EasyChair Preprint

№ 9761

The Roles of Stakeholders in Internet of Things: a Theoretical Framework

Latifah Almalki, Amany Alnahdi and Tahani Albalawi

February 22, 2023

The Roles of Stakeholders in Internet of Things: A Theoretical Framework

Latifah S. Almalki
Department of Computer Science
King Abdulaziz University
Jeddah, Saudi Arabia
ldasagalmalki@stu.kau.edu.sa

Amany K. Alnahdi
Department of Computer Science
King Abdulaziz University
Jeddah, Saudi Arabia
akalnahdi@kau.edu.sa

Tahani F. Albalawi
Department of Computer Science
Imam Mohammad Ibn Saud Islamic University
Riyadh, Saudi Arabia
tfalbalawi@imamu.edu.sa

***Abstract*—Internet of Things (IoT) technology is one example of a contemporary technology. Since millions of devices are sending data to the cloud quickly and in large quantities, a vast array of data is produced. The IoT is a subset of big data. Control, management, and monitoring of the network on the IoT needs to be improved. This study presents a theoretical framework for enhancing the work of humans as stakeholders in the security IoT domain. It aims to identify who the stakeholders are and then use mining algorithms to provide knowledge to each stakeholder based on their role in the business. A questionnaire was distributed to collect data about stakeholders. The framework uses a semi-supervised approach. A human-driven feature selection concept was used to determine the interrelationship between stakeholders and the attributes in the dataset. The objective of the study is improved decision-making to secure against security incidents in the IoT. The questionnaires were sent to IoT experts, and the responses were used to determine the stakeholders in the field of security.**

***Index Terms*—Stakeholders, Internet of Things, Feature Selection, Soft Clustering, Semi-supervised, Theoretical Framework.**

I. Introduction

IoT data analysis is described as the process by which the data gathered from diverse IoT devices is analyzed to produce results in a consistent fashion that can be comprehended and utilized to increase efficiency and develop and enhance operations. Integration between IoT and artificial intelligence (AI e.g., machine learning algorithms) can be used to support businesses and assist in improving security. Many recent studies have focused on human factors as a company's line of defense. Articles [1]–[3] look at employees from a variety of perspectives to demonstrate their effectiveness and their performance within organizations by using data mining algorithms. However, a new trend points to the importance of using AI and machine learning to assist employees with cyber security [4]. Being granted opportunities for to research more intelligent solutions can help employees accomplish their work effectively and improve decision-making. The stakeholders in this work refer to the human constituents or workforce who are interested in being aware of all security events in the networks, such as in the security operation center (SOC). There are many challenges examined by another survey [5] e.g., lack of experience or skills of stakeholders, insufficient automation, and the integration of multiple tools. This paper aims to shed light on the topic of IoT Role-Based Analytics for Security Data, which can be defined as the process of analyzing network data by using mining techniques based on the roles of stakeholders in IoT platforms. Figure 1 illustrates the general idea of this study.

The main contributions in this work are:

- A survey of IoT experts and interested parties conducted to identify the key stakeholders in the IoT security space.
- A presentation of the theoretical framework for selecting features from massive amounts of data using semi-supervised concepts, and
- A discussion of a forward-looking vision for IoT security stakeholders.

The rest of this article is set out as follows. Section II provides the context for the analysis of IoT data. Section III provides the proposed framework. Section IV presents the details of stakeholders, including a vision for the future. Section V provides the conclusion of this study.

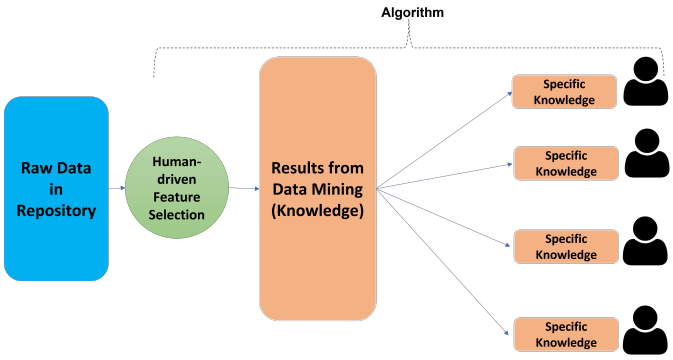

Fig. 1. Analysis data for stakeholders based on role in IoT environments

II. Background

Previously, the process of data analysis used to be challenging and expensive, but with the cheap cost of data storage and the technologies used to analyze and process data, many businesses, like Amazon, Microsoft, and Apple, have resorted to implementing analysis processes for IoT data. Massive volumes of data, up to terabytes per second, are produced

