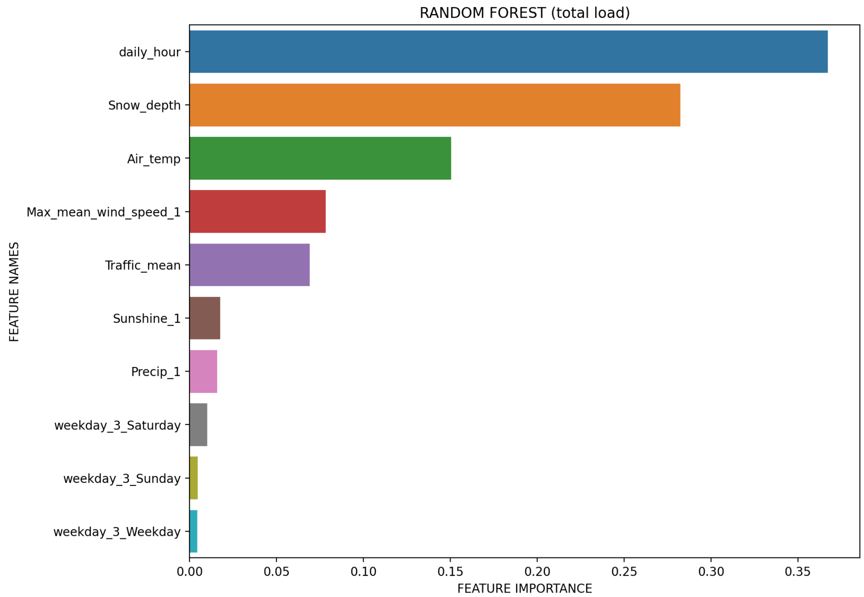

Figure 2: Random Forest - Total Load

2.3 AdaBoost

AdaBoost (*Adaptive Boosting*) is an ensemble learning technique that uses shallow trees as weak learners. The algorithm trains and deploys trees in a sequence, focusing on the mistakes made by the previous weak models and assigning more weight to those instances ("boosting") in subsequent iterations.

After tuning hyperparameters for predicting total load we selected a model with a learning rate equal to 0.01 maximum number of parameters before stopping boosting equal to 15.

We see that the AdaBoost model returned R-squared values or $R^2_{\text{train}} = 0.46$ and $R^2_{\text{test}} = 0.44$ and root mean square deviations of $RMSE_{\text{train}} = 8.43$ and $RMSE_{\text{test}} = 8.78$ respectively. This suggest some amount of overfitting. These values, are summarized in Table 3.

	$R2$	$RMSE$
Train	0.46	8.43
Test	0.44	8.78

Table 3: AdaBoost R^2 and $RMSE$
Total Load

Figure 3 shows us the feature importance for the load prediction. We see that the time of day (daily_hour), the depth of snow (Snow_depth) and air temperature (Air_temp) were essentially the only important predictors for our model.

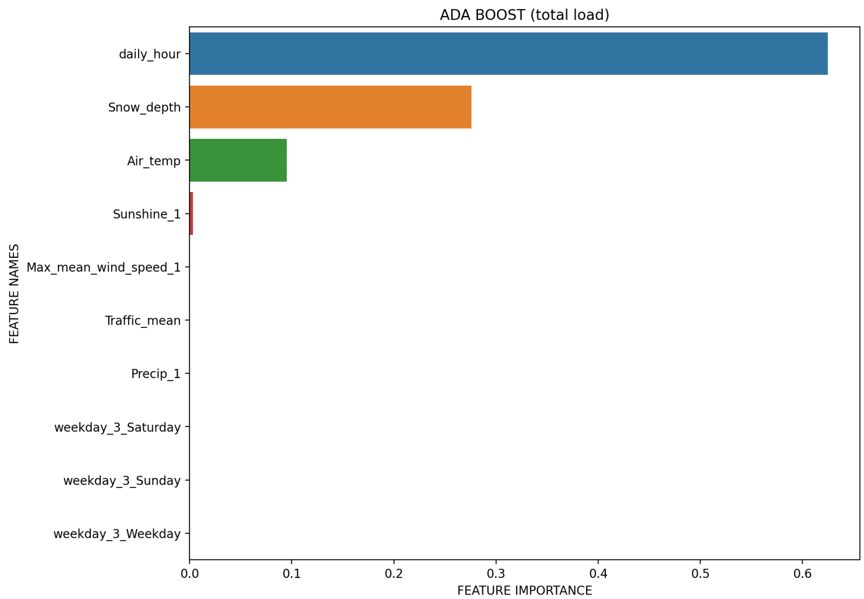

Figure 3: AdaBoost - Total Load

Similar results were obtained for the modeling of the total number of cars charging. Results are summarized in 4, feature importance for total cars was very similar to that of total load.

