SIMULATING ENSEMBLE LEARNING TECHNIQUE MODEL FOR DETECTING TEMPORAL ANOMALIES IN OLD PEOPLE'S HOME

Article · September 2021

DOI: 10.5281/zenodo.5504103

SIMULATING ENSEMBLE LEARNING TECHNIQUE MODEL FOR DETECTING TEMPORAL ANOMALIES IN OLD PEOPLE'S HOME

F U Ogban[1], E A Edim[1], A E Bassey[1], E U Oyo-ita[2], P Ana[2]
felixogban@unical.edu.ng, **edimemma@gmail.com**, ave.ebock@gmail.com, **emmanueloyoita@crutech.edu.ng, princeana@crutech,edu.ng**

[1]Computer Science Department, Faculty of Physical Science, University of Calabar
[2]Cross River University of Technology, Calabar Campus, Ekpo Abasi.

Abstract

The need to have a secure lifestyle for older persons at home is in demand more than ever. One of the major concerns for smart home systems is the capability of adapting to the user as personalizing the behaviorur of the home may provide improved comfort, control, and safety. The aim of this work is to (predict temporal anomalies among the aged living in older people home using ensemble learning algorithms) develop and simulate an ensemble learning algorithm for temporal anomaly detection of activities of daily living in old people's home. To achieve this goal, we applied two supervised learning algorithms which include the k-nearest neighbour (K-NN) and naïve Bayes (NB) algorithm. This two supervised learning algorithms were used to build the ensemble learning algorithm based on a 60day period secondary data collection, consisting of three of the four non-repetitive daily routine activities used for finding anomalies. The remaining data was used for testing the model. Rapidminer was used in model testing and the results showed that Ensemble learning model yielded better results than individual algorithms. This is because the two algorithms complemented each other as the strength of NB which is the stronger algorithm in the study; with capacity of handling missing value within a dataset, augmented the K-NN algorithm. The model's suitability was proven being that its average rate leads to increase in prediction accuracy. However, in this study NB algorithm was used in combination with K-NN as an ensemble learning technique, which improved k-nearest neighbour by increasing the ensemble learning model; producing a 95.0% level of accuracy as against the K-NN with an accuracy level of 71.67% singly. The result also revealed that NB algorithm is best when dealing with multivariate dataset and dataset with missing values as it individually outperformed the ensemble learning algorithm with an accuracy of 98.33% against 95.00% respectively. The ensemble learning algorithm has proven to have the capacity to handle missing values within data set, and so its average rate leads to better prediction accuracy.

Keywords: Temporal Anomalies, Aged people, Ensemble Learning Techniques, Simulation, Naïve Bayes, K- Nearest Neighbour, Supervised Learning,

1. Introduction

Anomalies are simply patterns or trends of older adults which are abnormal to their usual routine activities of daily living in a smart home whose data is collected via a wireless sensor network and stored in a central database for analysis. Areas of anomaly in these smart homes are: behavioural, spatial and temporal. However, not much has been done in the area of temporal anomaly; hence, this paper presents simulating an ensemble model for detecting temporal anomaly in old people's home based on the duration of non-repetitive generalized activity for a period of 60 days.

Not long ago, data collection on human daily living meant carrying out surveys or using witnesses in the location of other humans to record observations on the behavioural habits. However, at the advent of microprocessor miniaturization, computing power has been implemented in mobile devices and gadgets and has almost invaded every sphere of life. This birthed the evolution of wireless, modest and affordable sensors for information collation in daily 'smart' surroundings.

This information so collected when studied using machine learning techniques, do not only give insight that can be inter-leavened in our lives but can also produce computerized domain specific support to daily living environments, for example a smart home. However, researchers have been more focused on the use of smart home automation to observe occupant's ability to carry out key Activities of Daily Living (ADL) as well as instrumented ADLs (iADLS) due to its importance in independent living for the aged and health care management [1]. Such activities include: cook_breakfast, eat_breakfast, sleeping, showering and taking medications.

