A STRIDE CONCERNING EFFECTIVE RETAINMENT OF PRIVATENESS IN ASTUTE HOMES

[1]Manikya Sinha, [2]Salil Batra
[1]M. Tech Student, [2]Assistant Professor
[1,2]School Of Computer Science& Engineering
[1,2]Lovely Professional University, Phagwara, India

Abstract: Astute homes have reached a widespread popularity since the past decade. IOT, Data analytics, Cloud Computing are key technologies for astute homes. Numerous Astute home devices have been created till date. From the perspective of security, user's privacy has been a critical point of study. In several studies, it was analyzed that the classification algorithms have more accuracy in predicting privacy issues when combined with clustering algorithms for robustness. In the proposed approach K-means and Decision Tree classifier are used to provide the security of the user by detecting the activity of the user and intruder. The approach has achieved 94 % accuracy, precision and recall respectively.

Index Terms – Data Analytics, Machine Learning, K-means, Decision Tree Classifier, Smart Homes

I. INTRODUCTION

The idea for astute homes started in early 1900s when first independent electric or gas-controlled apparatuses were launched. Till date many astute home devices have been developed to ease our life. There are three generations described for astute home technologies [62] are wireless technology, artificial intelligence and robotics. Discovery of Internet of Things, Cloud Computing and Data analytics has revolutionized astute homes.

A. Internet-Of-Things (IOT)

An embedded network of smart devices focusing on providing collection and exchange of data wirelessly. IOT empowers today's generation like it incorporates smart homes, cloud computing etc. Gartner hype cycle recognize IOT as top emerging technologies [2]. IOT's architecture shown below [2]:

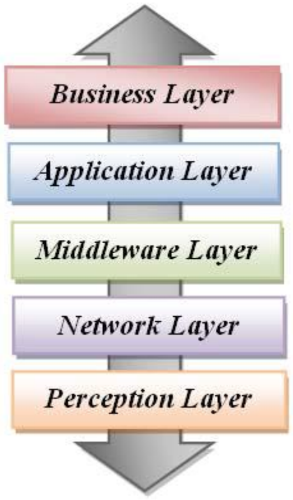


Figure 1: IOT's Generic Architecture

This architecture shows different layers like in OSI/TCP-IP model for Internet. Perception layer helps in gathering information. Network layer targets CIA. Middleware layer deal with services and low-level layer. Application layer manages application over the response of middleware. Business layer deals with user interaction.

B. Cloud Computing

The word "cloud computing" appeared first time in 1996. It's a technology that enables us to access system resources and services from any location through internet. Cloud provides many features such as device and location independency, 24 x 7 support, pay as you use, lower TCO, reliability, scalability and sustainability, secure storage management, easy and agile deployment, utility-based approach, highly automated, freeze internal resources and lower capital expenditure. Factors that make cloud computing as a compelling paradigm are: Advancement in processors, Virtualization, Distributed Storage, Automated Management and Broadband Internet Access. Cloud Computing has been decomposed into 3 models that are [4]:

1. Public Cloud

It's a global cloud that is generic in nature and is available for any user. This cloud is inexpensive in nature and all the hardware and software facilities are provided by the cloud vendor to the user.

2. Community Cloud

A cloud that is specific to any organizational work that is many organization’s combining together to achieve a common goal. This cloud helps in providing a place for companies to share their resources easily. This cloud is also called as Corporate Cloud. The ownership to cloud comes into picture when this model is used.

3. Hybrid Cloud

This cloud is basically a combination of public and community cloud this cloud is used by organizations where an organization wants to provides its features to the end users for free and also at the same time to maintain the security and privacy of the data. A figure shows different service models that are provided by cloud computing below [5]:

Table 1: Cloud Computing models comparison

CRITERIA	**SaaS**	**Paas**	**IaaS**
Consumer	End user	Application owner	Application owner
Service Type	Complicated application	Runtime scenario, Cloud Storage, Integration etc.	Cloud storage, Virtual server
Service level coverage	Application uptime, Application performance	Environment availability, Environment performance, no application coverage	Virtual server availability, time to provision, no platform or application coverage
Examples	Collaboratives, ERP	Application development, Decision support, Web, Streaming	Caching, Security, Legacy, System management

C. Data Analytics

It’s a process of inspecting, cleansing, transforming, and modeling the data. It is used for finding information, specific conclusions and decision-making data. It has multiple applications in different aspect of fields. Multiple tools have been developed for its purposes. Top 10 tools are R Programming, Tableau Public, Python, SAS, Apache Spark, Excel Rapid Miner, KNIME, Qlik View, Splunk [64].

