

Mining Frequent Patterns and Associations from the Smart meters using Bayesian Networks

Dr. G. Rama Subba Reddy

Associate Professor, Dept. of CSE, Mother Theresa Institute of Engineering & Technology, Palamaner, Andhra Pradesh, India.

Email: subbareddy1227@gmail.com

B. Dilip Chakravarthy

Assistant Professor, Dept. of CSE, Mother Theresa Institute of Engineering & Technology, Palamaner, Andhra Pradesh, India.

Email :dilip611@gmail.com

C. ReddiNeelima

Assistant Professor, Dept. of CSE, Mother Theresa Institute of Engineering & Technology, Palamaner, Andhra Pradesh, India.

reddineelimapg@gmail.com

G. Divya Zion

Assistant Professor, Dept. of CSE, Mother Theresa Institute of Engineering & Technology, Palamaner, Andhra Pradesh, India.

Email :Email :ziondivya@gmail.com

ABSTRACT

In today's world migration of people from rural areas to urban areas is quite common. Health care services are one of the most challenging aspect that is must require to the people with abnormal health. Advancements in the technologies lead to build the smart homes, which contains various sensor or smart meter devices to automate the process of other electronic device. Additionally these smart meters can be able to capture the daily activities of the patients and also monitor the health conditions of the patients by mining the frequent patterns and association rules generated from the smart meters. In this work we proposed a model that is able to monitor the activities of the patients in home and can send the daily activities to the corresponding doctor. We can extract the frequent patterns and association rules from the log data and can predict the health conditions of the patients and can give the suggestions according to the prediction. Our work is divided in to three stages. Firstly, we used to record the daily activities of the patient using a specific time period at three regular intervals. Secondly we applied the frequent pattern growth for extracting the association rules from the log file. Finally, we applied k-means clustering for the input and applied Bayesian network model to predict the health behavior of the patient and precautions will be given accordingly.

Keywords – Bayesian networks, Cluster analysis, FP pattern, Human activity prediction.

Date of Submission: April 10, 2018

Date of Acceptance: May 18, 2018

I. INTRODUCTION

The daily interaction of ubiquitous sensors and devices to connect physical and virtual things via a seamless network is known as Internet of things (IoT). To build a new process "anytime, anywhere any service for anybody", the IoT includes different heterogeneous techniques. There are some likelihood probabilities which are provided by IoT will make this possible to provide several applications related to it. From those, the smart home is highly developed research domain in smart automation systems with motive of improvement of users comfort and guarantee for their security and its condition at least costs of operation. As smart home is an automated environment it is able to monitor, detect and record everyday activities with the utilization of different sensors and communication technologies. The routine activities of users' create patterns that perform a crucial role in smart home environment. Those patterns are utilized in favor of recognizing users' activity which might be useful in the enhancement of smart home applications in terms of

efficiency and energy management, healthcare and security etc.

On the remote monitoring the identification of users' daily activities is done. It has applications in more domains like health care and daily care [1]. So research on domestic activity recognition is attaining high interest, particularly because of recent trends to move healthcare from hospitals to houses of patients and facilitate to live independently. Anyways, possible applications go far beyond healthcare. For instance, identifying daily life activities also helps to support home automation and energy savings in smart homes/buildings. Domestic activity recognition relies on creating an inference by information fusion from heterogeneous sensors and uncertainty relevance because of stochastic nature of human behavior and imperfect sensing equipment. To find a human activity in smart environment, various diverse sensors are included, ranged from number of sensors, sensors for measuring heartbeat, walking patterns, and environmental sensors. The sensor's choice and their placing for the identification of particular

activities are influenced by the way how it affects the activities of human and therefore guessed reasonable results for this activity. Prime constraints such as intrusiveness, privacy, energy efficiency, cost, and providing more preciseness activity recognition with less number of least price and non-intrusive sensors are to be taken into consideration. Those are the challenges in recent research.

