International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 3863–3871

Human Activity Patterns Prediction (HAPP) System for Smart Healthcare Applications

M.Rajalakshmi[1], Ankana Roy[2], Chirag Kumar[3] , Sahil Ansari[4],Survi Agarwal[5]

[1]*Assistant Professor, Department of Information Technology, SRMIST, Chennai, India*
[2,3,4,5]*UG Student, Department of Information Technology, SRMIST, Chennai, India*

Abstract

In recent days, individuals are relocated in urban rapidly from rural regions. Health care facility is the the most complex sector which is enormously influenced with the huge flood of individuals. Consequently, urban region society are taking enormous efforts for technological development for providing healthier environment society to individuals. This venture proposes a model Human Activity Patterns Prediction (HAPP) System for Smart Healthcare Application that uses a huge data set of smart home for knowledge gathering and finding the patterns of human behaviours for healthcare services provider. The proposed idea uses the various data mining techniques like Association rules, clustering, and prediction to quantify, examine the power consumptions variation that happens due to occupant's behaviour. HAPP system points out the requirement to examine the patterns of power used by the home appliances, since it is independent of human activity. The information collected from smart meters was processed and using the data mining techniques, the outcome were analysed and monitored.

Keywords—component, formatting, style, styling, insert

I INTRODUCTION

The survey demonstrates that, 60% of the total populace may relocate to urban region by the year 2040 [1]. This extraordinary statistical variation made current methodologies of health care services to remodel for providing the healthier society for the huge population. In the aspect of requirements and the problems of huge population, many houses were outfitted for using the electronic gadgets which produces gigantic volumes of indexical information from the fine – grained analysis which are examined to maintain the health care facilities. Progression in the technology of big data mining [9] provides methods for handling huge amount of data for significant bits of knowledge, can help us in seeing how individuals approach their life. Since individuals' habits are generally analysed by their daily activities, anomalous study reveals the individuals problem in decision making. The correlation between the usages of appliance in smart home with the regular activities detects the prospective health problems.

II EXISTING SYSTEM

The existing system uses multiple techniques to examine and detect the regular activities of individual. This increase the complexity of the algorithms used to solve the problem. Another major problem is that it includes only appliances which have been used for longer duration. Therefore, it misses out important appliances which are used for shorter duration but has a major impact on the results. The high cost due to use of sensors is one of the key obstacles in the implementation. When lesser number of appliances is taken into account for shorter duration, it decreases the accuracy of Rate of Identification of major health problems that can occur. The existing system is expensive [2] because of the use of sensors. The system causes lower accuracy of Rate of Identification because of less number of appliances taken into account. The current system is yet to predict the health issues accurately. It requires more input data and biological algorithms to get best and accurate results.

III PROPOSED METHODOLOGY

The proposed system will help in predicting better healthcare scenarios by taking into account more appliances used for shorter duration. The more the data, the better is the accuracy of the results. It also increases the time efficiency by using efficient algorithms. Lesser number of complex

ISSN: 2233-7857 IJFGCN

algorithms saves calculation time and gives results faster. The new system reduces the complexity by using fewer approaches. It lowers the cost of the system, making it easier to implement the model.

A Dataset

An artificial dataset has been used for this proposed work. The prototype for the appliances, used in the house, is monitored for 48 hours continuously. The recorded data is then converted to a CSV file. The dataset is recorded into columns as Date, Starting Time, Ending Time and the 15 prototype appliances. Each row indicates the recorded dataset for one hour. The requirement is to identify the human activity pattern [3] in the information gathered from smart meters [5]. For example, using electronic cooker, Mixer, Washing machine, air conditioner, using personal computer, charging mobile and laptop are the general ordinary schedules.

B System Architecture

The Figure.1 shows the system architecture which depicts the actual processing of data through different steps. First, the microcontroller connected to the home appliances. It provides the power consumption readings in UART format and the collected data is converted into CSV file. The recorded data is divided into 75% training and 25% testing set which is then processed using time, power and user analysis. It is an unsupervised learning method. The clusters are formed using K-Means Algorithm with the help of the Elbow Method. The Naïve Bayes is used to determine frequent pattern in energy consumption. And finally the result is predicted using FP Growth and Decision Tree Regressor Algorithm [4]. The results determine whether a specific machine is the explanation behind medical issues. The microcontroller is used for tasks like collecting the information i.e. power consumption data and converting it into csv format. After that the csv file is used for clustering and predicting data results and graph plotting using Bayesian Network.