D. Astute Homes

Also called as smart homes. It’s a branch of ubiquitous computing. It involves comfort, healthcare, safety, security and energy conservation for user. It uses telecommunication and web technologies like Internet to provide remote home control. A figure shown below [63]:

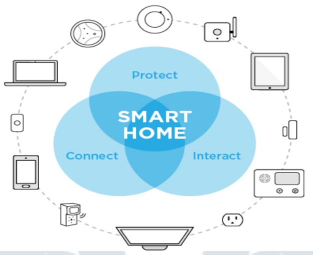


Figure 2: Smart home or Astute home Overview

IOT and Cloud Computing are the key technologies that power the Astute Home Automation. Challenges in Astute homes are Consumer Behavior, Lack of consumer awareness, High Implementation Costs, Complicated Installation, Data security and privacy, State regulation, Reliability and Comfortably, Power Consumption, Networking, Consumer Interaction, Communication.

II. Related Work

In previous studies many researches are done for security in astute homes. For instance, [58] created a cascaded algorithm that uses K–Means and C4.5 in it. The performance analysis for this model was measured by using five measures that are True Positive Rate, False Positive Rate, Precision, Accuracy and F–Measures Algorithm gives impressive detection accuracy in the experiment results. In another study [30], allocation prediction model based on k-means is developed where k-means clusters all the history location and then the time-matching method was used to calculate the history location in different time segments. In this way, a prediction trajectory is formed to predict object location. Another study [32], suggest a two-stage approach for the activity recognition of a single resident. For verifying the model, a set of naive models were used for comparison and tested to recognize set of daily activities. The approach demonstrated the superiority in predicting single-resident activities moreover k-means partitioned the set of activity records into two clusters and C4.5 is used to generate and train the model for activity recognition [32].

A new algorithm for detecting energy theft was used in a study. The algorithm relies on the predictability of customer’s normal and malicious usage patterns. The algorithm provides a high-performance even with a low-sampling rate which helps to preserve customer’s privacy [33]. Smartphone sensors are used for human activity recognition model. Four activities were considered still, walking, running, and vehicle and represented through some feature vectors. K-NN classier is used to recognize the activity performed by the user. Results showed the effectiveness of the implementation are in terms of accuracy and efficiency [35]. Another study suggested that better accuracy can be maintained while reducing false positives when model is composed of SVMs, decision trees and Naive Bayes. Firstly, SVM is trained to specify if the instance is an attack or normal traffic. Secondly attack traffic is routed through a decision tree for classification. Lastly Naive Bayes and the decision tree will then vote on any unclassified attacks [36]. A hybrid approach was created consisting of K-pattern clustering and neural network algorithm based on temporal relations. K-pattern clustering demonstrated its efficiency to group and identify the user's activities. The hybrid method of K-pattern and ANN is more accurate, extensible and adaptable in a dynamic environment such as an IoT network and is useful for smart home applications [40]. Another study uses a technique to detect and to subsequently reduce the number of IDS false alarms. A two-stage classification combined of SOM and k-means was developed. System has demonstrated its effectiveness in filtering all noisy and unnecessary IDS alerts which has usually contributed to more than 50% of false alarms from the common IDSs [42].

A comprehensive evaluation off 19 unsupervised bug detection algorithms on M10 datasets from different application domains has been performed for the first time in a study. The datasets have been made publicly available and therefore a foundation for a fair and standardized comparison for the community was introduced. Besides supporting the unsupervised bug detection research community, the study also suggests that the study and implementation is useful for researchers from neighboring fields [49]. In another study [50], a new approach for privacy preserving classification on vertically partitioned datasets in a multi-party scenario is used. The approach was applied to four standard machine learning datasets and results were observed. The findings show that the approach leads to similar classification performance as it would be achieved by applying traditional non-privacy preserving machine learning algorithms [50]. Present study an approach was performed for providing security in astute homes using K-means and decision tree classifier.