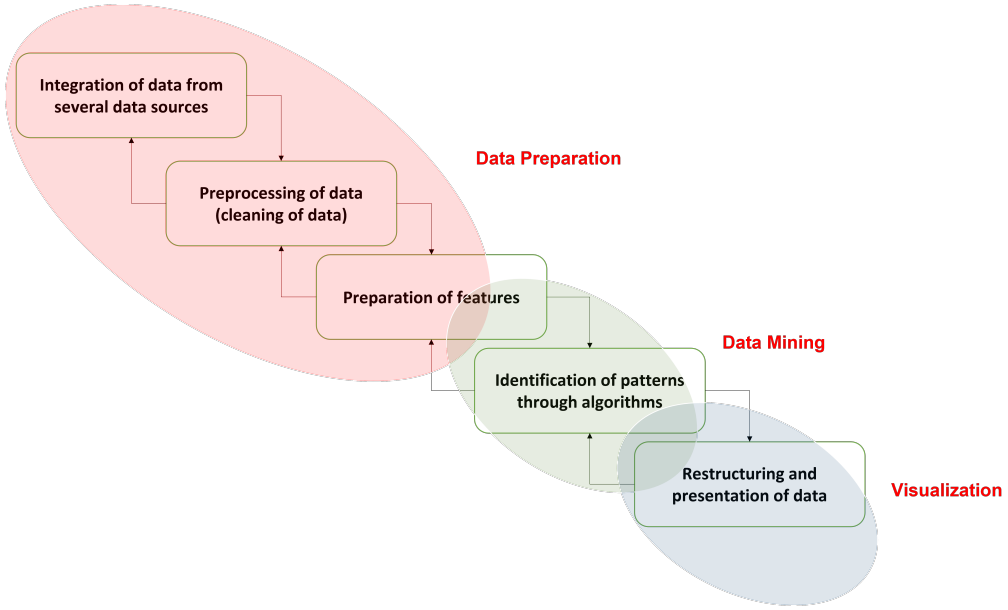

Fig. 2. The waterfall model showing the multiple stages involved in the process of data mining for extraction of knowledge

by devices. Because of this, the IoT can be viewed as a subset of big data. However, it differs from typical big data in that the data being sourced from a variety of incompatible devices means that the data are difficult to merge and require more effort to analyze. Several studies present improved big data analysis in the IoT domain. The authors in [6] and [7] surveyed the literature in the domain of big data analytics. The authors of [8] presented a system for analyzing IoT-captured home data. In [9], the researchers presented Stratum, a big data analytics solution delivered as a service. The authors in [10] examined one of the healthcare industry's IoT applications and the relationship between big data and analytical methods for this sector. In general, the size of the data makes it challenging for stakeholders to follow if it is not analyzed. The researchers in [11] summarized the challenges for data analytics in the next generation of IoT. For instance, data being produced continuously by smart devices. To address these problems and obstacles, embedded machine learning methods in IoT platforms would be beneficial. According to the research and articles that have been studied, a new intelligent solution in the IoT environment through data mining needs to be provided.

A. Overview of IoT Related Data Mining

Data mining is intended to extract hidden information or patterns by using a variety of technologies [12] e.g., classification, clustering, association rules, regression analysis, characteristics, etc. However, data mining in IoT data is more difficult than for traditional data. The reason depends on the differences between normal data mining and IoT data mining, such as tasks, goals, techniques employed, processes, computations, sources of data, and availability of resources [13] and also on the objectives for information that needs to be extracted. IoT data mining facilitates the transformation of enormous volumes of data collected from diverse surroundings into valuable insights. The waterfall model in Fig. 2 shows the general IoT data mining stages. Regular unstructured and impure raw data collected from infrastructure IoT includes sources such as smart home, healthcare, smart transport, industrial automation, etc. The next step, the preprocessing step, includes processes such as aggregation, feature selection, removing noise, normalization, reducing dimensions, etc. After data cleaning, the discovery pattern is calculated using an algorithm. The final step is the evaluation and representation of the results into understandable information that can be used to make decisions, or to implement automation or optimization.

