Big Data Mining activities from Multiple Smart Houses for Healthcare Applications

Bakare K. Ayeni[1]
Department of Computer Science,
Ahmadu Bello University, Zaria.
Email: bakarre@gmail.com

Abdulsalam A. Jamilu.[2]
Department of Computer Science,
Ahmadu Bello University, Zaria.

Abstract— *Millions of homes are being equipped with smart devices (e.g. smart meters, sensors etc.) which generate massive volumes of fine-grained and indexical data that can be analyzed to support smart city services. In this paper, we propose to refine a model and introduce distributed learning of big data mining from multiple smart houses in a near real-time manner. This will help health applications to promptly take actions such as sending alert to patients or care providers. Furthermore, our work addresses building a health ontology model that automatically map discovered appliances to potential activities. This means we can competently train the system and increase the accuracy of sensing human activities. After the pre-processing stage in the data mining process, in the data classification phase, m5rules, Decision table, Random forest, random tree, Multilayer Perceptron (MLP), Farthest first and Logistic Regression (LR) algorithms have been used. The success evaluation of data mining classification algorithms has been admitted through the data mining program WEKA. Multilayer Perceptron algorithm has been the best algorithm with the highest success percentage in the program; Farthest first has been the algorithm which has the lowest success percentage in the program. This study has indicated that data mining can be a useful tool in the medical field, Doctors can be provided with daily activities of the patients in the progress of their treatment. The proposed mechanism, of this research uses the UMass Trace Repository which provides network, storage, and other traces to the research community for analysis. The result of accuracy values of WEKA classifying Algorithms are presented in this paper.*

Keywords— Prediction, Big data, smart homes, smart devices, health care applications, Protégé, Medical Decision Support System

I. INTRODUCTION

Smart Houses have been under development for several years, and promise to bring comfort and peace of mind for the users. Smart Houses have been defined as any environment designed and built to help the user perform activities independently [5], [8]. They are designed for different types of users, from fully independent adults to older adults with disabilities and dementia. Thus, the function of a Smart House varies considerably depending on the target audience. Nonetheless, the overall purpose of a Smart House is to make best use of the comfort and ease of living of the inhabitant as much as possible. In order for the Smart House to assist the user, the house needs to adapt to its user. This is normally done by learning the behavior of the person. As a result, the Smart House aims to learn from the person's activity and actions, with a view to identifying repetitive patterns. This learning is done in order to model and predict the behavior of the person, which is achieved by collecting data from sensors installed throughout the Smart House, and then applying different algorithms and/or data mining techniques. In addition to learning from the user's activity and behavior, the Smart House also needs to learn from the context, which is called context awareness. The concept is that the Smart House knows the surroundings and the environment, knows where the activity takes place, and knows who and where the user is [22], [4]. This context awareness of the Smart House makes the prediction of the user's activity more accurate. There are successful applications for collecting annotations as a representation of electricity consumption data, and therefore make sense of past energy usage, as reported [25].

Many research projects are involved in development of user-friendly and convenient feedback tools for visualization of electricity consumption providing instantaneous consumption data, often through suggesting ambient displays aiming at emotional reactions [26], [14]. It can be found that both commercially and freely available resources use feedback tools including point of consumption devices such as kill a Watt electricity usage monitors, information dashboards, analysis interfaces, and online profiling and visualization tools such as Microsoft Hohm, PowerMeter by Google, ODEnergy, OPOWER or AlertMe. These tools offer precise quantitative measures of energy expenditures, historical and predictive charting facilities, cost breakdowns, and performance tracking. However, they require some effort to integrate them into home infrastructure, and they lack convenient feedback on real-time resource use. This might have an impact on the decision to discontinue development of some of them due to a lack of consumer uptake (PowerMeter service was ceased in 2011 and Hohm in 2012). However, new tools are constantly being developed to provide more detailed energy feedback. The itemized energy consumption from different appliances can be achieved by individually monitoring each of them. However, this strategy is expensive due to the hardware costs and complex infrastructure that may be difficult to deploy. In this context, there is a significant number of researches focused on appliance recognition based on non-intrusive appliance load monitoring approach (NIALM) [11], [6], [29] and [9]. It involves the use of machine learning algorithms and optimization techniques to recognize energy signatures of home devices. The challenge in NIALM is that individual appliances have very different energy signatures that are hard to distinguish unless very sensitive and high resolution meters are used. Therefore, this is an area of research which is still being thoroughly explored [32]. Based on NIALM, there have been research attempts devoted to load prediction on the individual household level [13], [2], [14]. They utilize smart meter data enriched with a set of household behavioral data (patterns of home appliances usage) and dwelling characteristics to benefit significant improvement in terms of the accuracy of the forecasts generated at the household level.

This paper presents design of refine model and introduces distributed learning of big data mining from multiple smart houses in a near real-time manner. With this, it is possible to efficiently train the system and increase the accuracy of detecting human activities to provide Patient and Doctor Feedback on usage patterns.