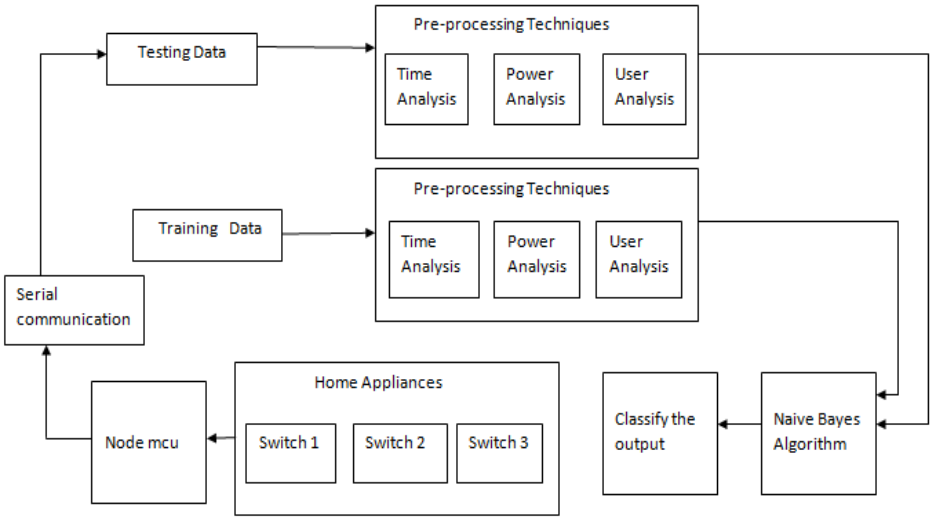


Figure 1. The Proposed architecture

C. Hardware Setup using Arduino

The Arduino Uno WiFi is an Arduino Uno with a coordinated Wi-Fi module (The ESP8266 WiFi).

In the proposed work, 10 different switches are connected to 10 different pins of Arduino UNO as shown in Figure. 2. When we put on the switch, based on a daily schedule of a regular household, the readings are calculated and displayed to user by the port 3.

ISSN: 2233-7857 IJFGCN

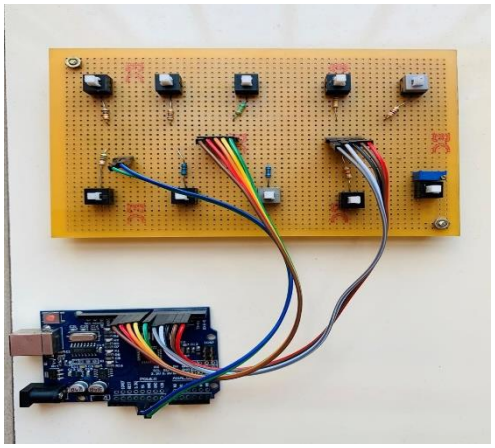


Figure. 2 System of Hardware Setup

IV DEEP LEARNING ALGORITHMS

Deep learning (DL) has turned into a quickly developing exploration field, rethinking best in class exhibitions in a wide scope of areas, for example, object recognition, speech recognition, image segmentation and machine interpretation. In present day producing frameworks, information driven machine health checking is picking up in growth because of the wide spread of cheap sensors and their functionalities with the Internet.

Types of Deep Learning Algorithms

- Supervised learning
- Unsupervised learning
- Reinforcement learning

A. K-Means Clustering

It is an unsupervised algorithm which is applied to survey data and considers perceptions of the information as items are dependent on areas and distance between different items. Objects are partitioned into commonly limited clusters (K) in such a way that objects inside each cluster continue as adjacent to one another. Each group is considered by its Centroid i.e., its middle point. The partings utilized in clusters in the phases don't generally highlight the spatial separations. In wide-running, the main goal for this issue of worldwide is a decision of beginning points.

In a dataset, the proper number of clusters K and a set of k starting point, the K-Means unsupervised algorithm finds the estimated number of particular clusters and their centroids. Centroid is the point where coordinates are gained by means for figuring every coordinates of the point, models allocated to the clusters.

