International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

Data Mining for Predicting Electricity Consumption by Classification

Akhil Anilkumar

Student, Dept. of Dual Degree Computer Applications, Sree Narayana guru institute of science and technology, Kerala, India

---***---

Abstract - Data analysis can be applied to power consumption data for predictions that allow for the efficient scheduling and operation of electricity generation. This work focuses on the parameterization and evaluation of predictive algorithms utilizing metered data on predefined time intervals. More specifically, electricity consumption as a total, but also as main usages/spaces breakdown and weather data are used to develop, train and test predictive models. A technical comparison between different classification algorithms and methodologies are provided. Several weather metrics, such as temperature and humidity are exploited, along with explanatory past consuming variables. The target variable is binary and expresses the volume of consumption regarding each individual residence. The analysis is conducted for two different time intervals during a day, and the outcomes showcase the necessity of weather data for predicting residential electrical consumption. The results also indicate that the size of dwellings affects the accuracy of model. , we applied a documented classification methods, that is Support Vector Machines (SVM), Random Forest, Stochastic Gradient Descent (SGD) and Logistic Regression, artificial neural network methods on the human resource data.

Keywords— neural network, Logistic Regression, SVM, stochastic gradient descent, Random Forest

1. INTRODUCTION

The system could provide residents with forecast electricity consumption as accurately as possible, so they can plan their activities wisely. The goal is to decrease the power consumption and bill. Energy consumption prediction in a particular building is usually influenced by many factors, such as the electrical appliances or devices in it, its geographic location, as well as the time range it is operational it is essential to guarantee that there is a constant tracking of the total consumption in each home

"Electric demand often considered as a function of weather variables and human social activities". Families have cycles of consumption on a daily and weekly basis If the consumption stays low for a consecutive number of days, it is more likely that next days will show a rise in total consumption. Similarly, the previous day consumption should also be a factor to predict the next day. House that consumes a higher than average amount of electricity, that the next day would result in lower consumption.

The first part focuses on exploiting the gathered weather data to produce building energy consumption forecasting models while the second aims to simulate the behavior of grid through the aggregation of consumption of all available homes. The selected dataset, named Home, includes both weather and consumption data collected for seven different householdsIn general, data mining is used to obtain new perspectives and capture hidden factors from unexploited information, which is available in the collected data

2. METHODOLOGY

Data sets

The selected dataset, named Home, includes both weather and consumption data collected for different household on a particular year .The labels of the Homes are identical to those introduced by the repository. Therefore, the homes will be referred as Home B, Home C and Home F. The ST load forecasting used for reducing costs and secure operation of power systems. In MT load forecasting the interest focuses on normal operation, while LT load forecasting is studied to ensure safer investments and long-term planning. , data mining is used to obtain new perspectives and capture hidden factors from unexploited information, which is available in the collected data

Data classification techniques

Data classification is that the process of organizing data into categories for its best and efficient use. Data classification techniques are methods that are Support Vector Machines (SVM), Random Forest, Stochastic Gradient Descent (SGD), Logistic Regression and artificial neural network.

Neural network

neural networks define to systems of neurons, either organic or artificial in nature.neural network is a computational learning system that uses a network of functions to understand and translate a data input of one form into a desired output,The concept of neural networks, which has its roots in artificial intelligence usually in another form. The concept of the artificial neural network was inspired by human biology and the way neurons of the human brain function together to understand inputs from human senses Neural networks adapt to changing input;A neural network is a series of algorithms that endeavors to identify underlying relationships in a set of data through a process that mimics the manner the human brain operates. during this sense then network generates the best attainable result without needing to regenerate the output criteria.

Random Forest

Random forest is a supervised learning algorithm the "forest" it builds, is an associate ensemble of decision trees, often trained with the “bagging” method. Random forest, consists of a large individual decision trees that operate as an associate ensemble. The overall idea of the bagging method is that a combination of learning models increases the result. it can be used for classification and regression problems.