II. DATA MINING

Developments in data mining systems have enabled data collection, database formation, data management and stress-free data transfer into the electronic environment, and reliable and cheap data storage. Data mining, as can be understood from its name, is a technique that tries to extract mine-like valuable and variant information from data stacks. The purpose in data mining is the data collected in databases or repositories to be analyzed by examining via mathematical and statistical methods, and the available rules, structures or some different unforeseen information to be revealed

Data Mining Application Programs

WEKA (Waikato Environment for Knowledge Analysis) is an open-source-code Data Mining application development program which has been developed on the Java platform which is used by several people in the world today and developed in Waikato University, New Zealand.

Data Mining Categorization Models

Human mentality tends to categorize and classify the objects, events, situations around. Thus, he can understand and talk about the objects and events better. Categorization process in data mining is the process of categorizing the available data according to the determined features and of predicting this data category when some new data are added.

Decision Tree: Decision Tree is one of the most frequently used categorization models to analyze data. A decision tree is composed of root, trunk and leaf joints. Considering that the structure of a tree develops from root to leaves, it's formed from top to bottom. The most outer joint is the root joint. Each inner joint of the tree is separated to make the best decision with the help of algorithms [24]. Tree leaves form the category tags, namely, categorical characteristics, making the data in the data set groups. **Artificial Neural Networks**: Artificial Neural Networks (ANN) normally enable people to learn from similar or different events they experience, face, observe, to get new knowledge, to generalize those events, relating them with each other, to learn from their mistakes if they make any, and to make decision using all of these in an event they come across. ANN is inspired from human's problem solving with the abilities of thinking, observing, learning from mistakes, trial-error, that is, in a more general speaking, learning. An artificial neural network is an information processing system based on human cognition simulation. It is composed of plenty of calculating neural units attached together. These units are called neurons. The neurons in the nervous system form the network getting attached. ANN is a model developed on layers. The neurons in the network are arranged all along the layers. A neuron in a layer is connected to all of the neurons in the next layer. Every neuron in the layers takes informative signals from all of the neurons in the former layer, and multiplies every informative signal with their massive values. Process is done with the activation function to get the output, collecting weighted inputs. The output of this function is transmitted to all of the neurons in the next layer. This process is completed after done by the neuron in the output layer, too [3].

Multilayer Perceptron: MLP (Multilayer Perceptron) is an artificial neural network model which is mostly used and learns best [10]. It is known that Multilayer Perceptron has a very strong function in classifying prediction problems [10]. Multilayer Perceptron (MLP) network is the most popular, effective, and easy to learn model for complex, multilayered networks [34]. A typical multilayer perceptron (MLP) network consists of a set of source nodes forming the input layer, one or more hidden layers of computation nodes, and an output layer of nodes. Multilayer Perceptron (MLP) is one of the most used ANNs, where 80% of ANNs researches focused on. It consists of a series of fully interconnected layers of nodes where there are only connections between adjacent layers. General structure is showed in Figure 1.

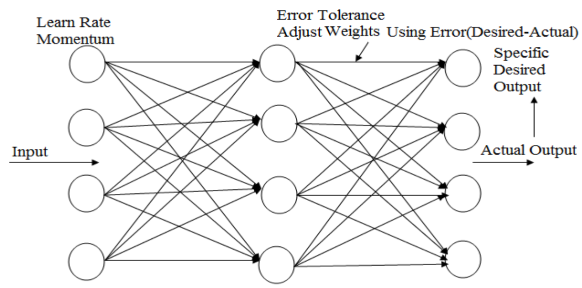

Fig.1: A Fully connected multilayer perceptron (MLP)

The first layer (the input layer) takes as inputs the various attribute values. The output of the nodes in the input layer, multiplied with the weights attached to the links, is passed to the nodes in the hidden layer. A hidden node collects the incoming weighted output values of the previous layers. Besides that, it receives also the weighted value of a bias node. The sum of the weighted input values is passed through a nonlinear activation function. The only requirements are that the output values of the function are bounded to an interval and that the nonlinear function can be differentiable. The output of a node in the hidden layer is fed into the nodes of the output layer. To each node in the output layer a class label is assigned.

Naive Bayes: Bayes classifiers are statistical classifiers. Bayes predicts the membership probabilities of the data, that is, their probability about belonging to a specific category. Bayes classifier is based on the Bayes theorem explained below:

A sample in a data set is composed of the input values X = {x1, x2, xm}. If it is pretended that the total number of the categories is m, the probability calculations are done with Equation 1 for the sample whose category is to be determined [19].

$$P(Ci|X) = \frac{P(X|Ci)P(Ci)}{P(X)} \tag{1}$$

p(x|Ci) : the probability for a sample from Category *i* to be *x*

P(Ci) : the first probability of Category *i*

p(x) : the probability for any sample to be *x*

P(Ci|x) : the probability for sample x to be from Category i (last probability)

Support Vector Machines: Support vector machines are a widespread method for binary categorization. SVMs can be seen as a perceptron which tries to find a hyper plane that split up the data. The perceptron simply tries to find a hyper plane. It does this process in a way that it separates the data of the hyper plane best. However, it is preferred for the hyper plane to separate the categories as much as possible because it is sought that the hyper plane generalizes the invisible data best with this way of separation.

