HUMAN ACTIVITY PREDICTION FOR HEALTH CARE APPLICATIONS USING SMART METER

S.Bakiyalakshmi[1], K.Malini[2], H.Fametha[3]
[1,2,3]Assistant Professor

Abstract

One of the most challenging aspects that is greatly affected by vast influx of people in urban areas to Health care services. As a result, smart cities around the world are provide heavily in digital transformation and millions of homes are being equipped with smart devices (e.g., smart meters, sensors, and so on), which provide massive volumes of fine-grained and indexical data that can be inspecting to support smart city health care services. In this paper, we propose a model that deploy smart home big data as a means of learning and predicting human activity patterns for health care applications. We also include the proposed concept as cluster analysis, prediction to measure and analyse energy usage by occupants' behaviour using association rules. Since people's habits are mostly identified by day to day life, discovering these routines allows us to recognize regular activities with association of appliances usage. For the evaluation of the proposed mechanism, this paper uses the U.K. Domestic Appliance Level Electricity data set—time series data of power consumption collected from 2010 to 2016.The data from smart meters are recursively mined in the quantum/data slice of 24 h, and the results are implemented using weak tool across successive mining exercises. The results of identifying human activity patterns from appliance usage are presented in detail in this paper along with the accuracy of short- and long-term predictions.

INDEX TERMS Big data, smart cities, smart homes, health care applications, cluster analysis, incremental data-mining, association rules.

I.INTRODUCTION

Based on varies surveys done by research scholars, it is stated that major percentage of human population will be living in urban area. This vast change in the living style might lead a big challenge to the health care environments. Hence major health care centres are trying to adopt a major digital transformation. This major digitalization of health care environment could be made easy by adopting the data mining technologies. This transformation with various smart devices will generate a huge amount of data, which could be processed using data mining to understand people's routine life styles. The changes in the usage of appliances at home might help us to predict the human health condition using certain smart devices. Any changes in the appliances usage is recorded and noticed for a while, that indicates the people's difficulties in taking care of themselves. The underlying correlation between appliance usage inside the smart home and routine activities can be used by health care applications to detect potential health problems. This will help the health care system to monitor continuously each individual 24/7 hours.

The proposed model use energy data from smart meters installed at homes to observe and analyzes reading to recognize normal and abnormal activities of human. for example, if the people often perform more than one activity at the same time such as preparing food, washing cloths and hearing music, which indicate multiple appliances are operated together Such endeavor, how- ever, is very challenging since it is not easy to detect usage dependencies among various appliances when their operation overlap or occur at the same time. Furthermore, deriving accurate prediction of human activity patterns is influenced by the probabilistic relationships of appliance usage events that have dynamic time intervals. It must be noted that in practice load disaggregation is carried by Non-Intrusive Appliance Load Monitoring (NALM) technique. NALM is a technique used to disaggregate a home's power usage into individual appliances and label them for further mining and analysis.

The data from smart meters are continuously mined in the quantum slice of 24 hours and then proposes a human activity pattern mining model based on appliance usage variations in smart homes. The model which utilize k-means clustering algorithms can identify appliance-to-appliance and appliance-to- time associations through incremental mining of energy consumption data [2]. This is not only determining activity routines, but also capable of detecting sudden changes of human activities that determine attention by a health provider. We propose a prediction model as Bayesian network to predict the use of multiple appliances and household energy consumption. It is capable of short-term predictions ranging from up to 24 hours and long-term predictions for days, weeks, months, or seasons. To adding accuracy of the system, the prediction model integrates probabilities of appliance- to-appliance and appliance-to-time associations, thus recognizing activities that occur in certain patterns more accurately. For experimental implementation of the proposed system, this research uses the UK-DALE, time series data of power consumption collected from 2010 to 2016 with mining exercises using weka tool.

Smart meter provides us current information to visualize, analyze, and optimize the energy consumption of buildings, to enable demand-response optimizations, and to identify the usage of appliances. They also can be used to help people to stay longer independent in their homes by detecting their activity and their behavior models to ensure their healthy level[3].