III. Methodology

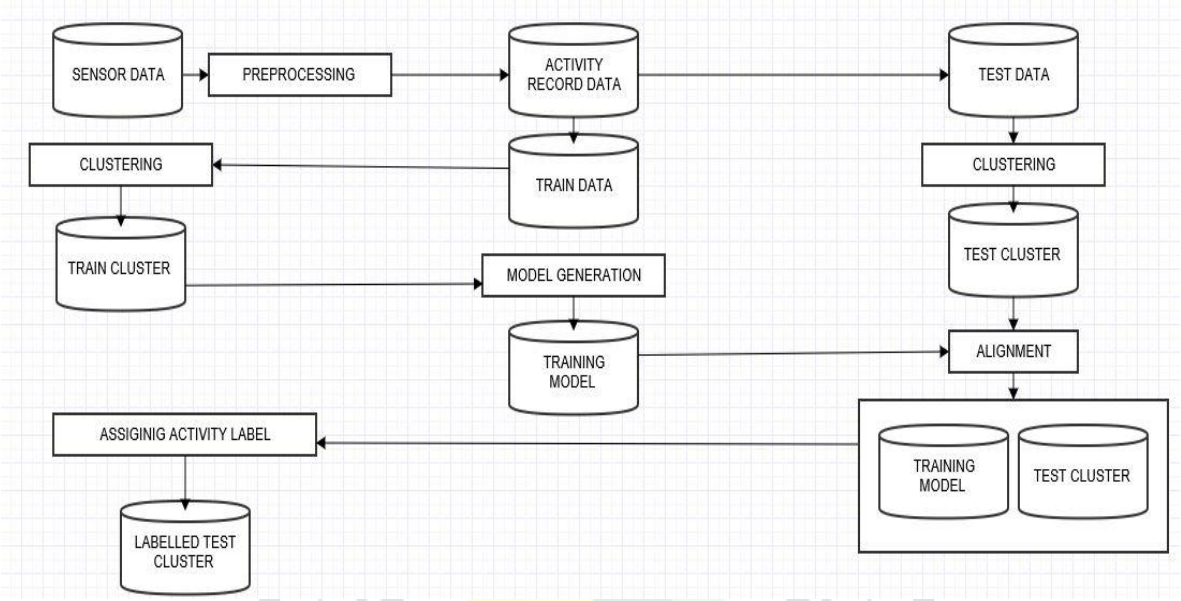


Figure 3: Showing Activity prediction steps

An approach is used for security in astute homes. This approach is used to detect the Invalid activity. As shown in figure 4, the approach is predicting an unknown activity also called as invalid activity after labeling test cluster. The approach consists of five steps preprocessing, clustering, model generation, alignment and assigning activity label. Step 1 is preprocessing, in this phase activity record data is extracted from the sensor data. Step 2 is clustering, where train and test cluster is extracted from train and test data which are divided in the ratio 70:30, here K-means is used for clustering. Step 3 is model generation where models are generated using train cluster here decision tree classifier is used for model generation. Step 4 is alignment where the test cluster is passed to the model and trained model is aligned with the test cluster. Step 5 is important where each aligned test clusters are assigned with the activity labels. K-means and Decision tree classifier are available in scikit-learn. The approach is performed on python. For predicting value of K elbow method is used.

A. K-means

A classic and unsupervised learning algorithm which is used to solve the clustering problem. The procedure follows a simple and easy way to cluster a given dataset through a certain number of clusters. First value of k is predicted by elbow method where k is number of clusters. After that number of centroids are initialized according to the number of cluster. Final step is a convergence step where firstly euclidian distances are calculated for every data point from each centroid and then the data points are assigned to the centroid with minimum euclidian distance until all the data points are assigned to the anyone centroid and no more changes happens [30].

B. Decision tree classifier

It is used for multi-label classification or binary classification and processes the data based upon a classification criterion described below:

1. *Classification Criteria*

If a target is a classification outcome taking on values 0, 1, K-1, for node representing a region with observations, let be the proportion of class k observations in node.

Common measures of impurity are Gini [64]:

$$p_{mk} = 1/N_m \sum_{x_i \in R_m} I(y_i = k)$$

be the proportion of class k observations in node m
Common measures of impurity are Gini :

$$H(X_m) = \sum_k p_{mk}(1 - p_{mk})$$

, Cross-Entropy:

$$H(X_m) = -\sum_k p_{mk} \log(p_{mk})$$

and Misclassification:

$$H(X_m) = 1 - \max(p_{mk})$$

where X_m is the training data in node m

C. Elbow Method

This method is used to find the number of clusters for the K-means. This method consists of n task. First task is defining the range for the value of k for example between 1 to 5. After setting ranges it will run a loop within that range and for every time when loop runs it will execute the k-means and calculate within sum of squares for each execution it will be second task. In task third will be to plot the set of within sum of squares that were calculated for each execution. Fourth task will be to visualize the curve the point where the line tends to bent or curve is the exact value for k. Example shown below figure 4:

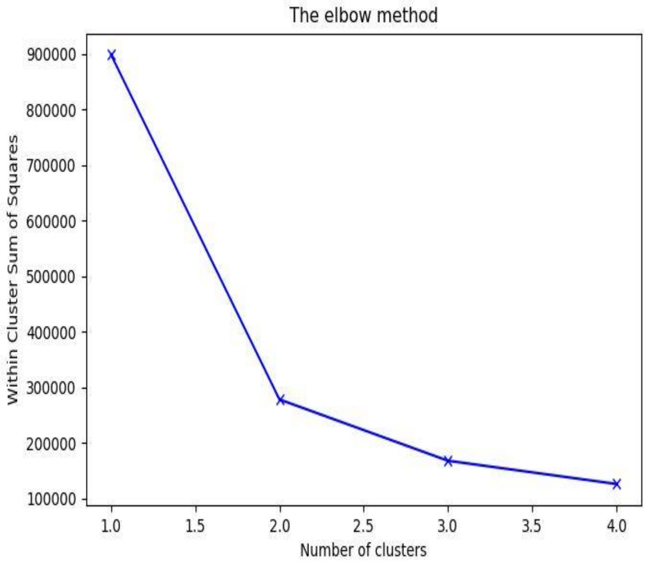
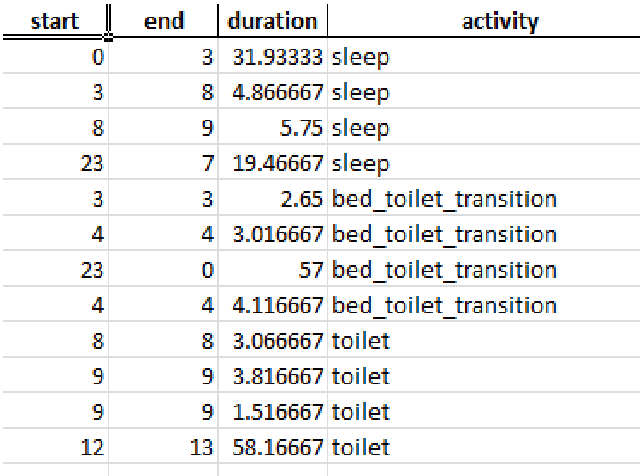


