Evaluation of Data Mining Algorithms for Predicting Behaviour and Visualizing Daily Activities of the Elderly People living in Smart Homes

Igene Oseghale Osezua
Dept. of Computer Science, University of Benin
P.M.B. 1154, Ugbowo, Benin City, Nigeria
ose.igene@uniben.edu
08122568075

Ehud Reiter
Department of Computer Science,
University of Aberdeen, King's College,
Aberdeen, AB24 3UE, Scotland, United Kingdom
e.reiter@abdn.ac.uk

Abstract

Over the years, there has been an increasing number of the elderly population in major countries of the world. This has led to the development of the Smart House technology that provides alternative means of effective health care. This automated environment is created by the deployment of Information and Communications Technology (ICT) which supports real-time monitoring of their daily activities as well as the implementation of Knowledge and Discovery in Databases (KDD) processes which enables researchers to study, analyze and predict human behaviour. In this study, we presented the development of the Smart Home and how this technology has helped to improve the well-being of the elderly people as well as the implementation of data analytical processes such as data mining and data visualization techniques used on datasets for discovering patterns and extracting non-trivial information that enables researchers to accurately predict human behaviour of the elderly people. We also highlighted current research works undertaken by various academic research groups where their framework and methodologies are highlighted. The data mining algorithms implemented in their works are also critically examined where their strengths, challenges and notable drawbacks are discussed to determine which of the models that has the potential to produce the highest prediction accuracy.

Key Words: Data Mining, Smart Home, Data Visualization, Human Behaviour, Daily Activities

1.0 Introduction and Background

There is a dramatic increase in the number of the elderly population and their life span especially in western countries like Britain, USA and most European countries which have greatly incurred heavy cost of having to support old people in care homes and hospitals [1]. According to a related article, it is believed that the striking increase in the number of elderly people have led to many major countries experiencing a very unstable situation where the amount of resources and services will not be able to accommodate and cater for such people [2]. Furthermore, it was forecasted according to the same paper that "*there were about 12 million people in the UK over 60 years of age, of whom 752,000 were aged between 85 and 89 and 376,000 exceeded 90 years of age*" [2]. In addition, Hine et al estimated that the number of people over 65s will reach about 12 million by 2011 and 13 million by 2026 [3]. The Health care system has also faced critical challenges due to issues with government approval processes, reimbursement policies and also the nonexistence of efficient deployment technologies [4]. To further support these facts, it is believed that by 2050, 28 percent of the European population would have risen over the age of 65 years, for example in countries like Sweden, where the share over 65 which was estimated at 17 percent as of 2005 is expected to increase to about 24 percent or more by the year 2050 [5] which could pose a serious threat to the social welfare systems for such Scandinavian countries in terms of providing social amenities most importantly health care for the elderly people where the larger share of elderly care is undertaken by their families.

These problems are attributed to the reality that as people age, their physical and mental faculties will begin to deteriorate and as such these effects will lead to increased risk of diseases and accidents in their homes and these problems have led to the increased necessity to be provide formal (health services) and informal health care through family members [2]. These factors have contributed to the critical need to support and provide support for the elderly population through the development of a "**Smart Home Environment**". Smart environments have provided the means of prolonging and improving the elderly people's lives through the converging of pervasive computing and machine learning technologies to provide health monitoring and response to alerts [6]. This technology will enable health carers and other medical practitioners to closely monitor the health status of the elderly and disabled people especially those that may be diagnosed with chronic diseases while also giving such patients the benefit of not having to pay extra expenses of receiving health care in the hospital or nursing homes. One of the major issues is detecting and uncovering patterns that will indicate the state of person's health status. Data mining has become a very important data analysis tool that has been extensively implemented in smart homes with various research groups using different data mining algorithms for analyzing datasets. These groups have used some of the data mining models in their experiments to be able make accurate predictions on the activities, movements and overall lifestyle of the elderly people. However these models (algorithms) implemented despite their advantages also have their respective drawbacks especially when it involves being able to effectively predict human behaviour. This study will explore the technological components that constitute the smart home as well as critically examining the data mining models implemented by various academic research groups and to determine which of the models that can potentially produce high prediction accuracy for future purposes.

2.0 Smart Home Technology and Ambient Assisted Living (AAL)

2.1. Smart Homes

A Smart home environment can be described as *"a physical world that is richly and invisibly interwoven with sensors, actuators, displays, and computational elements, embedded seamlessly in the everyday objects of our lives, and connected through a continuous network"* [7]. Smart Homes offer an ideal, safe and secure environment for the benefit of the elderly and disabled people through various functionalities which basically includes real-time monitoring of their activities, automating specific tasks, adapting to the changes in their daily living through prompts, detecting anomalies and alerting health carers for assistance [8]. Bierhoff et al stated that the number of real smart homes with the network and other additional applications are only limited to demonstration homes and that numerous homes worldwide have implemented home automation. However they added that this development is in its infancy stage due to socio-cultural factors, economic, costs, integration difficulties and lack of standardization which have contributed to the slow progress of smart home technologies [9].

2.2. Ambient Assisted Living (AAL)

An Ambient Assisted Living (AAL) environment can described as a consolidation of assistive technologies and services like telehealth and community alarms that support and monitor the elderly and disabled people using multi sensor technology [10] and undertakes daily activities, health status in addition to ensuring safety and security of the persons [11]. According to a research paper, "*AAL aims to prolong the time people can live in decent way in their own home by increasing their autonomy and self-confidence, the discharge of activities of daily living, to monitor and care for the elderly or ill person, to enhance the security and to save resources*" [12]. The European Commission which comprised of twenty (20) European Member States and three Associated States commissioned and implemented the AAL Joint Research and Development Funding Programme as a way to be able to proffer solutions to the crisis that the elderly population are facing in Europe due to health and age factors, limited health care resources and the fact that many of these people want to live in their homes instead of being treated in hospitals. This innovation funding programme was designed to be able to make use of intelligent products and providing remote services to the elderly population thereby promoting and extending their life span and healthy living through monitoring and assistance in their daily activities while also reducing innovation barriers of promising markets and reducing the social security costs as low as possible [13].