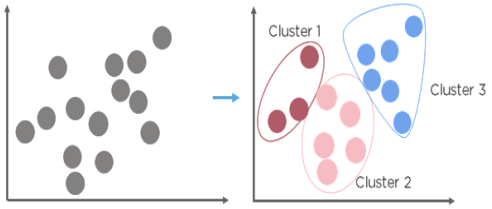


Figure 3. K-Means Clustering

Elbow Method:

The Elbow technique is a strategy for elucidation and validation of consistency inside clusters to help find the proper number of clusters in a dataset. The purpose of the elbow technique is to run clusters on the dataset for the value of k (state, k from 1 to 10 in the instances above), and for each estimation of k calculate the sum of squared errors (SSE).In this study, after applying the elbow

ISSN: 2233-7857 IJFGCN

method we concluded from the figure 1.6 that the optimum number of clusters is 3. Hence we further classified the data based on the above results.

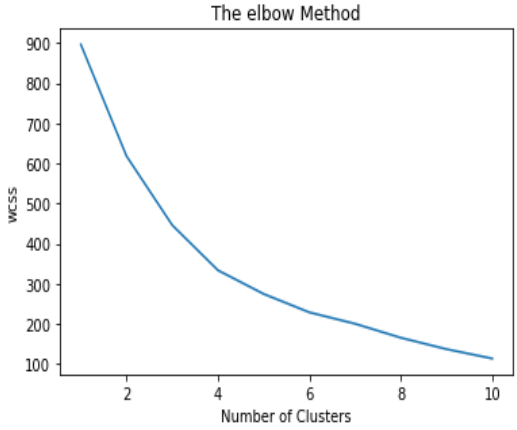
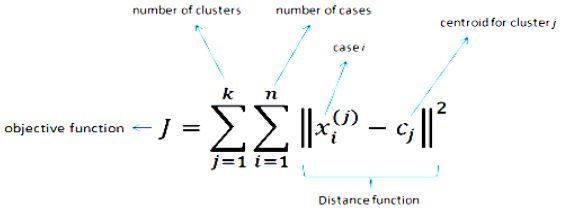


Figure. 4: Elbow Method

Three different colours (red, black, green) has been used to differentiate each cluster and using matplotlib library the graph is plotted as shown in Figure. 5. Clusters are formed based on the level of power consumption. Appliances consuming High, Medium and Low power is each clustered separately. Equation used for plotting clusters is given below.



$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$$

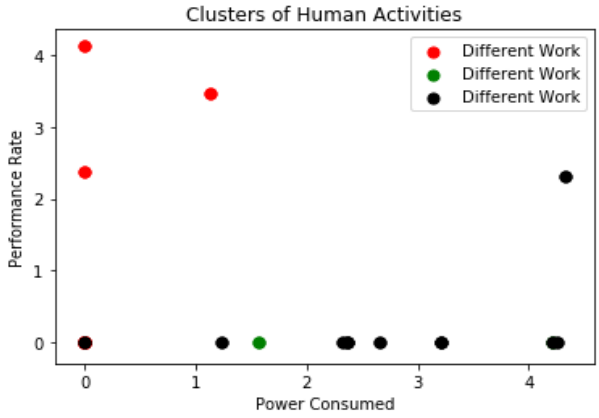


Figure. 5: Graph showing the clusters formed

B. Bayesian Network

A Bayesian system is a coordinated non-cyclic diagram in which each edge compares to a contingent reliance and every nodes relates to a one of a kind irregular variable. Bayesian systems are usable in fields where there is requirement for predicting and result is questionable. Rather than simply 'speculating', Bayesian network help its clients make smart, quantifiable and reasonable choices.

They are relevant in diagnosing, overseeing budgetary hazard and protection, displaying environments. They additionally help in checking and alarming, portfolio designation and expectation. For instance, in stocks, climate anticipating and sports wagering, and sensor combination. Since Bayesian classification system includes all the features for providing the better learning rate to report the problem along with understanding problem domain and predicting the consequences. To improve the bayesian network optimization , probabilistic and normal semantics were used to avoid overfitting. It predicts result of an intervention before intervening.

