Optimal Placement of New Electric Vehicle Chargers in the City of Los Angeles

Amber Guo
University of Southern California
amberguo@usc.edu

Daniel Rabinovich
University of Southern California
drabinov@usc.edu

Dhivya Senthil
University of Southern California
senthilm@usc.edu

Abstract

The City of Los Angeles has set goals to increase the percentage of zero-emission vehicles in the city-wide fleet to 25% by 2025 and 100% by 2050. To advance this goal, the City plans to install 150 new electric vehicle (EV) charging stations in 2021. This project utilizes census data, Google and Bing Map applications, and data published by the state of California in a linear regression model to determine which features of current EV charging stations and the neighborhoods that they are in most accurately predict the usage patterns for a particular station. Subsequently, we will apply our understanding of those predictive characteristics to a greedy max cover algorithm such as CELF to determine the optimal placement of new EV charging stations to maximize results including total power delivered, total charging time, and the number of charging sessions.

1 Introduction

Electric Vehicles (EVs) have become popular over the past decade. In California, there are over 250,000 EVs registered as of August 2020, which accounts for approximately 47% of the overall EV population [1]. The enthusiasm for EVs is unlikely to diminish, as statistics suggest a persistent increase in EV sales in the past four years [2]. Unlike traditional gasoline vehicles, EVs utilize electricity and thus contribute to the relief of environmental pressure by lowering carbon emissions. However, EV charging becomes a limiting factor for prospective car buyers. According to a national survey, 67% of prospective EV buyers state that more charging stations are necessary to promote EV ownership [3].

The City of Los Angeles actively participates in the “Green New Deal”, a plan to address climate change and carbon emissions. One of the plan’s detailed steps is to increase the percentage of zero-emission vehicles in the city-wide fleet, which includes private, commercial, and public transit vehicles, to 25% by 2025 and ultimately to 100% by 2050 [4]. To promote EV ownership, it is important to ensure the availability of public charging stations as part of the City's infrastructure development plan.

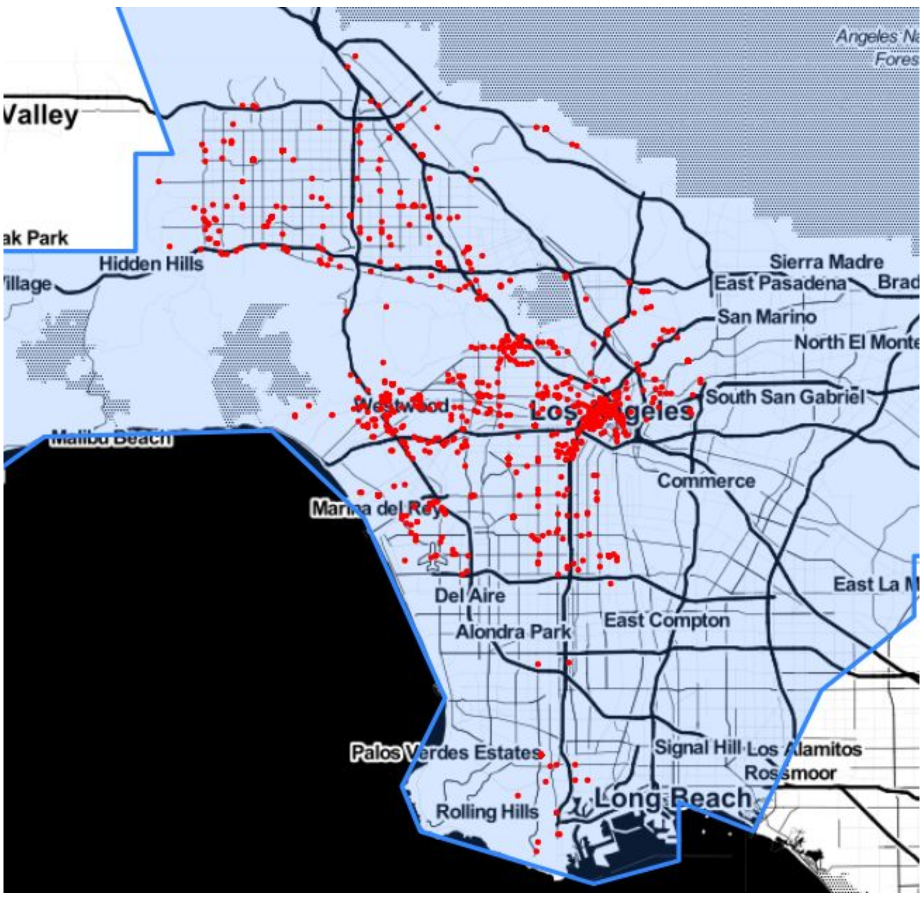

Figure 1: Existing EV charging stations in the City of Los Angeles.

Unfortunately, EV charging stations in the City of Los Angeles are not evenly distributed. In Figure 1, each red dot represents a charging station. Specifically, there are more charging stations congregated in the downtown Los Angeles area, whereas fewer stations are located in the southern and eastern parts of the city.

To advance the “Green New Deal”, the City of Los Angeles plans to install more public EV charging stations in 2021. Newly introduced chargers must be effective in increasing the availability of charging stations for EV owners, attracting more EV buyers, balancing the distribution of the EV chargers, and serving disadvantaged neighborhoods.

In this project, we propose the most promising charger locations based on the budget constraints of the City of Los Angeles. First, we present a predictive model that determines the features of prospective EV chargers. This model uses the data of frequency of use of existing EV chargers, charger locations, commute time, nearest amenities, and census. This model utilizes radial geo-analysis and the features of existing chargers to predict the placement of new EV chargers. Secondly, we examine the placement of new EV chargers using the predictions and the existing data as an optimization problem. We will optimize for the placement of new EV chargers in council districts; we model the task as a set cover problem, characterizing each existing charger with a radius and greedily analyzing a set of potential locations for new chargers to see which placements would serve the most people. As a result, the predictive model and optimization output the placement of the most promising new EV chargers based on the budget of the City of Los Angeles. We hope that the result will benefit existing EV owners as well as promote the City’s “Green New Deal”.