Some of the examples of activity recognition in smart homes works with D-S theory. Lee et al. proved that D-S theory of evidence could provide a way to incorporate and minimize the uncertainty's impact efficiently. The D-S theory of evidence together with a lattice structure is used for identifying basic human activities, such as flushing a toilet, for the purpose of assistive living. In this paper we argue that most of the important daily life activities can be recognized only with the use of energy monitors like smart meter data. By the year 2030 all homes in India are going to be provided a smart meter, with the same trends in rest of the developed nations. No additional hardware is needed because every home is going to be equipped with sensors so that recognition of domestic activity can be performed. The remainder of this paper is organized as follows. Section 2 describes about the related work and literature study of our work. In Section 3 we have presented our proposed work. Section 4 describes the results and discussion. Conclusions are drawn in Section 5.

II. RELATED WORK

Several researches proposed IoT related smart home environment to improve the protection, safety and comfort of residents with least costs of operation. Clearly, the utilization of sensors inside smart homes is indispensable to keep track of user activities. Activities Daily Living (ADL) of users is monitored and the general activity patterns are modeled as per the position of user in his/her environment. By that any normal or abnormal behavior of activity pattern will be identified. Furthermore other researchers had used to reduce the issues related to activity recognition with some different approaches in several real world activities. However, the diversity and complexity in activities are often very high in daily living.

In [3], the EM-algorithm is used for grouping of same objects. It is easy and fast but the efficiency of this relies on the number of input features, objects and along with iteration number. Jakkula et al. [4] suggest to divide with a centroid through k-means clustering approach. The scheme called distance measurement allots a score to cluster with the least value. Nevertheless the algorithm's efficiency is dependent on numerous clusters, cluster center selection and iterations. In [5] a hierarchical clustering algorithm used in a distributed environment evaluates its performance and preciseness by the application of valid measures such as entropy, time and coefficient of time. There is no need to determine number of clusters formerly and implementation is easy. Nevertheless the hierarchical algorithm yields clusters

with less quality and requires more time for execution when a large dataset was given. The SOM algorithm [6] gives the high accuracy in objects classification into related clusters. Additionally it yields better outcomes when compared to k-means and EM-clustering algorithm whenever random datasets are used. Nevertheless as 'k' (number of clusters) increases considerably there is decrement in performance otherwise with the use of huge dataset, this algorithm produces poor outputs.

Generally, there has some ambiguity while noisy data processing in prevailing clustering algorithms. Indeed that noise becomes tough to add an object to particular cluster because the algorithms' results will get affected by this noise. To overcome this problem we have an algorithm known as K-pattern clustering. On the other side, some tasks combine user behavior via activity recognition. Finding user activities usually resembles the gathering of observation order for new events recognition. Some models for predicting activity includes the prediction of sequential activity with the utilization of decision trees, the k-nearest neighbor and the Markov or Bayesian models. Alam et al. [7] uses probabilistic models like Hidden Markov Models (HMM) for user activities modeling. This method is widely used to identify the spatio-temporal relationships between the sensor data and also to find time series forecasting [8, 9]. Even so, the run-time is too long for big volume of data. In [10], On C4.5 classifier basis another classification model for recognition of activity is taken into consideration. This technique yields better results. Its performance in recognition preciseness is not much greater than the neural network algorithm; because of diversity and complexity of activities in real world [11]. Here we have discussed few more literature study of our work from last three years.

In 2015 K.Jack and K.William was given their work on domestic appliance-level electricity demand and whole-house demand from UK homes. A dataset that is possessed from disparate homes was adopted in this paper. Whichever the smart homes are equipped with smart devices contains huge number of meter readings. This meter readings changes from home to home on the users' equipment usage level basis. This paper presents an approach to data assembling from smart homes on utilization of installed appliances [12].

In 2016 M. S Hossain proposed a work by paper for showing patient's state recognition system for healthcare by using speech and facial expression. This paper demonstrates a model for addressing the complete framework on healthcare. It majorly deals with the theme of detecting a state of a patient to provide better recognition accuracy to get least cost model. This paper primarily relies on two kinds of inputs are audio and video which are attained in multi-sensory environment which gives average efficiency with 98 percent in detection. In 2016 M. U1Alam, N Roy, M. Petruska and A Zemp proposed their work on smart energy group variation on the basis of behavioral anomaly recognition. This paper

work proposes accessing a data analytic which classifies the abnormalities in utilization of energy based on the inhabitants' behavioral deformity. Research thesis significantly depends on identifying daily usage of appliances ranged between smart meter and smart plug data that keeps track of daily activity even at days and nights and then to learn every appliance consumption of energy. Majorly this paper's provide fundamental technologies which are non-intrusive health monitored are deployed at a very large scale without additional sensors need to be equipped in any home with multi-inhabitants.

