Heating and hot water industrial prediction system for residential district

Luc Dufour, Dominique Genoud [1]
Bruno Ladevie and Jean-Jacques Bezian [2]

1 Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland
2 UMR CNRS 5302,Mines-Telecom Albi, France
Email:(luc.dufour, dominique.genoud, antonio.jara)@hevs.ch,
(bruno.ladevie and jean-jacques.bezian)@mines-albi.fr

***Abstract*—This work presents a data-intensive solution to predict heating and hot water consumption. The ability to predict locally those flexible sources considering meteorological uncertainty can play a key role in the management of microgrid. A microgrid is a building block of future smart grid, it can be defined as a network of low voltage power generating units, storage devices and loads.The main novelties of our approach is to provide an easy implemented and flexible solution that used a supervised learning techniques. This paper presents an industrial methodology to predict heating and hot water consumption using time series analyzes and tree ensemble algorithm.The results are based on the data collected in a building in Chamoson (Switzerland) and simulations. Considering the winter season 2012-2013 for the training,the heating and hot water predictions is correctly estimated 90% +/- 1.2 for the winter season 2013-2014.**

Keywords—Heating consumption prediction; Hot water consumption prediction; Data intelligence analysis; Storage; Microgrid; Predictive energy management; KNIME;

I. Introduction

The insertion of variable production from wind and photovoltaic increases pressure on the network. For example, PV is the fastest-growing energy technology since 2002 with an average increase of 48% [1]. The majority plants are grid-connected systems and a high penetration of PV (like in the case of islands) raises issues for the grid-operators. Subsequently, forecasting the power output of the PV plant is necessary to assure the grid stability. The majority of the new installations are low powers and connected at low level voltage. For different point of views, it's interesting to pilot an part of energy consumption to manage the voltage and the frequency or to increase the autonomy of a district, to sell energy.The problematic is to predict locally the energy flux to anticipate the variations of the consumption and the decentralized production. Several intelligent districts were created, piloted most of the time by national energy company or are installed on campus academics. The aim is to detect the humans presence to increase the prediction.The models used the piloting and the predictions often remain the property of the projects partners [2]. We can cite the intrusive model [3] [4]. The neural networks are widely used to this purpose and manage these. [5] [6].

In this paper, two levels of prediction is realized. An anticipation level provide an prediction for one day every hour. An energy software enable to compute the different simulations for this anticipation level. The reactive level provide only a prediction for the next hour. The global consumption is not analyzed in this paper. A focus is done on the hot water and the heating. These two elements are a high potential of piloting. For this reactive level, the supervised learning techniques with the tree ensemble enable the heating and hot water consumption prediction for the next hour.a one hour prediction is computed for the heating and hot water consumption. Evaluation of the results and relevance of the predictions made on the basis of reference models such as detailed thermal models for heating . These reference models need a priori informations about the different buildings: activity sector, construction year, surface, people number. In this paper, the prediction methodology is a supervised learning techniques with a Tree Ensemble algorithm.
The paper is organized as follows. In Section II the prediction methodology is presented. The Section III describe the data set and the different test . In Section IV we present the results of the tree ensemble algorithm. Finally we conclude and discuss future directions of research in Section V.

II. Methodology

The reactive piloting consists to control the consumption of certain devices. It is a question to affect the energy by respecting the constraints of real time resource.

A. *Energy problem characterization*

Global equations: In this work,the energy prediction is computed every one hour for the heating and the hot water consumption. For a day j with a step of time of one hour, we have 24 predictions to realize.
The energy need for a district is the sum of the energy need for the sub-systems u at the moment i, for a D-Day, day of the week k, for one year n :