3.0 Research Design and Methodology

Current research groups that have designed and developed a smart home environment have made use of a set of components and technical terms that describes the architecture of a smart home or living laboratories, how the elderly person's lifestyle is measured and how the data generated are analyzed and visualized for discovery of patterns of human behaviour. They include the following:

3.1. Wireless Sensor Networks: '*A wireless sensor is a sensing module which houses one or many transceiver nodes and base stations and also uses a wide range of RF communications techniques'* [14]. A Wireless sensor network also comprises of communication and computing constituents that gives health professionals the power to observe human behaviour, and to respond to certain changes and events [4]. Sensor networks allow the monitoring of movements, locations, environmental parameters and physiological conditions of people and thus assisting in their well-being [15]. The data collected by the sensors are then relayed to a central station point where the data is stored in a database either online or offline and these datasets are then analyzed to detect change in behavioural patterns through various techniques such as data mining and data visualization [2]. A Sensor Network basically consists of four major components which includes a collection of sensors that are either localized or distributed, a wireless-based network, a central point or base station for the information gathering and clustering and a set of computing facilities to handle processes such as data mining and visualization, trend analysis and status querying [4]. A typical smart home scenario proposed by Raad and Yang in Figure 1 and Figure 2 details the various sensors implemented that not only detects the movement of individuals, captures and stores data in the database for analysis, it also triggers an alarm to assigned carers in case of an emergency that may involve the individual [16].

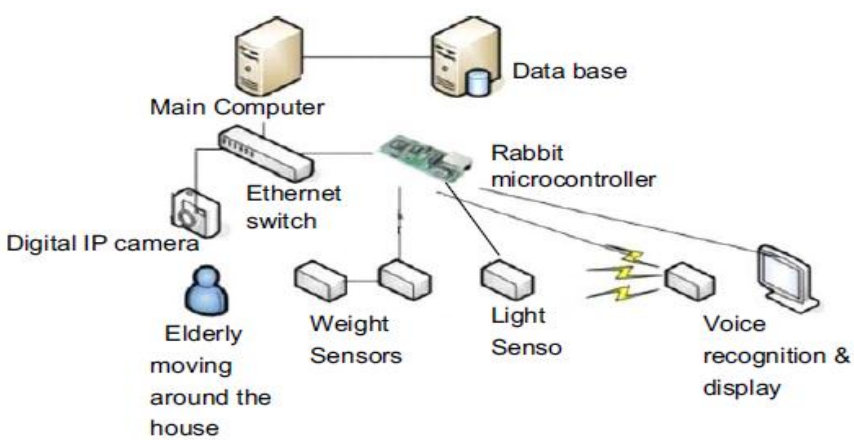

Fig 1: Overall smart home sensor layout.

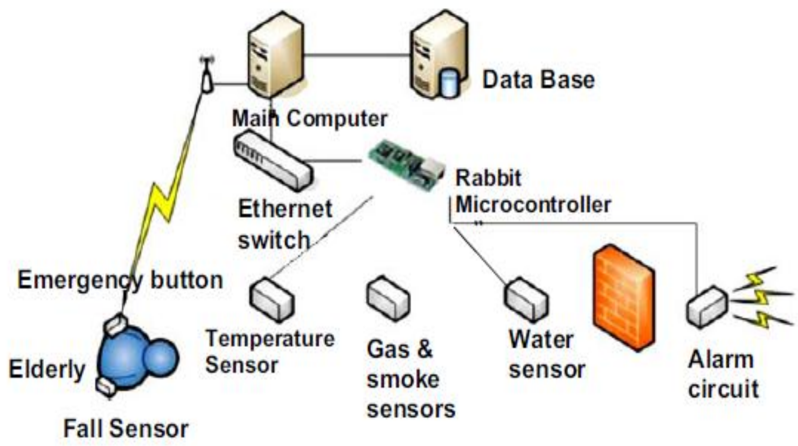

Figure 2: Layout of emergency sensors.

3.2. Data Mining: A component of Knowledge Discovery in Databases (KDD) is defined as the process of detecting and extracting hidden, unanticipated and potentially useful information from large amounts of datasets [15]. They are basically divided into two types of data mining models; the descriptive model which finds patterns and relations in the data and the predictive model which forecasts unknown values based on historical information [15]. The information that is extracted from large amounts of data '*can be stored in databases, data warehouses, OLAP or other repository information'* [17]. Different data mining models have been implemented by several research groups with the purpose of extracting meaningful information from the data and making accurate predictions.

3.3. Data Visualization: This is the process of communicating information distinctly and intelligibly by visualizing the data using through graphical means [18]. Data visualization could also be defined as *"the process of visually exploring data for pattern and trend analysis"* [19]. One of the most popular technologies that incorporates the ability to visualize complex data include the Online Analytical Processing (OLAP) whose functionality is the systematic exploration and handling of large datasets especially in domains such as smart homes. Data visualization includes other varieties such as information visualization, scientific visualization, information graphics and visual designs [20].