In 2015 C. Chelmis, J. Kolte and V. K Prasanna introduced their work on big data analytics to demand response: Clustering more time and space. This paper demonstrates that purpose of different representations of data on consumption of electricity. It also describes the use of behavior structures detected at (i) different times in a day, (ii) days-of-the-week or (iii) per annum or yearly for a clientele and similar one's are mined in clientele's with different features by exact clustering of time-series data.

In 2015 K. Gajowniczek and T. Zabkowski proposed work on techniques in data mining for identifying the features of household on smart meter data basis. In this the objective is regulating the usage patterns of household's structure, therefore by keeping the view of usage of smart metering systems it provides extra intellect on their utilization. Many of the techniques in machine learning which are unsupervised are processed in observing the usage patterns at multiple households. This work carries out the related solutions to smart metering systems which contributes to highly developed energy consciousness; suitable for predicting usage exactly gives input to demand response systems in households and suggests the users on energy saving periodically. This paper provided some outcomes showing that defining household features from smart meter data is exact and enables to extract usual trends in data.

2.1 DM algorithm in Healthcare

Healthcare covers detailed processes of the diagnosis, treatment and prevention of disease, injury and other physical and mental impairments in humans [13]. In many of the nations the healthcare industries are at rapid pace. These healthcare industries are treated as the places with good quality of data as they make huge data and also include electronic medical records, administrative reports and other benchmarking finding [14]. This data from healthcare is being under-utilized. The healthcare industries use different methods and are discussed in below:

2.1.1. Anomaly Detection

It is used in identifying the more important variations in a dataset [15]. Bo Lie et al [16] was used three different anomaly detection method, standard support vector data

description, density induced support vector data description and Gaussian mixture to measure the correctness of anomaly detection on a dataset of liver disorder which is uncertain and obtained from UCI. This method is calculated by UCI accuracy. The outcomes for a balanced dataset are attained with an average of 94%. For the similar dataset the average standard deviation is 2.63. The datasets which are uncertain are ignorable in available datasets. To resolve this anomaly detection provides a better way. In this paper we are not focusing on method's effectiveness.

2.1.2. Clustering

It is common theoretical task where one can find finite clusters collection for data description. Rui Veloso [17] was used a method called vector quantization in clustering in expecting readmissions in intensive medicine. This method used algorithms like k-means, k-medoids and x-means. The utilized datasets in this study are from patient's diagnosis process and results from laboratory. The evaluations of each and every algorithm are done with the use of Davies-Bouldin Index. The k-means, x-means and k-medoids obtained the better, fair and poor results respectively. Based on the results the researchers choose the good result which helps in characterization of different kinds of patients with high probability of readmission. This paper is only focused on the vector quantization method.

2.1.3. Classification

The process of discovering a predictive learning function which performs classification of data item into one of different predefined classes is known as Classification. The following subsections cover the classification related work.

2.1.4. Statistical

The MTS is widely applied in multivariable statistical analysis. The Mahalanobis distance (MD) is used to construct statistical judgments to differentiate a group from the rest of the other and the Mahalanobis space (MS) is to represent the observation's abnormality degree from the reference group that is already known. In the case of statistical classifiers, Su et al. used the Mahalanobis Taguchi System (MTS) to model the prediction pattern of pressure ulcers. The class with imbalance issues is more prevalent in healthcare datasets. Using of data mining algorithms are often affects with the skewed distribution when we use skews or datasets which are imbalanced.

2.1.5. Decision Tree

Many studies explored the method of decision tree for the analysis of clinical data. The authors Sharma & Om [18], Wang et al. [19] and Zolbanin et al.[20] was used the algorithm of decision tree for their research related work. To do a prediction it is necessary to have nature to

examine the data and making the tree and the rules of it are used. All the above mentioned works used decision tree for dataset to enhance the performance in terms of accuracy. The dataset used in this research work is a balanced dataset.