Logistic Regression:

It is a statistical technique for inspecting a dataset in which there are one or greater impartial variables that decide an outcome. For instance hours of reading will increase then the chance of passing checks will increase

Support Vector Machine (SVM)

It performs classification by finding the hyperplane that completely separates the vector into two non-overlapping classes. The vectors that outline the hyperplane are the support vectors.

Predicted data

By this complete evaluation we discover out the excellent data which discover the monthly and hourly power consumed data by the home that we consider.

3. EXPERMENTAL ANALYSIS

In the present structures they used solely few of facts mining methods for facts prediction. This work focuses on the parameterization and evaluation of predictive algorithms utilizing metered data on predefined time intervals. More specifically, electricity consumption as a total, but also as main usages/spaces breakdown and weather data are used to develop, train and test predictive models. Prediction model which includes usage patterns besides weather data. As long as the target is to predict the fluctuation of consumption of electricity in Home, In the proposed structures we use different algorithms that is Support Vector Machines (SVM), Random Forest, Stochastic Gradient Descent (SGD), Logistic Regression and artificial neural network. The following figures give an indication of the consuming behaviors for each of them. along the average hourly consumption, it is important to also examine the monthly averages.

Date & Time	use [kW]	gen [kW]	Furnace HRV [kW]	CellarOutlets [kW]	WashingMachine [kW]	FridgeRange [kW]	DisposalDishwasher [kW]	KitchenLights [kW]
01-01-2014 0.00	0	0	0.195337778	0.083204444	0.005686111	0.006891667	0.005568889	0.012153889
01-01-2014 0.30	0	0	0.182158333	0.036138889	0.005678889	0.094138333	0.005411667	0.0052
01-01-2014 1.00	0	0	0.134808333	0.047033889	0.005635	0.014786111	0.00551	0.003173333
01-01-2014 1.30	0	0	0.182125	0.071406667	0.005671667	0.082081111	0.005445	0.003071667

figure1 :A sample dataset of energy

The figure1 will be displayed sample energy Consumption of particular home at different time intervals our assumptions regarding the level of information that residents would be interested in. Ideally the outcome should be the exact consumption value, but this research is conducted with a broader motivation to suggest and highlight the factors that could affect a prediction either positively or negatively on individual home characteristics.

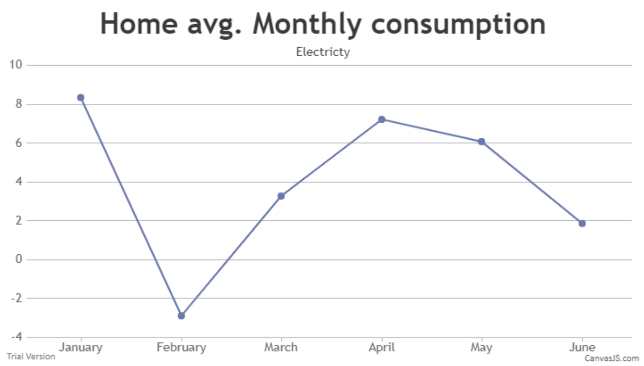

figure2 :average monthy consumption of home

Figure 2 will be displayed average energy Consumption of particular home at monthly basis regarding our data Its an energy consumption result obtain by classification

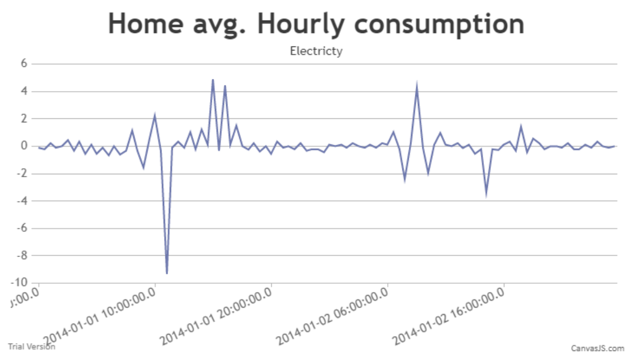

figure3: hourly consumption of month

Figure3 the hourly on a month which predict the fluctuation of consumption in explicit way so they can plan their activities wisely

4. CONCLUSION

This research was conducted mostly with a citizen centric approach; regarding the algorithms used, the basic conclusion is that the complexity of the model was not as high as to allow one of them to stand out. All the results are very close and sometimes identical.

