Estimation System of Occupant Behavior Against The Use Of Electricity Using Bayes Method And Decision Tree Algorithm

Almas Shabrina¹, Feri Fahrianto², Nenny Anggraini³
Department of Informatics Engineering
Faculty of Science and Technology
State Islamic University of Syarif Hidayatullah Jakarta
Ciputat, Indonesia
almas.shabrinaa@gmail.com

Abstract- Based on research by the year 2013, entitled Kleiminger Occupancy Monitoring Using Household Electricity Meters stated that digital power meter suitable for use as a sensor for occupancy detection average accuracy of your detection can reach 80%. Besides the estimated occupancy with future electric current sensors can identify electricity consumption usage patterns to estimate future electricity consumption (Silva, 2011) and a model of a daily routine for energy efficiency (Abreu, 2012). Huang in 2016 stated that the current occupancy detection system has several drawbacks, namely: (1) Cost of implementation is quite expensive (2) user privacy, (3) Accuracy of detection, and (4) Intrusiveness. As a result, it is feasible to study on other potential approaches to address the deficiencies in the system and the occupancy estimation can also estimate the model's daily routine to be one reason for the policy-making authorities for energy efficiency. In this study, the authors make the system estimates the consumption behavior of the household electric loads with maximum hypothesis Bayes methods appropriate probability (HMAP) and decision tree algorithm, using arduino microcontroller hardware and ethernet shield, with the value of the Positive Predictive Value (PPV) reached 64.5%.

I. INTRODUCTION

In the last five years there are several techniques of detection and estimation of occupancy in the literature stated the basic idea of the predicted occupancy or estimates are to install or activate sensors such as sensors CO2 concentration, cameras, motion sensors, light sensors, temperature sensors, acoustic sensors, humidity sensors, a passive infrared sensor (PIR), and RFID to collect a variety of information. Use of information occupancy in a building leads to behavior and energy efficient operation of the HVAC (Heating, ventilating, and air conditioning) so as to reduce the operating costs of the building and aims to achieve smart building sustainable (Sun, 2011) (Huang, 2011) (Li, 2012) (Yang, 2012) (Wang, 2014) (Masoudifar 2014)

The detection system housing that exist today have some drawbacks, according to Yang (2012) and Huang (2016) among others, (1) Cost of implementation is quite expensive, for example using the new infrastructure and requires the installation of devices and applications that too much, (2) user privacy, (3) Accuracy of detection, (4) occupants feeling disturbed because it was supervised by the sensors (intrusiveness). As a result, it is feasible to study on other potential approaches to address the deficiencies in the system and the estimated occupancy detection.

Based on research by Kleiminger in 2013 titled Household Occupancy Monitoring Using Electricity Meters stated that digital power meter suitable for use as an electric current sensor for occupancy detection average accuracy of your detection could reach 80% by adding a PIR sensor. In addition to electrical load occupancy detection to identify future electricity consumption usage patterns to estimate future electricity consumption (Silva, 2011) and a model of a daily routine for energy efficiency (Abreu, 2012).

From the results of the exposure analysis and theoretical simulations over which states that household merupaka user sector electric energy highs and high electric power consumption by households due to the behavior patterns of the use of electronic tools and the household sector have not been there a system that can estimate the behavior residents with detection accuracy, optimal cost and nature intrusiveness which is a problem of the system estimated occupancy, so the author uses sensors electric current, because these sensors have a cost effective in building the system, easy installation, and can protect the privacy of the household, for the detection accuracy Bayes methods chosen because it can calculate the probability of occurrence of an event based on the observation that the consumption of electricity load events, while the decision tree chosen for the system has been defined over the whole space models and parameter input learning of all training data.

Because of the lack of systems or smart meters in households to know the electronic equipment being used, activity, and behavior patterns of its inhabitants so that it can be used to estimate the model's daily routine to be one reason for the decision making authorities for energy efficiency, so that the authors draw the conclusion that the need for system estimation household occupant behavior to electrical load by simply using
the electric current sensor has a detection accuracy, able to maintain user privacy and intrusiveness or the user is not bothered by the presence of sensors.

1.1. Objectives

Objectives in this research are:
1. The system consists of tools that writing is designed as a substitute for smart meters.
2. Estimate by simply using current sensors.
3. The data was collected for 3 weeks, and the data known electric consumption patterns occupants of the house.
4. The method of estimation by the author using statistical quantitative probabilistic namely with Bayes method and decision tree algorithm.
5. The method of Bayes Hypothesis Appropri Maximum Probability (HMAP) was used to estimate the active equipment that is being used.
6. Be able to estimate the behavior and the activities being carried out occupant home, with the results estimated in the form of recess activities, entertainment or activity homework.