2.1.6. Swarm Intelligence

The authors Yeh et al. [21], Fei 2010 [22] and Abdi & Giveki [23] used swam intelligence method to generate their model of diagnosis. The algorithm PSO (Particle Swam Optimization) to detect optimal or approximate solutions in huge search spaces efficiently. All the above three authors strive for clearing the problem of optimization which includes with the characteristics of classification problems. If fewer features are used then the classification process becomes faster and more precise. From studies the PSO related approach proved to increase the complete results of classification as PSO is utilized for choosing related parameters in the classifiers which are included.

2.1.7. K-Nearest Neighbor

The k-nearest neighbor is an instance based classier method. The units of parameter comprise of samples which are utilized in this method and its algorithm and then considers that total instances related to points in n -dimensional space RN . This algorithm is very easy because the information in training data will never lose. Even so this paradigm is suitable only if training data set is big. Since this paradigm consumes more time while processing every sample of training set during classification of recent data and this procedure needs a long time for classification.

2.1.8. Bayesian Classifier

The Bayesian classifiers are well known in its computational efficiency and able to control skipped data efficiently and naturally. With this benefit they both (above mentioned authors) got fair accuracy from the developed models. As models are implemented by using Bayesian classifier also proved that the model is suitable because the approach's average rate leads to increase prediction accuracy and helps the authors to retrieve extra features from data with no over fitting. This approach will be good enough if the datasets are having with missed data.

III. PROPOSED WORK

The flow of our work is shown in Figure 1. It starts by data preprocessing such as data cleaning and data preparation and then we applied FP mining to discover the relationships between different appliances. This is used for identifying which appliances are working together. At that point, we used cluster analysis to find the associations between the appliance and time. After these two steps, now the system is able to extract the pattern of appliance and this will be given input to the Bayesian network for

short and long term forecasting. The yield of the framework is used by health care applications relying upon the intended use. For instance, a health care provider may be interested in identifying the activities with psychological debilitation of daily activities is important for reminding the patient when their behavior is not normal. We will briefly explain the theoretical concepts of the techniques used in the next section.

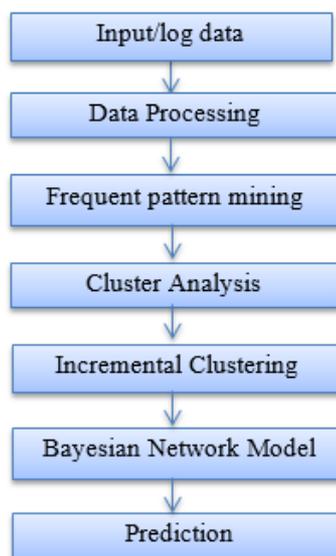


Figure 1: Flow of proposed work

3.1 Data Preprocessing

Data pre-processing is sometimes trivial, but non-trivial in data mining process. If there is irrelevant and noisy information in our dataset, the knowledge discovery process while training the data becomes more difficult. The quality of the input data is affecting the final outcomes. In order to get the accurate results at the end, data is to be preprocessed efficiently and ease in the data mining process that may either we use data cleaning, Integration, transformation and data reduction. The stages are shown in Figure 2. In our work we almost collecting the attributes values, hence the data should not be having any null values and the time we have considered our work is a stipulated one and that yields good results since, there is no inconsistency. We are collecting the patient details and the daily activities including various appliances present in the smart home. The values of these appliance usage are considered to be a binary values. For example if an appliance Television is either ON or OFF. The value '1' is for ON and '0' is for OFF.