ISSN: 2233-7857 IJFGCN

The frequent patterns in user's behaviour and device to-time relationship helps in finding out about the utilization of various devices and building the structure for usage patterns. The component uses Bayesian system which is a coordinated non-cyclic chart, where nodes refers to irregular factors and the edges show probabilistic conditions. A Bayesian network is defined by the probabilistic distribution as

np(x1,x2,...,xn)= Y p(xi|parents(xi))

As referenced above, that integration of devices probability with respect to time in hours develop a probabilistic prediction model. The architecture of Bayesian network [6]has one input node of unconditional probabilities forwarded to a output node. The posterior probability for the prediction model is defined as

p (.)= p (Hour) × p (Time of day) × p (Weekday) ×p (Week) ×p (Month) ×p (Season)

$$P(X_1,\ldots,X_n)=\prod_{i=1}^{n}P(X_i \mid X_1,\ldots,X_{i-1})=\prod_{i=1}^{n}P(X_i \mid Parents(X_i))$$

Finally using the results from above Bayesian network two graphs are plotted. Both the graphs are discussed in result section.

C. FP-Growth Pattern

In Data mining, the era of examining the successive sample in a huge database is costly and the FP Growth algorithm overcomes by all-inclusive and prefix tree arrangement which neglects the compacted, successive data. FP development has been utilized in this work as a result of the greatest preferred standpoint found in FP-Growth is the way that the calculation just needs to peruse the document twice, as opposed to apriori who understands it once for each emphasis. Another gigantic favourable position is that it evacuates the need to ascertain the sets to be checked, which is very preparing overwhelming, in light of the fact that it utilizes the FP-Tree. This makes it O(n) which is a lot quicker than apriori. The FP-Growth calculation stores in memory a conservative form of the database.

HEAT MAP

Seaborn Python , matplotlib dependent library is used to represent the statistical information. Since seaborn python is based over Matplotlib, the designs can be additionally changed utilizing Matplotlib instruments and rendered with any of the Matplotlib backend to create distribution quality figures. A heatmap is a two-dimensional graphical portrayal of information where the individual qualities that are contained in a framework are spoken to as hues. The seaborn python library permits the formation of commented on heat maps which can be changed utilizing Matplotlib according to the user's prerequisite. The heatmap is utilized for visual portrayal of true positives and true negatives of every machine. This will display the level of how much does each machine's capacity is truly influencing the outcomes and which are really negative. The heat map is shown in (Figure. 6).

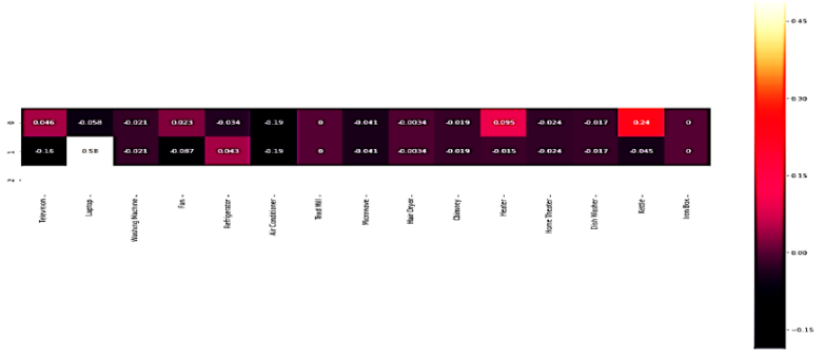


Figure. 6 Heat Map of appliances

ISSN: 2233-7857 IJFGCN

Sl. No.	Test Case Name	Test Case Description	Test Steps			Test Status (P/F)
			Steps	Expected Result	Actual Result	
1	Collect Input	To check what happens when the switches are connected to the power supply	Arduino Board with switches connected to power supply using USB	Switches when in the On/Off state should record data accordingly	48 hours data of a house is recorded and converted to a CSV file.	P
2	K-Means Clustering	To check effective number of clusters are formed	Using Elbow Method find the number of clusters; Plot the clusters formed	Effective number of clusters with visible differentiation b/w clusters	3 clusters formed with clear distinction b/w each cluster	P
3	Plotting Graphs	Graphs are plotted	Using Bayesian Network plot time slice and appliance graph	Graphs showing visible differentiation	Graphs showing the daily activities distinctively	P
4	Prediction Score	Accuracy of the results	Decision Tree Regressor Algorithm comparing the training and testing data	Accuracy for each comparison should be displayed	Each appliance shows their accuracy % of the rate of the prediction	P
5	Identify appliances	Appliances causing health issues identified	Threshold Value fond using trial and error; Using the values identify appliances causing health issues	Identify and display appliances causing health issues and by what amount	Appliances identified and displayed. Graph plotted showing the percentage of contribution of the appliance in the health issues.	P

Table. 1 Test Case Table

DECISION TREE REGRESSOR

Decision tree regressor is used to build relapse or grouping the subsets of the dataset. This helps to neglect the noisy data.