Research on independent living is however not limited to dementia sufferers. Many published works have also addressed the issue of independent living in a broader sense [2], [3], [4] The smart environment can identify, model and evaluate the performance progression of dementia of the Alzheimers type in the execution of activities of daily living. The smart home can either monitor and collect the activities information of the user by means of sensors or communicate and control its environment. The former approach is widely used for monitoring, anomalous behaviour detection, behaviour diagnosis and prediction of activities in an Ambient Intelligence (AmI) environment [5], [6], [7], while the latter approach is used to intervene and interact with the user as a means of preventing accidents and reminding the user of daily routine activities at the precise time.

In this paper, a 60 day period of data from the Tulin CASAS Washington State University [8] of four non-repetitive generalized activities of daily living of an older adult was used to build and simulate an ensemble model based on naïve bayes algorithm and k-nearest neighbour algorithm for detecting anomalies in the old people's home.

2. Related work

Anomaly simply put is the deviation from the normal patterns of older adults in carrying out their daily healthy living in a context aware environment. This context aware environments could be smart homes, Ambient Intelligent (AmI) environment to support the older adult in living a less dependent life.

These Smart homes are intentionally designed living spaces that provide interactive technologies and support systems to enable people – for example elderly persons to enjoy a high level of independence, activity participation or well-being than otherwise afforded [9], [10]. Longevity of life is often associated with high rate of vulnerability to diseases and injury [11]. Chronic diseases such as cancer, diabetes, arthritis, heart disease and chronic obstructive pulmonary disease are very common in older adults. Falls and injuries are even more common in elderly people [12]. It has been predicted that by 2035 the number of people with dementia will double [13] and by 2050 the number of full time carriers will have tripled [14]. With these current trends in population demographics, it is becoming increasingly difficult for governments worldwide to fully maintain their health and social care systems [15]. Therefore, the use of smart technologies, including smart-homes could arguably relieve the pressure on aged care health and social support services [11].

Reddy proposed the use of Bayesian networks as suitable for anomaly detection [16]; While Smolyakov et al, developed and implemented an automatic anomaly detection algorithm for meteorological time-series [17]; and Giannoni et al, proposed four different approaches for detecting anomaly, namely: Running Average Low-High Pass Filter, Univariate Gaussian Predictor, Seasonal ESD (Extreme Studentized Deviate) Algorithm and Local Density Cluster-based Outlier Factor (LDCOF). He discovered that LDCOF which is based on k-means clustering is best for multiple features or a multivariate dataset. They proved that K-means Clustering machine learning technique is best for creating a solution implemented using multiple features or a multivariate dataset [18]. However, since this paper is limited to the application of supervised machine learning technique; naïve bayes, k-nearest neighbour which is the supervised variant of the unsupervised k-means clustering algorithm was used.

Research in anomaly detection has presented a lot of solutions to effectively detect anomaly which can be behavioural, spatial and temporal for both categorical and continuous datasets. It has also been proven by recent researches that implementing solutions with individual machine learning technique for dataset inhibits the expected results [19]. Hence, ensemble learning techniques which allow for the implementation of two or more machine learning technique per time in creating a solution [17] is highly encouraged and recommended. This is because the shortcomings of either technique can be complimented by the corresponding technique. For example, in this paper the ensemble learning technique approach was used to build the model. This ensemble experts learning techniques used was Naïve Bayes algorithm and k-nearest neighbour algorithm which is the supervised variant of k-means clustering learning techniques.

3. Research Method

Data Collection

The data used in the study is secondary data collected from an online repository which consist data set of sensor events collated from the Washington State University, Centre for Advanced Studies in Adaptive Systems (WSU CASAS), Tulin smart apartment test bed from April to July of 2009 [8]. The apartment houses two married residents, though record of only one occupant is made public. The original data had ten generic activities of both repetitive and non-repetitive activity of daily living annotated by start time and end time of each routine activity occurrence having four attributes.

Data preparation and feature extraction

The data collected is contained in a text file and needed to be prepared by uploading it through the xlminer platform via the import text wizard. The xlminer converted the data from text file format into numeric and arranged in rows and columns. This resulted to 486,912 rows with 6 columns of data set. The dataset was then sorted and features were extracted into daily living activities with its corresponding start time, end time for each activity, the state, which is either the ‘begin’ or ‘end’ state per day. The duration (mins) of each activity was the focus of this study. This was derived as the difference between the start time of an activity and the end time of the activity.

