Project 1: Tesla Supercharging Stations Location Prediction

Author: Anand Sarda

DSC680 Applied Data Science

6/19/2025

Table of Contents

Business Problem

It is 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.

Background/History

For any new or prospective EV car buyer, there are still few concerns before taking their decision to go ahead and buy one. While the factors like environmental consciousness, brand, features, technology, design encourage them in the first place, the network of charging locations is a major factor. Based on car sales data, Tesla is clearly the market leader in EV market and perhaps the popular one. They have sold more cars than any other EV car maker. Based on surveys, EV car owners and prospective buyers consider not just the brand, features, design, or tech but most importantly the Tesla's supercharging stations network. Tesla's supercharging stations are the most reliable charging stations while the public charging stations often cause problems for the owners and not reliable most of the times. Therefore, it would be nice to build a prediction model for Tesla's supercharging locations expansion for prospective EV car buyers to confidently go ahead and make their decision.

Data Explanation

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 and few other datasets as supplemental datasets.

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 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.

Analysis

This should include the modeling analysis with accuracy and score (f1 score) calculated for all the models.

Accuracy**:** Accuracy represents the number of correctly classified data instance over the total number of data instances.

F1 Score: F1-Score is a metric which considers both precision and recall.

Precision: Positive predictive value

Recall: true positive rate

This should also include feature analysis using various methods to find the best features from the dataset. Best feature in the dataset which shows higher impact to the target variable "Purchase(likely to purchase)" compared to others present in the dataset.

Conclusion

Out of three possible models, will gauge which is the best model based on the scores to predict the purchase or the likelihood to purchase tesla car.

Will try to find the best features in the dataset to be able to come up with the prediction model so new and prospective EV car buyers can make that decision confidently.

Based on the code and exploratory data analysis and modeling, the Random forest with and without standard scalar turned out to be the best fit with high accuracy scores.

Assumptions

The datasets being considered may not have all the required features to support the model. I have taken the best possible dataset from the available sources. Also, the data for this model is being limited to United States. The data related to other countries is being excluded for this model. Also some other features which may not be relevant will also be excluded.

Limitations

The dataset considered for this prediction model is to be considered a fictional dataset as it may not represent real-world or factual data. The same goes for the supplemental datasets as well.

Challenges

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.

Future Uses/Additional Applications

While this may not exactly represent the real-world data, this model is still similar and can be run against real-world datasets to all other brands or EV manufacturers to get some useful insights.

Recommendations

This model predicts the purchase and relevant useful features that impact the likelihood of purchase with better accuracy with a caveat that the model should be regressed when more or better real-world data is available. The analysis should have been more robust with exploration and visualizations on more features.

Implementation Plan

As stated in the recommendations, this model can be implemented to predict the likelihood of an EV car purchase for a particular brand along with evaluation of other useful features that may impact the same outcome. While evaluating other features from the datasets, model must be ensured to re-evaluate for no slippage.

Ethical Assessment

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

References

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