	R^2	$RMSE$
Train	0.79	1.45
Test	0.61	1.96

Table 4: AdaBoost
R^2 and $RMSE$ - Total Number of Cars

2.4 Gradient Boosting

Gradient Boosting is an ensemble learning technique that similar to Ada Boost. The algorithm trains and deploys trees in a sequence but instead of adjusting weights at every interaction, this method attempts to fit the new predictor to the residual errors made by the previous predictor.

After tuning hyperparameters for predicting total load we selected a model with a learning rate equal to 0.01 maximum number of parameters before stopping boosting equal to 15.

We see that the Gradient Boosting model returned R-squared values or $R^2_{\text{train}} = 0.46$ and $R^2_{\text{test}} = 0.44$ and root mean square deviations of $RMSE_{\text{train}} = 8.43$ and $RMSE_{\text{test}} = 8.78$ respectively. This suggest some amount of overfitting. These values, are summarized in Table 5.

	$R2$	$RMSE$
Train	0.71	6.01
Test	0.53	7.77

Table 5: Gradient Boosting R^2 and $RMSE$
Total Load

Figure 4 shows us the feature importance for the load prediction. We see that this model had more variability in future importance with the time of day (daily_hour), the depth of snow (Snow_depth), air temperature (Air_temp), and traffic (traffic_mean) being the four most important predictors for the model.

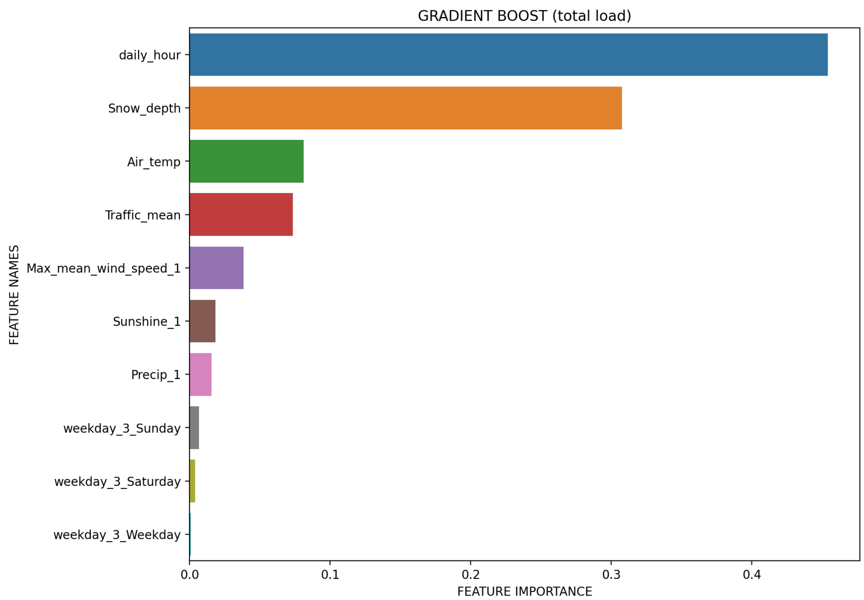

Figure 4: Gradient Boosting- Total Load

Similar results were obtained for the modeling of the total number of cars charging. Results are summarized in Table 6, feature importance for total cars was very similar to that of total load.

	$R2$	$RMSE$
Train	0.72	1.68
Test	0.53	2.14

Table 6: Gradient Boost - R^2 and $RMSE$
Total Number of Cars

We also evaluated time to run for the subset of our best models. All models ran very quickly, in under 0.018 s. This was likely due to our relatively small and low-dimension dataset with fewer than 10,000 observations and 8 features. Models did vary in time to run with Gradient Boosting being quickest (mean: 0.008 s), Random Forest intermediate (mean: 0.010 s) and AdaBoost longest (mean: 0.018 s). Given how quickly the models ran, however, these differences are of little practical importance. If more complex models or much larger training sets are used in future analyses, these differences could become more important.

