# Anomaly Detection for Energy Load Analysis

Jovita Chauvin, Kristen Jaskie, Andreas Spanias

***Abstract*— As household demand for energy increases there is a need to predict the amount of energy that is needed for consumers. In this paper, we used a large dataset of individual household load values gathered every hour to find patterns of consumption. We used Machine Learning algorithms for clustering and anomaly detection. Clustering algorithms are used to find normal energy usage patterns. Anomaly detection is used to find load irregularities. Through the use of these algorithms, we predict the average consumption of electricity per group of customers and detect anomalous usage.**

***Keywords*- energy analytics; machine learning; k-means clustering; anomaly detection.**

## I. Introduction

To provide energy effectively and efficiently, power companies must predict energy usage the goal of time series analytics is to identify patterns that describe the behavior of what is being analyzed, in this case, energy demand [1]. Anomaly detection finds patterns that differ from the expected [2]. Smart meters collect large amounts of utility data from their customers [3]. Energy consumption is not the same per person, nor per season [3].

K-means clustering is used to identify patterns of energy consumption by grouping consumers with similar energy usage together. The number of clusters is needed in order to lower cost and still represent the grouping of energy users [4].

There has been a trend towards smart renewable energy and energy conservation in recent years [5]. Smart meters are used to gather large amounts of data Some challenges in data collection involve errors in trying to collect the data and analyze the data accurately. For example, a misalignment with the sensors, like time delay, could lead to inaccurate data regarding the amount used and required [5]. This could be solved by analyzing larger amounts of data and grouping similar energy users together.

Anomaly detection helps determine inaccuracies in energy usage data. It could also aid in forecasting energy requirements for irregular events such as holidays. There are multiple algorithms that could aid in the detection of those anomalies. An Auto-Regressive Integrated Moving Average (ARIMA) model uses earlier data points and a random variable to forecast possible energy usage patterns [6].

To accurately determine the amount of energy usage, large amounts of energy usage data is necessary. By taking the average energy consumption of different groups of houses, we show that energy forecasting and anomaly detection are improved.

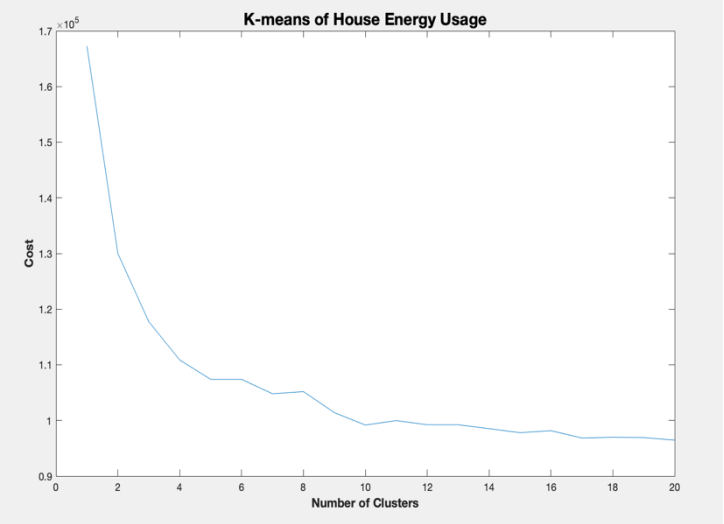


*Figure 1: K-means of London data house energy usage [7]. This graph shows that 6 clusters is the best number of clusters because the cost is low, but still represents the grouping of the data well.*

## Acknowledgment

This research is sponsored in part by NSF IRES Award number NSF IRES 185427.

## References

[1] P. Thongtra and A. Sapronova, "Time-Series Data Analytics Using Spark and Machine Learning," *Lecture Notes in Computer Science Foundations of Intelligent Systems*, pp. 509–515, Jun. 2017.

[2] M. Ved, "Outlier Detection and Anomaly Detection with Machine Learning," *Medium*, 06-Apr-2018. [Online]. Available: https://medium.com/@mehulved1503/outlier-detection-and-anomaly-detection-with-machine-learning-caa96b34b7f6. [Accessed: 13-Jun-2019].

[3] S. Singh and A. Yassine, "Big Data Mining of Energy Time Series for Behavioral Analytics and Energy Consumption Forecasting," *Energies*, vol. 11, no. 2, p. 452, Feb. 2018.

[4] Y. Wang, Q. Chen, C. Kang, and Q. Xia, "Clustering of Electricity Consumption Behavior Dynamics Toward Big Data Applications," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2437–2447, Sep. 2016.

[5] A. Marascu, P. Pompey, E. Bouillet, M. Wurst, O. Verscheure, M. Grund, and P. Cudre-Mauroux, "TRISTAN: Real-time analytics on massive time series using sparse dictionary compression," *2014 IEEE International Conference on Big Data (Big Data)*, Oct. 2014.

[6] P. Tiunov and P. Tiunov, "Time Series Anomaly Detection Algorithms," *Stats and Bots*, 08-Jun-2017. [Online]. Available: https://blog.statsbot.co/time-series-anomaly-detection-algorithms-1cef5519aef2. [Accessed: 13-Jun-2019].

[7] . R. Schofield, S. Tindemans, R. Carmichael, M. Woolf, M. Bilton, and G. Strbac, "Low Carbon London project: Data from the dynamic time-of-use electricity pricing trial, 2013," *UK Power Networks and EDF Energy*, Nov. 2015.