Figure 4: Showing curve for elbow method

From figure 4 it can infer that when k is 2 and the first point of curve can be seen therefore the chosen for the approach will be 2.

IV. RESULTS AND DISCUSSION

Dataset HH102 was taken from [32]. The features in the dataset are date, time, sensor name, and sensor value and activity name. After preprocessing and reconstructing the extracted features are start hour, end hour, duration (in minutes) and activity label that are included in newly dataset called as activity record. Activity record data consist of 22 activities in which the 21 activities are valid whereas 1 activity was invalid Bathe, Bed toilet transition, Cook, Dress, Eat, Enter home, Entertain guests, Groom, Hygiene, Leave home, Morning meds, Read, Relax, Sleep, Sleep out of bed, Take medicine, Toilet, Wash dishes, Watch TV, Work, Work at table are valid activities. Laptop Use is invalid activity.



start	end	duration	activity
0	3	31.93333	sleep
3	8	4.866667	sleep
8	9	5.75	sleep
23	7	19.46667	sleep
3	3	2.65	bed_toilet_transition
4	4	3.016667	bed_toilet_transition
23	0	57	bed_toilet_transition
4	4	4.116667	bed_toilet_transition
8	8	3.066667	toilet
9	9	3.816667	toilet
9	9	1.516667	toilet
12	13	58.16667	toilet

Figure 5: Showing data set sample

Figure 5 shows data set sample lookup. The features start, end defines the start and ending of activity are in hours whereas the duration shows for how long activity is being performed and is in minutes, activity defines the labels for activity.

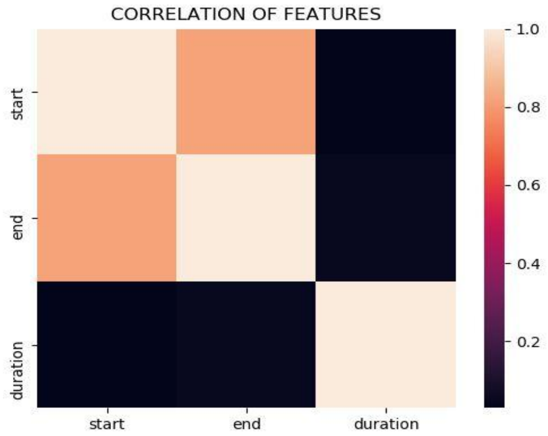


Figure 6: Showing Correlation of features

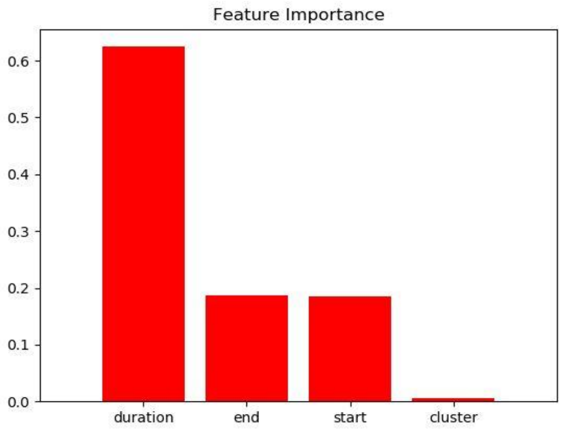


Figure 7: Showing feature importance after classification

Figure 6 shows the correlation between different features in the activity record data, it can be seen that all the features are properly correlated to itself. From figure 7 it can be seen that the duration has higher importance than other features after clustering.