3 Predicting Charging Scenarios

After assessing the performance of the three different models, as well as their key features, we proceeded to further investigate the use-case of the models by testing them with hypothetical future scenarios. As previously noted, there are two primary purposes for these models; first, to aid utilities in predicting hourly charging load to the electric system, and second, to aid EV users in predicting charging station activity and identifying ideal times for users to charge their cars.

As the key features identified were daily hour, snow depth, air temperature, wind speed and mean traffic rate, we used these features to set up the scenarios. Three scenarios were created: cold weather, warm weather, and heavy traffic. The data for each scenario were set up using a 24-hour cycle so that the full range of values present in the daily hour feature (0–23) were included. In the original dataset the following statistics were considered for the other key features: snow depth ranged from 0 cm to 39 cm with a mean of 3.49 cm, air temperature ranged from -13^oC to 31.3^oC with a mean of 6.27^oC, wind speed ranged from 0.3 km/h to 14 km/h with a mean of 3.23 km/h, and mean traffic rate ranged from 12.6 cars per hour to 2472.4 cars per hour with a mean of 699.6 cars per hour.

The cold weather scenario consisted of the following: average traffic conditions (cars per hour $= 699.6$), snow depth of 45 cm, wind speed of 20 km/h, and air temperature of -15^oC. This scenario was designed to simulate snowy, windy winter weather conditions. The warm weather scenario consisted of the following: average traffic conditions (cars per hour = 699.6), snow depth of 0 cm, wind speed of 1 km/h, and air temperature of 30^oC. This warm weather scenario was designed to simulate a pleasant, sunny, summer day. The high traffic scenario consisted of the following: average weather conditions (snow depth = 3.49 cm, wind speed = 3.23 km/h, air temperature = 6.27^oC) and a traffic rate of $3,000$ cars per hour. These data were transformed using the scaler fitted to the training set, and all three models were run to predict both targets: total load, and number of cars charging. The results the models produced were consistent with our expectations based on our initial testing analysis. The Random Forest and Ada Boost models were more closely in consensus while the Gradient Boosting model had greater variability.

The predictions for the number of cars charging were as follows. In the cold weather scenario, Ada Boost and Random Forest nearly identically predicted approximately 4cars charging at the start of the day (0 hours), dropping to 1 car charging at 4:00 up until 15:00 after which the models diverged with Random Forest predicting a greater number of cars charging. The Gradient Boosting model predicted a greater number of cars (11) charging at the start of the day and then dropped throughout the day, ending at 5 cars charging; a final value in between the two other models. Gradient Boosting started with a much higher prediction and finished the day in greater consensus with the other models.

The warm weather scenario again resulted in consensus between AdaBoost and Random Forest with approximately 3 cars charging at the start of the day and between 2 and 3 cars charging at the end of the day: mostly flat with little variation. The Gradient Boosting model varied considerably from the other two. It began with 11 cars charging at the start of the day, dropping to 2 cars charging at 13:00 and then spiking to 10 cars charging at 22:00. The heavy traffic scenario also resulted in AdaBoost and Random Forest and consensus with approximately 4 cars charging at the start of the day and 8 cars charging at the end of the day. Gradient Boosting also increased throughout the day but had higher overall predictions starting at 12 cars charging and ending at 25 cars charging. Overall, the models seem to suggest that greater charging activity will occur in cold weather scenarios vs warm weather scenarios, but the greatest impact to the number of cars charging are traffic conditions.