2 Related Work

As the usage of electric vehicles increases, so does the demand for electric vehicle chargers. This problem has long been anticipated with the advent of electric vehicles as a viable widespread transportation option, so there is a significant body of research on efficient infrastructure to supply charge for these new vehicles. This body of research often includes research on allocating resources for many different types of 'alternative fuel' sources other than traditional gasoline, but this report will only focus on methodology and research conducted specifically on electric vehicle chargers since each type of charging resource is subject to different constraints.

There are two types of electric vehicle chargers that are differentiated by the speed at which they can fully charge an electric vehicle, based on their power output. There are slow chargers, which can take on the scale of several hours to fully charge an electric vehicle. These are relatively inexpensive to install but are only truly useful in places where electric vehicle owners will leave their vehicles for long periods, such as workplaces, schools, industrial parking structures, etc. Fast chargers, on the other hand, can fully charge an electric vehicle in a quarter to half an hour, so they have fewer location constraints as they are convenient for use anywhere. They do, however, require a significant initial cost investment as well as spaces that can accommodate the hardware installation to support the high power output[5].

The problem of location-allocation of EV chargers is often mostly applied to fast chargers, because of the more complex set of constraints[6]. Though slow chargers are constrained by the type of location, the low installation cost means that quantity is not often a constraint. Therefore, an inefficiently placed slow charger is not considered a significant loss and is often not worth the expenditure of further research to improve efficiency. The only current data available for this project details the location and usage statistics of public level 2 chargers, which are classified as slow chargers; there is no available data from the city on the few public DC fast chargers that have been placed. The city of Los Angeles is aiming to add a certain number of slow chargers as well as a certain number of fast chargers, so the methods used in this project will be extrapolating the ideal placement of fast chargers from the placement and usage of slow chargers, which previous work seems to suggest will not be wholly accurate, as the use cases for both are intuitively and observationally different.

One of the most important use cases for fast chargers, to ensure the viability of widespread electric vehicle adoption, is its ability to cater to regular and sustained everyday use. This means that an electric vehicle must be able to effectively travel between two separate charging stations without running out of charge. This makes it essential for any research to capture the major flows of traffic, nodes where frequent travel begins and ends, to ensure that electric vehicle travel between them is supported. Given this information, two main objectives are commonly explored in research with regards to fast charger placement. If the goal is to capture all the flows of traffic, to achieve full coverage without a constraint on the number of fast chargers that can be placed, the problem is modeled as a cost minimization problem, where the minimum number of chargers that can be placed to achieve full coverage needs to be determined[7]. If there is a limit on the number of chargers that can be placed, the problem is modeled as a maximum cover problem, where chargers must be optimally placed to maximize the number of traffic flows that are captured[8]. Since the city of Los Angeles team has specified the numbers of chargers to be allocated, this project will use maximum cover algorithms and methodologies as appropriate for the specific constraints we have.

A review of a representative subsection of the existing literature on the subject noticed the discrepancy between research that has been conducted and the actual application of their results in the development of transportation infrastructure. The review noted that this discrepancy was often the result of modeling the problem under pure and predictable mathematical constraints for factors like availability at a certain time, consistent traffic, etc[5]. While these are necessary assumptions to model a problem, real-world constraints are far less predictable, since they are placed in locations that can restrict access, flows can be ill-captured due to uneven use of a certain charger on certain days, and other fluctuations that reflect the domain constraints of transportation and traffic. A large body of research exists on the topic, but many of the papers have problem formulations that are slight variations of one another. As a result, each paper makes little, if any, forward progress in developing a solution that is feasible for implementation in actual high usage areas. This project will attempt to take this discrepancy into account and minimize its impact by factoring in the multitude of constraints provided by actual domain professionals, such as zoning considerations, the objectives and resources of the city, etc. Additionally, this project will make accommodations for shortcomings that are already apparent in the data, such as the lack of usage data for fast chargers as well as no explicit data sets that capture traffic flows. The scope of the problem, as well as its formulation for practical purposes rather than academia, should hopefully mitigate concerns about lack of applicability.

3 Data

The predictive model uses the frequency data, Bing map and Google map data, and census data.

3.1 Predictive Model Datasets

We used data from the City of Los Angeles measuring FLO EV chargers [9]. FLO is a large network of EV charging stations and provides us information on chargers as well as their frequency of use. In the FLO dataset, we have information about 177 charging stations in the City of Los Angeles and their frequencies of use. Furthermore, the data of each charging session is also available, including charging duration, cost, and kilowatt-hours (kWh) of electricity delivered. The dataset of each charging session contains 8541 rows. We also used data from the National Renewable Energy Laboratory (NREL) [10]. Although there is a lack of frequency of use data, the dataset from NREL provided us with more locations of chargers in the City. This information was used to visualize the distribution of chargers in the City and to implement the CELF set cover problem in our optimization step.

3.1.2 Census Datasets

One source of features for the charging stations is data from the U.S. Census Bureau, accessed on Social Explorer. We used the American Community Survey (ACS) 5-year estimates for Los Angeles County, partitioned by census tract. We chose the 5-year estimates over the 1-year or 3-year estimates due to lower rates of missingness and a low need for currency.

We use 66 ACS elements as features of the EV charging stations. Additionally, we synthesized two other metrics as features; they will be discussed in more detail in the following subsection. In the current version of our predictive model, we use the value of the census tract the charger is located in as a feature. In our final model, we will implement radial geo-analysis, which will allow us to more accurately represent the area surrounding a charging station, especially if that area includes multiple census tracts.

3.1.3 Data Preprocessing

Our preprocessing consisted of two steps. First, we performed two sets of calculations on various census data to synthesize two metrics: concentrated disadvantage and the expected number of vehicles in the census tract. Second, we accounted for missing data.

Concentrated Disadvantage (CD) is a metric to determine how disadvantaged a prescribed area is related to the larger area of which it is a subset. For this project, we calculated the CD of each census tract in Los Angeles County compared to the entire county. Calculating this metric requires the following elements: percent of the population below the poverty line, percent of individuals on public assistance, percent of female-headed households, percent unemployment, and percent of the population younger than 18. Each of these elements is calculated from census data. Next, each element is converted to a z-score using the following formula:

$$z_i(x) = \frac{x_i - \mu_i}{\sigma_i}$$

Where x_i is the observed percentage of feature i for census tract x, μ_i is the mean percentage for feature i, and σ_i is the standard deviation of the percentages of feature i. Finally, the mean of the z-scores is taken for each census tract, and the result is the concentrated disadvantage.