3.4. Activities of Daily Living (ADL): This terminology generally describes the common daily activities that people performs in their homes. Many researchers believe that the automatic recognition of human activities of daily living (ADLs) is a very important constituent for systems that analyze and detects behavioural patterns and according to the scope of the Netcarity Integrated Project funded by the European Union [21], it is believed that such systems will greatly increase and improve the independence of the elderly people by enabling and providing automatic assistive services [22]. The basic ADLs that can be used to measure the medical status of a an individual includes eating and drinking, dressing and undressing, personal hygiene, going to the bathroom (Elimination), moving from the chair to bed and vice versa, sleeping, working and playing e.t.c [23]. There are two scales used to measure and determine the functionality of older people and it was developed by Lawton in 1969 [2]. The first scale was called Physical Self-Maintenance (PSMS) which had to do with measuring self-care ability in areas such as bathing, eating and drinking, dressing, grooming and the second scale was called Instrumental Activities of Daily Living (IADL) which mainly focused on measuring a set of more complex human behaviours which consisted of activities such as gardening, shopping, transportation, food preparations, receiving of telephone calls, listening to music and watching videos, laundering e.t.c [24]. Another model called Roper-Logan-Tierney Model was first developed in 1980 based on the initial work done by Nancy Roper in 1976 and was a very popular nursing model used most especially in the United Kingdom (UK) [25]. This model proposed another way to measure the functional status of a person though this model had to be refined several times in 1980, 1981 and 1983 by Roper, Logan and Tierney until it was finally published in 2000. This model consisted of basically five components namely activities associated with living which includes the lifespan, activities of living, independence continuum, and dependence factors influencing ADLs and individuality in living [2].

4.0 Related Works

Several academic institutions and industries have undertaken researches in the monitoring the human behaviour and the activities of daily living of people (in this case the elderly people) who desire to live independently in their homes. These researches focused on implementing data analysis techniques on the datasets collected from the sensors installed in various locations either in their homes or in living labs for the purpose of detecting patterns and changes in human behaviour, extracting and interpreting the information that will be beneficial to the well-being of the elderly people. This have been achieved using data analysis techniques such as data mining, predictive algorithms, data visualization and trend analysis on vast amounts of datasets acquired from smart homes. One of the current researches undertaken by Gil et al sought to answer some critical questions about how information can be ‘mined’ from the elderly people living in their homes [2]. The purpose was to find ways of improving the health care of the older people and their overall daily living by discovering and analyzing patterns and trends in individual's life in the domestic context and also by modeling the '**busyness**' of the activity of the individual. They used data mining techniques (decision trees) for creating a set of rules that model the lifestyle of the occupant and Online Analytical Processing (OLAP) for visualizing their daily activities, detect changes in their health status and being able to make predictions based on the information [15].

Stankovski and Trnkoczy demonstrated a way of making smart homes to be 'smarter' using decision trees and their corresponding models both descriptive and predictive [26]. Decision Tree Analysis are very popular and useful in determining the existence of patterns and extracting them into large databases and it's extensively used for exploratory data analysis (EDA) and predictive modeling applications [27] in addition to monitoring a sequence of actions taken by a person in the home and enables a monitoring system to react accordingly especially if the person requires immediate attention [28]. Their purpose was to tackle a generic issue on whether events or actions taken by an individual in smart house follow a usual trend and they sought to achieve this through the pre-processing the relevant data gathered from the sensors into an input for a decision tree model [26]. Choi et al presented a ubiquitous intelligent sensing system for smart homes that detects and analyzes the stress patterns experienced by any persons living in such smart homes and providing automatic services. They implemented the association rules of data mining in addition to linear support machines for analyzing home context patterns [29]. Luhr et al introduced a new way of detecting unnatural (anomalous) behavioural patterns or changes from people living in smart homes through the deployment of intertransaction association rule (IAR) mining which basically depicts "*association relationships that span outside traditional "market basket" intratransaction items or events in one or more domain specific dimension"* [30] and they considered this method to be a very efficient data mining algorithm.

The MavHome (Managing An Intelligent Versatile Home), a multi-disciplinary smart house research project was developed with the purpose of creating an intelligent environment that will maximize and improve on the daily living and well-being of the residents especially the older people by being aware of their activities, automating them and proactively responding to any changes in their conditions at a reduced cost [31]. This intelligent agent achieves this objective by using several prediction algorithms [32] and temporal rule data mining technique for learning, predicting and adapting to the residents and comprehending the conditions of the home through sensors and acting accordingly to ensure the inhabitant's maximum comfort [33]. The architecture of the MavHome as shown in Figure 3 basically comprised of four layers which includes Decision Layer (responsible for selecting actions based on the information given by the Information Layer), Information Layer (responsible for collection, storage and production of knowledge for decision making), Communication Layer (facilitates communication of queries and requests between agents) and the Physical Layer comprising of devices such as transducers and network hardware [32].

