Eric Zan
Divya Karthikeyan

Usage Behavior in Rural Microgrids

Introduction

Analyze power consumption and customer credit data from a solar powered micro-grid in rural Kenya. Results will attempt to identify usage behaviors to help evaluate the impacts of electrification and provide input for developing experiments to encourage and expand electrical usage in a sustainable manner.

Our Solution

- Provide user types to economists
- Spike prediction

Insights expected

- User types that are defined by their behavior
- Peak load times varying for user types
- Relationship between the power usage and credit transaction behaviour

Data

The data are provided by researchers at the Center for Information Technology Research in the Interest of Society (CITRIS) at UC Berkeley. The data set contains power consumption from specific households captured every three minutes.

• House ID
• Night wattage
• Cumulative power consumption throughout day
• Current Credit amount

Challenges

- Missing values
  - Some households credit
  - Some observations missing
- Time of observation is not consistently taken at the same times
- Continuous time series data

Solution

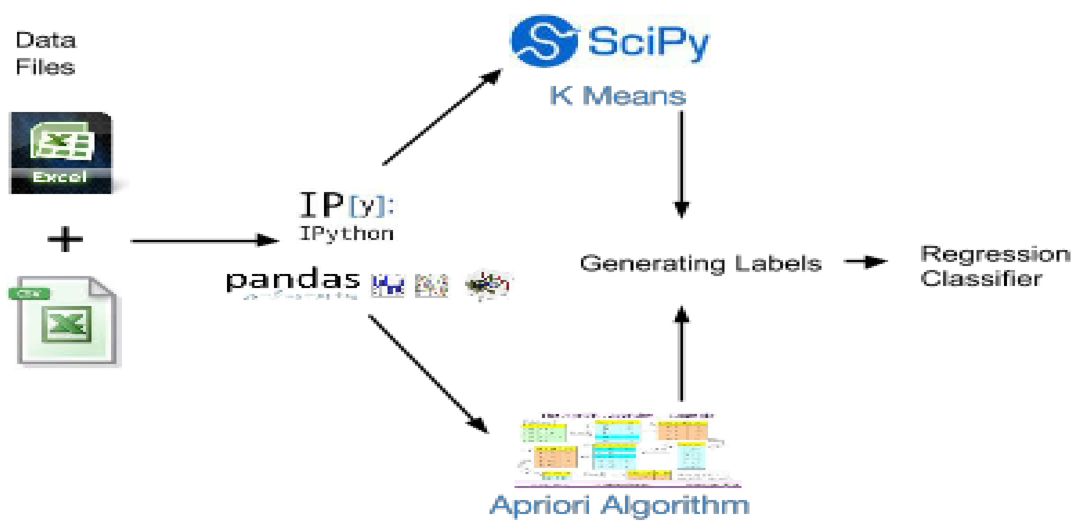
The first step was to use unsupervised learning and association rule learning methods to observe natural groups among the users in our dataset.
Apriori: Association rules among users, power consumption and credit transaction
K Means: Grouping users based on their power consumption and credit history

Eric Zan
Divya Karthikeyan

Process

K Means:

The K means clustering was performed separately on the power consumption and Credit transaction data. This optimum value of K was identified by beginning with the number of observations in the data set and doing down to 1 cluster.
Looked at graphs to determine which value of K returned best results. We observed one household to be an outlier from the others where the power usage is predominantly during the night.



Apriori:

Apriori association rule learning algorithm was used to identity rules between the Household and behaviour. Since the result only generated few rules of significance, we just used them as a means of supporting the validity of clusters we generated through K means. This result was used to aid in labelling the clusters. For eg: we saw a set of households to have a strong correlation with having a null power consumption in the night - Probably label could be *"Day users"*
Discovered data integrity issue when probabilities for night users did not match insights gained from K-means. We went back and inspected the data to find that the night wattage is not aggregated consistently.

Creating labels:

Based on the above mentioned algorithms, the users were clustered in two different sets one for their power usage trend and the other for their credit transaction behaviour and labels were created for the groups by observing the graphs.