The primary objective of IoT is to identify needs across interconnected networks. Different IoT devices can be connected to a network, enabling data to come in from various sources. Smart switches convey information about electricity use, smart thermostats transmit information about temperature variations, and smart watches report information about exercise activities. The volume of operations (read and write) on huge amounts of data collected from diverse sources makes it harder to select a suitable algorithm for any IoT environment.

B. Standards in IoT Environments

Standards organizations, for example, NTIA (National Telecommunications and Information Administration) or ISO (International Organization for Standardization), aim to describe the protocols, guidelines, regulations, and features that have been established and accepted by industry. The standards they establish provide the broad methods that contribute to efficient creation and management of systems. Services interact with and are integrated with many different entities, such as people, devices, and information resources [14]. Given that this is an emerging area, the IoT path is currently fraught with challenges. In particular, challenges exist with the complementarity of conventional systems [15], for instance,

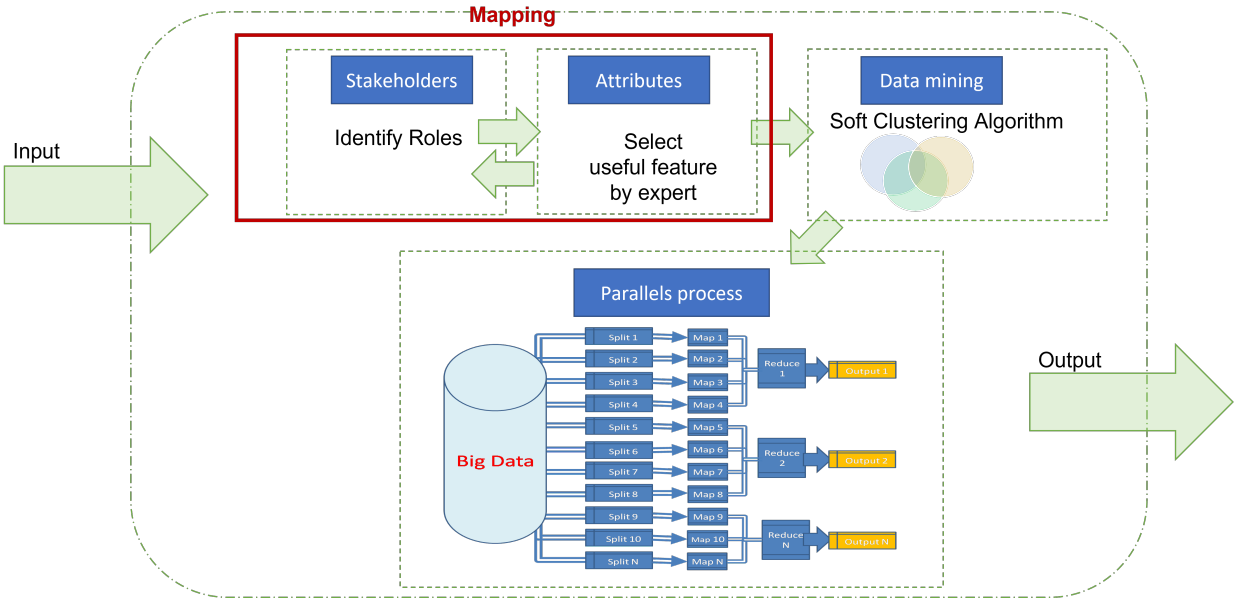

Fig. 3. The suggested framework for data analysis for stakeholders

the integration of SOCs (security operation centers) with IoT [16], and the integration of SIEM (security information and event management) with IoT [17]. This opens the way for the development of standards and the creation of new stakeholders. Section IV discusses potential future jobs for employees in the field of IoT security. Based on the results provide by the survey in [5], government, banking, finance, cybersecurity, technology, and telecommunication organizations are the ones most likely to have mandatory internal SOC use. It means that these organizations benefit the most from applying machine learning. The framework proposed to address this is given below.