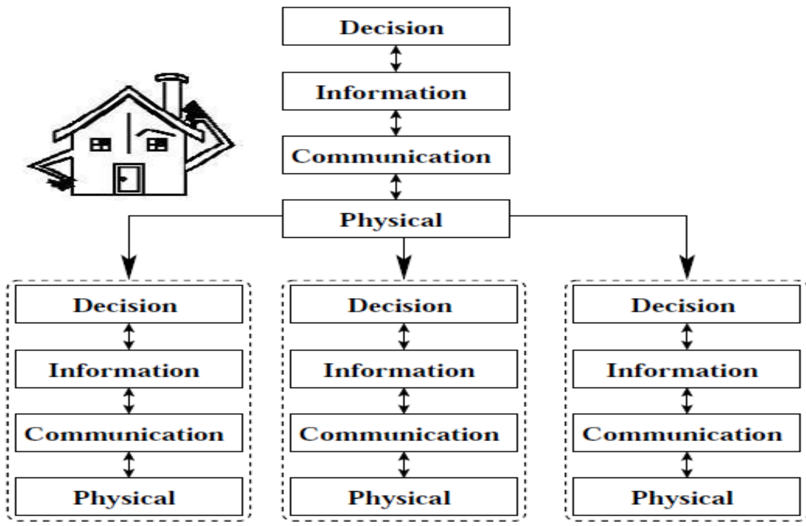

Figure 3: The MavHome Architecture**.**

Barger et al used a mixed model approach which comprised of implementing the clustering algorithm in the development of a probabilistic model for detecting patterns and analyzing behaviour from the datasets of the Medical Automation Research Center (MARC) smart house project [34]. Another research work developed by Massachusetts Institute of Technology (MIT) in partnership with TIAX, a leading collaborative product and technology development firm was the PlaceLab; *"a real home where the routine activities and interactions of everyday home life can be observed, recorded for later analysis, and experimentally manipulated"* [35]. The PlaceLab living house architecture included multiple non-invasive sensors installed in virtually every area of the house where the sensors are used to develop interface applications that will assist subjects in being healthy, active, helping them to save resources and to monitor the activities of various subjects for a limited period at various lengths of time [36]. One of the main focuses of the project was to determine how these new technologies affect the behaviour of people living in the PlaceLab [37].

5.0 Critical Analysis of Data Mining Models implemented in Current Research Works

Several research groups have studied the behaviour of elderly people living in smart home environments using various data mining techniques with the primary purpose of discovering trends, patterns and extracting meaningful information. Gil et al sought to use a data mining model (Decision Trees) and data visualization using Online Analytical Processing (OLAP) to achieve two major objectives; being able to enhance the home-based care and in general the quality of the elderly people's lives by constructing a **'busyness'** model comprising of their activities in their homes [2] and at various levels of granularity as an indication of their welfare and monitoring of Activities of Daily Living (ADL) by focusing on specific activities (i.e. bathing, taking breakfast). Their work presented detailed research methods and methodology on how their experimentation was achieved from the selection of home and various subjects, the installment of Passive Infrared Sensors (PIR), collection of information which is stored online and transformed into a data-ware house, utilizing OLAP to visualize data and allowing end users to manipulate the information, and finally conducting data analysis by detecting patterns and rules [15].

Their results generated from the both the OLAP and the data mining model revealed patterns and intrinsic details that can give an idea of the busyness level of the individual. Here the 'busyness' activity of the occupant is their central focus of interest rather than comprehending their activities of daily living (ADL) because they sought to measure the overall activities of the occupant, their daily interactions with objects and their movements in their dwelling. The concept of developing a 'busyness model' of the occupant specifically defines the measurement of the overall lifestyle such as their movements, interactions with objects, habits and activities of the person [2]. This model was preferable because it gave the researchers a clearer picture into the occupant's lifestyle and thus will enable them to effectively make predictions based on them but more importantly, this model helped to preserve the privacy of the occupant to a higher degree [2] (privacy has become another major issue when it comes to monitoring people and this could contribute to the difficulties in carrying data mining effectively).

Decision tree algorithms was implemented for the research perhaps due to its divide and conquer approach when it comes to classification of data according to and they are very useful in uncovering and extracting patterns from large datasets generated from the sensors for predictive modeling [27]. The rules which represented unequivocal events generated by this approach could be classified as a normal sequence of activities to be expected for example, it can be expected that during afternoon hours, a person will be in the living room doing activities like eating, watching television or having a conversation, while in the late evenings, it is expected that the occupant will be in the bedroom sleeping rather than being in any other location. While they preferred to use the pruned C4.5 decision tree algorithm method due to the accurate set of rules it produced and its advantage of being a robust model capable of handling noise data and learning disjunctive expressions [38], the rules applied were not strong enough in some certain scenarios. From the observations of the rules generated, there are bound to be deviations due to certain behaviour that may not be included in their rules; differences in lifestyle from one elderly person to another and also due to the irregularities in human behaviour as a result of emerging problems that can contribute to their health status or could be for other reasons that may not be health related. According to Nick Hine in a personal interview, these factors have contributed to the need for making the rules to adopt a fuzzy approach because the activities of daily living carried by people are never 100 percent [39]. Furthermore, George et al in their research made a comparative analysis of C4.5 classifier with association rule mining and based on the results concluded that while the former produced positive results, the latter was proven to be more accurate, flexible and more suitable for a smart home environment [40]. The implementation of OLAP using an application called ProClarity by the research group offered another approach in determining the health status of the occupant by visualizing the activity levels through graphical means [15]. The major advantage of this approach is that it gives the medical professionals (health carers) and the clients (the occupant and their relatives) a graphical idea of the health status by comparing the activities levels produced on a daily basis with other times (days or weeks) to determine any resulting changes that may have occurred which could either be due to physiological factors or other unknown factors. However the disadvantage of visualizing the activity levels of the occupant is that they may not present an accurate picture of their status; this can be attributed to the unpredictable nature of human behaviour and every individual may not behave in the same way in similar conditions.

Another research work was the MavHome (Managing An Intelligent Versatile Home), a more complex framework whose main objective was to create an intelligent environment that will improve the daily living and well-being of the elderly people by monitoring their activities, automating and responding to any changes [32]. The distinctive feature of the MavHome intelligent environment is its automation through learning and prediction by their interactions with devices (motion sensors) and discovering interactions that are considered as sequence of events [41]. Their work implemented a Temporal Rule data mining, a relatively new area of research that gained popularity in terms of its ability to process large amounts of complex data and comprises of classification, clustering, prediction and rule discovery and prediction [42]. This method was effective for uncovering patterns as well as gaining knowledge of the resident's daily activities and these discoveries are then used in the development of systems that serves as reminder assistants and helping them maintain their well-being by detecting deviations and tacking appropriate remedies [43].