Logistic Regression: The statistical analysis method which is used to express the relationship between a dependent variable and one or more than one independent variables numerically is called Regression Analysis. The purpose in Regression Analysis is to calculate to what extent the independent variables affect the dependent variable, that is, to predict the value of the dependent variable starting out of predicted variables [3]. The purpose of Logistic Regression Analysis is to calculate to what extent independent variables affect dependent variables.

Data Mining Clustering Model

K-Means is the dominant and mostly used clustering algorithm. It is used to divide into k number of categories which have been determined according to quality and characteristics of the records in the data set beforehand. Categorization is done by locating the records in the data set in the closest cluster centers. The success of the clustering process is determined by intra-cluster similarity to be maximum, inter-cluster to be at minimum. In other words, objects in the cluster are to be located in the closest way, clusters in the most distant way.

III. REVIEW OF EXISTING SYSTEMS

Smart Houses are a prominent field of research referring to environments adapted to assist people in their everyday life. Among the most relevant Smart Houses projects developed are the ACHE home in Colorado, USA [23], the MavHome in Texas, USA [8], the Gator Tech Smart House in Florida, USA [17], CareNet in the United Kingdom [31], the TERVA home in Finland [28], and others. There are many legal challenges that arise when developing a Smart House system.

The legal aspects vary according to the country in which the Smart House is being implemented. In Norway, a review [27] identified the following main legal challenges: data privacy, data access and management, stakeholders' interest, and informed consent of the users and/or the users' families. In order for a person to accept monitoring in his/her own home, legal regulations need to be established to assess patient-identifiable data [15].

Detecting human activities in smart homes by means of analyzing smart meters' data is studied in [7] the paper proposes two approaches to analyze and detect user's routines. One approach uses Semi-Markov-Model (SMM) for data training and detecting individual habits and the other approach introduces impulse based method to detect Activity in Daily Living (ADL) which focuses on temporal [1] the system use of frequent pattern mining, cluster analysis and prediction to measure and analyze energy usage changes sparked by occupants' behavior. Since people's habits are mostly identified by everyday routines, discovering these routines allows us to recognize anomalous activities that may indicate people's difficulties in taking care for themselves, such as not preparing food or not using shower/bath. The system addresses the need to analyze temporal energy consumption patterns at the appliance level, which is directly related to human activities analysis of activities that happen simultaneously.

MINING FREQUENT PATTERNS AND ACTIVITY PREDICTIONS FOR HEALTH CARE APPLICATIONS IN SMART HOMES

It starts by cleaning and preparing the data and then applying frequent pattern mining for discovering appliance-to-appliance associations, i.e., determining which appliances are operating together. Then, it uses cluster analysis to determine appliance to- time associations. With these two processes, the system is able to extract the pattern of appliance usage which is then used as input to the Bayesian network for short-term and long-term activities prediction (Fig. 1). The output of the system is utilized by specific health care applications depending on the intended use. For example, a health care provider might only interested in knowing activities related to cognitive impairment where tracking the sequence of daily activities is crucial for reminding the patient when abnormal behavior is detected.

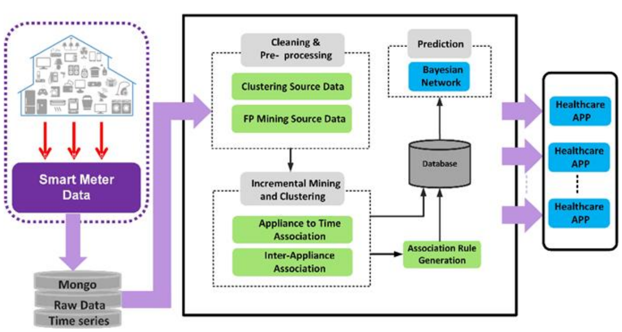

Fig.2: Model: Mining Frequent Patterns and Activity Predictions for Health Care Applications in Smart Homes [1]

Limitation of the existing system

To efficiently train the system and increase the accuracy of detecting human activities some of the common problems associated with existing system include:

- The system does not automatically map discovered appliances to potential activities
- The system does not provide prompt actions such as sending alert to patients or care providers.

IV. SYSTEM DESIGN AND IMPLEMENTATION OF THE PROPOSED SYSTEM

The proposed system presents distributed learning of big data mining from various smart houses in a near real-time manner. This will assist health applications to promptly take actions such as sending alert to patients or care providers .In addition, the system accommodate health ontology model to automatically map discovered appliances to potential activities. This means there is chance of efficiently train the system and increase the precision of sensing human activities (Fig. 2). We achieved this by using protégé tool to model the entities such as Kitchen Lights, Cooker, Bedroom Light, and Computer, Washing machine etc. thereby classifying them as active and non-active devices (Table 1).