The extraction of non-repetitive generic activity of daily living from the data set was carried out. This is made up of four (4) activities out of the ten the activities. The extracted activities also include the corresponding start time and end time of those activities. These activities include: cook_breakfast, cook_lunch, leave_home and enter_home. Five (5) attributes were used in this study. Two (2) of these attributes were derived from the initial data. the attributes include:

1. Approximate Duration: which was derived from the difference between the Start_Time and End_Time of the individual activity in Minutes?

2. State: derived from the approximate duration based on the following condition: duration (minutes) < 1 = False, OR duration (minutes) > 1 = True

The other regular attributes include:

1. Days
2. Start time
3. Adl (activities of daily living)

Modeling and simulation

In building and simulating the ensemble model for detecting temporal anomaly, Rapidminer, a predictive analysis tool was used. In using this tool, the example set made up 180 examples i.e. the rows, which were gotten from a 60day period of three different activities with one special attribute called the label and four regular attributes. This data set was added to the rapidminer repository via an operator named 'Add data' in a process environment. The data built was carried out via operators in a process. The first operator was the 'retrieve' operator, which is responsible for calling out data sets already saved in the rapidminer repository. The 'retrieve' operator has an 'out' port, it uses to connect to other operators via their ports or the 'res' port in the process environment. In building the ensemble model for this paper, the 'out' port of the retrieve operator was connected to the 'tra' port of the vote operator which is an operator capable oof implementing the more that one machine learning algorithm per time. The 'tra' port is simply what gives access to the data set for training while the 'mod' port on the right side of the vote operator, is the port responsible for the output of the model implementation. The 'mod' port is then connected to the 'res' port of the process environment. After this connection is set and ready, the run button is clicked on and the model is built. It is important to note that the vote operator is a sub process, and its at the sub process level naïve bayes algorithm and k-nearest algorithm are implemented via their operator calls as shown in the figure 1 below.