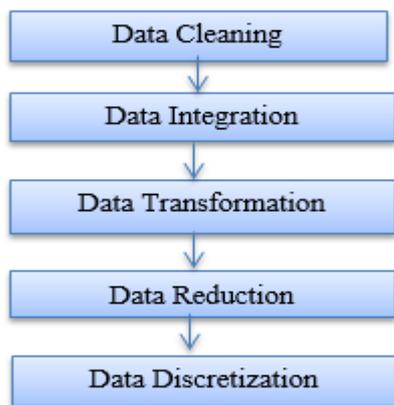


Figure 2: Data Pre-processing stages

3.2 Mining Frequent Patterns from Log Data

The main aim of our work is finding the patterns based on the human activities from the smart meters. There exists several activities for instance “Watching Television”, “Using Laptop”, “Using Micro oven”, “Using Washing Machine” etc. We need to detect the frequent patterns from given activities, which is essential in health care applications that monitors the sudden changes in the behavior of the patient. The doctor can able to see the log data and can give the suggestions/precautions to the patients while observing the frequent patterns/ association rules generated. In the essence, if the patient is taking rest for a long time, by seeing the association rule the doctor can give a suggestion to the patient and very important for the patients with heart diseases. We have set the different timing intervals 30 minutes to 60 minutes, the patient can upload the smart meter data and the frequent patterns can be generated accordingly. The numbers 1, 2, 3 represents the various daily activities. This concept was taken from [24], [25] which uses FP-growth technique using divide and conquer strategy. Let $\{I_1, I_2, \text{ and } I_3 \dots I_n\}$ be item sets containing ‘n’ number of items. I_k is referred as the ‘k’th item in the item set I_n . The association rules are generated in the form of $X \rightarrow Y$ in support and confidence framework. The preprocessing our data set using numeric to binary is shown in Figure 3.

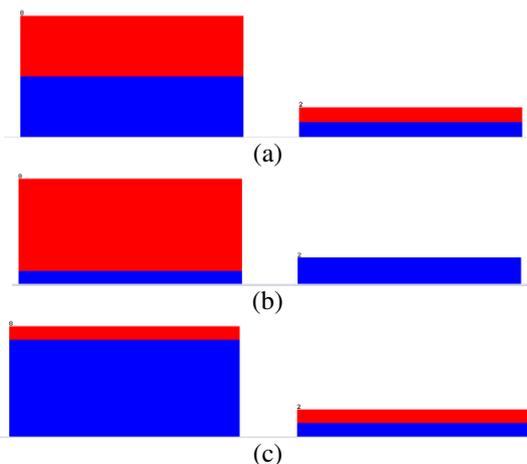


Figure 3: Result of Numeric to Binary Filter for our data set. (Only three attributes are shown in (a), (b), (c)).

3.3 Cluster Analysis

Determining the associations between the time and appliances is crucial to health applications which monitors the patient’s daily activities on timely basis. We used the cluster analyst to determine the usage of appliance with respect to the time. For example whether the patient is using the Television in a specific time in a whole day i.e. (00:00 – 23:59) and this activity can be monitored by the time of the day in three slots i.e. morning, afternoon, evening). The associations between the appliance and time is based on the underlying information captured by the smart meter data. In our work we used to manually upload this time series data instead of using smart meter such as sensor. The recorded information is stored in the data base and we can group as a group of products that the patient is using in a specific time.

For instance, if the patient is using a washing machine and micro oven at the same time in a morning at 07:00 to 08:00 then, these two data items comes under a group and forms a cluster. Whenever a new activity is uploaded, incremental clustering is performed and the clusters are formed accordingly and the corresponding clusters are shown in Figure 4. Finally we integrate the discovered patterns and the associations between time and appliances and will be given to machine learning model called Bayesian network model to predict the activities. Bayesian network is a directed acyclic graph, which contains node and edges. Node indicates random variables and edges denotes probabilistic dependencies. We considered this model because of its main feature “causality”. We have shown a sample Bayesian network model with ‘6’ nodes in Figure 5.

Attribute	Full Data (10.0)	0 (4.0)	1 (6.0)
Name	Subbareddy	Subbareddy	Subbareddy
DATE	4/8/2018	4/8/2018	4/8/2018
SHIFT	Morning	Morning	Morning
NO.OF	1	1	1
<u>MINUTES_binarized</u>			
<u>WM_binarized</u>	0	0	0
<u>BW_binarized</u>	0	0	0
<u>TM_binarized</u>	0	0	0
<u>YM_binarized</u>	0	0	0
<u>MO_binarized</u>	0	0	0
<u>NS_binarized</u>	0	0	0
<u>TV_binarized</u>	0	0	1
<u>LT_binarized</u>	0	0	1
WORK	LOW	LOW	LOW
FOOD	LOW	LOW	OVER
WORKOUT	LOW	LOW	LOW
REST	AVERAGE	AVERAGE	AVERAGE

Figure 4: Summary of Cluster analysis

Figure 4: Summary of Cluster analysis

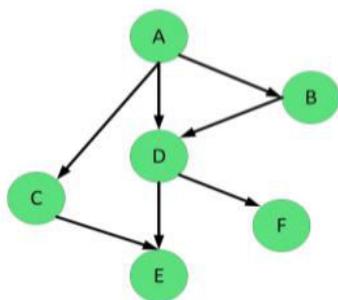


Figure 5: Bayesian network with '6' nodes.