$$B(t_i)_{,j,k,n} = \sum_{i=1,j=1,k=1,u=1}^{i=24,j=n,k=7,u=n} B(t_i)y_u$$

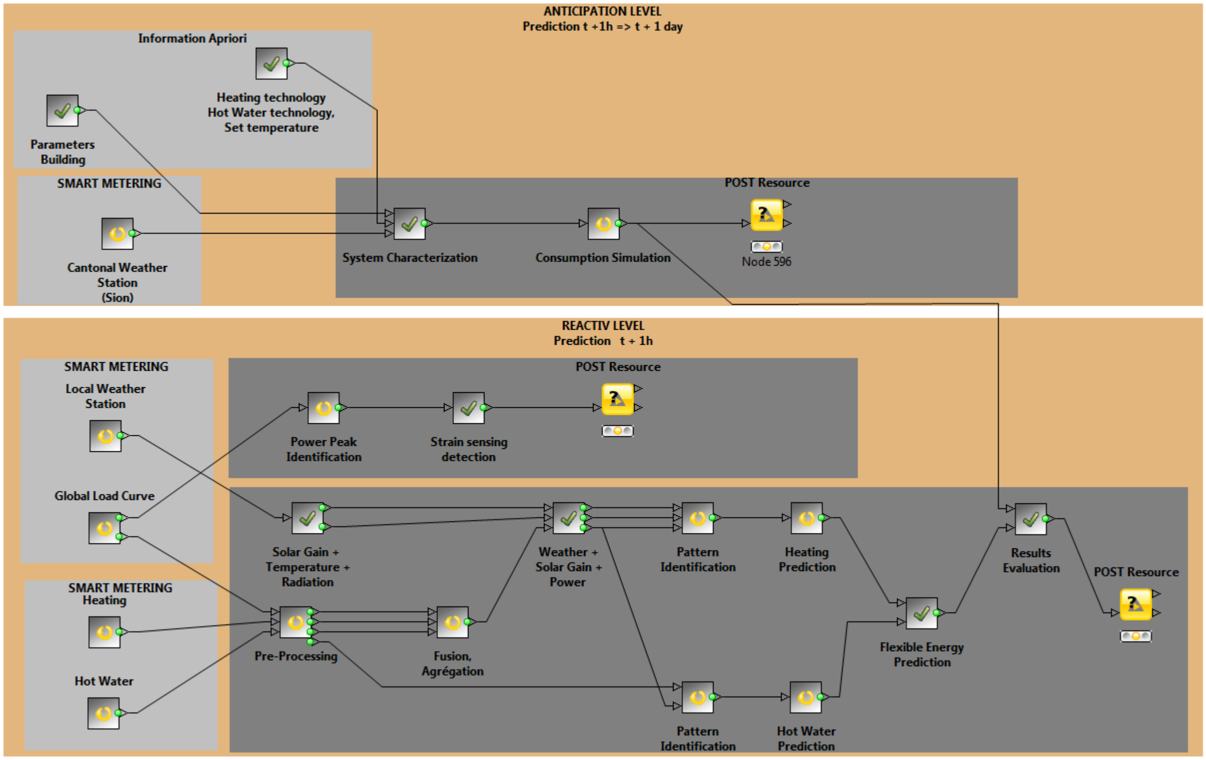

Fig. 1. Steps of the methodology implemented in KNIME

The total electric consumption is the sum of the electrical devices less the energy produced by the Renewable Decentralized Energies P_{RDE} less the available electric storage $P_s(t_i)y_u$:

$$C(t_i)_{,j,k,n} = \sum_{i=1,j=1,k=1,u=1}^{i=24,j=n,k=7,u=n}(\sum_{u=m,p=l}^{u=1,p=1} C_{devices}(t_i)) - P_{RDE}(t_i) - P_s(t_i))y_u$$

During every period of sampling k, every flexible equipment must be activated or deactivated. This decision is naturally expressed by a binary variable X (p,i). X (p,i) = 1 if the equipment p is activated during period i, 0 if the equipment is deactivated. In this paper, we consider two equipments which is shifted during a moment : the heating and the hot water.