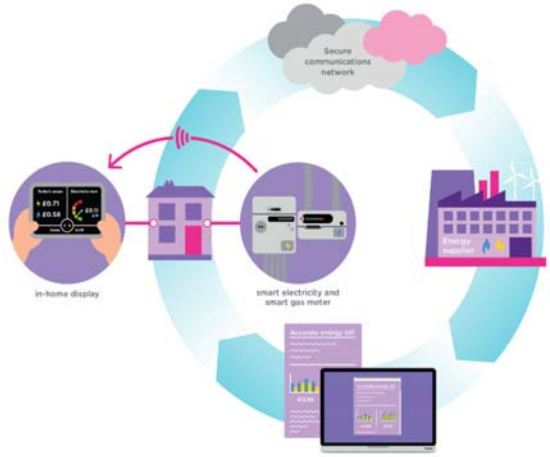


Figure1: Smart meter working system

Using data analytics for smart meters to detect and predict behavioral abnormality for remote health monitoring [4]. It uses everyday appliances usage from smart meter and smart plug data to trace regular activities and learn unique time segment groups of appliance's energy consumption. The study employs hierarchical probabilistic model-based detection to infer about discovered anomalous behavior. This in tern can be used to understand the criticality of some abnormal behaviors for Sustaining better health care. In an experimental demonstration for observing and measuring energy consumption of appliances is presented. The study aims to provide a portrait profile of activities of daily living for elderly patients independently living at home. The data is also used to mine important patterns of changes for short-term and long-term anomaly detection of urgent health conditions. The work in, uses Bayesian networks to predict occupant behavior from collected smart meters data. The study proposes behavior as a service based on a single appliance but does not pro- vide a model to be applied for real-world scenarios. Authors in and, used time-series multi-label classifier to forecast appliance usage based on decision tree correlations, however, the study takes only the last 24-hour window along with appliance sequential relationships. Clustering approach is used to identify the distribution of consumers' temporal consumption patterns, however, the study does not consider appliance level usage details. This might not be applicable for human activity recognition since specific activities require individual and multiple appliance to appliance and time associations. The work in considers the appliances' ON and OFF status to detect usage pattern using hierarchical and c-means clustering. However, the study does not consider the duration of appliance usage or the expected variations in the sequence of appliance usage. The work in proposes graphical model-based algorithm to predict human behavior and appliance interdependency patterns and use it to predict multiple appliance usages using a Bayesian model. The above-discussed approaches do not consider appliance level usage patterns, which is critical in determining human activity variations. Furthermore, our experiments are conducted using a much larger dataset than existing studies although there are similarities in data analytics techniques between the proposed study and existing work.

II.PROPOSED SYSTEM

Figure 1represents the proposed model. It starts by cleaning and preparing the data and then applying Rule mining for discovering appliance-to-appliance associations, i.e., determining which appliances are operating together.