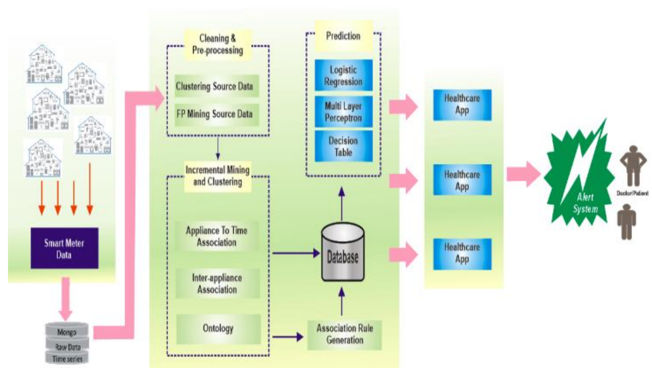

Fig. 3: Proposed Model: Mining Frequent Patterns and Activity Predictions for Health Care Applications in Smart Homes

Table 1. The Home Data Set

No	Name of the attribute
1.	**FurnaceHRV**
2.	**WashingMachine**
3.	**FridgeRange**
4.	**DisposalDishwasher**
5	**KitchenLights**
6	**Bedroom Outlets**
7	**BedroomLights**
8	**MasterOutlets**
9	**MasterLights**

A. Data Set Pre-processing Procedure

The data get prepared for the analysis before the data mining operation. The pre-processing for the Home data has been reported.

B. Data Cleaning

Since appliances cannot be 0 in a living room, Use, Gen, Grid, Duct Heater, Cellar Lights these data are deleted. Then, the 0 inputs in all of the observations are deleted. As a result, when the data out of 14 attribute data, as stated above, 5 attributes are deleted, 9 attribute data remain.

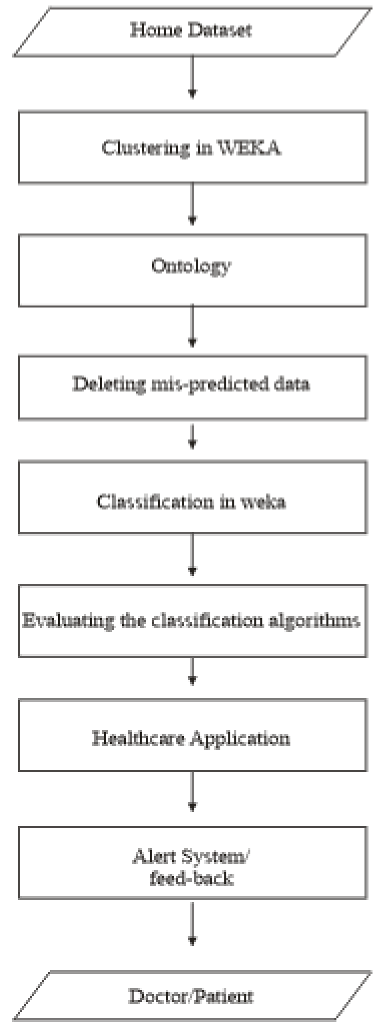

Fig. 4: Block diagram of the model applied on the home data set

C. Home Appliance using Ontology Knowledge

The proposed ontology is implemented using Protégé ontology development environment. The concepts of home energy management domain are organized into various hierarchies based on their functionality in domestic environment. A partial graph of this hierarchy of our implemented ontology is shown in Fig. (5). various approaches of ontology design have been proposed by researchers. We follow the methodology proposed by [30] to define ontology.

The focus of these ontologies is to provide semantic interoperability between heterogeneous components in smart home environments [16], [20].

This model addresses building a health ontology model to automatically map discovered appliances to potential activities. This means it is possible to competently train the system and increase the accuracy of sensing human activities, as in Fig.(5).

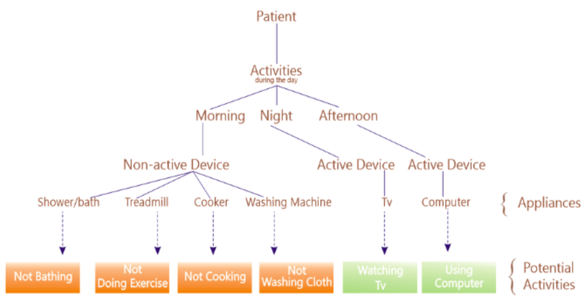

Fig. 5: Ontology model that maps discovered appliances to potentials activities .