Temporal rule mining and pattern discovery was extensively used according to Jakkula et al in their research work that involved obtaining temporal rules from a time series representation of activities detected in smart homes and corroborating this algorithm based on Allen's temporal logic (interval relations) by utilizing both real and synthetic data from MavHome [33]. In their series of experiments, X and Y were denoted as variables that described 'Events' that had taken place where the prediction is given for X giving information about event Y showing a relationship that is temporal through the following steps [33]:

Step A: Temporal relations were observed through the analysis of events X and Y with Y being the most current activity.

Step B: The occurrence of event X was then calculated by implementing the following formula in Equation (1) with Equation (2) denoting the evidence of X:

$$P(X|Y) = |After(Y,X)| + |During(Y,X)| + |OverlappedBy(Y,X)| + |MetBy(Y,X)| \quad (1)$$
$$+ |Starts(Y,X)| + |StartedBy(Y,X)| + |Finishes(Y,X)| + |FinishedBy(Y,X)|$$
$$+|Equals(Y,X)| / |Y| \ (1)$$

$$Evidence_X = P(X) \quad (2)$$

There were cases where the evidence needed to be placed in combination with multiple events that exhibited a temporal relationship with event X. In this scenario, they observed the start of events A and B to be able to establish X and how likely the event X will occur [33]. Based on the Equation (1), the evidence for B was derived to produce Equation (3) as follows:

$$P(B|A) = After(B,A)| + |During(B,A)| + |OverlappedBy(B,A)| + \quad (3)$$
$$|MetBy(B,A)| + |Starts(B,A)| + |StartedBy(B,A)| + |Finishes(B,A)|$$
$$+|FinishedBy(B,A)|+|Equals(B,A)| / |A|$$

Association and Multiplicative rules were also applied to create a new formula in Equation (4) that included previous calculated evidences of events that took place and then used to compute the most current event that occurred [36]. The evidence of B was now computed as shown below:

$$P(X|AUB) = P(X \cap (AUB)) / P(AUB) \quad (4)$$
$$= P(X \cap A) \cup P(X \cap B)/ P(\mathrm{A}) + P(B) - P(A \cap B)$$

[Association Rule]

$$= P(X|A).P(A) + P(X|B).\mathrm{P}(B) / P(A) + P(B) - P(A \cap B)$$

[Multiplication Rule]

Step C: This final step involved calculating the prediction of X using the formula in Equation (5).

$$Prediction_X = P(X) \quad (5)$$

However, there were notable disadvantages of the application of Allen's method and this was further validated in Morgan's work which indicated that the model was not strong because of its patterns producing varying interpretations (ambiguous) due to disturbances in interval boundaries [44]. One of the benefits of the implementation of this technique in MavHome is its ability to detect relationships and sequence of events thereby improving the learning, prediction and decision making of the smart house. However, another notable problem highlighted in their research was the application of data visualization in the temporal intervals although they came up with the solution of developing a smart interval visualization tool that will enable to detect patterns and visualize activity levels but more crucially, the visualized results served to benefit the occupants by improving their health and lifestyle [33].

Another prediction algorithm called SHIP (Smart Home Inhabitant Prediction) that was designed and implemented in the MavHome allowed the intelligent environment to gather information about its inhabitants to enable it to adapt and respond to any changes [32]. In this SHIP algorithm, commands of the occupant are collected as '*Actions*' and '*Matches*' and are basically composed of two steps. The first step consists of the match queue being brought up to date when a new action is recorded and it is further calculated where time (t) is given in state (s) to compute for l_t (s, a) which describes the length of the longest sequence that ends with a (action) and s (state) which then matches the history sequence to t (time) [32]. The second step involves the evaluation of the matches in the queue based on its length and frequency (the measure of frequency is denoted by the number of times the action (a) has been taken for current state (s)) [32]. From the Equation (6) below, the SHIP algorithm places matches on the combination of match length and normalized match frequency.

$$Rt\ (s, a) = \alpha\ \frac{l_t\ (s,a)}{\sum_t l_t\ (s,a_i)} + (1 -\ \alpha)\frac{f(s,a)}{\sum_t f(s,a_i)} \tag{6}$$

where a is returned by SHIP and is indicated by the match (s, a) with R_t (s, a) as its highest prediction value. While this algorithm produced high prediction accuracy outputs on both on their real (53.4% to 80%) and synthetic data (94.4%), its drawbacks included the history of actions recorded for an occupant needed to be processed offline to give a prediction, not practically viable over a long time period and can only predict the next event and not at the time it occurs [32].

A Smart House project developed by the Medical Automation Research Center (MARC), University of Virginia implemented based their model on previous works that utilized different analysis techniques such as clustering, neural networks, and the use of histograms and charts to infer activities by applying mixed models to their sensor data in developing a probabilistic model that was used to test the real data from their smart house [34]. In comparism to the trial project conducted by Gil et al, they both have a similar framework and objective of improving the elderly people's health by installing technologies that will be able remote monitoring and alerting health carers, relatives on any changes to their health status. The difference however is that while the former conducted a singular data mining approach aside from visualizing the occupant's daily activities, the latter sought to combine different approaches (clustering, rule based) to be able to detect and infer activities from the sensor data generated from simple sensors that was installed in their system (this was preferable to using video cameras as that would have invaded the privacy of the occupant).