Our second synthesized metric is the expected number of vehicles in the census tract. The census provides data on the number of households with one car, two cars, etc. up to five or more cars. By taking a weighted sum, we calculated the expected number of vehicles in the census tract.

3.1.4 Missing Census Data

Finally, we accounted for missing data whenever appropriate. We did not delete census tracts as a whole, rather we would ignore chargers for specific regressions when the census tract data was missing.

Of our three sources of geographic features, the census data are the only ones with missing values. We accounted for this by replacing missing values with the mean of the missing value for that tract's neighbors.

$$value_i(x) = mean(value_i(n)) \ \forall \ n \ \in N$$

Where N is the set of neighbors of the census tract x and i is the feature with the missing value. The above equation filled in approximately 90% of all missing values. Median house value, which was at first missing 107 (4.6%) entries, was only missing 13 (0.6%) after the above fill function was executed. The remaining missing values were filled in using forward fill with the census tracts sorted by their tract code.

3.1.5 Vehicle Fuel Type Count by Zip Code

Another source of geographical information is the "Vehicle Fuel Type Count by Zip Code" dataset published by the state of California [11]. The dataset contains counts of cars distinguished by make, model year, zip code, and fuel type. Included in the fuel types are "Battery Electric" and "Plug-in Hybrid" the two fuel types which use EV charging stations. We will combine all cars of those two fuel types based on zip code and use the total number of EVs in a zip code as a feature.

3.1.6 Map Data

Finally, we have the map data from the Google and Bing Map APIs. Both are map applications that provide geographical information on each charger, including the number of points of interest nearby, the distance to the nearest highway, and the driving distance between two points.

3.1.7 City of Los Angeles GIS data

We pulled new data from the city of Los Angeles to match the constraints for our optimization. Upon discussion, though placements are not actually limited to street corners, it was a valuable set of discrete points to start from, so we pulled datasets on Intersections in the city of LA[13]. We also used the General Plan Land Use data to calculate the zoning regions for Commercial and Multi-Family areas[14]. The goal is also constrained by placing a certain number of chargers in a council district, so we used Council District data from the city of LA as well[15].

3.2 Circular Geo-analysis

3.2.1 Motivations

To acquire a more accurate understanding of the area that an EV charger serves, we implemented a circular geo-analysis. Comparing the size of a census tract to the area served by an EV charger, as is done in Figure 2 below, demonstrates why this process is necessary.

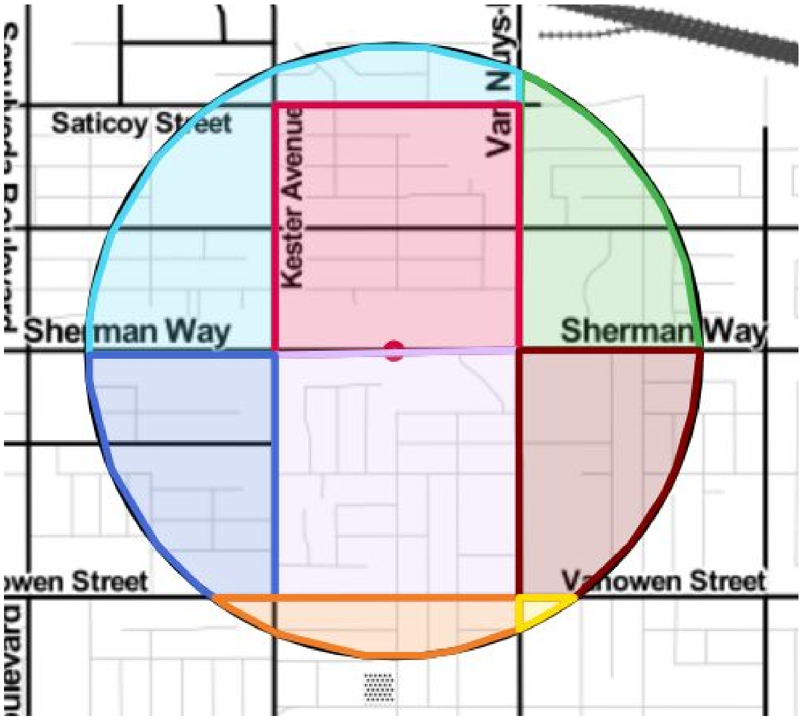

Figure 2: The eight census tracts within a 1 km radius of the charger at
20771 W Sherman Way, Los Angeles, CA

The circle in Figure 2 has a radius of 1 km, noticeably smaller than the average travel distance to a charging station, which we calculated to be 6.7 km.

We did our circular geo-analysis in three radii: 1km, 4km, and 6.7km. We chose 1km as it is a good estimate of walking distance. 4km is a good choice for an intermediate value between 1km and 6.7km. Finally, as stated above, we are interested in 6.7 km because this is the average travel distance for charger users.

3.2.2 Implementation

This involved calculating the values of features for a circle around a specific charger. Our first step was to distinguish absolute features from relative features. Absolute features, such as population, are counts, whereas relative features, such as median household income, are already averaged over the whole area. After making that distinction, we created the following two formulas to calculate the value of absolute features and relative features, respectively, for the circle around a charger *x*:

$$value_a(x) = \sum_{t \in T} \left(\left(\frac{area(t \cap c)}{area(t)} \right) (value_a(t)) \right)$$

$$value_r(x) = \sum_{t \in T} \left(\left(\frac{area(t \cap c)}{area(c)} \right) (value_r(t)) \right)$$

a signifies an absolute feature and *r* signifies a relative feature. *T* is the set of all census tracts that overlap with circle *c*. *overlap* is the intersection of a census tract and the circle around a charger. *overlap* can often be an entire census tract, as many of the tracts in Los Angeles are small.