The predictions for total charge also contained variation between the models. The cold weather scenario resulted in consensus between AdaBoost and Random Forest from the morning until 15:00, then the models diverged with Random Forest predicting a charge of 30 and AdaBoost predicting a charge of 10. The Gradient Boosting model had higher predictions overall, starting at 3 and peaking at 57 at 21:00. The warm weather scenario had considerable consensus among all three models, although Gradient Boosting still did have the most variability. All three models began high at the start, dropped in the morning till 15:00 then went back up creating a "U" shape with total charge ending the day between 8 and 11 among all three models. Random Forest started the highest at 14 while Gradient Boosting started the lowest at 5. The heavy traffic scenario resulted in consensus between Ada Boost and Random Forest. Both models started the day at around 15 and ended at 30. As in many cases before, the Gradient Boosting model had high predictions overall starting the day at 25 and ending at 68. In these three scenarios the Gradient Boosting predictions were much more variable than the other two models. All three models and scenarios, however, showed lower charging during the day with increased charging in the evening. Similar to the predictions for number of cars, the charging load predictions suggest that heavy traffic can drive higher numbers than weather. See figure 5.

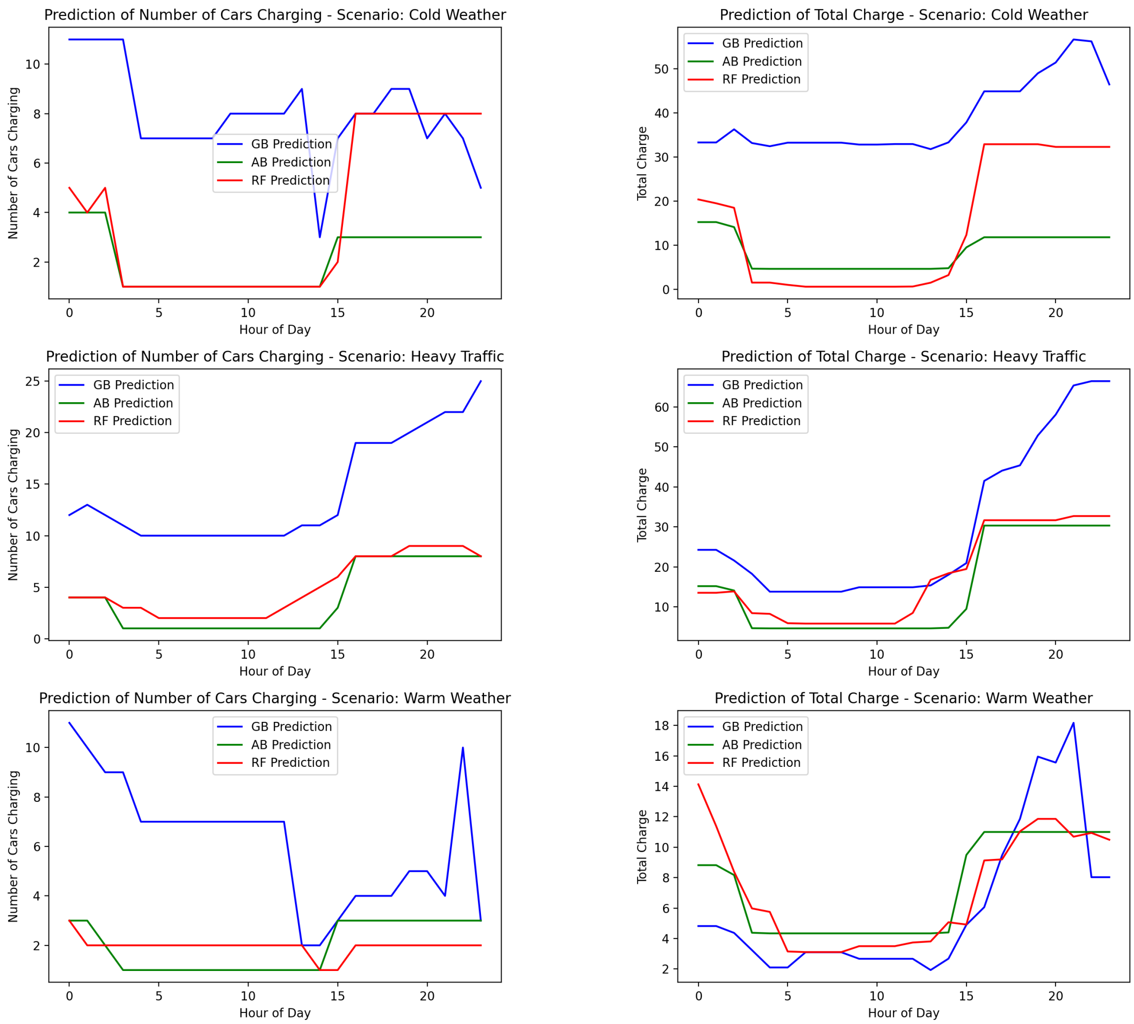

Figure 5: Predictions

4 Discussion

Through this analysis, we have shown that charging patterns vary strongly over the course of the day and can be highly dependent on a handful of key variables. Weather and traffic (to a lesser extent) influence car charging patterns, and if one were to model EV charging in the future, those variables should be included. One enhancement which should be considered for future modeling would be the incorporation of months into the dataset. In our analysis this was not done for two reasons. First, the number of cars increased over time in a trend likely indicative of increasing EV adoption, so including month as a feature would have reflected this temporal bias. Second, the data only consisted of 14 months of charging information therefore identifying patterns attributable to specific months of the year would have been difficult with such a limited set. In a future model, sampling over a greater period, perhaps 5 years, would help avoid some of these potential issues. We could also consider explicitly addressing temporal autocorrelation in our models, though again, we would need data from a longer time series to do this effectively.