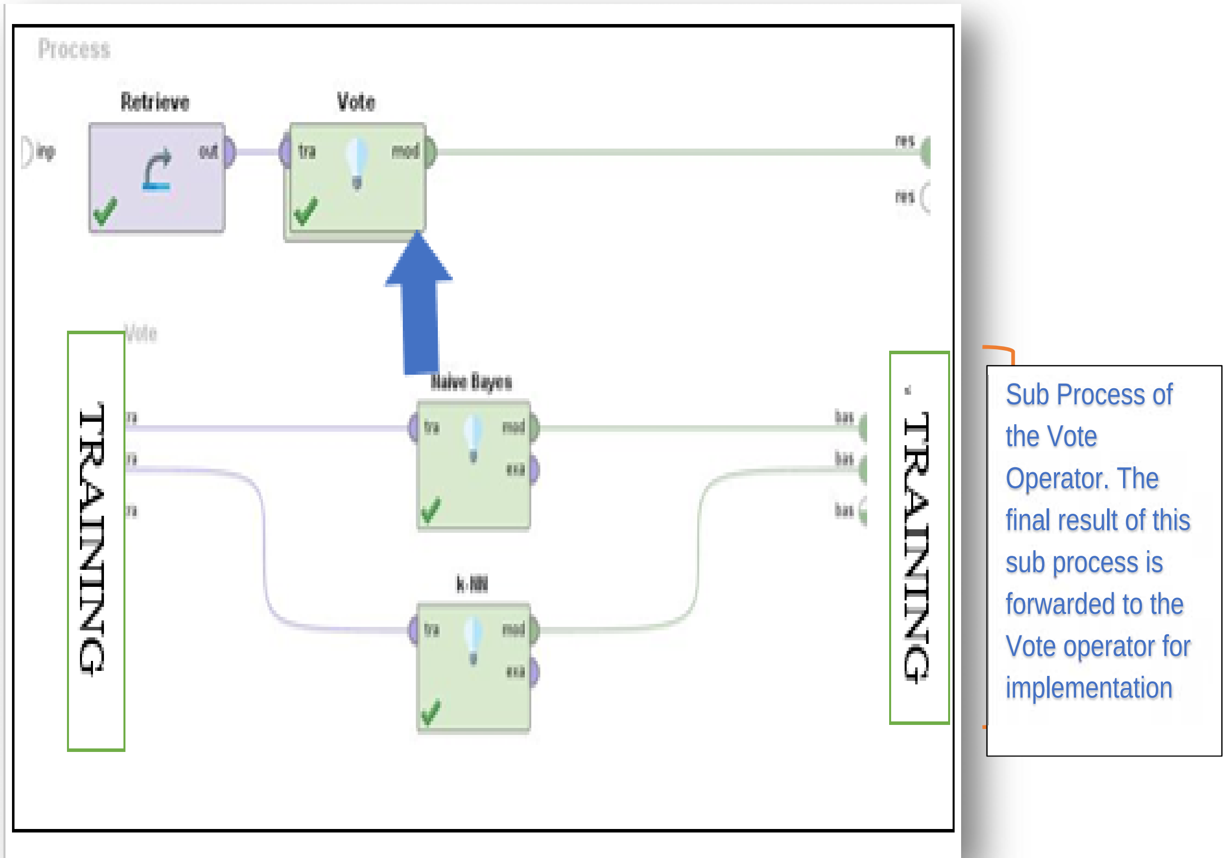

Figure 1: Design View of the Simulated Ensemble Model

4. Implementation

In implementing the model simulation, the operator called model simulator is called via its operator. This is done by dragging and dropping the operator in the process window. The model operator as shown in figure 2, has five ports; with three on the left side called ‘mod’, ‘tra’ and ‘tes’ respectively. The ‘mod’ port is connected to the ‘mod’ port of the vote operator; while the ‘tra’ port known as training is connected to the retrieve operator to feed in the training data already stored in the repository while the ‘tes’ port known for testing is being connected to the another retrieve operator for feeding in the test data also stored.

The model simulator has two ports on its right side namely: ‘sim’ i.e. simulation and ‘mod’ i.e. model ports. However, since we are more concerned about the simulation at this point, only the 'sim' port is connected to the 'res' port of the process environment as shown below.

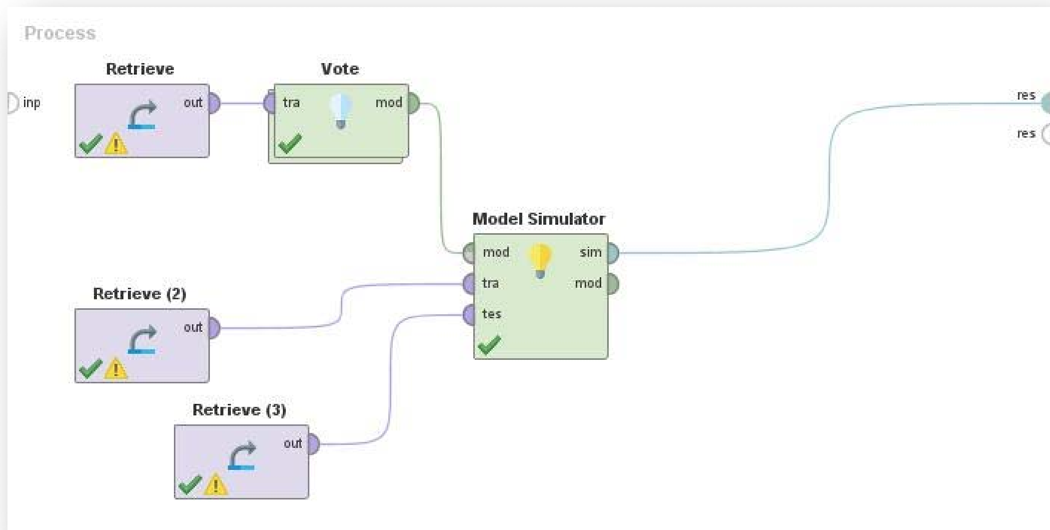

Figure 2: Process View of the Ensemble Model Simulator

5. Result and Performance Evaluation

The result of the model built from figure 1 and simulated in figure 2 as shown above is shown in figure 3 below' shows that the model's prediction is true with the confidence for this decision at 50% with the duration (mins) as the main attribute for this conclusion. However, 95% of predictions done by this model are correct and when the model says true, it covers 100% of the cases and its correct with 83.33% of all predictions for class true.