Circular geo-analysis was applied to all census features as well as the vehicle fuel type count by zip code. The Map APIs have this functionality built-in, so the above equations were not applied.

4 Predictive Model

We applied nonlinear transformation, feature selection on our labels. Furthermore, the machine learning methods we used to develop the predictive models were Lasso linear regression, random forest, and XGBoost. For each step, 10-fold cross-validation was performed on the dataset to ensure the accuracy of the model.

4.1 Non-linear transformation

Non-linear transformation was applied to our dataset for better statistical interpretability. 10 non-linear transformations were tested, including normal, gamma, dweibull, t, genextreme, gamma, lognormal, beta, and uniform transformations. The result shows that normal transformation outperformed other transformations with the lowest RSS (Residual sum of squares) score on most features.

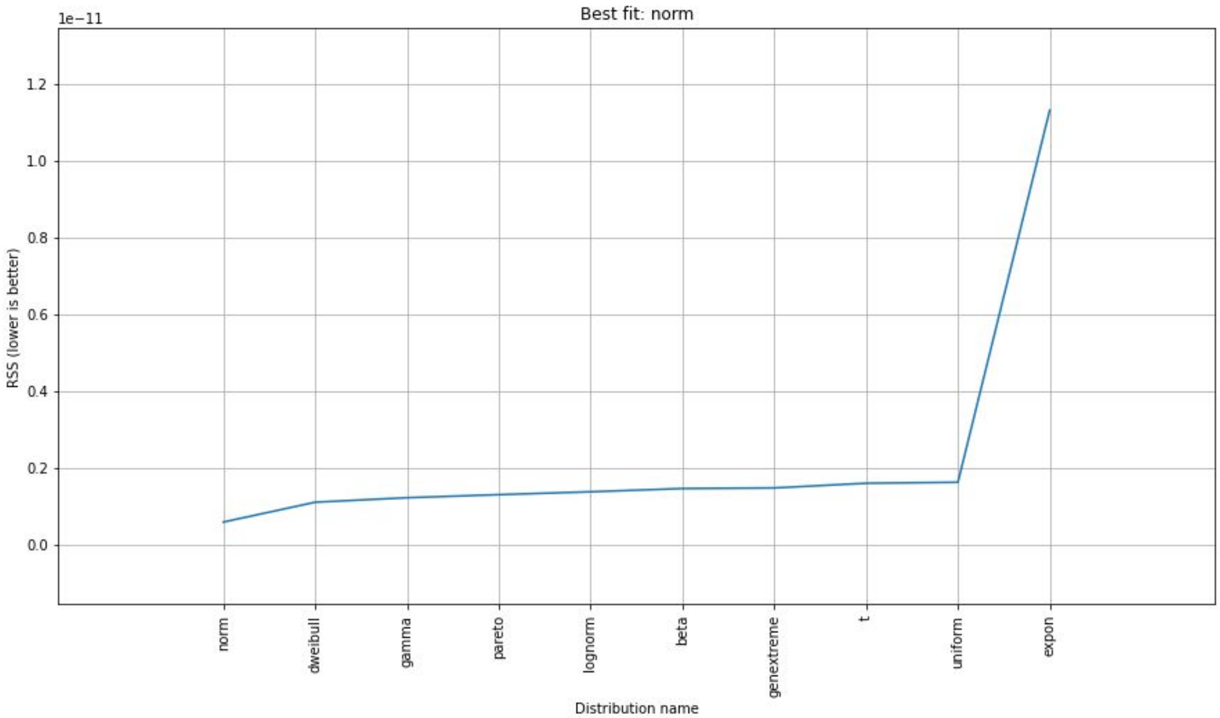

Figure 3: The graph of RSS scores for 10 tested transformations.

Additionally, a series of experiments were performed to test the performance improvement of transformation. The same dataset was put into different regressors with or without transformation to check if transformation helps increase the predictivity.

Table 1: R Squared and MSE Values for Predictive Models with or without nonlinear transformation

	Without transformation			With transformation		
Radius = 1km	Lasso Linear regression	Random Forest	XGBoost	Lasso Linear regression	Random Forest	XGBoost
R^2	-0.01411	0.12103	0.20002	0.04667	0.19194	0.27335
MSE	2979	2308	2512	2349	2318	2511

As a result, features that show compliance with normal distribution were transformed using PowerTransformer and Yeo-Johnson transform for better model performance.

4.2 Feature Selection

Feature selection was performed on our labels, where features that contribute most to the prediction power were selected. Irrelevant features were eliminated to reduce overfitting and improve accuracy.

Two ways of feature selection were examined: Recursive feature elimination and Sequential forward selection. Recursive feature selection is to select features by recursively considering smaller and smaller sets of features, i.e. in each iteration, the least important feature is pruned from the current set of features. Sequential forward selection, in the opposite, adds the most important feature into the current set in each iteration. Both methods were tested on our transformed dataset, where the most significant features were fed into three regressors.

Table 2: R Squared and MSE Values for Predictive Models using Recursive feature elimination and Sequential forward selection

	Recursive Feature Elimination			Sequential forward selection		
Radius = 1km	Lasso Linear regression	Random Forest	XGBoost	Lasso Linear regression	Random Forest	XGBoost
R^2	0.04654	0.18854	0.20473	0.05372	0.19245	0.28375
MSE	2780	2332	2570	2445	2435	2642

Based on the metrics, sequential forward selection outperformed recursive feature elimination. Next, sequential forward selection was applied to models with different radii. It is worth mentioning that there was not an upper cap for the number of features selected. Sequential forward selector was tuned to select the best subset of features.