Our models also showed clear differences in performance based on the algorithm used. The best models, produced by the Random Forest algorithm, explained about 60% of the variation; a good start, but certainly improvements can be made. Models produced by Gradient Boosting showed higher variability, potentially indicating that they should not be used for detailed predictions, and suggesting they may be overfit. As a whole, our models showed higher performance on training as opposed to test data, indicating that overfitting may be a more general concern. In future studies, simpler models could be used which might help alleviate this.

References

[1] Sørensen, Åse Lekang (2021), "Residential electric vehicle charging datasets from apartment buildings", Mendeley Data, V2, doi: 10.17632/jbks2rcwyj.2,

[2] Aresta, M., & Dibenedetto, A. (2021). The Carbon Dioxide Revolution. Springer, 262 pgs.

[3] Burlig, Fiona, James Bushnell, David Rapson, and Catherine Wolfram. 2021. "Low Energy: Estimating Electric Vehicle Electricity Use." AEA Papers and Proceedings, 111: 430-35. DOI: 10.1257/pandp.20211088

[4] IPCC, 2023: Summary for Policymakers. In: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, pp. 1-34, doi: 10.59327/IPCC/AR6-9789291691647.001

[5] Jin, C., Tang, J., & Ghosh, P. (2013). Optimizing electric vehicle charging: A customer's perspective. IEEE Transactions on vehicular technology, 62(7), 2919-2927. DOI: 10.1109/TSG.2012.2218834

[6] Kapustin, N. O., & Grushevenko, D. A. (2020). Long-term electric vehicles outlook and their potential impact on electric grid. Energy Policy, 137, 111103. https://doi.org/10.1016/j.enpol.2019.111103

[7] Neubauer, J., & Wood, E. (2014). The impact of range anxiety and home, workplace, and public charging infrastructure on simulated battery electric vehicle lifetime utility. Journal of Power Sources, 257, 12-20. https://doi.org/10.1016/j.jpowsour.2014.01.075

[8] Nour, M., Chaves-Ávila, J. P., Magdy, G., & Sánchez-Miralles, Á. (2020). Review of positive and negative impacts of electric vehicles charging on electric power systems. Energies, 13(18), 4675. https://doi.org/10.3390/en13184675

[9] Qian, K., Zhou, C., Allan, M., & Yuan, Y. (2010). Load model for prediction of electric vehicle charging demand. In 2010 International Conference on Power System Technology (pp. 1-6). IEEE. DOI: 10.1109/POWERCON.2010.5666587

[10] Sørensen, Å. L., Lindberg, K. B., Sartori, I., & Andresen, I. (2021). Residential electric vehicle charging datasets from apartment buildings. Data in Brief, 36, 107105. https://doi.org/10.1016/j.dib.2021.107105

[11] Teixeira, A. C. R., & Sodré, J. R. (2018). Impacts of replacement of engine powered vehicles by electric vehicles on energy consumption and CO2 emissions. Transportation Research Part D: Transport and Environment, 59, 375-384. https://doi.org/10.1016/j.trd.2018.01.004