III. FRAMEWORK

The proposed theoretical framework depends on a semi-supervised approach. The selection of appropriate relevant features requires an understanding of the feature domain, and this function is often designed for human experts who must be well-informed regarding the domain. Improvements in the clustering algorithm can be brought about by human-driven feature selection [18], which can help it distinguish between redundant and pertinent features. In specialized reference areas, the expert perspective is essential to fully comprehend the application context and role of various factors. Figure 3 illustrates the framework.

A. Mapping Between Roles and Features

The first step, defining roles by creating a table form, provides an overview of the domain for each stakeholder. There may be similar or shared duties. If a stakeholder holds any of the specified roles in a table, it will be assigned a group of features. Significant correlation between stakeholders and features can be found by an expert human. Figure 4 shows this process. Each stakeholder has many roles. These roles are governed by the rules of the organization. Every role can be represented by a set of features. They are rated and selected by the expert.

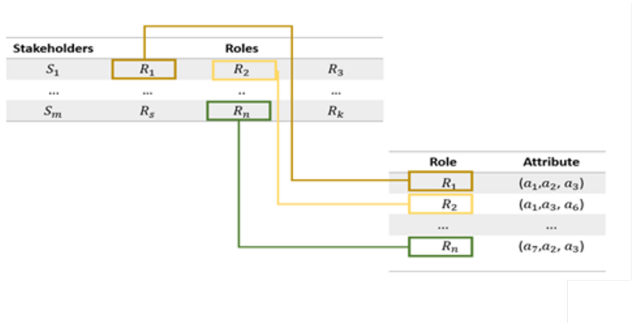

Fig. 4. Mapping roles to stakeholders

B. Clustering Algorithm

After the mapping process, soft clustering algorithms such as fuzzy c-mean (FCM) are used. The soft partitioning of data points into the intended cluster is the fundamental concept of fuzzy clustering. The primary benefit of FCM, which is flexible for multidimensional data analysis, is the complexity of its linear calculations related to the dimensions of the input data. Traditional FCM has some limitations, for instance, sensitivity to noise and outliers, and having to select the optimal number of clusters. Many articles solve these drawbacks by using extended FCM (EFCM). The authors in [19] provide a foundation for EFCM. In [20], they modify the distance metric to solve the problem of noise. In [18], they provide a new algorithm called Feature Selection EFCM (FS-EFCM) to cluster relevant features. In [21], they use EFCM to extract

the emotions of users on Twitter. In [22], this approach was extended based on entropy.

The FCM creates a matrix of fuzzy partition degrees and calculates membership for each point in the cluster. It then updates the center and membership for each iteration. The objective function is used to optimize the accuracy of clustering. The goal of this function is to minimize its value by increasing the similarity among all the data points within a cluster and reducing the similarity between different clusters. By minimizing this function by 1, the clustering algorithm can more effectively group similar data points together and separate dissimilar data points into different clusters, resulting in improved accuracy of the clustering process.

$$J = \sum_{i=1}^{C} \sum_{j=1}^{N} U_{ij}^{m} \left\| x_j - c_i \right\|^2 \tag{1}$$

Here in 1, C represents the center of the clusters, U_{ij} represents the membership degree of data points, and m represents the weighting exponent used to control the degree of fuzzy overlap between clusters. Typically, m is set to a value greater than 1. calculate the membership degree U_{ij} as follows:

$$U_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_j - c_i\|}{\|x_j - C_k\|} \right)^{\frac{2}{m-1}}} \tag{2}$$

Equation 3 is an essential step in the FCM algorithm as it is used to calculate the center of the cluster. It utilizes the membership degree U_{ij} and the data points to determine the new center of the cluster. The center of the cluster is recalculated for each iteration of the algorithm until convergence is reached. This step is crucial in ensuring that the clusters are accurately defined and that similar data points are grouped together.

$$C_i = \frac{\sum_{j=1}^{m} U_{ij}^{m} x_j}{\sum_{j=1}^{m} U_{ij}^{m}} \tag{3}$$

The steps for the FCM algorithm are:

1) Initialize matrix of membership of point U_{ij} and group of membership U.
2) Update the center of cluster C_i by using 3.
3) Update the membership degree U_{ij} by using 2.
4) If the distance between the center and point is small, then stop. Otherwise, back to step 2.