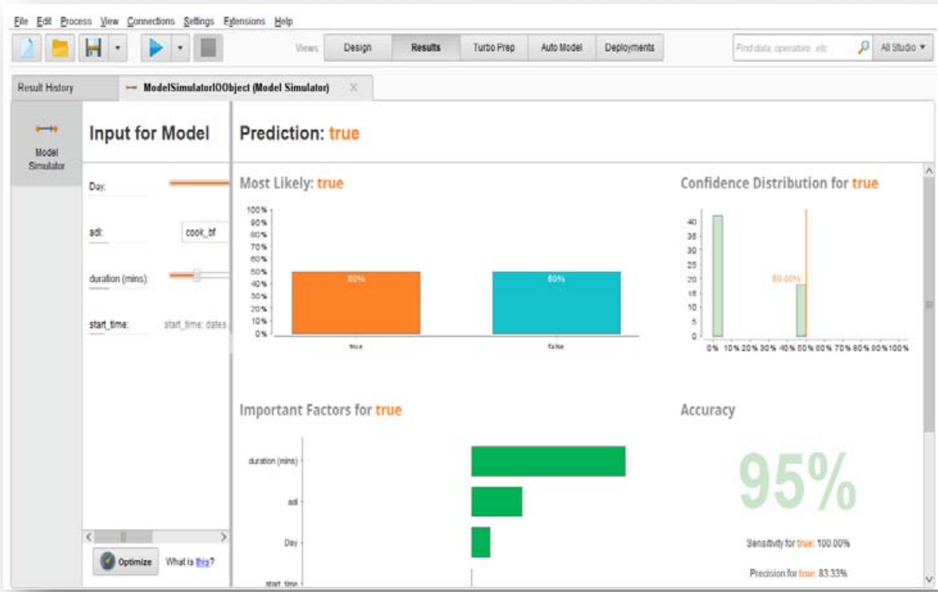

Figure 3: Result of the Ensemble Model Simulator

Performance evaluation was carried out on certain criteria, after building the model. This was to determine the level of efficiency of the models created and if they reflect the aim of the study. The results of these performance findings are shown in Table 1.

Table 1 Performance Evaluation between Naïve Bayes, K-Nearest Neighbour and Ensemble Learning Technique

S.NO.	Performance Criteria	Naïve Bayes Algorithm	K-Nearest Neighbour Algorithm	Ensemble Learning Algorithm
1.	Accuracy	98.33%	71.67%	95.00%
2.	Classification Error	1.67%	28.33%	5.00%
3.	Root_Mean_ Squared_Error (RMSE)	0.129 +/- 0.000	0.515 +/- 0.000	0.274 +/- 0.000
4.	Squared Error	0.017 +/- 0.128	0.266 +/- 0.428	0.075 +/- 0.115

The results from the comparison table above have shown that the accuracy performance criteria for the models was 98.33%, 71.6% and 95% for Naïve Bayes Algorithm, K-Nearest Neighbour Algorithm and Ensemble Algorithm respectively. It can be observed that, the accuracy for ensemble learning appears lower than that of Naïve Bayes

Algorithm, this is because the root_mean_squared_error (RMSE) which is the standard deviation of the predicted errors was recorded at 0.274 +/- 0.000 for ensemble technique and 0.129 +/- 0.000 for Naïve Bayes with squared error, which is the average of the squares of the errors; at 0.075 +/- 0.115 for ensemble technique and 0.017 +/- 0.128 for Naïve Bayes as shown in figure 4, 5 and 6 respectively. These errors which are inversely proportional to the rate of accuracy are the reason why the ensemble techniques accuracy is subtly lower than that of Naïve Bayes Algorithm.