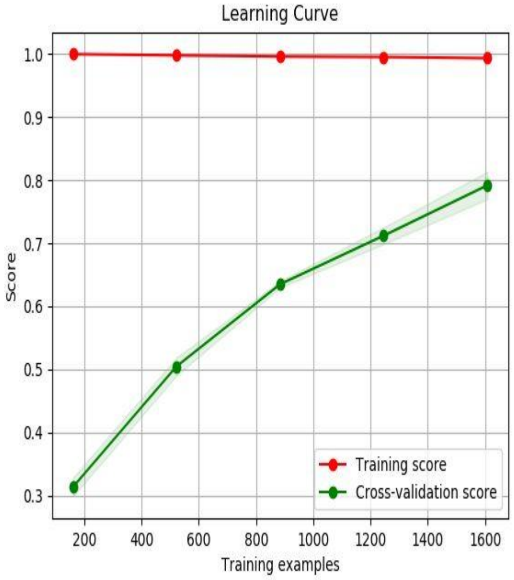
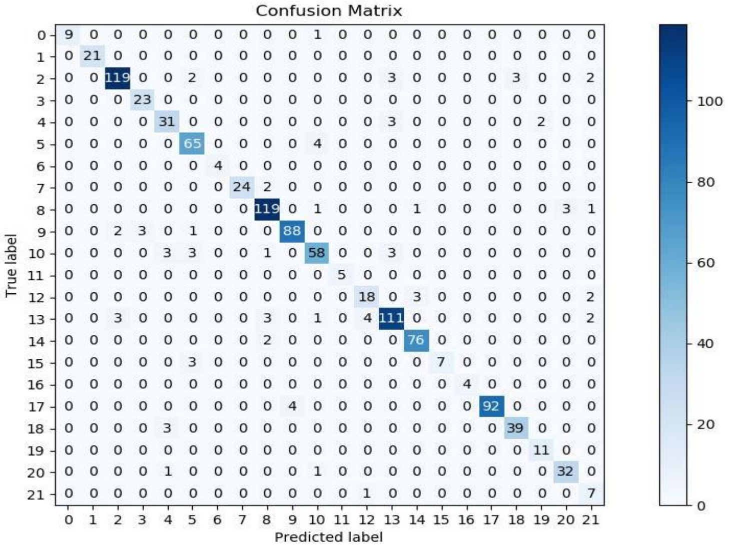


Figure 8: Showing training score



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
0	9	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	119	0	0	2	0	0	0	0	0	0	0	3	0	0	0	0	3	0	0	2
3	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	31	0	0	0	0	0	0	0	0	3	0	0	0	0	0	2	0	0
5	0	0	0	0	0	65	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	24	2	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	119	0	1	0	0	0	1	0	0	0	0	0	3	1
9	0	0	2	3	0	1	0	0	0	88	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	3	3	0	0	1	0	58	0	0	3	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	18	0	3	0	0	0	0	0	0	2
13	0	0	3	0	0	0	0	0	3	0	1	0	4	111	0	0	0	0	0	0	0	2
14	0	0	0	0	0	0	0	0	2	0	0	0	0	0	76	0	0	0	0	0	0	0
15	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	92	0	0	0	0
18	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	39	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0
20	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	32	0
21	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	7

Figure 9: Confusion Matrix

Figure 8 shows the training score is very high at the beginning and is constant and the cross-validation score is very low at the beginning and increases which means that the more classifier trained the better result will be achieved and also size of data is also an important aspect. Figure 9 shows confuse matrix in diagonal elements predicts the number of activities that are being predicted correctly for example, the invalid activity Laptop Use is encode as label 9 which can be seen in the figure that has been predicted 88 times correctly.

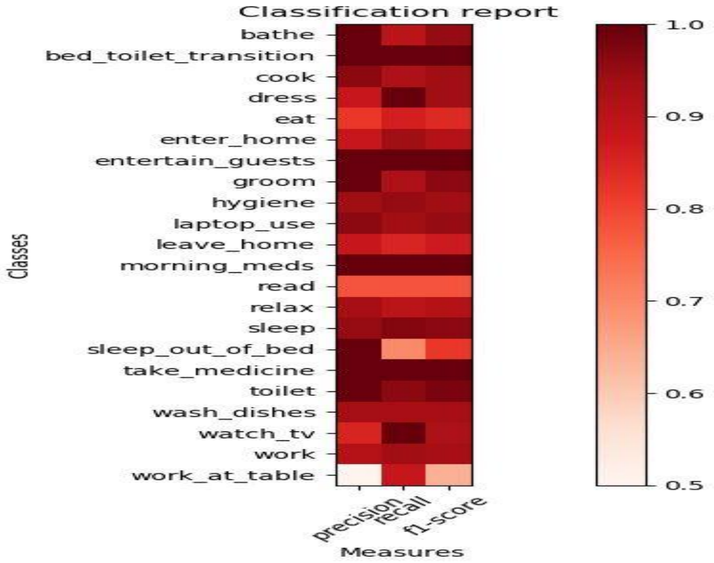


Figure 10: Showing classification report of used classifier

Figure 10 is a classification report of the decision tree classifier in which precision, recall and f1-score of each activity is represented properly for example it can be infer that activity 'bathe' is predicted more often whereas 'work_at_table' is predicted less.

