**Domain Background**

Advancements in electric automobile industries are contributing to the disruption of the energy industry, particularly for electric utilities. Electric vehicles (EV) sales are expected to increase by [117%](http://consumerfed.org/press_release/evs-charging-sales-beating-hybrid-intro-2017-expected-record-year/) in 2017. Countries like [Norway, Netherlands, and China](http://www.euractiv.com/section/electric-cars/news/norway-spearheads-europes-electric-vehicle-surge/) plan to put many millions of EVs on the road by 2020. Growing adoption of EV means that electricity consumption for EV owners could cause dramatic and unpredictable shifts in demand.

Utilities are gathering and analyzing data on how the growing popularity of EVs could potentially stress the electric grid’s infrastructure and develop a better-informed investment strategy as a result. For example, Con Ed’s SmartCharge New York program offered to *pay*[customers 5 cents per kWh](http://www.utilitydive.com/news/new-coned-ev-program-to-reward-customers-for-off-peak-charging/440639/) to charge during off-peak hours. While energy curtailment and time-of-use rates can partially support the economics of this program, the real value comes from the data the program will collect. Analysis on these data can involve disaggregation of EV consumption from other types of energy consumption and discriminating between EV owners and non-EV owners using Advanced Metering Infrastructure (AMI) data.

**Problem Statement**

The goal of this project is to analyze AMI data to predict which residences have electric vehicles. Furthermore, the goal is also to predict when, or during which intervals, the electric vehicles are being charged.

**Datasets and Inputs**

A labeled training [dataset](https://www.gridcure.com/wp2/wp-content/uploads/2016/05/EV_files-2.zip) containing two months of energy data taken at 30 minute intervals for 1590 houses has been downloaded from GridCure’s [website](https://www.gridcure.com/).

**Solution Statement**

**Benchmark Model**

In US Patent [US 9,576,245 B2](http://pdfpiw.uspto.gov/.piw?PageNum=0&docid=09576245&IDKey=D7EB0DBA1479%0D%0A&HomeUrl=http://patft.uspto.gov/netacgi/nph-Parser%3FSect1%3DPTO2%2526Sect2%3DHITOFF%2526p%3D1%2526u%3D%25252Fnetahtml%25252FPTO%25252Fsearch-bool.html%2526r%3D1%2526f%253), Fischer et al. describes a method for identifying EV owners. The machine-learning model predicted EV ownership status after defining features in the AMI data such as increases/decreases in power load by certain pre-determined amounts (e.g. 1-2 kWh), a particular frequency of such increases/decreases, and temporal spacing between such events. HVAC energy consumption is typically the largest load in the household and is typically seen during peak hours. Understanding the overall demand curve of residential energy was also an important consideration since large loads during off-peak hours could give a clearer signal of EV ownership. These signals get further strengthened by incorporating other datasets, such as additional monitoring services in the household, information on the appliances in the household, the size of the home, the geographical location of the home, weather data, etc.

According to [Utlity Dive](https://www.utilitydive.com/news/ontario-power-providers-tap-opower-for-efficiency-dsm-offerings/411688/), Opower, Inc, the company owning the patent, processed over 40% of all residential energy consumption data and nearly two-thirds of AMI data in the US. Opower had extracted other features from these AMI data that could be fed into their models, such as [usage disaggregation](https://blogs.oracle.com/utilities/data-algorithm-smart-grid-without-smart-meters) or [load curve archetypes](https://blogs.oracle.com/utilities/load-curve-archetypes) derived from unsupervised learning techniques like k-means clustering. Finally, engaging with these residential energy customers with personalized communications allowed Opower to obtain responses to EV rebate programs that could be used to label training data and improve the model.

The scope of this project is limited due the lack of access to many of these features. Furthermore, the bias-variance tradeoffs in training the model will be quite different from the one in the patent since the dataset under consideration is many orders of magnitude smaller. In any case, the patent also does not include information on obtained accuracy of the described model.

Therefore, random assignment will be used as a benchmark to determine whether the model

**Evaluation Metrics**

An unlabeled test [dataset](https://www.gridcure.com/wp2/wp-content/uploads/2016/05/EV_files-2.zip) is also available for download. Only GridCure has the ability to score the model on the test data set. Unlike Kaggle competitions, in this case there is no automated way of submitting predictions and obtaining a test score. Therefore, this project will be evaluated solely on a validation dataset separated from the training data.

**Project Design**

Attenuation in the same power load curve, within a predetermined time frame, may indicate that charging has stopped, due to the EV battery reaching a full-charge sate.

Time between full charges? When is the battery fully charged if it was only charged at home?

Markers for when the power unloads or loads? Is the slope similar?

Probablistic score? Logistic regression? Use k-means first? Neural network? Pre-trained?

Look at the charge rate of typical EV