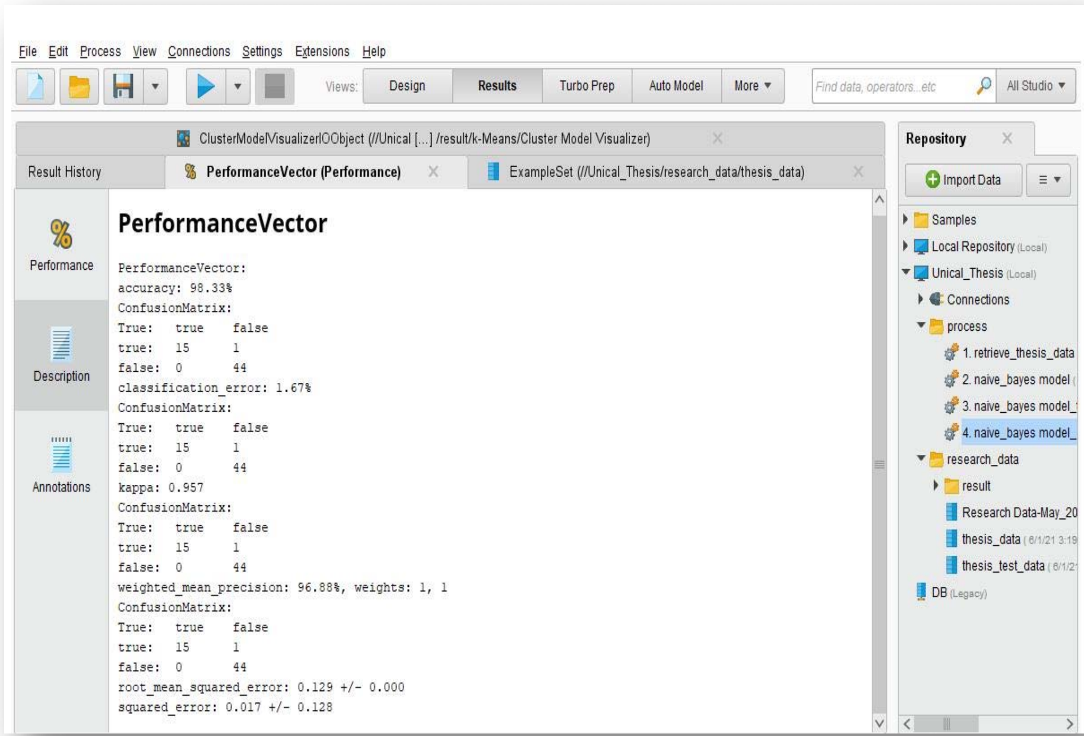

Figure 4 Description of Naïve Bayes Model Performance Evaluation

PerformanceVector

```
PerformanceVector:
accuracy: 71.67%
ConfusionMatrix:
True:   true    false
true:   0       2
false:  15      43
classification_error: 28.33%
ConfusionMatrix:
True:   true    false
true:   0       2
false:  15      43
kappa: -0.062
ConfusionMatrix:
True:   true    false
true:   0       2
false:  15      43
weighted_mean_precision: 37.07%, weights: 1, 1
ConfusionMatrix:
True:   true    false
true:   0       2
false:  15      43
root_mean_squared_error: 0.515 +/- 0.000
squared_error: 0.266 +/- 0.428
```

Figure 5 Description of the K-Nearest Neighbour Model Performance Evaluation

PerformanceVector

```
PerformanceVector:
accuracy: 95.00%
ConfusionMatrix:
True:   true    false
true:   15      3
false:  0       42
classification_error: 5.00%
ConfusionMatrix:
True:   true    false
true:   15      3
false:  0       42
kappa: 0.875
ConfusionMatrix:
True:   true    false
true:   15      3
false:  0       42
weighted_mean_precision: 91.67%, weights: 1, 1
ConfusionMatrix:
True:   true    false
true:   15      3
false:  0       42
root_mean_squared_error: 0.274 +/- 0.000
squared_error: 0.075 +/- 0.115
```

Figure 6 Description of the Ensemble Model Performance Evaluation

It also principally shows that ensemble learning is of a higher advantage to individual learning algorithm as k-nearest neighbour with an accuracy of 71.6% when ensembled with naïve Bayes had its accuracy rise to 95.00%/.

6. Conclusions

In conclusion, the results shown by Ensemble learning technique is advantageous over individual algorithms as the strength of the stronger algorithm augments the weakness of the other. For example, the strength of naïve bayes which is the stronger algorithm in the study with capacity of handling missing value within a dataset is seen and agrees with Reddy [16] where Bayesian networks model was used for anomaly detection and proved the model's suitability being that its average rate leads to increase in prediction accuracy, however, in this work Bayesian was used in combination with k-nearest neighbour as an ensemble learning technique, which improved k-nearest neighbour by increasing the model accuracy from 71.67% singly to a 95.00% as an ensemble learner.