II. REVIEW OF RELATED LITERATURE

A. Current transformers

Current transformers (CTs) is a sensor for detecting an alternating current (AC). With ratings up to 100A and accuracy of up to 50mA, this sensor is used to read the current on the building or home. This split type cores as easy to use because it does not need to disconnect and reconnect the existing installation. Just open and tuck the cable at this sensor device (Current transformers YHDC SCT-013 Data Sheet)

![Figure 2. 12 Split Core Current Transformer SCT-013](image)

This can be accomplished by the following series that consists of two main parts:
1. Sensor and load resistor
2. The biasing voltage divider (R1 & R2)

![Figure 2. 13 Circuit YHDC SCT-013-000](image)

B. Ethernet Shield

Ethernet shield Arduino module serves to connect to the Internet network. Wiznet shield embedded Ethernet W5100 ethernet chip. Wiznet W5100 ethernet provides network that can access the network either TCP or UDP. The use of this shield utilize Ethernet Ethernet library to write the code to connect to the Internet network.

![Figure 2. 14 Ethernet shield](image)

C. Bayes Method

Bayesian probability theory is a branch of mathematical statistics theory that allows us to create a model of the uncertainty of an event occurring by combining a general knowledge of the fact of observation. In accordance with the probability, if one looks at the incidence of B and have faith that there is a possibility B will appear, then the probability of B is called the prior probability. Once there is any additional information that the event A has emerged, for example, may be a change to the original estimate of the probability B to appear. The probability for B now is the result of the conditional probability of A and is called the posterior probability. Bayes’ Theorem is a mechanism to update the probability of the prior into posterior probabilities. The core of the methodology is the Bayes, Bayes theorem for events A and B, the conditional probability of A if B occurs, or can be written as follows

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Hypothesis Appropri Maximum Probability (HMAP) is a model simplification of Bayes method called Naive Bayes. HMAP is the one used in the machine learning as a method to get a hypothesis for a decision. HMAP stated hypotheses are based on a probability value based on the condition of prior (initial information) is known. In the present study the authors use Bayes method with the rules for dual evidence E1, E2, ..., Em and double hypothesis H1, H2, ..., Hn.

$$p(H_n|E_1, E_2, ..., E_m) = \frac{p(E_1, E_2, ..., E_m|H_n) \times p(H_n)}{\sum_{i=1}^{n} p(E_1, E_2, ..., E_m|H_i) \times p(H_i)}$$

To apply the equation above, it must be known to the conditional probability of all possible combinations of evidence-evidence for the entire hypothesis. In practical terms, this is not possible. Therefore, the above equation, replaced by equality.

D. Decision trees

A decision tree is a hierarchical data structure that implements (divide and conquer strategy) strategy of divide and conquer. This is an efficient nonparametric methods, which can be used for both classification and regression and how a learning algorithm that builds a decision tree from the training sample data can be converted into a set of simple rules that are easy to understand.
III. RESEARCH METHODOLOGY

Methods are used in this research are as follow:

A. Literature Reviews
At this step, literature review is done by collecting all books and articles that are relevant to this research. Several textbooks are used in this research in examples books that contain theory about UML, Embedded System, Electronics Circuits, and Database System. Articles that has similarity with this research also is collected, some article found in Journal and some articles found in proceeding of international conference.

B. Observation and Questionnaire
In this study, the observation site selection is done by a non-probability sampling at the house of the writer H. Afandi 19 RT 04 / RW: 001 Ex. Karadenan, district. Cibinong, Kab. Bogor. And implementation time in January 2017. The activities are recording the electrical current usage data and records every activity that involves the use of electronic equipment in the home to determine the electrical current consumption patterns toward activity undertaken, then as a reference for the system determines the accuracy of these estimates.

At this stage also obtained a list of electrical appliances and a large influx of the equipment, zoning and large idle currents that flow when there is no electrical equipment that is being actively used as well as the simple moving average electricity in the house. And also make some questionnaire to know relation between behavior against electricity devices

IV. ANALYSIS, DESIGN SYSTEM, IMPLEMENTATION AND TESTING SYSTEM

This chapter will discuss in detail and detail about the applications and systems that will be implemented by applying research methods that have been described in the previous chapter. In the previous chapter has discussed that the system development method used by the author is a development methodology prototyping system. There are 5 stages that are used in the method of prototyping (Roger S. Pressman 2010), namely: 1) Phase communications, 2) Requirements Gathering Phase, 3) Stage Building Systems, 4) Phase Encoding System, 5) System Testing Phase.

