

Capstone Project -1

Appliances Energy Prediction

Business Problem Description – Dataset contains the house temperature and humidity conditions were monitored with a ZigBee wireless sensor network. As per the description on UCI website, each wireless node transmitted the temperature and humidity conditions around 3.3 min, Then, the wireless data was averaged for 10 minutes periods. The energy data was logged every 10 minutes with m-bus energy meters. Combining this data with the weather data based on the date time columns

To find the key feature from the dataset which contributes to the most and the

1. Predict the appliance energy consumption.
 - a. With collected data of temperature and humidity (indoor and outdoor) sensors.
 - b. Weather data collected
 - c. Fuel price over the period of time.
2. Best prediction model with best parameter for future prediction of the appliance Energy.

Dataset Details –

There are 19735 and 29 attributes. Key features from the dataset are

Columns	Description
date time	year-month-day hour:minute:second
Appliances	energy use in Wh
lights	energy use of light fixtures in the house in Wh
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 %
To	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
Tdewpoint (from Chievres weather station)	Â°C
rv1	Random variable 1, nondimensional

Weather from the nearest airport weather station (Chievres Airport, Belgium) was downloaded from a public data set from Reliable Prognosis (rp5.ru). There are 2 random variables, which has be explored more.

Another dataset used is to about the fuel price on the year of 2016, integrating this with overall dataset will help understand the fuel price impact the appliance energy consumption.

Reference data source –

- <https://archive.ics.uci.edu/ml/datasets/Appliances+energy+prediction>
- <https://www.macrotrends.net/2516/wti-crude-oil-prices-10-year-daily-chart>

Approach –

1. Load the data which is already been gone through the data processing and EDA
 1. Data cleaning and feature engineering has been performed based on previous enhancements.
2. Experiments and Implement the best performing model -
 1. We will initially create the a benchmark model, using the linear regression algorithm
 1. We will calculate the RMSE and R^2 score.
 2. Lets update the data and scale the data.
 3. Now lets create the following models with key important features –
 - Regularized linear models as an improvement over Linear Regression.
 - Ridge Regression
 - Lasso Regression
 - Ensemble based Tree Regression models, which deal with number of features and outlier data.
 - Random Forests
 - Gradient Boosting
 - Extra Trees
 - Neural networks for non-linear relationships target feature and predictors.
 - Multi-Layer Perceptron

4. Comparing the Score and visualization on the performance of different algorithm, we will select the best performing algorithm
5. Hyper-parameter tuning the best Model observed from step4 –
 - Model tuning – Using the RandomizedSearchCV, tune to model and evaluate the model accuracy by calculating the root mean square error.
6. We will find out the important features contributing from the data set
 - Will generate the visualization of the features with high to low importance of the features.
- 7 Feature and Model Evaluation- we will try to clone the above best model clone and do a prediction only with the most important feature, to verify if there is any improvement in the model accuracy.