IV. RESULTS & DISCUSSION

Our complete work is summarized as follows. Initially the patients can register their details in the online. We are maintaining a list of doctors who are having specialized in various diseases. As a first step, the patient can send a request to the doctor for an appointment. Once the doctor can accept the request from the patient, the doctor can able to monitor the activities done by the patient at home. Meanwhile, the doctor can send the warning messages or regular messages to the patient to give proper suggestions to the patient by observing some weather conditions of the environment. Periodically, the patient can upload the daily activities from his login. The details can be sent to the doctor and the doctor can give the suggestions to the patient according to the patterns and the association rules

generated from the data given from the log file of the user. In our work we used a model that the patient can upload the data file but we can implement the things by using the sensor to automatically recognize the activities of the patient, who is present in the smart home. We can implement this technology using IoT using a better way. Here we have shown the details of the two patients and their log data, the association rules generated and the results of our cluster analysis. The activities are collected in the span of 60 minutes at three regular intervals in the morning, afternoon and evening. In our work we used the monitoring activities can be uploaded by any user in a specific home, since we are implemented this to give a better option to upload the details by any one member in the family. We can also add more security to capture the confidential information and to avoid the misuse. In our work we used few of the appliance that our smart sensor can able to capture the information. The status will be either ON or OFF. If the patient is using a specific appliance then the status will be ON, otherwise it is OFF. Figure 6 shows the details of two patient's sensor data i.e. regular activities of the patients at different time intervals. The activities of the patient at each time span is shown in Figure 7 and sum of the working time of the patient is shown in Figure 8 and status is shown in Figure 9.

Name	DATE	SHIFT	WM	BW	TM	YM	MO	NS	TV	LT
subbareddy	4/8/2018	morning	1	1	0	0	1	1	1	1
subbareddy	4/8/2018	afternoon	0	0	0	0	0	0	0	0
subbareddy	4/8/2018	evening	0	0	0	0	0	0	1	0
subbareddy	4/8/2018	morning	0	0	0	1	0	0	0	0
subbareddy	4/8/2018	morning	0	0	0	0	1	0	0	0
subbareddy	4/8/2018	afternoon	0	0	0	0	0	1	0	1
subbareddy	4/8/2018	evening	0	0	0	0	0	0	1	0
subbareddy	4/8/2018	evening	0	0	0	0	1	0	0	1
sudhakar	4/8/2018	morning	1	1	1	0	0	0	0	1
sudhakar	4/8/2018	afternoon	1	0	1	1	1	1	1	1

Figure 6: Patient Daily Activities at three different timing intervals.

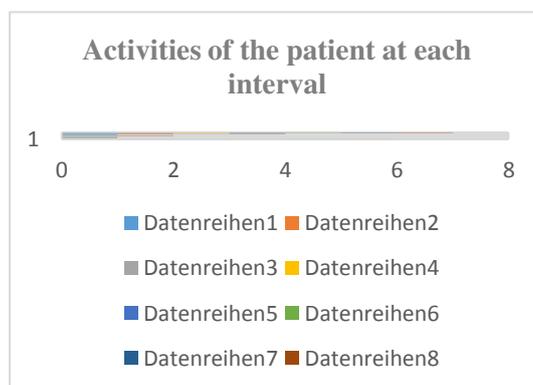


Figure 7: Activities of the patient at each time span.