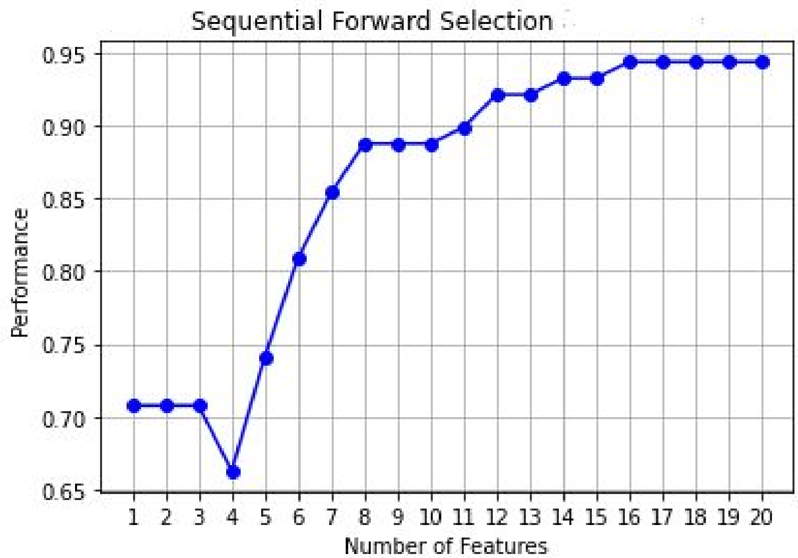

Figure 3: Visualization of sequential forward feature selection process

The five most important features in the best subset were taken out for displaying purpose.

Table 3: Five most important features for different models

Radius	Lasso Linear regression	Random Forest	XGBoost
1 km	30_49_percet_income_to_rent_1_km, commuters_by_car_truck_van_1_km, ev_car_count_1_km, households_retirement_income_1_km, housing_units_1_km	Commute_time_1_km, employed_1_km, High_school_grad_1_km, householder_high_school_degree_1_km, householder_less_than_high_school_1_km	Commute_time_1_km, Employed_1_km, High_school_grad_1_km, Householder_high_school_degree_1_km, householder_less_than_high_school_1_km

4 km	Commute_time_4_km, employed_4_km, ev_car_count_4_km, Householder_high_school_degree_4_km, householder_less_than_high_school_4_km	Commute_time_4_km, Employed_4_km, high_school_grad_4_km, Householder_high_school_degree_4_km, housing_units_4_km	Commute_time_4_km, Employed_4_km, high_school_grad_4_km, Householder_high_school_degree_4_km, housing_units_4_km
6.7 km	Children_living_with_single_parents_6.7_km, Commuters_by_car_truck_van_6.7_km, ev_car_count_6.7_km, housing_units_6.7_km, median_house_value_6.7_km	Average_household_size_6.7_km, Commute_time_6.7_km, Commuters_by_car_truck_van_alone_6.7_km, Employed_6.7_km, not_high_school_grad_6.7_km	Average_household_size_6.7_km, Commute_time_6.7_km, Commuters_by_car_truck_van_alone_6.7_km, Employed_6.7_km, not_high_school_grad_6.7_km

In Lasso linear regression, the best set of features often includes the number of EVs, commute time, educational level of householders. The most important features for random forest included commute time, number of employed people, and educational level. For XGBoost, the best set of features were similar, including commute time, number of employed people, and educational level. It is worth noting that for a 6.7 km radius, all three models output relatively different sets of features compared with sets from smaller radii (1km and 4km). For random forest and XGBoost, average household size and number of commuters became important features at a 6.7 km radius.

4.3. Models Results

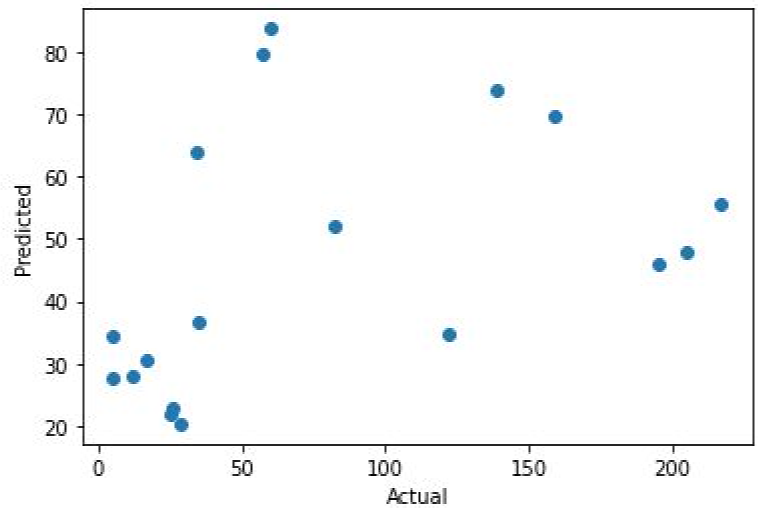

Figure 4: Predicted frequency vs. actual frequency using Lasso linear regression model. Radius = 1km. Number of features kept = 15

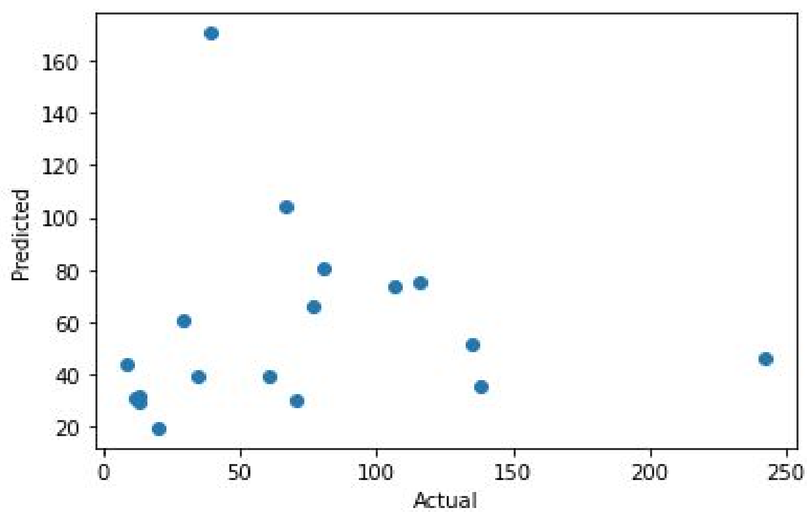

Figure 5: Predicted frequency vs. actual frequency using random forest model. Radius = 1km. Number of feature kept = 6