4.1 Scope
The first thing should be done in determining all elements needed by the system is define the scope. The scope that the researcher do in this research is in simple house. The researcher need to monitor all the electricity power from the electric company. This electricity monitoring system use sensor that captured data and send the data to a server automatically and periodically updated each 30 second. The monitoring system should be real-time and stored to the server. Afterwards the system able to show and visualize the graph of power used remotely, locally and the active electrical equipment and estimate occupant behavior.

Moreover, the authors describe what my activities which may take place at the housing and list all the electronic tools that may be used and correlated with the activities to be able to infer the behavior of residents. Observations carried out for 3 weeks starting from the date of 23 January to 10 February 2017 by recording changes in the electrical current every 30 seconds once, subsequent authors also monitor every activity carried out residents in accordance with the greater use of electric current recorded with the aim of creating a ground truth and occupancy transitions to be used as a testing parameters later.

4.2 Proposed Design System
In the present study the authors draw the hypothesis that the behavior of residents can be known through the consumption of electrical energy by measuring electrical current load and the time of use. From some other literary authors also conclude methods and algorithm appropriate for this settlement that is using the methods of machine learning - supervised with Bayes’ rule and decision tree algorithms (Decision tree).

Here is the proposed system is that the authors do,

Figure 4. 1 Analysis of the Proposed System

In Figure 4.1 describes the system power monitoring system based on the IOT. The system will monitor the load current of electricity by using SCT 013, henceforth serve as training data to be used in machine learning with the theorem Bayes is used to determine the electronic device is being actively used and the decision tree to determine the behavior is happening at that time in accordance with the training data set.

The following explanation of the functions of the proposed system components:
1. SCT 013 PLN current sensor of the cable measured and digitized by using Arduino.
2. Ethernet shield is used as a unit of Arduino connections and database servers for data delivery.
3. The Arduino is used as a data processing unit and a digital to analog converter and sends the data as serial data.
4. Database servers that send data to the web
5. For the next writing program code menggunakan PHP programming language.
6. Display Data visualization using realtime graph.
7. Users can monitor consumption of electric current load and the behavior of the housing. From the data passed to the database server and the database server will be pushed to the web server. By using a web browser to the user data can be visualized in real time.

4.3 Analysis of Needs

This step aims to define what the system should do to achieve the goals of the system. In this case to design the system estimates the behavior of residents against electrical load by the method of Bayes and decision tree algorithm to residents use to get information on how to use sensors electric current can be used as a detection and estimation of the activity and behavior of occupants of the house accurately and how the pattern the behavior of residents can be used as a reference to estimate the possibility of ongoing activity.

4.3.1 Functional Requirements

Build housing behavior estimation system to the electrical energy consumption based on measurement of electrical current load and the time of use must have the following functional requirements:

1. Being able to monitor and capture the flow of the house.
2. Data stored in the database in real time.
3. Visualization of consumption data with the electrical load.
4. Able to provide information about electricity consumption patterns in time series and berkalanjutan.
5. Able to provide information regarding the use of electronic tools that are active at that time.
6. Being able to provide an estimate of the behavior of residents.

4.3.2 Non-Functional Requirements

a) Analysis Hardware Requirements

1. Arduino Uno As the brain of the system that can process data and perform the entire process on the system.
2. Ethernet Shield Connecting to the network
3. Flow Sensor SCT 013 100A For a current sensor
4. Ethernet Cable / LAN Cable Allows ArduinoUno to be connected to the PC and the Internet.
5. Power Source USB 5 V 2.0 Give power to the Arduino Uno

![Figure 4. 7 Design database estimation system of the household behavior](image)

The following explanations use the tables in the database system estimates the behavior of occupants of the house to electrical load with Bayes method and decision tree algorithm.

4.4 System Implementation Phase

At this stage, the author divides into three stages: 1) Programming Tool, 2) Programming Probability Bayes, 3) Programming Decision Tree.

4.4.1 Programming Tool

Construction hardware used is as follows:

![Figure 4. 9 Construction Hardware](image)

At 4.9 image data obtained from the electric sensors will be translated by the microcontroller, and such data will be sent via Ethernet Shield uses wireless local network connection to the server to be stored in the database and will be displayed after the process as information on the web page. The system uses current estimates of this behavior that is SCT-013 sensor 100A-50mA to 50mA measurement of the ratio-100A, to obtain accurate results required calibration measurements. The sensors are calibrated to measure up to 30A AC. 30A is the RMS value of the maximum current.