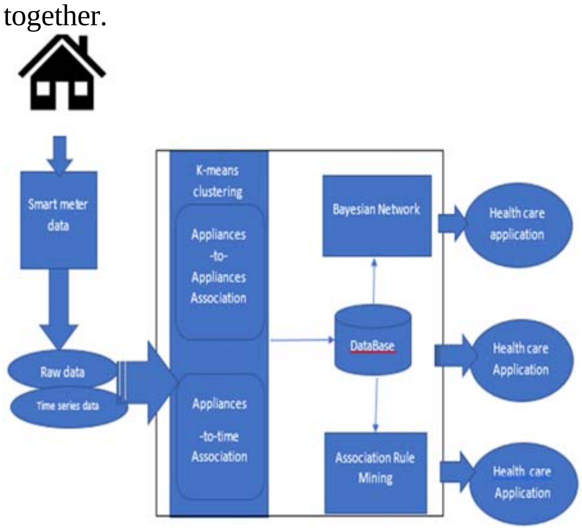


Figure2: Healthcare System architecture

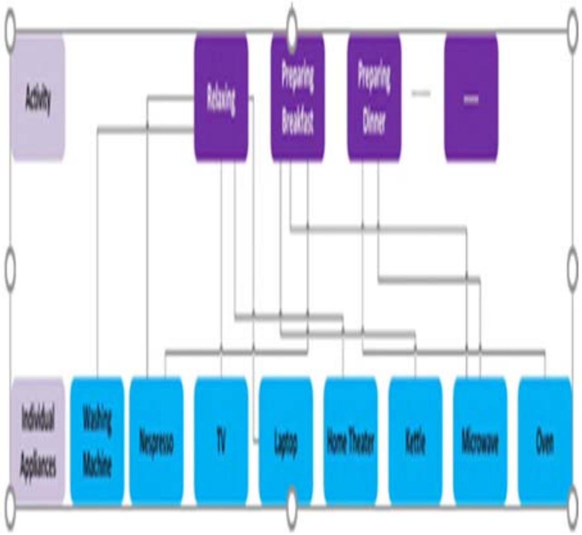
Then, it uses cluster analysis to determine appliance-to-time associations. With these two processes, the system can extract the pattern of appliance usage which is then used as input to the Bayesian network for short-term and long-term activities prediction. The output of the system is utilized by specific health care applications depending on the intended use. For example, a health care provider might only have 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.

III.DATA PREPARAION

The dataset used in this study is a collection of smart meters data from various houses in the United Kingdom (UK)[1]. This dataset includes 400 million raw records at time resolution of 6 seconds. In the first stage of the cleaning process we developed customized procedures to remove noises from the data and prepare it for mining. After cleaning and preparation, the dataset is reduced to 20 million. Additionally, we developed a synthetic dataset for preliminary evaluation of the model.

DERIVING FREQUENT PATTERN OF HUMAN ACTIVITIES

The aim is to discover human activity patterns from smart meters data. For example, activities such as ''Watching TV, Cooking, Using Computer, Preparing Food and Cleaning Dishes or Clothes'' are usually regular routines. Our aim is to detect the patterns of these activities so that a health care application, that monitors sudden changes in patient's behaviour, can send timely alert to health care providers. In pursuing such process, all appliances that are registered active during the 30-minute time interval are included into the source database for frequent pattern data mining [6]. The energy trace of appliances (TV, Oven and Treadmill) is related to human activities such as leisure/relaxation time, food preparation, and exercising. A simplified example which describes possible relationships between appliance usage and activities is shown in *figure3*. Extracting human activity patterns is not only discovering the individual appliance operation, but also the appliance-to-appliance associations; i.e., the patterns of activities that are combined such as washing clothes while exercising or watching TV.



(a)