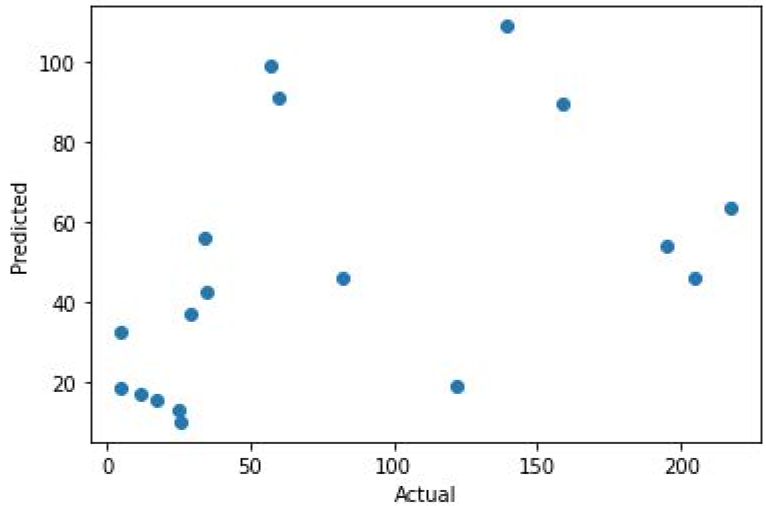

Figure 6: Predicted frequency vs. actual frequency using XGBoost model. Radius = 1km. Number of feature kept = 9

Table 3: R Squared and MSE Values for Predictive Models using Lasso, random forest, and XGBoost for different radii

Radius	Metrics	Lasso Linear regression	Random Forest	XGBoost
1km	R^2	0.02376	0.18245	0.29375
	MSE	2252	2075	2342
4km	R^2	0.05374	0.17452	0.32368
	MSE	2474	2085	2372
6.7km	R^2	0.09674	0.20834	0.33750
	MSE	2240	2182	2208

Overall, for different radii, XGBoost and random forest outperformed Lasso linear and in R^2 and MSE. Despite an exception for random forest at 4km radius, both XGBoost and random forest displayed a tendency of increasing predictability as radius increased.

When determining the performance of a model, it is crucial to take both interpretability and predictability into consideration. In general, Lasso linear regression has its advantage of being easily interpretable. However, it is unwise to trade interpretability with a suboptimal predictability. As a result, we claimed XGBoost to be the best method for our problem.

If we take the results from feature selection into account, it is obvious that commute time, number of commuters, size of household, number of employed people are important features that influence the frequency of use of a charger.

5 Optimization

5.1. Process

In addition to the regression model, we aimed to also take some preliminary steps towards the larger objective, which is to place fifteen new EV chargers in each council district. Since we had a fixed number of chargers to place, we modeled this as a maximal set cover problem; after placing all the existing chargers, we used the greedy CELF algorithm to determine new charger placements. We used a mixture of polygonal objects from different libraries and visualizations we used in other steps to perform and validate the optimization. Though an initial process for how to complete this was discussed and agreed upon, a few details became apparent during the actual implementation that necessitated a change in the procedure.

5.1.1 Placing Existing Chargers

We utilized all the existing charger locations available to us from the Chargepoint and the NREL datasets. The Chargepoint data set[12] had data about the driver's zip code associated with each charging session, so initially, each of these chargers was characterized with a coverage radius that was the mean of usage statistics for that particular charger. The NREL dataset[9] had no associated zip code data, so each of these chargers was characterized with the overall mean distance that users travel to use an EV charger, obtained from the Chargepoint dataset. All chargers were placed as circular Polygon objects, with their coordinates and their coverage radius. These were also specially transformed with respect to latitude and longitude coordinates, to make them fit for mapping on a flat plane.

5.1.2 Finding Valid Potential Points

To find a set of valid potential locations to test, we filtered based on the verbal constraints discussed with the LA data science team. The team mentioned that 15 new chargers need to be placed in each council district. Though more people in single-family households are likely to own electric vehicles, they also have more of an opportunity for private access to a charger, whereas in multi-family and commercial areas, there may be fewer owners, but more of a need for public provision of a charging station. Thus we filtered the intersection points by location, selecting for the ones within multi-family and commercial zones.

5.1.3 Categorizing by Council Location

Once a set of valid possibilities was established, we sorted all possibilities into their respective council districts. Each council had its own list of valid possibilities.

5.1.4 Placing New Chargers

After aggregating the coverage of already placed chargers (as the total covered area thus far), we ran the CELF algorithm for each council district's list. The CELF was constrained to choose 10 points from each list. If there were fewer than 10 options, then the algorithm would sort and return the options that were available, and if there were no options the algorithm would return an array of None. Since the algorithm uses a heap data structure, due to the functional details of our particular heap implementation, the algorithm returns a selection of IDs that map to a Polygon object that is the charger and associated radius.

```
0 : [4, 6, 3, 5, 0, 1, 2]
1 : [3, 2, 0, 1]
2 : [44, 39, 11, 41, 1, 43, 40, 6, 0, 10]
3 : [1, 0]
5 : [3, 12, 15, 7, 1, 11, 2, 4, 13, 8]
6 : [0, 38, 19, 30, 3, 34, 26, 32, 23, 13]
7 : [0]
8 : [0]
11 : [50, 11, 42, 23, 78, 67, 65, 4, 64, 60]
13 : [4, 0, 5, 1, 3, 2]
14 : [4, 26, 19, 0, 9, 1, 25, 18, 20, 7]
```