V EVALUATION AND RESULTS

Each of the units was tested as shown in Table. 1, against the K-Means Clustering [7], Bayesian Network and FP-Growth Pattern. Arduino board collected inputs from the switches and recorded the data correctly. K-Means Algorithm clustered the recorded data into 3 clusters using Elbow Method efficiently. Bayesian Network was the tested which gave accurate graph results. Both the hourly and appliance graph were found to be giving accurate results. Decision Tree Algorithm [8] was tested. It displayed prediction score [10] for each appliance successfully.

ISSN: 2233-7857 IJFGCN

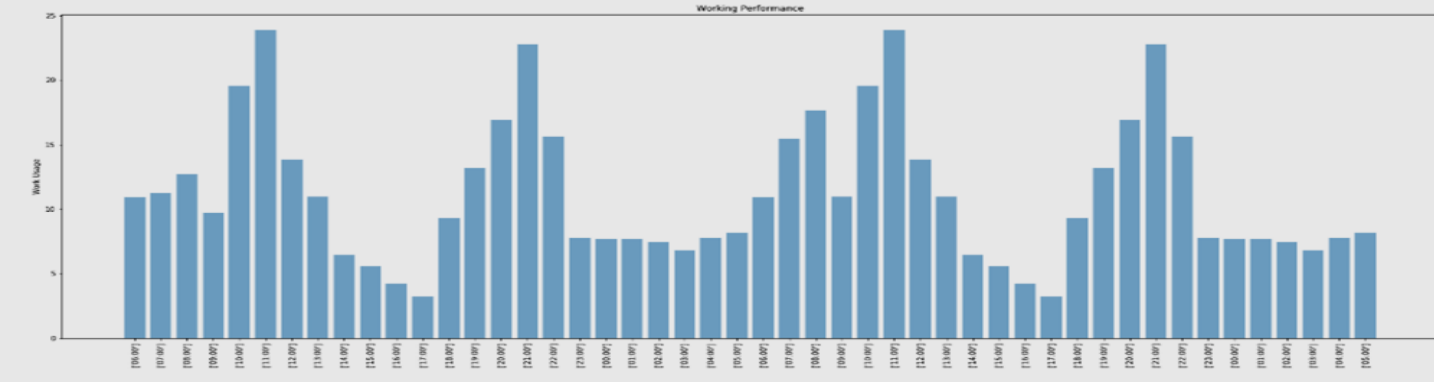


Figure. 7: Power Consumption vs. Hour Graph

As seen from the figure. 7, there are similarities and minute differences in power consumptions patterns of two days. Hence this helps us in determining if there is any noticeable change in power consumption hour wise then there might be problems with user's health. The second graph is being plotted using power consumption against individual.

The second graph as shown in Figure. 8 helps us in determining if there is underuse or overuse of any device which may lead to health problems or which might be result of health problems.

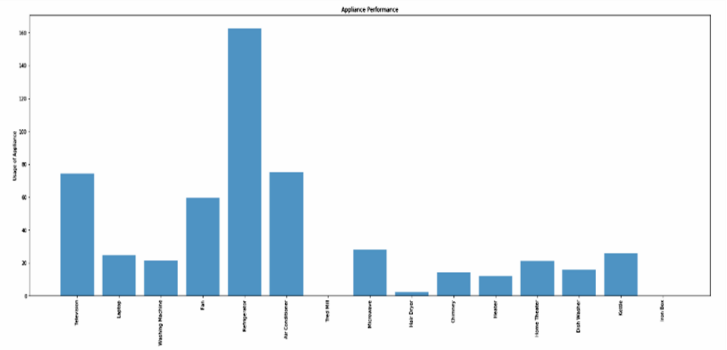


Figure. 8: Power Consumption vs. Appliance Graph

PREDICTION SCORE

After a few hit and preliminary runs, the most steady and precise outcomes were anticipated when limit control rate were kept between 0 to 20%. The anticipated score gave less fault rate and the outcomes were all the more persuading. Every expectation score as shown in Figure. 9 were determined utilizing Decision Tree technique to anticipate whether a specific machine is the explanation behind medical issues.