Heating Equation: The heating energy need is computed by the difference between the thermal heat loss and the energy gains. The thermal heat loss are by the envelope of the building $D_{Structure}(t_i)$ and by the ventilation $D_{Ventilation}(t_i)$ and the gains connected to the heat cleared by the occupants $G_{HumansGain}(t_i)$, the contribution of the solar gains $G_{SolarGain}(t_i)$, the cleared heat devices $G_{DevicesGain}(t_i)$ and by the contribution of energy stored in the house $D/G_{Storage}$:

$$B_{Heating/Cold}(t_i)y_u) = D_{Structure}(t_i) + D_{Ventilation}(t_i) - G_{Humans_Gain}(t_i) - G_{Solar_Gain}(t_i) - G_{Devices_Gain}(t_i) \pm D/G_{Storage}(t_i)y_u$$

Hot water Equation : The hot water energy consumption is the electrical consumption $CHW(t_i)$ multiplied by the energy efficiency R_{HW} less the production provided by the thermal panel.

$$CHW(t_i) = (R_{HW} * CHW(t_i) - C_{RDE_{Thermal}}(t_i))y_u$$

B. Data Pre-Processing

The power consumption varies 20% for reasons external and the load does not provide an ideal signature [7]. The load admittance, Y(t), can be calculated every second from the measured power P(t) and the voltage : $P_{Norm}(t) = 220^2 Y(t) = (220/V(t))^2 \times P(t)$ [7].
To reduce the effects of noise and transients on the measured data, we apply a median filter of size 10 to the time series of measured values.We have empirically determined the optimal size of the filter via a set of tests, by attempting to balance the elimination of spurious load level variations and the performance of the event detection algorithm, in terms of average number of undetected events.
After this step, the input data are three median values every second for the voltage, three for the amperage, three for the active and reactive power, three for the power factor and the accumulated active energy and a time reference for the global load curve. The edge identification method is computed which consists on detecting variations of active and reactive power with regards to a predefined threshold [8] [9]. The different work on this subject and our analysis working enable to know the different characteristics to describe a device [13], [12] [16].
In output, a list of active and reactive powers events by phases is created. This events are computed every 30 second to detect the consumption electricity peak. If the system has a renewable decentralized energy, it's possible

to apply the same methodology to identify the power peak production.

C. Technical constraints followed

In a contract with a supplier of energy, the maximum power which can be consumed is fixed. The subscription will be proposed by slices(edges) of consumption 6kW, 9kW, 12kW..In this case, the consumption in the building has to respect a strong constraint:

$$P(t_i)y_u \leq P_{max}(t_i)y_u$$
$$U(t_i)y_u \leq U_{max}(t_i)y_u$$

The objective of the control is to respect the constraint of energy capacity every reactive period i. $P_{ava}(i)$ is the available maximal power for the period i after subtraction of the flexible devices. The consumption of all the equipments does not have to exceed the power $P_{ava}(i)$:

$$\sum_{i=1,j=1,k=1,u=1}^{i=24,j=n,k=7,u=n} P(u) \times X(t_i, p) < P_{ava}(i)$$

D. Prediction Model

Heating input data: In our case, the aim is to estimate the heating consumption values every hour at time t(i). The input data are the heating and the global consumption at the time t(i-n). In these values, we have the difference between the thermal heat loss and the energy gains at the time t(i-n). The time enable to consider the humans presence in the building with the hour, the day of year, the month and the difference between the week-end and the week Day. Its important to note here that a model is created for the week-end and an another model is created for the weekday. A Boolean variable is added to identify the children holidays and an other for the public holidays. The weather data are added with the real outside temperature t(i-n) and the predicted outside temperature t(i). The weather station is based in Sion.

Hot Water input data: The input data are the hot water consumption at the time t(i-n). The hour, the day of year, the month and the difference between the week-end and the week Day are added. Its important to note here that a model is created for the week-end an another model for the weekday. a Boolean variable is added to identify the children holidays and the public holidays.

Mathematics prediction model : The Tree Ensemble Learner builds an ensemble of decision trees, as a variant of the random forest. Each of the decision tree models is trained on a different subset of rows and/or on a different subset of columns, randomly selected at each iteration. The output model is then an ensemble of differently trained decision tree models. The Tree Ensemble Predictor applies all decision trees to each data row and uses the simple majority vote for prediction. The Tree Ensemble Learner requires many settings, reflecting its complexity. Indeed, it has three setting tabs in the configuration window: one to select the attributes (that is the data columns), one to select the decision trees training parameters (entropy measure...), and one to select the ensemble settings.