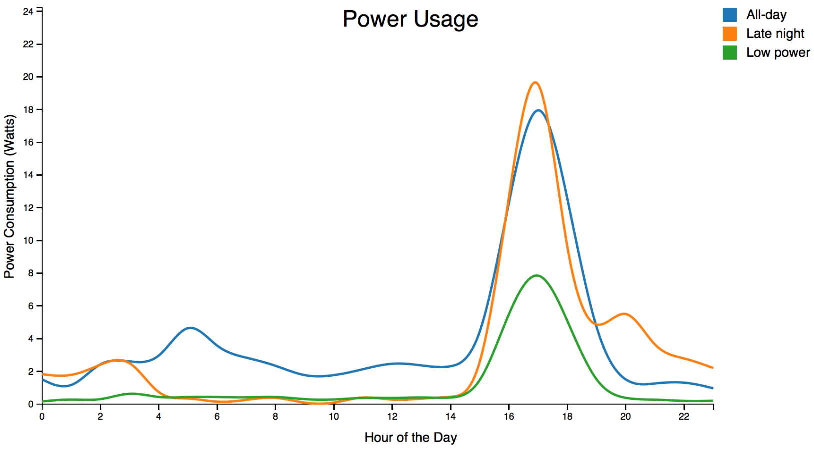
Regression:
A polynomial regression model was generated to identity the labels for User type. Since the user types are very close in behaviour and 2 peaks determine their type, the polynomial curve fit was not as accurate as expected which was a challenge during the regression prediction. The regression model requires further tuning to work as expected.

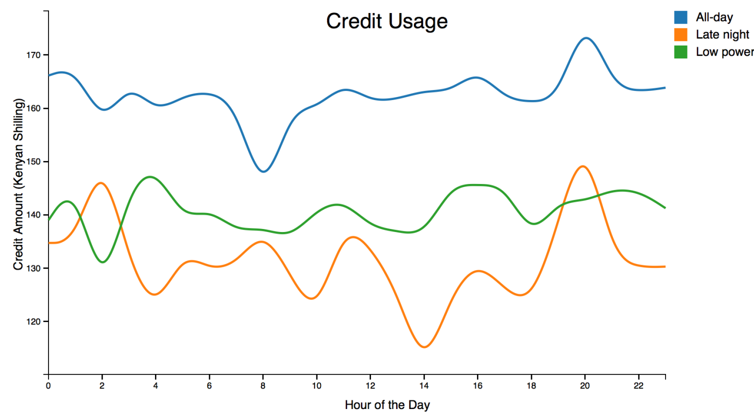
Results

We found that the households were clustered into the same groupclustwo different clusters based on power usage and credit transaction turned out to give us the same user groups. This means that the if we know the power usage of a household, we could predict the type of credit transaction behaviour they are most likely to have.

User groups based on Power usage and credit behaviour	Cluster Label
Household 1	All-day
Household 2	Low power
Household 3	Low power
Household 4	All-day
Household 5	Low power
Household 6	Low power
Household 7	Low power
Household 8	Low power
Household 9	All-day
Household 10	Late Night
Household 13	Low power

Eric Zan
Divya Karthikeyan





Based on the power usage and credit transaction the users are grouped into 3 types:

1. *All day:* User power throughout the day, peaking in the evening. They maintain a higher credit value in comparison to the other types. The credit peak occurs after their power usage has a peak value.
2. *Late night:* Minimal or zero power consumption during the day. Consumption peaks in the evening and is non zero through the night. Large variance in their credit behaviour. Peak credit additions happen before their maximum consumption in the evening.
3. *Low power:* Minimal or zero power consumption except during evenings. They maintain a relatively stable credit value in comparison with other types. Credit peaks occur during many times in the day.

Related Work

- Wang, Xiaozhe, et al. "A scalable method for time series clustering." *Unrefereed research papers* 1 (2004).
  - Helped justify rounding the time to the closest hour in order to perform time series clustering using K means since the length of each time series needs to have an equivalent euclidean distance.

- Thangsupachai, Noppol, et al. "Clustering Large Datasets with Apriori-based Algorithm and Concurrent Processing." Proceedings of the International MultiConference of Engineers and Computer Scientists. Vol. 1. 2011.
  - Apriori is used for feature identification to cluster upon in this reference. Since we wanted to see some natural clusters forming in the dataset, Apriori seemed to be a good means of identifying relationship within the dataset to help identifying features for clustering.

Future Work

- Gather weather data to cross reference usage.
- Gather user demographic data to cross reference with usage and credit.
- Regression tuning
- Determine if there is a correlation besides the time of day between the power and credit values