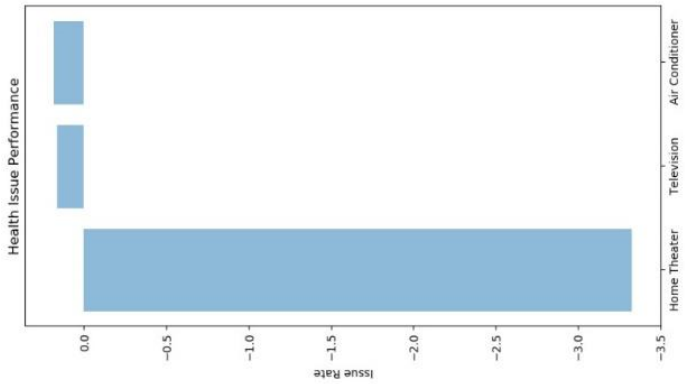


Figure. 9: Prediction Score Graph

ISSN: 2233-7857 IJFGCN

VI CONCLUSION AND FUTURE WORK

This work is intended to give the pattern of human activity from the dataset of smart meter placed in houses.

The pattern tells utilization of applications and restricted condition faced by the individual from the information gathered from the entire device with the relation to time. The results of this system showed great results. By using unsupervised learning, it could predict whether the appliances used are one of the reasons causing the health issues concerning the residents. Simpler algorithms like K-means and Bayesian Network helped us easily find out the pattern in the daily schedule of the residents in the house. Any noticeable change in the pattern or any deviation from the usage of appliances according to the day and season is taken into account. Using Decision Tree Regressor Technique, the accuracy of the test results were found compared to the training data. With the trial and error method, the threshold value was found to be 20%. Appliances with the prediction score below the threshold value is found to be causing medical issues. The result from the system shows great possibilities to help the health care services. It will help in providing better facilities and predict the cause behind the health issues concerned.

For future work, the square measure progressing can be used to redefine the prediction and network model for handling the huge data with various sensored appliances in an exceedingly close to period trend. This may encourage health applications to speedily take activities like making conscious patients or care providers. In he same manner, metaphysics model for health service can be précised to monitor the activities. This helps to increase the training dataset to predict more accurate results and help us find more patterns and exceptions.

There is another scope in future enhancement where the current system will predict the appliances causing harm to the individuals. Biological Symptoms can be added to the system where it predicts the possible health conditions the individuals in the house are suffering from. The residents can also be warned and made aware about the significance of living a healthy lifestyle and reducing the global carbon footprint.

REFERENCES

[1] Yi-Cheng Chen , Hsiu-Chieh Hung , Bing-Yang Chiang , Sheng - Yuan Peng , Peng-JunChen, "*Incrementally Mining Usage Correlations among Appliances in Smart Homes*", IEEE, 2015.

[2] Shamim Hossain, "*Cloud-Supported Cyber-Physical Localization Framework for Patients Monitoring*", IEEE, 2015.

[3] M. Shamim Hossain, Ghulam Muhammad, Wadood Abdul, Biao Song, B.B. Gupta, "*Cloud-assisted secure video transmission and sharing framework for smart cities*", IEEE/ACM, 2017.

[4] Jing Liao, Lina Stankovic and Vladimir Stankovic, *Detecting Household Activity Patterns from Smart Meter Data*", IEEE, 2014.

[5] A bdulsalam Yassine , Ali Asghar Nazari Shirehjini, *"Smart Meters Big Data: Game Theoretic Model for Fair Data Sharing in Deregulated Smart Grids"*,IEEE, 2015.

[6] S.Bakiyalakshmi, K.Malini, H.Fametha, *"Human Activity Prediction for Health care Application Using Smart Meter"*, Technical Research Organization India, 2018 .

[7] G.Nivetha,J.Pooja, P.Senthil , *"Mining Human Activity Patterns from Smart Home for Healthcare Applications"*, NCIRCST, 2018.

[8] Abdulsalam Yassine, Shailendra Singh, Atif Alamri, *"Mining Human Activity Patterns from Smart Home Big Data for Healthcare Applications",* IEEE, 2017.

[9] Sreelekshmi, Gopu Darsan, *"Mining Human Activity Patterns from Smart Home Big Data for Health care applications",* IRJET, 2017.

[10] Dr. Anjali Mathur, V Vaishnavi, K Jigeesha, KSVAG Sudheer, *"A framework using big data analysis on human activity patterns for Health prediction"*, IJMET, 2017.

ISSN: 2233-7857 IJFGCN