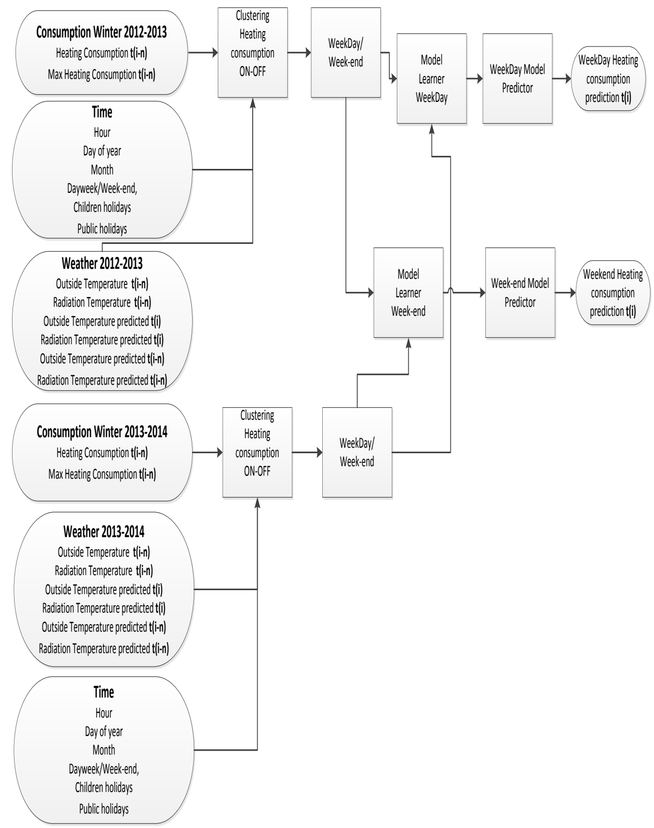

Fig. 2. Heating input datas for the prediction model

E. Performance prediction evaluation

ROC Curve : In order to create a ROC curve for a model, the input table is first sorted by the class probabilities for the positive class i.e. rows for which the model is certain that it belongs to the positive class are sorted to front. Then the sorted rows are checked if the real class value is the actually the positive class. If so, the ROC curve goes up one step, if not it goes one step to the right. Ideally, all positive rows are sorted to front, so you have a line going up to 100% first and then going straight to right. As a rule of thumb, the greater the area under the curve, the better is the model.

Anticipation level : The anticipation level used an energy software to predict the energy flux every hour for one day. This prediction enable to create an energy plan resources for the energy supplier. A comparison is realized between this reference model and our methodology. To build this model, we used the maximum of a priori information's in order to have the best simulation model. The software is Pleiade Comfie.