The study result also agreed with Analytics Vidha,[20] that Naïve Bayes algorithm is best when dealing with multivariate dataset and dataset with missing values as it individually outperformed the ensemble learner with an accuracy of 98.33% against 95.00%. This result differed with smolyakov [17] conclusion of ensemble learning technique outperforming individual techniques, however, the reason for the variance may be as a result of disparity in the type of data used, as real time nominal data which consist of missing values, was used in this study while artificially generated continuous data which is usually complete was used by Smolyakov [17]. The study simply confirmed the possibility of exceptions with ensemble learning techniques underperforming than individual algorithm in the use case of nominal values with missing dataset.

References

[1] Edemekong, P.F., Boomgars, D.L., Sukumaran, S. (2019). Activities of Daily Living (ADLs).

[2] Akhlaghinia, M. J., Lotfi, A., Langensiepen, C., and Sherkat, N. (2008). "A fuzzy predictor model for the occupancy prediction of an intelligent inhabited environment," in 2008 IEEE 16thInternational Conference on Fuzzy Systems (FUZZ-IEEE), 2008, pp.939-946.

[3] Hagras, H. (2007). "Embedding computational intelligence in pervasive spaces,"(2007) IEEE Pervasive Computing, vol.6, pp. 85-9, 07.2007.

[4] Helal, A., King, J., Bose, and R. (2009). "Assistive Environments for Successful Aging," in Advanced Intelligent Environments, Achilles D. Kameas, Victor

Callagan, Hani Hagras, Michael Weber and Wolfgang Minker (Eds.), Springer US, 2009.

[5] Sernal, A., Pigot, H., and Rialle, V. (2007). "Modeling the progression of Alzheimer's disease for cognitive assistance in smart homes," User Modeling and User-Adapted Interaction, vol. 17, pp. 415-38, 2007.

[6] Medjahed, H., Istrate,D., Boudy and Dorizzi, B. (2009) "A fuzzy logic system for home elderly people monitoring (EMUTEM)", in 10th WSEAS International Conference on Fuzzy Systems (FS'09), 2009, pp. 69-75.

[7] Cook, D.J. (2007) "Making sense of sensor data," IEEE Pervasive Computing, vol.6, pp.105-8, 04. 2007.

[8] CASAS (2009). http://casas.wsu.edu/datasets/

[9] McLean, A. (2011). Ethical frontiers of ICT and older users: cultural, pragmatic and ethical issues. Journal of Ethics and Information Technology 13: 313-326.

[10] Menschner, P., Prinz, A., Koene, P., Kobler, F., Altmann, M., Krcmar, H., Leimeister, J. (2011). Reaching into patients' homes-Participatory designed AAL services-The case of a patient centered nutrition tracking service. Electronic Markets 63-76.

[11] Ackerman, M. J. (2009). The smart home. J Med Pract Manage 25: 68-69.

[12] Morris, M. E. (2012) Preventing falls in older people. BMJ 345: e4919.

[13] Dementia across Australia 2011-2050 (2011). Deloitte Access Economics for Alzheimer's Australia.

[14] Caring for Older Australians (2011) Australia Productivity Commission Inquiry Report. Melbourne.

[15] Agoulmine, N., Deen, M.J., Jeong-Soo, L., Meyyappan, M. (2011). U-health smart home: Innovative solutions for the management of the elderly and chronic diseases. IEEE Nanotechnology Magazine 5: 6-11.

[16] Reddy, G. R. S., Charkravarthy, B. D., Neelima, C. R., Zion, G. D. (2018). Mining Frequent Patterns and Associations from the Smart meters using Bayesian Networks. International Journal Advanced Networking and Applications Volume:09 Issue:06, Pages: 3632-3639. ISSN: 0975-0290.

[17] Smolyakov, D., Sviridenko, N., Ishimtsev, V., Burikov, E. and Burnaev, E. (2019). Learning Esembles of Anomaly Detectors on Synthetic Data.

[18] Giannoni, F., Mancini, M., Marinelli, F., (2018). Anomaly Detection Models for IoT Time-Series

[19] Muthama, M. S., Mwangi, W., Calvin, O. (2018). Ensemble Network Intrusion Detection Model Based on Classification & Clustering for Dynamic Environment. International Journal of Engineering Research & Technology (IJERT). ISSN: N2278-0181 Vol.7 Issue 02, February-2018.

