IntelligEnSia based Electricity Consumption Prediction Analytics using Regression Method

Angreine Kewo
Informatics Engineering
De La Salle University
Manado, Indonesia
akewo@unikadelasalle.ac.id

Rinaldi Munir
Informatics Engineering
Institut Teknologi Bandung
Bandung, Indonesia
rinaldi@informatika.org

Aditya Kalua Lapu
Informatics Engineering
De La Salle University
Manado, Indonesia
aditya_lapu@yahoo.com

***Abstract*-Energy sustainability is one of the world focuses today. We have built our solution which is called IntelligEnSia (Intelligent Home for Energy Sustainability) that is focused on the prediction analytic using Web and Android technology platforms. In this case, to predict the energy consumption we applied three regression models: simple linear regression, KLM a and KLM b. All models can be applied to predict the next period of energy consumption based on the independent variable of X = day and dependent variables of Y = current, voltage, and power. It can be concluded that KLM a, has the smallest error accuracy among the proposed models. It means that, processing the data of similar period and category in a history, has bigger influence to the prediction value. Based on the testing, it is find out that the biggest error percentage among the models is relied on power, while the smallest is relied on current. These three models are valuable to help the decision maker in creating the better energy management in the city regarding the supply and availability.**

Keywords— IntelligEnSia, Electricity, Prediction, Linear regressio, KLM model

I. INTRODUCTION

Energy sustainability becomes one of the world's focuses since the vast majority of researchers believe that the climate is changing and it is fundamentally man-made. It is also a crucial problem in most of cities in Indonesia. The needs, availability, utilization and optimization of energy are the complex problems and finally become an ordinary matter. In our previous research stage, we have introduced an intelligent home system which is a system that may utilize and optimize the energy efficiently. This intelligent system can control, learn and adapt the behavior of energy consumption in each house in order to give recommendation to the energy management in the city [1]. We realized that the research of energy consumption prediction have been flourished and are increasing in order to evaluate specific energy saving potentials.

Some of those assessments examine specific countries or case studies, such as the UK. Schrock and Claridge (1989) have started to use a simple regression method of the ambient temperature to investigate a supermarket electricity use [3]. In addition, Braun *et al*. (2014) use a multiple regression analysis to investigate a supermarket based on gas and electricity data for 2012. It is find out that the electricity consumption is likely to increase by 2.1%, whereas the gas consumption will be drop by estimation of 13% [10].

In China, Tso and Yau (2006), presents three modeling techniques to predict electricity consumption: traditional regression analysis, decision tree and neural networks. It is appeared that decision tree and neural network models appear to be suitable alternatives to the stepwise regression model [6]. Lam *et al*. (2010) then optimize multiple regression analysis to study office buildings in different climates [7]. While in the same year, Li et al. (2010) were investigated 59 residential buildings using four different modeling methods to predict annual building energy consumption. The simulation results show that Support Vector Machine and General Regression Neural Network methods achieve better accuracy and generalization than Back Propagation Neural Network and Radial Basis Function Neural Network [9].

Furthermore, New and Parker (2012) have introduced a machine learning case study to predict future hourly residential electrical consumption which is closely related to our work. The result from residential data show that Feed Forward Neural Networks, Support Vector Regression and Linear Regression methods perform poorly and that Hierarchical Mixture of Experts with Least Squares Support Vector machines performs best – a technique not previously applied to this domain [13]. In 2013, Zahan and Kenett have developed a prediction model using the time series forecasting system of the SAS statistical software. It is concluded that energy supply, system and substitute where policy matter is necessary to be prepared [11].

Various researches mentioned above and all works in this area has their own objective(s), platform(s), situation, rules and process. In this work, we focused on the prediction analytics based on the user consumption profile using Web and Android technology platforms.

II. RESEARCH QUESTION AND OBJECTIVE

The research question here is focused on the question that is related to the energy's prediction analysis based on IntelligEnSia technology Web and Android platforms. We then construct the research question as how to predict the energy consumption based on regression methods using Web and Android technology platforms?

Thus, the objective in this research is to provide the prediction analytics using regressions method on Web and Android platforms.