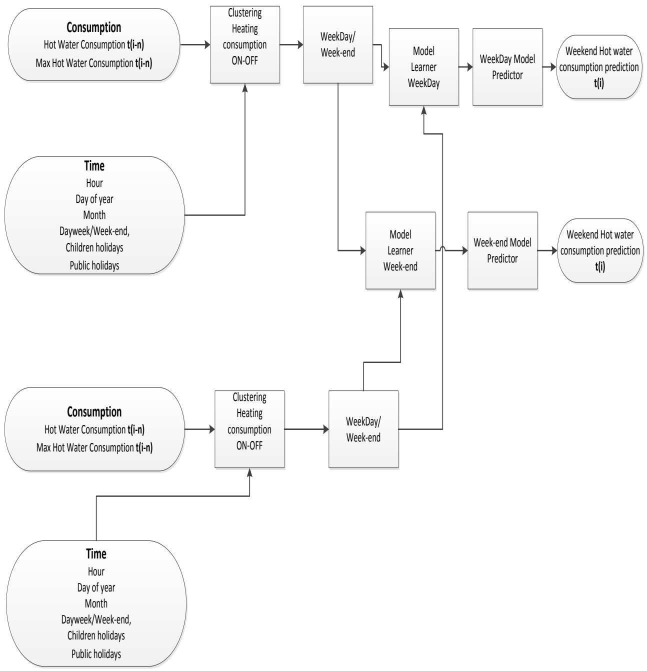

Fig. 3. Hot Water input datas for the prediction model

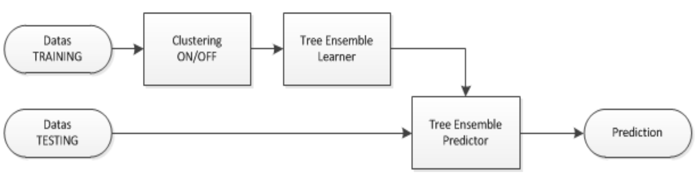

Fig. 4. Prediction Model

III. Experimental setting

In this section, we describe how our information system collects data, how the parameters are used in our analysis software and how the data set for prediction model.

A. Information system presentation

Different connections are tested between database (MongoDB, My SQL, Axibase...) and our data analysis software KNIME [24]. The connection is realized by the REST HTTP request directly with the data base or through an API. The predictions are sent by the POST HTTP request in JSON format.

A Schneider electric system collect the active and reactive power by phase from the global electric meter [22].This smart meter is the PowerLogic Series 800 PM810 of the Schneider Electrical Company [23]. As outputs, we have the amperage, voltage, active and reactive power and energy consumed with in a one second interval per phase. The same smart metering to collect the heating consumption. Plugs or Geroco [25] smart meters collect the hot water consumption with zig bee communication protocol.The data travels through a mod-bus communication.

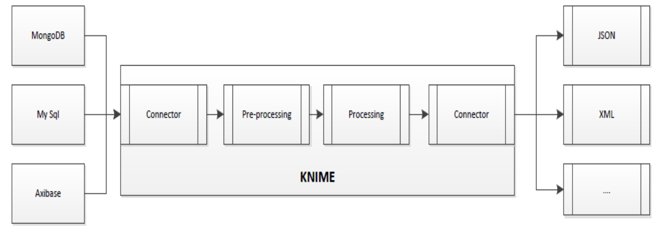

Fig. 5. Information flux through our analysis software

B. Building Characterization

The test house was built in 2000 and is isolated well. The house has a surface of $250m^2$. The heating system is a heat pump air-water with a 3kW nominal power. Its energy efficiency coefficient is 3.5. The hot water is too provided by a heat pump air-water. The energy efficiency coefficient of this heat pump is 3.7.

C. Dataset

The Schneider electric system collect the amperage, the voltage, the active and reactive power every second per phase. These information are available for the global meter and the heating. For the hot water, only the active power is collected. The data set begin in October 2012. The winter season 2012-2013 is the training data set. The testing data set is the winter season 2013-2014. Five years of data are available for the weather station in Sion. In a second test, the simulation heating consumption are the input data. The simulation heating consumption during the winter season 2012-2013 is the training data set. The real consumption during the winter season 2013-2014 is the testing data set.

IV. Classification results

When the variable under study cannot distinguish between the two groups, i.e. where there is no difference between the two distributions, the area will be equal to 0.5 (the ROC curve will coincide with the diagonal). When there is a perfect separation of the values of the two groups, i.e. there no overlapping of the distributions, the area under the ROC curve equals 1 (the ROC curve will reach the upper left corner of the plot). 30 Decision tree are created for this model. For the heating weekday prediction model, the area under curve is 0.9285. For the heating weekday prediction model, the area under curve is 0.8891. A statistics table on the attributes is created in the different tree learners. Each row represents one training attribute with these statistics: splits (level x) as the number of models, which use the attribute as split on level x (with level 0 as root split). Only the variables present in the level 0 are presented. The heating consumption the previous hour, the solar radiation predicted in the time t - 7 hour and the hour