Figure 7. Example of CELF index selections for R=400m

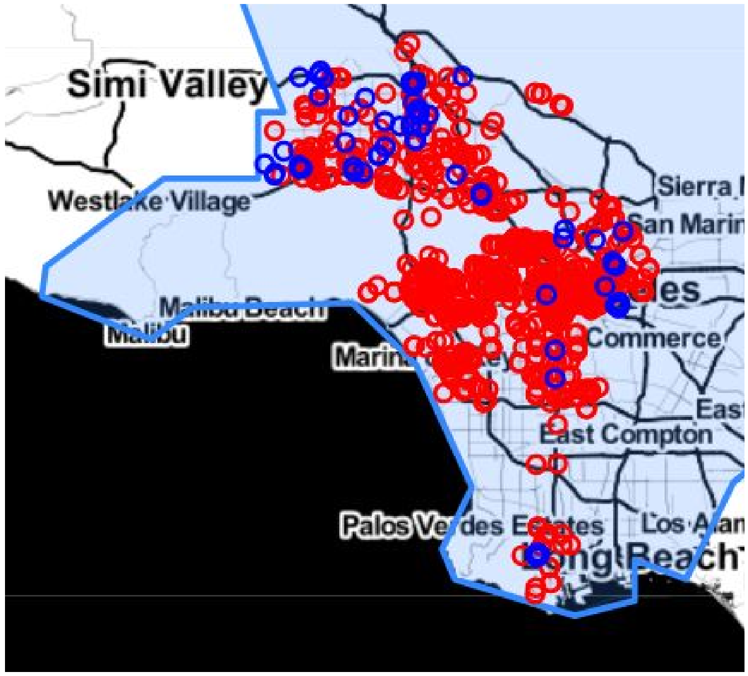

Figure 8. CELF selections for R=400 visualized

5.2. Modifications

5.2.1 Radial Characterization

The radial characterization in 5.1.1, though derived from our data, gave us no meaningful additional points. The results always appeared in the following format, where each element is an index.

```
council #: [len(potential pts in council district) - 1, 0, 1, 2, 3, 4, 5, 6, 7, 8]
```

Since we're using a heap implementation to return this, we can infer that this is because all the points had the same weight; the first selection is the last item to be entered and every subsequent selection is the next item on the heap. From this we can tell that the radius was too large, and the union of all existing chargers encompassed all smaller potential points such that their radii gave no additional benefit (so all the points had the same weight). This made sense in the context of Los Angeles' size; since the radial distance was an assumption to begin with, it seemed reasonable to assume a smaller size radius in order to test placement. Radii of 1000m, 2000m, 3000m, and 6000m were tested. These trials showed non-trivial results. Upon further discussion with the LA team, a radius of '5-minute walk' would be the most meaningful, so that users would be able to walk 5 minutes from their parked electric vehicle to their destination. This standard is widely agreed upon by city planners to be 400 ft[17]. The model above is characterized by 400 ft radii for all points.

5.2.2 Zoning

Initially this model was run over various radii using the single-family zoning regions - only further discussion with the LA team clarified that the intent is to fulfill the need for public charging as opposed to the likelihood of private charging. Some of the models are shown below. This is a proof of concept that both radius and zoning are modular, and running this model under different constraints will not be much more computationally expensive.

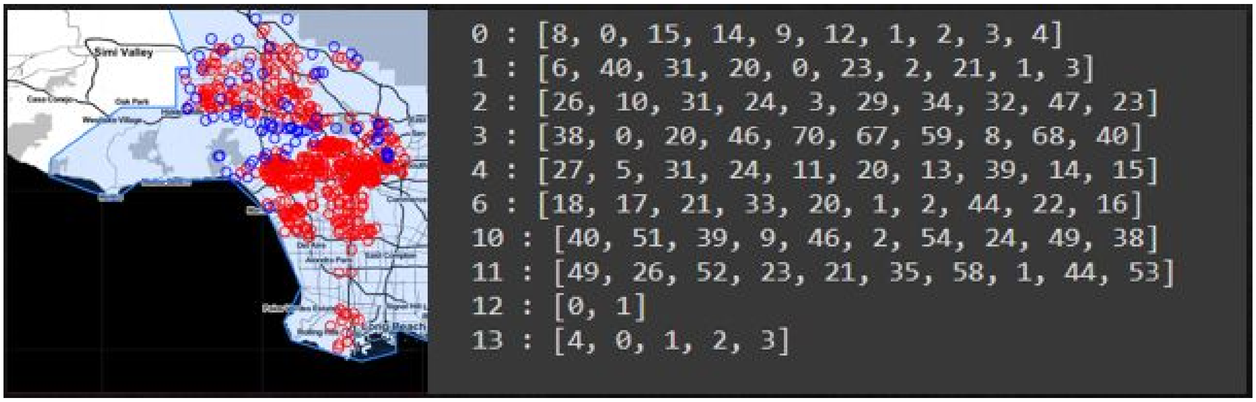

Fig 9. Visualization and index selections for Single Family at Radius = 3000m

6 Conclusions and Extensions

6.1. Conclusions

Our models used Lasso linear regression, gradient boosting (XGBoost) and random forest to predict charger usage using various datasets we obtained from census, from the City, and from sustainability laboratories. Both dataset transformation and feature selection were applied to our datasets for better performance. The XGBoost and random forest models performed better than Lasso linear regression, which is imaginable as it is hard for a real-world problem to have a strong linear correlation. Ultimately, with consideration about both predictability and interpretability, XGBoost was the preferred choice.

Our optimization process using the CELF algorithm shows that it could be a useful tool in modeling different sorts of basic constraints. Applying constraints such as points at intersections in muti-family and commercial zones, as well as characterizing chargers with a 'walking distance' radius, resulted in meaningful solutions that the placement team can take under advisement. The nature of the algorithm and the data sources makes it very easy to test out different radiuses, different zoning constraints, and depending on the availability of other data, like discrete street points, can also make it easy to experiment with different initial possible locations. At the very least, this is an effective proof of concept for modeling this problem in this way.

6.2. Modeling Limitations

6.2.1 Limitations of the Predictive Model

6.2.1.1 Limited Data

We did not have enough frequency of use data for chargers. What we had was that of ~180 chargers. At the preliminary stage of this research, we tried to find more data on related websites, including ChargePoint and PlugShare. Unfortunately, all of them required commercial license for accessing the API and frequency data, which went outside of our scope. As a result, our models did not have enough data and at a high risk of overfitting.

6.2.1.2 Limited Features

Our model relied heavily on census data. However, there was approximately 4.6% of census data missing for certain census tracts. Considering the small set of chargers we have, these missing data could potentially affect the performance of our models.

