

Machine Learning Nano Degree Capstone Project Proposal

Home Appliance Energy Prediction

1. **Domain Background** - This project is to predict energy consumption of different home appliance, so that it can be adequately monitored and planned. It's even more relevant in today's world when there is so much talk about the global warming, energy conservation and building of renewable sources for energy with minimal greenhouse effect.

There are quite a few works around this topic, both academic and otherwise, which are mentioned below:

Academic:

<http://dx.doi.org/10.1016/j.enbuild.2017.01.083>

<https://github.com/LuisM78/Appliances-energy-prediction-data>

Individual

<https://github.com/div3125/>

2. **Problem Statement** - Predict the energy consumption of the different appliance using techniques with even greater accuracy (over here better R2 score and Adjusted R2) than before which was .55 (<https://github.com/div3125/udacity-mlnd-capstone/blob/master/report.pdf>).
3. **Datasets and Inputs** - The dataset is obtained from UCI Machine Learning repository. It is donated by Luis Candanedo. His research paper and GitHub repository demonstrating his work can be viewed from links [1] and [2] respectively.

Dataset link:- <http://archive.ics.uci.edu/ml/datasets/Appliances+energy+prediction>

Luis Candanedo LinkedIn Profile:- <https://www.linkedin.com/in/luis-miguel-candanedo/>

Dataset information -

Data Points – 19375

Feature Variable - 28

Target Variable – 1

Feature Variables: -

date: year-month-day hour:minute:second

T1: Temperature in kitchen area, in Celsius

RH_1: Humidity in kitchen area, in %

T2: Temperature in living room area, in Celsius

RH_2: Humidity in living room area, in %

T3: Temperature in laundry room area

RH_3: Humidity in laundry room area, in %

T4: Temperature in office room, in Celsius

RH_4: Humidity in office room, in %
 T5: Temperature in bathroom, in Celsius
 RH_5: Humidity in bathroom, in %
 T6: Temperature outside the building (north side), in Celsius
 RH_6: Humidity outside the building (north side), in %
 T7: Temperature in ironing room, in Celsius
 RH_7: Humidity in ironing room, in %
 T8: Temperature in teenager room 2, in Celsius
 RH_8: Humidity in teenager room 2, in %
 T9: Temperature in parents' room, in Celsius
 RH_9: Humidity in parents' room, in %
 T_out: Temperature outside (from Chievres weather station), in Celsius
 Pressure: (from Chievres weather station), in mm Hg
 RH_out: Humidity outside (from Chievres weather station), in %
 Wind speed: (from Chievres weather station), in m/s
 Visibility: (from Chievres weather station), in km
 T_dewpoint: (from Chievres weather station), $^{\circ}\text{C}$
 rv1: Random variable 1, non-dimensional
 rv2: Random variable 2, non-dimensional
 Lights: energy use of light fixtures in the house in Wh

Target Variable

Appliances: energy use in Wh

Where indicated, hourly data (then interpolated) from the nearest airport weather station (Chievres Airport, Belgium) was downloaded from a public data set from Reliable Prognosis, rp5.ru. Permission was obtained from Reliable Prognosis for the distribution of the 4.5 months of weather data

4. **Solution Statement** - The solutions tried previously were mostly classical algorithms, hence in my attempt I am also exploring DeepLearning apart from classic ones to achieve better results than what has so far been.

- a. Regression Models – Simple and Polynomials
- b. Regularization Models - Ridge and Lasso
- c. Ensemble Models – RandomForest and Adaboost
- d. Deep Learning using Keras framework**

5. **Benchmark Models:** The author of the dataset used 4 models in his research which are listed below:

Multiple Linear Regression
 SVM with Radial Kernel
 Random Forest
 Gradient Boosting Machines (GBM)

Out of these 4 models, GBM was able to explain 97% of the variance in the training data and 57% in test data according the R2 score.

Training data - <https://github.com/LuisM78/Appliances-energy-prediction-data/blob/master/training.csv>

Testing data - <https://github.com/LuisM78/Appliances-energy-prediction-data/blob/master/testing.csv>

6. **Evaluation Metrics** - Metrics for Regression analysis are:

- Mean Absolute Error
- Mean Squared Error
- R2 Score

7. **Project Design:** The general sequence of steps are as follows-

- **Data Visualization:** Visual representation of data to find the degree of correlations between predictors and target variable and find out correlated predictors. Additionally, we can see ranges and visible patterns of the predictors and target variable.
- **Data Pre-processing:** Scaling and Normalization operations on data and splitting the data in training, validation and testing sets.
- **Feature Engineering:** Finding relevant features, engineer new features using methods like PCA if feasible.
- **Model Selection:** Experiment with various algorithms to find out the best algorithm for this use case.
- **Model Tuning:** Fine tune the selected algorithm to increase performance without overfitting.
- **Testing:** Test the model on testing dataset.