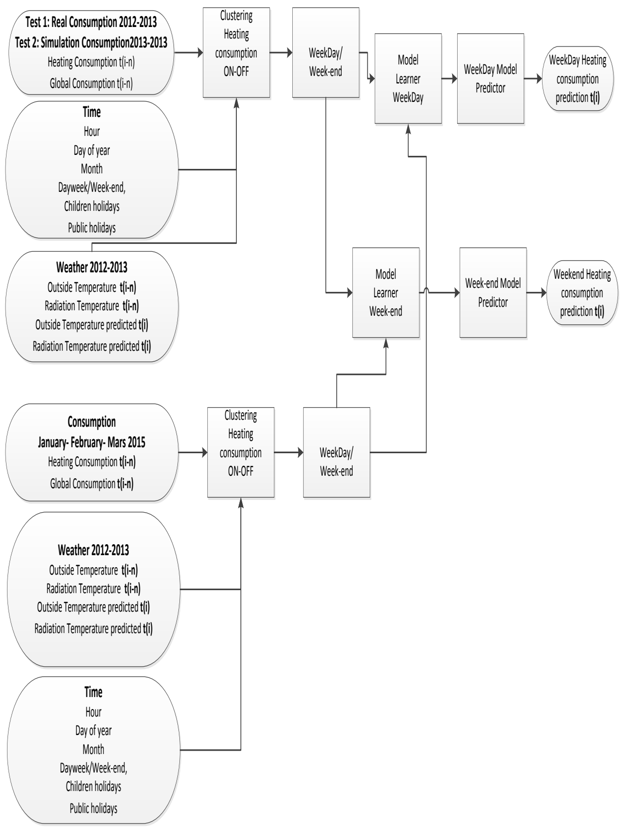

Fig. 6. Heating prediction test

Variables	Level 0	Level 1	Level 2
Solar Radiation Predicted (t-7)	5	2	1
Heating consumption (t-1)	4	3	10
Hour	2	3	5
Outside temperature (t-2)	2	2	0
Solar Radiation (t-9)	2	1	2
Heating consumption Max (t-1)	2	5	3
Outside temperature predicted (t)	1	2	3
Outside temperature (t-1)	1	1	2
Outside temperature (t-3)	1	1	1
Outside temperature (t-4)	1	2	0
Solar Radiation (t-8)	1	0	1
Solar Radiation (t-10)	1	2	4
Outside temperature predicted (t-1)	1	2	1
Solar Radiation predicted (t-5)	1	2	1
Solar Radiation predicted (t-6)	1	3	3
Solar Radiation predicted (t-8)	1	2	0
Solar Radiation predicted (t-9)	1	1	3
Solar Radiation predicted (t-10)	1	1	1

TABLE I. STATISTICS TABLE ON THE ATTRIBUTES USED IN THE DIFFERENT TREE LEARNERS FOR THE HEATING WEEKDAY PREDICTION MODEL

are the determined variable for the heating consumption prediction.

30 Decision tree are created for the hot water prediction model. For the heating weekday prediction model, the area under curve is 0.9191. For the hot water weekend prediction model, the area under curve is 0.8994. The hour and the hot water consumption the previous hour and the maximum of the hot water consumption in the time t - 6 hour are the determining variables for the heating consumption prediction.