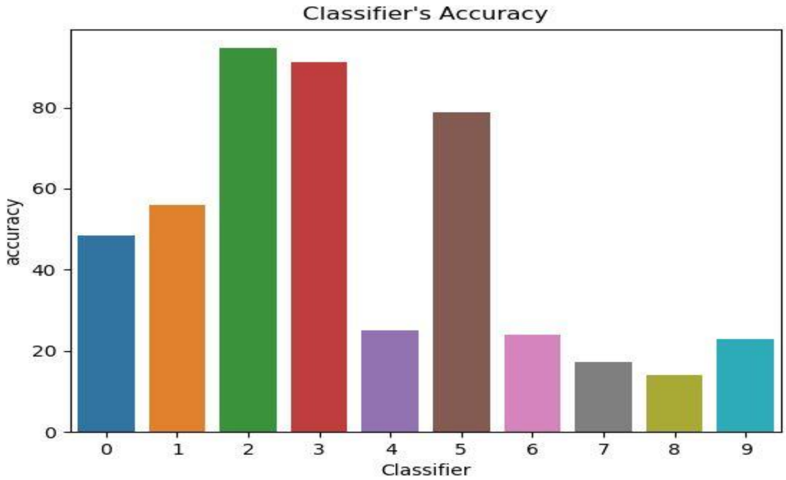


Figure 11: Showing Classifier Accuracy Comparison

Figure 11 shows that the accuracy of decision tree classifier is more than other classifiers. List of classifiers compared are KNN, SVC, Decision Tree Classifier, Random Forest Classifier, Ada Boosting Classifier, Gradient Boosting Classifier, Gaussian NB, Multinomial NB, Bernoulli NB and Linear Discriminant Analysis encoded as 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9. The accuracy of the classifier achieved is 94%

V. Conclusion and Future Work

It can be seen that Decision Tree Classifier having more accuracy than other classifiers. In detecting the invalid activity with 94 % accuracy, precision and recall respectively. It can also be seen that the approach is generic hence this approach can also be applied in other field for other applications for example terrorist analysis attack [31] or helping elderly people for carefree living [32]. The data set used is for single resident. In future the work will be extended to multi-resident dataset.

VI. Acknowledgement

I would like to thank my guide Mr. Salil Batra, School of Computer Science, Lovely Professional University, Phagwara for their guidance in completing my research work.

References

[1] L. G. Rios and J. A. I. Diguez, "Big Data Infrastructure for analyzing data generated by Wireless Sensor Networks," 2014 IEEE Int. Congr. Big Data, pp. 816–823, 2014.
[2] S. G. J, S. Jose, R. B. Raman, and P. Samuel, "Performance Analysis of Big Data Gathering in Wireless Sensor Network Using an EM Based Clustering Scheme," 2015 Fifth Int. Conf. Adv. Comput. Commun., vol. i, pp. 109–113, 2015.
[3] S. Rani, S. H. Ahmed, I. Member, R. Talwar, and J. Malhotra, "Can Sensors Collect Big Data ? An Energy Efficient Big Data Gathering Algorithm for WIRELESS SENSOR NETWORK," vol. XX, no. X, 2017.
[4] D. P. and K. Ahmed, "A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools," Int. J. Adv. Comput. Sci. Appl., vol. 7, no. 2, pp. 511–518, 2016.
[5] P. C. Neves, B. Schmerl, J. Cámara, and J. Bernardino, "Big data in cloud computing: Features and issues," Int. Conf. Internet Things Big Data, IoTBD 2016, pp. 307–314, 2016.
[6] Y. Ma, Y. Wang, J. Yang, Y. Miao, and W. Li, "Big Health Application System based on Health Internet of Things and Big Data," IEEE Access, vol. XX, no. c, pp. 1–1, 2016.
[7] B. Ragothaman, S. Prabha, E. Jose, and B. Sarojini, "A Survey on Big Data and Internet of
Things," Avinashilingam - UGC Spons. Two Day Natl. Conf. Internet Things, pp. 174–179, 2016.
[8] B. R. Chang, H. Tsai, Y. Chang, and C. Huang, "Multiple Big Data Processing Platforms," no. 3, pp. 207–211, 2016.
[9] S. Kraijak and P. Tuwanut, "a Survey on Iot Architectures , Protocols , Applications , Security , Privacy , Real-World."
[10] T. Reichherzer, A. Mishra, E. Kalaimannan, and N. Wilde, "A Case Study on the Trade-Offs between Security , Scalability , and Efficiency in Smart Home Sensor Networks," pp. 222–225, 2016.
[11] J. Bugeja, A. Jacobsson, and P. Davidsson, "On Privacy and Security Challenges in Smart Connected Homes," 2016.
[12] L. M. L. de Campos, R. C. L. de Oliveira, and M. Roisenberg, "Network Intrusion Detection System Using Data Mining," 2012, vol. 311, pp. 104–113.
[13] X. Wu, X. Zhu, G. Wu, and W. Ding, "Data Mining With Big Data" no. Ibm 2012, 2013.17
[14] H. Yuan, J. Wang, J. Liu, and S. Li, "Research of Zigbee and Big Data Analysis based Pulse Monitoring System for Efficient Physical Training 2 Zigbee based Pulse Monitoring and Analysis System," vol. 80, no. December, pp. 2357–2361, 2016.
[15] S. V Halde, "Efficient Collection of Big data in WIRELESS SENSOR NETWORK."
[16] A. Chakravorty, T. Wlodarczyk, and C. Rong, "Privacy Preserving Data Analytics for Smart Homes," 2013 IEEE Secur. Priv. Work., pp. 23–27, 2013.
[17] A. Jacobsson and P. Davidsson, "Towards a model of privacy and security for smart homes," IEEE World Forum Internet Things, WF-IoT 2015 - Proc., pp. 727–732, 2016.
[18] D. Pishva, "Internet of Things : Security and Privacy Issues and Possible Solution," ICACT Trans. Adv. Commun. Technol., vol. 5, no. 2, pp. 797–808, 2016.
[19] M. Sain, Y. J. Kang, and H. J. Lee, "Survey on Security in Internet of things : state of the art and challenges," pp. 699–704, 2017.

