

## **Project1/Milestone 1: Proposal & Data Selection**

Author: Anjani Bonda

Date: 6/10/2023

## Topic

Tesla Supercharging Locations Pattern/Prediction

## Business Problem

It is quite obvious from various surveys that people looking to buy an Electric Vehicle (EV) car consider below factors before making their first purchase.

- Brand
- Features & Technology
- Charging Stations
- Superior Design
- Environmentally Conscious
- Performance & Range

Of all the various brands in the current market, Tesla is clearly the market leader not just because of the brand value and design but mostly because of the reliable Tesla supercharging stations. The public charging stations had their share of problems as they sometimes don't work which is quite frustrating to the EV car owners. There are recent developments and collaboration between EV automakers to utilize Tesla supercharging stations which stands out to be the best and reliable ones so far with its large network.

One of the key features for prospective EV car buyers is charging stations. I will be exploring more on this aspect where it might be useful for upcoming EV buyers to make purchase based on the charging stations and their possible expansion pattern/prediction. The focus of this model is limited to USA.

## Datasets

The datasets are extracted from Kaggle website.

<https://www.kaggle.com/datasets/omarsobhy14/supercharge-locations>

<https://www.kaggle.com/datasets/richardg9/tesla-car-sales-quaterly>

The Tesla supercharge locations dataset is a treasure trove of information. It contains geographical coordinates, amenities, and other details for each supercharge location to analyze the data, discover optimal routes and uncover patterns for electrifying adventures.

This dataset has about 5876 records covering worldwide and about 2200+ for USA. Below are the various attributes of the dataset.

1. Supercharger: This feature represents the name or identifier of the Tesla Supercharger location. It helps identify and distinguish each Supercharger station in the dataset.
2. Street Address: This feature contains the specific street address where the Supercharger station is located. It provides the physical location information for each station.
3. \*\*City: \*\*This feature represents the city where the Supercharger station is situated. It helps identify the geographical location of each station.
4. State: This feature indicates the state or province where the Supercharger station is located. It provides additional regional information about each station's location.
5. \*\*Zip: \*\*This feature represents the postal code or ZIP code associated with the Supercharger station's address. It helps identify the precise location within a city or region.
6. Country: This feature indicates the country where the Supercharger station is situated. It provides information about the specific country in which each station is located.
7. Stalls: This feature represents the number of charging stalls available at the Supercharger station. It indicates the capacity of the station to accommodate multiple vehicles simultaneously.
8. kW: This feature represents the power capacity or kilowatt rating of the Supercharger station. It indicates the charging speed or power output available at each station.
9. GPS: This feature provides the GPS coordinates (latitude and longitude) of the Supercharger station. It offers precise location information for mapping and navigation purposes.

10. Elev(m): This feature represents the elevation or altitude of the Supercharger station above sea level. It provides information about the station's height relative to the surrounding area.
11. Open Date: This feature indicates the date when the Supercharger station was opened or made available for public use. It provides information about the timeline of station deployment and expansion.

We will be also using the Tesla quarterly car sales from 2013-2022 as a supplemental dataset.

## Methods

Below algorithms or model techniques will be utilized on the dataset to determine which features are related to our target variable "Purchase" (likely to purchase).

1. Logistic Regression
2. Random Forest
3. Decision Tree

Logistic regression is a statistical analysis method used to predict a binary outcome such as yes or no based on prior observation of the data set. Here, "Purchase" feature present in the dataset has only binary values and will be used as target for the model. This model falls under supervised learning as the data is well labelled and has a target variable, a column in the data representing values to predict from other columns in the data. Under supervised learning, this dataset falls under classification model as it reads the input and generates an output that classifies the input into two categories: one having purchase as "Yes" and "No". Decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.

Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time.

## Ethical Considerations

There are no possible ethical aspects to this model as the data is public info and doesn't really include any consumer related information.

## Challenges/Issues

Key challenge is to ensure if this data is good enough to build the prediction model and if we need more supplemental data to support the model. Might possibly need to explore more supplemental datasets to strengthen the model.

## Reference

Dataset1: <https://www.kaggle.com/datasets/omarsobhy14/supercharge-locations>

Dataset2: <https://www.kaggle.com/datasets/richardg9/tesla-car-sales-quaterly>

<https://amplifyxl.com/target-market-for-tesla/>

Random Forest: [https://en.wikipedia.org/wiki/Random\\_forest](https://en.wikipedia.org/wiki/Random_forest)