

# CEE 498DS Project 11: Building Energy Predictions - Project Report

This manuscript ([permalink](#)) was automatically generated from [cathyxinchangli/cee498ds-project11@f9ee654](#) on December 4, 2020.

## Authors

---

- **Xinchang 'Cathy' Li**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [cathyxinchangli](#)

Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign

- **Benjamin Smakic**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [mkbenja](#)

Department of Aeronautical & Vehicle Engineering, Royal Institute of Technology, KTH; Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign

- **Zhiyi Yang**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [zhiyiy2](#)

Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign

- **Mingyu Sun**

 [XXXX-XXXX-XXXX-XXXX](#) ·  [TBD](#)

Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign

# Abstract

---

This is the abstract. Testing to see if it will show up.

By utilizing modern electric meters, it possible to collect and store enormous amount of data about household energy consumption. This data can be used to predict energy consumption and help energy providers manage energy (electricity) output and plan for energy peaks/lows.

## Literature review

---

In the paper "Forecasting Residential Energy Consumption: Single Household Perspective" (Zhang, Grolinger & Capretz 2019), the authors attempt to predict energy consumption in residential households, with focus on single households.

According to the authors it is more difficult to predict single household energy (electricity) consumption, compared with e.g. workplace energy consumption. The reason is that single households often differ in energy consumption patterns while workplace patters tend to be more similar. Also, if big workplace buildings or multi-family residential buildings are analysed, any anomalies tend to cancel each other out (with a big enough dataset).

## Data set

The data set used originates from an electricity provider in London, Ontario, Canada. It consists of hourly smart-meter readings of electricity usage (kWh) of 15 households between 2014 and 2016.

Firstly, the data set used might not be sufficiently broad. Tracking only 15 households will most likely not capture a variety of electricity consumption patterns. However, it was deemed enough in this case. Furthermore, the data set comes from one city with a certain climate, which means that different environmental prerequisites are not considered. Perhaps using other cities from a different part of the world would lay a foundation for a more advanced ML-algorithm (Machine Learning algorithm). As it is now, the ML-algorithm might be inaccurate for other parts of the world. In addition to this, Zhang, Grolinger & Capretz state that residents of London, Ontario, Canada tend to heat their homes gas heating systems, which affects the electricity consumption drastically.

Secondly, the data set used is pre-processed in different ways. Any missing readings of electricity consumption is replaced with the average value of the previous reading and the next reading, missing weather condition is replaced with the weather condition of the previous hour etc. This is perfectly good way of replacing missing data. However, the authors do not give an explanation as to why this method was chosen, if there are any consequences and if there are other methods of making the data set complete. ### TBC

# References

---