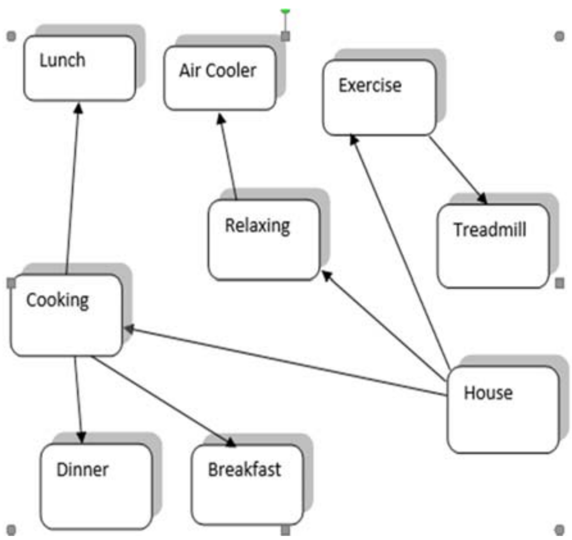


Figure 3: Example possible relationship inside a smart home

IV.DATA CLUSTERING: K-MEANS ALGORITHM

Discovering appliance-to-time associations is vital to health applications that monitor patients' activity patterns daily. In this section, a clustering analysis mechanism is used to discover appliance usage time with respect to hour of day (00:00 - 23:59), time of day (Morning, Afternoon, Evening, Night), weekday, week and/or month of the year. Appliance-to-time associations are underlying information in the smart meter time series data which include sufficiently close time-stamps, when relevant appliance has been recorded as active or operational. Using this data, we can group a class or cluster of appliances that are in operation simultaneously or overlapping. k-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem [8]. 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 centres, one for each cluster. These centres 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 centre. 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 centroids as barycentre of the clusters resulting from the previous step. After we have these k new centroids, a new binding must be done between the same data set points and the nearest new centre. A loop has been generated. As a result of this loop we may notice that the k centres change their location step by step until no more changes are done or in other words centres do not move any more. Finally, this algorithm aims at minimizing an objective function know as squared error function given by:

$$J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_i} \left(\left\| x_i - v_j \right\| \right)^2$$

where,

'$||x_i - v_j||$' is the Euclidean distance between x_i and v_j.

'c_i' is the number of data points in i^{th} cluster.

'c' is the number of cluster centres.

<u>ALGORITHM 1 K-MEANS ALGORITHM</u>

Let $X = \{x_1, x_2, x_3, \ldots\ldots, x_n\}$ be the set of data points and $V = \{v_1, v_2, \ldots\ldots, v_c\}$ be the set of centres.

1) Randomly select '*c*' cluster centres.
2) Calculate the distance between each data point and cluster centres.
3) Assign the data point to the cluster centre whose distance from the cluster centre is minimum of all the cluster centres.
4) Recalculate the new cluster centre using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

where, 'c_i' represents the number of data points in i^{th} cluster.
5) Recalculate the distance between each data point and new obtained cluster centres.
6) If no data point was reassigned then stop, otherwise repeat from step 3).

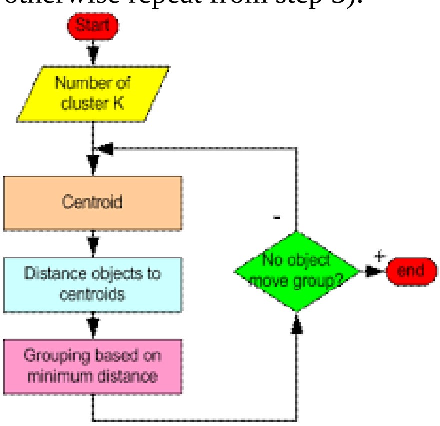


Figure 4: K-Means process flow

V.DATA CLASSIFICATION: APRIORI ALGORITHM

The Apriori Algorithmic an influential algorithm for mining frequent item sets for Boolean association rules. Frequent Itemset: The sets of items which has minimum support. Apriori Property: Any subset of frequent itemset must be frequent. Join Operation: To find Lk, a set of candidates k-item sets is generated by joining Lk-1with itself.Find the frequent item sets: the sets of items that have minimum support –A subset of a frequent itemset must also be a frequent itemset i.e., if {AB} isa frequent itemset, both {A} and {B} should be a frequent itemset –Iteratively find frequent itemsets with cardinality from 1 to k (k-itemset) Use the frequent itemsets to generate association rules.

ALGORITHM 2: APRIORI ALGORITHM

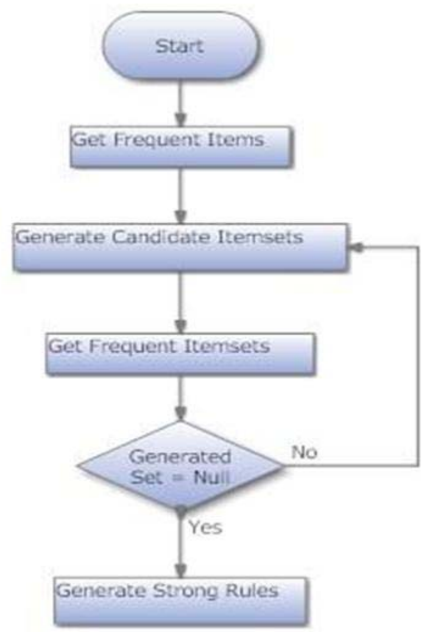
Join Step: $C_k$is generated by joining Lk-1with itself