This proposed method is supposed to contribute to allocating the information resulting from mining big data according to the responsibilities of the stakeholder, making decision-making easier and faster.

IV. Stakeholders and Roles

This section outlines the stakeholders covered in this study. To determine the roles, a questionnaire on the most important stakeholders in the field of IoT security was distributed. A total of 182 participants completed the questionnaire, but after analyzing the responses, the final sample size was 124. Table I shows the distribution of participants in different categories and it is illustrated in the pie charts in Fig. 5. In chart (a), it shows the distribution of participants in different fields such as academic, student, industrial and expert. In chart (b), it shows the distribution of participants in different age groups, such as 20-30, 31-40, and 41 or over. In chart (c), it shows the distribution of participants based on their experience.

TABLE I
DETAILS OF PARTICIPANTS

Field		
Option	**Frequency**	**Percent**
Academic	33	26.6%
Student	44	35.5%
Industrial	21	16.9%
IoT expert	14	11.3%
Other	12	9.7%
Total	124	100%
Age		
20 - 30	73	58.9%
31 - 40	28	22.6%
41 or above	23	18.5%
Total	124	100%
Experience		
Less than a year	38	30.6%
1 - 5	57	46%
6 - 10	14	11.3%
More than 10	15	12.1%
Total	124	100%

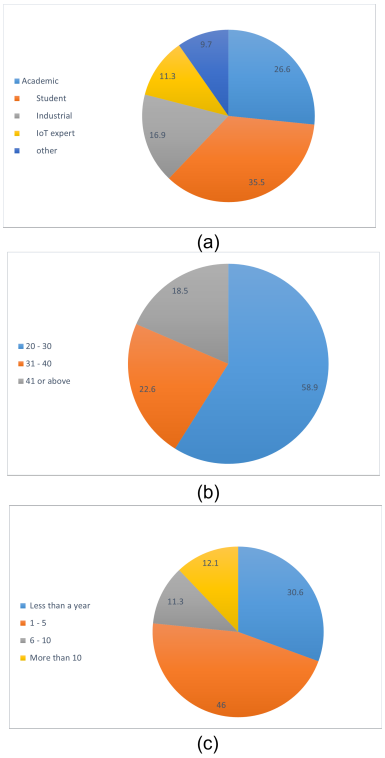

Fig. 5. Pie charts representing the distribution of participants. **(a)** represents the field of participants, **(b)** represents the age, and **(c)** represents the experience.

A. Importance of Identifying Stakeholders

A 5-point Likert scale was used. The available statements ranged from strongly agree (5) to strongly disagree (1). Table II outlines the five statements which were asked of participants. SPSS was used to perform statistical analysis on the data. The average for each phrase was calculated to determine the significance.

TABLE II
THE QUESTIONNAIRE: SENTENCES USED TO MEASURE THE IMPORTANCE OF STAKEHOLDER IDENTIFICATION

Code	Sentence
P1	Protecting IoT data is similar to the protection of other data on the Internet
P2	It is important to have companies specializing in the protection of IoT data
P3	It is important to develop a structure to identify/define job titles in the field of IoT
P4	Building a standardized structure in job titles helps companies to develop IoT data protection
P5	The presence of user data protection specialists is important for protecting users' data

TABLE III
DESCRIPTIVE STATISTICS

Code	Min.	Max.	Mean	Std. Deviation	Direction
P1	1	5	3.77	1.176	Agree
P2	1	5	4.41	0.773	Strongly Agree
P3	1	5	4.33	0.780	Strongly Agree
P4	1	5	4.22	0.825	Agree
P5	1	5	4.38	0.901	Strongly Agree

The results in Table III can be described as follows: a mean value of between 5 and 4.3 is equivalent to an evaluation of "Strongly Agree," 4.2 to 3.5 means "Agree," 3.4 to 2.7 means "Neutral," 2.6 to 1.9 means "Disagree," and 1.8 to 1 means "Strongly Disagree."