III. STATE OF ARTS

A. *IntelligEnSia*

IntelligEnSia is an Information Communication and Technology (ICT) innovative solution that can control and monitor the energy consumption through the internet connection on web and mobile platforms. It can also adapt and learn the user consumption profile. Based on the related data, IntelligEnSia can produce intelligence flexible reporting [1].We realize that for the sustainable energy matter, some solutions have been introduced and developed, such as Smart Grid technologies [8]. The Smart Grid is "*a combination of hardware, management and reporting software, built atop an intelligent communications infrastructure. In the world of the Smart Grid, consumers and utility companies alike have tools to manage, monitor and respond to energy issues*" [12]. Smart Grid technologies are optimizing the asset utilization by shifting peak load to peak-off times [14].

B. *Why IntelligEnSia?*

- Control the energy consumption at home and buildings.
- Learn and adapt the behavior or characteristics of the energy consumption at home and buildings.
- Provide intelligence reporting of the energy consumption for the better and effective energy management system in the city: resources, availability, consumption and characteristics peak load to peak off times. It is included the report of the list that have improved their consumption behavior.
- Help government and industries in decision making and creating policies regarding energy.
- Create the new better behavior of the people for the energy sustainability [1].

C. *IntelligEnSia Case Study*

IntelligEnSia's application: Controlling

Below is the example of IntelligEnSia web-based application.

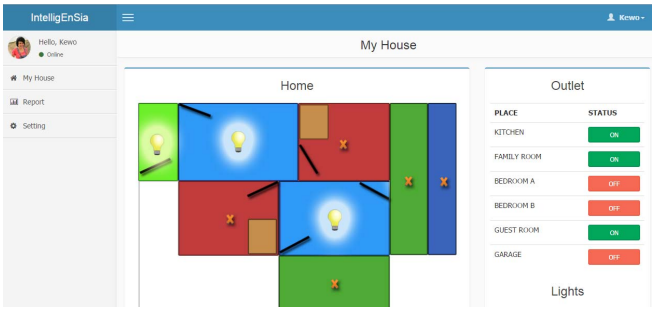

Figure 1. IntelligEnSia Application – layout [2]

Beside the house layout, there is a list of outlet and light from the house/building. Once the user login, he/she then may control from the distance the status of outlets or lights. Control here means to make an outlet(s) or light(s) on and off.

IntelligEnSia's application: Monitoring

Beside control feature, IntelligEnSia may optimize the data by providing a flexible reporting for the user. As displayed in the following figure, through the application customer can monitor the energy consumption in a flexible period of date, month and year. The data consist of voltage, current and power consumption figure which are displayed in the graph and chart forms.

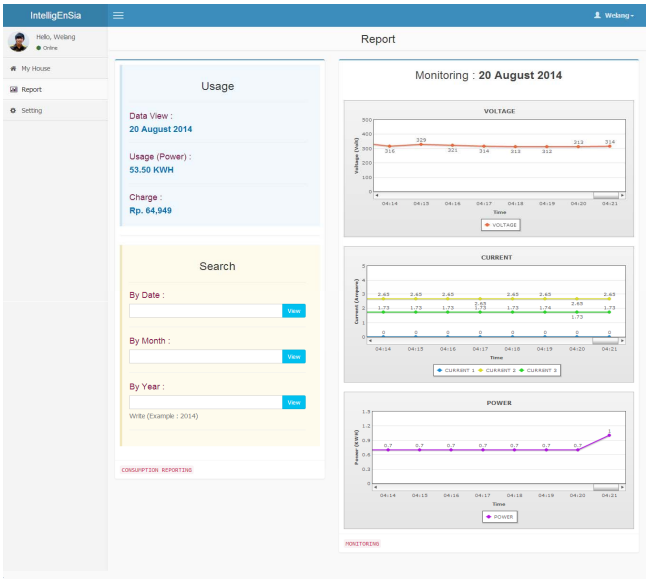

Figure 2. IntelligEnSia Data Reporting [2]

The mechanism of IntelligEnSia as followed:
Firstly, sensor will capture the condition of the light and wall-outlet. Then, the captured value of voltage (V) and Current (I) are being sent to the analog digital converter (ADC), where voltage and current will be used to find power consumption. It is reflected in one of the most well-known equation:

$$P = V \times I \qquad (1)$$

Where, P = Power (Watt),
V = Voltage (Volt)
I = Current (Ampere)

The result of applying equation (1) is presented in Power column of table 1.