4.4.2 Programming Bayes

Estimates and predictions are used to estimate the state will come through state testing in the past. In this study the estimates and the probability is raised to minimize the influence of the uncertainty of a problem. Estimates and predictions usability seen at the time of decision making. Good decision is a decision based on considerations - considerations that will occur at the time the decision was implemented.

\[
p(E|Hi) = \frac{p(E|Hi)p(Hi)}{\sum_{i=1}^{n} p(E|Hi)p(Hi)}
\]

- \(p(E|Hi)\) = probability hypothesis \(Hi\) correctly when given evidence \(E\)
- \(p(E|Hi)\) = probability of evidence \(E\), if known hypothesis \(Hi\) correctly
- \(p(Hi)\) = probability of hypothesis \(Hi\) (according to previous results) without regard to any evidence
- \(n\) = number of possible hypotheses.

The conditions at this great research value stored on the probability of the hypothesis already known, namely the amount of current to be estimated, whereas the probabilities that shape is not yet known. Therefore, the authors reverse the process at this Bayes method to derive great value that is stored on a probability value / value of what happened, which is a constituent part of the great value stored on the probability of the hypothesis. In the estimation process, the writer divides into six time zones, following the division of time zones. Distribution 11 time zones:

1. Night 11:01 p.m. to 3:00 1:53 A
2. Early morning 3:01 to 7:00 1.62 A
3. Morning 7:01 to 11:00 1:24 A
4. Noon 11:01 to 15:00 1:24 A
5. After noon 15:01 to 19:00 1:57 A
6. Evening 19:01 to 23:00 3:09 A

Here are the results of calculation of the probability that arose from the observation data in accordance with a predetermined zone.
4.4.3 Programming Decision Tree

In the process of using the author chose decision tree algorithm C4.5 decision tree algorithm which is to perform a classification or segmentation or clustering and predictive. Basic algorithm C4.5 is the formation of a decision tree (decision tree), with the branches of the decision tree is a classification question and the leaves are classes or the segments. C4.5 algorithm is a machine learning algorithm. With this algorithm, machine (computer) will be given a bunch of data to study the so-called learning dataset. Then the results of the study will then be used to process the data which is called a test dataset. Because C4.5 algorithms used to classify, so the results of the test dataset processing in the form of grouping data into its classes. In the C4.5 algorithm to construct a decision tree first thing to do is select the attributes as root. Then created a branch for each value in the root. The next step is to divide the case into each branch until all cases the branches have the same class. Below is a decision tree perila ku residents based on the equipment that is being actively used by time zone.

4.5 Phase Testing System

4.5.1 Independent Testing (Black Box Testing)

At this time aimed at testing determines the functionality of the system, Tests Self:
1. Send data from the microcontroller. Figure 4.36 OK
2. Displays the last 10 sensors recording. Figure 4.18 OK
3. Estimated weight of Bayes corresponding sensor zones installed and connected Showing weights Bayes database in accordance zone. Figure 4.37
4. Displays estimation active equipment consented time sensor mounted and connected database Displays a list of active equipment. Figure 4.38 OK
5. Householder behavior estimating system in real time sensor mounted and connected database Behavior residents are: Rest / Entertainment / Homework. Figure 4.39 OK
6. Looking at the chart the use of electric currents on that point. Graph that appears is the result of recording start 0001 hours. Figure 4.40 OK

Below is a picture of the system's functionality test results:

1. Figure 4.30 Zone A tool set
2. Figure 4. 31 Zone tool set B
3. Figure 4. 32 C zone tool set
4. Figure 4. 33 tool set D zone
5. Figure 4. 34 E zone tool set
6. Figure 4. 35 zone tool set F
7. Figure 4. 36 Update electric current
8. Figure 4. 37 Table electric current
9. Figure 4. 38 Displaying the weight of Bayes accordance zone
Of the above methods applied in the present study obtained Positive Predictive Value (PPV) of 64.5%. PPV of 64.5% value obtained is valid only for the scope of this research, the value of PPV could have been bigger or more for a different scope.

**REFERENCE**

23. Sudaryono, Margono dan, Rahayu, Pengembangan Instrumen Penelitian Pendidikan, Yogyakarta: Graha Ilmu