•Prune Step: Any (k-1)-itemset that is not frequent cannot be a subset of a frequent k-itemset •Pseudo-code:

Ck: Candidate itemset of size k

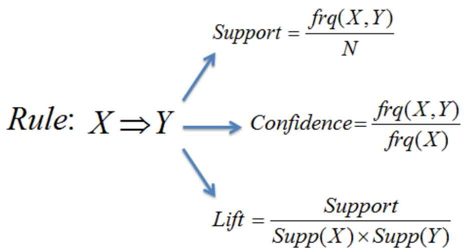
Lk: frequent itemset of size k L1= {frequent items};

```
for(k= 1; Lk!=∅; k++) do begin
    Ck+1= candidates generated from Lk;
    for each transaction t in database do
        increment the count of all candidates in Ck+1
            that are contained in t
        Lk+1=candidates in Ck+1 with min_support
    end
return∪kLk;
```



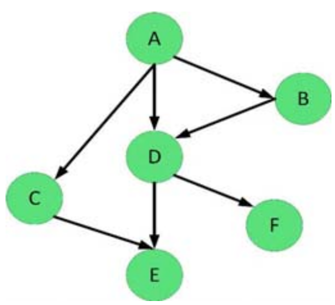
VI.ASSOCIATION RULES

Association rule find all sets of items (item sets) that have support greater than the minimum support and then using the large item sets to generate the desired rules that have confidence greater than minimum confidence. The lift of the rule is the ratio of the observed support to that expected if X and Y were independent.



VII.ACITIVITY PREDICTION: Bayesian network

We integrate the frequent patterns and appliance-to-time associations to learn about the use of multiple appliances and build the activity prediction model. The mechanism utilizes Bayesian network [7] which is a directed acyclic graph, where nodes represent random variables and edges indicate probabilistic dependencies. An example of Bayesian network, representing 6 random variables, is shown in **figure (6).** One of the key features of a Bayesian network is that it includes the concept of causality. For example, the link/arc between A to C in figure indicates that node A causes node C, which means that the directed graph in a Bayesian network is acyclic.



A Bayesian network, Bayes network, belief network, Bayes(ian) model or probabilistic directed acyclic graphical model is a probabilistic graphical model (a type of statistical model) that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG). For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases. Formally, Bayesian networks are DAGs whose nodes represent variables in the Bayesian sense: they may be observable quantities, latent variables, unknown parameters or hypotheses. Edges represent conditional dependencies; nodes that are not connected (there is no path from one of the variables to the other in the Bayesian network) represent variables that are conditionally independent of each other. Each node is associated with a probability function that takes, as input, a particular set of values for the node's parent variables, and gives (as output) the probability (or probability distribution, if applicable) of the variable represented by the node. For example, if m parent nodes represent m Boolean variables then the probability function could be represented by a table of 2^m entries, one entry for each of the 2^m possible combinations of its parents being true or false. Similar ideas may be applied to undirected, and possibly cyclic, graphs; such as Markov networks. Efficient algorithms exist that perform inference and learning in Bayesian networks. Bayesian networks that model sequences of variables (*e.g.* speech signals or protein sequences) are called dynamic Bayesian networks. Generalizations of Bayesian networks that can represent and solve decision problems under uncertainty are called influence diagrams. A Bayesian network is defined by the probabilistic distribution presented in equation (1)

$$p(x1,\ x2,\ \ldots,\ xn) = \prod_{i=1} p(xi|parents(xi)) \qquad (1)$$

As mentioned above, our probabilistic prediction model is constructed based on integrating probabilities for appliances- to-time associations in terms of hour of day (00:00 - 23:59), time of day (Morning, Afternoon, Evening, Night), week- day, week, month, season, and appliance-to-appliance associations. The topology of the resulting Bayesian network has only one level of input evidence nodes, accompanied by respective unconditional probabilities, converging to one output node. Equation (2) presents the posterior probability or marginal distribution for the proposed prediction model.