[20] Infographic-leaning Data science plan for 2017:https://www.analyticsvidhya.com

IJCSIS Call For Papers 2021-2022

https://sites.google.com/site/ijcsis/

The topics suggested by the journal can be discussed in term of concepts, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete ***unpublished papers***, which are ***not under review in any other conference or journal*** in the following, but not limited to, topic areas. All tracks are open to both research and industry contributions.

Ad Hoc & Sensor Network	**Knowledge based systems**
Ad hoc networks for pervasive communications	**Knowledge management**
Adaptive, autonomic and context-aware computing	**Location Based Services**
Advanced Computing Architectures and New Programming Models	**Management information systems**
Agent-based middleware	**Medical imaging**
Autonomic and self-managing middleware	**Micro/nano technology**
B2B and B2C management	**Middleware Issues**
BioInformatics	**Middleware services and agent technologies**
Bio-Medicine	**Mobile and Wireless Networks**
Biotechnology	**Mobile Computing and Applications**
Broadband and intelligent networks	**Mobile networks and services**
Broadband wireless technologies	**Multimedia Communications**
Cloud Computing and Applications	**Multimodal sensing and context for pervasive applications**
Collaborative applications	**Multisensor fusion**
Communication architectures for pervasive computing	**Natural Language Processing**
Communication systems	**Network management and services**
Computational intelligence	**Network Modeling and Simulation**
Computer and microprocessor-based control	**Network Performance; Protocols; Sensors**
Computer Architecture and Embedded Systems	**Networking theory and technologies**
Computer Business	**Neural Networks**
Computer Vision	**Neuro-Fuzzy and applications**
Computer-based information systems in health care	**Open Models and Architectures**
Computing Ethics	**Open Source Tools**
Context-awareness and middleware	**Operations research**
Cross-layer design and Physical layer based issue	**Optical Networks**
Cryptography	**Pattern Recognition**
Data Base Management	**Peer to Peer and Overlay Networks**
Data Mining	**Perception and semantic interpretation**
Data Retrieval	**Pervasive Computing**
Decision making	**Performance optimization**
Digital Economy and Digital Divide	**Positioning and tracking technologies**
Digital signal processing theory	**Programming paradigms for pervasive systems**
Distributed Sensor Networks	**Quality of Service and Quality of Experience**
E-Business	**Real-time computer control**
E-Commerce	**Real-time information systems**
E-Government	**Real-time multimedia signal processing**
Emerging signal processing areas	**Reconfigurable, adaptable, and reflective middleware approaches**
Enabling technologies for pervasive systems (e.g., wireless BAN, PAN)	**Remote Sensing**
Encryption	**RFID and sensor network applications**
Energy-efficient and green pervasive computing	**Scalability of middleware**
Event-based, publish/subscribe, and message-oriented middleware	**Security and risk management**
Evolutionary computing and intelligent systems	**Security middleware**

ISSN 1947-5500

Expert approaches **Fuzzy algorithms** **Fuzzy logics** **GPS and location-based applications** **Green Computing** **Grid Networking** **Healthcare Management Information Technology** **Human Computer Interaction (HCI)** **Image analysis and processing** **Image and multidimensional signal processing** **Image and Multimedia applications** **Industrial applications of neural networks** **Information and data security** **Information indexing and retrieval** **Information Management** **Information processing** **Information systems and applications** **Information Technology and their application** **Instrumentation electronics** **Intelligent Control System** **Intelligent sensors and actuators** **Internet applications and performances** **Internet Services and Applications** **Internet Technologies, Infrastructure, Services & Applications** **Interworking architecture and interoperability**	**Security, Privacy and Trust** **Security Systems and Technolgies** **Sensor array and multi-channel processing** **Sensor fusion** **Sensors and RFID in pervasive systems** **Service oriented middleware** **Signal Control System** **Signal processing** **Smart devices and intelligent environments** **Smart home applications** **Social Networks and Online Communities** **Software Engineering** **Software engineering techniques for middleware** **Speech interface; Speech processing** **Supply Chain Management** **System security and security technologies** **Technology in Education** **Theoretical Computer Science** **Transportation information** **Trust, security and privacy issues in pervasive systems** **Ubiquitous and pervasive applications** **Ubiquitous Networks** **User interfaces and interaction models** **Virtual reality** **Vision-based applications** **Web Technologies** **Wired/Wireless Sensor** **Wireless technology**