6.2.2 Limitations of the Optimization Algorithm

6.2.2.1 Current Limitations

Setting up the CELF for a real world naturally comes with its own assumptions that may or may not be faulty, but are sometimes acceptable. Selecting points and characterizing them with an appropriate radius was a challenge. There needs to be a discrete set of points that the algorithm can be applied on - the nature of this problem is that the possibilities tend to be less discrete, since there can be several possibilities along a single street. We made a limiting assumption in selecting only intersections, whereas in reality anywhere along the middle of the block would be acceptable. However, this level of precision is sufficient for advisory purposes - since this will probably only be used as a tool, approximate placement within one block is an acceptable range.

6.2.2.2 Future Limitations

This algorithm is really only appropriate for modeling the problem to this degree of complexity, with some room for growth. We can model a few additional constraints on the initial points, and filter them some more, and we can also model preferences towards certain selections by weighting pixels differently. However, if any additional constraints on the placement (rather than the initial selection) need to be placed, the greedy algorithm will no longer be appropriate. This is a concern when considering something like the place of fairness in resource allocation, and this algorithm is not very conducive to that.

6.3 Possible Extensions

6.3.1 More complete data for various chargers

We are interested in doing analysis for different charger types. Currently in the City of Los Angeles there are two types of chargers: fast and slow. We only have data for slow chargers. Because of the different nature of two chargers, i.e. charging duration, potential users, etc., we are expecting a different set of features to be important for fast and slow charger. A future extension could be made once frequency data of fast chargers become available.

6.3.2 Temporal forecasting

With temporal data of charging sessions, temporal forecasting is achievable as a possible extension. Temporal information of each session could be used to predict the expected usage of chargers in the next one month, 6 month, or 12 month, depending on the availability of the data. With temporal forecasting, it is feasible to adjust policies accordingly.

6.3.3 Detailed Optimization

6.3.3.1 Motivating Charger Radii

As discussed, we can spend some time looking into more factors that determine the radius of ‘service’, which can be defined as when an electric vehicle owner would choose to charge their vehicle at this station instead of charging it at another. Though walking distance seems appropriate, there are probably other factors we can explore as well. Also, we can add some nuance to the radii of existing chargers that we have data about, and in turn add some sophistication to characterizing new chargers as well. If we are able to group chargers by characteristics, with something like a K-means algorithm, we can also make some assumptions that they will be used in somewhat similar ways. We can characterize older chargers with a more accurate radius, and characterize new chargers with a radius that best fits with their group.

6.3.3.2 Adding Relevant Weights

We can work to integrate the two approaches used in our project more fully. We can use the features determined as most important from our feature selection and characterize each census tract with these features, weighting each ‘pixel’ of the census tract appropriately, basically replicating the radial geoanalysis process, but this time, in the interest of using that to weight the area of the circles when conducting CELF.

References

[1] U.S. Department of Energy. (2020). “Electric Vehicle Registrations by State” [Online]. Retrieved from: https://afdc.energy.gov/data/10962

[2] EVAdoption (2020). “EV Market Share by State” [Online]. Retrieved from: https://evadoption.com/ev-market-share/ev-market-share-state/

[3] Union of Concerned Scientists (2019). “Electric Vehicle Survey Findings and Methodology” [Online]. Retrieved from: http://www.jstor.com/stable/resrep24093

[4] City of Los Angeles (2019). “L.A.’s Green New Deal” [Online]. Retrieved from: https://plan.lamayor.org/

[5] “How Do Residential Level 2 Charging Installation Costs Vary by Geographic Location?” INL/MIS-15-35056; Idaho National Laboratory: Idaho Falls, ID, USA, 2015.

[6] Motoaki, Yutaka. “Location-Allocation of Electric Vehicle Fast Chargers—Research and Practice.” *World Electric Vehicle Journal*, vol. 10, no. 12, 2019, https://www.mdpi.com/2032-6653/10/1/12/htm.

[7] Wang, Y.-W.; Lin, C.-C. “Locating road-vehicle refueling stations”. Transp. Res. Part E Logist. *Transp. Rev*. 2009, 45, 821–829.

[8] Davidov, S.; Pantoš, M. “Planning of electric vehicle infrastructure based on charging reliability and quality of service”. *Energy* 2017, 118, 1156–1167

[9] “Electric Vehicle Charging Stations by locations” https://www.nrel.gov/about/ev-charging-stations.html

[10] “Flo Utilization Report”, “Flo Parks Utilization Report” https://drive.google.com/file/d/1-0TaPbqlZnWZwoC58WN67AfIfo5Qpp0G/view?usp=sharing

[11] “10/1/2020 Vehicle Fuel Type Count by Zip Code”, https://data.ca.gov/dataset/vehicle-fuel-type-count-by-zip-code/resource/ d304108a-06c1-462f-a144-981dd0109900

[12] ‘Chargepoint Data Set’ https://raw.githubusercontent.com/medhivya/ev-charger/main/chargepoint_coords_w_radius.csv (Sourced from LA Data Science Team)

[13] ‘Intersections in LA’ https://raw.githubusercontent.com/medhivya/ev-charger/main/Intersections.csv (Sourced from https://geohub.lacity.org/datasets/0372aa1fb42a4e29adb9caadcfb210bb_9)

[14] ‘General Plan Land Use - Multi Family and Commercial Zones’ https://services5.arcgis.com/7nsPwEMP38bSkCjy/arcgis/rest/services/GPLU/FeatureServer/0/query?where=CATEGORY%20%3D%20'COMMERCIAL'%20OR%20CATEGORY%20%3D%20'MULTIPLE%20FAMILY'&outFields=GPDESC,CATEGORY,Shape__Area,Shape__Length&outSR=4326&f=json (Sourced from https://geohub.lacity.org/datasets/e981169cdec3463892383021cb08dc2b_0)

[15] ‘Council Districts in LA’ 'https://raw.githubusercontent.com/medhivya/ev-charger/main/Council%20Districts%20(1).geojson' (Sourced from https://data.lacity.org/A-Well-Run-City/Council-Districts/5v3h-vptv)

[16] ‘LA County Boundaries’ 'https://raw.githubusercontent.com/ritvikmath/StarbucksStoreScraping/master/laMap.geojson'

[17] EV Studio (2020) *EV Studio* [Online]. Retrieved from https://evstudio.com/the-five-minute-walk-calibrated-to-the-pedestrian/.