$$p(.) = p(Hour) \times p(Time\ of\ day) \times p(Weekday) \times p(Week) \times p(Month) \times p(Season) \quad (2)$$

Table 4 shows sample of the training data. This data is derived from clustering and frequent patterns analysis where the probability of each appliance represents its operation during the specified period. This information is utilized by the Bayesian mechanism to determine and predict active appliances, operating simultaneously, using historical evidence from the cluster analysis (appliance-time association) and frequent pattern mining (appliance-appliance association). Furthermore, appliance prediction results establish the foundation for human activity prediction from the next hour up-to 24 hours (short-term) and days, weeks, or months (long-term). Next, we evaluate our model and provide result analysis [5].

NODE ROBABILITY FOR APPLIANCES

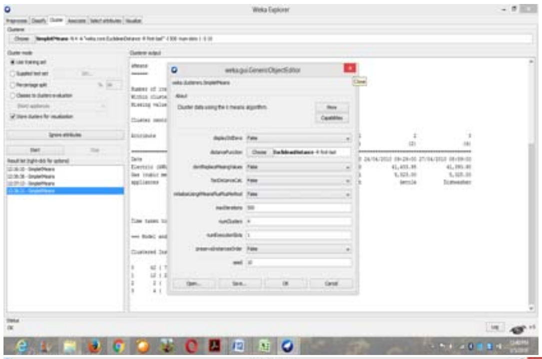
	Cluster Center/Centroid				
Appliance	00:00	00:30	01:00	01:30	02:00
2	0.0608	0.0602	0.1257	0.1414	0.1404
3	0.0338	0.0361	0.1078	0.1212	0.1149
4	0.1351	0.1205	0.1198	0.1010	0.0936
8	0.0270	0.0422	0.0719	0.1010	0.0979
10	0.0878	0.0783	0.0778	0.0657	0.0553
11	0.0068	0.0181	0.0180	0.0404	0.0511
12	0.6014	0.5060	0.4731	0.3939	0.4043
13	0.0000	0.0000	0.0000	0.0051	0.0043
15	0.0473	0.1386	0.0060	0.0303	0.0383

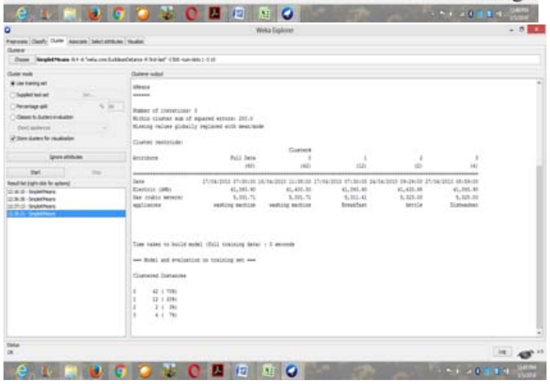
2 = Laptop, 3 = Monitor, 4 = Speakers
8 = Kettle, 10 = Running Machine, 11= Laptop2
12= Washing Machine, 13 = Dishwasher, 15 = Microwave

VIII.EXPERIMENT RESULT

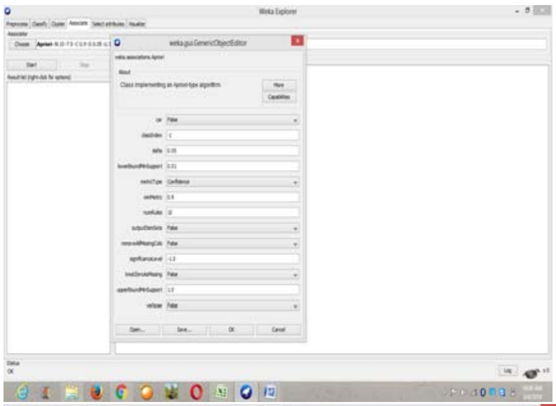
For the evaluation of the proposed model, we performed our experiments using the dataset UK-Dale along with the synthetic dataset to inspect intermediate and final results. The (UK-Dale) dataset includes time series data of power consumption collected between 2012 and 2015. The concept is implemented using weka tool and the results obtained is as follows:

CLUSTERING RESULT

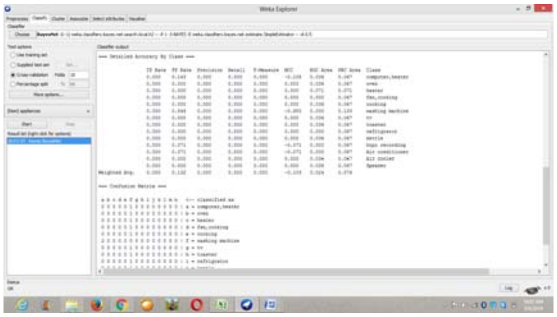




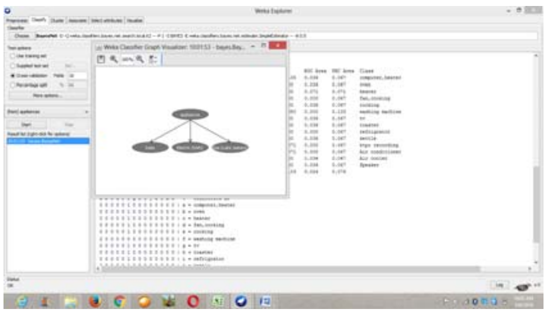
CLASSIFICATION RESULT



ASSOSIATION RULES



BAYSEIAN NETWORK RESULT



IX.CONCLUSION

In this Paper, we presented a model for predicting the human activity patterns from smart meter device that could be used in health application to track the wellbeing of everyone. The routine activities are learned by applying association rule such as appliances –to-appliance and appliances-to-time association. K-means clustering, and Association rule mining techniques are used to predict the human behavioural pattern and Bayesian network is used to present the experimental outputs. For future work, the system can promptly take action to send SMS to the patient's guardians or health care providers. In addition, the system can be trained using huge amount of data collected from multiple houses. It increases the accuracy of system prediction.

References

[1] United Nation. (2014). World Urbanization Prospect. [Online]. Available: http://dl.acm .org/citation.cfm?id=308574.308676

[2] Mining Human Activity Patterns From Smart Home Big Data for Health Care Applications ABDULSALAM YASSINE[1], (Member, IEEE), SHAILENDRA SINGH[2], AND ATIF ALAMRI[3,4], (Member, IEEE).

[3] A. Yassine, A. A. N. Shirehjini, and S. Shirmohammadi, ''Smart meters big data: Game theoretic model for fair data sharing in deregulated smart grids,'' IEEE Access, vol. 3, pp. 2743–2754, 2015.

[4] K. Gajowniczek and T. Ząbkowski, ''Data mining techniques for detect- ing household characteristics based on smart meter data,'' Energies, vol. 8, no. 7, p. 7407, 2015. [Online]. Available: http://www.mdpi.com/ 1996-1073/8/7/7407

[5] K. Basu, V. Debusschere, and S. Bacha, ''Appliance usage prediction using a time series based classification approach,'' in Proc. IEEE 38th Annu. Conf. Ind. Electron. Soc. (IECON), May 2012, pp. 1217–1222.

[6] J. Han, J. Pei, and M. Kamber, ''Data mining: Concepts and techniques,'' in Cluster Analysis: Basic Concepts and Methods, 3rd ed. San Mateo, CA, USA: Morgan Kaufmann, 2011, ch. 10, pp. 443–494. [Online]. Available: http://www.sciencedirect.com/science/book/978 0123814791

[7] D. Heckerman, ''Bayesian networks for data mining,'' Data Mining Knowl. Discovery, vol. 1, no. 1, pp. 79–119, 1997. [Online]. Available: http://dx.DOI.org/10.1023/A:1009730122752

[8] W. Haizhou and S. Mingzhou, ''Ckmeans.1d.dp: Optimal k-means clustering in one dimension by dynamic programming,'' R J., vol. 3, no. 2, pp. 29–33, 2011. [Online]. Available: http://journal.r-project.org/ archive/2011-2/RJournal_2011 2_Wang +Song .pdf