In ADC, the signal has been converted to discrete signal. The discrete signal then is processed by the microcontroller to be sent to the IP, where the signals are being formed into

digital data. Furthermore, the data are being sent to the webserver. The data then will be processed to be the information that may give recommendation to the user, in order to do some actions such as, turn off the light, unplug the electricity or give suggestion to be more efficient in using the energy [1].

The following is the system block diagram of IntelligEnSia:

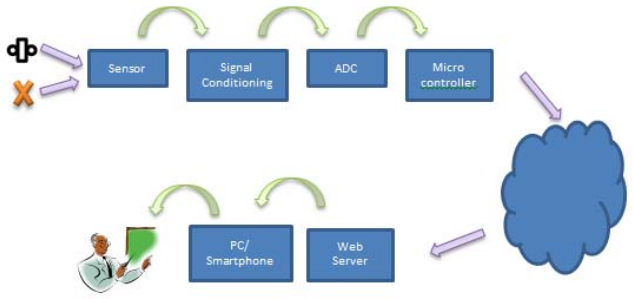

Figure 3. The System Block Diagram [1]

D. Energy Sustainability

The outline of Sustainability Energy definition proposed by the Sustainability of Energy Systems and Mobility Group, coordinated worldwide by Interuniversity Research Center on Sustainable Development (CIRPS), identifies five pillars for energy sustainability [5]:

• Renewability of energy resources;
• Efficiency in energy conversion, distribution, use;
• Lowering of environmental impact;
• Increasing of energy accessibility;
• Tailor making of energy systems to meet local social economic-environmental conditions.

E. Prediction

Prediction mostly been viewed as a quantitative product, a result of scientific research which by Fischoff (1994) is determined as set of probabilities associated with a set of future events.

It is required to understand the definition of the predicted event(s), as well as the expected likelihood of the event's occurrence. If prediction is easy to be seen in the light, then the goal of prediction is simply to develop a good prediction by evaluating the objective criteria(s), such as accuracy and skills. Ascher (1981) described that "Accuracy is a measure of how closely a specific prediction products conforms to the actual events [4].

F. Linear Regression

In this section, we emphasize on the method to predict the needs of energy for the coming period based on the user's consumption figure. There are many techniques for prediction, but we then decide to optimize regression method. It is because this simple technique may support IntelligEnSia computation in the cloud. It may support the data processing runs lighter and faster.

Linear regression method for prediction

1. The relationship between independent (X) and dependent (Y) variables. Based on that equation: can be found out how big is the contribution of X variable toward Y variable.
2. The general equation of simple linear regression is

$$Y = a + bX \quad (2)$$

3. a and be values have been determined through the following equation:

$$a = \frac{(\sum Y_i)(X_i^2)-(\sum X_i)(\sum X_i Y_i)}{n(\sum X_i^2)-(\sum X_i)^2} \quad (3)$$

$$b = \frac{n\sum X_i Y_i-(\sum X_i)(\sum Y_i)}{n(\sum X_i^2)-(\sum X_i)^2} \quad (4)$$

X = Independent variable (prevention cost, appraisal cost, internal failure cost, external failure cost)
Y = Dependent variable
a = Y value if $X = 0$
b = Gradient
n = Amount of sample

Linear regression is the simplest technique, which is used to establish and provide a baseline performance for predicting the energy consumption in a house/building. Edwards *et al.* (2012) stated that if a technique performs worse than the baseline predictor, then it is most likely not suitable for the residential data set [13].