Figure (5) Model is a partial view on Ontology Model that automatically map discovered appliances to potential activities

IV. RESULT AND DISCUSSION

The Model Applied on the Data Set: The model created in this paper, refines a model and introduce distributed learning of big data mining from multiple smart houses in a near real-time manner. This will help health applications to promptly take actions such as sending alert to patients or care providers. The prediction performance rates of the model and the programs have been evaluated by applying the WEKA on the data set of the designed

Table 2. Simple *K*-Means Algorithm Clustering Classes result

Attribute Clusters	Clusters Instances	Percentage
Cluster 1 → yes	13504	77%
Cluster 0 → no	4016	23%

The data set has been transformed into the Arff (Attribute Relation File Format) (.arff) which WEKA can read. The Home data has been clustered with K-Means clustering algorithm, and the result of clustering is that the new data set is to be created by deleting the data whose classes are mis-predicted. The total Number of the Samples is 17520 in all operations conducted in WEKA, where the number of the data whose classes are mis-predicted is 4016 and the number of the data whose classes are predicted is 13504 as shown in table (2) above.

```
@relation "HomeA-meter2-weka.filters

@attribute "FurnaceHRV [kW]" numeric
@attribute "WashingMachine [kW]" numeric
@attribute "FridgeRange [kW]" numeric
@attribute "DisposalDishwasher [kW]" numeric
@attribute "KitchenLights [kW]" numeric
@attribute "BedroomOutlets [kW]" numeric
@attribute "BedroomLights [kW]" numeric
@attribute "MasterOutlets [kW]" numeric
@attribute "MasterLights [kW]" numeric

@data
0.195338,0.005686,0.006892,0.005569,0.012154,0.020452,0.004899,0.046563,0.010393
0.182158,0.005679,0.094138,0.005412,0.0052,0.020571,0.008869,0.051677,0.009884
0.134808,0.005635,0.014786,0.00551,0.003173,0.020516,0.004901,0.053619,0.010183
0.182125,0.005672,0.082081,0.005445,0.003072,0.020506,0.004844,0.050543,0.009996
0.092988,0.00557,0.031901,0.005401,0.003154,0.020412,0.004841,0.049052,0.009927
```

Fig. 6: Data collected for home appliances in Arff

K-Means Clustering Algorithm:

K-Means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroid as bar center of the clusters resulting from the previous step. After we have these k new centroid, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we may notice that the k centers change their location step by step until no more changes are done or in other words centers do not move any more. Finally, this K-means algorithm aims at minimizing an objective function namely sum of squared error (SSE). SSE is defined as:

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} |p - m_i|^2 \quad (2)$$

Where E is sum of the square error of objects with cluster means for k cluster. *p* is the object belong to a cluster C_i and mi is the mean of cluster C_i.

Algorithm: *k*-means

The *k*-means algorithm for partitioning, where each cluster center is represented by the mean value of the objects in the cluster.

Input:

k: the number of clusters,

D: a data set containing *n* objects.

Output: A set of *k* clusters.

Initialize all weights and biases in *network*;
while terminating condition is not satisfied {
for each training tuple X in D {
// Propagate the inputs forward:
for each input layer unit j {
$O_j = I_j$; // output of an input unit is its actual input value
for each hidden or output layer unit j {
$I_j = \sum_i w_{ij} O_i + \theta_j$; //compute the net input of unit j with respect to the previous layer, i
$O_j = \frac{1}{1+e^{-I_j}}$; } // compute the output of each unit j
// Backpropagate the errors:
for each unit j in the output layer
$Err_j = O_j(1 - O_j)(T_j - O_j)$; // compute the error
for each unit j in the hidden layers, from the last to the first hidden layer
$Err_j = O_j(1 - O_j) \sum_k Err_k w_{jk}$; // compute the error with respect to the next higher layer, k
for each weight w_{ij} in *network* {
$\Delta w_{ij} = (l) Err_j O_i$; // weight increment
$w_{ij} = w_{ij} + \Delta w_{ij}$; } // weight update
for each bias θ_j in *network* {
$\Delta \theta_j = (l) Err_j$; // bias increment
$\theta_j = \theta_j + \Delta \theta_j$; } // bias update
} }

Fig.7: Backpropagation algorithm [33]

Method:

(a) Arbitrarily select *k* objects from *D* as the initial cluster centers;

(b) Repeat

(c) Repeat assigning each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;

(d) Update the cluster means, i.e., calculate the mean value of the objects for each cluster;

(e) Until no alteration;

Steps for K-Means Method:

Step(i): Select the number of clusters. Let this number be k.

Step(ii): Pick k seeds as centroid of the k clusters. The seeds may be picked randomly unless the user has some insight into the data.

Step(iii): Compute the Euclidean distance of each object in the Dataset from each of the centroid.

Step(iv): Allocate each object to the cluster ,it is nearest to base on the distances computed in the previous step.

Step(v): Compute the centroid of the clusters by computing the means of the attribute Values of the objects in each cluster.

Step(vi): Check if the stopping criterion has been met. If yes, go to step(vi). If not, go to step (iii).

Step(vii): One may decide to stop at this stage or to split a cluster or combine two clusters heuristically until a stopping criterion is met.

Algorithm: Backpropagation. Neural network learning for classification or numeric prediction, using the backpropagation algorithm.