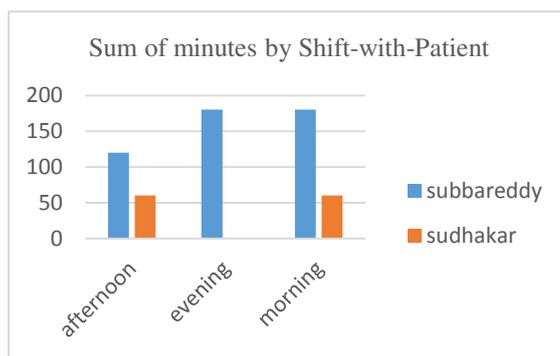


Figure 8: Sum of the working time by shift.

WORK	FOOD	WORKOUT	REST
OVER	OVER	LOW	OVER
LOW	LOW	LOW	LOW
LOW	LOW	LOW	AVERAGE
LOW	LOW	AVERAGE	LOW
LOW	AVERAGE	LOW	LOW
LOW	AVERAGE	LOW	AVERAGE
LOW	LOW	LOW	AVERAGE
LOW	AVERAGE	LOW	AVERAGE
OVER	LOW	AVERAGE	AVERAGE
AVERAGE	OVER	OVER	OVER

Figure 9: Status of the patient for various activities.

V. CONCLUSION

Health care services are one of the most challenging aspect that is must require to the people with abnormal health. Data mining play a prominent role towards healthcare industries, particularly in prediction of different kinds of diseases. While predicting diseases the medical diagnosing is extensively used. As conclusion there should not be any data mining method for resolving the risks in healthcare data sets. We have to generate a hybrid model which supports for resolving the issues that are mentioned. At first, we used to record the daily activities of the patient using a specific time period at three regular intervals. Subsequently, we applied the frequent pattern growth for extracting the association rules from the log file. Finally, we applied k-means clustering for the input followed by Bayesian network model to predict the health behavior of the patient and precautions will be given accordingly. This helps in obtaining higher accuracy among classifiers which is most significant in medical diagnosis by taking features of data with care. To make enhancement in predictions along with real time sensor meter data by using hybrid models is considered as our further work.

REFERENCES

[1] H. Zheng, H. Wang, and N. Black, "Human activity detection in smart home environment with self-adaptive neural networks," in Proc. of IEEE International Conference on Networking, Sensing and Control, pp. 1505-1510, 2008.

[2] E.M. Tapia, S.S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous

sensors," Proceeding of Pervasive, vol. LNCS 3001, Springer Berlin Heidelberg, pp.158–175, 2005.

[3] Rashidi, P.; Cook, D.J. Transferring Learned Activities in Smart Environments, In Proceedings of the International Conference on Intelligent Environments, Barcelona, Spain, 20–21 July 2009.

[4] Jakkula, V.R.; Cook, D.J. Using Temporal Relations in Smart Environment Data for Activity Prediction. In Proceedings of the International Conference on Machine Learning, Corvallis, OR, USA, 20–24 June 2007.

[5] Riabov, A.; Liu, Z.; Wolf, L.; Yu, S.; Zhang, L. Clustering Algorithms for Content-Based Publication-Subscription Systems. In Proceedings of the 22nd International Conference on Distributed Computing Systems, Vienna, Austria, 2–5 July 2002; pp. 133–142.

[6] Li, C.; Biswas, G. Unsupervised Learning with Mixed Numeric and Nominal Data. IEEE Trans. Knowl. Data Eng. 2002, 4, 673–690.

[7] Alam, M.R.; Reaz, M.B.I.; Mohd Ali, M.A. SPEED: An Inhabitant Activity Prediction Algorithm for Smart Homes. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 2011, 4, 985–990.

[8] Liao, L.; Fox, D.; Kautz, H. Location-Based Activity Recognition Using Relation Markov Network. In Proceedings of the International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, 30 July–5 August 2005; pp. 773–778.

[9] Miskelly, F.G. Assistive Technology in Elderly Care. Age Ageing 2001, 6, 455–458.

[10]. Building Classification Models: ID3 and C4.5. Available

[11]. Weka 8: Data Mining Software in Java.

[12] W. Haizhou and S. Mingzhou, "Ckmeans.1d.dp: Optimal k-means clustering in one dimension by dynamic programming,"RJ. vol.3,no.2, pp. 29–33, 2011.