Analytics Process

The following are the required steps in order to solve the problem:

1. Find the values of X^2, Y^2, dan X.Y
2. Find $\Box X$, $\Box Y$, $\Box X^2$, $\Box Y^2$, dan $\Box X.Y$
3. Find a and b values through the equation (3) and (4).
4. Through the simple linear regression equation: $Y = a + bX$ (2), please count the required value of Y based on the previous a and b values, also given variable of X.
5. Find RMSE (*Root – Mean – Square - Error*) with equation:

$$_{RMSE} = (\frac{1}{n} \sum_{i=1}^{n} | f(xi) - yi |^2)^2 \quad (5)$$

The f(xi) and yi are variables x (independent) and variable y (dependent) that are used to calculate the RMSE. An example of calculating regression model RMSE: Current model i = 17–22. The f(xi) and yi that used to calculate the RMSE can be shown in the tabel below.

Table 1. Example calculation of RMSE

i	X	y	Regression	Deviation	Deviation Square
1-16	1-16	-	-	-	-
17	17	2,279	2,2796	-0,0006	0,00000
18	18	2,274	2,2798	-0,0058	0,00003
19	19	2,279	2,28	-0,001	0,00000
20	20	2,279	2,2802	-0,0012	0,00000
21	21	2,285	2,2804	0,0046	0,00002
22	22	2,283	2,2806	0,0024	0,00001

Based on the table above, RMSE=

RMSE	0,0000138236
Error	0,003718015

G. KLM Regression Method

For the in depth research, we propose three regression model. The first prediction model is based on the linear regression as explained in the previous section. The second and third models are the improved regression model, which is named KLM (Kewo – Lapu – Munir) models.

In this case, we focused on the prediction period of day, users are working people and the location is Manado. Thus, in context of day prediction, it is required some identifications and classification of:

1. Day's category: Monday to Friday = working day
   Saturday to Sunday = weekend
2. Week : I, II, III, IV
3. Month : January = Decline
   February-November = Normal
   December = Surplus

KLM a model: In predicting the next period (day x), the model will make a regression based on the similar period of x in the history.

KLM b model: In predicting the next period (day x), we combine the regression result of simple linear regression and KLM a model.

For example: We want to predict the 10^{th} day of the current month and that day is Monday.

For KLM a model, the regression will be conducted to all Monday in the pass within the six months. While for KLM b model, the regression will be executed to all previous days in the current month and to all Mondays in the pass within the six months.

IV. EXPERIMENTS AND RESULTS

A. Implementation of Linear regression method

To analyze the applied method, we did a system simulation and prediction performance testing to a house of working people in Manado. We optimize 22 days data on March 2015 as displayed in the following table. Then, we wish to predict the energy consumption on day 23 by applying the three models as discussed in the previous section:

1st Regression Model: Simple Linear Regression

a. Based on the table below, the equation (1) is applied as follows:

Table 2. 22 days energy consumption data

Day	Current	Voltage	Power	Cost
(X)	(Y)			
1	2,283	223	509,109	381831,75
2	2,275	222	505,05	378787,5
3	2,275	221	502,775	377081,25
4	2,275	222	505,05	378787,5
5	2,275	222	505,05	378787,5
6	2,275	222	505,05	378787,5
7	2,284	223	509,332	381999
8	2,283	223	509,109	381831,75
9	2,274	222	504,828	378621
10	2,273	222	504,606	378454,5
11	2,279	222	505,938	379453,5
12	2,278	221	503,438	377578,5
13	2,275	222	505,05	378787,5
14	2,285	223	509,555	382166,25
15	2,283	223	509,109	381831,75
16	2,279	222	505,938	379453,5
17	2,279	222	505,938	379453,5
18	2,274	221	502,554	376915,5
19	2,279	222	505,938	379453,5
20	2,279	222	505,938	379453,5
21	2,285	223	509,555	382166,25

22	2,283	223	509,109	381831,75
23	2,28	222	505,321	382298,60

b. Determining a and b values

From the data in the table above, we try to create a model based on equation (2). To determine a and b, we applied equation (3) and (4). Based on those equations, here are the results of each current, voltage, power and cost:

Current a = 2,27; b = 0,02
Voltage a = 222,064; b = 0,01
Power a = 505,48; b = 0,06

c. Prediction Result

To predict the value of day 23 we recall equation (2) and substitute the a and b values with the result from the previous step (B).