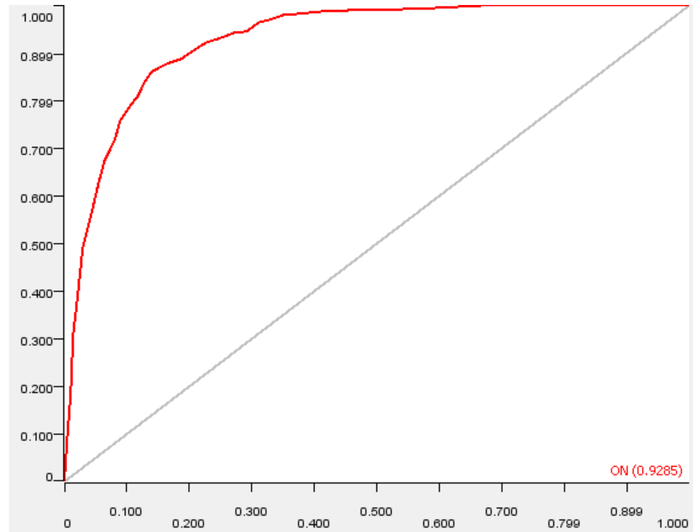

Fig. 7. ROC Curve for the heating weekday prediction model

Variables	Level 0	Level 1	Level 2
Hour	5	5	10
Hot water consumption (t-1)	5	3	15
Hot water consumption max (t-6)	5	1	3
Hot water consumption (t-4)	3	6	3
Hot water consumption (t-6)	3	4	7
Hot water consumption (t-5)	2	5	3
Hot water consumption max (t-1)	2	10	12
Hot water consumption (t-2)	1	2	3
Hot water consumption (t-7)	1	2	4
Hot water consumption max (t-2)	1	0	3
Hot water consumption max (t-3)	1	7	6
Hot water consumption max (t-5)	1	2	3

TABLE II. STATISTICS TABLE ON THE ATTRIBUTES USED IN THE DIFFERENT TREE LEARNERS FOR THE HOT WATER WEEKDAY PREDICTION MODEL

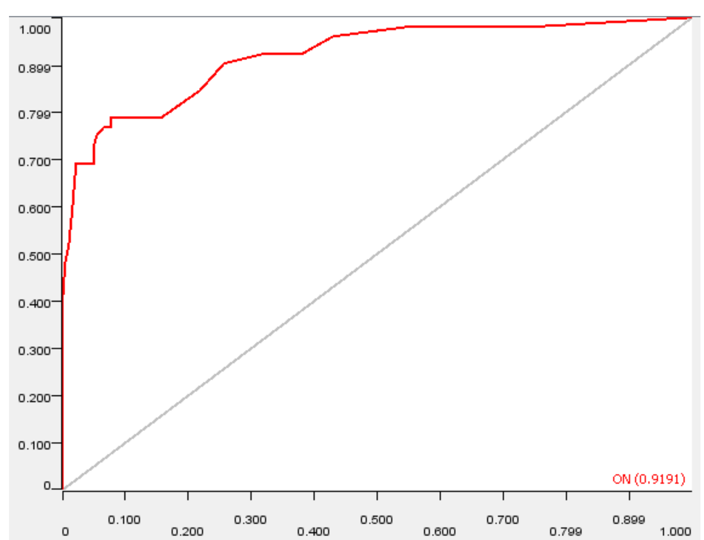

Fig. 8. ROC Curve for the hot water weekday prediction model

V. Conclusions

The prediction work in our approach focuses on the heating and the hot water consumption. Using an ON-OFF clustering and tree ensemble algorithm on a dataset from the winter season 2012-2013, the heating and the hot water consumption prediction is estimated correctly around 91% +/- 1.2 for the winter season 2013-2014. For the heating prediction model, the hour, the heating consumption the previous hour, the maximum heating consumption the previous hour and the solar radiation predicted are the determining variables. For the hot water prediction model, the hour, the hot water consumption the previous hour are the determining variables. The results can be improved if the local weather data is given as an input.

VI. acknowledgements

This research is a part of the I-BAT project and was financed by The Ark Energy. It is too the collaboration through a thesis between the Mines Telecom Albi with the RAPSODEE research institute (CNRS UMR 5302) and the HES SO Valais with the information system research institute in Sierre.

References

[1] M. Cococcioni, E. DAndrea, B. Lazzerini, One day ahead forecasting of energy production in solar photovoltaic installations: An empirical study, Intelligent Decision Technologies 6 , 2012

[2] N.W.A. Lidula, A.D. Rajapakse,ELSEVIER (2010) Microgrids research: A review of experimental microgrids and test systems. *Renewable and Sustainable Energy Reviews 15 (2011) 186202*.

[3] Victoria Moreno, Antonio F Skarmeta, Antonio J Jara, How to intelligently make sense of real data of smart cities, IEEE 2015

[4] Victoria Moreno, Luc Dufour, Antonio F Gmez Skarmeta, Antonio J Jara, Dominique Genoud, Bruno Ladevie and Jean-Jacques Bezian, Big data: the key to energy efficiency in smart buildings, 2015, Soft computing manuscrit.