Their work demonstrated how mixed model was used to split their data collected over a period of 65 days into training and test datasets but they specifically focused on categorizing the 'days' into 'work days' and 'off days'. They implemented a clustering approach using Ward's method to split the 65 days into two clusters of 28 'work days' and 37 'off days' and on the time interval of 7am to 7pm where each observation was placed into their respective cluster and eventually combined. One of the results of their analysis indicated that there was a low uncertainty when the clusters occurred uniformly overtime [34]. Although their research did not explicitly state how efficient and effective this model is in analyzing their datasets and if the results generated show an accurate assessment of the state of the individual's well-being. However this mixed model approach could potentially be applied in Gil et al's research work by creating clusters based on their segmentation of time zones with each time interval indicating a specific event carried out by the occupant as illustrated in Table 1 where for example for the hour of '**7:00am to 9.00am**', clusters can be designated for a specific or series of activities that took place in the '**early morning**'.

Table 1: Segmentation of time into zones.

Zone	Hour
Sleeping	00:00 - 7:00 a.m.
Early morning	7:00 - 9.00 a.m.
Late morning	9:00 - 12:00 p.m.
Lunch	12:00 - 4:30 p.m.
Afternoon	4:30 - 7:00 p.m.
Evening	7:00 - 10:30 p.m.
Late evening	10:30 - 12:00 midnight

6.0 Discussion

Research groups have sort to improve on the prediction accuracy when analyzing vast amounts of data generated by sensors used in monitoring people's lifestyle by utilizing a combination of one or more data mining algorithms to achieve better prediction accuracy. Some of the researchers have in addition to the data mining models implemented in their work, have also made use of data visualization applications to further give a pictorial/graphical representation of the lifestyle of the individual. However from the evaluation of the various algorithms utilized by these research groups, it can be deduced that one model that may work in a particular problem scenario may not yield a similar outcome in another scenario. One of the major challenges that research groups have faced in discovering and effectively interpreting the data from smart homes is the variability in human behaviour. This is also attributed to the fact that research groups do not have access to the datasets generated by their counterparts that would have allowed them to validate their data mining models they utilized in their own work. For example, the formulas that were generate in relation to the temporal rule data mining implemented in the MavHome [33] could have been validated on another research scenario to determine how high and consistent the prediction accuracy is. This particular issue is as a result of the sensitive nature of the data which in most cases cannot be shared between research groups due to privacy and insurance concerns.

Despite the advantages and disadvantages that have been documented in these research works, it is worth noting that some of these data mining models can be combined with other models depending on the scope and complexity of the datasets relating to the individuals (in this case the elderly people). The best and logical approach will be to implement a 'mixed model' that involves a combination of specific predictive algorithms such as decision trees (known for its simplistic and efficient approach), the application of temporal rule data mining that deals with a sequence of events that are time related in collaboration with using data visualization techniques such as Online Analytical Processing (OLAP) as was highlighted in the work of Gil et al [15] which can help to increase the likelihood of not only achieving higher prediction accuracy but will also allow researchers to have a graphical presentation of the lifestyle of the elderly people where specific trends and non-trivial patterns can be inferred from.

7.0 Conclusion

In this study, we presented the Smart home technology, its various components and potential long term benefits it provides especially for the elderly and disabled people. This technology enables an automated environment that not only promotes independence; it also improves their health and quality of life. This study also discussed the various smart home research projects developed by various research groups and the data mining models that were implemented to discover patterns and behaviour of people living in them. However we also discovered that it is a challenge for these groups to effectively analyze the daily activities of the elderly and disabled people living in smart homes. Each of these research groups have used either a specific data mining model or have combined more than one data mining algorithms to uncover trends and patterns from large amounts of data and based on the information that is extracted will enable them to make effectively predict human behaviour.

One of the major challenges of data mining involving people living in smart homes that these research groups have faced is dealing with the variability and unpredictable human behaviour (people having different lifestyle patterns) while trying to understand and make sense of the data. Another crucial issue noted in this study had to do with determining which data mining model is most effective in discovering hidden patterns in human behaviour and producing an accurate assessment of the well-being of the elderly people. This is because each of these data mining techniques have their respective advantages and drawbacks.

References

[1] Naeem, U. and Bigham, J. A Comparison of Two Hidden Markov Approaches to Task Identification in the Home Environment. *Proceedings of the 2nd International Conference on Pervasive Computing and Applications*, Birmingham, UK, pp. 383-388, 2007.

[2] Gil, N. M., Hine, N. A., Arnott, J. L., Hanson J., Curry, R. G., Amaral, T. and D.Osipovic, D. Data visualization and data mining technology for supporting care for older people. *Proceedings of 9th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2007)*, Tempe, Arizona, USA, pp.139-146, 2007.

[3] Hine, N., Judson, A., Ashraf, S., Arnott, J., Sixsmith, A., Brown, S., and Garner, P. Modeling the Behaviour of Elderly People as a Means of Monitoring Well Being. *International Conference on User Modeling (UM 2005)*, Lecture Notes in Computer Science (LNCS), Vol.3538, Edinburgh, Scotland, Springer-Verlag, pp. 241-250, 2005.

[4] Fern´andez-Luque, F., Zapata, J., Ruiz, R., and Iborra, E. A Wireless Sensor Network for Assisted Living at Home of Elderly People. In J. Mira et al. (eds.), *IWINAC 2009, Part II, LNCS 560,* Springer-Verlag Berlin Heidelberg, pp. 65-74, 2009.

[5] Bengtsson, T. and Scott, K. Population Ageing: A Threat to European Welfare? In S. Gustavsson et al. (eds.), *How Unified Is the European Union?*, Springer-Verlag Berlin Heidelberg, pp. 117-137, 2009.