B. Determining Stakeholders

The following set of options are available to choose from: IoT security director, IoT security engineer, IoT malware analyst, IoT security consultant, IoT security specialist, IoT cyber security analyst, incident analyst, incident responder, and auditor. The data becomes clearer after identifying the most important stakeholders whose roles will be related to the features of the dataset. Table IV displays the results for the frequencies of each stakeholder.

The questionnaire results demonstrated that the top four most important stakeholders are: IoT security engineer, IoT security specialist, IoT cyber security analyst, and IoT security manager, respectively. Figure 6 shows the frequency of stakeholders. The responsibilities of IoT engineering personnel include updating security systems, identifying threats and vulnerabilities, and testing security systems to monitor and enhance performance. IoT security specialists' responsibilities include integrating security analytic systems, detecting assaults on IoT nodes in real-time, and fighting off IoT security threats. Network and IoT infrastructure security are the main responsibilities of an IoT cyber security analyst. A deep understanding of infrastructure and programs is necessary for the manager roles in IoT security. They support, train, and monitor operational services, including ensuring system availability, integrity, and confidentiality. They are also responsible for network and system risk management, including protection against infiltration of an IoT network.

TABLE IV
STAKEHOLDER FREQUENCIES

Stakeholders	Responses	Percent
IoT security manager	89	13%
IoT security engineer	100	15%
IoT malware analyst	68	10%
IoT security consultant	64	10%
IoT security specialist	95	14%
IoT cyber security analyst	89	13%
Incident analyst	59	9%
Incident responder	51	9%
Auditor	45	7%
Total	660	100%

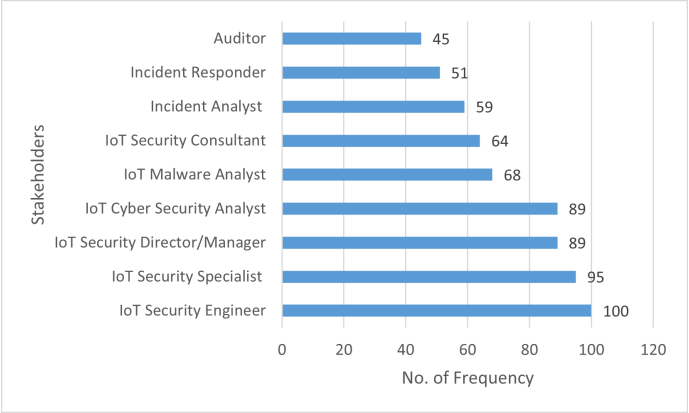

Fig. 6. Frequency of stakeholders based on the survey

C. New Stakeholders in Future

All aspects of the security sector are developing. A question was asked in the questionnaire on the future employment expectations of stakeholders in IoT security. This section provided a summary of the jobs that exist or will be targeted by businesses in the future as follows:

- A data protection specialist with a role to safeguard the confidentiality of data and ensure its integrity.
- A specialist in proposing solutions and solving domain-related problems.
- IoT governance risk and compliance consultant.
- IoT security awareness content creators. It was noted that lack of awareness is one of the reasons for safety weaknesses.
- IoT pen analysts, who aim to analyze numerous IoT device components to assist with increasing the security of the device.

V. CONCLUSION

This paper coined the term IoT Role-Based Analytics for Security Data and discussed a theoretical model for analyzing security IoT datasets based on clustering techniques. To achieve that, the roles of stakeholders were determined, then a correlation between roles and features was found. The final step in the framework was to use data mining clustering to extract information. The challenges faced can be summarized as dealing with high-dimensional data and different types of data. In future work, an algorithm through which the framework can be applied should be created, considering the challenges in the big data environment that need solving. One of the recommendations proposed in this study is to create a dataset with features that are more suitable to cover all roles to achieve the objective of the framework. Real-time data analysis is one of the major challenges in the area of IoT. More real-time analysis experiments are required on this topic.