The steps involved are expressed in terms of inputs, outputs, and errors,

Input:

- *D*, a data set consisting of the training tuples and their associated target values;
- *l*, the learning rate;
- *network*, a multilayer feed-forward network.

Output: A trained neural network.

Method:

The steps are described by;

Initialize the weights: The weights in the network are initialized to small random numbers (e.g., ranging from -1.0 to 1.0, or -0.5 to 0.5). Each unit has a bias associated with it. The biases are similarly initialized to small random numbers. Each training tuple, *X*, is processed by the following steps.

Propagate the inputs forward: First, the training tuple is fed to the network's input layer. The inputs pass through the input units, unchanged. That is, for an input unit, *j*,

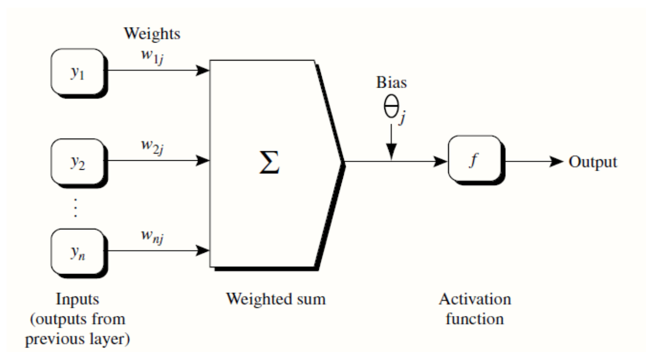

Fig.8 [33]: Hidden or output layer unit *j*: The inputs to unit *j* are outputs from the previous layer. These are multiplied by their corresponding weights to form a weighted sum, which is added to the bias associated with unit *j*. A nonlinear activation function is applied to the net input. (For ease of explanation, the inputs to unit *j* are labeled *y*1, *y*2,:::, *yn*. If unit *j* were in the first hidden layer, then these inputs would correspond to the input tuple .*x*1, *x*2, : : : , *xn*/.) its output, θ_j , is equal to its input value, *Ij* . Next, the net input and output of each unit in the hidden and output layers are computed. The net input to a unit in the hidden or output layers is computed as a linear combination of its inputs. To help illustrate this point, a hidden layer or output layer unit is shown in Figure 9.4. Each such unit has a number of inputs to it that are, in fact, the outputs of the units connected to it in the previous layer. Each connection has a weight. To compute the net input to the unit, each input connected to the unit is multiplied by its corresponding weight, and this is summed. Given a unit, j in a hidden or output layer, the net input, Ij , to unit j is:

$$I_j = \sum_i w_{ij} O_i + \theta_j, \tag{3}$$

Where wij is the weight of the connection from unit i in the previous layer to unit j; Oi is the output of unit i from the previous layer; and _j is the **bias** of the unit. The bias acts as a threshold in that it serves to vary the activity of the unit.

Terminating condition: Training stops when

All $\triangle wij$ in the previous epoch are so small as to be below some specified threshold, or the percentage of tuples misclassified in the previous epoch is below some threshold, or a pre-specified number of epochs has expired.

When the classifying algorithms whose success percentages are the highest are applied on the new data set, the results are as in Table 3.

Table 3 Accuracy Values of WEKA Classifying Algorithms

Classifying Algorithms	Accuracy (%)	Time taken to build model (Sec)
Multilayer Perceptron (MLP)	96.46	14.14
Logistic Regression (LR)	91.05	0.59
Decision table	91.03	0.72
m5rules	81.47	10.94
Random forest	71.59	17.14
Random Tree	66.80	0.33
Farthestfirst	1.00	0.06

The classifying accuracy values of this application have given very good results. The algorithm which has the lowest accuracy value is Farthestfirst with 1.00%, Random Tree with 66.80%, Random forest of 71.59%, Logistic Regression with 91.05 In this application, Logistic Regression, 91.05%, Decision Table with 91.03%, Multi-Layer Perception 96.46% have given the highest and the closest results (Fig. 9) in which their algorithm is with a high success of percentage. The level of appliance used is shown in (Fig. 10).