[20] V. Sivaraman, H. H. Gharakheili, A. Vishwanath, R. Boreli, and O. Mehani, "Network-level security and privacy control for smart-home IoT devices," 2015 IEEE 11th Int. Conf. Wirel. Mob. Comput. Netw. Commun. WiMob 2015, pp. 163–167, 2015.
[21] S. Jaiswal and D. Gupta, "Security Requirements for Internet of Things (IoT)," pp. 419–427.
[22] M. R. Alam, S. Member, M. B. I. Reaz, and M. A. M. Ali, "A Review of Smart Homes – Past , Present , and Future," no. June 2015, 2012.
[23] S. Latif, A. Shabani, A. Esser, and A. Martkovich, "Analaytics of Residential Electrical Energy Profile," no. December 2006, pp. 3–6, 2017.
[24] Y. Lee, W. Hsiao, Y. Lin, and S. T. Chou, "Privacy-Preserving Data Analytics in Cloud-Based Smart Home with Community Hierarchy," vol. 63, no. 2, pp. 200–207, 2017.
[25] C. Perera, C. Mccormick, A. K. Bandara, B. A. Price, and B. Nuseibeh, "Privacy-by-Design Framework for Assessing Internet of Things Applications and Platforms," 2016.
[26] Y. Liu, G. Zhang, W. Chen, and X. Wang, "An Efficient Privacy Protection Solution for Smart Home Application Platform," pp. 2281–2285, 2016.
[27] Z. Guan, G. Si, X. Du, P. Liu, Z. Zhang, and Z. Zhou, "Protecting User Privacy Based on Secret Sharing with Fault Tolerance for Big Data in Smart Grid," 2017.
[28] T. Song et al., "A Privacy Preserving Communication Protocol for IoT Applications in Smart Homes," vol. 4662, no. c, 2017.
[29] K. Kaur, A. Syed, A. Mohammad, and M. N. Halgamuge, "Review : An Evaluation of Major Threats in Cloud Computing Associated with Big Data," pp. 368–372, 2017.
[30] Y. Hu, X. Zhu, and G. Ma, "Location Prediction Model Based on K-means Algorithm," pp. 681–687, 2018.
[31] P. Gupta, A. S. Sabitha, and T. Choudhury, "Terrorist Attacks Analysis Using Clustering Algorithm," pp. 317–328.
[32] Y. Liu, D. Ouyang, Y. Liu, and R. Chen, "A novel approach based on time cluster for activity recognition of daily living in smart homes," Symmetry (Basel)., vol. 9, no. 10, 2017.
[33] P. Jokar, N. Arianpoo, and V. C. M. Leung, "Electricity theft detection in AMI using customers' consumption patterns," IEEE Trans. Smart Grid, vol. 7, no. 1, pp. 216–226, 2016.
[34] J. Dahmen, D. J. Cook, X. Wang, and W. Honglei, "Smart secure homes: a survey of smart home technologies that sense, assess, and respond to security threats," J. Reliab. Intell. Environ., vol. 3, no. 2, pp. 83–98, 2017.
[35] F. Concone, S. Gaglio, G. Lo Re, and M. M. B, "AI*IA 2017 Advances in Artificial Intelligence," vol. 10640, pp. 58–71, 2017.
[36] K. Goeschel, "Reducing False Positives in Intrusion Detection Systems Using Data Mining Techniques Utilizing Support Vector Machines , Decision Trees , and Naive Bayes for Off-Line Analysisng Data-Mining Techniques Utilizing Support Vector Machines , Decision Trees , a," SoutheastCon 2016, pp. 1–6, 2016.
[37] C. W. Tsai, C. F. Lai, M. C. Chiang, and L. T. Yang, "Data mining for internet of things: A survey," IEEE Commun. Surv. Tutorials, vol. 16, no. 1, pp. 77–97, 2014.
[38] S. T. M. Bourobou and Y. Yoo, "User activity recognition in smart homes using pattern clustering applied to temporal ANN algorithm," Sensors (Switzerland), vol. 15, no. 5, pp. 11953–11971, 2015.
[39] S. Ganapathy, K. Kulothungan, S. Muthurajkumar, and M. Vijayalakshmi, "Intelligent feature selection and classification techniques for intrusion detection in networks : a survey," pp. 1–16, 2013.
[40] H. D. Kuna, R. García-Martinez, and F. R. Villatoro, "Outlier detection in audit logs for application systems," Inf. Syst., vol. 44, pp. 22–33, 2014.