[6] Cook, D. J., Schmitter-Edgecombe, M., Crandall, A., Sanders, C., and Thomas, B. Collecting and Disseminating Smart Home Sensor Data in the CASAS Project. *Proceedings of the CHI Workshop on Developing Shared Home Behaviour Datasets to Advance HCI and Ubiquitous Computing Research*, [Online] 2009. Available from: http://www.eecs.wsu.edu/~cook/pubs/chi09.1.pdf [Accessed 19 March 2010]

[7] Edwards, W. K. and Grinter, R. E. At home with ubiquitous computing: Seven challenges. In *Proceedings of third international conference ubiquitous computing (LNCS)*, Atlanta, Georgia, Springer pp. 256-272, 2001.

[8] Edge M., Taylor B., Dewsbury G. and Groves M. The Potential for Smart Home Systems in Meeting the Care Needs of Older People with Disabilities. *Housing Update*, Gerontology Research Centre, Canada 10.1, pp. 6-7, 2000.

[9] Bierhoff, I., Berlo, A., Abascal, J., Allen, B., Civit, A., Fellbaum, K., Kemppainen, E., Bitterman, N., Freitas, D., and Kristiansson, K. Smart home environment. In Roe, P.R.W. (ed.), *Towards an Inclusive Future: Impact and wider potential of information and communication technologies*, Brussels: COST 219ter, pp. 111-149, 2007.

[10] Savage R., Yon, Y., Campo, M., Wilson, A., Kahlon, R., and Sixsmith, A. Market Potential for Ambient Assisted Living Technology: The Case of Canada. In M. Mokhtari et al. (eds.), *ICOST 2009, LNCS 5597*, Springer-Verlag Berlin Heidelberg, pp. 57-65, 2009.

[11] Sánchez, P. N., and Molina, J. M. A Smart Solution for Elders in Ambient Assisted Living. In J. Mira et al. (eds.), *IWINAC 2009, Part II, LNCS 5602*, pp. 95-103, 2009.

[12] Jara, A. J., Zamora, M. A., and Gómez-Skarmeta, A. F. An Architecture for Ambient Assisted Living and Health Environments. In S. Omatu et al. (eds.), *IWANN 2009, Part II, LNCS 5518*, Springer-Verlag Berlin Heidelberg, pp. 882–889, 2009.

[13] AAL. Ambient Assisted (AAL) Joint Programme: Objectives, [Online] 2010. Available from: http://www.aal-europe.eu/about-aal [Accessed 20 March 2010]

[14] Mukhopadhyay, S. C., Gaddam, A., and Gupta, G. S. Wireless Sensors for Home Monitoring - A Review. *Recent Patents on Electrical Engineering 2008*, Bentham Science Publishers Ltd, [Online], 1.1, pp. 32-39, 2007.

[15] Gil, N. M., Hine, N. A., and Judson, A. Lifestyle Monitoring System to Improve the Well-being of the Elderly. *Advances in System Information*, Medellin, Junio, 3.1, pp.39-44, 2006.

[16] Raad, M. and Yang, L. A ubiquitous smart home for elderly. *Information System Frontiers*, Springer Science + Business Media, LLC 11.5, pp. 529-536, 2008.

[17] Nasereddin, H. H. O. Stream Data Mining. *International Journal of Web Applications*, [Online] 1.4, 2009. Available from: http://www.dline.info/IJWA%20sample%20paper.pdf [Accessed 16 April 2010]

[18] Friedman, V. Data Visualization and Infographics, [Online] 2008. Available from: http://www.smashingmagazine.com/2008/01/14/monday-inspiration-data-visualization-and-infographics/ [Accessed 8 March 2010]

[19] Shekhar S., Chang-Tien, L., Zhang, P., and Liu, R. Data Mining for Selective Visualization of Large Spatial Datasets. *ICTAI: Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence*, IEEE Computer Society, Washington DC, USA, pp. 41-48, 2002.

[20] Friendly, M. Milestones in the history of Thematic Cartography, Statistical Graphics, and Data Visualization, [Online] 2009. Available from: http://www.math.yorku.ca/SCS/Gallery/milestone/milestone.pdf [Accessed 11 April 2010]

[21] Netcarity. Ambient Technology to Support Older People at Home, [Online] 2008. Available from: http://www.netcarity.org/About.11.0.html [Accessed 15 April 2010]

[22] Libal, V., Ramabhadran, B., Mana, N., Pianesi, F., Chippendale, P., Lanz, O., and Potamianos, G. Multimodal Classification of Activities of Daily Living Inside Smart Homes. In S. Omatu et al. (eds.), IWANN 2009, Part II, LNCS 5518, Springer-Verlag Berlin Heidelberg, pp. 687-694, 2009.

[23] Gilbert Guide. Activities of Daily Living: What are ADLs and IADLs, [Online] 2010. Available from: http://www.caring.com/articles/activities-of-daily-living-what-are-adls-and-iadls [Accessed 15 April 2010]

[24] Graf, C. The Lawton Instrumental Activities of Daily Living (IADL) Scale, [Online] 2007. Available from: http://www.annalsoflongtermcare.com/article/7453 [Accessed 25 March 2010]

[25] Roper, N., Logan, W., and Tierney, A. The Roper-Logan-Tierney Model of Nursing. Churchill Livingstone, Edinburgh, Scotland, UK, 2000.

[26] Stankovski, V. and Trnkoczy, J. Application of Decision Trees to Smart Homes. In Augusto, J. C. and Nugent, C. D. (eds.), *Designing Smart Homes, LNAI 4008,* Springer-Verlag Berlin Heidelberg, pp. 132 – 145, 2006.

[27] Myles, A. J., Robert N. Feudale, R. N., Woody, N. A., Liu, Y., and Brown, S. D. An Introduction to Decision Tree Modeling. *J. Chemometrics,* 18, pp. 275-285, 2004.