Current:
Y_{23} = 2,280 Ampere
Voltage:
Y_{23} = 222,064 Volt
Power:
Y_{23} = 507,0642 Watt

d. Implementation method Of Root mean square error (RMSE) for accuraccy

For accuracy, error prediction is defined to measure the difference between the measured and estimated values. In this case, we use equation (5):

Current:
E_{RMS} = 0,003 %
Voltage:
E_{RMS} = 0,64 %
Power:
E_{RMS} = 2,19 %

2nd Regression Model: KLM a

a. The equation (1) is applied in the table.

b. Determining a and b values

Current a = 2,27; b = 0,01
Voltage a = 222,35; b = 0,01
Power a = 379.42; b = -21,14

c. Prediction Result

Current:
Y_{23} = 2,275 Ampere
Voltage:
Y_{23} = 222,064 Volt
Power:
Y_{23} = 505,386 Watt

d. Implementation method Of Root mean square error (RMSE) for accuraccy

Current:
E_{RMS} = 0,002 %
Voltage:
E_{RMS} = 0,4 %
Power:
E_{RMS} = 1,06 %

3rd Regression Model: KLM b

a. The equation (1) is applied in the table.

b. Determining a and b values

Current a = 2,27; b = -0,01
Voltage a = 222,25; b = -0,01
Power a = 506,20; b = -0,05

c. *Prediction Result*

Current:
Y_{23} = 2,2752 Ampere
Voltage:
Y_{23} = 222,0136 Volt
Power:
Y_{23} = 505,0006 Watt

d. Implementation method Of Root mean square error (RMSE) for accuraccy

Current:
E_{RMS} = 0,003 %
Voltage:
E_{RMS} = 0,45 %
Power:
E_{RMS} = 1,49 %

The summary of all RMSE from 3 regression models, are presented as followed:

Table 3. RMSE result of 3 models

Basic Linear Regression	Current	0,003 %
	Voltage	0,64 %
	Power	2,19 %
KLM a	Current	0,002 %
	Voltage	0,4 %
	Power	1,06 %
KLM b	Current	0,003 %
	Voltage	0,45 %
	Power	1,49 %

B. Implementation of Prediction in IntelligEnSia System

We then applied these methods as the new requirements of IntelligEnSia as displayed in the interface design below:

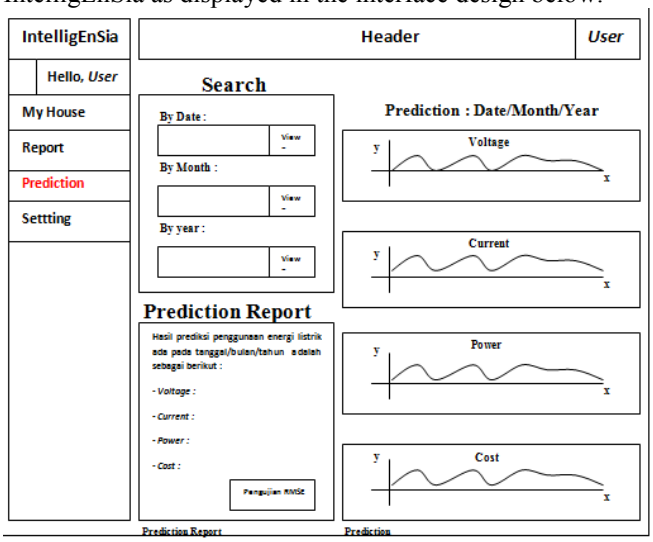

Figure 4. IntelligEnSia's prediction design

V. CONCLUSIONS

In conclusion, a linear regression method can be applied to predict the next period of energy consumption based on the independent variable of X = day and dependent variables of Y = Current, voltage, power and cost. The different between the prediction value and the real value of current, voltage, and power and cost are relatively small. It can be concluded that KLM a has the smallest error accuracy than basic linear regression and KLM b. It means that model from the similar period and category in a history has big influence to the prediction value. Based on the testing, the biggest error percentage of all models is on power, while the smallest one is on current. It is also shown that KLM b has closer result to the KLM a, the smallest error accuracy, than basic linear regression to KLM a.