The accuracy of the model increases when the consumption follows similar paths throughout the months Home C is twice as big as Home B, and this may indicate that the bigger a home is, the more difficult it is to predict its consumer behavior. Occupancy in such larger houses can also deviate much more in comparison to typical-size homes. Transforming the consumption from a real number into a binary class achieves better results when the point of division is just the mean of all instances. Next in performance is when the division is based on a seasonal mean and finally on a monthly mean value; the first two seems to be the more reasonable Every research 'suffers' from threats that might question the validity of the approach and its results. In this case, a main threat is that of sufficiently deseasonalising data. The factor of time could possibly be analyzed to more explanatory variables. Another significant factor is access to accurate weather data. Since the model examines if weather data can accommodate forecasting, these data should be accurate. However, obviously for forecasting on day-ahead period the weather data themselves are also acquired via prediction. Since the problem is not treated as a time-series one, it is not clear which method against over-fitting is more appropriate. Research has shown that decision tree classifiers improve generalization accuracy via pre-pruning In general, in time-series problems there is a different way to use cross-validation than the classic

ACKNOWLEDGEMENT

In the name of almighty, I would like to extend my heartfelt thanks to our HOD Mrs. Kavitha C.R, Department of a Dual Degree Master of Computer Applications for the helps extended to me throughout my course of my study. I am deeply grateful to my guide Mrs. prajisha Assistant Professor, Department of a Dual Degree Master of Computer Applications for the valuable guidance

REFERENCES

[1] D. Chen, D. Irwin (2017) Weatherman: Exposing Weather-based Privacy Threats in Big Energy Data, 2017 IEEE Int'l Conf. Big Data (Big Data).

[2] A. Mishra, D. Irwin, P. Shenoy, J. Kurose, T. Zhu (2012) SmartCharge: Cutting the Electricity Bill in Smart Homes with Energy Storages, e-Energy '12 Proc. 3rd Int'l Conf. on Future Energy Systems: Where Energy, Computing and Communication Meet, Article No. 29.

[3] S. Barker, A. Mishra, D. Irwin, P. Shenoy, J. Albrecht (2012) SmartCap: Flattening Peak Electricity Demand in Smart Homes, 2012 IEEE Int'l Conf. Pervasive Computing and Communications

[4] O.M. Longe, K. Ouahada, S. Rimer, H. Zhu, H.C Ferreira (2015) Effective Energy Consumption in Smart Homes, AFRICON 2015.

[5] A. Azadech, S.F Ghaderi, S. Sohrabkhami (2007) Forecasting electrical consumption by integration of Neural Network, time-series and ANOVA, Applied Mathematics and Computation, Vol. 186, No. 2, pp. 1753-1761.

[6] A. Ahmad, M. Hassan, M. Abdullah, H. Rahman, F. Hussin, H Abdullah, R. Saidur (2014) A review on applications of ANN and SVM for building electrical energy consumption forecasting, Renewable and Sustainable Energy Reviews, Vol. 33, pp. 102-109

[7] A. Sauhats, R. Varfolomejeva, O. Linkevics, R. Petrecenko, M. Kunickis, M. Balodis (2015) Analysis and prediction of electricity consumption by smart meter data, IEEE 5th Int'l Conf. Power Engineering, Energy and Electrical Drives (POWERENG).

[8] X. Li, C.P. Bowers, T. Schnier (2010) Classification of Energy Consumption in Buildings With Outlier Detection, IEEE Transactions on Industrial Electronics, Vol. 57, No 11.

[9] M. Espinoza, C. Joye, R. Belmans, B.D Moor (2006) Short-Term Load Forecasting, Profile Identifications, and Customer Segmentation: A Methodology Based on Periodic Time Series, IEEE Transactions on Power Systems, Vol. 20, No 3.

[10] J.Contreras, R. Espinola, F. Nogales, A. I. Conejo (2003) ARIMA Model to Predict Next-Day Electricity Prices, IEEE Transactions on Power Systems, Vol. 18, N