Research Paper1 :

Topic: Data Driven prediction of model of energy use of appliances

Data Source:

1. Data from home-
2. sensor data from home - Temp, Humidity
3. components of energy consumption: Appliances, Light
4. Data from weather: recorded from nearest Airport

Data format:

1. Timeseries data collected over a period of 6 months
2. sensor data collected every 10 minutes

Target Parameter:

1. Energy consumption

EDA:

1. Average Electricity consumption and frequency of variety of power consumption
2. Correlation diag between electricity consumption of appliances, Room Temperature, Relative Humidity are measured.
3. Relationship between indoor temperature and weather conditions outside {Pressure, Temperature, Windspeed}
4. positive correlation energy consumption vs lights, Appliances & T2, Appliances & Outdoor Temp, NSM & Appliances
5. Negative correlation appliances & outdoor humidity

Data Filtering

1. Parameter selection based on which features are improving the accuracy of model.

Boruta Algorithm:

1. Comparision of random parameter (as base) with other parameters
2. Ranks the variables NSM: high ranked, week status- weak

Correaltion with RMSE:

1. RFE algorithm applied and dummy variables are introduced for week of day and week status. Number of optimal parameter are 34.

Performance of regression model: {LM, GBM, SVM, Random Forest are used}

1. 10 cross validation are applied for making the model robust.
2. Since the RMSE for linear model are not normally distributed hence is not accurate to use.

Four Parameter used for performance evaluation of model:

1. RMSE
2. R2
3. MAE
4. MAPE

SVR radial kernel - {sigma and cost function are tuning parameters apart from 12 parameters}

Random Forest model - Tree based model

1. Tree based on random set of predictors for removing the correlation between trees & improve the prediction.
2. predicts optimal number of trees (300) - {RMSE stops improving after that}
3. Number of optimal variables for the model calculated as 18.

GBM model:

1. Improve prediction of information from first trees.
2. Require selection of optimal paramter for number of trees(10900) & tree depth (5)

Model Selection:

1. 5 Model from 10 cross validation & 3 repeats.
2. RFE & GBM- similatr performance on RMSE & R2.
3. SVR is better than LM.

variable importance measured by residual sum of squares GBM is best.

Evaluating further GBM & variable importance:

1. subsets of variables as model variables - RMSE parameter is computed

Reseach Paper3:

Perdiction of Appliance use in smart homes:

1. The paper predicts the energy consumption on the next day. Two basic predcitors and one stochastic predictor is proposed.
2. Predictors gives better performance than other approaches.
3. Two processings are proposed to improve the prediction, segmentation, aggregation of data.
4. Data was collected from European countrries including Central and Eastern Europen Countries(REMODECE database)
5. Hourly data was analysed for appliances over a year.
6. The performance of the predictor is evaluated using e(h), which is 1 if the appliance is actually consuming the electricity,else 0. P(h) be the prediction provided by the predictor a , which is equal to 1 if the appliance is acually consuming the electricity else 0. Then with the formula the precision of the predictor is calculated for any time.

proposed algorithm for assessing a predictor a involves the following steps:

1. Set the time window dimension to n hours within the period for which the historical data was registered where n goes from 24 to 364 * 24;
2. Compute the predictions for the data corresponding to the historical sliding time window;
3. Compute the predictor precision pa(h) based on the “next day” data for all possible hours h and compute an average precision for the predictor.

Prediction with different predictors:

1. "Will Always cosume" and "Will never consume predictors" It refers to the probability of the appliance will consume and vice versa

ARMA predictor:

1. In the algorithm the current value of a time variable is made a function of its past values and is expresses as sum of weights.
2. This ARMA model was used in order to predict the next day energy consumption.

Proposed Predictor:

1. Proposed predictor specifies the probability of the appliance to consume on an hourly base

Improving the Prediciton precison:

1. temporal segmentation, that considers each day of the week as a partition was done.
2. the hourly predictions are made considering the proposed predictor. A k-Means clustering algorithm is applied in order to group the similar consumption days.
3. Each cluster is defined by its cluster center and clustering proceeds by assigning each of the input data to the cluster with the closest centre.

Prediction precison after clustering:

1. After applying the iterative k-Means algorithm, two clusters are obtained. In the presented case, cluster C1 groups weekdays data and cluster C2 gathers Saturday and Sunday data

Research Paper 2:

1. Engineering and hybrid approaches use thermodynamic equations to estimate energy use, the AI-based approach uses historical data to predict future energy use under constraints.
2. Ease of use and adaptability to seek optimal solutions in a rapid manner, the AI-based approach has gained popularity in recent years.
3. Approaches for building energy use prediction, conducts an in-depth review of single AI-based methods such as multiple linear regression, artificial neural networks, and support vector regression, and ensemble prediction method
4. Combining multiple single AI-based prediction models improves the prediction accuracy manifold. This paper elaborates the principles, applications, advantages and limitations of these AI-based prediction methods and concludes with a discussion on the future directions of the research on AI-based methods for building energy use prediction.