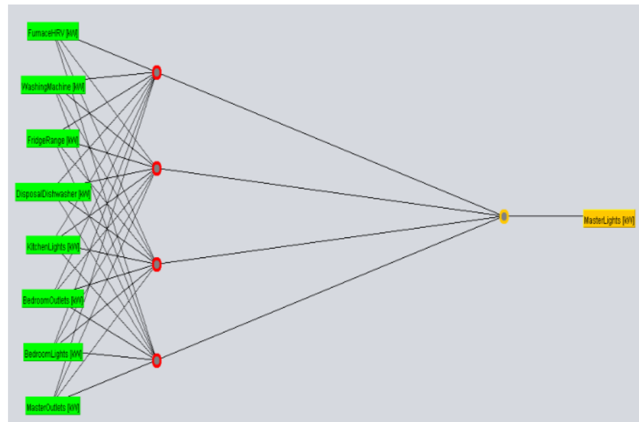

Fig. 9: Partial view of Multilayers of various Appliances

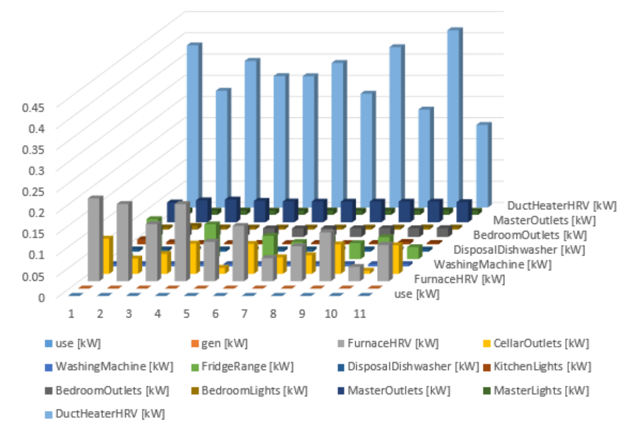

Fig.10: Level of appliance used

V. CONCLUSION

In this study, the model created with the data mining program WEKA using the home data that we have pre-processed has been applied. In the model applied, the data have been separated into two different clusters with *K*-Means clustering algorithm. After this clustering process, clustering operation has been realized on the remaining data by deleting the data whose classes are mis-predicted, in other words, mis-clustered. The Decision Trees, Multilayer Perceptron, m5rules, Decision table, Random forest, Random Tree, Farthest first and Logistic Regression algorithms out of the classifying algorithms in the WEKA program have been used in the process of classifying. The success rates of these algorithms have been compared. When the results are evaluated in general, the success in WEKA are between 91-100%. This has demonstrated that the success of the applied model is very good. When the comparison is done in terms of the algorithms, the Multilayer Perceptron algorithm has been the best algorithm with the highest success percentage in the program with 96.46%, the farthest first algorithm has been the algorithm whose success percentage is the lowest with 1.00%. This study has indicated that data mining can be a useful tool in the medical field because Doctors can be provided with daily activities of the patients in the progress of their treatment.

FUTURE WORK

Possible work could be in developing a system Using biometric / wearable device for smart houses to detect activities in the houses like standing, lying , sitting, walking, going up and down the stairs, working on computer or using body sensor networks (BNS) for continuous patient monitoring living in Smart Houses.

REFERENCES

[1] Abdulsalam Yassine, Shailendra Singh and Atif Alamri, (2017) Mining Human Activity Patterns from Smart Home Big Data for Healthcare Applications

[2] Aung, Z.; Williams, J.; Sanchez, A. (2012); Toukhy, M.; Herrero, S. towards Accurate Electricity Load Forecasting in Smart Grids. In Proceedings of the 4th International Conference on Advances in Databases, Knowledge and Data Applications, DBKDA, Saint Gilles, Reunion Island, 29 February–5 March; pp. 51–57.

[3] Beh Boon, C.; Jafri, M. Z. M.; Lim Hwee, S. (2011): "Mangrove Mapping in Penang Island by Using Artificial Neural Technique", Open Systems (ICOS), IEEE Conference on, (2011) 245-249.

[4] Brdiczka, O.; Langet, M.; Maisonnasse, J.; Crowley, J.L. (2009) Detecting human behavior models from multimodal Observation in a smart home. IEEE Trans. Autom. Sci. Eng., 6, 588–597.

[5] Chan, M.; Estève, D.; Escriba, C.; Campo, E. (2008) A review of smart homes—Present state and future challenges. Comput. Methods Programs Biomed., 91, 55–81.

[6] Chang, H.H. (2012) Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses. Energies, 5, 4569–4589.

[7] Clement J. , J. Ploennigs, and K. Kabitzsch, (2003) Detecting Activities of Daily Living with Smart Meters. Springer, Germany, 11 2014, ch. Advance Technology and Societal Change, pp. 143–160. [Online]. Available: https://link.springer.com/chapter/10.1007/ 978-3-642-37988-8 10

[8] Das, S.K.; Cook, D.J.; Battacharya, A.; Heierman, E.O.; Lin, T.Y. (2002) The role of prediction algorithms in the MavHome smart home architecture. IEEE Wirel. Commun., 9, 77–84.

[9] Dent, I. (2015) Deriving Knowledge of Household Behaviour from Domestic Electricity Usage Metering. Ph.D. Thesis, University of Nottingham, Nottingham, England, UK, 1 July.

[10] Delen, D., Walker, G., Kadam, A. (2005) “Predicting breast cancer survivability: a comparison of three data mining methods”, Artificial Intelligence in Medicine, vol 34, , pp113-127

[11] Firth, S.; Lomas, K.; Wright, A.; Wall, R. (2008) Identifying trends in the use of domestic appliances from household electricity consumption measurements. Energy Build., 40, 926–936.