[5] Simon . H ,Neural networks: a comprehensive foundation,Prentice Hall,1999

[6] Lane, Stephen H and Flax, Marshall G and Handelman, David A and Gelfand, Jack J,(1990)Multi-layer perceptrons with B-spline receptive field functions,pp 684-692

[7] Hart, G.W. (1992) Nonintrusive appliance load monitoring. *Proceedings of the IEEE*. vol. 80, no. 12, pp. 1870-1891.

[8] L. Dufour,D. Genoud, B. Ladevie,JJ. Bezian,(2014) Verification method for home electrical signal disaggregation, *ESIOT, IEEE*

[9] L. Dufour,D. Genoud,B. Ladevie, JJ. Bezian,(2014) Identification method for home electrical signal disaggregation, *SET , IEEE*

[10] Farinaccio, L. (1999) The disaggregation of whole-house electric load into the major end-uses using a rule-based pattern recognition algorithm. *Concordia University*, PhD Thesis.

[11] Marceau, M.L., Zmeureanu, R. (2000) Nonintrusive load disaggregation computer program to estimate the energy consumption of major end uses in residential buildings. *Energy Conversion and Management*. vol. 41, no. 13, pp. 1389-1403.

[12] Srinivasan, D. and Ng, W. S. and Liew, A.C. (2006) : Neural-network-based signature recognition for harmonic source identification, Vol. 21,no. 1,pp. 398-405.

[13] Patel, S., Robertson, T., Kientz, J., Reynolds, M., Abowd, G. (2007) At the Flick of a Switch: Detecting and Classifying Unique Electrical Events on the Residential Power Line. *UbiComp 2007: Ubiquitous Computing. Lecture Notes in Computer Science*. Vol. 4717, Springer Berlin Heidelberg, pp. 271-288.

[14] Zoha, A., Gluhak, A., Imran, M.A., Rajasegarar, S. (2012) Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey. *Sensors*. vol. 12, no. 12, pp. 16838-16866.

[15] IEEE (2006) Utility-scale smart meter deployments, plans and proposals. *The Edison Foundation Institute for Electric Efficiency*.

[16] El Guedri, Mabrouka (2009) Caractérisation aveugle de la courbe de charge électrique: Détection, classification et estimation des usages dans les secteurs résidentiel et tertiaire

[17] Luan, S. W., Teng, J. H., Chan, S. Y., Hwang, L. C. (2009). Development of a smart power meter for AMI based on ZigBee communication. *In Power Electronics and Drive Systems, PEDS 2009. International Conference on*. pp. 661-665, IEEE.

[18] Froehlich, J., Larson, E., Gupta, S., Cohn, G., Reynolds, M., Patel, S. (2011). Disaggregated end-use energy sensing for the smart grid. *Pervasive Computing, IEEE*. Vol. 10, no. 1, pp. 28-39.

[19] T. Cover,P. Hart : Nearest Neighbor Pattern Classification(1967), journal=The Edison Foundation Institute for Electric Efficiency,*The Edison Foundation Institute for Electric Efficiency* pp. 21-27.

[20] A. Blake and A. Zisserman,Visual Reconstruction, *MA : MIT Press*,1987
Proceedings of the 1990 conference on Advances in neural information processing systems 3,

[21] John Shafer, Rakesh Agrawal, Manish Mehta,SPRINT, A scalable parallel classifier for data URL = http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.152rep.pdf.

[22] Power meter series 800 data sheet,URL = http://www2.schneider-electric.com/resources/sites/ SCHNEIDER ELECTRIC/content/live/FAQS/140000/FA140345/enUS/PM80020DataSheet.pdf.

[23] Power meter series 800 user guide,URL = http://download.schneider-electric.com/files?p File Id= 27600253p File Name=63230-500-225A2 PM800 User Guide EN.pdf.

[24] Knime,URL = http://knime.com/.

[25] Ecowizz,URL = https://www.ecowizz.net/.