REFERENCES

[1] Ani, U., He, H. & Tiwari, A. Human factor security: evaluating the cybersecurity capacity of the industrial workforce. *Journal Of Systems And Information Technology*. **21**, 2-35 (2019)

[2] Dimovski, V., Grah, B. & Colnar, S. Modelling the industrial workforce dynamics and exit in the ageing society. *IFAC-PapersOnLine*. **52**, 2668-2673 (2019)

[3] Sarker, A., Shamim, S., Zama, M. & Rahman, M. Employee's performance analysis and prediction using K-means clustering decision tree algorithm. *Global Journal Of Computer Science And Technology*. (2018)

[4] Smith, G. The intelligent solution: automation, the skills shortage and cyber-security. *Computer Fraud Security*. **2018**, 6-9 (2018)

[5] Crowley, C. & Pescatore, J. Common and best practices for security operations centers: Results of the 2019 SOC survey. *SANS, Bethesda, MD, USA, Tech. Rep*. (2019)

[6] Bi, Z. & Cochran, D. Big data analytics with applications. *Journal Of Management Analytics*. **1**, 249-265 (2014)

[7] Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. & Rong, X. Data mining for the internet of things: literature review and challenges. *International Journal Of Distributed Sensor Networks*. **11**, 431047 (2015)

[8] Yassine, A., Singh, S., Hossain, M. & Muhammad, G. IoT big data analytics for smart homes with fog and cloud computing. *Future Generation Computer Systems*. **91** pp. 563-573 (2019)

[9] Bhattacharjee, A., Barve, Y., Khare, S., Bao, S., Kang, Z., Gokhale, A. & Damiano, T. Stratum: A bigdata-as-a-service for lifecycle management of iot analytics applications. *2019 IEEE International Conference On Big Data (Big Data)*. pp. 1607-1612 (2019)

[10] Saheb, T. & Izadi, L. Paradigm of IoT big data analytics in the healthcare industry: A review of scientific literature and mapping of research trends. *Telematics And Informatics*. **41** pp. 70-85 (2019)

[11] Zikria, Y., Ali, R., Afzal, M. & Kim, S. Next-generation internet of things (iot): Opportunities, challenges, and solutions. *Sensors*. **21**, 1174 (2021)

[12] Zhang, K. Research on data mining security under the background of big data era. *8th International Conference On Management And Computer Science (ICMCS 2018)*. pp. 236-239 (2018)

[13] Savaglio, C. & Fortino, G. A simulation-driven methodology for IoT data mining based on edge computing. *ACM Transactions On Internet Technology (TOIT)*. **21**, 1-22 (2021)

[14] Lee, E., Seo, Y., Oh, S. & Kim, Y. A Survey on Standards for Interoperability and Security in the Internet of Things. *IEEE Communications Surveys Tutorials*. **23**, 1020-1047 (2021)

[15] Kar, S., Chakravorty, B., Sinha, S. & Gupta, M. Analysis of stakeholders within IoT ecosystem. *Digital India*. pp. 251-276 (2018)

[16] Weissman, D. & Jayasumana, A. Integrating IoT monitoring for security operation center. *2020 Global Internet Of Things Summit (GIoTS)*. pp. 1-6 (2020)

[17] Roldán, J., Boubeta-Puig, J., Martınez, J. & Ortiz, G. Integrating complex event processing and machine learning: An intelligent architecture for detecting IoT security attacks. *Expert Systems With Applications*. **149** pp. 113251 (2020)

[18] Di Martino, F. & Senatore, S. Balancing the user-driven feature selection and their incidence in the clustering structure formation. *Applied Soft Computing*. **98** pp. 106854 (2021)

[19] Li, C., Becerra, V. & Deng, J. Extension of fuzzy c-means algorithm. *IEEE Conference On Cybernetics And Intelligent Systems, 2004.*. **1** pp. 405-409 (2004)

[20] Kumar, N., Kumar, H. & Sharma, K. Extension of FCM by introducing new distance metric. *SN Applied Sciences*. **2**, 1-21 (2020)

[21] Di Martino, F., Senatore, S. & Sessa, S. A lightweight clustering–based approach to discover different emotional shades from social message streams. *International Journal Of Intelligent Systems*. **34**, 1505-1523 (2019)

[22] Ramathilagam, S., Devi, R. & Kannan, S. Extended fuzzy c-means: an analyzing data clustering problems. *Cluster Computing*. **16**, 389-406 (2013)