[28] Augusto, J. C., and Nugent, C. D. Smart Homes can be Smarter. In J. C. Augusto and C.D. Nugent (eds.), *Designing Smart Homes, LNAI 4008,* Springer-Verlag Berlin Heidelberg, pp. 1-15, 2006.

[29] Choi Jonghwa, Shin Dongkyoo, and Shin Dongil. Ubiquitous Intelligent Sensing System for a Smart Home. In D.-Y. Yeung et al. (eds.), *SSPR & SPR 2006, LNCS 4109,* Springer-Verlag Berlin Heidelberg, pp. 322-330, 2006.

[30] Luhr, S., West, G., and Venkatesh, S. An Extended Frequent Pattern Tree for Intertransaction Association Rule Mining. *Technical Report,* Department of Computing, Curtin University of Technology, [Online] 2005. Available from: http://www.computing.edu.au/~svetha/papers/papers2005/reports/luehr_EFP-Tree_TR2005_01.pdf [Accessed 1 April 2010]

[31] Mehta, R., Cook, D. J. and Holder, L. B. Identifying Inhabitants of an Intelligent Environment Using a Graph-Based Data Mining System, [Online] 2003. Available from: http://www.aaai.org/Papers/FLAIRS/2003/Flairs03-061.pdf [Accessed 2 April 2010]

[32] Cook, D. J., Youngblood, M. G., Heierman, E. O. III, Gopalratnam, K., Rao, S., Litvin, A., Khawaja, F. MavHome: An Agent-Based Smart Home. *Proceedings of the First IEEE International Conference on*

Pervasive Computing and Communications, March 23rd -26th, IEEE Computer Society, Washington DC, USA, pp. 521-524, 2003.

[33] Jakkula, V. R., and Cook, D. J. Mining Sensor Data in Smart Environment for Temporal Activity Prediction. *Proceedings of the First International Workshop on Knowledge Discovery from Sensor Data* [Online] 2007. Available from: http://eecs.wsu.edu/~vjakkula/ICTA.pdf [Accessed 2 April 2010]

[34] Barger, T., Brown, D., Alwan, M. Health Status Monitoring Through Analysis of Behavioural Patterns. Lecture Notes in Artificial Intelligence (LNCS/LNAI), *Proceedings of the 8th Congress of the Italian Association for Artificial Intelligence (IAAI) on Ambient Intelligence*, Springer-Verlag, Pisa, Italy, [Online] 2003. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.103.760&rep=rep1&type=pdf [Accessed 18 April 2010]

[35] Intille, S. S., Larson, K., Beaudin, J. S., Tapia, E. M., Kaushik, P., Nawyn, J., McLeish, T. J. The PlaceLab: A Live-In Laboratory for Pervasive Computing Research (Video). In *Proceedings of Pervasive 2005 Video Program*, [Online], 2005. Available from: http://web.media.mit.edu/~intille/papers-files/PervasiveIntilleETAL05.pdf [Accessed 22 March 2010]

[36] House_n Research Group. The PlaceLab, [Online] 2004. Available from: http://architecture.mit.edu/house_n/placelab.html [Accessed 14 March 2010]

[37] Massachusetts Institute of Technology (MIT). The PlaceLab: An MIT and TIAX Initiative, [Online] 2003. Available from: http://web.media.mit.edu/~kll/A_PlaceLab_Sept1-2004.pdf [Accessed 18 March 2010]

[38] Ratanamahatana, C. A., and Gunopulos, D. Scaling up the Naive Bayesian Classifier: Using Decision Trees for Feature Selection. *In proceedings of Workshop on Data Cleaning and Preprocessing at IEEE International Conference on Data Mining*, Maebashi, Japan, [Online] 2002. Available from: http://www.cs.ucr.edu/~ratana/DCAP02.pdf [Accessed 15 April 2010]

[39] Hine, N. Data mining elderly people in Smart homes for MSc project. Interviewed by Igene, O. 4 March, 2010.

[40] Papamatthaiakis, G., Polyzos, G. C., and Xylomenos, G. Monitoring and Modeling Simple Everyday Activities of the Elderly at Home. *Proceedings of the IEEE Consumer Communications and Networking Conference (CCNC)*, [Online] 2010. Available from: http://www.mm.aueb.gr/publications/2010-ARCH-CCNC.pdf [Accessed 17 April 2010]

[41] Das, S. K. and Cook, D. J. Designing Smart Environments: A Paradigm Based on Learning and Prediction. In S.K. Das and D. J. Cook (eds.), *Pattern Recognition and Machine Intelligence 2005, LNCS 3776.* Springer-Verlag, Berlin Heidelberg, pp. 80-90, 2005.

[42] Galushka, M., Patterson, D., and Rooney, N. Temporal Data Mining for Smart Homes. In J. C. Augusto and C. D. Nugent (eds.), *Designing Smart Homes, LNAI 4008.* Springer-Verlag Berlin Heidelberg, pp. 85-108, 2006.

[43] Jakkula, V. R., and Cook, D. J. Using Temporal Relations in Smart Home Data for Activity Prediction. *Proceedings of the ICML Workshop on the Induction of Process Models,* [Online] 2007. Available from: http://wwwkramer.in.tum.de/ipm07/ipm07_jakkula_cook.pdf [Accessed 26 March 2010]

[44] Mörchen, F. A better tool than Allens relations for expressing temporal knowledge in interval data. In *Proceedings Temporal Data Mining Workshop (TDM), The Twelveth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Philadelphia, PA, USA, [Online] 2006. Available from: http://www.mybytes.de/papers/moerchen06tdm.pdf [Accessed 22 April 2010]