[13] J.-J. Yang, J. Li, J. Mulder, Y. Wang, S. Chen, H. Wu, Q. Wang, and H. Pan, "Emerging information technologies for enhanced healthcare," *Comput. Ind.*, vol. 69, pp. 3–11, 2015.

[14] N. Wickramasinghe, S. K. Sharma, and J. N. D. Gupta, "Knowledge Management in Healthcare," vol. 63, pp. 5–18, 2005.

[15] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Mag.*, pp. 37–54, 1996.

[16] B. Liu, Y. Xiao, L. Cao, Z. Hao, and F. Deng, "SVDD-based outlier detection on uncertain data," *Knowl. Inf. Syst.*, vol. 34, no. 3, pp. 597–618, 2013.

[17] R. Veloso, F. Portela, M. F. Santos, Á. Silva, F. Rua, A. Abelha, and J. Machado, "A Clustering Approach for Predicting Readmissions in Intensive Medicine," *Procedia Technol.*, vol. 16, pp. 1307–1316, 2014.

[18] N. Sharma and H. Om, "Data mining models for predicting oral cancer survivability," *Netw. Model. Anal. Heal. Informatics Bioinforma.*, vol. 2, no. 4, pp. 285–295, 2013.

[19] K.-J. Wang, B. Makond, and K.-M. Wang, "An improved survivability prognosis of breast cancer by using sampling and feature selection technique to solve imbalanced patient classification data.," *BMC Med. Inform. Decis. Mak.*, vol. 13, p. 124, 2013.

[20] H. M. Zolbanin, D. Delen, and A. Hassan Zadeh, "Predicting overall survivability in comorbidity of cancers: A data mining approach," *Decis. Support Syst.*, vol. 74, pp. 150–161, 2015.

[21] W.-C. Yeh, W.-W. Chang, and Y. Y. Chung, "A new hybrid approach for mining breast cancer pattern using discrete particle swarm optimization and statistical method," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 8204–8211, 2009.

[22] S. W. Fei, "Diagnostic study on arrhythmia cordis based on particle swarm optimization-based support vector machine," *Expert Syst. Appl.*, vol. 37, no. 10, pp. 6748–6752, 2010.

[23] M. J. Abdi and D. Giveki, "Automatic detection of erythematous-squamous diseases using PSO-SVM based on association rules," *Eng. Appl. Artif. Intell.*, vol. 26, no. 1, pp. 603–608, 2013

[24] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, Dallas, TX, USA, 2000, pp. 1–12

[25] J. Han, J. Pei, Y. Yin, and R. Mao, "Mining frequent patterns without candidate generation: A frequent-pattern

tree approach," *Data Mining Knowl. Discovery*, vol. 8, no. 1, pp. 53–87, 2004.

AUTHOR'S BIOGRAPHY



Dr. G. RamaSubba Reddy received his Ph.D in Computer science & Engineering from Sunrise University. He received his M.E in Computer Science & Engineering from Sathyabama University, Chennai. Presently he is working as an Associate Professor and Head of the Department in Computer Science & Engineering, Mother Theresa Institute of Engineering & Technology. Palamaner, Andhra Pradesh, INDIA. His current research focus is on Wireless sensor networks and Cloud computing.



Mr. B. Dilip Chakravarthy received his M.Tech in Computer Science & Engineering from JNTUA Ananthapuramu. Presently he is working as an Assistant Professor, Department in Computer Science & Engineering, Mother Theresa Institute of Engineering & Technology. Palamaner, Andhra Pradesh, INDIA. His current research focus is on Data Mining and artificial Intelligence.



Mrs. C. Reddi Neelima her M.Tech in Computer Science & Engineering from JNTUA Ananthapuramu. Presently he is working as an Assistant Professor, Department in Computer Science & Engineering, Mother Theresa Institute of Engineering & Technology. Palamaner, Andhra Pradesh, INDIA. Her current research focus is on Cloud Computing and Data Mining.



Mrs. G. Divya Zion, a PhD Research Scholar at school of Computer science engineering, VIT university Vellore. Currently working as Assistant Professor in the Dept. of CSE with 5 years of experience. She Obtained Masters Degree at G. Pulla Reddy Engineering college (Autonomous) in the year 2012, and Bachelors Degree at St. Johns College of Engineering and Technology in the year 2010.