Appliances Energy Prediction

OR568

Project Proposal Report

Prof. Jie Xu

by

Shravan Chintha

Shreyasi Thota

Meghana Tunga

Prediction Problem:

Understanding the consumption of electricity in daily home appliances such as washing machines, televisions, microwaves etc., which constitutes the major part of electricity demand of a low energy household, could provide major insights in utilization of electricity. Studying this data is important in finding the key factors that could influence the electricity consumption in appliances and thereby work on those factors to decrease the consumption of electricity by appliances. The problem predicts the appliances usage of electricity based on various factors that could influence the consumption of electricity.

The electricity consumption in low energy houses is determined by two main factors, the number of electrical appliances in the house and the usage of appliances by the occupants of the house. There are many factors that could influence the usage by appliances, some of the factors that could influence are the indoor environmental factors near the vicinity of the appliances such as temperature, humidity, light, vibrations etc. The occupancy level of house in different locations could also help in determining the usage of appliances. Developing prediction models for this problem can be useful for many applications such as detecting abnormal energy usage patterns, determining energy demand, to use in building performance simulation etc.

Dataset:

The dataset chosen for this prediction problem is **Appliances energy prediction** dataset from UCI Machine Learning Repository. This dataset helps in predicting the energy utilization in appliances. Data is collected by measuring the temperature and humidity using wireless sensor networks. Electric energy consumption in domestic buildings is characterised by the type and number of appliances and utilization of appliances by the residents.

The data was captured every ten minutes in order to record any quick changes in the energy consumption. The house temperature and humidity conditions were monitored with a ZigBee wireless sensor network. Each wireless node transmitted the temperature and humidity conditions around 3.3 min and the wireless data was averaged for 10 minutes periods. The weather data from nearest airport weather station is merged with this to evaluate its impact on energy consumption of appliances. Two random variables are used to in this dataset to test the regression model and to eliminate the non-predictive attributes.

date	Appliances	lights	T1	RH_1	T2	RH_2	T3	RH_3
1/11/16 17:00	60	30	19.89	47.59667	19.2	44.79	19.79	44.73
1/11/16 17:10	60	30	19.89	46.69333	19.2	44.7225	19.79	44.79
1/11/16 17:20	50	30	19.89	46.3	19.2	44.62667	19.79	44.93333
1/11/16 17:30	50	40	19.89	46.06667	19.2	44.59	19.79	45
1/11/16 17:40	60	40	19.89	46.33333	19.2	44.53	19.79	45
1/11/16 17:50	50	40	19.89	46.02667	19.2	44.5	19.79	44.93333
1/11/16 18:00	60	50	19.89	45.76667	19.2	44.5	19.79	44.9
1/11/16 18:10	60	50	19.85667	45.56	19.2	44.5	19.73	44.9

Attribute Information:

date time – represents the date and time when the temperature was calculated

Appliances - shows the energy utilization in Watts

lights - shows the energy utilization of light fixtures in the house in Watts

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 parent's room, in Celsius

RH_9 - Humidity in parent's 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

rv2, Random variable 2

After studying various predictor variables in the dataset, to predict the target variable Appliances, some of the regression models could be chosen. Some of the modelling techniques that could be applied to predict Appliances in this dataset are Multiple linear regression model, Random Forest algorithm, Gradient boosting machines.

Measuring performance of models:

To measure the performance of the models that are developed to predict the energy consumption, we are planning to use RMSE (Root Mean Squared Error) and R-Squared indices.

RMSE:

RMSE is defined as the square root of mean of square of all the errors. It is the standard deviation of Residuals. RMSE is most commonly used metric for numerical predictions. Since our data consists of only numerical values, it is ideal performance metric for our prediction problem.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \widehat{yi})^2}$$

R-Squared:

R2, also called coefficient of determination is the proportion of variance in the dependent variable that is predicted from the independent variable. It is used to evaluate the fit of a model by analyzing how differences in one variable can be explained by a difference in second variable.

$$R^2 = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}$$

The lower the value of the RMSE the better the model and the higher the R^2 the better the model.