Furthermore, the limitation of this work happens when the missing data is occurred due to the unstable internet connection. In fact, it has been solved by another work of IntelligEnSia's missing data solution using Lagrange interpolation.

This research then may help to achieve IntelligEnSia's objective in providing intelligence reporting in context of energy availability and resources for the better and effective energy management system in the city and helping the decision maker: government and related industry, to produce decision and policy in energy management.

For future research and more accurate result, it is recommended to enclose all related factors, such as weather, temperature, amount of people in the house, user behavior, socio economics and electricity off duration in order to examine their relationship with the consumption profile. It is also required to analyze and optimize the other prediction methods and compare them, in order to obtain better result in energy consumption prediction.

ACKNOWLEDGEMENT

The authors would like to thank De La Salle University Manado and IEEE Region 8, under the project contest of IEEE Young Engineers Humanitarian Challenge 2013.

REFERENCES

[1] A. Kewo, P. Manembu, and C. Sengkey, "IntelligEnSia for Energy Sustainability, Case Study: Manado, Indonesia", European Modeling Symposium, pp.443-447, Manchester, November 2013.

[2] B. Welang, "Sistem Pengendalian dan Monitoring Energi Listrik Secara Real Time", Tugas Akhir, Teknik Informatika, De La Salle University Manado, 2014.

[3] D, Sarewitz., Pielke, R., and Byerly, R . *Prediction: Science, Decision Making and The Future of Nature*, Island Press: Washington DC, 2000.

[4] DW. Schrock and DE. Claridge, "Predicting Energy Usage in a Supermarket", 6th Symposium on Improving Building Systems in Hot and Humid Climates, vol.51, 1989.

[5] F. Orecchini. Energy sustainability pillars. *Int. J. Hydrog. Energy.* 2011, Vol. *36*, pp. 7748-7749.

[6] G. Tso and K. Yau, "Predicting Electricity Energy Consumption: A Comparison of Regression Analysis, Decision Tree and Neural Networks", Energy, vol.32, pp.1761-1768, 2007.

[7] J. C. Lam, K. K. W. Wan, D. Liu, and C. L Tsang, "Multiple Regression Models for Energy Use in Air-Conditioned Office Buildings in Different Climates", Energy Convers Manage, vol.51, No. 12, pp.2692-2697, 2010.

[8] Kates, R.W.; Clark, W.C.; Corell, R.; Hall, J.M.; Jaeger, C.C.; Lowe, I.; McCarthy, J.J.; Schellnhuber, H.J.; Bolin, B.; Dickson, N.M.; *et al*. Environment and development: Sustainability science. *Science* 2001, vol. *292*, pp.641-642.

[9] Q. Li, R. Pen, and Q. Meng, "Prediction Model of Annual Energy Consumption of Residential Buildings", International Conference on Advances in Energy Engineering (ICAEE), pp.223-226, Beijing, June 2010.

[10] M. R. Braun, H. Altan and S.B.M Beck, "Using Regression Analysis to Predict the Future Energy Consumption of a Supermarket in the UK", Apply Energy, vol.130, pp.305-313, 2014.

[11] M. Zahan and R. Kennet, "Modeling and Forecasting Energy Consumption in the Manufacturing Industry in South Asia", International Journal Of Energy Economics and Policy, vol. 3, no.1, pp.87-98, 2013.

[12] Podmore, R., and Robinson, M, The Role of Simulators for Smart Grid Development, IEEE Transactions on Smart Grid, 2010, Vol. 1., Nr. 2.

[13] R.E Edwards, J. New, and Parker, L, "Predicting Future Hourly Residential Electrical Consumption: A Machine Learning Case Study. Energy and Buildings", Energy and Buildings, 2012, Vol. 49, No. 0, pp.591-603.

[14] S. Heinen, D. Elzinga, S. Kim, and Y. Ikeda. *Impact of Smart Grid Technologies on Peak Load to 2050*, International Energy Agency: Paris, 2011.