[41] C. Guo, Y. Zhou, Y. Ping, Z. Zhang, G. Liu, and Y. Yang, "A distance sum-based hybrid method for intrusion detection," 2013.
[42] G. C. Tjhai, S. M. Furnell, M. Papadaki, and N. L. Clarke, "A preliminary two-stage alarm correlation and filtering system using SOM neural network and K -means algorithm," Comput. Secur., vol. 29, no. 6, pp. 712–723, 2010.
[43] S. Of, M. Learning, C. For, and A. M. Botnet, "A Study Of Machine Learning Classifiers For Anomaly-Based Mobile Botnet Detection. pp 251-265," vol. 26, no. 4, pp. 251–265, 2013.
[44] Q. Song, "Assistant Detection of Skewed Data Streams Classification in Cloud Security," pp. 60–64, 2010.
[45] C. Defense, "Classification Algorithms in Permission Based," 2014.
[46] O. Banos, J. Galvez, M. Damas, and H. Pomares, "Window Size Impact in Human Activity Recognition," pp. 6474–6499, 2014.
[47] G. Suarez-tangil, J. E. Tapiador, P. Peris-lopez, and S. Pastrana, "Power-aware anomaly detection in smartphones : An analysis of on-platform versus externalized operation," Pervasive Mob. Comput., 2014.
[48] M. K. Nagle and S. K. Chaturvedi, "Feature Extraction Based Classification Technique for Intrusion Detection System," vol. 8, no. 2, pp. 23–38, 2013.
[49] M. Goldstein, M. Goldstein, and S. Uchida, "A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data," PLoS One, no. April, pp. 1–31, 2016.
[50] N. Schlitter, "Distributed privacy preserving classification based on local cluster identifiers," Proc. 11th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. Trust. - 11th IEEE Int. Conf. Ubiquitous Comput. Commun. IUCC-2012, pp. 1265–1272, 2012.
[51] D. Choujaa and N. Dulay, "Activity Recognition from Mobile Phone Data: State of the Art, Prospects and Open Problems," vol. V, p. 32, 2009.
[52] F. Oroumchian, "Clustering Scheme for Intrusion Detection ," pp. 553–558, 2005.
[53] E. Fotiadou and N. Nikolaidis, "Activity-based methods for person recognition in motion capture sequences," Pattern Recognit. Lett., vol. 49, pp. 48–54, 2014.
[54] K. Stroeh, E. Roberto, M. Madeira, and S. K. Goldenstein, "An approach to the correlation of security events based on machine learning techniques," no. 1, pp. 1–16, 2013.
[55] C. S. Tucker, "DETC2013-13155," pp. 1–13, 2018.
[56] S. Samarah, M. G. Al Zamil, A. F. Aleroud, M. Rawashdeh, M. F. Alhamid, and A. Alamri, "An Efficient Activity Recognition Framework: Toward Privacy-Sensitive Health Data Sensing," IEEE Access, vol. 5, no. c, pp. 3848–3859, 2017.
[57] S. Y. Yerima, S. Sezer, and I. Muttik, "Android Malware Detection Using Parallel Machine Learning Classifiers," 2014 Eighth Int. Conf. Next Gener. Mob. Apps, Serv. Technol., pp. 37–42, 2014.
[58] M. Solanki, "Intrusion Detection System by using K-Means clustering , C 4 . 5 , FNN , SVM classifier," vol. 3, no. December, 2014.

[59] A. Aravkin and M. Wolf, “Analytics for understanding customer behavior in the energy and utility industry,” IBM J. Res. Dev., vol. 60, no. 1, pp. 1–13, 2016.
[60] B. S. Sheykhahmadloo and S. Mehrnoosh, “Proposing A New Model to Improve Alert Detection in Intrusion Detection Systems,” vol. 5, no. 4, pp. 76–82, 2017.
[61] http://scikit-learn.org/stable/modules/tree.html#tree-algorithms-id3-c4-5-c5-0-and-cart
[62] http://www.sersc.org/journals/IJSH/vol10_no8_2016/18.pd
[63] https://www.protection1.com/smart-home/
[64] https://www.proschoolonline.com/blog/top-10-data-analytics-tools/