[12] Gustafsson, A.; Gyllensward, M. (2005) The power-aware cord: energy awareness through ambient information display. In Proceedings of the CHI EA ’05, Portland, OR, USA, 2–7 ; pp. 1423–1426

[13] Ghofrani, M.; Hassanzadeh, M.; Etezadi-Amoli, M.; Fadali, M.F. (2011) Smart meter based short-term load forecasting for residential customers. In Proceedings of the North American Power Symposium (NAPS), Northeastern University, Boston, MA, USA, 4–6; pp. 1–5.

[14] Gajowniczek, K.; Ząbkowski, T. (2014) Short term electricity forecasting using individual smart meter data. Procedia Comput. Sci., 35, 589–597.

[15] Grguric, A. (2012) ICT towards elderly independent living; research and development center, Ericsson Nikola Tesla d.d.: Krapinska, Croatia,.

[16] Guo, T., Wang X. Hang, Pung, H. K., Zhang, D. Q., An ontology-based context model in intelligent environments, communication networks and distributed systems modeling simulation conference, 270-275

[17] Helal, S.; Mann, W.; El-Zabadani, H.; King, J.; Kaddoura, Y.; Jansen, E. (2005) The gator tech smart house: A programmable pervasive space. Computer, 38, 50–60.

[18] Helal, S.; 2002, B.; Lee, C.; Kaddoura, Y.; Ran, L.; Giraldo, C.; Kuchibhotla, S.; Mann, W. ()2003 Enabling location-aware pervasive computing applications for the elderly. In Proceedings of the First IEEE

International Conference on Pervasive Computing and Communications (PerCom 2003), Fort Worth, TX, USA, 23–26 March; pp. 531–536.

[19] Jiangtao, R.; Sau Dan, L.; Xianlu, C.; Ben, K.; Cheng, R.; Cheung, D. (2009): "Naive Bayes Classification of Uncertain Data", Data Mining,. ICDM '09. Ninth IEEE International Conference on, (2009) 944-949.

[20] Kim, E., Cho, J. (2006): An Ontology-Based Context Model in a Smart Home. Computational Science and Its Applications - ICCSA 2006, Lecture Notes in Computer Science,

[21] Korhonen, I.; Lappalainen, R.; Tuomisto, T.; Kööbi, T.; Pentikäinen, V.; Tuomisto, M.; Turjanmaa, V. Schilit, B.;

[22] Adams, N.;Want, R. (1994) Context-aware computing applications. In Proceedings of the 1994 First Workshop on Mobile Computing Systems and Applications, Santa Cruz, CA, USA, 8–9 December; pp. 85–90.

[23] Mozer, M.C. (1998) The neural network house: An environment that adapts to its inhabitants. Proc. AAAI Spring Symp. Intell. Environ., 110–114.

[24] Quinlan, J. R. (1993), C4.5: Programs for Machine Learning, Morgan Kaufmann, San Mateo, California.

[25] Rollins, S.; Banerjee, N.; Choudhury, L.; Lachut, D. (2014) A system for collecting activity annotations for home energy management. Pervasive Mob. Comput., 15, 153–165.

[26] Rodgers, J.; Bartram, L. (2011) Exploring ambient and artistic visualization for residential energy use Feedback. IEEE Trans. Vis. Comput. Graph., 17, 2489–2497.

[27] Sánchez, V.G.; Pfeiffer, C.F. (2016) Legal Aspects on Smart HouseWelfare Technology for Older People in Norway. In Intelligent Environments 2016: Workshop Proceedings of the 12th International Conference on Intelligent Environments; IOS Press: Amsterdam, the Netherlands,.

[28] TERVA (1998): Wellness monitoring system. In Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Hong Kong, China ; Volume 4, pp. 1988–1991.

[29] Tina, G.M.; Amenta, V.A. (2014) Consumption awareness for energy savings: NIALM algorithm efficiency evaluation. In Proceedings of the 5th International Renewable Energy Congress (IREC), Hhammamet, Tunisia, 25–27.

[30] Uschold, U. Gruninger, M. (1996): Ontologies: Principles, methods and applications, Knowledge Engineering Review, 11(2),.Web Ontology Language http://www.w3.org/TR/owl-features/

[31] Williams, G.; Doughty, K.; Bradley, D. (1998) A systems approach to achieving CarerNet-an integrated and Intelligent tele care system. IEEE Trans. Inf. Technol. Biomed, 2, 1–9.

[32] Zeifman, M.; Roth, K. (2011) Nonintrusive appliance load monitoring: Review and outlook. Consumer Electronics. IEEE Trans. Consum. Electron, 57, 76–84.

[33] J. Han and M. Kamber, “Data Mining: Concepts and Techniques,” Morgan Kaufmann, 2012.

[34] Zhao, Zhizhong, et al. "Application and comparison of BP neural network algorithm in MATLAB." Measuring Technology and Mechatronics Automation (ICMTMA), 2010 International Conference on. Vol. 1. IEEE, 2010.